

**Final Technical Report
AASERT F49620-95-1-0366**

*Innovative seismic array analysis for studies of
wave propagation in the earth*

Principal Investigator:
Prof. Gary L. Pavlis
Department of Geological Sciences
Indiana University
1001 East 10th Street
Bloomington, IN 47405

Grant period: May 1, 1995 - April 30, 1999

20000420 152

REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-00-

Public reporting burden for this collection of information is estimated to average 1 hour per response, including gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Service, Paperwork Project, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Project, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302.

1. sources,
act of this
Jefferson
3.

0144

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE 9 Mar 00	3. REPORT TYPE AND DATES COVERED Final Technical Report 1 May 95 to 30 Apr 99
4. TITLE AND SUBTITLE Innovative seismic array analysis for studies of wave propagation in the earth			5. FUNDING NUMBERS F49620-95-1-0366
6. AUTHOR(S) Professor Gary L. Pavlis			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Geological Sciences Indiana University 1001 East 10th Street Bloomington, IN 47405			8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR/NM 801 N. Randolph St, Rm 732 Arlington, VA 22203-1977			10. SPONSORING/MONITORING AGENCY REPORT NUMBER F49620-95-0366
11. SUPPLEMENTARY NOTES			
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 words) We have developed a new seismic array data processing method to produce slowness vector estimates and an objective measure of their uncertainties in the form of statistical confidence intervals. The slowness vector, which is typically transformed into bearing and velocity, is a key parameter used for identifying seismic phases and for event source location. Our method, multi-wavelet beamforming, is closely related to both time-domain and frequency-domain beamforming. The major advantage of multi-wavelet beamforming is that it produces multiple estimates of the slowness vector that are approximately statistically independent. First, a set of wavelet transforms is applied to the data in a manner analogous to the use of the windowed Fourier transform. Next, for each wavelet transform, we calculate semblance, a measure of signal coherence, for a range of possible slowness vectors. Then, the slowness vector estimate associated with that transform is the vector that produces the largest semblance value. The multiple slowness vector estimates can be treated as samples from a probability distribution, whose "center" we estimate using the mean, the median, and an M-estimator. Uncertainty intervals are calculated for these estimators by applying the jackknife statistical method. The intervals for the mean estimator appear to be true statistical confidence intervals, but the estimates can be biased by a directional noise field in low signal-to-noise circumstances. The median estimates are less biased by a directional noise field, but sometimes underestimate the uncertainty. The M-estimator produces less-biased estimates while appearing to estimate correctly their uncertainty.			
14. SUBJECT TERMS Workshop, conference, Defense related signal processing			15. NUMBER OF PAGES
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

Accomplishments

Lorie Bear, Ph.D.

This AASERT award was a supplement to an AFOSR grant in nuclear verification research used primarily to support PhD dissertation work by Lorie Bear at Indiana University. Ms. Bear completed her Ph.D. dissertation defense in December of 1997. The following is the title of her dissertation:

Bear, Lorie (1997). The use of multiwavelets for seismic array data processing, Ph.D. dissertation, Indiana University, Bloomington, Indiana.

Her dissertation is actually three papers: All have now been published. Reprints are attached. A summary of the main technical advances provided by her work can be gleaned from reading the abstracts of these papers. The relevant citations are:

Refereed Papers:

Bear, L. K., G. L. Pavlis, and G. H. R. Bokelmann (1999). Multiwavelet analysis of three-component seismic arrays: Application to measure effective anisotropy at Pinon Flats, California, *Bull. Seism. Soc. Amer.*, **89**, 693-705.

Bear, L. K. and G. L. Pavlis (1999). Multichannel estimation of time residuals from seismic data using multiwavelets, *Bull. Seism. Soc. Amer.*, **89**, 681-692.

Bear, L.K. and G.L. Pavlis (1997) Estimation of slowness vectors and their uncertainties using multiwavelet seismic array processing, *Bull. Seism. Soc. Amer.*, **87**, 755-769.

Published Abstracts:

Bear, L. K. and G. L. Pavlis (1997). High-resolution multiwavelet seismic array processing, *EOS Trans. Amer. Geophys. Union*, **78**, S217.

Bear, L. K. and G. L. Pavlis (1996). Multiwavelet estimation of bearing and slowness uncertainty from broadband seismic arrays, *Seism. Res. Letters*, **67**, 32.

Bear, L. and G. L. Pavlis (1995). Variable bandwidth analysis of broadband seismic array data, *EOS Trans. Amer. Geophys. Union*, **76**, S203.

Reports:

Bear, L.K., and G.L. Pavlis (1996). Multiwavelet beamforming--A new seismic array data processing method, *Proceedings of the 18th annual seismic research symposium on monitoring a comprehensive test ban treaty*, Report PL-TR-96-2153, 657-665.

Brent Foshee

The residual funds in this grant after Dr. Bear completed her Ph.D. degree were used to support Mr. Foshee who is an M.S. student in this department. Mr. Foshee was supported through this grant in two nonadjacent semesters in the fall of 1997 and in 1998.

Unfortunately, Mr. Foshee has still not completed his M.S. thesis and fear he may never do so as he has taken a job away from Bloomington. His M.S. thesis work is directed at understanding an observed systematic deviation of P and S wave phase velocities measured on a high-frequency array in Turkmenistan (An alpha array is being planned for deployment at the same site for the CTBT.) Mr. Foshee is attempting to model the observed anomaly in terms of a dipping fault contact between media with different Vp/Vs ratios.

Estimation of Slowness Vectors and Their Uncertainties Using Multi-Wavelet Seismic Array Processing

by Lorie K. Bear and Gary L. Pavlis

Abstract We have developed a new seismic array data processing method to produce slowness vector estimates and an objective measure of their uncertainties in the form of statistical confidence intervals. The slowness vector, which is typically transformed into bearing and velocity, is a key parameter used for identifying seismic phases and for event source location. Our method, multi-wavelet beamforming, is closely related to both time-domain and frequency-domain beamforming. The major advantage of multi-wavelet beamforming is that it produces multiple estimates of the slowness vector that are approximately statistically independent. First, a set of wavelet transforms is applied to the data in a manner analogous to the use of the windowed Fourier transform. Next, for each wavelet transform, we calculate semblance, a measure of signal coherence, for a range of possible slowness vectors. Then, the slowness vector estimate associated with that transform is the vector that produces the largest semblance value. The multiple slowness vector estimates can be treated as samples from a probability distribution, whose "center" we estimate using the mean, the median, and an M-estimator. Uncertainty intervals are calculated for these estimators by applying the jackknife statistical method. The intervals for the mean estimator appear to be true statistical confidence intervals, but the estimates can be biased by a directional noise field in low signal-to-noise circumstances. The median estimates are less biased by a directional noise field but sometimes underestimate the uncertainty. The M-estimator produces less-biased estimates while appearing to estimate correctly their uncertainty.

Introduction

After the Geneva "Conference of Experts" of 1958, seismic arrays quickly gained importance as a fundamental tool for detection and location of small seismic events (see Husebye and Ruud, 1989). A variety of methods have been developed to determine the direction of approach and phase velocity of a seismic waveform as it crosses an array. This indispensable information is generally determined in the form of a slowness vector (slowness = $1/\text{velocity}$) consisting of east-west and north-south components. Until now, an objective measure of the estimate's uncertainty, which is fundamental to determining whether it has the necessary precision for a given application, has been lacking. Existing methods determine a *single* estimate for the slowness vector and do little to address the question of how good that estimate is. To our knowledge, the only previous attempt to quantify uncertainties in array slowness vector measurements is a somewhat *ad hoc* method devised by Bratt and Bache (1988) that uses arbitrary measures of the quality of the signal to determine solution uncertainties. Our seismic

array processing method, which we call multi-wavelet beamforming, is important because it provides an objective method for quantifying errors in the slowness vector estimates. Because our method works directly with the observed data and uses no arbitrary free parameters, it is objective and holds great promise for automated calculation and appraisal of array solutions.

The primary difference between multi-wavelet beamforming and other array processing methods is that the new method produces multiple slowness component estimates that are approximately statistically independent. These estimates can be treated as samples from a probability distribution and used to calculate confidence intervals using standard statistical methods. This new array processing method combines aspects of both time-domain and frequency-domain beamforming and incorporates concepts drawn from such diverse topics as wavelet transforms (Daubechies, 1992), multiple spectral estimation (Thomson, 1982; Lilly and Park, 1995), and statistics (Miller, R., 1974).

Integral Transforms

The characteristics of a seismic signal can be described fundamentally by time location and frequency content. These two properties can be studied simultaneously by using a moving-window Fourier transform. The windowed Fourier transform of signal $s(t)$ for frequency f and time t can be written as

$$\begin{aligned} \mathcal{F}[s](f, t) &= \int_{t-T/2}^{t+T/2} s(\xi) g(\xi - t) \cos[2\pi f(\xi - t)] d\xi \\ &\quad + i \int_{t-T/2}^{t+T/2} s(\xi) g(\xi - t) \sin[2\pi f(\xi - t)] d\xi \quad (1) \\ &= \int_{t-T/2}^{t+T/2} s(\xi) g_{f,e}(\xi - t) d\xi + i \int_{t-T/2}^{t+T/2} s(\xi) g_{f,o}(\xi - t) d\xi, \end{aligned}$$

where $g(t)$ is a chosen taper of length T (Kumar and Foufoula-Georgiou, 1994). The integration kernel for this transform is then $\mathbf{g}_f(t) = g(t) \cos(2\pi ft) + ig(t) \sin(2\pi ft)$.

Within the last 10 years, there has been widespread interest in developing a special class of kernel functions called wavelets. Wavelets have been independently introduced in various fields of study since the 1960s (Daubechies, 1992). They were first used in geophysics by Morlet *et al.* (1982). Wavelets are functions developed to have time-frequency localization. That is, the time lengths of the functions match the scale of the frequencies to be studied (Kumar and Foufoula-Georgiou, 1994). When used as the integration kernels for a transform, they provide an automatic scaling of the time window to be studied. In a real sense, they provide a natural bridge between the time and frequency domains.

It is well known that a finite time function cannot also be finite in frequency due to the uncertainty principle (Slepian, 1983). We can at best hope to concentrate its energy into a frequency band of interest. The work of Slepian (1983) and Thomson (1982) introduced a novel approach to finding finite time functions with their energy concentrated in a frequency band of the form $-W < f < W$. These functions have been applied in a range of techniques commonly referred to as multi-taper or multi-window methods (e.g., Park *et al.*, 1987). A generalization of this approach was recently developed by Lilly and Park (1995). The Lilly and Park wavelets are real, discrete time series (w_m) with M samples and sampling rate Δt . They are designed to concentrate energy within a frequency range defined by a center frequency f_c and a bandwidth $2f_w$ (where $f_w \leq f_c$).

Because the wavelets are real time series, energy in the frequency domain must appear in both the positive and negative frequencies. Thus, any frequency band of interest is defined by $|f \pm f_c| \leq f_w$. The fraction of the total energy contained in this frequency band is

$$\lambda = \frac{\int_{-(f_c+f_w)}^{(f_c+f_w)} |W(f)|^2 df - \int_{-(f_c-f_w)}^{(f_c-f_w)} |W(f)|^2 df}{\int_{-1/2\Delta t}^{1/2\Delta t} |W(f)|^2 df},$$

where

$$W(f) = \Delta t \sum_{m=-P+1}^R w_m e^{i2\pi f m \Delta t} \quad (2)$$

and where P is the closest integer $\geq M/2$ and R is the closest integer $\leq M/2$.

Lilly and Park (1995) calculate a set of wavelets by rewriting equation (2) as an eigenvalue equation of the form $\mathbf{A}\mathbf{w} = \lambda\mathbf{w}$, where

$$\begin{aligned} A_{mn} &= \frac{\sin[2\pi(f_c + f_w)\Delta t(m - n)]}{\pi(m - n)} \\ &\quad - \frac{\sin[2\pi(f_c - f_w)\Delta t(m - n)]}{\pi(m - n)} \quad (3) \end{aligned}$$

and solving for $\mathbf{w}$. This equation has M orthogonal solutions of eigenvectors $\mathbf{w}^{(k)}$ and associated eigenvalues λ_k that we order so that $\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_M$. The wavelets $\mathbf{w}^{(k)}$ are normalized such that $\sum_{m=1}^M (w_m^{(k)})^2 = 1$. We only use the wavelets $\mathbf{w}^{(k)}$ that have λ_k close to 1, since these functions have almost all their spectral energy within the frequency band of interest. The number of such wavelets is dependent on the width of the frequency band and the time length of the wavelets (Lilly and Park, 1995). This is a direct consequence of the uncertainty principle.

The wavelets occur as even and odd pairs where the pairs emphasize different portions of the frequency band (Fig. 1). Since the pairs of wavelets are functions that are 90° out of phase, they can be combined into complex functions of the form $\mathbf{w}_f^{(j)} = \{w_{f,e}^{(j)} + iw_{f,o}^{(j)}\}$, where f is the center frequency, $w_{f,e}^{(j)}$ is the j th even wavelet, and $w_{f,o}^{(j)}$ is the j th odd wavelet.

The multi-wavelet transform with kernel $\mathbf{w}_f^{(j)}$ and time length $T = M\Delta t$ is

$$\begin{aligned} \mathcal{W}^{(j)}[s](f, t) &= \int_{t-T/2}^{t+T/2} s(\xi) w_{f,e}^{(j)}(\xi - t) d\xi \\ &\quad + i \int_{t-T/2}^{t+T/2} s(\xi) w_{f,o}^{(j)}(\xi - t) d\xi. \quad (4) \end{aligned}$$

Note that equations (1) and (4) have the same form. To understand the nature of the wavelet transform, we compare the real parts of the windowed Fourier and multi-wavelet

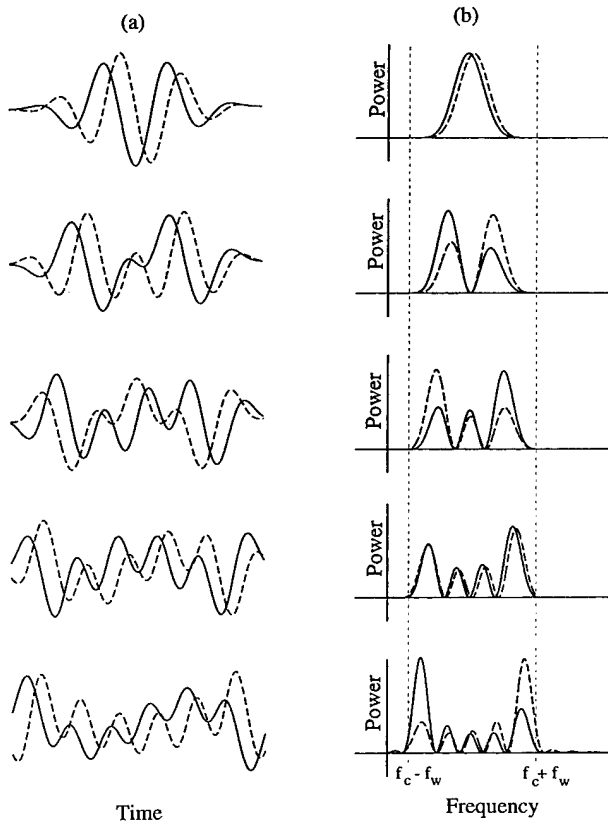


Figure 1. Ten real Lilly and Park wavelets as calculated from equation (2): (a) in the time domain and (b) their frequency power spectra. Wavelets from top to bottom go from higher- to lower-energy fraction (λ) values.

integration kernels shown in Figure 2. The taper used for both Fourier kernels is a 2-sec-long, zeroth-order 4π prolate taper (Park *et al.*, 1987). The wavelets in Figures 2c and 2d are scaled such that $f_{c(c)} = 1/k f_{c(d)}$ and $f_{w(c)} = 1/k f_{w(d)}$, where $k = 4$. A fundamental difference between wavelets and more conventional Fourier methods is that wavelets vary the time and frequency scales in a manner that matches the time length of a wavelet to the frequencies it contains. This is important for optimal resolution of broadband signals that are characteristic of modern seismological data. It is shown in Appendix A that if the number of samples stays the same and $\Delta t_{(c)} = k \Delta t_{(d)}$, then the wavelets in Figures 2c and 2d are exact scaled versions of one another. Thus, we can simultaneously scale the time length and the frequency bandwidth by simply changing the sampling rate instead of having to recalculate the wavelets each time.

It is obvious that signal analysis using either type of transform can only resolve features in the time domain that are on the order of length T . For the windowed Fourier transform, it is critical to note that the taper length T is an arbitrary free parameter. The choice of the taper also directly affects the resolution limit in the frequency domain since the frequency spectra of $\mathbf{g}_f$ is the frequency spectra of the taper

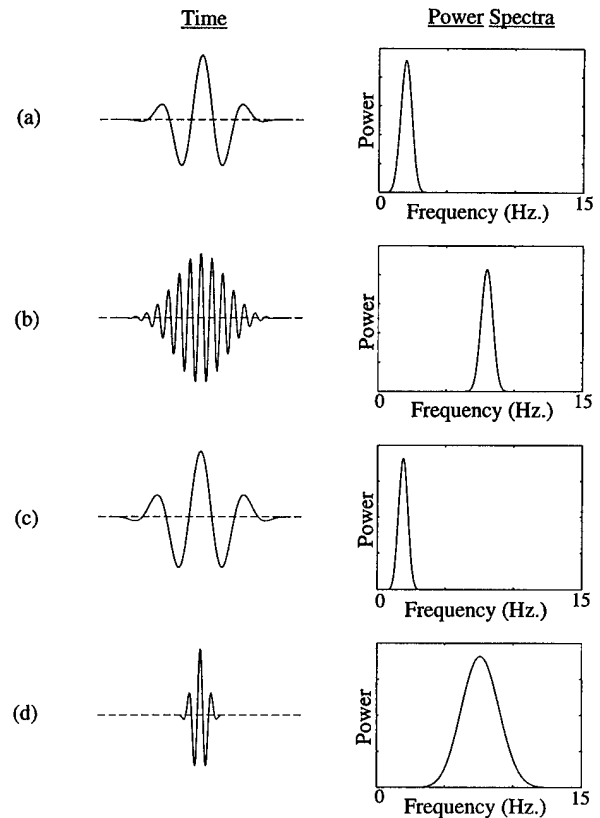


Figure 2. Four integration kernels for moving-window time-frequency analysis: (a) a windowed Fourier kernel with $f = 2$ Hz and $T = 2$ sec, (b) a windowed Fourier kernel with $f = 8$ Hz and $T = 2$ sec, (c) a Lilly and Park wavelet with $f = 2$ Hz and $T = 2$ sec, and (d) a Lilly and Park wavelet with $f = 8$ Hz and $T = 0.5$ sec. For (a) and (b), we used a 2-sec long, zeroth-order 4π prolate taper (Park *et al.*, 1987). Note that (a) and (c) are nearly identical, while (b) and (d) differ drastically. In (d), frequency resolution has been sacrificed for improved time resolution.

shifted to be centered about f . Once the taper is chosen, the frequency resolution is fixed regardless of which frequencies are being studied (Kumar and Fouloula-Georgiou, 1994).

In contrast, the time length T of the wavelet integration kernel is automatically scaled to be compatible with the scale of the center frequency in the time domain. This scaling, however, comes with a price. If the resolution in the time domain is increased by shortening the time interval T , then the resolution in the frequency domain must decrease. We would argue that for seismological applications the increased resolution in the time domain is a major advantage in the study of impulsive arrivals and more than offsets any information loss due to the lower-frequency resolution.

Beamforming

Classical beamforming (e.g., Kvaerna and Doornbos, 1986) uses windowed Fourier transforms to slant-stack array

data in the frequency domain. For a given slowness vector $\mathbf{u} = (u_{ew}, u_{ns})$ and array station n , there is a plane-wave time delay $\tau(\mathbf{u}, n) = -\mathbf{u} \cdot \mathbf{x}_n$, where $\mathbf{x}_n$ is the vector position of the station with respect to a local Cartesian system. For N stations, the $N \times N$ spectra covariance matrix $\mathbf{C}(f, t, \mathbf{u})$ has components

$$C_{nm}(f, t, \mathbf{u}) = \mathcal{F}[s_n][f, t + \tau(\mathbf{u}, n)] \cdot \mathcal{F}[s_m]^*[f, t + \tau(\mathbf{u}, m)], \quad (5)$$

where $\mathcal{F}[s_n]$ is the windowed Fourier transform for station n and the asterisk denotes the complex conjugate transpose. The average station signal power can be written as

$$P_{\text{avg}}(f, t, \mathbf{u}) = |\mathbf{d}|^2 \cdot \text{tr}[\mathbf{C}(f, t, \mathbf{u})], \quad (6)$$

where tr is the trace of the matrix and $\mathbf{d}$ is an $N \times 1$ vector of station weights such that $\sum_{n=1}^N d_n = 1$. We will be using uniform weighting such that $d_n = 1/N$. The power of the beam is defined to be

$$P_{\text{beam}}(f, t, \mathbf{u}) = \mathbf{d}^T \mathbf{C}(f, t, \mathbf{u}) \mathbf{d}, \quad (7)$$

where the superscript T denotes transpose.

A common measure of the overall coherence between the station signals is semblance (Husebye and Ruud, 1989):

$$S(f, t, \mathbf{u}) = \frac{P_{\text{beam}}(f, t, \mathbf{u})}{P_{\text{avg}}(f, t, \mathbf{u})}. \quad (8)$$

Semblance values can range between zero and one. If the signals stack exactly, then the semblance is one. For white noise, S is approximately $1/N$. The semblance values are calculated for a range of slowness vectors and contoured in what we call the slowness grid.

The slowness grid pattern for a given array geometry, frequency band, and any coherent plane-wave signal is completely predictable. A standard display for this "beam pattern" is the slowness grid for an impulse signal with zero slowness. Beam patterns for two frequency bands are shown in Figure 3. In all cases, the highest semblance value appears at the center of the beam pattern. This is what we expect since the station signals should be most coherent when stacked with zero time lags.

The slowness grid for a plane-wave signal with slowness vector $\mathbf{u}_0$ would have the same pattern translated to be centered on $\mathbf{u}_0$. Consequently, the standard method to estimate $\mathbf{u}$ for an event of unknown origin is to search for the point in the slowness grid with the largest semblance value. A problem with this approach is that the resolution is limited by the grid spacing Δu . Thus we follow a methodology used by Harvey (1994) where each slowness component is determined by a center of mass calculation. That is, for the east-west slowness estimate, we calculate

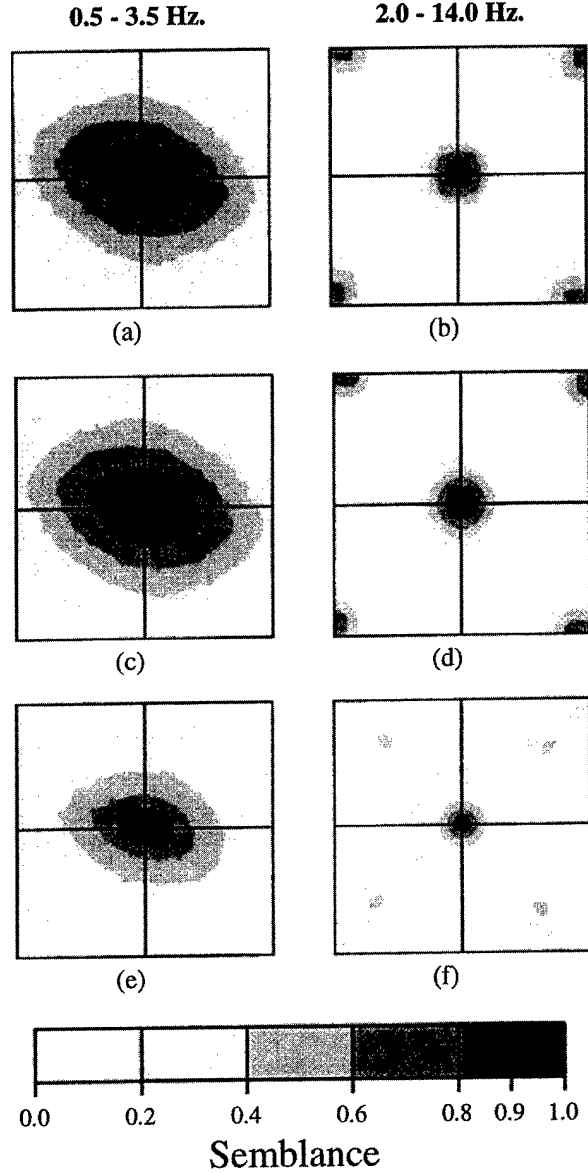


Figure 3. Beam patterns for the Geyokcha array in two different frequency bands. The beam patterns are oriented so that north is up and east is to the right and the slowness values range from -0.3 to 0.3 sec/km in both the north-south and east-west directions. (a) through (d) were calculated using the corresponding kernel functions of Figure 2; (e) and (f) were calculated by averaging the beam patterns from the five complex wavelets for each frequency band. The beam patterns for the 2- to 14-Hz frequency band exclude the data from stations SEH, SWH, and NH.

$$u_{ew} = \frac{\sum_{(u_k, u_j) \in A} u_k m(u_k, u_j)}{\sum_{(u_k, u_j) \in A} m(u_k, u_j)}, \quad (9)$$

where

$$m(u_k, u_j) = \begin{cases} 0 & S(f, t, u_k, u_j) < 0.9 S_{\text{peak}}(f, t), \\ S(f, t, u_k, u_j) & S(f, t, u_k, u_j) \geq 0.9 S_{\text{peak}}(f, t) \end{cases}$$

A is the set of all slowness vectors in the slowness grid, and S_{peak} is the largest semblance value calculated using the slowness vectors in A . The equation for the north-south slowness estimate is similar.

