Geophys. J. Int. (1999) **136,** 295–308

Vertical profile of effective turbidity reconstructed from broadening of incoherent body-wave pulses—I. General approach and the inversion procedure

A. A. Gusev and I. R. Abubakirov

Institute of Volcanic Geology and Geochemistry and Kamchatka EMSD, 9 Piip Blvd, Petropavlovsk-Kamchatsky, 683006 Russia.
E-mail: ivgg@svyaz.kamchatka.su

Accepted 1998 May 20. Received 1998 May 12; in original form 1997 July 2

SUMMARY

A method is developed for the reconstruction of a non-uniform distribution of scattering properties in the upper layers of the Earth using data on broadening of an incoherent body-wave group or pulse along a number of rays. The theoretical basis for this reconstruction is a linear integral formula after Bocharov (1985, 1988), which is employed to design a linear inversion procedure. The inversion is performed in terms of a single scalar parameter of effective turbidity. This parameter presents an adequate generalization of the common turbidity parameter used in the isotropic scattering case; it describes, simultaneously, scattering attenuation, pulse broadening and backscattering or coda formation. As a preliminary step, necessary conditions of applicability of the transport equation approach for the analysis of regional high-frequency seismic waves are verified. A new compact derivation of Bocharov's formula is then presented. A linear least-squares inversion procedure for recovering a layered turbidity structure is proposed and tested on synthetic data of onset-to-peak delays of incoherent body-wave pulses. A few practical aspects of the application of the general approach to seismological data are analysed, including the correctness of the low-angle approximation, the use of peak delay observations instead of pulse centroid, the effects of a realistic spatial spectrum of inhomogeneity field, the potential bias produced by intrinsic loss, and the distortions produced by a non-spherical (double dipole) source radiation pattern. The latter point is considered as critically important, as one can expect significant data contamination by nodal arrivals. An efficient robust estimation procedure is designed and tested that is capable of suppressing distortions from nodal and near-nodal data.

Key words: inhomogeneous media, inverse problem, inversion, scattering, seismic waves.

INTRODUCTION

The spatial distribution of random-scattering properties of the Earth is a little-studied field of seismology. There is a general understanding of a fast decay of scattering capability with depth, based on the analysis of teleseismic *P* waves (e.g. Aki 1973; Flatte & Wu 1988) and regional *S* waves including codas (Rautian *et al.* 1981). Gusev (1995) proposed an approximately power-law decay of turbidity or scattering coefficient with depth, but this result was obtained in a somewhat indirect way, from the interpretation of coda shapes and attenuation.

Generally speaking, to determine the spatial distribution of turbidity from seismological data the entire wave envelope can be used, from an arrival up to late coda; in practice, however, either 'direct' (more accurately speaking, forward-scattered) waves, or backscattered waves (coda) are analysed. For 'direct' waves, scattering manifests itself most obviously as pulse broadening; that is, as an increase in the duration of the incoherent 'direct' wave group with distance. For coda, scattering manifests itself merely in its existence, and the amplitude of the coda immediately reflects the strength of scattering. To determine the scattering parameters in the first case, one can use the rate of pulse-width increase with distance (Gusev & Lemzikov 1983, 1985; Sato 1989); in the second case, one can use the relative coda amplitude normalized to the amplitude of the direct wave (Aki & Chouet 1975; Aki 1980). Using a model of a uniformly scattering medium to analyse the Kamchatka data, Gusev & Lemzikov (1983, 1985), and then later more

accurately Abubakirov & Gusev (1990), were able to demonstrate an acceptable agreement between the observational estimates of turbidity determined by these two approaches.

When studying a non-uniformly scattering medium, the pulse broadening of a 'direct' body wave is the most direct approach, since the broadening effect, being produced by forward scattering, reflects the properties of a tubular volume of the medium in the vicinity of the ray. This approach may permit the reconstruction of a spatial turbidity distribution based on the data on broadening for a multitude of rays. A similar problem is known in radioastronomy: the analysis of the broadening of radio pulses from pulsars can in principle yield a 3-D structure of the scattering plasma in the intergalactic medium (Bocharov 1987, 1990).

It would be very convenient for the application of such an approach if the scattering properties of the Earth could be specified by a single parameter describing both forward scattering manifested in pulse broadening and backscattering manifested in coda formation. Fortunately, such a parameter, namely effective turbidity g_e, can indeed be introduced. The theoretical background to the method that permits one to relate the value of the pulse width to the values of effective turbidity along a ray was developed by Bocharov (1985, 1987, 1988). His integral formula is linear with respect to the unknown profile of g_e along the ray; therefore, the arising tomographic problem is also linear.

This general approach will be developed in some detail below. Attention is given to many critical questions that should be addressed when one tries to apply the theoretical results, obtained with assumptions that are not quite realistic, to the interpretation of real seismological data. In the companion paper (Gusev & Abubakirov 1999), the new technique is applied to the inversion of data on the pulse broadening of body waves from local Kamchatka earthquakes as expressed in onset-to-peak delay time.

In a preliminary study, Gusev & Abubakirov (1996a) made the first successful attempt to perform the inversion of vertical effective turbidity structure from pulse delays. Compared to the method described there, the present work has some radical improvements, in particular, realistic rays instead of straight ones are used, and residual-dependent data weighting is applied.

ON EFFECTIVE TURBIDITY

Before formulating the inverse problem for effective turbidity, one must be able to solve the forward problem; that is, to determine the expected parameters of the pulse from given scattering properties of the medium. A problem arises: in terms of which parameter of the medium should one pose the forward and inverse problems? Analyses of scattering based on coda waves have frequently been carried out assuming an isotropically scattering medium; that is, the radiation pattern/directivity of scattered energy (known under the compact and convenient name of 'scattering indicatrix' in optics) is assumed to be spherically symmetric. The single parameter of such a model is the value of the turbidity (scattering coefficient) g, or of the mean free path $l = 1/g$.

These parameters are defined through the loss of energy to scattering for a plane monochromatic wave. For a wave propagating along the z-axis of a Cartesian coordinate system, the amplitude $A(z)$ of a direct wave decays according to the equation

$$A^2(z) = A_0^2 \exp(-gz) = A_0^2 \exp(-z/l) = A_0^2 \exp(-2\pi fz/cQ_{sc}), \tag{1}$$

where $A_0 = A(0)$, c is the wave velocity, f is the wave frequency, and Q_{sc} is the quality factor related to scattering loss. In the study of non-stationary problems, such as pulses, one should imagine a finite frequency band, of width Δf, around f. The inverse of Δf is the characteristic time t_Δ. When Δf is sufficiently wide, t_Δ is small and often can be neglected in comparison with the other temporal parameters of the problem; then (1) still holds. In practice, we are interested in the application of our results to the interpretation of high-frequency seismic body waves, so f is in the range 0.5–30 Hz. All discussion will be developed for the acoustic case, neglecting $P \to S$ or $S \to P$ conversion.

In terms of inhomogeneity structure of the medium, the case of a spherical indicatrix corresponds to a very small scatterer/heterogeneity size $a \ll 1/k$, or $ka \ll 1$, or $a \ll \lambda$ ($k = 2\pi f/c$, $\lambda = c/f$ is the wavelength). The a parameter in this inequality defines the correlation distance of the random heterogeneity field; the particular form of correlation function is irrelevant in this case. This assumption of small a is unlikely to hold for the Earth, where we can expect a wide spectrum of heterogeneity sizes. The inadequacy of the isotropic scattering model is manifested most clearly by its inability to predict any rate of pulse broadening with distance: according to this model, the direct pulse must be delta-like at any distance where it is above the noise. This means that the observed duration of the body-wave group must be short (determined by the source process duration of an event) and independent of distance, in contradiction to observations. Therefore, one must reject the isotropic scattering model and pass to more complicated ones, with non-spherical indicatrix; that is, with non-isotropic scattering.

How should the parameter g be generalized for these more realistic models? Let us first consider the case where the opposite inequality, $ka \gg 1$ or $a \gg \lambda$, is true. This corresponds to large-scale inhomogeneities, and the scattering indicatrix is a narrow lobe along the wave vector of an incident ray (prominent 'forward' scattering) with an angular width of the order of $1/ka$. A well-known example of an inhomogenity field is one with the Gaussian autocovariance function (referred to as the Gaussian-ACF case):

$$\langle \varepsilon(x)\varepsilon(x+d)\rangle = \langle \varepsilon^2 \rangle \exp(-d^2/a^2), \tag{2}$$

where $\varepsilon(x) = (c(x) - \langle c \rangle)/\langle c \rangle$ is the fractional wave-velocity perturbation and $\langle \varepsilon^2 \rangle$ is its mean square ($\langle\ \rangle$ represents ensemble average here and below). For this field, in the case $ka \gg 1$, the mean square angle of scattering is

$$\theta_0^2 = 2/k^2a^2. \tag{3}$$

In the case of forward scattering, the broadening of an incoherent wave pulse with distance is the characteristic property of the problem. To discuss it, we introduce a simple integral duration parameter of the mean delay (centroid):

$$\langle T \rangle = \int_{t_d}^{\infty} (t - t_d) E(t)\, dt \Big/ \int_{t_d}^{\infty} E(t)\, dt, \tag{4}$$

where t_d is the direct wave arrival time and $E(t)$ is the mean square amplitude, or instantaneous power of signal within the

band Δf. In the simplest case of low-angle scattering, point source and a uniformly scattering Gaussian-ACF medium, Williamson (1972) found that

$$\langle T\rangle = \frac{D}{12c} r^2, \tag{5}$$

where r is the distance travelled and D is 'the mean square angular dispersion of the ray per unit distance' parameter employed by Williamson. The 'ray diffusion coefficient', employed in a similar context by Chernov (1975), is $D_{\mathrm{Ch}} = D/4$. Both parameters describe the medium's capability of scattering, and neither is quite suited to our aims. Thus we define here the 'effective turbidity' parameter as $g_{\mathrm{e}} = 2D_{\mathrm{Ch}} = D/2$, so that

$$\langle T\rangle = \frac{g_{\mathrm{e}}}{6c} r^2. \tag{6}$$

Therefore, the value of $\langle T\rangle$ increases as the distance squared, and the intercept of this trend may be used for the practical determination of g_{e}.

