TWO METHODS OF DECONVOLUTION: POWER SPECTRUM SMOOTHING AND PARSIMONIUS DECONVOLUTION

Bob Godfrey and Jon Claerbout

Abstract

Two independent methods of deconvolving seismic data are presented. The first section gives the theoretical development of each method and the second section follows with some examples.

Power spectrum smoothing consists of estimating the bubble spectrum by dividing a smoothed raw data spectrum by a super-smoothed raw data spectrum. The time domain bubble is computed using the minimum phase factorization of the resultant spectral ratio.

Parsimonius deconvolution is an iterative gradient descent algorithm. A norm measuring the parsimony or spikiness of a trace provides the gradient direction along which descent is made. The optimation process is subject to the constraint that the observed seismic trace is composed of random noise plus the convolution of a reflectivity series with a causal waveform.

Theory of Power Spectrum Smoothing

We wish to decompose the seismic trace y_t, into a bubble b_t and reflectivity series x_t. Each trace is formed by the process illustrated in equation (1) where n_t is noise.

$$y_t = b_t * x_t + n_t \tag{1}$$

If the bubble is considered stationary across a group of traces and if the noise is uncorrelated with reflectivity, then adding trace spectrums over all channels gives

$$\overline{YY} = (\overline{BB})(\overline{XX}) + \overline{NN} \tag{2}$$

where $\overline{YY} = \frac{1}{N} \sum_{ich} (\overline{YY})_{ich}$

$$\overline{XX} = \frac{1}{N} \sum_{ich} (\overline{XX})_{ich}$$

$$\overline{NN} = \frac{1}{N} \sum_{ich} (\overline{NN})_{ich}$$

Capital letters refer to frequency domain. In deriving (2) we used the uncorreltaed noise assumption, namely

$$E\,(NX) = 0$$

where the expectation operator is approximated by a finite sum over channel. Excluding the noise term from (2), (implying high signal-noise ratio), (2) is rewritten

$$\overline{BB} = \frac{\overline{YY}}{\overline{XX}} \tag{3}$$

Utilizing the same type of philosophy used in predictive deconvolution, we want the desired reflectivity to have an autocorrelation which is a heavily smoothed version of the seismic trace autocorrelation. Eq. (3) now becomes

$$\hat{\overline{BB}} = \frac{\overline{YY}}{\langle\langle \overline{YY} \rangle\rangle} \tag{4}$$

where angle brackets denote autocorrelation smoothing and $\hat{\overline{BB}}$ is an estimate of the "true" $\overline{BB}$.

The forward wavelet represents a physical waveform and therefore must be finite in length (causal). The presumed length of b_t is considered a constraint and any algorithm computing b_t must incorporate the constraint. If the denominator in (4) was smoothed so that $<< Y\bar{Y} >> = 1$, then the autocorrelation of b_t would equal that of y_t . Reducing the amount of smoothing reduces the autocorrelation length of b_t . Alternatively, either the numerator or ratio in (4) can be smoothed in order to reduce the length of b_t . Incorporating this feature into (4), we obtain an estimate of the bubble spectrum

$$\hat{B\bar{B}} = \frac{< Y\bar{Y} >}{<< Y\bar{Y} >>} \quad \text{or} \quad \hat{B\bar{B}} = < \frac{Y\bar{Y}}{<< Y\bar{Y} >>} > \tag{5}$$

An alternate way of viewing (5) is contained in the following paragraph.

The average spectrum of raw field data recorded using an air-gun array commonly exhibits sharp notches and peaks. [see figure 3]. These features can be associated with a waveform of duration $1/\Delta f$, where Δf is the frequency separation between adjacent notches or peaks. Deconvolved data should be free of such features yet the same trend between raw and deconvolved data should be retained. Eq. (5) gives an estimate of a source waveform spectrum which when used to deconvolve data will preserve trends but eliminate gross features between raw and deconvolved spectrums.

The complex spectrum, $\hat{B}$ can be computed from (5) using a minimum phase factorization technique and each trace deconvolved using

$$X = \frac{Y}{\hat{B}} \tag{5a}$$

Eq. (5a) is stable at all frequencies since a necessary condition for $\hat{B}$ to be minimum phase is that $|\hat{B}| \neq 0$ on the unit circle.

Parsimonious Deconvolution

Theory for Single Channel

Claerbout, [SEP 13 - Parsimonius Deconvolution] has shown that minimizing a spikiness norm $Sn(x_t)$, of a seismic trace achieved a subjectively good deconvolution, where

$$Sn(x_t) = \ell n \sum_t |x_t|^n - \frac{\sum_t |x_t|^n \ell n |x_t|^n}{\sum_t |x_t|^n} \tag{6}$$

Two constraints imposed on the minimization of $Sn(x_t)$ are (considering a single-channel, noiseless process)

$$\text{i)} \quad y_t = x_t * b_t \tag{6a}$$

ii) b_t is causal or finite

Previous algorithms had problems with stability and the present algorithm was developed to solve this problem. Since we're working with forward wavelets the algorithm had to be formulated in the frequency domain since the computation of x_t , given y_t and a current estimate of b_t is extremely time-consuming if the system of equations in (6) is solved by Levinson recursion. The minimization proceeds as follows.

