The effect of small-scale heterogeneity on the arrival time of waves

Jesper Spetzler and Roel Snieder*

Department of Geophysics, Utrecht University, PO Box 80.021, NL-3508 TA, *Utrecht, the Netherlands. E-mail: spetzler@geo.uu.nl*

SUMMARY

Small-scale heterogeneity alters the arrival times of waves in a way that cannot by explained by ray theory. This is because ray theory is a high-frequency approximation that does not take the finite frequency of wavefields into account. We present a theory based on the first-order Rytov approximation that predicts well the arrival times of waves propagating in media with small-scale inhomogeneity with a length scale smaller than the width of Fresnel zones. In the regime for which scattering theory is relevant we find that caustics are easily generated in wavefields, but this does not influence the good prediction of finite frequency arrival times of waves by scattering theory. The regime of scattering theory is relevant when the characteristic length of heterogeneity is smaller than the width of Fresnel zones. The regime of triplications is independent of frequency but it is more significant the greater the magnitude of slowness fluctuations.

Key words: caustics, Fréchet derivatives, ray theory, Rytov approximation, scattering.

1 INTRODUCTION

Ray theory is valid only if the wavelengths of the waves and the associated widths of Fresnel zones are much smaller than the characteristic length of heterogeneity. For example, in geophysics when working with surface wave tomography it is common to use ray theoretical schemes, which offer a computationally effective solution to the forward problem. This approximation, however, poses a problem from a theoretical point of view because the length scale of inhomogeneity in high-resolution models is comparable to the widths of Fresnel zones (Passier & Snieder 1995). Other domains where scattering is considered to be important are ocean acoustics (Kuperman *et al.* 1998), medical imaging (Baba *et al.* 1989; King & Shao 1990) and non-destructive testing experiments (Haque *et al.* 1999).

Several different approaches to scattering theory are reported in the literature. Marquering *et al.* (1998) described how to calculate sensitivity kernels based on the first-order Born approximation and surface wave mode coupling. Marquering *et al.* (1999) developed a sensitivity kernel formulation of the perturbed time starting with the cross-correlation function. Jensen & Jacobsen (1997) explained how a linearized inversion of time–distance helioseismic data is established by introducing an approximate Gaussian sensitivity kernel. Yomogida (1992) utilized the Born approximation and then the Rytov approximation to derive the sensitivity kernel. Woodward (1992) introduced the finite frequency effect on wave paths, and the concept of Rytov and Born wave paths for transient, reflected and refracted wavefields were explained. Snieder & Lomax (1996)

* Now at: Department of Geophysics, Colorado School of Mines, Golden, CO 80401, USA.

computed a frequency-averaging function from the first-order Rytov approximation. Fehler *et al.* (2000) applied the Rytov approximation to simulate multiple forward wave scattering in Gaussian random media, and the results are compared with those from finite difference solutions of the wave equation.

We follow the idea of Snieder & Lomax (1996) that the phase shift of the scattered wavefield due to a perturbation of the medium can be expressed as the integration of a sensitivity kernel multiplied by the slowness perturbations over the complete model space. In addition to this, we transform the phase shift expression obtained into a time-shift expression so that scattering theory is directly applicable to the interpretation of arrival time data.

Scattering theory includes non-ray-geometrical phenomena. In brief, time residuals due to scattering theory are altered by slowness perturbations surrounding the geometrical ray, and the maximum sensitivity to slowness perturbations is largest just beside the geometrical ray. In contrast, ray-theoretical time delays are only sensitive to the slowness field on the ray path.

The regime of scattering theory is determined from a 2-D numerical experiment wherein the frequency of the waves, the magnitude of slowness perturbations, the offset of the receivers and the length scale of the slowness perturbation field are controlled. We compare the residual times for ray theory and scattering theory with time delays computed with a finite difference solution of the acoustic wave equation. Because we have control over the parameters in the numerical experiment, the regimes of ray theory, scattering theory and triplications are investigated. Furthermore, we show with another 2-D numerical experiment that the regimes of scattering theory and triplications remain valid in a more complex medium (namely Gaussian random media).

In addition, we show that triplications are easily generated in wavefields if the slowness perturbations are sufficiently large. Although the scattering approach is based on a linearization of the phase (i.e. the first-order Rytov approximation), the 'observed' time delays estimated from wavefields with caustics are well predicted by scattering theory.

In Section 2, we explain how to derive the widths of Fresnel zones, the focal length of converging 2-D wavefields and the time-shifts due to ray theory and scattering theory. In Section 3, we describe the numerical experiment that is used to determine the regime for scattering theory and present the experiment for models of Gaussian random media. In Section 4, we summarize the results of the numerical experiments and define the different regimes of ray theory, scattering theory and triplications. In Section 5, we give examples, taken from global seismology, ocean acoustic and medical imaging, where scattering theory is important.

2   THEORY

In this section, we present the theory applied to the investigation of the influence of small-scale heterogeneity on travel-times. The theory is derived for two distinct source geometries: the plane wave (plw) source, and the point source (ps). First, we derive the widths of the Fresnel zones. We then discuss the focal length of converging wavefields in 2-D slowness perturbation fields. Finally, we deduce the first- and second-order linearized ray theories and the first-order linearized scattering theory for 2-D experiments.

2.1   The widths of Fresnel zones

Fresnel zones are defined in terms of the difference in propagation path lengths for rays with nearby paths. All points of a ray taking a detour compared with the ballistic ray are inside the Fresnel zone if the difference in length of propagation paths for the ballistic ray and the detour ray is less than or equal to a certain fraction of the wavelength λ (e.g. Kravtsov 1988). This is the first Fresnel zone, which physically signifies constructive interference of the scattered wavefield produced by single-point scatterers inside the Fresnel zone. We prefer to keep the formulae as general as possible, so the Fresnel zone is defined as the set of points that give single scattered waves with a detour smaller than the wavelength divided by a number n.

Let $x \in [0; L]$ denote the ballistic propagation distance between the source and the receiver, with the source–receiver separation indicated by L. For a plane wave in a homogeneous medium, the detour d is

$$d = \sqrt{(L - x)^2 + q^2(x)} + x - L, \tag{1}$$

where $q(x)$ is the perpendicular distance to the geometrical ray at position x along the ray. The Fresnel zone condition is that $d \leq \lambda/n$. To estimate the boundaries of Fresnel zones, we use the sign of equality in the Fresnel zone condition. To leading order in $q(x)/(L - x)$, the perpendicular distance from the ballistic ray is then isolated from eq. (1) as

$$q(x) = \sqrt{\frac{2\lambda(L - x)}{n}}. \tag{2}$$

The width, W^{plw}, of the Fresnel zone is twice the perpendicular distance from the ballistic ray. Hence,

$$W^{\mathrm{plw}}(x) = \sqrt{\frac{8\lambda(L - x)}{n}}. \tag{3}$$

The maximum width $L_{\mathrm{F}}^{\mathrm{plw}} = \sqrt{8\lambda L/n}$ is obtained at the initial wavefront ($x = 0$), while the widths of Fresnel zones for plane waves to first-order approximation go to zero at the receiver position ($x = L$).

The widths of Fresnel zones for point sources in homogeneous media are calculated in the same way as in the plane wave case. The detour is

$$d = \sqrt{(L - x)^2 + q^2(x)} + \sqrt{x^2 + q^2(x)} - L$$

$$\approx \frac{1}{2} q^2(x) \frac{L}{(L - x)x}. \tag{4}$$

The Fresnel zone condition is again used to compute the widths of Fresnel zones. Hence,

$$W^{\mathrm{ps}}(x) = \sqrt{\frac{8\lambda x(L - x)}{nL}}, \tag{5}$$

which is maximum at $x = L/2$. The maximum width $L_{\mathrm{F}}^{\mathrm{ps}} = \sqrt{2\lambda L/n}$. Notice that $L_{\mathrm{F}}^{\mathrm{ps}} = L_{\mathrm{F}}^{\mathrm{plw}}/2$.

