Multichannel Seismic Deconvolution

Jérôme Idier and Yves Goussard, *Member, IEEE*

Abstract— **This paper deals with Bayesian estimation of two-dimensional (2-D) stratified structures from echosounding signals. This problem is of interest in seismic exploration, but also for non-destructive testing or medical imaging. The proposed approach consists of a multichannel Bayesian deconvolution method of the 2-D reflectivity based upon a theoretically sound prior stochastic model. The Markov–Bernoulli random field representation introduced in [1] is used to model the geometric properties of the reflectivity, and the emphasis is placed on representation of the amplitudes and on deconvolution algorithms. It is shown that the algorithmic structure and computational complexity of the proposed multichannel methods are similar to those of single-channel B-G deconvolution procedures, but that explicit modeling of the stratified structure results in significantly better performances. Simulation results and examples of real-data processing illustrate the performances and the practicality of the multichannel approach.**

I. Introduction

This paper deals with deconvolution of two-dimensional (2-D) signals which represent the inner structure of layered media. This problem arises when the structure of an unknown propagation medium is estimated from acoustic measurements performed at its boundaries, and is a basic issue in many areas of applied physics such as geophysics, nondestructive testing, and ultrasonic medical imaging. Assume that the general orientation of the layers is horizontal, i.e., that most of them are only slightly bent from the horizontal direction. We also assume that incident acoustic waves are emitted from the upper boundary, and that the reflected waves are observed on the same boundary (see Fig. 1). Under reasonable assumptions, and after adequate preprocessing of the data, the observations constructed from the reflected waves can be considered as the noise-corrupted convolution product of two signals: the wavelet, and the reflectivity. The wavelet is a one-dimensional (1-D) vertical signal which represents the waveshape that scatters through the medium, and the reflectivity, defined as the vertical logarithmic derivative of the acoustic impedance, is a 2-D signal which characterizes the medium.

The object of multichannel deconvolution is to estimate the 2-D reflectivity from the observations and from the knowledge available on the wavelet and on the noise. Here, it is assumed that the wavelet and the probability distribution of the noise are perfectly known. Even under these simplifying assumptions,

Manuscript received July 22, 1992; revised April 1, 1993.

J. Idier is with the Laboratoire des Signaux et Systèmes (CNRS/ESE/UPS), École Supérieure d'Électricité, Plateau de Moulon, 91192 Gif-sur-Yvette Cédex, France.

Y. Goussard was with the Laboratoire des Signaux et Systèmes (CNRS/ESE/UPS), École Supérieure d'Électricité, Plateau de Moulon, 91192 Gif-sur-Yvette Cédex, France. He is now with the École Polytechnique, Biomedical Engineering Institute, Montreal, Québec, Canada, H3C 3A7.

IEEE Log Number 9210722.

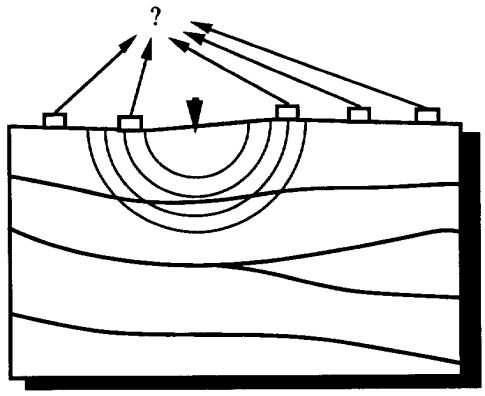


Fig. 1. Seismic exploration: The problem is to characterize a layered structure from acoustic measurements performed at its surface.

multichannel deconvolution presents serious difficulties for two major reasons:

1) Estimation of the reflectivity is an *ill-posed inverse* problem; this means that even an accurate input–output model of the physical process is not sufficient to provide a reliable estimate of the reflectivity. In order to obtain an acceptable behavior of the estimates, introduction of *a priori* information on the solution is necessary.
2) In most application domains mentioned earlier, and particularly for geophysical applications, very large amounts of data must be processed, which imposes that the estimation algorithms present a low numerical complexity. Clearly, proper choice of prior information is critical, as it should both capture the essential characteristics of the solution and yield numerically tractable estimation methods. Here, the major characteristic of the unknown medium is its stratified structure, to which we add the simplifying assumption that the layers are homogeneous in the vertical direction.

A common approach to the multichannel deconvolution problem consists of breaking it into independent vertical 1-D deconvolution problems. This is possible because the wavelet is a 1-D vertical signal. In the vertical direction, a 1-D reflectivity signal appears as a sparse spike train where each spike (i.e., reflector) corresponds to a boundary between two adjacent homogeneous layers. Several types of methods have been developed for deconvolution of such signals. In the early 1980's, Mendel and his students introduced a Bayesian approach which proved very fruitful [2], [3]. It consists of performing the maximum *a posteriori* (MAP) estimation of the 1-D reflectivity which is modeled as a Bernoulli–Gaussian (B-G) random process [4], [5]. B-G processes can adequately represent sparse spike trains, and yield tractable estimation

0196–2892/93$03.00 © 1993 IEEE

procedures, thanks to their Gaussian component. In addition, B-G representations present a hierarchical structure: the geometry of the reflectivity, i.e., the location of the reflectors, is controlled by the Bernoulli part of the model, while amplitude of the reflectors is defined conditionally to their location.

In two dimensions, the reflectors, which are located along boundaries between layers, generally form smooth lines with a preferential orientation in the horizontal direction. Application of 1-D restoration methods to such 2-D signals is clearly suboptimal, as the essential information of continuity of the layers is not accounted for. Capturing such properties with a mathematical model presents important difficulties and to our knowledge, existing multichannel methods such as [6] or [7] strongly rely on empirical treatments. The goal of this paper is to propose a multichannel restoration method based upon a theoretically sound prior model of the 2-D reflectivity. This method retains the essential characteristics of single-channel B-G deconvolution. It is a Bayesian technique, and the 2-D statistical prior model comprises two parts organized in a hierarchical manner: the higher level of the model controls the geometry of the reflectivity and accounts for stratification, while the lower level represents the amplitude of the reflectors and is defined conditionally to the geometric component of the model. Here, the Markov–Bernoulli random field (MBRF) representation introduced in [1] is used to model the geometric properties of the reflectivity, and the emphasis is placed on representation of the amplitudes and on the corresponding restoration algorithms. Two amplitude models are proposed. The first one is an exact 2-D extension of the 1-D B-G representation, while the second one is more consistent with the physical properties of the underlying phenomena. For each prior model, a restoration algorithm which implements a suboptimal MAP estimator of the reflectivity is derived, and its general characteristics and performances are investigated. It is shown that the algorithmic structure and computational complexity of the multichannel methods are similar to those of 1-D B-G deconvolution procedures, but that explicit modeling of the stratified structure results in significantly better performances.

The paper is organized as follows: the modeling assumptions and problem statement are given in Section II. In Section III, background material on Markov random fields and MBRF's is presented, and two forms of complete prior reflectivity models are proposed. The corresponding MAP restoration algorithms are presented in Section IV. In order to highlight the similarity between these algorithms and single-channel B-G deconvolution techniques, all restoration procedures are jointly derived from a *unique* algorithmic structure. In Section V, simulation results and examples of real-data processing illustrate the performances and the practicality of the multichannel approach. Concluding comments are given in Section VI.

II. Problem Formulation

As stated in the introduction, it is assumed that the raw data have been adequately preprocessed so that the observations can be considered as the noise-corrupted convolution product of the wavelet and of the reflectivity. Therefore, one has

$$\boldsymbol{z} = h * r + n \tag{1}$$

where $\boldsymbol{z}, h$ and $\boldsymbol{r}$, respectively, denote the observed dataset, the wavelet, and the reflectivity. All data are assumed to be discrete; $\boldsymbol{z}$ and $\boldsymbol{r}$ are 2-D arrays (z_{ij}) and (r_{ij}), $1 \leq i \leq I, 1 \leq j \leq J$, and wavelet $\boldsymbol{h}$ is 1-D vector (h_i), $0 \leq i \leq p$, which is assumed to be invariant in both horizontal and vertical directions for the sake of simplicity. Additional noise $\boldsymbol{n} = (n_{ij}), 1 \leq i \leq I, 1 \leq j \leq J$ represents unmodeled physical phenomena (structural noise), imperfections of the measurement system (observation noise), and modeling errors. It is assumed that $\boldsymbol{n}$ is independent from $\boldsymbol{r}$, white, Gaussian, zero mean with variance equal to σ_n^2. Here, only *simple* deconvolution is considered and consequently, $\boldsymbol{h}$ and σ_n^2 are assumed to be known.

Since $\boldsymbol{h}$ is a 1-D vertical signal, 2-D convolution product of (1) can be separated into independent 1-D convolution products, and the input–output equation can be rewritten as

$$\forall\, j \in [1, J], \quad \boldsymbol{z}_j = h * \boldsymbol{r}_j + \boldsymbol{n}_j \tag{2}$$

where $\boldsymbol{z}_j, \boldsymbol{r}_j$ and $\boldsymbol{n}_j$ denote the jth column of $\boldsymbol{z}, r$ and $\boldsymbol{n}$, respectively. Equation (2) can be put in the following matrix form:

$$\forall\, j \in [1, J], \quad \boldsymbol{z}_j = \boldsymbol{H}\boldsymbol{r}_j + \boldsymbol{n}_j \tag{3}$$

in which matrix $\boldsymbol{H}$ contains shifted samples of the wavelet.

The object of the deconvolution procedure is to estimate $\boldsymbol{r}$. Here, only MAP-type estimators are considered, and the solution $\hat{\boldsymbol{r}}$ is defined by

$$\hat{\boldsymbol{r}} \triangleq \underset{\boldsymbol{r}}{\operatorname{argmax}} P(\boldsymbol{R} = r | Z = z). \tag{4}$$

Throughout the rest of the paper, random variables and their realizations will be, respectively, denoted by corresponding upper and lowercase letters; in addition, the shorthand notation $P(x|y)$ will be used for conditional probability $P(X = x|Y = y)$ whenever unambiguous. Applying Bayes' rule to (4) immediately yields

$$\hat{\boldsymbol{r}} = \underset{\boldsymbol{r}}{\operatorname{argmax}} P(\boldsymbol{z}|r)P(r). \tag{5}$$

A straightforward consequence of the statistics of the noise is that $P(\boldsymbol{z}|r)$ is white Gaussian, with mean value $\boldsymbol{H}r$ and variance σ_n^2. Using (3), $\boldsymbol{P}(z|r)$ can be expressed as

$$P(z|r) = \frac{1}{\sqrt{2\Pi\sigma_n^{IJ}}} \times \exp\left\{-\frac{1}{2\sigma_n^2}\sum_{j=1}^{J}(\boldsymbol{z}_j - \boldsymbol{H}r_j)'(\boldsymbol{z}_j - \boldsymbol{H}r_j)\right\} \tag{6}$$

where the prime sign denotes the transpose operator. Equation (6) can be further factored as

$$P(\boldsymbol{z}|r) = \prod_{j=1}^{J} \frac{1}{\sqrt{2\Pi\sigma_n^I}} \times \exp\left\{-\frac{1}{2\sigma_n^2}(\boldsymbol{z}_j - \boldsymbol{H}r_j)'(\boldsymbol{z}_j - \boldsymbol{H}r_j)\right\} \tag{7a}$$

$$= \prod_{j=1}^{J} \boldsymbol{P}(\boldsymbol{z}_j|\boldsymbol{r}_j). \tag{7b}$$

If the prior density of the 2-D reflectivity can also be factored as

$$P(r) = \prod_{j=1}^{J} P(\boldsymbol{r}_j) \tag{8}$$

then it is clear that the optimization problem defined by (5) can be separated into J independent 1-D maximization problems which each correspond to MAP restoration of a vertical reflectivity sequence. The assumption that $P(\boldsymbol{r})$ can be factored according to (8) is implicit when 1-D Bayesian methods, such as B-G deconvolution, are used for restoration of 2-D reflectivity sections. However, (8) indicates that vertical reflectivity signals $\boldsymbol{r}_j$ are statistically independent from each other, and that no account is taken of the strong lateral organization of the stratified propagation medium. In order to correct for this deficiency, we now propose prior probability distributions of the reflectivity which explicitly model its layered structure.

III. Prior Models of the Reflectivity

In recent years, Markov random fields (MRF's) have emerged as a powerful tool for accounting for local interactions and inhomogeneities with a statistical model. Since the notions of continuity of the layers and of lateral correlation between reflectivity columns which characterize stratified media are essentially *local*, MRF's appear as a particularly suitable class of models for specifying our priors. Some background material on MRF's is presented hereafter.

A. Markov Random Fields

Consider a stochastic process $\boldsymbol{X} = \{X_n, 1 \leq n \leq N\}$ defined on a lattice $\Lambda = \{s_n, 1 \leq n \leq N\}$. Each element s_n of the lattice, or *site*, is in one-to-one correspondence with elementary variable X_n, and the ordering of the sites is arbitrary. The Markov property expresses the fact that the conditional distribution of any elementary variable X_n only depends upon a—preferably small—set of neighbors of s_n. In order to formalize this property, define a system $\mathcal{N} \triangleq \{\boldsymbol{N}_n, 1 \leq n \leq N\}$, where $\boldsymbol{N}_n$ is any subset of Λ. $\mathcal{N}$ is a neighborhood system if the following two conditions hold:

- $\forall n, s_n \notin \boldsymbol{N}_n$;
- $\forall m, \forall n, s_n \in \boldsymbol{N}_m \Leftrightarrow s_m \in \boldsymbol{N}_n$.

$\boldsymbol{X}$ is a MRF with respect to the neighborhood system $\mathcal{N}$ if and only if, for any site s_n and for any joint realization $\boldsymbol{x} = (x_1, \cdots, x_N)$,

$$P(x_n|x_1, \cdots, x_{n-1}, x_{n+1}, \cdots, x_N) = P(x_n|x_m, s_m \in N_n). \tag{9}$$

Any set of sites which either consists of a single site or in which every pair of distinct sites are neighbors is called a *clique*.

MRF's as defined above represent the most general class of Markov processes on finite lattices. Several subclasses with specific properties have been introduced in the past two decades. For an exhaustive overview, see [8].

Property (9) provides an easy way of checking whether a random process is a MRF and, if it is, to determine its neighborhood system. On the other hand, specification of a neighborhood system and of local conditional probability measures $P(x_n|x_m, s_m \in N_n)$ generally does not yield a consistent joint probability measure $P(\boldsymbol{x})$. In order to ensure the validity of the statistical model, two approaches may be employed. The first one is based upon the equivalence between MRF's and Gibbs random fields. Since Gibbs distributions can be explicitly expressed using local functionals (namely, Gibbs potentials), the Markov–Gibbs equivalence provides a general way of specifying local conditional probability measures which are consistent with a valid MRF structure. This approach has been widely employed, particularly for image processing and computer vision applications. However, it presents two limitations: 1) the Markov–Gibbs equivalence holds only if $P(\boldsymbol{x})$ never vanishes, or, in other words, if MRF $\boldsymbol{X}$ does not present any forbidden configuration; 2) the corresponding estimation methods are generally computationally intensive (e.g., simulated annealing [9] or GNC [10]), which is a severe disadvantage in the application domains which are in the scope of this study.

