STATISTICAL INVERSION FROM REFLECTIONS OF SPHERICAL WAVES BY A RANDOMLY LAYERED MEDIUM

M. ASCH ^{*}, W. KOHLER [†], G. PAPANICOLAOU [‡], M. POSTEL [§] AND B. WHITE [¶]

1. Introduction. Waves reflected by randomly layered media are very noisy because of the intense multiple scattering they undergo before returning to the surface. As a result, it is difficult to extract useful information about the medium from observations of reflected waves. In [1] and the references cited there we showed that if the large scale variations of medium properties, such as density and sound speed, can be distinguished well from their small scale fluctuations due to the layering, then it is possible to use reflection data to solve some statistical inverse problems. More specifically, the statistical properties of reflected signals arising from plane or spherical wave sound pulses incident on such a random medium can be calculated accurately when the pulse width is intermediate between the two scales characterizing the medium properties, with the ratio between the two scales acting as a small parameter in an asymptotic analysis. Within the framework of this three-scale theory (where the scale of macroscopic variations is much larger than the pulse width which in turn is much larger than the scale of the random layering) we formulated and solved in [1] and [2] statistical inverse problems for plane wave pulses. That is, from the reflected acoustic pressure or velocity measured at the surface of the randomly layered medium we recovered the large scale variations in the medium properties.

In the plane wave case it is necessary to have available several independent realizations of reflected signals in order to compensate for their very noisy structure. We did this with numerical simulations in [1] and [2]. However, for a strictly layered medium it is unrealistic to expect to have available more than one realization of the reflected signal. One can argue that plane waves and layered media are useful idealizations that in reality are confined to large but finite regions and that independent realizations correspond to reflections from statistically similar topographies in nearby locations. This kind of argument is appealing but very difficult to substantiate even in a qualitative way.

The purpose of this paper is to show that the statistical inversions of [1] and [2] can be carried out with a single set of measurements that arise from a spherical pulse emitted by a point source. Instead of multiple independent realizations we now use the observed signals at different locations on the surface of the randomly layered half space in order to compensate for the noise generated by the multiple scattering. The asymptotic theory of the forward problem suggests that this should be possible but the point source reflection problem is sufficiently complicated that even with numerically generated reflected signals that are carefully controlled so as to conform to the theory it is far from clear that inversion can be implemented. The inversion algorithm for point source reflection data given here is not a simple extension of the one in the plane

^{*} Laboratoire d'Analyse Numérique, Université Paris XI, 91 405 Orsay, France

[†] Department of Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061

[‡] Department of Mathematics, Stanford University, Stanford CA 94305

[§] Laboratoire d'Analyse Numérique, UPMC et CNRS UA 189, Paris 75252 Cedex, France

[¶] Exxon Research and Engineering Company, Route 22 East, Annandale, NJ 08801

wave case because arrival times are not related to penetration depths in a direct way. Moreover, most of the detailed decisions in the implementation of the algorithm are based on our experience with extensive numerical simulations. The theory, important as it is in our whole approach, cannot give more than general guidance. This work is, therefore, a successful, carefully implemented feasibility study of inversions from point source reflection data. It is not, at this stage, the presentation of a polished and robust algorithm for inversion.

Our study is based entirely on numerically generated reflected signals from a layered random medium whose statistical properties are within the scope of the three-scale model described briefly above. What about real signals as they arise in reflection seismology, for example? There is evidence [3] that the three scale model is compatible with well log data and that a length scale of about 3 to 6 meters can be assigned to the layering for a variety of topographies. If the macroscopic properties vary on a scale of hundreds of meters and a nominal sound speed is 3km/sec then a pulse with frequency content in the 20-40 Hz range has a width of about 100 meters so the three scale model is not a bad idealization. Intrinsic dissipation is important, of course, but is neglected here. It can, however, be incorporated in some form and the theory for it has been worked out in [1]. Other effects, such as departure from perfect layering and full elastic wave propagation are more difficult to model and to simulate and remain to be worked out before inversions can be attempted. Another important issue that we do not address is the identification of embedded discontinuities (or sharp changes) in the medium properties, along with the slowly varying properties. We assume here that the macroscopic properties of the medium are smooth. Discontinuities require additional theoretical developments, some of which are related to the O'Doherty-Anstey theory [4,5,6,7,8], as well as extensions of the statistical methodology used in the inversion algorithm that have to be worked out.

To make the paper reasonably self contained and to emphasize numerical aspects we review briefly in section 2 plane wave inversions. We summarize the necessary theory that relates the local power spectral density of the reflected signals to the mean sound speed profile and then formulate the inversion as a maximum likelihood problem for the estimated local power spectral density. The maximization problem can be solved for one macroscopic segment of the sound speed profile at a time, in the plane wave case. This layer peeling works well and efficiently but does not generalize to the point source case. In Appendix C, we compare it to a global optimization method to gain some insight into the loss of efficiency that is inevitable in the point source case where a combined layer peeling-optimization approach is used.

In section 3 we review briefly the necessary theory of the forward problem for spherical pulse reflection by randomly layered media. The local power spectral density is again related to the mean sound speed profile by a set of equations that are now much harder to solve than in the plane wave case. We do not, in fact, solve them in our inversion algorithm but use only a low frequency (or weak fluctuation) approximation. This makes the computations in the inversion manageable but limits severely the effectiveness of the theory. Our inversion results would surely be much better if we had used the full equations relating the local power spectral density and the mean sound speed profile rather than the low frequency approximation. In this section we also describe briefly the code that generates reflected signals from a pulsed point source with sufficient accuracy to capture their statistical properties correctly.

Section 4 contains the new results which show that the mean sound speed profile

can be recovered reasonably well from a single set of measurements of reflected signals, taken at different observation points on the surface of the randomly layered half space on which a pulsed spherical wave is incident. We first describe the inversion algorithm precisely, which is a maximum likelihood method as in the plane wave case that is carried out by a combination of layer peeling and global maximization. This is done iteratively with only single variable maximizations. We then describe carefully the numerically generated data and explain how they must be used so as to comply with the low frequency approximation that relates the power spectral density with the mean sound speed profile. Before showing the results on three different sets of reflected data we discuss two important modifications of the algorithm that are simple but decisive. One is a multigrid iteration for the estimation of the mean sound speed profile, starting from a crude guess and refining it successively. The other is the imposition of some reasonable bounds on the variations of the mean sound speed profile in the likelihood maximization steps.

Inversion from point source reflection data is a very hard problem and it is surprising that it can be done at all. We note again that our results are only a feasibility study, which involves several simplifications so as to make it tractable.

2. Plane wave inversion. We begin with a brief review of the forward problem for the reflection of plane wave pulses normally incident on a randomly layered half space, $z < 0$, [9]. The acoustic pressure $p(t, z)$ and velocity $u(t, z)$ satisfy the continuity and momentum equations

$$(2.1) \quad \begin{aligned} \frac{1}{K} p_t + u_z &= 0 \\ \rho u_t + p_z &= 0 \end{aligned}$$

Here ρ is the material density and K the bulk modulus. As in [1] we assume for simplicity that the density has no random variation

$$(2.2) \quad \rho(z) = \begin{cases} \rho_0, & z > 0, \\ \rho_1, & z < 0 \end{cases}$$

with ρ_0 and ρ_1 constants. For the bulk modulus we assume that

$$(2.3) \quad K^{-1}(z) = \begin{cases} K_0^{-1}, & z > 0, \\ K_1^{-1}(z) \left(1 + \nu\left(\frac{z}{\epsilon^2}\right)\right), & z < 0 \end{cases}$$

with K_0 a constant, $K_1(z)$ a smooth deterministic function of z and $\nu(s)$ a bounded stationary random function with mean zero, representing the fluctuations in K^{-1} . Note that they vary on the scale ϵ^2 , where ϵ is a small parameter. If z is measured in kilometers and the fluctuations vary on the scale of a few meters then a value of ϵ around 0.05 captures the scale separation we wish to model. We assume that the random function $\nu(s)$ has a correlation length of order one so that the correlation length of $\nu(z/\epsilon^2)$ is of order ϵ^2 in kilometers (about 2.5 meters for $\epsilon = 0.05$). The mean sound speed c is given by

$$(2.4) \quad c(z) = \begin{cases} c_0 = \sqrt{\frac{K_0}{\rho_0}}, & z > 0 \\ \bar{c}(z) = \sqrt{\frac{K_1(z)}{\rho_1}}, & z < 0 \end{cases}$$

For $t < 0$ a normally incident plane wave solution in $z > 0$ has the form

$$(2.5) \quad \begin{aligned} u(t, z) &= \frac{1}{\sqrt{\epsilon}} \frac{1}{\sqrt{\rho_0 c_0}} f\left(\frac{t + z/c_0}{\epsilon}\right) \\ p(t, z) &= -\frac{1}{\sqrt{\epsilon}} \sqrt{\rho_0 c_0} f\left(\frac{t + z/c_0}{\epsilon}\right) \end{aligned}$$

Here f is the pulse shape function which is assumed to vanish for negative arguments and to have support that is of order one in the macroscopic t units that are seconds. With $\epsilon = 0.05$, the pulse width is about 50 *msec* or, with a speed of 3*Km/sec*, 150 meters. The multiplicative factor $1/\sqrt{\epsilon}$ in (2.5) makes the total energy of the incident plane wave pulse independent of ϵ . Continuity of p and u at the interface $z = 0$ makes (2.1) and (2.5) a complete problem. We are interested in $p(t, 0)$ or $u(t, 0)$ for $t > 0$ and this involves the solution of a complicated random scattering problem because of the form (2.3) of K^{-1} .

We will consider the reflected pressure $p_{refl}(t, 0)$, at $z = 0$ and $t > 0$, which is the total pressure minus the incident pressure (2.5). After a time of order ϵ , the duration of the incident pulse, the two are the same. Of particular interest is the two-time reflected average pressure intensity.

$$(2.6) \quad I(t, \bar{t}) = \frac{1}{\rho_0 c_0} \langle p_{refl}(t + \frac{\epsilon \bar{t}}{2}, 0) p_{refl}(t - \frac{\epsilon \bar{t}}{2}, 0) \rangle$$

with the angular brackets denoting statistical average. The factor $1/\rho_0 c_0$ is a normalization.

For simplicity, we will assume in the sequel that there is no macroscopic discontinuity at $z = 0$ so that $\rho_0 = \rho_1$ and $K_0^{-1} = K_1^{-1}(0)$.

Note that the time offset in (2.6) is proportional to the pulse width ϵ . The reason for this is that for time offsets of more than a few pulse widths the reflected signals are essentially uncorrelated. Moreover, in the absence of discontinuities in the medium, $\langle p_{refl}(t, 0) \rangle$ is essentially zero except for a time of order ϵ near $t = 0$ when the reflection from the interface $z = 0$ is felt.

