Geophysica (2023), 58 (1), 3–32

Use of Amplitudes in Velocity and Joint Velocity-Attenuation Surface Wave Geotomography

William Menke[1,*] *and Charlotte Rhoads*[2]

[1]Lamont Doherty Earth Observatory of Columbia University, Palisades, New York, USA
[2]Vassar College, Poughkeepsie, New York, USA
[*]Corresponding mail address: menke@ldeo.columbia.edu

(Submitted: 2022-09-12; Accepted: 2023-06-12)

Abstract

*In the past, geotomographic imaging of phase velocity using Rayleigh waves has mostly relied upon travel time data (or, equivalently, path-averaged phase velocity data). Over the last decade or two, the wide availability of observations from well-calibrated seismometers has enabled the use of surface wave amplitudes to supplement travel time data. Amplitudes vary in a heterogeneous Earth due to lensing effects, with high velocity patches leading to defocusing and lower amplitudes and low velocity patches leading to focusing and higher amplitudes. Amplitude data complements travel time data, because while the latter is due to the phase velocity along the propagation path, the former is due to the second derivative of phase velocity perpendicular to it. In this study, we quantify the benefit of supplementing traditional travel time geotomography with amplitude data. Key to our formulation is the representation of the spatially-varying phase velocity using cubic splines, which allow second derivatives to be stably and efficiently computed. We find that high-quality (*10 : 1 *signal-to-noise ratio) amplitude data improves recovery of the phase velocity by 95 % and poor-quality (*1 : 1 *s.n.r.) by 75 %. Furthermore, the improvement is excellent irrespective of whether the azimuthal coverage of sources is poor or good, and irrespective of whether the data are sparse and the problem underdetermined, or plentiful and overdetermined. We also examine the viability of a joint inversion for phase velocity and intrinsic attenuation factor. This problem is not as well-behaved as the velocity-only one. Its sensitivity to amplitude noise is higher, with 1 % amplitude noise leading to up to 4 % phase velocity error; furthermore, the attenuation factor is well-recovered only when the problem is overdetermined. A squeezing analysis indicates that phase velocity and attenuation factor can significantly trade off.*

Keywords: tomography; surface waves; attenuation

1 Introduction

Rayleigh wave geotomography is a standard geophysical tool for imaging the velocity structure of the Earth (e.g., *Nishimura and Forsyth*, 1989; *Lee and Nolet*, 1997; *Li and Detrick*, 2006; *Nettles and Dziewonski*, 2008; *Shen and Ritzwoller*, 2016; *Ekström*, 2017). Most commonly, the quantity that is imaged is the phase velocity of the wave (or its phase slowness, the reciprocal of velocity), viewed as a local property of a particular place on the Earth's surface, and at a particular frequency of observation. Rayleigh wave phase velocity has a direct correlation to the shear wave velocity in the layers beneath the Earth's surface. Post-processing of images for a suite of frequencies allows determination of three-dimensional shear velocity structure of the shallow part of the Earth. However,

the phase velocity maps, themselves, are often used directly in more qualitative interpretations of Earth structure, because they are considered less prone to artifacts associated with non-uniqueness. The most commonly used data are path averages of phase velocities, or equivalently, travel times between sources (earthquakes) and receivers (seismometers), or in the case of ambient noise studies, between pairs of receivers (e.g., *Bensen et al.*, 2007; *Zha et al.*, 2014; *Ekström*, 2017). Another application of Rayleigh waves is the imaging of the Earth's attenuation structure (*Dalton and Ekström*, 2006a,b; *Bowden et al.*, 2017; *Dalton et al.*, 2017). Here the attenuation factor (defined later), or equivalently, the quality factor, is imaged. Attenuation factor is a quantity distinct from phase velocity that provides useful information about Earth structure, and especially about its thermal structure, for attenuation factor increases strongly with temperature (e.g., *Cammarano et al.*, 2003).

