Geophys. J. Int. (2000) **141,** 820–828

Noise reduction for broad-band, three-component seismograms using data-adaptive polarization filters

Zhijun Du,[1,*] G. R. Foulger[1] and Weijian Mao[2]
[1] *Department of Geological Sciences, University of Durham, Durham,* DH1 3LE, *UK*
[2] *Western Geophysical, Isleworth, Middlesex,* TW7 5AB, *UK*

Accepted 2000 March 5. Received 2000 February 21; in original form 1999 September 28

SUMMARY

We develop a data-adaptive polarization filter that can spectacularly reduce microseismic noise contamination in three-component broad-band seismograms. The filter uses a multitaper spectral analysis method for computing the data spectral density matrix, which is defined as an ensemble average of outer products of the spectrum and its Hermitian adjoint. Under the assumption that strong noise in three-component, broad-band seismograms is additive white noise, and that its spectral density can be determined from seismogram segments without signals, that is, a pre-signal arrival time window, we construct a data-adaptive filter from a spectral density matrix that has been decontaminated of noise. Since the noise corrupting the seismograms is complicated and stochastic, the resulting residual due to the real, non-stationary nature of microseismic noise can leave small-amplitude, quasi-sinusoidal, background oscillations after filtering. These oscillations can be removed by subsequent application of an optimum Wiener filter. Application of the filter to synthetic data with real noise superimposed suppresses the noise by about three orders of magnitude at the expense of less than 5 per cent corruption of the original seismogram in amplitude. Application to several real recordings of teleseismic earthquakes on a three-component broad-band seismic station in Iceland shows that excellent signal-to-noise recovery is possible, rendering such data usable for both arrival time and waveform analysis. This technique may potentially increase by an order of magnitude the volume of usable data collected in seismic experiments in noisy environments, for example, on oceanic islands.

Key words: data-adaptive filter, microseismic noise, optimum filter, polarization, spectrum.

1 INTRODUCTION

When using broad-band seismic data, it is common practice to apply a bandpass (e.g. Butterworth) filter to attenuate low- and high-frequency noise to improve the signal-to-noise ratio. However, this can be successful only for a few cases. In general, background noise has a frequency band that overlaps with that of the seismic signals. Superimposed on the background noise is signal-generated noise. This is noise that is the result of multiple reflections and refractions of *P* and *S* body waves at crustal interfaces and inhomogeneities, and local conversion of body waves to surface waves.

Often it is difficult to give an objective mathematical definition of the characteristics of noise. In most instances, the noise is in some sense a random and stochastic process. Considering patterns of particle motion, seismic signal polarization properties have been used to detect seismic signals in noisy seismograms. Montalbetti & Kanasewich (1970) designed a time-domain polarization filter based on the measurement of rectilinearity and the direction of particle motion from three-component teleseismic recordings using correlation techniques. Their non-linear polarization filter is useful for enhancing *P*, *pP*, *sP*, *S* and *Ps* and other compressional or shear phases. Unfortunately, this technique implicitly assumes that the waveforms have essentially the same polarization over all or most frequencies. In reality, seismograms are superpositions of different types of waves and there is no guarantee that the polarization or phase is constant in frequency. Samson (1977, 1983a,b) describes a method of estimating particle polarization as a function of frequency. For processing multichannel geophysical data, Samson & Olson (1981) have designed a frequency-domain data-adaptive polarization filter. This filter can be used to enhance the waveforms of pure-state data of arbitrary shape and requires little prior knowledge of the

* Now at: Institute of Theoretical Geophysics, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK.
E-mail: Zhijun.Du@itg.cam.ac.uk

spectral content or temporal features of the signal. However, in the presence of strong background noise, this filter is not effective in revealing seismic signals embedded in random noise from three-component seismograms (Fig. 1). A detailed analysis of frequency-dependent polarization was given by Park *et al.* (1987) using three-component seismic data. By applying a series of prolate, spheroidal functions as tapers, they developed a multitaper spectral analysis method. The method is suitable for the analysis of non-stationary processes such as seismic waveforms. Recently, Reading (1996) reported a successful application of the method in picking converted waves from broad-band three-component seismograms collected in New Zealand. Multitaper analysis not only enables well-constrained, smoothed spectral estimation with very low variance, but also provides excellent spectral leakage resistance.

In this paper, we incorporate the multitaper spectral method of Park *et al.* (1987) into the algorithm for the estimation of the spectral density matrix in order to derive a data-adaptive polarization filter. Samson (1983a,b) suggested an approach for estimating the unknown background noise that can be used to improve the polarization filter of Samson & Olson (1981). We assume that the noise in three-component seismograms is a random, stationary process that has no correlation with the signal, although the noise is anisotropic and has polarizations that mix with the signal at some frequencies. We derive a data-adaptive filter using a spectral matrix that has been 'cleaned' of the noise using the approach of Samson (1983a,b). By processing three-component seismic data, we illustrate the procedure whereby such a data-adaptive polarization filter can be constructed. As a test, a synthetic data example is shown. A number of examples of the extraction of seismic signals from noisy data recorded in Iceland are also given.

2 METHOD

According to Samson & Olson (1981), a polarization filter can be expressed analytically as

$$\mathbf{y}(t) = \frac{1}{2\pi T} \int_{-\infty}^{\infty} P^g(\omega)\mathbf{z}(\omega) \exp(i\omega t)\, d\omega\,, \tag{1}$$

where the exponent g is a positive number that can be increased to enhance the rejection of unpolarized noise. $\mathbf{z}(\omega)$ is the spectral representation of $\mathbf{x}(t)$, the data vector of the time-series. Suppose we have three-component seismograms of the form

$$\mathbf{x}(t) = (x^{(1)}(t), x^{(2)}(t), x^{(3)}(t))\,, \quad t = 0, N-1\,, \tag{2}$$