What we call multi-wavelet beamforming can be couched in the same mathematical constructs as classical beamforming by substituting

$$C_{nm}^{(j)}(f, t, \mathbf{u}) = \mathcal{W}^{(j)}[s_n][f, t + \tau(\mathbf{u}, n)] \cdot \mathcal{W}^{(j)}[s_m]^*[f, t + \tau(\mathbf{u}, m)] \quad (10)$$

for the spectra covariance matrix. In practice, though, the formation of the covariance matrix is inefficient when taking time delays into account.

It is straightforward to incorporate the time delays when forming the slowness grids in the time domain. In standard time-domain beamforming (Pavlis and Mahdi, 1996), frequency-domain localization is normally accomplished by applying a bandpass filter, with center frequency f , to the data. If $\phi_f(t)$ is the time-reversed filter of length T , then the filtered signal can be written

$$C[s](f, t) = \int_{t-T/2}^{t+T/2} s(\xi) \phi_f(\xi - t) d\xi. \quad (11)$$

The power at each station is summed over some arbitrary time L such that

$$P_n(f, t, \mathbf{u}) = \int_{t-L/2}^{t+L/2} |C[s_n][f, \xi + \tau(\mathbf{u}, n)]|^2 d\xi, \quad (12)$$

and the average summed station power is

$$P_{\text{avg}}(f, t, \mathbf{u}) = \frac{1}{N} \sum_{n=1}^N P_n(f, t, \mathbf{u}). \quad (13)$$

The beam for the array is defined as

$$b(f, t, \mathbf{u}) = \frac{1}{N} \sum_{n=1}^N C[s_n][f, t + \tau(\mathbf{u}, n)], \quad (14)$$

so the beam power summed over time length L is

$$P_{\text{beam}}(f, t, \mathbf{u}) = \int_{t-L/2}^{t+L/2} |b(f, \xi, \mathbf{u})|^2 d\xi. \quad (15)$$

The semblance is then defined as in equation (8).

The summing over time L is done to smooth the power estimates as the signal cycles back and forth from local maxima to local minima. This averaging is necessary to produce a stable solution but introduces an arbitrary free parameter

in the form of the length L . We originally used this same approach in multi-wavelet beamforming, setting L equal to the length of the wavelet. We found, however, that the statistical independence of the multi-wavelet slowness component estimates is seriously violated by this averaging. As a result, we use a different approach that exploits the slow variability of the envelope of a function (Kanasewich, 1981). The station power used in multi-wavelet beamforming is calculated as

$$P_n^{(j)}(f, t, \mathbf{u}) = |\mathcal{W}^{(j)}[s_n][f, t + \tau(\mathbf{u}, n)]|^2. \quad (16)$$

Since the wavelets are even and odd functions, the wavelet transform can be considered a convolution, or filtering, operation where $w_{f,e}$ and $w_{f,o}$ are two filters with 90° phase differences. This implies that the real and imaginary parts of the wavelet transform also have a 90° phase difference. The envelope of any function g can be calculated as

$$E(t) = |g(t) + ig_H(t)|^2, \quad (17)$$

where g_H is the Hilbert transform of g . The Hilbert transform introduces a 90° phase shift to g , just as the imaginary part of the wavelet transform introduces a 90° phase shift to the real part of the wavelet transform. Thus, the wavelet power estimate behaves like an envelope function and varies relatively slowly. This implies that the power estimate will be relatively stable.

Analogous to the time-domain method, then, we calculate

$$P_{\text{avg}}^{(j)}(f, t, \mathbf{u}) = \frac{1}{N} \sum_{n=1}^N P_n^{(j)}(f, t, \mathbf{u}), \quad (18)$$

and the beam power is defined as

$$P_{\text{beam}}^{(j)}(f, t, \mathbf{u}) = |b^{(j)}(f, t, \mathbf{u})|^2 = \left| \frac{1}{N} \sum_{n=1}^N \mathcal{W}^{(j)}[s_n][f, t + \tau(\mathbf{u}, n)] \right|^2. \quad (19)$$

The semblance is then defined as

$$S^{(j)}(f, t, \mathbf{u}) = \frac{P_{\text{beam}}^{(j)}(f, t, \mathbf{u})}{P_{\text{avg}}^{(j)}(f, t, \mathbf{u})}. \quad (20)$$

Jackknife Method

The primary advantage of multi-wavelet beamforming over other methods is that it enables us to produce multiple slowness component estimates using equation (9)—one estimate for each complex wavelet. These estimates can be thought of as samples drawn from some unknown distribution. We are particularly interested in estimating a parameter

θ that defines a "center" of the distribution and the variance in that estimate.

For an estimator $\hat{\theta}$, the jackknife statistical method provides a computationally simple way to determine its variance with no *a priori* information about the distribution. If we have J independent samples $X_1, X_2, \dots, X_J$ from the distribution, then the delete-one estimate, $\hat{\theta}_{(-i)}$, is a function of $J - 1$ samples such that

$$\hat{\theta}_{(-i)} = \hat{\theta}(X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_J). \quad (21)$$

These delete-one estimates can be used to calculate the variance of $\hat{\theta}_{\text{all}} = \hat{\theta}(X_1, \dots, X_J)$ by using the formula

$$s^2 = \frac{J-1}{J} \sum_{i=1}^J (\hat{\theta}_{(-i)} - \hat{\theta}_{(c)})^2$$

where

$$\hat{\theta}_{(c)} = \frac{1}{J} \sum_{i=1}^J \hat{\theta}_{(-i)} \quad (22)$$

(see Thomson and Chave, 1991). This estimate is known to be conservative. Even if the data are not identically distributed, it has been shown that the expected value of s^2 is always larger than the true variance (Efron and Stein, 1981).

We use the jackknife variance to create confidence intervals for our estimates of the slowness components. If $\hat{\theta}_{\text{all}}$ has minimal bias, then it has been shown that $(\hat{\theta}_{\text{all}} - \theta)/s$ is asymptotically standard normally distributed (Chave and Thomson, 1989). Work by Hinkley (1977) has suggested that if the distribution of the data is close to normal, then the usual method of producing a 95% confidence interval can be used:

$$\theta = \hat{\theta}_{\text{all}} \pm s \cdot t_{0.975}, \quad (23)$$

where $t_{0.975}$ is the 0.975 quantile of the Student t distribution with $J - 1$ degrees of freedom.

In any real application, there is some limit on how well the slowness vector can be estimated. Possible limiting factors include the data sampling rate as compared to the aperture of the array and the slowness grid cell size. Such factors as these combine to place a floor on the width of the confidence interval. How this floor should be determined is probably dependent on details of a given array and should be considered on a case-by-case basis.

Statistical Considerations

It is important to determine whether the slowness estimates calculated from multi-wavelet beamforming can be considered statistically independent samples from a distribution, as required by the jackknife method. The true slowness component u_{true} for a pure plane-wave arrival is a con-

stant, so there must be some "noise" component in the recorded signal that provides the variability in the slowness estimates such that

$$u_{\text{estimate}} = u_{\text{true}} + u_{\text{noise}}. \quad (24)$$

If the expected value for u_{noise} is zero (i.e., the noise is non-directional), then the true slowness u_{true} can be considered the "center" for the distribution of u_{estimate} . If the noise field is directional, the "center" of the distribution will be biased toward that direction to some extent.

Since u_{true} does not affect the variability in the estimate, we will now only consider the noise field. Seismic "noise" generally falls into two categories: ambient noise (including both natural and cultural sources) and event-generated noise (or coda). Both types can be highly colored in frequency content and directional, but we will make the standard assumption that the noise is Gaussian with mean zero and variance one and that the noise samples are all mutually independent. With this assumption, it is shown in Appendix B that the slowness component estimates are approximately independent. We note that it is the necessity of satisfying the independence assumption that leads to the sample-by-sample power estimates given by equation (16). Time averaging (as in equation 12), which is standard practice in time-domain beamforming, would completely invalidate the independence assumption and thus invalidate the jackknife confidence intervals.

Application

To demonstrate our approach, we analyze data recorded at the Geyokcha seismic array (Al-Shukri *et al.*, 1995). This array operated in Turkmenistan on the northern border of Iran from August 1993 to November 1994. We used a subset of the array that consists of 12 stations with three-component Streckeisen STS-2 sensors arranged as shown in Figure 4. All the data used here come from the triggered data stream that was recorded at 250 samples/sec. We used only the vertical components and removed the DC bias from each trace.

Three different frequency bands were used to study these data: 0.5 to 3.5, 1 to 7, and 2 to 14 Hz. A set of Lilly and Park wavelets was calculated using equation (2) for the band of 2 to 14 Hz ($f_c = 8$ Hz, $f_w = 6$ Hz) and a sampling rate of 250 samples/sec. If each wavelet is 125 samples long, then there are 10 real wavelets (shown in Fig. 1) with energy fraction greater than 0.9. These wavelets are combined into complex wavelets and used to produce five estimates for the slowness vector. These five estimates take advantage of the fact that the quantile values of the Student t distribution change relatively slowly for 4 or more degrees of freedom. These same wavelets, with sampling rates of 125 samples/sec ($f_c = 4$ Hz, $f_w = 3$ Hz) and 62.5 samples/sec ($f_c = 2$ Hz, $f_w = 1.5$ Hz), can also be used for the other two frequency bands (Appendix A).

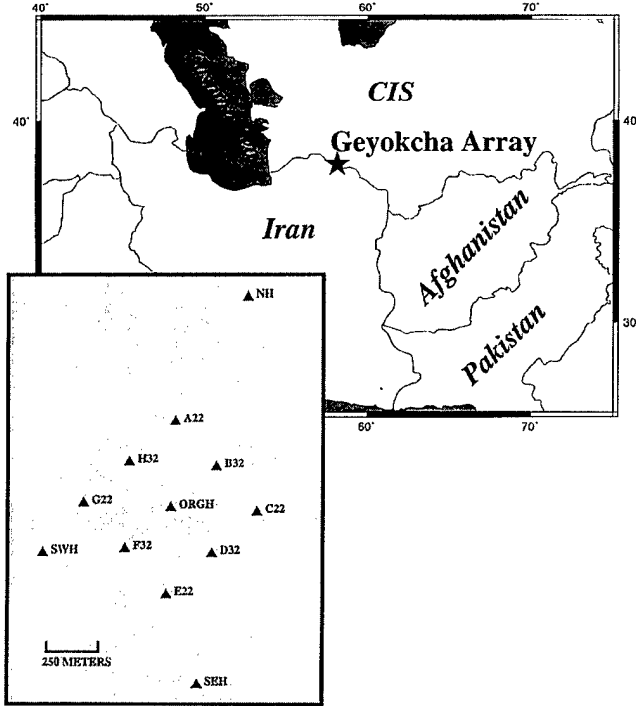


Figure 4. The broadband stations of the Geyokcha array. The star on the map shows the location of the array in central Asia. The inset map shows the array geometry.

A Simulation

We first apply our processing method to a controlled simulation with 100 similar synthetic events. We started by using a trace with a very high signal-to-noise ratio recorded at station ORGH of the array. This trace had a maximum first-arrival amplitude of $a = 1820$ nm/sec. We then produced a simulated plane-wave signal as if this waveform traveled across the array with a given azimuth of 166° and slowness of 0.158 sec/km (Fig. 5a). [These were the values obtained for the true event by the time-domain method as implemented by Harvey (1994)].

We separated a long segment of noise recorded at the array into 100 data sets such that each trace had the same number of samples as the plane-wave traces. The synthetic traces for each station n were then defined as

$$s_{n,t} = \frac{3\sigma}{a} p_{n,t} + \eta_{n,t}, \quad (25)$$

where $p_{n,t}$ was the plane-wave trace for station n , $\eta_{n,t}$ was a noise trace recorded at station n , and σ is a measure of the noise amplitude defined as

$$\sigma = \left(\frac{\sum_{t=1}^T (\eta_{n,t} - \bar{\eta}_n)^2}{T - 1} \right)^{1/2}, \quad (26)$$

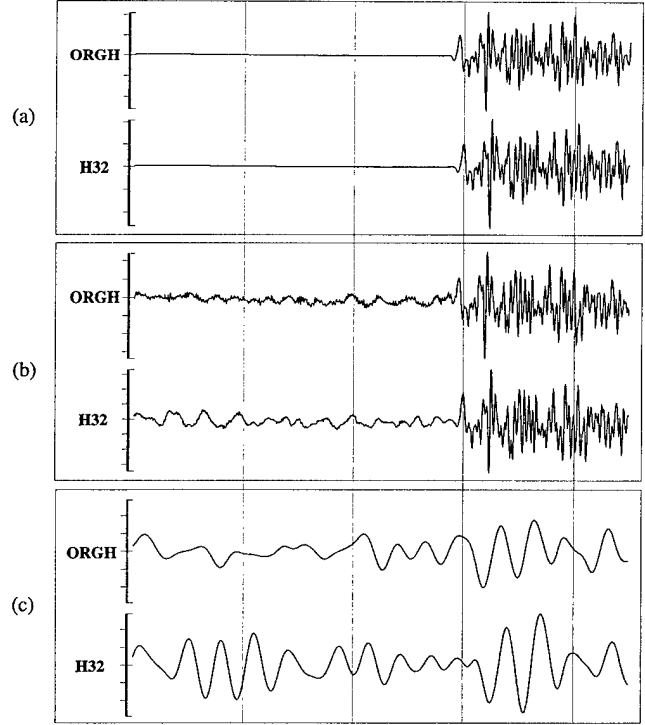


Figure 5. Vertical records from two stations for one of the synthetic events. Shown here are (a) the plane-wave traces, (b) the plane-wave traces with noise added (the synthetic traces), and (c) the synthetic traces filtered to the frequency band 0.5 to 3.5 Hz. The light dotted lines denote time in seconds.

where T is the number of samples. This is, on average, comparable to a 3-to-1 signal-to-noise ratio for the broadband signal (Fig. 5b).

There is a question as to which time window to use when calculating the slowness grids. The way in which the wavelets interact with a particular phase and the underlying noise field is unpredictable. It is not optimal to always place the time window about a phase in the same position relative to a time pick. A coherence maximization scheme appears to hold the most promise. The method we use in this article (which is by no means the only possibility) is as follows. First, we produce slowness grids over a suite of times about the phase of interest. Next, we pick the peak semblance values $S_{\text{peak}}^{(j)}(f, t)$ for each time. Then, we average the values for the J wavelets,

$$S_{\text{avg}}(f, t) = \frac{1}{J} \sum_{j=1}^J S_{\text{peak}}^{(j)}(f, t), \quad (27)$$

and determine the time point, t_{max} , where S_{avg} attains its maximum. Since semblance is a measure of coherence, it is intuitively appealing to assume that the maximum semblance occurs where there is the strongest signal. Thus, we use the slowness grids calculated at t_{max} . We performed our signal analysis in the frequency band 0.5 to 3.5 Hz. (Fig.

5c). Note that the signal-to-noise ratio is lower in this frequency band. We calculated semblance values over a 50-by-50 slowness grid with values between -0.3 and 0.3 sec/km.

Our goal in creating 100 samples of the "same" event was to test whether jackknife intervals calculated for the slowness component estimates would approximate 95% confidence intervals. For each of the 100 sets of estimates, we produced jackknife confidence intervals for the mean and the median. For the east-west slowness component, 96 of the 100 jackknife intervals for the mean contained the known value (Fig. 6a), while 90 of the 100 intervals for the median contained the known value (Fig. 6b). For the north-south slowness component, 98 of the 100 intervals for the mean contained the known value (Fig. 6c), while 87 of the 100 intervals for the median did (Fig. 6d).

There is a noticeable bias in the mean estimates, particularly for the north-south slowness component. Vernon *et al.* (1994) observed that the noise field at Geyokcha has a high-velocity, directional component with a northeastern bearing. Thus, it is not surprising that at these low signal-to-noise ratios the noise field would bias the slowness estimates. Perhaps a more interesting observation is that the bias is hardly detectable in the median estimates. We feel that these observations illustrate how the complex wavelets each

interact with the data differently. Some of the wavelets interfere constructively with the noise field, while others interfere destructively. The median, unlike the mean, naturally minimizes the effects of the outlying biased estimates by only using the inner samples of the sorted data. However, the median is inefficient in its use of data, which probably explains the poorer performance of the median jackknife intervals.

Hinkley (1977) suggested that for small data sets from significantly non-normal distributions, the validity of the Student t jackknife intervals as confidence intervals was compromised. We used a quantile-quantile plot to compare our set of slowness estimates to the normal distribution. If y is the quantile of order p , it is defined as $P[X \leq y] = p$. For a discrete ordered sample set $X_1 \leq X_2 \leq \dots \leq X_n$, the sample quantile of order p is defined as $y = X_{[np]+1}$, where $[np]$ denotes the largest integer $< np$ (Dudewicz, 1976). The sample quantiles for the slowness component estimates are plotted against the quantiles from a standard normal distribution in Figure 7. Where the plot is linear, the standardized distributions are the same (Kleiner and Graedel, 1980).

It is interesting to compare the normal portion of the slowness values with the multi-wavelet average beam pattern (Fig. 3e) centered on the known slowness vector (0.0382

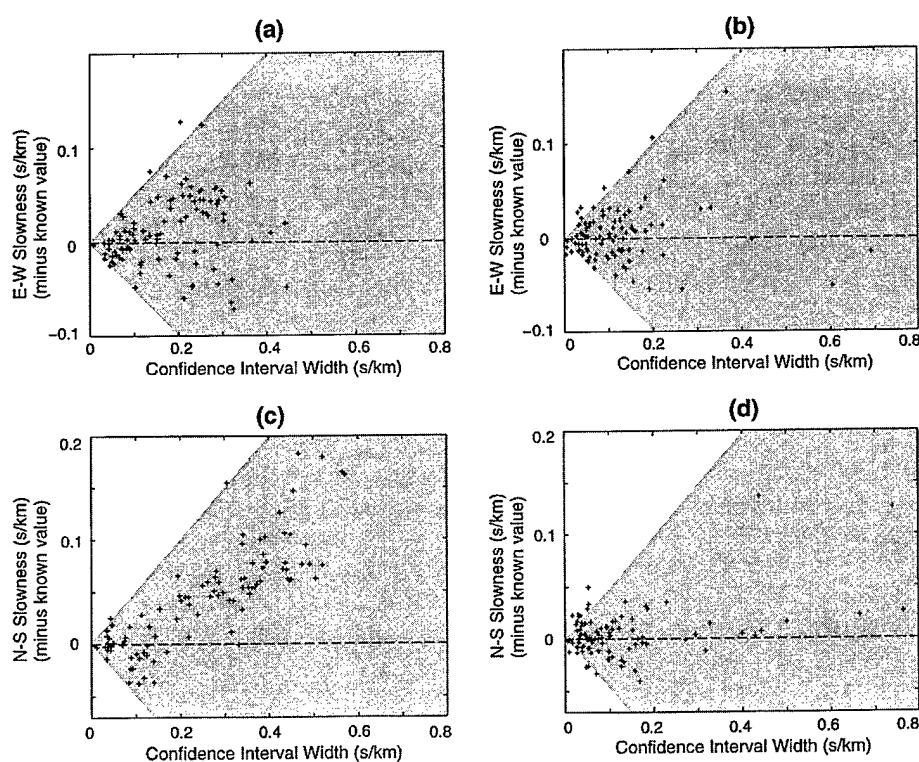


Figure 6. Plots of estimates (minus the known values) from estimator $\hat{\theta}$ versus the confidence interval width for (a) east-west slowness, $\hat{\theta}$ = mean, (b) east-west slowness, $\hat{\theta}$ = median, (c) north-south slowness, $\hat{\theta}$ = mean, and (d) north-south slowness, $\hat{\theta}$ = median. The shaded area signifies where the jackknife intervals contain the known value (plotted as a dashed line).

sec/km, -0.1533 sec/km). The beam pattern components are plotted for Figure 7a as $S(f, t, u_{ew}, -0.1533)$ versus u_{ew} and for Figure 7b as $S(f, t, 0.0382, u_{ns})$ versus u_{ns} . Note how the portion of the distribution of slowness values that matches the normal model corresponds to the peak of the beam pattern in a systematic way. We suspect that this results from interference effects at these low signal-to-noise ratios. The filtering effect of the wavelets can either enhance the noise by filtering out signal or enhance the signal by filtering out noise. Thus the filtering process can effectively change the signal-to-noise ratio. When the wavelet interaction with the data diminishes the effects of the noise field, the slowness estimates are normally distributed about the true value. However, when the wavelet interaction with the noise field leads to an effective signal-to-noise ratio much less than unity, the estimated slowness values can fall anywhere on the slowness grid. A class of robust methods called M-estimators was developed specifically to deal with statistical distributions like those for the slowness values shown in Figure 7 [compare, for example, with the quantile-quantile plots of Fig. 6 in Chave and Thomson (1989)]. In the presence of this type of data outliers, M-estimators are known to be useful in simultaneously reducing the bias and decreasing the uncertainty estimates derived from such data.

M-Estimators

M-Estimators are the generalization of what seismologists commonly call residual weights (Anderson, 1982). Details of the mathematical background for M-estimators can be found in Chave *et al.* (1987) and Press *et al.* (1992). In our case, this involves a weighted mean calculation in which the weights are adjusted repeatedly in an iterative procedure. In practice, application of an M-estimator always involves three choices: (1) a robust method for estimating the scale of the spread of the data, (2) an assumption about the underlying statistical distribution of the data, and (3) a choice for the weighting function.

In our case, we chose to use a form of the median absolute difference (MAD) as a robust estimate of scale for the spread of the data. That is, we calculate

$$s = \text{median} \{ \|\mathbf{d}_i - \boldsymbol{\theta}\| \}, \quad (28)$$

where the $\mathbf{d}_i$ are the measured slowness vectors from each of the multi-wavelets and $\boldsymbol{\theta}$ is the current estimate of the center of the underlying distribution. This measure of scale is appropriate as long as the expected value of s approximately satisfies for any $\mathbf{z}$

$$P[\|\boldsymbol{\mu} - \mathbf{z}\| \leq E[s]] \cong \frac{1}{2}, \quad (29)$$

where $\boldsymbol{\mu}$ is the true center of the distribution.

From Figure 7, we assume that the underlying data distribution is approximately Gaussian but with heavier tails. For this assumption, equation (29) holds.

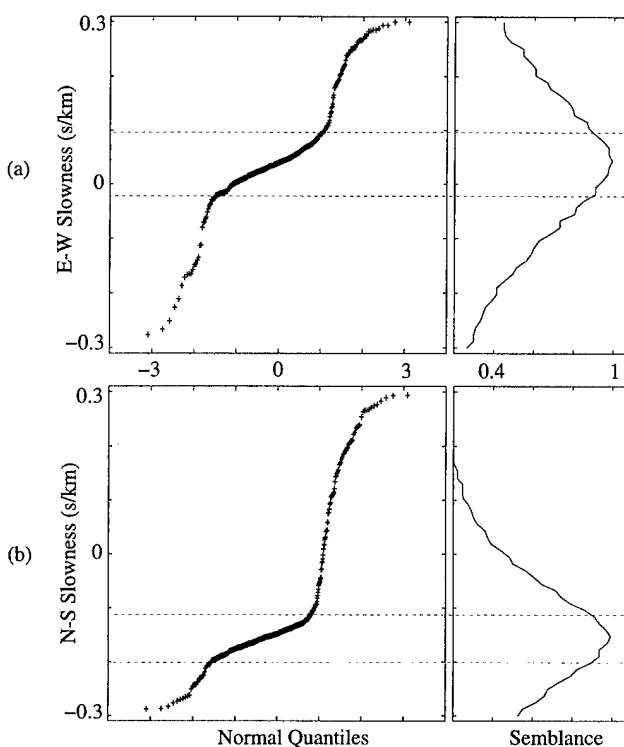


Figure 7. Quantile-quantile plots compared to one-component beam patterns for the (a) east-west and (b) north-south slowness components. The dashed lines denote where the semblance value is 0.9.

Finally, we chose to use Thomson's (1977) redescending formula for calculating the weights:

$$w(x) = \exp\{-e^{\beta(x - \beta)}\}, \quad (30)$$

where $x = \|\mathbf{d} - \boldsymbol{\theta}\|/s$ and β is a parameter that determines how far the data must be from $\boldsymbol{\theta}$ before it is downweighted. We chose to use $\beta = 3$, which roughly corresponds to the downweighting of data more than 2 standard deviations from $\boldsymbol{\theta}$.

The downweighting of outlying data can be expected to change the effective degrees of freedom of the data. A simple way to see this is if one of n data points were assigned a weight of zero, then the estimate would effectively be determined from $n - 1$ data samples. An accepted definition of the effective degrees of freedom for the M-estimator is the sum of the data weights (Anderson, 1982). We will use the closest integer less than the sum of the data weights minus one for the effective degrees of freedom.

Confidence Ellipses

Our arbitrary decomposition of the slowness vector data into north-south and east-west components for the calculation of confidence intervals could give a misleading picture of the actual estimate uncertainty since the direction of maximum error is most likely not in one of these cardinal direc-

tions. A standard presentation of error for two-dimensional data is to calculate confidence ellipses (Press *et al.*, 1992). The ellipse provides a way to define a direction of maximum uncertainty (the major axis) while delineating a compact region of uncertainty in two-dimensional space. We determined the major axis by performing a linear regression on the weighted data translated so that the origin is at the estimated center $\hat{\theta}$. The linear regression line points in the direction of greatest spread in the data. We then calculated jackknife confidence intervals along the major and minor axes by using the weighted average (with the M-estimator weights) as the estimator $\hat{\theta}$.

The M-estimator results for the 100 synthetic events are shown in Figure 8. To create a display similar to that of Figure 6, we have plotted the confidence intervals and estimates for the major and minor axes of the confidence ellipses. For the major axis, 96 of the 100 jackknife intervals contained the known value, while 94 of the 100 intervals for the minor axis contained the known value. We can see that the estimates are not severely biased like those of the mean, while the jackknife intervals still behave as 95% confidence intervals. Thus, the M-estimator does a better job than either the mean or the median. All further results presented were determined using this M-estimator.