First, transform (6a) (caps refer to frequency domain)

$$Y = XB \tag{7}$$

Next, recall the total derivative of a function, in this case Y

$$dY = \frac{\partial Y}{\partial X} dX + \frac{\partial Y}{\partial B} dB \tag{8}$$

Using (7) in (8) and setting $dY = 0$

$$\frac{dB}{B} = \frac{dx}{X} \tag{9}$$

Define the gradient g_t , as

$$g_t = \frac{\partial Sn(x_t)}{\partial x_t}$$

We wish to march dX in the direction $-G$, an amount $d\alpha$ since we're trying to minimize Sn . Letting $dX = -d\alpha G$, (9) becomes

$$\frac{dB}{B} = \frac{d\alpha\ G\bar{X}}{X\bar{X}} \tag{10}$$

Equation becomes numerically unstable if $|G|$ remains finite as $|X| \to 0$. To ensure stability (10) is rewritten as

$$\frac{dB}{B} = \frac{d\alpha\ G\bar{X}}{X\bar{X}+\varepsilon} \tag{10a}$$

where $0 < \varepsilon < \infty$

Now we are certain that $\frac{dB}{B} \to 0$ as $|X| \to 0$. To simplify notation, define

$$\bar{X}' = \frac{\bar{X}}{X\bar{X}+\varepsilon}$$

Now (10a) is written compactly as

$$\frac{dB}{B} = d\alpha \; \bar{GX}' \tag{11}$$

A discussion on what values the parameter ε was allowed to have is in the section *Examples*.

Assuming that $\bar{GX}'$ is a slowly varying function of α and that α is small, (11) is integrated to give:

$$\int_{B^k}^{B^{k+1}} \frac{dB}{B} = \int_0^{\alpha} \bar{GX}' \; d\alpha \approx (\alpha \bar{GX}')^k$$

$$\ell n\left(\frac{B^{k+1}}{B^k}\right) = (\alpha \bar{GX}')^k$$

$$B^{k+1} = B^k \exp\left[(\alpha \bar{GX}')^k\right] \tag{12}$$

where k refers to iteration index

Eq. (12) is the essential equation for computing a new bubble B^{k+1}, given some previous estimate B^k.

Causality of b_t must still be enforced in some stable manner. It is advantageous at this point to consider noisy multichannel data since the presence of noise will provide some explicit stabilization in the computation of x_t given b_t and y_t.

Theory for Multichannel Noisy Data

Copying down (1) using ich to denote the channel number we have

$$y_{ich,t} = b_t * x_{ich,t} + n_{ich,t} \tag{13}$$

The bubble b_t is assumed stationary across a common shot gather and the noise $n_{ich,t}$ is assumed random from trace to trace. There are two problems to solve when noise is present on a channel.

First, assuming we have an estimate of b_t, how do we compute x_t? Dropping the subscript ich, temporarily, the classical approach [see Taylor et al, 1977] is to minimize an L-2 error norm, E, defined as

$$E = \sum_t n_t^2 + \lambda \sum_t x_t^2 \tag{14}$$

where λ is a non-negative Lagrange multiplier. Rewriting (14) in the frequency domain, using Rayleigh's theorem gives

$$E = \sum_\omega (N\bar{N})_\omega + \lambda \sum_\omega (X\bar{X})_\omega \tag{15}$$

From (13) we know that

$$\bar{N} = \bar{Y} - \overline{BX} \tag{16}$$

Substituting (16) into (15) and differentiating E ω.r.t. $\bar{X}$, a specific frequency component, we obtain the result

$$X = \frac{\bar{B}Y}{B\bar{B}+\lambda} \tag{17}$$

The stabilizing influence of λ , commonly called the pre-whitening parameter, is evident from (17) since the denominator is positive definite ($\lambda>0$) . An estimate of the noise/signal ratio is used when choosing this parameter. A scheme for determining λ is discussed under *EXAMPLES*.

Replacing the term $G\bar{X}'$ by $\sum_{ich} (G\bar{X}')_{ich}$ in (12) solves the problem of how to compute B^{k+1} from B^k for multichannel noisy data. Since (12) was derived using noiseless data, the replacement of $G\bar{X}'$ by the channel sum must be qualified. If channel noise is uncorrelated with reflectivity, then it can be shown that the above replacement is approximately valid.

Summarizing the results, we have an algorithm for deconvolving multichannel noisy data

$$\text{form B :} \quad B^{x+1} = B^k \exp \alpha \sum_{ich} (G\bar{X}')_{ich} \tag{18}$$

$$\text{deconvolve : channels} \quad X_{ich} = \frac{\bar{B}Y_{ich}}{B\bar{B}+\lambda} \tag{19}$$

By truncating B^k in the time domain after each iteration, causality of b_t is ensured.

EXAMPLES

Parameters Needed

1. Power Spectrum Smoothing

Smoothing the raw data spectrum is done by multiplying a symmetrical triangular window with the autocorrelation of the data. The length of the triangle must be specified and therefore two parameters are required, Kshort= $\frac{1}{2}$ -length for super-smoothing and Klong= $\frac{1}{2}$ -length for either ratio or numerator smoothing. Kshort equals the number of points to the second zero crossing of the autocorrelation function.

2. Parsimonius Deconvolution

A summary of the parameters needed are

n = norm coefficient (6)

ε = perturbation stabilization (10a)

λ = pre-whitening coefficient (17)

α = length to march along gradient (18)

nb = allowed length of bubble

t_0 = starting value for $b(t) = \delta(t-t_0)$

With the exception of the first example, the parameter n was initially set to 2.5 for a prescribed number of iterations and then reset to 2.0. The iteration was halted when either all channels had a convergent norm or when some norms were diverging and others converging. In all cases, the norm was reset from 2.5 to 2.0 when each channel norm (with the exception of one or two channels) had converged.

Two values for ε were used.

$$\text{(i)} \quad \varepsilon_1 = \frac{\underset{\omega}{\text{largest}}(X\bar{X}) - \underset{\omega}{\text{smallest}}(X\bar{X})}{10}$$

$$\text{(ii)} \quad \varepsilon_2 \to \infty$$

The practical implementation of ε_2 consists of recognizing that $(X\bar{X}+\varepsilon_2)\to\varepsilon_2$. Then (10a) is rewritten as

$$\frac{dB}{B} = \frac{d\alpha\ G\bar{X}}{X\bar{X}+\varepsilon_2} \approx \frac{d\alpha\ G\bar{X}}{\varepsilon_2} = (d\alpha)\ G\bar{X}$$

where the constant $d\alpha/\varepsilon_2$ has been reset to $d\alpha$. Empirically it was observed that ε_1 and ε_2 gave similar results, but ε_1 gave quicker convergence. All examples shown use $\varepsilon=\varepsilon_1$.