The second approach for specifying a valid statistical model consists of repeatedly applying Bayes' rule so as to factor the joint probability measure of $\boldsymbol{X}$ as

$$P(\boldsymbol{x}) = P(x_1) \prod_{n=2}^{N} P(x_n|x_1, \cdots, x_{n-1}) \tag{10}$$

and of assuming that each conditional probability $P(x_n|x_1, \cdots, x_{n-1})$ only depends upon a small subset $\boldsymbol{P}_n$ of *predecessors* of current site s_n. Then it is straightforward to show that $\boldsymbol{X}$ is a valid MRF whose set of *maximal* cliques is $\{\boldsymbol{P}_n \cup \{s_n\}, 1 \leq n \leq N\}$ (*maximal* cliques are those which are not subsets of any larger ones). Such MRF's are referred to as unilateral MRF's (UMRF's). One of the drawbacks of UMRF's is that the ordering of the sites required by the *unilateral* characterization (10) is arbitrary and may introduce privileged directions which are not related to any physical property of the system under study. On the other hand, by construction, probability distributions of UMRFs can always be expressed as

$$P(\boldsymbol{x}) = P(x_1) \prod_{n=2}^{N} P(x_n|x_m, s_m \in \boldsymbol{P}_n) \tag{11}$$

and this factored form is particularly suitable for recursive data processing. UMRF's present several other interesting features. For more details see, e.g., [1], [11], and [12].

As mentioned earlier, it is always desirable that a prior model be as simple as possible while adequately capturing the essential characteristics of the system under study. For MRF's, this simplicity requirement applies to the structure of the neighborhood system and to the parameterization of the joint probability distribution. For layered media characterization, an important step toward simplicity can be made by structuring the model in a hierarchical manner. This can be achieved by explicitly introducing an auxiliary random field $\boldsymbol{G}$ which represents the geometric characteristics of the medium (i.e., the

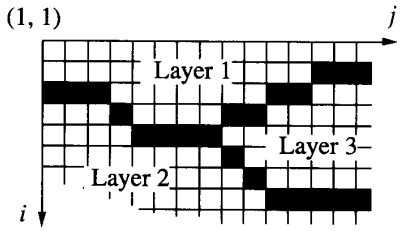


Fig. 2. Discrete binary representation of layer boundaries: Black squares denote the presence of a layer boundary on the corresponding site.

location of the layers or of their boundaries), and by defining the probability distribution of reflectivity $\boldsymbol{R}$ conditionally to $\boldsymbol{G}$. Specifying the priors is now equivalent to defining $P(\boldsymbol{G} = g, R = r)$ which is factored as

$$P(\boldsymbol{g}, r) = P(\boldsymbol{g})P(\boldsymbol{r}|\boldsymbol{g}). \tag{12}$$

As noted in the introduction, such a hierarchical structure is present in the Bernoulli–Gaussian model used for 1-D deconvolution. In addition, $\boldsymbol{G}$ plays a role similar to that of *hidden variables* in speech recognition [13] and of *line processes* in Markov modeling of images [9]. Such stochastic models have proved very effective in accounting for discontinuities (such as, here, the layer boundaries) and in reducing the size of the neighborhood sets. Here, a Markov–Bernoulli random field (MBRF) representation is used to describe the statistical properties of the geometric properties of the complete prior model. MBRF's belong to the class of UMRF's and were specifically designed to model the geometry of stratified structures. Unlike other UMRF's, the enumeration of the sites of MBRF's is chosen in accordance with the general orientation of the layers, and does not introduce any *artificial* privileged direction. Before proceeding with the derivation of the complete prior model, we give a more precise definition of MBRF's and highlight their main features.

B. Markov–Bernoulli Random Fields

In this subsection, most results are stated without proof. For the complete derivation of MBRF's and investigation of their properties, the reader is referred to [1]. MBRF models comprise two kinds of binary variables: *location* variables and *transition* variables. The role of location variables $\boldsymbol{Q} = \{Q_{ij}\}$, which are sampled on a rectangular lattice Λ°, is to indicate the position of the layer boundaries. Location variables are set to one if the corresponding site belongs to a boundary, and to zero otherwise, as shown in Fig. 2. The role of transition variables $\boldsymbol{T}$ is to indicate explicitly whether adjacent location variables belong to the same layer boundary or not. They are set to one in the first case and to zero otherwise. Since the general orientation of the layers is horizontal, transition sites are defined only between every pair of diagonally and horizontally adjacent location sites, according to the scheme represented in Fig. 3.

Let $\Lambda^/, \Lambda^-$ and $\Lambda^\backslash$ denote the sets of diagonally ascending, horizontal and diagonally descending transition sites, and let $\boldsymbol{T}^/ = \{T^/_{ij}\}, \boldsymbol{T}^- = \{T^-_{ij}\}$ and $\boldsymbol{T}^\backslash = \{T^\backslash_{ij}\}$ denote the corresponding transition variables. The probability distribution

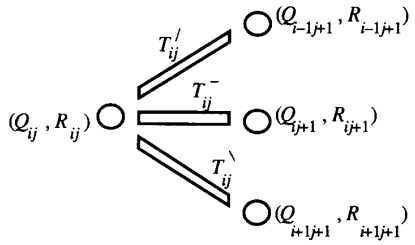


Fig. 3. Introducing transition variables T_{ij}: Transition variables are placed between pairs of reflectors which are either diagonally or horizontally adjacent. They are meant to indicate whether two adjacent layer boundary sites belong to the same layer boundary.

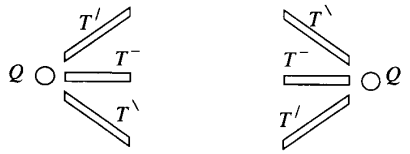


Fig. 4. Generic cells on lattice Λ: The unilateral joint characterization of $\boldsymbol{T}, \boldsymbol{Q}$ is based on the decomposition of $P(\boldsymbol{t}, \boldsymbol{q})$ as a product of local probabilities on these cells.

of the complete geometric model $\boldsymbol{G} \triangleq \{\boldsymbol{T}, Q\}$ defined on nonrectangular lattice $\Lambda \triangleq \Lambda^\circ \cup \Lambda^/ \cup \Lambda^- \cup \Lambda^\backslash$ is given by

$$P(\boldsymbol{t}, q) \triangleq \prod_{j=1}^{J} \prod_{i=1}^{I} \tau(t^/_{ij}, t^-_{ij}, t^\backslash_{ij} | q_{ij}) \times \tau(q_{ij} | t^/_{i+1j-1}, t^-_{ij-1}, t^\backslash_{i-1j-1}) \tag{13}$$

where τ is an invariant probability measure defined on the basic cells depicted in Fig. 4. Note that in the above equation, lattice boundary effects have been neglected. For the sake of clarity, the same approximation is made throughout the rest of the paper, but rigorous expressions corresponding to a *free boundary* assumption can be found in [1].

From (13), it can be shown that $\{\boldsymbol{T}, Q\}$ is a valid UMRF with a simple neighborhood structure, and that its maximal cliques correspond to the basic cells shown in Fig. 4. $\{\boldsymbol{T}, Q\}$ can also be considered as a first-order vector Markov chain, since (13) can be rewritten as

$$P(\boldsymbol{t}, q) = P(\boldsymbol{q}_1) \prod_{j=1}^{J-1} P(\boldsymbol{t}_j | \boldsymbol{q}_j) P(\boldsymbol{q}_{j+1} | \boldsymbol{t}_j) \tag{14}$$

where each probability in the RHS of (14) is defined as

$$P(\boldsymbol{q}_1) = \prod_{i=1}^{I} \tau(q_{i1}) \tag{15a}$$

$$P(\boldsymbol{t}_j | \boldsymbol{q}_j) = \prod_{i=1}^{I} \tau(t^/_{ij}, t^-_{ij}, t^\backslash_{ij} | q_{ij}) \tag{15b}$$

$$P(\boldsymbol{q}_{j+1} | \boldsymbol{t}_j) = \prod_{i=1}^{I} \tau(\boldsymbol{q}_{ij+1} | t^/_{i+1j}, t^-_{ij}, t^\backslash_{i-1j}). \tag{15c}$$

However, in order to obtain an adequate geometric prior model [1], probability measure τ must fulfill the following

separability property:

$$\tau(t^{/}, t^{-}, t^{\backslash}) = \tau(t^{/})\tau(t^{-})\tau(t^{\backslash}). \tag{16}$$

Under this condition, each column $\boldsymbol{Q}_j, \boldsymbol{T}_j^{/}, \boldsymbol{T}_j^{-}$ and $\boldsymbol{T}_j^{\backslash}$ is white and Bernoulli distributed marginally from the rest of the field. The respective characteristic parameters of the Bernoulli distributions are given by

$$\lambda = \tau(q = 1) \tag{17a}$$

$$\begin{cases} \mu^{/} = \tau(t^{/} = 1) \\ \mu^{-} = \tau(t^{-} = 1) \\ \mu^{\backslash} = \tau(t^{\backslash} = 1). \end{cases} \tag{17b}$$

The marginal Bernoulli distribution of columns $\boldsymbol{Q}_j$ indicates that MBRF's are consistent with the Bernoulli distribution of reflector locations which is assumed in 1-D B-G deconvolution methods.

If (16) holds, the joint probability of $\{\boldsymbol{T}, Q\}$ is left invariant by simultaneous reversion of horizontal and vertical indices. In addition, if τ presents a horizontal symmetry, i.e.,

$$\tau(q, t^{/}, t^{-}, t^{\backslash}) = \tau(q, t^{\backslash}, t^{-}, t^{/}) \tag{18}$$

then $P(\boldsymbol{t}, q)$ is left unchanged by independent reversion of horizontal or vertical indices. From a practical standpoint, such symmetry properties are desirable, since the direction in which the sites are indexed is not connected to any physical property and therefore should not affect the prior model. In order to fully define τ, it is sufficient to impose that

$$\tau(t^{/} = 0, t^{-} = 0, t^{\backslash} = 0 | q = 0) = 1 \tag{19}$$

$$\tau(q = 1 | t^{/} = 0, t^{-} = 0, t^{\backslash} = 0) = \varepsilon. \tag{20}$$

Equation (19) simply indicates that isolated transition variables cannot be set to one, and this is consistent with their definition. Equation (20) authorizes some discontinuities (corresponding to geological faults or inclusions, for example) along layer boundaries. Through an appropriate application of Bayes' rule, it is possible to deduce all values of $\tau(q, t^{/}, t^{-}, t^{\backslash})$ from (17b), (19), and (20). Therefore, the set of parameters $\{\mu^{/}, \mu^{-}, \mu^{\backslash}, \varepsilon\}$ fully describes the family of MBRF's. These four parameters take their values between zero and one, and each of them can be given a physical interpretation: $\mu^{/}, \mu^{-}$, and $\mu^{\backslash}$ are the probabilities of occurrence of upward, horizontal, and downward transitions, while ε represents the probability of occurrence of discontinuities along layer boundaries. Characteristic parameter λ introduced in (17a) is related to $\{\mu^{/}, \mu^{-}, \mu^{\backslash}, \varepsilon\}$ by the following relationship:

$$\lambda = 1 - (1 - \mu^{/})(1 - \mu^{-})(1 - \mu^{\backslash})(1 - \varepsilon). \tag{21}$$

It should be stressed that MBRF's $\{\boldsymbol{T}, Q\}$ also present a hierarchical structure, and that transition variables play the role of hidden variables or line processes. These transition variables are crucial in the derivation of a tractable model, as they express the relationship between adjacent columns of location variables using *local* interactions. The hierarchy between $\boldsymbol{T}$ and $\boldsymbol{Q}$ is underlined by the fact that marginally from $\boldsymbol{Q}, T$ remains a UMRF with local neighborhood system, whereas the converse is not true: marginally from $\boldsymbol{T}$, Q is not a MRF with local neighborhood sets [1].

C. Amplitude Field R

We now turn to the derivation of the complete reflectivity model $(\boldsymbol{T}, Q, R)$ which of course should retain the attractive features of MBRF's, such as their UMRF structure or the locality of the interactions. At this point, it is useful to specify the prior distribution of the amplitude of the reflectors independently from one another. Here, it is assumed that this distribution is Gaussian. The validity of this hypothesis has been severely questioned, especially in the field of geophysics [14]. However, it is retained in this study because it considerably simplifies the derivation of the compete prior model and of the corresponding restoration algorithms.

Another open question concerns the independence of the reflector amplitudes in the vertical direction. The validity of such an assumption for different types of applications will not be discussed here, but two models will be proposed. The first one is based upon the assumption of independence of the reflector amplitudes in the vertical direction, and constitutes a true 2-D extension of the 1-D B-G model. In the second model, correlation is introduced in the vertical direction on the basis of simple physical considerations.

1) White Gaussian Reflectivity Sequences (M-B-G I): We start by precisely defining the amplitude distribution of the reflectors independently from one another. Since no distinction is made *a priori* between reflectors, there is no reason for assuming any heterogeneous distribution of the amplitudes, and we assume that the following condition is in force:

Homogeneity condition: Marginally from the rest of the field, the amplitude of each reflector is sampled from a unique probability distribution.

Since it was assumed that the reflector amplitudes are normally distributed, the marginal probability density function of the amplitudes of the reflectors is Gaussian with zero mean and variance equal to σ^2.

We now proceed with the explicit specification of conditional model $P(\boldsymbol{r}|t, q)$ which together with MBRF $P(\boldsymbol{t}, q)$ constitutes the complete prior model. Clearly, $P(\boldsymbol{r}|t, q)$ should preserve the UMRF and vector Markov chain structures of $P(\boldsymbol{t}, q)$ and also introduce proper interactions between reflectors located on the same layer boundary. We first extend the first-order vector Markov chain structure (14) of $(\boldsymbol{T}, Q)$ to the conditional distribution $P(\boldsymbol{r}|t, q)$ as follows

$$P(\boldsymbol{r}|t, q) = P(\boldsymbol{r}_1|\boldsymbol{q}_1) \prod_{j=2}^{J} P(\boldsymbol{r}_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}). \tag{22}$$

Since reflectivity values r_{ij} are assumed to be correlated only with reflectors located on the same boundary, we naturally factor the conditional probabilities in the RHS of (22) according to

$$P(\boldsymbol{r}_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}) = \prod_{i=1}^{I} P(r_{ij}|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}). \tag{23}$$

Now it only remains to specify the scalar conditional probability functions in the RHS of (23). Such functions must be defined in order to introduce proper correlation between $\boldsymbol{r}_{ij}$ and reflectivity values in the previous column. More precisely,

the proper form of correlation depends on the local geometry of the layers, which is fully described by elements of the binary vectors $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$. In order to clarify this point, we first give the following definition:

Definition: If transition variable $t^{/}_{i+1j-1}$ (respectively, $t^{-}_{ij-1}, t^{\backslash}_{i-1j-1}$) is set to one, then reflector r_{ij} is referred to as a successor of r_{i+1j-1} (respectively, r_{ij-1}, r_{i-1j-1}); symmetrically, the latter is referred to as a predecessor of r_{ij}.

