Geophys. J. Int. (2004) **158,** 187–210 doi: 10.1111/j.1365-246X.2004.02300.x

Statistics of traveltimes and amplitudes in random media

A. M. Baig* and F. A. Dahlen

Department of Geosciences, Princeton University, Princeton, NJ 08544*, USA. E-mail: abaig@eos.ubc.ca*

Accepted 2004 February 27. Received 2004 February 13; in original form 2003 May 8

SUMMARY

In this study, we build on the results of the study of 3-D wave propagation in weakly heterogeneous media conducted by Baig *et al.* We measure traveltimes and amplitudes from 'ground-truth' seismograms, computed using a numerical wave propagation code, and compare the measurements with approximate finite-frequency and ray-theoretical values for these quantities. Ray-theoretical traveltimes become invalid whenever the scale length of the 3-D heterogeneity is smaller than half the maximum width of the Fresnel zone; in contrast, ray-theoretical amplitudes have a much more restricted range of validity: the scale length should not be less than one Fresnel-zone maximum width. Finite-frequency theory gives better results for amplitudes, suffering no observable degradation for small-scale media for the weakest heterogeneity considered, but suffering appreciable misfit in more strongly heterogeneous media. Using these finite-frequency expressions for traveltime, we derive expressions for the expected variances of traveltimes and amplitudes that act, in most cases, as extensions to the ray-theoretical expressions. Finally, we propose using the amplitude variance as a criterion for delineating the validity of these linear approximations. For traveltimes, provided that one rejects waveforms that do not yield a good cross-correlation traveltime, the remaining data are linearly related to the model over the values of theoretical amplitude variance that we probe in this experiment. Amplitudes do not behave as well: when the theoretical amplitude variance rises above 0.1, significant non-linearities start to invalidate our linear approximation.

Key words: body waves, diffraction, inhomogeneous media, ray theory, traveltime, wave propagation.

1 INTRODUCTION

Recent tomographic images (Inoue *et al.* 1990; Su & Dziewonski 1992, 1997; Pulliam *et al.* 1993; Grand 1994; Masters *et al.* 1996, 2000; Grand *et al.* 1997; Van der Hilst *et al.* 1997; Vasco & Johnson 1998; Boschi & Dziewonski 2000; Montelli *et al.,* 2004a) of the mantle have managed to elucidate some of the geodynamic processes governing convection in this part of the Earth. These images rely on traveltime residuals of many *P* and *S* seismic body wave phases as their main source of data, inverting these data for 3-D velocity heterogeneity. In a recent publication, Tibuleac *et al.* (2003) have argued for the feasibility of adding amplitude anomalies as another source of data that provides insight into mantle structure. The process of translating these traveltime delays into structure, with the exception of the study by Montelli *et al.* (2004a), was performed using ray theory, confining any sensitivity to structure along the ray. However, there is a strong frequency dependence to traveltimes and amplitudes, as elucidated by a number of theoretical studies on the Fréchet (or, more whimsically, banana-doughnut) kernels for these observables (Marquering *et al.* 1999; Dahlen *et al.* 2000; Hung *et al.* 2000; Zhao *et al.* 2000; Dahlen & Baig 2002; Spetzler & Sneider 2001b; Baig *et al.* 2003; Montelli *et al.* 2004b). Although ray theory may be thought of as a limiting regime of this finite-frequency theory, exactly where this transition occurs in parameter space is of interest to us. Baig *et al.* (2003) found that, if the maximum width of the first Fresnel zone exceeds two structural scale lengths, ray theory becomes an inaccurate tool in predicting traveltime delays. If amplitudes are to be used in the future to augment teleseismic data sets, an investigation into the behaviour of amplitudes over this transition is also warranted. Furthermore, we should note that traveltime delays and amplitude perturbations are just two quantities that one can observe on a given seismogram, and that each sample of the waveform provides differing constraints on Earth structure, leading into the realm of waveform inversion.

*Now at: Department of Earth and Ocean Sciences, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

The first section after this Introduction describes in moderate detail our procedure for computing synthetic traveltimes and amplitudes in various random media. We use a pseudo-spectral technique to solve the wave equation and generate 'ground-truth' seismograms in several weakly heterogeneous random media. Measured traveltime and amplitude perturbations are obtained by comparing these observed waveforms with seismograms for the same source–receiver path in the homogeneous model. Section 3 contrasts this 'nuts-and-bolts' discourse of the numerics with a brief theoretical discussion on ray-theoretical and finite-frequency approaches to calculating the first-order dependence of traveltimes and amplitudes on weak structure. The scatterplots that we use to compare these theories with the measured data are introduced in the next section. Switching gears, Section 5 contains a lengthy theoretical discussion on how one can use these finite-frequency expressions to calculate the ensemble distributions of these data. As traveltimes and amplitudes are influenced by the velocity anomalies in the medium, the variances of these observational anomalies can provide important insight into the spectrum of velocity heterogeneity, as Gudmundsson *et al.* (1990) pointed out by investigating the distribution of ISC traveltimes to infer the spectrum of velocity anomalies in the mantle. Although we feel that we need to more completely sample path space in our numerical models to obtain an accurate measure of the ensemble variances, we can robustly measure best-fitting lines for scatterplots of theoretical versus measured traveltimes, so we also derive theoretical expressions for the slopes of these best-fitting lines. We argue that these slopes act as proxies for the validity of our expressions for variance. Finally, in the penultimate section, we derive a condition for the validity of these linear theories and compare this condition with the amount of non-linearity that we observe in our seismograms.

2 3-D NUMERICAL WAVE PROPAGATION

Baig *et al.* (2003) generated an extensive suite of synthetic seismograms for their study of acoustic wave propagation in random media. As we will utilize their data set, and even augment it, we shall review their numerical technique in the following section. In addition, we shall use this opportunity to introduce the so-called 'mantle in a box' models of random slowness perturbations. A more detailed discussion of the numerics is given by Baig *et al.* (2003).

2.1 The 'mantle in a box'

As we are interested in numerically simulating wave propagation in random media, we need to discuss the recipe for creating these media. For most of our computations, the media are confined to a $256 \times 256 \times 256$ grid-point cube. At a grid spacing of $\Delta x = 30$ km, this corresponds to a range of 7650 km for each side. A few computations were carried out to a propagation distance of 19 170 km, in a $256 \times 256 \times 640$ box. We did not vary the density, ρ, from 1000 kg m^{-3}, only considering deviations from the homogeneous background slowness of $\sigma = 125$ μs m^{-1} (or a background velocity of $\sigma^{-1} = 8$ km s^{-1}). The perturbations to this background slowness, $\delta\sigma(\mathbf{x})$, are a single realization of the ensemble of random media that have zero mean,

$$\langle \delta\sigma(\mathbf{x}) \rangle = 0, \tag{1}$$

and a specific autocorrelation function,

$$\langle \delta\sigma(\mathbf{x}) \delta\sigma(\mathbf{x}') \rangle = \sigma^2 N(r). \tag{2}$$

The single argument of N, $r = \|\mathbf{x} - \mathbf{x}'\|$, ensures that the autocorrelation is isotropic; there is no tendency for anomalies to be elongated in any one direction.

The autocorrelation is related, via a Fourier sine transform, to the power spectrum, $\Phi(k)$, of the medium (Tatarskii 1961; Sato & Fehler 1997):

$$N(r) = \sqrt{\frac{2}{\pi}} \int_0^\infty k \Phi(k) \sin(kr)\, dk, \qquad \Phi(k) = \sqrt{\frac{2}{\pi}} \int_0^\infty r N(r) \sin(kr)\, dr. \tag{3}$$

We create our perturbation field by first specifying its power spectrum, assigning each wavenumber a random phase, and then inverse Fourier transforming into the spatial domain. There are two types of media that we consider, Gaussian and exponential, with autocorrelation functions of the form

$$N_{\mathrm{g}}(r) = \varepsilon^2 \exp(-r^2/a^2) \tag{4}$$

for Gaussian media, and

$$N_{\mathrm{e}}(r) = \varepsilon^2 \exp(-r/a) \tag{5}$$

for exponential media. We shall adhere to the convention of assigning a subscript 'g' to functions pertinent to Gaussian media, reserving a subscript 'e' for the counterpart functions in exponential media. Two new quantities have been introduced in eqs (4) and (5): ε is the root mean square of the fractional perturbations of slowness,

$$\varepsilon^2 = \frac{\langle \delta\sigma^2 \rangle}{\sigma^2}; \tag{6}$$

whereas a is a measure of the minimum distance between two relatively uncorrelated points in the slowness field, and it is therefore known as the correlation length. Furthermore, the power spectrum of a Gaussian medium is given by

$$\Phi_{\mathrm{g}}(k) = \frac{\varepsilon^2 a^3}{2\sqrt{2}} \exp(-k^2 a^2/4), \tag{7}$$

whereas an exponential medium has a power spectrum of

$$\Phi_{\mathrm{e}}(k) = \frac{2\sqrt{2}\varepsilon^2 a^3}{\sqrt{\pi}(1 + k^2 a^2)^2}. \tag{8}$$

For a given value of a, an exponential medium has a much smaller scale structure than a Gaussian model, since the high wavenumbers in the power spectrum decay at a much slower rate. Typical realizations of a Gaussian and an exponential 'mantle in a box' are shown in Fig. 1, with $a = 600$ km in both cases.

2.2 Numerical wave propagation

Simulation of acoustic wave propagation is accomplished by solving the following system of four equations for pressure, p, and the three components of particle velocity, $\mathbf{u}$ (Morse & Ingard 1968, sections 6.2 and 7.1):

$$\partial_t \mathbf{u}(\mathbf{x}, t) = -\rho^{-1} \nabla p(\mathbf{x}, t), \tag{9}$$

$$\partial_t p(\mathbf{x}, t) = -\rho[\sigma + \delta\sigma(\mathbf{x})]^{-2}[\nabla \cdot \mathbf{u}(\mathbf{x}, t) - m(t)\delta(\mathbf{x} - \mathbf{s})]. \tag{10}$$

This system of equations is numerically integrated using the pseudospectral code of Hung & Forsyth (1998). All of the spatial derivatives, $\nabla p(\mathbf{x}, t)$ and $\nabla \cdot \mathbf{u}(\mathbf{x}, t)$, are computed by multiplication in the wavenumber domain, necessitating discretely Fourier transforming $p(\mathbf{x}, t)$ and

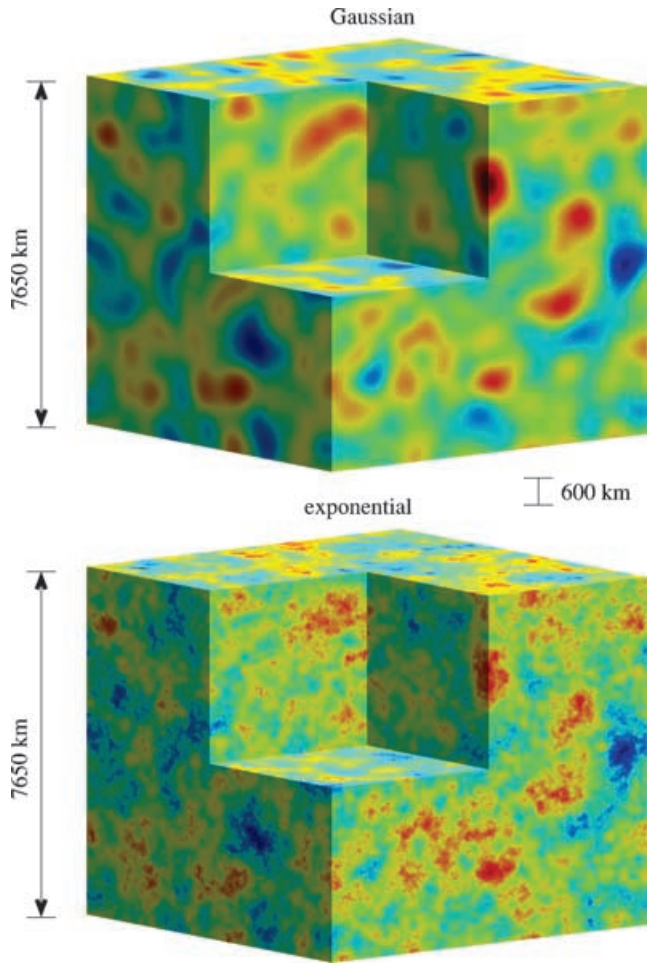

Figure 1. Two 'mantle in a box' models of randomly distributed slowness perturbations with a Gaussian (above) or an exponential (below) correlation function. Both media have a correlation length $a = 600$ km, but the exponential model features much smaller-scale structure than its Gaussian counterpart, because there is much more power at higher wavenumbers in this case. The fact that the same random seed was used in the generation of these models is reflected in the large-scale similarities of the anomalies.

$\mathbf{u}(\mathbf{x}, t)$ at every time step. This operation is advantageous in comparison with finite-difference techniques in that it leads to less numerical dispersion than the local approximation of these derivatives. The system (9) and (10) is then evolved in time, using a fourth-order Runge–Kutta scheme. With our choice of time step, $\Delta t = 0.25$ s, we more than satisfy the von Neumann stability criterion (Kosloff & Baysall 1982; Kosloff *et al.* 1984):

$$\Delta t \leq \frac{2\sigma \Delta x}{\sqrt{3}\pi} \approx 1.4 \text{ s}. \tag{11}$$

Finally, to minimize the effects of the boundaries of our $7650 \times 7650 \times 7650$ km^3 box, the code implements the absorbing boundary scheme of Cerjan *et al.* (1985), which effectively suppresses waves from reflecting at near normal incidence but is less effective for waves that graze this absorbing layer.

To initiate wave propagation in eqs (9) and (10), we need to include the source term, $m(t)\delta(\mathbf{x} - \mathbf{s})$. Effectively, this creates a point source at $\mathbf{s}$ with a source-time (moment, not moment-rate) function, $m(t)$, which we take to be a Gaussian pulse:

$$m(t) = \exp\left[-2\pi^2 \left(\frac{t}{\sigma\lambda} - \frac{1}{2}\right)^2\right], \tag{12}$$

where the introduction of the characteristic wavelength, $\lambda = 200$ km, adds another length scale into our analysis. The pressure response in the absence of slowness perturbations at a time t and a point $\mathbf{r}$, a distance $L = \| \mathbf{r} - \mathbf{s}\|$away from the source, is well known (Morse & Ingard 1968, section 7.1),

$$\begin{aligned} p_{\text{syn}}(t) &= \frac{\rho \dot{m}(t - \sigma L)}{4\pi L} \\ &= -\frac{\pi\rho}{\sigma\lambda L}\left(\frac{t}{\sigma\lambda} - \frac{L}{\lambda} - \frac{1}{2}\right) \exp\left[-2\pi^2 \left(\frac{t}{\sigma\lambda} - \frac{L}{\lambda} - \frac{1}{2}\right)^2\right], \end{aligned} \tag{13}$$

and will be used in Section 2.3 for our prescriptions for measuring traveltimes and amplitudes in eqs (16) and (17), respectively.

