Probing through Cloudiness: Theory of Statistical Inversion for Multiply Scattered Data

Benjamin White and Ping Sheng

Exxon Research & Engineering Company, Annandale, New Jersey 08801

Marie Postel and George Papanicolaou

Courant Institute of Mathematical Sciences, 251 Mercer Street, New York, New York 10012
(Received 23 January 1989)

Wave multiple scattering is responsible for making a random medium cloudy in appearance and opaque in the sense of structure delineation. For a randomly layered medium such as the Earth's subsurface, however, knowledge about the generic behavior of multiple scattering enables us to construct a theory of statistical inversion which can recover from a signal data set the slowly varying mean character of a medium with single amplitude only 10^{-3} that of the multiple-scattering noise. Inversion accuracy improves systematically with the availability of statistically redundant data.

Wave multiple scattering is responsible for the cloudy and opaque appearance of many disordered materials. For the purpose of structure delineation or detection of targets embedded in random media, such cloudiness has long been recognized as a major obstacle to obtaining useful data. Recently, however, progress in understanding the statistical character of multiply scattered waves has enabled new gains on this difficult problem. In particular, the diffusive nature of the multiply backscattered light was utilized to advantage in deciphering the dynamical information about opaque colloidal suspensions.[1]

In this work, we present a theory of statistical inversion for pulse-reflection data from a randomly layered medium. It is shown that, by taking into account the statistical property of the reflection power spectrum[2-5] from first principles, one may be able to recover the statistical mean of the medium property even from data dominated by strong multiple scattering. It thus demonstrates for the first time that the cloudiness barrier may be penetrated. In addition, since the slow variation of the average medium property is usually difficult to reconstruct due to the insensitivity of a pulse to structures much larger than its width, the present theory therefore also represents a capability which is complementary to that of the traditional inversion approach.[6,7] Application of the theory to numerically simulated data demonstrates that our method is able to recover from a single data set, with reasonable accuracy, large-scale variations whose scattering amplitudes are only 10^{-3} times those of the multiple-scattering noise. With the availability of additional sets of statistically redundant data, the theory improves the inversion accuracy systematically. In what follows, we will present the theory in terms of a randomly layered elastic medium, such as the Earth's subsurface. However, both the results and the approach are generally applicable to electromagnetic waves as well.[8]

Our model for the inversion theory may be described as follows: Since nature generally presets a continuous range of length scales, the probing pulse width, defined to be 1, naturally sets the dividing line that separates the medium structure into two scale categories, ϵ and $1/\epsilon$, where ϵ is a small number. The small-scale structures are denoted as fluctuations, and the large-scale structures are associated with the variation of the mean. For elastic-wave (or electromagnetic-wave) propagation, we express the bulk modulus $\kappa(z)$ and the density $\rho(z)$ of the medium as

$$\kappa^{-1}(z) = \kappa_0^{-1}(z)[1 + \sigma_\kappa(z)], \qquad (1a)$$

$$\rho(z) = \rho_0(z)[1 + \sigma_\rho(z)], \qquad (1b)$$

where $\kappa_0^{-1}(z), \rho_0(z)$ vary on the $1/\epsilon$ scale and $\sigma_\kappa(z)$, $\sigma_\rho(z)$ vary on the ϵ scale for $z > 0$, and $\kappa_0^{-1}(z) = \kappa_0^{-1}(0)$, $\rho(z) = \rho_0(0)$, $\sigma_\kappa(z) = \sigma_\rho(z) = 0$ for $z < 0$. Figure 1 gives a schematic illustration of the sound speed $c(z) = [\kappa(z)/\rho(z)]^{1/2}$ for such a medium where the
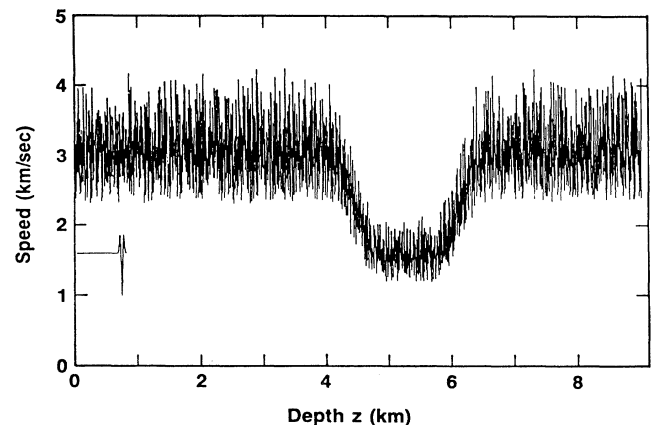