Then, depending on the existence and uniqueness of successors and predecessors, the elementary conditional probabilities in the RHS of (23) can be easily separated into four categories (C1)–(C4):

(C1) If $q_{ij} = 0$, then $r_{ij} = 0$: there is no reflector at position (i, j).

(C2) If $q_{ij} = 1$ and if r_{ij} is the unique successor of a unique predecessor $r_{i+di,j-1}(-1 \leq di \leq 1)$, then r_{ij} is sampled from a first-order AR process, conditionally to $r_{i+di,j-1}$

$$\boldsymbol{r}_{ij} = \boldsymbol{a} r_{i+di,j-1} + \boldsymbol{n}_r \tag{24}$$

where $a \in [0, 1]$ controls the degree of correlation between amplitudes of reflectors along the same boundary. n_r follows a Gaussian distribution with zero mean and variance equal to $(1 - a^2)\sigma^2$, so as to fulfill the homogeneity condition and ensure that the variance of the AR process defined by (24) is equal to σ^2. This case corresponds to interactions along a single layer boundary.

(C3) If $q_{ij} = 1$ and if r_{ij} has no predecessor, then r_{ij} is sampled from the basic Gaussian distribution $\mathcal{N}(0, \sigma^2)$. This happens in three cases: i) for the reflectors of the first column $\boldsymbol{r}_1$; ii) when a reflector appears within Λ° with probability ε, according to (20); iii) at horizontal and vertical boundaries of Λ°, because of the free boundary assumption.

(C4) If q_{ij} = 1 and if r_{ij} has more than one predecessor, or symmetrically when r_{ij} is not a unique successor, then r_{ij} is sampled from the basic Gaussian distribution $\mathcal{N}(0, \sigma^2)$. This configuration corresponds to the most difficult cases, and it would seem natural to correlate r_{ij} with its predecessors. However, the relevant form of correlation is difficult to determine. For instance, one might think of a mere linear combination of the values of the predecessors. Unfortunately such a choice does not guarantee the whiteness of reflectivity columns, and not even the fulfillment of the homogeneity condition. On the other hand, nonlinear combinations are incompatible with the assumed homogeneous Gaussianity. In fact, we have found that the proposed choice is the only consistent one.

It should be stressed that $P(r_{ij}|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1})$ does not depend on complete vectors $\boldsymbol{t}_{j-1}, \boldsymbol{q}_j$ and $\boldsymbol{r}_{j-1}$, but only on a small number of their elements. It is easy to check that

$$\begin{aligned} &P(r_{ij}|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}) \\ &\quad = P(r_{ij}|q_{ij}, t^{/}_{i-1j-1}, t^{-}_{i-1j-1}, t^{\backslash}_{i-1j-1}, t^{/}_{ij-1}, t^{-}_{ij-1}, t^{\backslash}_{ij-1}, \\ &\qquad t^{/}_{i+1j-1}, t^{-}_{i+1j-1}, t^{\backslash}_{i+1j-1}, r_{i-1j-1}, r_{ij-1}, r_{i+1j-1}). \end{aligned} \tag{25}$$

Since the conditional probabilities in the RHS of (22) are Gaussian, they are fully characterized by their first two moments. We now give the expressions of the corresponding mean vectors $\boldsymbol{m}_j$ and covariance matrices $\boldsymbol{M}_j$ which will be used in the derivation of the deconvolution algorithms. Define

$$\boldsymbol{m}_j = (m_{ij}) \triangleq \begin{cases} E\{\boldsymbol{R}_1|\boldsymbol{q}_1\} & \text{if } j = 1 \\ E\{\boldsymbol{R}_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}\} & \text{for any } j > 1 \end{cases} \tag{26}$$

$$\boldsymbol{M}_j = (M^j_{ik}) \triangleq \begin{cases} E\{\boldsymbol{R}_1\boldsymbol{R}_1'|\boldsymbol{q}_1\} - \boldsymbol{m}_1\boldsymbol{m}_1' & \text{if } j = 1 \\ E\{\boldsymbol{R}_j\boldsymbol{R}_j'|\boldsymbol{q}_j, \\ \quad \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}\} - \boldsymbol{m}_j\boldsymbol{m}_j' & \text{for any } j > 1 \end{cases} \tag{27}$$

Then, in the four cases defined above, we obtain the following expressions of these quantities:

$$\begin{array}{lll} \text{(C1)}: & m_{ij} = 0, & M^j_{ii} = 0 \\ \text{(C2)}: & m_{ij} = a r_{i+di,j-1}, & M^j_{ii} = (1 - a^2)\sigma^2 \\ \text{(C3)\&(C4)}: & m_{ij} = 0, & M^j_{ii} = \sigma^2 \end{array} \tag{28}$$

and $\boldsymbol{M}_j$ is a diagonal matrix (this is obvious from the conditional statistical independence stated in (23)).

By recursion along the columns, it is almost immediate to check that the above definition of the conditional probability guarantees that the homogeneity property holds and that the reflectivity columns are white. In addition, since each column $\boldsymbol{Q}_j$ is Bernoulli distributed, it follows that each reflectivity column $(\boldsymbol{Q}_j, \boldsymbol{R}_j)$ is a B-G process, with characteristic parameters (λ, σ^2). The complete model $\boldsymbol{X} = (\boldsymbol{T}, \boldsymbol{Q}, \boldsymbol{R})$ will be referred to as M-B-G I. Remarkably enough, it can easily be checked that the B-G model is a special case of the M-B-G I representation, in which lateral correlation is canceled by choosing an underlying MBRF with parameters

$$\mu^{/} = \mu^{-} = \mu^{\backslash} = 0 \Rightarrow \lambda = \varepsilon. \tag{29}$$

2) White Gaussian Log-Impedance Sequences (M-B-G II) : Model M-B-G I provides a true 2-D extension of the 1-D B-G representation. However, it is difficult to give a physical interpretation to the choice of amplitude distribution in the case of multiple junctions, i.e., to condition (C4). In order to correct for this deficiency, we introduce a new kind of amplitude variables Y_{ij} which are related to reflectivity variables R_{ij} by first-order differentiation in the vertical direction:

$$R_{ij} = \frac{1}{2}(Y_{ij} - Y_{i-1j}). \tag{30}$$

Note that variables Y_{ij} must be defined even for $i = 0$ to ensure that (30) holds for every R_{ij} on Λ°. Equation (30) can be rewritten in matrix form as

$$\begin{aligned} &\boldsymbol{R}_j = \frac{1}{2}\boldsymbol{D}\boldsymbol{Y}_j, \\ &\text{where } \boldsymbol{D} \triangleq \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & & & 0 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \\ &\text{and } \boldsymbol{Y}_j \triangleq (Y_{ij}), 0 \leq i \leq I. \end{aligned} \tag{31}$$

Variables Y_{ij} have an appealing physical interpretation: it is shown in [15] that the relationship between reflector

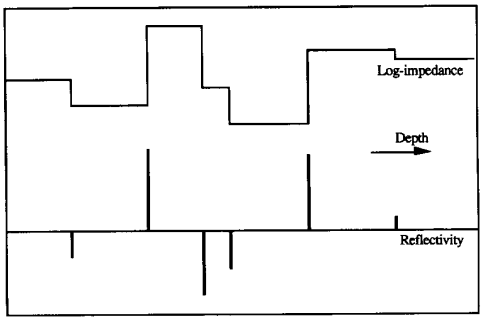


Fig. 5. Correspondence between reflectivity $\boldsymbol{R}$ and log-impedance $\boldsymbol{Y}$: Reflectivity columns are half the first-order differences of the log-impedance sequences, so sparse reflectivity columns $\boldsymbol{R}_j$ correspond to blocky log-impedance sequences $\boldsymbol{Y}_j$.

amplitudes R_{ij} and acoustic impedance values I_{ij} can be reliably approximated as:

$$R_{ij} \approx \frac{1}{2} \ln\left(\frac{I_{ij}}{I_{i-1j}}\right) \tag{32}$$

when reflectivity values are small with respect to one. Comparison of (30) and (32) shows that variables Y_{ij} are approximately equal to the logarithm of acoustic impedance, to within an arbitrary additive constant. Therefore, $\boldsymbol{Y}$ will be referred to as the field of *log-impedance*. Obviously, sparse reflectivity columns $\boldsymbol{R}_j$ correspond to log-impedance columns $\boldsymbol{Y}_j$ that are constant within layers (see Fig. 5). This characteristic, which is called *blockyness* in [15], has been accounted for by several authors in the form of Markov [15] or semi-Markov [16] chains. These models generally assume that the values of the acoustic impedance of different layers are independent. The same assumption is made in this subsection.

In order to derive the second model, referred to as M-B-G II, local interactions are defined using impedance (or log-impedance) variables, which have a physical meaning. The reflectivity model is deduced afterwards from (30). The following set of assumptions is used:

(A1) The first-order vector Markov chain structure (14) of geometric model $(\boldsymbol{T}, Q)$ extends to the whole field $(\boldsymbol{T}, Q, Y)$ as follows:

$$P(\boldsymbol{y}|\boldsymbol{t}, q) = P(\boldsymbol{y}_1|\boldsymbol{q}_1) \prod_{j=2}^{J} P(\boldsymbol{y}_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{y}_{j-1}). \tag{33}$$

(A2) If a layer is created at column j, its log-impedance value is sampled from a Gaussian distribution with arbitrary mean (zero for sake of simplicity) and variance $2\sigma^2$. This happens in five cases, graphically represented in Fig. 6: i) in the first column $\boldsymbol{Y}_1$; ii) at horizontal and vertical boundaries of $\Lambda°$, because of the free boundary assumption; iii) when new layers appear due to layer boundary splitting; iv) when a new boundary is created with probability ε (the upper and lower new layers are assigned independent log-impedance values); v) symmetrically, when a boundary vanishes.

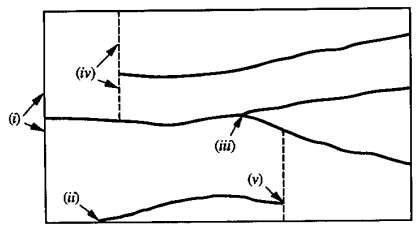


Fig. 6. Special cases to be handled in the M-B-G II approach: When the seismic section is examined from left to right, new log-impedance values are introduced in columns where new layers arise. This may happen in five different cases i)–v), described in Section III–C2. In the example depicted here, the arrows point at the beginning of each layer and the labels refer to the relevant case.

(A3) If a layer of column j already exists in column $j-1$, the associated log-impedance value y is sampled from a first-order AR process, conditionally to the previous value y_p:

$$y = a y_p + n_p \tag{34}$$

where n_p follows a Gaussian distribution with zero mean and variance equal to $(1-a^2)2\sigma^2$, and where $a \in [0, 1]$ controls the degree of correlation between log-impedance values across the same layer. It may seem simpler to impose $a = 1$, i.e., $y = y_p$, so that log-impedance is rigorously constant within layers. But it is preferable to introduce some degree of variability from one column to the next, so as to account for possible slow changes across the reflectivity section, and in order to prevent instabilities of the estimation procedures induced by modeling errors.

Note that (A2) and (A3) yield an implicit specification of the conditional probabilities in the RHS of (33); explicit specification of these conditional probabilities would have been much more difficult. Given such specifications, it is straightforward to check that the following two properties hold:

(P1) In each column $\boldsymbol{Y}_j$, the log-impedance is constant within each layer and independently distributed among different layers.

(P2) The marginal distribution of log-impedance values is Gaussian with zero mean and variance $2\sigma^2$. The marginal distribution of the reflector amplitudes is also Gaussian with zero mean, and their variance is still σ^2. However, within a column, the reflector amplitudes are no longer independent. The amplitudes of adjacent reflectors are correlated, and the correlation coefficient is equal to $-\sigma^2/2$. More precisely, the covariance matrix $\boldsymbol{P}_j = (P^j_{ik}) \triangleq \mathrm{E}\{\boldsymbol{R}_j \boldsymbol{R}j' | \boldsymbol{q}_j\}$ of reflectivity column $\boldsymbol{R}_j$ given the location column $\boldsymbol{Q}_j = \boldsymbol{q}_j$ and marginally from the rest of the field is defined by:

$$\begin{cases} q_{ij} = 1 \Rightarrow P^j_{ii} = \sigma^2, \\ \forall i, \forall k, k > i, (q_{ij}, \cdots, q_{kj}) \\ \quad = (1, 0, \cdots, 0, 1) \Rightarrow P^j_{ik} = P^j_{ki} = -\sigma^2/2, \\ \text{all other entries of } \boldsymbol{P}_j \text{ are equal to zero} \end{cases} \tag{35}$$

which gives matrix $\boldsymbol{P}_j$ the following structure:

$$\boldsymbol{q}_j = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} \Rightarrow P_j$$

$$= \sigma^2 \begin{bmatrix} 0 & & & & & & & & & \\ & 1 & 0 & \cdots & 0 & -1/2 & & & & \\ & 0 & 0 & \cdots & 0 & 0 & & & & \\ & \vdots & \vdots & \ddots & \vdots & \vdots & & & & \\ & 0 & 0 & \cdots & 0 & 0 & & & & \\ & -1/2 & 0 & \cdots & 0 & 1 & 0 & -1/2 & & \\ & & & & & 0 & 0 & 0 & & \\ & & & & & -1/2 & 0 & 1 & 0 & \cdots \\ & & & & & & & 0 & 0 & \\ & & & & & & & \vdots & \vdots & \ddots \end{bmatrix} \tag{36}$$

Finally, it is useful to determine whether the Markov chain structure imposed on $(\boldsymbol{T}, Q, Y)$ implies that $(\boldsymbol{T}, Q, R)$ is also a Markov chain, i.e., whether (33) implies (22). Rigorously, this implication does not hold because linear equation (31) cannot be inverted exactly: $\boldsymbol{y}_j$ can be deduced from a reflectivity column $\boldsymbol{r}_j$ only to within an arbitrary additive constant. However, as we assume that the log-impedance is zero mean, we can always choose the constant so that the mean value of $\boldsymbol{y}_j$ is zero. As a consequence, we shall consider that (22) is also fulfilled for the M-B-G II model. Mathematically, this corresponds to the following approximation:

$$P(\boldsymbol{r}_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}) \approx P(\boldsymbol{r}_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \tilde{\boldsymbol{y}}_{j-1}) \tag{37}$$

where $\tilde{\boldsymbol{y}}_{j-1})$ is the most probable log-impedance sequence compatible with reflectivity sequence $\boldsymbol{r}_{j-1}$. Log-impedance $\tilde{\boldsymbol{y}}_{j-1}$ can be obtained by inversion of the following linear system:

$$\begin{cases} \boldsymbol{D}\tilde{\boldsymbol{y}}_{j-1} = 2\boldsymbol{r}_{j-1} \\ \tilde{\boldsymbol{q}}'_{j-1}\tilde{\boldsymbol{y}}_{j-1} = 0 \end{cases} \text{ where } \tilde{\boldsymbol{q}}_{j-1} \triangleq \left[\frac{1}{\boldsymbol{q}_{j-1}}\right]. \tag{38}$$