For eqs (3) and (5), the reference medium is homogeneous, which means that ray bending is not taken into account. However, it is possible to compute the boundaries of Fresnel zones in heterogeneous media as well (Pulliam & Snieder 1998).

2.2   Estimation of caustics in 2-D slowness perturbation fields

We discuss in this section at which point caustics, also called triplications, start to develop in the special case of a slab with a perturbed slowness field depending only on depth and in the general case of Gaussian random media. The general theory for the formation of caustics is explained thoroughly in Spetzler & Snieder (2001).

First, we consider the case that caustics develop when an initially plane wave propagates in the x-direction through a slab with a depth-dependent slowness field $u_1(z)$. The set-up of this experiment is shown in Fig. 1. The vertical slab is at an offset x_l from the source, and the slab width is denoted W. The slowness field to the left and right of the slab is set to the constant reference slowness field u_0.

First it is assumed that caustics develop before the ballistic wavefield leaves the slab. We use eq. (6) in Spetzler & Snieder
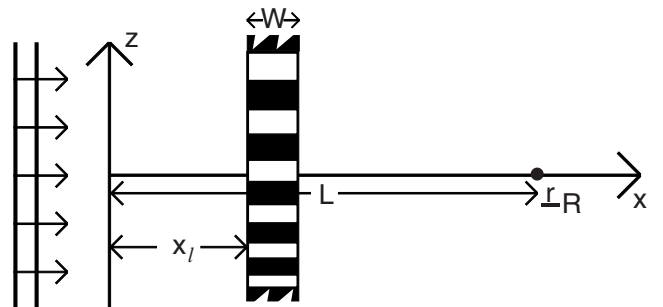

Figure 1. Explanation of the variables used in the experiment with a vertical slab of heterogeneity. A plane wave is incident from the left.

(2001), where the integration along the reference ray is carried out from x_l to x_{caus} (for $x_{\mathrm{caus}} - x_l \leq W$), to determine when caustics form inside the slab. Hence,

$$x_{\mathrm{caus}}(z) = x_l + \sqrt{\dfrac{-2}{\dfrac{\partial^2}{\partial z^2}\left(\dfrac{u_1}{u_0}\right)(z)}} \quad \text{(inside slab)}. \qquad (6)$$

If triplications develop after the waves pass the slab (i.e. $x_{\mathrm{caus}} > x_l + W$), then the propagation length of plane wavefields at which caustics start to occur is

$$x_{\mathrm{caus}}(z) = x_l + \frac{1}{2}W - \dfrac{1}{W\dfrac{\partial^2}{\partial z^2}\left(\dfrac{u_1}{u_0}\right)(z)} \quad \text{(after slab)}. \qquad (7)$$

Let the distance between the source and receiver be denoted L. If $x_{\mathrm{caus}}(z) < L$, triplications will be present in the recorded wavefield.

Next, we discuss the formation of triplications in Gaussian random media. The autocorrelation function of a Gaussian random medium is given by

$$\langle u_1(\mathbf{r}_1)u_1(\mathbf{r}_2)\rangle = (\varepsilon u_0)^2\, e^{-(r/a)^2}, \qquad (8)$$

where ε is the rms value of slowness perturbation fluctuations, a denotes the autocorrelation length (or roughly the length scale of slowness perturbations) and $r = |\mathbf{r}_1 - \mathbf{r}_2|$.

Spetzler & Snieder (2001) showed for an incoming plane wave in Gaussian random media that the formation of caustics is significant when

$$\frac{L}{a} \geq 0.52\,\varepsilon^{-2/3}, \qquad (9)$$

with the source–receiver distance denoted L. For wavefields emitted by point sources in Gaussian random media, the non-dimensional number L/a for the condition that triplications develop in the recorded wavefield is given by

$$\frac{L}{a} \geq 1.12\,\varepsilon^{-2/3}, \qquad (10)$$

(see Spetzler & Snieder 2001). The conditions for the formation of caustics in eqs (9) and (10) are independent of the wavelength but they depend on the rms value of slowness fluctuations.

2.3   Time-shift derivations

We apply two approaches to derive the residual times in a 2-D isotropic perturbed slowness medium. First, we explain how the time delay due to first- and second-order ray perturbation theory is estimated. Second, we show how the time-shift based on first-order linearized scattering theory can be written as a linear function of the 2-D slowness perturbation field. Third, we discuss the properties of the scattering theory.

2.4   The ray-geometrical time-shift

According to second-order ray perturbation theory (e.g. Snieder & Sambridge 1992), the traveltime is the sum of three components, namely

$$T = T_0 + T_1 + T_2. \qquad (11)$$

T_0 is the contribution from the reference ray in the reference medium, T_1 is the time-shift due to the slowness perturbation field along the reference ray (based on Fermat's principle) and the term T_2 is a more complicated expression that accounts for the deflection of the ray by the slowness perturbation. A complete explanation of how to calculate T_2 is given in Snieder & Sambridge (1992).

In the numerical experiment presented here, we compute the ray-theoretical time delay δt due to a perturbed slowness medium to first order. This implies that the time-shift is expressed as a linear function of the slowness anomaly $u_1(\mathbf{r})$ along the reference ray. Hence,

$$\delta t = T_1$$
$$= \int_{\mathrm{Ref\,ray}} u_1(\mathbf{r})\,ds. \qquad (12)$$

In our experiment the reference slowness is constant, so the reference ray is a straight line between the source and receiver.

In the numerical examples shown here, the second-order traveltime perturbation T_2 is much smaller than the first-order traveltime perturbation T_1. For this reason the first-order traveltime perturbation is used for the ray-geometrical traveltime.

Ray theory is valid when the characteristic length a of heterogeneity is much larger than both the wavelength λ and the widths of the Fresnel zones L_{F}. Hence, in non-dimensional numbers, the condition for ray theory is written as

$$\frac{\lambda}{a} \ll 1 \quad \text{and} \quad \frac{L_{\mathrm{F}}}{a} \ll 1 \qquad (13)$$

(see Menke & Abbot 1990).

2.5   Single-scattering theory

We show that the time perturbation $\langle\delta t\rangle(L)$ for the receiver at the offset L is written as an integration over the slowness perturbation field $u_1(x, z)$ multiplied by a sensitivity kernel $K(x, z)$:

$$\langle\delta t\rangle(L) = \int_{-\infty}^{\infty}\int_{0}^{L} u_1(x, z)K(x, z)\,dx\,dz. \qquad (14)$$

The first-order perturbation of the phase of the wavefield follows from the Rytov approximation. The unperturbed wavefield is denoted by p_0. The Born approximation gives the first-order perturbation p_1 of the wavefield. According to Beydoun & Tarantola (1988) and Snieder & Lomax (1996), the phase shift $\delta\varphi$ is given by

$$\delta\varphi = \mathscr{I}m\left\{\frac{p_1}{p_0}\right\}. \qquad (15)$$

The condition for the validity of the Rytov approximation is that $k_0 L(u_1/u_0)^2 \ll 1$, where k_0 is the wavenumber. Comparing this condition with the condition for the Born approximation $k_0 L u_1/u_0 \ll 1$ (Snieder & Lomax 1996), we see that the Rytov approximation has validity for a larger slowness perturbation parameter than does the Born approximation.