Fix a $t > 0$, not close to zero, and a small ϵ . Since $I(t, \bar{t})$ is essentially zero for large $\bar{t}$ we can introduce its (essentially local) Fourier transform

$$(2.7) \quad \Lambda(t, \omega) |\hat{f}(\omega)|^2 = \int e^{i\omega \bar{t}} I(t, \bar{t}) d\bar{t}$$

in which Λ is the normalized local power spectral density. The normalization is $|\hat{f}(\omega)|^2$ with $\hat{f}(\omega)$ the Fourier transform of the pulse shape function $f(t)$. The two-time intensity function can be written as

$$(2.8) \quad I(t, \bar{t}) = \frac{1}{2\pi} \int |\hat{f}(\omega)|^2 \Lambda(t, \omega) e^{-i\omega \bar{t}} d\omega$$

The main thrust of our theoretical work in [9], [1] is that in the limit $\epsilon \rightarrow 0$ the local power spectral density can be calculated by solving a system of partial differential equations where

$$(2.9) \quad \Lambda(t, \omega) = W^1(0, t, \omega)$$

and the $W^N(z, t, \omega)$, $N \geq 0$ satisfy the equations

$$(2.10) \quad \frac{\partial W^N}{\partial z} + \frac{2N}{\bar{c}(z)} \frac{\partial W^N}{\partial t} - \frac{2\alpha\omega^2 N^2}{\bar{c}^2(z)} \{W^{N+1} - 2W^N + W^{N-1}\} = 0$$

for $-L < z \leq 0$, with

$$(2.11) \quad W^N(-L, t, \omega) = \delta(t)\delta_{N,0}$$

Here the mean sound speed $\bar{c}(z)$ is given by (2.4) and $\alpha > 0$ is the noise intensity level of the fluctuations

$$(2.12) \quad \alpha = \frac{1}{4} \int_0^\infty \langle \nu(s)\nu(0) \rangle ds$$

The length L is arbitrary, provided that for any given $t > 0$ for which we want to calculate $\Lambda(t, \omega)$ it satisfies

$$(2.13) \quad L > c_{max} \frac{t}{2}$$

with c_{max} the maximum speed $\bar{c}(z)$ in $z \leq 0$. Because of the hyperbolic nature of the equations (2.10) it is easy to see (and explained in the references) that the choice of L satisfying (2.13) does not affect $\Lambda(t, \omega)$ given by (2.9).

There is no quick and simple way to explain the result (2.9)-(2.13) that relates the local power spectral density $\Lambda(t, \omega; \bar{c}(\cdot))$, the mean sound speed profile $\bar{c}(z)$ and the noise intensity level α . But we will now make several remarks that will help explain the nature of this relationship.

From the definition (2.6) and (2.7) it is clear that $\Lambda(t, \omega)$ is a local Fourier transform but it is not necessarily positive as it would have to be if $p_{refl}(t, 0)$, the reflected pressure, were a stationary process in t so that $I(t, \bar{t})$ were independent of t . However, in the limit $\epsilon \rightarrow 0$, and hence when ϵ is small, $\Lambda(t, \omega)$ given by (2.9)-(2.13) is indeed positive. For a general profile $\bar{c}(z)$ it cannot be computed explicitly but for $\bar{c}(z) = \bar{c}$, a constant, it has the form

$$(2.14) \quad \Lambda(t, \omega) = \frac{\frac{\alpha\omega^2}{\bar{c}}}{\left(1 + \frac{\alpha\omega^2}{\bar{c}}t\right)^2}$$

In terms of the localization length [10] at frequency ω

$$(2.15) \quad l(\omega) = \frac{\bar{c}^2}{2\alpha\omega^2}$$

we can write (2.14) in the form

$$(2.16) \quad \Lambda(t, \omega) = \frac{1}{2} \frac{\bar{c}l(\omega)}{\left(l(\omega) + \frac{\bar{c}t}{2}\right)^2}$$

As shown in [10], the localization length at frequency ω is a measure of the depth of penetration of a time harmonic wave with this frequency into a randomly layered medium with uniform sound speed $\bar{c}$ and noise level α for the fluctuations. Wave

energy does not penetrate much below this length. In the time domain, the local power spectral density at a fixed time t , $\Lambda(t, \omega)$ in (2.16), has a maximum $\omega_{max} = \omega_{max}(t)$ that depends on time. From (2.16) the maximum is calculated to be

$$(2.17) \quad \omega_{max} = \sqrt{\frac{\bar{c}}{\alpha t}}$$

In a more physical way [10] we can say that the maximum of the local power spectral density at time t occurs for that $\omega = \omega_{max}$ for which

$$(2.18) \quad l(\omega_{max}) = \frac{\bar{c}t}{2}$$

Thus, for the frequency for which the localization length equals the mean travel time we have the maximum contribution to the noise spectrum of the reflected signal.

It is because of wave localization and its manifestations in the time domain described above that signals reflected by randomly layered media are so noisy. We can interpret (2.10) as a hierarchy of equations for moments associated with the scattering problem [1]. The infinite hierarchy ($N \geq 0$ in (2.10)) indicates that the second moment that we are interested in (I of (2.6) or Λ of (2.7)) cannot be computed separately from all higher moments (the W^N , $N > 2$ in (2.10)). This is another manifestation of localization.

We can describe briefly how the mean sound speed profile $\bar{c}(z)$ can be estimated from observations of $p_{refl}(t, 0)$ or

$$(2.19) \quad R_f(t) = \frac{1}{\sqrt{\rho_0 c_0}} p_{refl}(t, 0)$$

in which dependence of the pulse shape function f is indicated. Our inversion strategy is based on one more fact about the reflected signal $R_f(t)$, in addition to (2.9)-(2.13). It is that as ϵ tends to zero $R_f(t)$ becomes approximately a Gaussian process. We have not been able to prove this but we have some good heuristic indications that it is true [9] and extensive numerical simulations that corroborate it very well [11] [1]. From the Gaussian property of $R_f(t)$ we conclude that

$$(2.20) \quad \frac{1}{|\hat{f}(\omega)|^2} \int e^{i\omega \bar{t}} R_f \left(t + \frac{\epsilon \bar{t}}{2} \right) R_f \left(t - \frac{\epsilon \bar{t}}{2} \right) d\bar{t} = \hat{\Lambda}(t, \omega)$$

is approximately, when ϵ is small, an exponential random variable with mean $\Lambda(t, \omega)$ given by (2.9)-(2.13), when $\bar{c}(z)$ is known. Moreover, for distinct $0 < t_1 < t_2 < \dots < t_{N_t}$ and $0 < \omega_1 < \omega_2 < \dots < \omega_{N_f}$, where N_t and N_f are integers, the random variables $\{\hat{\Lambda}(t_j, \omega_l)\}$ are independent with exponential distribution having mean $\{\Lambda(t_j, \omega_l)\}$.

Our inversion strategy is now this: Depending on the available data, fix a set of time points $\{t_j\}$ and frequencies points $\{\omega_l\}$ as above. For each realization of $R_f(t)$ that is available, estimate $\hat{\Lambda}(t_j, \omega_l)$ from (2.20). This is actually a very delicate step that must be done carefully as we discuss in [1], Appendix E. Then form

$$(2.21) \quad O(\bar{c}) = \prod_{realiz} \prod_{j=1}^{N_t} \prod_{l=1}^{N_f} \frac{e^{-\hat{\Lambda}(t_j, \omega_l) / \Lambda(t_j, \omega_l; \bar{c}(.))}}{\Lambda(t_j, \omega_l; \bar{c}(.))}$$

where the first product is over different independent realizations. This is the likelihood functional for the estimates $\hat{\Lambda}$, given a known mean speed profile $\bar{c}(z)$. We now choose $\bar{c}(z)$ in order to maximize this functional. This is a rather usual maximum likelihood estimation *except* that now the maximization must be done over the profiles $\bar{c}(\cdot)$ which in turn determine $\Lambda(t, \omega; \bar{c}(\cdot))$ in (2.21) via the partial differential equations (2.10)-(2.11) and the relation (2.9).

The most convenient way to solve the maximization problem for (2.21), and thus estimate $\bar{c}(z)$, is to assume that it is piece-wise linear over a few macroscopically large layers and then maximize O over a finite set of speeds $\bar{c}^1, \bar{c}^2, \dots, \bar{c}^{N_z}$. These speeds are approximations of $\bar{c}(z)$ at successively larger depths numbered from 1 to N_z . Moreover, because of the hyperbolic nature of (2.10)-(2.11) the maximization can be done one layer at a time with increasing depth. This avoids the difficult problem of finding the maximum of a complicated function of several variables. Physically this layer peeling process makes sense because there is a direct relation between the sound speed profile up to a certain depth and the smallest time before which the rest of the medium is not felt in the reflected signal $R_f(t)$.

In the point source case described in the next section and in section 4 this layer peeling approach, that is quite successful for plane waves ([2],[12], [1]), has no simple extension and some form of multivariate (or global) maximization is necessary. In Appendix C we compare layer peeling to the global maximization algorithm described in section 4 but restricted to the plane wave case. This provides a benchmark for the performance of the algorithm in the one case where another well-studied method is available.

3. The forward problem for spherical waves. In this section we will summarize the basic facts from [1] that we use in our inversions from point source data described in the next section

The three dimensional acoustic equations are

$$(3.1) \quad \begin{aligned} \rho \frac{\partial \mathbf{u}}{\partial t} + \nabla p &= \epsilon^{1/2} f\left(\frac{t}{\epsilon}\right) \delta(x) \delta(y) \delta(z - z_s) \mathbf{e} \\ K^{-1} \frac{\partial p}{\partial t} + \nabla \cdot \mathbf{u} &= 0 \end{aligned}$$

where ρ and K are given by (2.2) and (2.3) and $p(t, x, y, z)$, $\mathbf{u}(t, x, y, z)$ are the acoustic pressure and velocity fields. In the momentum equations (3.1) the right hand side is a pulsed point source, which is in the homogeneous medium $z > 0$ when $z_s > 0$, as we assume. The unit vector $\mathbf{e}$ is the source directivity and we take it vertical $\mathbf{e} = (0, 0, 1)$ as in [1]. As in the plane wave case the pulse width is of order ϵ and the multiplicative factor $\epsilon^{1/2}$ in the source term in (3.1) makes the total energy released by the pulsed point source independent of ϵ . The small parameter $\epsilon > 0$ has the same meaning as in the plane wave case. We are dealing with a three scale model problem as described in section 2.