The first equation expresses the relationship between $\tilde{\boldsymbol{y}}_{j-1}$ and $\boldsymbol{r}_{j-1}$ (see (30) and (31)); the second equation imposes that the mean-value of log-impedance $\tilde{\boldsymbol{y}}_{j-1}$ be zero. Inversion of (38) can be trivially computed in two steps:

1) $\overline{y}_{0j-1} = 0$ and $\overline{y}_{ij-1} = \overline{y}_{i-1j-1} + 2r_{ij-1}$, for any $i > 1$ (39a)

2) For any $i, \tilde{y}_{ij-1} = \overline{y}_{ij-1} - \dfrac{\tilde{\boldsymbol{q}}'_{j-1}\overline{\boldsymbol{y}}_{j-1}}{\tilde{\boldsymbol{q}}'_{j-1}\tilde{\boldsymbol{q}}_{j-1}}$. (39b)

As for the M-B-G I model, we now compute the mean values and covariance matrices which characterize the Gaussian conditional probabilities in the RHS of (22) and (33). We first define the conditional mean-value vector $\boldsymbol{s}_j$ and covariance matrix $\boldsymbol{S}_j$ of $\boldsymbol{Y}_j$ as

$$\boldsymbol{s}_j = (s_{ij}) \triangleq \begin{cases} \mathrm{E}\{\boldsymbol{Y}_1|\boldsymbol{q}_1\} & \text{if } j = 1 \\ \mathrm{E}\{\boldsymbol{Y}_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{y}_{j-1}\} & \text{for any } j > 1 \end{cases} \tag{40}$$

$$\boldsymbol{S}_j = (S^j_{ik}) \triangleq \begin{cases} \mathrm{E}\{\boldsymbol{Y}_1\boldsymbol{Y}'_1|\boldsymbol{q}_1\} - \boldsymbol{s}_1\boldsymbol{s}'_1 & \text{if } j = 1 \\ \mathrm{E}\{\boldsymbol{Y}_j\boldsymbol{Y}'_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{y}_{j-1}\} - \boldsymbol{s}_j\boldsymbol{s}'_j & \text{for any } j > 1 \end{cases} \tag{41}$$

Then, using assumptions (A2) and (A3), we easily obtain

$$\begin{array}{lll} \text{Case (A2):} & s_{ij} = 0 & S^j_{ii} = 2\sigma^2 \\ \text{Case (A3):} & s_{ij} = ay_p, & S^j_{ii} = (1-a^2)2\sigma^2 \end{array} \tag{42}$$

$\boldsymbol{s}_j$ is a piecewise-constant vector whose constant sections correspond to the position of each layer. $\boldsymbol{S}_j$ is a block-diagonal matrix whose *constant* diagonal blocks take their values according to (42) and are in correspondence with the constant sections of $\boldsymbol{s}_j$. For instance, (37) gives the following structure for $\boldsymbol{s}_j$ and $\boldsymbol{S}_j$:

$$\boldsymbol{s}_j = a \begin{bmatrix} y_{p1} \\ y_{p1} \\ 0 \\ 0 \\ 0 \\ y_{p2} \\ y_{p3} \\ y_{p3} \\ \vdots \end{bmatrix}, \boldsymbol{S}_j = 2\sigma^2$$

$$\times \begin{bmatrix} 1-a^2 & 1-a^2 & & & & & & & \\ 1-a^2 & 1-a^2 & & & & & & & \\ & & 1 & 1 & 1 & & & & \\ & & 1 & 1 & 1 & & & & \\ & & 1 & 1 & 1 & & & & \\ & & & & & 1-a^2 & & & \\ & & & & & & 1-a^2 & 1-a^2 & \cdots \\ & & & & & & 1-a^2 & 1-a^2 & \\ & & & & & & \vdots & & \ddots \end{bmatrix} \tag{43}$$

Then $\boldsymbol{m}_j$ and $\boldsymbol{M}_j$ (defined in (26) and (27)) are deduced from $\boldsymbol{s}_j$ and $\boldsymbol{S}_j$ using (31) and take the following form:

$$\begin{aligned} \boldsymbol{m}_j &= \frac{1}{2}\boldsymbol{D}\,\boldsymbol{s}_j \\ \boldsymbol{M}_j &= \frac{1}{4}\boldsymbol{D}\,\boldsymbol{S}_j\,\boldsymbol{D}'. \end{aligned} \tag{44}$$

IV. Multichannel M-B-G Deconvolution

In the previous section, a new class of 2-D statistical representations of layered reflectivity sections has been introduced. More precisely, two models have been proposed. The M-B-G I representation can be thought of as the most natural extension of 1-D B-G models, while the M-B-G II representation has more physical consistency. In this section, we address the problem of deconvolution of reflectivity sections modeled as M-B-G fields. The Bayesian framework defined in Section II is

used, and efficient algorithms are derived for the computation of MAP estimates of the reflectivity. Whenever possible, the algorithms are given in a unique formulation which is valid for both M-B-G I and M-B-G II priors, and extends previous results relevant to 1-D B-G deconvolution [17].

A. MAP Estimation

1) Exact Expression of the MAP Criterion: Let $J(\boldsymbol{x}) \triangleq P(\boldsymbol{x}|z)$ denote the *a posteriori* likelihood of a realization of a M-B-G field $\boldsymbol{X} = (T, Q, R)$ given the observed data $\boldsymbol{z}$. Using Bayes' rule, one obtains

$$J(\boldsymbol{x}) \propto P(\boldsymbol{z}|x)P(\boldsymbol{x}) \Rightarrow J(\boldsymbol{x}) \propto P(\boldsymbol{z}|r)P(\boldsymbol{x}) \quad (45)$$

since $P(\boldsymbol{z}|x)$ and $P(\boldsymbol{z}|r)$ are identical, as indicated by (1). This again reflects the fact that the geometric part ($\boldsymbol{T}$, Q) of M-B-G models is a hidden variable of the estimation problem. Using (14) and (22), prior probability $P(\boldsymbol{x})$ can be factored as

$$\begin{aligned} P(\boldsymbol{x}) &= P(\boldsymbol{r}_1|\boldsymbol{q}_1)P(\boldsymbol{q}_1)\prod_{j=2}^{J} P(\boldsymbol{r}_j|\boldsymbol{q}_j, \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}) \\ &\quad \times P(\boldsymbol{q}_j|\boldsymbol{t}_{j-1})P(\boldsymbol{t}_{j-1}|\boldsymbol{q}_{j-1}) \\ &= P(\boldsymbol{q}_1, \boldsymbol{r}_1)\prod_{j=2}^{J} P(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j|\boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) \end{aligned} \quad (46)$$

which together with the expression of $P(\boldsymbol{z}|r)$ given in (7), yields the following factored form of $J(\boldsymbol{x})$:

$$\begin{aligned} J(\boldsymbol{x}) \propto\ & P(\boldsymbol{z}_1|\boldsymbol{r}_1)P(\boldsymbol{q}_1, \boldsymbol{r}_1) \\ & \prod_{j=2}^{J} P(\boldsymbol{z}_j|\boldsymbol{r}_j)P(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j|\boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}). \end{aligned} \quad (47)$$

In order to compute the exact MAP solution, the *a posteriori* likelihood criterion $J(\boldsymbol{x})$ must be maximized over the whole state-space of $\boldsymbol{X}$. One can take advantage of the factored form given in (47) and maximize $J(\boldsymbol{x})$ in a recursive manner using a Viterbi algorithm [18]. From a theoretical standpoint, application of such a procedure does not present any difficulty, as ($\boldsymbol{T}$, Q) is a hidden Markov chain with a finite number of states. However, the state-space of vector Markov chain ($\boldsymbol{T}$, Q) has such a large dimension that the computational cost of the corresponding Viterbi algorithm is beyond the capabilities of computers available now or in the foreseeable future. To our knowledge, no other exact optimization procedure can reasonably be implemented, and maximization of $J(\boldsymbol{x})$ must be carried out in some suboptimal manner. In the sequel, we propose a suboptimal *column-recursive* restoration algorithm which takes advantage of the factorization of $J(\boldsymbol{x})$ given in (47).

2) Suboptimal Maximization of the MAP Criterion: We start by defining partial likelihood criteria J_j as

$$\begin{cases} J_1(\boldsymbol{q}_1, \boldsymbol{r}_1) \triangleq P(\boldsymbol{q}_1, \boldsymbol{r}_1, \boldsymbol{z}_1) \\ \forall\ j \in [2, J], J_j(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) \\ \quad \triangleq P(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j, \boldsymbol{z}_j|\boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}). \end{cases} \quad (48)$$

Bayes' rule allows us to rewrite (47) as a function of the partial criteria:

$$J(\boldsymbol{x}) \propto J_1(\boldsymbol{q}_1, \boldsymbol{r}_1)\prod_{j=2}^{J} J_j(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}). \quad (49)$$

Obviously, maximization of the partial likelihood criteria independently from one another is impossible, as each argument $\boldsymbol{q}_j$ and $\boldsymbol{r}_j$, $2 \leq j \leq J$, enters the expression of more than one partial criterion. This simply reflects the 2-D nature of the M-B-G prior model. However, recursive formula (49) suggests the following maximization procedure:

$$\text{(1) Initial step:} \quad (\hat{\boldsymbol{q}}_1, \hat{\boldsymbol{r}}_1) = \underset{\boldsymbol{q}_1, \boldsymbol{r}_1}{\operatorname{argmax}}\ J_1(\boldsymbol{q}_1, \boldsymbol{r}_1) \quad (50a)$$

$$\text{(2) } j \in [2, J]: \quad (\hat{\boldsymbol{t}}_{j-1}, \hat{\boldsymbol{q}}_j, \hat{\boldsymbol{r}}_j) = \underset{\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j}{\operatorname{argmax}}\ J_j(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j, \hat{\boldsymbol{q}}_{j-1}, \hat{\boldsymbol{r}}_{j-1}) \quad (50b)$$

In the above algorithm, the data are processed in a column-recursive manner: reflectivity column $\boldsymbol{r}_j$ and corresponding hidden binary vectors $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$ are determined from the current observed trace $\boldsymbol{z}_j$ and from the estimates of the previous reflectivity columns. As indicated by (50), each recursion involves maximization of a single partial criterion. The procedure is clearly suboptimal since i) partial criterion J_j is maximized only with respect to $\boldsymbol{t}_{j-1}$, $\boldsymbol{q}_j$ and $\boldsymbol{r}_j$: all previously estimated quantities are frozen and $\boldsymbol{q}_{j-1}$ and $\boldsymbol{r}_{j-1}$, which also enter the expression of J_j, are set to the values $\hat{\boldsymbol{q}}_{j-1}$ and $\hat{\boldsymbol{r}}_{j-1}$ obtained at the previous recursion; ii) determination of the jth reflectivity column is based on observations only up to trace $\boldsymbol{z}_j$, and not on the whole set of observed data. For instance, $\hat{\boldsymbol{r}}_j$ is evaluated without accounting for $\boldsymbol{z}_{j+1}$ which is undoubtedly very informative about $\boldsymbol{r}_j$ because of strong lateral correlation in the layered medium. These drawbacks are common to all such *decision-directed* procedures, which nonetheless have been shown to perform well in a 1-D framework as long as the operating conditions are not too extreme [19]–[21]. The degree of suboptimality of procedure (50) will be further commented upon in Section V on the basis of deconvolution results. On the other hand, the proposed procedure is much simpler than global maximization of $J(\boldsymbol{x})$, since a set of 1-D maximization problems has been substituted for the original 2-D optimization problem.

B. Maximization of Partial Likelihood Criteria

1) Detection-Estimation Strategies: The methodological problems associated to the maximization of partial criteria J_j are essentially similar to those encountered in 1-D B-G deconvolution. Procedure (50) involves a detection step (determination of binary variables $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$) and an estimation step (determination of amplitudes $\boldsymbol{r}_j$). Due to conditionally Gaussian distribution of the amplitudes, the estimation step is rather easy to solve and the solution is available in closed form. The detection step has a crucial importance, because it presents great algorithmic difficulties, but also because the quality of the detection has a major

impact on the overall performances of the restoration method. In this respect, it should be underlined that partial criteria J_j correspond to joint probability distributions of binary and amplitude variables, and that maximization of such criteria has been reported to lack robustness and to often produce a large number of false alarms in 1-D B-G problems [3], [20]. The problem can be alleviated by performing detection and estimation sequentially: detection is carried out first through maximization of a modified *marginal* likelihood function, and estimation is performed in second step using the result of the detection operation. The sequential approach can be derived as follows: using Bayes' rule, the expression of partial criteria is rewritten as

$$\begin{cases} J_1(\boldsymbol{q}_1, \boldsymbol{r}_1) = P(\boldsymbol{q}_1, z_1) P(\boldsymbol{r}_1 | z_1, \boldsymbol{q}_1) \\ \forall\, j \in [2, J], J_j(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) \\ \quad = P(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, z_j | \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) \\ \quad\quad P(\boldsymbol{r}_j | z_j, \boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) \end{cases} \tag{51}$$

J_j appears as the product of two probability distributions. The first one can be expressed as

$$\begin{cases} M_1(\boldsymbol{q}_1) \triangleq \int J_1(\boldsymbol{q}_1, \boldsymbol{r}_1) d\boldsymbol{r}_1 = P(\boldsymbol{q}_1, z_1) \\ \forall\, j \in [2, J], \boldsymbol{M}_j(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) \\ \quad \triangleq \int J_j(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{r}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) d\boldsymbol{r}_j \\ \quad = P(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, z_j | \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) \end{cases} \tag{52}$$

and is proportional to the joint posterior distribution of hidden binary vectors $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$. It can be considered as a partial *marginal* conditional likelihood in which the amplitudes are treated as nuisance parameters and are integrated out of the detection problem. It constitutes the partial detection criteria which are maximized in the first step of the sequential approach. In (51), the second term of the product represents the posterior density of the amplitude variables conditionally to the knowledge of the binary variables. It is Gaussian under the assumptions stated in Section III and the solution can be written in closed form. Its computation constitutes the second step of the sequential approach, which can be summarized as follows:

$$\text{(1) First column:} \begin{cases} \text{detection}: \ \hat{\boldsymbol{q}}_1 = \underset{\boldsymbol{q}_1}{\text{argmax}} \boldsymbol{M}_1(\boldsymbol{q}_1) \\ \text{estimation}: \ \hat{\boldsymbol{r}}_1 = \underset{\boldsymbol{r}_1}{\text{argmax}}\, J_1(\hat{\boldsymbol{q}}_1, \boldsymbol{r}_1) \end{cases} \tag{53a}$$

$$\text{(2) } j \in [2, J]: \begin{cases} \text{detection}: \ (\hat{\boldsymbol{t}}_{j-1}, \hat{\boldsymbol{q}}_j) \\ \quad = \underset{\boldsymbol{t}_{j-1}, \boldsymbol{q}_j}{\text{argmax}} \\ \quad\quad \boldsymbol{M}_j(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \hat{\boldsymbol{q}}_{j-1}, \hat{\boldsymbol{r}}_{j-1}) \\ \text{estimation}: \ \hat{\boldsymbol{r}}_j \\ \quad = \underset{\boldsymbol{r}_j}{\text{argmax}}\, J_j(\hat{\boldsymbol{t}}_{j-1}, \hat{\boldsymbol{q}}_j, \boldsymbol{r}_j, \hat{\boldsymbol{q}}_{j-1}, \hat{\boldsymbol{r}}_{j-1}) \end{cases} \tag{53b}$$

As simulation results (not reported here) have confirmed the superiority of the sequential approach over the simultaneous scheme, only the former is described in the sequel. Transposition of the following derivations to simultaneous detection-estimation is fairly straightforward and can be performed in a way similar to the one described in [17].

