A Differential Semblance Criterion for Inversion of Multioffset Seismic Reflection Data

William W. Symes

May, 1992

TR92-18

A Differential Semblance Criterion for Inversion of Multioffset Seismic Reflection Data

W.W. Symes

Abstract

Mean-square error leading to least-squares inversion of multioffset reflection seismograms is insensitive to velocity trend information except in the immediate vicinity of a kinematically correct model. In contrast, *differential semblance* retains sensitivity to velocity trend changes over a wide range of models. The differential semblance criterion combines mean-square error with the mean-square differences of inverted models from datasets at neighboring shot positions (or offsets, or slownesses, ...). Differential semblance compares model estimates at nearby acquisition parameters which are similar even when the model velocity trends are incorrect. Because the method inverts the data, so that the estimated model amplitudes are meaningful, simple differences between the (unstacked) model estimates give a reliable measure of velocity error. A mathematical investigation indicates that the differential semblance criterion is smooth and convex over a large range of velocity models. Numerical simulation using synthetic data sets verifies this contention.

Introduction

Least-squares (or more generally, least-error or maximum-likelihood) inversion has been advocated over the last decade as a general approach to the inversion of reflection seismograms, for several reasons. It is capable of reflecting directly almost any physics of seismic wave propagation; it can be modified to incorporate nonseismic constraints; and it possesses an elegant statistical justification (Tarantola [1987]).

Unfortunately, least squares inversion using gradient (quasi-Newton) optimization generally fails to yield large-scale velocity trends, as was first noticed in numerical experiments (see Kolb *et al.* [1986], Gauthier *et al.* [1986] and Tarantola *et al.* [1990]). We review the reasons for this failure below; for lengthier explanation, see Santosa and Symes [1989]. In fact the mean-square fit error, which serves as the objective function or criterion in least-squares inversion, is nonconvex, and iterative local optimization methods tend to stagnate far from useful velocity estimates. Accordingly many recent papers on least-error inversion advocate the use of global (non-gradient) optimization methods such as Monte-Carlo search, simulated annealing or genetic algorithms (e.g. Sen and Stoffa [1991], Scales *et al.* [1991], Mosegard and Tarantola [1991]).

This paper presents a smooth convex variant of the mean square error criterion. The principal novelty is the introduction of an enlarged space of infeasible (incoherent) models. A differential measure of event semblance (or coherency) calibrates infeasibility in a natural way, and gives the inversion method based on the new criterion its name: differential semblance optimization (DSO).

DSO actually encompasses a number of algorithms. In previous papers (Symes [1990], [1991a, b]; Symes and Carazzone [1989], [1991]) we have explored algorithms in this class appropriate for plane-wave data and layered media. We have given both theoretical and numerical evidence that the plane-wave differential semblance method yields efficient and accurate velocity and reflectivity estimates. The numerical evidence includes successful treatment of field data sets.

The present paper describes a version of the DSO criterion suitable for application to common-source reflection seismograms using a general 2D (two-dimensional nonlayered) acoustic model. By mathematical argument and numerical example, we establish the smoothness and convexity properties of

the differential semblance criterion. We do not explicitly address the design of an inversion algorithm based on the criterion in this paper. We begin by describing the acoustic model in the form needed for the rest of the discussion. Then we review the sensitivity analysis of the seismogram, and its implications for least-squares inversion. We define the differential semblance criterion, and explain its analytical and numerical properties. Finally, we illustrate these properties using an example drawn from the Institut Français du Pétrole Marmousi model (Versteeg and Grau [1991]).

The Acoustic Model

Throughout we denote by x the two- or three-dimensional spatial variable, i.e. $x = (x_1, x_2)$ or (x_1, x_2, x_3), and by z the depth, i.e. the last spatial coordinate, so $z = x_2$ (2D) or $z = x_3$ (3D). The constant-density linear acoustic model connects the (spatially varying) sound velocity $v(x)$, the pressure field $p(x,t)$, and a body force divergence ("source") $f(x,t)$ through the acoustic wave equation

$$\frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = f \ . \tag{1}$$

We assume that f is transient and known and p is causal: $p \equiv 0$ for $t << 0$. We also assume (for simplicity) that the fluid occupies the half-space $\{z > 0\}$, and that the surface $\{z = 0\}$ is free, i.e., $p \equiv 0$ there.

It is very important in what follows that f depends on another parameter x_s, so we write: $f(x,t,x_s)$. Thus p depends on x_s as well. For common-source seismic data x_s is naturally the source position; however x_s could also be construed as offset for application to common-offset data, or as slowness (ray parameter, incidence angle) for application to plane wave data.

The seismogram is the sampling of the pressure field at a set $\{x_r\}$ of receiver points in the plane $\{z = z_r\}$. The seismogram depends on the velocity distribution v. We base this discussion on the perturbational approximation to the seismogram, in which v is split as $v = v_b + v_r$, with v_b a smooth background velocity and v_r a rough or oscillatory perturbation. Using regular first-order perturbation theory we write $S = S_b + S_r$ where S_b is the background seismogram with $v = v_b$ and S_r is the perturbation due to v_r. If v_b is sufficiently smooth, which we assume, then S_b consists of the direct wave, plus possibly refractions. We limit our attention to reflections here; so we

assume that S_b has been subtracted or windowed ("muted") out (a nontrivial step in practice!) and identify S with S_r.

S is thus the sampling at the receiver locations of the pressure field perturbation δp, which solves

$$\frac{1}{v_b^2}\frac{\partial^2 \delta p}{\partial t^2} - \nabla^2 \delta p = \frac{2v_r}{v_b}\nabla^2 p \tag{2}$$

plus appropriate side conditions. Note the appearance of the *reflectivity* $r = v_r/v_b$; in the sequel we shall use r rather than v_r to represent the rough part of the model.

As shown in Cohen and Bleistein [1977], Beylkin [1985], Rakesh [1988], Percell [1989], S can be approximated as an oscillatory integral of the form

$$\begin{aligned} & S[v_b]r(x_s; x_r, t) \\ & \cong \int dt' \int dx f(x, t-t', x_s) \\ & \quad \int d\xi A[v_b](x, x_r, t', \xi) e^{i\phi[v_b](x,x_r,t',\xi)} \cdot \hat{r}(\xi) \ . \end{aligned} \tag{3}$$

The notation is chosen to emphasize the following points:

(1) The seismogram S depends *linearly* on the reflectivity r, and *nonlinearly* on the background velocity v_b.

(2) The seismogram is a function of the source parameter x_s, the receiver coordinate x_r, and time t;

(3) The *amplitude* or A and the *phase* ϕ also depend on a wave vector ξ of the same dimensionality as the space coordinates.

(4) A and ϕ depend further on a source coordinate x and are convolved in x and t against the source distribution $f(x, t, x_s)$.

(5) The seismogram S, the amplitude A, and the phase function ϕ depend on v_b. The phase function ϕ is related to the travel time of certain reflected rays.

For space dimension n, the symbol A behaves for large $|\xi|$ like

$$A_0(x_s, x_r, t, \xi/|\xi|)|\xi|^{\frac{n-1}{2}}$$

for a smooth function A_0 (the "principal symbol"). A_0 is non-zero over a sector in $\xi/|\xi|$ determined by the ray geometry (hence by v_b). The phase function of ϕ is positively homogeneous of degree 1 in ξ.

Sensitivity of the Seismogram and Properties of the Mean-Square Error

With the conventions established so far, we can state a simple version of the least-squares inversion problem:

> *Find* v_b, r *to minimize the* misfit function J_{LS} *defined by*
>
> $$J_{LS}\left[v_b, r; S_{\text{data}}\right] = \frac{1}{2}\int\int\int dx_s dx_r dt \left|S[v_b]\cdot r - S_{\text{data}}\right|^2$$

Here we understand the integral sign to mean integral or sum, as appropriate. To avoid writing an excessive number of integral signs in the sequel, we introduce the standard notations for the L^2 inner product and norm:

$$\begin{aligned}\langle\psi,\phi\rangle &= \int \bar{\psi}\phi \\ \|\psi\| &= \langle\psi,\psi\rangle^{\frac{1}{2}}\end{aligned}$$

where the integration variables are understood from context, and replaced by scaled sums in case the functions ψ, ϕ are discretely sampled in one or more variables.

In this notation,

$$J_{LS}[v_b, r; S_{\text{data}}] = \frac{1}{2}\|S[v_b]r - S_{\text{data}}\|^2 \,. \tag{4}$$

In a typical reflection seismic model in 2D, v_b might be represented by a few tens to a few hundreds of parameters, while r requires perhaps $10^5 - 10^6$ parameters for a useful degree of resolution. Thus the least-squares problem is computationally very large, and efficient minimization algorithms are required. By far the most efficient numerical optimization techniques are the descent methods related to Newton's method — when they work.

These iterations take steps predicted by the linearized model/data relation so rely for their effectiveness on a close relation between the misfit function and its quadratic approximation. Accordingly, we now examine (somewhat formally) the response of S to perturbations in v_b and r.

From the oscillatory integral expression (3) the perturbation of S due to a change δv_b in v_b is

$$\delta_{v_b} S \cdot r = \int d\xi (i\delta\phi \cdot A + \delta A) e^{i\phi} \hat{r} \ . \tag{5}$$

This is an oscillatory integral of the same form as (3), with a different symbol. In fact $\delta\phi$ is also homogeneous of order 1, exactly as is ϕ. Therefore the symbol (amplitude) in the above integral grows as $|\xi|^{\frac{n+1}{2}}$ as $|\xi| \to \infty$, i.e. at a more rapid rate than A. It follows that for at least some oscillatory r, smooth δv_b

$$\|\delta_{v_b} S\| >> \|S\| \ .$$

On the other hand, since S is linear in r, one might expect that $\delta_r S$ and S are comparable: perturbation with respect to r does not change the phase ϕ, hence does not magnify high-frequency components.

The following example illustrates this disparity in sensitivity to the two types of model perturbation. We consider a layered model, i.e. $v_b = v_b(z)$, $r = r(z)$, z = depth coordinate, see Figure 1. We display in Figure 2(a) the unperturbed seismogram; in Figure 2(b) the perturbation $\delta_{v_b} S$ for this example, with $\delta v_b = 0.05 * (v_b - 1500 m/s)$; and in Figure 2(c) the perturbation $\delta_r S$ with $\delta_r = .05 * r$, plotted on the same scale. In both cases the source is $f(x,t) = F(t)\delta(x - x_s)$, with $F(t)$ a Ricker wavelet with peak frequency 15 Hz. The source and receivers are punctual (no arrays); the source depth is 8m, receiver depth is 12m, near offset 100m, far offset 1800m, and receiver interval is 50m. Trace length is 2s. The simulation method was (2,4) finite differences (see e.g. Dablain [1986]). Grid spacing in both spatial directions was 16m, and the time step was 4ms. The top surface of the model is free ($p \equiv 0$).

Taking this reasoning one step further, one sees immediately that

$$\|\delta^2_{v_b} S\| >> \|\delta_{v_b} S\| \ ,$$

that is, that S is very nonlinear in v_b.

Since

$$\begin{aligned}
\delta_{(v_b,r)} J_{LS} &= \langle \delta_{(v_b,r)}[S[v_b]\cdot r], S[v_b]r - S_{\text{data}} \rangle \\
&= \langle \delta_{v_b} S[v_b]r + S[v_b]\delta r, S[v_b]r - S_{\text{data}} \rangle \\
\delta^2_{(v_b,r)} J_{LS} &= \|\delta_{v_b} S[v_b]r + S[v_b]\delta r\|^2 \\
&\quad + \langle \delta^2_{v_b} S[v_b]r + 2\delta_{v_b} S[v_b]\delta r, S[v_b]r - S_{\text{data}} \rangle
\end{aligned}$$

one might well expect that

$$\left|\delta^2_{v_b} J_{LS}\right| >> \left|\delta^2_r J_{LS}\right|$$

and this is indeed the case for some oscillatory $r, \delta r$ of the same magnitude and smooth $v_b, \delta v_b$. Thus the Hessian is extremely ill-conditioned.