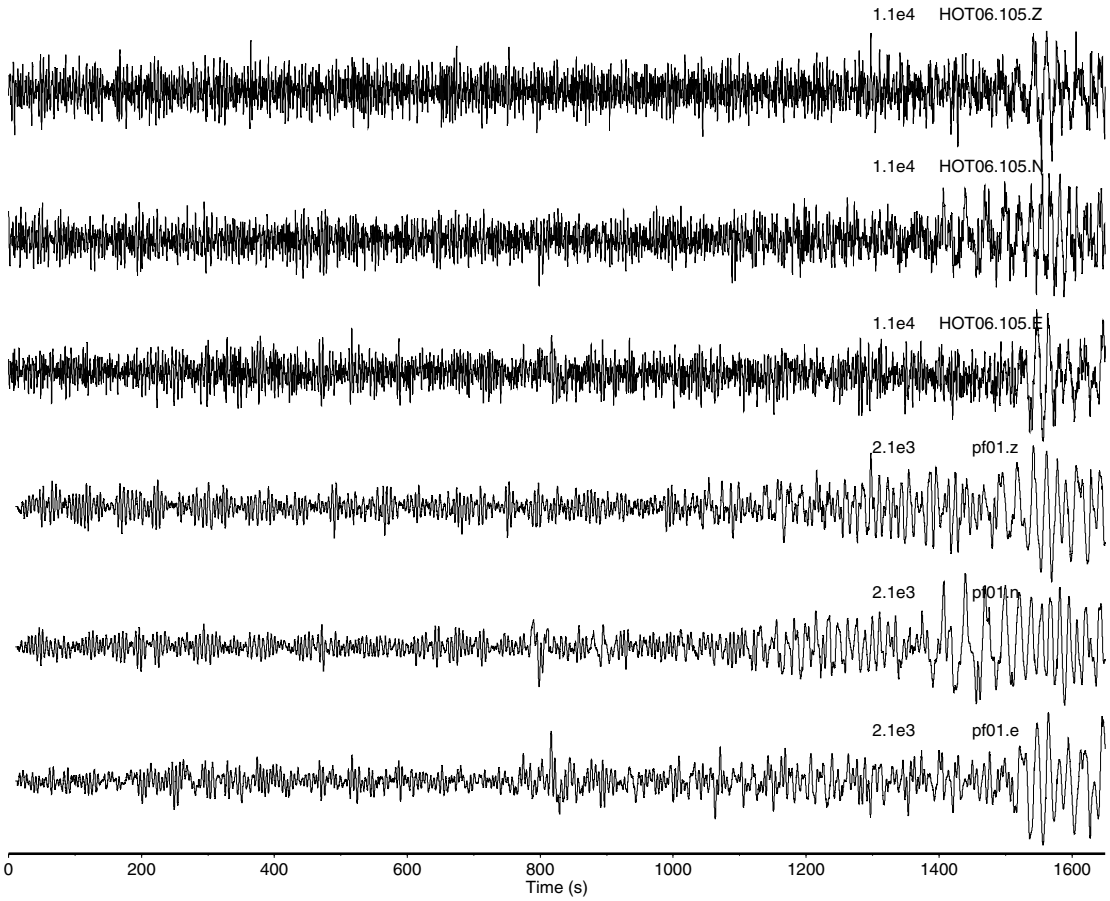

Figure 1. The use of data-adaptive polarization filters suggested by Samson & Olson (1981) to filter three-component seismograms of the Southern Xinjiang earthquake, China, $\Delta = 55.6^\circ$, $M_s = 5.8$, recorded at station HOT06 in Iceland. The top three traces are three-component broad-band seismograms in the order Z, N and E. The bottom three seismograms have been filtered with a sliding polarization filter, with length 65 s [$v = 6$ and $g = 6$; v was defined by Samson & Olson (1981)—see the Appendix for an explanation]. The seismograms are not corrected for instrument response, and so the peak-to-peak amplitude is given in arbitrary units related to the voltage output of the seismometer ('digital counts').

where $x^{(1)}(t)$ is the vertical component, positive upwards, and the coordinate system is right-handed. In eq. (1), $P(\omega)$ has a value of $0 \le P \le 1$ and is the estimation of the degree of polarization,

$$P = \frac{n(Tr\mathbf{S}^2) - (Tr\mathbf{S})^2}{(n-1)(Tr\mathbf{S})^2}, \tag{3}$$

where $\mathbf{S}$ is the spectral density matrix and Tr denotes the trace. It is clear that $P(\omega)$ is rotationally invariant. If $\mathbf{S} = \sigma\mathbf{I}$, where $\mathbf{I}$ for the three-component seismic data is the 3×3 identity matrix, and σ is real and non-negative, then $P = 0$. This represents the most random or unpolarized state.

We wish to estimate $\mathbf{S}(\omega) = E\{\mathbf{z}(\omega)\mathbf{z}^{\dagger}(\omega)\}$, where E denotes the expectation operator and † denotes the Hermitian adjoint, directly from the time-series $\mathbf{x}(t)$. For a chosen frequency band of non-trivial width $2W$, we use a set of K prolate spheroidal wavefunctions (i.e. eigentapers) $w^k(N, W)$, $k = 0, 1, \ldots, K-1$ (Park *et al.* 1987):

$$\mathbf{S}(\omega) = \frac{1}{K}\,\mathbf{M}^{\dagger}(\omega) \cdot \mathbf{M}(\omega), \tag{4}$$

where

$$M(\omega) = \begin{vmatrix} z_0^{(1)}(\omega) & z_0^{(2)}(\omega) & z_0^{(3)}(\omega) \\ z_1^{(1)}(\omega) & z_1^{(2)}(\omega) & z_1^{(3)}(\omega) \\ \vdots & \vdots & \vdots \\ z_{K-1}^{(1)}(\omega) & z_{K-1}^{(2)}(\omega) & z_{K-1}^{(3)}(\omega) \end{vmatrix}, \tag{5}$$

where

$$\mathbf{z}(\omega) = (2\pi)^{-1} \sum_{t=0}^{N-1} w_t^k(N, W)\mathbf{x}(t)\exp(-i\omega t). \tag{6}$$

The single-taper estimation of $\mathbf{S}$ used by Samson & Olson (1981) has a relatively large variance, increasing as a larger fraction of the data is discarded where cosine or boxcar tapers were used. To counteract this, Samson & Olson (1981) smoothed their single tapers by applying a moving average to $\mathbf{S}$. This reduces the variance but results in a short-range loss of frequency resolution and therefore an increase in bias in the estimation of $\mathbf{S}$. The shorter the time-series, the greater the bias in the estimation of $\mathbf{S}$.

In eq. (6) the data are multiplied by several leakage-resistant tapers (Park *et al.* 1987). This yields several tapered time-series from one record. $\mathbf{z}(\omega)$ is formed by summing several 'eigenspectra' from a Fourier transform of each of these time-series. Therefore, $\mathbf{S}$ has a very smooth spectrum, which is the sum of the eigenspectra of the three components of motion (see eq. 4).