2) Explicit Forms for Detection-Estimation: We now give the detailed expressions of criteria M_j used for detection of binary vectors $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$, and of the estimated reflectivity values $\hat{\boldsymbol{r}}_j$. These expressions are valid for both M-B-G I and M-B-G II prior models, as the two representations only differ by the structure of the mean vector and covariance matrix which characterize the Gaussian part of the model. Moreover the derivations are also valid for 1-D B-G deconvolution since, according to (29), the B-G model is a special case of the M-B-G I model.

When vectors $\boldsymbol{t}_{j-1}, \boldsymbol{q}_j$ (respectively, $\boldsymbol{q}_1$) and $\boldsymbol{r}_{j-1}$ are known, it is clear by inspection of (51) that maximizing J_j with respect to the reflectivity values is equivalent to maximizing *a posteriori* density $P(\boldsymbol{r}_j | z_j, \boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1})$ (respectively, $P(\boldsymbol{r}_1 | z_1, \boldsymbol{q}_1)$). Since the prior distribution of $\boldsymbol{r}_j$ is Gaussian, and since the relationship between $\boldsymbol{r}_j$ and z_j is linear with additive Gaussian noise (see (3)), $P(z_j | \boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1})$ (resp. $P(z_1 | \boldsymbol{q}_1)$) is also Gaussian and the estimate of $\boldsymbol{r}_j$ is given by the classical MAP formulas in a linear and Gaussian setting:

$$\hat{\boldsymbol{r}}_j = \boldsymbol{m}_j + \boldsymbol{M}_j \boldsymbol{H}' (\boldsymbol{H} \boldsymbol{M}_j \boldsymbol{H}' + \sigma_n^2 I_I)^{-1} (z_j - \boldsymbol{H} \boldsymbol{m}_j) \tag{54}$$

where I_N denotes the identity matrix of size (N, N); $\boldsymbol{m}_j$ and $\boldsymbol{M}_j$, respectively, denote the prior mean value and covariance matrix of $\boldsymbol{r}_j$ which are given by (28) in the M-B-G I case and by (42)–(44) in the M-B-G II case. The mean value $\boldsymbol{b}_j$ and covariance matrix $\boldsymbol{B}_j$ which characterize $P(z_j | \boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1})$ (respectively, $P(z_1 | \boldsymbol{q}_1)$) are expressed as

$$\boldsymbol{b}_j = \boldsymbol{H} \boldsymbol{m}_j \tag{55a}$$

$$\boldsymbol{B}_j = \boldsymbol{H} \boldsymbol{M}_j \boldsymbol{H}' + \sigma_n^2 \boldsymbol{I}_I \tag{55b}$$

and (54) can be written in the following simpler form:

$$\hat{\boldsymbol{r}}_j = \boldsymbol{m}_j + \boldsymbol{M}_j \boldsymbol{H}' \boldsymbol{B}_j^{-1} (z_j - \boldsymbol{b}_j). \tag{56}$$

In order to derive the expression of marginal detection criteria M_j, the formulas given in (52) are transformed using Bayes' rule. One has

$$\begin{cases} M_1(\boldsymbol{q}_1) \propto P(\boldsymbol{z}_1 | \boldsymbol{q}_1) P(\boldsymbol{q}_1) \\ \forall\, j \in [2, J], \\ M_j(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j, \boldsymbol{q}_{j-1}, \boldsymbol{r}_{j-1}) \propto P(z_j | \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}, \boldsymbol{q}_j) \\ \quad P(\boldsymbol{q}_j | \boldsymbol{t}_{j-1}) P(\boldsymbol{t}_{j-1} | \boldsymbol{q}_{j-1}) \end{cases} \tag{57}$$

As noted above, the conditional distributions of z_j which appear in (57) are Gaussian with mean value $\boldsymbol{b}_j$ and covariance matrix $\boldsymbol{B}_j$. They can therefore be written in the following logarithmic form, to within an additional constant indicated by the sign "#":

$$\begin{cases} \ln(P(\boldsymbol{z}_1 | \boldsymbol{q}_1)) \# - e_1/2 - \frac{1}{2} \ln(|\boldsymbol{B}_1|) \\ \forall\, j \in [2, J], \ln(P(\boldsymbol{z}_j | \boldsymbol{t}_{j-1}, \boldsymbol{r}_{j-1}, \boldsymbol{q}_j)) \\ \quad \# - e_j/2 - \frac{1}{2} \ln(|\boldsymbol{B}_j|) \end{cases} \tag{58}$$

where

$$e_j \triangleq (\boldsymbol{z}_j - \boldsymbol{b}_j)' \boldsymbol{B}_j^{-1} (\boldsymbol{z}_j - \boldsymbol{b}_j). \tag{59}$$

We now need to find the explicit forms of the conditional probabilities of the geometric part of the model. This can be

done by using the factored form of the MBRF distribution (15) and the definition of basic measure τ given in (16), (17), (19), (20). The result can be expressed as a function of five scalar quantities defined as follows:

I_j: number of nonzero samples of Bernoulli sequence $\boldsymbol{q}_j$
I_j^{ε}: number of nonzero samples of $\boldsymbol{q}_j$ *with no predecessor*
$I_j^{/}$: number of nonzero samples of Bernoulli sequence $\boldsymbol{t}_j^{/}$
I_j^{-}: number of nonzero samples of Bernoulli sequence $\boldsymbol{t}_j^{-}$
$I_j^{\backslash}$: number of nonzero samples of Bernoulli sequence $\boldsymbol{t}_j^{\backslash}$.

It should be noted that there is an implicit dependence between I_j^{ε} and $\boldsymbol{t}_{j-1}$, and that the five quantities defined above do not form a redundant set, as they are related through inequalities only. The conditional distributions of the binary quantities take the following logarithmic form:

$$\ln(P(\boldsymbol{q}_1))\#I_j \ln\left(\frac{\lambda}{1-\lambda}\right) \tag{60a}$$

$$\forall\, j \in [2,J], \ln\big(P(\boldsymbol{q}_j|\boldsymbol{t}_{j-1})\big)\#I_j^{\varepsilon}\ln(\varepsilon) - I_j\ln(1-\varepsilon) \tag{60b}$$

$$\forall\, j \in [2,J], \ln\big(P(\boldsymbol{t}_{j-1}|\boldsymbol{q}_{j-1})\big)\#I_{j-1}^{/}\ln\left(\frac{\mu^{/}}{1-\mu^{/}}\right) + I_{j-1}^{-}\ln\left(\frac{\mu^{-}}{1-\mu^{-}}\right) + I_{j-1}^{\backslash}\ln\left(\frac{\mu^{\backslash}}{1-\mu^{\backslash}}\right) - (I_j - I_j^{\varepsilon})\ln(\varepsilon) \tag{60c}$$

Note that derivation of (60c) makes use of the number of nonzero samples of $\boldsymbol{q}_{j-1}$ which do not have any successor. Such a quantity need not be defined explicitly, since it is equal to $I_{j-1} + I_j^{\varepsilon} - I_j$. From (57)–(59), one finally obtains the logarithmic form L_j of marginal detection criteria $\boldsymbol{M}_j$ as an explicit function of $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$:

$$L_1(\boldsymbol{q}_1)\# - \frac{e_1}{2} - \frac{1}{2}\ln(|\boldsymbol{B}_1|) - I_1\ln\left(\frac{1-\lambda}{\lambda}\right) \tag{61a}$$

$$\forall\, j \in [2,J], L_j(\boldsymbol{t}_{j-1},\boldsymbol{q}_j)\# - \frac{e_j}{2} - \frac{1}{2}\ln(|\boldsymbol{B}_j|) - I_j\ln\big(\varepsilon(1-\varepsilon)\big) + 2I_j^{\varepsilon}\ln(\varepsilon) - I_{j-1}^{/}\ln\left(\frac{1-\mu^{/}}{\mu^{/}}\right) - I_{j-1}^{-}\ln\left(\frac{1-\mu^{-}}{\mu^{-}}\right) - I_{j-1}^{\backslash}\ln\left(\frac{1-\mu^{\backslash}}{\mu^{\backslash}}\right) \tag{61b}$$

In the case of a M-B-G I model, $L_1(\boldsymbol{q}_1)$ is identical to the detection criterion used for B-G deconvolution. This is not surprising, as no lateral interactions are accounted for in the determination of the first reflectivity column. It can also be checked that when the parameters of the MBRF are chosen so that lateral correlation vanishes (see (29)), all criteria L_j become identical and equal to the B-G detection criterion. This indicates that the column-recursive approach introduced in Section IV–A2 is a 2-D extension of B-G deconvolution.

In order to determine the jth reflectivity column according to (53b), one must first maximize detection criterion L_j with respect to $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$ and second evaluate $\boldsymbol{r}_j$ using (56). Since the number of possible configurations of $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$ is finite, L_j always admits a maximum. Unfortunately, this maximum cannot be expressed in closed form, and optimal detection requires that L_j be evaluated for all possible configurations of the binary variables. Practical implementation of such a scheme presents two major difficulties. First, the total number of possible configurations of $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$ varies from 2^I to 8^I where I denotes the number of samples per trace. In most applications, I ranges between a few hundred and a few thousand, and this corresponds to a totally unrealistic number of trials. Clearly, only a subset of the state-space of $(\boldsymbol{t}_{j-1},\boldsymbol{q}_j)$ can be explored. Second, evaluation of L_j for a single configuration of $\boldsymbol{t}_{j-1}$ and $\boldsymbol{q}_j$ is computationally burdensome, as (I,I)-dimensional matrix $\boldsymbol{B}_j$ must be inverted.

In order to circumvent these difficulties, we propose a deterministic suboptimal maximization procedure which can be seen as an extension of the SMLR-type methods used in B-G deconvolution [3]. The core of the method consists of numerically efficient formulas which relate the criterion values and amplitude estimates of two *neighboring* configurations $(\boldsymbol{t}_{j-1},\boldsymbol{q}_j)$ and $(\bar{\boldsymbol{t}}_{j-1},\bar{\boldsymbol{q}}_j)$. At this point, neighboring sequences need not be defined precisely, as the formulas are valid as long as $(\boldsymbol{t}_{j-1},\boldsymbol{q}_j)$ and $(\bar{\boldsymbol{t}}_{j-1},\bar{\boldsymbol{q}}_j)$ are different from one another. Then, the state-space of $(\boldsymbol{t}_{j-1},\boldsymbol{q}_j)$ is explored by scanning all neighboring sequences of the current configuration, selecting the one which maximizes the criterion and iterating the process until a, possibly local, maximum is reached. Neighborhoods are defined in Section IV–B4 so as to achieve an acceptable balance between the numerical efficiency of the procedure and the size of the explored subset.

3) Core Algorithm: The main computational burden for the evaluation of the detection criterion and of the amplitude estimates lies in the inversion of (I,I)-dimensional matrix $\boldsymbol{B}_j$. The general idea for deriving the algorithm is to express the relations between criteria values and amplitude estimates of two neighboring sequences using auxiliary quantities with much smaller dimensions. Let $j \in [2,J]$ denote the index of the current column, and consider two different realizations $(\boldsymbol{t}_{j-1},\boldsymbol{q}_j)$ and $(\bar{\boldsymbol{t}}_{j-1},\bar{\boldsymbol{q}}_j)$ of $(\boldsymbol{T}_{j-1},\boldsymbol{Q}_j)$. All quantities referred to $(\bar{\boldsymbol{t}}_{j-1},\bar{\boldsymbol{q}}_j)$ are denoted by a superscript sign "-", e.g., $\overline{\boldsymbol{m}}_j, \overline{\boldsymbol{M}}_j, \overline{\boldsymbol{B}}_j$, etc. The difference between the prior mean values and covariance matrices of $(\boldsymbol{t}_{j-1},\ \boldsymbol{q}_j)$ and $(\bar{\boldsymbol{t}}_{j-1},\bar{\boldsymbol{q}}_j)$ are expressed as

$$\overline{\boldsymbol{m}}_j - \boldsymbol{m}_j = \boldsymbol{u}_j \tag{62a}$$

$$\overline{\boldsymbol{M}}_j - \boldsymbol{M}_j = \boldsymbol{U}_j\boldsymbol{S}_j\boldsymbol{U}_j' \tag{62b}$$

where the difference of the two symmetric definite nonnegative matrices $\boldsymbol{M}_j$ and $\overline{\boldsymbol{M}}_j$ is factored using generalized square-root factorization: $\boldsymbol{U}_j$ and $\boldsymbol{S}_j$ are a (I,K)-dimensional matrix and a (K,K)-dimensional signature matrix, respectively, and K is the rank of $\overline{\boldsymbol{M}}_j - \boldsymbol{M}_j$. In the sequel, it is assumed that K is small with respect to I, and that $\boldsymbol{u}_j$ is sparse with $O(K)$ nonzero entries. This assumption seems reasonable, as K may be interpreted as a measure of the distance between the two *neighboring*—hence fairly similar—sequences $(\boldsymbol{t}_{j-1},\boldsymbol{q}_j)$ and $(\bar{\boldsymbol{t}}_{j-1},\bar{\boldsymbol{q}}_j)$. This also shows that the choice of neighboring sequences sets a tradeoff between the completeness of the search and the complexity of the algorithm. From a practical standpoint, determination of $\boldsymbol{S}_j$ is trivial in the case of the M-B-G I model (as both $\boldsymbol{M}_j$ and $\overline{\boldsymbol{M}}_j$ are diagonal), but

may require spectral factorization of real symmetric matrix $\overline{\boldsymbol{M}}_j - \boldsymbol{M}_j$ in the case of the M-B-G II model. This can be performed using standard mathematical libraries.