Examples with Real Data

An Event Swarm

We have gained some insights into how multi-wavelet beamforming behaves with a controlled simulation. We now investigate how it behaves with a set of similar seismic events with varying signal-to-noise conditions. The recordings of four events at station ORGH are shown in Figure 9. These events took place over a 20-min period on 26 October 1993 and have similar waveforms. We analyzed the first arrivals of these events in the frequency range of 2 to 14 Hz. The performance of the Geyokcha array at these frequencies improves when the outlying stations of SEH, SWH, and NH are removed, so we conducted our analysis with the nine inner stations.

The slowness estimates and their confidence ellipses are plotted in Figure 10. Based on the area of the confidence ellipses, we rank the estimates from best to worst: event 2, event 1, event 3, and event 4. This ordering is consistent with ordering by signal-to-noise ratios as seen in the seismograms and the average semblance plots of Figure 9. We note that the slowness vector for event 2 is statistically distinct from those of events 1 and 3. On further inspection of the signals in Figure 9, we can see that for events 1, 3, and 4 there is a second upward impulse closely following the initial upward impulse that does not appear to be present on event 2. This is possibly due to a difference in the sources for event 2 and the other three events, but it would most likely have been dismissed as due to the noise field without the statistical evidence for distinct slowness vectors.

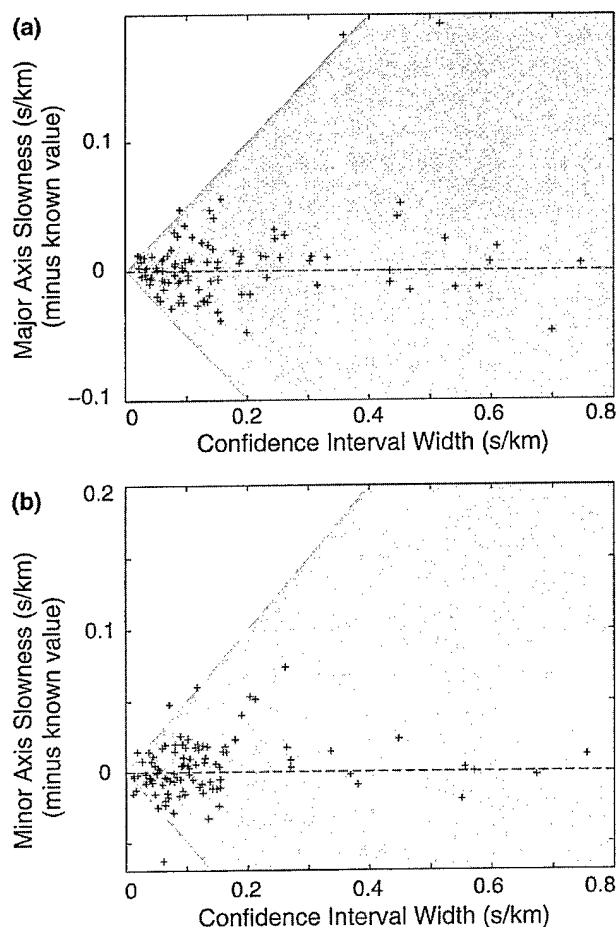


Figure 8. Plots of estimates (minus the known value) from estimator $\hat{\theta} = \text{M-estimator}$ versus the confidence interval width for the (a) major and (b) minor axes of the confidence ellipse. The shaded area signifies where the jackknife intervals contain the known value (plotted as a dashed line).

Impulsive versus Emergent Signals

Both the simulated plane-wave and the swarm events have impulsive first-arrival phases. Many regional phases, however, are highly emergent, and it is important to see how this method responds to such phases. We compare average semblance plots for an emergent and an impulsive event in the frequency band 1 to 7 Hz in Figure 11. The character of these two types of events is preserved in the average semblance plot. Choosing the time window for estimating the slowness components of the emergent signal's first arrival(s) is not as straightforward as with the impulsive signals. Since the wavelets are each 1-sec long, we certainly cannot resolve semblance peaks closer than 0.25 sec apart. The initial 2.5 sec of the emergent event yields a series of six semblance peaks spaced at intervals of approximately 0.4 sec. The maximum for each cycle is marked by an arrow on the average semblance plot of Figure 11a. We analyzed each of these local maxima as a distinct arrival and refer to them by order

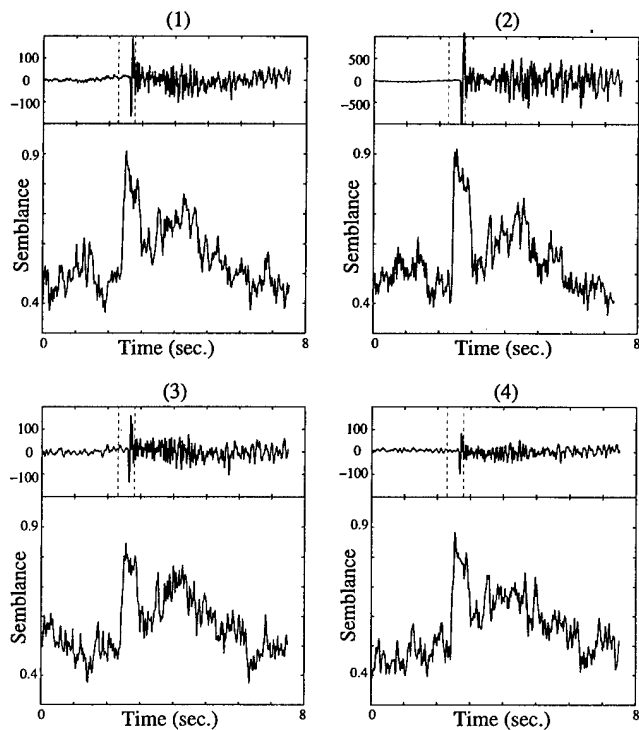


Figure 9. Vertical records from station ORGH for four similar events. Note the seismogram scale difference for event 2. Under each trace is a plot of the average semblance versus time. The dotted lines delineate the time window for $t_{\max}$.

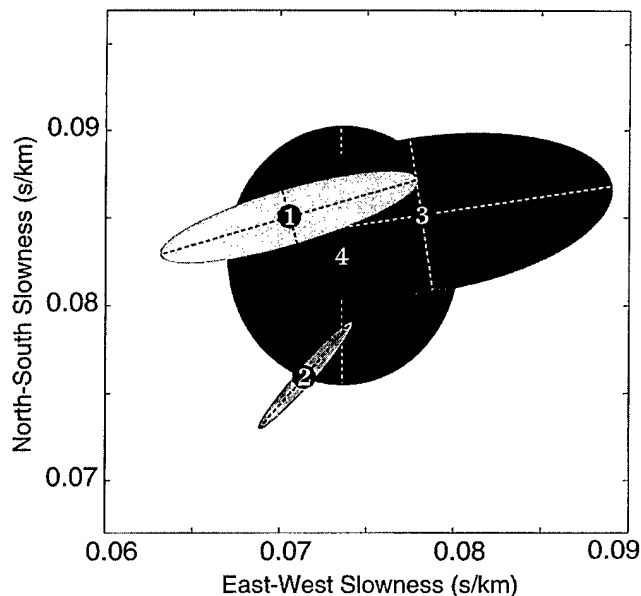


Figure 10. Plot of slowness vectors and confidence ellipses for the four similar events of Figure 9. Note the coordinate system scale.

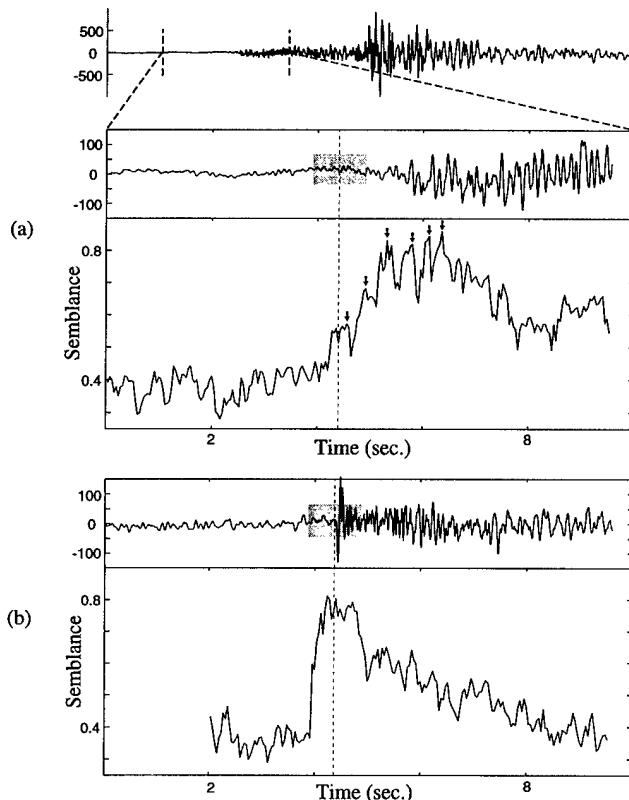


Figure 11. Vertical records from station ORGH for (a) an emergent event and (b) an impulsive event. The shaded boxes delineate the time length of the wavelets used. Under each trace is a plot of the average semblance versus time. The light dotted line denotes the time of the first arrival. The arrows in (a) correspond to the six arrivals whose slowness vector estimates are plotted in Figure 12.

of arrival. The slowness estimates and their confidence ellipses are plotted in Figure 12. The character of the signal (Fig. 11a) appears to change between the third and fourth arrival, and the slowness vector estimates also appear to cluster into slowness vectors with more easterly bearings for arrivals before and including the third arrival and vectors with more westerly bearings for those after the third arrival. However, when we look at the confidence ellipses, there is no statistical differentiation between any of the slowness vector estimates. Thus, this clustering effect must be considered marginal at best.

Conclusions

We have demonstrated that multi-wavelet beamforming has a distinct advantage over existing seismic array processing methods in that it produces multiple estimates of the slowness vector that can be used to calculate objective statistical confidence intervals. We compared confidence intervals calculated using the jackknife statistical method for three estimators of the "center" of the distribution of slow-

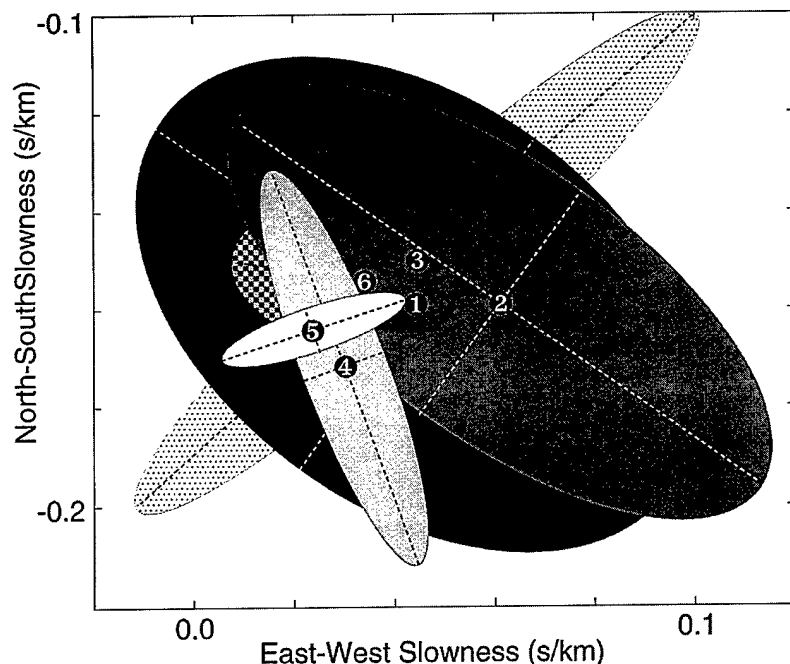


Figure 12. Plot of slowness vectors and confidence ellipses for six arrivals from the emergent event of Figure 11a. Note the coordinate system scale.

ness components—the median, the mean, and an M-estimator.

The jackknife intervals obtained from the mean estimator are conservative since they weight all data equally, even gross outliers, and our simulation suggests that they closely approximate true 95% confidence intervals. However, the mean estimates are severely biased by outlying values. On the other hand, the median effectively reduced the bias in the slowness estimates introduced from the noise field in low signal-to-noise conditions. However, the jackknife intervals determined using the median underestimated the uncertainty closer to 10% of the time.

The best estimator we investigated was the M-estimator. Our simulation results suggest that the M-estimator effectively reduced the bias in the slowness vector estimates, while still producing jackknife intervals that closely approximate true 95% confidence intervals. The M-estimator had a further advantage in that it dealt with the slowness vector data samples as two-dimensional objects instead of two one-dimensional components. This allowed us to determine confidence ellipses for the data instead of the somewhat artificial confidence intervals along the east-west and north-south components.

For the real data example of the swarm of events, we noted that, as would be expected, the area of the confidence ellipses had a negative correlation with the signal-to-noise ratios of the data. We were also able to determine that the source for one event was statistically different from the sources of two of the other events. For the example of the emergent event, we discovered an underlying series of average semblance peaks spaced at intervals of approximately 0.4 sec. We treated these peaks as distinct phase arrivals and observed that the slowness estimates clustered according to

a noted change in signal characteristic. However, their confidence ellipses all mutually overlapped, rendering this observation as statistically unconvincing.

It should be relatively straightforward to implement multi-wavelet beamforming as an automated tool. The computational machinery built up for time-domain beamforming can easily be changed to produce the slowness grids and vector estimates for our method. The number of computations should also be tractable since the wavelet transforms only need to be calculated at a single time for each slowness grid.

Acknowledgments

The data we examined here were collected under the IRIS (a facilities program of the National Science Foundation) Joint Seismic Program. We are thankful to a long list of people who were responsible for the collection of these data. This work was supported by the Air Force Office of Scientific Research (under the AASERT program), Grant No. F49620-95-1-0366.

References

- Al-Shukri, H. J., G. L. Pavlis, and F. L. Vernon (1995). Site effect observations from broadband arrays, *Bull. Seism. Soc. Am.* **85**, 1758–1769.
- Anderson, K. R. (1982). Robust earthquake location using M-estimates, *Phys. Earth Planet. Interiors* **30**, 119–130.
- Bratt, S. R. and T. C. Bache (1988). Locating events with a sparse network of regional arrays, *Bull. Seism. Soc. Am.* **78**, 780–798.
- Chave, A. D. and D. J. Thomson (1989). Some comments on magnetotelluric response function estimation, *J. Geophys. Res. B, Solid Earth Planets* **94**, 14215–14225.
- Chave, A. D., D. J. Thomson, and M. E. Ander (1987). On the robust estimation of power spectra, coherences, and transfer functions, *J. Geophys. Res. B, Solid Earth Planets* **92**, 633–648.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania.

- Dudewicz, E. J. (1976). *Introduction to Statistics and Probability*, Holt, Rinehart and Winston, New York.
- Efron, B. and C. Stein (1981). The jackknife estimate of variance. *Ann. Stat.* **9**, 586–596.
- Harvey, D. (1994). *dbap (a public domain computer program)*, JSPC, University of Colorado, Boulder.
- Hinkley, D. V. (1977). Jackknife confidence limits using Student *t* approximations, *Biometrika* **64**, 21–28.
- Husebye, E. S. and B. O. Ruud (1989). Array seismology; past, present, and future developments, in *Observatory Seismology*, J. J. Litchiser (Editor), University of California Press, Berkeley, 123–153.
- Kanasewich, E. R. (1981). *Time Sequence Analysis in Geophysics*, University of Alberta Press, Edmonton, Alberta, Canada.
- Kleiner, B. and T. E. Graedel (1980). Exploratory data analysis in the geophysical sciences, *Rev. Geophys. Space Phys.* **18**, 699–717.
- Kumar, P. and E. Fofoula-Georgiou (1994). Wavelet analysis in geophysics: an introduction, in *Wavelets in Geophysics*, E. Fofoula-Georgiou and P. Kumar (Editors), Academic Press, San Diego, California, 1–43.
- Kvaerna, T. and D. J. Doornbos (1986). *An Integrated Approach to Slowness Analysis with Arrays and Three-Component Stations*, NORSAR 2-85/86.
- Lilly, J. M. and J. Park (1995). Multiwavelet spectral and polarization analyses of seismic records, *Geophys. J. Int.* **122**, 1001–1021.
- Miller, K. S. (1974). *Complex Stochastic Processes*, Addison-Wesley, Reading, Massachusetts.
- Miller, R. G. (1974). The jackknife—a review, *Biometrika* **61**, 1–15.
- Morlet, J., G. Arens, E. Fourgeau, and D. Giard (1982). Wave propagation and sampling theory: Part II, Sampling theory and complex waves, *Geophysics* **47**, 222–236.
- Park, J., C. R. Lindberg, and F. L. Vernon (1987). Multitaper spectral analysis of high-frequency seismograms, *J. Geophys. Res. B, Solid Earth Planets* **92**, 12675–12684.
- Pavlis, G. L. and H. Mahdi (1996). Surface wave propagation in central Asia: Observations of scattering and multipathing with the Kyrgyzstan broadband array, *J. Geophys. Res.* **101**, 8437–8455.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C: The Art of Scientific Computing*, Second Ed., Cambridge Univ. Press, Cambridge, U.K.
- Slepian, D. (1983). Some comments on Fourier analysis, uncertainty and modeling, *Soc. Ind. Appl. Math. Rev.* **25**, 379–393.
- Thomson, D. J. (1977). Spectrum estimation techniques for characterization and development of WT4 waveguide, I, *Bell Syst. Tech. J.* **56**, 1769–1815.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis, *Proc. IEEE* **70**, 1055–1096.
- Thomson, D. J. and A. D. Chave (1991). Jackknifed error estimates for spectra, coherences, and transfer functions, in *Advances in Spectrum Analysis and Array Processing*, S. Hayken (Editor), Prentice-Hall, Englewood Cliffs, New Jersey, 58–113.
- Vernon, F. L., D. J. Harvey, G. L. Pavlis, W. Y. Kim, G. A. Abers, and G. S. Wagner (1994). Array studies in the Joint Seismic Program (abstract), *EOS* **75**,

and $\lambda_2 \mathbf{w}_2 = \mathbf{A}_2 \mathbf{w}_2$, where $\mathbf{A}_1$ and $\mathbf{A}_2$ are the same size. If we compare these two matrices, we find

$$\begin{aligned} A_{mn,1} &= \frac{\sin [2\pi(f_{c,1} + f_{w,1}) \Delta t_1 (m - n)]}{\pi(m - n)} \\ &\quad - \frac{\sin [2\pi(f_{c,1} - f_{w,1}) \Delta t_1 (m - n)]}{\pi(m - n)} \\ &= \frac{\sin [2\pi \frac{1}{k} (f_{c,2} + f_{w,2}) (k\Delta t_2) (m - n)]}{\pi(m - n)} \\ &\quad - \frac{\sin [2\pi \frac{1}{k} (f_{c,2} - f_{w,2}) (k\Delta t_2) (m - n)]}{\pi(m - n)} = A_{mn,2}. \end{aligned} \quad (\text{A1})$$

Since λ and $\mathbf{w}$ are dependent on $\mathbf{A}$, $\mathbf{w}_1$ and $\mathbf{w}_2$ are exact scaled versions of one another.

Appendix B

We are interested in showing that the slowness component estimates are approximately statistically independent. We start by looking at the multi-wavelet transform. Ignoring dependence on time and frequency for now, the multi-wavelet transform can be written

$$\mathcal{W}^{(j)}[\eta_n] = \sum_{t=1}^M w_t^{(j)} \eta_{n,t}, \quad (\text{B1})$$

where $w^{(j)}$ is the j th complex Lilly and Park wavelet with M samples. Assume $\{\eta_{n,t}\}$ is a Gaussian noise segment from station n with zero mean and unit variance. It follows that $z_1 = \mathcal{W}^{(j)}$ is a complex Gaussian with zero mean and probability density function

$$f(z_1) = \frac{1}{\pi E[|z_1|^2]} \exp\left(\frac{-|z_1|^2}{E[|z_1|^2]}\right), \quad (\text{B2})$$

where $E[\cdot]$ denotes expected value (Miller, K., 1974).

$$\text{If } \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} \mathcal{W}^{(j)}[\eta_n] \\ \mathcal{W}^{(k)}[\eta_m] \end{bmatrix}$$

and

$$\mathbf{R} = \begin{bmatrix} E[|z_1|^2] & E[z_1 z_2^*] \\ E[z_2 z_1^*] & E[|z_2|^2] \end{bmatrix},$$

then the joint probability density function for $\mathbf{z}$ (Miller, K., 1974) is

$$f(\mathbf{z}) = \frac{1}{\pi^2 \det(\mathbf{R})} \exp(-\mathbf{z}^* \mathbf{R}^{-1} \mathbf{z}). \quad (\text{B3})$$

If $E[z_1 z_2^*] = E[z_2 z_1^*] = 0$, then it is easy to show that $f(\mathbf{z})$

Appendix A

We will show that if the sampling rate and the frequency band width are scaled simultaneously as stated below, then the Lilly and Park wavelets are exact scaled versions of one another. We start with two frequency bands with center frequencies $f_{c,1} = \frac{1}{k} f_{c,2}$ and bandwidths $f_{w,1} = \frac{1}{k} f_{w,2}$, where k is an integer. We also choose to decimate the data by k (i.e., $\Delta t_1 = k\Delta t_2$). If we keep the number of samples that define the wavelets constant, we have two equations $\lambda_1 \mathbf{w}_1 = \mathbf{A}_1 \mathbf{w}_1$

$= f(z_1)f(z_2)$. This implies that z_1 and z_2 are independent. So we will now investigate

$$E[\mathcal{W}^{(j)}[\eta_n](\mathcal{W}^{(k)}[\eta_m])^*] = \sum_{y=1}^M \sum_{p=1}^M w_y^{(j)} (w_p^{(k)})^* E[\eta_{n,t} \eta_{m,p}]. \quad (\text{B4})$$

Case 1: $n = m, j = k$

By the assumptions made about the noise field, we know that

$$E[\eta_{n,t} \eta_{m,p}] = \delta_{mn} \delta_{tp}, \text{ where } \delta_{xy} = \begin{cases} 0 & x \neq y \\ 1 & x = y \end{cases}. \quad (\text{B5})$$

So for this case,

$$E[\mathcal{W}^{(j)}[\eta_n](\mathcal{W}^{(j)}[\eta_n])^*] = \sum_{t=1}^M w_t^{(j)} (w_t^{(j)})^* = 2. \quad (\text{B6})$$

The sum is equal to 2 because the real wavelets are individually normalized to have an $L2$ norm of unity.

Case 2: $n = m, j \neq k$

In this case,

$$E[\mathcal{W}^{(j)}[\eta_n](\mathcal{W}^{(k)}[\eta_n])^*] = \sum_{t=1}^M w_t^{(j)} (w_t^{(k)})^* = 0 \quad (\text{B7})$$

by the orthogonality of the real wavelets.

Case 3: $n \neq m$

For all values of t and p , $E[\eta_{n,t} \eta_{m,p}] = 0$. Thus in this case,

$$E[\mathcal{W}^{(j)}[\eta_n](\mathcal{W}^{(k)}[\eta_m])^*] = 0, \quad (\text{B8})$$

even if $j = k$.

From these three cases, we see that the transforms are independent in the presence of Gaussian noise.

We will now reinstate the frequency and time dependence to the wavelet transform. The frequency dependence does not change any of the previous statistical arguments, since we only use the jackknife method on estimates determined from a single set of wavelets. There are, however, differences in the time delays used from one slowness estimate to the next (otherwise the J slowness estimates would all be exactly the same!). This dependence does not affect case 1 or case 3, but it does change case 2. We define

$$\xi_{jk}(s) = E[\mathcal{W}^{(j)}[\eta_n](f,0)(\mathcal{W}^{(k)}[\eta_n](f,s\Delta t))^*] = \sum_{q=s+1}^M w_q^{(j)} (w_{q-s}^{(k)})^*, \quad (\text{B9})$$

where the time delay $\tau = s\Delta t$. This is the cross-correlation function, and it affects the argument for independence as the off-diagonals of the matrix $\mathbf{R}$. As an informal order of magnitude argument, we plot the sample offsets s against

$$\rho_{jk}(s)^2 = |\xi_{jk}(s)|^2 / E[|\mathcal{W}^{(j)}[\eta_n](f,0)|^2] E[|\mathcal{W}^{(k)}[\eta_n](f,s\Delta t)|^2], \quad (\text{B10})$$

which is the ratio of the magnitudes of the off-diagonals with the magnitudes of the diagonals of $\mathbf{R}$. We can think of $\rho_{jk}(s)^2$ as the square of the correlation coefficient for $\mathcal{W}^{(j)}$ and $\mathcal{W}^{(k)}$. Using the same Lilly and Park wavelets as in Figure 1, we can see in Figure B1 that these ratios are small for small time shifts. Thus we will assume that $\xi_{jk}(s)$ is close enough to zero that $f(\mathbf{z}) \approx f(z_1)f(z_2)$, which would imply that the wavelet transforms are statistically independent. Equation (9) defines the slowness component estimates as functions of semblance where the semblance for the j th Lilly and Park wavelet is

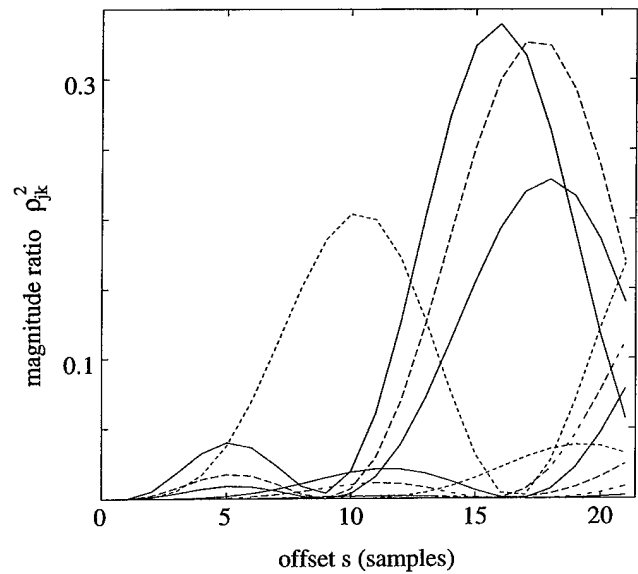


Figure B1. Plot of the magnitude ratio ρ_{jk}^2 versus the sample offset s , where the relative time delay between complex wavelets j and k is $s\Delta t$. Ratios for all possible combinations of pairs of the five wavelets are plotted.

$$S^{(j)}(\mathbf{u}) = \frac{\left| \frac{1}{N} \sum_{n=1}^N W^{(j)}[\eta_n][f, t + \tau(\mathbf{u}, n)] \right|^2}{\frac{1}{N} \sum_{n=1}^N \left| W^{(j)}[\eta_n][f, t + \tau(\mathbf{u}, n)] \right|^2}. \quad (\text{B11})$$

Since the slowness component estimates are ultimately functions of the independent wavelet transforms, they are then themselves independent.