FIG. 1. Model of sound speed $c(z)$ plotted as a function of depth z. The dip in $c_0(z)$ is noted to be buried below ~ 1000 random layers. Also shown is the Ricker wavelet whose width is ~ 50 random layers. Each layer in the model has a thickness of 3 m.

large-scale structure consists of a single dip, and the small layers are of constant thickness, l_0 with $\sigma_\kappa, \sigma_\rho$ constant within each layer and varying randomly from layer to layer with a uniform distribution in the interval $[-0.3, 0.3]$. Also shown is the pressure pulse used to probe such a medium. The goal of our inversion is to recover from the pulse backscattering data at $z = 0$ the mean speed profile $c_0(z) = [\kappa_0(z)/\rho_0(z)]^{1/2}$ of the medium.

The propagation of a normally incident pulse is governed by the equations

$$\rho \dot{v} = -\partial p/\partial z , \qquad (2a)$$

$$\dot{p} = -\kappa \, \partial v/\partial z . \qquad (2b)$$

Here p is the pressure, v is the displacement velocity, and the overdot denotes time derivative. By using the velocity profile shown in Fig. 1, we have numerically solved Eq. (2) by using a fourth-order-accurate scheme[9] to obtain $p(t)$ at $z = 0$. The pulse shape used is the so-called Ricker wavelet given by

$$f(y) = 2[4\pi^2(y-0.5)^2 - 0.5]\exp[-4\pi^2(y-0.5)^2] , \qquad (3)$$

where $y = (t-z/c)/20$ and lies in $[0,1]$, t has the unit of second, and $f = 0$ for y outside of $[0,1]$. For $l_0 \simeq 3$ m and $c_0 \simeq 3000$ m/sec such as those used in Fig. 1, the dominant frequency content of $f(y)$ is 40 Hz, and the pulse width is $\sim 2c_0/\text{frequency} = 150$ m so that $\epsilon \simeq 0.02$. The effect of the macroscopic dip is represented by a tiny modulation of the backscattered signal. To extract the structure we must therefore use the statistical approach based on the analysis of the power spectrum of the backscattered data.

Because of the nonstationarity of the back-reflected signal, its power spectrum $S(t, \omega)$ depends on both the frequency ω and the center of the time window t at which the power spectrum is evaluated. $S(t, \omega)$ is related to the time correlation function of $p(t)$,

$$\langle p(t')p(t+t') \rangle = (2\pi)^{-1} \int_{-\infty}^{\infty} d\omega \exp(-i\omega t')S(t,\omega) , \qquad (4a)$$

where the angular brackets denote configurational averaging. From the linearity of the problem, we have

$$p(t) = (2\pi)^{-1} \int_{-\infty}^{\infty} d\omega \exp(i\omega t)f(\omega)R(\omega) . \qquad (4b)$$

Here $f(\omega)$ is the frequency spectrum of the Ricker wavelet, and $R(\omega)$ is the reflection coefficient. Since a randomly layered medium is impenetrable, we can write $R(\omega) = \exp[i\psi(\omega)]$. By substituting Eq. (4b) into Eq. (4a), it is straightforward to show that

$$S(t,\omega) \cong |f(\omega)|^2 N(t,\omega) , \qquad (5a)$$

with

$$N(t,\omega) = (2\pi)^{-1} \int_{-\infty}^{\infty} dh \exp(-iht)u(h,\omega) \qquad (5b)$$

and

$$u(h,\omega) = \langle \exp\{i[\psi(\omega-h/2) - \psi(\omega+h/2)]\} \rangle$$

$$= \langle \exp(i\bar{\psi}) \rangle , \qquad (5c)$$

where h is the difference between the frequencies.