By placing the source near the edge of the box, we are able to maximize the distance, L, at which we can place 2-D fan arrays of receivers. In practice, we place three other arrays between the source and this most distant array. A quick parameter count reveals that there are three dimensionless numbers governing acoustic wave propagation in isotropic random media: ε, a/λ, and L/λ. For our $256 \times 256 \times 256$ runs, we have 32 data sets of seismograms recorded in Gaussian and exponential media, with four different values of correlation length ($a = 0.75\lambda$, 1.5λ, 2.25λ , 3λ) and four different values of relative heterogeneity strength ($\varepsilon = 0.01$, 0.02, 0.03, 0.04). Two long propagation runs in a $256 \times 256 \times 640$ box were calculated for $\varepsilon = 0.01$ at two scalelengths, $a = 0.75\lambda$ and $a = 3\lambda$. Each of these data sets was divided into four 13×13 arrays, distinguished on the basis of propagation distance from the source. Normally, our receivers were situated in arrays $L = 7.5\lambda$, 15λ, 22.5λ, and 30λ away from the source, although these array distances stretched out to $L = 21.2\lambda$, 42.4λ, 63.6λ, and84.5λ in our $256 \times 256 \times 640$ models. For each run, we doubled our data by considering the 'mirror images' of each medium, reversing the signs of the perturbation field $\delta\sigma\ (\mathbf{x}) \rightarrow -\delta\sigma(\mathbf{x})$, so that we preserve our location in parameter space. This augmentation is an admittedly crude way to improve our statistics by a meagre two-fold increase in path space, but we hope to remove thereby any biases in our data caused by the source being in a highly anomalous region of the slowness field.

2.3 Measured traveltimes and amplitudes

We have a number of seismograms from which we measure both traveltimes and amplitudes. In order to measure the traveltimes, we compare the waveform in the unperturbed medium, $p_{\text{syn}}(t)$, with the observed waveform, $p_{\text{obs}}(t)$, in the heterogeneous medium. Furthermore, we need to make the assumption that the recorded waveform is basically the unperturbed pulse shifted in time by a traveltime delay, δT, and scaled by an amplitude factor, $1 + \delta(\ln A)$:

$$p_{\text{obs}}(t) \approx [1 + \delta(\ln A)] p_{\text{syn}}(t - \delta T). \tag{14}$$

Under these assumptions, Ruff (1989) advocated measuring both traveltimes and amplitudes by minimizing a cost function of the form

$$\text{cost} = \int_{t_1}^{t_2} \{p_{\text{obs}}(t + \delta T) - [1 + \delta(\ln A)]\, p_{\text{syn}}(t)\}^2 dt \tag{15}$$

over some window, $[t_1, t_2]$. Provided that this window encapsulates all of the unperturbed pulse, such that $p_{\text{syn}}(t_1) = p_{\text{syn}}(t_2) = 0$, the traveltime one obtains from such a procedure is identical, to first order, to the lag at which the maximum of the cross-correlation function between the unperturbed and perturbed pulse occurs:

$$\int_{t_1}^{t_2} p_{\text{syn}}(t) p_{\text{obs}}(t + \delta T)\, dt = \text{maximum}. \tag{16}$$

Under the same restrictions on t_1 and t_2, the amplitude that minimizes the cost function (15), given δT, is also, within first-order, given by the same expression as that found by Dahlen & Baig (2002) using a slightly different argument:

$$\delta(\ln A) = \sqrt{\frac{\int_{t_1}^{t_2} p_{\text{obs}}^2(t + \delta T)\, dt}{\int_{t_1}^{t_2} p_{\text{syn}}^2(t)\, dt}} - 1. \tag{17}$$

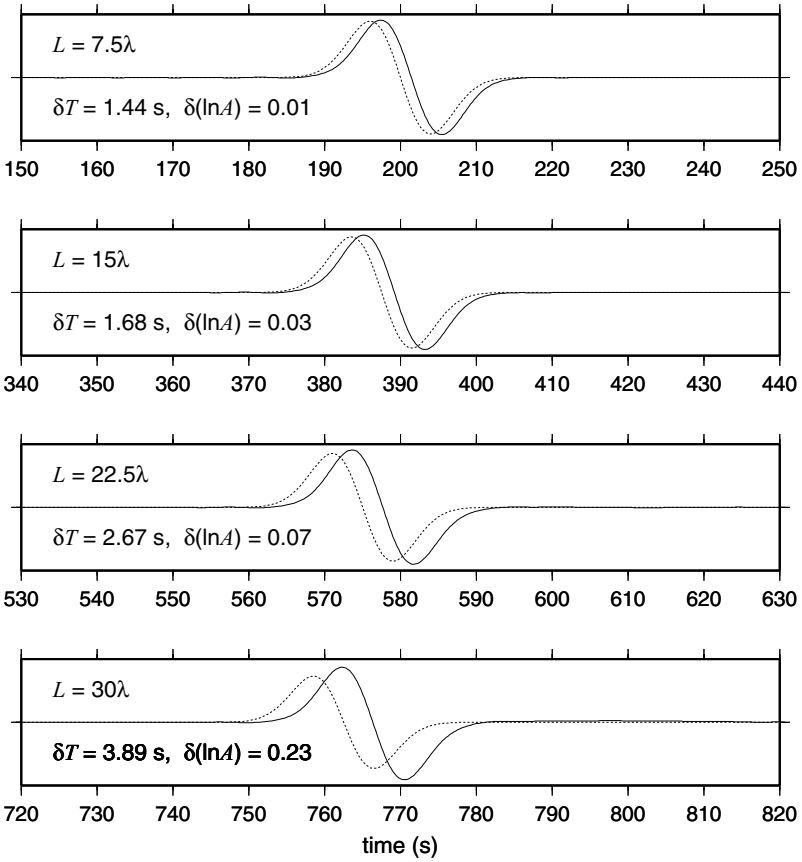

Figure 2. A set of seismograms arranged in increasing propagation distance from $L = 7.5\lambda$ at the top to $L = 30\lambda$. The perturbed seismograms (solid lines) are generated in a Gaussian medium with $\varepsilon = 0.01$ and $a = 2.25\lambda$. To obtain the measured values of δT and $\delta(\ln A)$, these waveforms are compared against the unperturbed seismograms (dashed lines).

We shall measure traveltimes and amplitudes using eqs (16) and (17) respectively, rather than by a grid-search minimization of the cost function in eq. (15).

In Fig. 2, we show a number of seismograms (solid lines) from a medium with a relatively long scale length ($a = 2.25\lambda$) and weak Gaussian perturbations ($\varepsilon = 0.01$), compared against the unshifted, unperturbed seismograms for reference (dashed lines). Unlike some waveforms in our data set, these seismograms are very well behaved and we can obtain robust measurements of both traveltimes and amplitudes. However, since effects such as multipathed arrivals appearing within the cross-correlation window, $[t_1, t_2]$, corrupt our measurements, we chose to reject seismograms where these non-linearities interfere with the incident wavefront. We visually inspected each traveltime fit to ensure that we are indeed lining up our synthetic pulse, $p_{\rm syn}(t)$, with the first-arriving energy on $p_{\rm obs}(t)$. We did not, however, try to assess the quality of our amplitude measurements, trusting that if we obtained a reliable traveltime, the amplitude would also be fairly robust. The adoption of automatic rejection criteria, like those advocated by Ritsema & van Heijst (2002), does not significantly improve upon our visual criterion.

3 LINEAR THEORIES

3.1 Ray theory

The standard workhorse of global tomography has traditionally been ray theory. If contributions due to scattering are ignored, the first-order perturbation in traveltime due to a heterogeneous slowness distribution, $\delta\sigma(\mathbf{x})$, embedded in a homogeneous background slowness field σ, like one of our 'mantle in a box' models, may be expressed as an integral of slowness perturbations along the unperturbed ray:

$$\delta\hat{T} = \int_0^L \delta\sigma(\mathbf{x})\, d\ell. \tag{18}$$

In the above relation, L is the distance between the source and receiver and $d\ell$ is an increment of arclength along the straight source–receiver ray path. Note that we shall adhere to the notation of 'hatting' purely ray-theoretical quantities such as $\delta\hat{T}$ throughout the rest of this paper. Amplitude measurements, could, in principle, augment traveltimes in global tomographic inversions, provided the linear relationship between these data and slowness anomalies can be ascertained. Dahlen & Baig (2002) showed that this relationship can be expressed as a line integral

as well:

$$\delta(\ln \hat{A}) = \frac{-1}{2\sigma L}\int_0^L \ell(L-\ell)\nabla_\perp^2 \delta\sigma(\mathbf{x})\,d\ell + \frac{1}{\sigma L}\int_0^L \delta\sigma(\mathbf{x})\,d\ell - \frac{1}{2\sigma}[\delta\sigma(\mathbf{s}) + \delta\sigma(\mathbf{r})], \tag{19}$$

where $\nabla_\perp^2 = \nabla^2 - \partial_\ell^2$ is the Laplacian evaluated only in the normal plane of the ray, and the slowness perturbations at the source and receiver, $\delta\,\sigma(\mathbf{s})$ and $\delta\,\sigma(\mathbf{r})$ respectively, arise as endpoint contributions in the evaluation. The presence of the cross-path Laplacian $\nabla_\perp^2$ operating on the slowness perturbation field makes any ray-based amplitude inversions inherently more complicated than those based only on traveltime data.

3.2 Finite-frequency kernels

The previous section provided formulae that are only valid when the slowness anomalies are voluminous enough so that the wavelength of the probing wave can be neglected. In general, this requires *a priori* knowledge of the slowness perturbation field that one is trying to determine. To overcome this limitation, one can use the Born approximation to incorporate first-order scattering effects into the expressions for traveltimes and amplitudes (Dahlen *et al.* 2000; Dahlen & Baig 2002). One of the effects of extending the theory to finite frequencies is that contributions to the observables are no longer confined to the ray path and the correct linear relationships are necessarily written as volume integrals over all space:

$$\delta T = \iiint_{\Re^3} K_{\mathrm{T}}(\mathbf{x})\,\delta\sigma(\mathbf{x})\,d^3\mathbf{x} \tag{20}$$

and

$$\delta(\ln A) = \iiint_{\Re^3} K_{\mathrm{A}}(\mathbf{x})\,\delta\sigma(\mathbf{x})\,d^3\mathbf{x}, \tag{21}$$

where the kernel functions, K_{T} and K_{A} in practice confine most of the sensitivity to the vicinity of the ray path.

For a source at $\mathbf{s}$ and a receiver at $\mathbf{r}$, the traveltime kernel is given by Dahlen *et al.* (2000):

$$K_{\mathrm{T}}(\mathbf{x}) = \frac{\sigma}{2\pi}\frac{L}{L'L''}\frac{\int_0^\infty \omega^3|\dot{m}(\omega)|^2 \sin\left[\omega\sigma(L'+L''-L)\right]d\omega}{\int_0^\infty \omega^2|\dot{m}(\omega)|^2 d\omega}, \tag{22}$$

where $L = \|\mathbf{r}-\mathbf{s}\|$, $L' = \|\mathbf{x}-\mathbf{s}\|$, and $L'' = \|\,\mathbf{r}-\mathbf{x}\|$. The source-time function, $m(t)$, appears in this expression through the power spectrum of its derivative, $|\dot{m}(\omega)|^2$, which for our numerical experiments is

$$|\dot{m}(\omega)|^2 = \frac{(\omega\sigma\lambda)^2}{2\pi}\exp\left(-\frac{\omega^2\sigma^2\lambda^2}{4\pi^2}\right), \tag{23}$$

and acts to band-limit the kernel. For the same geometry, the amplitude kernel is (Dahlen & Baig 2002)

$$K_{\mathrm{A}}(\mathbf{x}) = \frac{\sigma}{2\pi}\frac{L}{L'L''}\frac{\int_0^\infty \omega^2|\dot{m}(\omega)|^2 \cos[\omega\sigma(L'+L''-L)]\,d\omega}{\int_0^\infty |\dot{m}(\omega)|^2\,d\omega}, \tag{24}$$

differing only from the traveltime kernel via the powers of ω in the numerator and denominator and the sin $\rightarrow$ cos substitution in the path-length detour term. These seemingly minor differences manifest themselves in the difference between the structure of the kernels along the ray. Also, note that our convenient choice of $|\dot{m}(\omega)|^2$ allows the integrals in eqs (22) and (24) to be analytically evaluated, such that

$$K_{\mathrm{T}}(\mathbf{x}) = \frac{4}{3\lambda}\frac{L}{L'L''}\exp\left[-\frac{\pi^2}{\lambda^2}(L'+L''-L)^2\right] He_5\left[\frac{\pi}{\lambda}(L'+L''-L)\right] \tag{25}$$

and

$$K_{\mathrm{A}}(\mathbf{x}) = \frac{4\pi}{\sigma\lambda^2}\frac{L}{L'L''}\exp\left[-\frac{\pi^2}{\lambda^2}(L'+L''-L)^2\right] He_4\left[\frac{\pi}{\lambda}(L'+L''-L)\right], \tag{26}$$

where $He_n(x)$ is the nth order Hermite polynomial.

Fig. 3 compares both of these kernels: the traveltime kernel has absolutely no sensitivity along the ray, whereas the amplitude kernel is maximally sensitive along the ray path in a cross-section in the normal plane of the ray. Physically, this difference is a manifestation of the single scattering (Born) approximation: a scatterer on the ray will generate a scattered wave arriving in phase with the ballistic wavefront, such that it cannot shift the waveform, but, by the same token, it maximizes the constructive or destructive interference. No traveltime anomaly would be observed in this approximation for this hypothetical scatterer, although an amplitude anomaly would be present. In this way, traveltime and amplitude data provide complementary information on the distribution of scatterers.

Finally, we should note that these kernels exhibit the expected behaviour in the high-frequency limit. For all background media, the traveltime kernel reduces to the 1-D ray-theoretical integral of the slowness perturbations along the ray (Dahlen *et al.* 2000). It has not previously been verified that the amplitude kernels universally converge to the general equivalent of eq. (19), but in the case of a homogeneous background medium this is indeed the case, as shown in Appendix A.

4 SCATTERPLOTS

Baig *et al.* (2003) advocated that scatterplots of theoretical versus measured traveltimes provide an indication of the efficacy of a given theory in approximating these data. Fig. 4 shows a collection of four scatterplots from two realizations, one Gaussian and the other exponential, with

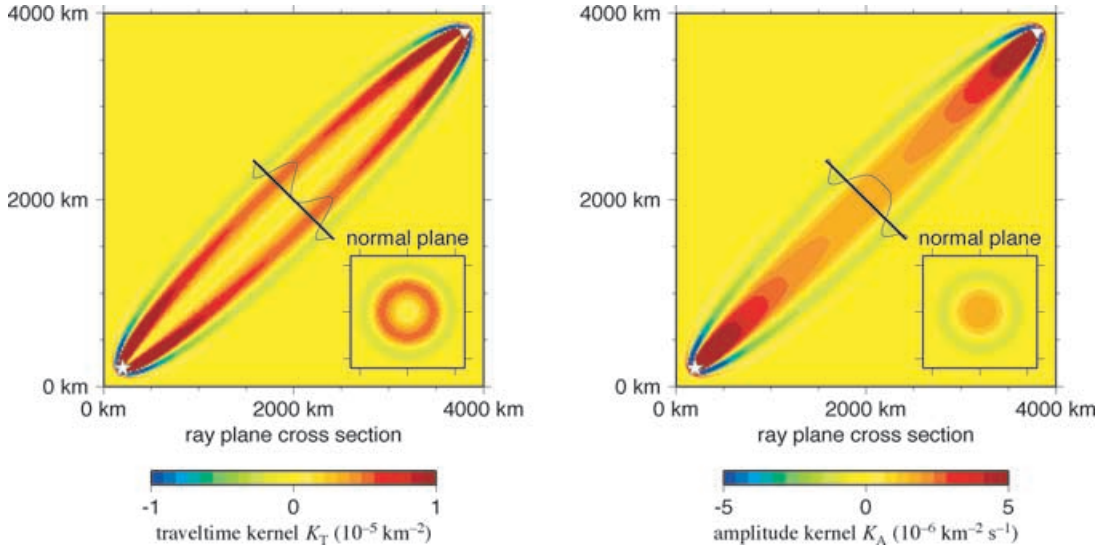

Figure 3. A traveltime kernel, K_T (left), and an amplitude kernel, K_A (right), for a background homogeneous medium. The source, **s**, and receiver, **r**, are respectively denoted by the star and the triangle in each plot. The insets in the lower right corners show cross-sections through each kernel in the normal plane of the ray. The traveltime kernel, K_T, is zero along the background ray, whereas the amplitude kernel, K_A, is maximal on this line. The variation of the kernel across the ray is also shown by the solid curve on each plot. For both kernels, the wavelength, λ, is 200 km.