Snieder & Lomax (1996) demonstrated that the Born approximation to the solution of the acoustic wave equation with constant density is

$$p_1(\underline{r}_{\mathrm{R}}, \omega) = \omega^{3/2}\sqrt{\frac{u_0}{2\pi}}\,e^{i\frac{\pi}{4}}\int_{\mathrm{V}}\frac{u_1(\underline{r})}{\sqrt{|\underline{r}_{\mathrm{R}} - \underline{r}|}}\,p_0\,e^{ik_0|\underline{r}_{\mathrm{R}} - \underline{r}|}\,d^2\underline{r} \qquad (16)$$

for an incident plane wave that is given by $p_0 = \exp(ik_0 x)$. The receiver is at the position (L, z_j). We assume that $(z - z_j)/(L - x) \ll 1$ and set $|\underline{r}_R - \underline{r}|$ and $1/\sqrt{|\underline{r}_R - \underline{r}|}$ to the first- and zeroth-order Taylor approximations, respectively (Snieder & Lomax 1996); see Fig. 2 for a definition of the geometric variables:

$$|\underline{r}_R - \underline{r}| \approx (L - x) + \frac{(z - z_j)^2}{2(L - x)} \quad \text{and} \quad \frac{1}{\sqrt{|\underline{r}_R - \underline{r}|}} \approx \frac{1}{\sqrt{L - x}}.$$

(17)

We insert these two small-angle approximations into eq. (16) and define the full-space integration as a double integration going from 0 to L for the offset x and from $-\infty$ to ∞ for the perpendicular distance z from the geometrical ray:

$$p_1(\underline{r}, \omega) = \omega^{3/2} \sqrt{u_0/(2\pi)}\, e^{i\pi/4}\, e^{ik_0 L}$$

$$\times \int_0^L \frac{1}{\sqrt{L - x}} \int_{-\infty}^{\infty} u_1(x, z)\, e^{ik_0 \frac{(z-z_j)^2}{2(L-x)}}\, dz\, dx.$$

(18)

Using eq. (15) the phase shift at the receiver position (L, z_j) is then given by

$$\delta\varphi(L, \omega) = \omega^{3/2} \sqrt{\frac{u_0}{2\pi}} \int_{-\infty}^{\infty} \int_0^L u_1(x, z)$$

$$\times \frac{\sin\left(k_0 \frac{(z-z_j)^2}{2(L-x)} + \frac{\pi}{4}\right)}{\sqrt{L - x}}\, dx\, dz.$$

(19)

So far all the calculations in this section have been performed in the frequency domain. In spite of this, we can express the linearized phase perturbation as a linearized time delay, supporting this statement by mathematically representing waves as $A(\omega)\exp(i\varphi(\omega))$, where the amplitude $A(\omega)$ and the phase $\varphi(\omega) = \omega t$ depend on the angular frequency ω. The phase shift is then expressed as

$$\delta\varphi = \omega\delta t,$$

(20)

where δt is the time perturbation, which is a function of frequency. Hence, the theoretical time-shift due to single scattering is

$$\delta t(L, \omega) = \sqrt{\frac{u_0 \omega}{2\pi}} \int_{-\infty}^{\infty} \int_0^L u_1(x, z) \frac{\sin\left(k_0 \frac{(z-z_j)^2}{2(L-x)} + \frac{\pi}{4}\right)}{\sqrt{L - x}}\, dx\, dz.$$
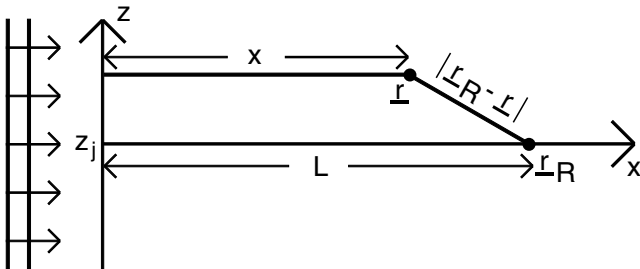
(21)

Wavefields are never monochromatic so we need to frequency-average the time-shift. For example, the time perturbation can be calculated for a frequency band in the range $v_0 - \Delta v$ to $v_0 + \Delta v$, where v_0 is the central frequency and Δv is the half-width of the frequency band. Moreover, to account for the variation of the frequency spectrum in the range of frequency integration, we introduce the normalized amplitude spectrum $A(v)$ of the recorded wavefield. The normalization condition for the amplitude spectrum is that $\int_{v_0-\Delta v}^{v_0+\Delta v} A(v)dv = 1$. The frequency band averaged time-shift is calculated as

$$\langle \delta t \rangle(L) = \int_{v_0-\Delta v}^{v_0+\Delta v} A(v)\delta t(L, v)dv$$

$$= \int_{-\infty}^{\infty} \int_0^L u_1(x, z)\sqrt{u_0} \int_{v_0-\Delta v}^{v_0+\Delta v} A(v)\sqrt{v}$$

$$\times \frac{\sin\left(\pi v u_0 \frac{(z-z_j)^2}{(L-x)} + \frac{\pi}{4}\right)}{\sqrt{L - x}}\, dv\, dx\, dz.$$

(22)

Comparing eq. (22) with eq. (14), we identify the sensitivity kernel for an incoming plane wave as

$$K^{\text{plw}}(x, z) = \sqrt{u_0} \int_{v_0-\Delta v}^{v_0+\Delta v} A(v)\sqrt{v}\, \frac{\sin\left(\pi u_0 v \frac{(z-z_j)^2}{(L-x)} + \frac{\pi}{4}\right)}{\sqrt{L - x}}\, dv.$$

(23)

For point sources, the sensitivity kernel in eq. (23) is modified by taking the point-source geometry into account. The solution to the zeroth-order wavefield in the far field is $p_0 = -(1/\sqrt{8\pi kr})\exp i(kr + \pi/4)$, where k is the wavenumber and r is the propagation length. This solution for the source geometry contains the geometrical spreading factor $1/\sqrt{r}$, yielding the sensitivity kernel for a point source,

$$K^{\text{ps}}(x, z) = \sqrt{u_0 L} \int_{v_0-\Delta v}^{v_0+\Delta v} A(v)\sqrt{v}\, \frac{\sin\left(\pi u_0 v L \frac{(z-z_j)^2}{x(L-x)} + \frac{\pi}{4}\right)}{\sqrt{x(L-x)}}\, dv.$$

(24)

2.6 The properties of the sensitivity kernel

The sensitivity kernels for plane waves in eq. (23) and for point sources in eq. (24), assuming a constant frequency spectrum over the range of integration, are shown in Fig. 3. The perpendicular distance to the geometrical ray is plotted on the horizontal axis and the sensitivity to slowness perturbations is plotted on the vertical axis. The figure shows that the maximum sensitivity to the slowness perturbation field is off the path of the geometrical ray. This phenomenon has been observed by several authors (see Marquering *et al.* 1998, 1999; Snieder & Lomax 1996; Yomogida 1992). Thus, scattering theory deviates from ray theory, which predicts that the traveltime is sensitive only to the slowness field on the ray. We see furthermore that the sensitivity kernel has sidelobes with a decreasing amplitude away from the ray path. This means that finite frequency time
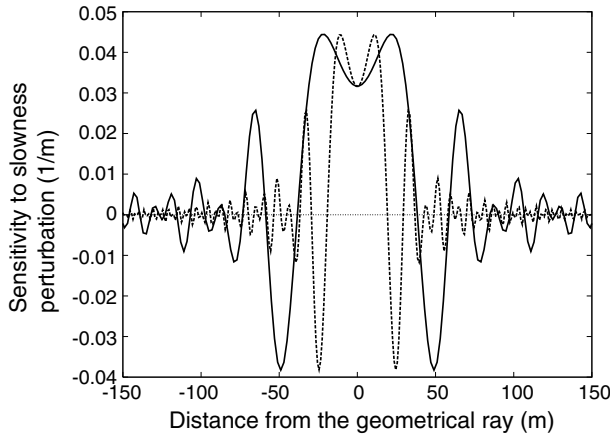

Figure 2. Definition of the geometric variables for an incoming plane wave in a 2-D medium with a constant reference slowness.

Figure 3. The sensitivity kernel for an incident plane wave (solid line) and for a point source (dashed line). The reference slowness is 2.5×10^{-4} s m^{-1}, the constant frequency band is between 150 and 250 Hz and the offset is 100 m. The sensitivity kernel for a plane wave is computed at the initial wavefront, whereas the sensitivity kernel for a point source is evaluated at the half-distance between the source and receiver. The maximum width of the positive, central lobe of the sensitivity kernel for a plane wave is twice the maximum width of the positive, central lobe of the sensitivity kernel for a point source.

perturbations are sensitive to slowness perturbations surrounding the ray path. In three dimensions, the sensitivity kernel is even zero on the ray, which is well illustrated in Marquering *et al.* (1999).