In some cases, the slowness estimates may differ drastically. This would violate our assumption that the time shifts

are small and thus the assumption of independence. This does not pose any problems in practice, however, since the jackknife confidence intervals in these cases would be very large anyway. A practical solution would be simply to consider such measurements ill constrained.

Department of Geological Sciences
1005 East Tenth Street
Indiana University
Bloomington, Indiana 47405

Manuscript received 18 July 1996.

Multi-wavelet Analysis of Three-Component Seismic Arrays: Application to Measure Effective Anisotropy at Piñon Flats, California

by Lorie K. Bear,* Gary L. Pavlis, and Götz H. R. Bokelmann

Abstract We develop and apply a new technique to determine array-averaged particle motions from three-component seismic array data. The method is based on multi-wavelets, which are an extension of multi-taper spectral methods, and is a hybrid of Fourier and time-domain methods of array processing. Particle motions are determined by a time-domain principal-component method. A complex singular value decomposition is used on wavelet transformed signals assembled into multiple matrices (one for each wavelet). The eigenvector of the largest singular value of each matrix is used to estimate the phase between individual signals. We determine the relative phase between components to estimate an average particle motion ellipse for the array. The estimation procedure is made more stable by the redundancy inherent in the multi-wavelets and by M-estimators applied to individual phase factors in the complex plane. The method is applied to data from three-component array experiments conducted at Piñon Flats, California, in 1990 and 1991. We find remarkable departures of *P*-wave particle motions from the pure longitudinal motion expected for an isotropic media. Anomalies as large as 40° are measured from some azimuths. The azimuthally varying particle-motion anomalies are frequency dependent, generally increasing in magnitude as frequency increases. Borehole measurements from sensors at 153 and 274 m depth below the array show a pattern indistinguishable from the surface sensors. The data are fit with a dipping, transversely isotropic medium with a symmetry plane having a strike of 70° and a dip of 30° to the northwest. We attribute our results to three superimposed effects: (1) an anisotropy of the near surface induced by preferential weathering of the granodiorite bedrock along joints, (2) a larger scale anisotropy induced by structural and intrinsic anisotropy related to the Santa Rosa mylonite, and (3) near-surface scattering.

Introduction

In the past 8 years, a wide range of experiments have been fielded using arrays of three-component seismic stations. These arrays span apertures from 0.1 to 1000 km and provide new data on wave propagation by sampling the three-dimensional wave field at many length scales. This article focuses on quantifying the relationship of particle motions and phase velocities (slowness vectors)—a unique capability of a three-component array. This article has two distinct contributions. First, the methodology for three-component array analysis that we introduce here is new and utilizes some fundamentally different approaches from previous work. Second, the observations we make in applying this new methodology are remarkable. We find strong evidence for large departures of *P*-wave particle motions from those

predicted for a stratified, isotropic Earth—the prevailing theoretical model of the Earth in seismology.

The first part of this article introduces a new processing method for three-component arrays that simultaneously estimates three features of the incident wave field: (1) best-fitting particle motion ellipses for each station, (2) an average particle motion ellipse for the entire array, and (3) the slowness vector for a best-fitting plane wave traveling across the array. This method incorporates aspects of time- and frequency-domain beamforming (e.g., Pavlis and Mahdi, 1996; Kvaerna and Doornbos, 1986), and principal-component analysis (Vidale, 1986). The principal-component analysis is applied to all the station and component data simultaneously, much like the multi-channel detector described in Wagner and Owens (1996). Our analysis is performed on data that has been multi-wavelet transformed (similar to a windowed Fourier transform) so that the anal-

*Present address: Exxon Production Research, P.O. Box 21879, Houston, Texas.

ysis reflects the signal's behavior in a particular time window and frequency band.

The methodology developed in this article is closely related to that described in a companion article (Bear and Pavlis, 1999, this issue), and we will lean on some of the theoretical development contained therein. Both of these methodologies are dependent on the robust measurement of relative phase: either from station to station or between components of a single station. The companion article uses the robust phase measurement between stations to determine static time residuals such as would be used for seismic tomography. Here, we instead use the phase measurements between components to determine particle motion ellipses.

The most extensive recent work on three-component seismic array methodology can be found in a comprehensive suite of articles by Wagner and Owens (Wagner, 1994, 1997; Wagner and Owens, 1993, 1995, 1996). Our work, in many respects, leans heavily on these previous articles, but our approach differs in two main ways. First, our approach utilizes a recent innovation we will refer to as the multi-wavelet transform (Lilly and Park, 1995). Wagner and Owen's work focuses primarily on Fourier methods. The multi-wavelets provide a hybrid between conventional time and Fourier domain array processing methods that we will argue have some significant advantages. Second, Wagner and Owen's articles focus on so-called high-resolution methods. We experimented with high-resolution methods (Bear and Pavlis, 1997b) but found utilizing them in combination with the multi-wavelet transform led to serious stability problems. Instead of trying to maximize resolution, our approach focuses on robustness that we achieve through redundancy inherent in the multi-wavelets and the inherent redundancy of array data. We would argue this is an important strength of this methodology over other approaches that have been applied to this problem.

On the observational side, the closest previous work to this article was that done by Bokelmann (1995a, 1996). Bokelmann used data from GERESS, an array in southern Germany composed of 25 single-component stations. Five of these stations have co-located three-component sensors. Bokelmann determined the slowness vector by processing the vertical-component array data. He then used conventional particle motion analysis methods applied to each of the five three-component stations at GERESS and averaged the results. Our results differ in that all the data we examine are from full three-component seismic arrays. As a result, we determine a best slowness vector and best-fitting particle motion ellipse simultaneously from all the available data. In addition, the methodology developed here can be applied in multiple frequency bands to examine the frequency dependence of particle motion relative to a measured phase velocity. On the other hand, our results can be directly related to Bokelmann's (1995a) as we apply the same inversion procedure below to determine an anisotropic model to fit our observational data.

Method

The approach we use here is closely related to ideas described in a companion article by Bear and Pavlis (1999). Some background is found in that article that will not be repeated here for the sake of brevity. The primary difference is in how we exploit the relative phase between a group of signals. Consider a pure harmonic signal at frequency f . In this simple situation, the phase differences between stations m and n can be considered equivalent to a time difference between the two signals through the relation

$$e^{i(\theta_m - \theta_n)} = e^{i2\pi f(t_m - t_n)}. \quad (1)$$

This relation was used in Bear and Pavlis (1999) to determine time residuals. On the other hand, the vector formed from the three components of a given station for a pure harmonic signal

$$[r_1 e^{i(2\pi f t + \phi_1)}, r_2 e^{i(2\pi f t + \phi_2)}, r_3 e^{i(2\pi f t + \phi_3)}] \quad (2)$$

defines an ellipse of instantaneous particle motion in three dimensions. The details of the ellipse are determined by the magnitude and relative phase of the three different components. It is this relation that is most applicable to the problem at hand. The key point is that with signals that are localized in frequency, the relative phase is the key quantity to be measured.

For an array of N three-component stations and K time samples, we start by creating the $3N \times K$ complex data matrix $\mathbf{A}(f, t, \mathbf{p})$, which can be partitioned into three $N \times K$ matrices:

$$\mathbf{A}(f, t, \mathbf{p}) = \begin{bmatrix} \mathbf{A}^x(f, t, \mathbf{p}) \\ \mathbf{A}^y(f, t, \mathbf{p}) \\ \mathbf{A}^z(f, t, \mathbf{p}) \end{bmatrix},$$

where

$$A_{nk}^c(f, t, \mathbf{p}) = \mathcal{W}[s_n^c][f, t_k + \tau(\mathbf{p}, n)]. \quad (3)$$

$\mathcal{W}$ denotes an integral transform for studying the signal behavior near frequency f , s_n^c is the data recorded on the c th component of the n th station, and the $\tau(\mathbf{p}, n)$ are time shifts to account for travel-time differences between stations due to a plane-wave arrival with slowness vector $\mathbf{p}$, or due to some more complex travel-time function [e.g., τ may contain a static correction relative to a plane-wave model determined as described in Bear and Pavlis (1999)]. The integral transform we choose to use is the multi-wavelet transform developed by Lilly and Park (1995). We make this choice largely because of an important advantage it inherently has in applications to modern broadband data: the multi-wavelet functions are "wavelets" in the sense that their timescales can be adjusted to match the frequency scale being resolved.

That is, wavelets for lower frequency bands are naturally longer in time than those for higher frequency bands.

The rows of $\mathbf{A}$ consist of K samples of the wavelet transformed data, each over a slightly different time window. If we assume that the signal of interest propagates with a constant slowness vector over the times involved, then each column of $\mathbf{A}$ provides the same information on the signal plus the effect of noise. We use times $t_k = [t + (k - 1)\Delta t]$, $k = 1, \dots, K$, where Δt is chosen so that $t_K - t_1$ is the length of one-half to one cycle of the center frequency f . This choice is not unique, but we have found it useful for reducing the computational load of this procedure.

The complex matrix $\mathbf{A}$ can be written as a singular value decomposition

$$\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\dagger, \quad (4)$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Lambda}$ is a real diagonal matrix. We assume the SVD is organized such that $\lambda_1 \geq \lambda_2 \dots \geq \lambda_{\min(N,K)}$. $\mathbf{U}$ can be considered a rotation of the data such that the column

$$\mathbf{u}_1(f, t, \mathbf{p}) = \begin{bmatrix} \mathbf{u}_1^x(f, t, \mathbf{p}) \\ \mathbf{u}_1^y(f, t, \mathbf{p}) \\ \mathbf{u}_1^z(f, t, \mathbf{p}) \end{bmatrix} \quad (5)$$

points in the direction of largest energy. We then perform the beamforming process by searching through possible slowness vectors $\mathbf{p}$. We choose the best slowness vector by maximizing the coherence measure

$$S(f, t, \mathbf{p}) = \max_{c=x,y,z} \frac{\|\lambda_1(f, t, \mathbf{p}) \mathbf{d} \cdot \mathbf{u}_1^c(f, t, \mathbf{p})\|^2}{\|\mathbf{d}\|^2 \|\lambda_1(f, t, \mathbf{p}) \mathbf{u}_1^c(f, t, \mathbf{p})\|^2}, \quad (6)$$

where $d_n = 1/N$ for $n = 1, \dots, N$. This measure is analogous to semblance (Husebye and Ruud, 1989) in that the denominator measures the average of the station powers and the numerator measures the power of the stack.

The matrix $\mathbf{A}$ can also be partitioned by stations such that

$$\mathbf{A}(f, t, \mathbf{p}) = \begin{bmatrix} \mathbf{A}^1(f, t, \mathbf{p}) \\ \vdots \\ \mathbf{A}^N(f, t, \mathbf{p}) \end{bmatrix},$$

where

$$\mathbf{A}_{ck}^n(f, t, \mathbf{p}) = \mathcal{W}[s_n^c][f, t_k + \tau(\mathbf{p}, n)]. \quad (7)$$

Each of the partition matrices can be considered as a portion of a 3×3 covariance matrix

$$\mathbf{R}(f, t, \mathbf{p}) = \frac{1}{K} \mathbf{A}^n(f, t, \mathbf{p}) \cdot \mathbf{A}^{n\dagger}(f, t, \mathbf{p}). \quad (8)$$

We showed in Bear and Pavlis (1997a) that the real and imaginary parts of the multi-wavelet transforms behave as the analytic filtered signal. Thus $\mathbf{R}(f, t, \mathbf{p})$ is analogous to the 3×3 covariance matrix used by Vidale (1986) for principal-component analysis of the signal's particle motion. We have already determined the N principal components in

$$\mathbf{u}_1(f, t, \mathbf{p}) = \begin{bmatrix} \mathbf{u}_1^1(f, t, \mathbf{p}) \\ \vdots \\ \mathbf{u}_1^N(f, t, \mathbf{p}) \end{bmatrix} \quad (9)$$

so that the complex values in the eigenvector associated with the largest singular value, $\mathbf{u}_1^n$, describe the particle motion for the signal at station n over the time and frequency windows used.

The use of the multi-wavelet transform is particularly appropriate here, since we are interested in a phenomenon that is expected to vary significantly both in time and in frequency. The real and imaginary parts of the kernels to the multi-wavelet transforms are specifically designed as real, discrete, finite time series that have their energy concentrated within a frequency band defined by the center frequency f and a bandwidth $2f_w$, where $f_w \leq f$ (Lilly and Park, 1995; Bear and Pavlis, 1997a). The form of the multi-wavelets we used here can be found in Figure 1 of the companion article by Bear and Pavlis (1999). The multi-wavelet functions occur in even and odd pairs, where each pair emphasizes a different portion of the time and frequency windows. The lengths of the time and frequency windows determine how many transforms can be used (Lilly and Park, 1995). In this study, we produced five usable transforms for each fre-

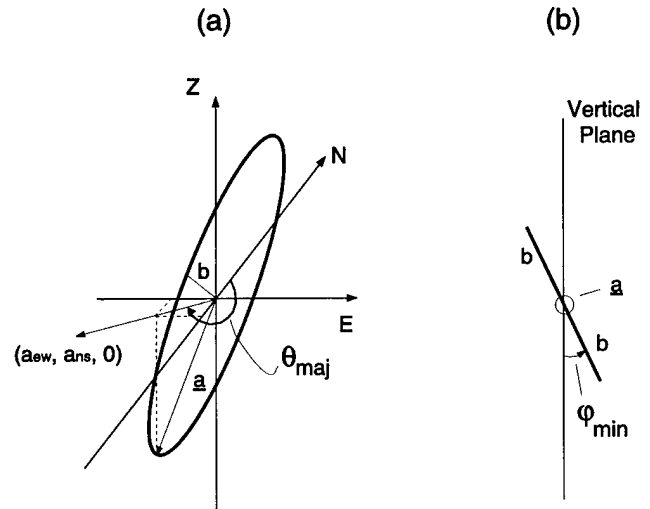


Figure 1. Picture of an ellipse in (a) three-dimensional view and (b) a view where the major axis $\mathbf{a}$ is pointing out of the page. By convention, $\phi_{\min}$ is positive when the minor axis is rotated clockwise and negative when rotated counter-clockwise. In this case, θ_{maj} would be approximately 240° , and $\phi_{\min}$ would be approximately -25° .

quency band following Bear and Pavlis (1997a). We emphasize, however, that this choice is not unique. Longer analysis windows and/or different frequency bandwidths could be used to define different wavelets with a differing number of usable functions.

It is important to emphasize that our method is a multi-channel procedure. It determines estimates of elliptical particle motions for each of the N stations in the context of producing the most coherent beam for all three components of the entire array. It is also important to remember that we have N ellipse estimates over a small spatial area for each of the five multi-wavelet transforms. This means that we can parlay this redundancy into more stable results for each station (by using the five pieces of data from the different transforms) and for the array as a whole (by using the $5N$ pieces of data from all the stations and the different transforms). Note, however, that the absolute phase estimates determined using different multi-wavelet functions cannot be compared directly, due to the relative phase differences between the complex wavelets that we have found no way to unambiguously resolve.

An ellipse can be completely described by the spatial orientations and lengths of the major and minor axes. We characterize the particle motion by three main observables (Fig. 1). First, there is the linearity of particle motion for station n . We use a standard measure called rectilinearity defined as $e_n = 1 - b_n/a_n$, where $2a_n$ and $2b_n$ are the major and minor axes lengths, respectively. Second, we measure the azimuth of the projection of the major axis onto the horizontal plane that we will refer to with the symbol $\theta_{maj,n}$. Finally, we compute the angle between the minor axis and the vertical plane $\phi_{min,n}$ when looking down the major axis. For P waves propagating in a homogeneous, isotropic, horizontally layered medium, the ellipticity in the particle motion will be induced mainly by P -to- S conversions at the layer contacts. Thus, the elliptical motion should be contained in the radial-vertical plane such that $\theta_{maj,n}$ should be the same as the propagation azimuth (determined from the slowness vector) and $\phi_{min,n}$ should be zero. We will refer the deviation of the major axis projection angle from the propagation azimuth ($\theta_{maj,n} - \theta_p$) as the major axis skew, and the angle between the minor axis and the vertical plane $\phi_{min,n}$ as the minor axis tilt.

We need similar values to describe the average elliptical particle motion for the entire array. There are multiple ways that one could average the results obtained for each station from the individual multi-wavelet transforms. We chose a dual averaging scheme that is made robust by using an M-estimator (Bear and Pavlis, 1997a; Chave *et al.*, 1987). The M-estimator is used at each averaging step because it removes the effects of any outlying values. For the rectilinearity e , we average over the e_n determined from each multi-wavelet transform separately and then average the values for the five multi-wavelet transforms to obtain the final value. We determine an average array major axis by first averaging all the station major axes determined using a given

multi-wavelet transform. This leaves us with five samples of the array major axis orientation—one for each transform. In this article, we are primarily interested in determining the azimuth of the projection of the major axis onto the horizontal plane. To compute this angle, we normalize the lengths of the five major axis samples to one and perform another average to determine the array major axis. We then compute θ_{maj} , which is the azimuth of the projection of the array major axis onto the horizontal plane. A similar process is applied to determine the array minor axis. The only difference is that the station minor axes are all projected onto the plane perpendicular to the array major axis. This assures that the array minor axis is perpendicular to the array major axis as is necessary for an ellipse.

Piñon Flat

Numerous seismic arrays have been deployed at Piñon Flat because of its characterization as a hard-rock site (e.g., Vernon *et al.*, 1991; Al-Shukri *et al.*, 1992; Owens *et al.*, 1991). Piñon Flat is a nearly planar erosional surface in the San Jacinto Mountains, California, located 12 km northeast of the San Jacinto fault system and 25 km southwest of the San Andreas fault system (Fig. 2). It has approximately 60 m of weathered material floored by plutonic rocks of granodiorite composition (Wyatt, 1982; Fletcher *et al.*, 1990).

We analyzed local event data from two seismic arrays

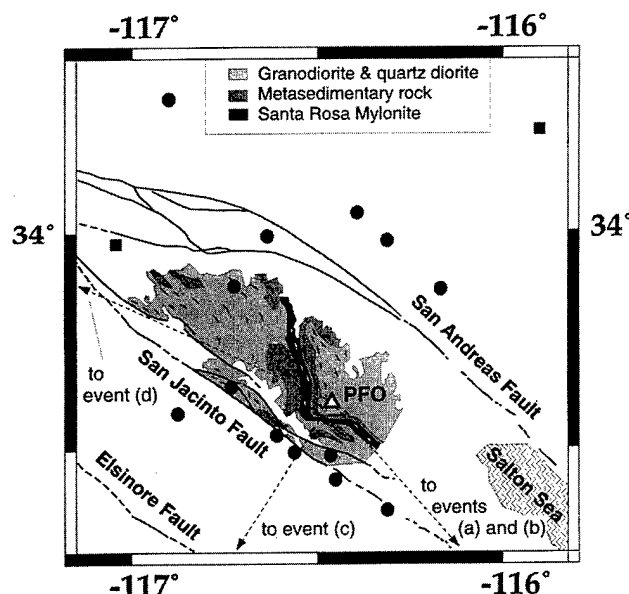


Figure 2. Map of area surrounding Piñon Flat. The geology is modified from Dibblee (1981). The locations for the events recorded by PFO-HF are plotted as filled circles, and those recorded by PFO-BB are plotted as filled squares. Events (a), (b), and (d) were recorded by PFO-HF and have θ_{loc} values of 139° (1), 139° (2), and 296° , respectively. Event (c) was recorded by PFO-BB and has $\theta_{loc} = 213^\circ$.

at Piñon Flat—a very small aperture high-frequency array (PFO-HF) and a larger aperture broadband array (PFO-BB). We focused on studying local events due to some intriguing observations that had been made earlier with data recorded with the PFO-HF array. Vernon *et al.* (1998) plotted the raw station particle displacements for the first *P* arrival in a three-dimensional display for one of the same events we analyzed here. They note that (1) the motions are strongly elliptical and (2) the major axes of the motions are skewed from the great circle path backazimuth.

The PFO-HF array consists of two borehole sensors and 58 surface sensors (Fig. 3). All stations were equipped with 2-Hz, three-component seismometers. The nominal sensor spacing is 7 m in the grid and 21 m along the arms (Owens *et al.*, 1991). This array operated between 18 April and 27 May 1990 and recorded triggered data at 250 samples/sec. We studied 16 events from PFO-HF, chosen for their coverage of arrival azimuths, signal-to-noise ratio, and data recovery. The event locations are plotted as filled circles in Figure 2, and source parameters are listed in Table 1. Note that no events were recorded by this array in a 90° gap to the east of the array.

We applied the processing described in the Method section to the surface sensors in the frequency bands 7 to 21 Hz (time length 0.428 sec) and 2 to 6 Hz (time length 1.5 sec). These bands were chosen to coincide with the change in signal behavior noted by both Vernon *et al.* (1998) and Wilson (1997) at approximately 8 Hz. The vertical-component beams [determined using a time-domain beamforming program called dbap (Harvey, 1994)] for these 16 events are plotted in Figure 4. This figure also shows the window over which the transforms were applied for the processing. The windows were chosen automatically by an algorithm that searched for the maximum coherence (equation 6). For most events, this window only overlaps with the first few cycles of the *P* wave, but for more emergent events, it sometimes is chosen at a later time. For this figure and the discussion that follows, we define three different angles: θ_{loc} is the backazimuth predicted by a great circle path from the source to the receiver, θ_p is the backazimuth measured by slowness analysis from the array, and θ_{maj} is the backazimuth measured from the average *P*-wave particle motion major axis.

We also independently processed data from the two borehole sensors in the 7- to 21-Hz frequency band. We determined estimates for the best-fit particle motion ellipses treating these data as a two-station array with a fixed slowness vector equal to that determined by the surface array for each event. Due to data problems, we did not process the events with $\theta_{loc} = 43^\circ, 177^\circ, 296^\circ$, and 321° for the borehole sensors.

We were not able to analyze data from the PFO-HF array in frequency bands lower than 2.0 Hz due to its small aperture and to its use of high-frequency sensors. The larger aperture of the PFO-BB array and the use of broadband sensors allowed us to study the wave-field behavior in a lower frequency band of 0.75 to 2.25 Hz (time length 4.0 sec).

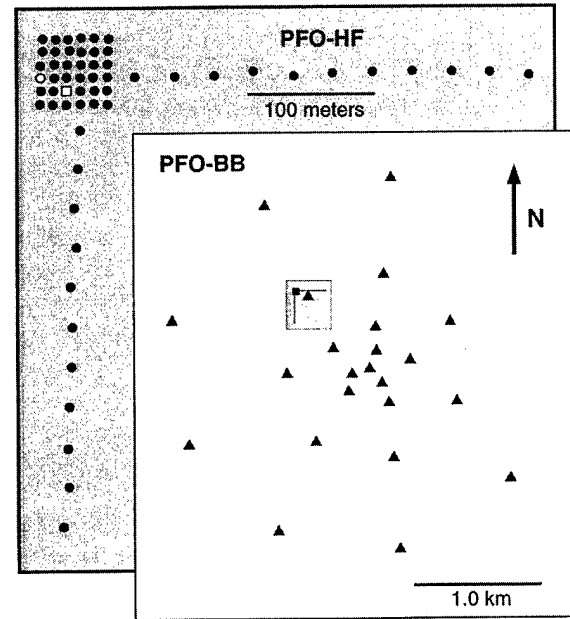


Figure 3. Station locations for the two arrays at Piñon Flat. All the symbols in the shaded region denote positions of surface sensors for PFO-HF. The open symbols also denote the approximate positions of the borehole sensors. The borehole sensor at the open circle is at 153 m depth, while the borehole sensor at the open square is at 274 m depth. The stations of PFO-BB are shown in the unshaded region as filled triangles. The relative position and scale of PFO-HF to PFO-BB is shown schematically in the unshaded region.