We denote by $p_{refl}(t, \mathbf{x}, 0)$ the reflected pressure at the interface $z = 0$, with $\mathbf{x} = (x, y)$ the horizontal coordinates. It is the pressure field $p(t, \mathbf{x}, 0)$ minus the incident pressure. The latter is the field that would result at $(\mathbf{x}, 0)$ if the medium was uniform. Since the pulse width is of order ϵ , the pressures p , and p_{refl} are the same after an interval of order ϵ past the arrival time at $(\mathbf{x}, 0)$. The quantity of interest is

the two-time intensity function

$$(3.2) \quad I(t, \mathbf{x}, \bar{t}) = \lim_{\epsilon \downarrow 0} \left\langle p_{\text{refl}} \left(t + \frac{\epsilon \bar{t}}{2}, \mathbf{x}, 0 \right) p_{\text{refl}} \left(t - \frac{\epsilon \bar{t}}{2}, \mathbf{x}, 0 \right) \right\rangle$$

and, as in the plane wave case (2.7) we define the normalized local power spectral density $\Lambda(t, \mathbf{x}, \omega)$ by

$$(3.3) \quad \Lambda(t, \mathbf{x}, \omega) \omega^2 |\hat{f}(\omega)|^2 = \int I(t, \mathbf{x}, \bar{t}) e^{i\omega \bar{t}} d\bar{t}.$$

The angular brackets denote statistical averaging and the offset in time in (3.2) is proportional to the pulse width ϵ . The intensity I vanishes when $\bar{t}$ is large so that (3.3) is a local Fourier transform. The scaling of Λ by the Fourier transform of the pulse shape function $\hat{f}(\omega)$ differs in (3.3) from (2.7) because we are in the point source case ([1], section 2.4).

Let us assume as in the previous section that the mean medium is matched at $z = 0$: $\rho_0 = \rho_1$ and $K_0 = K_1(0)$. Then the main thrust of [1] is the following result that relates the local power spectral density $\Lambda(t, \mathbf{x}, \omega)$ and the mean speed profile $\bar{c}(z)$

For $N \geq 0$, $\mathbf{x} \in R^2$, $\boldsymbol{\kappa} \in R^2$, $\kappa = |\boldsymbol{\kappa}|$, $-L < z \leq 0$, $t \in R^1$ and $\omega \in R^1$, let $W^N(z, t, \mathbf{x}, \boldsymbol{\kappa}, \omega)$ be the solution of the system of partial differential equations

$$(3.4) \quad \begin{aligned} \frac{\partial W^N}{\partial z} + 2N \left\{ \frac{1}{\bar{c}(1 - \bar{c}^2 \kappa^2)^{1/2}} \frac{\partial W^N}{\partial t} + \frac{\bar{c} \boldsymbol{\kappa}}{(1 - \bar{c}^2 \kappa^2)^{1/2}} \cdot \nabla_{\mathbf{x}} W^N \right\} \\ - 2 \frac{\omega^2 \alpha N^2}{\bar{c}^2 (1 - \bar{c}^2 \kappa^2)} \left\{ W^{N+1} - 2W^N + W^{N-1} \right\} = 0, \end{aligned}$$

$$(3.5) \quad W^N|_{z=-L} = \delta_{N0} \delta(t) \delta(\mathbf{x}),$$

Then

$$(3.6) \quad \Lambda(t, \mathbf{x}, \omega) = \frac{1}{4(2\pi)^2} \iint_{R^2} d\boldsymbol{\kappa} W^1(0, t, \mathbf{x}, \boldsymbol{\kappa}, \omega).$$

Compared to the plane wave case (2.9)-(2.11), the determination of the local power spectral density Λ is now much more complicated because W^1 depends on the slowness vector $\boldsymbol{\kappa}$ ($|\boldsymbol{\kappa}|$ has the dimension of $(\text{speed})^{-1}$) and after solving the system (3.4)-(3.5) we must integrate over $\boldsymbol{\kappa}$ to get Λ . This is nothing more than a plane wave decomposition of a spherical problem that was followed in [1]. It is important to note also that as $\kappa = |\boldsymbol{\kappa}|$ varies in (3.4) the factor $1 - \bar{c}^2(z) \kappa^2$ may vanish for some z , making (3.4) singular. We showed in [1] that (3.4)-(3.5) is well posed in general despite such singularities that correspond to turning points in the plane wave decomposition. However the numerical solution of (3.4)-(3.5) is much more complicated now and the singularities are part of the problem.

We should also note here that the depth L at which (3.5) is imposed does not affect the local power spectral density Λ in (3.6) provided that again condition (2.13) holds. This is shown in [1] and the demonstration relies on the adjoint system to (3.4)-(3.5) that is introduced in Appendix A.

The system adjoint to (3.4)-(3.5) that is presented in Appendix A is more convenient for numerical computations but that is not easily done for variable mean speed

profiles $\bar{c}(z)$. In our inversion algorithm we use the small ω approximation of $\Lambda(t, \mathbf{x}, \omega)$ which is derived in detail in Appendix B. As mentioned in the introduction, an efficient way to solve for Λ given $\bar{c}$, either from (3.4)-(3.5) or from the adjoint system (A.9)-(A.11) would greatly enhance our approach to inversions.

In [1] we carried out extensive numerical simulations that compare directly the statistics of solutions of (3.1) with the theoretical prediction for Λ given by (3.6). The code that solves directly the scattering problem (3.1) is presented in detail in section 4 of [1]. It is a "layer code" which means that it is based on successive multiplication of transfer matrices over the microscopic layers of the scattering medium, followed by Fourier synthesis over the horizontal slowness $\kappa \in R^2$ and the frequency $\omega \in R^1$.

In Figures (3.1) to (3.3) we compare estimates¹ $\hat{\Lambda}$ averaged over 20 realizations of the local power spectral density $\Lambda(t, r, \omega)$, $r = |\mathbf{x}|$, which is the solution of (3.4)-(3.5) when $\bar{c}(z) = \bar{c}$, a constant, and the medium is matched at $z = 0$: $\rho_0 = \rho_1$ and $K_0 = K_1(0)$. In this case Λ is given by the following explicit formula ([1] p. 549)

$$(3.7) \quad \Lambda(t, r, \omega) = \frac{1}{4(2\pi)^2} \frac{1}{\bar{c}(\bar{c}t)^2} \frac{\frac{\alpha\omega^2}{\bar{c}^2(1-(\frac{r}{\bar{c}t})^2)^{1/2}}}{\left[1 + \frac{\alpha\omega^2}{\bar{c}^2(1-(\frac{r}{\bar{c}t})^2)^{1/2}}\bar{c}t\right]^2}.$$

which is the point source analog of (2.14). The parameters used for the simulations are described in detail in section 4.4. The signal is generated by sending a spherical pulse of about 50 ms duration into a slab of 1600 layers each 2.5 meters thick. The point source is at the interface $z = 0$. The reflected signals are recorded at 20 receivers spaced 200 meters apart, starting at 200 meters from the source. The first arrival of the reflected signal is clearly detected at the right time but its amplitude is unreliable. On the other hand the tail of the spectrum behaves as predicted by the theory. The bottom of the slab is felt around $t = 8.3$ sec for the nearest receiver which is the homogeneous two way travel time $2 \times 1600 \times 0.05^2$. After this time the theory and the computed statistics should, and do, disagree.

4. Inversion with spherical wave data.

4.1. Description of the algorithm. From reflected signals recorded at different offsets from a point source, which is located at the interface $z = 0$, we estimate the local power spectral density on a discrete set of points in time, frequency and offset. The points are chosen so that the estimates are independent random variables and we know from the theory and the numerical simulations that they are exponentially distributed. The exponential distribution and independence of the estimates of the local power spectral density for widely spaced time and frequency windows was seen clearly in the plane wave case in [1] (Section 6 and Appendix E) and [13]. In these two papers, moreover, we present algorithms to numerically estimate (2.7) based either on a windowed Fourier transform or on a wavelet transform.

In this paper estimates of the local power spectral density are obtained using the windowed Fourier transform algorithm of [1] directly on the point source reflection data. A new feature now is that for widely spaced receivers the measurements are also independent. We can therefore replace the averaging of the estimates over measurements made with independent realizations of the random medium by measurements

¹ Estimated from the direct simulations according to the windowed Fourier transform method described in Appendix E of [1] or by some other local Fourier transform approach.

FIG. 3.1. The local power spectral density as a function of time for a fixed r and ω

of reflections from a single random medium made at different receiver locations along the surface.

We compare these estimates with the theory, that is, the local power spectral density that results from the transport equations (3.4)-(3.6) as described in the previous section and in Appendix B. To make this comparison into an inversion algorithm we introduce the likelihood functional for the estimates $\hat{\Lambda}$, given the mean sound speed profile $\bar{c}(z)$, in a manner similar to the one for plane waves in section 2. The main difference between the likelihood functional

$$\begin{aligned}
 (4.1) \quad & \text{Prob.dens.}\{\hat{\Lambda}(t_i, x_m, \omega_l), \quad m = 1, \dots, N_x; i = n_t^m, \dots, N_t; l = 1, \dots, N_\omega | \bar{c}(\cdot)\} \\
 &= \prod_{m=1}^{N_x} \prod_{i=n_t^m}^{N_t} \prod_{l=1}^{N_\omega} \frac{e^{-\hat{\Lambda}(t_i, x_m, \omega_l) / \Lambda(t_i, x_m, \omega_l; \bar{c}(\cdot))}}{\Lambda(t_i, x_m, \omega_l; \bar{c}(\cdot))}
 \end{aligned}$$

and the one for the plane wave case (2.21) is that the product over measurements made from independent realizations has been replaced in (4.1) by a product over measurements from different receivers. The limits in the three products in the likelihood function (4.1) must obey several criteria that will be described in detail later. Actually, it is more convenient in computations to work with the negative logarithm of the likelihood function which we call the objective function

$$\begin{aligned}
 (4.2) \quad O(\bar{c}(\cdot)) &= -\log[\text{Prob.dens.}\{\hat{\Lambda}(t_i, x_m, \omega_l), m = 1, \dots, N_x; i = n_t^m, \dots, N_t; l = 1, \dots, N_\omega | \bar{c}(\cdot)\}] \\
 &= \sum_{m=1}^{N_x} \sum_{i=n_t^m}^{N_t} \sum_{l=1}^{N_\omega} \left\{ \frac{\hat{\Lambda}(t_i, x_m, \omega_l)}{\Lambda(t_i, x_m, \omega_l; \bar{c}(\cdot))} + \log(\Lambda(t_i, x_m, \omega_l; \bar{c}(\cdot))) \right\}.
 \end{aligned}$$

The minimization of the log-likelihood function, the global inversion problem, is done over piecewise linear mean speed profiles that pass through the points $\{\bar{c}^k, k = 1, \dots, N_z\}$ that correspond to macroscopic layers. The mean power spectral densities in (4.2) must be computed using (3.4)-(3.6) for such piecewise linear mean speed profiles. The choice of the discretization points is not so simple here and will be discussed in detail in sections 4.2 and 4.5.