The parameter, α is chosen so that $|B^{k+1}|$ is allowed a fixed maximum excursion from $|B^k|$ at each iteration. This means

$$\max_{\omega} \left| \frac{|B^{k+1}| - |B^k|}{|B^k|} \right| = \text{fact} \qquad (20)$$

where $0 < \text{fact} < 1$

By substituting (18) into (20) a relation between α and fact can be deduced. Selecting a value for fact fixes α .

A scheme for selecting λ follows. Essentially, we provide some mathematical equivalents for the statement "λ is small compared to $B\bar{B}$" . Using (16) we define noise/signal ratio as

$$r(\omega) \equiv \frac{N\bar{N}}{X\bar{X}} = \frac{(Y-BX)(\bar{Y}-\bar{B}\bar{X})}{X\bar{X}} \qquad (21)$$

Now use (17) to eliminate X from the righthand side of (21), giving

$$r(\omega) = \frac{\lambda^2}{B\bar{B}} \qquad (22)$$

By demanding that a high percentage of $r(\omega)$ be small, say equal to 0.001, λ is determined. In practice, $(1/B\bar{B})$ was sorted and the value corresponding to the index=120 (for 128 data points) or index=490 (for 512 data points) was multiplied by 0.001 to yield λ^2. If a large number of $B\bar{B} \to 0$, however, $\lambda \to 0$ according to this scheme and (19) becomes unstable. Either a lower index value must be chosen or a new scheme used.

Synthetic Data

Five channels of synthetic data, each 128 points long are shown in Fig. 1. By convolving the true wavelet b_t (23 points) with the true reflectivity x_t, the traces y_t were synthesized. The estimated reflectivity and bubble using parsimonius deconvolution (PD) are shown in Fig. 1a and 1b.

Fig. 1a represents 11 iterations with n=2.0, whereas Fig. 1b represents 21 iterations, the first 11 with n=2.5 and the latter 10 with n=2.0. One explanation of the disparity between results is illustrated in Fig. 1c. The gradient is plotted as a function of y for both n=2.5 and n=2.0. The lower norm not only attaches a larger gradient to small data values (which tends to drive them to zero) but attaches a gradient of the same sign to *more* large data values (which tends to increase their magnitude).

Spectrum smoothing deconvolution (SSD) is shown in Figs. 1d and 1e. Both cases used Kshort=6 and Klong=50, however, Fig. 1d used numerator smoothing to implement Klong whereas Fig. 1e used ratio smoothing. One explanation for the poor results is that in SSD, one assumes $\hat{b}_t$ is minimum phase whereas in fact b_t wasn't minimum phase.

Common Shot Gather

Fig. 2a is the same original common shot gather that Claerbout (SEP 13) used. The average amplitude spectrum is shown in Fig. 3a.

Fig. 2b shows the SSD of the gather using Kshort=15 and no numerator smoothing. The areas labeled bubble in Fig. 2a have been largely removed

but a faint resonance parallel to the sea floor is evident. Fig. 3b illustrates the smoothed appearance of the deconvolved average amplitude spectrum.

Fig. 2c shows the PD of the gather using 11 iterations of which the first 6 used n=2.5, followed by 5 with n=2.0. Other parameters are as follows: fact=0.3, nb=200 and t_o=50. Reducing the lag tended to rid the gather of the faint water bottom precurssor evident in Fig. 2c. Again, the areas labeled bubble have been removed but the resonance mentioned above persists. The high frequency present in the bubble (Fig. 3c) from 62.5 Hz onwards is a computational artifact that results because the raw data was absent of energy in these frequencies.

Finally, Fig. 2d is the same gather deconvolved using a technique being studied by Will Gray (see report in SEP 14). All three deconvolved gathers are also shown in variable area format in his report.

All three methods give very similar results and since spectrum smoothing ism't an iterative type algorithm, we conclude that it's the preferred method of deconvolution for this particular data set.

Earthquake Data

The final example is a set of earthquakes originally deconvolved by R. A. Wiggins (1977) using minimum entropy deconvolution (MED). Fig. 4 shows both the original earthquake suite and the result using PD. Ten iterations were used with the first five having n=2.5 and the last five, n=2.0. The deconvolved data is of much lower frequency than the raw data, in common with MED. Wiggins was able to show how a refracted and reflected wave were resolved better on traces 9 and 10 using MED, however, PD hasn't accomplished this.

SSD was also attempted on this data but the results were characteristic of that for the synthetic data and are omitted here.

Conclusion

When parsimonius and smoothed spectrum deconvolution are tried on synthetic data, the former method gives a better result. When the methods are applied to one particular common shot gather, however, the results are

qualitatively very similar

An explanation of the last result is that the field data contained a minimum phase bubble and therefore spectrum smoothing should work well. To test this hypothesis, an air-gun source signature, obtained from Robert Brune of the USGS, was factored and the resultant minimum phase wavelet closely approximated the original signature. This wavelet was then used to construct some synthetic data using a uniform probability density function to determine reflection coefficients and arrival times. Parsimonius deconvolution again performed better than spectrum smoothing although the disparity between results wasn't as pronounced as in earlier synthetics. The effects of using other probability distributions hasn't been studied yet.

Data recorded using a non-minimum-phase wavelet (e.g. maxi-pulse) would provide a good test of the deconvolution methods. Parsimonius deconvolution is favored simply because of the minimum phase assumption required in spectrum smoothing. On the other hand, the results above are inconclusive on what to expect from data recorded using a minimum-phase wavelet.

References

[1] Taylor, H. L., Bonks, S. C., and McCoy, J. G., "Deconvolution with the ℓ_1 Norm," Presented at the 47th Annual Meeting of SEG, (1977), Submitted for publication in *Geophysics*.