It has long been understood that seismic amplitudes depend on slowness, as well as attenuation, through two effects (discussed further below), one related to lensing (geometrical spreading) and the other to local conservation of energy. In their attenuation inversion, *Dalton and Ekström* (2006b) account for these effects by correcting amplitude data for them, using previously-published slowness maps derived from travel times, alone. Furthermore, *Dalton and Ekström* (2006a) show that a global dataset of surface wave amplitudes contains useful information about slowness. Their slowness image based on amplitude data, only, bears a string resemblance to a slowness image derived from travel times, alone.

These findings raise two interesting questions. The first is the viability of using travel time and amplitude data together to determine slowness, under the assumption that the effect of attenuation on amplitudes are small. Such an assumption would be justified in some areas of the Earth, such as cratons, where the attenuation factor is known by independent means to be small (e.g., *Cafferky and Schmandt*, 2015). The advantage of using both data types is that they are complementary, and may allow a higher-resolution image than can be gained from travel times, alone. The second is whether a joint inversion for slowness and attenuation factor is feasible, or rather, whether it is feasible when the travel time data do not completely constrain the slowness image. In that case, one requires amplitude data to do double-duty; on the one hand, they needs to supplement travel times in the slowness inversion, on the other hand, they need to provide information about attenuation. Whether it can do both is unclear.

We address these questions using a series of numerical experiments that allow us to gauge the sensitivity of the geotomography to measures of data quality, such as noise level, degree of over-determinedness, and angular spread of sources.

2 *The physics underlying travel time and amplitude*

Rayleigh surface waves are a dispersive phenomenon, with a phase slowness, $u(x,y,\omega)$ that is a function of both geographical position, (x,y), and angular frequency, ω. However, in this paper we assume that measurements are being performed at a specific frequency, $\omega = \omega_0$, isolated by Fourier transforming the seismograms. Consequently, we drop explicit mention of frequency and write phase slowness simply as, $u(x,y)$. For

wavelengths small compared to the scale of smooth heterogeneities in phase slowness, Rayleigh waves can be shown to obey the premise of classical ray theory: that the wavefront advances in the (x,y) plane normal to itself at a speed given by the local phase velocity, so that wavefront normals trace out ray paths (*Wang and Dahlen*, 1995). This behavior leads to a two-dimensional version of the well-known ray equation (*Aki and Richards* (2009); our Equation C.2).

Rayleigh waves have a complicated depth, z, dependence, but for large depths, their decay is exponential. The decay rate can be shown to increase with phase slowness (Appendix, Section I); that is, Rayleigh wave energy is squeezed towards the surface as phase slowness is increased (or, equivalently, as the velocity of the medium is decreased).

Under the ray approximation, travel time, T, increases along the ray according to $\mathrm{d}T = u\,\mathrm{d}s$, where s is arc length, and the overall travel time is the integral of phase slowness, $u(x,y)$, along the ray path. The linear relationship between T and u explain why most imaging uses slowness, as contrasted to velocity, $v = 1/u$ as the imaged quantity; the relationship between travel time and velocity are nonlinear.

This is a two-dimensional analog of the behavior of body waves under the ray approximation. Supposing that a spatially-constant perturbation in phase slowness, $u = u_0 + \Delta u$, leads to travel time perturbation, $T = T_0 + \Delta T$, the ray integral implies $\Delta T/T_0 = \Delta u/u_0$. Consequently, in a heterogenous medium, travel time perturbations are never larger than the largest slowness perturbation, $|\Delta T|/T_0 \leq \max|\Delta u|/u_0$ (when both are measured in percent).