Table 1
Information about the Events Plotted in Figure 2

θ_{loc}	Date/Time	Array	#sta	Δ (degrees)	Depth (km)	m_i
7	1990117:12:52:42	PFO-HF	51	0.44	0.0	2.3
18	1990132:23:54:47	PFO-HF	55	0.39	4.8	2.5
43	1990109:20:24:58	PFO-HF	51	0.36	4.2	2.4
139(1)	1990137:17:02:50	PFO-HF	53	0.99	6.9	3.3
139(2)	1990137:19:32:50	PFO-HF	53	0.99	7.0	3.4
153	1990139:09:48:20	PFO-HF	53	0.27	12.5	2.1
177	1990136:01:14:16	PFO-HF	57	0.17	11.1	2.4
181	1990122:11:34:57	PFO-HF	53	0.12	7.0	2.1
217	1990132:19:42:47	PFO-HF	50	0.14	9.7	1.4
239	1990134:05:05:21	PFO-HF	57	0.14	11.8	2.6
266	1990138:12:05:43	PFO-HF	51	0.35	13.3	1.7
280	1990130:07:23:31	PFO-HF	57	0.23	15.7	2.1
296	1990110:03:24:59	PFO-HF	51	1.16	3.5	3.6
321	1990137:18:44:27	PFO-HF	53	0.35	17.3	1.4
333	1990124:02:23:22	PFO-HF	55	0.79	10.2	2.6
339	1990139:06:30:58	PFO-HF	53	0.41	5.0	2.7
36	1991057:17:08:30	PFO-BB	20	0.78	1.6	2.7
213	1991088:04:55:20	PFO-BB	23	1.80	6.0	3.0
308	1991077:03:58:26	PFO-BB	20	0.61	11.7	2.6

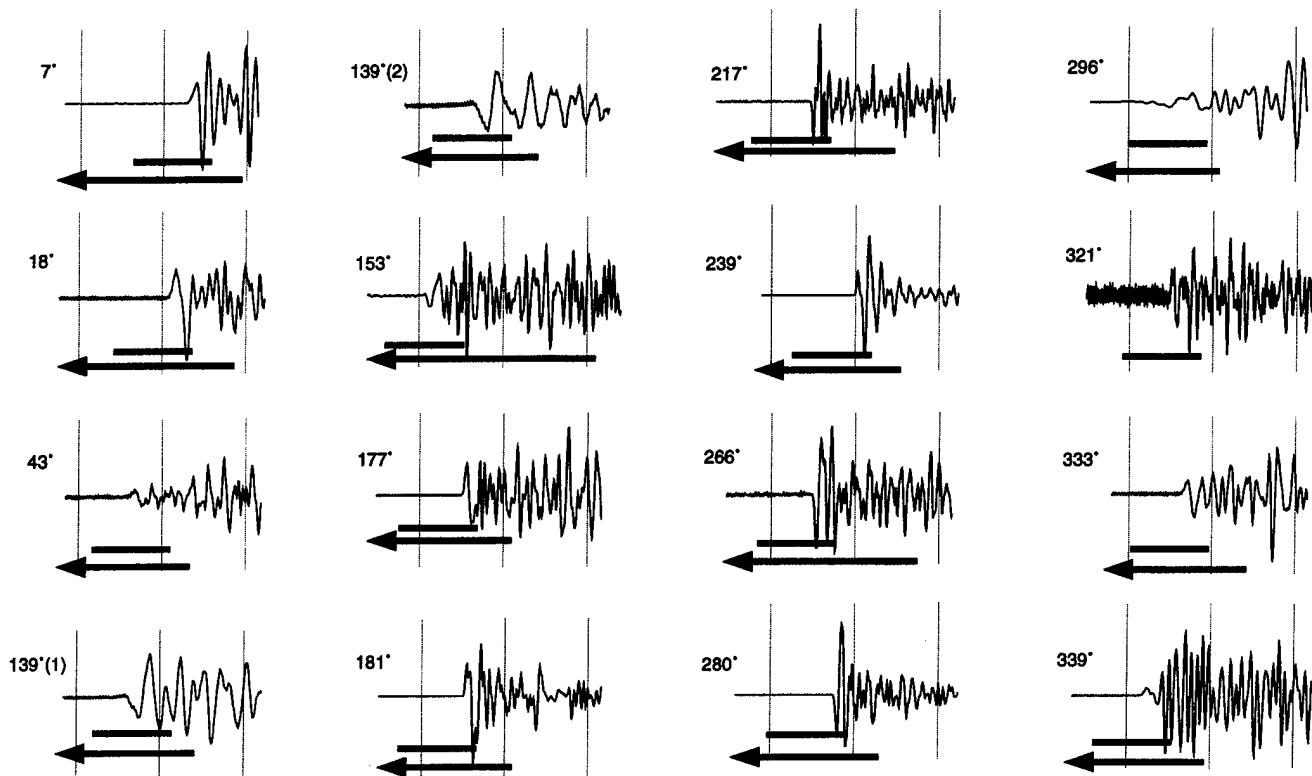


Figure 4. Vertical-component beams for events recorded by PFO-HF normalized to peak amplitude. The labels refer to each event's θ_{loc} , and the thin vertical lines denote 0.5-sec marks. The placement of the multi-wavelet transforms are shown as thick horizontal lines. The entire time window is shown for the 7- to 21-Hz transforms. The arrowheads specify that the remainder of the time windows for the 2- to 6-Hz transforms are earlier in time. Note that the time windows generally overlap with the signal for no more than three cycles of the transform's center frequency.

This array consisted of 28 three-component broadband Streckeisen STS-2 sensors arranged in five concentric circles named A through E (Al-Shukri *et al.*, 1992). The three sensors of the outer E-ring had a large influence over the beam pattern of this array, and the loss of any particular station caused severe distortions in the beam pattern. Because two different E-ring stations were down for two of the three events we studied from this array, we decided not to use any of the data from the outer E-ring. We also had to throw out two of the D-ring stations due to data problems. Thus, we ended up with the 23 stations shown in Figure 3. This reduced array had an effective aperture of 3 km. We present results for three events whose locations are shown as filled squares in Figure 2. Detailed source parameters are listed in Table 1. Unfortunately, these were the only local events that had good enough signal-to-noise ratios for processing in the frequency band of interest (0.75 to 2.25 Hz).

Observations

In Figure 5, we plot the array major axis skew ($\theta_{maj} - \theta_p$) versus the propagation direction, θ_p , and the location skew, ($\theta_{loc} - \theta_p$), versus θ_p for the three frequency bands. The hor-

izontal and vertical lines are error bars for θ_{maj} (vertical lines in Fig. 5a) and θ_p (all other lines). The error bars for θ_{maj} were determined by first finding a sphere around the array major axis that contained at least three of the five samples determined by averaging the station major axes for each multi-wavelet transform. The range of azimuths covered by the projection of the sphere onto the horizontal plane was then used as an error estimate. A similar process was used to determine the error for θ_p from the five slowness vector samples. This is similar to the methodology for quantifying slowness vector errors described by Bear and Pavlis (1997a), but the time-averaging required in equation (7) implies that the samples used to produce these error bars are not statistically independent. Note the systematic variations in the array major axis and location skews with propagation direction.

These regular variations can also be seen in individual station results. Figure 6 shows the major axis skews for individual stations from a subset of the events we examined. (The arms of the PFO-HF array are not shown so that the grid data can be seen more clearly.) We see the data show a consistent background skew that defines the array average with a superimposed smoothly varying pattern of deviations

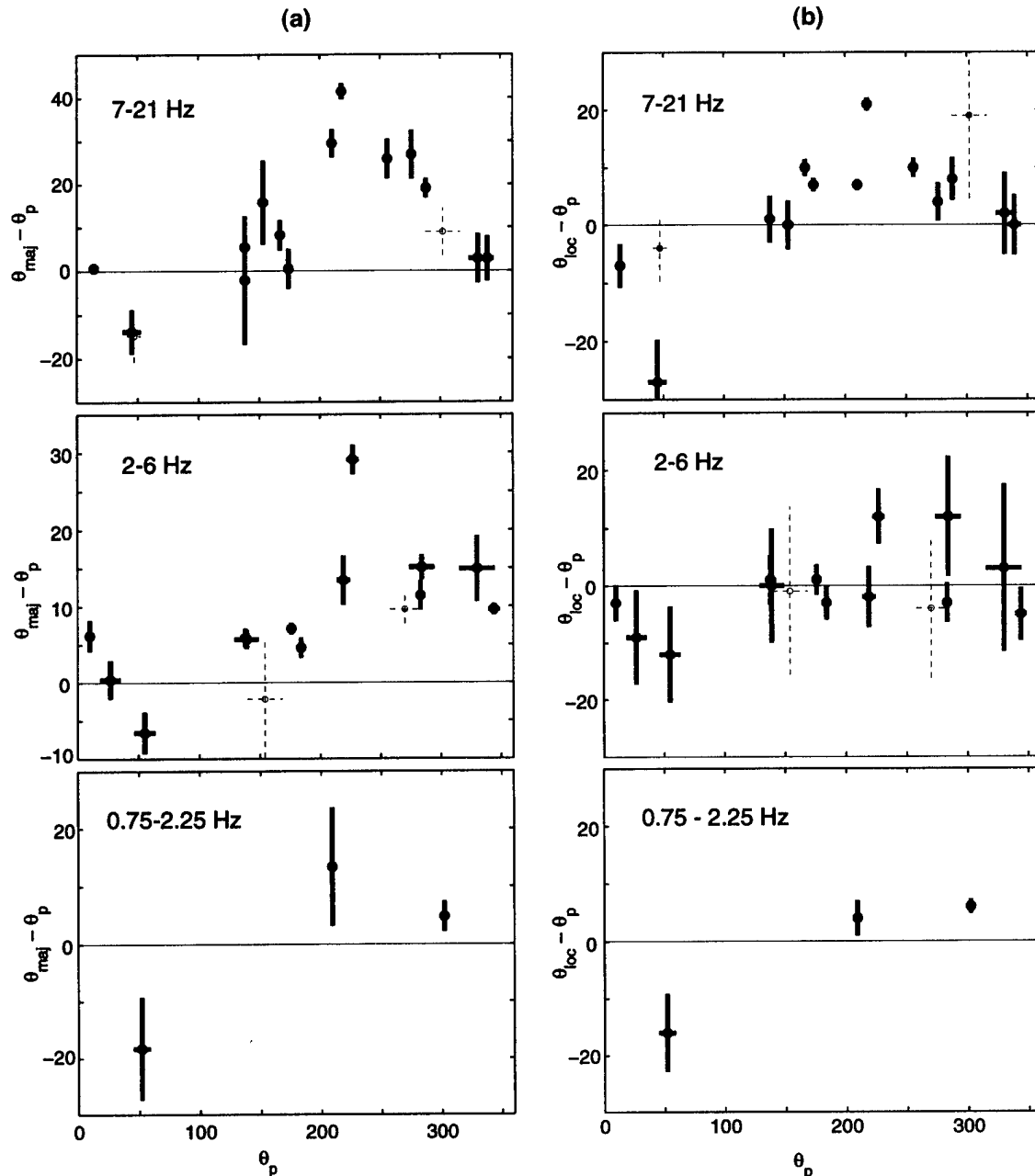


Figure 5. Plots of (a) $(\theta_{\text{maj}} - \theta_p)$ and (b) $(\theta_{\text{loc}} - \theta_p)$ versus θ_p with error bars. The open circles with dashed error bars are associated with events that have at least 20 stations with signal-to-noise ratios of 2:1 or less. Note the persistent pattern over all three frequency bands, particularly in (a).

within the array. The consistency, in fact, is even better for the PFO-HF data than it may appear at first glance. There are five stations—X0Y1, X3Y5, X3Y6, X4Y0, and X4Y2—that have consistently more negative skews than the other stations. We suspect that their behavior is due to problems with one of the three sensors of those stations. The individual station results for the PFO-BB data show a sense of skew consistent with the array average, but we do not see the smoothly varying trends that are apparent from the PFO-HF

data. This is an important observation as it indicates that there is an overall pattern that distorts particle motions for the entire PFO-BB array, but the smoothly varying patterns seen in the PFO-HF data occur at scale lengths smaller than the station spacing of the PFO-BB data.

The rectilinearity, e , for the three frequency bands is plotted in Figure 7 versus propagation azimuth θ_p . Note that the particle motion is markedly more elliptical in the 7- to 21-Hz band than in either of the other frequency bands. This

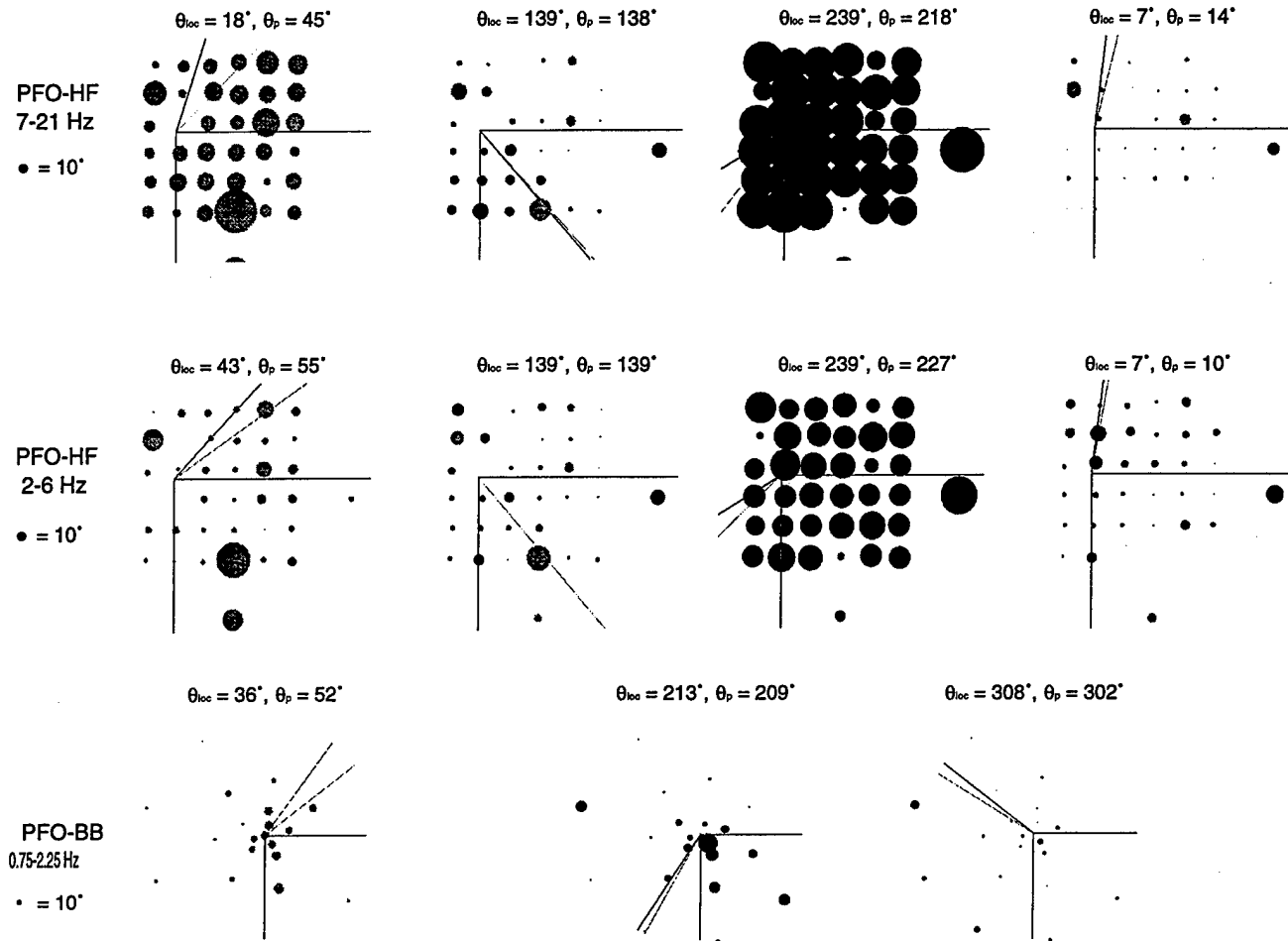


Figure 6. The absolute magnitudes of $(\theta_{maj,n} - \theta_p)$ for the raw station data. A sphere of radius $\|(\theta_{maj,n} - \theta_p)\|$ is plotted at the position of each station in the array. A black sphere is plotted if $(\theta_{maj,n} - \theta_p)$ is positive; a gray sphere, if $(\theta_{maj,n} - \theta_p)$ is negative. The scales for the radii are shown to the left. Only the grid of PFO-HF is shown due to space and resolution limitations. The axes point to the east and south, the solid lines point in the direction of θ_{loc} , and the dashed lines point in the direction of θ_p .

matches the difference in the behavior of the PFO-HF data above and below 8 Hz noted by Anderson (1993) and Vernon *et al.* (1998). In Figure 8, we plot the station minor axis tilts ($\phi_{min,n}$) for the PFO-HF data in the 7- to 21-Hz frequency band. We do not show the comparable figures for the other frequency bands because the orientations of the minor axes are not meaningful for signals that are linearly polarized. We note that there is a consistent pattern to the orientations of the minor axes across the array. This is remarkable considering there is no spatial smoothing in this processing scheme. The patterns in magnitude and sign for the minor axis orientations are very similar to those shown in Wilson and Pavlis (1999) for spectral amplitude variations.

We suggest that the elliptical particle motions of the signals in the 7- to 21-Hz band (see Fig. 7) is caused mainly by near-surface focusing and scattering effects. Further demonstration of this comes from the two borehole instruments, which are both well below the weathered layer (Fletcher *et*

al., 1990). We note that these signals are much more linearly polarized than those recorded at the surface. The rectilinearity values range from 0.8275 to 0.9894 and track well with the curve for the 2- to 6-Hz frequency band in Figure 7. On the other hand, the station major axis skews ($\phi_{maj,n} - \theta_p$) versus θ_p do not appear to be a strictly near-surface affect. In Figure 9, we plot the polarization properties for the two borehole stations along the averaged surface sensor values from Figure 5a. The strong agreement of these results indicates a more deep-seated source for the observed skews.

Discussion

Our analysis of the data from Piñon Flat revealed five significant observations.

1. The difference between the propagation azimuths and the

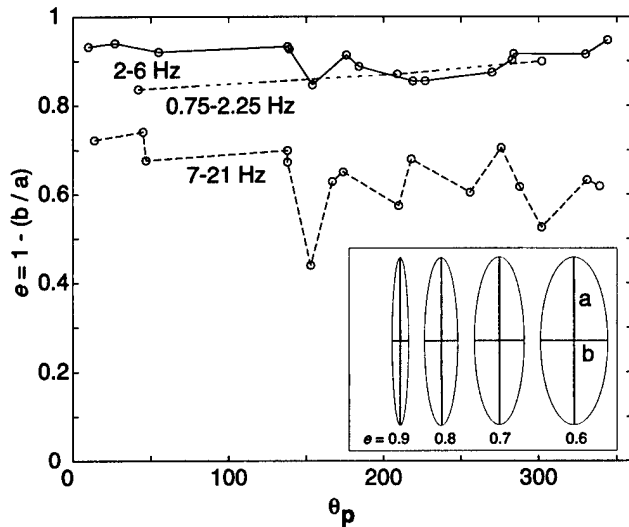


Figure 7. This figure shows the rectilinearity, e , versus the propagation azimuth θ_p for all the events in the three frequency bands.

- array major axis skews have a regular pattern that is consistent in all three frequency bands (Fig. 5a).
2. The difference between the propagation azimuths measured by array analysis and the location azimuths also have a regular pattern very similar to that of the array major axis skews. This pattern holds in all three frequency bands, though it is most pronounced in the 7- to 21-Hz band (Fig. 5b).
 3. The particle motions in the 2- to 6-Hz and 0.75- to 2.25-Hz frequency bands are nearly linear (Fig. 7).
 4. The particle motions in the 7- to 21-Hz frequency band for the surface sensors are much more elliptical (Fig. 7). The tilts in the minor axes from the vertical plane, though, are not random. There is a very distinct pattern in magnitude and sign across the array (Fig. 8).
 5. The particle motions for the borehole sensors in the 7 to 21-Hz frequency band are nearly linear, but the pattern of major axis skews is nearly indistinguishable from the surface sensors. (Fig. 9).

These observations are not what is expected for standard Earth models. Most theoretical seismology assumes the Earth is a horizontal stack of homogeneous, isotropic layers. But such a model cannot account for any of the observations listed above. Lin and Roecker (1996), using data from the PFO-BB experiment, measured differences between location and propagation azimuths for regional events analogous to the plots in Figure 5b. They modeled this behavior with a single dipping layer with moderate success. The difference between their study and this one, however, is that they did not investigate particle motion variation with respect to the propagation azimuths. They looked only at phase velocity variations from that expected from independently determined event locations. This is a significant point about our

results compared to most previous work—we are measuring P -wave particle motion relative to phase velocity vectors measured by the same array (e.g., observation 1). For even a complex stack of nonparallel dipping layers, conventional ray theory for isotropic media, which is based on a high-frequency limit (Aki and Richards, 1980, pp. 84–105), would predict no difference between the P -wave particle motion direction and the azimuth defined by the array slowness vector. This means that the values plotted in Figure 5a should be identically zero.

Because of the inadequacy of conventional layer models, we tried modeling the behavior of these data with a simple anisotropic medium. The model we use is one level of complexity above a dipping isotropic layer. That is, we treat the entire volume beneath the array as a dipping, transversely isotropic medium. Using the inversion method described by Bokelmann (1995a,b, 1996), the 7- to 21-Hz data were fit with a medium with the following properties: (1) the normal vector to the plane of symmetry dips 60° from the vertical and points 20° west of north, and (2) the medium is characterized by $\eta = 0.65$ and $\tau = 0.3$ [defined in Bokelmann (1995a)] The results are shown in relation to the data in Figure 9. (Results from inversion of the 2- to 6-Hz data were similar but are not shown for the sake of brevity.)

The transversely isotropic model yielded a variance reduction of 57% for the 7- to 21-Hz data. Although this is a significant improvement over a dipping layer model, it still has some serious inadequacies. This model fits the gross pattern of these data, but comparison to Figure 5b shows we are not completely fitting the data within our measured error bars. There are two explanations for this: (1) the error bars underestimate the real uncertainties in the particle motion major axes, or (2) the model is inadequate. Although the error estimates we obtain here have theoretical weaknesses (discussed earlier), we doubt they are drastically in error. Furthermore, the consistency of individual station particle motions (Fig. 6) argue against this. The implication is that the model we have determined is an oversimplification.

The anisotropy implied by this model is exceptionally strong. We compute that S -wave splitting would theoretically approach 30% in some directions for such a medium. This is much larger than reported S -wave splitting for the Anza region by Peacock *et al.* (1988). They observed S -wave splitting in this region of the order of 2% and less. We note, however, that very little of their data were from PFO. Furthermore, S -wave splitting measurements are based on an effect accumulated on the entire path from a source to a receiver, while what we measure is a local effect controlled primarily by the elastic properties of the earth directly under the array.

What is the source of the observed particle motion deviations we observe here, and does the anisotropic model we determined have any relationship to reality? The answer is ambiguous and points out a fundamental weakness in our existing theoretical models for wave propagation in the real Earth. That is, the earth is unquestionably a heterogeneous

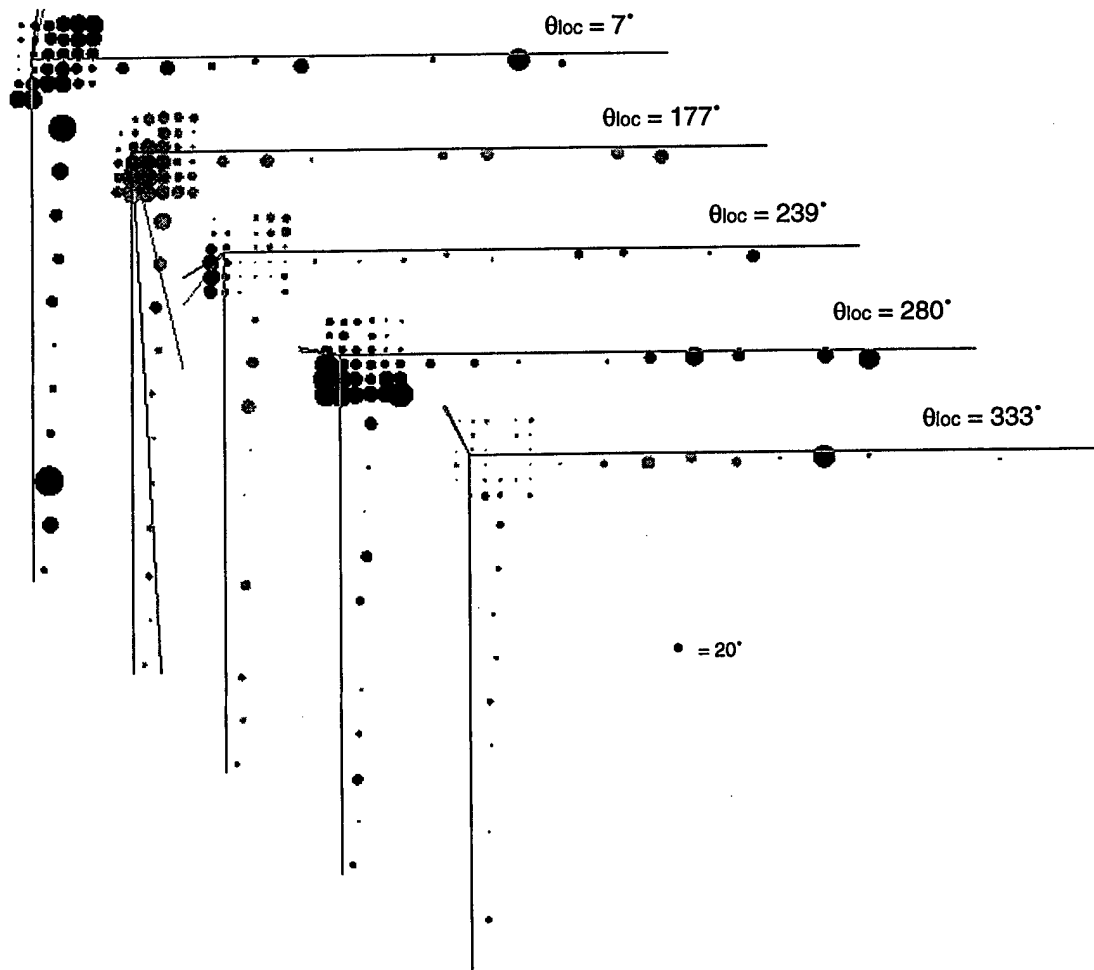


Figure 8. PFO-HF station data for $\phi_{\min,n}$ in the 7- to 21-Hz frequency band for five location azimuths. A sphere of radius $\|\phi_{\min,n}\|$ is plotted at the position of each station in the array. A scale is shown in the lower right corner. A black sphere is plotted if $\phi_{\min,n}$ is positive; a gray sphere, if $\phi_{\min,n}$ is negative (see Fig. 1 for definition of positive and negative). Note the regular patterns that exist across the array.

media with seismic properties that fluctuate over a huge range of scale lengths ranging from the grain size of a given rock (~ 1 mm) to thousands of kilometers. Seismic wavelengths range from the order of a few meters to thousands of kilometers. A poorly understood, fundamental problem is how inhomogeneity of a given scale is expressed observationally. We all understand that a rock with a preferred orientation of minerals like a schist is intrinsically anisotropic because the fabric that defines the anisotropy is at a scale far below the smallest observed seismic waves. Fabric at intermediate scales, however, can induce anisotropic effects that are more difficult to sort out. It has been known for more than three decades [based on landmark work by Backus (1962)] that layered sedimentary rocks are anisotropic at wavelengths that are large compared to the scale of the layering. Thus, a layered model over some range of wavelengths must pass from conventional, high-frequency limit behavior to a behavior more analogous to that of an intrin-

cally anisotropic material. Theoretical progress has been made toward relating anisotropic effects of different scales (see, e.g., Werner and Shapiro, 1998), but the problem remains poorly understood at best. [For a good fundamental physical understanding of this issue, the reader is referred to Chapter 1 of the book by Helbig (1994).] The issue this raises for this article is that to understand our results, we need to review what we understand about heterogeneity of the Earth within the vicinity of Piñon Flat at a range of relative scales. We organize this discussion in order of increasing scale length.

