

## Estimation of the discrete Fourier transform, a linear inversion approach

Mauricio D. Sacchi\*, and Tadeusz J. Ulrych\*

### ABSTRACT

Spatio-temporal analysis of seismic records is of particular relevance in many geophysical applications, e.g., vertical seismic profiles, plane-wave slowness estimation in seismographic array processing and in sonar array processing. The goal is to estimate from a limited number of receivers the 2-D spectral signature of a group of events that are recorded on a linear array of receivers. When the spatial coverage of the array is small, conventional  $f$ - $k$  analysis based on Fourier transform leads to  $f$ - $k$  panels that are dominated by sidelobes. An algorithm that uses a Bayesian approach to design an artifacts-reduced Fourier transform has been developed to overcome this shortcoming. A by-product of the method is a high-resolution periodogram. This extrapolation gives the periodogram that would have been recorded with a longer array of receivers if the data were a limited superposition of monochromatic plane waves.

The technique is useful in array processing for two reasons. First, it provides spatial extrapolation of the array (subject to the above data assumption) and second, missing receivers within and outside the aperture are treated as unknowns rather than as zeros.

The performance of the technique is illustrated with synthetic examples for both broad-band and narrow-band data. Finally, the applicability of the procedure is assessed analyzing the  $f$ - $k$  spectral signature of a vertical seismic profile (VSP).

### INTRODUCTION

$F$ - $K$  spectral analysis plays a fundamental role in the interpretation of 2-D geophysical signals. The conventional analysis based on 2-D Fourier transformation might result in poorly resolved spectral panels caused by the presence of sidelobes that tend to mask the signals. This is more noticeable when the

spatial aperture of the signal is small compared with the range of wavenumbers we are seeking. The combined effect of the sidelobes and noise make the problem even more severe.

Many of the 1-D high-resolution spectral analysis techniques that are cited in the literature can be extended easily to the 2-D case. A review of these methods can be found in Kay and Marple (1981). Two-dimensional extensions of these procedures are given in Marple (1987). We are not going to discuss these procedures in detail since that would exceed the scope of the paper. However, one can say that the power spectrum estimate is computed by honoring a few lags of the autocorrelation function (1-D or 2-D depending the problem). Examples of these approaches are 2-D parametric spectral analysis (Lim and Malik, 1982) and the Capon maximum likelihood method (Capon, 1969). Hybrid techniques can be designed by applying any 1-D high-resolution spectral method to the rows of an auxiliary array composed of column vectors obtained with the discrete Fourier transform (DFT). Since the auxiliary array is complex, high-resolution algorithms in complex form are necessary (Marple, 1987). The hybrid scheme serves to improve the resolution of only one of the variables (temporal frequency,  $f$ , or wavenumber,  $k$ .) In array processing, the spatial resolution is limited by the small number of receivers relative to the number of time samples of each trace. In other words, the spatial coverage dictates the requirement of high-resolution methods. Ulrych and Walker (1981) and more recently Swingler and Walker (1989) applied linear prediction to extrapolate the data and simulate a longer array that can be analyzed using conventional spectral analysis.

Finally, we would especially like to mention the eigenvalues based or subspace techniques. These techniques have been borrowed from the field of sonar processing (Bienvenu and Koop, 1983) and have been successfully applied in geophysics in different contexts, i.e., velocity analysis (Biondi and Kostov, 1989; Kirilin, 1992), wavefield separation (Mari and Glangaud, 1990; Rutty and Jackson, 1992). Subspace methods can certainly provide high resolution but their performance is severely affected when the number of signals are over or underestimated. In fact, these methods are based on a 2-D extension of

Manuscript received by the Editor November 28, 1994; revised manuscript received June 15, 1995.

\*Department of Geophysics and Astronomy, The University of British Columbia, 129-2219 Main Mall, Vancouver, B.C. V6T 1Z4 Canada.

© 1996 Society of Exploration Geophysicists. All rights reserved.

Pisarenko's harmonic spectral analysis (Pisarenko, 1973) which, as is well known, is very sensitive to the number of harmonics that compose the process. A similar problem is also encountered in parametric spectral estimation, where a knowledge of the number of parameters is vital in obtaining reliable estimates.

Since the techniques described above provide a high-resolution power spectral density (PSD) estimate, only the autocorrelation function of the data can be reconstructed. In this paper, we propose a new approach to the problem of determining a high-resolution 2-D PSD estimate. First, we use the data as the constraints rather than the autocorrelation function (an approach first suggested to us by Colin Walker). Second, our unknown is an unwrapped DFT, or in other words a DFT where truncation artifacts (sidelobes) are mitigated. This is an important distinction with respect to conventional spectral analysis where the target is the PSD. It is straightforward to see that the PSD is obtained as a by-product of the procedure using the unwrapped DFT.

The high-resolution DFT is computed solving an inverse problem. The proper regularization of the problem is derived using Bayes's rule to combine prior information with the data likelihood. We propose to model the prior distribution of spectral amplitudes by means of the Cauchy probability density function (pdf). Techniques using the Bayesian approach are also popular in image reconstruction (see, for instance, Skilling and Bryan, 1984).

It must be pointed out that the Cauchy pdf has also been adopted in Crase et al. (1990) and Amundsen (1991) who used this pdf in an attempt to diminish errors resulting from outliers. In our approach the Cauchy pdf is used to model the unknown parameters. The Cauchy pdf serves to model a sparse distribution of parameters, which is indeed the case when the data consist of a limited number of monocromatic plane waves.

### LINER INVERSION OF THE DISCRETE FOURIER TRANSFORM

#### Problem definition and details

For simplicity, we will start with the 1-D DFT since extensions to higher dimensions are straightforward. Consider a  $N$ -sample time or spatial series  $x_0, x_1, x_2, \dots, x_{N-1}$ . The DFT of the discrete series is given by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi nk/N} \quad k = 0, \dots, N-1, \quad (1)$$

and similarly, the inverse DFT is given by

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{i2\pi nk/N} \quad n = 0, \dots, N-1. \quad (2)$$

Let us suppose that we want to estimate  $M$  spectral samples where  $M > N$ . A standard approach to solving this problem is by means of zero padding. Defining a new time series consisting of the original series plus a zero extension for  $n = N, \dots, M-1$ , we can estimate  $M$  spectral samples using the DFT. This procedure helps to remove ambiguities caused by discretization of the Fourier transform but, as is well known, it does not reduce the sidelobes created by the temporal/spatial

window or improve the resolution. Let us therefore consider the estimation of  $M$  spectral samples but using equation (1) without zero padding. In other words, we want to estimate the DFT using only the available information. At this point, it is interesting to note that the underlying philosophy is similar to Burg's maximum entropy method (MEM) (Burg, 1975). However, in Burg's MEM the target is a PSD estimate, a phase-less function.