Equation (5c) tells us that the kernel of information in the reflected data lies in the correlation of the reflection phases at two different frequencies. In fact, for a single-frequency wave phase $\psi(\omega)$ is uniformly distributed in the interval $[0, 2\pi]$ and is therefore not expected to yield useful information. The use of a multifrequency pulse as a probe is therefore seen as necessary. From the definition of $u(h,\omega)$ we see immediately that $u(0,\omega) = 1$. Also, since h is the conjugate variable to t [Eq. (5b)], the width of u as a function of t is expected to govern the time scale of nonstationarity in the reflected data. To achieve our goal of inversion, we must now relate the time variation of ψ [and therefore $N(\tau,\omega)$] to the property of the medium. From previous works on pulse scattering and localization,[3,5] we get

$$\frac{\partial U}{\partial \tau} = \left[h\frac{\partial}{\partial\bar{\psi}} + 2\omega^2 a(\tau)(1-\cos\bar{\psi})\frac{\partial^2}{\partial\bar{\psi}^2} \right] U , \qquad (6a)$$

where $a(\tau) = \bar{a}/c_0(\tau)$,

$$\bar{a} = \int_0^\infty \langle n(z)n(z+s) \rangle ds , \qquad (6b)$$

and $n(z) = [\sigma_\rho(z) - \sigma_\kappa(z)]/2$. In the above τ is the two-way travel time defined by

$$\tau(z) = 2\int_0^z ds\, c_0^{-1}(s), \quad z > 0 , \qquad (6c)$$

which may be viewed as a transformed distance variable. In order to evaluate $u(h,\omega)$, the initial condition on U is

$$U|_{z=0} = \exp(i\bar{\psi}) . \qquad (7a)$$

Then u is given by

$$u(h,\omega) = \lim_{z\to\infty} U(z,\bar{\psi},h,\omega) . \qquad (7b)$$

The limit in Eq. (7b) exists and is independent of $\bar{\psi}$, which follows from the assumption that $\bar{a}$ and c_0 are constant for $z \to \infty$.

The physical interpretation of Eq. (6a) is as follows. When the pulse front travels into a random medium, the variation in $\bar{\psi}$ as a function of τ consists of two parts. The first term on the right-hand side of Eq. (6a) states that $\Delta\bar{\psi} = h\Delta\tau$. Since we know that $\bar{\psi} = 0$ for $h = 0$, this term merely expresses the first-order effect due to the frequency difference: If h is small, then the phase difference $\bar{\psi}$ is also small in a given increment $\Delta\tau$, and vice versa. The presence of this wavelike term ($\partial U/\partial\tau = h^- \partial U/\partial\psi$ has a wave solution) is important in making Eq. (6a) casual in character. The second term, $(\Delta\bar{\psi})^2 \propto \omega^2 a(\tau)\Delta\tau$, is clearly diffusive in nature and represents the net effect of multiple scattering by random

fluctuations. The diffusion constant is noted to be proportional to $\omega^2 a(\tau)$. Here $a(\tau)$ represents the medium characteristics as sensed by the reflection power spectrum. It contains integrated information about the small-scale fluctuations $(\tilde{a})$ and the slowly varying mean speed $c_0(z)$. The aim of our inversion scheme is to regard the noise spectrum $N(\tau,\omega)$ [Fourier transform of $u(h,\omega)$] as input from which to evaluate $a(\tau)$.