[2] Wiggins, R. A., "Minimum Entropy Deconvolution," Proceedings of the International Symposium on Computer Aided Seismic Analysis and Discrimination," (1977), Falmouth, Mass., IEEE Computer Society, pp. 7-14

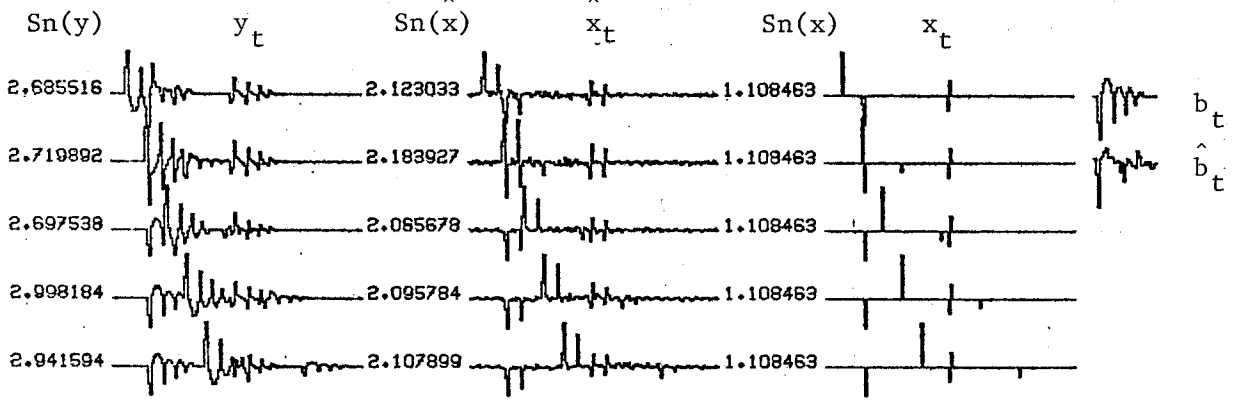

Fig 1a: $\hat{x}_t$ is the deconvolved estimate of y_t . Y_t was formed by convolving b_t with x_t . Eleven iterations with n=2.0 were used to give $\hat{x}_t$. The computed bubble, $\hat{b}_t$ is shown underneath b_t .

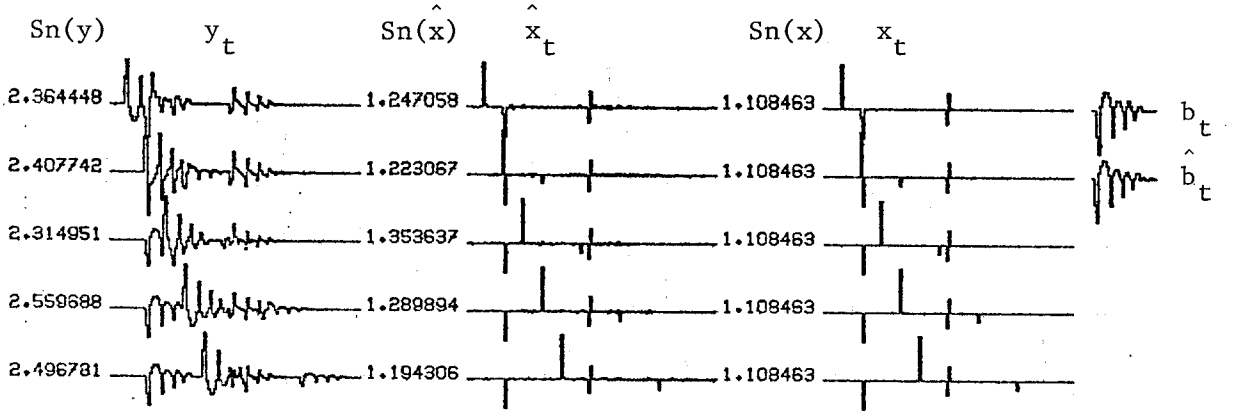

Fig 1b: Same as 1a, except $\hat{x}_t$ was found by using 11 iterations with n=2.5 followed by 10 iterations with n=2.0

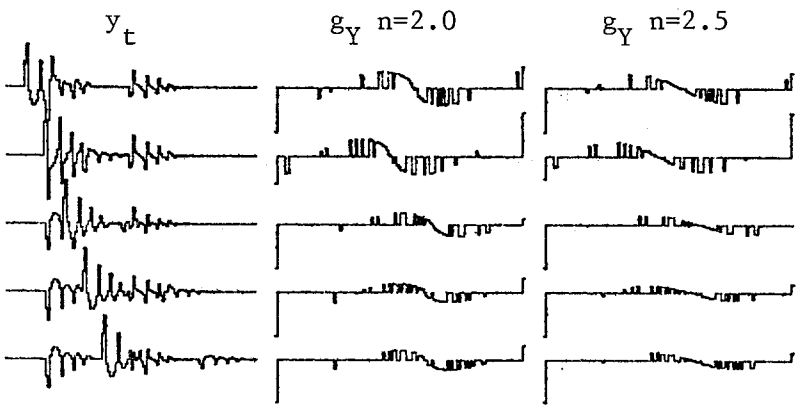

Fig. 1c: Illustrates how different norms "see" y_t differently. The abscissa of the two plots on the right is y and the ordinate is g . Note: on all plots, positive ordinates point down the page.