It should be noted in connection with this and further similar results related to multiple low-angle scattering, that the 'low-angle' condition is set with respect to the entire cumulative deflection angle, and not to the scattering angle in a single act of scattering. However the quadratic trend (6) holds true well outside the strict limits of the 'low-angle' condition: numerical modelling by Gusev & Abubakirov (1996b) showed it to be approximately true up to sufficiently large distances, when the cumulative angle is of the order of 1 radian. Also, the approximately quadratic mode of the trend is not limited to the case of Gaussian ACF, and applies to any narrow indicatrix with finite second moment ($\theta_0^2 < \infty$).

The notion of effective turbidity ['equivalent isotropic turbidity' of Gusev & Lemzikov (1983, 1985)] is of key importance. It should not be confused with the common, 'true', non-isotropic turbidity g_{n}, defined by eq. (1) if one replaces g by g_{n}. To introduce a pair of parameters like g_{n} and g_{e} is not novel at all: they are in fact identical to two concepts very well known in neutron transport theory: those, respectively, of the 'true' and 'transport' cross-sections of a unit volume of the medium. We will also use effective and 'true' mean free paths $l_{\mathrm{e}} = 1/g_{\mathrm{e}}$ and $l_{\mathrm{n}} = 1/g_{\mathrm{n}}$.

As was noted by Abubakirov & Gusev (1990), the mean cosine of cumulative angular deflection θ of a ray decreases with distance r as

$$\langle \cos\theta\rangle = \exp(-2g_{\mathrm{e}} r) = \exp(-Dr). \tag{7}$$

Thus, after propagating a distance of the order of $r_0 = l_{\mathrm{e}}$, the rays suffer significant angular deflection (of the order of one to two radians). Therefore, $r_0 = l_{\mathrm{e}}$ is the critical distance. Up to r_0, the wave energy propagates within some gradually widening tube or, rather, 'horn', along the initial ray. At distances of the order of r_0 and larger, the low-angle approximation breaks down. As the rays continue to propagate, they soon 'forget' their initial direction and begin to wander almost isotropically [see Gusev & Abubakirov (1996b) for examples of both stages]. In the asymptotic case of large propagation time $t \gg l_{\mathrm{e}}/c$, the behaviour of rays is essentially a sort of random walk. In this case, the theory (Ishimaru 1978) predicts the space–time distribution of wave energy according to the diffusion law, similar to the case of isotropic scattering. The asymptotic energy density distribution for this case coincides with that for the case of isotropic scattering, with the value of g equal to

$$g_{\mathrm{e}} = g_{\mathrm{n}}(1 - \langle\cos\theta\rangle), \tag{8}$$

where g_{n} is the already introduced common, non-isotropic scattering coefficient or turbidity, and $\langle\cos\theta\rangle$ is the average cosine of the scattering angle:

$$\langle\cos\theta\rangle = \int_{4\pi} \cos\theta\, m(\Omega)\, d\Omega, \tag{9}$$

where $d\Omega$ is solid angle element, and $m(\Omega)$ is the indicatrix function mentioned above, scaled so as to produce the integral of the value unity over the entire sphere (and therefore to represent the probability density for a scattered ray direction).

Note that for a narrow indicatrix ($ka \gg 1$), $1 - \langle\cos\theta\rangle \approx \langle\theta^2\rangle/2 = \theta_0^2/2$, and

$$g_{\mathrm{e}} = (\theta_0^2/2) g_{\mathrm{n}}. \tag{10}$$

[For the Gaussian-ACF case (see eq. 2) this gives $g_{\mathrm{e}} = g_{\mathrm{n}}/k^2a^2$.] Now imagine two model Gaussian-ACF media with different but narrow indicatrices, such that the values of a and g_{n} are different, but the values of g_{e} are equal. Then (1) the trends of mean pulse delay with distance (and even pulse shapes for the multiple scattering case) will coincide in the two media; and (2) the same source will produce the same level of (late) coda in both media. A difference between the full envelopes for the cases with different $g_{\mathrm{n}}/g_{\mathrm{e}}$ ratios will still be present, but it will be manifested only in 'early' codas and usually is not easy to observe [see results of the numerical modelling in (Gusev & Abubakirov 1996b)]. This explains why we consider g_{e} as the main practical parameter for the specification of a scattering medium. Note that by using the two types of observations (pulse delay and coda level), one can try to estimate this parameter via two different, independent procedures.

Now let us consider more general and more realistic situations when either ka is comparable to unity, or the Earth medium cannot be characterized by a single characteristic heterogeneity size a. An example is a self-similar fractal inhomogeneity field. The limit of applicability of the g_{e} concept in such a case is yet to be demonstrated rigorously, but is expected to be fairly wide. To produce clear pulse broadening, it is sufficient that the real indicatrix be moderately elongated, and this assumption seems to be in qualitative agreement with the observed coda shapes and observed broadening phenomena (see discussion in Gusev & Abubakirov 1996b). The comparison of modelled and observed records performed in that study suggests that the characteristic width of a real indicatrix is about 30°–40°.

Despite its key role in the theory, the true turbidity parameter g_{n} is seemingly much more difficult to determine from observations than g_{e}. This could, in fact, be expected: g_{n} accounts for scattering into all angles, including very small ones, in which case it is practically impossible to distinguish between direct and scattered wave energy. For example, if there is a hundred-fold increase in the amount of scattering into all angles below 10^{-4} rad, the value of g_{n} will be changed radically, but g_{e} will be modified only slightly, because the contribution of small angles to the integral (9) is small (as $d\Omega = \sin\theta\, d\theta\, d\phi$, and $\sin\theta$ is small at small θ).

One of the best-known manifestations of scattering is scattering attenuation. Let us discuss how the effective turbidity g_{e} parameter is related to this phenomenon. We note immediately that if scattering losses are defined in the 'seismological'

manner, using the integral over the whole observed pulse to calculate energy and to judge its decay, then any small deflections of the ray that do not cause a decrease in this energy are irrelevant to such 'seismological' loss. In such a case, the example described in the previous paragraph shows not only that the value of true turbidity g_n is difficult to measure, it also means that this parameter is irrelevant to losses and therefore is useless for specifying attenuation. In the isotropic scattering case, however, the value of $g = g_e = g_n$ defines scattering loss in quite a consistent way. Recalling that both $1/g$ and $1/g_e$ are the distances at which the ray forgets its initial direction and is deflected by an angle of the order of 1–2 rad, it becomes clear that the g_e parameter is the natural candidate for describing scattering losses.

Note that the implicit, practical seismological definition of loss is not strict: it does not specify accurately either the time interval for integrating pulse energy or the critical angle such that deflections in excess of it are considered 'loss'. One seemingly reasonable outcome is to consider as lost all energy deflected by more than 90°. Unfortunately, this definition disagrees with the reference isotropic scattering case: in this case, when the distance is equal to $l = 1/g$, loss is just 1/e by definition. However, the lost energy is distributed evenly over frontal and back hemispheres, and hence only half of it goes to the back hemisphere, in contradiction with the possible generalized definition given above. The problem is not of key importance because it relates to the coefficient of the order unity. Despite the fact that some details of this sort are not yet settled, one can quite safely consider g_e as the parameter approximately representing scattering loss. [The only problem may arise at small dimensionless ('optical') distances $L_e = g_e r$ where, in the case of low-angle scattering, seismological loss is not proportional to L but represents an infinitely small value of the order higher that L, which is, in practice, equal to zero. In seismological applications, however, optical distances are usually comparable to unity, and also the assumption of low-angle scattering is far from being correct, hence this problem is not very important.]

To sum up, the single g_e parameter defines three main properties of real, non-isotropically scattered high-frequency seismic waves: pulse broadening, relative coda level and 'seismological' scattering loss. It is natural to formulate the inversion problem in terms of this parameter.

RADIATION TRANSFER/ENERGY TRANSPORT THEORY AND STOCHASTIC RAY PATHS: THEORETICAL BACKGROUND FOR THE FORWARD PROBLEM

In seismology, a stochastic representation of wave propagation phenomena, in particular the description of wavefields by their energy instead of amplitude, is usually done on a purely empirical basis, despite several decades of history. Such a representation has, however, a long tradition in optics, acoustics and radiophysics, and there are well-defined criteria for its applicability. Before any data analysis in terms of random waves and random media is performed, it is a good idea to check whether these criteria are satisfied.

Thus, we shall try to check whether the propagation of high-frequency seismic waves through the complex Earth medium can be treated using the energy transport equation (ETE). See texts such as Ishimaru (1978) or Rytov *et al.* (1978) for details on this classic equation, initially introduced by Boltzmann for diffusion of particles. This equation is referred to in optics of stellar or planetary atmosphere as the 'radiation transfer equation', and is known in neutron physics as the 'neutron transport equation'. The ETE is an effective approach if one wishes to substitute the general wave equation treatise of a wave propagation problem by the study of wave energy propagation (that is, one neglects phase information, assuming the waves to be incoherent, the phases random and the energy additive). It should be noted that most problems solved with ETE in optics, neutron theory and radiophysics are stationary, hence the use of ETE in seismology, where non-stationarity is a characteristic property of the problem, leads to a number of problems unexplored within traditional applications of ETE.

Rytov *et al.* (1978) list the following necessary conditions that are to be satisfied for ETE to be applicable:

$$k \gg 2g_n \quad \text{or} \quad \lambda \ll l_e/\pi \quad \text{or} \quad Q_{sc,n} \gg 0.5, \tag{11a}$$

$$k^2 a^2 \langle \varepsilon^2 \rangle \ll 1, \tag{11b}$$

$$a \ll C_1, \tag{11c}$$

$$a \ll C_2, \tag{11d}$$

where $Q_{sc,n}$ is defined by eq. (1) when g is replaced by g_n, and C_1 and C_2 are the characteristic scales describing the spatial variation of wavefield intensity along and across the ray. To check inequalities (11a) and (11b) we will assume the ACF of inhomogeneities to be Gaussian. This is permissible for rough estimates because, despite the presence of many sizes of inhomogeneities in the medium, for the neighbourhood of any particular frequency f the most important contribution is the scattering due to inhomogeneities of a size close to the wavelength (Wu & Aki 1985). Let $\theta_0 = 0.5$ rad in agreement with observations (Gusev & Abubakirov 1987), then $ka \approx 3$, or $a \approx 0.5\lambda$. At $\theta_0 = 0.5$, $g_e/g_n \approx 0.5$. With this value and $g_e \approx 0.01\ \mathrm{km}^{-1}$ (Sato 1978), the inequality (11a), at $c = 5\ \mathrm{km\ s}^{-1}$, gives $f \gg 0.013$ Hz. This condition holds for scattered body waves at regional distances.