Moreover, it also follows that the growth rate of J_{LS} as one moves v_b away from the minimizer must be many times the overall size of J_{LS} itself. Therefore the growth cannot be sustained over a large change in v_b, and J_{LS} *saturates.* Consequently J_{LS} tends to be very non-convex, with a very small region of convexity near the global optimum model. See Gauthier *et al.* [1986], Figure 1 and Symes and Carazzone [1989], Figure 4 for actual pictures of J_{LS} illustrating these features.

The Differential Semblance Criterion

In this paragraph we will pose the least-squares inverse problem in a larger space of *infeasible models*, with a feasibility constraint incorporated in a "soft" form *via* a penalty functional. Implicit reparameterization of the model *via* partial optimization then leads to a nearly quadratic, convex objective in the space of background velocities, hence to a well-behaved inversion. We defer the scientific justification of this approach to the last paragraph.

We begin with two observations:

(i) For *fixed* v_b, J_{LS} is perfectly convex in r — in fact, quadratic!

(ii) If the set of shot parameter values $\{x_s\}$ reduces to a *singleton*, e.g. only one point source record is used, the minimum value of J_{LS} is essentially independent of v_b.

The first observation is obvious, as we have approximated the seismogram by its linearization in r. However, it is also likely very nearly correct even without the linearization, as r consists of high-frequency components and the seismogram is known to be nearly linear in these. See e.g. Tarantola [1986].

To verify the second observation, we present an example based on the model of Figure 1. We have carried out single gather inversion for r at constant (1500m/s) background velocity. The reflectivity model presented in Figure 3 achieves a data misfit of 6.9% rms. The predicted seismogram and its misfit are displayed in Figure 4, both plotted on the same scale as the data seismogram (Figure 2(a)). In contrast inversion at the correct background velocity (Figure 1) yields the reflectivity displayed in Figure 5, with approximately the same level of rms misfit (6%). The inversions were performed using the conjugate residual algorithm, as described in the "Numerical Exploration" section below.

From these observations we conclude that the inversion of a single shot record is feasible, and constrains *only* r, *not* v_b. Since this task is practical, it suggests the expedient of viewing r as a function of the shot parameter x_s

$$r = r(x, x_s) \; .$$

Of course, if S_{data} is noise free,

$$S_{\text{data}} = S[v_b^*]r^* \tag{6}$$

then $r(x, x_s) \equiv r^*(x)$ is amongst the minimizers of

$$\left\| S[v_b^*]r(\cdot, x_s) - S_{\text{data}}(\cdot, x_s) \right\|^2$$

and has the addition property of *coherence*, or independence of x_s, which we can express as

$$\frac{\partial r}{\partial x_s} \equiv 0 \; . \tag{7}$$

We call the measure of coherence embodied in this equation *differential semblance*, as it measures differentially the semblance of reflectivity estimates from neighboring shots. Only coherent reflectivity estimates have any ultimate meaning, since there is only one earth!

This notion of coherence, or shot-independence of model estimates, is actually the basic principle of velocity analysis. It is illustrated indirectly in the

example of Figures 3–5: as the exact model is layered, shot gathers simulated over it are identical to Figure 2, and independent of x_s. The reflectivity estimates from *any* layered velocity model at various shot locations are therefore horizontal translates of each other. Two neighboring reflectivity estimates are similar, in this setting, exactly when both are *flat*. Apart from numerical and edge diffraction artifacts, the reflectivity estimate from the exact velocity profile (Figure 5) is quite flat, whereas the estimate from the constant (wrong) profile (Figure 3) is quite non-flat. Thus comparison between reflectivity estimates at neighboring shot locations distinguishes correct from incorrect velocity models. This principle applies equally well to nonlayered models: then individual shot-gather reflectivity estimates will not be flat, but collections of traces representing the same surface location ("common x," "common image," or CDP gathers) will be flat when the correct velocity model is used. A strongly nonlayered illustration of this contention is discussed in the "Numerical Exploration" section below.

We quantify the coherence-of-inverted-reflectivities notion into a misfit function by combining misfit versions of equations (6) and (7):

$$\|S[v_b]r - S_{\text{data}}\|^2 + \sigma^2 \|\partial r/\partial x_s\|^2 \tag{8}$$

where r is now allowed to depend explicitly on x_s — with such dependence penalized by the second term, weighted by a parameter σ^2.

The function (8) is quadratic in r, so the minimization with respect to r presents no difficulties, in principle. On the other hand, as a functional of *both* v_b and r, it is still quite non-convex, for the same reasons as is J_{LS}. Together these two observations suggest *elimination* of r: that is, we define a function of v_b only by

$$\boxed{J_\sigma\big[v_b; S_{\text{data}}\big] = \min_r \frac{1}{2}\left\{\left\|S[v_b]\cdot r - S_{\text{data}}\right\|^2 + \sigma^2\left\|\frac{\partial r}{\partial x_s}\right\|^2\right\}.}$$

It is a remarkable fact that this function is smooth — in fact, nearly quadratic — in its dependence on v_b, despite its rather close relation with the least squares function! Minimization of J_σ over a class of smooth background velocities v_b is the *differential semblance optimization (DSO) problem.* Note that for noise-free data $S_{\text{data}} = S[v_b^*]r^*$, J_σ attains the value 0 for $v_b = v_b^*$, which is clearly its global minimum, and that this minimum is reached by

setting $r = r^*$ on the right-hand side. That is, the global minimum is achieved at the correct velocity — and, implicitly, at the correct reflectivity. Since J_σ is smooth, it is necessarily convex near a consistent global minimizer. That is, if the data is noise-free, then J_σ is necessarily convex near v_b^*, and remains convex when S_{data} is perturbed by small amounts of noise. We conjecture that J_σ is strongly convex for near-consistent data S_{data} and proper choice of σ^2, over a large subset of background velocity models. We will present some numerical evidence for this later in the paper, and have given a proof for the related (but simpler) plane/wave layered medium case (see Symes [1990], [1991a, b]; Symes and Carazzone [1989]).

Smoothness of the DSO Objective Function

In this section, we will outline the reasons for thinking that J_σ is smooth and might be minimized quite efficiently. We give only the formal skeletons of arguments here. Guy Chavent (personal communication) has pointed out that the partial optimization over r amounts to an *implicit reparameterization* of r by time — see Symes [1990], Appendix, and Clement [1991]. The results of the plane-wave/layered medium DSO, cited above, were obtained using explicit reparameterization by (travel) time, which is straightforward in that case.

Before starting we take care of a few technical details. The first is that the normal operator

$$S^T S$$

is a *pseudodifferential operator of order* $n-1$ *in dimension* $n(=1, 2,$ *or* $3)$ if the source is impulsive, $f(x,t) = \delta(x - x_s)\delta(t)$, under some ray-geometric restrictions (no caustics in the incident wavefront). That is, $S^T S$ is given by an oscillatory integral of the form

$$S^T S\, r(x_s, x) = \int d\xi\, b(x_s, x, \xi) e^{ix\cdot\xi} \hat{r}(x_s, \xi)$$

where $b \sim b_0(x_s, x, \xi/|\xi|)|\xi|^{n-1}$ as $|\xi| \to \infty$, and $\hat{r}(x_s, \xi)$ is the Fourier transform in x of $r(x_s, x)$.

The theory of pseudodifferential operators is an indispensible tool in developing a precise and effective understanding of the reflection seismic inverse

problem. Good references are Taylor [1980] and Hörmander [1983]. For the following discussion, the reader needs to understand that

(a) A pseudodifferential operator of order k (roughly) scales the Fourier transform of the function to which it is applied by $|\xi|^k$. That is, it "acts" like a k^{th} derivative operator. Here k may be negative or even non-integral. A pseudodifferential operator of order zero does not alter the rate of growth of the Fourier components, so acts as a bounded operator on L^2.

(b) If the symbol b of a pseudodifferential operator

$$B[v]u(x) = \int d\xi b(v, x, \xi) e^{i\xi \cdot x} \hat{u}(\xi)$$

depends smoothly on parameters v, then so does the operator $B[v]$, and all derivatives with respect to v are operators of the same order. That is, differentiation with respect to parameters does not result in increased weighting of high-frequency components. Note the contrast with the behaviour of oscillatory integrals of the type defining S, as in equations (3) and (5), in which the phase also depends on parameters. This contrast underlies the entire theory developed in these pages.

In fact $S^T S$ is a family of pseudodifferential operators in x, parameterized by x_s, of order $n-1$. It is a slight technical headache that such a family of pseudodifferential operators in x is *not* a pseudodifferential operator in x *and* x_s; however this is not an essential complication (e.g., Taylor [1975], Appendix) and we shall ignore it here. The pseudodifferential representation of $S^T S$ is an immediate consequence of the FIO calculus (Duistermaat [1973]), and is mentioned explicitly in Beylkin [1985], Symes [1986], and Rakesh [1988] for example. As shown in Percell's 1989 thesis, this conclusion is *false* when caustics are present in the incident wave-front — a generic occurrence in heterogeneous media. It is possible to recover the pseudodifferential nature of $S^T S$ by modifying the definition of S by a microlocal (phase space) suppression or muting of the reflectivity r. Without going into details, we assume that this has been done. Then $S^T S$ operator is *elliptic*, i.e. acts as an invertible Fourier multiplier at high spatial frequencies, over a conic sector of wave vectors (the "reflection aperture") determined by the relative positions of sources and receivers and the ray geometry of the background velocity

field. Outside of the reflection aperture, which varies with location in the subsurface, S^TS suppresses high-frequency components (these correspond to *off-cable reflections*). That is, the symbol b satisfies

$$\begin{aligned} b_0(x_s, x, \xi/|\xi|) &> 0 \text{ within the reflection aperture} \\ b_0(x_s, x, \xi/|\xi|) &= 0 \text{ outside of the reflection aperture .} \end{aligned}$$

The high-frequency components of r outside the inversion aperture must be constrained *a priori* in solving equations involving S^TS. To accomplish this goal in a well-scaled way, we first modify the definition of S: we assume that the source has point-support, and in its time dependence is a low frequency perturbation of the $(\frac{3-n}{2})$-th derivative of $\delta(t)$: thus

$$f(x,t) = \text{const.} \quad \begin{cases} \delta(x - x_s) t_+^{-\frac{3}{2}} , & n = 2 \\ \delta(x - x_s)\delta(t) , & n = 3 \\ \qquad + f_0(x,t) \end{cases}$$

(the distribution $t_+^{-\lambda}$ is defined in Gel'fand and Shilov [1958], for example), where f_0 is a smooth function. This amounts to assuming that f, while bandlimited below, behaves as $t_+^{-\frac{3}{2}}$ $(n = 2)$ or $\delta(t)$ $(n = 3)$ across the upper part of the passband of the seismic signals. Practically, this assumption is realized by preprocessing the data to re-scale it in the frequency domain; i.e., subject it to a spiking deconvolution followed by a suitable filter. Then, to some degree of approximation, the data will behave as if it were the response to the $(\frac{3-n}{2})$-derivative of $\delta(t)$ at "high" frequencies.

With this modification, S^TS is a *pseudodifferential operator* of order 2, independently of the dimension.

We chose a regularizing operator W, pseudodifferential of order 2 in x and depending parametrically on x_s, so that

$$S^TS + \lambda^2 W$$

is *uniformly elliptic* for each x_s as long as $\lambda^2 > 0$. A simple choice is

$$W = I - \nabla_x^2 .$$

This choice is suboptimal, as it also affects the components within the reflection aperture, but for small λ^2 this is probably of little consequence. It will

be important in the sequel to write $W = C^T C$, with C a pseudodifferential operator of order 1. This is certainly possible for the simple choice just given with $C = (I - \nabla_x^2)^{\frac{1}{2}}$.