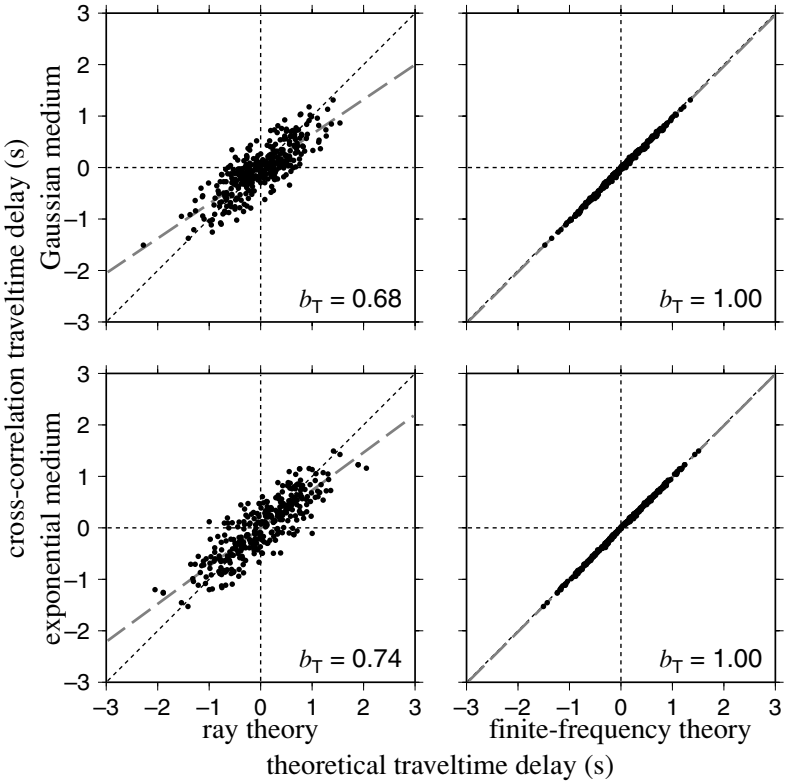

Figure 4. Four scatterplots of either ray-theoretical (left) or finite-frequency-theoretical (right) traveltimes against measured traveltimes for the same source–receiver pairs. The comparisons are performed in two different media with $\varepsilon = 0.01$ and $a = 0.75\lambda$, one Gaussian (top) and the other exponential (bottom), for a receiver array situated at a distance $L = 7.5\lambda$ from the source. The best-fitting line (grey, dashed) is drawn through the scatter in each subplot, and its slope, b_T, is shown in the lower left corner.

the same medium characteristics, $\varepsilon = 0.01$ and $a = 0.75\lambda$, and at the same propagation distance, $L = 7.5\lambda$. For each medium, the measurements are compared with both linearized ray theory and finite-frequency Born theory delays. Furthermore, the best-fitting line through the scatter is drawn in grey dashes. Looking at these two particular media, finite-frequency theory seems to model the observations well. Ray theory is inferior in these regimes, as shown by both the increase in the scatter around the line of best fit and the tendency to over-predict the magnitude of the delays. In general, when ray theory breaks down, the slope through the scatter tends to decrease from the ideal value of unity. This under-prediction is a manifestation of wavefront healing (Nolet & Dahlen 2000; Hung *et al.* 2001; Baig *et al.* 2003).

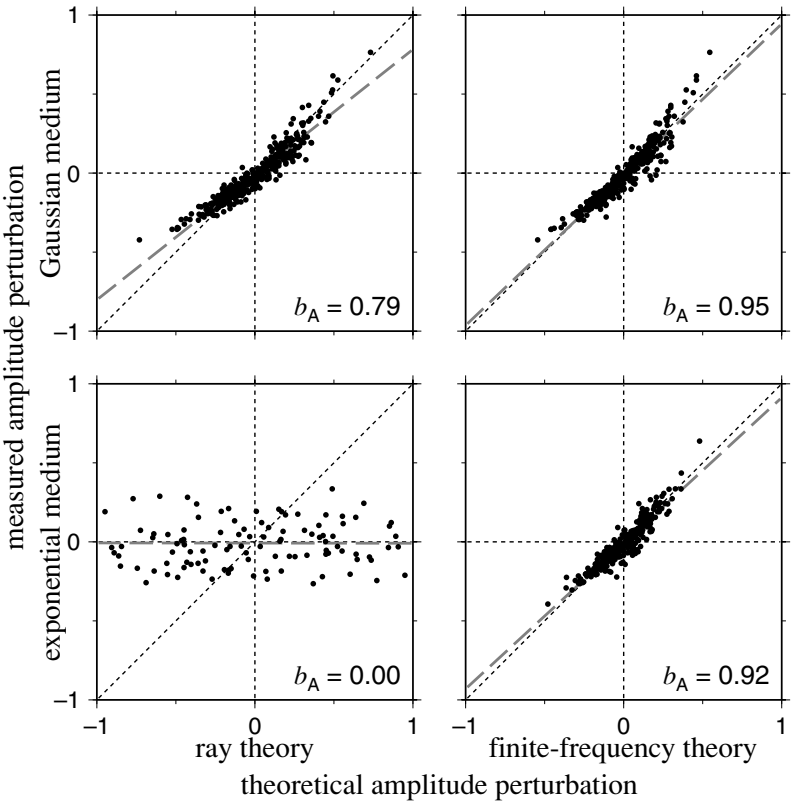

Figure 5. Scatterplots, as in Fig. 4, but for amplitude instead of traveltime data. Each scatterplot compares either ray theory or finite-frequency theory with the measured amplitude anomalies in either Gaussian or exponential media at receivers 22.5λ from the source. The heterogeneity spectrum in both media has $a = 2.25\lambda$ and $\varepsilon = 0.01$. Finally, the grey dashed line represents the line of best fit through the scatter, and the slope of this line, b_A, is given in the lower left corner of each subplot.

Scatterplots of theoretical amplitudes against measurements also prove very instructive. Fig. 5 shows examples of these scatterplots, for ray theory and finite-frequency kernels in both Gaussian and exponential media. In these plots, the regimes are defined by $\varepsilon = 0.01$, $a = 2.25\lambda$, and $L = 30\lambda$. For the Gaussian media, ray theory does a little worse than finite-frequency theory at predicting the amplitudes. The scatter in both of the Gaussian plots has a clearly arcuate character, perhaps indicating that amplitudes are inherently more non-linear than traveltimes. However, the grey, dashed, best-fitting line through the scatter is much shallower for the ray-theoretical plot than its finite-frequency counterpart, reflecting the overestimation of the variance by assuming the wavelength is negligible. The scatter in the finite-frequency plot in the exponential medium seems to also have an arcuate character, but the agreement between observation and prediction is far superior to that in the corresponding ray-theory plot, where there is absolutely no correlation between the theoretical and measured amplitudes. In fact, the slope of the best-fitting line in this case is zero, within error. This zero slope is a consequence of the fact that the ray-theoretical variance of amplitudes is divergent in an exponential medium (see Section 5.2.1). The term involving $\nabla_\perp^2$ in the ray-theoretical amplitude expression (19) acts to non-physically amplify the effect of the smallest scale anomalies in the calculation, causing the lack of correspondence between theory and measurement in this case

5 WAVEFIELD STATISTICS

Although Baig *et al.* (2003) used the slopes of their scatterplots as a diagnostic for wavefront healing, we shall first derive finite-frequency expressions for traveltime and amplitude variances, as these quantities appear more ubiquitously in the literature (Chernov 1960; Gudmundsson *et al.* 1990; Boyse & Keller 1995; Iooss *et al.* 2000). In fact, Gudmundsson *et al.* (1990) used the ray-theoretical expressions for traveltime variance in Gaussian media to try to constrain the statistical character of heterogeneity in the mantle. After these variances have been derived, we will derive finite-frequency expressions for traveltime and amplitude scatterplot slopes.

5.1 Traveltime variances

The ray-theoretical expressions for traveltime variance in both Gaussian and exponential media are given by several authors (Chernov 1960; Boyse & Keller 1995; Iooss *et al.* 2000). In Gaussian media, the high-frequency variance, $\langle \delta \hat{T}^2 \rangle$, is

$$\langle \delta \hat{T}^2 \rangle_g = \sqrt{\pi} \langle \delta \sigma^2 \rangle a L . \tag{27}$$

In an exponential medium, only the pre-multiplicative constant is different:

$$\langle \delta \hat{T}^2 \rangle_{\mathrm{e}} = 2\langle \delta\sigma^2 \rangle a L. \tag{28}$$

However, traveltimes, to first order, obey eq. (20) rather than eq. (18), and so we would expect deviations from ray theory whenever the scale of the heterogeneity becomes fine enough. The variance for finite-frequency traveltimes is given by

$$\begin{aligned}\langle \delta T^2 \rangle &= \iiint_{\Re^3} \iiint_{\Re^3} K(\mathbf{x}) K(\mathbf{x}+\mathbf{h}) \langle \delta\sigma(\mathbf{x}) \delta\sigma(\mathbf{x}+\mathbf{h}) \rangle d^3\mathbf{x} d^3\mathbf{h} \\ &= \langle \delta\sigma^2 \rangle \iiint_{\Re^3} C_{\mathrm{T}}(\mathbf{h}) N(\|\mathbf{h}\|)\, d^3\mathbf{h},\end{aligned} \tag{29}$$

where $C_{\mathrm{T}}(\mathbf{h})$ is the 3-D autocorrelation of the kernel:

$$C_{\mathrm{T}}(\mathbf{h}) = \iiint_{\Re^3} K_{\mathrm{T}}(\mathbf{x}) K_{\mathrm{T}}(\mathbf{x}+\mathbf{h})\, d^3\mathbf{x}. \tag{30}$$

At this step, we find it necessary to implement a paraxial approximation, effectively assuming that the kernels are long enough such that we may neglect variations in their structure along the ray path. The kernels in eq. (30) then become

$$K_{\mathrm{T}}(\mathbf{x}) = \frac{\sigma\Gamma}{2\pi} \frac{\int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin\left[\frac{1}{2}\omega\sigma\Gamma q^2\right] d\omega}{\int_0^\infty \omega^2 |\dot{m}(\omega)|^2 d\omega}, \tag{31}$$

$$K_{\mathrm{T}}(\mathbf{x}+\mathbf{h}) = \frac{\sigma\Gamma}{2\pi} \frac{\int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin\left[\frac{1}{2}\omega\sigma\Gamma(q^2+h^2+2qh\cos\psi)\right] d\omega}{\int_0^\infty \omega^2 |\dot{m}(\omega)|^2 d\omega}, \tag{32}$$

where $\Gamma = L/[\ell(L-\ell)]$; q is length of the component of $\mathbf{x}-\mathbf{s}$ perpendicular to the ray; h is the projection of $\mathbf{h}$ perpendicular to the ray in the the plane containing $\mathbf{s}$, $\mathbf{r}$, and $\mathbf{x}$; ψ measures the angle between $\mathbf{h}$ and this plane; and h_3 is the ray-parallel projection of $\mathbf{h}$ (see Fig. 6). We are now able to deal with some of the integrals in eq. (30) by first interchanging the frequency integrals in the kernels with the spatial integrals, and then evaluating the $d\psi$ integral before the dq radial integral, leaving the $d\ell$ integral along the ray. This yields

$$C_{\mathrm{T}}(\mathbf{h}) = C_{\mathrm{T}}^{+}(h) + C_{\mathrm{T}}^{-}(h), \tag{33}$$

where

$$C_{\mathrm{T}}^{\pm}(h) = \frac{\sigma}{4\pi R_{\mathrm{T}}^2} \int_0^\infty \int_0^\infty \frac{\omega^3 \omega'^3 |\dot{m}(\omega)|^2 |\dot{m}(\omega')|^2}{\omega \pm \omega'} \int_0^L \Gamma \sin\left[\frac{1}{2}\sigma\Gamma\left(\frac{\omega\omega'}{\omega\pm\omega'}\right)h^2\right] d\ell\, d\omega\, d\omega'. \tag{34}$$

The quantity

$$R_{\mathrm{T}} = \int_0^\infty \omega^2 |\dot{m}(\omega)|^2\, d\omega \tag{35}$$

acts as a normalization for the outer double integral in eq. (34).

5.1.1 Gaussian media

Once we have the autocorrelation of the kernel, we can derive the traveltime variance by defining

$$\langle \delta T^2 \rangle^{\pm} = \langle \delta\sigma^2 \rangle \iiint_{\Re^3} C_{\mathrm{T}}^{\pm}(h) N(\|\mathbf{h}\|)\, d^3\mathbf{h}. \tag{36}$$

First, consider Gaussian media:

$$\begin{aligned}\langle \delta T^2 \rangle_{\mathrm{g}}^{\pm} &= \langle \delta\sigma^2 \rangle \iiint_{\Re^3} C_{\mathrm{T}}^{\pm}(h) \mathrm{e}^{-\|\mathbf{h}\|^2/a^2} d^3\mathbf{h} \\ &= 2\pi\sqrt{\pi} a \langle \delta\sigma^2 \rangle \int_0^\infty C_{\mathrm{T}}^{\pm}(h) \mathrm{e}^{-h^2/a^2}\, h\, dh.\end{aligned} \tag{37}$$

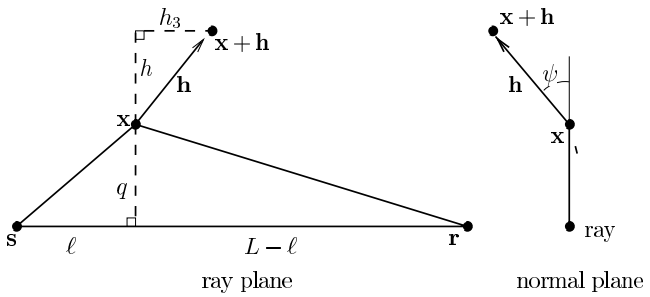

Figure 6. These cross-sections, in the ray plane (left) and normal plane (right), describe the geometry used in the derivation of the autocorrelation kernel. The source, receiver, and scatterer are situated at $\mathbf{s}$, $\mathbf{r}$, and $\mathbf{x}$ respectively. The autocorrelation is a function of the spatial 'lag', $\mathbf{h} = (h\sin\psi, h\cos\psi, h_3)$.