As for the inequality (11b), at $ka = 3$, even with $\langle \varepsilon^2 \rangle$ set to its practical upper bound of (10 per cent)^2, it gives $0.09 \ll 1$ which is fully acceptable. In (11c), C_1 is of the order l_e, and because of $a \approx 0.5\lambda$, (11c) holds simultaneously with (11a), which has been already checked. Finally, in (11d) C_2 is of the order r (for a radiator of double-dipole type). Therefore the condition (11d) can be violated near the source at distances r of the order of a, that is of λ. This distance is too small for the usual conditions of regional observations. Hence, the necessary conditions are satisfied and the assumption of applicability of ETE for the analysis of regional body waves can be considered reasonable.

We have already noted that ETE describes both the propagation of random waves and the propagation of particles in a medium with random scatterers (Ishimaru 1978); thus, in theory, the choice of description of wave propagation phenomena, through intensity of radiation or through the density or flux of particles, is a matter of convenience only. The description through particles is the basis of the numerical Monte Carlo method (for example Gusev & Abubakirov 1987, 1996b), but it can be successfully applied for analytical calculations as well (Williamson 1972). An important and non-trivial fact, which should be emphasized, is that the description of the propagation of random waves by a flux of particles is

a formally well-founded approach and not some rough estimation scheme. This fact is related to the strong analogy between wave and particle ('photon' or 'phonon') propagation, discussed formally in, for example Williamson (1975), Uscinsky (1977), Dashen (1979), Bocharov (1985, 1990). For the low-angle case (or within the parabolic approximation), this analogy has been demonstrated formally to be an equivalence.

Therefore, the parameters of the scattered body-wave pulse may be determined as parameters of the distribution function of the moments of arrival of delta-like pulses/particles along different randomly shaped rays; each of this multitude of rays is formed by one single realization of a random medium. If a ray/particle trajectory can be considered as a smooth random curve, then the corresponding approach is named 'the ray diffusion technique' (Chernov 1975). However, the assumption of smoothness is not critical. Rays/particle trajectories can be assumed piecewise linear instead of curved, as was proposed by Williamson (1972, 1975) under the special name of the 'stochastic ray-path method'. In a difference from the ray diffusion approach, Williamson (1975) discusses only the case of low-angle scattering. However, the correspondence between a full wave description, ETE and a description through an ensemble of particles or piecewise-linear rays is of general validity and is not limited to small angles. The case of small angles is nevertheless very important, because in this case analytical results can be obtained, whereas in the general case numerical modelling should be used. Below we shall combine these two approaches.

BOCHAROV'S FORMULA AND AN EXAMPLE OF ITS APPLICATION

Andrei Bocharov of Moscow derived a formula for the mean delay $\langle T \rangle$ of a pulse radiated by an instant point source and propagating through a non-uniformly scattering medium, assuming low-angle scattering. He developed the stochastic ray-path approach of Williamson (1972), who had already studied the cases of constant-g_e medium and of a scattering layer. The final formula was published in Bocharov (1988), but its systematic derivation, rather lengthy, was published only in preprint form (Bocharov 1985, 1987). In the Appendix we present a short new proof for Bocharov's formula. The final result for mean pulse delay $\langle T \rangle$ along a ray of length S can be written as

$$c\langle T \rangle = F = \frac{1}{S} \int_0^S g_e(u)(S-u)u \, du, \tag{12}$$

where the integral can be evaluated along either a perturbed or an unperturbed/mean ray (the difference is negligible within the approximation used). It contains one and only one medium parameter of effective turbidity, and is valid for low-angle scattering of arbitrary mean multiplicity. An additional requirement is that g_e must change negligibly over a distance comparable to the value of the side shift of the ray; that is, within 'the mean ray tube' sounded by random rays, which seems rather natural. The 'weighting function' factor under the integral to modify $g_e(u)$ is a symmetric inverted parabola giving small weight to contributions from the ends of the ray, and a large weight to its middle part. For constant g_e, (12) reproduces Williamson's result (6). It should be noted that the integral representation for F also holds for the case of variable velocity. Therefore, for the two changes that a pulse undergoes during its propagation through a non-uniformly scattering medium, the broadening is described by the integral (12), while scattering attenuation is related to another integral, the effective optical length of the ray:

$$L_e = \int_0^S g_e(u) \, du. \tag{13}$$

This integral also predicts cumulative ray deflection. The L_e parameter replaces the gz combination in (1) and the $g_e r$ combination in (7). In the case of isotropic scattering, L_e is identical to the mean multiplicity of scattering.

Bocharov's formula is linear, but its predictions are nevertheless sometimes far from evident. In order to acquire some understanding of the general properties of pulse broadening in a non-uniformly scattering medium let us investigate some simple idealized case. We choose the particular case of a receiver located on top of a scattering layer of thickness H and effective turbidity $g_{e,l}$ that overlays a transparent half-space, and a series of model sources located along the vertical ray directed downwards from the receiver. The medium velocity is set constant, equal to 4 km s^{-1}. Fig. 1 illustrates the application of Bocharov's formula to this structure, in common and log–log scales. For ray hypocentral distances $r < H$, the curves are based on eq. (6), and for $r > H$ on the theoretical formula which is implied from (12) for this case:

$$\langle T \rangle = \frac{g_{e,l} H^2}{cr} \left(\frac{r}{2} - \frac{H}{3} \right). \tag{14}$$

The critical value of $\langle T \rangle$, for a source situated just on the lower boundary of the layer, is denoted by a symbol on each graph of Fig. 1.

The first row shows the results for a fixed layer thickness of 50 km and varying scattering in the layer: $g_{e,l} = 0.0071$, 0.02 and 0.071 km^{-1}. At a fixed distance, $\langle T \rangle$ is proportional to $g_{e,l}$, as expected. What is unexpected is the remarkable increase of $\langle T \rangle$ during propagation through the non-scattering medium (asymptotically, threefold at large r), as immediately follows from eq. (14). This effect looks counterintuitive. However, the result is accurate, and the value of the asymptotic limit has already been obtained by Williamson (1972). Williamson and Bocharov use a common assumption: that the initial problem with random rays connecting two points can be replaced by the assumedly equivalent problem of rays starting at a fixed direction and arriving at a spherical surface. We checked the possibility that this assumption could fail in the case studied. Fortunately, for the asymptotic case of a very thin layer, one can perform the whole derivation without this assumption, and arrive at the same result of three-fold difference.

The actual origin of the problem is when the correct idea that a transparent medium does not affect energy loss [represented by the integral (13)] is applied (unlawfully) to the integral (12) of a different structure. The quadratic increase of mean delay with distance, as compared to the linear increase of loss, should have been taken as a warning signal. The real cause of the problem is easy to find: the 'horn' formed by randomly perturbed rays in the layer continues to widen as the receiver moves deeper into the transparent half-space; thus the average side deflection of the ray increases, and the mean delay with it. Generally, all this means that, whereas the contributions of segments of the ray into the $\langle T \rangle$ value are indeed additive because of the linear structure of the integral (12), such

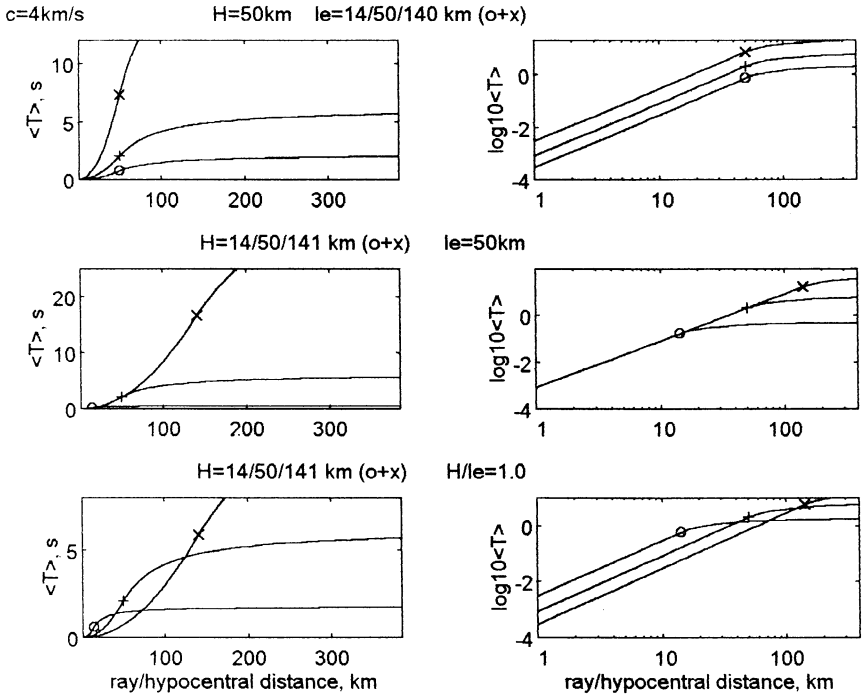

Figure 1. The behaviour of mean peak delay observed by a receiver on top of a layer of constant effective turbidity of thickness H underlain by a transparent half-space, from a series of sources distributed along a down-going ray. Right column: natural scale; left column: log–log scale. Top row: three values of effective mean free path/turbidity and a constant layer thickness of 50 km; middle row: three values of the layer thickness and a constant effective turbidity of 1/50 km; bottom row: three values of layer thickness, and identical values of optical thickness for these layers—$L_e = H/l_e = 1.00$.

contributions for the integral (13) cannot be calculated along each segment independently and then added: this simple kind of additivity does not hold. It should be noted that this lack of simple additivity has an important outcome for the analysis of g_e values obtained from an inversion: if one compares the g_e estimates for a thick layer with similar estimates for its sublayers, the sum of the latter need not be near the former (as might be expected for, for example, g_e values estimated from amplitude decay).