Under the ray approximation, phase slowness perturbations cause lensing (geometrical spreading). Patches of locally high slowness cause rays to bend towards one another (focusing) and patches of locally low slowness cause rays to bend away from one another (defocusing) (Fig. 2.1A). As the time-averaged energy flux, $\langle F \rangle$, between neighboring rays is constant, and as energy is proportional to the square of amplitude, A^2, focusing causes amplitude to increase and defocusing causes it to decrease. This effect can be quantified by the second derivative of the slowness, measured perpendicular to the ray path (*Dahlen and Tromp*, 1998, and Appendix, Section I). A positive second derivative corresponds to a patch of low slowness, which leads to defocusing and an amplitude decrease. A negative second derivative corresponds to a patch of high slowness, which leads to focusing and an amplitude increase. An idiosyncrasy of focusing and defocusing is that their effect on amplitudes is 'non-local,' in the sense that a perturbation in slowness can lead to a perturbation in slowness that grows with distance. For instance, once an initially-plane wave is defocused, the distance between neighboring rays increases indefinitely with propagation distance, causing the amplitude to decrease indefinitely. Mathematically, this effect is represented by the $(s - s')$ factor in Equation (C.18), which grows in proportion to the distance between slowness perturbation and observer. Consequently, amplitude perturbations $|\Delta A|/A_0$, can be much larger than the slowness perturbations, $|\Delta u|/u_0$, that cause them.

In addition to being proportional to A^2, energy flux depends on the phase slowness of the wave and on local material properties such as density, ρ. This 'local conservation of

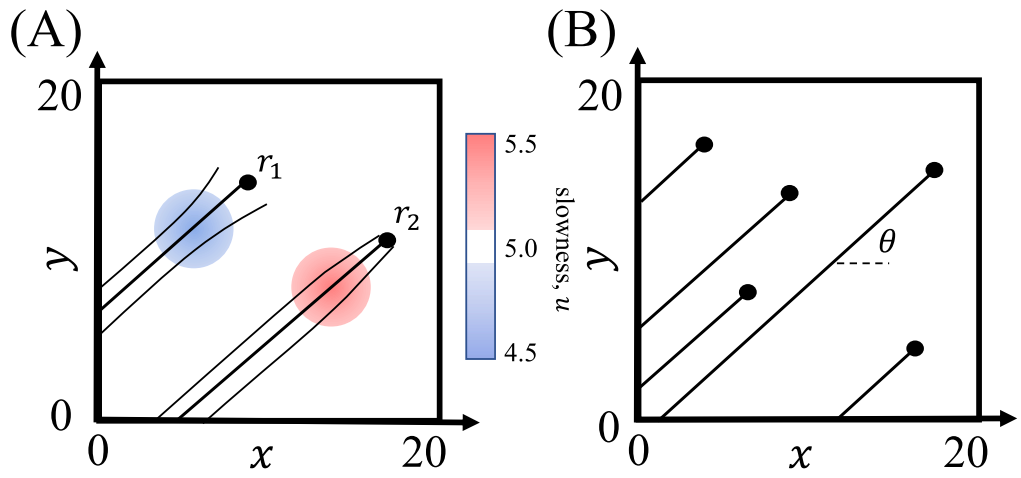

Fig. 2.1: (A) Rays (curved lines) are focused by high slowness regions (red), causing high amplitudes, and defocused by low slowness regions (blue), causing low amplitudes. (A) Rays (lines) from a distant earthquake, all with orientation, θ, impinging upon stations (dots).

energy' effect is in addition to lensing, as can be understood by considering a case where no lensing occurs, such as a plane wave normally incident upon a planar interface between two different media.

Neglecting reflections, the energy flux on the two sides of the interface is equal. Consequently, amplitude must change to balance changes in material properties and phase slowness. For a plane shear wave in a homogeneous medium, the flux is $\langle F \rangle = \frac{1}{2}\omega^2 \rho u^{-1} A^2$ (*Aki and Richards*, 2009). The condition that the flux be constant across an interface between media with two different phase slownesses (but the same density) is $\partial \langle F \rangle / \partial u = 0$ or $\mathrm{d}A/A = \gamma \, \mathrm{d}u/u$, with $\gamma = \frac{1}{2}$ (see Appendix, Section I). The situation for surface waves is more complicated, because the vertical distribution of energy changes across the interface, too. It is shown in Section I of the Appendix that in this case $\gamma = 1$ (where A now refers to the amplitude at $z = 0$).