The radial *dh* integral, transverse to the ray, can be analytically evaluated, yielding the general expression for $\langle \delta T^2 \rangle_{\rm g}^{\pm}$ in a Gaussian medium:

$$\langle \delta T^2 \rangle_{\rm g}^{\pm} = \frac{\sqrt{\pi} a L \langle \delta \sigma^2 \rangle}{2 R_{\rm T}^2} \int_0^\infty \int_0^\infty \int_0^1 \frac{\omega^2 \omega'^2 |\dot{m}(\omega)|^2 |\dot{m}(\omega')|^2}{1 + \beta^{\pm} x^2 (1-x)^2} \, dx \, d\omega \, d\omega', \tag{38}$$

where

$$\beta^{\pm} = \frac{4L^2}{\sigma^2 a^4} \left(\frac{1}{\omega} \pm \frac{1}{\omega'} \right)^2. \tag{39}$$

As the bandwidth of $\dot{m}(\omega)$ becomes sufficiently high frequency, the inner x integral tends to unity and the value of the traveltime variance tends to the ray-theoretical value. To quantify how this transition occurs we need to specify the source-time function, so we will use the one defined for our pseudospectral runs in eq. (12). In this case, eq. (38) reduces to

$$\langle \delta T^2 \rangle_{\rm g} = \sqrt{\pi} a L \langle \delta \sigma^2 \rangle \int_0^\infty \int_0^\infty \frac{u^4 {\rm e}^{-u^2/4\pi^2} u'^4 {\rm e}^{-u'^2/4\pi^2}}{(\int_0^\infty \tilde{u}^4 {\rm e}^{-\tilde{u}^2/4\pi^2} d\tilde{u})^2} \frac{F^+(u,u';D) + F^-(u,u';D)}{2} \, du \, du', \tag{40}$$

where

$$D = \frac{a}{\sqrt{\lambda L}} \tag{41}$$

is the so-called 'doughnut-hole parameter' (Baig *et al.* 2003), or the ratio of the anomaly scale length to the width of the first Fresnel zone, and where

$$F^{\pm}(u,u';D) = \int_0^1 \frac{dx}{1 + 4D^{-4} \left(\frac{1}{u} \pm \frac{1}{u'} \right)^2 x^2 (1-x)^2}. \tag{42}$$

Hence, comparing eqs (40) and (27), the full expression for the variance is of the form of the ray-theoretical expression multiplied by a function solely of the doughnut-hole parameter, $\Theta_{\rm g}(D)$:

$$\langle \delta T^2 \rangle_{\rm g} = \langle \delta \hat{T}^2 \rangle_{\rm g} \Theta_{\rm g}(D). \tag{43}$$

This variance, shown in Fig. 7, is mildly reminiscent of an error function in that it begins at zero at $D = 0$ and asymptotically approaches unity as D increases. Unlike the error function though, this function behaves non-linearly in D near the origin. In fact, this work is replete with pseudo-error functions like $\Theta_{\rm g}(D)$, and they are summarized in Table 1.

5.1.2 Exponential media

The derivation for the traveltime variance in an exponential medium is slightly more complicated, and this complexity necessitates that we go into more detail in reproducing our derivation here. Ultimately, we obtain an expression similar to eq. (43) consisting of ray theory multiplied by a function of D. The equivalent of eq. (37) is

$$\langle \delta T^2 \rangle_{\rm e}^{\pm} = \langle \delta \sigma^2 \rangle \iiint_{\Re^3} C_{\rm T}^{\pm}(h) {\rm e}^{-\sqrt{h^2 + h_3^2}/a} d^3\mathbf{h}. \tag{44}$$

Focusing on the along-ray, dh_3, integral in eq. (44):

$$\int_{-\infty}^{\infty} \exp\left(-\sqrt{h^2 + h_3^2}/a \right) dh_3 = 2h K_1(h/a), \tag{45}$$

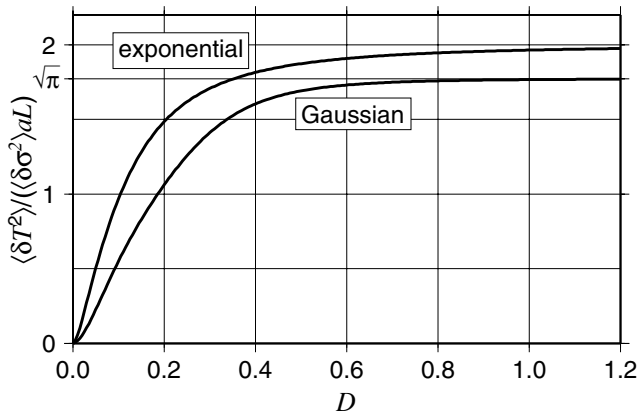

Figure 7. Comparison of traveltime variances for Gaussian and exponential media. The plotted variances have been normalized by $\varepsilon^2 \sigma^2 a L$ to isolate the dependence on D. For large values of D, these curves tend asymptotically to their ray-theoretical values of $\sqrt{\pi}$ and *2* for Gaussian and exponential media, respectively.

Table 1. Each of the variances derived in this paper can be decomposed into a frequency-independent expression (in all cases except amplitude variances in exponential media, this is the ray-theory expression) multiplied by a function of the doughnut-hole parameter, D, which describes the frequency dependence. The word 'slopes' in the last four rows of this table refers to the expressions for the ensemble slopes of ray theory versus measurement scatterplots, all of which are only functions of D. We also supply the number of the equation that defines all of these quantities as well as the number of the figure in which they are plotted.

Function	Frequency-independent expression	Finite-frequency dependence	eq.	Fig.
Gaussian traveltime variance	$\sqrt{\pi}aL\varepsilon^2\sigma^2$	$\Theta_{\rm g}(D)$	(40)	7
Exponential traveltime variance	$2aL\varepsilon^2\sigma^2$	$\Theta_{\rm e}(D)$	(51)	7
Gaussian amplitude variance	$\frac{4\sqrt{\pi}L^3}{15a^3}\varepsilon^2$	$\Xi_{\rm g}(D)$	(66)	8
Exponential amplitude variance	$\frac{L^3}{2a^3}\varepsilon^2$	$\Xi_{\rm e}(D)$	(71)	8
Gaussian traveltime slopes	–	$\langle b_{\rm T}\rangle_{\rm g}(D)$	(81)	9
Exponential traveltime slopes	–	$\langle b_{\rm T}\rangle_{\rm e}(D)$	(85)	9
Gaussian amplitude slopes	–	$\langle b_{\rm A}\rangle_{\rm g}(D)$	(90)	11
Exponential amplitude slopes	–	$\langle b_{\rm A}\rangle_{\rm e}(D)$	(91)	11

where K_1 is a modified Bessel function of order 1 (not to be confused with our kernels). To evaluate the *dh* integral, we need to use another representation of K_1 (Gradshteyn & Ryzhik 2000, eq. 8.432.6):

$$K_1(h/a) = \frac{1}{4}\frac{h}{a}\int_0^\infty \frac{\exp(-t - h^2/4ta^2)}{t^2}dt. \tag{46}$$

Upon evaluating the *dh* integral, followed by the *dt* integral, we arrive at

$$\langle \delta T^2\rangle_{\rm e}^{\pm} = \frac{\langle\delta\sigma^2\rangle a}{R_{\rm T}^2}\int_0^\infty\int_0^\infty \omega^2\omega'^2|\dot m(\omega)|^2|\dot m(\omega')|^2\int_0^L H(\gamma^\pm)\,d\ell\,d\omega\,d\omega', \tag{47}$$

where

$$H(\gamma^\pm) = 1 - \gamma^\pm[\mathrm{ci}(\gamma^\pm)\sin(\gamma^\pm) - \mathrm{si}(\gamma^\pm)\cos(\gamma^\pm)] \tag{48}$$

and

$$\gamma^\pm = \frac{\omega \pm \omega'}{2\omega\omega'\sigma\Gamma a^2}. \tag{49}$$

The functions si(x) and ci(x) are the sine and cosine integrals, respectively (Abramowitz & Stegun 1965, section 5.2):

$$\mathrm{si}(x) = \int_0^x \frac{\sin x'}{x'}dx' - \frac{\pi}{2}, \qquad \mathrm{ci}(x) = \int_0^x \frac{\cos x'}{x'}dx'. \tag{50}$$

As we did in the case of Gaussian media, we substitute our source-time function, eq. (12), into the expression for the traveltime variance:

$$\langle \delta T^2\rangle_{\rm e} = 2aL\langle\delta\sigma^2\rangle\int_0^\infty\int_0^\infty \frac{u^4{\rm e}^{-u^2/4\pi^2}u'^4{\rm e}^{-u'^2/4\pi^2}}{\left(\int_0^\infty \tilde u^4{\rm e}^{-\tilde u^2/4\pi^2}d\tilde u\right)^2}\,\frac{H^+(u,u';D) + H^-(u,u';D)}{2}\,du\,du', \tag{51}$$

$$H^\pm(u,u',D) = 1 - \int_0^1 \gamma^\pm(x)\left[\mathrm{ci}(\gamma^\pm(x))\sin\left(\gamma^\pm(x)\right) - \mathrm{si}\left(\gamma^\pm(x)\right)\cos\left(\gamma^\pm(x)\right)\right]dx, \tag{52}$$

and we have recast $\gamma^\pm$ in terms of u,u', and D:

$$\gamma^\pm(x) = \frac{x(1-x)|u \pm u'|}{2uu'D^2}. \tag{53}$$

So, again, we have an expression for the traveltime variance that increases from zero to the ray-theoretical value as D increases:

$$\langle \delta T^2\rangle_{\rm e} = \langle\delta\hat T^2\rangle_{\rm e}\Theta_{\rm e}(D), \tag{54}$$

where $\Theta_{\rm e}(D)$ is a function that increases from zero at $D = 0$ to one as $D \to \infty$. This variance, as a function of the doughnut-hole parameter, D, is plotted in Fig. 7.

5.2 Amplitude variances

5.2.1 Ray theory

In the ray-theoretical limit, amplitudes obey eq. (19), but, for the purposes of determining the ray-theory variance, we only need to analyse the term involving the $\nabla_\perp^2$ operator, since this operator will make this term dominant. Boyse & Keller (1995) calculated the variance of amplitudes in this high-frequency limit, but we shall review their derivation. Using the dominant contribution to the amplitude perturbation for $L \gg a$,

$$\delta(\ln\hat A) \approx -\frac{1}{2\sigma L}\int_0^L \nabla_\perp^2\delta\sigma(\mathbf{x})\,d\ell, \tag{55}$$

we can construct the variance as

$$\frac{\langle \delta \hat{A}^2 \rangle}{A^2} = \frac{1}{4\sigma^2 L^2} \int_0^L \int_0^L \ell(L-\ell)\ell'(L-\ell')\nabla_\perp^2 \nabla_\perp'^2 \langle \delta\sigma(\mathbf{x})\delta\sigma(\mathbf{x}')\rangle \, d\ell \, d\ell'. \tag{56}$$

The first step in evaluating this quantity is determining the effect of the $\nabla_\perp^2 \nabla_\perp'^2$ operator on the slowness field along the ray:

$$\nabla_\perp^2 \nabla_\perp'^2 \langle \delta\sigma(\mathbf{x})\delta\sigma(\mathbf{x}')\rangle = 8\langle \delta\sigma^2\rangle [s^{-2}\ddot{N}(s) - s^{-3}\dot{N}(s)], \tag{57}$$

where $s = |\ell - \ell'|$, and the dots on $\dot{N}$ and $\ddot{N}$ denote differentiation with respect to s. To evaluate the double integral in eq. (56), we transform to difference and centre-of-mass coordinates to get the following result:

$$\frac{\langle \delta \hat{A}^2 \rangle}{A^2} = \frac{2}{15} L^3 \varepsilon^2 \int_0^\infty \left[s^{-2}\ddot{N}(s) - s^{-3}\dot{N}(s) \right] ds. \tag{58}$$

For a Gaussian medium, this expression becomes

$$\frac{\langle \delta \hat{A}^2 \rangle_{\mathrm{g}}}{A^2} = \frac{4\sqrt{\pi}}{15} \frac{L^3}{a^3} \varepsilon^2, \tag{59}$$

but, for an exponential medium, the integral in eq (58) diverges such that $\langle \delta \hat{A}^2 \rangle_{\mathrm{e}} / A^2 \to \infty$. Finally, we should note that our expression for the amplitude variance is greater, by a factor of two, than that presented by Boyse & Keller (1995), who appear to have made an error in evaluating the effect of the $\nabla_\perp^2 \nabla_\perp'^2$ operator on $N(s)$.

5.2.2 Finite frequencies

To obtain expressions for the variance of finite-frequency amplitudes, we can simply replicate the procedure in Section 5.1, utilizing the same paraxial approximation, but replacing K_{T} with K_{A} in eq. (30). So, for amplitudes we have

$$\frac{\langle \delta A^2 \rangle}{A^2} = \varepsilon^2 \sigma^2 \iiint_{\Re^3} C_{\mathrm{A}}(\mathbf{h}) N(\|\mathbf{h}\|) \, d^3\mathbf{h}, \tag{60}$$

where $C_{\mathrm{A}}(\mathbf{h})$ is the autocorrelation of an amplitude kernel:

$$C_{\mathrm{A}}(\mathbf{h}) = \iiint_{\Re^3} K_A(\mathbf{x}) K_A(\mathbf{x}+\mathbf{h}) \, d^3\mathbf{x} = C_{\mathrm{A}}^+(h) - C_{\mathrm{A}}^-(h), \tag{61}$$

and $C_{\mathrm{A}}^{\pm}(h)$ is given as

$$C_{\mathrm{A}}^{\pm}(h) = \frac{\sigma}{4\pi R_{\mathrm{A}}^2} \int_0^\infty \int_0^\infty \frac{\omega^2 \omega'^2}{\omega \pm \omega'} |\dot{m}(\omega)|^2 |\dot{m}(\omega')|^2 \int_0^L \Gamma \sin\left[\frac{\sigma\Gamma}{2} \left(\frac{\omega\omega'}{\omega \pm \omega'} \right) h^2 \right] d\ell \, d\omega \, d\omega'. \tag{62}$$

We have a new normalization factor corresponding to amplitude kernels:

$$R_{\mathrm{A}} = \int_0^\infty |\dot{m}(\omega)|^2 \, d\omega. \tag{63}$$

5.2.3 Gaussian media

With the autocorrelation of the amplitude kernels derived, we now have all of the ingredients necessary to determine the amplitude variance in a given isotropic random medium. As with the traveltime variances, will shall first consider the case of a Gaussian medium. Inserting the autocorrelation function for this type of medium yields the following result:

$$\frac{\langle \delta A^2 \rangle_{\mathrm{g}}}{A^2} = \frac{8\sqrt{\pi}}{L^2 a^3 R_{\mathrm{A}}^2} \varepsilon^2 \int_0^\infty \int_0^\infty |\dot{m}(\omega)|^2 |\dot{m}(\omega')|^2 I(\omega, \omega') \, d\omega \, d\omega', \tag{64}$$

where

$$I(\omega, \omega') = \int_0^L \frac{\ell^2(L-\ell)^2}{\left[1 + \frac{4}{\sigma^2\Gamma^2 a^4}\left(\frac{1}{\omega} + \frac{1}{\omega'}\right)^2\right]\left[1 + \frac{4}{\sigma^2\Gamma^2 a^4}\left(\frac{1}{\omega} - \frac{1}{\omega'}\right)^2\right]} \, d\ell. \tag{65}$$