A first step toward simplification consists of relating the inverses of $\overline{\boldsymbol{M}}_j$ and $\boldsymbol{M}_j$. Equations (55b) and (62b) immediately yield

$$\overline{\boldsymbol{B}}_j = \boldsymbol{B}_j + \boldsymbol{H}\boldsymbol{U}_j\boldsymbol{S}_j\boldsymbol{U}_j'\boldsymbol{H}' \tag{63}$$

and by application of the matrix inversion lemma, we obtain

$$\overline{\boldsymbol{B}}_j^{-1} = \boldsymbol{B}_j^{-1} - \boldsymbol{B}_j^{-1}\boldsymbol{H}\boldsymbol{U}_j\boldsymbol{V}_j^{-1}\boldsymbol{U}_j'\boldsymbol{H}'\boldsymbol{B}_j^{-1} \tag{64}$$

with

$$\boldsymbol{V}_j \triangleq \boldsymbol{S}_j + \boldsymbol{U}_j'\boldsymbol{H}'\boldsymbol{B}_j^{-1}\boldsymbol{H}\boldsymbol{U}_j. \tag{65}$$

Compared to direct inversion of (55b), inversion of the (K, K) matrix $\boldsymbol{V}_j$ has been substituted for that of $\boldsymbol{B}_j$. Computation of $\ln(|\boldsymbol{B}_j|)$ can also be carried out iteratively as

$$\ln(|\overline{\boldsymbol{B}}_j|) = \ln(|\boldsymbol{B}_j|) + \ln \operatorname{abs}(|\boldsymbol{V}_j|). \tag{66}$$

Equation (66) is a straigthforward consequence of the following theorem stated and proven in [22]:

Theorem: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ denote two matrices with respective dimensions (n, m) and (m, n). Then we have

$$\forall \lambda \in \mathbb{R}, |\lambda \boldsymbol{I}_m - \boldsymbol{B}\boldsymbol{A}| = \lambda^{m-n}|\lambda \boldsymbol{I}_n - \boldsymbol{A}\boldsymbol{B}|.$$

The set of equations (64)–(66) provides a consistent iterative scheme, since L_j and $\hat{\boldsymbol{r}}_j$ can be computed from matrix and inner products involving $\boldsymbol{B}_j^{-1}$ only, from $\ln(|\boldsymbol{B}_j|)$ and from quantities that can be easily deduced from $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$, such as $\boldsymbol{b}_j, \boldsymbol{I}_j, \boldsymbol{I}_j^{\varepsilon}, \cdots$ In particular, matrix $\boldsymbol{B}_j$ need not be stored or updated. However, these formulas are still quite burdensome because each updating of L_j requires several (I, I) matrix products.

Fortunately, the computational load can be highly reduced if the iterations are performed on auxiliary quantities $e_j, \boldsymbol{w}_j$ and $\boldsymbol{A}_j$, jointly defined as

$$\left[\begin{array}{c|c} \boldsymbol{A}_j & \boldsymbol{w}_j \\ \hline \boldsymbol{w}_j' & e_j \end{array}\right] \triangleq \left[\begin{array}{c} \boldsymbol{H}' \\ \hline \boldsymbol{z}_j' - \boldsymbol{b}_j' \end{array}\right] \boldsymbol{B}_j^{-1}[\boldsymbol{H}|\boldsymbol{z}_j - \boldsymbol{b}_j]. \tag{67}$$

From (64), it is possible to obtain updating equations for $e_j, \boldsymbol{w}_j$ and $\boldsymbol{A}_j$:

$$\begin{aligned} \overline{e}_j = e_j &- 2\boldsymbol{w}_j'\boldsymbol{u}_j - \boldsymbol{w}_j'\boldsymbol{W}_j\boldsymbol{w}_j + 2\boldsymbol{w}_j'\boldsymbol{W}_j\boldsymbol{A}_j\boldsymbol{u}_j \\ &+ \boldsymbol{u}_j'\boldsymbol{A}_j\boldsymbol{u}_j - \boldsymbol{u}_j'\boldsymbol{A}_j\boldsymbol{W}_j\boldsymbol{A}_j\boldsymbol{u}_j \end{aligned} \tag{68}$$

$$\overline{\boldsymbol{w}}_j = \boldsymbol{w}_j - \boldsymbol{A}_j\boldsymbol{u}_j - \boldsymbol{A}_j\boldsymbol{W}_j\boldsymbol{w}_j + \boldsymbol{A}_j\boldsymbol{W}_j\boldsymbol{A}_j\boldsymbol{u}_j \tag{69}$$

$$\overline{\boldsymbol{A}}_j = \boldsymbol{A}_j - \boldsymbol{A}_j\boldsymbol{W}_j\boldsymbol{A}_j \tag{70}$$

with

$$\boldsymbol{W}_j \triangleq \boldsymbol{U}_j\boldsymbol{V}_j^{-1}\boldsymbol{U}'_j \tag{71}$$

$$\boldsymbol{V}_j = \boldsymbol{S}_j + \boldsymbol{U}'_j\boldsymbol{A}_j\boldsymbol{U}_j. \tag{72}$$

In conjunction with (66), (68)–(72) constitute a new iterative scheme of very high practical interest. First, $\boldsymbol{B}_j^{-1}$ does not appear any more in the updating procedure; this does not represent any saving in storage requirement (since $\boldsymbol{A}_j$ is used instead $\boldsymbol{B}_j^{-1}$) but the computational load is greatly reduced, as all products with matrix $\boldsymbol{H}$ have disappeared. Second, evaluation of $L_j(\bar{\boldsymbol{t}}_{j-1}, \bar{\boldsymbol{q}}_j)$ according to (61b) do not require the computation of $\overline{\boldsymbol{w}}_j$ or $\overline{\boldsymbol{A}}_j$, but only of $\overline{e}_j$ and $\ln(|\overline{\boldsymbol{B}}_j|)$. This only requires inversion of (K, K)-dimensional matrix $\boldsymbol{V}_j$, computation of its determinant, and products of matrices with dimensions (I, K) at most.

Starting from a current combination $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$, it is possible to compare the likelihood values of several sequences $(\bar{\boldsymbol{t}}_{j-1}, \bar{\boldsymbol{q}}_j)$ by repeated applications of (71)–(72), (66), (68) and (61b). Once a new combination $(\bar{\boldsymbol{t}}_{j-1}, \bar{\boldsymbol{q}}_j)$ has been selected as the one which maximizes L_j among the trials, then $\boldsymbol{w}_j$ and $\boldsymbol{A}_j$ are to be updated using (69) and (70) before a new iteration is performed. Finally, estimation of the reflectivity values is directly obtained at very low cost as a by-product of the detection step since (56) can be rewritten as

$$\hat{\boldsymbol{r}}_j = \boldsymbol{m}_j + \boldsymbol{M}_j\boldsymbol{w}_j. \tag{73}$$

In the above derivations, the special case of column $(\boldsymbol{Q}_1, \boldsymbol{R}_1)$ has been set aside for sake of clarity. In fact, the algorithm remains valid for $j = 1$, except that $\boldsymbol{q}_1$ must be substituted for $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$.

4) Suboptimal Iterative Maximization Procedure: As previously mentioned, the detection step is necessarily suboptimal because exhaustive exploration of the possible binary sequences is intractable. The aim of the present subsection is to propose a suboptimal iterative exploration scheme which fully takes advantage of the available updating equations. In essence, it consists in starting from an initial solution and jumping from the current solution $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$ to a better one among a restricted set of neighboring combinations $(\bar{\boldsymbol{t}}_{j-1}, \bar{\boldsymbol{q}}_j)$. The process is iterated until no neighbor is more likely than the current solution, which is then chosen as the final one.

Since the criterion value is guaranteed to increase at each iteration, and since the number of possible binary combinations is finite, the procedure necessarily converges in a finite—unknown—number of iterations. Of course, no guarantee of convergence towards the global optimum is available. From a practical point of view, the degree of suboptimality and the average number of iterations greatly depend on the choice of the initial point and of the neighborhood system between binary sequences. As empirically confirmed in Section V, the following neighborhood system gives very acceptable results:

Definition Two admissible realizations $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$ and $(\bar{\boldsymbol{t}}_{j-1}, \bar{\boldsymbol{q}}_j)$ $(j > 1)$ are neighbors if one of the two following conditions holds:

1) *$\bar{\boldsymbol{t}}_{j-1}$ and $\boldsymbol{t}_{j-1}$ differ from no more than one triplet of transition variables $(t^{/}_{ij-1}, t^{-}_{ij-1}, t^{\backslash}_{ij-1})$.*
2) *$\bar{\boldsymbol{t}}_{j-1}$ is equal to $\boldsymbol{t}_{j-1}$ and $\bar{\boldsymbol{q}}_j$ differ from $\boldsymbol{q}_j$ at one location only.*

Naturally, $\boldsymbol{q}_1$ and $\bar{\boldsymbol{q}}_1$ are neighbors if they differ at one location only. An *admissible realization* $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$ is such that $P(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$ is strictly positive, and that no crossing ($t^{\backslash}_{ij-1} = 1$ and $t^{/}_{i+1j-1} = 1$) is present, because they are physically meaningless.

The above definition is quite intuitive: given a current combination $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$, its neighbors are obtained by considering successively i) every possible arrangements of each identified boundary of column $j-1$, including cancellation; ii) every possible creation of a new boundary in column j. It is easy to see that each realization $\boldsymbol{q}_1$ holds exactly I neighbors, while each realization $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$ holds a number of neighbors between I and $8I$.

In the M-B-G I case, jumping from neighbor to neighbor gives values of K which are most often equal to one and sometimes to two or three, and substitution of M-B-G II for M-B-G I adds one unit to K, which then varies between two and four. Such low values gives full efficiency to the M-B-G algorithm and entire compatibility with practical implementation.

Finally, the initial combination is chosen as the *a priori* most probable one. In practical situations, μ^- is greater than $\mu^/$ and $\mu^\backslash$, ε is small compared to the other parameters, and the resulting value of λ is lower than $\frac{1}{2}$. Hence the most probable first column $\boldsymbol{q}_1$ is a uniformly zero Bernoulli sequence and the most probable combination $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$ holds only horizontal transitions from the previous column $\boldsymbol{q}_{j-1}$:

$$q_{ij-1} = 1 \Rightarrow t_{ij}^- = 1 \Rightarrow q_{ij} = 1.$$

Initialization of the algorithm with a uniformly zero Bernoulli sequence can be performed easily since we have in this case

$$\boldsymbol{m}_j = 0, \boldsymbol{M}_j = 0 \Rightarrow \boldsymbol{b}_j = 0, \boldsymbol{B}_j = \sigma_n^2 I_I \Rightarrow \begin{bmatrix} \boldsymbol{A}_j | \boldsymbol{w}_j \\ \boldsymbol{w}_j{}' | e_j \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}' \\ \boldsymbol{z}'_j \end{bmatrix} \frac{1}{\sigma_n^2} [\boldsymbol{H} | \boldsymbol{z}_j]. \quad (74)$$

To start from another combination $(\boldsymbol{t}_{j-1}, \boldsymbol{q}_j)$, the algorithm is first initialized with a zero Bernoulli sequence. Then, the boundaries are introduced one after the other through several iterations of (66) and (68)–(72).

When (29) is applied to the M-B-G I model, the above definition of neighboring sequences and the choice of initial solution are such that the exploration scheme exactly reduces to the classical B-G iterative procedure [4], [17]. Finally, not only the prior M-B-G model, but also the choice of MAP criterion and the way it is maximized provide a true 2-D extension of the 1-D B-G approach.

In order to sum up the whole procedure designed in Section IV, Table I depicts the sequential detection-estimation algorithm for multichannel M-B-G deconvolution.

TABLE I
SEQUENTIAL DETECTION-ESTIMATION ALGORITHM FOR M-G-B DECONVOLUTION. THE TABLE IS GIVEN FOR COLUMNS $\boldsymbol{x}_j$, $2 \leq j \leq J$. IF $\boldsymbol{x}_1$ IS NOT KNOWN, IT CAN BE ESTIMATED USING A B-G DECONVOLUTION (SEE SECTION IV-B AND [17])

I- NULL SEQUENCE

Compute e_j, $\mathbf{w}_j$, $\mathbf{A}_j$, (74)

set $\ln(|\mathbf{B}_j|) = I \ln(\sigma_n^2)$ and $I_j = I_j^\varepsilon = I_j^/ = I_j^- = I_j^\backslash = 0$.

II- INITIAL SEQUENCE $(\mathbf{t}_{j-1}, \mathbf{q}_j)$

Iterate on $\ln(|\mathbf{B}_j|)$, e_j, $\mathbf{w}_j$ and $\mathbf{A}_j$. (62, 71-72, 66, 68-70)

evaluate $(I_j, I_j^\varepsilon, I_j^/, I_j^-, I_j^\backslash)$ for $(\mathbf{t}_{j-1}, \mathbf{q}_j)$ and compute L_j. (61)

III- NEIGHBORHOOD EXPLORATION

For every neighboring $(\bar{\mathbf{t}}_{j-1}, \bar{\mathbf{q}}_j)$ of current sequence $(\mathbf{t}_{j-1}, \mathbf{q}_j)$,

compute $\ln(|\mathbf{B}_j|)$ and e_j, (62, 70-71, 66, 68)

evaluate $(\bar{I}_j, \bar{I}_j^\varepsilon, \bar{I}_j^/, \bar{I}_j^-, \bar{I}_j^\backslash)$ and $\bar{L}_j$. (61)

Go to stage V if all $\bar{L}_j$ are lower than L_j.

IV- UPDATING STAGE

Substitute the optimal $(\bar{\mathbf{t}}_{j-1}, \bar{\mathbf{q}}_j)$ for the current sequence,

update $\mathbf{w}_j$ and $\mathbf{A}_j$. (69-70)

Go to stage III.

V- FINAL STAGE

Deduce $\hat{\mathbf{r}}_j$ from $\mathbf{w}_j$. (73)

TABLE II
CPU TIMES ON A 22 MIPS WORKSTATION. THE CPU TIME OF M-B-G II DECONVOLUTION IS ABOUT TWICE THAT OF M-B-G I AND THREE TIMES THAT OF GAUSSIAN OR B-G DECONVOLUTION. ICM IS EASY TO IMPLEMENT BUT CONVERGENCE IS SLOW COMPARED TO EQUIVALENT GRADIENT DESCENT TECHNIQUES. NOTE THAT THE ELAPSED TIMES FOR FIGS. 10(d) AND (e) ARE GIVEN FOR ONE FORWARD SCAN ONLY. SINCE FIVE FORWARD AND BACKWARD SCANS HAVE BEEN PERFORMED, ACTUAL ELAPSED TIMES ARE TEN TIMES LARGER

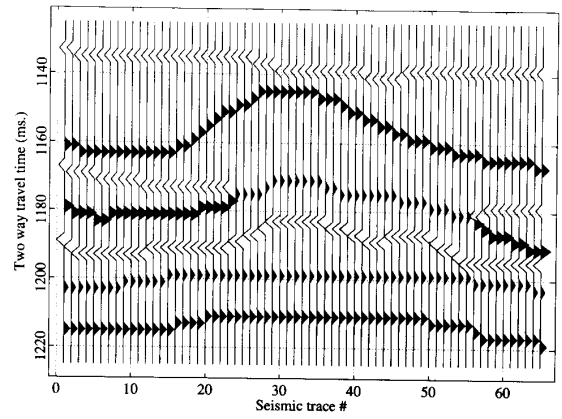
figure #	8a	8b	8c	8d		10c	10d	10e
method	G	ICM	B-G	M-B-G I	M-B-G II	B-G	M-B-G I	M-B-G II
CPU time (s.)	10.0	292.8	10.2	11.5	29.4	85.5	132.8	258.8

V. RESULTS

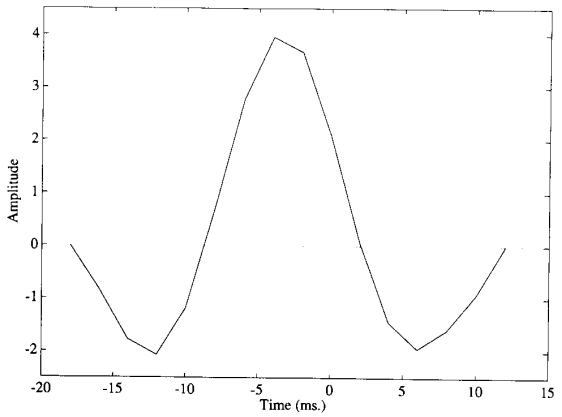
In this section, the performances of the proposed methods are illustrated by results obtained on simulated and real data. As we have restricted the scope of the study to simple deconvolution, the question of the choice of the *hyperparameters* (i.e., the various tuning parameters of the methods) is left unanswered, and some empirical rules must be devised so as to specify their values.