Finally, Rayleigh waves lose amplitude by the process of anelastic attenuation; that is, the passage of the wave causes inelastic deformation within the Earth that converts elastic wave energy into heat. Because the mechanism is microscopic, the energy loss due anelasticity need obey the principle that the effect of propagation through a distance, $2s$, equals the effect of propagation through a distance, s, convolved with itself (*Richards and* W. *Menke*, 1983). This condition implies that amplitude, $A(\omega)$, must decay as $A(\omega) = A_0 \exp(-\alpha s)$, where A_0 is an initial amplitude and α is a decay rate (or 'attenuation factor'). In general, the attenuation factor may vary both with angular frequency, ω, and position, (x,y). In the context of ray theory, amplitude decays as $\mathrm{d}A/A = \mathrm{d}\ln A = -\alpha \, \mathrm{d}s$, where s is arc-length along the ray. Thus, $\ln A$ and $-\alpha$ in the attenuation relationship play roles completely analogous to T and u in the travel time relationship. Thus, measurements of amplitude can be used to image attenuation factor, $\alpha(x,y)$ (for fixed angular frequency, ω_0), in the same way that measurements of travel

time, T, can be used to image slowness, $u(x,y)$. As an aside, we note that some authors use quality factor, Q, not attenuation factor, α, to quantify anelastic attenuation. They are related by $Q = \frac{1}{2}\omega u/\alpha$.

Images of attenuation factor provide information about the Earth that is complementary to images of slowness; taken together, they can provide more information about the Earth than either does alone. For instance, although both an increase in temperature and a decrease in chemical depletion can cause a region of high slowness, only the former causes a region of high attenuation factor. An important difference between the imaging of slowness and attenuation is that amplitude is affected by both slowness and attenuation, while travel time is affected only by slowness. Consequently, imaging attenuation factor more complicated than imaging slowness.

We now describe a set of numerical experiments designed to explore the viability of two types of geotomography: (1) slowness-only geotomography using travel time and amplitude data; and (2) joint slowness-attenuation-factor geotomography using travel time and amplitude data.

3 Design of numerical experiments

We consider a rectangular study region in which spatially-varying slowness and attenuation factor are the unknown model parameters. The rectangular region contains receivers (seismometers) at which travel time and amplitude of Rayleigh waves are measured. The sources (earthquakes) are outside of – and distant from – the study region, so that rays impinging upon the study region are parallel (that is, plane waves). This geometry is similar to the one used by *Li and Detrick* (2006) in their study of Rayleigh waves in Iceland.

In actual practice, seismic surface wave studies are conducted over a wide range of scale, from global studies tens of thousands of kilometers in scale utilizing frequencies of a few mHz (e.g., *Nettles and Dziewonski*, 2008) to environmental studies just tens of meters in scale utilizing frequencies of tens of Hz (e.g., *West and* W. *Menke*, 2000). Consequently, we do not assign any realistic distance units to our models. Instead, a pixel is taken to be 1×1 distance units in size. Similarly, as slowness can be measured in a variety of units (e.g., s/km, ms/m, etc.) we do not assign realistic slowness units, either. To simplify plotting, we assign the background slowness to be 5 units, and perturbations in slowness to have an absolute maximum of 1 unit. This leads to a ± 20 % variation in slowness, which is about the maximum encountered in realistic applications. We choose the background value of the attenuation factor so that the decrease in amplitude across the study region is no more than a factor of ten, in keeping with the reality that very-highly-attenuated waves in practice are unobservable. Modeled signal-to-noise levels were informed by the fact that travel times can be measured much more accurately than amplitudes. Modern cross-correlation techniques can reduce travel time error to less than 0.01 % (*Waldhauser and Ellsworth*, 2000), whereas we infer from the stated error levels in *Dalton and Ekström* (2006b) that achieving an error of 1 % for amplitudes is challenging. Most of our images consist of a 21×21 grid of pixels. This number is a compromise between having enough

pixels to represent non-trivial patterns in $u(x,y)$ and $\alpha(x,y)$ and computation time (which grows with the number of unknown pixels).