Substituting eq. (12) for $|\dot{m}(\omega)|$ gives us

$$\frac{\langle \delta A^2 \rangle_{\mathrm{g}}}{A^2} = \frac{4\sqrt{\pi}}{15} \frac{L^3}{a^3} \varepsilon^2 \int_0^\infty \int_0^\infty \frac{u^2 \mathrm{e}^{-u^2/4\pi^2} u'^2 \mathrm{e}^{-u'^2/4\pi^2}}{\left(\int_0^\infty \tilde{u}^2 \mathrm{e}^{-\tilde{u}^2/4\pi^2} \, d\tilde{u} \right)^2} \tilde{I}(u, u'; D) \, du \, du', \tag{66}$$

where

$$\tilde{I}(u, u'; D) = 30 \int_0^1 \frac{x^2(1-x)^2}{\left[1 + \frac{4}{D^4}\left(\frac{1}{u} + \frac{1}{u'}\right)^2 x^2(1-x)^2\right]\left[1 + \frac{4}{D^4}\left(\frac{1}{u} - \frac{1}{u'}\right)^2 x^2(1-x)^2\right]} dx. \tag{67}$$

Much like the finite-frequency expressions for traveltime variance, we have an expression of the form of the ray-theoretical value for amplitude variance multiplied by a function of the doughnut-hole parameter. By incorporating all of the D dependence into another one of these quasi-error

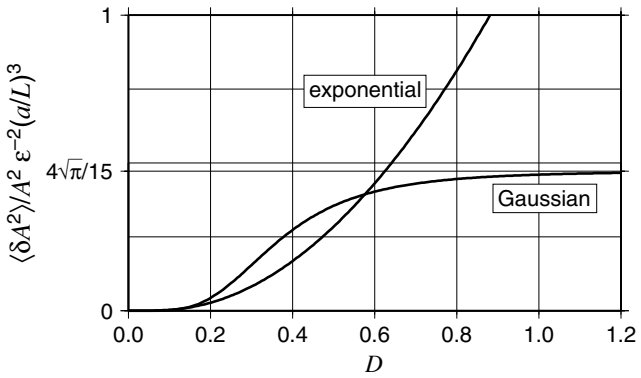

Figure 8. Curves of amplitude variance plotted versus the doughtnut-hole parameter, D, for both Gaussian and exponential media. These variances are normalized by $\varepsilon^2\,(L/a)^3$ to isolate the D dependence. For the Gaussian curve at high values of D, the curve tends asymptotically to the ray-theoretical value of $4\sqrt{\pi}/15$; in contrast, the exponential curve diverges in this regime as approximately $1.42D^2$.

functions, $\Xi_g(D)$, the variance of finite-frequency amplitudes can be written as

$$\frac{\langle \delta A^2 \rangle_{\rm g}}{A^2} = \frac{\langle \delta \hat{A}^2 \rangle_{\rm g}}{A^2}\, \Xi_{\rm g}(D). \tag{68}$$

This expression, plotted in Fig. 8, increases quadratically from zero at $D = 0$, but then tends asymptotically to unity as $D \to \infty$.

5.2.4 Exponential media

The equivalent of eq. (64) in an exponentially correlated medium is given as

$$\frac{\langle \delta A^2 \rangle_{\rm e}}{A^2} = \frac{\varepsilon^2 \sigma^2}{2\sigma a L R_{\rm A}^2} \int_0^\infty \int_0^\infty |\dot{m}(\omega)|^2 |\dot{m}(\omega')|^2 \int_0^L ([\omega + \omega']X^+ - |\omega - \omega'|X^-)\ell(L - \ell)\, d\ell\, d\omega\, d\omega', \tag{69}$$

where

$$X^\pm = {\rm ci}(\gamma^\pm)\sin(\gamma^\pm) - {\rm si}(\gamma^\pm)\cos(\gamma^\pm), \tag{70}$$

and $\gamma^\pm$is the same as in eq. (49). Again, to get this expression in the form of a function of the doughnut-hole parameter, we will insert our source-time function, eq. (12). This substitution yields the result

$$\frac{\langle \delta A^2 \rangle_{\rm e}}{A^2} = \frac{1}{2}\varepsilon^2 \frac{L^3}{a^3} \int_0^\infty \int_0^\infty \frac{u^2 {\rm e}^{-u^2/4\pi^2} u'^2 {\rm e}^{-u'^2/4\pi^2}}{\left(\int_0^\infty \tilde{u}^2 {\rm e}^{-\tilde{u}^2/4\pi^2} d\tilde{u}\right)^2} J(u, u'; D)\, du\, du', \tag{71}$$

where

$$J(u, u'; D) = D^2 \int_0^\infty x(1 - x)([u + u']X^+ - |u - u'|X^-) dx. \tag{72}$$

As the doughnut-hole parameter $D \to \infty$, $X^\pm \to \pi/2$, and the amplitude variance increases quadratically with D, reflecting the ray-theoretical divergence of this quantity. The divergence of this expression can be observed in Fig. 8. Although we cannot, in this case, write the amplitude variance as a finite ray-theoretical value times a function of D, we can still isolate the dependence on the doughnut-hole parameter into a function $\Xi_{\rm e}(D)$, and write the amplitude variance as

$$\frac{\langle \delta A^2 \rangle_{\rm e}}{A^2} = \frac{1}{2}\varepsilon^2 \frac{L^3}{a^3}\, \Xi_{\rm e}(D). \tag{73}$$

Removing the D^2 factor from $\Xi_{\rm e}(D)$ leaves us with yet another function that rises non-linearly from the origin to a finite value; in this case, $D^{-2}\Xi_{\rm e}(D) \to 1.42$ as $D \to \infty$.

5.3 Traveltime scatterplot slopes

The best-fitting slope through the ray-theoretical scatterplot can be calculated in a similar fashion to that of traveltime variance. The formula for the this slope, $b_{\rm T}$, for a single realization, is well-known (see, for example, Taylor 1982 section 8.2):

$$b_{\rm T} = \frac{n \sum_{j=1}^n (\delta T_j \delta \hat{T}_j) - \left(\sum_{j=1}^n \delta T_j\right)\left(\sum_{j=1}^n \delta \hat{T}_j\right)}{n \sum_{j=1}^n \left(\delta \hat{T}_j^2\right) - \left(\sum_{j=1}^n \delta \hat{T}_j\right)^2}, \tag{74}$$

where n is the number of fitted data points, and δT_j and $\delta \hat{T}_j$ are the measured and ray-theoretical traveltimes respectively. To extend this expression beyond a single realization of a random medium, we can use the ergodic assumption to replace the summations over data points

by ensemble averages. Upon performing this replacement, the expression (74) simplifies somewhat, inasmuch as

$$\langle \delta \hat{T} \rangle = 0, \tag{75}$$

leaving the ensemble slope, $\langle b_{\rm T} \rangle$, as the ratio of the covariance between ray theory and measured traveltimes over the ray-theoretical variance:

$$\langle b_{\rm T} \rangle = \frac{\langle \delta T \delta \hat{T} \rangle}{\langle \delta \hat{T}^2 \rangle}. \tag{76}$$

Expressions for $\langle \delta \hat{T}^2 \rangle$ have already been presented in eqs (27) and (28), so it remains to derive expressions for $\langle \delta T \delta \hat{T} \rangle$.

The covariance can be computed in a similar manner to the variance in Section 5.1, as the integral of the autocorrelation function of the medium multiplied by the cross-correlation of two kernels. In this case, however, one of these kernels will be the ray-theoretical kernel of a delta function along the background ray, while the other is our finite-frequency expression. Hence,

$$\langle \delta T \delta \hat{T} \rangle = \varepsilon^2 \sigma^2 \iiint_{\Re^3} B_{\rm T}(\mathbf{h}) N(\|\mathbf{h}\|) d^3\mathbf{h}, \tag{77}$$

with

$$\begin{aligned} B_{\rm T}(\mathbf{h}) &= \textstyle\int_0^L K_{\rm T}(\mathbf{x}+\mathbf{h})\, d\ell \\ &= \frac{\sigma}{2\pi R_{\rm T}} \int_0^L \Gamma \int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin\left(\frac{1}{2}\omega\sigma\Gamma h^2\right) d\omega\, d\ell, \end{aligned} \tag{78}$$

after applying the usual paraxial approximation to $K_{\rm T}(\mathbf{x}+\mathbf{h})$. Since eq. (78) is remarkably similar to, and even less complicated than, eq. (34) for the equivalent kernel autocorrelation function $C_{\rm T}(\mathbf{h})$ for the traveltime variance, the remaining analysis for the traveltime slope mirrors that for its variance.

5.3.1 Gaussian media

Inserting eq. (4) for the autocorrelation function in eq. (77) gives us the general result for the covariance between measured and ray-theoretical traveltimes in a Gaussian medium:

$$\langle \delta T \delta \hat{T} \rangle = \frac{\sqrt{\pi} a \varepsilon^2 \sigma^2}{R_{\rm T}} \int_0^\infty \omega^2 |\dot{m}(\omega)|^2 \int_0^L \left[1 + \frac{4}{\omega^2\sigma^2 a^4 \Gamma^2}\right]^{-1} d\ell\, d\omega. \tag{79}$$

Dividing out the ray-theoretical traveltime variance, which appears conveniently as a multiplicative factor, gives us the general expression for the scatterplot slope of a Gaussian medium:

$$\langle b_{\rm T} \rangle_{\rm g} = \frac{1}{L R_{\rm T}} \int_0^\infty \omega^2 |\dot{m}(\omega)|^2 \int_0^L \left[1 + \frac{4}{\omega^2\sigma^2 a^4 \Gamma^2}\right]^{-1} d\ell\, d\omega. \tag{80}$$

For comparison with our slopes measured from the traveltimes in our pseudospectral propagation runs, we insert eq. (12) for $\dot{m}(\omega)$ to obtain

$$\langle b_{\rm T} \rangle_g(D) = \int_0^\infty \frac{u^4 \exp(-u^2/4\pi^2)}{\int_0^\infty \tilde{u}^4 \exp(-\tilde{u}^2/4\pi^2)\, d\tilde{u}} \int_0^1 \left[1 + \frac{4x^2(1-x)^2}{D^4 u^2}\right]^{-1} dx\, du. \tag{81}$$

Like our expressions for variance, this expression resembles a pseudo-error function that rises non-linearly from the origin and tends asymptotically to unity as the doughnut-hole parameter increases.

5.3.2 Exponential media

To derive the expression for the slope of a ray-theoretical traveltime versus measurement scatterplot, it suffices to follow the analysis of the previous subsection for Gaussian media, appropriately replacing the autocorrelation function. In this way, we arrive at the following expression for the slope:

$$\langle b_{\rm T} \rangle_{\rm e} = \frac{1}{L R_{\rm T}} \int_0^\infty \omega^2 |\dot{m}(\omega)|^2 \int_0^L [1 - \gamma X(\gamma)]\, d\ell\, d\omega, \tag{82}$$

where

$$X(\gamma) = \mathrm{ci}(\gamma)\sin(\gamma) - \mathrm{si}(\gamma)\cos(\gamma) \tag{83}$$

and

$$\gamma = (2\omega\sigma\Gamma a^2)^{-1}. \tag{84}$$

Upon inserting eq. (12) for $\dot{m}(\omega)$ we get

$$\langle b_{\rm T} \rangle_e(D) = \int_0^\infty \frac{u^4 \exp(-u^2/4\pi^2)}{\int_0^\infty \tilde{u}^4 \exp(-\tilde{u}^2/4\pi^2)} \left[1 - \int_0^1 \gamma X(\gamma; D)\right] dx\, d\omega, \tag{85}$$

where we have recast the factor γ in eq. (84) as

$$\gamma = \frac{x(1-x)}{2D^2u}. \tag{86}$$

This expression for the ensemble slope for an exponential medium is yet another function of D which rises from zero at $D = 0$ and approaches unity as $D \to \infty$.

5.3.3 Measured slopes

The scatterplot slopes are quantities which, with our limited suite of numerical simulations, we can measure much more robustly than the ensemble average of the variance. Our dubious assessment of our ability to measure these variances is based on our feeling that we have undersampled the number of realizations required to get a true measure of the ensemble variance. To correct this problem would require several runs in different media with the same power spectra and, although we have been blessed with access to several powerful computer clusters, we feel the issue of getting a good numerical estimate of the ensemble variance is somewhat secondary to the point we wish to make on the effect of wavefront healing regarding this quantity.

Scatterplot slopes, on the other hand, can be measured fairly robustly by looking only at one medium realization, since the slopes themselves are inherently normalized by the ray-theoretical variances. For each regime, we actually have collected data from two realizations, considering one medium and its 'mirror image' where we simply reverse the sign of the perturbation field, $\delta\sigma \to -\delta\sigma$. The advantage of looking at one medium together with its mirror image, beyond doubling our data, is that we are able to do a much better job of satisfying condition (75).

In Fig. 9, we have plotted the measured slopes from scatterplots of ray theory versus measured traveltimes. The functions $\langle b_T \rangle_g(D)$ and $\langle b_T \rangle_e(D)$, given by eqs (81) and (85), are plotted for reference as a solid line on the appropriate subplots. In general, our formulae for the ensemble slopes agree quite well with our measurements, to within the ± 2 standard deviation error bars most of the time. The increase in the size of these error bars as ε increases is reflective of the fact that many more data were rejected in these regimes. Our finite-frequency scatterplot slopes are plotted in Fig. 10, together with the expected theoretical slope, $b = 1$. Like the ray-theoretical slopes, these slopes agree with the theoretical value of unity within two standard deviations for all but the most strongly heterogeneous media. The agreement of these slopes with our measurements gives us confidence that, not only are our expressions for the ensemble slope accurate, but that we are not algebraic klutzes, and have also derived the correct behaviour for the traveltime variance.

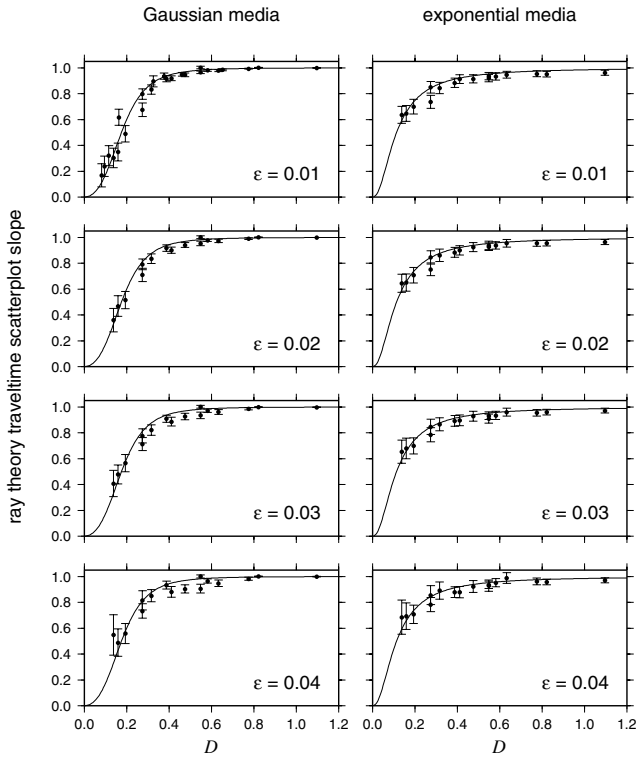

Figure 9. Slopes of all the ray theory versus measurement scatterplots in our study plotted versus the doughnut-hole parameter, D. The subplots separate the Gaussian media, on the left, from the exponential media, on the right. Furthermore, the vertical segregation is organized such that ε increases from the top to the bottom. On each plot, we have plotted the ensemble slopes, either $\langle b_T \rangle_g(D)$ or $\langle b_T \rangle_e(D)$, for reference as the solid line. Error bars represent two standard deviations.