The matrix $\mathbf{S}$ is a non-negative Hermitian matrix that can be expanded in the form (Samson & Olson 1981)

$$\mathbf{S} = \sum_{j=1}^{3} \lambda_j \mathbf{u}_j \mathbf{u}_j^{\dagger}, \tag{7}$$

where the λ_j are the eigenvalues of $\mathbf{S}$ and $\mathbf{u}_j$ are the eigenvectors. If $\mathbf{S}$ has only one non-zero eigenvalue at ω (i.e. $\lambda_1 \neq 0$, $\lambda_2 =$, $\lambda_3 = 0$), then

$$\mathbf{S} = \lambda_1 \mathbf{u}_1 \mathbf{u}_1^{\dagger}. \tag{8}$$

This represents the completely polarized, or pure, state. From eq. (3), for the pure state, P has a maximum value of $P = 1$. Note that since $\mathbf{u}_1$ has complex components, this representation also allows for the analysis of elliptically polarized waves.

In the presence of strong background noise, a pure-state polarization may not be seen at any frequency nor be expected at some frequencies. Study of the degree of polarization using eq. (7) shows that when the eigenvector associated with the largest eigenvalue does not point in the direction of the signal polarization, the polarization filter (eq. 1) loses its ability to discriminate between the pure state and noise. The reason for this failure is that the noise may assume large values of P (given by eq. 3), even in a window prior to the arrival of the signal where only polarized noise is present. In order to isolate signal frequencies embedded in a spectrum of random white noise, an effective determination of $\mathbf{S}$ is required, so that P selectively enhances only the polarized seismic signals.

A general model for $\mathbf{z}$, the spectral representation of the observation $\mathbf{x}$(t), in the presence of noise, according to Samson (1983a), can be written as

$$\mathbf{z} = a\mathbf{v} + \mathbf{e}, \tag{9}$$

where a is a frequency-dependent complex scalar. $\mathbf{v}$ remains constant and denotes the direction of the signal polarization, and $\mathbf{e}$ is the noise spectrum. The physics behind eq. (9) assumes that seismic signal is in the pure state, but seismic recordings also contain white noise.

If $\mathbf{e}$ represents additive random noise, which has no correlation with the signal, the spectral density matrix $\mathbf{S}$ can be rewritten as

$$\mathbf{S} = E\{a^2\}\mathbf{v}\mathbf{v}^{\dagger} + E\{\mathbf{e}\mathbf{e}^{\dagger}\}. \tag{10}$$

On the right-hand side of eq. (10), the first term is the optimal estimation of the pure state described by eq. (8), and the second term is the noise spectral density matrix,

$$\mathbf{N}(\omega) = E\{\mathbf{e}(\omega)\mathbf{e}^{\dagger}(\omega)\}, \tag{11}$$

where $\mathbf{N}(\omega)$ is a non-singular Hermitian matrix. The determination of $\mathbf{N}(\omega)$ is difficult since almost nothing is known about $\mathbf{e}$. However, if a theoretical model of $\mathbf{N}(\omega)$ is available, similar procedures as are used to whiten noise can be applied to 'clean' $\mathbf{S}(\omega)$. Following Samson (1983a), we reconstruct $\mathbf{S}(\omega)$ by $\mathbf{N}^{-1/2}(\omega)\mathbf{z}(\omega)$ instead of using $\mathbf{z}(\omega)$. We obtain

$$\mathbf{A}(\omega) = \mathbf{N}^{-1/2}(\omega)\mathbf{S}(\omega)\mathbf{N}^{-1/2}(\omega). \tag{12}$$

From eq. (10), we find that $\mathbf{A}$ satisfies $P(\mathbf{A}) = 0$ (eq. 3) if only noise is present ($\mathbf{S} = \mathbf{e}\mathbf{e}^{\dagger}$), and $P(\mathbf{A}) = 1$ if $\mathbf{S} = \{a^2\}\mathbf{v}\mathbf{v}^{\dagger}$ [i.e. $\mathbf{A} = (\mathbf{N}^{-1/2}\mathbf{v})(\mathbf{v}^{\dagger}\mathbf{N}^{-1/2})$].

In general, the spectra of $P(\mathbf{S})$ and $P(\mathbf{A})$ are different, although they have the same extremes as described above. This is because

$$Tr\mathbf{S}(\omega) = E\{\mathbf{z}(\omega)\mathbf{z}^{\dagger}(\omega)\}, \tag{13}$$

but

$$Tr\mathbf{A}(\omega) = E\{\mathbf{z}^{\dagger}(\omega)\mathbf{N}^{-1}(\omega)\mathbf{z}(\omega)\}. \tag{14}$$

In this way, the spectrum of $P(\mathbf{A})$ is constructed, which has dominant signal power for the polarized signals.