In Section V-A, these rules are outlined and used on simulated data. It should be emphasized that the hyperparameters were not optimized for each specific example. Therefore, the results are quite typical of the average capabilities of the deconvolution methods. Finally, Section V-B provides an example of real seismic data processing.

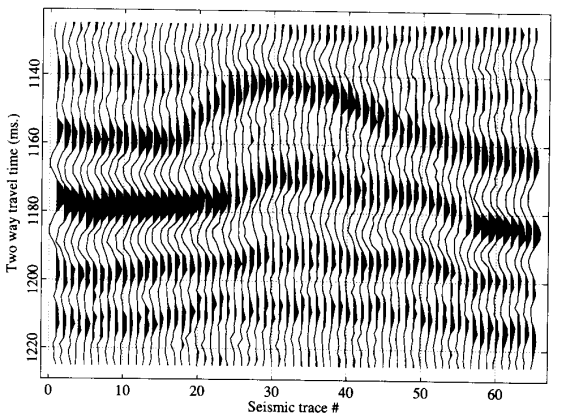
CPU times spent on computing some of the results are collected in Table II. Times have been measured on a standard 22 MIPS workstation. Table II shows that the computational complexity remains acceptable: the volume of computations of M-B-G II deconvolution is about twice that of M-B-G I deconvolution and three times that of Gaussian or B-G deconvolution.



(a)



(b)



(c)

Fig. 7. Simulated data: (a) Synthetic reflectivity used for the simulations. (b) Seismic wavelet (i.e., vertical blur function). (c) Observation sequences obtained by convolution between the reflectivity columns (a) and the wavelet (b), and addition of white noise (SNR = 10 dB).

A. Synthetic Data

1) First Example: Fig. 7(a) shows a synthetic reflectivity section, sampled on 65 columns of 51 points, with a sampling period of 2 ms. Conventionally, reflectivity pulses are represented as triangles of variable height, and positive areas of signals are shaded in black so as to enhance the figure. The seismic wavelet depicted in Fig. 7(b) is nonminimum phase and has a very poor spectral content, which makes deconvolution a very difficult problem. Fig. 7(c) shows the noise-corrupted convolution product between the reflectivity section and the wavelet, which corresponds to the observed data. The signal-to-noise ratio (SNR), defined as the ratio of the mean power of the noiseless observations to the additional white Gaussian noise variance, was set to 10 dB.

Deconvolution results are presented in Fig. 8. In order to compare our approach to standard methods, two techniques based upon Gaussian prior models of the reflectivity were implemented. The first one used a simple white Gaussian (G) model, which allowed us to estimate the reflectivity trace by trace using the following 1-D equations:

$$\hat{\boldsymbol{r}}_j = \underset{\boldsymbol{r}_j}{\operatorname{argmax}} \|\boldsymbol{z}_j - \boldsymbol{H}\boldsymbol{r}_j\|^2 + \beta\|\boldsymbol{r}_j\|^2 = (\boldsymbol{H}'H + \beta\boldsymbol{I}_I)^{-1}\boldsymbol{H}'\boldsymbol{z}_j. \tag{75}$$

The above estimator was implemented using the fast Kalman filter proposed in [23] (see also [2], [4]). The value of regularization parameter β was chosen so as to minimize the least-squares error $\|\hat{r} - r\|^2$ between actual and estimated solutions. The result is shown in Fig. 8(a). In spite of the optimal choice of β, the solution is strongly oversmoothed even though no horizontal nor vertical correlation are present in the model. In fact, this approach seems unable to provide any acceptable tradeoff between noise cancellation and high resolution because of the very poor spectral content of the wavelet.

The second Gaussian model was truly two-dimensional and introduced lateral correlations. More specifically, Gaussian interactions were introduced in the log-impedance section, which allowed us to determine the corresponding estimate of the reflectivity through minimization of

$$\sum_{j=1}^{J} \|\boldsymbol{z}_j - \boldsymbol{H}r_j\|^2 + \beta^{/}\sum_{i,j}(y_{ij} - y_{i-1j+1})^2 + \beta^{-}\sum_{i,j}(y_{ij} - y_{ij+1})^2 + \beta^{\backslash}\sum_{i,j}(y_{ij} - y_{i+1j+1})^2 \tag{76}$$

where $\boldsymbol{r}_j$ fulfills (31). Minimization was carried out using an ICM algorithm [24], and here again, we relied on the least squares estimation error $\|\hat{\boldsymbol{r}} - \boldsymbol{r}\|^2$ to select the best values of parameters β. The result is depicted in Fig. 8(b). Compared to 1-D Gaussian deconvolution, the least squares estimation error $\|\hat{\boldsymbol{r}} - \boldsymbol{r}\|^2$ is about 10 % lower. Visually, the improvement is hardly noticeable, and the tradeoff between noise cancellation and high resolution remains far from acceptable. We now restrict our attention to B-G and M-B-G deconvolution.

Fig. 8(c) shows the result obtained with B-G deconvolution in the case of a marginal likelihood estimator [17]. The parameter values were chosen empirically as follows:

$$\begin{cases} \text{SNR} = 10 \text{ dB} \\ \lambda = 0.12 \end{cases} \tag{77}$$

The parameter values (77) correspond to the known "true" ones. According to empirical results reported in [25], such

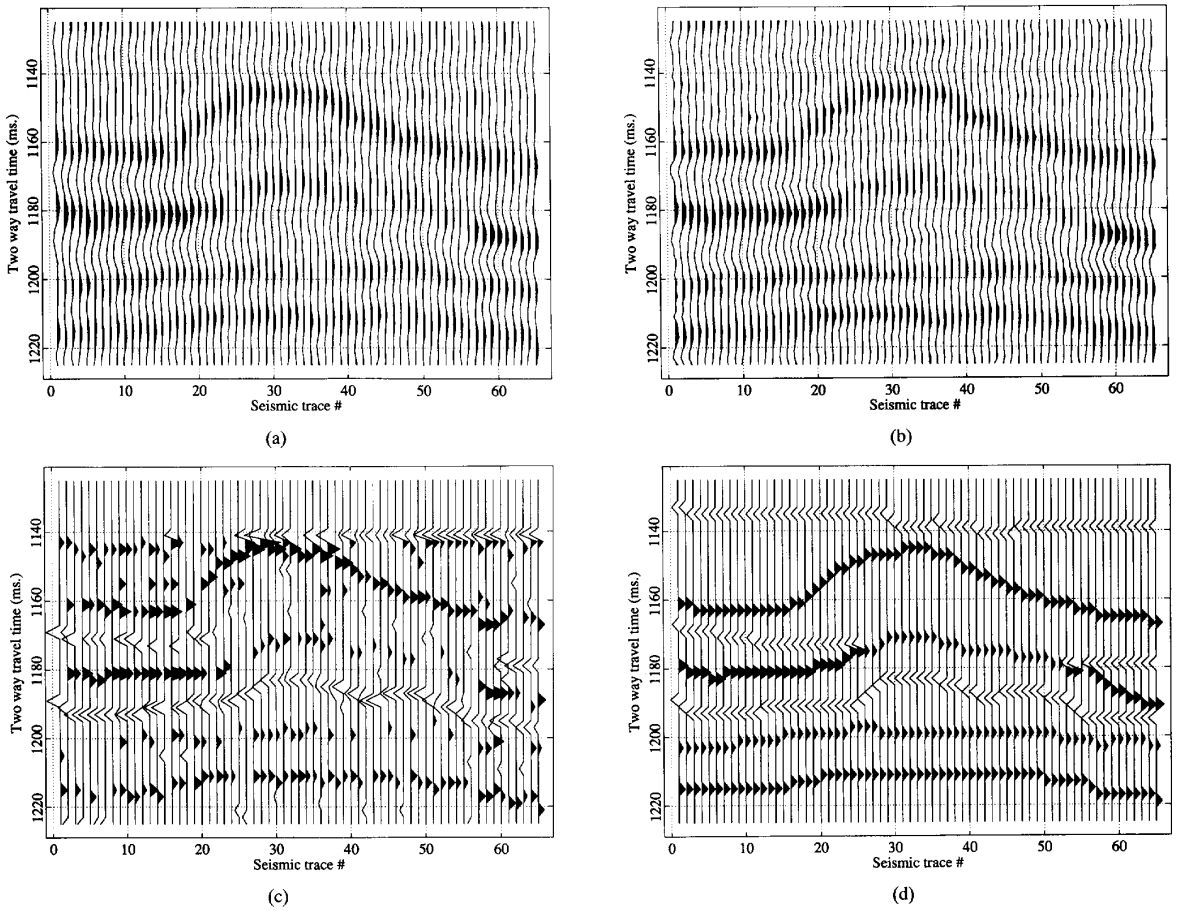


Fig. 8. Deconvolution results: (a) 1-D linear deconvolution using a white Gaussian prior and a fast Kalman filtering technique [23] provides a strongly oversmoothed result. (b) 2-D linear deconvolution using a Gauss–Markov random field and ICM restoration [24] still lacks resolution. (c) B-G deconvolution misses the horizontally stratified organization of the medium. (d) M-B-G I deconvolution provides a nearly perfect result (M-B-G II performs very similarly to M-B-G I).

a choice is nearly optimal with respect to B-G capabilities. Visual comparison between Fig. 8(c) and Fig. 7(a) shows that the original reflectivity section is very roughly restored. Many details are lost: thin layers or weak reflectors are hardly detected; some columns are strongly distorted by artifacts such as false alarms or doubled detection. More generally horizontal continuity is not preserved, which makes the stratified structure of the medium quite difficult to perceive from Fig. 8(c).

Fig. 8(d) depicts the M-B-G I estimate, obtained for the following values of parameters:

$$\begin{cases} \text{SNR} = 7 \text{ dB} \\ \mu^{-} = 0.084, \mu^{/} = \mu^{\backslash} = 0.020, \varepsilon = 0 \Rightarrow \lambda = 0.12 \\ a = 0.999 \end{cases} \quad (78)$$

Contrarily to B-G deconvolution, M-B-G I deconvolution requires an under-evaluation of the SNR (say, 2 to 5 dB less than the "true" value), in order to avoid a large number of errors during the detection step. Underevaluation of the SNR corresponds to overregularization of the solution, which means that the prior model is given more confidence. Because the sequential structure of the multichannel algorithm does not correspond to optimal processing of the data (see Section III-D), overregularization partly makes up for this limitation. To a lesser extent, the same is true of M-B-G II deconvolution. Here we have not represented M-B-G II results, since they can hardly be distinguished from M-B-G I (Fig. 8(d)), for the same parameter values (78).

Compared to Fig. 8(c), the accuracy of the estimates is greatly improved by M-B-G deconvolution. Differences between Fig. 8(d) and the original reflectivity section (Fig. 7(a)) are actually negligible. Even small details, such as the double junctions in columns 24 and 55, are restored properly. This excellent result illustrates the improvement that can be obtained from accounting stratification in the field of seismic deconvolution.

In these tests, it was assumed that the first reflectivity column is known, so that M-B-G deconvolution can be correctly initialized. In practice, this favorable situation occurs

when well log information can be used for initialization. If no borehole data are available, restoration of the first column can be obtained from B-G deconvolution, as described in Section IV. Since in this case the result is not reliable, it is advisable to process the data several times to and fro, and to initialize each new scan with the last estimated column given by the previous scan. Such a procedure is used for real data processing in Section V-B.

2) Second Example: The previous example was well suited to accounting for horizontal boundary continuity, since no fault nor other types of discontinuity were present in the synthetic seismic section; this allowed us to set the discontinuity parameter ε to zero. Fig. 9(a) is similar to Fig. 7(a) except for an inclusion in the upper left corner of the reflectivity section.

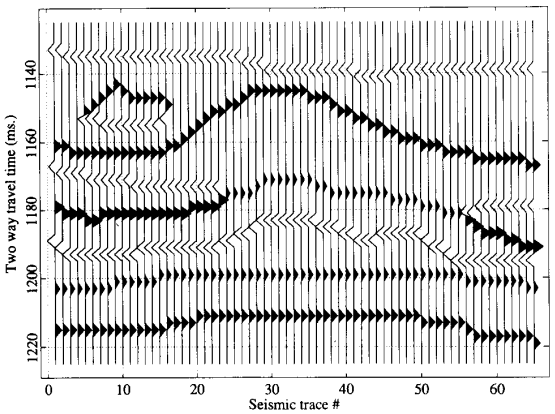
Figs. 9(b) and (c) show the results obtained with M-B-G deconvolution, respectively, M-B-G I and II. In both cases, the parameter values were chosen as follows:

$$\begin{cases} \text{SNR} = 7 \text{ dB} \\ \mu^{-} = 0.084, \mu^{/} = \mu^{\backslash} = \varepsilon = 0.020 \Rightarrow \lambda = 0.138 \\ a = 0.999 \end{cases} \quad (79)$$

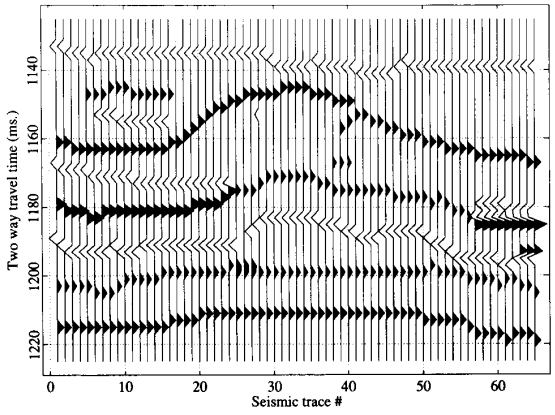
The only difference between (78) and (79) is the nonzero value of ε which authorizes the presence of discontinuities.

Both versions provide a satisfactory estimate; in particular, the inclusion is well detected. However, the quality of the results is significantly lower than in the first example. This is not surprising since rigorous horizontal continuity can no longer be assumed; this tends to increase the ill-posed character of the problem because a lot of discontinuous possibilities now compete for the estimated solution.