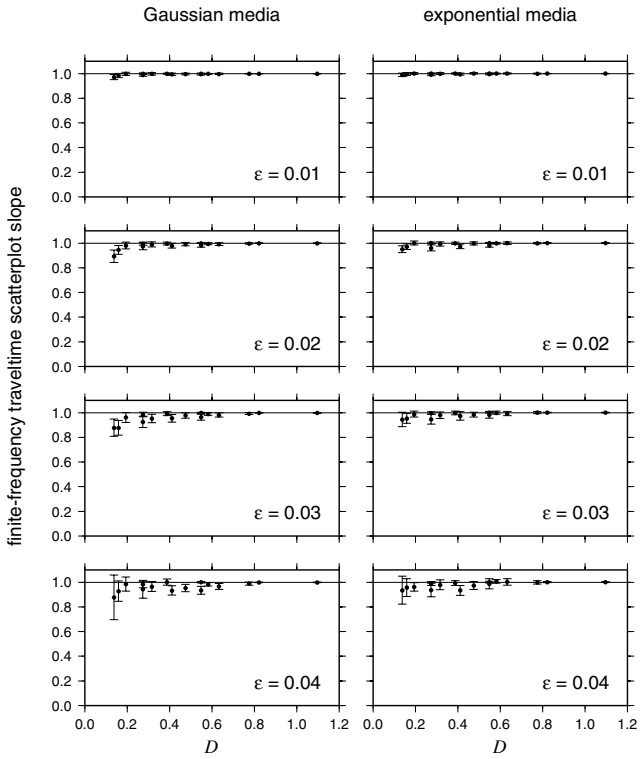

Figure 10. The above plot shows each finite-frequency versus measurement scatterplot in our study, organized into subplots in exactly the same fashion as in Fig. 9. Unlike in Fig. 9, the theoretical slope is unity, so we plot a solid line at slope = 1. Again, the uncertainty in the measured slopes is reflected by their two-standard-deviation error bars.

5.4 Amplitude scatterplot slopes

5.4.1 Theoretical scatterplot slopes

We can formulate expected values for the best-fitting slopes of the ray-theoretical versus measured amplitudes in an almost identical way to how we dealt with traveltime slopes. However, as the form of the ray-theoretical amplitude is dominated by a term with a $\nabla_\perp^2$ operator, evaluating the covariance in the numerator of the slope is a little more complicated. Replacing the ray-theoretical and finite-frequency traveltime kernels with their amplitude counterparts in eqs (77) and (78) gives us the following expressions for the covariance of measured amplitudes, δA, against ray-theoretical amplitudes, $\delta\hat{A}$:

$$\langle \delta(\ln A)\delta(\ln \hat{A})\rangle = \varepsilon^2\sigma^2 \iiint_{\Re^3} B_{\rm A}(\mathbf{h}) N(\|\mathbf{h}\|)\, d^3\mathbf{h}, \tag{87}$$

where

$$\begin{aligned} B_{\rm A}(\mathbf{h}) &= -\frac{1}{2\sigma L}\int_0^L \ell(L-\ell)\nabla_\perp^2 K_{\rm A}(\mathbf{x}+\mathbf{h})\, d\ell \\ &= \frac{\sigma}{4\pi R_{\rm A}}\int_0^L \Gamma \int_0^\infty \left[2\omega^3\Gamma|\dot{m}(\omega)|^2 \sin\left(\frac{1}{2}\omega\sigma\Gamma h^2\right) + h^2\omega^4\sigma\Gamma^2|\dot{m}(\omega)|^2\cos\left(\frac{1}{2}\omega\sigma\Gamma h^2\right)\right] d\omega\, d\ell. \end{aligned} \tag{88}$$

If we insert eq. (4) to evaluate the covariance for a Gaussian medium, and then divide out the ray-theoretical variance, we obtain the following expression for the ensemble slope, $\langle b_{\rm A}\rangle_{\rm g}$:

$$\langle b_{\rm A}\rangle_{\rm g} = \frac{1}{L^3 R_{\rm A}}\int_0^\infty |\dot{m}(\omega)|^2 \int_0^L \left[\Gamma + \frac{4}{\omega^2\sigma^2 a^4\Gamma}\right]^{-2} d\ell\, d\omega. \tag{89}$$

For our Gaussian source-time function (12), this equation becomes

$$\langle b_{\rm A}\rangle_g(D) = \int_0^\infty \frac{u^2\exp(-u^2/4\pi^2)}{\int_0^\infty \tilde{u}^2\exp(-\tilde{u}^2/4\pi^2)}\left(30\int_0^1 x^2(1-x)^2\left[1+\frac{4x^2(1-x)^2}{D^4u^2}\right]^{-2} dx\right) du. \tag{90}$$

Yet again, we have an expression, this time for the slope of an amplitude scatterplot, that increases from zero at $D = 0$ and tends asymptotically to unity as D increases. In exponential media, when one normalizes the covariance by the divergent ray-theoretical variance to obtain this slope, the result is zero:

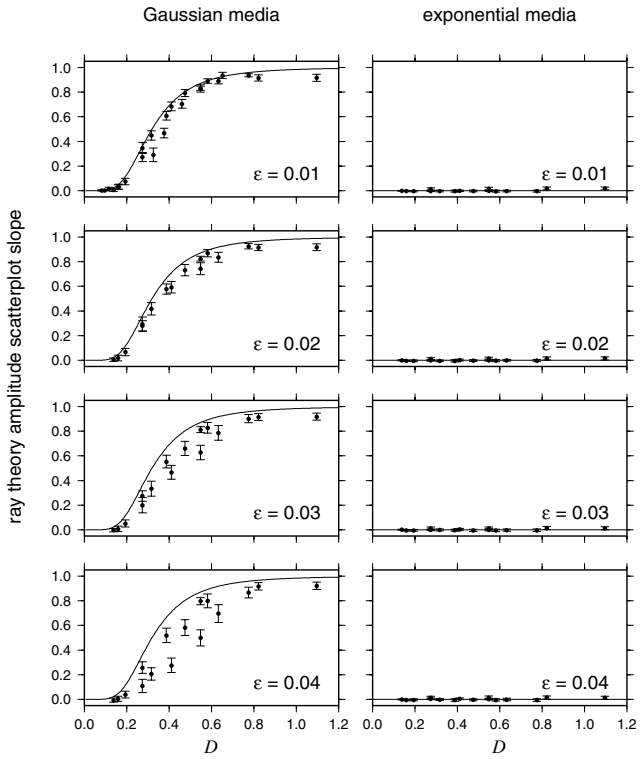

Figure 11. The slopes of each measurement versus ray theory amplitude scatterplot plotted versus D. The subplots are arranged as in Figs 9 and 10, with increasing values of ε from top to bottom, and with Gaussian media on the left and exponential media on the right. On each subplot, the ensemble theoretical average slopes, either $\langle b_{\mathrm{A}} \rangle_{\mathrm{g}}$ or $\langle b_{\mathrm{A}} \rangle_{\mathrm{e}} = 0$, are plotted as the solid line; the error bars represent two standard deviations.

$$\langle b_{\mathrm{A}} \rangle_{\mathrm{e}} = 0, \tag{91}$$

reflecting the fact that the high-frequency approximation exaggerates the effect of the small-scale structure of the medium.

5.4.2 Measured scatterplot slopes

In Gaussian media, the agreement of the measured scatterplot slopes with the theoretical expression (90), like in the case of traveltimes, can be used as a proxy for the accuracy of our curves of amplitude variance. Again, we do not believe we have sampled enough paths to get a true estimate of the ensemble variance, but the slopes behave more robustly than the variances. We cannot draw the same conclusions for measured slopes in the exponential media, since agreement with the theoretical value of zero, in this casan only be seen as an indictment of the inadequacy of ray theory in these models.

In Fig. 11 we have plotted the measured slopes for each scatterplot against D. In Gaussian media, the slopes for the weakly heterogeneous $\varepsilon = 0.01$ case seem to fit the ray-theoretical curve well. However, with increasing ε, our measured slopes start to fall well below the theoretical ensemble average. In the exponential media, the complete failure of ray theory is shown by all of the measured slopes being zero within error. The values of all of our finite-frequency scatterplot slopes, shown in Fig. 12, should not deviate significantly from unity. However, we see that this is not the case as, in both Gaussian and exponential media, these values tend to fall below one, especially as ε increases.

This discrepancy, we suspect, is the result of a non-linearity in the physics of scattering, coupled with a limitation introduced by our measurement procedure. By inspecting eq. (17), one can see that our method of measuring amplitudes restricts us from obtaining values less than -1. These cases would correspond to the observed pulse having reversed polarity with respect to the incident pulse, and the δT would align with a minimum of the cross-correlation function. However, our first-order perturbation theories cannot capture such restrictions, and over-predict the amount of defocusing experienced by the waveform. Because the failure of the Born approximation to conserve energy (Sato & Fehler 1997 section 4.2.1) non-linearly biases our measured amplitudes to lower than their predicted values, the data in scatterplots like those in Fig. 13 begin to collect above the $y = -1$ line, biasing the measured slopes to shallower values. We shall investigate the connection between non-linearities and these shallow slopes in greater detail in the next section.

6 VALIDITY OF LINEARITY

Several authors have placed bounds on the validity of linear ray theory, by determining the probability that a ray travelling a given distance in a given medium will encounter a caustic (Kulkarny & White 1982; White *et al.* 1988; Spetzler & Snieder 2001a). Using the approach of

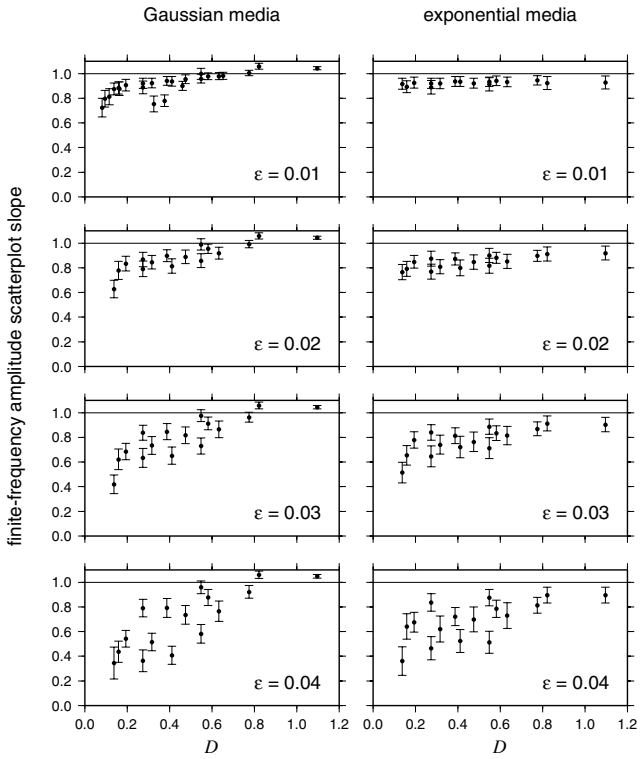

Figure 12. This plot is similar to Fig. 11, with the difference that measured amplitudes are being compared with their finite-frequency theoretically predicted values. As in Fig. 10, the solid line at slope = 1 in all plots indicates the theoretical expectation that scatter in plots like Fig. 5 should fall on the 45° line. Again, the error bars are two standard deviations.

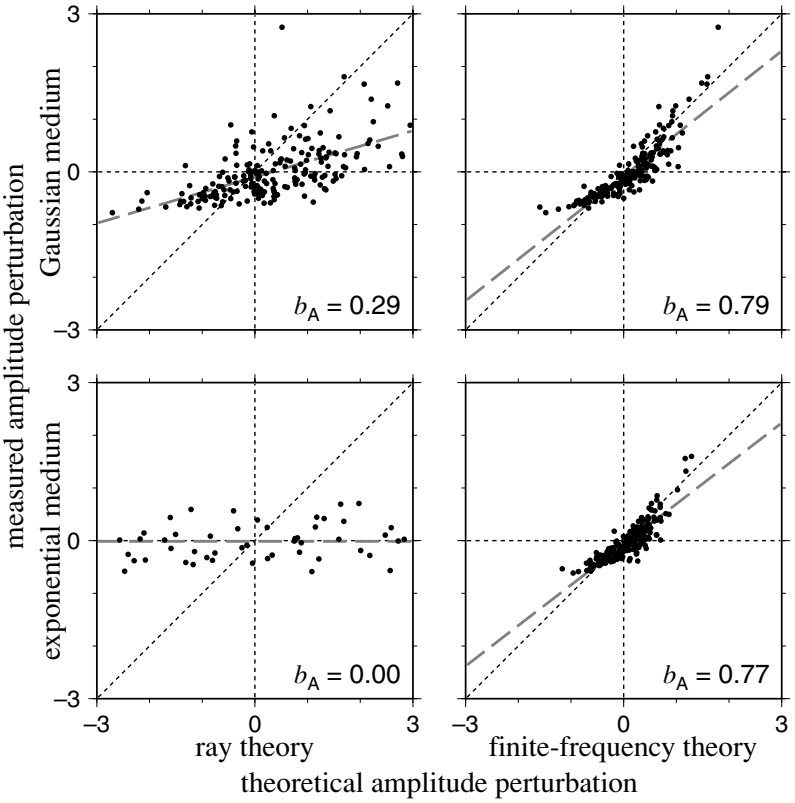

Figure 13. Collection of scatterplots for amplitudes, like the ones shown in Fig. 5, but in the less well-behaved regime of $\varepsilon = 0.02$, $a = 1.5\lambda$, and $L = 15\lambda$. The over-prediction of negative anomalies by both of our first-order theories biases the lines of best fit (grey, dashed) to shallower slopes than we would expect based on our theories (with the exception of ray theory in the exponential medium).

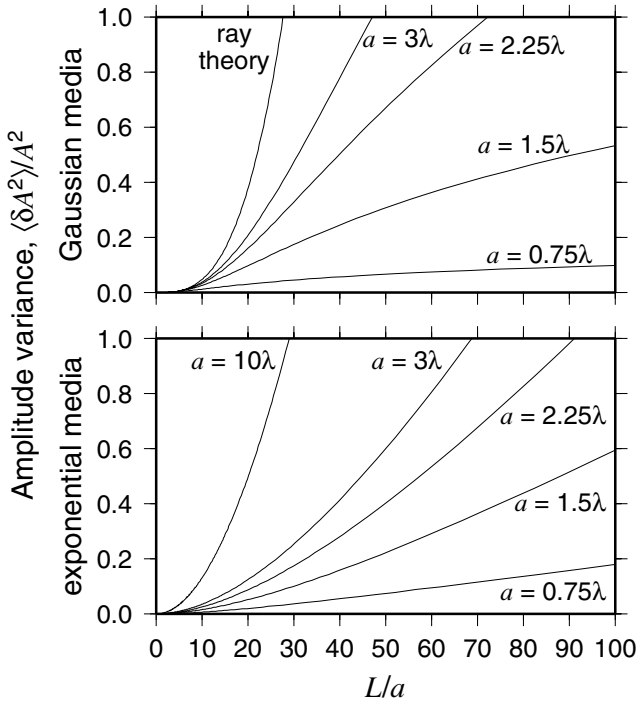

Figure 14. The amplitude variance is plotted against propagation distance (normalized by scale length) for several values of a/λ as indicated in both Gaussian (top) and exponential (bottom) media, with $\varepsilon = 0.01$. In addition, the ray-theoretical ($a/\lambda \to \infty$) amplitude variance is plotted for reference in the Gaussian media. The fact that these curves become steeper with increasing a/λ indicates that linearity is valid for small-scale media at farther propagation distances, if we refer to the criterion (97).