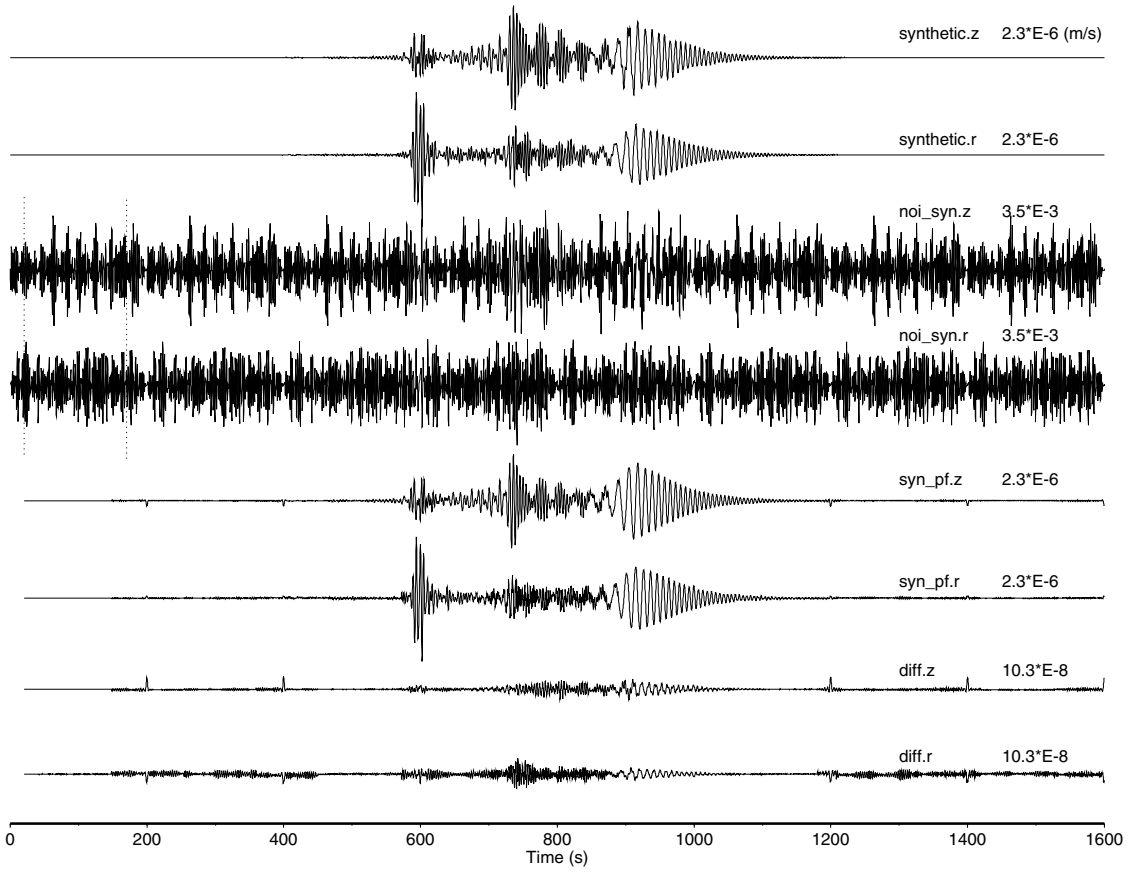

Figure 2. The use of data-adaptive polarization filters to extract synthetic seismic signals from real microseismic noise. Top pair of traces: synthetic radial and vertical seismograms generated by a point double-couple source 25° away. The synthetic seismograms are scaled to a moment of $M_0 = 6.5 \times 10^{23}$ N m. Second pair of traces: the seismograms have been corrupted by sequentially adding real microseismic noise taken from a pre-signal window 200 s long prior to an $Ms = 6.0$ teleseismic event at an epicentral distance of $\Delta = 59°$. Third pair of traces: seismic signals extracted from the second pair of traces using the data-adaptive polarization filtering. Bottom pair of traces: the original minus the recovered seismograms (top pair of traces minus third pair). Numbers indicate maximum peak-to-peak amplitude on each trace. Note the different scales used. The lengths of the sliding and noise windows are 150 s, $K = 4$ and $g = 6$. The dotted lines indicate the noise window.

3 DATA EXAMPLES

In analysing the data presented in the examples given here, we used a discrete representation of eq. (1). Also, for a long data series, sliding-window filters have to be used. The procedures for discretization are discussed in detail in the Appendix. In practice, three-component seismograms can be filtered first to extract the linearly polarized waves (e.g. *P*, *S* and their reverberations *pP*, *sP* and *Ps* etc.), then filtered to enhance waves of specific polarization (e.g. Love wave) and to extract elliptically polarized waves (e.g. Rayleigh wave) by moving the sliding window down along the traces.

3.1 Synthetic data

Fig. 2 shows the application of data-adaptive polarization filtering to synthetic signals embedded in real noise. The seismograms were calculated for a continental-type, layered structure using the modal summation technique (Panza 1985), with an upper frequency of 1.0 Hz (upper two traces). The noise is taken from a pre-signal 200-s-long time window prior to a teleseismic event ($M_s = 6.0$, $\Delta = 69.4°$) (Fig. 3) recorded by the broad-band station HOT06, located in the northwest fjords, Iceland, operated as part of the Iceland Hotspot Project (e.g. Du & Foulger 1999). The second two traces of Fig. 2 show the seismograms, which have been contaminated by sequentially adding the noise to the top two traces. The third pair of traces of Fig. 2 show the seismograms recovered by applying the filters to the seismograms shown in the middle two traces. The differences between the original and recovered seismograms are shown in the bottom two traces.

Comparing the second two traces to the third pair shows that an excellent signal enhancement is achieved. The noise has been reduced by over three orders of magnitude. The distortion of the original seismogram is less than 5 per cent in amplitude, as shown in the bottom two traces, which illustrate the differences between the original and the recovered seismograms. The amplitude changes most probably arise because the filter gain is time- and frequency-dependent. The signal-generated noise, the least polarized waves, has also been removed. However, filtering inevitably distorts the signal. For example, wave polarities and onset times can be quite severely affected by the filter properties of seismic recording systems (Scherbaum 1996). Our retrieval of ground motion from digital seismograms is based on the hopeful assumption at the practical level that we can recover the signal by applying additional

