Multichannel Blind Deconvolution of Seismic Signals

Kjetil F. Kaaresen and Torfinn Taxt

Abstract — A new algorithm for simultaneous wavelet estimation and deconvolution of seismic reflection signals is given. To remove the inherent ambiguity in this blind deconvolution problem, we introduce relevant a priori information. Our major assumption is sparseness of the reflectivity, which corresponds to a layered earth model. This allows non-minimum phase wavelets to be recovered reliably and closely spaced reflectors to be resolved. To combine a priori knowledge and data, we use a Bayesian framework and derive a maximum a posteriori estimate. Computing this estimate is a difficult optimization problem which is solved by a sub-optimal iterative procedure. The procedure alternates steps of wavelet estimation and reflectivity estimation. The first step only requires a simple least squares fit, while the second step is solved by the iterated window maximization algorithm recently proposed by Kaaresen. This enables better efficiency and optimality than established alternatives. The resulting optimization method can easily handle multichannel models with only a moderate increase of the computational load. Lateral continuity of the reflectors is achieved by modeling local dependencies between neighboring traces. Major improvements in both wavelet and reflectivity estimates are obtained by taking the wavelet to be invariant across several traces. The practicality of the algorithm is demonstrated on synthetic and real seismic data. An application to multivariate well log segmentation is also given.

I. INTRODUCTION

In reflection seismology, a short duration acoustic wave is transmitted into the ground and the reflected energy due to impedance changes in the earth is measured. A two or three dimensional section of the earth is mapped by using a large number of different source and receiver positions at the surface. For simplicity, we consider a 2-D (poststack) section, corresponding to a number of roughly vertical and parallel 1-D traces. Under simplifying assumptions, one such trace can be modeled as a noise corrupted 1-D convolution between the vertical reflectivity function of the earth and an unknown wavelet [1]. To increase resolution and enable identification of closely spaced reflectors, deconvolution is routinely used to (partially) remove the effect of the wavelet. Due to various effects such as high frequency absorption, receiver ghosting, etc., the effective wavelet is usually unknown and must be estimated together with the reflectivity. Such *blind* deconvolution is considerably more difficult than *simple* deconvolution where the wavelet is known a priori. The problem is highly ill-posed and a priori knowledge must be introduced to achieve a unique solution.

Conventional deconvolution procedures are often based on the (implicit) assumption that both the reflectivity and noise are stationary Gaussian processes. This leads to computationally convenient linear solutions such as Wiener filtering [1], [2] or minimum variance deconvolution [3]. However, Gaussianity is not realistic for the sparse nature of many actual reflectivities, and less than optimal resolution improvement may result. Furthermore, a Gaussian reflectivity distribution does not allow the phase of the wavelet to be estimated and a questionable minimum phase assumption is commonly used. Homomorphic filtering techniques [4], [5] can estimate the phase of the wavelet, but tend to be highly sensitive to noise. Methods based on higher order statistics [6] can identify an arbitrary wavelet, as long as the reflectivity is a stationary, white, and non-Gaussian random process. These have theoretically pleasing asymptotic properties, but will often be computationally intensive and require large amounts of data to give satisfactory estimates [7]. It is thus of interest to introduce more a priori information in order to improve the estimates for finite amounts of data.

In this paper, we formulate the problem in a Bayesian framework. Pertinent a priori information is introduced by considering the unknowns as realizations of stochastic variables with certain *prior* distributions. Then, the unknowns are estimated by the maximum a posteriori (MAP) estimate, which is the maximizer of the *posterior* (conditional) probability density of the unknowns given the observed data. This will remove the inherent ambiguity in the blind deconvolution problem by selecting the (a priori) most probable estimate among all those approximately equally compatible with the observed data. Our major a priori assumptions are:

(a) The reflectivities are sparse, i.e. only a limited number of randomly located coefficients are non-zero. In particular, we shall use the Bernoulli-Gaussian distribution [3], and generalizations thereof. This corresponds to a layered earth model. Provided the layers are sufficiently homogeneous, the reflectivity will be approximately zero everywhere except at layer boundaries.

K. F. Kaaresen is with the Department of Mathematics, University of Oslo, Box 1053 Blindern, N-0316 Oslo, Norway. E-mail: kjetilka@math.uio.no

Torfinn Taxt is with the Department of Informatics, University of Oslo, Gustadalleen 23, Box 1080 Blindern, N-0316 Oslo, Norway.

(b) The reflectivities tend to be continuous in the horizontal direction, i.e. we assume the layers of the earth to be mainly horizontal and continuous.

(c) The wavelet is common to several traces.

Assumption (a) follows the pioneering works of Mendel and coworkers [3], [8], [9], [10], [11]. Thus, our computational solutions will also be related to theirs, but with some important differences, as discussed below. Assumption (b) is related to those used by Lavielle [12] and Idier and Goussard [13] for (simple) *multichannel* deconvolution. It aims at improving the estimates by using information from neighboring traces. Assumption (c) is simple, but often realistic and leads to large improvements when applicable.

The two last assumptions can be used in combination or separately, whereas the first is essential to our approach. This sparsity assumption removes the ill-posed nature of the problem, but the price to pay is a posterior with a difficult shape and many local maxima. Following Mendel [3], we attack the problem by a two-stage procedure. First we fix the wavelet and maximize with respect to the reflectivity. Then we fix the reflectivity and maximize with respect to the wavelet. Iterating these two steps will at least lead to a local maximum of the posterior. Although convergence to the truly global maximum may depend on the starting point, the simulations presented in this paper indicate that a satisfactory estimate will usually be found. This iterative method is commonly referred to as a *block component method* (BCM) [3], but can also be considered as an instance of the iterated conditional modes technique [14] with vector valued variables.

The reflectivity estimation step is computationally hard and must be solved by sub-optimal methods [9]. Here, we use the iterated window maximization (IWM) algorithm recently proposed in [15]. IWM relies on local updating and pre-storing of some key quantities to achieve an efficient iterative search. Similar to iterated conditional modes, the search increases the posterior at each step, but instead of modifying only one variable at each step, various sub-groups of variables are updated simultaneously. This allows complicated "transitions", such as moving several reflectors, to take place in one step. Thereby the problem of convergence to local maxima is greatly reduced. The resulting optimality is higher than for the single most likely replacement (SMLR) detector [9] which is the major component of Mendel's BCMs. Also, IWM uses a finite impulse response (FIR) representation of the wavelet. This makes the wavelet estimation step particularly simple. It is only necessary to solve a certain linear system with a Toeplitz structure and dimensions equal to the length of the wavelet. In contrast, Mendel's state-space solutions use an iterative Marquardt–Levenberg optimization to estimate the parameters of the state-space model. The FIR solution of

IWM will also be faster in many cases. In particular, for wavelets which can only be fitted by high-order models, Mendel's state-space solutions may be too slow to be practical.

In the following, the overall BCM is discussed first. Then we focus on the separate solutions to reflectivity and wavelet estimation problems. The solutions are given for the single-channel situation first. Then, we consider multichannel extensions. Finally, synthetic and real data results are presented and discussed. In addition to seismic results, an application of the method to multivariate segmentation of filtered and noisy well log data is also given.

II. MODEL AND BLOCK COMPONENT METHOD

Our outset is the convolutional model

$$z(n) = \sum_{k=0}^{K} x(n-k)h(k) + e(n), \quad n = 1, 2, \ldots, N, \quad (1)$$

where z is an observed seismic trace, x is the reflectivity of the earth, h is the seismic wavelet, and e is additive noise. The FIR wavelet, h, is non-zero only for $k = 0, 1, \ldots K$. To avoid separate treatment of some border conditions, we also assume that observed record contains no truncated reflections, i.e. $x(n) = 0$ for $n \leq 0$ and for $n > N - K$. Due to the sparsity assumption, $x(n)$ can be represented more efficiently by two vectors; $\mathbf{t}$ representing the (time) positions of the reflectors and $\mathbf{a}$ representing their amplitudes. The relation between x, $\mathbf{a}$ and $\mathbf{t}$ is

$$x(n) = \sum_{i=1}^{M} a_i \delta(n - t_i), \quad (2)$$

where M denotes the (stochastic) number of reflectors. Inserting (2) in (1), gives

$$z(n) = \sum_{i=1}^{M} h(n - t_i)a_i + e(n), \quad n = 1, 2, \ldots, N. \quad (3)$$

The two equivalent convolutional models (1) and (3) are central. The first will be used for wavelet estimation, while the second is better suited for reflectivity estimation. Model (3) can be written in matrix form as

$$\mathbf{z} = \mathbf{Ha} + \mathbf{e}, \quad (4)$$

with obvious interpretation of $\mathbf{z}$ and $\mathbf{e}$ and $H_{ni} = h(n - t_i)$ for $n = 1, 2, \ldots, N$ and $i = 1, 2, \ldots, M$. Let also $\mathbf{h}$ be a vector consisting of the samples of $h(k)$ for $k = 0, 1, \ldots K$.[1]

---

[1] To allow a common index range for the vector and sequence we shall index $\mathbf{h}$ from zero. Similar conventions will be used for other vectors and matrices with the same dimension as $\mathbf{h}$.
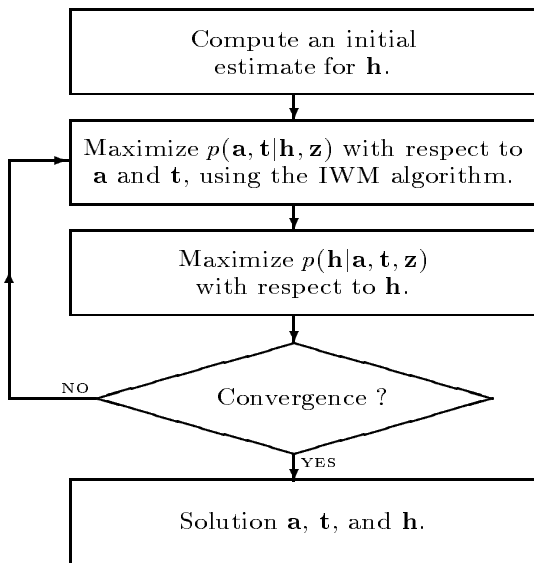
Fig. 1. The overall structure of the block component method.