Spetzler & Snieder (2001a), we shall consider perturbed paraxial rays to determine where the first caustic is likely to occur. Perturbed paraxial rays in a homogeneous background medium satisfy (Dahlen & Baig 2002, eq. 147)

$$d(\delta \mathsf{Q})/d\ell = \sigma^{-1}\delta \mathsf{P} - \sigma^{-2}\delta\sigma \mathsf{I}, \qquad d(\delta \mathsf{P})/d\ell = \sigma^{-1}\ell\nabla_\perp\nabla_\perp\delta\sigma, \tag{92}$$

where $\delta\mathsf{P}$ and $\delta\mathsf{Q}$ are perturbations to P and Q, the 2×2 matrices that decompose the Ricatti version of the dynamic ray-tracing equations for the traveltime Hessian $\mathsf{M} = \mathsf{P} \cdot \mathsf{Q}^{-1}$ (Červený 1985). The 2×2 identity matrix is denoted as I. A caustic is signified by a singularity in M, i.e.

$$\det(\mathsf{Q} + \delta\mathsf{Q}) = 0. \tag{93}$$

Integrating eqs (92) subject to the initial conditions for a point, $\delta\mathsf{Q} = 0$ and $\delta\mathsf{P} = \mathsf{I}$, gives an expression for $\delta\mathsf{Q}$:

$$\delta\mathsf{Q}(L) = \sigma^{-2}\int_0^L \ell(L-\ell)\nabla_\perp\nabla_\perp\delta\sigma(\ell)d\ell - \sigma^{-2}\mathsf{I}\int_0^L \delta\sigma(\ell)\,d\ell. \tag{94}$$

With the corresponding expression for a point source in a homogeneous medium, $\mathsf{Q}(L) = \sigma^{-1}L\mathsf{I}$, the condition (93), to first order in $\delta\sigma$, is

$$L - \sigma^{-1}\int_0^L \delta\sigma(\ell)\,d\ell + \frac{1}{2\sigma}\int_0^L \ell(L-\ell)\nabla_\perp^2\delta\sigma(\ell)\,d\ell = 0, \tag{95}$$

or, recalling eq. (19) without the endpoint terms:

$$\delta(\ln \hat{A}) = 1. \tag{96}$$

This interpretation of caustic formation in terms of amplitudes allows us to use our formulae for $\langle \delta A^2 \rangle / A^2$ to infer the region of validity for linear finite-frequency theories. For ray theory in a Gaussian medium, this constraint gives us

$$\frac{\langle \delta \hat{A}^2 \rangle}{A^2} = \frac{4\sqrt{\pi}}{15}\left(\frac{L}{a}\right)^3 \varepsilon^2 \ll 1, \tag{97}$$

or $L/a \ll 1.28\,\varepsilon^{-2/3}$. This expression captures the fact, observed by Kulkarny & White (1982), White *et al.* (1988), and Spetzler & Snieder (2001a), that the relevant dimensionless number for the validity of linear ray theory is $\varepsilon^{2/3}(L/a)$. If we extend this amplitude variance criterion to finite frequencies using eqs (66) and (71), D also becomes a relevant parameter governing the accuracy of the linear approximation.

To observe how these finite-frequency considerations might influence the range of validity of linearity, we plot the theoretical amplitude variance for a few values of a/λ against L/a in Fig. 14, keeping $\varepsilon = 0.01$ constant. The ray-theoretical curve rises as a cubic function of L/a in the Gaussian medium, but the finite-frequency curves breaks away from this ray-theoretical curve, so that the amplitude variance eventually increases linearly with L/a. This behaviour would indicate that the linear approximation is valid over longer propagation distances for finite-frequency waves than in ray theory. In the exponential medium, there is no ray-theoretical curve to refer to, but, for the finite-frequency curves,

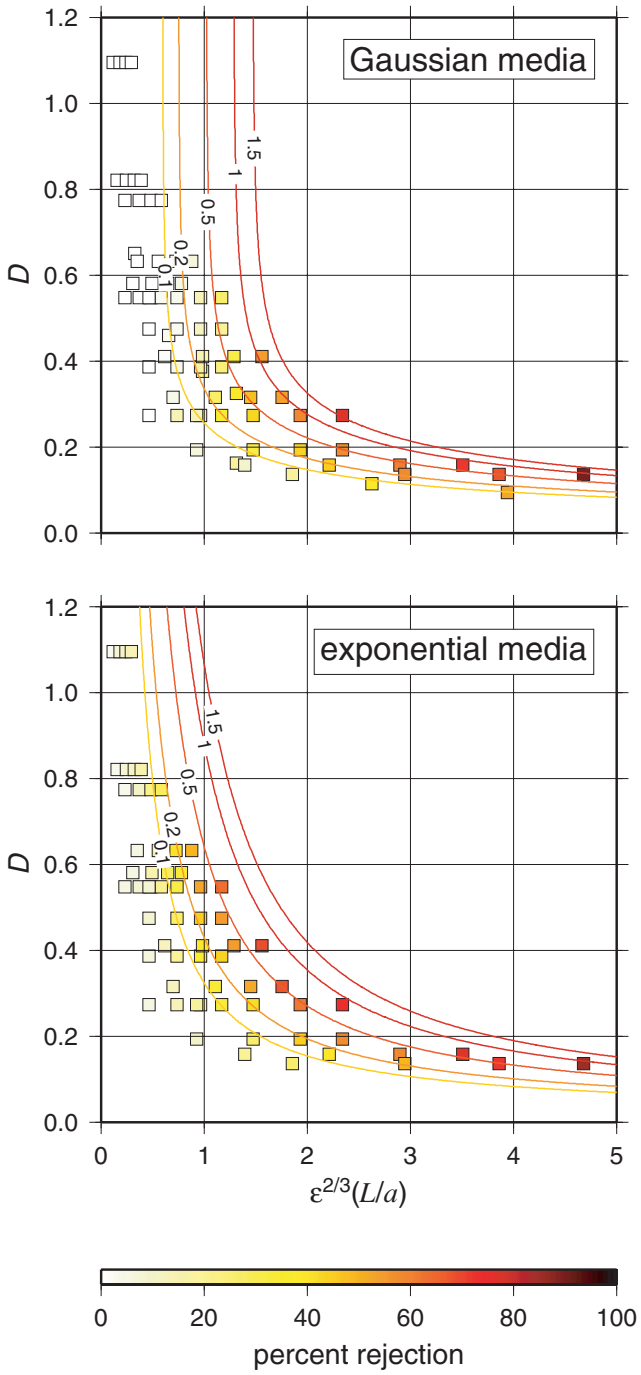

Figure 15. For each Gaussian (top) and exponential (bottom) regime, we plot the percentage of rejected seismograms as a function of D and $\varepsilon^{2/3}(L/a)$, the colours of the squares getting redder as more seismograms are rejected. Also plotted are lines of constant amplitude variance, to see how $\langle \delta A^2 \rangle / A^2$ might work to organize our rejected data. For these lines, the redder colours indicate higher variances.

the same trend as we observed for Gaussian media of small amplitude variances at farther propagation distance for shorter-scale media is present as well. Increasing the parameter a/λ brings these curves closer to the y-axis, again reflecting the divergence of ray theory.

As a test of our criterion, we try to measure the non-linearities we observe in our data and see how they plot in D, $\varepsilon^{2/3}(L/a)$ space. We take the percentage of seismograms that we are compelled to reject in each regime as a rough measure of the failure of linearity. In Fig. 15, we plot the percentage of rejected seismograms against D and $\varepsilon^{2/3}(L/a)$. Contours of constant $\langle \delta A^2 \rangle / A^2$ in both types of media are shown. These iso-variance lines tend asymptotically to their ray-theoretical values in the Gaussian medium for increasing D, but exhibit no such behaviour in the exponential medium, again reflecting the failure of ray theory in these regimes. The data seem fairly well organized along these contours, but not perfectly: in general, as one travels along one of these iso-variance lines, more data are rejected with decreasing D. However, we see some indication that the region of validity for these data is enlarged for low values of small-scale media by a general trend of

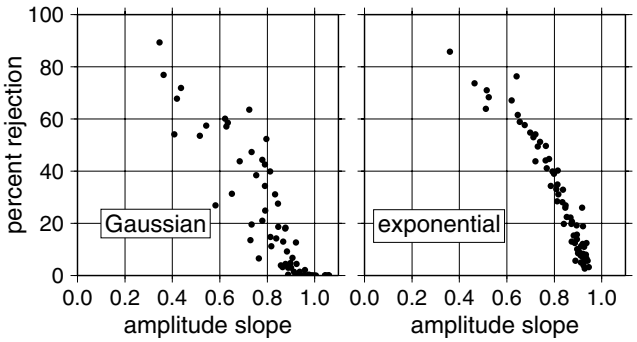

Figure 16. For each regime, we plot the slope of the finite-frequency amplitude scatterplot against the percentage of rejected data in that regime. Gaussian media are plotted on the left, and exponential media on the right. The good correlation in the exponential media and the more scattered correlation in the Gaussian media indicate that non-linearity is controlling the tendency for the measured slopes to fall below unity.

decreasing rejection rate for decreasing D. We should assure the reader that neither D nor $\varepsilon^{2/3}(L/a)$ were monitored in our interactive rejection process.

The slight imperfection in the fit of the contour lines to the percentage rejection rate also might indicate that, although our criterion (97) is necessary, it is not sufficient beyond the ray-theory limit. Ideally, one would construct a criterion based on comparing the variance of the first term in the Taylor expansion for an observable with the variance of the second term: where this second term is no longer negligible with respect to the first, linearity is violated. Iooss *et al.* (2000) computed this criterion for traveltimes in ray theory, agreeing with the scaling we found of $\varepsilon^{2/3}(L/a) \ll$ a constant of order one. In order to compute a similar criterion beyond ray theory, we would need to compute the variance of second-order finite-frequency traveltimes, and we feel that the effort required to do this would be beyond the scope of this paper.

Finally, in Fig. 16, we note that there is a rough correlation between the number of rejected data in a given regime and the value of the best-fitting slope of the finite-frequency amplitude scatterplot in that regime. We feel that this agreement lends credence to the argument we presented in Section 5.4.2, relating the misfit of these slopes to the inability of the Born approximation to account for energy scattering out of the ballistic wavefront. This effect causes our predicted finite-frequency amplitudes to overestimate the measurement, biasing our data below the 45° line. However, our measurement procedure restricts us from measuring amplitudes less than -1, so the scatter collects above this lower limit such that the best-fitting line sees this cut-off and fits the data with a shallow slope, as in Fig. 13.

7 DISCUSSION

To summarize, we have developed formulae that enable us to extend previously derived ray-theoretical expressions for the variances of traveltimes and amplitudes to finite frequencies. In the case of exponential media, we were able to derive a finite-frequency expression for the amplitude variance, where the corresponding ray-theoretical analysis diverges. The only relevant parameter to describe the frequency dependence is the doughnut-hole parameter $D = a/\sqrt{\lambda L}$. Comparing our expressions for scatterplot slopes against our theoretical slopes gives us an indirect indication of how well our expressions for variance predict the actual variances we should observe, if we could do a more complete job of collecting enough different paths to approach the true ensemble average.

Our curves for the scatterplot slopes can tell us what structural scale lengths can be modelled using ray theory, without having to worry about finite-frequency wavefront healing effects. For traveltime data, so long as $a \gtrsim 0.5\sqrt{\lambda L}$, ray theory suffices to accurately compute these data in both Gaussian and exponential media. This limit is, however, higher for amplitude data: in Gaussian media ray theory only predicts adequate amplitudes for $a \gtrsim \sqrt{\lambda L}$; and in exponential media, ray theory completely fails. Our finite-frequency kernels seem to account for most of the traveltime data over all the range of ε that are examined in this study, but they do not do as well for amplitudes, except in the most weakly heterogeneous media that we consider.

Using the ray-theoretical concept of caustic formation as a starting point to determining where the linear assumptions in our Fréchet kernel formalism break down, we try to extend this criterion to finite frequencies by employing our expressions for the theoretical amplitude variance. When we plot the amount of data we reject for a given regime against the amplitude variance, there seems to be broad agreement, although we tend to reject more data if we keep the theoretical amplitude variance constant but decrease D, indicating that, although our condition may be necessary, it is not quite sufficient. In practice, when $\langle \delta A^2 \rangle / A^2 \approx 0.1$, we begin to observe some waveforms that significantly misfit the data. Past this cut-off, the remaining traveltimes in our data set still exhibit a linear relationship to the model. The same cannot be said of amplitudes: the growing misfit in our finite-frequency amplitude scatterplot slopes as we reject more and more data provides us with evidence that a linear theory, like the first Born approximation, accounts inadequately for the real amplitude variations in the data. However, when the theoretical amplitude variance, $\langle \delta A^2 \rangle / A^2$, is below 0.1, linearity in the model proves to be a valid assumption for amplitudes.

ACKNOWLEDGMENTS

We would like to thank Colin Thompson and Jereon Tromp for providing constructive reviews of the original manuscript. Financial support for this work has been provided by the US National Science Foundation under grant EAR-0105387. Many of our pseudospectral wave propagation calculations were performed on Hans-Peter Bunge's *Geowulf* computer cluster. Discussions with Brian Schlottmann, Jean-Paul Ampuero, and Guust Nolet have been very constructive.

REFERENCES

Abramowitz, M. & Stegun, I.A., eds, 1965. *Handbook of Mathematical Functions,* Dover, New York.

Baig, A.M., Dahlen, F.A. & Hung, S.-H., 2003. Traveltimes of waves in random media, *Geophys. J. Int.,* **153,** 467–482.

Boschi, L. & Dziewonski, A.M., 2000. Whole earth tomography from delay times of P, PcP and PKP phases: lateral heterogeneity in the outer core or radial anisotropy in the mantle?, *J. geophys. Res.,* **105,** 13 675–13 696.

Boyse, W. & Keller, J.B., 1995. Short acoustic, electromagnetic, and elastic waves in random media, *J. Opt. Soc. Am. A,* **12,** 380–389.

Cerjan, C., Kosloff, D., Kosloff, R. & Reshef, M., 1985. A non-reflecting boundary condition for discrete acoustic and elastic wave equations, *Geophysics,* **50,** 705–708.

Červený, V., 1985. The application of ray tracing to the numerical modeling of seismic wavefields in complex structures, in, *Seismic Shear Waves, Part A: Theory, Handbook of Geophysical Exploration,* Vol. 15A, pp. 1–124, ed. Dohr, G.P., Geophysical Press, London.

Chernov, L.A., 1960. *Wave Propagation in a Random Medium,* McGraw-Hill, New York.

Dahlen, F.A. & Baig, A.M., 2002. Fréchet kernels for body-wave amplitudes, *Geophys. J. Int.,* **150,** 440–466.

Dahlen, F.A., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite frequency traveltimes—I. Theory, *Geophys. J. Int.,* **141,** 157–174.

Gradshteyn, I.S. & Ryzhik, I.M., 2000. *Tables of Integrals, Series, and Products,* 6th edn, Academic Press, San Diego.

Grand, S.P., 1994. Mantle shear structure beneath the americas and surrounding oceans, *J. geophys. Res.,* **99,** 11 591–11 621.

Grand, S.P., Van der Hilst, R.D. & Widiyantoro, S., 1997. Global seismic tomography: a snapshot of convection in the earth, *GSA Today,* **7,** 1–7.

Gudmundsson, O., Davies, J.H. & Clayton, R.W., 1990. Stochastic analysis of global traveltime data: mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.,* **102,** 25–43.