Bedrock at Piñon Flat is a granodiorite with grain sizes of the order of a few millimeters and no appreciable fabric (i.e., no intrinsic anisotropy). At the scale of 1 to 100 m, however, the situation is drastically different. The near surface at Piñon Flat is an ancient weathering profile that has altered the original granodiorite to a depth of at least 60 m (Fletcher *et al.*, 1990). The geologic processes that created

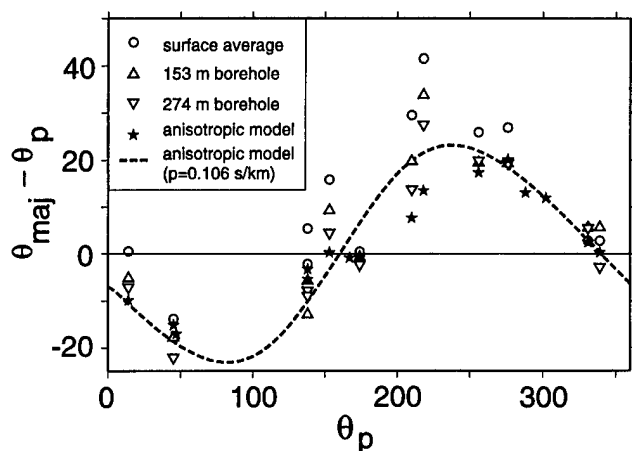


Figure 9. This figure plots deviations of the particle motion azimuths determined from P waves from that expected for a pure longitudinal wave (major axis skews) as a function of the measured phase velocity azimuth. Data from both the surface array average (circles) are shown here in relation to single-station particle motions measured in the two boreholes in the PFO-HF experiment (upright and inverted triangles). All measured results plotted are for the 7- to 21-Hz band. The star symbols and the dashed lines are predictions of major axis skews for the anisotropic model discussed in the text. The stars show the actual predicted value that takes into account the measured phase velocity across the array. The dashed line shows the prediction for the anisotropic model at a constant ray parameter, but varying azimuth. It illustrates the sinusoidal pattern that characterizes this dipping, transversely isotropic model.

this weathering layer are unique to granitic rocks (Ollier, 1969) and produce an extremely heterogeneous medium. Granitic rocks are weathered preferentially along preexisting joints because the primary agent of weathering is chemical attack by water. Water flow is focused on the enhanced permeability zone around a joint in the rock leading to concentrations of weathering along these surfaces. The result is largely unaltered corestones surrounded by rings of progressively more altered material. Preferred orientation of joints, which is almost universally observed, will lead to a characteristic fabric of the near surface. We suggest this may lead to an effectively anisotropic media at the higher frequencies we are observing. At these frequencies, the wavelength of the P waves we record are of the same order of magnitude as the entire weathered layer. We suggest this leads to bulk elastic properties of the near surface that are a major factor in producing all the particle motion deviations we observe, particularly in the upper frequency band of 7 to 21 Hz. At the same time, the large variations in properties within the weathered layer undoubtedly strongly scatter the incident wave field as argued by Vernon *et al.* (1998). At the frequencies we are working with here, the scattering and induced anisotropy could well be thought of as essentially

the same phenomenon—distortion of observed ground motion induced by near-surface heterogeneity.

The weathered layer, however, is probably not the whole story. It does not fully explain the borehole data, and it is hard to reconcile with the data from the 0.75- to 2.25-Hz band. How can we obtain nearly the same major axis skews in the borehole data when they are located well below the weathered layer? One explanation is that although at the lowest frequencies the boreholes are only a fraction of a wavelength below the surface, at the highest frequency observed here (21 Hz), the deepest borehole is only about 1 wavelength below the surface. Hence, it is conceivable that the borehole data are impacted by anisotropic properties of the near surface even though these sensors are located well below the weathered layer. This cannot be addressed, however, without more extensive modeling with synthetic seismograms that can properly model anisotropic media at finite wavelengths. This is beyond the scope of this article.

The next level of heterogeneity is structure at the scale of a geologic map. A rock unit called the Santa Rosa mylonite wraps around Piñon Flat (Fig. 2). Parcel (1981) argues the Santa Rosa mylonite was formed by right-lateral, horizontal transport and that the bend in this rock unit at Piñon Flat was induced by a deflection of the shear zone caused by interaction with the plutonic body that floors Piñon Flat. The Santa Rosa mylonite is a very strongly anisotropic rock. Kern and Wenk (1990) found these rocks to be transversely isotropic with a 5 to 19% anisotropy at surface pressures decreasing to 5 to 12% anisotropy at 600 MPa. Unfortunately, the sense of the anisotropy is the opposite of that determined from our inversion of these data. That is, the model we determined is a transversely isotropic medium dipping to the northwest with the fast axis perpendicular to the plane of symmetry. If we had pure Santa Rosa mylonite with its foliation plane dipping in the same northwesterly direction, the normal to the foliation plane would be in the slow axis, not the fast axis. What this means remains ambiguous because the subsurface geometry of the Santa Rosa mylonite beneath Piñon Flats is not known. Dibblee's (1981) maps (see Fig. 2) show the mylonites wrapping around the west and south side of Piñon Flats. Measured foliations dip eastward swinging to the north with angles ranging from 30 to 60°, suggesting the mylonites wrap around and underneath the granodiorite from both the south and west sides. How this complex geometry would map into our data is not at all clear.

The overall conclusion we reach is that the total effect we observe is probably the superposition of at least three processes: (1) near-surface anisotropy introduced by preferential weathering along joints in the granites, (2) a larger scale fabric induced by the Santa Rosa mylonite, and (3) near-surface scattering. The last process is probably most important in the highest frequency band (7 to 21 Hz) and probably contributes to the exceptionally large major axis skews seen from azimuths near 240°. There are several fundamental ambiguities that prevent us from fully sorting this

out with available data. First, we have only a loose idea of the subsurface geometry of the Santa Rosa mylonite. The best guess of the actual geometry (Fig. 2) is that the mylonite wraps around the area where these data were collected. We know of no existing program capable of modeling such a complex medium even if we had better constraints on its geometry. Second, our concepts of the interactions of the wave field with the near surface at this site are largely conjecture. The near-surface material at this site has heterogeneity on a wide range of scales from at least 0.001 to 100 m. We have considered attempting to model the near-surface material, but this is fraught with ambiguity for two reasons: (1) limited knowledge of subsurface structure and (2) fundamental questions about the validity of existing computer codes to properly model such a wildly heterogeneous medium.

The phenomenon we observe here is only observable with a three-component array. A question this article leaves hanging is: How common is this type of departure from the standard model of wave propagation in a layered Earth? We suggest that the methodologies developed here, when applied to three-component array data, can provide fundamental new observations on anisotropy within the Earth. Broadband data from arrays of varying scale have the potential to provide a new way to measure crustal anisotropy through application of the techniques described in this article.

Acknowledgments

We are grateful to all the individuals responsible for fielding the Piñon Flat array experiments and assembling the results into a workable data set. Funding for this work was provided by an AASERT grant from the Air Force Office of Scientific Research (Number F49620-95-1-0366) and by the IRIS Joint Seismic Program.

References

- Aki, K. and P. G. Richards (1980). *Quantitative Seismology Theory and Methods*, Vol. 1, Freeman, San Francisco, 557 pp.
- Al-Shukri, H., T. Owens, G. Pavlis, S. Roecker, F. Vernon, and G. Wagner (1992). Data report for the 1991 Piñon Flat broadband array experiment, IRIS/PASSCAL Data Report.
- Anderson, P. N. (1993). Animated visualization techniques for three-component seismic array data, *Ph.D. Thesis*, Indiana University.
- Backus, G. E. (1962). Long-wave elastic anisotropy produced by horizontal layering, *J. Geophys. Res.* **67**, 4427–4440.
- Bear, L. K. and G. L. Pavlis (1997a). Estimation of slowness vectors and their uncertainties using multi-wavelet seismic array processing, *Bull. Seism. Soc. Am.* **87**, 755–769.
- Bear, L. K. and G. L. Pavlis (1997b). High-resolution multiwavelet seismic array processing, *EOS* **78**, S216.
- Bear, L. K. and G. L. Pavlis (1999). Multi-channel estimation of time residuals from broadband seismic data using multi-wavelets, *Bull. Seism. Soc. Am.* **89**, 681–692.
- Bokelmann, G. H. R. (1995a). P-wave array polarization analysis and effective anisotropy of the brittle crust, *Geophys. J. Int.* **120**, 145–162.
- Bokelmann, G. H. R. (1995b). Azimuth and slowness deviations from the GERESS regional array, *Bull. Seism. Soc. Am.* **85**, 1456–1463.
- Bokelmann, G. H. R. (1996). Local effects on P-wave polarization, in "Seismic Anisotropy, E. Fjaer, R. M. Holt, and J. S. Rathore (Editors), Society of Exploration Geophysicists, 763 pp.
- Chave, A. D., D. J. Thomson, and M. E. Ander (1987). On the robust estimation of power spectra, coherences, and transfer functions, *J. Geophys. Res.* **92**, 633–648.
- Dibblee, T. W. (1981). Geology of the San Jacinto Mountains and adjacent areas, in *Geology of the San Jacinto Mountains: Annual Field Trip Guidebook No. 9*, A. R. Brown and R. W. Ruff (Editors), South Coast Geological Society, Santa Rosa, California, 1–47.
- Fletcher, J. B., T. Fumal, H. Liu, and L. C. Carroll (1990). Near-surface velocities and attenuation at two boreholes near Anza, California, from logging data, *Bull. Seism. Soc. Am.* **80**, 807–831.
- Harvey, D. (1994). *dbap (a public domain computer program)*, Joint Seismic Program Center, University of Colorado, Boulder.
- Helbig, K. (1994). *Foundations of Anisotropy for Exploration Seismics*, Oxford, New York, 486 pp.
- Husebye, E. S. and B. O. Ruud (1989). Array seismology; past, present, and future developments, in *Observatory Seismology*, J. J. Lithiser (Editor), University of California Press, Berkeley, 123–153.
- Kern, H. and H.-R. Wenk (1990). Fabric related anisotropy and shear wave splitting in rocks from the Santa Rosa mylonite zone, California, *J. Geophys. Res.* **95**, 11213–11223.
- Kvaerna, T. and D. J. Doornbos (1986). An integrated approach to slowness analysis with arrays and three-component stations, *NORSAR 2-85/86*.
- Lilly, J. M. and J. Park (1995). Multiwavelet spectral and polarization analyses of seismic records, *Geophys. J. Int.* **122**, 1001–1021.
- Lin, C. H. and S. W. Roecker (1996). P-wave backazimuth anomalies observed by a small-aperture seismic array at Piñon Flat, southern California: implications for structure and source location, *Bull. Seism. Soc. Am.* **86**, 470–476.
- Ollier, C. (1969). *Weathering*, Elsevier, New York.
- Owens, T., P. N. Anderson, and D. E. McNamara (1991). Data report for the 1990 Piñon Flat grid experiment: an IRIS Eurasian Studies Program passive source experiment, IRIS/PASSCAL Data Report #91-002.
- Parcel, R. F. (1981). Structure and petrology of the Santa Rosa shear zone in the Pinyon Flat area Riverside County, California, in *Geology of the San Jacinto Mountains: Annual Field Trip Guidebook No. 9*, A. R. Brown and R. W. Ruff (Editors), South Coast Geological Society, Santa Rosa, California, 139–150.
- Pavlis, G. L. and H. Mahdi (1996). Surface wave propagation in central Asia: observations of scattering and multipathing with the Kyrgyzstan broadband array, *J. Geophys. Res.* **101**, 8437–8455.
- Peacock, S., S. Crampin, D. C. Booth, J. B. Fletcher (1988). Shear wave splitting in the Anza seismic gap, southern California: temporal variations as possible precursors, *J. Geophys. Res.* **93**, 3339–3356.
- Vernon, F. L., J. Fletcher, L. Carroll, A. Chave, and E. Sambera (1991). Coherence of seismic body waves from local events as measured by a small-aperture array, *J. Geophys. Res.* **96**, 11981–11996.
- Vernon, F. L., G. L. Pavlis, T. J. Owens, D. E. McNamara, and P. E. Anderson (1998). Near-surface scattering effects observed with a high-frequency array at Pinyon Flats, California, *Bull. Seism. Soc. Am.* **88**, 1548–1560.
- Vidale, J. E. (1986). Complex polarization analysis of particle motion, *Bull. Seism. Soc. Am.* **76**, 1393–1405.
- Wagner, G. S. (1994). Analysis of multidimensional JSP data, *IRIS Newslett.* (summer), 13–16.
- Wagner, G. S. (1997). Local wave propagation in the San Jacinto Fault Zone, southern California: observations from a three-component seismic array, *J. Geophys. Res.* **102**, 8285–8311.
- Wagner, G. S. and T. J. Owens (1993). Broadband bearing-time records of three-component seismic array data and their application to the study of local earthquake coda, *Geophys. Res. Lett.* **20**, 1823–1826.
- Wagner, G. S. and T. J. Owens (1995). Broadband eigen-analysis for three-component seismic array data, *IEEE Trans. Signal Process.* **43**, 1738–1741.

- Wagner, G. S. and T. J. Owens (1996). Signal detection using multi-channel seismic data, *Bull. Seism. Soc. Am.* **86**, 221–231.
- Werner, U. and S. A. Shapiro (1998). Intrinsic anisotropy and thin multi-layering—two anisotropy effects combined, *Geophys. J. Int.* **132**, 363–373.
- Wilson, D. C. (1997). Near-surface site effects in crystalline bedrock: azimuthal dependence, scale lengths of spectral variation, and change in spectral character with depth, *Master's Thesis*, Indiana University, 27 pp.
- Wilson, D. C. and G. L. Pavlis (1999). Near-surface site effects in crystalline bedrock: a comprehensive analysis of spectral amplitudes determined from a dense, three-component seismic array, *Earth Interactions*, in press.
- Wyatt, F. (1982). Displacement of surface monuments: horizontal motion, *J. Geophys. Res.* **87**, 979–989.

Department of Geological Sciences
Indiana University
1001 East 10th Street
Bloomington, Indiana 47405
(L. K. B., G. L. P.)

Institute of Geophysics
Ruhr-University Bochum
D-44780 Bochum, Germany
(G. H. R. B.)

Manuscript received 11 July 1998.

Multi-channel Estimation of Time Residuals from Broadband Seismic Data Using Multi-wavelets

by Lorie K. Bear* and Gary L. Pavlis

Abstract We describe a new multi-channel procedure for estimating arrival time residuals from seismic array data. It incorporates aspects of three traditional array processing methods: frequency-domain beamforming, time-domain beamforming, and principal-component analysis. We start by applying the multi-wavelet transform to the data, which yields a suite of narrow-band seismograms. We use the multi-wavelet transform, instead of the windowed Fourier transform, for superior control over both the time and frequency resolution. We employ a beamforming procedure that uses principal component analysis on the transformed, time-aligned data. The values in the principal component vector and value pair are used to calculate a measure of coherence analogous to semblance. A measure of the misfit of the data to our plane-wave model is contained in the phase differences in the principal component vector. The phase differences can be converted directly to time residuals, but they are only resolvable to a fraction of the analyzing wavelength. Hence, our method is a staged process that moves from lower to higher analyzing frequency bands. We present two data examples that illustrate the wide range of spatial and temporal scales over which this approach can be applied. First, we determine time residuals for the deep-focus Bolivian earthquake of 1994 for a set of broadband stations spread over most of southern California. The time residuals had a range of 2 sec, and after their removal, we were able to stack the data to over 1.0 Hz. Second, we study a local event recorded by high-frequency sensors at an array in Turkmenistan with an aperture of less than 1 km. We found that the time residuals only had a range of 0.02 sec, but by removing them, we significantly improved the stack of the data for the arrival's dominant frequency.

Introduction

The past decade has witnessed a revolution in seismic instrumentation brought on by the IRIS-PASSCAL program. Two features of this facility are proving especially important: (1) the portable broadband seismometer and (2) the sheer number of matched instruments that can be deployed simultaneously. The first has dramatically improved the range of temporal frequencies we can work with. The second expands our capabilities in recording spatial frequencies. That is, by deploying many sensors over a mix of scale lengths, we can investigate the wave field at different spatial scales. As a result, a large suite of data is now accumulating that provides unprecedented, precision measurements of the seismic wave field that are rapidly changing our understanding of wave propagation in the Earth. We assert that the speed of change brought on by this revolution in data quality has led to a gap in processing procedures that can

take full advantage of these new data. This article and a companion article (Bear *et al.*, 1999, this issue) investigate some new ways to exploit these features of modern experimental data. Both focus on coherent processing of seismic array data to solve two different, but fundamentally inter-related problems.

Seismic arrays have a long history in nuclear monitoring (e.g., Husebye and Ruud, 1989). Originally, arrays were viewed as three or more matched instruments deployed in a regular pattern with a centralized recording and processing system. In the modern era, however, a large class of recent experiments are arrays in a less restricted sense. We define an array as any group of stations with reliable timing whose signals can be combined by a coherent processing method based on some assumed model of wave propagation. With this definition, what constitutes an array becomes frequency dependent. At one end, the global seismic network is an array for free oscillations. That is, by definition, the whole Earth is only a few wavelengths in size for free oscillation,

*Present address: Exxon Production Research, P.O. Box 21879, Houston, Texas.

and the signals are coherent in the sense that they can be modeled very accurately from existing global Earth models (e.g., Lay and Wallace, 1995, pp. 154 to 171). At the other end of the seismic spectrum are high frequencies of the 1- to 100-Hz band that overlap the observational data range of exploration seismic methods. Between these end members are the traditional short-period and long-period recording bands that were the focus of seismology before the digital age. Modern instruments are called "broadband" because they cover all or a large portion of the classical short-period and long-period bands. The traditional short-period records spanned the range of approximately 0.5 to 2 Hz, and long-period records spanned the range of approximately 0.02 to 0.08 Hz. Both are less than one decade in frequency. In contrast, modern instruments commonly record three or more decades in frequency with "broadband" meaning approximately 0.01 to 10 Hz to most people.

A key element of broadband is the range of scales implicit in broadband data. One cycle of a signal at 10 Hz is four orders of magnitude shorter than one cycle at 0.01 Hz. The same is true for wavelength. Given that four orders of magnitude is about the difference in size between your entire body and the thickness of a single strand of your hair, it is clear that processing methods applied to modern broadband seismic data need to adapt their scale to the frequency band being analyzed. The approach described in this article and its sibling (Bear *et al.*, 1999) does this by utilizing a new technique, developed by Lilly and Park (1995), we refer to as the multi-wavelet transform (Bear and Pavlis, 1997). The multi-wavelets are a set of special functions derived from an optimization condition that makes the functions as concentrated in time and frequency as possible. They are closely related to the prolate spheroidal functions (Slepian, 1983) that are the foundation of the multi-taper spectral methods of Thomson (1982). The approach, however, is more akin to the newly emerging field of wavelet transforms [see, for example, Strang and Nguyen (1996) or Kumar and Fouloula-Georgiou (1994)] that build on the pioneering work by Daubechies (1992). The result is something that should be viewed as somewhat of a hybrid between conventional Fourier methods and wavelet transforms.

The methods described in this article and its sibling (Bear *et al.*, 1999) can be viewed as a type of multi-channel cross-correlation technique that exploits the wide bandwidth of modern seismic arrays. In this article, we correlate corresponding channels across the aperture of a given seismic array to produce a set of travel-time differences from a reference model (in our case, a simple plane-wave time shift). In the companion article, we take this one step further to analyze particle motions from three-component seismic arrays. That is, we can do time shifts using methods in this article to more precisely align data from multiple stations, and then we can use a similar approach to determine average polarization properties from the entire array and individual station polarizations relative to the array average.

The essence of the approach introduced in this article is

simple. We start by filtering to a low enough frequency band that the time residuals are small with respect to the center frequency of the band. We apply beamforming in the complex, multiwavelet transform domain to find a best-fit plane wave. The misfits to the plane-wave model appear as complex phase shifts in the data. The phase shifts are converted into time residuals and added to the plane-wave time delays. We repeat this process through a series of staged complex transform filters that refine the residuals using successively higher frequency bands. The approach can be thought of as a hybrid of the time-domain and spectral methods of correlation. It is time domain like in its use of explicit time shifts to align data for stacking. At the same time, it is frequency domain like through the use of phase shifts computed from a series of complex numbers. This has a potential advantage over traditional methods in three ways: (1) precise control of time-frequency resolution trade-offs in different frequency bands are possible through the multi-wavelet functions, (2) the procedures can be made more robust through redundancy inherent in the multi-wavelet transform (Bear and Pavlis, 1997), and (3) variations in signal-to-noise ratio in different bands can, in principle, be handled.

The ideas developed in this article are important because determination of relative arrival times by coherent signal processing is a ubiquitous problem in modern seismology. In the days of analog seismograms, picking was a manual process. Today, timing of phases is commonly done by either manual picking or some form of cross-correlation. This article presents an alternative approach that can, we suspect, make more effective use of modern broadband data. Although this method is analogous to cross-correlation, the net result is fundamentally different from any existing procedure we are aware of because it is truly a multi-channel method. VanDecar and Crosson (1990) developed a method they refer to as a multi-channel cross-correlation method, but there is a fundamental difference between what we describe here and their method. VanDecar and Crosson's (1990) method looks at correlations between all pairs of related channels and then determines a best-fitting set of time delays for the whole array using a least-squares procedure. Our method is more direct because we work only on the original waveforms and process them directly to produce a set of optimal time shifts. Furthermore, although we have only limited experience with the algorithm to date, we suspect it will prove more robust than cross-correlation-based methods like that of VanDecar and Crosson (1990) because of our extensive use of robust M-estimators as an intrinsic part of the methodology and because we consider all the data simultaneously. We argue that this robustness may make our approach well suited to automated retrieval of teleseismic *P*-wave residual—the fundamental data of seismic tomography. This is an important, potential cost-saving benefit to experimental programs given the now routine application of *P*-wave tomography as a tool for study of crust and upper mantle structure.

We demonstrate the use of this approach on arrays of

two dramatically different scales. First, we apply the methodology to seismograms from the great Bolivian earthquake of 1994 recorded by a suite of broadband stations spread throughout southern California. This array has an aperture of over 300 km. Without the static corrections we determined, these data will not stack at frequencies over 0.2 Hz, but with the static corrections, we obtain coherent stacks to frequencies over 2 Hz. Second, we apply the method to high-frequency (1 to 100 Hz) data recorded by a three-component array with an aperture under 1 km. Here we found statics improved the array performance by allow the stacking of higher frequencies but much less dramatically than for the southern California example. We found that even with statics, the array performance fell off dramatically at frequencies over 35 Hz. We suggest this may reflect strong near-surface scattering into high-frequency surface-wave modes and strongly variable, localized site effects.

Method with Single-Component Data

Beamforming

To put what we have done in perspective, it is useful to review the ideas used in standard frequency-domain beamforming with an array of single-component stations. Conventional frequency-domain beamforming (e.g., Kvaerna and Doornbos, 1986) uses windowed Fourier transforms to form the spectral covariance matrix $\mathbf{R}(f, t)$ with entries

$$R_{mn}(f, t) = \mathcal{F}[s_n](f, t) \cdot \mathcal{F}[s_m]^\dagger(f, t). \quad (1)$$

$\mathcal{F}[s_n](f, t)$ is the windowed Fourier transform for signal s recorded at station n at frequency f and time t , and † denotes the complex conjugate transpose. The plane-wave time delay for an arrival with slowness vector $\mathbf{p} = (p_e, p_s)$ and station n is $\tau(\mathbf{p}, n) = -\mathbf{p} \cdot \mathbf{x}_n$, where $\mathbf{x}_n$ is the vector position of the station with respect to a local Cartesian system. Thus, the power for a slant stack of the measured signals assuming slowness vector $\mathbf{p}$ is

$$P_{\text{beam}}(f, t, \mathbf{p}) = \mathbf{e}^\dagger \mathbf{R}(f, t) \mathbf{e},$$

where the steering vector is

$$\mathbf{e} = \frac{1}{N} [e^{2\pi i f \tau(\mathbf{p}, 1)} \dots e^{2\pi i f \tau(\mathbf{p}, N)}]. \quad (2)$$

The use of a simple phase-shift for the steering vector means that we assume the signal behaves as a sinusoid over the times involved. With a time window that is long compared to the dominant period of a signal, this can be inaccurate because the phase shifts may involve many cycles of the dominant signal. Consequently, a more accurate algorithm is to apply the transform to the time-shifted data such that the covariance matrix is a function of the slowness vector $\mathbf{p}$ and

$$R_{mn}(f, t, \mathbf{p}) = \mathcal{F}[s_n][f, t + \tau(\mathbf{p}, n)] \cdot \mathcal{F}[s_m]^\dagger[f, t + \tau(\mathbf{p}, m)]. \quad (3)$$

The power of the beam is then

$$P_{\text{beam}}(f, t, \mathbf{p}) = \mathbf{d}^\dagger \mathbf{R}(f, t, \mathbf{p}) \mathbf{d}, \quad \text{where } d_n \equiv \frac{1}{N}. \quad (4)$$

In this case, the steering vector is not needed because $\mathbf{R}$ is computed by moving the window in time rather than depending on phase shifts. Clearly, this method is more accurate, particularly when the phase shifts are large, but is more expensive to compute.