FIG. 3.2. The local power spectral density as a function of time for a fixed r and ω

In fact making this choice of macroscopic discretization points is a major difference between the point source and plane wave case that complicates matters. In the point source case there is no general link between the measurements of the reflected pressure sampled in the time domain and the unknown speed profile discretized on a given vertical grid. This can be seen from equations (A.12) and (A.14) where the dependence of travel-time on depth disappears once we are back to the level of the lateral space variables, after integration with respect to the slowness κ . Therefore, the simple layer peeling idea used for plane wave data cannot be used here. Nevertheless we would still like to use the hyperbolicity to somehow reduce the computations. We want to take advantage of the physical observation that the effects of a layer at depth z will be felt first at zero offset. From (A.12) we see that the travel time τ is monotonically increasing with slowness κ at fixed $z < 0$. Therefore there is a time, $\tau(z, 0)$, before which the effects below depth z are not felt. In other words, for each parameter of the piecewise linear profile $\bar{c}(z_k)$, $k = 1, \dots, N_z$ we can associate a time t_{N_k}

$$(4.3) \quad t_{N_k} \leq \tau(z_k, 0) < t_{N_k+1}$$

before which the quantity $\bar{c}^k$ has no influence on the local power spectral density for any $\mathbf{x}$ because W^1 is not influenced for any κ . In view of (A.14) this means that $\Lambda(t_i, x_m, \omega_l)$ does not depend on $\bar{c}^k$ for all m, l and all i such that $t_i < t_{N_k}$.

As mentioned in the introduction, we will not use equations (A.12)-(A.14) to compute the local power spectral density but rather the small frequency approximation² (B.6)-(B.7) derived in detail in Appendix B.

With these conditions on the time sampling that determine the macroscopic layers in the piecewise linear mean speed profile we are led to an inversion algorithm whose complexity is intermediate between global optimization and layer peeling. First we recall the standard relaxation method [14] for multivariate optimization. Let $\{\bar{c}_j^k, k = 1, \dots, N_z\}$ be the vector of the unknown parameters at the end of j th iteration. We

² This is a simplification that reduces the computations involved in solving repeatedly the equations (A.12)-(A.14) but limits the effectiveness of our method and we will try to remove it in future work.

FIG. 3.3. The local power spectral density as a function of time for a fixed r and ω

introduce N_z one-dimensional functions by

$$(4.4) \quad F_j^k(c) = O(\bar{c}_j^1, \dots, \bar{c}_j^{k-1}, c, \bar{c}_{j-1}^{k+1}, \dots, \bar{c}_{j-1}^{N_z})$$

For a given precision ζ we can find a discretized mean speed profile minimizing O with the following algorithm:

1. $j \leftarrow j + 1$
 - $k \leftarrow k + 1$
 - find $\bar{c}_j^k$ that maximizes $F_j^k(c)$
2. compare $\|\bar{c}_j - \bar{c}_{j-1}\| / \|\bar{c}_j\|$ and ζ and go back to the first step if necessary

We now use the observations we made above and note that maximizing $F_j^k(c)$ is equivalent to maximizing $G_j^k(c)$, which is the partial objective function where only the time steps where the spectrum depends on $\bar{c}_k$ are considered. In other words we include in the objective function only the time samples later than t_{N_k} defined by (4.3):

$$(4.5) \quad G_j^k(c) = \sum_{m=1}^{N_x} \sum_{i=N_k}^{N_t} \sum_{l=1}^{N_\omega} \left\{ \frac{\hat{\Lambda}(t_i, x_m, \omega_l)}{\Lambda(t_i, x_m, \omega_l; \bar{c}_j^1, \dots, \bar{c}_j^{k-1}, c, \bar{c}_{j-1}^{k+1}, \dots, \bar{c}_{j-1}^{N_z})} + \log(\Lambda(t_i, x_m, \omega_l; \bar{c}_j^1, \dots, \bar{c}_j^{k-1}, c, \bar{c}_{j-1}^{k+1}, \dots, \bar{c}_{j-1}^{N_z})) \right\}.$$

The Global Inversion algorithm can now be stated as follows. Let m be the receiver offset index varying from 1 to N_x , let l be the frequency index varying from 1 to N_ω and denote by $\{\Lambda_{n,m,l}^j\}$ the local power spectral density at points of the discretization that is, the solution of (3.4)-(3.6) for $(t_n, \mathbf{x}_m, \omega_l)$ using a mean speed profile intermediate between the discretized solutions at iteration $j-1$ and j in the algorithm below.

1. choose an initial profile $\bar{c}_0^1, \bar{c}_0^2, \dots, \bar{c}_0^{N_z}$ and set $j = 0$
2. until convergence do iteration j
 - (a) at step k we know
 - part of the solution $\bar{c}_j^n, n = 1, k-1$
 - $\Lambda_{n,m,l}^j$ for all $n < N_k$ which depend at most on $\bar{c}_j^p, p = 1, k-1$.
 - (b) for different trial values c of $\bar{c}_j^k$

- compute $\Lambda_{n,m,l}^j$ for all $n \geq N_k$ using $(\bar{c}_j^1, \dots, \bar{c}_j^{k-1}, c, \bar{c}_{j-1}^{k+1}, \dots, \bar{c}_{j-1}^{N_z})$ as test profile
 - repeat until we find the maximum of $G_j^k(c)$ defined in (4.5)
 - (c) once we have $\bar{c}_j^k$ we use (A.12) to find N_{k+1} such that $t_{N_{k+1}} \leq \tau(z_k, 0) < t_{N_{k+1}+1}$
 - (d) use the partial solution $\bar{c}_j^n, n = 1, k$ to compute
 - $\Lambda_{n,m,l}^j$ for all $n < N_{k+1}$ which depend at most on $\bar{c}_j^p, p = 1, k$.
 - (e) if $k < N_z$ start next step $k \leftarrow k + 1$ until N_z
3. test for overall convergence: if $\frac{\sqrt{\sum_{k=1}^{N_z} (\bar{c}_{j-1}^k - \bar{c}_j^k)^2}}{\sqrt{\sum_{k=1}^{N_z} \bar{c}_j^{k2}}} > \zeta$ then $j \leftarrow j + 1$

As noted above, $\{\Lambda_{n,m,l}^j\}$ should be computed by solving either (A.9)-(A.11) or (A.12)-(A.14) but we use here the small frequency solution, described in detail in Appendix B, and more precisely (B.6)-(B.7). Some drawbacks in using the low frequency approximation are described in section 4.3.

The computational advantages of using a relaxation method are many. It is most appropriate because it uses the hyperbolicity of the problem and also for other, less intrinsic reasons. It reduces a multivariate minimization problem to several single variable minimizations, which are much more stable problems. There are single variable minimization schemes which do not require differentiation of the function. The one we use, Brent's algorithm from Numerical Recipes [15], can handle both situations. However, efficient multivariate minimization schemes require knowing the gradient of the function. We have considered the possibility of computing derivatives of the objective function with respect to the discrete model parameters, the $\bar{c}_j^k$, by differentiating the transport equations and then solving them. In the plane case wave case, where a simpler algorithm is used, both methods have been tried. They give the same results and performance is similar. Using the optimization routine Dbrent of Numerical Recipes [15] the derivative method requires slightly fewer iterations but each one takes longer than a Brent iteration since it computes the derivative of the function. By iteration we mean solving the transport equation using a trial profile. In the point source case computing the gradient of the objective function with respect to the discretized mean speed profile is very cumbersome and costly and we avoid it.

4.2. Discretization of the unknown velocity profile. We are trying to recover a function of depth z using time-sampled data. In the plane wave case we solve this problem by first getting the mean sound speed profile as a function of time and then transforming it to a function of the z variable through the unique travel-time versus depth relation. This is explained in detail in Appendix C. For point source data there is no such relation because travel-time is now defined in the wavenumber domain and not in physical space so the unknown mean speed profile must be sought as a function of depth. However, if we choose a fixed depth sampling rate, the number of relevant time samples will vary with the value of the unknown speed. It might even happen, if the value of the velocity is large enough, that depth sampling will become too fine to be felt by the time samples available. To overcome this difficulty we let the unknown discrete mean velocity profile be given on a variable depth grid. This depth grid is known because it is chosen so that the number of time samples per depth sample is constant and, to fix ideas, we take it equal to N_{tz} at normal incidence $\kappa = 0$.

At the k -th step of iteration j we are looking for the value c that makes the profile

$c_j^k = (\bar{c}_j^1, \dots, \bar{c}_j^{k-1}, c, \bar{c}_{j-1}^{k+1}, \dots, \bar{c}_{j-1}^{N_z})$ maximize the function $G_j^k(c)$. For each trial value c of $\bar{c}_j^k$ we need to shift the remaining part of the profile to ensure that there is a constant number of time steps between two depth steps. For $n \geq k$ the values of $\bar{c}_j^n$ are not changed but those of z_n are. At $\kappa = 0$ we get directly the depth grid by integrating (A.12) when $\kappa = 0$. Using the piecewise linearity of $\bar{c}(z)$ we get

$$(4.6) \quad \begin{aligned} \tau_k &= \tau_{k-1} + 2 \frac{z_k - z_{k-1}}{\bar{c}_k - \bar{c}_{k-1}} \log \frac{\bar{c}_k}{\bar{c}_{k-1}} & \text{if } \bar{c}_k \neq \bar{c}_{k-1} \\ \tau_k &= \tau_{k-1} + 4 \frac{z_k - z_{k-1}}{\bar{c}_k + \bar{c}_{k-1}} & \text{otherwise} \end{aligned}$$

which, with δ_t equal to time-sampling interval, inverts into

$$\begin{aligned} z_k &= z_{k-1} + \frac{N_{tz}\delta_t}{2} \frac{\bar{c}_k - \bar{c}_{k-1}}{\log(\bar{c}_k/\bar{c}_{k-1})} & \text{if } \bar{c}_k \neq \bar{c}_{k-1} \\ z_k &= z_{k-1} + \frac{N_{tz}\delta_t}{4} (\bar{c}_k + \bar{c}_{k-1}) & \text{otherwise} \end{aligned}$$

The time steps that feel depth z_k are those for which $i \geq N_k$ with N_k such that $\delta_t(N_k - 1) \leq \tau_{k-1} < \delta_t N_k$. For these time steps we compute $\tilde{c}^i$ by interpolation between $\bar{c}^{n-1}$ and $\bar{c}^n$ with n such that $\tau_{n-1} \leq \delta_t i < \tau_n$.