We first note that, if $a\omega^2$ is small, then Eq. (6a) may be solved perturbatively[10] to obtain

$$N(\tau,\omega) = \omega^2 a(\tau) + O(a^2\omega^4) , \qquad (8a)$$

from which we can get $a(\tau)$ as

$$a(\tau) = \lim_{\omega \to 0} \frac{S(\tau,\omega)}{\omega^2 |f(\omega)|^2} . \qquad (8b)$$

Equation (8b) states that if the pulse has low-frequency content, then these frequencies would be sufficient to determine $a(\tau)$ uniquely. This is so because low-frequency waves are special, due to their sensitivity to the large structure whose scales are comparable to the wavelengths. At the same time, they are also most insensitive to fluctuations. However, in many instances $|f(\omega)|^2$ has no low-frequency content, e.g., the Ricker wavelet given by Eq. (3), and the determination in the manner of Eq. (8b) would not be feasible. The inversion must therefore rely on the redundant information about $a(\tau)$ contained in the spectrum values at different higher frequencies and different times. The fact that the higher-

frequency waves also contain information about $a(\tau)$ is not surprising. In this case although the scattering signals from large structures are small, yet the multiple-scattering–induced diffusive behavior makes the resulting power spectrum sensitive to the local $c_0(z)$ as evidenced by Eq. (6a). The effect of multiple scattering is therefore seen to be twofold. On the one hand, it has the well-known effect of scrambling the fine structures so as to make them nonresolvable; yet, on the other hand, *it retains the information about the local mean property of the medium and spreads such information among all frequencies, thus making them available for inversion.*

To extract $a(\tau)$ from the data, we first note that the spectrum $S(\tau,\omega) = |f(\omega)|^2 N(\tau,\omega)$, calculated from Eq. (6a) through the Fourier transform of $u(h,\omega)$, represents the configurationally averaged expected value. In reality, we can only obtain from data an estimate of $S(\tau,\omega)$, denoted as $\hat{S}(\tau,\omega)$, by sampling the data over a finite time window T. If M is the number of equally spaced data points of $p(\tau)$, then the sampling time interval is T/M, and $\hat{S}$ is given by

$$\hat{S}(\tau,\omega_j) = \left| \frac{1}{\sqrt{T}} \frac{T}{M} \sum_{l=0}^{M-1} \exp\left(-\frac{2\pi j l}{M} \right) p(\tau+s_l) \right|^2 , \qquad (9)$$

where $s_l = Tl/M - T/2$, and $\omega_j = 2\pi j/T$. A well-known theorem in signal-processing theory[11] tells us that as $T,M \to \infty$ and $T/M \to 0$, $\hat{S}(\tau,\omega)$ does not converge uniquely to $S(\tau,\omega)$. Instead, the values of $\hat{S}(\tau,\omega)$ (for different configurations) converge to an exponential distribution with $S(\tau,\omega)$ as the mean, i.e.,

$$P\{\hat{S}(\tau_k,\omega_j) | a(\tau)\} = \exp[-\hat{S}(\tau_k,\omega_j)/S(\tau_k,\omega_j)a(\tau)]/S(\tau_k,\omega_j)a(\tau) , \qquad (10a)$$

where $P\{\hat{S}|a(\tau)\}$ denotes the probability of having a value $\hat{S}$, given $a(\tau)$. Moreover, each frequency ω_j is independent. A likelihood function $L[a(\tau)]$ may therefore be written as

$$L[a(\tau)] = -\sum_{k=1}^{L} \sum_{j=1}^{M} \ln[P\{\hat{S}(\tau_k,\omega_j) | a(\tau)\}] . \qquad (10b)$$

Maximization of L would yield the optimal $a(\tau)$ consistent with the data $\hat{S}$. Within the same formalism we can also include the effect of additive white noise. In that case $P\{\hat{S}|a(\tau)\}$ would still be described by an exponential distribution but with a mean of $S(\tau,\omega) = |f(\omega)|^2 N(\tau,\omega) + \sigma^2(\tau)$. By substituting this $S(\tau,\omega)$ expression into Eq. (10b), we get a likelihood function of a and σ^2.