Multi-wavelet Transforms

For classical beamforming, the use of the windowed Fourier transform is standard, but a key concept of this article is that there is a better choice. We will be considering our data over a wide range of frequency bands and prefer to use transforms where the time lengths of the kernel functions match the scale of the frequencies to be studied. This scaling could be accomplished with Fourier transforms by scaling the time length of the windowing taper to match the frequencies to be studied. Instead, we choose to use the complex multi-wavelet transforms (Lilly and Park, 1995; Bear and Pavlis, 1997) that are genetically related to the multitaper spectral estimation method (Thomson, 1982).

The real and imaginary parts of the integration kernels for the multi-wavelet transforms are specifically designed as real, discrete, finite time series w_m that have their energy concentrated within a frequency band defined by a center frequency f_c and a bandwidth $2f_w$, where $f_w \leq f_c$. As can be seen in Figure 1, these functions occur in even and odd pairs, where each pair emphasizes a different portion of the time and frequency windows. The multi-wavelets are defined as $\mathbf{w}_f^{(j)} = \{w_{f,e}^{(j)} + i w_{f,o}^{(j)}\}$, where f is the center frequency, $w_{f,e}^{(j)}$ is the j th even wavelet, and $w_{f,o}^{(j)}$ is the j th odd wavelet. Each multi-wavelet transform can then be written

$$\begin{aligned} \mathcal{W}^{(j)}[s](f, t) = & \int_{(t-T/2)}^{(t+T/2)} s(\xi) w_{f,e}^{(j)}(\xi - t) d\xi \\ & + i \int_{(t-T/2)}^{(t+T/2)} s(\xi) w_{f,o}^{(j)}(\xi - t) d\xi. \end{aligned} \quad (5)$$

That is, the multi-wavelet transformed data, $\mathcal{W}^{(j)}[s](f, t)$, are produced by convolution with the complex pair of basis functions, $\mathbf{w}_f^{(j)}$.

By using the multi-wavelet transforms in place of the Fourier transform, we can define the elements of the covariance matrix for the j th wavelet to be

$$R_{mn}^{(j)}(f, t, \mathbf{p}) = \mathcal{W}^{(j)}[s_n][f, t + \tau(\mathbf{p}, n)] \cdot \mathcal{W}^{(j)*}[s_m]^\dagger[f, t + \tau(\mathbf{p}, m)]. \quad (6)$$

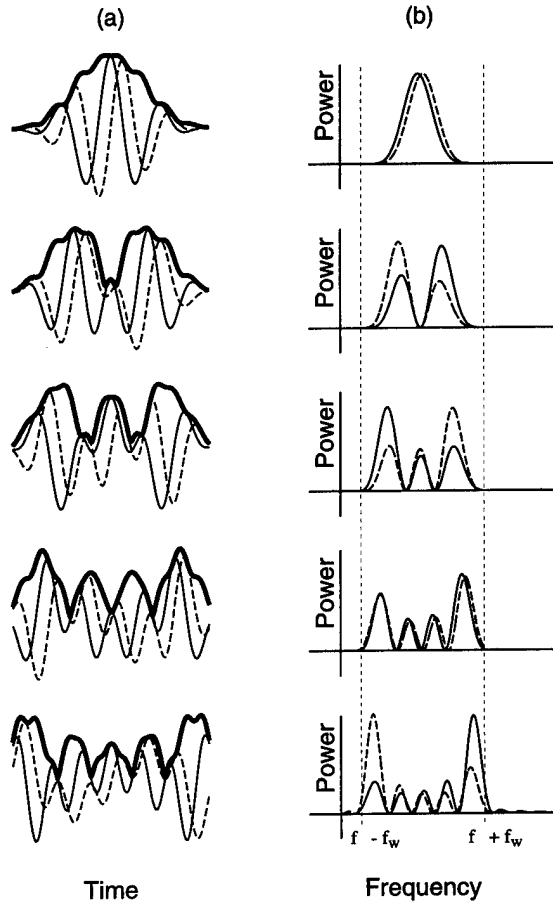


Figure 1. Ten real Lilly and Park wavelets plotted as even and odd pairs, (a) in the time domain and (b) their frequency power spectra. The thick lines in (a) are the upper portions of the envelopes for the multi-wavelets that roughly describe the weighting of each pair in the time domain. The solid and dashed thin lines correspond to the even and odd wavelets, respectively.

The number of multi-wavelet transforms that can be constructed for a given frequency band is dependent upon the desired time and frequency resolution, as specified in Lilly and Park (1995). Lilly and Park (1995) used different basis functions for different frequency bands, but Bear and Pavlis (1997) show that the same basis functions can be used for progressively lower frequency bands by using a series of decimation stages.

Principal-Component Analysis

With the multi-wavelet transform, an estimate of the covariance matrix $\mathbf{R}^{(j)}(f, t, \mathbf{p})$ can be obtained from the outer conjugate product of the complex vector $\mathbf{a}^{(j)}(f, t, \mathbf{p})$ whose entries are the transformed seismic data at each station

$$a_n^{(j)}(f, t, \mathbf{p}) = \mathcal{W}^{(j)}[s_n][f, t + \tau(\mathbf{p}, n)]. \quad (7)$$

With this definition, $\mathbf{d} \cdot \mathbf{a}^{(j)}(f, t, \mathbf{p})$ is the complex array

beam for wavelet j , and equation (4) can still be used to compute the power in the beam. $\mathbf{R}^{(j)}(f, t, \mathbf{p})$, for a fixed j and time t , is of rank one. One standard way to increase the rank is to use a fixed time window and average the power of the transformed data over frequency (e.g., Wagner and Owens, 1996). For our analysis, however, we choose to fix the frequency band and average the power over K points in time as

$$\mathbf{R}^{(j)}(f, t, \mathbf{p}) = \frac{1}{K} \mathbf{A}^{(j)}(f, t, \mathbf{p}) \mathbf{A}^{(j)\dagger}(f, t, \mathbf{p}),$$

where

$$\mathbf{A}^{(j)}(f, t, \mathbf{p}) = [\mathbf{a}^{(j)}(f, t_1, \mathbf{p}) \dots \mathbf{a}^{(j)}(f, t_K, \mathbf{p})]. \quad (8)$$

The rows of each of these matrices are time-shifted (using slowness vector $\mathbf{p}$), complex valued signals yielded by convolution with the j th pair of multi-wavelets. We note that the t_k are not necessarily adjacent time samples nor evenly spaced in time. We used $t_k = [t + (k - 1)\Delta t]$, where Δt was chosen so that $t_K - t_1$ was the length of one-half of one cycle of the center frequency. The value of Δt could be as small as the sample rate, but for low frequencies, this is ill-advised as the number of columns in $\mathbf{A}^{(j)}$ can become large. More discussion on the structure of the complex matrix $\mathbf{A}^{(j)}$ appears in the Appendix.

Any complex matrix can be written as a singular value decomposition:

$$\mathbf{A}^{(j)} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^\dagger, \quad (9)$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Lambda}$ is a real diagonal matrix. Note that

$$\begin{aligned} \mathbf{R}^{(j)}(f, t, \mathbf{p}) &= \frac{1}{K} \mathbf{A}^{(j)}(f, t, \mathbf{p}) \mathbf{A}^{(j)\dagger}(f, t, \mathbf{p}) \\ &= \frac{1}{K} \mathbf{U} \mathbf{\Lambda} \mathbf{V}^\dagger \mathbf{V} \mathbf{\Lambda} \mathbf{U}^\dagger = \frac{1}{K} \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^\dagger. \end{aligned} \quad (10)$$

Thus, finding the singular value decomposition (SVD) of $\mathbf{A}$ is equivalent to finding the eigenvalue decomposition of $\mathbf{R}$.

We assume the SVD is organized such that $\lambda_1 \geq \lambda_2 \dots \lambda_{\min(N, K)}$. $\mathbf{U}$ can be considered as a rotation of the data such that the column $\mathbf{u}_1$ points in the direction of largest energy (for further discussion, see the Appendix). If the signals from each station are exactly lined up in time and where identical in form, then the values in $\mathbf{u}_1$ would all be the same. If there are time shifts between stations, then the complex values of $\mathbf{u}_1$ will have different associated phase angles. Similarly, amplitude variations between stations will be reflected in the modulus of the components of $\mathbf{u}_1$.

We can relate the variations in phase angles between stations m and n to time differences by

$$e^{i(\theta_m - \theta_n)} = e^{i2\pi(f_m t_m - f_n t_n)}. \quad (11)$$

If we set f_m and f_n to the center frequency of the wavelet transform, the relation between phase angle and time difference is direct. These phase differences, which we determine from $\mathbf{u}_1$, reflect the deviations of the data from the propagation model (here deviations from a plane-wave fit). Thus, the related time differences are the time residuals that we are trying to determine.

We note that this approach incorporates aspects of three traditional array processing methods: frequency-domain beamforming, time-domain beamforming, and principal-component analysis. It is frequency-like through its use of the complex phase, but it is also timelike in that the data are directly shifted in time based on a best-fit plane-wave arrival. Finally, it is a principal-component method because it uses only the largest singular value and associated left eigenvector for determining the phase angles between stations. In fact, this approach is similar in many respects to the multi-channel detector described in Wagner and Owens (1996).

Methodology

Our method is a staged process in which the final time residuals are determined by progressively accumulating time residuals from the lowest to the highest frequencies. Each stage is predicated on choosing a frequency band for the multi-wavelet transforms to be applied to the data. The frequencies for a particular stage must be low enough so that the data stack fairly well even with the remaining time residuals but not so low that the phase differences between station signals are unresolvable (i.e., the maximum spread of the time residuals should generally be between 1/15 and 1/4 the center period). Because each stage decreases the spread of the remaining time residuals, we naturally progress from lower to higher frequencies. Our method is fundamentally broadband in the real sense of the word.

For an L staged process, there will be a set of center frequencies for the multi-wavelet transforms $f_1 < f_2 < \dots < f_L$. During the I th stage, we form the $N \times K$ matrices $\mathbf{A}^{(j)}(f, t, \mathbf{p})$, where

$$\mathbf{A}_{nk}^{(j)}(f, t, \mathbf{p}) = \mathcal{W}^{(j)}[s_n][f, t_k] + \tau(\mathbf{p}, n) + \sum_{i=1}^{I-1} \tau_i(n). \quad (12)$$

The superscript j again denotes the j th member of a set of J multi-wavelet transforms with center frequency f . We denote the time residuals determined from the previous $I - 1$ iterations as $\tau_i(n)$. We determine the singular value decompositions of the $\mathbf{A}^{(j)}(f, t, \mathbf{p})$ and pick out the largest singular value $\lambda_1^{(j)}(f, t, \mathbf{p})$ and the associated left eigenvector $\mathbf{u}_1^{(j)}(f, t, \mathbf{p})$. To measure the coherence of the stack, we use

$$\begin{aligned} S^{(j)}(f, t, \mathbf{p}) &= \frac{|\lambda_1^{(j)}(f, t, \mathbf{p}) \mathbf{d} \cdot \mathbf{u}_1^{(j)}(f, t, \mathbf{p})|^2}{\|\mathbf{d}\|^2 \|\lambda_1^{(j)}(f, t, \mathbf{p}) \mathbf{u}_1^{(j)}(f, t, \mathbf{p})\|^2} \\ &= \frac{|\sum_{k=1}^N u_{1k}^{(j)}|^2}{\sum_{k=1}^N |u_{1k}^{(j)}|^2}, \end{aligned} \quad (13)$$

where $\mathbf{u}_{1k}^{(j)}$ is the k th component of $\mathbf{u}_1$. This measure is directly comparable to semblance commonly used in time-domain beamforming (e.g., Husebye and Ruud, 1989). The numerator is the power of the beam, and the denominator is the average signal power at each station.

For each time t and a set of possible slowness vectors B , we determine the peak coherence values $S_{\text{peak}}^{(j)}(f, t)$ from a grid search through the values calculated using equation (13). We choose the time to use in estimating the slowness vector by averaging the peak coherences $S_{\text{peak}}^{(j)}(f, t)$ for the J multi-wavelet transforms and finding the time point t_{max} where the average peak coherence attains a maximum value. We calculate estimates for the east-west and north-south components of the slowness vector separately. We compute the east-west slowness components by

$$p_{ew}^{(j)} = \frac{\sum_{(p_k, p_i) \in B} p_k m(p_k, p_i)}{\sum_{(p_k, p_i) \in B} m(p_k, p_i)},$$

where

$$m(p_k, p_i) = \begin{cases} 0 & \text{if} \\ S^{(j)}(f, t_{\text{max}}, p_k, p_i) < 0.9 S_{\text{peak}}^{(j)}(f, t_{\text{max}}), \\ S^{(j)}(f, t_{\text{max}}, p_k, p_i) & \text{if} \\ S^{(j)}(f, t_{\text{max}}, p_k, p_i) \geq 0.9 S_{\text{peak}}^{(j)}(f, t_{\text{max}}). \end{cases} \quad (14)$$

The north-south components are computed in a similar fashion. Equation (14) involves a center of mass calculation for points with values above the threshold $0.9 S_{\text{peak}}^{(j)}(f, t_{\text{max}})$ and is preferable to a pure peak measure due to the discrete nature of the set of slowness vectors. The final estimate for the slowness vector $\mathbf{p}_{\text{est}}$ is found by applying a robust M-estimator (Bear and Pavlis, 1997; Chave *et al.*, 1987) to the separate $\mathbf{p}^{(j)}$ measurements.

In this article, every time we produce an estimate from a set of values, we use an M-estimator to protect the results from being severely biased by any outlying values. Outlying values in this case could be due to spectral nulls in the signal that correspond to a particular wavelet's frequency content. In cases of estimation from a set of station and/or component data, outliers could be due to increased local noise or the breakdown of a given station or station's component. Obviously, these types of outliers do not reflect the data process itself and would generally be considered bad points to be thrown away. The M-estimation procedure provides an automatic way of throwing away such points without requiring human intervention.

Phase Shifts to Time Residuals

The values in $\mathbf{u}_1^{(j)}$ are complex numbers with associated magnitudes and phase angles. If we plot these numbers (Fig. 2), we expect them to cluster in the complex plane (i.e., the signals almost match, even with no shifting). We produce a

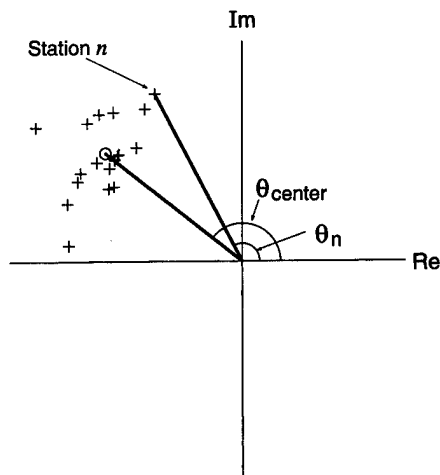


Figure 2. Plot of values of a single-component $\mathbf{u}_1$ in the complex plane. The estimate of the center point (circle) was determined by applying a robust estimator to the data values (crosses).

robust M-estimate of the center of the distribution of the complex values of $\mathbf{u}_1$ and define its phase angle $\theta_{\text{center}}^{(j)}$ as the zero time residual. We then convert the phase angle for the n th station to a time residual by

$$e^{i(\theta_n^{(j)} - \theta_{\text{center}}^{(j)})} = e^{i2\pi[f_c \tau_1^{(j)}(n)]}, \quad (15)$$

where f_c is the center frequency of the multi-wavelet transforms. Finally, we calculate the time residuals $\tau_1^{(j)}(n)$ using a robust M-estimator on the separate $\tau_1^{(j)}(n)$.

Method with Three-Component Data

The method we use for three-component data is a direct extension of the single-component case. The largest change is the addition of a data rotation to place the maximum amount of signal energy along one axis.

Beamforming

The beamforming computation for the l th stage is essentially the same as in the single-component case except the matrices $\mathbf{A}^{(j)}$ are now of dimension $3N \times K$. The matrices can be partitioned by the three components such that

$$\mathbf{A}^{(j)}(f, t, \mathbf{p}) = \begin{bmatrix} \mathbf{A}^{(j),1}(f, t, \mathbf{p}) \\ \mathbf{A}^{(j),2}(f, t, \mathbf{p}) \\ \mathbf{A}^{(j),3}(f, t, \mathbf{p}) \end{bmatrix},$$

where

$$\begin{aligned} \mathbf{A}_{nk}^{(j),c}(f, t, \mathbf{p}) \\ = \mathcal{W}^{(j)}[s_n^c] \left[f, t_k + \tau(\mathbf{p}, n) + \sum_{i=1}^{l-1} \tau_i(n) \right] \end{aligned} \quad (16)$$

and where s_n^c denotes the signal recorded on the c th component at the n th station. The left eigenvectors can also be partitioned into

$$\mathbf{u}_1^{(j)}(f, t, \mathbf{p}) = \begin{bmatrix} \mathbf{u}_1^{(j),1}(f, t, \mathbf{p}) \\ \mathbf{u}_1^{(j),2}(f, t, \mathbf{p}) \\ \mathbf{u}_1^{(j),3}(f, t, \mathbf{p}) \end{bmatrix}. \quad (17)$$

Thus, we can replace equation (13) with

$$\begin{aligned} \mathbf{S}^{(j)}(f, t, \mathbf{p}) \\ = \max_{c=1,2,3} \frac{\|\lambda_1^{(j)}(f, t, \mathbf{p}) \mathbf{d} \cdot \mathbf{u}_1^{(j),c}(f, t, \mathbf{p})\|^2}{\|\mathbf{d}\|^2 \|\lambda_1^{(j)}(f, t, \mathbf{p}) \mathbf{u}_1^{(j),c}(f, t, \mathbf{p})\|^2}. \end{aligned} \quad (18)$$

Axis Rotation

The left eigenvectors $\mathbf{u}_1^{(j)}$ from equation (17) combine relative phase and amplitude information for all three data components. If we group the values of $\mathbf{u}_1^{(j)}$ by station, then we have N complex three-vectors for each multi-wavelet transform. These three-vectors are analogous to the principal eigenvectors determined in Vidale's (1986) principal-component analysis method. Vidale used the complex analytic signals, determined using a Hilbert transform, to produce the 3×3 covariance matrix. We are using the multi-wavelet transformed data, but in Bear and Pavlis (1997), we showed that the multi-wavelet transform behaves in a manner similar to a Hilbert transform. The three-vectors can be written as $(r_1 e^{2\pi i \phi_1}, r_2 e^{2\pi i \phi_2}, r_3 e^{2\pi i \phi_3})$ and can be considered as three-dimensional phasors that define ellipses of instantaneous particle motion. The best single-component recording of the signal at a particular station would have been accomplished by orienting the recording component along the major axis of the particle motion ellipse for that station. We can construct this best recording by changing the coordinate system of our three-component data so that one of the new coordinate axes lines up with the major axis of the particle motion ellipse. We choose a best overall coordinate system for the array by aligning one of the new axes with the direction determined using a robust M-estimator on the major axes from all the stations. (More details on this procedure are given in the companion article by Bear *et al.*, 1999.)

After rotation, we perform the rest of the processing using an equivalent single-component array with one component aligned along this direction. Thus, we are back to the single-component case, and the phase shift to time residual step is accomplished just as before. We note, however, that at each stage, we choose to recalculate the best particle motion major axis. We do this because the polarization can be strongly dependent on frequency (Park *et al.*, 1987).

Data Examples

Deep-Focus Bolivian Earthquake

Our first data example is a set of recordings of the 9 June 1994 Bolivian deep-focus event (Wu and Beck, 1995;

Goes and Ritsema, 1995). We created a large aperture array of 19 three-component broadband stations situated in southern California as shown in Figure 3. These stations have been pulled from three networks: TERRAscope (Wainger, 1988), a subset of ANZA (Fletcher *et al.*, 1987), and a temporary network (Ichinose *et al.*, 1996). This array encompasses most of southern California, with an aperture of approximately 300 km.

The epicenter for the Bolivian event is at a backazimuth of 121° from the array, and slowness analysis for the full ANZA array also suggests an azimuth of approximately 121° . Thus, we restricted our range of slowness vectors to those within 3° of 121° to reduce processing time. We started our processing in a frequency band with $f_c = 0.05$ Hz ($f_w = 0.0375$ Hz) and found that the major axis for the particle motion also suggested an azimuth of 121° . We then progressed through frequency bands of $f_c = 0.2$ Hz ($f_w = 0.15$ Hz) and $f_c = 0.5$ Hz ($f_w = 0.375$ Hz). The effects of the removals of the time residuals at each iterative step are shown in Figure 4. This figure illustrates four interesting features. First, we note that the signals without time residuals removed would obviously not stack at the dominant frequency in this window of 0.5 Hz. Second, we note that to deal with time residuals as large as those seen in Figure 4 and to be able to stack to the dominant frequency, we need the instruments to be able to record frequencies from approximately 0.05 to 1.0 Hz. Instrumentation capable of this has only become common in the last decade. Third, we note

that stations USC and RPV have significant differences in time residuals and signal shape. Yet these two stations are at similar elevations and relatively close to one another. Finally, we note that the signal alignment process is not necessarily smooth. In Figure 4b, we see that the signal for GSC has shifted past both stations USC and RPV to the left and then must shift back toward the right in Figure 4c.

In Figure 5, we examine spectral properties of the array stack for the Bolivian event both with and without the time residuals removed. The spectral ratios were determined by taking the power spectrum of the array beam at each frequency [determined using the multi-taper method of Thomson (1982)] and dividing it by the median power spectra of the individual signals at each array station. The higher the spectral ratios, the better the data stacks. Because of the very wide bandwidth of these data, we computed spectral ratios for time windows of two different lengths—100 and 10 sec. The former provides improved frequency resolution for the



Figure 3. The locations of the 19 broadband, three-component instruments from the TERRAscope and ANZA networks and a temporary network in southern California that form our large aperture array. The thick arrow in the lower right denotes the direction of arrival of the Bolivian event.

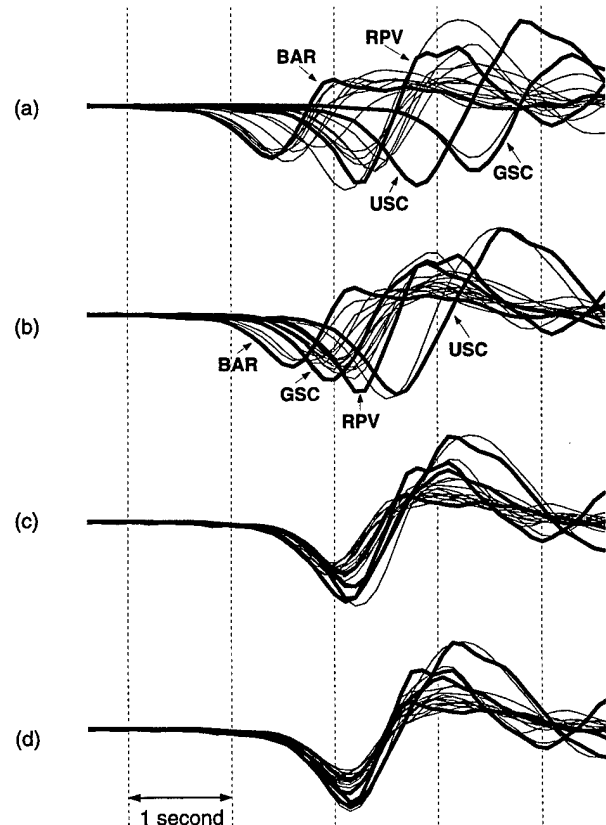


Figure 4. The iterative removal of time residuals from the Bolivian event. All the station signals are plotted with data from four stations emphasized with thicker lines. Shown are the vertical components after the removal of the plane-wave time delays for (a) the original data, (b) with the time residuals from the 0.05-Hz frequency band removed, (c) with the time residuals from (b) and from the 0.2-Hz frequency band removed, and (d) with the time residuals from (b) and (c) and from the 0.5-Hz frequency band removed.

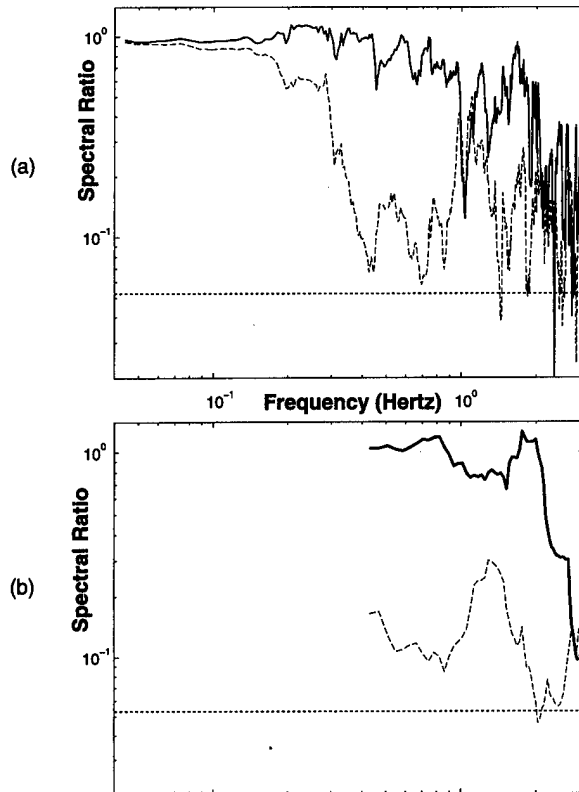


Figure 5. Power spectral ratios for the Bolivian event using (a) a 100-sec-long analysis window and (b) a 10-sec-long analysis window. The bold curves are the spectral ratios for the beam with the time residuals removed. The dashed curves are the spectral ratios for the beam without the time residuals removed. The dotted lines denote the theoretical pure white noise floor at $1/N$.

low frequencies, while the latter provides improved time resolution at the higher frequencies. As expected, the removal of the time residuals has vastly decreased the power loss in the stack at the higher frequencies. Without corrections for statics, this group of stations would not function as an array for frequencies greater than 0.2 Hz. With the removal of the time residuals, we get coherent stacks to 2.0 Hz, which is within 1 Hz of the point where the data intersect the noise floor. This benefit to stacking with careful prealignment has been known for some time (Bungum and Husebye, 1971; Cox, 1973), but its importance seems to be less widely recognized.