Using the above method we compare in Figure (4.1) the two different discretizations of the unknowns. Suppose we are comparing the two linearly increasing profiles represented on the top Figure. The two other graphs show the discretization points both in depth and time domains. The superimposed horizontal lines show the time discretization of the data (here 0.4). In the middle Figure, we keep the travel time increment between two discretization points constant. The n -th unknown depends for both profiles on the same data points in time. On the other hand, in the bottom graph we keep the depth increment constant. We see for instance that the 10-th unknown located at abscissa $z = 2$ will depend for profile one on data points later than 3.2, and for profile two on data points later than 3.8 only. That is, two fewer time samples will enter the objective function.

4.3. Comments on the low frequency approximation. The small frequency approximation must be used carefully in the inversion scheme because the local power spectral density Λ is now computed with an iterative method. For a given test mean speed profile there might be, and there are, points in the (t, r, ω) space where the resolution of the characteristics (B.7) is insufficient. If we just drop these "bad" points and do not include them in the summation defining the objective function (4.5), it becomes discontinuous and, therefore, very hard to minimize. In fact, its minimum might be different from the theoretical minimum obtained using all the available data. We cannot improve much the resolution of the characteristics because some of the points that are hard to compute might be inaccessible, beyond a turning point for instance. We overcome this problem by interpolating between the solutions computed for the two nearest test profiles. This is a rough approximation but in practice it is applied on fewer than ten per cent of the points entering the summation. In any case, the error made with such points is much less than the value of the function itself and it is therefore better to include them with the approximate value than drop them entirely. There are also points for which the local power spectral density cannot be computed, if we start far from the true profile for example, but their contribution

should be negligible. This modification of the initial algorithm has been carefully tested and turns out to be very useful when the initial guess is far from the solution. In that case the algorithm has to try quite a few "unrealistic" profiles before getting near the solution and the objective function is very irregular. The approximation we use smoothens it, even if artificially. In the results we present in the last section we use a multigrid approach to determine an initial profile and we use this regularization only for the first steps that correspond to coarsely discretized profiles.

4.4. Description of the data. The inversion algorithm is applied to point source data generated by solving numerically the forward problem using a layer code. The details of the numerical simulation can be found in [[1],section 5]. The model consist of a stack of 3200 thin layers between two half-spaces of constant density and bulk modulus. The physical parameters are normalized by their value in these homogeneous half-spaces. In the stack, the dimensionless density is also equal to one and the dimensionless bulk modulus is varying from layer to layer so that $K^{-1} = K_0^{-1}(1 + \nu)$, where ν is a random process uniformly distributed between -0.3 and 0.3. The thickness of the layers is 0.0025km, which corresponds to a scale factor $\epsilon = 0.05$ and the point source is directed vertically and is located just above the origin on the interface between the upper homogeneous half-space and the stack of layers. The incident pulse is rapidly decaying in time with the following form in the frequency domain:

$$(4.7) \quad \hat{f}(\omega) = \exp\left(-\frac{(\epsilon\omega/2\pi)^2}{2}\right) \quad \text{for } \omega \in [\omega_{min}, \omega_{max}]$$

The reflected signals are recorded at the interface between the stack and the upper half space at 40 receivers spaced by 0.2km which corresponds to 4 pulse widths and should be adequate for treating the measurements as statistically independent ones. Estimates of the local power spectral density of these signals are obtained using the algorithm described in [[1],App E]. Since we use a physically realistic pulse shape we cannot expect to recover the very small frequency content of the local power spectral density of the reflected signal. In fact, after estimating the local power spectral density using the windowed Fourier Transform, we have to normalize it by dividing by $\omega^2|\hat{f}(\omega)|^2$ before using it in the likelihood function. Moreover, since we are using a small frequency approximation to compute the theoretical local power spectral density given the mean speed profile, we must apply the criterion described in Appendix B (B.10), which limits the range of times, frequencies and receiver offsets

$$(4.8) \quad \frac{\alpha\omega^2}{c_1^2(1 - (\frac{r}{c_1 t})^2)^{1/2}} \leq 0.01$$

The range of frequencies that we can use in the point source inversion scheme is quite limited actually. This is an unfortunate side effect of using the low frequency expansion of Appendix B rather than the full solution of the transport equations (3.4)-(3.6), especially since in the plane wave case the wider range of frequencies influence significantly the quality of the results. To use higher frequencies we must solve the full W-equations or at least extend the small frequency expansion solution to a higher order, where the admissibility criterion is less restrictive.

In general, with the maximum likelihood approach consecutive frequencies, time windows and receivers should not be too closely spaced. Then the local power spectral density estimates are approximately independent random variables exponentially

distributed around their mean, which is the solution of the W-equations. We end up with the following triple entry dataset

$$\begin{aligned} t_i &= i\delta_t \text{ with } \delta_t = 0.19635 \text{ and } 1 \leq i \leq 64 \\ r_m &= m\delta_r \text{ with } \delta_r = 0.2 \text{ and } 1 \leq m \leq 20 \\ \omega_l &\in \{0.5125, 0.5625, 0.6125, 0.6625, 0.7125\} \end{aligned}$$

4.5. Modified algorithm. A well-chosen initial discretization of the unknown mean speed profile is crucial for the inversion algorithm. As explained above, it is not determined on a uniform z grid but rather by the number of time samples between two consecutive depth points from the normal incidence travel time relation. For a given dataset this fixes the number of unknowns of the problem. If the discretization is too fine, there is too much dependent information in the likelihood functional (4.4), it is hard to minimize and has several local minima. The resulting mean speed profile is very unstable and irregular as a function of depth. If the profile discretization is too coarse then some important features may be overlooked.

To handle this problem we made two modifications to the initial algorithm. First, we use prior knowledge on the unknown profile, which is that it must be a smooth function of the z variable. In the piecewise linear discretization of the profile we forbid fluctuations that are too large by constraining the unknowns. The admissible values are based on the initial guess, which is another reason why it should not be too far from the solution, with the following bounds:

$$c_{min} \leq \bar{c}_j^k \leq c_{max}$$

with

$$c_{min} = (1 - \gamma) \min\{\bar{c}_j^{k-1}, \bar{c}_{j-1}^k, \bar{c}_{j-1}^{k+1}\}$$

and

$$c_{max} = (1 + \gamma) \max\{\bar{c}_j^{k-1}, \bar{c}_{j-1}^k, \bar{c}_{j-1}^{k+1}\}$$

where γ is a parameter which decreases with the level of allowed fluctuations.

Second, we discretize progressively the unknown profile by a multigrid method, applied along with the inversion algorithm described in the previous section. The first time we run the inversion program we determine only one parameter in the profile, its linear deviation from a constant value equal to one. This is a simple but crucial step because it gives the trend of the profile. In the subsequent inversion runs we progressively determine finer discretized profiles, starting each time from the previous result.

4.6. Numerical results. Clearly there are many parameters to adjust in the inversion algorithm and we have done extensive numerical experiments to study their influence on the results. We will describe here only two observations that play an important role. In the numerical experiments the mean speed profile is equal to 1km/sec down to 1.5km, increases slowly up to 1.5km/sec at a depth of 2.5km and then remains constant.

The number of the receivers and their spacing are important parameters. In fact, understanding and exploiting this feature is the main point of this study. In our previous work on inversions with plane wave data, spatial dependence of the local power spectral density estimates was replaced by averaging over independent realizations. In Figure (4.2) we compare the inversion profiles obtained using data from only a few consecutive receivers at a time. Since there is not much information in such a data set we start with an initial guess close to the solution. In each plot we show curves corresponding to the exact profile, the initial guess profile and the solution obtained with a few of the 32 receivers, the ones numbered 10 to 41 and located on the interface at 1.8 to 8km from the source. In the last plot at the bottom right corner all the 32 receivers are used. The initial profile is constant down to the fifth unknown corresponding to a depth around $z = 1.7\text{km}$ and the travel time has 16 discrete points.

Since the results obtained using datasets from different receivers are quite different, this numerical experiment supports the theoretical claim that signals measured at widely spaced locations are independent random processes. Moreover, the result obtained using all the receivers fits best the exact mean speed profile, since it catches quite accurately the height of the two constant states in the profile. This also supports the claim of independence.

In the second set of numerical experiments we studied the robustness of the multigrid algorithm described in the previous section. In Figure (4.3) we illustrate the first step of the inversion scheme with multigrid. The mean speed profile has the coarsest possible discretization, which is a piecewise linear function from depth 0 to a depth corresponding to 40 time steps and then a constant. The unknown is the speed in the constant part, below the linear part. We show the variations of the objective function as a function of this unknown speed for three different independent datasets and with different configurations of receivers. We see that the choice of the receivers is critical only for those near the source. If we use only receivers with offsets larger than 2km, 4km or 6km, the range of the objective function varies, since fewer terms will contribute to the summation but its minimum remains between 1.4 and 1.6. Moreover, even though the range of the objective function varies from one dataset to the other, the abscissa of the minimum is about the same. However, if we include all the receivers the minimum position varies more from one dataset to the other, and its abscissa is lower than when near receivers are dropped. From this numerical experiment and from the form of the local power spectral density estimates given in Figures (3.1) to (3.3) we conclude that we should use data only from receivers that are at least 2km from the source. The data from receivers near the source seems to be unreliable and incompatible with that from the rest of the receivers.

Figure (4.4) shows the second step of the multigrid inversion scheme. Now the unknown speed profile is discretized with two parameters. It is linear from depth 0 to a depth corresponding to 20 time steps, starting with the speed $c(0) = 1$. in the upper half-space then linear from the intermediate depth to a depth corresponding to 40 time steps below which it remains constant. The two unknowns are the speeds at depths corresponding to 20 and 40 time steps. The value of the objective function is shown as a function of two variables along with its level lines. Each plot corresponds to an independent dataset and produces a minimum at a different location, but the main point of this particular experiment is to show that whatever the dataset may be,

that is the particular realization of the random medium, the objective functional has a global minimum, which is, however, rather flat.

Finally we compare the full inversion process for the 3 independent realizations. Each group of three Figures (4.5) to (4.8) corresponds to a successive level of discretization of the velocity profile. The plots on the left come dataset number 1, the middle plots from dataset number 2, and the ones on the right from dataset number 3. In each plot there are three curves: The initial profile, the exact profile and the solution for this discretization level. The linear interpolation of the solution becomes the initial guess at the finer discretization level. After the first two multigrid levels the velocity is kept unchanged down to a depth around $z = 1$. Since the receivers nearest to the source are not included in the minimization process there is not enough resolution at the shallow depth to increase the discretization. The 4-th and last multigrid level corresponds to a discretization of the profile with three times steps per depth sample.