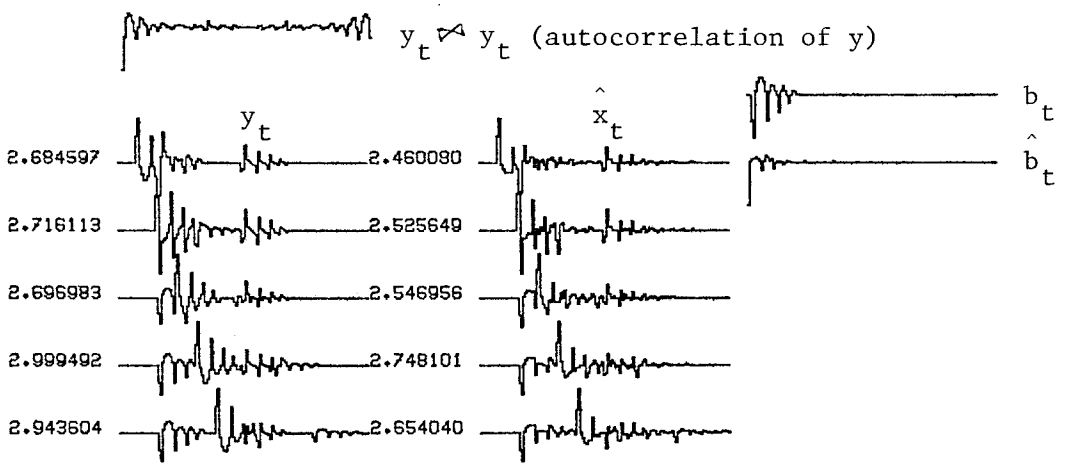

Fig. 1d: Deconvolving y_t using spectrum smoothing. Numerator smoothing was used.

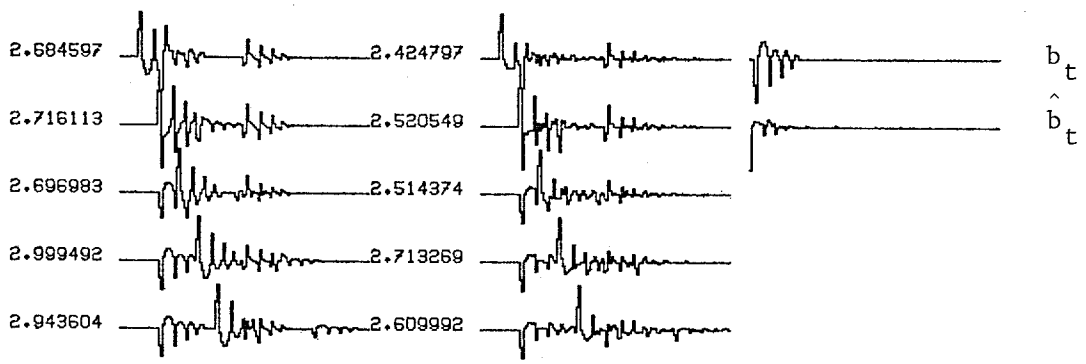

Fig. 1e: Same as Fig. 1d except that ratio smoothing was used. Both results are poor compared to parsimonius deconvolution.

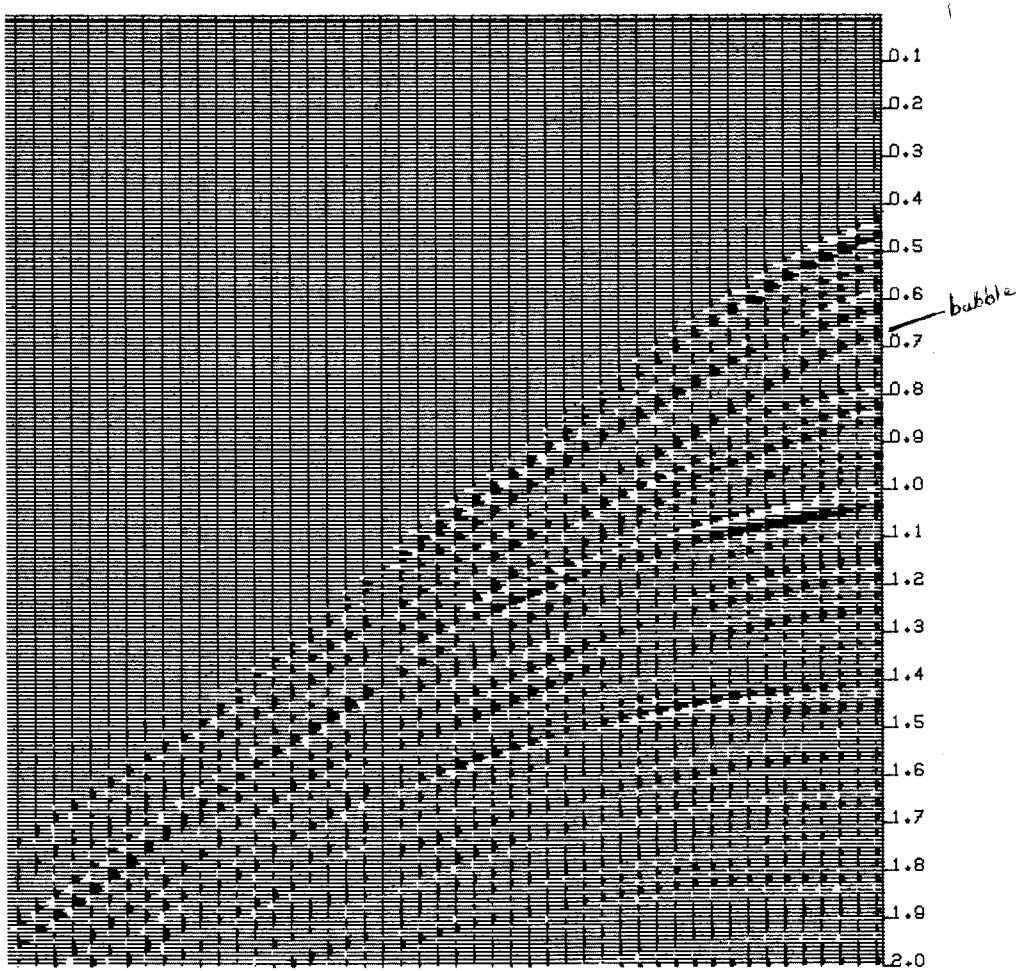

Fig 2a: A common shot gather plotted after spherical divergence correction and relative amplitude scaling have been applied to the data. The area referred to by "bubble" is parallel to the sea-floor and is interpreted as source waveform