The width of the positive, central lobe of the sensitivity kernel for plane waves is computed by setting the sine function in eq. (23) equal to zero. Hence,

$$\sin\left(\pi u_0 v \frac{(z - z_j)^2}{(L - x)} + \frac{\pi}{4} \right) = 0 . \tag{25}$$

We isolate $z - z_j$ and multiply by 2 in order to calculate the width $W_{\text{sens}}^{\text{plw}}(x)$ of the positive, central lobe:

$$W_{\text{sens}}^{\text{plw}}(x) = \sqrt{3\lambda(L - x)} . \tag{26}$$

In the same manner, we derive from eq. (24) the width $W_{\text{sens}}^{\text{ps}}(x)$ of the positive central lobe of the sensitivity kernel for point sources:

$$W_{\text{sens}}^{\text{ps}}(x) = \sqrt{\frac{3\lambda x(L - x)}{L}} . \tag{27}$$

Next we compare the widths $W_{\text{sens}}^{\text{plw}}(x)$ and $W_{\text{sens}}^{\text{ps}}(x)$ with the widths of Fresnel zones in eqs (3) and (5), respectively. Except for different factors 3 and $8/n$, the two kinds of expressions have the same dependence on λ, x and L. Equating these factors enables us to obtain an estimate of the number n in eqs (3) and (5) for the width of Fresnel zones. We find that

$$n = \frac{8}{3} \tag{28}$$

in two dimensions. In three-dimensions, $n = 2$, which can be derived by comparing the width of the positive, central lobe of 3-D sensitivity kernels with the widths of Fresnel zones. The value of n is important because Fresnel zones are physically interpreted as the positive interference of waves with a detour less that λ/n. In two dimensions, this difference in propagation

length must not exceed $3\lambda/8$ for the first Fresnel zone. We interpret the width of the positive, central lobe of sensitivity kernels as the width of the Fresnel zone.

Finally, we show with the stationary phase approximation (Bleistein 1984) that the integration of the product of the slowness perturbation field and the sensitivity kernel for plane waves over space is equivalent to eq. (12), which is valid for first-order ray perturbation theory. Although we use the sensitivity kernel for plane waves in eq. (23) in the derivation, the result is also valid for the sensitivity kernel for point sources in two dimensions as well as for a point source or a plane incoming wave in three dimensions. We assume that the slowness perturbation field depends only on the propagation distance from the source. Thus, by making use of the 2-D sensitivity kernel for plane waves, it follows that, for $z_j = 0$,

$$\int_0^L \int_{-\infty}^{\infty} u_1(x) K^{\text{plw}}(x, z) \, dz \, dx$$

$$= \sqrt{u_0} \int_{v_0 - \Delta v}^{v_0 + \Delta v} A(v) \sqrt{v} \int_0^L \frac{u_1(x)}{\sqrt{L - x}}$$

$$\times \int_{-\infty}^{\infty} \sin\left(v\pi u_0 \frac{z^2}{(L - x)} + \frac{\pi}{4} \right) dz \, dx \, dv$$

$$= \frac{\sqrt{u_0}}{2i} \int_{v_0 - \Delta v}^{v_0 + \Delta v} A(v) \sqrt{v} \int_0^L \frac{u_1(x)}{\sqrt{L - x}}$$

$$\times \int_{-\infty}^{\infty} \left(e^{i(v\pi u_0 \frac{z^2}{(L - x)} + \frac{\pi}{4})} - e^{-i(v\pi u_0 \frac{z^2}{(L - x)} + \frac{\pi}{4})} \right) dz \, dx \, dv$$

$$\approx \frac{\sqrt{u_0}}{2i} \int_{v_0 - \Delta v}^{v_0 + \Delta v} A(v) \sqrt{v} \int_0^L \frac{u_1(x)}{\sqrt{L - x}} \left(2i \sqrt{\frac{L - x}{v u_0}} \right) dx \, dv$$

$$= \int_{v_0 - \Delta v}^{v_0 + \Delta v} A(v) \, dv \int_0^L u_1(x) \, dx$$

$$= \int_0^L u_1(x) \, dx , \tag{29}$$

which is directly comparable to the result for ray theory in eq. (12).

3 SET-UP OF THE NUMERICAL EXPERIMENT

In order to scrutinize the discrepancy between ray theory and scattering theory, we have constructed a 2-D numerical experiment with a slab of slowness perturbations that varies increasingly rapidly as a function of depth. The wavefield is initialized by a plane wave source. The name of this experiment is the 'sweep experiment' because the function of the slowness perturbation field resembles the sweep source function used in exploration seismology. The heterogeneous slab has a width W with its left-hand side at offset x_l from the source, as shown in Fig. 1. The slowness field is defined as

$$u(z) = \begin{cases} u_0 & \text{outside the slab}, \\ u_0 + \sqrt{2}\varepsilon u_0 \sin\left(\frac{(z + z_0)^4}{k} \right) & \text{inside the slab}, \end{cases} \tag{30}$$

where u_0 is the reference slowness and ε is the rms value of slowness perturbation fluctuations. The two parameters

$z_0 = 300$ m and $k = 1.5 \times 10^{10}$ m^4 are used to adjust the sweep function to the situation in which scattering theory becomes significant. In every sweep experiment, $u_0 = 2.5 \times 10^{-4}$ s m^{-1}, $x_l = 20$ m and $W = 20$ m. The size of this experiment is 100 m $\times$ 600 m, with the horizontal source position at zero offset, and the vertical receiver array is at $L = 100$ m offset. In the sweep experiment in Figs 4(a) and (b), $\varepsilon = 0.017$ and 0.035, respectively.

In addition to the sweep experiment, with another numerical experiment we demonstrate the validity of our scattering theory in a realization of the Gaussian random model. The auto-correlation function for Gaussian media is given in eq. (8). For one case, we applied the plane wave source as in the sweep experiment, and in the other case we used a point source to verify that the sensitivity kernel due to point sources in eq. (24) is correct. The Gaussian random media experiments in Figs 4(c) and (d) measure 200 m $\times$ 230 m and 100 m $\times$ 130 m for the plane wave source and the point source, respectively. For the incident plane wave, the initial wavefront is at $x = 0$, and the offset of the vertical array receivers is $L = 200$ m. For the point source, the source at zero offset is located at 65 m depth, and the vertical receiver array is at $L = 100$ m offset.