5. Conclusions. We have presented an algorithm in section 4 that estimates the mean sound speed profile $\bar{c}(z)$ of a randomly layered medium from measurements of the reflected pressure at different locations on the surface, when a pulsed spherical wave is incident upon it.

The main assumption underlying our inversion strategy is the three-scales model of the randomly layered medium for which we have developped a detailed asymptotic theory. In the three-scales model the typical size of the random layers is small compared to the incident pulse width which is in turn small compared to propagation distances. We also assume that the mean sound speed profile $\bar{c}(z)$ that we wish to recover from reflection data has no deterministic discontinuities or sharp variations.

The main shortcoming in the implementation of inversion algorithm is the use of the low frequency approximation of the local power spectral density Λ given the mean sound speed profile $\bar{c}(z)$ (Appendix B). This is because we do not have, at present, an efficient way to solve the hyperbolic system (3.3)-(3.6) when $\bar{c}(z)$ is given. This system must be solved repeatedly during the inversion algorithm.

The inversion strategy that we have presented is only a feasibility study. We have shown that it is possible to recover the mean sound speed profile $\bar{c}(z)$ from a single realization of point source reflection data. This is the main result of our paper. The inversion algorithm is complicated and depends sensitively on many implementation-level details. We are far from having a robust ready-to-use inversion algorithm even when the signals come from a random medium well within the scope of the three scale model.

6. Acknowledgment. The work of W. Kohler was supported by AFOSR grant F49620-92-J-0289. The work of G. Papanicolaou by NSF grant DMS9308471 and by AFOSR grant F49620-92-J-0098. The numerical computations were done at IDRIS, France, under project 940303.

REFERENCES

- [1] M. Asch W. Kohler G. Papanicolaou M. Postel and B. White. Frequency content of randomly scattered signals. *SIAM Review*, 33:pp 526–629., 1991.
- [2] G. Papanicolaou M. Postel P. Sheng and B. White. Frequency content of randomly scattered signals. Part II: Inversion. *Wave Motion*, 12:527–549., 1990.

- [3] B. White P. Sheng and B. Nair. Localization and backscattering spectrum of seismic waves in stratified lithology. *Geophysics*, 55:pp. 1158–1165., 1990.
- [4] R. F. O'Doherty and N. A. Anstey. Reflections on amplitudes. *Geophysical Prospecting*, 19:pp. 430–458., 1971.
- [5] R. Burridge G. Papanicolaou and B. White. One-dimensional wave propagation in a highly discontinuous medium. *Wave Motion*, 10:pp.19–44, 1988.
- [6] P. Lewicki and G. Papanicolaou. Reflection of wave fronts by randomly layered media. *to appear in Wave Motion*,.
- [7] J.F. Clouet and J.P. Fouque. Spreading of a pulse travelling in random media. *to appear in The Annals of Applied Probability*.
- [8] R. Burridge P. Lewicki and G. Papanicolaou. Pulse stabilization by a strongly heterogenous medium. *to appear in Wave Motion*,.
- [9] R. Burridge G. Papanicolaou P. Sheng and B. White. Probing a random medium with a pulse. *SIAM J. Appl. Math.*, 49:582–607., 1989.
- [10] P. Sheng B. White Z.Q. Zhang and G.Papanicolaou. Wave localization and multiple scattering in randomly-layered media. In Ping Sheng, editor, *Scattering and Localization of Classical Waves in Random Media*, pages pp. 563–619. World Scientific, 1990.
- [11] M. Asch G. Papanicolaou M. Postel P. Sheng and B. White. Frequency content of randomly scattered signals. Part I. *Wave Motion*, 12:429–450, 1990.
- [12] B. White P. Sheng M. Postel and G. Papanicolaou. Probing through cloudiness: Theory of statistical inversion for multiply scattered data. *Phys. Rev. Lett.*, 63:2228–2231, 1989.
- [13] J.F. Clouet J.P. Fouque and M. Postel. Spectral analysis of randomly scattered signals using wavelet transform. *Submitted to Wave Motion*, 1993.
- [14] P.G. Ciarlet. *Analyse Numérique Matricielle et Optimisation*. Masson, 1982.
- [15] W.H. Press B.P. Flannery S.A. Teukolsky W.T. Vetterling. *Numerical Recipes, The Art of Scientific Computing*. Cambridge University Press, 1986.

A. Some properties of the transport equations. In this appendix we derive some simple properties of the system of transport equations (3.4)-(3.6) that are used to compute the local power spectral density of the reflected signals from the mean speed profile.

We start with the transport equations with z as the time-like independent variable:

$$(A.1) \quad \frac{\partial W^N}{\partial z} + 2N \left\{ \frac{1}{\bar{c}(1 - \bar{c}^2 \kappa^2)^{1/2}} \frac{\partial W^N}{\partial t} + \frac{\bar{c} \kappa}{(1 - \bar{c}^2 \kappa^2)^{1/2}} \cdot \nabla W^N \right\} - 2 \frac{\omega^2 \alpha}{\bar{c}^2 (1 - \bar{c}^2 \kappa^2)} \Delta_D W^N = 0,$$

$$W^N|_{z=-L} = \delta_{N0} \delta(t) \delta(\mathbf{x}),$$

with the notation

$$\Delta_D = N^2 \left\{ W^{N+1} - 2W^N + W^{N-1} \right\}$$

The mean speed profile is $\bar{c}(z)$ and the noise intensity level is α as in sections 2 and 3. When there is no discontinuity in the effective medium properties at $z = 0$ we get the local power spectral density from the solution of (A.1) by integrating over the slowness vector κ

$$(A.2) \quad \Lambda(t, \mathbf{x}, \omega) = \frac{1}{4(2\pi)^3} \iint_{R^2} d\kappa W^1(0, t, \mathbf{x}, \kappa, \omega),$$

For numerical computations and for the low frequency expansion of Λ it is more convenient to work with the adjoint transport equation which we get as follows. Introduce the inner product

$$(A.3) \quad \langle U(z), W(z) \rangle = \int dt \int d\mathbf{x} \sum_N U^N(z, t, \mathbf{x}) W^N(z, t, \mathbf{x}).$$

and suppose that $U^N(z, t, \mathbf{x})$ satisfies the transport equations

$$(A.4) \quad \frac{\partial U^N}{\partial z} + 2N \left\{ \frac{1}{\bar{c}(1 - \bar{c}^2 \kappa^2)^{1/2}} \frac{\partial U^N}{\partial t} + \frac{\bar{c}\boldsymbol{\kappa}}{(1 - \bar{c}^2 \kappa^2)^{1/2}} \cdot \nabla U^N \right\} + 2 \frac{\omega^2 \alpha}{\bar{c}^2 (1 - \bar{c}^2 \kappa^2)} \Delta_D^* U^N = 0$$

in $-L \leq z \leq 0$ with

$$(A.5) \quad U^N(0, t, \mathbf{x}) = U_0^N(t, \mathbf{x})$$

the given initial solution and with Δ_D^* the operator adjoint to Δ_D

$$(A.6) \quad \Delta_D^* U_N = (N+1)^2 U^{N+1} - 2N^2 U^N + (N-1)^2 U^{N-1}$$

Then

$$\frac{d}{dz} \langle U(z), W(z) \rangle = 0$$

and so, by integrating from $z = 0$ to L

$$(A.7) \quad \sum_N \int \int dt d\mathbf{x} W^N(0, t, \mathbf{x}) U^N(0, t, \mathbf{x}) = \sum_N \int \int dt d\mathbf{x} W^N(-L, t, \mathbf{x}) U^N(-L, t, \mathbf{x}) = U^0(-L, 0, 0).$$

Therefore, if

$$U^N(0, t, \mathbf{x}) = \delta_{NN_0} \delta(t - t_0, \mathbf{x} - \mathbf{x}_0)$$

then

$$(A.8) \quad W^{N_0}(0, t_0, \mathbf{x}_0) = U^0(-L, 0, 0),$$

which means that the solution of the W-equations (A.1) can be obtained also by solving the adjoint U-equations (A.4). The adjoint equations are more convenient for numerical computations in connection with inverse problems.

Since we want to compute the local power spectral density through (A.2) we see that we need only solve the adjoint transport equations with the initial conditions zero except for $N_0 = 1$. After a change of variables: $t \rightarrow t_0 - t$ and $\mathbf{x} \rightarrow \mathbf{x}_0 - \mathbf{x}$ we get

$$(A.9) \quad \frac{\partial U^N}{\partial z} - 2N \left\{ \frac{1}{\bar{c}(1 - \bar{c}^2 \kappa^2)^{1/2}} \frac{\partial U^N}{\partial t} + \frac{\bar{c}\boldsymbol{\kappa}}{(1 - \bar{c}^2 \kappa^2)^{1/2}} \cdot \nabla U^N \right\} + \frac{2\omega^2 \alpha}{\bar{c}^2 (1 - \bar{c}^2 \kappa^2)} \Delta_D^* U_N = 0$$

in $z < 0$ with

$$(A.10) \quad U^N|_{z=0} = \delta(t)\delta(\mathbf{x})\delta_{N1}.$$

and the local power spectral density (A.2) is now given by

$$(A.11) \quad \Lambda(t, \mathbf{x}, \omega) = \frac{1}{4(2\pi)^2} \iint d\boldsymbol{\kappa} U^0(-L, t, \mathbf{x}, \boldsymbol{\kappa}, \omega)$$

where L is larger than the depth the waves penetrate up to time $t/2$.

To solve the U-equations numerically it is convenient to transform the depth variable z to travel time and straighten the characteristics. Let

$$(A.12) \quad \tau(z, \kappa) = -2 \int_0^z \frac{ds}{\tilde{c}(s) (1 - \tilde{c}^2 \kappa^2)^{1/2}}.$$

We then get

$$(A.13) \quad \frac{\partial U^N}{\partial \tau} + N \left\{ \frac{\partial U^N}{\partial t} + \tilde{c}^2 \boldsymbol{\kappa} \cdot \nabla U^N \right\} - \frac{\omega^2 \alpha}{\tilde{c} (1 - \tilde{c}^2 \kappa^2)^{1/2}} \Delta_D^* U_N = 0$$

with the same initial condition (A.10) at $\tau = 0$. In these equations the mean speed profile $\tilde{c}$ has a tilda on top of it so as to emphasize that it now depends on the travel time τ and the slowness κ through the inverse function of (A.12) which we denote by $\tilde{z}(\tau, \kappa)$. The local power spectral density Λ (A.11) is now determined by

$$(A.14) \quad \Lambda(t, \mathbf{x}, \omega) = \frac{1}{4(2\pi)^2} \iint d\boldsymbol{\kappa} U^0(t/2, t, \mathbf{x}, \boldsymbol{\kappa}, \omega)$$

Using this formulation we carried out extensive numerical computations, as reported in section 5 of [1]. However, they are very time consuming and further studies showed that the discretization of the slowness $\boldsymbol{\kappa}$ and of $\boldsymbol{\lambda}$, the dual variable of $\mathbf{x}$, should depend on the mean speed profile $\tilde{c}$. For these two reasons this numerical method was not used in the inversion algorithm. We replaced it by the small frequency approximation of Λ presented in the next appendix.