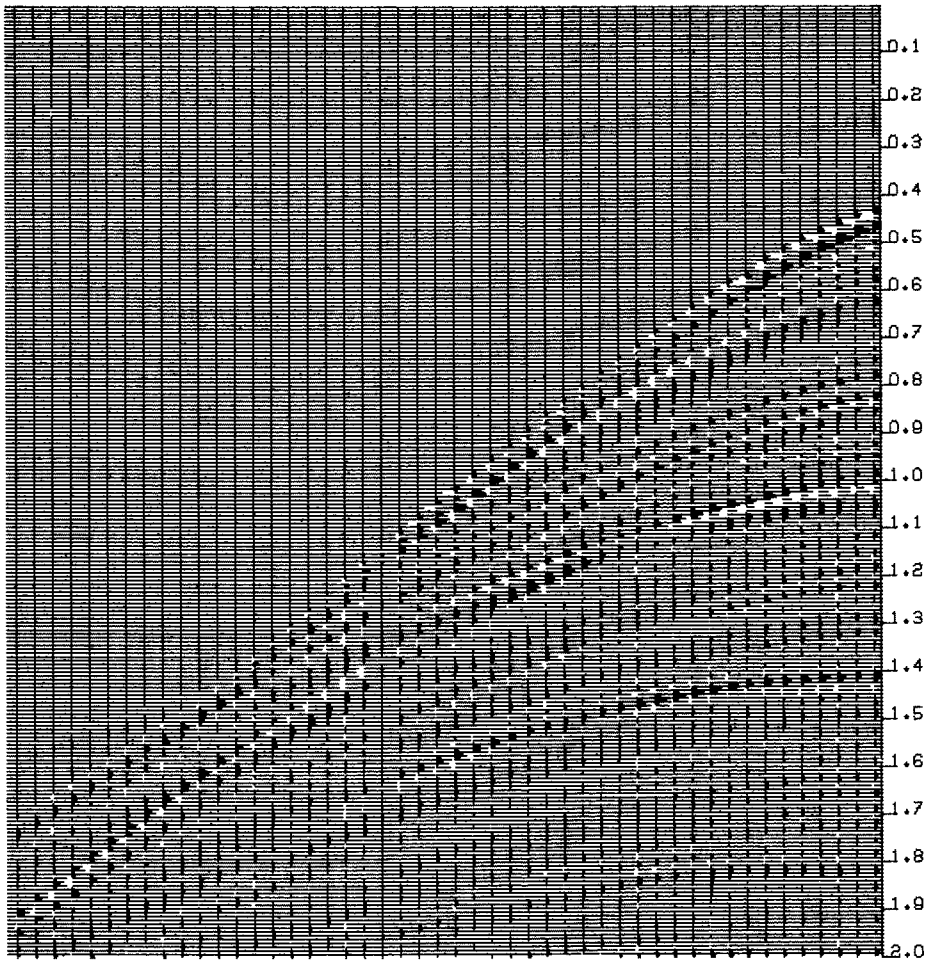

Fig 2b: SSD of Fig. 2a. Note absence of the bubble.

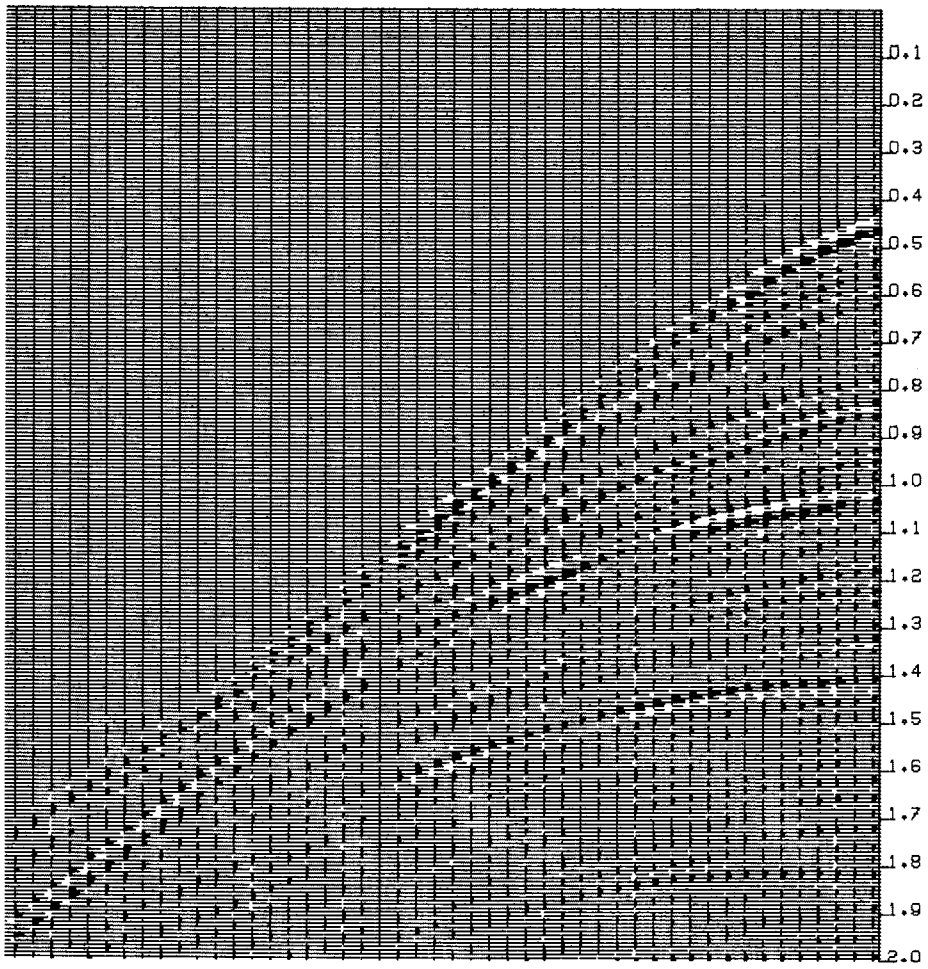

Fig. 2c: PD of Fig. 2a. A faint water bottom precurssor is now evident on 4-5 traces. The bubble has been largely removed.