To avoid biasing our results by the discretization, we also impose the condition  $M \gg N$ . Rewriting equation (2) as

$$x_n = \frac{1}{M} \sum_{k=0}^{M-1} X_k e^{i2\pi nk/M} \quad n = 0, \dots, N-1, \quad (3)$$

gives rise to a linear system of equations

$$\mathbf{y} = \mathbf{F}\mathbf{x}, \quad (4)$$

where the vector  $\mathbf{y} \in \mathbf{R}^N$  and  $\mathbf{x} \in \mathbf{C}^M$  denote the available information and the unknown DFT, respectively. Equation (4) is a linear underdetermined problem which, as is well known, can be satisfied by many different solutions. Uniqueness is imposed by defining a regularized solution,  $\hat{\mathbf{x}}$ , which is obtained solving the problem (see, for example, Tikhonov and Goncharsky, 1987) expressed by

$$\min \{J(\mathbf{x}) = \Phi(\mathbf{x}) + \|\mathbf{y} - \mathbf{F}\mathbf{x}\|_2^2\}. \quad (5)$$

The regularizer  $\Phi(\mathbf{x})$  serves to impose a particular feature on the solution. In the next section, we explore a Bayesian approach to compute the regularizer. In equation (5)  $\|\cdot\|_2$  stands for the  $\ell_2$  norm.

#### The Gaussian regularization

Throughout this paper, we will consider data contaminated with noise that is distributed as  $N(0, \sigma_n^2)$ . We use the Gaussian distribution not only because this leads to an easily manageable least-squares type solution but also because it is the most parsimonious manner of describing "noise" in probabilistic terms. Therefore, invoking the Gaussian hypothesis, the conditional distribution of the data is given by

$$p(\mathbf{y}|\mathbf{x}, \sigma_n) = \left(\frac{1}{2\pi\sigma_n^2}\right)^{(N-1)/2} e^{-\frac{1}{2\sigma_n^2}\|\mathbf{y}-\mathbf{F}\mathbf{x}\|_2^2}. \quad (6)$$

Consider a prior distribution for  $p(\mathbf{x}|\mathbf{a})$ , conditional on a set of parameters  $\sigma_x$ . According to Bayes's rule [see, for instance, Loredó (1990) for a review of Bayesian methods], the a posteriori distribution of the vector of parameters is given by

$$p(\mathbf{x}|\mathbf{y}, \sigma_x, \sigma_n) = \frac{p(\mathbf{x}|\sigma_x)p(\mathbf{y}|\mathbf{x}, \sigma_n)}{p(\mathbf{x}|\mathbf{x}, \sigma_x, \sigma_n)}. \quad (7)$$

From the point of view of a frequentist, where probabilities are derived from the relative frequencies of outcomes, assigning a prior density to a model may appear incorrect. However, from a Bayesian perspective, the prior assigns the degree of plausibility that a model is correct. Once the prior probability is assigned, the problem remaining is how to choose a model. A natural rule is to compute the so called maximum a posteriori (MAP) estimator  $\hat{\mathbf{x}}$ . This estimator maximizes  $p(\mathbf{x}|\mathbf{y}, \sigma_x, \sigma_n)$  for given  $\sigma_x$  and  $\sigma_n$ . Let us assume that the prior

distribution of  $X_k$  is Gaussian, and since  $X_k$  is a complex variable, the real and imaginary parts are independent variables with Gaussian distributions. Then

$$p(\Re(\mathbf{x}), \Im(\mathbf{x})|\sigma_x) = \left(\frac{1}{2\pi\sigma_x^2}\right)^{M-1} e^{-\frac{1}{2\sigma_x^2}(\|\Re(\mathbf{x})\|_2^2 + \|\Im(\mathbf{x})\|_2^2)}$$

$$= \left(\frac{1}{2\pi\sigma_x^2}\right)^{M-1} e^{-\frac{1}{2\sigma_x^2} \sum_{k=0}^{M-1} (\Re(X_k)^2 + \Im(X_k)^2)} \quad (8)$$

The last equation describes the joint distribution of  $2(M - 1)$  random variables that represent the joint distribution of  $M - 1$  complex variables (Johnson and Kotz, 1972). Since  $\|\Re(\mathbf{x})\|_2^2 + \|\Im(\mathbf{x})\|_2^2 = \mathbf{x}^H \mathbf{x}$ , equation (8) may be regarded as the joint distribution of the complex variable  $\mathbf{x}$  we designate by  $p(\mathbf{x})$ . The MAP solution that maximizes the a posteriori probability also minimizes the following cost function

$$J_{gg}(\mathbf{x}) = \lambda \|\mathbf{x}\|_2^2 + \|\mathbf{y} - \mathbf{F}\mathbf{x}\|_2^2 \quad (9)$$

$$= \lambda \mathbf{x}^H \mathbf{x} + (\mathbf{y} - \mathbf{F}\mathbf{x})^H (\mathbf{y} - \mathbf{F}\mathbf{x}),$$

where the subscript *gg* stands for the Gauss-Gauss model. The scalar  $\lambda = \sigma_n^2/\sigma_x^2$  is also known in inverse theory as the tradeoff or ridge regression parameter. Equation (9) is the objective function of the problem. The first term represents the model norm, while the second term is the misfit function. The hyper-parameter  $\lambda$  enables us to move our estimate along the tradeoff curve. Taking derivatives and equating to zero yield

$$\hat{\mathbf{x}} = (\mathbf{F}^H \mathbf{F} + \lambda \mathbf{I}_M)^{-1} \mathbf{F}^H \mathbf{y}. \quad (10)$$