We use the following distributional assumptions: The noise $\mathbf{e}$ is zero mean Gaussian and white, independent of $\mathbf{a}$ and $\mathbf{t}$, and has variance σ_e^2. The conditional distribution for $\mathbf{a}$ given $\mathbf{t}$ is also zero mean Gaussian, having independent components with variance σ_a^2. The prior for $\mathbf{t}$ is a geometric process, i.e. $t_{i+1} - t_i$ is geometrically distributed with parameter λ and the increments are independent (modifications will be considered later). Note that the distribution of $(\mathbf{a}, \mathbf{t})$ is equivalent to a Bernoulli-Gaussian distribution for x. Finally, we take the prior for $\mathbf{h}$ to be Gaussian, independent of $\mathbf{a}$, $\mathbf{t}$ and $\mathbf{e}$, and having independent components with variance σ_h^2.

Together with the convolutional model (3) these distributional assumptions define the posterior density, $p(\mathbf{a}, \mathbf{t}, \mathbf{h} | \mathbf{z})$. The purpose of the BCM is to maximize this function with respect to $\mathbf{a}$, $\mathbf{t}$ and $\mathbf{h}$. To maximize with respect to $\mathbf{a}$ and $\mathbf{t}$ for fixed $\mathbf{h}$, we use elementary conditional probability and factor the posterior as $p(\mathbf{a}, \mathbf{t}, \mathbf{h} | \mathbf{z}) = p(\mathbf{a}, \mathbf{t} | \mathbf{h}, \mathbf{z}) p(\mathbf{h} | \mathbf{z})$. Since $\mathbf{h}$ is fixed, the second factor is constant and can be ignored. Maximization of the first factor is precisely the *simple* deconvolution problem solved in [15], and reviewed in the next section. This is the first step of the BCM. The second step is to maximize with respect to $\mathbf{h}$ for fixed $\mathbf{a}$ and $\mathbf{t}$. In this case, we factor the posterior $p(\mathbf{a}, \mathbf{t}, \mathbf{h} | \mathbf{z}) = p(\mathbf{h} | \mathbf{a}, \mathbf{t}, \mathbf{z}) p(\mathbf{a}, \mathbf{t} | \mathbf{z})$. Again the second factor is constant, and we only need to maximize the first factor. The problem thus reduces to finding the MAP estimate of the wavelet in the simple situations with known reflectivity. The solution is derived in Section IV. Iterating these two steps gives the overall solution as depicted in Fig. 1.

The structure of our BCM differs from Mendel's in that we only estimate the reflectivity sequence and the wavelet. Other parameters are treated as inputs which should be specified by the user or estimated from other sources. In the simplest setting there is effectively only one such parameter. This is a "sparseness" parameter which will be discussed later. Conceivably, estimation of this parameter could be included as a third step of the BCM, but we prefer not to do so. In our opinion, fixation of the sparseness parameter presents exactly the necessary a priori input to *regularize* [16] the originally highly ill-posed blind deconvolution problem. This is supported by studies based on a simplified model [17] which shows that Mendel's approach may produce non-meaningful parameter estimates at the border of the parameter space. Fixing the sparseness parameter avoids such anomalies.

III. REFLECTIVITY ESTIMATION

A. Iterated Window Maximization

We now consider the wavelet as fixed and use the IWM algorithm to maximize $p(\mathbf{a}, \mathbf{t} | \mathbf{h}, \mathbf{z})$. The optimization can be separated into two sub-problems: Detection of the reflector positions $\mathbf{t}$, and estimation of the reflector amplitudes $\mathbf{a}$ for given $\mathbf{t}$. Due to the Gaussian assumptions on $\mathbf{a}$ and $\mathbf{e}$, the last problem is easy. Maximizing the posterior with respect to $\mathbf{a}$ for given $\mathbf{t}$ only amounts to a linear least squares fit. The optimal value determined by this fit, $\hat{\mathbf{a}}$, will depend on $\mathbf{t}$. Inserting $\hat{\mathbf{a}}$ into $p(\mathbf{a}, \mathbf{t} | \mathbf{h}, \mathbf{z})$, results in a function of $\mathbf{t}$ only, $p(\hat{\mathbf{a}}, \mathbf{t} | \mathbf{h}, \mathbf{z})$. Unfortunately, this function is highly non-linear with numerous local maxima. Exact maximization is beyond reach for realistic signal sizes. Instead, IWM proceeds iteratively through a series of small perturbations to $\mathbf{t}$. Only changes which increase $p(\hat{\mathbf{a}}, \mathbf{t} | \mathbf{h}, \mathbf{z})$ are accepted.

Since each evaluation of $p(\hat{\mathbf{a}}, \mathbf{t} | \mathbf{h}, \mathbf{z})$ requires refitting of the amplitudes, the efficiency of this operation is vital. The solution is to constrain each update to a local window. To compare two $\mathbf{t}$ values which are identical outside the window, only the amplitudes within the window need to be recomputed. This reduces the dimension of the least-squares fit to the number of reflectors within the window. Fast inversion of the linear system would be of no use, however, if compiling the coefficients was time consuming. Fortunately, a simple relationship exists between the coefficients and two correlation functions. Precomputing and storing the correlation functions allows repeated initialization to take place with minimal effort during the iterations.

The necessary formulas are now stated, cf. [15] for derivation. Assume a window w has been chosen. Partition $\mathbf{t}$ in the components inside the window, $\mathbf{t}^w$, and the components outside the window, $\mathbf{t}^{\overline{w}}$. Partition $\mathbf{a}$ likewise and let $\mathbf{H}^w$ be the columns of $\mathbf{H}$ corresponding to reflectors inside the window and $\mathbf{H}^{\overline{w}}$ be the columns corresponding to reflectors outside the window. Define the

inverse signal-to-noise ratio $\gamma_a = \sigma_e^2/\sigma_a^2$ and introduce the matrices

$$\mathbf{S}^w = (\mathbf{H}^w)'\mathbf{H}^w + \gamma_a\mathbf{I} \qquad (5)$$

and

$$\mathbf{v}^w = (\mathbf{H}^w)'\mathbf{z} - (\mathbf{H}^w)'\mathbf{H}^{\overline{w}}\mathbf{a}^{\overline{w}}. \qquad (6)$$

The optimal value of $\mathbf{a}^w$, given $\mathbf{a}^{\overline{w}}$, $\mathbf{t}$ and $\mathbf{z}$ is

$$\hat{\mathbf{a}}^w = (\mathbf{S}^w)^{-1}\mathbf{v}^w. \qquad (7)$$

The $(\mathbf{t}^w, \hat{\mathbf{a}}^w)$ pair corresponding to the largest value of the posterior is the one which yields the larger value of the criterion

$$l(\mathbf{t}) = (\mathbf{v}^w)'\hat{\mathbf{a}}^w + g(\mathbf{t}). \qquad (8)$$

Here, the first term measures the fit to the data, whereas $g(\mathbf{t})$ acts as a "penalty" term which favors configurations that are compatible with the prior assumptions on $\mathbf{t}$. Under the Bernoulli-Gaussian assumption this term assumes the particularly simple form

$$g(\mathbf{t}) = -\theta M^w, \qquad (9)$$

where M^w is the number of reflectors inside the window and the "sparsity parameter" θ is discussed in Section III-C. Note also, for later use, that equation (8) is valid for general $p(\mathbf{t})$. In this case the penalty term is determined by the prior for $\mathbf{t}$ by a relationship on the form

$$g(\mathbf{t}) = \theta_1 \ln p(\mathbf{t}) - \theta_2 M^w, \qquad (10)$$

for suitable parameters θ_1 and θ_2.

The correlation functions which should be computed and stored prior to the iterations are

$$c_{hh}(n) = \sum_{k=n}^{K} h(k-n)h(k), \quad n = 0, 1, \ldots, K \qquad (11)$$

and

$$c_{hz}(n) = \sum_{k=\max(1,n)}^{\min(N,n+K)} h(k-n)z(k), \quad n = 1, 2, \ldots, N. \qquad (12)$$

Using these the initialization is performed by

$$S_{ij}^w = c_{hh}(|t_i^w - t_j^w|) + \gamma_a\delta(i-j) \qquad (13)$$

and

$$v_i^w = c_{hz}(t_i^w) - \sum_l a_l^{\overline{w}}c_{hh}(|t_i^w - t_l^{\overline{w}}|), \qquad (14)$$

for $i, j = 1, 2, \ldots M^w$. In the last sum only reflectors producing "overlapping" reflections with those inside the window need to be taken into account, i.e. the sum can be confined to l such that $|t_i^w - t_l^{\overline{w}}| \le K$. In equation (13), $\delta(\cdot)$ denotes the Dirac delta function which is one if its argument is zero and zero elsewhere.

The major computational effort in comparing two competing $\mathbf{t}^w$ values is the inversion in equation (7). Provided the number of reflectors within the window, M^w, is not large, the comparison will thus be fast. The matrix $\mathbf{S}^w$ is symmetric positive definite, so it is only necessary to initialize e.g. the upper triangle and Cholesky decomposition [18] can be used to solve the linear system. Since M^w is typically very small, e.g. 1-3, it will be even faster to treat each of the low-order cases separately and use the simple symbolic solutions of the linear system.

B. The iterative search

Based on the formulas of the preceding section many slightly different search procedures can be constructed. The procedure used to process the data presented here is as follows: The window positions are repeatedly scanned systematically through all currently existing reflectors (components of $\mathbf{t}$). For each reflector a fixed size window is centered at that reflector. Within this window, new "candidate" reflector configurations are obtained by systematically trying all changes in pre-determined *transition set*. The transition set includes the following one-reflector transitions: Insert a new reflector in the window, delete the center reflector, move the center reflector with one or more samples in either direction. The criterion (8) is computed for the original reflector configuration and for all new candidate configurations. If a candidate configuration improves on the original one, a replacement is made. In this case both $\mathbf{t}$ and $\mathbf{a}$ are updated. If no improvements are found, $\mathbf{a}$ is still updated with the refitted amplitudes corresponding to the original value of $\mathbf{t}$. The procedure is repeated until a complete scan fails to change either $\mathbf{t}$ or $\hat{\mathbf{a}}$.