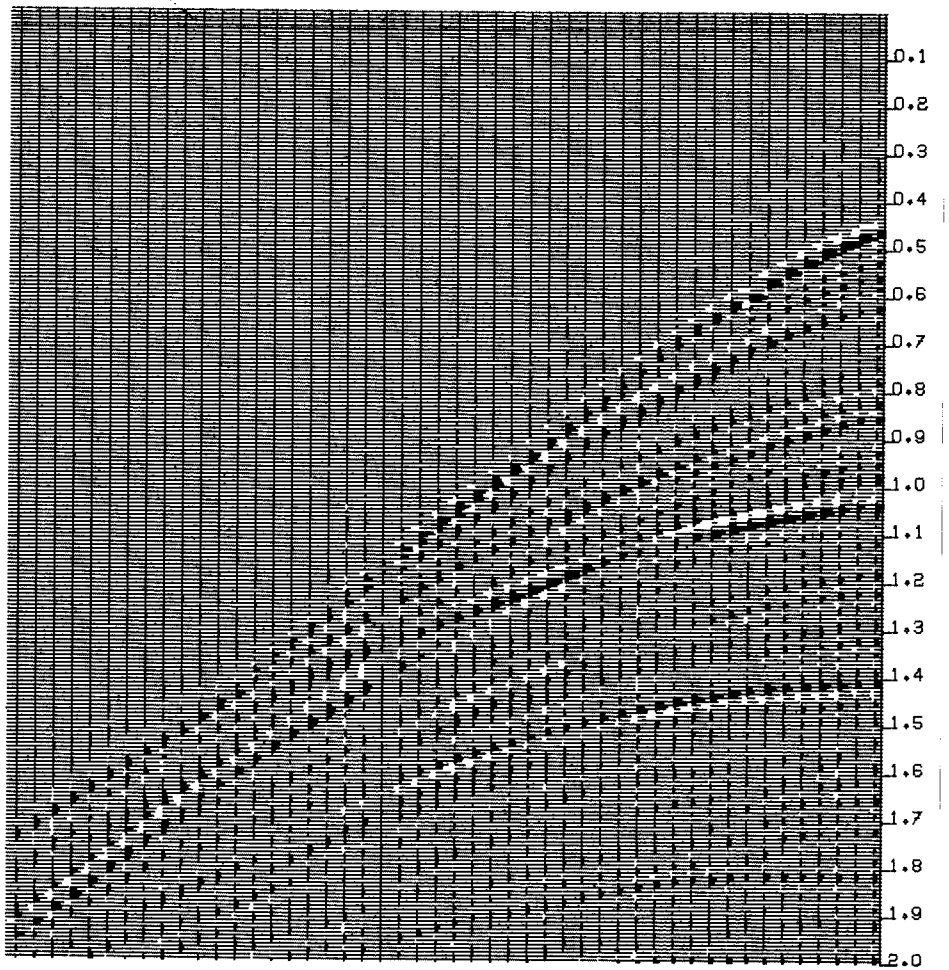

Fig. 2d: Deconvolution of Fig. 2a using a method being studied by Will Gray. The result is very similar in appearance to both Fig. 2b and Fig. 2c.

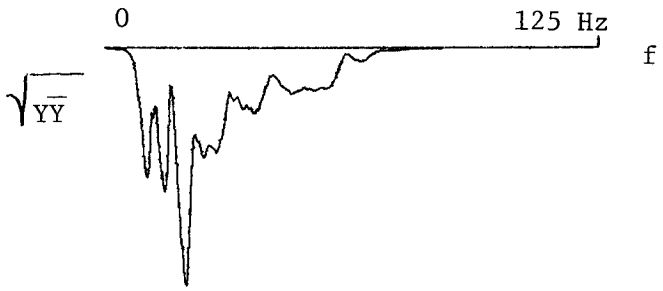

Fig 3a: Average amplitude spectrum of Fig. 2a

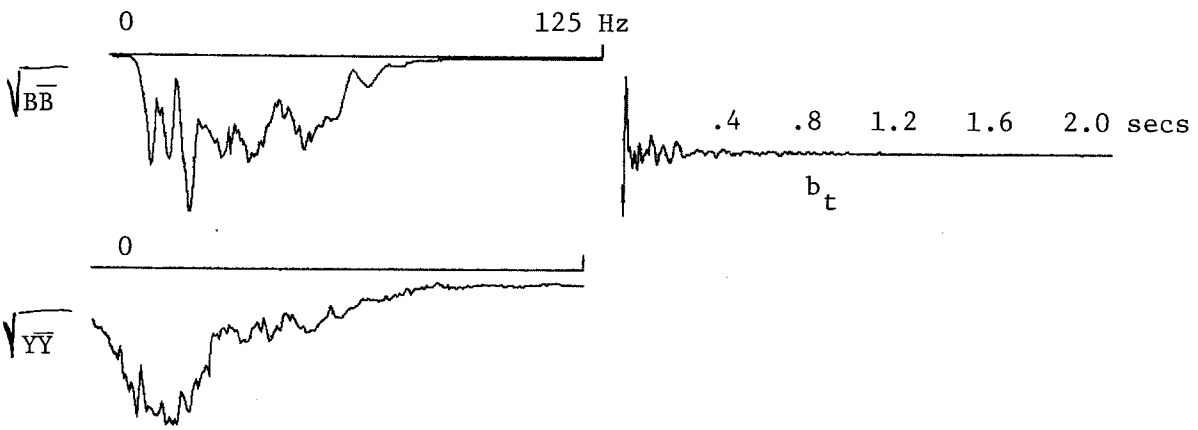

Fig. 3b: Top figure shows amplitude spectrum and shape of b_t computed using SSD. Bottom figure shows average amplitude spectrum of Fig. 2b.