To actually implement this inversion process, however, we note that since the time-domain data are causal in nature, an approximate layer-peeling approach can greatly simplify the numerical optimization scheme. To accomplish this we represent the unknown functions $a(\tau)$, $\sigma^2(\tau)$ by piecewise linear approximations between the points (τ_i,a_i) for $a(\tau)$ and (τ_i,σ_i^2) for $a^2(\tau)$, where $a_i = a(\tau_i)$ and $\sigma_i^2 = \sigma^2(\tau_i)$. If we assume that $a(\tau)$, $\sigma^2(\tau)$ are known up to time τ_{i-1}, i.e., all a_j, σ_j^2 for $j < i$

are known, then the hyperbolicity of Eq. (6a) ensures that the power spectral density at time τ_i depends only on these known values plus the next a_i and σ_i. One thus reduces the optimization of a many-variable function to a series of simpler optimizations for a two-variable function.

We can also add a Bayesian[12] term $-\ln P_\beta(a_i | a_{i-1})$ to $L(a_i,\sigma_i^2)$ if there is some *a priori* knowledge about $a(\tau)$. For example, if we know that $a(\tau)$ is a slowly varying function and therefore a_i is most likely to deviate from a_{i-1} within a certain amount, then this information may be incorporated into our inversion scheme by letting

$$P_\beta(a_i | a_{i-1}) = \frac{1}{\beta a_{i-1}(2\pi)^{1/2}} \exp\left[-\frac{(a_i - a_{i-1})^2}{2(\beta a_{i-1})^2} \right] , \qquad (11)$$

where β represents the *a priori* knowledge about the function $a(\tau)$. If β is set to be very large, then the Bayesian term would have no effect. On the other hand, a suitable value of β in the joint minimization of $L - \ln P_\beta$ can improve the inversion accuracy.
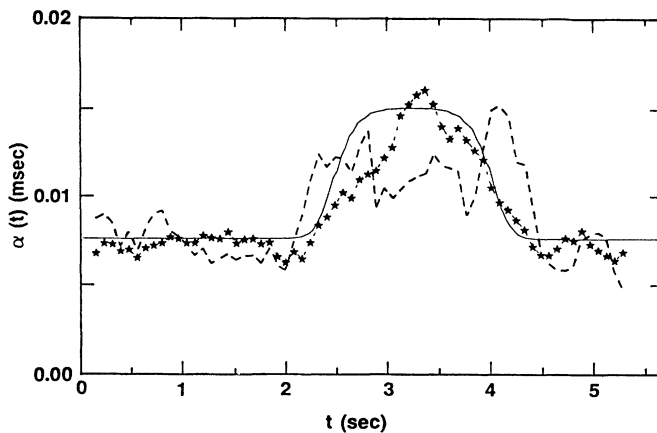
FIG. 2. Comparison of the exact profile, denoted by the solid line, with the inversion using one realization, denoted by the dashed line, and six realizations, denoted by the dashed-star line. The horizontal axis label τ is the two-way travel time. The $\alpha(\tau)$ of the vertical axis is defined by $\bar{a}/c_0(\tau)$, where $\bar{a}$ is given by Eq. (6b) and $c_0(\tau)$ denotes the main speed at the position of the two-wave travel τ.