B. Small frequency approximation. In this appendix we show how to get a small frequency approximation for the local power spectral density Λ . We also compare this approximation with the exact result in the case of a homogeneous background and give a criterion for its validity.

We start with the adjoint transport equations (A.9, A.11) and look for a small frequency expansion of the solution U

$$(B.1) \quad U = V_0 + \omega^2 V_2 + \omega^4 V_4 + \dots$$

The system of equations satisfied by each term of the expansion is determined successively. The first one decouples and allows us to compute V_0 .

$$\frac{\partial V_0^N}{\partial z} - 2N \left\{ \frac{1}{\tilde{c} (1 - \tilde{c}^2 \kappa^2)^{1/2}} \frac{\partial V_0^N}{\partial t} + \frac{\tilde{c} \boldsymbol{\kappa}}{(1 - \tilde{c}^2 \kappa^2)^{1/2}} \cdot \nabla V_0^N \right\} = 0$$

in $z < 0$ with

$$V_0^N|_{z=0} = \delta(t)\delta(\mathbf{x})\delta_{N1}.$$

This gives $V_0^N = 0$ for $N \neq 1$. To compute V_0^1 we introduce the auxiliary unknown $Y(z, s, \boldsymbol{\kappa}) = V_0^1(z, \bar{t}(z, s, \boldsymbol{\kappa}), \bar{\mathbf{x}}(z, s, \boldsymbol{\kappa}), \boldsymbol{\kappa})$ which is V_0^1 evaluated along the characteristics corresponding to $N = 1$

$$(B.2) \quad \begin{aligned} \bar{t}(z, s, \boldsymbol{\kappa}) &= -2 \int_s^z \frac{du}{\bar{c}(u)\sqrt{1 - \bar{c}(u)^2 \boldsymbol{\kappa}^2}} \\ \bar{\mathbf{x}}(z, s, \boldsymbol{\kappa}) &= -2 \int_s^z \frac{\bar{c}(u)\boldsymbol{\kappa} du}{\sqrt{1 - \bar{c}(u)^2 \boldsymbol{\kappa}^2}} \end{aligned}$$

The function Y satisfies

$$\begin{aligned} \frac{dY}{dz} &= 0 \\ Y|_{z=0} &= \delta(\bar{t}(0, s, \boldsymbol{\kappa}))\delta(\bar{\mathbf{x}}(0, s, \boldsymbol{\kappa})) \end{aligned}$$

which leads to the solution

$$(B.3) \quad V_0^1(z, t, \mathbf{x}, \boldsymbol{\kappa}) = \delta(t - \bar{t}(z, 0, \boldsymbol{\kappa}))\delta(\mathbf{x} - \bar{\mathbf{x}}(z, 0, \boldsymbol{\kappa}))$$

To get the next term in the expansion we calculate V_2^N

$$\begin{aligned} \frac{\partial V_2^N}{\partial z} - 2N \left\{ \frac{1}{\bar{c}(1 - \bar{c}^2 \boldsymbol{\kappa}^2)^{1/2}} \frac{\partial V_2^N}{\partial t} + \frac{\bar{c}\boldsymbol{\kappa}}{(1 - \bar{c}^2 \boldsymbol{\kappa}^2)^{1/2}} \cdot \nabla V_2^N \right\} \\ = -\frac{2\alpha}{\bar{c}^2(1 - \bar{c}^2 \boldsymbol{\kappa}^2)} \Delta_D^* V_0^N \end{aligned}$$

in $z < 0$ with

$$V_2^N|_{z=0} = 0$$

with $\Delta_D^* V_0^N$ defined by (A.6). The $N = 0$ component, which enters in the computation of Λ through (A.11), satisfies

$$\begin{aligned} \frac{\partial V_2^0}{\partial z} &= -\frac{2\alpha}{\bar{c}^2(1 - \bar{c}^2 \boldsymbol{\kappa}^2)} V_0^1 \\ V_2^0|_{z=0} &= 0 \end{aligned}$$

whose solution is

$$V_2^0(z, t, \mathbf{x}, \boldsymbol{\kappa}) = -\int_0^z \frac{2\alpha}{\bar{c}^2(1 - \bar{c}^2 \boldsymbol{\kappa}^2)} V_0^1(s, t, \mathbf{x}, \boldsymbol{\kappa}) ds$$

We expand Λ in even powers of ω

$$(B.4) \quad \Lambda(t, \mathbf{x}) = \omega^2 \Lambda_2(t, \mathbf{x}) + \omega^4 \Lambda_4(t, \mathbf{x}) + \dots$$

and use (B.3) in (A.11) to get the order ω^2 term

$$\begin{aligned} \Lambda_2(t, \mathbf{x}, \omega) &= \frac{1}{4(2\pi)^2} \iint d\boldsymbol{\kappa} \\ &\int_0^{-L^*} -dz' \frac{2\alpha}{\bar{c}(z')^2(1 - \bar{c}(z')^2 \boldsymbol{\kappa}^2)} \delta(t - \bar{t}(z', 0, \boldsymbol{\kappa}))\delta(\mathbf{x} - \bar{\mathbf{x}}(z', 0, \boldsymbol{\kappa})) \end{aligned}$$

This triple integral can be computed by a change of variables from $(\boldsymbol{\kappa}, z')$ to $(\tilde{\mathbf{x}}, \tilde{t})$ given by

$$\begin{aligned}\tilde{t} &= \bar{t}(z', 0, \boldsymbol{\kappa}) \\ \tilde{\mathbf{x}} &= \bar{\mathbf{x}}(z', 0, \boldsymbol{\kappa})\end{aligned}$$

This allows us to evaluate the delta functions and get

$$(B.5) \quad \Lambda_2(t, \mathbf{x}) = \frac{1}{\det J(\boldsymbol{\kappa}, z)} \frac{2\alpha}{\bar{c}(z)^2 (1 - \bar{c}(z)^2 \kappa^2)} \Big|_{t=\bar{t}(z, 0, \boldsymbol{\kappa}), \mathbf{x}=\bar{\mathbf{x}}(z, 0, \boldsymbol{\kappa})}$$

where $\mathbf{J}$ is the 3x3 jacobian matrix of the change of variables and its determinat is

$$\det J(\boldsymbol{\kappa}, z') = 8 \frac{\sqrt{1 - \bar{c}(z')^2 \kappa^2}}{\bar{c}(z')} \int_0^{z'} \frac{\bar{c}(s) ds}{\sqrt{(1 - \bar{c}(s)^2 \kappa^2)^3}} \int_0^{z'} \frac{\bar{c}(s) ds}{\sqrt{(1 - \bar{c}(s)^2 \kappa^2)}}$$

The final expression for Λ_2 is then

$$(B.6) \quad \Lambda_2(t, \mathbf{x}) = \frac{\alpha}{4\bar{c}(z) (1 - \bar{c}(z)^2 \kappa^2)^{3/2} \int_0^z \frac{\bar{c}(s) ds}{\sqrt{(1 - \bar{c}(s)^2 \kappa^2)^3}} \int_0^z \frac{\bar{c}(s) ds}{\sqrt{(1 - \bar{c}(s)^2 \kappa^2)}}$$

where z and $\boldsymbol{\kappa}$ are such that

$$(B.7) \quad t = \bar{t}(z, 0, \boldsymbol{\kappa}) \quad \text{and} \quad \mathbf{x} = \bar{\mathbf{x}}(z, 0, \boldsymbol{\kappa})$$

This is the formula that we use in our inversions in this paper.

To compute the ω^4 term of the expansion we should also know V_2^1 which is needed in $\Delta^* V_2^0$. It satisfies

$$\begin{aligned}\frac{\partial V_2^1}{\partial z} - 2 \left\{ \frac{1}{\bar{c}(1 - \bar{c}^2 \kappa^2)^{1/2}} \frac{\partial V_2^1}{\partial t} + \frac{\bar{c}\boldsymbol{\kappa}}{(1 - \bar{c}^2 \kappa^2)^{1/2}} \cdot \nabla V_2^1 \right\} \\ = \frac{2\alpha}{\bar{c}^2 (1 - \bar{c}^2 \kappa^2)} V_0^1\end{aligned}$$

in $z < 0$ with

$$V_2^1 \Big|_{z=0} = 0$$

Using (B.3) we get

$$V_2^1(z, t, \mathbf{x}, \boldsymbol{\kappa}) = \int_0^z \frac{2\alpha dz'}{\bar{c}(z')^2 (1 - \bar{c}(z')^2 \kappa^2)} \delta(t - \bar{t}(z', 0, \boldsymbol{\kappa})) \delta(\mathbf{x} - \bar{\mathbf{x}}(z', 0, \boldsymbol{\kappa}))$$

We can now consider the next term in the ω expansion which satisfies

$$\begin{aligned}\frac{\partial V_4^N}{\partial z} - 2N \left\{ \frac{1}{\bar{c}(1 - \bar{c}^2 \kappa^2)^{1/2}} \frac{\partial V_4^N}{\partial t} + \frac{\bar{c}\boldsymbol{\kappa}}{(1 - \bar{c}^2 \kappa^2)^{1/2}} \cdot \nabla V_4^N \right\} \\ = -\frac{2\alpha}{\bar{c}^2 (1 - \bar{c}^2 \kappa^2)} \Delta_D^* V_2^N\end{aligned}$$

in $z < 0$ with

$$V_4^N \Big|_{z=0} = 0$$

For the $N = 0$ component, which enters in the computation of Λ , we have

$$(B.8) \quad V_0^4(z, t, \mathbf{x}, \boldsymbol{\kappa}) = \int_0^z V_2^1(z', t, \mathbf{x}, \boldsymbol{\kappa}) dz'$$

Combining (A.11) and (B.8) leads to a quadruple integral expression for the ω^4 term of Λ in (B.4)

$$(B.9) \quad \Lambda_4(t, \mathbf{x}, \omega) = \frac{1}{4(2\pi)^2} \iint d\boldsymbol{\kappa} \int_0^{-L^*} dz \int_0^z dz' \frac{2\alpha}{\bar{c}(z')^2 (1 - \bar{c}(z')^2 \kappa^2)} \delta(t - \bar{t}(z', 0, \boldsymbol{\kappa})) \delta(\mathbf{x} - \bar{\mathbf{x}}(z', 0, \boldsymbol{\kappa}))$$

In numerical computations this is treated simultaneously with (B.6). It is a quadrature with respect to z of $\Lambda_2(t, \mathbf{x})$ multiplied by $H(z - z')$, the Heaviside function, where the argument z' is the solution of the characteristic equations (B.7) that enter also in the first order term (B.6).