We can write equation (10) in another form using the following identity

$$(\mathbf{F}^H \mathbf{F} + \lambda \mathbf{I}_M)^{-1} \mathbf{F}^H = \mathbf{F}^H (\mathbf{F} \mathbf{F}^H + \lambda \mathbf{I}_N)^{-1}, \quad (11)$$

where  $\mathbf{I}_M$  and  $\mathbf{I}_N$  represent  $M \times M$  and  $N \times N$  identity matrices, respectively. Recalling equation (11) and that  $\mathbf{F} \mathbf{F}^H = \frac{1}{M} \mathbf{I}_N$ , we end up with the Fourier transform estimate for the Gauss-Gauss model

$$\hat{\mathbf{x}} = \left(\frac{1}{M} + \lambda\right)^{-1} \mathbf{F}^H \mathbf{y}. \quad (12)$$

The result is nothing else than the DFT of  $x_n$ , modified by a scale factor. The solution expressed by equation (12) becomes

$$\hat{X}_k = \frac{1}{1 + \lambda M} \sum_{n=0}^{N-1} x_n e^{-i2\pi nk/(M-1)}, \quad (13)$$

which is the DFT of the windowed time series [equation (1)] and is equivalent to padding with zeros in the range  $n = N, \dots, M - 1$ . It is easy to see that the Gauss-Gauss model yields a scaled version of the DFT. Hence, the associated periodogram exhibits a resolution that is proportional to the inverse of the length of the time series. The periodogram becomes

$$P_k = \hat{X}_k \hat{X}_k^*, \quad k_i = 0, \dots, M - 1. \quad (14)$$

It is interesting to point out that Oldenburg (1976) arrived to the same observation using the Backus and Gilbert formalism

that is a completely different approach to inversion. Specifically, when the first Dirichlet criterion of the Backus and Gilbert theory is used, the resulting expression is equivalent to that shown in equation (13). This is a simple consequence of using this criterion; other spreading functions lead to different results. An important conclusion that may be drawn from Oldenburg's paper is that the resolution is fixed and cannot be modified by moving our estimate along the tradeoff curve. An intuitive explanation is as follows; since the resolution of the PSD is controlled by the width of the sidelobes that are independent of the scaling factor, the DFT computed with the Gauss-Gauss model will lead to a PSD estimator that is equivalent to the periodogram of the truncated time series.

### Regularization by the Cauchy-Gauss model

In the previous section, we showed that the Gauss-Gauss regularization does not serve our purpose of obtaining a high-resolution PSD estimator. We have found that a very appealing idea is the use of a regularization derived from a distribution that mimics a sparse distribution of spectral amplitudes. The reason is simple and can be interpreted as follows: In general, the interpretation of spatio-temporal data is involved with a finite number of events that exhibit spatial continuity. If the data show great complexity, time and/or spatial windows are chosen such that particular events may be more easily distinguished and mapped. Because of the limited number of such events, it is appropriate to require a model that consists of the minimum number of events that can satisfy the data. This is the classic idea of simplicity or parsimony and quite opposite in philosophy to norms that describe the "structure" of a surface by means of, for example, the smallest, flattest, or smoothest models. A "long tailed" distribution, like the Cauchy pdf, will induce a model consisting of only a few elements different from zero. This pdf has been used by Crase et al. (1990) and Amundsen (1991) to increase the robustness of the inversion of seismic data. In their approach, the "long tails" of the pdf permit the damping of the influence of outliers in the inversion. In our approach, the Cauchy pdf is used to impose a particular feature on our model, that of sparseness. The Cauchy pdf is given by

$$p(X_k|\sigma_c) = \frac{1}{2\pi\sigma_c^2} \frac{1}{\left(1 + \frac{X_k X_k^*}{2\sigma_c^2}\right)}, \quad (15)$$

where  $\sigma_c$  is a scale parameter. This distribution does not possess finite moments, but  $\sigma_c$  plays a role similar to that of variance. When we combine a Cauchy prior with the data likelihood the cost function becomes

$$J_{cg}(\mathbf{x}) = S(\mathbf{x}) + \frac{1}{2\sigma_n^2} (\mathbf{y} - \mathbf{F}\mathbf{x})^H (\mathbf{y} - \mathbf{F}\mathbf{x}), \quad (16)$$

where the subscript *cg* stands for the Cauchy-Gauss model. The function,  $S(\mathbf{x})$ , which is expressed by

$$S(\mathbf{x}) = \sum_{k=0}^{M-1} \log \left(1 + \frac{X_k X_k^*}{2\sigma_c^2}\right) \quad (17)$$

is the regularizer imposed by the Cauchy distribution and is a measure of the sparseness of the vector of spectral powers

$P_k = X_k X_k^*$ ,  $k = 0, \dots, M - 1$ . The constant  $\sigma_c$  controls the amount of sparseness that can be attained by the inversion. The sparseness of the estimate will also depend on the noise level since  $\sigma_n$  may inhibit a reliable sparse solution.

Taking derivatives of  $J_{cg}(x)$  and equating to zero yields the following result

$$\hat{\mathbf{x}} = (\lambda \mathbf{Q}^{-1} + \mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H \mathbf{y}, \quad (18)$$

where  $\lambda = \sigma_n^2 / \sigma_c^2$  and  $\mathbf{Q}$  is a  $M \times M$  diagonal matrix with elements given by

$$Q_{ii} = 1 + \frac{X_i X_i^*}{2\sigma_c^2}, i = 0, \dots, M - 1. \quad (19)$$

Although expression (18) resembles the damped least-squares solution, we note that  $\mathbf{Q}$  is nonlinearly related to the DFT of the data. Equation (18) can be written using the following identity

$$\mathbf{F}^H (\lambda \mathbf{I}_N + \mathbf{F} \mathbf{Q} \mathbf{F}^H) = (\lambda \mathbf{Q}^{-1} + \mathbf{F}^H \mathbf{F}) \mathbf{Q} \mathbf{F}^H. \quad (20)$$

The forms  $(\lambda \mathbf{Q}^{-1} + \mathbf{F}^H \mathbf{F})$  and  $(\lambda \mathbf{I}_N + \mathbf{F} \mathbf{Q} \mathbf{F}^H)$  are positive definite leading to the following identity

$$(\lambda \mathbf{Q}^{-1} + \mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H = \mathbf{Q} \mathbf{F}^H (\lambda \mathbf{I}_N + \mathbf{F} \mathbf{Q} \mathbf{F}^H)^{-1}. \quad (21)$$

Using identity (21) we can write equation (18) as

$$\hat{\mathbf{x}} = \mathbf{Q} \mathbf{F}^H (\lambda \mathbf{I}_N + \mathbf{F} \mathbf{Q} \mathbf{F}^H)^{-1} \mathbf{y}. \quad (22)$$

We stress that, from the theoretical point of view of uniqueness and convergence, the operators given by equations (18) and (22) are equivalent. However, from the point of view of computational advantages, the following observations apply.