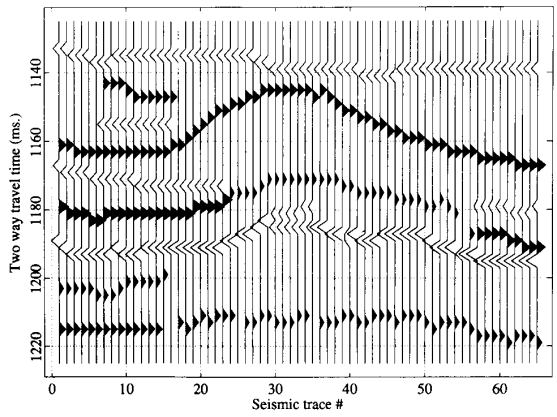
Further analysis of Figs. 9(b) and (c) shows that M-B-G I and II do not behave the same way with respect to detection. Because it is based on a more pertinent prior model, M-B-G II is more robust toward detection errors (boundaries are more accurately tracked in Fig. 9(c) than in Fig. 9(b), to the detriment of weak reflectors (for instance the boundary located at 1200 ms). M-B-G I is more sensitive, but additional tests have shown that it often lacks robustness: a single error can jeopardize the whole estimation by being replicated and amplified from one column to the next. Such a cumulative effect of detection errors is the main drawback of sequential *decision-directed* procedures. In this respect, substitution of M-B-G II for M-B-G I provides a substantial improvement. Further improvement could be expected from introducing a detection delay from column to column, as Goussard and Demoment did for recursive B-G deconvolution [20].



(a)



(b)



(c)

Fig. 9. Second synthetic example: (a) Synthetic reflectivity, identical to Fig. 7(a) except for an inclusion in the upper left corner of the section. (b) and (c), respectively, show M-B-G I and II estimates.

B. Actual Seismic Data

The last example comes from an actual seismic experiment. Fig. 10(a) shows a stacked seismic section of 41 traces. The time sampling interval is equal to 2 ms.

Application of multichannel deconvolution requires that the seismic wavelet is known. In order to estimate the unknown wavelet, a preliminary step of blind single-channel deconvolution was carried out. It is based on a generalized maximum-likelihood approach with B-G prior modeling and MA description of the wavelet, as described in [25] (an equivalent approach available for ARMA parameterization can be found in [5]). The identified nonminimum phase wavelet is shown in Fig. 10(b).

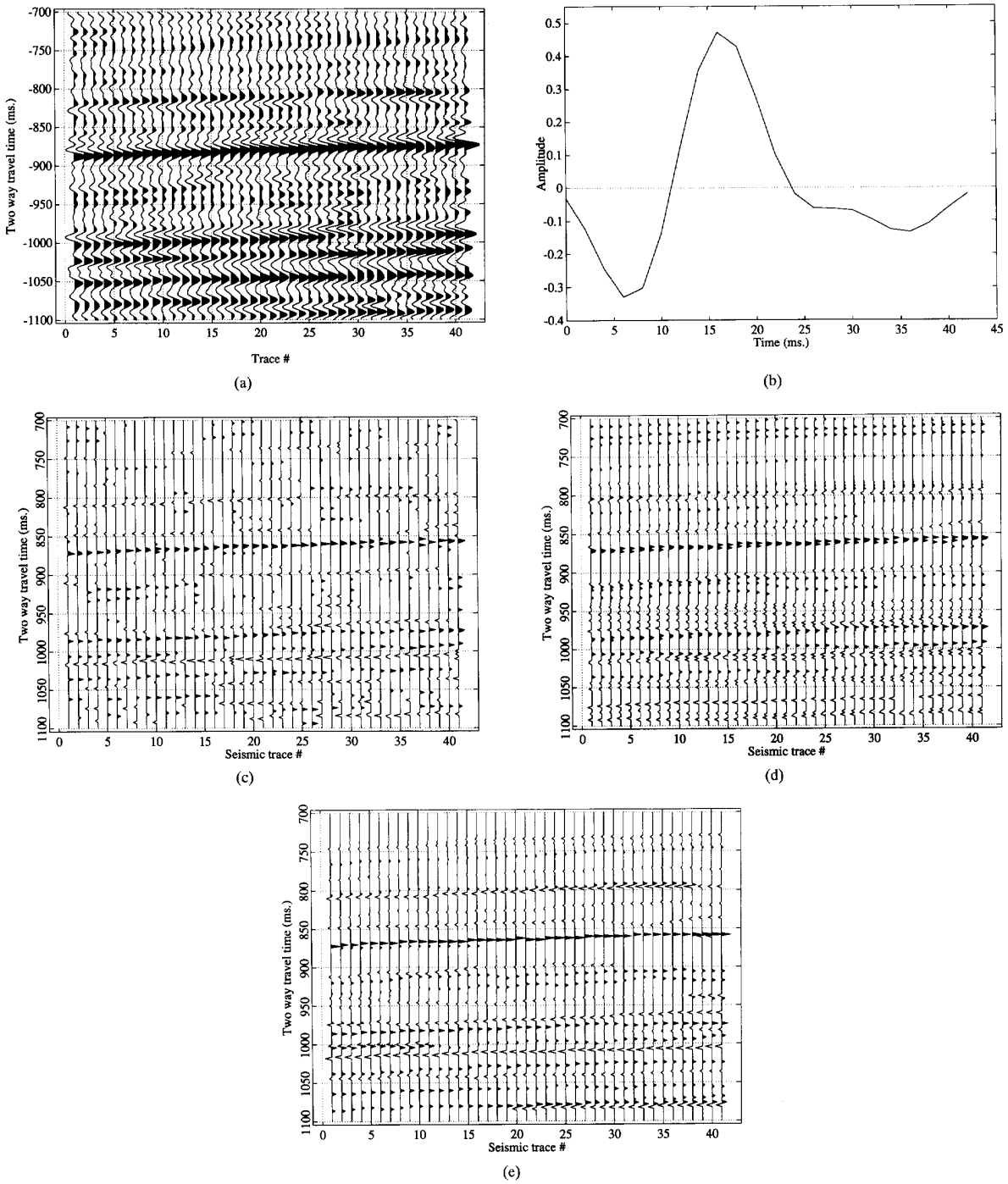


Fig. 10. Actual seismic example: (a) shows a stacked seismic section. (b) Identified nonminimum phase wavelet (i.e., vertical blur function) using a generalized maximum-likelihood approach [25]. (c), (d), and (e), respectively, show B-G, M-B-G I, and II estimates.

Fig. 10(c) shows the result obtained with B-G deconvolution for the following parameter values:

$$\begin{cases} \text{SNR} = 4 \text{ dB} \\ \lambda = 0.2 \end{cases} \tag{80}$$

Except for the strongest reflectors, the difference between successive reflectivity sequences does not allow any stratigraphic interpretation. Nevertheless, a lot of reflectors reappear from trace to trace, which indicates that the investigated medium is horizontally stratified.

For M-B-G deconvolution, parameter values were empirically chosen as follows:

$$\begin{cases} \mathrm{SNR} = 4\ \mathrm{dB} \\ \mu^{-} = 0.105, \mu^{/} = \mu^{\backslash} = 0.050, \varepsilon = 0.008 \Rightarrow \lambda = 0.2 \\ a = 0.99 \end{cases} \tag{81}$$

In order to avoid initialization instabilities, the data were scanned five times. Figs. 10(d) and (e), respectively, show the deconvolution results obtained then with M-B-G I and II. As expected, both results preserve horizontal continuity and parsimoniously locate boundaries. In this respect, M-B-G II performs better than M-B-G I: for instance, the latter splits the strong positive reflector at about 860 ms, whereas the former locates a single spike. Another unrealistic aspect of Fig. 10(d) is the presence of meaningless discontinuities, whereas M-B-G II better preserves horizontal continuity.

VI. Conclusion

In this paper, we addressed multichannel deconvolution of echosounding signals in order to estimate two-dimensional stratified structures. The proposed approach constitutes a true 2-D extension of Bernoulli–Gaussian deconvolution [7]. Since the latter was designed in a 1-D context only, prior information was reduced to the spikiness of reflectivity signals in the vertical direction [3]–[5]. In two dimensions the medium is assumed to be stratified into roughly horizontal layers, so that reflectors generally form smooth lines with a preferential orientation in the horizontal direction. Single-channel deconvolution can hardly account for such essential information and, to our knowledge, existing multichannel methods strongly rely on empirical treatments [6], [7]. Instead, we proposed a Bayesian multichannel deconvolution method based on MBRF prior modeling [1]. Two versions were presented, called M-B-G I and II. The first one provides an exact 2-D extension of B-G deconvolution, while the second accounts for impedance representation, which leads to a more consistent prior model and to better deconvolution results.

Special attention was paid to computational complexity: the algorithmic structure of M-B-G multichannel deconvolution is very similar to sequential application of B-G single-channel deconvolution, but significantly better performances were obtained on both simulated and real seismic data. This is a consequence of the 2-D nature of the prior model.

Perspectives of improvement mainly concern the introduction of a detection delay, in order to make up for the inherent lack of robustness of sequential decision. Another issue is to cope with blind and adaptive multichannel deconvolution.

Acknowledgment

The authors wish to thank Total Exploration, TRITON France, and Compagnie Générale de Géophysique for providing the real seismic data presented in the paper.

References

[1] J. Idier and Y. Goussard, "Markov modeling for Bayesian restoration of 2-D layered structures," LSS Internal Report GPI-89/06, 1990; also, *IEEE Trans. Inform. Theory*, , vol. 39, no. 4, 1993..

[2] J. M. Mendel, J. Kormylo, F. Aminzadeh, J. S. Lee, and F. Habibi-Ashrafi, "A novel approach to seismic signal processing and modeling," *Geophysics*, , vol. 46, pp. 1398–1414, 1981.

[3] J. Kormylo and J. M. Mendel, "Maximum-likelihood detection and estimation of Bernoulli–Gaussian processes," *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 482–488, 1982.

[4] J. M. Mendel, *Optimal Seismic Deconvolution*. New York: Academic, 1983.

[5] J. Goutsias and J. M. Mendel, "Maximum-likelihood deconvolution: An optimization theory perspective," *Geophysics*, , vol. 51, pp. 1206–1220, 1986.

[6] C. Y. Chi and J. M. Mendel, "Multichannel maximum-likelihood deconvolution," in*Proc. 55th Ann. Meet. Soc. Exploration Geophysicists*, Washington, DC; also Internal Report, University of Southern California, 1985.

[7] J. Brac, P. Y. Déquirez, F. Hervé, C. Jacques, P. Lailly, V. Richard, and D. Tran Van Nhieu, "Inversion with *a priori* information: An approach to integrated stratigraphic interpretation," in *Proc. 58th Ann. Meet. Soc. Exploration Geophysicists*, Anaheim, 1988.

[8] H. Derin and P. A. Kelly, "Discrete-index Markov-type random processes," *Proc. IEEE*, vol. 77, pp. 1485–1510, 1989.

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

[10] A. Blake and A. Zisserman, *Visual Reconstruction*. Cambridge, MA: MIT Press, 1987.

[11] D. K. Pickard, "Unilateral Markov fields," *Adv. Appl. Prob.*, vol. 12, pp. 655–671, 1980.

[12] J. Goutsias, "Mutually compatible Gibbs random fields," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1233–1249, 1989.

[13] L. R. Rabiner and B. H Juang, "An introduction to hidden Markov models," *IEEE ASSP Mag.*, pp. 4–16, Jan. 1986.

[14] A. T. Walden and J. W. J. Hosken, "The nature of the non-Gaussianity of primary reflection coefficients and its significance for deconvolution," *Geophys. Prospecting*, vol. 34, pp. 1038–1066, 1986.

[15] R. Godfrey, F. Muir, and F. Rocca, "Modeling seismic impedance with Markov chains," *Geophysics*, vol. 45, pp. 1351–1372, 1980.

[16] J. Goutsias and J. M. Mendel, "Optimal simultaneous detection and estimation of filtered discrete semi-Markov chains," *IEEE Trans. Inform. Theory*, vol. 34, pp. 551–568, 1988.

[17] Y. Goussard, G. Demoment, and J. Idier, "A new algorithm for iterative deconvolution of sparse spike trains," in *Proc. Int. Conf. ASSP*, Albuquerque, 1990, pp. 1547–1550.

[18] G. D. Forney, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–278, 1973.

[19] A. K. Mahalanabis, S. Prasad, and K. P. Mohandas, "Recursive decision-directed estimation of reflection coefficients for seismic data deconvolution," *Automatica*, vol. 18, pp. 721–726, 1982.

[20] Y. Goussard and G. Demoment, "Recursive deconvolution of Bernoulli–Gaussian processes using a MA representation," *IEEE Trans. Geosci. Remote Sensing*, vol. 27, pp. 384–394, 1989.

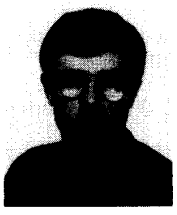
[21] J. Idier and Y. Goussard, "Stack algorithm for recursive deconvolution of Bernoulli–Gaussian processes," *IEEE Trans. Geosci. Remote Sensing*, vol. 28, pp. 975–978, 1990.

[22] P. Lascaux and R. Théodor, *Analyse numérique appliquée à l'art de l'ingénieur*. Paris: Masson, 1986.

[23] G. Demoment, R. Reynaud, and A. Herment, "Range resolution improvement by a fast deconvolution method," *Ultrasonic Imaging*, vol. 6, pp. 435–451, 1984.

[24] J. Besag, "On the statistical analysis of dirty pictures," *J. Roy. Stat. Soc. B*, vol. 48, pp. 259–302, 1986.

[25] Y. Goussard, *Déconvolution de processus aléatoires non-gaussiens par maximisation de vraisemblances*, Thèse de Doctorat, Université de Paris-Sud, Centre d'Orsay, 1989.



Jérôme Idier was born in France in 1966. He received the diploma degree in electrical engineering from the École Supérieure d'Électricité in 1988 and the Ph.D. degree in physics from the Université de Paris-Sud, Orsay, in 1991.

He is now with the Centre National de la Recherche Scientifique, assigned to the Laboratoire des Signaux et Systèmes. His major area of interest is inverse problems in signal and image processing, and more specifically, modeling problems for imaging purposes.

Yves Goussard (M'89) was born in Paris, France, in 1957. He graduated from the École Nationale Supérieure de Techniques Avancées in 1980 and received the Docteur-Ingénieur and Ph.D. degrees from the Université de Paris-Sud, Orsay, France, in 1983 and 1989, respectively.

From 1983 to 1985, he was a Visiting Scholar at the Department of Electrical Engineering and Computer Science at the University of California, Berkeley. In 1985, he was appointed a Chargé de Recherche at CNRS, Gif-sur-Yvette, France, and in 1992, he joined the Biomedical Engineering Institute of the École Polytechnique, Montreal, Canada, where he is now an Associate Professor. During the academic year 1990–1991, he was on sabbatical leave at the Department of Electrical Engineering-Systems, University of Southern California, Los Angeles. After some work on nonlinear systems identification and modeling, his interests moved toward ill-posed problems in signal and image processing with application to biological systems.

Dr. Goussard is a member of RS^TLYC and LSD^2.