Because the mean value of slowness perturbations in finite sized realizations of Gaussian random media is not necessary zero, the value of the constant reference slowness u_0 differs in the two Gaussian random media experiments. It can be
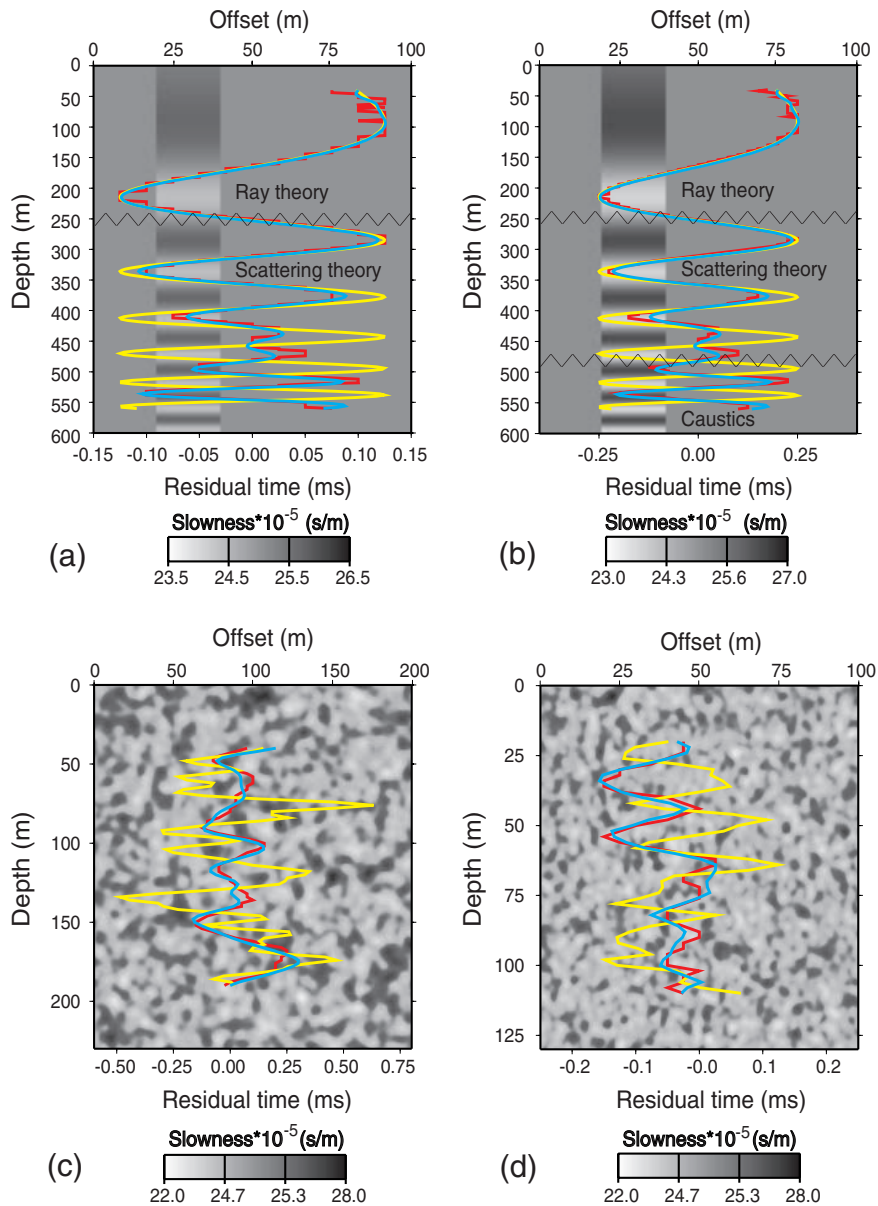

Figure 4. Performance of ray theory versus scattering theory in the numerical experiments. The slowness field is shown with greyscale shades. Time residuals for first-order ray perturbation theory (yellow line) and for scattering theory (blue line) are compared with the 'observed' time-shifts (red line). (a) The sweep experiment using a plane wave with $\varepsilon = 0.017$. (b) The sweep experiment using a plane wave with $\varepsilon = 0.035$. (c) The Gaussian random media experiment with $\varepsilon = 0.025$ and $a = 7$ m. The wavefield is initialized by a plane wave source. (d) The Gaussian random media experiment with $\varepsilon = 0.025$ and $a = 3$ m, but with a different slowness medium than that in (c). The wavefield is initiated by a point source.

proven that

$$\text{if} \quad \int N(\mathbf{r})d\mathbf{V} \neq 0 \quad \text{then} \quad \langle u_1 \rangle \neq 0, \tag{31}$$

where $N(\mathbf{r})$ is the autocorrelation of the random medium (Müller *et al.* 1992). We have chosen the reference slowness field to be equal to the mean value of the slowness field in the two Gaussian random media, e.g. $u_0 = \langle u \rangle$. Due to the more severe grid condition for using a point source than an incoming plane wave in the numerical experiment, we used different realizations of a Gaussian random medium in the two Gaussian experiments. As a result, the reference slowness is given by $u_0 = 2.470 \times 10^{-4}$ s m^{-1} for the test with an incoming plane wave, and 2.456×10^{-4} s m^{-1} for that with a point source. However, $\varepsilon = 0.025$ in both experiments.

To ascertain whether ray theory or scattering theory is dominant, we compare the theoretical residual times for ray theory and scattering theory with the 'observed' data determined in the following way. First, synthetic data for a reference model and a perturbed model are computed with a finite differences solution of the wave equation (FD code). For the reference model, the slowness field is set to the constant u_0, and, for the perturbed model, the sweep model or the Gaussian random medium is applied. The 'observed' residual times are then obtained by comparing the waveforms in the filtered reference wavefields with the waveforms in the filtered perturbed wavefields. By filtering we mean that the FD data are band-pass filtered in the same frequency range over which the sensitivity kernels are averaged. The first extremum of the waveform is used as a measuring point to obtain the absolute traveltime for each set of filtered waveforms. The 'observed' delay time is then defined as the difference between the absolute traveltime for the filtered reference wavefields and for the filtered perturbed wavefields.

4 RESULTS

In Figs 4(a) and (b) we show the traveltime changes for the sweep experiment due to an incident plane wave. The frequency band of the recorded wavefield is between 150 and 250 Hz.

The 2-D slowness field is shown with greyscales in both figures. The time delays due to first-order ray perturbation theory are plotted with a yellow line, residual times computed with the Rytov approximation are shown with a blue line, and the 'observed' time-shifts are shown with a red line. In both examples of the sweep experiment we used the sensitivity kernel in eq. (23). It is observed in Figs 4(a) and (b) that the FD time delays have some small but abrupt oscillations that are due to errors in the picking of the 'observed' data.

In Figs 4(a) and (b), we mark with a jagged black line the transition zone where ray theory breaks down in favour of scattering theory based on the condition for ray theory in eq. (13) that the width of Fresnel zones in eq. (26) is less than the local length scale a of slowness perturbations in the sweep experiment. For a central wavelength $\lambda = 20$ m and $x = 70$ m (the central distance of the heterogeneous slab from the receiver), we have $W_{\mathrm{sens}}^{\mathrm{plw}} = 65$ m. For comparison, in the centre of the transition zone ($z = 250$ m) the half-wavelength of the sweep function in eq. (30) is about 61 m. We conclude from these two experiments that, in general, the non-dimensional number

$L_{\mathrm{F}}^{\mathrm{plw}}/a$ for the regime of scattering theory is

$$\frac{L_{\mathrm{F}}^{\mathrm{plw}}}{a} > 1, \tag{32}$$

where $L_{\mathrm{F}}^{\mathrm{plw}}$ is the maximum width of Fresnel zones for plane waves.

In Figs 4(a) and (b), below the transition zone from ray theory to scattering theory the time-shifts computed with first-order ray perturbation theory cannot fit the 'observed' residual times; moreover, these ray-theoretical time delays are even out of phase with both the 'observed' time-shifts and the finite frequency time delays for depths between 450 and 560 m. The time delays due to single-scattering theory predict the FD time-shifts rather well. The scattering theory predicts not only the order of magnitude of the 'observed' time delays, but it also gives the correct result for depths below 450 m in the sweep experiment, where the FD time-shifts are anticorrelated with the ray-theoretical residual times.