The (regularized) differential semblance functional J_σ is defined by minimizing over r the (regularized) quadratic

$$Q_\sigma = \frac{1}{2}\left\{ \left\| Sr - S_{\text{data}} \right\|^2 + \lambda^2 \langle r, Wr \rangle + \sigma^2 \left\| \frac{\partial r}{\partial x_s} \right\|^2 \right\} .$$

A minimizer of Q_σ is a solution of the normal equations

$$Nr := [S^T S + \lambda^2 W - \sigma^2 \, \partial^2/\partial x_s^2] r = S^T S_{\text{data}} .$$

The operator N is (essentially) an elliptic pseudodifferential operator of order 2 in x, x_s. Standard techniques show that N is invertible, under reasonable restrictions on r, and that r depends stably on S_{data} in suitable norms.

Since $S = S[v_b]$, the solution of the normal equations also depends on v_b: $r = r[v_b, S_{\text{data}}]$ also. The dependence of r on v_b is quite erratic — this is another consequence of our analysis of the least-squares problem in the last section. In any case we can write $r = N^{-1} S^T S_{\text{data}}$ and obtain

$$\begin{aligned} J_\sigma &= \min_r Q_\sigma \\ &= \frac{1}{2} \langle S^T S_{\text{data}}, (I - N^{-1}) S^T S_{\text{data}} \rangle \\ &= \frac{1}{2} \langle S_{\text{data}}, \, S(I - N^{-1}) S^T S_{\text{data}} \rangle . \end{aligned}$$

The operator $S(I - N^{-1})S^T$ is — essentially — pseudodifferential, for the same reasons as is the normal operator. Its symbol is a smooth function of v_b, whence follows the smoothness of J_σ, by virtue of the property of pseudodifferential operators depending on a parameter, noted earlier.

Note that the quantity $S^T S_{\text{data}}$ is *not* a smooth functional of v_b: it is given by an oscillatory integral with phase depending nontrivially on v_b. The same is true of $N^{-1} S^T S$. We can write

$$\begin{aligned} \langle S^T S_{\text{data}}, N^{-1} S^T S_{\text{data}} \rangle &= \frac{1}{4} \left\{ \| (I + N^{-1}) S^T S_{\text{data}} \|^2 \right. \\ &\quad \left. - \, \| (I - N^{-1}) S^T S_{\text{data}} \|^2 \right\} . \end{aligned}$$

Thus the smoothness of J_σ is an amusing instance of the observation that a non-smooth function taking values in a Hilbert space may have a smooth norm.

We assume that we apply to the normal equations an algorithm yielding, after a finite number of steps, an estimate of r. We will consider only Krylov space methods such as conjugate gradient iteration (Golub and Van Loan [1983], Ch. 10). The normal equations are very similar in nature to the Laplace equation, and in particular preconditioning is required to produce rapid convergence of the Krylov sequence. Preconditioning amounts to replacing the normal equations by the *preconditioned normal equations*

$$Mr = S^* S_{\text{data}}$$

where $M = GN$, $S^* = GS^T$, and G is the solution operator of the Laplace equation in (x, x_s) with suitable boundary conditions. (In fact, S^* is the adjoint of S with respect to a so-called Sobolev scalar product.) This is the system solved in the numerical work reported in the next section. The operator M is now of order zero, i.e. does not enhance high-frequency components (nor does it suppress them). In a qualitative sense at least, M is as well-conditioned as possible.

Numerical Exploration

In this section we present some preliminary numerical results, in which we have evaluated the coherency optimization functional over line segments in the space of background velocity models, to display directly its smooth and convex nature.

For reasons of computational cost, we have restricted our experiments to two-dimensional models. Since we have already established the ability of DSO to estimate layered velocity models, we present here a thoroughly nonlayered example, abstracted from the Marmousi model (Versteeg and Grau [1991]).

As is evident from the expression

$$\begin{aligned} Q_\sigma &= \|Sr - S_{\text{data}}\|^2 + \lambda^2 \langle r, Wr \rangle + \sigma^2 \|\partial r / \partial x_s\|^2 \\ N &= S^T S + \lambda^2 W - \sigma^2 \, \partial^2 / \partial x_s^2 \\ M &= GN \end{aligned}$$

it is necessary to approximate

(i) the seismogram ("forward map") S;

(ii) its adjoint S^T;

(iii) the regularizing operator W;

(iv) $\partial/\partial x_s$;

(v) the solution operator G of the Laplace equation in x and x_s, with suitable boundary conditions.

We shall examine each of these approximations briefly.

We approximated the seismogram ("forward modeling") operator S *via* finite difference approximate solution of the wave equations for the pressure field $p(x,t)$ and its perturbation $\delta p(x,t)$, which were displayed in Section 2. We used a centered difference scheme of order 4 in space and 2 in time; see for example Dablain [1986] for discussion and references. The seismogram was extracted from the pressure perturbation field $\delta p(x,t)$ by sampling at the receiver locations, with bilinear interpolation when these are not grid points.

The adjoint operator S^T is approximated *via* solution of the adjoint state equations, as explained for example in Tarantola [1987]. For reliable convergence of the algorithm described here it is essential that the discrete versions of S and S^T actually be adjoint, i.e. that

$$\langle Sr, \phi \rangle_t = \langle r, S\phi \rangle_z$$

where $\langle\ ,\ \rangle_t$ and $\langle\ ,\ \rangle_z$ are time- and depth-domain discrete L^2 inner products. This can be accomplished *via* discrete implementation of the adjoint state technique.

The regularizing operator W is a scaled version of the discrete (2D) Laplace operator in x. This operator is conveniently implemented in the (spatial) Fourier domain. It is applied to reflectivities (r), which are regarded as multiply periodic. The source-direction derivative $\partial/\partial x_s$ is approximated by the midpoint rule.

The 3D Laplace operator H in x and x_s is a weighted sum of W and $(\partial/\partial x_s)^T\ \partial/\partial x_s = -\partial^2/\partial x_s^2$. As noted before, W is approximated by multiplication in the Fourier domain. The operator $-\partial^2/\partial x_s^2$ is tridiagonal, when

$\partial/\partial x_s$ is discretized by the midpoint rule. Consequently the application of the inverse operator G involves the solution of a system of tridiagonal equations, one for each spatial wave vector.

To compute the DSO objective function, we must minimize the functional Q_σ defined above over reflectivities. Since this functional is quadratic in reflectivity, it is natural to use a member of the conjugate-gradient family of algorithms, which are in some respects the most efficient iterative algorithms for quadratic minimization. We used conjugate residual iteration (Golub and Van Loan [1983], Ch. 10), motivated by the monotone decrease in the (normal) residual achieved by this algorithm.

The program is structured to operate in out-of-core mode: the simulation and adjoint simulation (migration) require only the data for a single shot to reside in-core at any one time, and the tridiagonal solution (G) is implemented *via* a paging scheme. The code is written in FORTRAN, except for the i/o system which is written in C. The use of XDR portable binary data encoding has proved vital to this project, enabling us to use machines of several different architectures interchangeably. The bulk of our work has been carried out with Sun Sparc stations, Stardent Titan, Cray Y-MP and Connection Machine 2 machines.

We used this program to compute some sample values of the DSO objective function in an example derived from the Marmousi model (see Bourgeois *et al.* [1991] for an extensive discussion of this model). As explained in "The Acoustic Model" above, we split the Marmousi velocity field v into a smooth background field v_b and a rough reflectivity $r = \delta v/v$. The smooth background field v_b was extracted from v by averaging on a length scale of 300m, using a trapezoidal kernel. Then $\delta v = v - v_b$. Grey-scale plots of v_b and r are presented in Figure 6.

We simulated a survey across this model of 40 shots spaced 50m apart, the westmost (leftmost) shot being located 4000m from the western edge. Thus the eastmost (rightmost) shot is located just short of 6000m, on top of the block of high-velocity material extending up to the sea-floor near the middle of the model. We used a Ricker wavelet source time dependence, with peak frequency of roughly 12.5 Hz, and recorded into an array ("cable") extending to the right (east) of each shot. In other respects, our shooting parameters were as described in Bourgeois *et al.* [1991]: the source is a superposition of 6 point sources at 8m depth spaced 6m apart; the far receiver is offset 2525m from the source center, the near receiver 150m, the receiver spacing is

25m, the receiver depth 12m. We were able to obtain acceptable simulations with a spatial grid with spacing 16m in both directions, and a time sample increment of 1.5ms. We recorded 1.5s of data.

A typical shot (common source) gather appears as Figure 7. Since the theory developed above applied only to precritical reflections, we have muted (windowed) the data to suppress the direct wave and other postcritical energy arriving early in each trace. The mute is incorporated into the definition of the simulation and adjoint simulation operators.

The main computational task in computing the values of J_σ is estimation of the reflectivity r by least-squares solution of the linearized inverse problem, i.e. minimization of Q_σ or solution of the normal equations. This solution is of course only achieved approximately by conjugate residual iteration, and the resulting estimate for r substituted in the expression for J_σ. The accuracy of r is measured by the normal residual decrease. In all cases we carried out ten steps of conjugate residual iteration and in so doing reduced the normal residual by a factor of roughly 20.

The curves in Figure 8 interpolate samples of the DSO objective function at velocities v_h defined by

$$v_h(x,z) = (1-h)1500m/s + hv_b(x,z)$$

with $h = 0.0, 0.25, 0.5, 0.625, 0.75, 0.875, 1.0$ and 1.125. We set the differential semblance weight $\sigma^2 = 1.0$, and the regularization weight $\lambda^2 = 10^{-4}$. With these parameter choices, this one-dimensional slice through the DSO objective function appears to exhibit the properties predicted by the theoretical analysis:

- smoothness, at least at the sample scale;
- convexity near the minimum, and quasiconvexity (absence of secondary critical points) otherwise;
- minimum achieved at the target velocity field ($h = 1$).

At least if restricted to this one-dimensional slice of velocity model space, DSO would apparently recover the smoothed Marmousi model in a few steps of a quasi-Newton algorithm.

The calculations presented here were performed on the Connection Machine 2 operated by the Advanced Computing Laboratory, Los Alamos National Laboratory. The code was written in CM Fortran and C, and compiled

with the CMF 1.0 "slicewise" compiler. The experiments required a total of 40 CPU hours on one 16k quadrant of the CM2 to produce the data of Figure 8. The CM2 delivered roughly 250 MFLOPS in the finite difference portions of the code. Careful optimization and use of the Datavault i/o channel yielded a throughput equivalent to roughly 200 MFLOPS.

Discussion and Conclusion

Let us state precisely the scientific justification for our approach, i.e. the sense in which we hope to solve the velocity inversion problem *via* differential semblance optimization. First, for noise-free data, as noted above, the "exact" model is amongst the minima of J_σ. Thus the velocity estimator obtained by minimization of J_σ is consistent with the model. Second, if our convexity conjecture is correct, then the velocity estimate is a *stable* function of the data insofar as kinematic information resolves velocity; in particular it is well-defined near the set of consistent data ("near" in the mean-square sense). Third, because of the smoothness (proved) and convexity (conjectured) of J_σ, rapidly convergent Newton-type iteration will produce the estimator: it is efficiently computable.

We believe that this triple of properties (consistency, stability, and computability of the velocity estimator) form a satisfactory substitute for the statistical inverse theory (Tarantola [1987]). These are *mathematical* properties, and imply that the model recovered by minimization of J_σ will be quite accurate (at least in its ability to reproduce the data) when the data noise is small in a well defined sense. Smallness of the noise, both random and model-error, can be verified — by solving the minimization problem! Thus the approach outlined in this paper is justified if reflection seismic data are near-consistent, i.e. "low noise". We can cite two sorts of evidence for believing that high-quality marine data, at least, is low-noise. First, in a few cases, least-squares inversion has succeeded (with lots of human intervention) in producing data fits in the 70%–95% RMS range (Cao *et al.* [1990], Kolb and Helgesen [1989], Carazzone and Srnka [1989]), using severely simplified models (in particular layered media). Second, reflection seismic data appears to be highly reproducible, when sufficiently rigorous control of experimental parameters is exercised. That is, the truly irreproducible or "random" components of data noise are possibly quite small, at least in some instances, so

that any additional data misfit must be due to inadequate deterministic physical modeling. Thus we argue that high-quality reflection seismic data may be quite close to that predicted by a suitably sophisticated wave-propagation model, and that extensions of the technique proposed here may be capable of adjusting the parameters to verify such a fit.