More terms in the expansion (B.4) could be computed in a similar way. However, with the ω^6 term the complexity starts to increase because the equations in (A.9) are fully coupled by the non-zero terms in $\Delta^* W^N$. This requires solving simultaneously several characteristic problems of the form (B.7).

In this paper the inversions are done with the ω^2 approximation only. To estimate the range of validity of this approximation, which is a small frequency as well as a small noise, α , approximation, we use the uniform background formula (3.7) for the local power spectral density. It is easy to see that when $\bar{c}(z) = \bar{c}$, a constant, the first term in the small frequency expansion, (B.6), is the first term in the expansion of (3.7)

$$(B.10) \quad \Lambda(t, r, \omega) \approx \frac{1}{4(2\pi)^2} \frac{1}{\bar{c}(\bar{c}t)^2} \frac{\alpha\omega^2}{\bar{c}^2(1 - (\frac{r}{\bar{c}t})^2)^{1/2}}$$

The left side of Figure (B.1,B.2) compares (B.6), or (B.10), and (3.7), for two frequencies ($\omega = 0.6875$ and 1.6875 respectively) and for five different receiver positions, as a function of time. The right side of the Figure shows the error $\frac{\alpha\omega^2}{\bar{c}^2}(1 - (\frac{r}{\bar{c}t})^2)^{-1/2}$ as a function of time. We see that if the error is kept below some reasonable level, say 0.01, then the agreement between the two sets of curves is excellent. We will use this criterion even when the background velocity $\bar{c}$ is not constant. In that case the numerical solution of (B.6-B.7) is done as follows. The profile $\bar{c}(z)$ is given by a piecewise linear function of z . For a given pair of variables $(t, \mathbf{x})$ we get an initial guess for the solution of (B.7) by using a constant profile with speed equal to the maximum speed in the variable one.

$$(B.11) \quad z = \frac{1}{2}t \, c_{max} \sqrt{1 - r^2/t^2} \quad \boldsymbol{\kappa} = \frac{1}{t \, c_{max}^2} \mathbf{x}$$

We also change unknowns from κ to θ to keep it bounded and the argument of the square roots in (B.6) positive.

$$(B.12) \quad \kappa = |\boldsymbol{\kappa}| = \frac{1 + \sin(\theta)}{2 \, c_{max}}$$

The derivatives of $\bar{t}$ and $\bar{\mathbf{x}}$ with respect to z and $\boldsymbol{\kappa}$ (or z and θ) can easily be computed and the evaluated numerically by quadratures. The numerical solution of (B.7) is obtained using the Newton-Raphson algorithm.

C. Validation of the inversion algorithm on plane wave data. As we noted briefly in section 2, inversions with plane wave data have been carried out extensively and the maximum likelihood approach is quite effective when the hyperbolicity of the transport equations is exploited using layer peeling, which we describe below. In this appendix we will compare the point source inversion algorithm of section 4 to this layer peeling algorithm applied to plane wave inversions. The only essential difference between the point source inversion algorithm of section 4 and its application to plane wave data here is that we do not use the low frequency approximation in the plane wave case because solving the transport equations numerically is relatively easy. We also do not use the multigrid variant of the method.

In the plane wave case the adjoint form of the transport equations is (A.13) with $\kappa = 0$:

$$(C.1) \quad \begin{aligned} \frac{\partial U^N}{\partial \tau} + 2N \frac{\partial U^N}{\partial t} + 2\omega^2 \frac{\alpha}{\tilde{c}(\tau)} \Delta_D^* U_N &= 0 \\ U^N(0, t, \omega) &= \delta(t) \delta_{N,1}. \end{aligned}$$

Here $\tau = \tau(z)$ is the travel time

$$(C.2) \quad \tau(z) = \int_0^z \frac{ds}{\tilde{c}(s)},$$

which is negative when z is negative, $\tilde{z}(\tau)$ is its inverse function and

$$(C.3) \quad \tilde{c}(\tau) = \tilde{c}(\tilde{z}(\tau)).$$

The noise intensity is denoted by α as before. The local power spectral density is now given by

$$(C.4) \quad \Lambda(t, \omega) = U^0(\tau = -t/2, t, \omega)$$

where U^0 is the $N = 0$ component of the solution U^N of (C.1). This equation is solved numerically by a finite difference scheme described in detail in [2] but simpler methods will also work quite well.

Using the windowed Fourier transform we get from the reflection data a set estimates $\{\hat{\Lambda}(t_i, \omega_l), i = 1, \dots, N_t; l = 1, \dots, N_\omega\}$ of the local power spectral density and construct the likelihood function

$$(C.5) \quad \begin{aligned} &\prod_{\text{realiz}} \text{Prob.dens.} \{ \hat{\Lambda}(t_i, \omega_l), \quad i = 1, 2, \dots, N_t; l = 1, 2, \dots, N_f | \tilde{c}(\cdot) \} \\ &= \prod_{\text{realiz}} \prod_{i=1}^{N_t} \prod_{l=1}^{N_f} \frac{e^{-\hat{\Lambda}(t_i, \omega_l) / \Lambda(t_i, \omega_l; \tilde{c}(\cdot))}}{\Lambda(t_i, \omega_l; \tilde{c}(\cdot))}, \end{aligned}$$

which includes the product over the different, independent realizations. The unknown profile $\tilde{c}$ is recovered by maximizing (C.5). There are several ways to do this. The best one is layer peeling, described in [2],[1]. Very briefly, we first determine $\tilde{c}(\tau)$ as a function of travel time instead of depth and then convert it to a function of depth by inverting (C.3) and (C.2). The layer peeling method uses the hyperbolicity of (C.1), i.e., the fact that $\Lambda(t, \omega; \tilde{c}(\cdot))$ depends on $\tilde{c}(\tau)$ only for travel times $-\tau \leq t/2$. Using a piecewise linear representation for $\tilde{c}(\cdot)$, for each fixed t there is an index j such

that $\Lambda(t_j, \omega_l; \bar{c}(\cdot))$ depends on $\bar{c}^1, \bar{c}^2, \dots, \bar{c}^j$ but not on $\bar{c}^{j+1}, \dots, \bar{c}^{N_t}$. We may therefore determine $\bar{c}^j$ recursively. That is, if $\bar{c}^1, \bar{c}^2, \dots, \bar{c}^{j-1}$ have been determined, $\bar{c}^j$ can be determined by maximizing the likelihood and so (C.5) is to be maximized only over the single variable $\bar{c}^j$.

With the algorithm of section 4 we have a global inversion problem for the discretized mean speed profile $\{\bar{c}^k, k = 1, N_z\}$ that maximizes the likelihood function or minimizes the negative log-likelihood function

$$\begin{aligned}
 O(\bar{c}^1, \dots, \bar{c}^{N_z}) &= - \sum_{realiz} \log[\text{Prob.dens.}\{\hat{\Lambda}(t_i, \omega_l), i = 1, \dots, N_t; l = 1, \dots, N_\omega | \bar{c}\}] \\
 (C.6) \qquad &= \sum_{realiz} \sum_{i=1}^{N_t} \sum_{l=1}^{N_\omega} \left\{ \frac{\hat{\Lambda}(t_i, \omega_l)}{\Lambda(t_i, \omega_l; \bar{c}(\cdot))} + \log(\Lambda(t_i, \omega_l; \bar{c}(\cdot))) \right\}.
 \end{aligned}$$

The algorithm works well but is much slower than layer peeling. It does not give exactly the same results as layer peeling but this is as expected because the unknown speeds do not depend on the same variables in the two cases and are computed on different grids. The numerical experiments were done with a set of 20 realizations. The estimates $\hat{\Lambda}$ of the local power spectral density were obtained on the time-frequency grid $\delta t = 0.163$, $\delta \omega = 0.72$ and $\omega_0 = 0.261$. We used fifteen frequencies in the maximum likelihood function, $N_\omega = 15$.

In the first group of four plots, figures (C.1) and (C.2), we compare the layer peeling and global inversion schemes. In each plot there are three curves: The exact profile and the layer peeling solution are the same in all four plots. In the first plot we show the solution by global inversion for a reduced set of unknowns $\bar{c}^i, i = 1, 10$, computed with four iterations using the first twenty time estimates of the local spectral density. In the second the global inversion result is obtained with the first twenty unknown speeds $\bar{c}^i, i = 1, 20$. The algorithm converges after nine iterations. The third and fourth plots show the tenth and fourteenth (and last) iterations in the global inversion scheme for the full set of unknowns $\bar{c}^i, i = 1, 30$ for which we use estimates of the local spectral density for sixty times. Convergence requires fourteen iterations with the criterion described above, the discrete L^2 norm of the relative difference between two consecutive solutions, and error level equal to 0.01.

The second group of plots, figure (C.3), shows nine of the fourteen iterations that lead to the second plot of the previous figure (C.2). The initial profile was flat $\bar{c}^i = 1, i = 1, 30$

FIG. 4.1. Comparison of the profile discretizations with constant depth increment and constant time increment

FIG. 4.2. Comparison of mean speed profile obtained using different sets of receivers

FIG. 4.3. *Objective as a function of one speed parameter for 3 independent realizations and different sets of receivers*

FIG. 4.4. *Objective as a function of two speed parameters for 3 independent realizations*

FIG. 4.5. Inversion results for 3 independent datasets. After the first multigrid inversion step

FIG. 4.6. Inversion results for 3 independent datasets. After the second multigrid inversion step

FIG. 4.7. Inversion results for 3 independent datasets. After the third multigrid inversion step

FIG. 4.8. Inversion results for 3 independent datasets. After the fourth multigrid inversion step

FIG. B.1. *Small frequency approximation for $\omega = 0.6875$*

FIG. B.2. *Small frequency approximation for $\omega = 1.6875$*

FIG. C.1. Comparison of layer peeling and global inversion schemes for partial sets of unknowns

FIG. C.2. Comparison of layer peeling and global inversion schemes for the full sets of unknowns

FIG. C.3. *Evolution of the solution with the global inversion scheme*