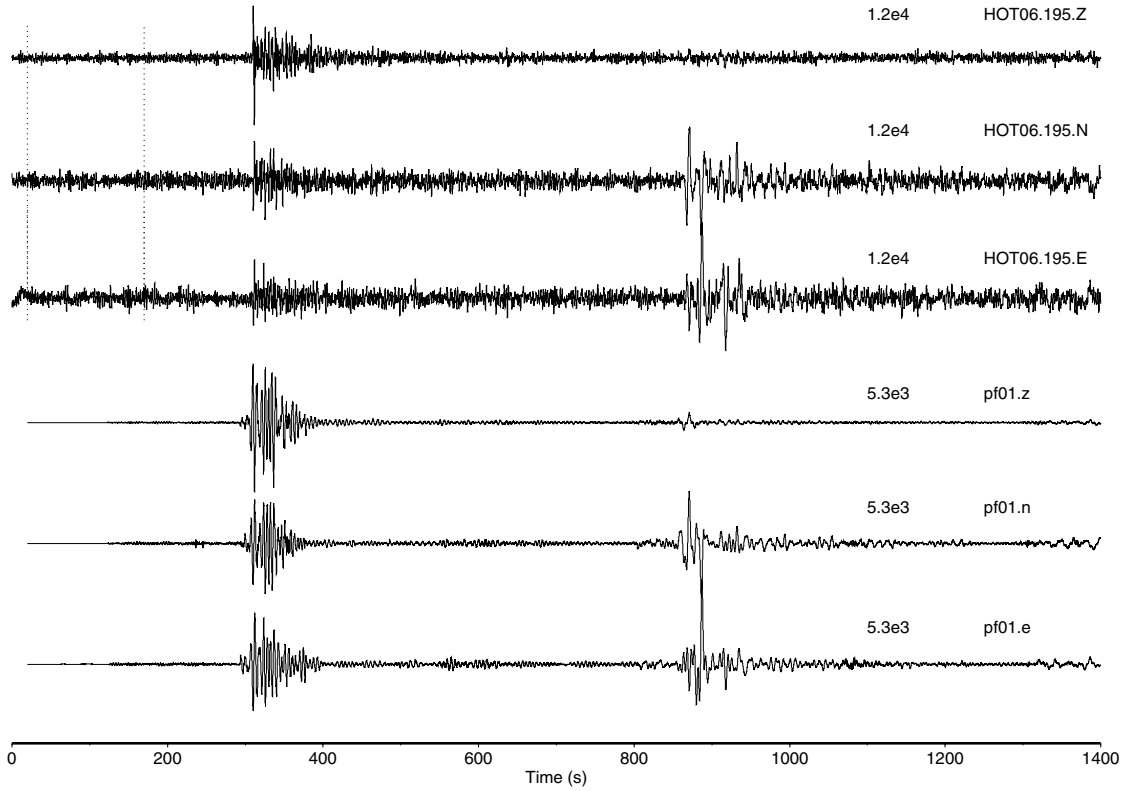

Figure 3. Top three traces are three-component broad-band seismograms in the order Z, N and E for the event of 97195001 in the Kuril Islands, $M_s = 6.0$ and $\Delta = 69.4^\circ$, recorded at station HOT06 in Iceland. The bottom three traces have been filtered with a sliding polarization filter (lengths of sliding and noise windows = 150 s, $K = 4$ and $g = 4$). The dotted lines indicate the noise window. The peak-to-peak amplitude is given in arbitrary units as in Fig. 1

'appropriate' filtering. In fact, seismic signals are not band-limited. At the very least, the less filtering that is applied to extract signals the better the signals are preserved. In view of these inevitable real signal distortions, the differences between the filtered and 'real' seismograms shown in Fig. 2 are really very small. Nevertheless, the wave polarities and onsets can be confidently identified in the data after filtering, whereas this was not possible before filtering nor using other causal or optimum filters.

In practice we used a sliding window 150 s long, and the noise spectral density matrix, $\mathbf{N}(\omega)$, was computed from a pre-signal time window of the same length as the sliding window, in the time interval 20–170 s. The filter parameters, K (the total number of prolate spheroidal eigentapers) and g (a positive number used to enhance the rejection of unpolarized noise; see Samson & Olson 1981), were 4 and 6 respectively.

3.2 Real data

During the two-year period 1996–1998, a 35-station network of broad-band seismic stations was operated over the whole of Iceland as part of the Iceland Hotspot Project (e.g. Du & Foulger 1999). The objective of the network deployment was to collect a large set of seismic recordings with which to study the structure of the crust and upper mantle beneath the Iceland hotspot. More than 2400 $M_s \geq 5.5$ events that occurred worldwide were recorded. However, the oceanic location of Iceland means that oceanic microseismic noise is strong on the seismograms, and we embarked upon the present noise-reduction study in order to increase the amount of data usable from this massive data set. We illustrate the efficacy of the filter for processing three-component broad-band data using the following two examples. Fig. 3 shows a relatively high-quality earthquake, recorded at station HOT06 in Iceland, in the Kuril islands, 1997 July 14, $M_s = 6.0$. The epicentral distance was $\Delta = 69.4^\circ$. In applying the filter we used a 150-s-long sliding window, and selected a noise window prior to the arrival of the signal, in the time interval 20–170 s. The filter parameters used were $K = 4$ and $g = 4$.

The second data example used (Fig. 4) is the three-component seismograms recorded for the event of 1997 April 15 in the Southern Xinjiang, China, $\Delta = 55.6^\circ$, $M_s = 5.8$, recorded at station HOT06. These seismograms are typical of noisy recordings. The filter parameters used for this event were $K = 6$ and $g = 6$, and the length of the sliding window was 65 s. The noise window was 130 s. Compared to Fig. 1, which shows the application of the filter of Samson & Olson (1981) to the same data, the filter used here demonstrates far better signal-enhancing ability for suppressing noise.

Since the noise corrupting the seismograms is a complicated stochastic process, our assumption that the process is stationary can only be regarded as a first-order approximation. We have found that the length of the pre-signal noise window must be 1–3 times that of the sliding window. In order to characterize better the noise spectrum for very non-stationary, noisy data, it is necessary to use an even longer noise window.