We have described a modified least-squares principle for velocity and reflectivity estimation. We have explained how this principle avoids the nonconvexity of least-squares inversion. Finally we have offered some preliminary numerical evidence in support of the conjecture that this functional is smooth and has no secondary minima over a large domain in model space. Should this conjecture hold, quasi-Newton methods should be adequate to approximate the global minimum, at a kinematically correct velocity model. Construction of such a quasi-Newton method requires efficient and reasonably accurate calculation of the gradient of the DSO objective function, and algorithms to set the differential semblance weight σ^2 and the regularizing weight λ^2. Calculation of the gradient is addressed in Kern and Symes [1992]. A crude attempt at inversion of (part of) the Marmousi data set, using a less accurate gradient calculation, appears in Symes [1991c], and some further discussion of the underlying mathematics in Symes [1991d].

Acknowledgements: This work was partly supported by the National Science Foundation, the Office of Naval Research, the State of Texas, and the sponsors of the Rice Inversion Project: Amoco Research, Conoco Inc., Mobil Research and Development Corp., Exxon Production Research Co., Earth Modeling Systems, and Cray Research Inc.

REFERENCES

Beylkin, G., Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized radon transform, *J. Math. Phys.*, 26, 99–108, 1985.

Bourgeois, A., P. Lailly, and R. Versteeg, The Marmousi model. In R. Versteeg and G. Grau, editors, *Practical Aspects of Inversion: the Marmousi Experience*, The Hague, 1991. EAEG.

Cao, D., S. Singh, and A. Tarantola, Simultaneous inversion for background velocity and impedance maps, *Geophysics*, 55, 458–469, 1990.

Carazzone, J., and L. Srnka, Elastic inversion of Gulf of Mexico Data. In *Proc. 59th Annual International Meeting*, pages 956–958, Dallas, 1989. Society of Exploration Geophysicists. Expanded abstract.

Clement, F., A traveltime-based formulation for seismic inversion. In Cohen, Halpern, and Joly, editors, *Mathematical and Numerical Aspects of Wave Propagation Phenomena.* SIAM, Philadelphia, 1991.

Cohen, J.K., and N. Bleistein, An inverse method for determining small variations in propagation speed, *SIAM J. Appl. Math.*, 32, 784–799, 1977.

Dablain, M.A., The application of high-order differencing to the scalar wave equation, *Geophysics*, 51, 54–66, 1986.

Duistermaat, J., Fourier integral operators. Lecture notes, Courant Institute, New York, 1973.

Gauthier, O., A. Tarantola, and J. Virieux, Two-dimensional nonlinear inversion of seismic waveforms, *Geophysics*, 51, 1387–1403, 1986.

Gel'fand, I.M., and G.E. Shilov, *Generalized Functions*, volume I, Academic Press, New York, 1958.

Golub, G., and C. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 1983.

Helgesen, J. and P. Kolb, Inversion of 1-D acoustic medium from real seismic data. In *Proc. 59th Annual International Meeting*, pages 1000–1003, Society of Exploration Geophysicists, 1989. Expanded abstract.

Hörmander, L., *The Analysis of Linear Partial Differential Operators*, volume I, Springer Verlag, Berlin, 1983.

Kern, M. and W. Symes, Implementation of differential semblance optimization for 2D velocity estimation, Technical Report 92-3, Department of Mathematical Sciences, Rice University, Houston, TX, 1992.

Kolb, P., F. Collino, and P. Lailly, Prestack inversion of a 1d medium. In *IEEE 74*, pages 498–506, 1986.

Mosegard, K. and A. Tarantola, Monte Carlo analysis of geophysical inverse problems. In *Proc. 61st Annual International Meeting and Exposition*, Society of Exploration Geophysicists, 1991.

Percell, C., The effect of caustics in acoustics inverse scattering experiments, Technical Report 89-3, Department of Mathematical Sciences, Rice University, Houston, Texas, U.S.A, 1989.

Rakesh, A linearized inverse problem for the wave equation, *Comm. on P.D.E.*, 13, 573–601, 1988.

Santosa, F. and W. Symes, An analysis of least-squares velocity inversion, Geophysical Monograph 4, Soc. of Expl. Geophys., Tulsa, 1989.

Scales, J.A., M. Smith, and T.L. Fischer, Global optimization methods for highly nonlinear inverse aspects of wave propagation phenomena. In Cohen, Halpern, and Joly, editors, *Mathematical and Computational Aspects of Wave Propagation Phenomena.* SIAM, Philadelphia, 1991.

Sen, M.K., and P. Stoffa, Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics*, 56, 1624-1636, 1991.

Symes, W., Stability and instability results for inverse problems in several-dimensional wave propagation. In R. Glowinski and J.-L. Lions, editors, *Computing methods in applied science and engineering VI*, North-Holland, 1986.

Symes, W., Velocity inversion: a case study in infinite-dimensional optimization, *Math. Prog.*, 48, 71–102, 1990.

Symes, W., Layered velocity inversion: a model program from reflection seismology, *SIAM J. Math. Anal.*, 22, 680-716, 1991a.

Symes, W., A differential semblance algorithm for the inverse problem of

reflection seismology, *Computers and Math. with Appl.*, 22(4/5), 147-178, 1991b.

Symes, W., Non-interactive estimation of the Marmousi velocity model by differential semblance optimization: Initial trials. In G. Grau and R. Versteeg, editors, *The Marmousi Experience: Proceedings of the EAEG Workshop on Practical Aspects of Inversion*, The Hague, 1991c.

Symes, W., The reflection inverse problem. In Cohen, Halpern, and Joly, editors, *Mathematical and Numerical Aspects of Wave Propagation Phenomena*. SIAM, Philadelphia, 1991d.

Symes, W., and Carazzone, J., Velocity inversion by coherency optimization, Technical Report 89-8, Department of Mathematical Sciences, Rice University, Houston, TX, 1989. (To appear in Proc. of Workshop in Geophysical Inversion, ed. J.B. Bednar, SIAM).

Symes, W., and Carazzone, J., Velocity inversion by differential semblance optimization, *Geophysics*, 56(5), 654–663, 1991.

Tarantola, A., A strategy for nonlinear elastic inversion of seismic reflection data, *J. Geophysics*, 51, 1893–1903, 1986.

Tarantola, A., *Inverse Problem Theory*, Elsevier, 1987.

Tarantola, A, E. Crase, M. Jervis, Z. Konen, J. Lindgren, K. Mosegard, and M. Noble, Nonlinear inversion of seismograms: State of the art. In *Proc. 60th Annual International Meeting and Exposition*, San Francisco, 1990. Society of Exploration Geophysicists. Expanded abstract S13.7.

Taylor, M., Reflection of singularities of solutions to systems of differential equations, *Comm. on Pure and Applied Math.*, 28, 457–478, 1975.

Taylor, M., *Pseudodifferential Operators*, Princeton University Press, Princeton, New Jersey, 1980.

Versteeg, R. and G. Grau, editors, *Practical Aspects of Inversion: the Marmousi Experience*, EAEG, The Hague, 1991.

Figure 1(a). Depth profile of layered velocity model used in perturbation examples.

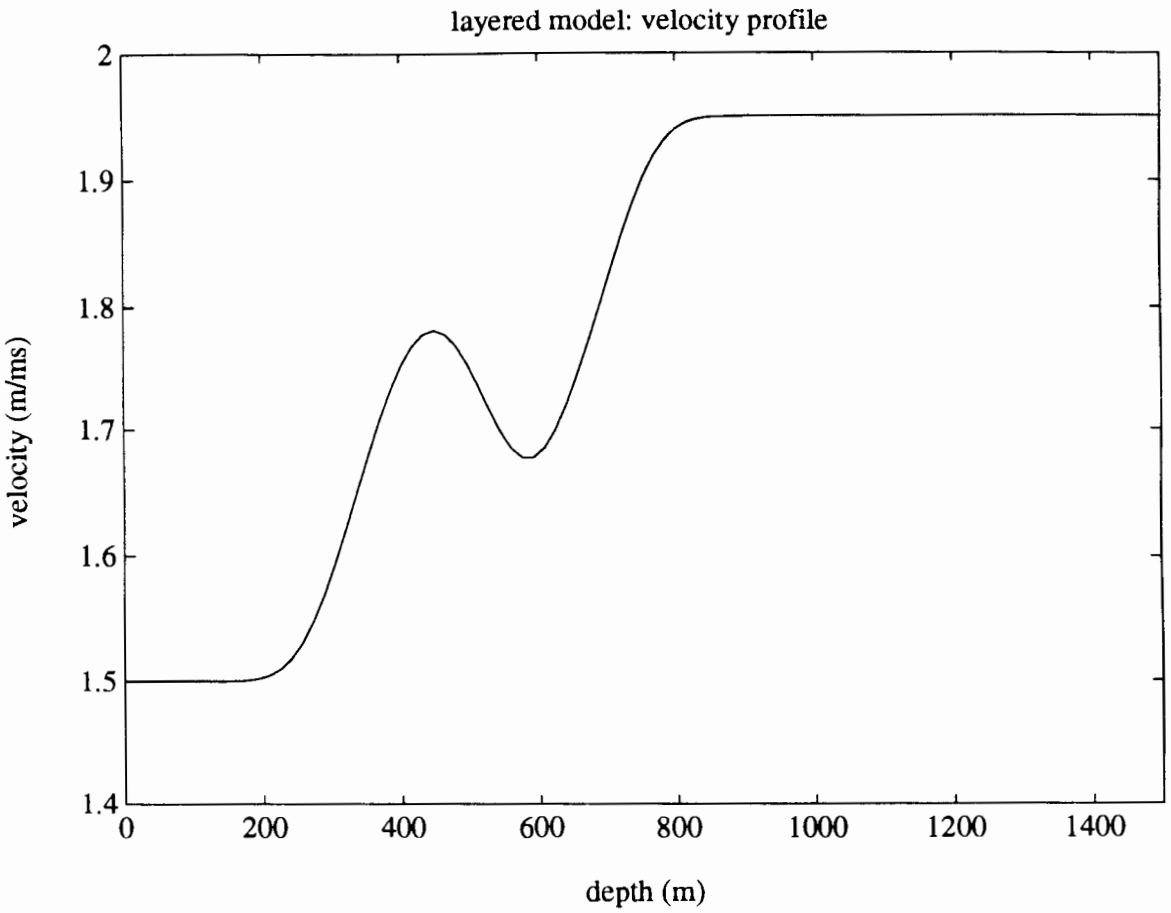
layered model: velocity profile
velocity (m/ms)
depth (m)
2
1.9
1.8
1.7
1.6
1.5
1.4
0
200
400
600
800
1000
1200
1400

Figure 1c

Figure 1(b). Depth profile of reflectivity.