Hung, S.-H. & Forsyth, D.W., 1998. Modelling anisotropic wave propagation in oceanic inhomogeneous structures using the parallel multidomain pseudo-spectral method, *Geophys. J. Int.,* **133,** 726–740.

Hung, S.-H., Dahlen, F.A. & Nolet, G., 2000. Fréchet kernels for finite frequency traveltimes—II. Examples, *Geophys. J. Int.,* **141,** 175–203.

Hung, S.H., Dahlen, F.A. & Nolet, G., 2001. Wavefront healing: a banana-doughnut perspective, *Geophys. J. Int.,* **146,** 289–312.

Inoue, H., Fukao, Y., Tanabe, K. & Ogata, Y., 1990. Whole mantle P-wave travel time tomography, *Phys. Earth planet. Inter.,* **59,** 294–328.

Iooss, B., Blanc-Benon, P. & Lhuillier, C., 2000. Statistical moments of travel times at second order in isotropic and anisotropic random media, *Waves Random Media,* **10,** 381–394.

Kosloff, D.D. & Baysall, E., 1982. Forward modelling by a Fourier method, *Geophysics,* **47,** 1402–1412.

Kosloff, D.D., Reshef, M. & Loewenthal, D., 1984. Elastic wave calculations by the Fourier method, *Bull. seism. Soc. Am.,* **74,** 875–891.

Kulkarny, V.A. & White, B.S., 1982. Focusing of waves in turbulent inhomogeneous media, *Phys. Fluids,* **25,** 1770–1784.

Marquering, H., Dahlen, F.A. & Nolet, G., 1999. Three-dimensional sensitivity kernels for finite frequency travel times: The banana-doughnut paradox, *Geophys. J. Int.,* **137,** 805–815.

Masters, G., Johnson, S., Laske, G. & Bolton, H., 1996. A shear-velocity model of the mantle, *Phil. Trans. R. Soc. Lond.,* A, **354,** 1385–1411.

Masters, G., Laske, G., Bolton, H. & Dziewonski, A.M., 2000. The relative behavior of shear velocity, bulk sound speed and compressional velocity in the mantle: implications for chemical and thermal structure, in, *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale,* Vol. 117, pp. 63–87, eds Karato, S., Forte, A.M., Liebermann, R.C. Masters, G. & Stixrude, L., Geophysical Monograph, American Geophysical Union, Washington, DC.

Montelli, R., Nolet, G., Dahlen, F.A., Masters, G., Engdahl, E.R. & Hung, S.-H., 2004a. Finite-frequency tomography reveals a variety of plumes in the mantle, *Science,* **303,** 338–343.

Montelli, R., Nolet, G., Dahlen, F.A., Masters, G. & Hung, S.-H., 2004b. Global P and PP traveltime tomography: rays versus waves, *Geophys. J. Int.,* in press.

Morse, P.M. & Ingard, K.U., 1968. Theoretical Acoustics, Princeton University Press, Princeton.

Nolet, G. & Dahlen, F.A., 2000. Wavefront healing and the evolution of seismic delay times, *J. geophys. Res.,* **105,** 19 043–19 054.

Pulliam, R.J., Vasco, D.W. & Johnson, L.R., 1993. Tomographic inversions for mantle P wave velocity structure based on the minimization of *l*2 and *l*1 norms of Intervational Seismic Center travel time residuals, *J. geophys. Res.,* **98,** 699–734.

Ritsema, J. & van Heijst, H.-J., 2002. Constraints on the correlation of P- and S-wave heterogeneity in the mantle from P, PP, PPP, and PKPab traveltimes, *Geophys. J. Int.,* **149,** 482–289.

Ruff, L.J., 1989. Mulit-trace deconvolution with unknown trace scale factors: omnilinear inversion of P and S waves for source-time functions, *Geophys. Res. Let.,* **16,** 1043–1046.

Sato, H. & Fehler, M.C., 1997. *Seismic Wave Propagation and Scattering in the Heterogeneous Earth,* Springer-Verlag, New York.

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

Spetzler, J. & Sneider, R., 2001b. The effect of small-scale heterogeneity on the arrival time of waves, *Geophys. J. Int.,* **145,** 786–796.

Su, W.-J. & Dziewonski, A.M., 1992. On the scale of mantle heterogeneity, *Phys. Earth planet. Inter.,* **74,** 29–54.

Su, W.-J. & Dziewonski, A.M., 1997. Simultaneous inversions for 3-D variations in shear and bulk sound velocity in the mantle, *Phys. Earth planet. Inter.,* **100,** 135–156.

Tatarskii, V.I., 1961. *Wave Propagation in a Turbulent Medium,* McGraw-Hill, New York.

Taylor, J.R., 1982. *An Introduction to Error Analysis,* University Science Books, Mill Valley, CA.

Tibuleac, I.M., Nolet, G., Michaelson, C. & Koulakov, I., 2003. P wave amplitudes in a 3D earth, *Geophys. J. Int..,* **155,** 1–10.

Van der Hilst, R.D., Widiyantoro, S. & Engdahl, E.R., 1997. Evidence for deep mantle circulation from global tomography, *Nature,* **386,** 578–584.

Vasco, L.W. & Johnson, L.R., 1998. Whole earth structure estimated from seismic arrival times, *J. geophys. Res.,* **103,** 2633–2671.

White, B., Nair, B. & Bayliss, A., 1988. Random rays and seismic amplitude anomalies, *Geophysics,* **53,** 903–907.

Zhao, L., Jordan, T.H. & Chapman, C.H., 2000. Three-dimensional Fréchet differential kernels for seismic delay times, *Geophys. J. Int,* **141,** 558–576.

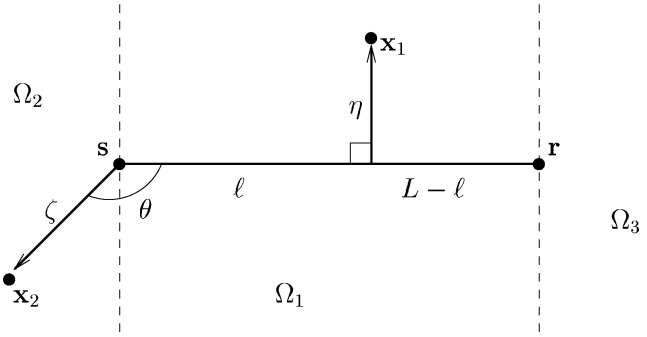

Figure A1. Two scatterers are described in different coordinate systems depending on whether they are in the infinite cylinder centred on the ray, Ω_1, or in one of the endcap regions, Ω_2 or Ω_3, which we conceptualize as hemispheres of infinite radius centred on the source, **s**, or receiver, **r**, respectively. In the cylinder, a scatterer, $\mathbf{x}_1$, is located by its projection on the ray, ℓ, and its distance away from the ray η. In Ω_2 the scatterer, $\mathbf{x}_2$, is a distance ζ away from the source, and $\mathbf{x}_2 - \mathbf{s}$ makes an angle θ with the ray. In both cases, the third coordinate is an azimuthal angle, ϕ, about the ray (not shown).

APPENDIX A: REDUCTION TO RAY THEORY FOR AMPLITUDE KERNELS

In Dahlen & Baig (2002), a hefty portion of the analysis was devoted to demonstrating how their finite-frequency amplitude kernel expression reduced to the corresponding ray-theoretical expression in the high-frequency limit ($\omega \to \infty$). The ray-theoretical analysis gave an amplitude perturbation in a medium with a homogeneous background slowness, given here in eq. (19). However, upon evaluating the high-frequency limit of K_{A}, Dahlen & Baig (2002) only recovered one of these terms:

$$\delta(\ln A)_{\mathrm{DB}} = \lim_{\omega\to\infty} \iiint_{\Re^3} K_{\mathrm{A}}(\mathbf{x})\delta\sigma(\mathbf{x})\, d^3\mathbf{x} \approx \frac{-1}{2\sigma L}\int_0^L \ell(L-\ell)\nabla_\perp^2\delta\sigma(\mathbf{x})\, d\ell. \tag{A1}$$

At the time, we justified this discrepancy by noting that, for a heterogeneous medium, the term containing the $\nabla_\perp^2\delta\sigma$ factor should dominate the unrecovered terms that do not involve differentiating the slowness perturbations. However, upon further inspection, it emerges that these missing terms are a symptom of the limitations of the paraxial approximation—by extending this approximation to a higher order, these terms can be recovered.

First, we need to split the integral over all of space, $\Re^3$, in eq. (A1) into three regions, as shown in Fig. A1: Ω_1 is a cylinder of infinite radius whose axis is coincident with the ray; and Ω_2 and Ω_3 are hemispherical endcaps that are also of infinite radius, but are centred on the source and receiver respectively. In Ω_1, the location of a point is given in a cylindrical coordinate system, (η, ϕ, ℓ), corresponding to radius, polar angle, and along-ray distance. In Ω_2 and Ω_3, independent spherical coordinate systems, radius, co-latitude, and azimuth, (ζ, θ, ϕ), are used.

In the vicinity of the ray, in Ω_1, we can extend the paraxial approximation to higher order as follows:

$$\frac{L}{L'L''} \approx \Gamma\left[1 - \frac{\eta^2}{2}\left(\frac{L^2 - 2L\ell + 2\ell^2}{\ell^2(L-\ell)^2}\right)\right] \tag{A2}$$

and

$$\Delta T \approx \frac{\sigma\eta^2\Gamma}{2}\left[1 - \frac{\eta^2}{4}\frac{L^2 - 3L\ell + 3\ell^2}{\ell^2(L-\ell)^2}\right], \tag{A3}$$

for a point a distance η away from the ray. This approximation allows us to add some terms to the expression for the contribution to the amplitude perturbation from the ray-centred cylinder:

$$\delta(\ln A)_{\Omega_1} = \int_0^L \frac{\int_0^\infty \omega^2|u(\omega)|^2\,[I_1(\ell,\omega) + I_2(\ell,\omega) + I_3(\ell,\omega)]\, d\omega}{\int_0^\infty |u(\omega)|^2 dw}\, d\ell, \tag{A4}$$

where I_1 gives us the leading-order term derived by Dahlen & Baig (2002) as eq. (A1); I_2 arises from the approximation (A2) for the geometrical spreading factors; and I_3 is a consequence of the correction term for

$$\cos(\omega\Delta T) \approx \cos\left(\frac{\omega\sigma\Gamma\eta^2}{2}\right) + \frac{\omega\sigma\eta^4\Gamma(L^2 - 3L\ell + 3\ell^2)}{8\ell^2(L-\ell)^2}\sin\left(\frac{\omega\sigma\Gamma\eta^2}{2}\right). \tag{A5}$$

Although the approximations (A2) and (A3) are only strictly valid in the vicinity of the ray, we can extend the integration out over the whole region Ω_1, since this over-extension introduces an error term one factor of ω less than the contribution from the neighbourhood of the ray. So, for high frequencies,

$$\begin{aligned} I_2 &\approx \frac{\sigma\Gamma\delta\sigma(\mathbf{x})}{4\pi}\frac{L^2 - 2L\ell + 2\ell^2}{\ell^2(L-\ell)^2}\int_0^{2\pi}\int_0^\infty \eta^3\cos\left(\frac{\omega\sigma\Gamma\eta^2}{2}\right) d\eta\, d\phi \\ &= \lim_{t\to\infty}\left[\frac{L^2 - 2L\ell + 2\ell^2}{\omega^2 L\ell(L-\ell)}\frac{\delta\sigma(\mathbf{x})}{\sigma}\,(1 + \cos\omega t - \omega t\sin\omega t)\right] \end{aligned} \tag{A6}$$

and

$$I_3 \approx \frac{\omega\sigma^2\Gamma^2\delta\sigma(\mathbf{x})}{16\pi}\frac{L^2-3L\ell+3\ell^2}{\ell^2(L-\ell)^2}\int_0^{2\pi}\int_0^{\infty}\eta^5\sin\left(\frac{\omega\sigma\Gamma\eta^2}{2}\right)d\eta d\phi$$
$$= \lim_{t\to\infty}\left[\frac{-L^2+3L\ell-3\ell^2}{\omega^2 L\ell(L-\ell)}\frac{\delta\sigma(\mathbf{x})}{\sigma}\left(1+2\cos\omega t-\omega^2 t^2\cos\omega t+2\omega t\sin\omega t\right)\right]. \tag{A7}$$

The total contribution to the amplitude perturbation from Ω_1 is then given as

$$\delta(\ln A)_{\Omega_1} = \frac{-1}{2\sigma L}\int_0^L \ell(L-\ell)\nabla_\perp^2\delta\sigma(\mathbf{x})\,d\ell + \frac{1}{\sigma L}\int_0^L \delta\sigma(\mathbf{x})\,d\ell, \tag{A8}$$

where the highly oscillatory terms in eqs (A6) and (A7) make no contribution, by an application of the Riemann–Lebesgue lemma to the $d\omega$ integral in eq. (A4).

If we now turn our attention to the source endcap, Ω_2, we can write its contribution to the amplitude perturbation as

$$\delta(\ln A)_{\Omega_2} = \int_0^{2\pi}\int_{\frac{\pi}{2}}^{\pi}\int_0^{\infty} K_{\mathrm{A}}(\mathbf{x})\delta\sigma(\mathbf{x})\zeta^2\sin\theta\,d\zeta\,d\theta d\phi. \tag{A9}$$

In the vicinity of the source, the following approximations hold:

$$\frac{L}{L'L''}\approx\frac{1}{\zeta}, \qquad \Delta T\approx\sigma\zeta(1-\cos\theta). \tag{A10}$$

In this limit, the contribution to the amplitude perturbation becomes

$$\delta(\ln A)_{\Omega_2} \approx \frac{\sigma\,\delta\sigma(\mathbf{s})}{\int_0^\infty|u(\omega)|^2 d\omega}\int_{\frac{\pi}{2}}^{\pi}\int_0^{\infty}\int_0^{\infty}\omega^2|u(\omega)|^2\cos\left[\omega\sigma_{\mathrm{s}}\zeta(1-\cos\theta)\right]\zeta\sin\theta\,d\omega\,d\zeta\,d\theta$$
$$= -\frac{\delta\sigma(\mathbf{s})}{2\sigma}, \tag{A11}$$

where we have abused our approximations (A10) by carrying the $d\zeta$ integral out to infinity, although, as in the previous case, we thereby incur an error of higher order in the limit $\omega\to\infty$. We obtain the exact endpoint contribution we need for the amplitude perturbation; the same analysis can be done for the receiver endcap (or we can just appeal to reciprocity) to obtain

$$\delta(\ln A)_{\Omega_3} = -\frac{\delta\sigma(\mathbf{r})}{2\sigma}.$$

So, for a homogeneous background medium, the slowness amplitude kernel does indeed capture all of the behaviour of first-order ray theory:

$$\lim_{\omega\to\infty}\iiint_{\Re^3} K_{\mathrm{A}}(\mathbf{x})\delta\sigma(\mathbf{x})\,d^3\mathbf{x} = \frac{-1}{2\sigma L}\int_0^L \ell(L-\ell)\nabla_\perp^2\delta\sigma(\mathbf{x})d\ell + \frac{1}{\sigma L}\int_0^L\delta\sigma(\mathbf{x})d\ell$$
$$-\frac{1}{2\sigma}[\delta\sigma(\mathbf{s})+\delta\sigma(\mathbf{r})]. \tag{A12}$$

We have no proof that, for a general background medium, the exact finite-frequency expressions completely reduce to ray theory in the infinite-frequency limit; however, our results for the simple homogeneous background encourage us to believe that this is a plausible conjecture.