A Local Earthquake

Our second example is a local earthquake recorded by a high-frequency array. This event was recorded on 21 January 1994 at the Geyokcha array on the border between Iran and Turkmenistan. We use a subset of 36 stations from the array equipped with triaxial 4.5-Hz natural period sensors. The location of the array and the station configuration are shown in Figure 6. Note that this array is configured as a

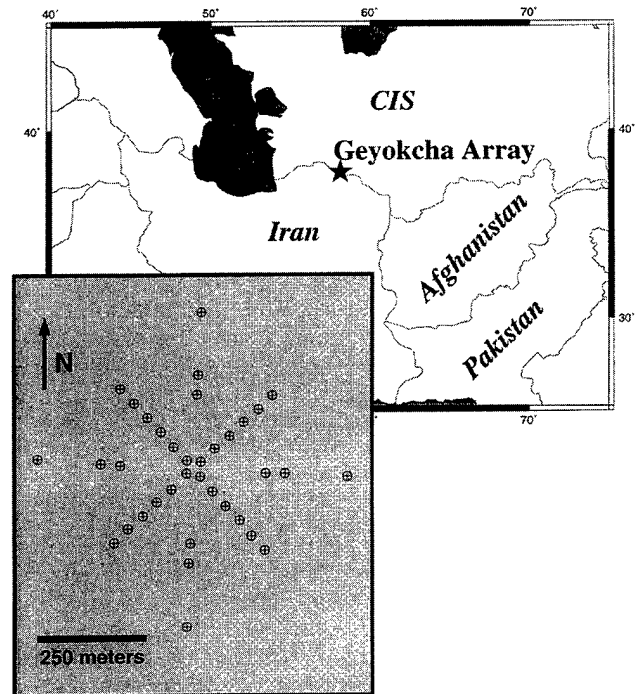


Figure 6. The 36 high-frequency stations of the Geyokcha array. The star on the map shows the location of the array in central Asia. The inset map shows the array geometry.

square, 600 m on a side. This is more than 300 times smaller than the California array.

The Geyokcha array is located on mudstone bedrock (P velocity ≈ 2.5 km/sec) and has 50 m of topographic relief with a general slope from northwest to southeast. These elevation differences produce static time delays at each station, and if the topography were exactly planar, then the delays could be written

$$\tau_s(n) = \mathbf{a} \cdot \mathbf{x}_n \quad (19)$$

where $\mathbf{a}$ has units of time per distance and defines the time delays due to topography with respect to direction. We can write the total time delays as

$$\begin{aligned} \tau_{\text{tot}}(\mathbf{p}, n) &= \tau(\mathbf{p}, n) + \tau_s(\mathbf{a}, n) \\ &= -\mathbf{p} \cdot \mathbf{x}_n + \mathbf{a} \cdot \mathbf{x}_n, \end{aligned} \quad (20)$$

which is indistinguishable from

$$\tau'(\mathbf{p}', n) = \mathbf{p}' \cdot \mathbf{x}_n = -(\mathbf{p} - \mathbf{a}) \cdot \mathbf{x}_n. \quad (21)$$

Equation (21) implies that planar changes in topography, if not corrected for before beamforming, can change the calculated slowness vector significantly. For the event we are examining, the backazimuth changed by 10° and the apparent velocity increased by 2.5 km/sec after we applied

the topographic corrections. We note that other planar effects in the Earth (e.g., a dipping structure) would also have similar effects on the slowness vector such that the time delays can be written

$$\tau_{\text{tot}}(\mathbf{p}, n) = \tau(\mathbf{p}, n) + \sum_{p=1}^P \tau_s(\mathbf{a}_p, n). \quad (22)$$

The only question is how much effect they have on the calculated slowness vector. It is conceivable that an equation could be derived to solve for the summed effect because the $\mathbf{a}_p$ are static in time while $\mathbf{p}$ varies with arrival time (analogous to the relation in reflection seismology between residual statics and normal moveout corrections). We do not pursue this path though, because this array is small.

We started our processing in a frequency band with $f_c = 8.0$ Hz ($f_w = 6.0$ Hz) and iterated through $f_c = 15.625$ Hz ($f_w = 11.72$ Hz) and $f_c = 31.25$ Hz ($f_w = 23.4375$ Hz). The effects of the removals of the time residuals at each iteration step are shown in Figure 7. In Figure 7a, we can see that there are still significant time residuals even after the topographic corrections have been applied. We can also see that there would be significant power loss at the dominant frequency (25 Hz) if the data were stacked without applying the static corrections we determined.

The spectral ratios plotted in Figure 8 were calculated in the same manner as those for the Bolivian event in Figure 5, except we only required a single time window of length 1.25 sec to resolve all the frequencies of interest. We see that the beam for the data without any static corrections differs little from the corrected beam below 14 Hz. A major portion of the topographic effects appears to have been absorbed into the shifting of the slowness vector when trying to produce a best beam in the 2- to 14-Hz frequency band. The removal of the time residuals is necessary, though, to significantly improve the stack in the 14- to 35-Hz band.

Why both beams lose power so precipitously above 35 Hz is not understood. We would expect to be able to stack to approximately 50 Hz, because the incident wave field arrives at near-vertical incidence (13 km/sec—much like a teleseism) and the wavelengths for frequencies between 35 and 50 Hz are still comparable to the dimensions of the array. When we look at the data filtered to frequencies above 40 Hz, we can see that there is significant signal but that the signals differ drastically from station to station both in amplitude (by an order of magnitude) and in waveform. This loss in signal coherence above 35 Hz may be due to very localized near-surface effects. Wilson (1997) notes for an array somewhat smaller than Geyokcha that the spatial scale length of signal spectral fluctuation decreases with higher frequencies. He argues that these fluctuations are due to near-surface scattering at the base of the weathered layer. The approximate wavelength for ground roll (assuming a velocity of 1.0 km/sec) at a frequency of 35 Hz is 29 m. This is a scale much smaller than the size of the array and

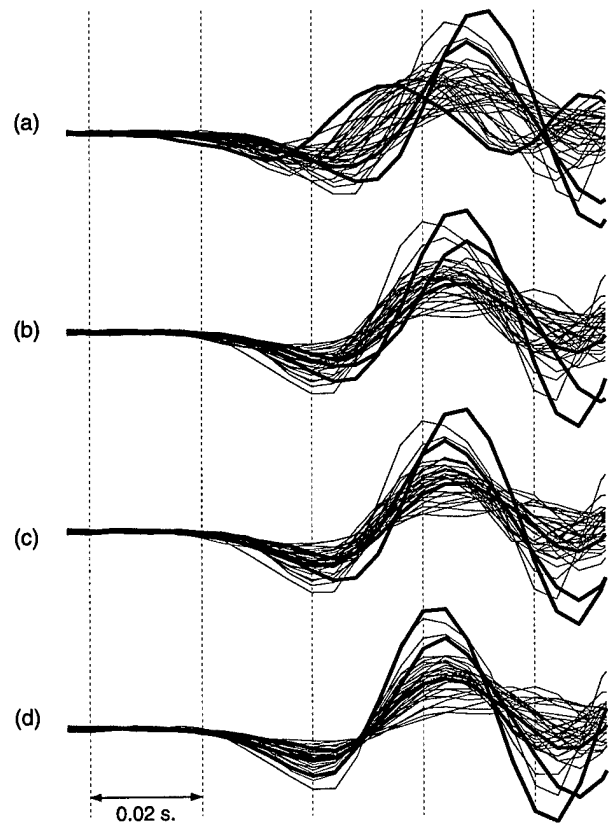


Figure 7. The iterative removal of time residuals from a local event near the Geyokcha array. The signals from three stations are emphasized. Shown are the vertical components after the removal of the topographic and plane-wave time delays for (a) the original data, (b) with the time residuals from the 8-Hz frequency band removed, (c) with the time residuals from (b) and from the 15.625-Hz frequency band removed, and (d) with the time residuals from (b) and (c) and from the 31.25-Hz frequency band removed.

comparable to the interstation spacing. We suggest that the loss in beam power may be indicative of scattering into high-frequency Rayleigh waves or very localized near-surface resonances that vary under scales of 50 m.

Discussion

We have introduced a new method for calculating time residuals directly from recorded seismic signals that incorporates aspects of three traditional array processing techniques: frequency-domain beamforming, time-domain beamforming, and principal-component analysis. All the station and component data are used simultaneously in determining an arrival's slowness vector for a best-fit plane wave and in calculating the time residuals that represent the misfit to the plane-wave model. Our method works like a multi-channel cross-correlation procedure but does not directly use any correlation functions.

Our method is broadband in the real sense of the word.

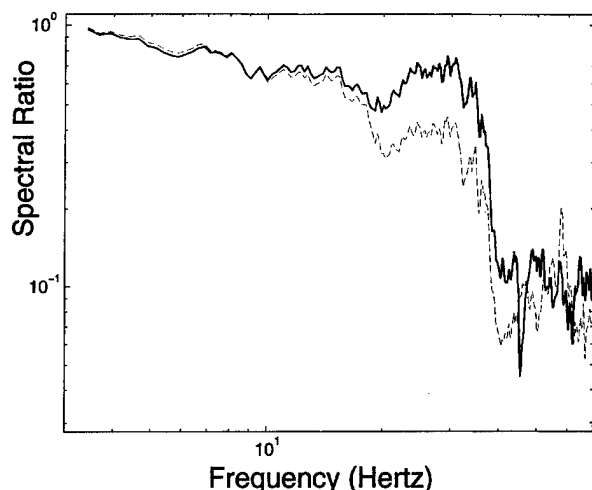


Figure 8. Power spectral ratios for a local event. The bold curve is the spectral ratio for the beam after topography effects and time residuals are removed. The dashed curve is the spectral ratio for the beam without any corrections. The bottom axis is at the $1/N$ level predicted for pure white noise.

We start by processing at low enough frequencies so that array beamforming will find a coherent stacking direction despite any time residuals. The time residuals appear as phase shifts in the complex transformed data. The phase shifts are converted to times, removed from the data, and we progress to a higher frequency band to remove remaining time residuals. Since the only information needed by this process is the frequency band for the multi-wavelet transforms and a range of times and slowness vectors, it should be easy to automate for routine processing.

The type of data necessary for applying this method depends on the scale of the time residuals (generally related to the scale of the array) and the frequencies contained in the signal. The Bolivian event had time residuals of 2 sec across the large aperture California array and a dominant frequency of 0.5 Hz for the first arrival. To remove the time residuals so that the data would stack to the dominant frequency, we needed a recording of the signal from approximately 0.05 to 1.0 Hz. This is an extremely wide bandwidth, and the use of this method would require the use of broadband instrumentation for analysis of teleseismic signals.

On the other hand, the time residuals for the local event at Geyokcha were only on the order of 0.02 sec, and the dominant frequency of the processed arrival was 25 Hz. Thus, we only needed frequencies above 2 Hz, and the high-frequency sensors of the array were acceptable. A major difficulty with these data were topographic effects. The topographic corrections for the Geyokcha array were of the same order as the time residuals we determined from our procedure. Furthermore, because of the planar slope of the topography in the array, these corrections produced a large shift in the measured slowness vector. In this case, the removal

of the topographic effect was critical to producing unbiased time residuals. Had we not made this correction, both the estimated slowness vector and residuals would have contained this bias. This situation has an exact parallel in reflection seismology processing. The elevation corrections we applied are completely analogous to what are usually called geometric statics, and the corrections we estimate are analogous to higher order static corrections computed by a range of methods (residual statics, refraction statics, etc.).

There are circumstances that may cause difficulties for this method. As noted earlier, the recording bandwidth for the data can pose a limitation. If the instrumentation in use does not record to low enough frequencies, an initial step to remove the largest time residuals by another method may be necessary. This may also be the case for data that has a low signal-to-noise ratio in the lowest frequencies. With teleseismic signals, it is common to find low signal-to-noise ratios in the microseism band (15- to 5-sec period). For such signals, it would probably be necessary to skip this band. In this situation, it may prove more practical to start the process from a set of initial picks made by an automated picker or by a human analyst. For a system designed to process large amounts of data, these issues need to be addressed, but considering them in detail is beyond the scope of this article.

We also need to stress that the time residuals determined here are generally not the same as those determined by standard methods. The direction of arrival is determined independently from the source location and is reflected in the slowness vector used to determine the plane-wave time delays $\tau(\mathbf{p}, n)$. These plane-wave delays are removed from the data before the time residuals are calculated. If the slowness vector azimuth is the same as the azimuth from the array to the source, then the time residuals determined by this method will be comparable to those determined using a standard method. Otherwise, they will be distinctly different. Comparable time residuals, in this case, could be determined by adding the plane-wave time delays to the time residuals, then removing the travel path time ΔT from each station.

Acknowledgments

We are grateful to Frank Vernon for putting together the data from the Bolivian earthquake and sharing it with us. We are equally grateful to the long cast of characters who helped field the Geyokcha data and assemble the results into a workable data set. The funding for this work was provided by an AASERT grant from the Air Force Office of Scientific Research (Number F49620-95-1-0366) and by the IRIS Joint Seismic Program.

References

- Bear, L. K. and G. L. Pavlis (1997). Estimation of slowness vectors and their uncertainties using multi-wavelet seismic array processing, *Bull. Seism. Soc. Am.* **87**, 755–769.
- Bear, L. K., G. L. Pavlis, and G. H. R. Bokelmann (1999). Multi-wavelet analysis of three-component seismic arrays: application to measure effective anisotropy at Piñon Flats, California, *Bull. Seism. Soc. Am.* **89**, 693–705.

- Bungum, H. and E. S. Husebye (1971). Errors in time delay measurements, *Pageoph* **91**, 56–70.
- Chave, A. D., D. J. Thomson, and M. E. Ander (1987). On the robust estimation of power spectra, coherences, and transfer functions, *J. Geophys. Res.* **92**, 633–648.
- Cox, H. (1973). Resolving power and sensitivity to mismatch of optimum array processors, *J. Acoust. Soc. Am.* **54**, 771–785.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.
- Fletcher, J. B., L. C. Haar, T. Hanks, L. M. Baker, F. L. Vernon, J. Berger, and J. Brune (1987). The digital array at Anza, California: processing and initial interpretation of source parameters, *J. Geophys. Res.* **92**, 369–382.
- Goes, S. and J. Ritsema (1995). A broadband P wave analysis of the large deep Fiji Island and Bolivia earthquakes of 1994, *Geophys. Res. Lett.* **22**, 2249–2252.
- Husebye, E. S. and B. O. Ruud (1989). Array seismology; past, present, and future developments, in *Observatory Seismology*, J. J. Litchiser (Editor), University of California Press, Berkeley, 123–153.
- Ichinose G., S. Day, H. Magistrale, T. Prush, F. Vernon, and A. Edelman (1996). Crustal thickness variations beneath the Peninsular Ranges, southern California, *Geophys. Res. Lett.* **23**, 3095–3098.
- Kumar, P. and E. Fouloula-Georgiou (1994). Wavelet analysis in geophysics: An introduction, in *Wavelets in Geophysics*, E. Fouloula-Georgiou and P. Kumar (Editors), Academic, San Diego, California, 1–43.
- Kvaerna, T. and D. J. Doornbos (1986). An integrated approach to slowness analysis with arrays and three-component stations, *NORSAR* **2** 85/86.
- Lay, T. and T. C. Wallace (1995). *Modern Global Seismology*, Academic, New York, 521 pp.
- Lilly, J. M. and J. Park (1995). Multiwavelet spectral and polarization analyses of seismic records, *Geophys. J. Int.* **122**, 1001–1021.
- Park, J., F. L. Vernon, and C. R. Lindberg (1987). Frequency dependent polarization analysis of high-frequency seismograms, *J. Geophys. Res.* **92**, 12664–12674.
- Slepian, D. (1983). Some comments on Fourier analysis, uncertainty and modeling, *Soc. Indust. App. Math. Rev.* **25**, 379–393.
- Strang, G. and T. Nguyen (1996). *Wavelets and Filter Banks*, Wellesley, Cambridge, 490 pp.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis, *Proc. IEEE* **70**, 1055–1096.
- VanDecar, J. C. and R. S. Crosson (1990). Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bull. Seism. Soc. Am.* **80**, 150–169.
- Vidale, J. E. (1986). Complex polarization analysis of particle motion, *Bull. Seism. Soc. Am.* **76**, 5:1393–1405.
- Wagner, G. S. and T. J. Owens (1996). Signal detection using multi-channel seismic data, *Bull. Seism. Soc. Am.* **86**, 221–231.
- Wainger, L. A. (1988). Caltech sets up Terrascope to collect real-time seismic data on earth motions, *Earth Space* **1**, 11–12.
- Wilson, D. C. (1997). Near-surface site effects in crystalline bedrock: azimuthal dependence, scale lengths of spectral variation, and change in spectral character with depth, *Master's Thesis*, Indiana University, 27 pp.
- Wu, J. and S. Beck (1995). A very broadband study of the 1994 deep Bolivia earthquake sequence, *Geophys. Res. Lett.* **22**, 2237–2240.

Appendix: Further Discussion on the Complex Matrix $A^{(j)}$

Figure A1 shows graphically the steps leading to the creation of the complex matrix $A^{(j)}(f, t, \mathbf{p})$. We start with the raw time signals (Fig. A1a) and use multi-wavelet beam-forming (equation 14) to perform a grid search over possible slowness vectors ($\mathbf{p}$) to determine the best-fit plane-wave

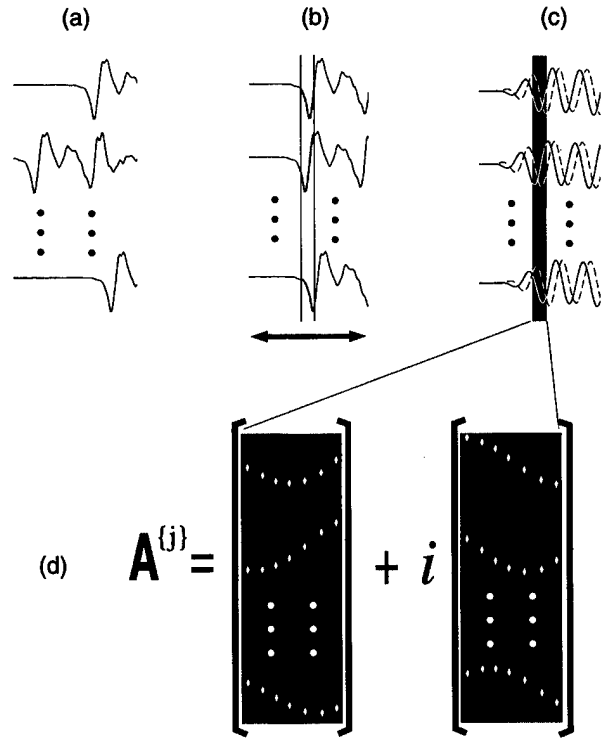


Figure A1. A graphical representation of the creation of the complex matrix $A^{(j)}$ showing 3 of N signals. (a) Start with the raw data— $s_n(t)$. (b) Remove the plane-wave time delays from (a)— $s_n[t + \tau(\mathbf{p}, n)]$. (c) Apply the multi-wavelet transform to (b)— $W^{(j)}[s_n][f, t + \tau(\mathbf{p}, n)]$ (solid line is the real part; dashed line is the imaginary part). (d) The complex matrix $A^{(j)}$ is formed from the data in (c). The black windows delineate the time window of analysis. The two-headed arrow in (b) delineates the portion of the signal that contributes to the transformed signal in the analysis window of (c).

time delays $\tau(\mathbf{p}_{\text{est}}, n)$. We remove these time delays from the data (Fig. A1b) and then apply the multi-wavelet transforms with center frequency f (Fig. A1c). We choose a time window of analysis starting at time t and pick the values of the transformed data at time samples $t_1, \dots, t_K$ within this window to form the matrix $A^{(j)}(f, t, \mathbf{p})$ (Fig. A1d). Note that because the transformed data are complex, the values in $A^{(j)}(f, t, \mathbf{p})$ are complex.

The matrix $A^{(j)}$ has a physical interpretation that is basic to the understanding of our method of time-residual estimation. The entries in each row are analogous to time samples taken from the filtered (complex) analytic signal, because the real and imaginary parts of the multi-wavelet transform have a similar orthogonal relationship (Bear and Pavlis, 1997). The complex values in each column, which can be written $r_n e^{i\theta_n}$, provide information on how misaligned the signals are from one station to another (seen in the phase angles θ_n), so for K time samples, there are K estimates of this misalignment.

We can investigate the physical meaning of performing principal-component analysis on the matrix $\mathbf{A}^{(j)}$ in the following manner. We consider the case where the data at each station consists of a known, perfectly coherent signal S with slowness vector $\mathbf{p}$ and "random noise" v . We define the transformed data vectors for the signal and noise respectively as

$$\mathbf{w} = [\mathcal{W}^{(j)}[S][f, t + \tau(\mathbf{p}, 1)] \dots \mathcal{W}^{(j)}[S][f, t + \tau(\mathbf{p}, N)]]^T \quad (\text{A1})$$

and

$$\boldsymbol{\eta}(t) = [\mathcal{W}^{(j)}[v][f, t + \tau(\mathbf{p}, 1)] \dots \mathcal{W}^{(j)}[v][f, t + \tau(\mathbf{p}, N)]]^T, \quad (\text{A2})$$

where the superscript T denotes the matrix transpose. This means that $\mathbf{d} \cdot \mathbf{w} = \mathcal{W}^{(j)}[S][f, t + \tau(\mathbf{p}, 1)]$ and hopefully $\mathbf{d} \cdot \boldsymbol{\eta}(t) \approx 0$ (where $\mathbf{d}$ is as defined in equation 4). If we assume the signal propagates with a constant slowness vector over the analysis time window, then we can write

$$\mathbf{A}^{(j)}(f, t, \mathbf{p}) = [c_1 \mathbf{w} + \boldsymbol{\eta}(t_1) \dots c_K \mathbf{w} + \boldsymbol{\eta}(t_K)], \quad (\text{A3})$$

where the c_k are complex constants.

If there is no noise, then $\mathbf{A}^{(j)}$ is of rank one and $\mathbf{u}_1 \equiv \mathbf{w}$. If the noise is, on average, nondirectional, then $\mathbf{u}_1$ should point in a direction similar to that of $\mathbf{w}$. If no averaging occurs, then the direction of $\mathbf{u}_1$ will be biased in the direction of the noise at that time. A cartoon of the situation is shown in Figure A2, where $N = 3$ and the data are real.

Real signals, however, differ from this simplified model in two ways. First, amplitude fluctuations can occur with no change in waveform. This has no effect on this method because we only consider the phase. The second more pervasive problem is introduced by scattering. Scattering introduces waveform fluctuations within the scale of a wavelength. This leads to variations in waveform shape from station to station. These variations in waveform shape changes the complex coefficients we use to determine phase

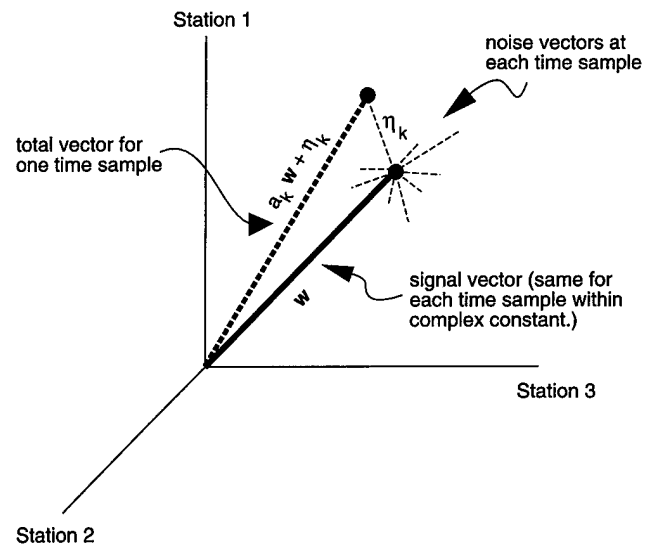


Figure A2. Averaging of the station data through time allows the influence of the noise to be mitigated. For the averaged data, $\mathbf{u}_1$ should point roughly along $\mathbf{w}$. When only using data from time t_k , $\mathbf{u}_1$ will point in the direction $c_k \mathbf{w} + \boldsymbol{\eta}(t_k)$.

shifts. As a result, some bias is inevitable as we are forced by the uncertainty principle to accept some smoothing in time. In our case, this is defined by two time scales: (1) the length of the analyzing wavelet and (2) the window used in the principal-component analysis. This is, in fact, a major strength of this approach as the multi-wavelets are the most compact functions possible with a specified frequency bandwidth.

Department of Geological Sciences
Indiana University
1001 East 10th Street
Bloomington, Indiana 47405

Manuscript received 11 July 1998

AIR FORCE OFFICE OF SCIENTIFIC
RESEARCH (AFOSR)
NOTICE OF TRANSMITTAL TO DTIC. THIS
TECHNICAL REPORT HAS BEEN REVIEWED
AND IS APPROVED FOR PUBLIC RELEASE
IWA AFR 190-12. DISTRIBUTION IS
UNLIMITED.
YONNE MASON
STINFO PROGRAM MANAGER