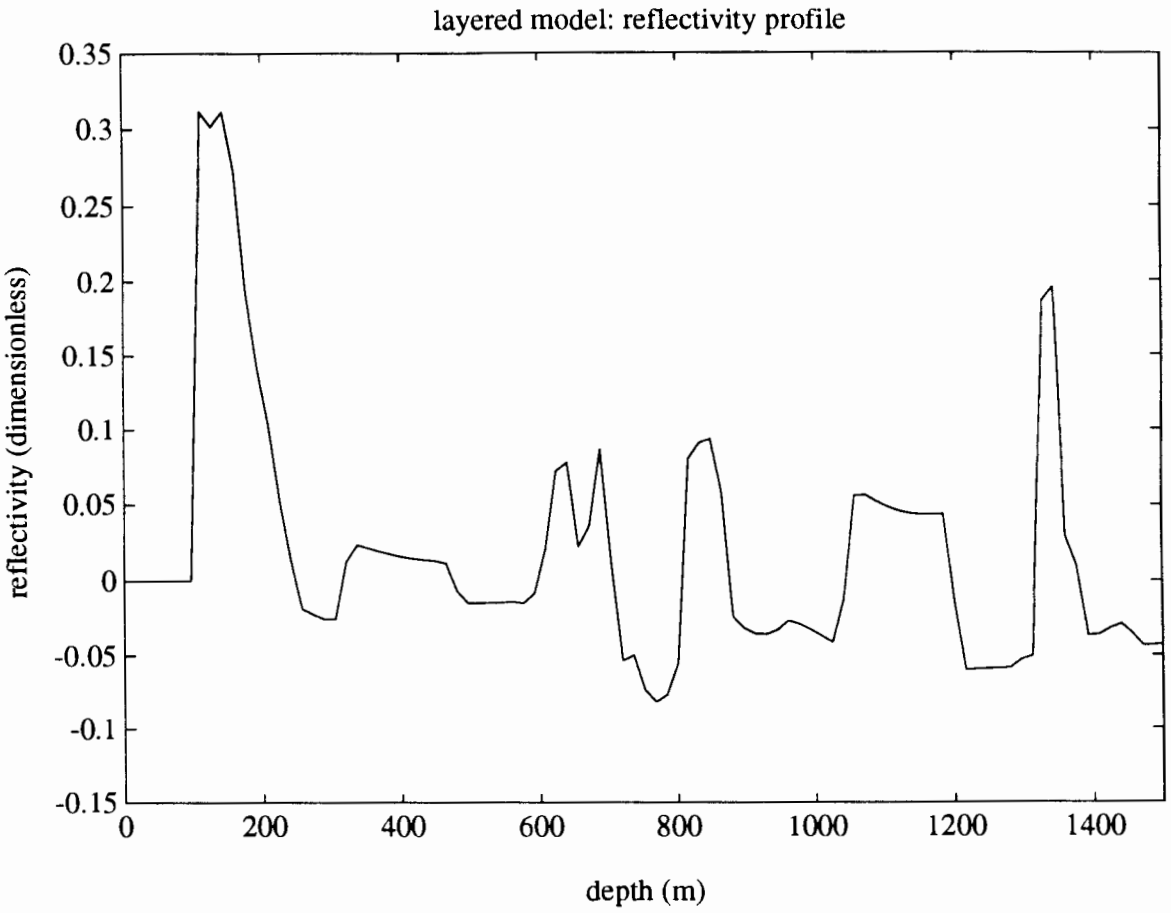
layered model: reflectivity profile
reflectivity (dimensionless)
depth (m)
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
0
200
400
600
800
1000
1200
1400

Figure 2(a). Common source seismogram (shot gather) for model of Figure 1. Source time dependence is 15 Hz Ricker wavelet. Source depth is 8m, receiver depth 12m. Near offset is 150m, far offset 1800m. Vertical axis unit is milliseconds.

data1.sep

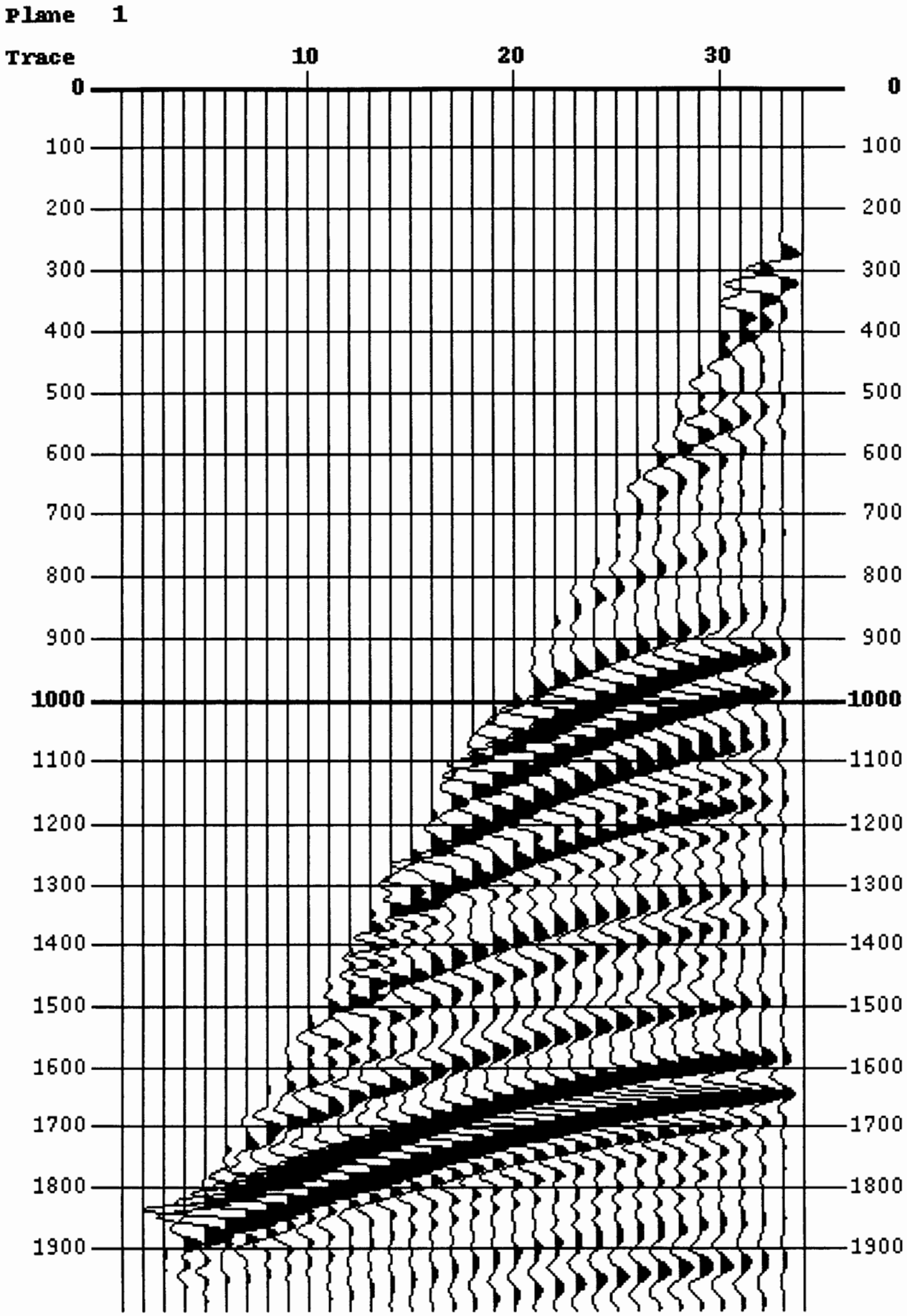
Plane 1
Trace
10
20
30
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

Figure 2(b). Perturbation of Figure 2(a) resulting from 5% perturbation of velocity. Notice that the size of the event perturbation increases with time, reflecting the accumulation of travel-time errors. Amplitude scale same as 2(a).

fdvtest

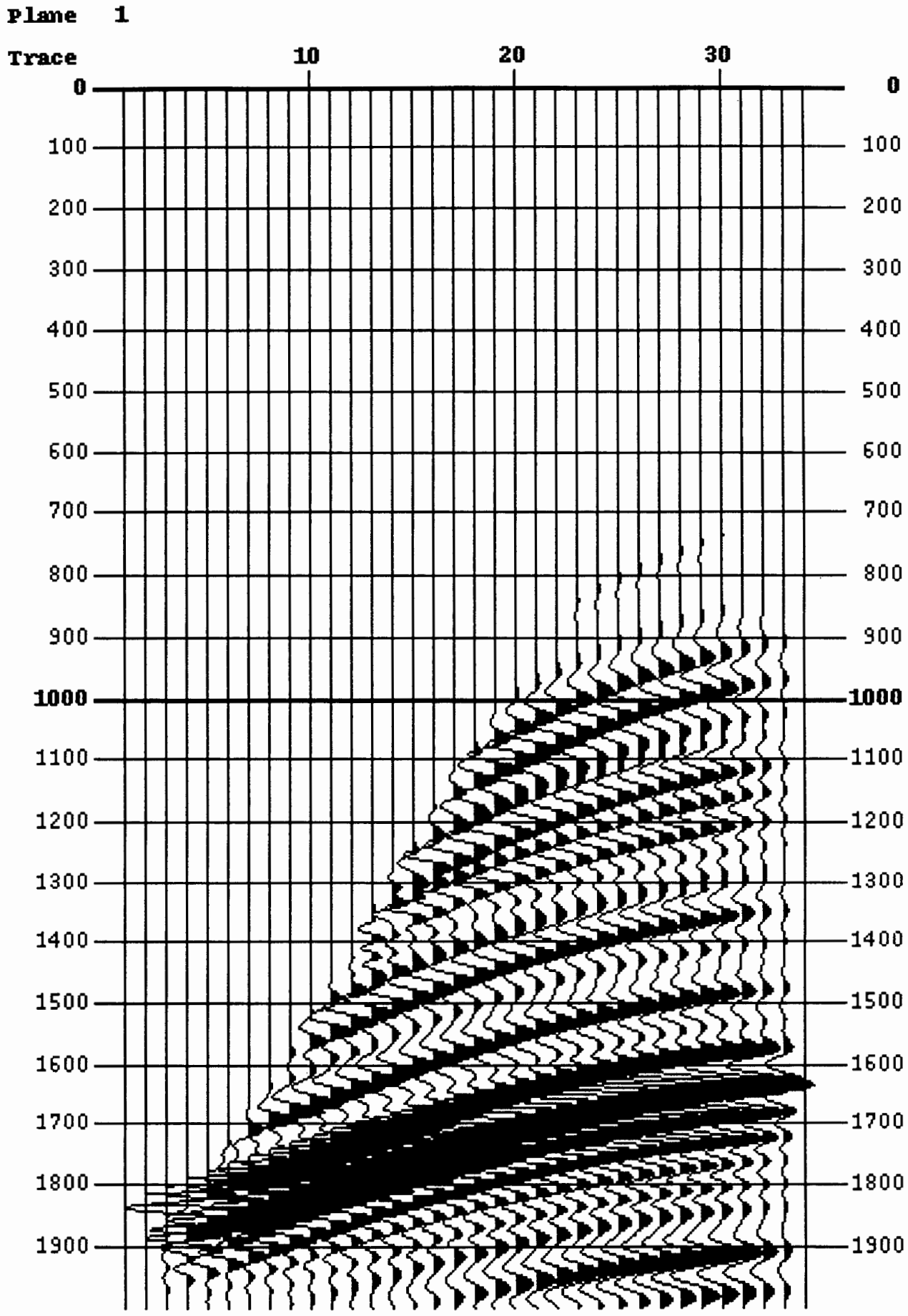
Plane 1
Trace
10
20
30
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

Figure 2(c). Perturbation of Figure 2(a) resulting from 5% perturbation of reflectivity. Amplitude scale same as 2(a).