We have implemented an efficient numerical scheme to calculate $S(\tau,\omega)$ from a given set of $\{a_i\}$. The maximization of α is done by a hybrid method due to Powell,[13] which is a combination of the steepest-descent and the general Newton's methods. By using the numerically calculated data, the inversion is performed with $\Delta\tau = 2$ (in units of pulse width $l_0/\bar{c}_0\epsilon$), $M = 6$-60 frequencies, and $\beta = 0.4$. The results are compared with the actual $\alpha(\tau)$ profile in Fig. 2. It is seen that in the initial flat region the inversion results hover around the correct value. The positions of the rise and the fall of the bump profile are both recovered fairly well, but the flat plateau is not accurately portrayed. In view of the fact that the structure is buried below 1000 scattering layers, and that the pulse scattering from the dip structure [calculated with bare $c_0(z)$ with no fluctuations] yields a signal that is only 10^{-3} times that of the multiple-scattering noise, this recovery is indeed remarkable and represents an extension of the previous inversion capability. Moreover, the probing pulse is noted to have essentially no low-frequency content comparable to that of the structure. Our results are also insensitive to additive white noise. In fact, setting σ^2 to zero or making it finite and large in the data yields practically identical inversion results.

The inversion accuracy can be improved systematically if there are multiple sets of statistically redundant data. This is possible in many instances since the correlation time of the small-scale fluctuations (e.g., in ocean and atomosphere) are usually short compared to those of the large structures. Therefore, by taking multiple data sets over a period of time in which many small-scale fluctuations can occur but during which the larger structures remain almost static, one can utilize the redundancy to configurationally average the data and thereby obtain a more accurate inversion of the large structure. This effect is illustrated in Fig. 2, where the inversion using data averaged over six configurations shows definite improvement over the single-configurational result. For the Earth's subsurface, such configurational averaging would also be possible if the small fluctuations have shorter transverse correlation lengths than the large trends so that redundant reflection data can be taken at multiple sites on the surface.

In summary, we have presented the first statistical inversion theory in which the knowledge of multiple scattering is incorporated as an integral part. The theory is demonstrated to work remarkably well. Since for the present theory one prefers to use strongly multiply scattered data so that the power spectrum can be well approximated by its generic behavior, our inversion method is thus complementary to the conventional approaches both in the selection of data as well as in the information sought.

[1]D. J. Pine, D. A. Weitz, P. M. Chaikin, and E. Herbolzheimer, Phys. Rev. Lett. 60, 1134 (1988).

[2]P. Sheng, Z.-Q. Zhang, B. White, and G. Papanicolaou, Phys. Rev. Lett. 57, 1000 (1986).

[3]B. White, P. Sheng, Z.-Q. Zhang, and G. Papanicolaou, Phys. Rev. Lett. 59, 1918 (1987).

[4]R. Burridge, G. Papanicolaou, and B. White, SIAM J. Appl. Math. 47, 146 (1987).

[5]P. Sheng, B. White, Z.-Q. Zhang, and G. Papanicolaou, in "Scattering and Localization of Classical Waves in Random Media," edited by P. Sheng (World Scientific, Singapore, to be published).

[6]R. Burridge, Wave Motion 2, 305 (1980).

[7]Z. S. Agranovich and V. A. Marchenko, The Inverse Problem of Scattering Theory (Gordon and Breach, New York, 1956).

[8]Relevant electromagnetic wave equations can be obtained from Eq. (2) by the following substitutions: $\epsilon/c \rightarrow \rho$, $c/\mu \rightarrow \kappa$, electric field $\rightarrow v$, and magnetic field $\rightarrow p$. Here c is the speed of light, ϵ is the dielectric constant, and μ is the magnetic permeability.

[9]A. Bayliss, K. E. Jordan, B. J. LeMesuier, and E. Turkel, Bull. Seismol. Soc. Am. 76, 1115 (1986).

[10]G. Papanicolaou, M. Postel, P. Sheng, and B. White, Wave Motion (to be published).

[11]E. Parzen, Technometrics 3, 167 (1961).

[12]W. Feller, An Introduction to Probability Theory and Its Applications (Wiley, New York, 1978), Vol. 1, p. 124.

[13]M. J. D. Powell, Harwell Report No. R.6469, 1970 (unpublished).