Since only some components of $\mathbf{a}$ are refitted in each step, the vector as a whole will not be truly globally optimal for the current value of $\mathbf{t}$ during the iterations. However, because refitting is performed with systematically changing window positions, the global fit will improve "concurrently" with the changing of $\mathbf{t}$. And as soon as $\mathbf{t}$ ceases to change, the fit will converge fast to the global optimum given the final value of $\mathbf{t}$ [15]. Note, however, that this optimality of the amplitudes is relative to the final $\mathbf{t}$, which may still be sub-optimal.

The optimality of the algorithm depends on the richness of the transition set. A limited transition set increases the risk that the algorithm will terminate in a local maximum for $\mathbf{t}$. On the other hand, if too many transitions must be tried at each step, the algorithm is slowed down. Choosing the transition set is thus a cost/optimality trade-off. Though good results were obtained with the one-reflector transitions above, the implementation used here was also extended by a set of two-reflector transitions for increased optimality. These consisted of joining two reflectors into one, splitting one into two, and moving two simultaneously.

C. Parameter values and window size

In the Bernoulli-Gaussian case, the algorithm depend on two parameters, γ_a and θ. The MAP criterion determines both as functions of the distributional parameters: σ_e^2, σ_a^2, and λ. However, even if the true values of the distributional parameters were known, it is not obvious that plugging in these gives the best performance. In particular for θ, it is argued in [15] that the value determined by the MAP criterion generally can not be used. Instead, we prefer to consider θ as a "tuning parameter". From equation (9) it is clear that this parameter determines the penalty connected with the number of reflectors. This penalty is of crucial importance since the fit to the data [as measured by the first term of equation (8)] will always increase by adding another reflector. Setting $\theta = 0$ will therefore lead to a dense estimate with reflectors at all positions. Increasing θ leads to sparser estimates. A simple solution is therefore to adjust θ until the sparsity nature of the estimate is consistent with a priori assumptions. (Other solutions are to "train" the algorithm to maximize performance on a large synthetic data set, as measured by a realistic loss function [19], or to use "L-curve" analysis from regularization theory [20].) As a rule of thumb it is argued in [15] that θ should be chosen proportional to the noise variance, $\theta = \chi \sigma_e^2$. The simulations given here (and numerous others) show that selecting the proportionality constant χ in the range $20 - 50$ usually works well for blind deconvolution. Note that the argument depends on correctness of the assumed model. In practice the "effective" noise level must also include various forms of model mismatch, such as a slightly time varying wavelet, etc.

The inverse signal-to-noise ratio γ_a appearing in equation (13), is less important. Its effect is to penalize large amplitudes [15]. If estimates or a priori knowledge about the noise and amplitude variances are at hand, it can be determined from its definition. Another simple solution which usually works well, is to use $\gamma_a = 0$ which corresponds to a non-informative prior [21] for the amplitudes with infinite variance. This approach was taken in all examples presented here.

The window size primarily affects the execution speed. Large windows will include many reflectors and make the inversion in equation (7) costly. On the other hand, very small windows may reduce the optimality of the algorithm. Though convergence to globally optimal amplitudes given the final $\mathbf{t}$ is still guaranteed, small windows may reduce the ability of the algorithm to escape from local maxima for $\mathbf{t}$. Simulation experiments show that a window size approximately equal to the length of the wavelet (excluding small samples at each end), will usually give efficient execution combined with no significant loss of optimality. This approximate rule was used for all examples presented here.

IV. Wavelet estimation

We now determine the MAP estimator of the wavelet for fixed reflectivity, i.e. the maximizer of $p(\mathbf{h}|\mathbf{a}, \mathbf{t}, \mathbf{z})$. Rewriting the convolutional model (1) in matrix form gives

$$\mathbf{z} = \mathbf{X}\mathbf{h} + \mathbf{e}, \tag{15}$$

where $X_{nk} = x(n - k)$ for $n = 1, 2, \ldots N$ and $k = 0, 1, \ldots, K$. Due to the Gaussian assumptions on $\mathbf{h}$ and $\mathbf{e}$, the MAP estimator in this linear model has the well known form (see e.g. [22])

$$\hat{\mathbf{h}} = \mathbf{B}^{-1}\mathbf{u} \tag{16}$$

where

$$\mathbf{B} = \mathbf{X}'\mathbf{X} + \gamma_h \mathbf{I} \quad \text{and} \quad \mathbf{u} = \mathbf{X}'\mathbf{z}. \tag{17}$$

The parameter $\gamma_h = \sigma_e^2/\sigma_h^2$ is an inverse signal-to-noise ratio. All points from the discussion on γ_a in the previous section apply. In particular, the non-informative limit $\gamma_h = 0$ usually works well, and was used for all reconstructions presented here. Note that in this case the solution (16) reduces to the standard least-squares estimate.

Computing $\mathbf{B}$ by the matrix multiplication indicated by equation (17) would be unnecessarily burdensome. Instead we take advantage of the assumption that $x(n) = 0$ for $n \leq 0$ and for $n > N - K$. This implies that $\mathbf{B}$ is Toeplitz, i.e. each diagonal band is constant. Using equation (17) and the definition of $\mathbf{X}$ shows that the k'th band, b_k, is given by

$$b_k = \sum_{n=1}^{N} x(n - k)x(n) + \gamma_h \delta(k), \tag{18}$$

for $k = -K, \ldots, 0, \ldots, K$. Note that since $\mathbf{B}$ is symmetric, it is only necessary to compute the non-negative bands. The components of $\mathbf{u}$ are immediately seen to be

$$u_k = \sum_{n=1}^{N} x(n - k)z(n), \quad k = 0, 1, \ldots, K. \tag{19}$$

The entire procedure for wavelet estimation is thus to initialize $\mathbf{B}$ and $\mathbf{u}$ by equations (18) and (19) and then perform the inversion in equation (16) by Levinson recursion [18].

The simplicity of the above estimate depends on a Gaussian prior for $\mathbf{h}$, but the particular correlation structure is not crucial. With a completely general covariance matrix, $\mathbf{\Sigma}_h$, the necessary modification is to replace the first part of equation (17) by $\mathbf{B} = \mathbf{X}'\mathbf{X} + \sigma_e^2 \mathbf{\Sigma}_h^{-1}$. This would allow more sophisticated features of the wavelet to be modeled. For example, one may encourage the estimate to be continuous and to have smaller samples at

the end. Incorporating more a priori information in this fashion is most important when the data is scarce. In seismic applications the data available for wavelet estimation is usually abundant. We therefore tend to favor the simpler non-informative approach above. Using a non-Toeplitz $\mathbf{\Sigma}_h$ would also destroy the Toeplitz property of $\mathbf{B}$.

V. MULTICHANNEL REFLECTIVITY ESTIMATION

Consider now a seismic section consisting of many parallel (and vertical) traces. Due to horizontal continuity of the reflecting surfaces of the earth, the traces will be dependent. For simplicity, we assume that these inter-trace dependencies can be modeled by the prior for the reflector positions, $p(\mathbf{t})$. Arguably, some continuity should also be expected for the amplitudes. However, as will be seen, confining the dependencies to $\mathbf{t}$ leads to major simplifications. Also, some simulation experiments indicate that the difficult problem is to estimate $\mathbf{t}$. If a correct estimate of $\mathbf{t}$ is obtained, the required continuity of the amplitudes will usually be enforced by the data.

The multichannel version of the convolutional model (3) is

$$z_c(n) = \sum_{i=1}^{M_c} h_c(n - t_{c,i}) a_{c,i} + e_c(n), \qquad (20)$$

where $c = 1, 2, \ldots, C$ indexes the different channels or traces. The wavelet h_c may or may not depend on c. We assume, as before, independent Gaussian distributions for the amplitudes and the noise, but the variances may depend on c, i.e. $\mathrm{Var}\, a_{c,i} = \sigma_{a,c}^2$ and $\mathrm{Var}\, e_c(n) = \sigma_{e,c}^2$.

A. Generalization of IWM

To adapt the IWM algorithm to this model, we take advantage of the fact that the derivation of equations (5)-(10) in [15] does not depend on the form of $\mathbf{H}$. In order to use previous results, we reduce the multichannel model to the matrix form (4) but with a new definition of $\mathbf{H}$. To accomplish this, concatenate the data, reflector positions, amplitudes, and noise for each trace into the vectors: $\mathbf{z}_c$, $\mathbf{t}_c$, $\mathbf{a}_c$, and $\mathbf{e}_c$. Define also the matrices $\mathbf{H}_c$ by $H_{c,ni} = h_c(n - t_{c,i})$. Finally, introduce the following aggregate quantities:

$$\mathbf{z} = \begin{pmatrix} \frac{1}{\sigma_{e,1}} \mathbf{z}_1 \\ \vdots \\ \frac{1}{\sigma_{e,C}} \mathbf{z}_C \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_C \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} \frac{1}{\sigma_{a,1}} \mathbf{a}_1 \\ \vdots \\ \frac{1}{\sigma_{a,C}} \mathbf{a}_C \end{pmatrix},$$

$$\mathbf{e} = \begin{pmatrix} \frac{1}{\sigma_{e,1}} \mathbf{e}_1 \\ \vdots \\ \frac{1}{\sigma_{e,C}} \mathbf{e}_C \end{pmatrix}, \quad \mathbf{H} = \begin{pmatrix} \frac{\sigma_{a,1}}{\sigma_{e,1}} \mathbf{H}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \frac{\sigma_{a,C}}{\sigma_{e,C}} \mathbf{H}_C \end{pmatrix}. \qquad (21)$$

Note that $\mathbf{a}$ and $\mathbf{e}$ are weighted such that all components have equal (unit) variance. Thus, with the definitions (21), the distributional assumptions from the single-channel situation are fulfilled. It is also easily seen that with these definitions model (20) can be written in the matrix form (4). Thus, equations (5)-(10) are valid in terms of definitions (21). Using equations (5)-(7) and the block diagonal form of $\mathbf{H}$ shows that computation of $\hat{\mathbf{a}}^w$ reduces to a separate problem for each trace. The solution can be written as

$$\hat{\mathbf{a}}_c^w = (\mathbf{S}_c^w)^{-1} \mathbf{v}_c^w, \qquad (22)$$

where

$$\mathbf{S}_c^w = (\mathbf{H}_c^w)' \mathbf{H}_c^w + \gamma_{a,c} \mathbf{I}, \qquad (23)$$

$$\mathbf{v}_c^w = (\mathbf{H}_c^w)' \mathbf{z}_c - (\mathbf{H}_c^w)' \mathbf{H}_c^{\overline{w}} \mathbf{a}_c^{\overline{w}}, \qquad (24)$$

and $\gamma_{a,c} = \sigma_{e,c}^2 / \sigma_{a,c}^2$. The superscript w should be interpreted as before, i.e. $\mathbf{H}_c^w$ is the columns of $\mathbf{H}_c$ corresponding to reflectors inside the window, etc. Equations (22)-(24) are exactly as in the single-channel situation, except for the dependence on c. Initialization of the matrices $\mathbf{S}_c^w$ and $\mathbf{v}_c^w$ can thus be performed completely analogously to (13)-(14), but if h_c depends on c, one set of correlation functions on the form (11)-(12) will be needed for each trace.