fdrtest

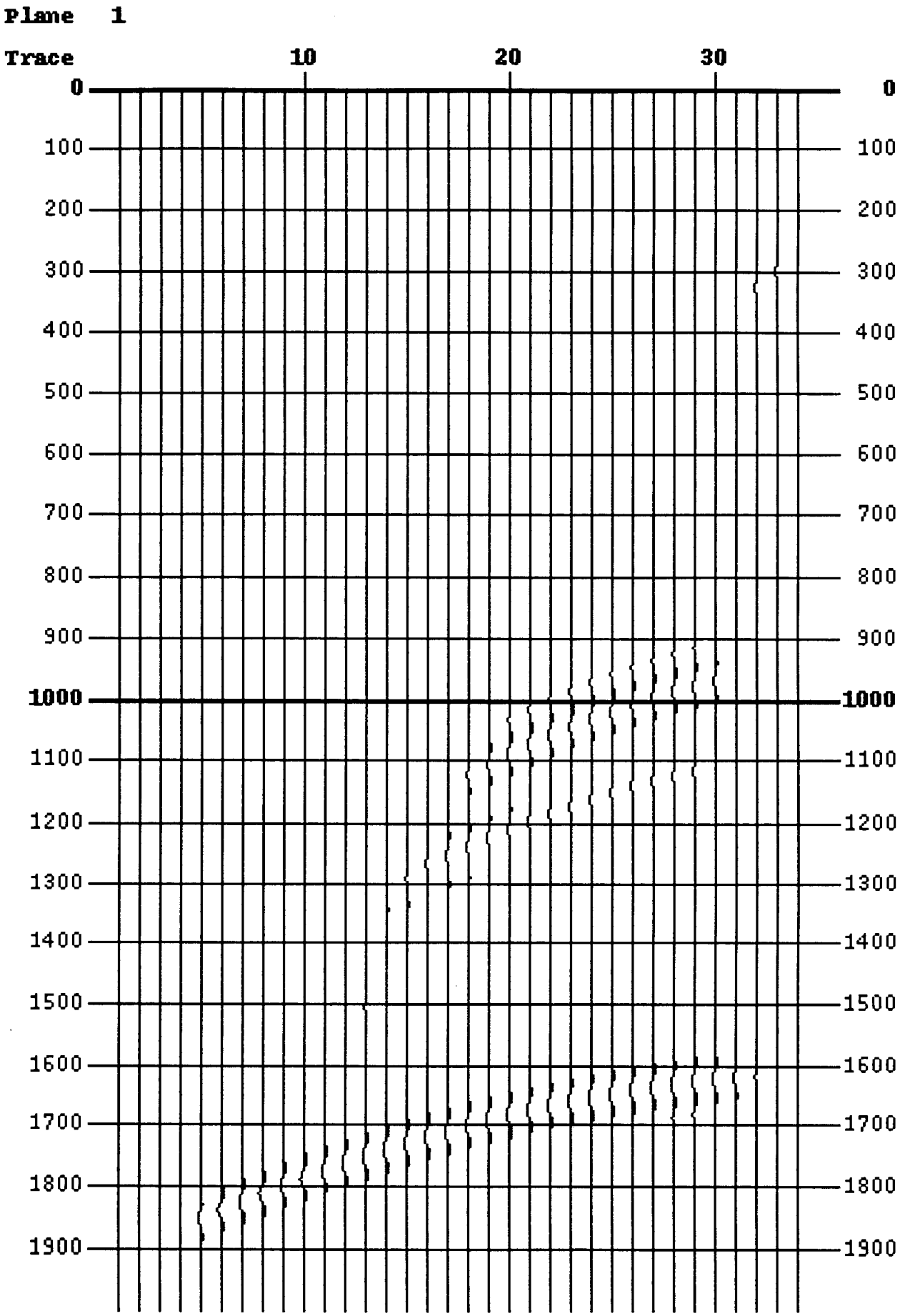
Plane 1
Trace
10
20
30
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

Figure 3. Inverted reflectivity from data of Figure 2(a), with constant background velocity (= 1500m/s). Note slope of major events.

re_cc2

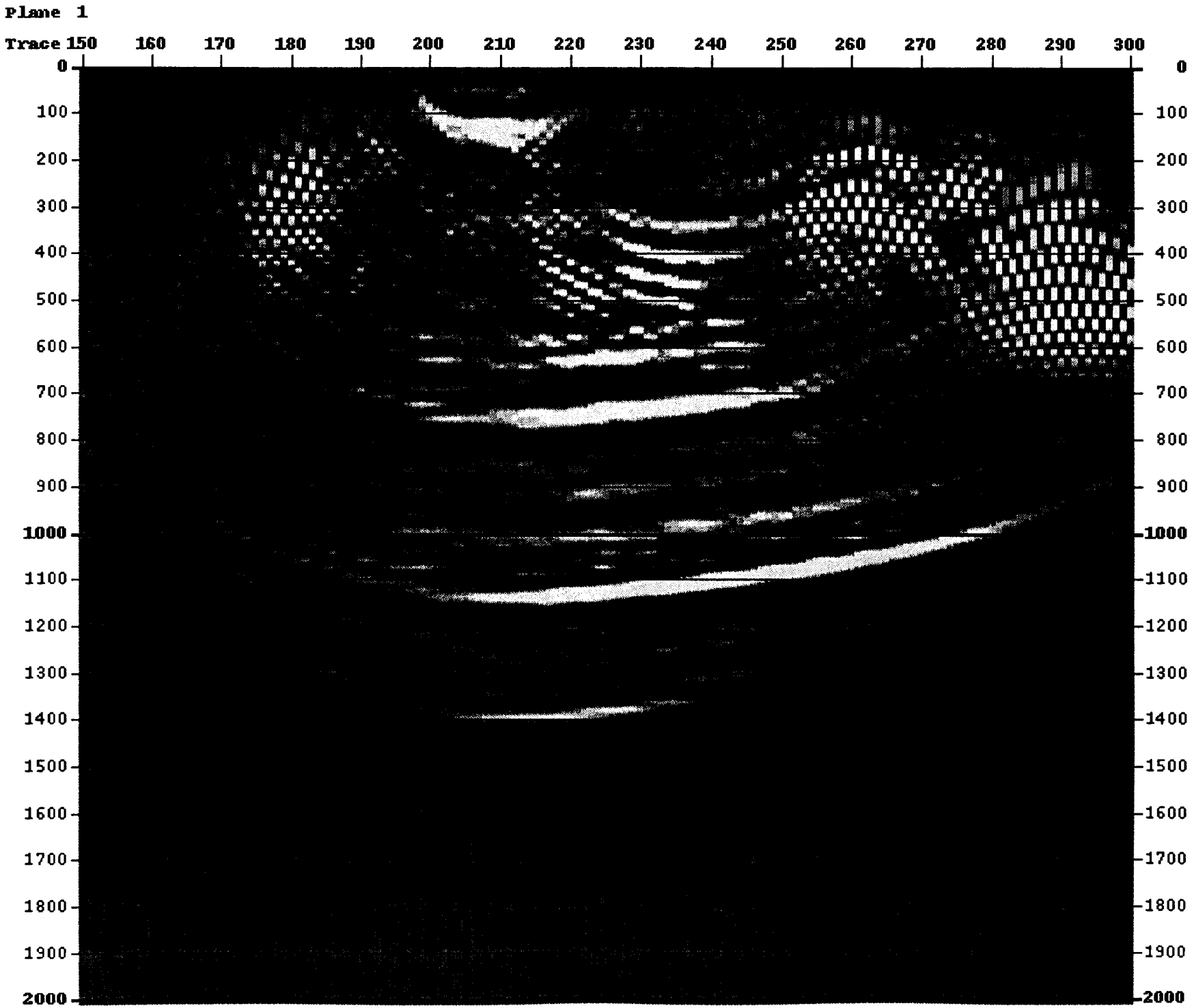
Plane 1
Trace 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000

Figure 4(a) Predicted data from constant velocity and reflectivity of Figure 3.

data2_cc

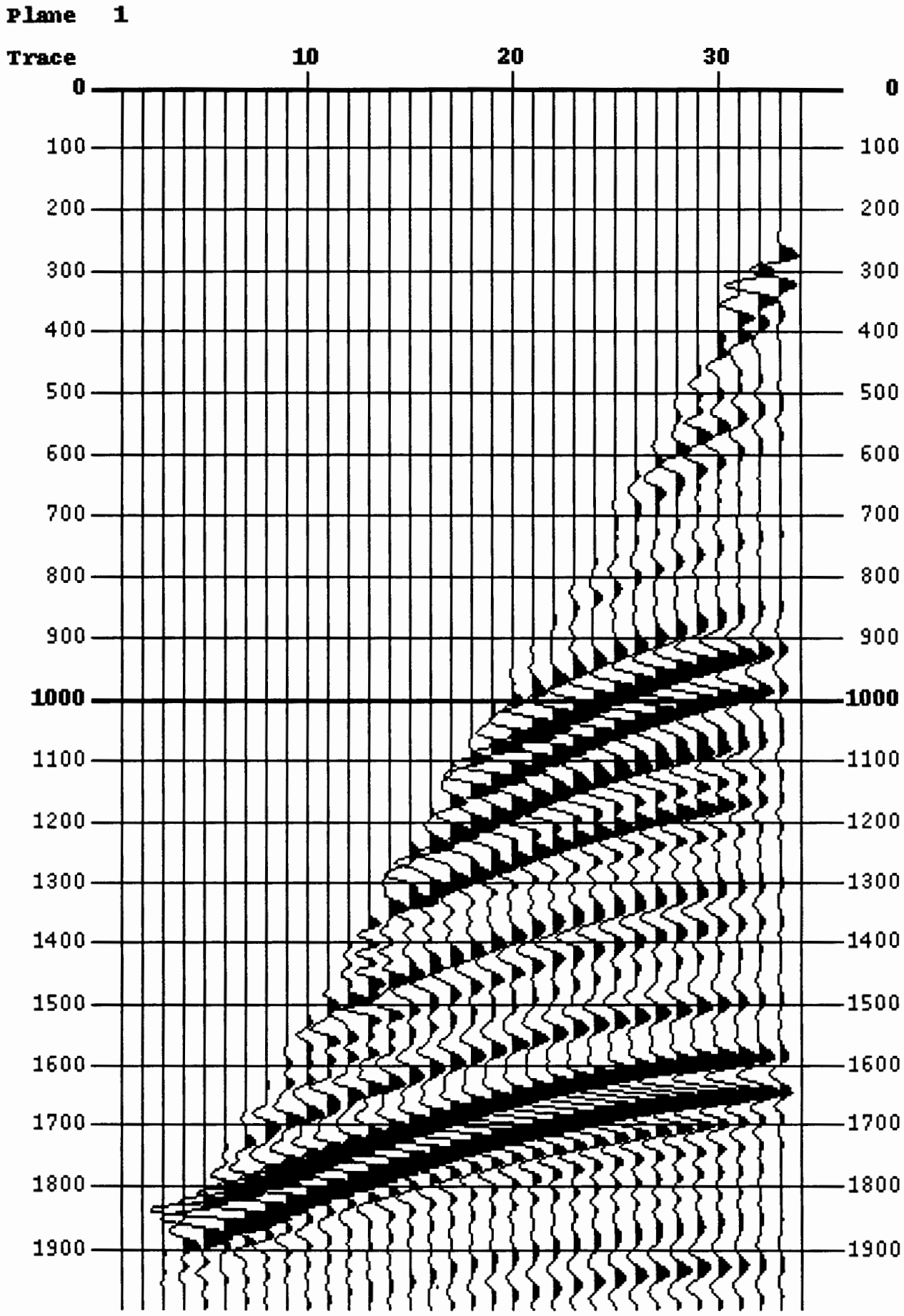
Plane 1
Trace
10
20
30
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

Figure 4(a). Error seismogram: difference between predicted data (Figure 4(a)) and input data (Figure 2(a)). Plotted on same scale as Figure 2(a). RMS error level is 6%.