The second row illustrates how the values of $\langle T \rangle$ depend on distance: they increase as the square of distance travelled within the layer, and farther saturate as discussed above. The third row shows the effect of layers of a fixed effective optical length $L_e = g_{e,1}H$, but of different H. Here, again, the difference in nature between integrals (12) and (13) is clearly manifested. The L_e parameter does not immediately determine pulse width: at constant L_e, $\langle T \rangle$ grows linearly with H. This is an important fact: whereas both $\langle T \rangle$ and L_e increase with distance, the precise modes of increase are different. In the primitive case of uniform effective turbidity, $\langle T \rangle$ increases as distance squared, whereas L_e grows only linearly. As these two manifestations of scattering are not proportional to one another, integral loss cannot be judged from integral pulse delay, and formal inversion is needed to relate them.

PRACTICAL ASPECTS OF THE APPLICATION OF BOCHAROV'S FORMULA TO THE DATA ANALYSIS

Bocharov's formula (12) has been derived for an ideal case; when it is applied to the analysis of real seismological data several problems are encountered, to be discussed below.

Curved mean rays

To apply the derived formula in seismology, one must move from constant to variable mean velocity, and thus from straight unperturbed rays to curved ones. This does not seem to produce serious problems: the initial geometry with a straight unperturbed ray can be transformed into one with a smoothly curved ray producing errors that are small in the first approximation. The formula for F remains valid, but now we cannot simply set $\langle T \rangle = F/c$. However, as variations of velocity are moderate in seismological applications, to obtain a reasonable approximation we can merely substitute c by the ray-averaged velocity.

Multi-ray propagation

The general approach of the present paper is complementary to the common representation of the Earth's inhomogeneity by flat layers. Therefore, in some cases, the traditional explanation of pulse shapes based on multi-ray propagation within a layered average structure is a competitor to the present approach. There are cases, however, when the results of an analysis in terms of a random structure that ignores the deterministic ray propagation in a layered structure may be completely misleading. An example case among regional observations is when, within the window chosen for the analysis, we observe phases such as P_n followed by P_g, or S_n followed by L_g. The arrival of a fast head wave gives an incorrect reference to judge the value of the delay of wave energy that propagated, mainly, along lower-velocity paths. In order to avoid this kind of difficulty, in the practical inversion below we will use only uprising rays.

Low-angle scattering and real scattering angles

A small degree of deviation of rays from their unperturbed position is the formal condition of the applicability of the present theory. When studying pulse broadening one can verify this condition directly. The pulse width should simply be much smaller than the traveltime t_d, and this can be checked immediately with real data. It should be realized, however, that in this case, as well as in many other similar cases, the real domain of applicability of the asymptotic theory is much wider. As can be seen from the results of numerical modelling (Gusev & Abubakirov 1996b), even values of optical distance comparable to unity are quite acceptable.

Relation between $\langle T \rangle$ and measured parameters of the pulse

To determine the observed value of the $\langle T \rangle$ parameter one could analyse the digital records directly. We are interested here in a technique that is applicable both to analogue and digital data. In this case, it is convenient to measure the delay time between the arrival and the peak of a body-wave group, denoted t_m. To relate this parameter to $\langle T \rangle$, assume that the average value of t_m, or $\langle t_m \rangle$, coincides with the delay of the maximum of the theoretical mean pulse shape function. For the case of a uniformly scattering medium, large multiplicity, and low-angle scattering, the relationship between $\langle T \rangle$ and $\langle t_m \rangle$ can be determined using Williamson's (1972) formula for the pulse shape function: the result (Gusev & Abubakirov 1996b) is

$$\langle t_m \rangle / \langle T \rangle = 6/10.8 = 0.55 . \quad (15)$$

In practice, however, one cannot lean upon the assumption of Gaussian-ACF medium, nor that of large multiplicity, nor that of low-angle scattering. We can use, however, the results of numerical modelling, for example, those after Gusev & Abubakirov (1996b). Comparing these results to observed body-wave envelopes, these authors have found that the most probable model of a random scattering inhomogeneity field can be specified by the power-law k-spectrum, with the value of the exponent γ equal to 3–4. In this case, the deviation of pulse shapes from Williamson's (1972) formula can be significant. At a fixed value of the mean delay, the particular manner of the shape difference is a shift of the position of the peak for the power-law case to earlier times (to the onset). As a result, t_m decreases by a factor of 1.7 for $\gamma = 4$ and by a factor of 3 for $\gamma = 3$, compared to the low-angle Gaussian-ACF case of Williamson (1972).

With practical inversion in view, this shift of the peak can be approximately accounted for by introducing a correction coefficient. The most probable value of γ is near to 3.7 (Sato 1990; Gusev & Abubakirov 1996b); the corresponding value of the correction coefficient is about 2.0, and this value may be used for data analysis. There is an additional complication as well, related to the fact that the ratio $\langle t_m \rangle / \langle T \rangle$ depends on the actual degree of non-uniformity of scattering properties along the ray. Judging by the limiting case of a thin random phase screen in a non-scattering medium (Williamson 1972), non-uniformity may cause some additional decrease of t_m with respect to $\langle T \rangle$. (In the present study we neglect effects of this kind.) For all these reasons, we cannot fix an accurate value of the correction coefficient at present. In such a situation we prefer to use the relationship (15) for the inversion, and thus to obtain apparent estimates valid for the reference, Gaussian-ACF medium model. In data interpretation, however, we will introduce the above-mentioned correction coefficient, its value set to 2.0.

One can ask whether our approach is greatly handicapped by using a seemingly 'primitive' t_m value instead of a 'strict' mean delay (as the direct estimate of the parameter $\langle T \rangle$). It should be emphasized that at least in the case of regional observations, both the individual measured t_m value and the mean peak delay $\langle t_m \rangle$ in general are the types of parameters that have important advantages as compared to $\langle T \rangle$ when specifying broadening of an observed pulse. The reason is that observed (and even numerically modelled) estimates of $\langle T \rangle$ tend to become unstable, because they include integration (4), performed in practice along rather slowly decaying coda multiplied by time. Any erratic spikes in the coda can and do distort the estimate of $\langle T \rangle$ radically. In contrast, the t_m values are robust, being insensitive to this kind of bias. Their larger scatter can be compensated by an increase in data volume.

Possible bias caused by intrinsic loss

Intrinsic loss shortens the pulse, thus reducing the pulse broadening caused by scattering (Sato 1989). For the Gaussian-ACF medium with Williamson's pulse shape, this effect can be easily estimated, and its magnitude, for this particular case, is not large for typical observations. For more realistic media with power-law inhomogeneity spectra, the decrease of the peak delay is considerably smaller than for Gaussian-ACF media of similar effective turbidity. (The cause is that for shorter pulses, the relative pulse shortening produced by intrinsic loss is smaller.) Also, for such analysis one should use the correct values of intrinsic Q (Q_i): neither the Q_i estimates based on the interpretation of coda in assumedly uniformly scattering media, nor the Q values recovered from body-wave attenuation data will do. Both these kinds of estimates represent the summary effect of scattering and intrinsic losses, and therefore can serve as no more than upper bounds for the latter; in fact they may significantly overestimate it (Dainty 1981; Gusev & Abubakirov 1996b).

On the constant term in observed t_m values

It is common to view a seismic signal as a result of consecutively acting linear operators, each describing one of the following processes or effects: earthquake source; propagation path (with excluded near-station effects); near-station structure; and instrument filter. This representation implies that each of these (physical) operators makes an additive non-negative contribution to the first power moment (centroid, $\langle T \rangle$) of the pulse (as the first moment of convolution is a sum of those for operands). Approximately, the contributions to other measures of pulse broadening, such as t_m, will also be additive. In an analysis of the effective turbidity of a medium we are interested solely in the path effect, so that source, near-station and instrument terms should be treated as biasing factors. The instrument term can be estimated from the system response (and then subtracted, as was done in Gusev & Lemzikov 1983, 1985). The source term can be roughly estimated based

on corner frequency, or on visual signal period. However, the near-station term presents a difficult problem. Its effect is often hard to notice when studying seismic station data because the art of choosing a good station location is, to a large degree, just the art of finding a spot with negligible near-station resonances. With field data, near-station effects on pulse duration are often clear and prominent. All this means that, for completely reliable inversion, one should preferably estimate either the station term, or merely the entire constant term directly from data. For 'good' stations, the station term must be relatively small and the total additional delay is close to the value estimated from the source and instrument effects combined. With some risk, one can use this estimate to constrain the inversion.

On the effect of a non-spherical source radiation pattern

The theories of Williamson (1972) and Bocharov (1988), as well as the numerical results of Gusev & Abubakirov (1996b), are valid, strictly speaking, only for sources with spherical symmetry (isotropic). More accurately, the radiation pattern must vary only slightly over angles of the order of the cumulative angle of multiple scattering. This may be acceptable in astrophysics, but clearly is not true for small earthquakes at regional distances, where cumulative scattering angles are of the order of unity, and double-couple sources have radiation lobes of comparable angular size. The most unpleasant case is when the station (component) is in a nodal direction: the direct arrival is then absent, and forward-scattered energy comes from neighbouring directions only. This energy will be the most depressed for small delay times, and with increasing time, more and more energy will leak to this ray direction. (Finally, late coda will arrive, without any observable amplitude depression: this is the reason for its high efficiency for magnitude evaluation.) This mechanism will make the observed peak position strongly delayed compared with the case of an isotropic source.

A qualitatively similar but less pronounced picture is observed for rays that are nearly nodal. Thus peak delay data from small earthquakes, as compared with the ideal case of an isotropic source, will look as if they are contaminated by a considerable proportion of large positive errors ('outliers'). For *P* waves, as compared to *S* waves, the holes in the radiation pattern are more prominent and the apparent contamination can be expected to be more pronounced.