Inserting (21) in (5) and (6) also shows that the criterion (8) for comparing two candidate values of $\mathbf{t}$ becomes

$$l(\mathbf{t}) = \sum_{c=1}^{C} \sigma_{e,c}^{-2} (\mathbf{v}_c^w)' \hat{\mathbf{a}}_c^w + g(\mathbf{t}). \qquad (25)$$

In conclusion, the multichannel case is handled by computing the amplitudes separately for each trace exactly as before. The competing $\mathbf{t}$ values are compared by computing a sum consisting of one term for each trace. Each term is on the form considered before, but weighted by the inverse of the noise variance for that trace. The dependencies across the traces are completely contained in the term $g(\mathbf{t})$, which is defined in terms of the prior distribution for $\mathbf{t}$ by (10).

B. Window Selection

The above theory is valid for general windows, i.e. a window can be any selection of sample points from one or more of the traces. The simplest possibility is to select windows which cover an interval in only one trace. In this case the sum in equation (25) reduces to one term. (Equation (25) is still valid in the case when one or more traces have no reflectors inside the window, provided an empty vector is interpreted as zero.) For the seismic data results presented here, this approach was used. The window positions were systematically scanned through the first trace. Then, the second trace was scanned. When all traces had been scanned once, another scan of the

first trace was performed, and so on. Note that this gives higher optimality than the recursive solution proposed by Idier and Goussard [13]. Since the recursive solution only process each trace once, it has the weakness that data from the later traces will not influence the estimate in the earlier traces. Additionally, the iterative scan proposed here will be almost as fast as a recursive strategy since all traces will converge "concurrently" towards their solution.

If there are strong dependencies across the traces, windows covering only one trace may not be sufficient. For an extreme example, consider a prior which enforces the deterministic relation $\mathbf{t}_1 = \mathbf{t}_2 = \ldots = \mathbf{t}_C$, i.e. the reflector positions are the same for all traces. This prior was used for processing the multivariate well log data presented here. Since a fundamental assumption for the IWM algorithm is that changes to $\mathbf{t}$ can only be made within the window, the windows must in this case cover an interval in all traces simultaneously. Thus, in this case (25) will contain one term from each trace. The transition sets from the single-channel situation can, however, be used unchanged.

The most difficult situation is when the prior includes strong but not deterministic inter-trace dependencies. Determination of suitable strategies for selection of windows and transitions in this situation is basically an open problem. Note, however, that for the seismic data results presented later, inclusion of some rather simple multi-trace updates gave clear improvements compared to only single-trace updates.

C. Priors

The derived optimization method gives large flexibility in modeling the prior for $\mathbf{t}$. In theory, any prior can be used, but to achieve efficient execution the priors must allow fast evaluation and also ensure a sparse estimate. Markov random field priors have been considered by Lavielle [12] and Idier and Goussard [13]. Any of these could be used in the present framework. The actual prior used for the seismic data results presented here is most closely related to those of Lavielle [12] and is described next.

The relationship (10) could be used to model $p(\mathbf{t})$ and transfer to a value for $g(\mathbf{t})$. However, we prefer to model $g(\mathbf{t})$ directly, thinking in terms of "penalizing" unwanted/unlikely configurations. (In Markov random field terms, we model the associated *energy function* [23]). To encourage horizontal continuity of the estimated reflectors, we used the penalty function

$$g(\mathbf{t}) = -\theta M^w + \nu^- M^- + \nu^{\backslash} M^{\backslash} + \nu^{/} M^{/} - \nu^{|} M^{|} - \nu^{||} M^{||},$$
(26)

where

$$
\begin{aligned}
M^- &= \#\{t_{c,i} = t_{c+1,j}\}, \\
M^{\backslash} &= \#\{t_{c,i} = t_{c+1,j} - 1\}, \\
M^{/} &= \#\{t_{c,i} = t_{c+1,j} + 1\}, \quad (27) \\
M^{|} &= \#\{t_{c,i} = t_{c,j} + 1\}, \\
M^{||} &= \#\{t_{c,i} = t_{c,j} + 2\}.
\end{aligned}
$$

The first term in equation (26) penalizes the number of reflectors and is on the form considered before. The second term encourages horizontal continuity. The number M^- counts how many times two reflectors in adjacent traces are placed the same horizontal position, while the parameter ν^- determines the strength of the encouragement. (The counts in equation (27) should be taken over all c, i, and j that make sense.) Similarly, the third and fourth term also encourage horizontal continuity, but for dipping or raising layers. The parameters ν^-, $\nu^{\backslash}$, and $\nu^{/}$ must not be chosen too large relative to θ. To see why, consider insertion of a reflector at a horizontal position where one of the neighboring traces also has a reflector. Inserting the reflector will add one to the count M^w and one to the count M^-. The overall change to $g(\mathbf{t})$ will therefore be $\nu^- - \theta$. If this number is positive the reflector will certainly be inserted (since inserting a reflector always improves the fit to the data). Choosing $\nu^- > \theta$ will therefore "prohibit" blind ends in a line of reflectors. Similarly, choosing $\nu^- > \theta/2$ will prohibit single holes. The same comments apply to $\nu^{\backslash}$ and $\nu^{/}$. Note also that if one trace has several vertically closely spaced reflectors, inserting a reflector in a neighboring trace can possibly add to all of M^-, $M^{\backslash}$ and $M^{/}$, i.e. the reflector can similarly take part in both horizontal and dipping and rising layers. This could lead to clustering of large numbers of reflectors, and is the reason for including the last two terms in the penalty function (26). These remove the problem by penalizing vertically close reflectors. For the examples presented here, the parameters were heuristically selected as $\nu^- = \nu^{\backslash} = \nu^{/} = 0.7\theta$, and $\nu^{|} = \nu^{||} = 2\theta$. (We thus allowed reflector chains with blind ends, but not single holes).

Computation of the penalty function (26) presents no problems. To compare two competing $\mathbf{t}$ values, it is only necessary to compute the changes to the counts (27). Since the algorithm only modifies a few reflector positions at each step, the changes can be determined by examining a few neighboring positions.

VI. MULTICHANNEL WAVELET ESTIMATION

Assuming wavelet invariance across traces implies that multichannel wavelet estimates can be obtained by simple averaging operations. To see this, define the aggre-

gate quantities:

$$\mathbf{z} = \begin{pmatrix} \frac{1}{\sigma_{e,1}}\mathbf{z}_1 \\ \vdots \\ \frac{1}{\sigma_{e,C}}\mathbf{z}_C \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} \frac{1}{\sigma_{e,1}}\mathbf{e}_1 \\ \vdots \\ \frac{1}{\sigma_{e,C}}\mathbf{e}_C \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \frac{1}{\sigma_{e,1}}\mathbf{X}_1 \\ \vdots \\ \frac{1}{\sigma_{e,C}}\mathbf{X}_C \end{pmatrix}. \tag{28}$$

Here $\mathbf{z}$ and $\mathbf{e}$ correspond to the definitions in the preceding section and $X_{c,nk} = x_c(n-k)$. Since we assume $\mathbf{h}$ to be common to all traces, model (15) is valid with definitions (28). Also, since the noise from different traces is taken to be independent, $\mathbf{e}$ has the same distribution as in the single-channel situation. Thus, equations (16)-(17) still give the MAP estimate, now in terms of the aggregate quantities (28). Inserting definitions (28) in equation (17) gives

$$\mathbf{B} = \sum_{c=1}^{C} \sigma_{e,c}^{-2} \mathbf{X}_c' \mathbf{X}_c + \gamma_h \mathbf{I} \tag{29}$$

and

$$\mathbf{u} = \sum_{c=1}^{C} \sigma_{e,c}^{-2} \mathbf{X}_c' \mathbf{z}_c. \tag{30}$$

From equation (29) the k'th band of $\mathbf{B}$ is seen to be

$$b_k = \sum_{c=1}^{C} \sigma_{e,c}^{-2} b_{k,c} + \gamma_h \delta(k), \tag{31}$$

where, in analogy with equation (18), $b_{k,c} = \sum_{n=1}^{N} x_c(n-k)x_c(n)$. Similarly, using equation (30) gives

$$u_k = \sum_{c=1}^{C} \sigma_{e,c}^{-2} u_{k,c}, \tag{32}$$

where $u_{k,c} = \sum_{n=1}^{N} x_c(n-k)z_c(n)$ is analogous to equation (19).

Thus, MAP estimation of $\mathbf{h}$ based on all traces takes place as before, except that $\mathbf{B}$ and $\mathbf{u}$ must be initialized by equations (31) and (32). These are on the form considered previously, except that one term is added for each trace. Note in particular that the dimension of the Toeplitz system which must be inverted does not depend on the number of traces. It is still determined by the wavelet length. If the wavelet is slowly varying across the traces the above sums could be modified in an obvious way to (weighted) averages of only the few closest traces.