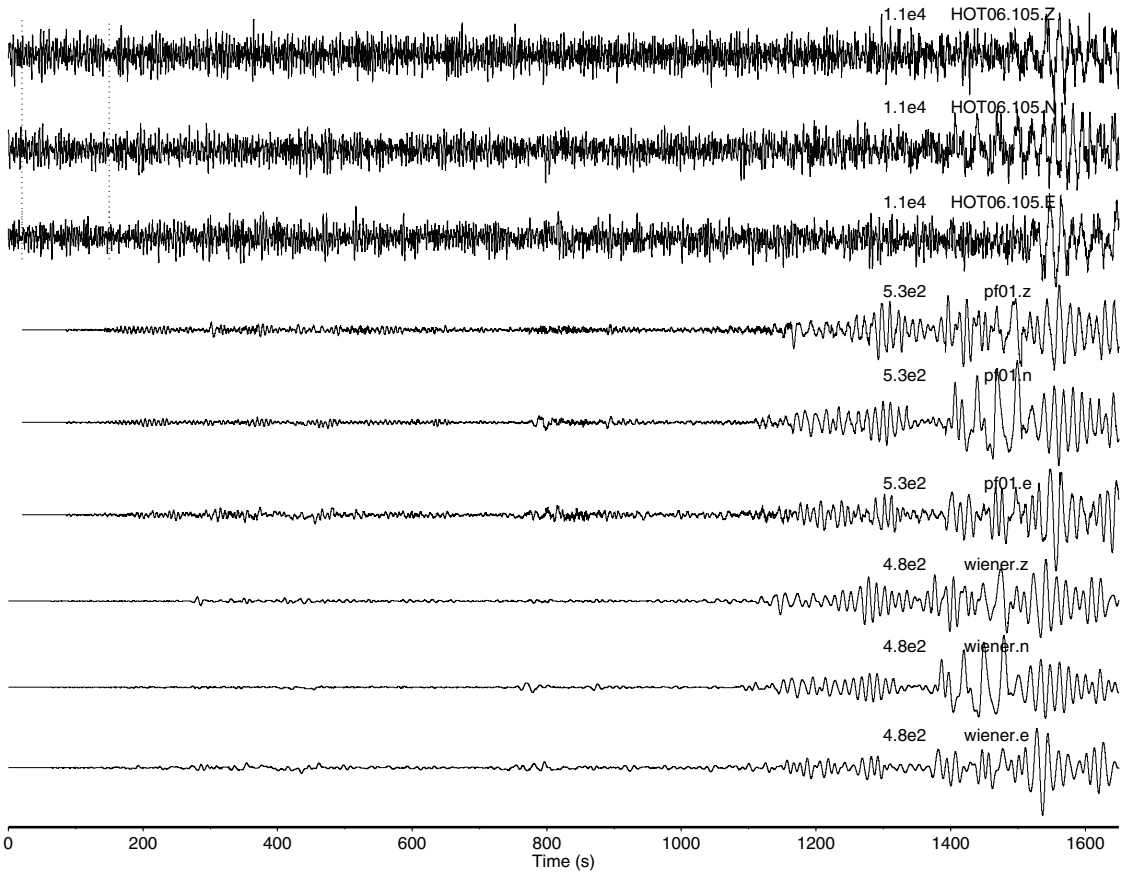

Figure 4. Top three traces are three-component broad-band seismograms in the order Z, N and E for the same event as shown in Fig. 1. The middle three traces have been filtered with a sliding polarization filter (length of sliding window = 65 s, length of noise window = 130 s, $K = 6$ and $g = 6$). The dotted lines indicate the noise window. The bottom three traces are the results of the optimum filtering. The predominant period of the noise is assumed to be 6 s.

The effective length of the sliding window is data-dependent and is related to the length of the signal. More trials may be needed before an optimal length is found, if the signal is obscured by very strong noise. The polarization filter enhances any wavelet that shows some polarization, even if the wavelet is linearly polarized with all the power in one channel. If a longer sliding window is used, a problem may arise in that anisotropic noise may be highly polarized over several narrow frequency bands where signal polarization is present. This can consequently increase the difficulties in detecting the signal. The safest procedure is to start with a long sliding window and then gradually decrease its length. In practice, it is usually not difficult to find an optimal length for the sliding window that enables the filter to adapt to the different frequency contents of the polarized signals.

3.3 Further reducing noise with the optimum filter

Figs 5 and 6 are two additional data examples. Although the signals are clearly visible after polarization filtering, the onsets and first motions are nevertheless still difficult to pick confidently. Note in particular the small-amplitude, quasi-sinusoidal oscillations that occur for the duration of the records. These oscillations obscure the wave onsets. The use of causal filters to reduce the noise further may distort the polarity onsets (Douglas 1997). To suppress the noise further we therefore apply an optimum filter, that is, an extended version of the optimum Wiener filter, as suggested by Douglas (1997). We construct a noise model with the dominant period the same as that of the main noise component, that is, the autocorrelation computed for a pure sinusoid, and we construct a signal model using the autocorrelation function of the impulse response of our recording system. As shown in the bottom three traces of Figs 5 and 6, the optimum filter removes the quasi-sinusoidal noise, as it is designed to do, but preserves the main features of the signals. The bottom three traces of Fig. 4 shows the results of applying optimum filtering to its middle three traces.

4 DISCUSSION AND CONCLUSIONS

The examples shown above indicate the efficiency of the polarization filtering, which can be used to extract useful signals from seismograms that are so noisy they would be unusable otherwise. The polarization filter can be used to enhance selectively seismic waveforms of arbitrary shapes. Even some impulsive wavelets with discontinuities can be successfully filtered with little distortion, since the filter is data adaptive