To process data of this kind one can employ a robust estimation method. The method selected is a modification of the 'uniform reduction' technique of Jeffreys (1961); see also Mosteller & Tukey (1976). In this technique, a weight is ascribed to each least squares equation that depends on its residual, and the weighting procedure is applied in an iterative manner until weights stabilize. The particular weighting function we use is

$$W(\delta) = 1/[(1 + q(|\delta|/\sigma)^{\beta}], \tag{16}$$

where δ is the residual, σ is the rms residual determined at the previous iteration, and q and β are adjustable parameters. The parameter q was set to 2.3; that is, to the upper 1 per cent quantile of the normal law, and hence when the residual is as large as 2.3σ, its weight is reduced to 0.5. By manipulating β we define how gradual or how abrupt is the switching of the value of the weight from about unity at small residuals to about zero at large ones. Setting $\beta = 4$ we achieve a compromise: switching is sufficiently abrupt so that small deviations are hardly suppressed, whereas large deviations are suppressed strongly. It should be noted that in this case the value of the average weight over all data can be viewed as the fraction of non-suppressed data. On the other hand, switching is sufficiently gradual to block prohibitively slow convergence of the iterative procedure. The described weighting scheme does not affect the 'regular' sub-population of data, and is capable of suppressing a large proportion of asymmetric outliers. The neglect of this point was the main cause of the bias that distorted the estimates of Gusev & Abubakirov (1996a).

THE INVERSION ALGORITHM

Based on the above approach, we designed a data processing procedure aimed at the practical inversion of the vertical profile of effective turbidity g_e under a seismic station. We assume that the data are the t_m measurements for a number of small earthquakes, with their hypocentres below the station. The inversion procedure begins with choosing some parametric model of effective turbidity structure. For a given vector of parameters, one can calculate theoretical estimates of data from Bocharov's formula and compare them to observations. Inversion is then reduced to the choice of an optimal parameter vector providing the best fit between theoretical and observed values in terms of some criterion, in the simplest example that of least squares. Bocharov's formula is linear with respect to $g_e(s)$. This makes it possible to formulate the inverse problem as a linear least-squares problem, and we will discuss this case.

Let us write down the equations that present the basis for inversion. Let $t_{m,j}$ $(j = 1, 2, \ldots N)$ denote the jth observed t_m value, and let the theoretical expression that relates t_m to the vector of (unknown) parameters $\mathbf{p} = \{p_i\}$ of a particular parametric model be written as $T_j(\mathbf{p})$. Then we can write down the equation that relates $t_{m,j}$ to unknown $\mathbf{p}$ as

$$T_j(\mathbf{p}) + \kappa_j = t_{m,j}, \tag{17}$$

where κ_j is the true misfit that combines random error and model inadequacy. However, if one departs from equations of this kind and uses no *a priori* weighting, one implicitly ascribes the same accuracy to t_m measurements made at different distances, and this assumption is incorrect. The mean value of t_m is expected to increase as the distance squared in a uniform medium. In such a case, the assumption of constant relative error is much more reasonable, and the more so when one notes that the range of t_m values is sufficiently wide, up to two orders in magnitude. The increase of t_m data scatter with distance can readily be seen on the records. The average observed t_m increases approximately in proportion to distance in real cases. To implement the idea of the constant relative error in this situation, we will assume that the absolute error of t_m is proportional to traveltime, and therefore will divide both sides of (17) by the value of traveltime $t_{d,j}$. Thus we introduce the new variable $y_j = t_{m,j}/t_{d,j}$, define $Z_j(\mathbf{p}) = T_j(\mathbf{p})/t_{d,j}$, and arrive at a modified equation:

$$Z_i(\mathbf{p}) + \varepsilon_j = y_{m,j}, \tag{18}$$

where now the variance of the error term ε_j can be assumed constant, independent of j.

Let us now consider the simplest parametric model—that of a piecewise-constant vertical effective turbidity profile, or PCP. Let i be the number of a layer, counting from the surface down, h_i the depth of the top of a layer, g_{ei} the effective turbidity in a layer, M the total number of layers plus 1, so that $i = M$ corresponds to the lower half-space, and m the value of i for the layer that contains the source. For $i = 1, \ldots, M$, $p_i = g_{ei}$; for $i = 0$, $p_0 = g_{e0} = t_{m,0}$.

To find the particular form of $Z_j(\mathbf{p})$ let us evaluate the integral (12) for our case. We temporarily omit the j subscript. Let s be the coordinate along the unperturbed ray of the total length S, and let $g_e(s)$ be equal to constants g_{e1}, g_{e2}, ... g_{eK} within corresponding ray pieces $(0, s_1)$, (s_1, s_2), ..., (s_{m-1}, S). These pieces are not assumed to be straight lines, and may be segmented; their lengths are assumed to be known from ray calculations. (We do not assume any specific correspondence between the constant-turbidity layers discussed here and constant-velocity layers that specify rays.) For this case, (12) yields

$$(c/S^2)\langle T\rangle = g_{e0} + g_{e1}U(0, \zeta_1) + g_{e2}U(\zeta_1, \zeta_2) + \ldots + g_{em}U(\zeta_{m-1}, 1), \quad (19)$$

where $\zeta_i = s_i/S$, and

$$U(p, q) = (p^2/2 - p^3/3) - (q^2/2 - q^3/3). \quad (20)$$

For a given ray, s_i, S and ζ_i are known. To account for variable velocity, we simply set $c = S/t_d$ in (19) as explained above.

Now taking into account the assumed relationship $\langle t_m\rangle/\langle T\rangle = 6/10.9 = 0.55 = B$, for $Z_j(\mathbf{p}) = T_j(\mathbf{p})/t_{d,j}$, we can write down

$$Z_j = \sum_{i=0}^{M} a_{ij} g_{ej}, \quad (21)$$

where the coefficients a_{ij} are

$$a_{0j} = 1/t_d,$$

$$a_{ij} = BSU(\zeta_{i-1}, \zeta_i) \quad \text{for } i \le m, \quad (22)$$

and

$$a_{ij} = 0 \quad \text{for } m+1 < i < M.$$

At given h_i, and $N > M$, N equations (21) represent an over-determined linear system, to be solved by least squares. Actual t_m data are very noisy, so a practical inversion needs $N \gg M$. Now denote the a_{ij} matrix as $\mathbf{A}$, the Z_j vector as $\mathbf{Z}$, and the matrix of residual-dependent weights as $\mathbf{W}$, then for the least squares estimate $\mathbf{p}'$ we obtain the standard result

$$\mathbf{p}' = (\mathbf{A}'\mathbf{W}\mathbf{A})^{-1}\mathbf{A}'\mathbf{W}\mathbf{Z}. \quad (23)$$

In a practical implementation, this result is recalculated in an iterative manner several times, with the diagonal $\mathbf{W}$ matrix being adjusted on each iteration employing rule (16) until convergence is reached. The starting state for $\mathbf{W}$ is the unity matrix.

Within the framework of the assumed PCP model and simple least squares, one tries to estimate the values of g_{ei} as independent parameters. How detailed such a reconstruction may be is limited by the effective rank k of $\mathbf{A}$. In this approach, the choice of a particular set of layer boundaries may have an unwanted effect on the result. In a more general inversion procedure, such as for example SVD, the number of constant-turbidity layers can be made large, causing the values of g_{ei} to become significantly correlated between layers. One advantage of this approach is that the shape of the estimated profile does not critically depend on the assumed set of layer depths. This approach is left for the future. As a result of inversion, we obtain the estimates of parameters, of their errors (more precisely, of the covariance matrix), and also of the residual error and average weight, which are useful for comparing different models.

The source-to-station rays are assumed to be uprising, and hence in practical inversion a threshold value is set for the angle of incidence at a source to select such data. The first step of the procedure is therefore to calculate seismic rays through a known velocity structure, and to check this selection criterion.

We have described the simplest case, when the problem is linear and the standard linear weighted least-squares procedure, made robust by means of iterative residual-dependent weighting, is sufficient. The generalization to non-linear cases introduces no fundamental changes, although it may add technical complications. In the simplest case one might apply a grid search over 'non-linear' unknowns, and combine it with least squares over the rest of the variables.

TESTING THE INVERSION PROCEDURE ON A SYNTHETIC EXAMPLE

The inversion procedure described above was tested using a series of synthetic data sets. These were generated to emulate expected properties of data. One particular property is the large scatter of the 'regular' t_m data, related to the fact that the position of the maximum peak in a body-wave group varies greatly among various components and among records of events with comparable hypocentre locations. Another critical property is the presence of a substantial fraction of the 'outlier' subpopulation in the observed t_m data, produced by nodal or near-nodal arrivals, as explained above. This subpopulation must be present even when one works with vector data; it will be more evident when using each of the three components separately, as we are planing to do. The assumed observational setup is similar to that of the practical inversion of our companion paper: source depth interval 25–250 km, hypocentre density decays with depth, data volume 250. The velocity structure is a constant-velocity crust ($H_{Moho} = 35$ km, $c_S = 3.5$ km s^{-1}) over a constant-velocity mantle ($c_S = 4.7$ km s^{-1}), and the S-wave effective turbidity structure is a four-layer one: $h = 0$–10 km, $g_e = 0.01$ km^{-1}; $h = 10$–35 km, $g_e = 0.005$ km^{-1}; $h = 35$–100 km, $g_e = 0.002$ km^{-1}, $h = 100$–250 km, $g_e = 0.0005$ km^{-1}. The velocity structure fixed in the inversion was the true one; the effective turbidity structure to invert consisted of three unknowns, g_{e1}, g_{e2} and g_{e3}, for the layers $h = 0$–20 km, $h = 20$–80 km and $h = 80$–250 km. The statistical structure of data was assumed as follows: 60 per cent is the 'regular subpopulation', assumed to be distributed according to the exponential law with the mean equal to the forward-calculated $\langle t_m\rangle$; and 40 per cent is the 'outlier' subpopulation, distributed according to the saw-tooth/triangular distribution density, with its maximum at $t_m = 0$ and linear decay up to the t_m value equal to traveltime. The exponential law of the 'regular' subpopulation imitates very noisy data: for this law, the coefficient of variation (standard deviation/mean) is equal to unity. The fraction of the 'outlier' subpopulation is assumed to be rather large. Its density decays in a manner

reflecting the expected behaviour of near-nodal data: very large delays, corresponding to truly nodal data, are relatively rare.