VII. Experimental Results

A. Implementation Issues

To apply the procedure in practice, the BCM iterations need a starting point. One possibility is to start from some conventional, possibly minimum phase, wavelet estimate. An even simpler solution, which made the overall procedure self contained, was used here. The BCM was started from an initial *reflectivity* estimate formed by inserting reflectors at all local maxima of the data record. (A point was considered as a local maxima if it had larger absolute value than the 40 nearest samples.) In spite of the crudeness of this initial estimate, all presented reconstructions converged fast to sensible values.

Also, the length D of the wavelet must be decided. In the computer simulations we used the true length (51 samples) of the FIR wavelet used to generate the data. In the real data examples we performed a trial run with a large value, and then reduced the value to remove small samples at each end which appeared to contain only noise. (A suitable alignment of the wavelet within its finite size record was obtained by shifting the initial reflectivity estimate described above.)

Combining non-informative priors for $\mathbf{h}$ and $\mathbf{a}$ expose an inherent indeterminacy in the convolutional model (1). The model is insensitive to increasing $\mathbf{h}$ with a factor and decreasing $\mathbf{a}$ with the same factor. The ambiguity was resolved by simply normalizing $\mathbf{h}$ to unit energy after each step of wavelet estimation.

B. Computer Simulations

For the first example, we generated a single trace consisting of 1000 samples (Fig. 2a). The true reflectivity was sampled from a Bernoulli-Gaussian distribution with reflector density $\lambda = 0.05$, and convolved with a wavelet borrowed from the seismic data in the next section. Finally, white Gaussian noise was added to achieve signal-to-noise ratio 15 dB (as defined by mean power of noiseless signal divided by noise variance, SNR $= \lambda \sigma_a^2 \sum_k h^2(k)/\sigma_e^2$). The blind deconvolution algorithm was tested on this data set using two different values for the sparsity parameter; $\theta = 20\sigma_e^2$ and $\theta = 50\sigma_e^2$. In both cases all major reflectors were recovered correctly (Fig. 2c,e), though a few more small reflectors were missed for the larger θ-value. We conclude that the exact θ value is not critical for these data. The wavelet estimates (Fig. 2d,f) were highly accurate, and almost identical in the two cases. This can be compared to the much poorer estimate (Fig. 2b) obtained by a homomorphic deconvolution method, using phase unwrapping as discussed in [24]. Inserting this estimate in a Wiener filter as in [24] produced estimates which could not resolve the individual reflectors.

In the next example, the data was made more difficult by decreasing the SNR to 7 dB (Fig. 3a). The algorithm was tested on two data sets, one consisting of only one trace and one consisting of 10 traces. The different traces were independent. Thus no continuity was assumed in the reflectivity estimation step, but a common wavelet estimate was used for all traces. The sparsity parameter
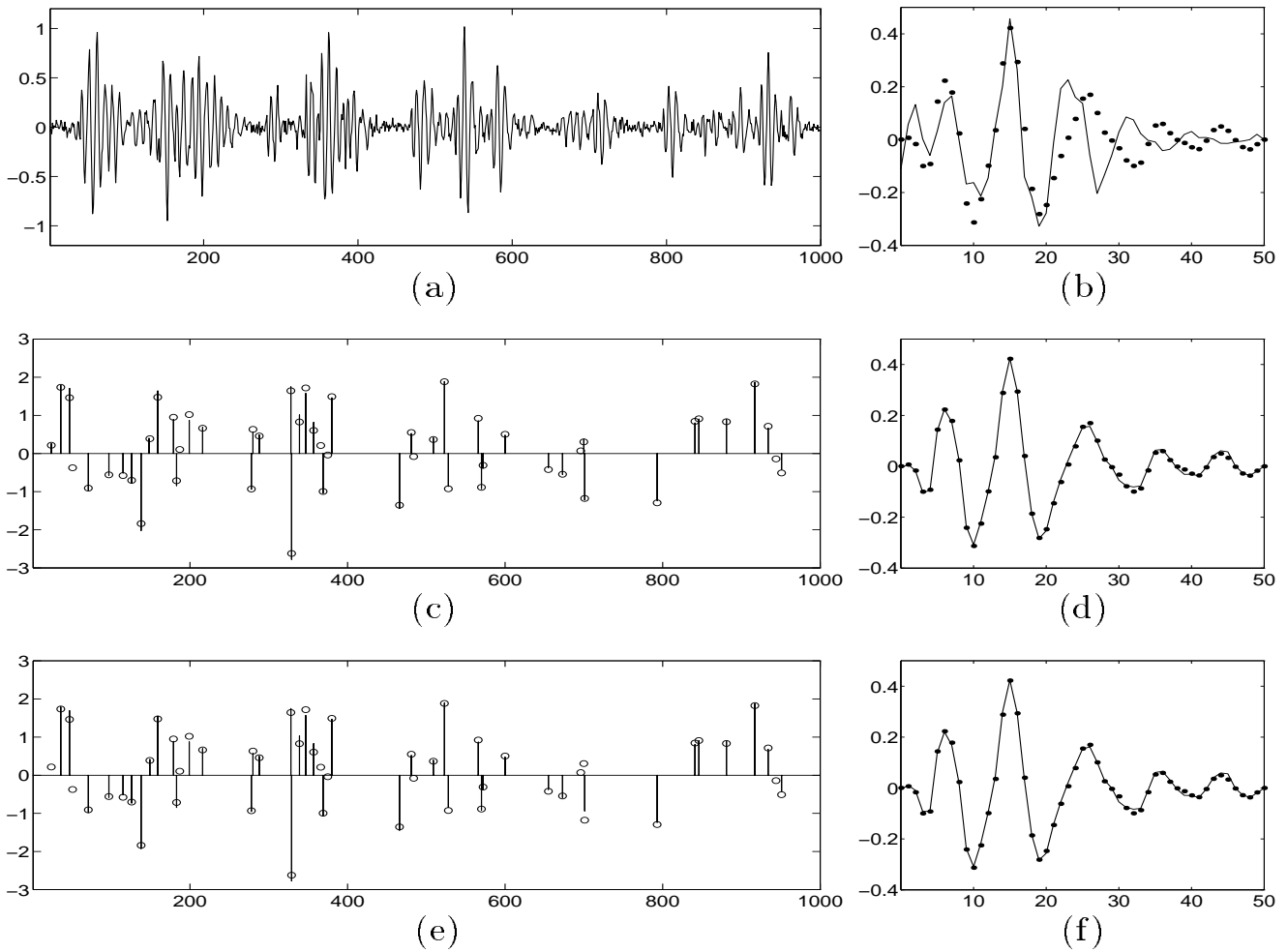
Fig. 2. Deconvolution of synthetic low noise data: (a) Data with SNR = 15 dB. (b) Wavelet estimate obtained by a homomorphic deconvolution method for comparison with the following estimates from the algorithm proposed here. (c) Reflectivity estimate obtained with sparsity parameter $\theta = 20\sigma_e^2$. (d) Wavelet estimate corresponding to (c). (e) Reflectivity estimate obtained with $\theta = 50\sigma_e^2$. (f) Wavelet estimate corresponding to (e). Estimated reflectors are shown with bars and true reflectors are shown with circles. Estimated wavelets are shown with solid lines and true wavelets are shown with dots. The horizontal unit is digitizing samples.

was fixed at $\theta = 20\sigma_e^2$. In the reflectivity estimate based on only one trace (Fig. 3c), most of the large reflectors were correctly recovered, but the estimate also contained several false detections and many small reflectors were missed. The wavelet estimate (Fig. 3d) was also considerably poorer than in the previous example. Using 10 traces, however, gave a much improved wavelet estimate (Fig. 3f), which again was almost perfect. This improved wavelet estimate also lead to a clear improvement in the corresponding reflectivity estimate (Fig. 3e). This illustrates that using more traces in the wavelet estimation step can improve both wavelet and reflectivity estimates. Furthermore, it is interesting to note that the quite moderate errors in the wavelet estimate in Fig. 3d worsen the corresponding reflectivity estimate significantly. Again,

the estimate from the homomorphic method (Fig. 3b) was much poorer, even though it was based on all 10 traces.

The final synthetic example is a data set which does not meet our model assumptions exactly. Instead of generating the data from equation (1), we include a "backscatter" term to yield the more realistic model