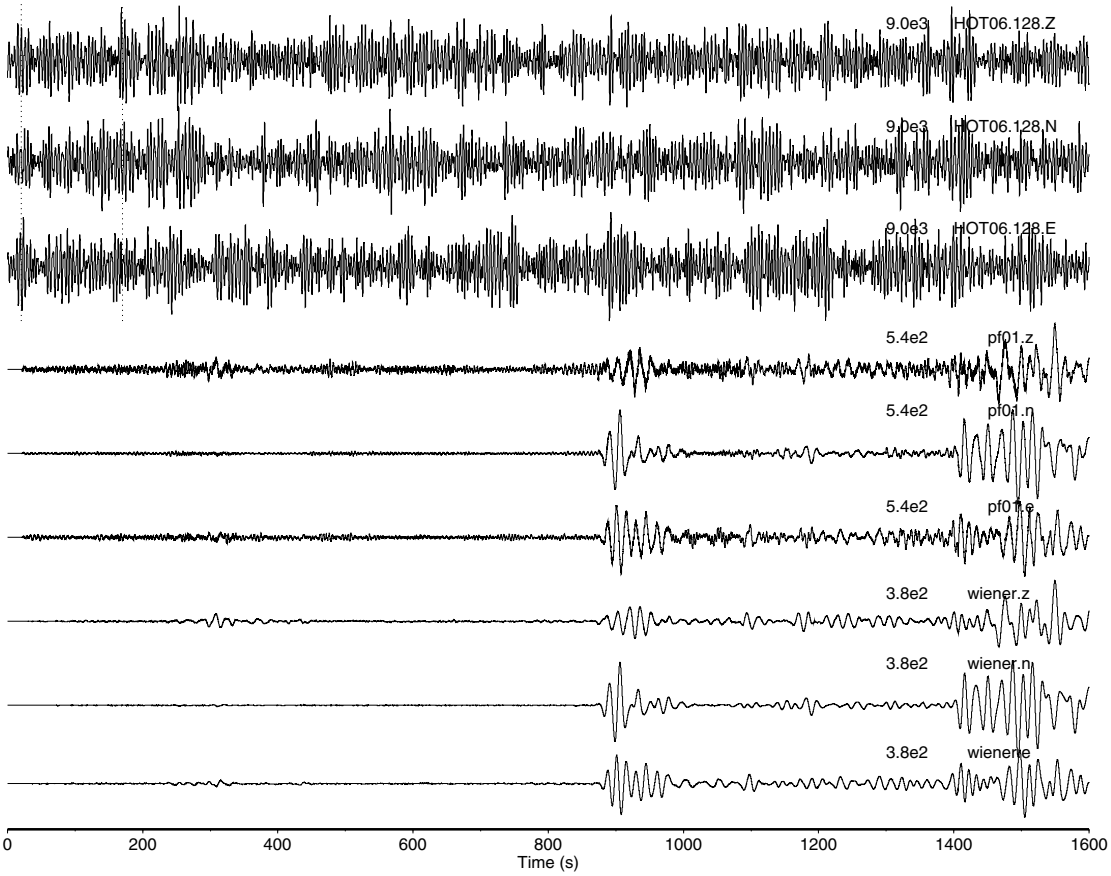

Figure 5. Same as Fig. 4 except for the event of 97128001, in the border region of India and Bangladesh, $M_s = 5.7$ and $\Delta = 74.2°$. The middle three traces have been filtered with a sliding polarization filter (length of sliding window = 150 s, length of noise window = 150 s, $K = 4$ and $g = 6$). The bottom three traces are the results of the optimum filtering. The predominant period of the noise is assumed to be 5 s.

and the seismic signal does not need to be narrow band. The polarization filter is computationally very efficient because it relies on the scalar invariants of the spectral matrix. We only need to compute an inverse matrix (see eq. 14) at the beginning of each application. There is no requirement for large amounts of computer storage because we segment the seismograms into overlapping sliding windows. To filter three-component seismograms with $3 \times 18\,000$ points takes no more than 15–20 s.

The design of a data-adaptive polarization filter requires the noise to be defined by parameters that allow some means of separating the noise from the signal. In this paper, we consider three-component broad-band seismograms where we define noise as additive random noise, and assume that its spectral properties can be determined from the part of the seismograms where signals are apparently absent. Although some of the data show quasi-sinusoidal wave trains after polarization filtering, this can be remedied by applying another optimum (Wiener) filter (Douglas 1997).

Non-stationary noise is responsible for the quasi-sinusoidal oscillations in the polarization-filtered data. The sample biases in filter parameters, which were determined from a finite number of samples, may also cause such oscillations, as discussed by Samson (1983b). Samson (1983b) has suggested using asymptotic expansions to correct for sample biases. We did not find, however, that asymptotic correction significantly improves the final results.

We are currently using the method described in this paper for processing the data of the Iceland Hotspot Project. Since polarization filtering can preserve better signal polarities and onsets with trivial amplitude distortions, we are also extending the method to seismic source spectrum studies. One area where polarization filtering is particularly useful for us is in filtering waveforms (*S* and Rayleigh waves) to improve them for use in waveform tomography (e.g. Du & Panza 1999) to determine the structure of the upper mantle beneath Iceland. Application of the method can be extended further for processing other kinds of geophysical data (e.g. seismic exploration and geomagnetic data) to recover signals from strong background noise.

ACKNOWLEDGMENTS

We thank Bruce Julian for useful discussions. We appreciate helpful suggestions and reviews from David Gubbins and Neil D. Selby, which helped to improve the manuscript. We acknowledge Alan Douglas for providing us with suggestions for using the optimal Wiener filter. This research was funded by Natural Environment Research Council (NERC) grants GST/02/1238 and GR3/10727 and a University of Durham Special Projects Grant. WM benefited from a contract No. RAD 23(90) of Amoco (UK) when he worked in the University of Leeds.

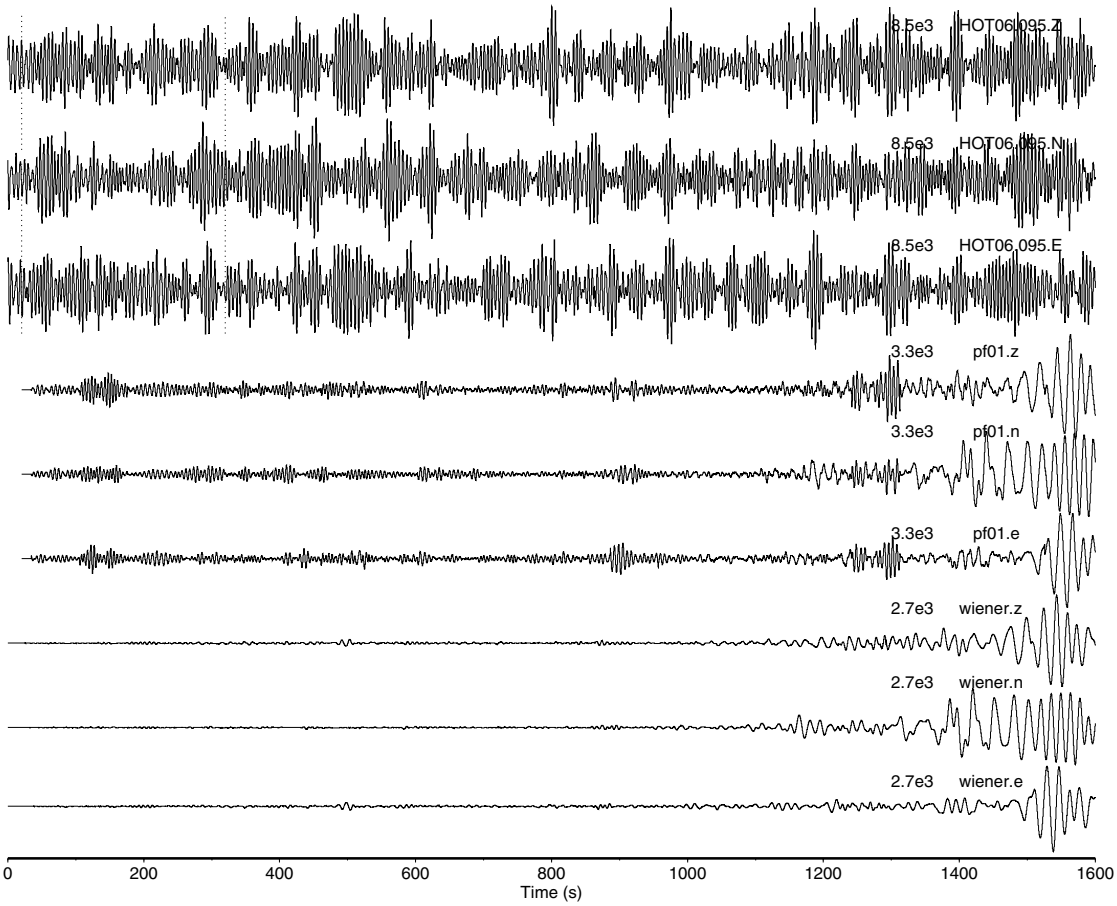

Figure 6. Same as Fig. 4 except for the event 97095002, in the Southern Xinjiang, China, $M_s = 5.6$ and $\Delta = 57^\circ$. The middle three traces have been filtered with a sliding polarization filter (length of sliding window = 100 s, length of noise window = 300 s, $K = 4$ and $g = 6$). The bottom three traces are the results of the optimum filtering. The predominant period of the noise is assumed to be 5 s.