- 1) Whereas equation (18) demands the inversion of an  $M \times M$  matrix, equation (22) requires the inversion of an  $N \times N$  matrix.
- 2) The operator  $(\lambda \mathbf{I}_N + \mathbf{F} \mathbf{Q} \mathbf{F}^H)$  in equation (22) is Toeplitz Hermitian provided that the time series is uniformly discretized and a fast solver like Levinson's recursion can be used.
- 3) In the case of nonuniform discretization, a Cholesky decomposition is appropriate. This type of algorithm is also appropriate in the case of gapped data.

As we have mentioned, the Cauchy-Gauss model leads to an algorithm that resembles the minimum norm solution of equation (4). This is particularly true when  $\sigma_c$  is large compared to the spectral amplitudes that we are seeking. In this case, the functional  $S(x) \approx K + \mathbf{x}^H \mathbf{x} / (2\sigma_c^2)$  where  $K$  is a constant. Thus, minimizing  $J_{cg}(x)$  is equivalent to minimizing  $J_{gg}(x)$ . In the contrary case, the algorithm will seek a DFT with a sparse distribution of spectral amplitudes,  $X_k X_k^*$ , leading to an enhancement of the spectral peaks and reducing windowing effects or sidelobes. In the Gauss-Gauss regularization, the scale parameters  $\sigma_n$  and  $\sigma_x$  reduce to a single hyper-parameter  $\lambda$ , which completely defines the distribution of the unknown model. On the other hand, in the Cauchy-Gauss regularization, we have two independent hyper-parameters,  $\sigma_n$  or  $\lambda$ ,  $\sigma_c$ . When the power of the noise is known, the optimum  $\sigma_c$  is computed using any fitting criterion, e.g., the  $x^2$

criterion. The simultaneous estimation of  $\sigma_c$  and  $\sigma_n$  is not an easy task since it demands the computation of the marginal joint pdf of  $\sigma_c$  and  $\sigma_n$  for a given solution  $\hat{\mathbf{x}}$ .

The method that we employ to solve equation (22) is very simple.

- 1) Begin with an initial model that can be the DFT of the finite-length time series  $\mathbf{x}^{(0)}$ .
- 2) Select the hyperparameters  $\sigma_c$  and  $\sigma_n$ .
- 3) Iteratively solve equation (22),

$$\hat{\mathbf{x}}^{(\mu)} = \mathbf{Q}^{(\mu-1)} \mathbf{F}^H (\lambda \mathbf{I}_N + \mathbf{F} \mathbf{Q}^{(\mu-1)} \mathbf{F}^H)^{-1} \mathbf{y}, \quad (23)$$

where  $\mu$  is the iteration number.

- 4) The procedure is stopped when  $J_{cg}$  does not change with the number of iterations.
- 5) The misfit is computed. A new value of  $\sigma_c$  is used if the misfit is not satisfactory.

Since the Hessian matrix of  $J_{cg}$  is positive, the uniqueness of the solution is guaranteed. The algorithm converges after about ten iterations. Good resolution may, however, be attained in less than ten iterations.

## NUMERICAL SIMULATIONS

### Hybrid 2-D estimator of the DFT

We present a hybrid procedure based on standard Fourier analysis in the temporal variable, while for the spatial variable, we invert in the wavenumber space using the Cauchy-Gauss regularization. Usually the length of the temporal window is sufficient to achieve high resolution with simple methods based on standard Fourier analysis while the aperture of the array limits the spatial resolution. The 2-D algorithm works as follows:

- 1) Each record is transformed to the frequency-offset domain using the FFT.
- 2) High-resolution analysis is performed at each frequency that comprises the signal using the Cauchy-Gauss regularization.
- 3) The amplitude in the  $f$ - $k$  space is plotted to identify the spatio-temporal structure of each source.
- 4) Alternatively, the data outside the original aperture may be extrapolated to simulate a longer array, and any 2-D spectral technique may be used in conjunction with the extended data set.

Undesired components can be masked before mapping back the unwindowed DFT to space-time. This is demonstrated in detail in the broad-band example below.

### First example; Spatio-temporal spectrum of narrow-band signals

The algorithm is applied to estimate the spatio-temporal spectrum of a signal received by a passive array of receivers. This problem frequently arises in radar and sonar processing (Bienvenu and Kopp, 1983). The goal is to estimate the direction of arrival and the temporal spectral signature of a set of sources impinging from different angles on a uniform array

of  $N$  receivers. In seismology, the problem has been studied particularly to detect plane-wave signals and estimate the slowness vector. We assume that the array of receivers is linear, but the method can be easily generalized to any distribution of receivers.