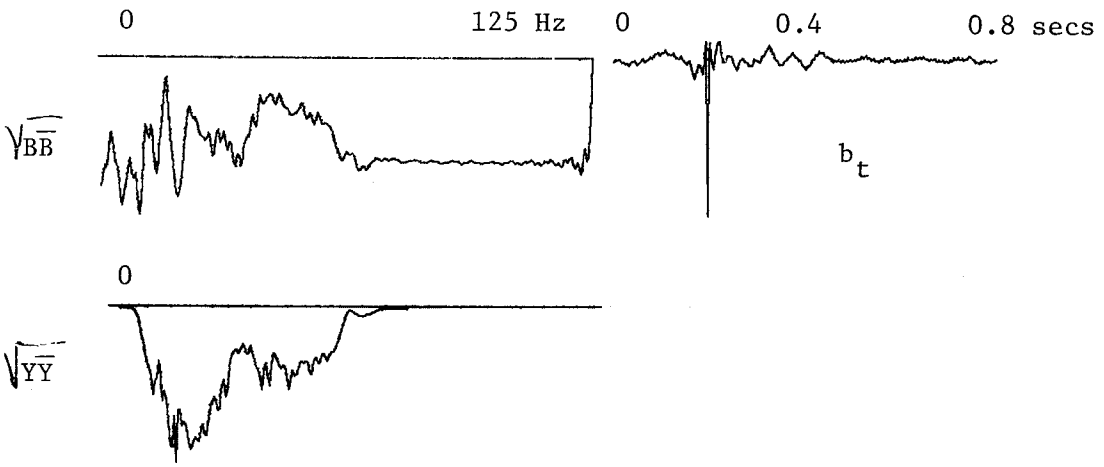

Fig. 3c: Top figure shows amplitude spectrum and shape of b_t computed using PD. Bottom figure shows average amplitude spectrum of Fig. 2c.
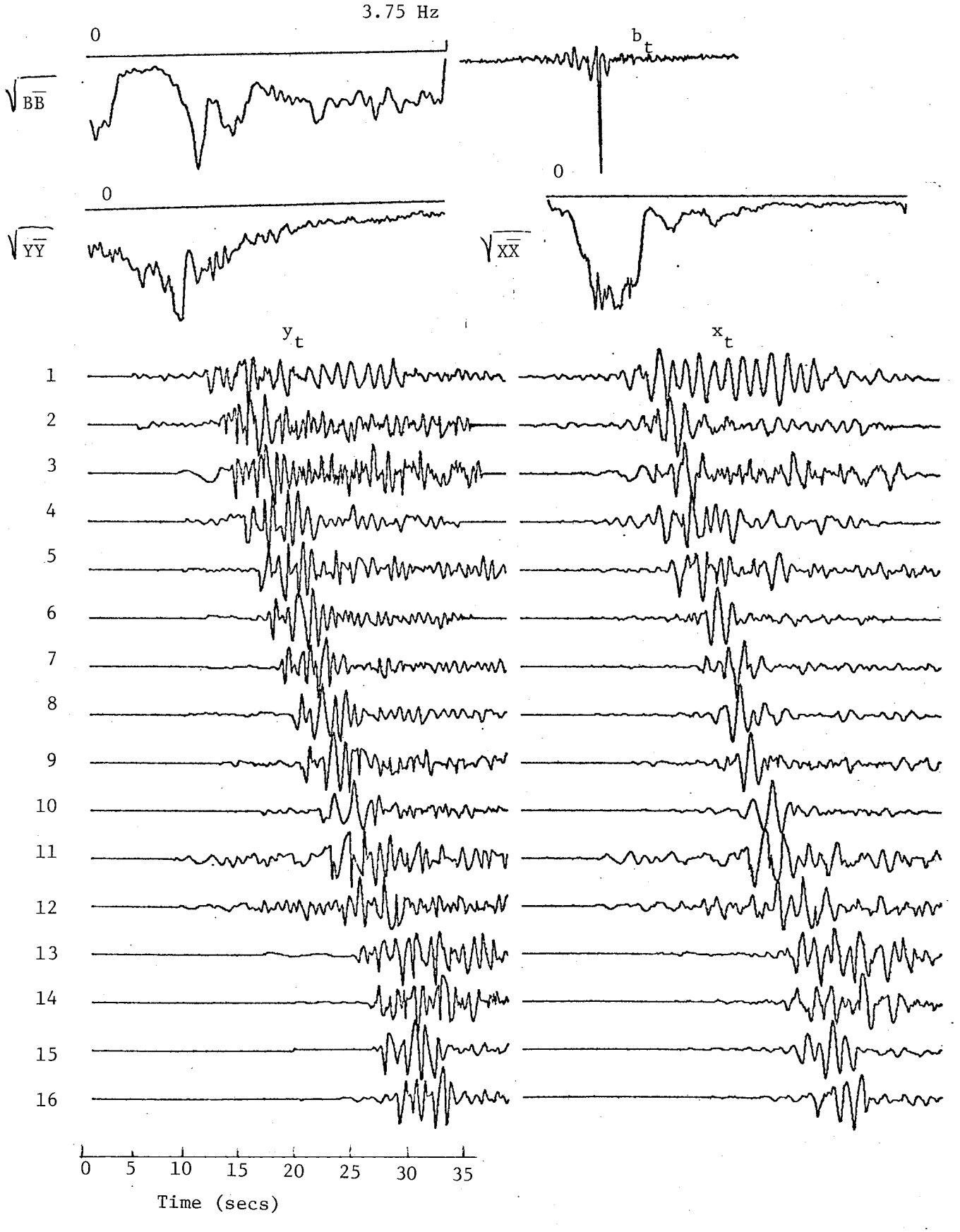
Fig. 4: Parsimonius deconvolution of earthquake data [see Wiggins]. The PkP phase is the dominant event.