REFERENCES

Douglas, A., 1997. Bandpass filtering to reduce noise on seismograms: is there a better way?, *Bull. seism. Soc. Am.,* **87,** 770–777.

Du, Z.J. & Foulger, G.R., 1999 The crustal structure beneath the northwest fjords, Iceland, from receiver functions and surface waves, *Geophys. J. Int.,* **139,** 419–432.

Du, Z.J. & Panza, G.F., 1999. Amplitude and phase differentiation of synthetic seismograms: a must for waveform inversion at regional scale, *Geophys. J. Int.,* **136,** 83–98.

Montalbetti, J.F. & Kanasewich, E.R., 1970. Enhancement of teleseismic body phases with a polarization filter, *Geophys. J. R. astr. Soc.,* **21,** 119–129.

Panza, G.F., 1985. Synthetic seismograms: the Rayleigh waves modal summation, *J. Geophys.,* **58,** 125–145.

Park, J., Vernon III, F.L. & Lindberg, C.R., 1987. Frequency dependent polarization analysis of high-frequency seismograms, *J. geophys. Res.,* **92,** 12 664–12 674.

Reading, A.M., 1996. Deep lithospheric structure from multi-phase seismic tomography: the subduction zone beneath southern North Island, New Zealand, *PhD thesis,* University of Leeds.

Samson, J.C., 1977. Matrix and stokes vector representations of detectors for polarized waveforms: theory, with some applications to teleseismic waves, *Geophys. J. R. astr. Soc.,* **51,** 583–603.

Samson, J.C., 1983a. Pure states, polarized waves, and principal components in the spectra of multiple, geophysical time-series, *Geophys. J. R. astr. Soc.,* **72,** 647–664.

Samson, J.C., 1983b. The reduction of sample-bias in polarization estimators for multichannel geophysical data with anisotropic noise, *Geophys. J. R. astr. Soc.,* **75,** 289–308.

Samson, J.C. & Olson, J.V., 1981. Data-adaptive polarization filters for multichannel geophysical data, *Geophysics,* **46,** 1423–1431.

Scherbaum, F., 1996. *Of Poles And Zeros,* Kluwer Academic, Dordrecht.

APPENDIX A: THE DISCRETE REPRESENTATION OF THE POLARIZATION FILTER

According to Samson and Olson (1981), the discrete representation of eq. (1) can be written as

$$\mathbf{y}(l) = \frac{1}{N} \sum_{j=0}^{N-1} \mathbf{z}(j) P^g(j) \exp(2\pi i j l / N), \tag{A1}$$

where $l = 0, 1, \ldots, N-1$ and N is the number of points in the individual time-series or the filter length if a one-step filter is preferred.

In practice, the use of a sliding filter is more attractive. If we segment $\mathbf{x}(t)$ into overlapping windows and filter it section by section, the relationship between the windowed data $\mathbf{x}_s(t)$ and

$\mathbf{x}(t)$ is

$$\mathbf{x}_s(l)=\mathbf{x}[l+(s-1)(n-m+1)]\,, \tag{A2}$$

where $l=0, 1, \ldots, n-1$. n is the length of the window and m gives the amount of window overlap. $s=1, 2, \ldots,$ $(n-1)+(s-1)(n-m+1)\leq N$. N is the length of the time-series $\mathbf{x}(t)$. The filtered $\mathbf{y}_s(l)$ is given by

$$\mathbf{y}_s(l)=\frac{1}{n}\sum_{j=0}^{n-1}\mathbf{z}_s(j)P_s^g(j)\exp(2\pi ijl/n)\,, \tag{A3}$$

where

$$\mathbf{z}_s(j)=\sum_{j=0}^{n-1}w_j^k(n,\,W)\mathbf{x}_s(l)\exp(-2\pi ijl/n) \tag{A4}$$

and

$$\sum_{j=0}^{n-1}|w_j^k(n,\,W)|^2=1\,. \tag{A5}$$

The optimal K value for $w^k(n, W)$ depends on the concentration of spectral energy in the frequency band $2W=2K/n$. Since higher-order tapers have weak leakage resistance and the spectral information discarded by the first two tapers can be retrieved by a few later tapers, a value of $K=3-6$ can be used (Park *et al.* 1987). The spectral smoothing parameter, v, used by Samson & Olson (1981) as the number of degrees of freedom in spectral estimation, is replaced here by K.

The filtered data, $\mathbf{y}(l)$, are given by

$$\mathbf{y}[l+(s-1)(n-m+1)]=\mathbf{y}_s(l)\,. \tag{A6}$$

The amount of overlap M depends on the rate of change of P_s^g and the effective width of $2W$, which relates to the parameter K ($K=nW$ gives the time-bandwidth of the prolate taper that concentrates spectral energy in frequency bands of width $2W$). According to Samson & Olson (1981), a truncation point n_t can be approximated by the criterion

$$\sum_{l=0}^{n_t}|\omega(l)|\Big/\sum_{l=0}^{(n-1)/2}|\omega(l)|=0.8\,, \tag{A7}$$

where

$$\omega(l)=P_E^g\delta_{0l}+(1-P_E^g)\cdot\frac{K}{\sqrt{2}n}\sum_{m=1}^{(n/2-1)}\exp[-\pi/2(l/N-m)^2K^2]\,, \tag{A8}$$

where P_E is the expectation of P_s. Samson & Olson (1981) built the P_E as a function of v (see Fig. 1 in Samson & Olson 1981). The relationship between P_E and K can be obtained by analogy with Samson & Olson (1981).