We simulate three narrow-band linear events with the following features. First, we model two sinusoids with unit amplitude and with normalized wavenumbers of 0.30 and 0.25 units and a normalized frequency of 0.20 units, respectively. The third wave ( $f = 0.35$ ,  $k = -0.25$ ) has an amplitude that is 25% below the amplitude of the first and second waves. The temporal extension of each channel is 150 samples, which represents one order of magnitude above the aperture of the array (15 receivers). Gaussian noise with standard deviation,  $\sigma_n = 0.1$  was added to the composite record. The noise represents 40% of the amplitude of the third wave. Each channel was tapered with a Hamming window. The spatio-temporal spectrum computed using the periodogram is illustrated in Figure 1a. The contour lines correspond to normalized amplitudes ranging from 0 to -40 dB, with an interval of -5 dB. The  $f$ - $k$  plane is dominated by sidelobes caused by truncation in space and time. This is more noticeable for the wavenumber, since the aperture of the array is one order smaller than the length of the time series. The data were processed with the hybrid procedure based on the Cauchy-Gauss model. The parameters  $\sigma_n$  and  $\sigma_c$  were chosen to reject the noise. The number of iterations varies from frequency to frequency, usually  $\approx 10$  iterations are sufficient to reach the minimum of the objective function  $J_{cg}$ . The resulting high resolution  $f$ - $k$  panel is portrayed in Figure 1b. Contour lines range from 0 to -40 dB as in Figure 1a. There is a clear enhancement of the spatial resolution and a suppression of the background noise. Since the contour lines exhibit similar width in the  $k$  and  $f$  direction, we can infer that the normalized aperture is of the same order as the length of the time series. In other words, the  $f$ - $k$  panel portrayed in Figure 1b corresponds to an equivalent array of approximately 150 receivers.

## Second example: Broad-band applications

The algorithm was tested with two broad-band linear events impinging an array of 25 receivers. The source is modeled with a Ricker wavelet with central frequency,  $f = 20$  Hz. The first event has slowness  $1.2 \times 10^{-5}$  s/m and the second  $-1.2 \times 10^{-5}$  s/m. Since we used a different polarity for each wave, there is destructive interference at near offset traces (Figure 2a.) We have changed the sign convention of the temporal DFT to obtain positive wavenumbers for waves propagating with positive slowness.

We first examine the noiseless case. Figure 2a shows the data while Figure 2b illustrates the conventional PSD computed with the DFT. Since the slowness of each event is nearly identical, the conventional  $f$ - $k$  panel cannot distinguish the existence of two signals. The positive and negative wavenumber quadrants were masked to estimate the wave with negative slowness (Figure 2c) and the wave with positive slowness (Figure 2d). Although both wavefields were decomposed, the decomposition is not correct as is clearly shown by the character of the wavelet that varies with offset in both figures.

The same procedure was carried out using the Cauchy-Gauss inversion scheme. We set  $\sigma_n = 0$  in equation (22) since we are willing to fit the data exactly. The parameter  $\sigma_c$  is 0.1% of the maximum amplitude encountered in the raw periodogram which is computed from the finite-length DFT. The PSD computed with the Cauchy-Gauss DFT is shown in Figure 2e. We note that there is a clear enhancement of the spatial resolution. The wavefields with negative and positive slowness are portrayed in Figures 2f and 2g, respectively. These panels were computed after masking the corresponding wavenumber quadrant. In comparing Figures 2f and 2g with Figures 2c and 2d it is clear that the high-resolution scheme enables us to discriminate both events correctly. Figures 2f and 2d also show that the wavelet does not suffer substantial changes with offset. In Figures 2h and 2i, we show the original aperture plus the extrapolated aperture computed with the

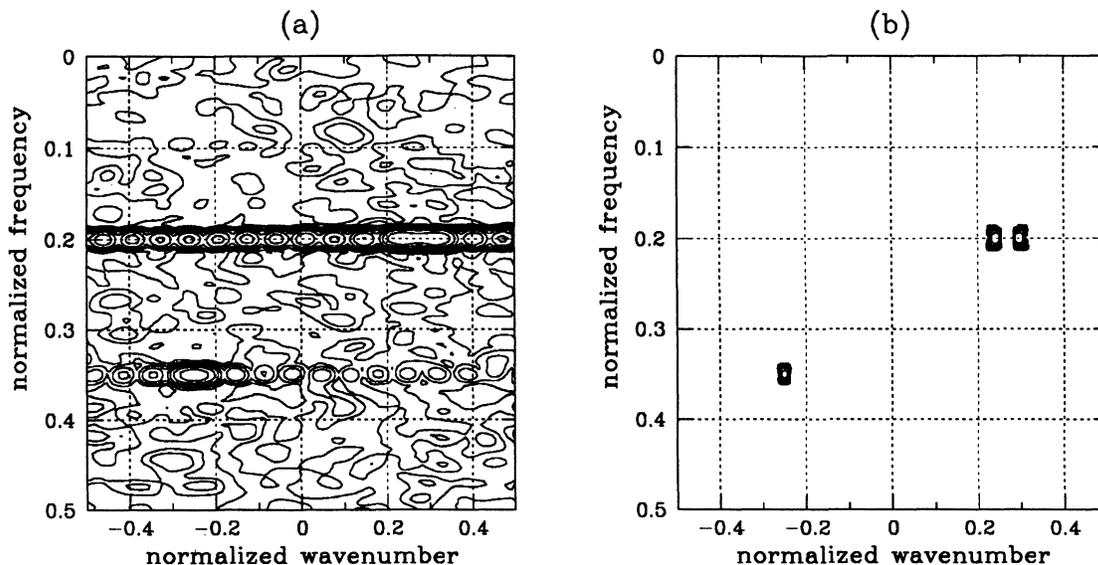


FIG. 1. 2-D spectrum of three narrow-band signals of normalized frequency-wavenumber pairs: (0.2, 0.3), (0.2, 0.25), and (0.35, -0.25). (a) Conventional 2-D estimator obtained with DFT. (b) 2-D estimator obtained with the Cauchy-Gauss model.