The results of inversion are given in Table 1. It includes: true average values of effective turbidity for the assumed three layers, calculated from the 'true' four-layer profile; inversion estimates obtained without any noise (these should be seen as the target values for efficient inversion; their deviations from true values reflect the error caused by the difference between 'real' and 'assumed' layering); three inversions for data sets with 100 per cent 'regular' error, and no weighting applied; and 10 inversions representing the main simulation. For each data set, two inversions are presented: one (A) with respect to g_{e1-3} (three unknowns, t_{m0} fixed as zero) and another (B) with respect to g_{e1-3} and t_{m0} (four unknowns). The four bottom lines represent averages and standard deviations over nine acceptable inversions (the 3A/3B variant is rejected), for the A and B cases. The columns contain: variant code; ψ, rms residual error in fitting y_j multiplied by 100; w_{av}, average data weight/fraction of unsuppressed data; estimates of g_{e1-3} (in 1/1000 km units) and of t_{m0} (s); L_{200}, the total optical thickness L_e of the upper 200 km of the Earth, measured along an oblique ray with an angle of incidence of 45° at a depth of 200 km. CV and σ values are coefficients of variation and standard deviations provided by the least-squares procedure or calculated from its results.

Fig. 2 illustrates the inversion of the data set 1A. (a) shows the spatial data distribution. A first glance at plots (b) and (d) with raw $t_{m,j}$ data gives an impression of wild scatter, suggesting zero information content as regards to distance variation of t_m. However, the robust estimation procedure applied manages (in this particular test, and in most others) to effectively screen out the outliers. Plot (c) shows the graph of the weighting function (16), whose width has been self-adjusted during 40 successive iterations, to arrive at the final value that corresponds to the standard deviation of 0.035, not far from the true value of about 0.0315. In the process, about one-third of data (marked ×) acquired weights below 0.3 and were thus practically excluded from the inversion.

The analysis of numerical estimates given in the Table 1 may be summed up as follows:

(1) The assumption of an arbitrary set of layer boundaries may cause errors of up to 20 per cent for the layer g_e estimates, even without any noise. However, the estimate of L_e is not distorted by this factor.

Table 1. The results of inversion for simulated data sets.

No.	100ψ	w_{av}	g_1	g_2	g_3	t_{m0}	L_{200}	CV($g_{1,2,3}$)			$\sigma(t_{m0})$	CV(L_{200})
		True values										
–	–	–	7.5	3.25	0.676	0.0	0.523					
		Inversion: no noise										
00	0.11	1	7.98	2.667	0.670	0.0	0.521					
		Inversions: 'regular' noise, no outliers										
01A	3.10	1	9.0	2.36	0.82	0.0	0.54	0.21	0.19	0.35	0.0	0.051
01B	3.10	1	8.0	2.34	0.27	0.03	0.52	0.48	0.19	0.35	0.113	0.095
02A	3.28	1	6.2	3.58	0.51	0.0	0.52	0.42	0.14	0.60	0.0	0.061
02B	3.28	1	2.4	3.60	0.50	0.11	0.43	2.71	0.14	0.61	0.181	0.159
03A	3.17	1	7.7	2.83	0.77	0.0	0.54	0.24	0.16	0.40	0.0	0.051
03B	3.17	1	0.2	2.83	0.78	−0.07	0.60	0.42	0.16	0.40	0.115	0.106
		Inversions: both 'regular' noise and outliers										
1A	3.49	0.64	13.2	2.65	0.83	0.0	0.42	0.17	0.19	0.35	0.0	0.055
1B	3.48	0.64	18.0	2.69	0.82	−0.16	0.78	0.24	0.19	0.35	0.132	0.106
2A	5.03	0.67	18.2	2.79	0.99	0.0	0.83	0.17	0.26	0.50	0.0	0.087
2B	4.93	0.67	6.3	2.82	0.92	0.36	0.53	1.09	0.26	0.53	0.187	0.173
3A	9.98	0.78	23.2	6.57	0.14	0.0	1.06	0.24	0.22	6.17	0.0	0.148
3B	9.93	0.78	11.4	6.42	0.12	0.38	0.77	0.98	0.22	7.41	0.346	0.284
4A	2.95	0.65	8.9	2.37	0.56	0.0	0.50	0.24	0.19	0.47	0.0	0.051
4B	2.97	0.65	−1.2	2.50	0.53	0.29	0.26	5.00	0.18	0.50	0.169	0.146
5A	4.41	0.68	10.5	3.84	0.26	0.0	0.61	0.23	0.18	1.63	0.0	0.067
5B	4.51	0.68	5.1	3.77	0.31	0.19	0.48	0.92	0.19	1.40	0.146	0.121
6A	4.52	0.69	10.5	2.71	2.34	0.0	0.87	0.24	0.27	0.20	0.0	0.074
6B	4.60	0.69	9.2	2.73	2.37	0.04	0.84	0.73	0.27	0.20	0.182	0.165
7A	2.15	0.65	6.1	2.10	0.66	0.0	0.43	0.21	0.14	0.32	0.0	0.036
7B	2.14	0.65	1.1	2.07	0.66	0.15	0.30	2.89	0.14	0.31	0.094	0.081
8A	2.89	0.65	8.2	2.84	0.50	0.0	0.51	0.18	0.16	0.56	0.0	0.041
8B	2.93	0.65	6.5	2.82	0.50	0.06	0.47	0.44	0.17	0.57	0.089	0.076
9A	2.75	0.67	6.7	3.13	0.06	0.0	0.42	0.26	0.13	3.61	0.0	0.043
9B	2.74	0.67	0.8	3.04	0.94	0.19	0.28	3.65	0.14	2.67	0.084	0.076
10A	3.30	0.62	7.7	2.06	0.81	0.0	0.49	0.25	0.23	0.40	0.0	0.055
10B	4.36	0.65	0.5	1.86	0.00	0.35	0.37	9.87	0.35	0.35	0.162	0.134
av/A	3.49	0.65	10.0	2.72	0.78	0.0	0.56	0.21	0.19	0.89	0.0	0.056
sd/A	0.95	0.02	3.7	0.54	0.65	0.0	0.17	0.03	0.04	1.10	0.0	0.016
av/B	3.62	0.66	5.1	2.70	0.78	0.16	0.47	2.75	0.21	0.76	0.138	0.119
sd/B	0.99	0.02	5.9	0.55	0.66	0.16	0.21	3.13	0.06	0.79	0.040	0.037

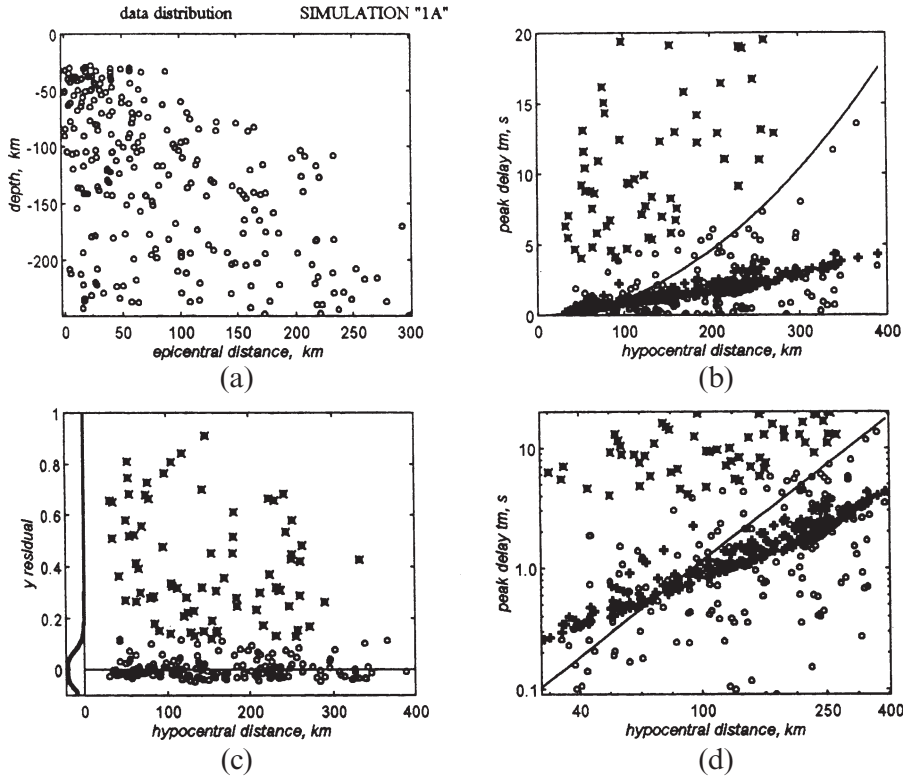

Figure 2. Inversion of the simulated data set 1A (see Table 1). (a) Input data distribution over depth and epicentral distance. (b) t_m versus hypocentral distance: ○, data points, ×, rejection of a data point, i.e. its weight is set to 0.3 or below; +, fitted theoretical t_m values; the smooth line shows a quadratic trend, characteristic of a constant-turbidity medium. (c) Residuals of y versus hypocentral distance: ○ and × as in (b). On the left, along the vertical axis, the graph of the weighting function (16), whose parameter σ was self-adjusted, is shown. (d) Similar to (b), but on a log–log scale.

(2) In the case of purely 'regular' error, layer g_e estimates seem reasonable for the 'constrained-t_{m0}' inversion. 'Free-t_{m0}' inversion may produce major errors in the g_{e1} value for the upper layer. Note that there are no sources inside this layer.

(3) The proposed robust inversion procedure that uses residual-dependent weighting (16), and adjusts these weights iteratively (see eq. 34 and comments thereon), performs quite well. Most outliers are rejected and the residual error is reduced, approximately, to a level (0.035) comparable to that for the reference case of purely 'regular' error (0.032). In one case in 10 (variant 3A/3B), the procedure fails. This case is identified by its very high residual error of about 0.10, and it is likely that such an identification will be possible in a real inversion. Thus we used only the remaining nine inversions to judge the accuracy of inversion, represented by the averages below. The average fraction of rejected data is about 65 per cent, in reasonable agreement with the modelled data structure.

(4) The systematic error of g_{e1} is larger in the case of the 'free-t_{m0}' inversion (-36 per cent) than in the case of 'constrained-t_{m0}' inversion ($+25$ per cent). The estimate of t_{m0} is biased (formal estimate $+0.16 \pm 0.05$ s against the true value of zero). One can conclude from these facts that it would be very important to avoid estimating of t_{m0} in a real inversion. In the following we do not discuss further the results of the 'free-t_{m0}' inversion.

(5) Large errors in g_{e1} are associated with variant nos 1, 2, 4 and 5; all these have relatively high residual error. This fact suggests that the results obtained in inversions with relatively large residual errors are less reliable.