A smooth representation of slowness is vital for properly computing the effect of lensing (geometrical spreading), because stable estimates of the second derivative are needed. We use the form of cubic interpolation described by *Keys* (1981) to represent slowness and attenuation factor. This method represents a smooth function by a rectangular, $N_x \times N_y$ grid of control points, with uniform spacing $(\Delta x, \Delta y)$ on the intervals $0 \le x \le x_{\max}$ and $0 \le y \le y_{\max}$. Consequently, the total number of slowness model parameters is $N_x N_y$ and attenuation factor control points is $N_x N_y$. The total number of model parameters is either $M = N_x N_y$, when only slowness is estimated, or $M = 2N_x N_y$, when both slowness and attenuation factor are estimated. Only the 4×4 grid of control points surrounding a given point, (x_0, y_0), contributes the estimate of the function and its derivatives at that point, so the interpolation is a 'local' one. The spacing of the control points imposes smoothness on the estimated solution, making other forms of smoothness regularization unnecessary. Although the interpolated function can be evaluated indefinitely finely, allowing for high-resolution images, here we present only images with grid spacing $(\Delta x, \Delta y)$; that is, images of the control points.

We use simple slowness and attenuation factor test functions, chosen to make easy visual assessment of the success of an experiment (Fig. 3.1):

$$\begin{aligned} u(x,y) &= u_0 + u_1 \sin\left(\frac{\pi x}{x_{\max}}\right) \sin\left(\frac{\pi y}{y_{\max}}\right) \\ \alpha(x,y) &= \alpha_0 + \alpha_1 \sin\left(\frac{2\pi x}{x_{\max}}\right) \sin\left(\frac{\pi y}{y_{\max}}\right) \end{aligned} \tag{3.1}$$

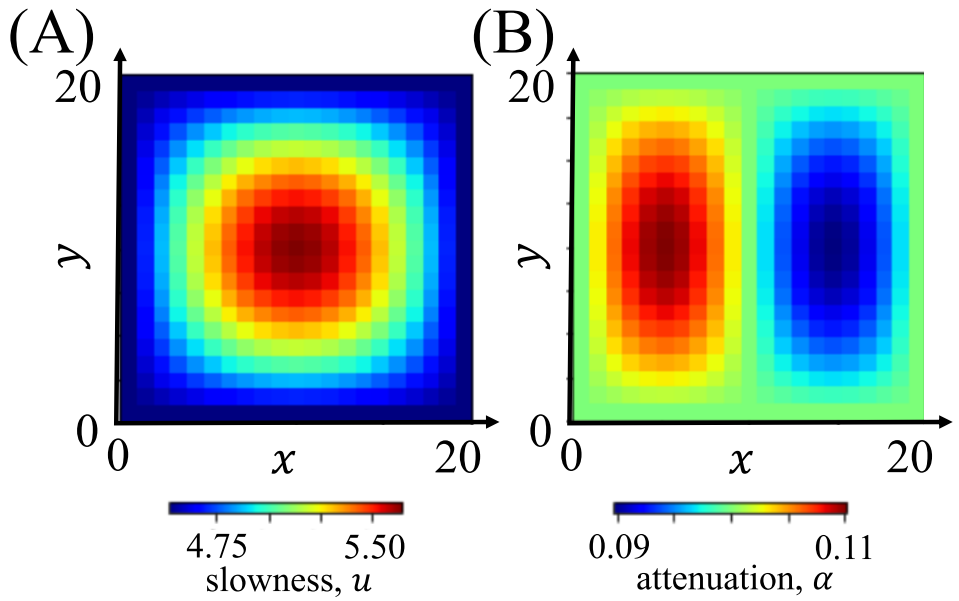

Fig. 3.1: Test images of (A) slowness, $u(x,y)$ and (B) attenuation, $\alpha(x,y)$ that are used in the numerical experiments. Each image has $21 \times 21 = 441$ pixels.