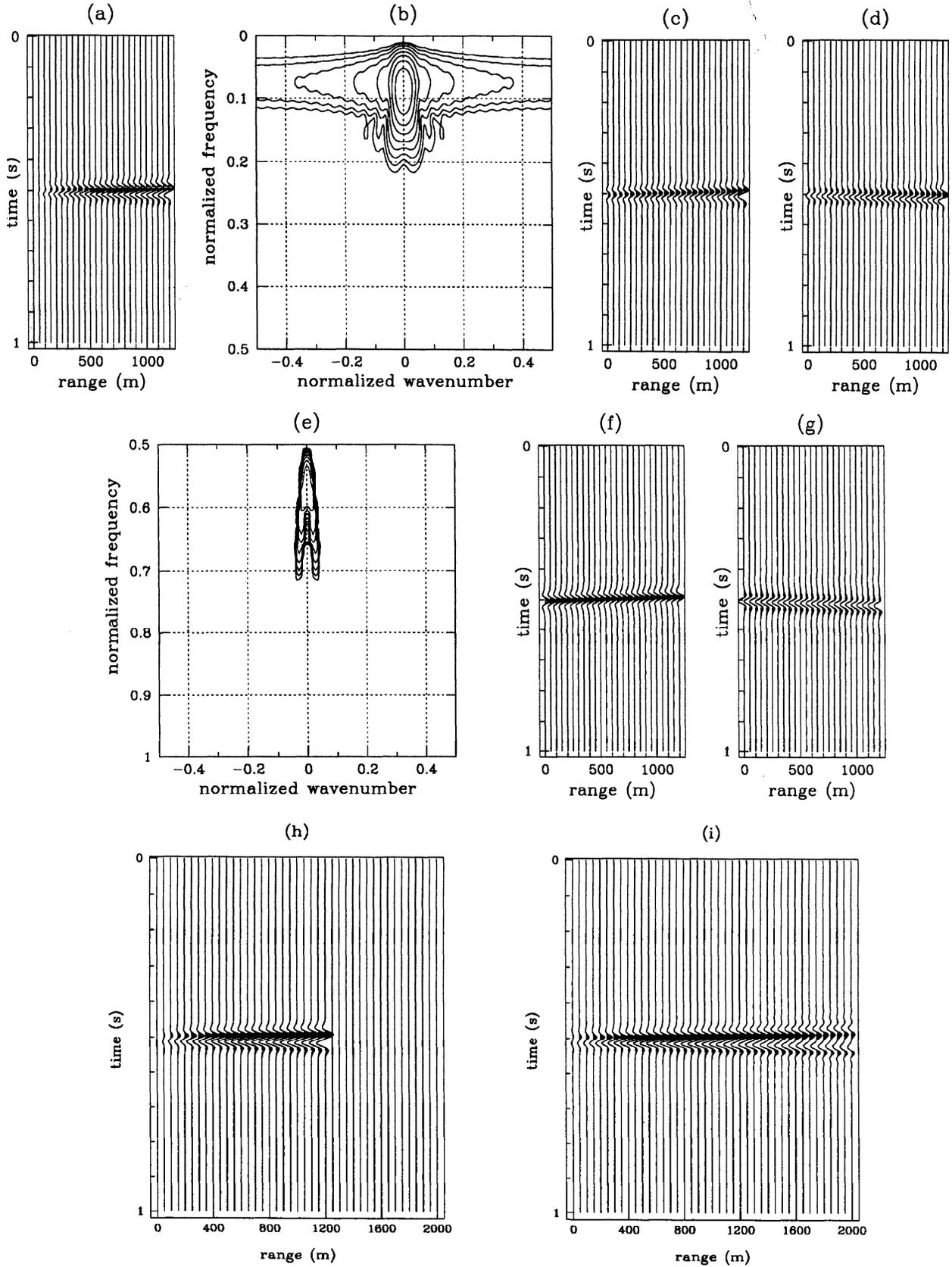


FIG. 2. (a) Synthetic data. (b) Spatio-temporal spectrum using DFT. (c)-(d) Wavefield decomposition after masking the positive and negative quadrant respectively. (e) High-resolution f-k plane computed with the Cauchy-Gauss model. (f)-(g) Wavefield decomposition. Original aperture plus extrapolated aperture: (h) Conventional solution using the DFT. (i) Cauchy-Gauss regularization.

conventional DFT and with the high resolution DFT, respectively. It is evident that the high resolution DFT correspond to a longer array of receivers. The reliability of the extrapolated data depends on how reliable is the sparse assumption to model the DFT.

Finally, the data were contaminated with Gaussian noise ( $\sigma_n = 0.1$ ). Figure 3a portrays the data while Figure 3b shows the PSD computed using the DFT (conventional  $f-k$  analysis). Figures 3c and 3d show the wavefield separation using the conventional DFT analysis. Not only is the character of the wavelet changing with offset, but also the signal-to-noise ratio has not been improved.

The Cauchy-Gauss PSD is illustrated in Figure 3e. When compared with Figure 3b, we note that the resolution has been improved and the noise is attenuated substantially. It must be pointed out that, like in the broad-band case, the parameters  $\sigma_c$  and  $\sigma_n$  are chosen to suppress the noise. The Cauchy-

Gauss DFT was used to reconstruct the  $t-x$  space after masking the right and left quadrant of the  $f-k$  domain and the results are portrayed in Figures 3f and 3g. These figures show an accurate separation of each wavefield and an important signal-to-noise ratio enhancement.

APPLICATION TO VERTICAL SEISMIC PROFILING

VSP data are particularly suitable for our algorithm. The VSP data are composed of two principal linear wavefields: the downgoing and upgoing waves. The downgoing waves have the higher amplitude and tend to mask the upgoing waves. The data correspond to 26 traces of a VSP and are portrayed in Figure 4a. As in the previous examples, we compute the 2-D PSD with the conventional DFT and with the DFT computed by means of the Cauchy-Gauss inversion. The results are portrayed in Figures 4b and 4c, respectively. The wavefield