(6) Estimated values for coefficients of variation of g_{e1} g_{e2} and g_{e3}, equal on average to 21, 19 and 89 per cent respectively, should be compared to actual systematic errors for g_{e1}, g_{e2} and g_{e3} (of $+20$, $+2$ and $+16$ per cent respectively) and to the actual values of the coefficient of variation of the estimates (calculated over the nine variants) equal to, respectively, 37, 20 and 83 per cent. To be on the safe side when analysing the results of real inversions, we should consider the estimates based on calculated coefficients of variation as somewhat too optimistic, and those based on comparing different data sets as more realistic. This applies to the estimates of L_e as well: whereas the estimates practically lack any systematic error, their accuracy of 5 per cent calculated from the least-squares results is as much as 3.5 times below the real scatter of 17 per cent.

(7) On the whole, despite very strong noise, including both 'regular' dispersion and 'outlier' errors, reasonable estimates for the effective turbidity structure were obtained, with a general accuracy of about 40 per cent for an individual data set, and of the order of 20 per cent for averaged results of inversion, based on 4–5 independent data sets.

DISCUSSION

The new approach developed here on the basis of Bocharov (1987, 1990) is essentially a kind of tomography with respect to the spatial effective turbidity distribution, and many aspects of other tomographic problems and techniques are relevant here. Its specific value can be perceived more clearly when one

realizes that for traditional velocity tomography with unknown values in cells, the characteristic length of inhomogeneity must be greater or comparable to the cell size. The opposite is necessary to study inhomogeneity as a random field, namely that many inhomogeneities should occupy a single cell, so that the notion of effective turbidity value in a cell is meaningful. A similar difference arises with respect to wavelength: ray tomography is efficient for inhomogeneities much larger than the wavelength, and deterministic diffraction tomography can do no more than weaken somewhat this inequality. Effective turbidity tomography is not limited in this aspect, and recovers the integral effect of all inhomogeneity sizes at the wavelength of the sounding wavefield, but is weighted towards those inhomogeneities whose size is comparable to the wavelength (Wu & Aki 1985).

It should be emphasized that, whereas velocity tomography tries to estimate mean velocity in a cell, effective turbidity tomography, essentially, looks for its variance (there is a correspondence between velocity variance and effective turbidity, very simple in the Gaussian-ACF case, see for example Sato 1989). Therefore, the fields of application for velocity tomography and effective turbidity tomography are different, and not simply complementary: sometimes one can estimate both mean and variance, sometimes only one of the two, and sometimes neither.

Viewed in general, the approach can be applied to recover 1-D, 2-D or 3-D effective turbidity structure if data sets of adequate size and quality can be collected. The particular inversion procedure developed here can be used to reconstruct vertical effective turbidity profiles if a considerable relative vertical range of hypocentral depths is observed; for example, one can use distributions over such ranges as $h = 25$–250 km or $h = 2$–20 km (using higher frequencies in the second case). The procedure is somewhat primitive and aimed at gross reconstruction; if data permit the recovery of detailed structure, more powerful inversion algorithms are preferable.

CONCLUSIONS

We have presented an approach for the reconstruction of a non-uniform spatial distribution of scattering properties, on the basis of body-wave pulse-broadening data. First of all, we are interested in the determination of the vertical structure of scattering properties. The assumed experimental data are measurements of onset-to-peak delays. The reconstruction procedure is a linear inversion of tomographic style, based on the solution of the forward problem after Bocharov. To pose the problem formally, we had to introduce the concept of effective turbidity—an important parameter that defines most manifestations of scattering: broadening of incoherent pulse, scattering loss and coda formation. It is in terms of this parameter that the inverse problem is formulated.

As a prelude to applying this approach to real data we analysed many practical aspects of the inversion and revealed a number of potential problems but none that would prevent meaningful inversion. The most unpleasant is the effect of nodal arrivals that are characterized by delayed peak energy. This delay may be confused with effects of scattering-related broadening. To overcome this obstacle, we proposed a robust estimation procedure based on residual-dependent weighting and tested it on synthetic data.

ACKNOWLEDGMENTS

A discussion with Andrei Bocharov was very useful. The support of the Russian Foundation for Basic Research (project 93-05-08514) is gratefully acknowledged.

REFERENCES

Abubakirov, I.R. & Gusev, A.A., 1990. Estimation of scattering properties of lithosphere of Kamchatka based on Monte-Carlo simulation of record envelope of a near earthquake, *Phys. Earth. planet. Inter.,* **64,** 52–67.

Aki, K., 1973. Scattering of P-waves under Montana LASA, *J. geophys. Res.,* **78,** 1334–1346.

Aki, K., 1980. Scattering and attenuation of shear waves in the lithosphere, *J. geophys. Res.,* **85,** 6496–6504.

Aki, K. & Chouet, B., 1975. Origin of coda waves: source, attenuation and scattering effects, *J. geophys. Res.,* **80,** 3322–3342.

Bocharov, A.A., 1985. *'Stochastic ray-path' method: point source in random medium,* Preprint 978, IKI AN SSSR, Moscow (in Russian).

Bocharov, A.A., 1987. *Space tomography: developing the foundations of the method,* Preprint 1277, IKI AN SSSR, Moscow (in Russian).

Bocharov, A.A., 1988. Mean delay and broadening of a pulse produced by scattering in the random inhomogeneous medium, *Izv. vuzov, Radiofizika,* **31,** 1407–1409 (in Russian).

Bocharov, A.A., 1990. Reconstruction of spatial distribution of the turbulent medium from pulse broadening: the 'space tomography' method, *Izv. vuzov, Radiofizika,* **33,** 395–402 (in Russian).

Chernov, L.A., 1975. *Waves in Random Media,* Nauka, Moscow (in Russian).

Dainty, A.M., 1981. A scattering model to explain seismic Q observations in the lithosphere between 1 and 30 Hz, *Geophys. Res. Lett.,* **8,** 1126–1128.

Dashen, R., 1979. Path integrals for waves in random media, *J. math. Phys.,* **20,** 894–920.

Flatté, S.M. & Wu, R.-S., 1985. Small-scale structure in the lithosphere and astenosphere deduced from arrival time and amplitude fluctuations at NORSAR, *J. geophys. Res.,* **93,** 6601–6614.

Gusev, A.A., 1995. Vertical profile of turbidity and coda Q, *Geophys. J. Int.,* **123,** 665–672.

Gusev, A.A. & Abubakirov, I.R., 1987. Monte-Carlo simulation of record envelope of a near earthquake, *Phys. Earth. planet Inter.,* **49,** 30–36.

Gusev, A.A. & Abubakirov, I.R., 1996a. Study of the vertical turbidity profile of the lithosphere based on the inversion of body wave pulse broadening data, *Volkanol. seism.,* **4,** 81–90 (see also English edition: 1997, **18,** 453–464).

Gusev, A.A. & Abubakirov, I.R., 1996b. Simulated envelopes of non-isotropically scattered body waves as compared to observed ones: another manifestation of fractal inhomoheneity, *Geophys. J. Int.,* **127,** 49–60.

Gusev, A.A. & Abubakirov, I.R., 1999. Vertical profile of effective turbidity reconstructed from broadening of incoherent body-wave pulses—II. Application to Kamchatka data, *Geophys. J. Int.,* **136,** 309–323 (this issue).

Gusev, A.A. & Lemzikov, V.K., 1983. Estimation of scattering parameters of shear waves in the crust and upper mantle of Kamchatka according to observations of 'Shipunski' station, *Vulkanol. Seism.,* No. 1, 94–108.

Gusev, A.A. & Lemzikov, V.K., 1985. Properties of scattered elastic waves in the lithosphere of Kamchatka: parameters and temporal variations, *Tectonophysics,* **112,** 137–153.

Ishimaru, A., 1978. *Wave Propagation and Scattering in Random Media,* Vols 1 and 2, Academic, San Diego.

Jeffreys, H.K., 1961. *Theory of Probability,* Oxford University Press, Oxford.

Mosteller, F. & Tukey, J.W., 1977. *Data Analysis and Regression,* Addison-Wesely, Reading.

Rautian, T.G., Khalturin, V.I., Zakirov, M.S., Zemtsova, A.G., Proscurin, A.P., Pustovitenko, B.G., Pustovitenko, A.N., Sinelnikova, L.G., Filina, A.G. & Shengelia, I.S., 1981. *Experimental Studies of Seismic Coda,* Nauka, Moscow (in Russian).

Rytov, S.M., Kravtsov, Yu.A. & Tatarskii, V.I., 1978. *Vvedenie v Statisticheskuyu Radiofiziku, II. Sluchainye Polya,* Nauka, Moscow (in Russian); also English edition: *Introduction to Statistical Radiophysics,* Vols 1–4, Springer, New York.

Sato, H., 1978. Mean free path of S-waves under the Kanto district of Japan, *J. Phys. Earth,* **26,** 185–198.

Sato, H., 1989. Broadening of seismogram envelopes in the random inhomogeneous lithosphere based on the parabolic approximation: Southeastern Honshu, Japan, *J. geophys. Res.,* **94,** 17 735–17 747.

Sato, H., 1990. Unified approach to amplitude attenuation and coda excitation in the randomly inhomogeneous lithosphere, *Pure appl. Geophys.,* **132,** 93–119.

Uscinski, B.J., 1977. *The Elements of Wave Propagation in Random Media,* McGraw-Hill, New York.

Williamson, I.P., 1972. Pulse broadening due to multiple scattering in the interstellar medium, *Mon. Not. R. astr. Soc.,* **157,** 55–71.

Williamson, I.P., 1975. The broadening of pulses due to multipath propagation of radiation, *Proc. R. Soc.* A, **343,** 131–147.

Wu, R.-S. & Aki, K., 1985. Elastic wave scattering by a random medium and the small-scale inhomogeneities in the lithosphere, *J. geophys. Res.,* **90,** 10 261–10 273.

APPENDIX A: SHORT PROOF OF BOCHAROV'S FORMULA

We present a short new proof for Bocharov's formula. To find the mean delay $\langle T \rangle$ (4) of forward-scattered energy propagated through a scattering medium with Gaussian ACF, first consider an ensemble of random stochastic ray paths (Fig. A1a) that connect a source O set at $(0, 0, 0)$ and a receiver R set at $(0, 0, z)$. Following Williamson (1972), one can move to an equivalent problem, considering instead the ray paths that depart from O along the Z-axis and terminate on a sphere of radius OR. The related error is small for sufficiently smooth effective turbidity structures; the meaning of 'smoothness' will be explained in more detail later.