$$z = x * h + b * h + e. \tag{33}$$

The backscatter, b, models reflections due to small inhomogenities within the major layers. The school represented by Mendel [3] models the backscatter term explicitly, and estimates the backscatter together with the other components of their BCM's. It is argued [10], [25] that this is necessary in order to avoid too many false de-
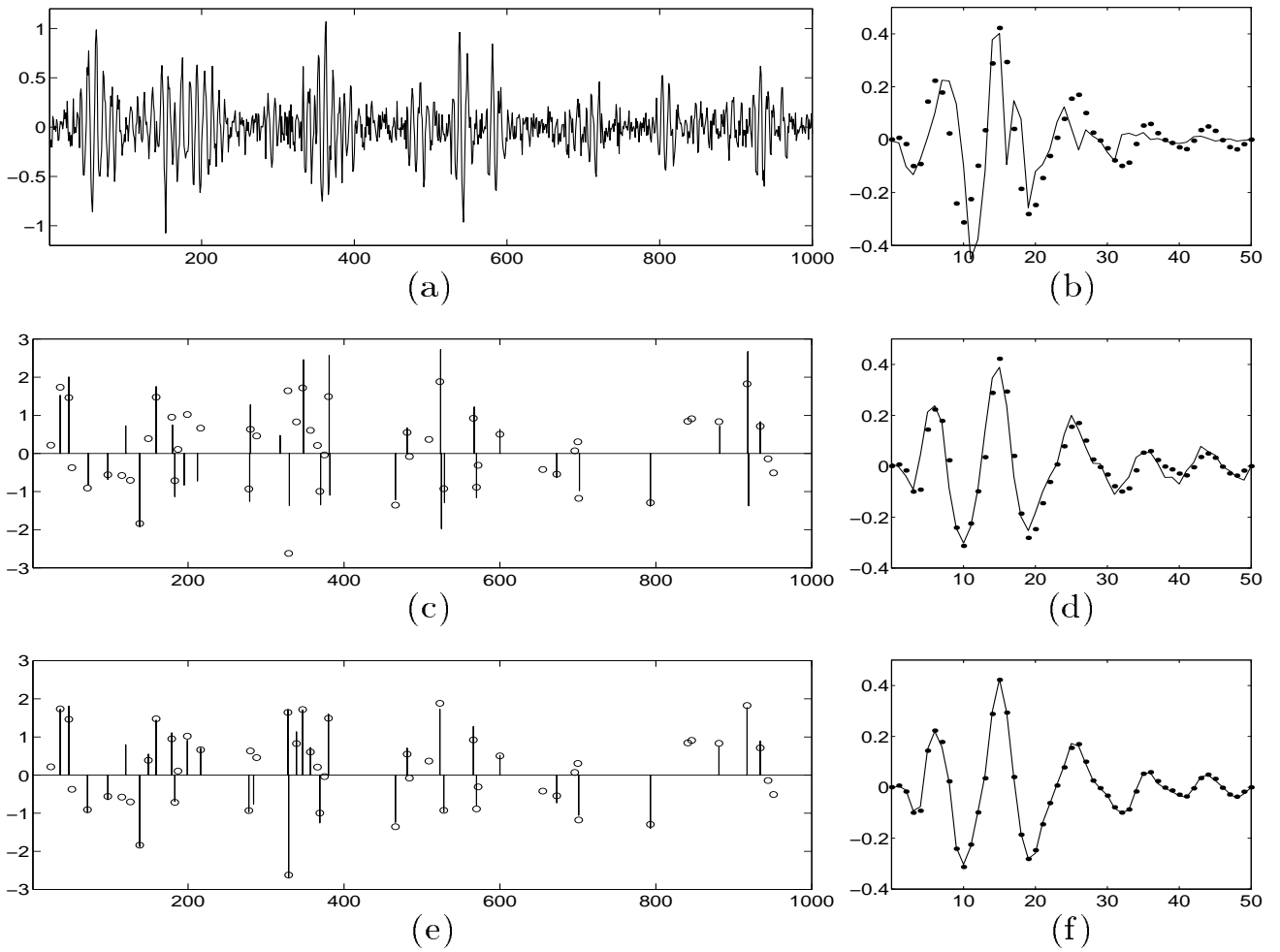
Fig. 3. Deconvolution of synthetic high noise data: (a) Data with SNR = 7 dB. (b) Wavelet estimate obtained by a homomorphic deconvolution method (based on 10 traces) for comparison with the following estimates from the algorithm proposed here. (c) Reflectivity estimate based on only one trace. (d) Wavelet estimate corresponding to (c). (e) Reflectivity estimate based on 10 traces (only the first one shown). (f) Wavelet estimate corresponding to (e).

tections. In the present setup it is clear that the problem of too many false reflections can be avoided by appropriate choice of the sparsity parameter θ. We therefore prefer to absorb the backscatter in the noise term instead. As a consequence, the noise will be colored. If the power spectrum of the noise is known (e.g. estimated from data known to contain no major reflections), white noise could be restored by a simple pre-filtering [26]. However, in the following we examine the behavior of the algorithm in colored noise conditions without attempting to "whiten" the noise spectrum.

Two data sets were generated from the backscatter model (33). For both, b was taken to be Gaussian, white, and with variance $\sigma_b^2 = \sigma_e^2$. (Since we have normalized the wavelet to unit energy, this made the power of the backscatter term and the white noise term equal.) The combined power of backscatter and white noise were then adjusted to obtain a given noise level. The first data set consisted of one trace with SNR 15 dB, while the other had 10 independent traces with SNR 7 dB. Thus, except for the fact that the noise was colored, the first data set was identical to that in Fig. 2a, while the second was identical to that in Fig. 3a. For both data sets the sparsity parameter was set to $\theta = 50(\sigma_e^2 + \sigma_b^2)$. The estimates from the low noise trace were excellent (Fig. 4a,b) and comparable to those for white noise (Fig. 2e,f). For the high noise data set only the largest reflectors were recovered, and some false detections were also present (Fig. 4c). This estimate appears inferior to the corresponding white noise result (Fig. 3e). However, part of the difference is due to the higher sparseness parameter used in the colored noise case. Reducing θ would enable smaller reflectors to be detected, but might also introduce more false detections. In spite of the reduced
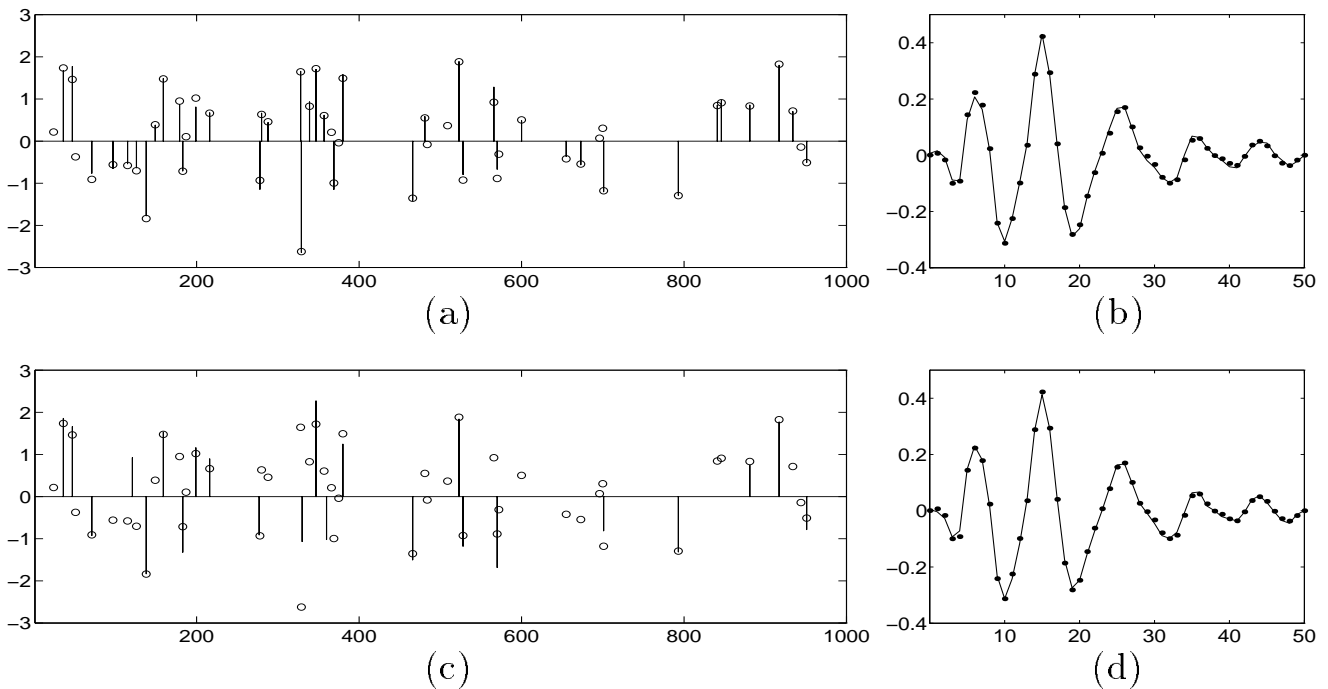
Fig. 4. Deconvolution of synthetic colored noise data: (a) Reflectivity estimate based on one trace with SNR = 15 dB. (b) Wavelet estimate corresponding to (a). (c) Reflectivity estimate based on 10 independent traces with SNR = 7 dB (only the first one shown). (d) Wavelet estimate corresponding to (c).

quality of the reflector estimate, the wavelet estimate for high colored noise (Fig. 4d) was still almost perfect. This illustrates that even when noise conditions are too adverse for good reflector estimates, averaging over several traces still enables excellent wavelet estimates. Note, however, that neighboring traces of real seismic data will often be strongly correlated. This reduces the effective amount of independent information available for wavelet estimation. It can thus be expected that more than 10 traces will be needed to achieve a similar improvement as obtained here.

Processing the above examples on a 60 MHz Pentium PC required from 0.9 to 7 seconds for one trace of 1000 samples (execution times for the multi-trace data sets were divided by the number of traces). The differences are mainly related to the density of the reflectivity estimates. Denser estimates increase the average number of reflectors in each window, and thereby slow down computation of equations (7)-(8). As a result, the high noise estimates were faster to compute since these had recovered fewer small reflectors. Also, processing 10 traces required less time per trace than processing one. This is partly because the wavelet estimation step was common to all traces, but also because more traces lead to faster convergence. The two 10 trace examples above required 5 and 7 iterations of the outer BCM, while the one trace examples ranged from 8 to 12.