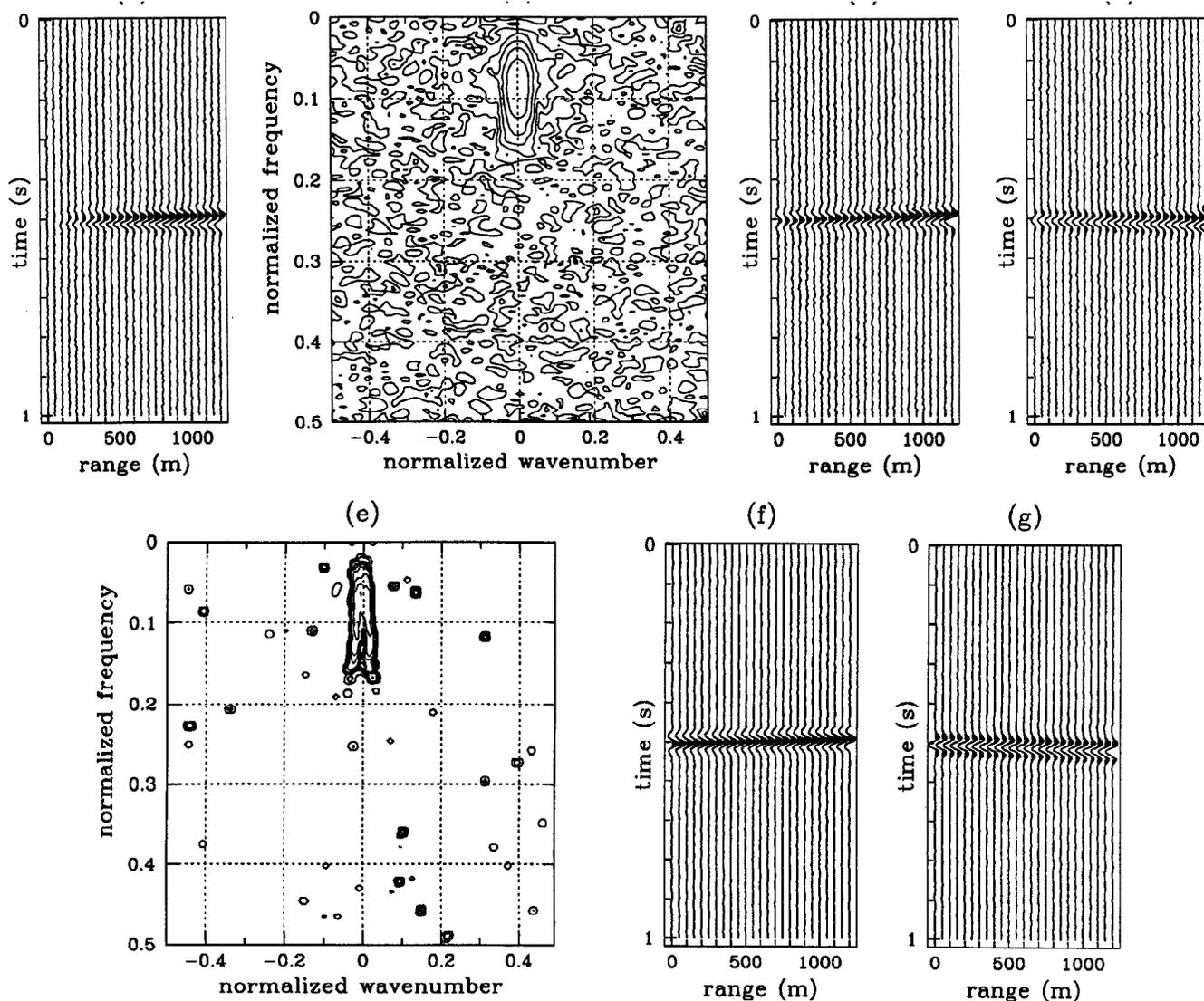


FIG. 3. (a) Synthetic data contaminated with noise. (b) Spatio-temporal spectrum using DFT. (c)-(d) Wavefield decomposition after masking the positive and negative quadrant, respectively. (e) High-resolution  $f-k$  plane computed with the Cauchy-Gauss model. (f)-(g) Wavefield decomposition.