diff_cc2

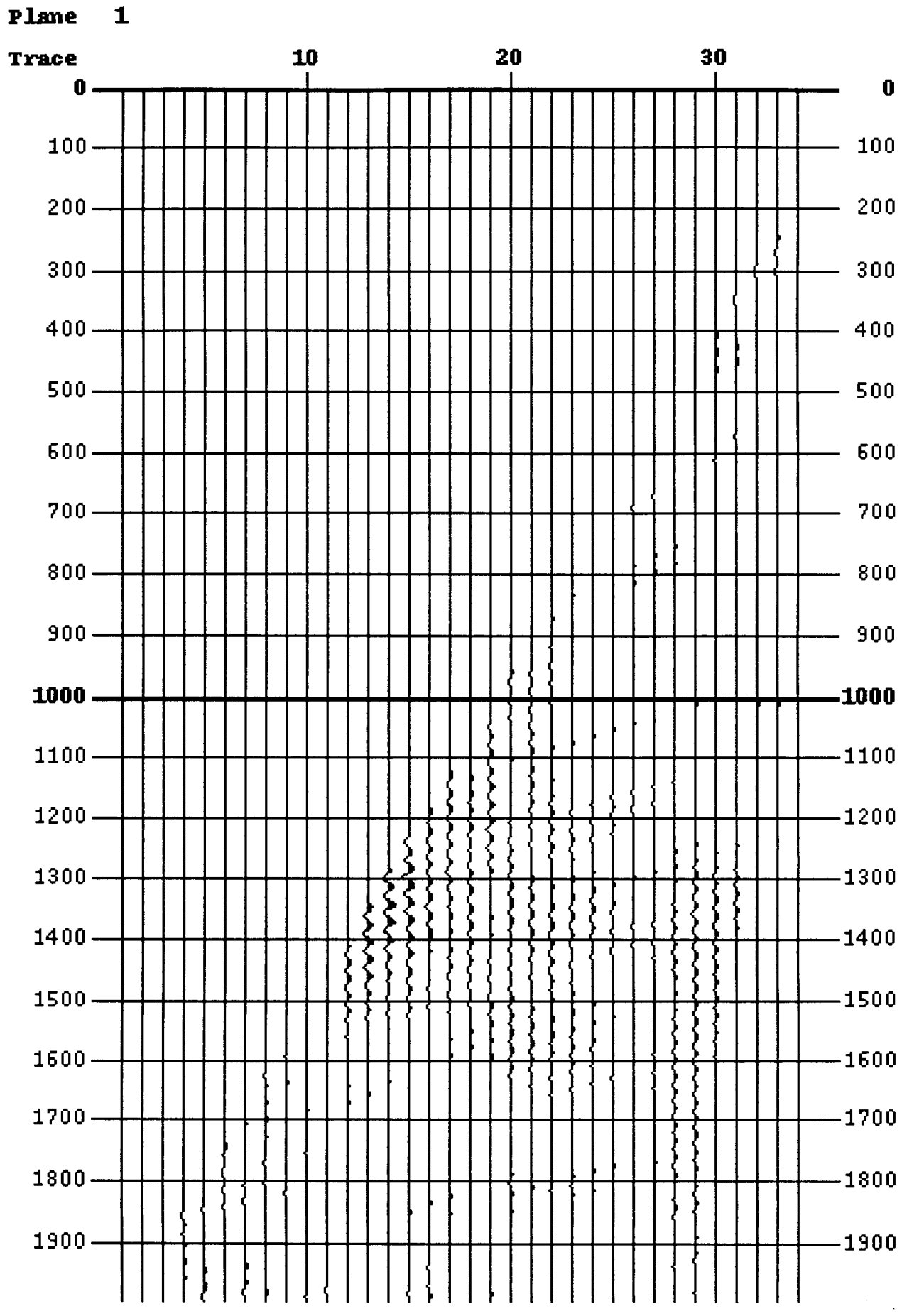
Plane 1
Trace
10
20
30
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900

Figure 5. Inverted reflectivity from data of Figure 2(a), with correct background velocity. Note that, apart from evident aperture effects (edge diffractions), major events are essentially *flat.*

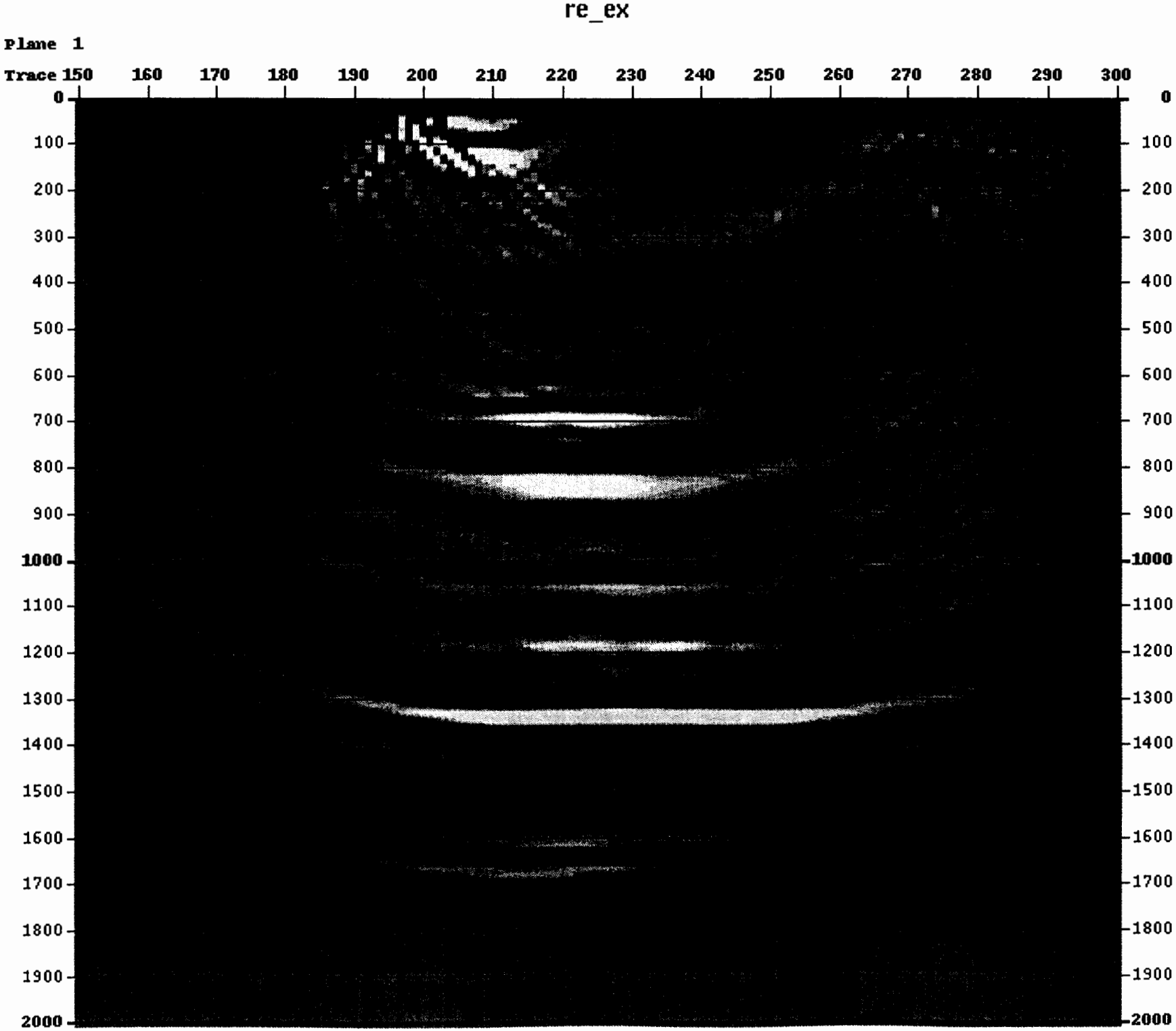
re_ex
Plane 1
Trace 150 160 170 180 190 200 210 220 230 240 250 260 270 280 290 300
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000

Figure 6(a). The central portion of the Marmousi velocity model, smoothed and plotted in a grey scale with low velocity (1500m/s) as black and high velocity (3700m/s) as white. The length scale of the smoothing was 300m. This model was used as the background velocity v_b in the experiments reported in the text.

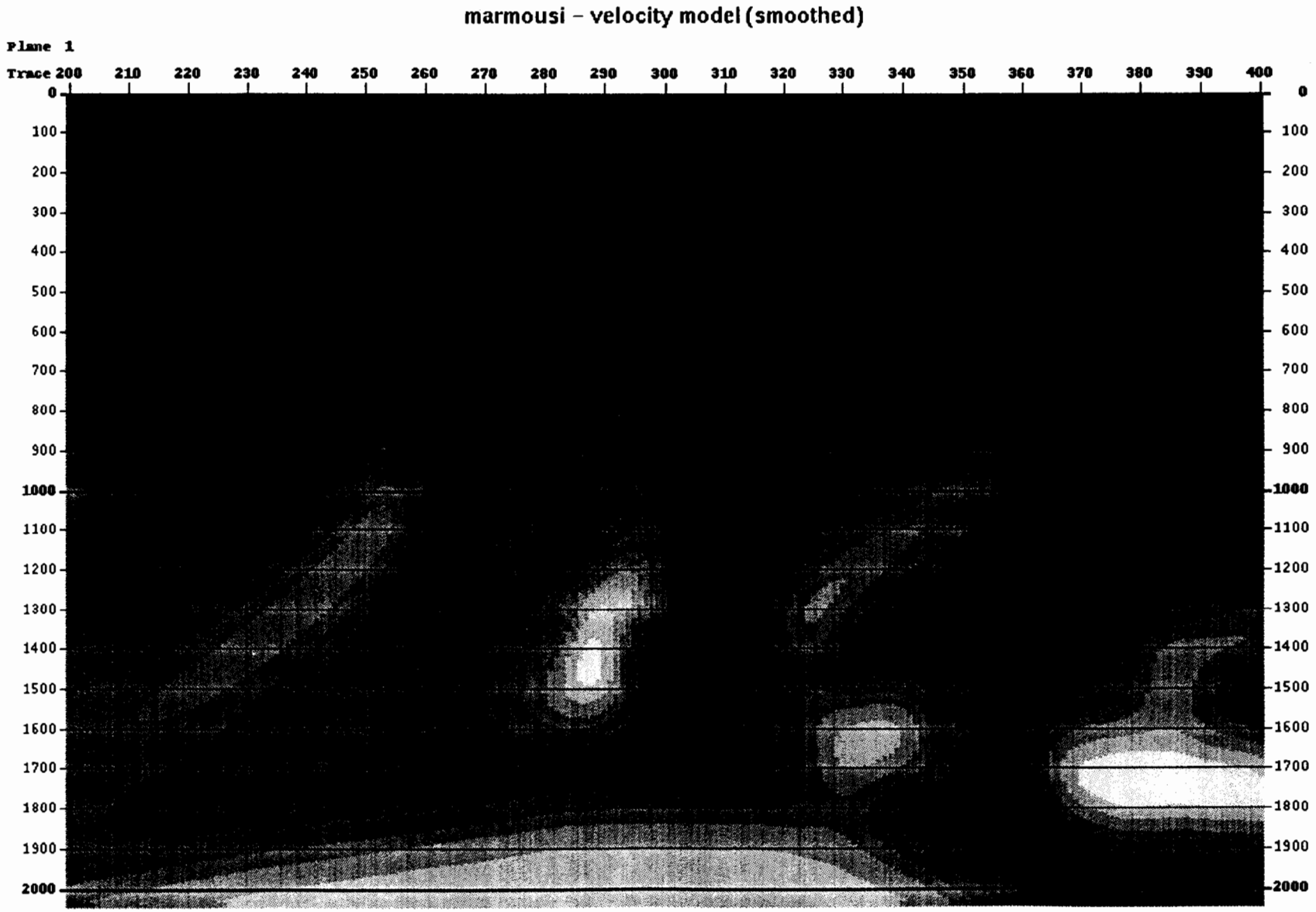
marmousi – velocity model (smoothed)
Plane 1
Trace 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000

Figure 6(b). Grey scale plot of the oscillatory part (nondimensional reflectivity) of the Marmousi velocity, obtained by subtracting from the original model the smoothed model displayed in Figure 2(a). Lightest areas are most positive (0.37), darkest areasmost negative (-0.22).