C. Seismic Data Results

We now consider application to a seismic section recorded in the Fram Strait west of Svalbard. A near-stack section obtained by averaging the 10 smallest off-sets was used, and a small subsection was selected for illustrative purposes. To make the data comply better with the white noise assumption, some low-frequency noise was suppressed by a moderate high-pass filtration. The pre-processed data are shown in Fig. 5a. The strong reflection between 1500 and 1650 msec. is the seafloor reflection, while the compound reflection between 1700 and 1950 msec. is from a structure of interest. Four different reflectivity estimates were computed. The first two differ only in the value of the sparseness parameter, θ, which were selected heuristically. Fig. 5c corresponds to a relatively small value and includes many weak reflections, while Fig. 5d corresponds to a larger value and show only major reflections. In both cases the reflectivities were taken to be independent. In the next reconstruction (Fig. 5e) horizontal continuity was modeled as described in Section V-C, while θ was kept as in 5d. As expected, this estimate shows better continuity between neighboring traces than the two preceding ones. For this estimate only single-trace transitions were used. Due to the strong dependencies introduced by the chosen prior, this may not give sufficient optimality. To investigate this, the single-trace transition set was extended by
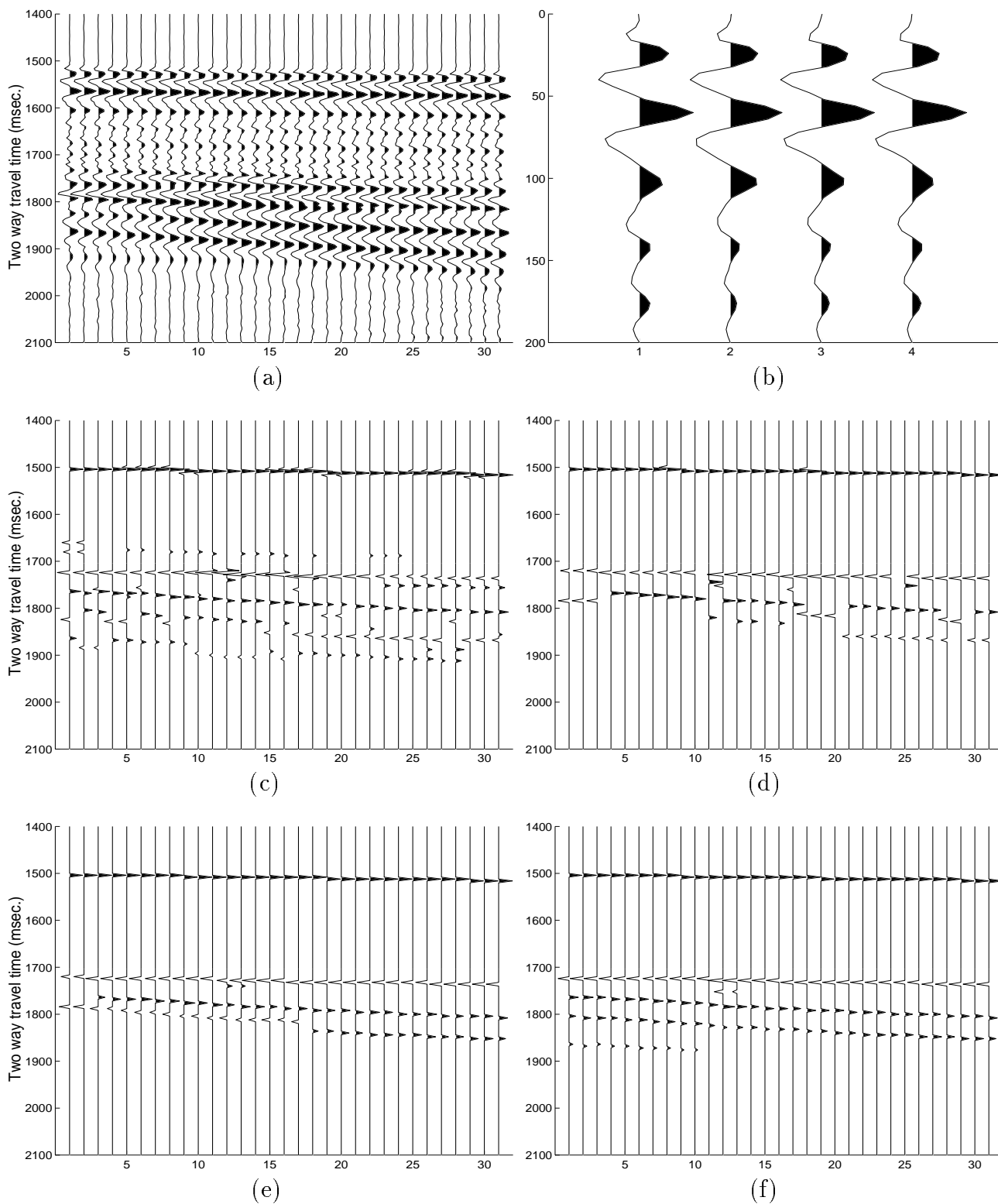
Fig. 5. Deconvolution of seismic data: (a) Seismic section. (b) Wavelet estimates corresponding to the reflectivity estimates in (c)-(f). (c) Reflectivity estimate obtained with small sparsity parameter θ. (d) Reflectivity estimate obtained with large θ. (e) As (d), but with horizontal continuity assumption. (f) As (e), but with multi-trace transitions.

some multi-trace transitions. The idea of these were to locate each connected line of reflectors and then examine various moves of the entire line. This lead to further increase of the posterior. The resulting estimate (Fig. 5f) also appears more plausible geologically.

The wavelet estimates corresponding to each of the four different reflectivity estimates are shown in Fig. 5b. The differences are minute, and they correspond well with the seafloor reflection.

All reconstructions in Fig. 5, were completed within 7 to 48 seconds on a 60 MHz Pentium PC. As with the simulated data, the sparser estimates gave the faster execution.

D. Application to Well Log Segmentation

In well logging, various variables, such as acoustic velocity and level of gamma radiation, are measured at regular intervals in an existing bore hole. Assuming a layered earth model implies that one such variable can be approximated by a step function. Since the measurements usually do not represent exact point values, but rather are weighted averages over a small area [27], they can reasonably be modeled as filtered and noise-corrupted versions of the true step function. This is the model considered here, except that the sparse reflectivities, equation (2), must be substituted by a step function of the form

$$x(n) = \sum_{i=1}^{M} a_i \left[u(n - t_{i-1}) - u(n - t_i) \right]. \qquad (34)$$

Here $u(n)$ denotes the unit step function which is 0 for negative arguments and 1 otherwise. The basic IWM algorithm can easily be adapted to deconvolution of step functions [26]. (It is only necessary to modify correlation functions (11)-(12) and equations (13)-(14) used for initialization of $\mathbf{S}$ and $\mathbf{v}$. Otherwise, the algorithm can be used as before.) Incorporating this step function modification in the present framework for blind multichannel deconvolution gives no further complications.

Since all variables (channels) relate to exactly the same geologic structure, it is reasonable to require $\mathbf{t}$ to be the same for all channels. It is thus necessary to perform updates in all channels simultaneously, as discussed in Section V-B. To determine the weighting of each log variable in equation (25) we scaled the total variance of all variables to unity and then assumed equal error variance. For the segment lengths we used a geometric prior, giving a penalty function on the simple form of equation (9) with M^w now signifying the number of segments. A selected part of a well log together with the segmented solutions and recovered filter functions is given in Fig. 6. The given segmentation was based on visual adjustment of the sparsity parameter θ and appears to have captured the major structures in the data well.
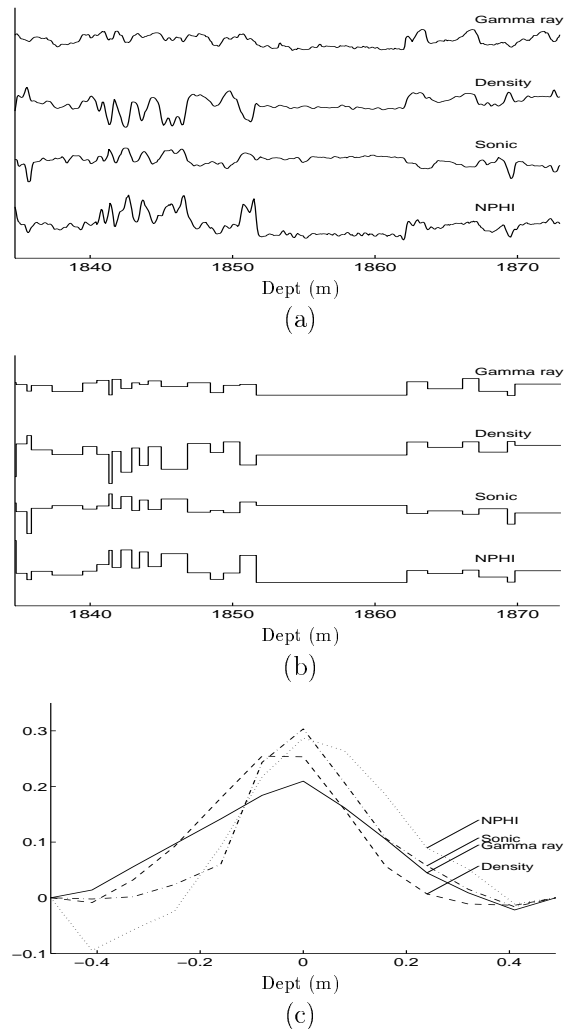

Fig. 6. Segmentation of multivariate well log data: (a) Data. (b) Segmented solutions. Note that the segment boundaries are common to all variables. (c) Estimated filter functions.

Depending on the objective a finer segmentation could be obtained with a smaller θ. Multivariate segmentation and estimation of the filter functions, based on the entire log consisting of 4 variables with 3467 samples each, required 13 sec. on a 60 MHz Pentium PC.

Other approaches to well log segmentation, c.f. e.g. [28], [29], [30], [31], have mostly been concerned with the simpler model without filtering. Failing to account for filtering will systematically bias the amplitude of the segments. In particular, short segments with large amplitudes will be estimated much too close to the neighboring segments [26]. Optimization in the more realistic model with filtering, has been considered by Goutsias and Mendel [32]. However, their solution is quite different from ours and requires that the values of x are restricted to a finite set. Furthermore, none of the above references consider multivariate segmentation.

VIII. DISCUSSION

A. Optimality

The proposed method is not guaranteed to locate the global maximum of the posterior. We have observed in some simulations that the BCM apparently has converged to a local maximum where both wavelet and reflectivity estimates are far off. Not surprisingly, the problem is related to the wavelet shape with narrow-band wavelets being the most troublesome. Perhaps more important, the problem only occurred for short record lengths. Increasing the record length and, in particular, using several traces, always lead to reliable wavelet estimates. This suggest that the problem is not serious in practice since typical seismic data sets consist of a large number of traces.

Even though the true wavelet is basically recovered, the estimated reflector positions may not correspond to the their global optimum due to the sub-optimality of the reflectivity estimation step. The approach given here is clearly more optimal than established alternatives based on the SMLR algorithm [3], [10], [11], since this algorithm updates only one sample at each step. Specifically, our optimization scheme gives higher optimality than the multichannel SMLR solution of Idier and Goussard [13]. Improved optimality results from updating several reflectors in one trace simultaneously, using iterative instead of recursive scanning of the traces, and allowing multitrace updates.