In Fig. 5, we plot the focus position x_{caus} of a plane wavefield passing through the sweep model with $\varepsilon = 0.071$, which is on purpose a larger value than that applied in the sweep experiments in Figs 4(a) and (b). Given a depth z, the focal length (solid line) in Fig. 5 is computed with eqs (6) and (7), while the dashed line marks the source–receiver distance. Where the focal length of the converging wavefield is smaller than the distance between the source and receiver, caustics develop before measuring the wavefield at the receivers. Additionally, in Fig. 6 we show six snapshots (taken at the absolute traveltimes $t = 0, 5, 10, 15, 20$ and 25 ms) of the wavefield that propagates through the sweep model for $\varepsilon = 0.071$. The thin slab of inhomogeneity is marked in the figure with a black box
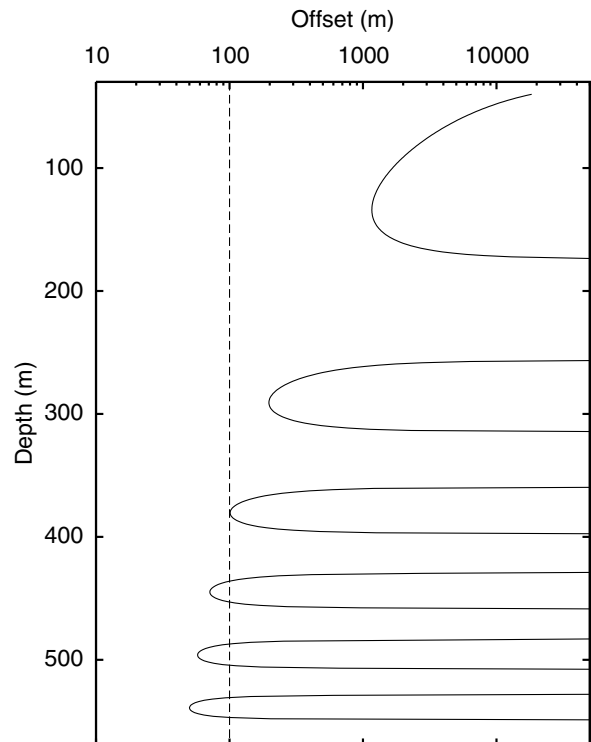

Figure 5. The focal length of a plane wavefield is calculated for the case of the sweep experiment with $\varepsilon = 0.071$. If the focal length (solid line) of the converging wavefield is to the left of the receiver position (dashed line) then caustics will occur in the recorded data.
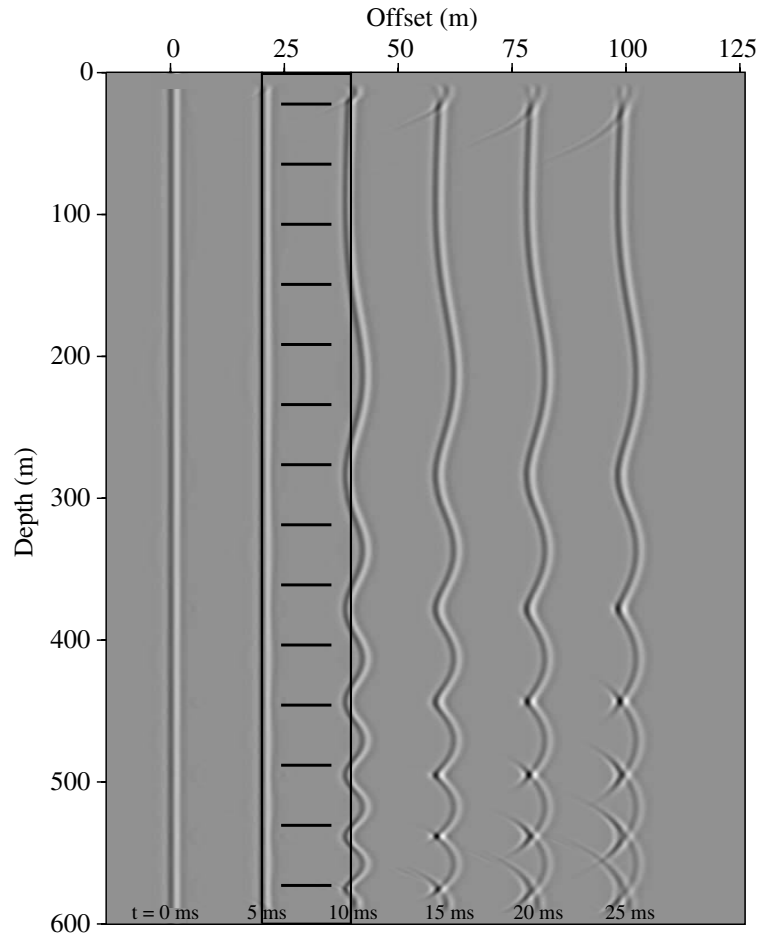
Figure 6. Snapshots of plane wave propagation in the sweep experiment with $\varepsilon = 0.071$. The slab of slowness heterogeneities is shown with the black box and horizontal black stripes. The absolute traveltimes $t = 0$, 5, 15, 20 and 25 ms are marked at the respective wavefronts. The triplications become clear in the wavefronts for $t = 15$, 20 and 25 ms.

and horizontal black stripes. For the two earliest snapshots, the wave propagates in a constant-slowness field. For $t = 10$ ms, the wave has just passed through the slowness perturbation field, so the wavefront has been deflected by slowness hetero-geneities. The first two triplications occur between 40 and 60 m offset. Note the high energy density at the kinks in the wave-front; these kinks are associated with the caustics. In the two latest snapshots, for $t = 20$ and 25 ms, the caustics are much more clear as they give rise to half-bow-tie-shaped wavefronts (triplications) behind the ballistic wavefield. The distances at which the caustics in Fig. 6 start to generate correspond well with the focal lengths that are predicted in Fig. 5.

For the sweep experiment with $\varepsilon = 0.017$, no caustics are produced in the wavefield, while for $\varepsilon = 0.035$, triplications occur before recording the wavefield at receivers at depths below 480 m. The zone with caustics is indicated in Fig. 4(b). At this point, we must reconsider the validity of the Rytov approxi-mation for transient waves where triplications are present in the recorded wavefield, such as in results of the sweep experiment as shown in Fig. 4(b). In comparing the Rytov approximation with the Born approximation, Beydoun & Tarantola (1988), Brown (1967), DeWolf (1967), Fried (1967), Hufnagel & Stanley (1964), Keller (1969), Sancer & Varvatsis (1970) and Taylor (1967) concluded that the Rytov approximation has validity for a larger range of the slowness perturbation parameter than does the Born approximation. They, however, do not investigate the validity of

the Born and Rytov approximations when non-linear effects such as the development of triplications become operative. In this study, we have tested the validity of the Rytov approximation in the regime of caustics. We computed the perturbed wave-fields for the sweep experiment with $\varepsilon = 0.017$, 0.035, 0.071, 0.11 and 0.14, and estimated the FD time delays by using the first extremum of the filtered waveforms. For the sweep experiment with $\varepsilon = 0.017$, the theory for caustics in 1-D slowness pertur-bation fields predicts that triplications would not be recorded in the data, but for larger ε triplications would always occur in the measured wavefield. We have shown the sweep experiment with the two lowest values of ε in Figs 4(a) and (b), but the sweep experiments with larger ε are not shown here. In brief, we find that the Rytov approximation does a good job even in areas with a strong development of triplications. We therefore propose that the validity of the Rytov approximation of ballistic waves extends into the regime where caustics are present in data.

In order to demonstrate the validity of the single-scattering theory in more complex media, we use the Gaussian random media experiment, where scattering is significant. The relative rms value of slowness fluctuations ε is given by 0.025, and the length scale of slowness anomalies, a, is 7 m for the incoming plane wave experiment and 3 m for the point source experi-ment. We use the same colour convention as in the sweep experiment for the residual times computed with ray theory, scattering theory and the FD code. Results for this experiment,

for an incident plane wave and a point source, respectively, are shown in Figs 4(c) and (d), where we make use of the sensitivity kernel formulation in eqs (23) and (24) to compute the scattering theoretical time-shifts.

For an incident plane wave, the frequency is from 150 to 250 Hz, so according to eq. (32), ray theory breaks down when the characteristic length of slowness anomalies is smaller than $L_F^{plw} = 110$ m. In this case, the length scale of slowness perturbations $a = 7$ m, so the 'observed' time delays should be strongly dominated by scattering. This is indeed what is observed in Fig. 4(c). The finite frequency residual times fit the 'observed' time delays correctly while ray theory does not account for the traveltime deviations.