The projection of a stochastic ray path onto the ZOX plane is sketched in Fig. A1(b). Shown are the unperturbed ray OR_0, the rotated unperturbed ray OR_1 and a perturbed ray OR, all of the same length S. The straight-line distance from O to R is equal to r. Let the X-coordinates of R and R_1 be x and x_1, and denote QR_0 as p. Note that the angle ψ is assumed small, so that $\psi \approx x_1/\mathrm{S} \approx x/S$. The individual delay is $RR_1/c \approx (S-r)/c$. Clearly,

$$S - r = S - \frac{S-p}{\cos\psi} \approx p - \frac{1}{2}S\psi^2 \approx p - \frac{x^2}{2S}. \tag{A1}$$

Now assume the ray path to consist of a large number N of linear segments Δs_i $(i = 1, 2 \ldots N)$ of equal length Δs, so that the along-ray distance from O to the end of the Δs_i segment is $s_i = i\Delta s$ and $S = N\Delta s$. Fig. A1(c) shows the first four segments. Denote by θ_i the angle of scattering, or angular deflection between segments Δs_{i-1} and Δs_i. Denote by Δx_i and Δp_i the contributions of the segment Δs_i into x and p. It is evident

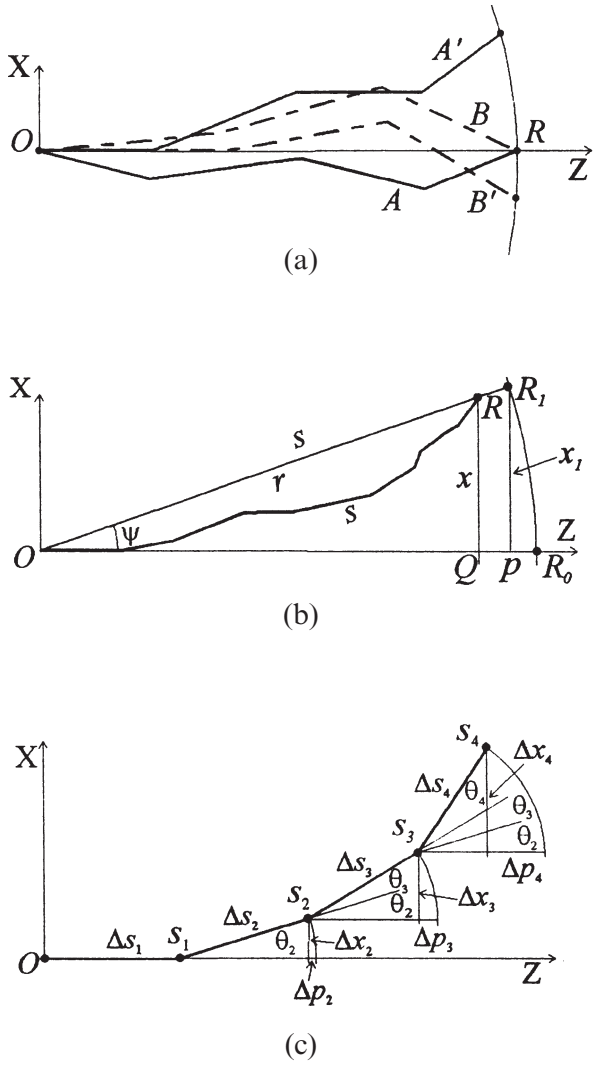

Figure A1. A sketch to illustrate the derivation of Bocharov's formula. (a) Two ray paths A, B from a source O to receiver R, and their approximate rotated equivalents A', B'. (b) Unperturbed (OR_0), rotated unperturbed (OR_1) and rotated perturbed (OR) rays. (c) Representation of a perturbed ray as a chain of constant-length segments.

from the figure that

$$\begin{aligned} \Delta x_i &= \Delta s_i \sin\left(\sum_{k=2}^{i}\theta_k\right) \approx \Delta s_i \sum_{k=2}^{i}\theta_k, \\ \Delta p_i &= \Delta s_i - \Delta s_i\left(\cos\left(\sum_{k=2}^{i}\theta_k\right)\right) \approx \frac{1}{2}\Delta s_i\left(\sum_{k=2}^{i}\theta_k\right)^2. \end{aligned} \tag{A2}$$

Summing these contributions with respect to i, we obtain

$$\begin{aligned} x &= \sum_{i=2}^{N}(N+1-i)\theta_i\Delta s_i, \\ p &= \frac{1}{2}\sum_{i=2}^{N}\Delta s_i\left(\sum_{k=2}^{i}\theta_k\right). \end{aligned} \tag{A3}$$

Reformulating (1) in terms of particles, we obtain the probability of a scattering event to take place on a ray element du as $dP(du) = g_n(u)du$. Now, introducing the error of the order $(\Delta s)^2$, assume a scattering event, if it takes place somewhere on Δs_i, to be shifted to the endpoint of this segment (as shown

in Fig. A1c); for the probability of this event we obtain

$$P_{sc} = P(\Delta s_i) \approx g_n(s_i)\Delta s\,. \tag{A4}$$

For each scattering event, define the 'angle vector' $\boldsymbol{\theta}_i = (\theta_i^{(x)}, \theta_i^{(y)})$, where $|\boldsymbol{\theta}_i| \ll 1$, that connects the two points of a unit sphere that are the ends of unit vectors of the incident and of the scattered ray segments, and let the distribution of $\boldsymbol{\theta}_i$ be (see eq. 9)

$$P(\boldsymbol{\theta}_i) = P(\theta_i^{(x)}, \theta_i^{(y)}) = m(\boldsymbol{\theta}_i, s_i)\,. \tag{A5}$$

When no scattering occurs, we can formally assume $\boldsymbol{\theta}_i$ to be distributed as a delta function. For an arbitrary segment Δs_i, the act of scattering either occurs on it once, with probability P_{sc} (A4) or does not occur, with probability $1 - P_{sc}$. Multiple scattering over Δs_i may safely be neglected. In the former case, the distribution of the vector $\boldsymbol{\theta}_i$ $(= \{\theta_i^{(x)}, \theta_i^{(y)}\})$ is defined by eq. (A5); otherwise, it equals zero. Formally, one may write the following mixed distribution law for $\boldsymbol{\theta}_i$ as

$$p(\boldsymbol{\theta}_i) = (g_n(s_i)\Delta s_i)m(\boldsymbol{\theta}_i, s_i) + (1 - g_n(s_i)\Delta s_i)\delta(\boldsymbol{\theta}_i)\,. \tag{A6}$$

Now we recall that $\langle|\boldsymbol{\theta}_i|^2\rangle = \theta_0^2$ and from (A6) we calculate the mean square length of the vector $\boldsymbol{\theta}_i$:

$$\langle|\boldsymbol{\theta}_i|^2\rangle = g_n(s_i)\theta_0^2(s_i)\Delta s_i\,. \tag{A7}$$

The right-hand side of (A7) can be expressed through (10) as

$$g_n(s_i)\theta_0^2(s_i)\Delta s_i = 2g_e(s_i)\Delta s_i\,. \tag{A8}$$

(At this point, the initial problem with piecewise-linear ray paths has essentially been reduced to the simpler problem of smooth ray paths; that is, to Chernov's 'ray diffusion'.) To evaluate (A3) we now note that for an axisymmetric indicatrix $m(\boldsymbol{\theta}_i, s_i)$,

$$\begin{aligned} &\langle\theta_i^{(x)}\rangle = \langle\theta_i^{(y)}\rangle = 0\,; \\ &\langle(\theta_i^{(x)})^2\rangle = \langle(\theta_i^{(y)})^2\rangle = \frac{1}{2}\langle|\boldsymbol{\theta}_i|^2\rangle = g_e(s_i)\Delta s_i\,, \end{aligned} \tag{A9}$$

and we may assume that $\langle\theta_i^{(x,y)}\theta_i^{(x,y)}\rangle$. Substituting this into (A3) we obtain $\langle x\rangle = 0$, $\langle p\rangle$, $\langle x^2\rangle$ and finally $\langle S - r\rangle = \langle p\rangle - \langle x^2\rangle/2S$:

$$\begin{aligned} \langle p\rangle &= \frac{1}{2}\sum_{i=2}^{N} g_e(s_i)(S - s_i - \Delta s_i)\Delta s_i \\ \langle x^2\rangle &= \sum_{i=2}^{N} g_e(s_i)(S - s_i - \Delta s_i)^2\Delta s_i \\ \langle S - r\rangle &= \frac{1}{2S}\sum_{i=2}^{N} g_e(s_i)s_i(S - s_i)\Delta s_i \\ &\quad + \frac{1}{2S}\sum_{i=2}^{N} g_e(s_i)(S - 2s_i - \Delta s_i)(\Delta s_i)^2\,. \end{aligned} \tag{A10}$$

Now as N tends to infinity we may neglect all terms of second order in Δs_i to obtain the integral

$$\langle S - r\rangle = \frac{1}{2S}\int_0^S g_e(u)u(S - u)\,du\,. \tag{A11}$$

With the error of second order in ψ one can replace the perturbed ray-length variable u [defined on $(0, S)$] by the length along the straight line OR [defined on $(0, r)$, and again denoted u]:

$$\frac{1}{2S}\int_0^S g_e(u)u(S - u)\,du \approx \frac{1}{2r}\int_0^r g_e(u)u(r - u)\,du\,. \tag{A12}$$

A similar result is true for the ZOY plane. As was shown by Williamson (1972), the contributions to $\langle S - r\rangle$ (that is, to $c\langle T\rangle$), calculated for each of the two projections of a 3-D ray path onto the ZOX and ZOY planes can merely be added. Denoting $\langle S - r\rangle$ as F, we obtain for the total delay

$$c\langle T\rangle = F = \frac{1}{S}\int_0^S g_e(u)(S - u)\,udu\,, \tag{A13}$$

which is just Bocharov's result. In his original derivation, Bocharov (1985, 1987, 1988) used a more general approach, first writing down the theoretical multiple-integral representation for the distribution of T, and then calculating $\langle T\rangle$ from it.