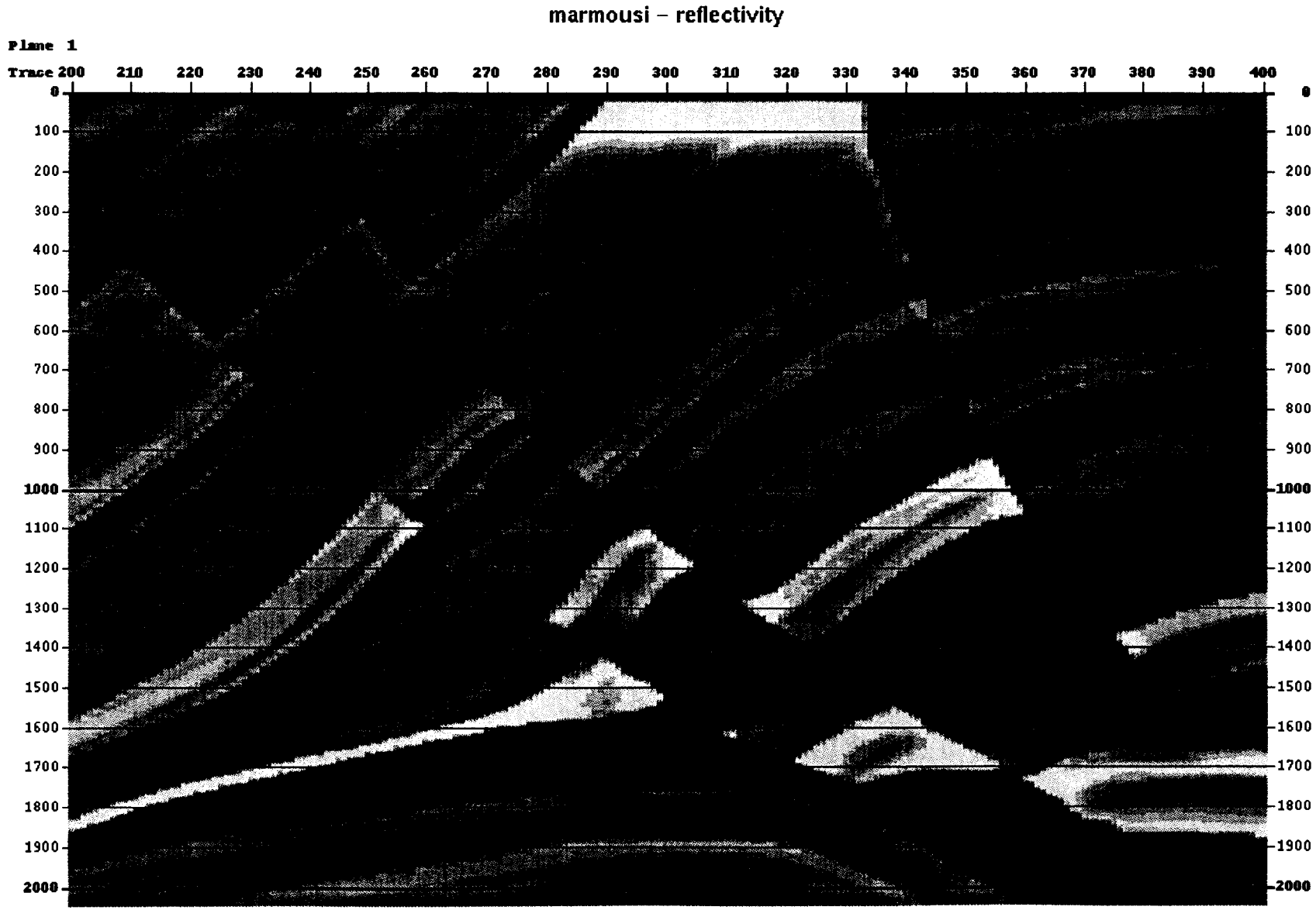
marmousi – reflectivity
Plane 1
Trace 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370 380 390 400
0 100 200 300 400 500 600 700 800 900 1000 1100 1200 1300 1400 1500 1600 1700 1800 1900 2000

Figure 6.b

Figure 7. Common source data set for shotpoints at 4000m, 5000m, 6000m and 7000m from west edge of model (Figure 6). The source is an array of 6 point sources, spaced 6m apart and 8m deep, with 12.5 Hz Ricker time dependence. The receiving array consists of 96 point receivers with an interval of 25m and a near offset of 150m. A mute with zero offset intercept 200m/s and velocity 1400m/s has been applied.

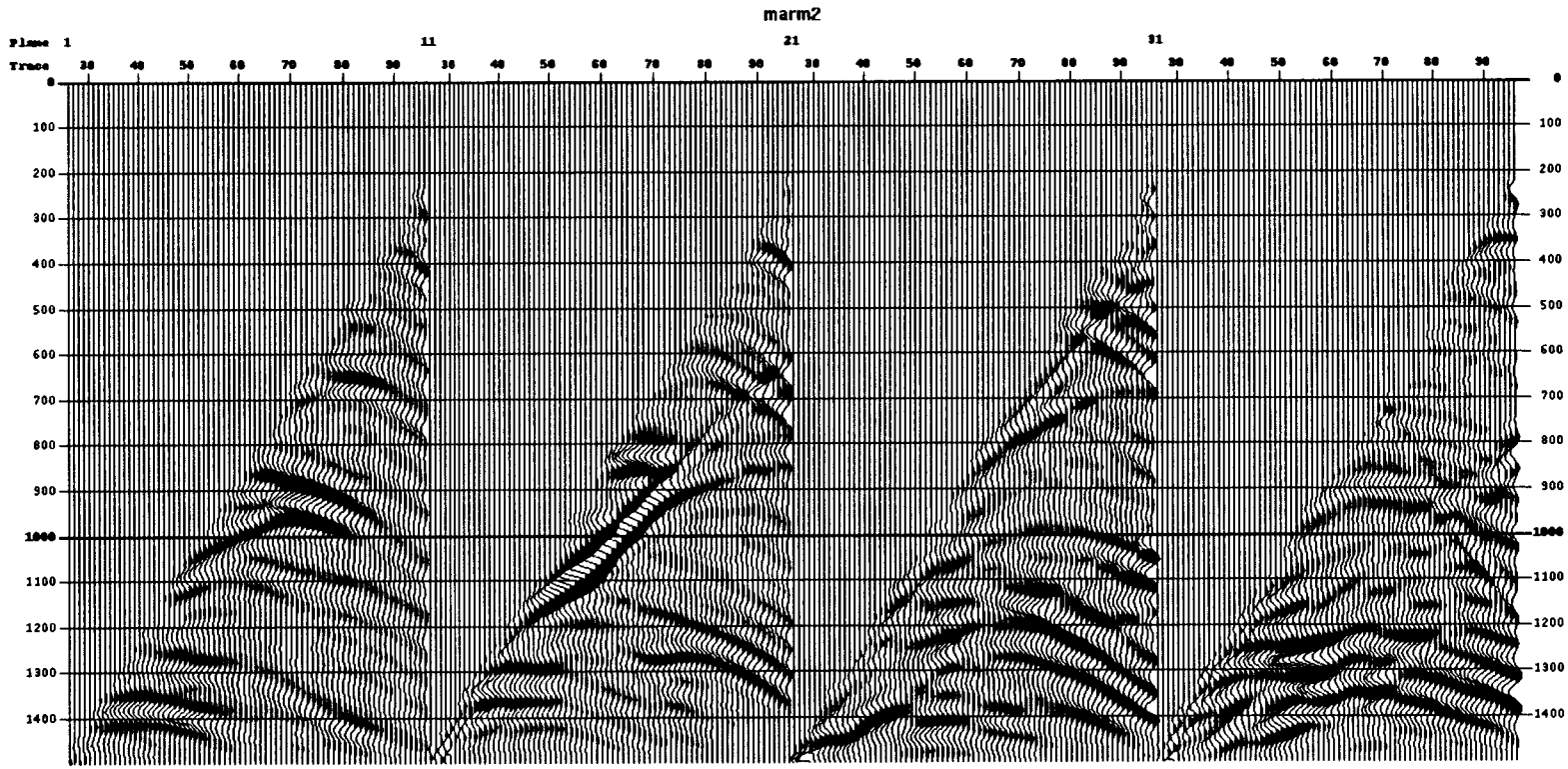
marm2
Plane 1
11
21
31
Trace
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
1400

Figure 8. The value $J[v_h]$ of the DSO criterion plotted against h. Here $v_h = 1500(1 - h) + hv_b$ so that the target model (used to generate the data) is placed at $h = 1$. The top (solid) curve is $J[v_h]$. The dashed curve is the mean-square error component of $J[v_h]$, and the dotted curve is the differential semblance component (that is, $\|\partial r/\partial x_s\|^2$ with r evaluated at v_h). The sample points in h were $h = 0.0$, 0.25, 0.5, 0.625, 0.75, 0.875, 1.0, and 1.125.

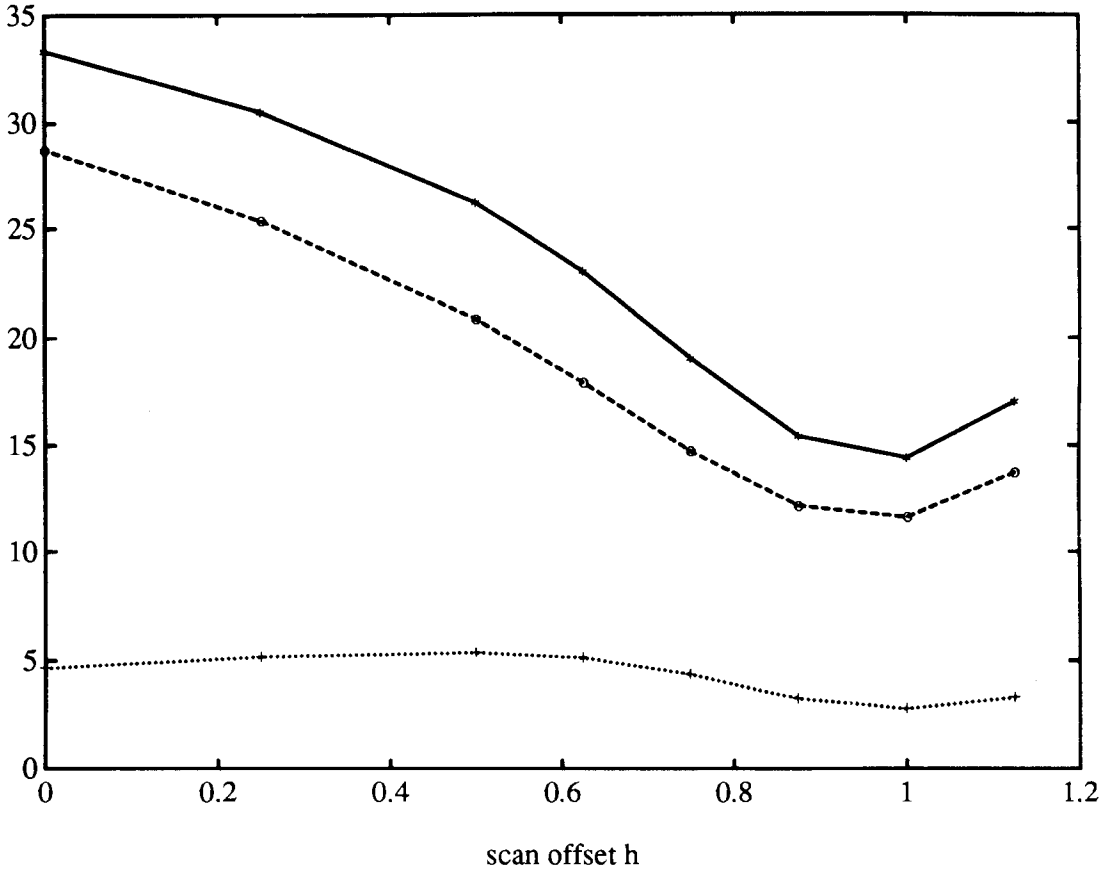

Figure 8