Fig. 4(d) shows results of the Gaussian random media experiment for a point source with frequencies ranging from 200 to 400 Hz. Using the sensitivity kernel in eq. (24) to compute the residual times due to scattering theory, we compute the maximum width of Fresnel zones as $L_F^{ps} = 31.9$ m for $\lambda = 13.6$ m and $L = 100$ m. The length scale of slowness anomalies, $a = 3$ m, is thus about 10 per cent of L_F^{ps}. In Fig. 4(d), scattering theory predicts the 'observed' residual times well but the ray-theoretical time-shifts do not fit the FD time delays. As in the experiments in Figs 4(a), (b) and (c) where a plane wave is applied, we conclude that the regime of scattering theory for wavefields emitted by a point source is significant when

$$\frac{L_F^{ps}}{a} > 1 \, . \tag{33}$$

We have ascertained whether the recorded wavefields for plane waves and point sources in the Gaussian random media experiments contain triplications. By inserting $\varepsilon = 0.025$ in the condition for caustic formation in eqs (9) and (10), we find that

$$\frac{L}{a} \geq \begin{cases} 6.1 & \text{for plane waves}, \\ 13.1 & \text{for point sources}. \end{cases} \tag{34}$$

With the autocorrelation lengths $a = 7$ m for the plane wave and $a = 3$ m for the point source, this implies that caustics are present in the recorded wavefields in the Gaussian experiments in Figs 4(c) (a plane wave with $L = 200$ m) and (d) (a point source with $L = 100$ m).

The non-dimensional numbers for the regime of ray theory, scattering theory and triplications are summarized in Table 1. Notice that four parameters determine when these three distinct regimes are significant. These four parameters are the wavelength λ, the source–receiver distance L, the relative rms value of slowness fluctuations ε and the length scale a of inhomogeneity.

Table 1. The non-dimensional numbers λ/a, L_F/a and L/a that describe the regime of ray theory, scattering theory and triplications. $\times$ indicates that this parameter is not of relevance for the physical effects considered.

Regime	λ/a	L_F/a	L/a
Ray theory	$\ll 1$	$\ll 1$	$\times$
Scattering theory	$\times$	> 1	$\times$
Triplications	$\times$	$\times$	$> \begin{cases} 0.52\varepsilon^{-2/3} & \text{(plw)} \\ 1.12\varepsilon^{-2/3} & \text{(ps)} \end{cases}$

5 APPLICATION OF THE REGIME OF SCATTERING THEORY

We consider the implications for three examples taken from seismology, ocean acoustic and medical imaging for which scattering theory is important. The source in all three cases is a point source, so the condition for scattering theory due to point sources (eq. 33) is that $L_F^{ps}/a > 1$. The example from seismology is global surface wave tomography, where the surface waves propagate in two dimensions on a sphere, while in the case of ocean acoustic or medical imaging the wave propagation is in 3-D Cartesian space. Thus the dimension of the wave propagation in each particular experiment must be considered.

The maximum width of Fresnel zones for point sources occurs at the half-source–receiver distance. We find that

$$L_F^{ps} = \sqrt{\frac{3\lambda}{2} \tan\left(\frac{\Delta}{2}\right)} \quad \text{(a sphere)} \quad \text{and}$$

$$L_F^{ps} = \sqrt{\lambda L} \quad \text{(3-D Cartesian space)} \, . \tag{35}$$

(See appendix A for a derivation of the width of Fresnel zones for surface waves.) For the widths of the Fresnel zone on the sphere, we have used $n = 8/3$ according to eq. (28). On the sphere both the wavelength and the epicentral distance are measured in radians. The parameter Δ denotes the epicentral distance between the source and receiver. For L_F^{ps} in a 3-D Cartesian space, we applied eq. (5) for $n = 2$.

In global surface wave tomography (Trampert & Woodhouse 1995, 1996), a characteristic propagation distance is about 145°, and the wavelength is about 700 km for Love waves at 150 s. For a high-resolution, global surface wave experiment, slowness anomalies have a length scale as small as 1000 km. We find that $L_F^{ps} = 4600$ km, so scattering theory is important. Scattering theory will be even more significant for larger wavelengths and source–receiver distances.

In ocean acoustic, Hodgkiss *et al.* (1999) and Kuperman *et al.* (1998) carried out a time-reversed mirror experiment wherein the source–receiver distance was 6.3 km and the characteristic wavelength $\lambda = 3.4$ m (sound speed in sea water with 3.5 per cent salinity at 20 °C is 1522 m s^{-1}, and the characteristic frequency of acoustic waves in the experiment was 445 Hz). The width of the Fresnel zone is then 146 m, which is larger than the surface bottom depth in their experiment. This means that any heterogeneity within the region of the experiment is smaller than the width of the Fresnel zone, so scattering theory is significant.

In medical imaging (Baba *et al.* 1989; King & Shao 1990), ultrasound is applied to scan the brain, the chest, the foetus, etc. The velocity of the employed waves varies between 1440 m s^{-1} (fat) and 1675 m s^{-1} (collagen), while the frequency is in the MHz range, so let the frequency $\nu = 30$ MHz. The wavelength then varies between 48 and 56 μm and the average distance L between the transducer and receiver instrument is about 20 cm. The width of the Fresnel zone $L_F^{ps} \approx 3$ mm, which is greater than the diameter of blood vessels and cell structures in the body. Scattering theory is thus important.

6 CONCLUSIONS

We have shown that first-order ray theory breaks down in favour of linearized scattering theory in predicting the traveltime shifts

of waves in heterogeneous media when the length scale of slowness heterogeneity is smaller than the widths of Fresnel zones. The condition for the regime of scattering theory depends on the frequency content of the recorded wavefield and the propagation length of the ballistic wave between the source and receiver.

The scattering theory presented in this paper is based on the first-order Rytov approximation, so the scattering theoretical time delays are well defined for finite frequencies. Physically, this means that finite frequency time-shifts have the maximum sensitivity to slowness fluctuations off the path of the ray and, moreover, are sensitive to the slowness fluctuation field in the whole space of wave propagation. In contrast, ray-theoretical residual times are dependent on only the slowness fluctuation field which is on the geometrical ray.

Scattering theory can predict the residual traveltimes of waves in inhomogeneous media even if triplications are present in the recorded wavefield. We have presented a condition for the regime of caustics in heterogeneous media, both with a 1-D slowness field and with a Gaussian random medium for initially plane waves and point sources. This condition is independent of the frequency of the recorded wavefield. Not surprisingly, we have found that the greater the magnitude of the slowness fluctuations, the more easily triplications develop in the wave-fields. However, the $\varepsilon^{2/3}$ dependence for slowness fields described by Gaussian random media is non-trivial. Notice that in the numerical experiments the Rytov approximation provides an accurate estimate of the time-shift, regardless of whether caustics have developed or not.

The numerical experiments carried out in this paper are kept as general as possible. The results for the regimes of scattering theory and triplications are therefore applicable to domains such as seismology, ocean acoustic, non-destructive testing and medical imaging where wave phenomena are important.

ACKNOWLEDGMENTS

REFERENCES

Baba, K., Satch, S., Sakamoto, S., Okai, T. & Shiego, I., 1989. Development of an ultrasonic system for three-dimensional recon-struction of the fetus, *J. perinat. Med.,* **17,** 19–24.

Beydoun, W.B. & Tarantola, A., 1988. First Born and Rytov approximations: modeling and inversion conditions in a canonical example, *J. acoust. Soc. Am.,* **83,** 1045–1055.

Bleistein, N., 1984. *Mathematical Methods for Wave Phenomena,* Academic Press, Orlando, FL.

Brown, W.P., 1967. Validity of the Rytov approximation, *J. opt. Soc. Am.,* **57,** 1539–1543.

DeWolf, D.A., 1967. Validity of Rytov's approximation, *J. opt. Soc. Am.,* **57,** 1057–1059.

Fehler, M., Sato, H. & Huang, L., 2000. Envelope broadening of outgoing waves in 2D random media: a comparison between the Markov approximation and numerical simulations, *Bull. seism. Soc. Am.,* **90,** 914–928.