A further increase in optimality is possible by using more advanced transitions than those employed here (particularly multi-trace transitions). Alternatively, optimality may be increased by stochastic relaxation and simulated annealing [12], [23], [33]. Although simulated annealing is sometimes misleadingly referred to as a globally optimal method, this is only for an unrealizable amount of computation [23]. In practice, the degree of optimality is governed by the computing time allowed. Since any level of optimality can also be attained by extending the transition set of IWM, the relevant question is which method gives the highest optimality for a given amount of computation. For a specific single-channel problem we have found that the deterministic IWM algorithm by far surpasses the stochastic approach [19], but more research is needed to compare the two approaches in the multichannel situation.

An intermediate solution is to apply stochastic relaxation to the IWM framework. Common stochastic solutions only update one sample of the reflectivity at each step. This is a likely reason for the slow convergence experienced in [19]. Relaxing the IWM formulas would allow stochastic updating of more than one sample at each step, and thereby speed up convergence. To reduce the sub-optimality related to the outer BCM, one might also consider stochastic relaxation of the wavelet estima-

tion step as in [33].

Stochastic relaxation in the IWM framework or more complicated multi-trace transitions are topics for future research. We do, however, think that the procedure described here is sufficiently optimal for most practical purposes. An exception is when particularly strong crosstrace continuity is assumed. On the other hand, one should be careful about enforcing continuity too strongly. Although this will generate aesthetically pleasing estimates, the uncertainty may be concealed.

B. Significance for Seismic Deconvolution

To our knowledge, the true reflectivity of the seismic data presented here is not known. It is not clear whether the increased continuity in Fig. 5e,f (with horizontal continuity modeling) is due to true reflectors or artifacts. It is, for example, possible that the observed reflections do not originate from distinct boundaries, but result from a gradual increase of sound velocity instead. In this case the layered earth model is not appropriate, and a noncontinuous chaotic estimate simply expresses the fact that the model does not fit these data. Note, however, the two reflecting surfaces at approximately 1720 and 1760 msec. in Fig. 5f. These show up consistently in all four estimates. A possible explanation is that these correspond to the upper and lower boundary of sediments containing free gas which are trapped below sediments containing gas hydrate. This interpretation is consistent with the conclusions reached in [34] for these data (BGR 31-74, CMP 697-727). With careful interpretation, it appears that the method arrived at here would be a useful tool for analysis of these data. The clear difference between Fig 5e and 5f indicates that the increased optimality of the given method, as compared to SMLR based alternatives [13], is significant for multichannel seismic deconvolution.

The given method can also be used for wavelet estimation only. A conventional (non-sparse) mixed-phase deconvolution can be obtained by using the recovered wavelets as input to, e.g., a Wiener filter. The simulation results demonstrate that much better wavelet estimates are produced by the given MAP method than by a homomorphic method. (The latter represents a class of mixed-phase wavelet estimation methods that have received much attention.) Since the degree of wavelet compression that can be obtained with Wiener filtering is quite sensitive to the correctness of the input wavelet [35], the experienced improvement is clearly significant. Note, however, that the synthetic test data fulfilled the distributional assumptions of the MAP method exactly. If these assumptions are seriously violated, the homomorphic method (and other more general methods) can be expected to compete more favorably.

C. More advanced priors

An advantage of the presented optimization method, is the large flexibility in choice of prior model for **t**. Here, the prior was used to model sparsity and horizontal continuity. More advanced priors are topics for future research. For example, a model for multiple reflections could be built into the prior. One might also consider working with pre-stack data. In this case, the prior should express the fact that traces with a common midpoint should have (almost) the same reflector positions. This aspect would resemble the prior used for well log segmentation here.

IX. Conclusions

We have presented a practical and complete solution for multichannel blind deconvolution based on a model for sparse reflectivities. Unlike Gaussian based models this allows the phase of the wavelet to be recovered. Application to two-dimensional data was considered here, and extension to three dimensions is straightforward. The estimation problem was formulated in a maximum a posteriori framework. This lead to a difficult optimization problem which we solved by an iterative and suboptimal block component method. The computational load depends on the degree of optimality required. For practical purposes, sufficient optimality can normally be achieved while still retaining fast execution.

The presented computer simulations show that excellent estimates are obtained when the underlying model is appropriate. Slight deviations such as a moderate amount of colored noise are not detrimental. Violation of wavelet invariance is likely to be more serious. Modifications for other model assumptions, such as a slightly time variant wavelet or non-stationary noise, can be considered along the lines of [26], but sparseness of the reflectivity is essential. For estimation of non-sparse reflectivities other methods should be used.

References

[1] E. A. Robinson and S. Treitel, *Geophysical Signal Analysis*, Englewood Cliffs, NJ: Prentice-Hall, 1980.

[2] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, New York: Wiley, 1949.

[3] J. M. Mendel, *Optimal Seismic Deconvolution: An Estimation-Based Approach*, New York: Academic Press, 1983.

[4] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1989.

[5] J. M. Tribolet, *Seismic Applications of Homomorphic Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1979.

[6] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, 1991.

[7] G. D. Lazear, "Mixed-phase wavelet estimation using fourth-order cumulants," *Geophysics*, vol. 58, no. 7, pp. 1042–1051, 1993.

[8] J. J. Kormylo, *Maximum-likelihood seismic deconvolution*, Ph.D. thesis, Univ. of Southern California, 1979.

[9] J. J. Kormylo and J. M. Mendel, "Maximum likelihood detection and estimation of Bernoulli-Gaussian processes," *IEEE Trans. Inform. Theory*, vol. 28, no. 3, pp. 482–488, 1982.

[10] C.-Y. Chi, J. M. Mendel, and D. Hampson, "A computationally fast approach to maximum-likelihood deconvolution," *Geophysics*, vol. 49, no. 5, pp. 550–565, 1984.

[11] J. Goutsias and J. M. Mendel, "Maximum likelihood deconvolution: An optimization theory perspective," *Geophysics*, vol. 51, no. 6, pp. 1206–1220, 1986.

[12] M. Lavielle, "2-D Bayesian deconvolution," *Geophysics*, vol. 56, no. 12, pp. 2008–2018, 1991.

[13] J. Idier and Y. Goussard, "Multichannel seismic deconvolution," *IEEE Trans. Geosci. Remote Sensing*, vol. 31, no. 5, pp. 961–979, 1993.

[14] J. Besag, "On the statistical analysis of dirty pictures," *Journ. Royal Statist. Soc. Ser. B*, vol. 48, no. 3, pp. 259–302, 1986.

[15] K. F. Kaaresen, "Deconvolution of sparse spike trains by iterated window maximization," *IEEE Trans. Signal Processing*, vol. 45, no. 5, pp. 1173–1183, 1997.

[16] C. W. Groetsch, *Inverse Problems in the Mathematical Sciences*, Wiesbaden: Vieweg, 1993.

[17] E. Gassiat, F. Monfront, and Y. Goussard, "On simultaneous signal estimation and parameter identification using a generalized likelihood approach," *IEEE Trans. Inform. Theory*, vol. 38, no. 1, pp. 157–162, 1992.

[18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, New York: Cambridge University Press, 2 edition, 1992.

[19] K. F. Kaaresen, "Maximum a posteriori deconvolution of sparse spike trains," Tech. Rep. 4, Dept. of Mathematics, University of Oslo, Norway, 1996.

[20] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Stat. Comput.*, vol. 13, pp. 1142–1150, 1992.

[21] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*, New York: Springer-Verlag, 2 edition, 1988.

[22] J. A. Hartigan, *Bayes Theory*, New York: Springer-Verlag, 1983.

[23] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, pp. 721–741, 1984.

[24] T. Taxt, "Comparison of cepstrum-based methods for radial blind deconvolution of ultrasound images," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr., to appear*, vol. 44, no. 3, 1997.

[25] J. M. Mendel, *Maximum-Likelihood Deconvolution: A Journey into Model-Based Signal Processing*, New York: Springer-Verlag, 1990.

[26] K. F. Kaaresen, "Efficient maximum a posteriori deconvolution of sparse structures," Tech. Rep. 13, Dept. of Mathematics, University of Oslo, Norway, 1996.

[27] E. S. Robinson and C. Coruh, *Basic Exploration Geophysics*, New York: Wiley, 1988.

[28] R. Godfrey, F. Muir, and F. Rocca, "Modeling seismic impedance with Markov chains," *Geophysics*, vol. 45, no. 9, pp. 1351–1372, 1980.

[29] P. N. Chouinard and K. V. Paulson, "A Markov-Gauss algorithm for blocking of well logs," *Geophysics*, vol. 53, no. 8, pp. 1118–1121, 1988.

[30] A. Moghaddamjoo, "Constraint optimum well-log signal seg-
     mentation," *IEEE Trans. Geosci. Remote Sensing*, vol. 27,
     no. 5, pp. 633–641, 1989.

[31] S. Radhakrishnan, G. Srikanth, and C. H. Mehta, "Segmen-
     tation of well logs by maximum likelihood estimation: The al-
     gorithm and fortran-77 implementation," *Computers & Geo-
     sciences*, vol. 17, no. 9, pp. 1173–1196, 1991.

[32] J. Goutsias and J. M. Mendel, "Optimal simultaneous detec-
     tion and estimation of filtered discrete semi-Markov chains,"
     IEEE Trans. Inform. Theory, vol. 34, no. 3, pp. 551–568,
     1988.

[33] Q. Cheng, R. Chen, and T.-H. Li, "Simultaneous wavelet
     estimation and deconvolution of reflection seismic signals,"
     IEEE Trans. Geosci. Remote Sensing, vol. 34, no. 2, pp. 377–
     384, 1996.

[34] K. Andreassen, *Seismic Reflections Associated with Subma-
     rine Gas Hydrates*, Ph.D. thesis, University of Tromsø, Nor-
     way, 1995.

[35] G. Hayward and J. E. Lewis, "Comparison of some non-
     adaptive deconvolution techniques for resolution enhancement
     of ultrasonic data," *Ultrasonics*, vol. 27, pp. 155–164, 1989.