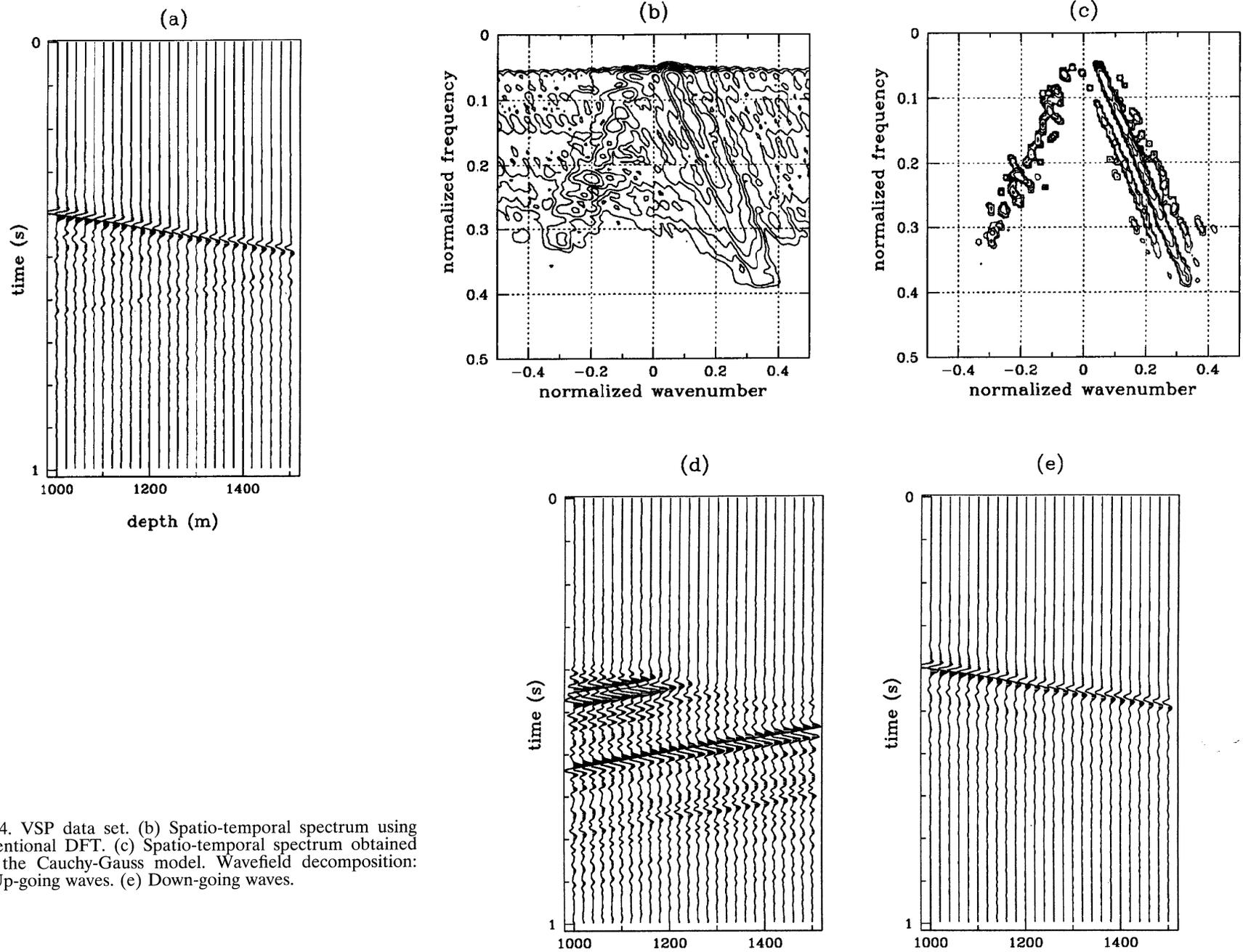


FIG. 4. VSP data set. (b) Spatio-temporal spectrum using conventional DFT. (c) Spatio-temporal spectrum obtained with the Cauchy-Gauss model. Wavefield decomposition: (d) Up-going waves. (e) Down-going waves.

separation carried out using the Cauchy-Gauss DFT is shown in Figure 4d (upgoing waves) and Figure 4e (downgoing waves). It is clear from a comparison of the  $f-k$  results in Figures 4b and 4c that the Cauchy-Gauss  $f-k$  result is considerably superior not only in allowing a better identification of different wavefields, but also from the point of view of the design of suitable filters for event separation.

#### DISCUSSION AND CONCLUSIONS

The high-resolution technique for the estimation of the power spectrum presented in this paper is based on the application of an algorithm that seeks a sparse solution to the ubiquitous problem of spectrum estimation from a finite set of data. What makes the algorithm very attractive is that, since the sparseness measure is minimized subject to data constraints, phase information is also recovered and allows the extrapolation of the signal outside the original window or aperture, depending on the problem. The latter is consistent with the idea of simulating a longer array and then estimating the PSD using the DFT.

Another attractive feature of the method is that the background noise may be considerably attenuated after tuning the hyperparameters. The synthetic examples and the VSP example show the viability of the technique in the processing of real data. The technique that we have presented has wide applicability and a wide range of problems suggest themselves. An example is multiple suppression from normal moveout corrected gathers, where the aperture of the array inhibits the correct identification of signals with close moveout.

We would like to mention that our algorithm is not intended as an alternative to the pivotal FFT algorithm. We do, however, suggest its application in cases where effects related to the ubiquitous aperture limitation problem give rise to difficulties in the identification and interpretation of closely spaced events.

Finally, we would like to stress the importance of using prior information to improve the resolution of 2-D spectral estimators. We have to recognize, however, that there is a resolution tradeoff. In our problem, if the data do not consist of a limited number of plane waves, the sparse assumption is doomed to failure. This facet is constantly present in any inverse problem.

#### ACKNOWLEDGMENTS

MDS appreciates the support of a UBC graduate fellowship. This research was also supported by a NSERC grant to TJU.

#### REFERENCES

- Amundsen, L., 1991, Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion: *Geophysics*, 56, 2027-2035.
- Bienvenu, G., and Kopp, L., 1983, Optimality of high-resolution array processing using the eigensystem approach: *IEEE, Trans. Acoust. Speech Signal Process.*, **ASSP-31**, 1235-1248.
- Biondi, B. L., and Kostov, C., 1989, High-resolution velocity spectra using eigenstructure methods: *Geophysics*, 54, 832-842.
- Burg, J. P., 1975, Maximum entropy spectral analysis: Ph.D. Thesis, Stanford University.
- Capon, J., 1969, High-resolution, frequency-wavenumber spectrum analysis: *Proc. IEEE*, 57, 1408-1418.
- Cruse, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A., 1990, Robust elastic nonlinear waveform inversion: Application to real data, *Geophysics*, 55, 527-538.
- Johnson, N. L., and Kotz, S., 1970, *Distributions in statistics: 1 and 2*: John Wiley & Sons, Inc.
- Kay, S. M., and Marple, L. M., 1981, Spectrum analysis-A modern perspective: *Proc. IEEE*, 69, 1380-1419.
- Kirlin, R. L., 1992, The relationship between the semblance and the eigenstructure velocity estimators: *Geophysics*, 57, 1027-1033.
- Lim, J. S., and Malik, N. A., 1982, A new algorithm for two-dimensional maximum entropy power estimation: *IEEE Trans. Acoust. Speech Signal Process.*, vol. **ASSP30**, 788-797.
- Loredo, T. J., 1990, From Laplace to supernova SN 1987A: Bayesian inference in astrophysics, in Fougere, P., Ed., *Maximum entropy and Bayesian methods*: Kluwer Academic Publishers, 81-142.
- Mari, J. L., and Glangeaud, F., 1990, Spectral matrix filtering applied to VSP filtering: *Rev. Institute Francais du Petrole*, 45.
- Marple, S. M., 1987, *Digital spectral analysis, with applications*: Prentice-Hall Inc.
- Oldenburg, D. W., 1976, Calculation of Fourier transforms by the Backus-Gilbert method: *Geophys. J. Roy. Astr. Soc.*, 44, 413-431.
- Pisarenko, V. P., 1973, The retrieval of harmonics from a covariance function: *Geophys. J. Roy. Astr. Soc.*, 33, 347-366.
- Rutty, M. J., and Jackson, G. M., 1992, Wavefield decomposition using spectral matrix techniques: *Expl. Geophys.*, 23, 293-298.
- Skilling, J., and Bryan, R. K., 1994, Maximum entropy image reconstruction: General algorithm: *Mon. Not. Roy. Astr. Soc.*, **211**, 111-124.
- Swingler, D., and Walker, R., 1989, Line-array beamforming using linear prediction for aperture interpolation and extrapolation: *IEEE Trans. Acoust. Speech Signal Process.* **ASSP-37**.
- Tikhonov, A. H., and Goncharsky, A. V., 1987, *Ill-posed problems in the natural sciences*: MIR Publ.
- Ulrych, T. J., and Walker, C. J., 1981, High-resolution, two-dimensional power spectral estimation, in Findley, D. F., Ed., *Applied time series II*: Academic Press, Inc.