Fried, D.L., 1967. Test of the Rytov approximation, *J. opt. Soc. Am.,* **57,** 268–269.

Haque, S., Lu, G.Q., Goings, J. & Sigmund, J., 1999. Characterization of interfacial thermal resistance by acoustic micrography imaging, *Proc. 1999 Annual Power Electronics Seminar at Virginia Tech, September 19–21,* 375–382.

Hodgkiss, W.S., Song, H.C., Kuperman, W.A., Akal, T., Farla, C. & Jackson, D., 1999. A long-range and variable focus phase-conjugation experiment in shallow water, *J. acoust. Soc. Am.,* **105,** 1597–1604.

Hufnagel, R.P. & Stanley, N.R., 1964. Modulation transfer function associated with image transmission through turbulent media, *J. opt. Soc. Am.,* **54,** 52–61.

Jensen, J.M. & Jacobsen, B.H., 1997. Rapid exact linearized inversion of time-distance helioseismic data in the periodic approximation, in *Proc. Interdisciplinary Inversion Workshop 5, Methodology and Applications in Geophysics, Astronomy, Geodesy and Physics,* pp. 57–67, ed. Jacobsen, B.H., Århus University, Denmark.

Keller, J.B., 1969. Accuracy and validity of the Born and Rytov approximation, *J. opt. Soc. Am.,* **59,** 1003–1004.

King, D.L. & Shao, M.Y., 1990. Three-dimensional spatial registration and interactive display of position and orientation of real-time ultrasound images, *J. Ultrasound Medicine,* **9,** 525–532.

Kravtsov, Y.A., 1988. Rays and caustics as physical objects, in *Progress in Optics,* Vol. XXVI, pp. 227–348, ed. Wolf, E., Elsevier, Amsterdam.

Kuperman, W.A., Hodgkiss, W.S., Song, H.C., Akal, T., Farla, C. & Jackson, D., 1998. Phase conjugation in the ocean: experimental demonstration of an acoustic time-reversal mirror, *J. acoust. Soc. Am.,* **103,** 25–40.

Marquering, H., Nolet, G. & Dahlen, F.A., 1998. Three-dimensional waveform sensitivity kernels, *Geophys. J. Int.,* **132,** 521–534.

Marquering, H., Dahlen, F.A. & Nolet, G., 1999. The body wave traveltime paradox: bananas, doughnuts and three-dimensional delay-time kernels, *Geophys. J. Int.,* **137,** 805–815.

Menke, W. & Abbot, D., 1990. *Geophysical Theory,* Columbia University Press, NY.

Müller, G., Roth, M. & Korn, M., 1992. Seismic wave traveltimes in random media, *Geophys. J. Int.,* **110,** 29–41.

Passier, M.L. & Snieder, R., 1995. Using differential waveform data to retrieve local S-velocity structure or path-averaged S-velocity gradients, *J. geophys. Res.,* **100,** 24 061–24 078.

Pulliam, J. & Snieder, R., 1998. Ray perturbation theory, dynamic ray tracing, and the determination of Fresnel zones, *Geophys. J. Int.,* **135,** 463–469.

Sancer, M.I. & Varvatsis, A.D., 1970. A comparison of the Born and Rytov methods in *Proc. IEE,* 140–141.

Snieder, R. & Lomax, A., 1996. Wavefield smoothing and the effect of rough velocity perturbations on arrival times and amplitudes, *Geophys. J. Int.,* **125,** 796–812.

Snieder, R. & Sambridge, M., 1992. Ray perturbation theory for travel times and raypaths in 3-D heterogeneous media, *Geophys. J. Int.,* **109,** 294–322.

Spetzler, J. & Snieder, R., 2001. The formation of caustics in two- and three-dimensional media, *Geophys. J. Int.,* **144,** 175–182.

Taylor, L.S., 1967. On Rytov's methods, *Radio Sci.,* **2,** 437–441.

Trampert, J. & Woodhouse, J.H., 1995. Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int.,* **122,** 675–690.

Trampert, J. & Woodhouse, J.H., 1996. High resolution global phase velocity distributions, *Geophys. Res. Lett.,* **23,** 21–24.

Woodward, M.J., 1992. Wave-equation tomography, *Geophysics,* **57,** 15–26.

Yomogida, K., 1992. Fresnel zone inversion for lateral heterogeneities in the Earth, *PAGEOPH,* **138,** 391–406.
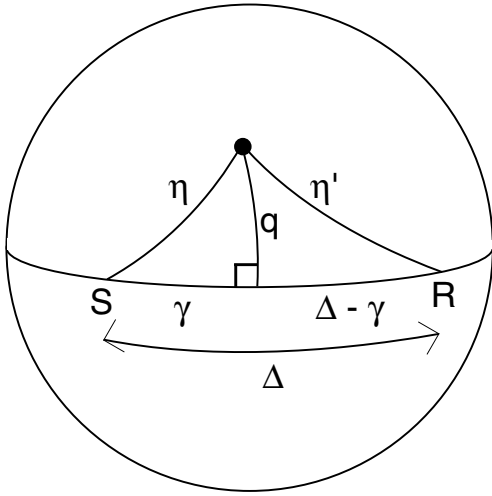
Figure A1. Explanation of the variables used to construct the Fresnel zone due to a point source on a sphere.

APPENDIX A: THE WIDTHS OF FRESNEL ZONES ON A SPHERE

According to Fig. A1, the epicentral distance between the source and receiver is denoted by Δ, and the epicentral distance between the source and scatterer point and the scatterer point and receiver are marked as η and η', respectively. The half-width of the Fresnel zone at offset γ is denoted q. Using the law of cosines on a sphere to relate η to q and γ, we obtain

$$\cos(\eta) = \cos(q)\cos(\gamma) + \sin(q)\sin(\gamma)\cos\left(\frac{\pi}{2}\right)$$

$$= \cos(q)\cos(\gamma) . \tag{A1}$$

Isolating η from eq. (A1) and assuming that the ray deflection q is small gives

$$\eta = \arccos(\cos(q)\cos(\gamma))$$

$$\approx \arccos\left(\cos(\gamma) - \frac{1}{2}q^2\cos(\gamma)\right)$$

$$\approx \gamma + \frac{q^2}{2\tan(\gamma)} . \tag{A2}$$

Similarly, we have for η'

$$\eta' = (\Delta - \gamma) + \frac{q^2}{2\tan(\Delta - \gamma)} . \tag{A3}$$

The detour $\eta + \eta' - \Delta$ is calculated as

$$\eta + \eta' - \Delta = \frac{q^2}{2}\left(\frac{1}{\tan(\gamma)} + \frac{1}{\tan(\Delta - \gamma)}\right)$$

$$= \frac{q^2}{2}\frac{\sin(\Delta)}{\sin(\gamma)\sin(\Delta - \gamma)} . \tag{A4}$$

The condition for Fresnel zones on a sphere that the detour is less than the wavelength divided by a number n is given by

$$\eta + \eta' - \Delta \leq \frac{\lambda}{n} , \tag{A5}$$

where λ is the wavelength measured in radians. The detour in eq. (A4) is inserted into the Fresnel zone condition in eq. (A5), where the sign of equality is applied for the Fresnel zone boundary. Thereby the half-width q of the Fresnel zone is given by

$$q = \sqrt{\frac{2\lambda\sin(\gamma)\sin(\Delta - \gamma)}{n\sin(\Delta)}} , \tag{A6}$$

which has the largest value for $\gamma = \Delta/2$. For that case,

$$q = \sqrt{\frac{\lambda}{n}\tan\left(\frac{\Delta}{2}\right)} . \tag{A7}$$

The maximum width L_F^{ps} of Fresnel zones due to point sources is the half-width q in eq. (A7) multiplied by 2, thus

$$L_F^{ps} = \sqrt{\frac{4\lambda}{n}\tan\left(\frac{\Delta}{2}\right)} , \tag{A8}$$

where L_F^{ps} and λ are measured in radians.