Our experiments have N_r receivers (seismometers), randomly drawn from a 2D uniform distribution on $(\Delta x, \Delta y)$, and N_θ sources (earthquakes), whose azimuths, θ, is are randomly drawn on a uniform distribution on the interval, $0 \leq \theta < \theta_{\max}$. In most experiments, we set $\theta_{\max} = 360°$, but sometimes we use a smaller value that corresponds to a suboptimal arrangement of sources. As each source-receiver combination contributes two data, travel time and amplitude, the number of data in an experiment is $N = 2N_r N_\theta$. We compute these data numerically, according to the physics described in the previous section, and perturb them with normally-distributed random noise, to simulate real data (Fig. 3.2).

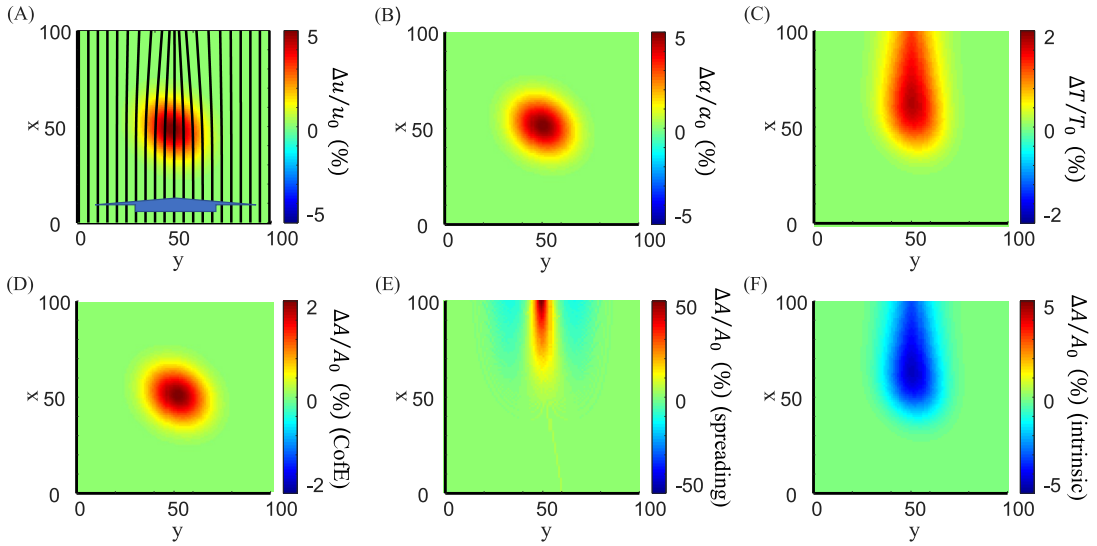

Fig. 3.2: Exemplary Earth model and corresponding data. (A) Slowness anomaly, $\Delta u/u_0$ (colors), showing rays (curves) from a distant earthquake propagating in the y-direction. (B) Attenuation anomaly, $\Delta\alpha/\alpha_0$. (C) Traveltime anomaly, $\Delta T/T_0$. (D) The part of the amplitude anomaly, $\Delta A/A_0$ due to conservation of energy (CoE). (E) The part of the amplitude anomaly, $\Delta A/A_0$ due to geometrical spreading. (E) The part of the amplitude anomaly, $\Delta A/A_0$ due to intrinsic attenuation.

The goal in geotomography is to estimate the model parameters (slowness and attenuation factor control points, generically the vector, $\mathbf{m}$) from the data (travel time and amplitudes, generically the vector, $\mathbf{d}$). This is enabled through writing the relationships:

$$\begin{aligned} \text{travel time} &= \text{ray integral of slowness} \\ \text{amplitude} &= \text{ray integral of second derivative of slowness} \\ &+ \text{slowness perturbation at the receiver} \\ &+ \text{ray integral of attenuation factor} \end{aligned} \tag{3.2}$$

as a matrix equation of form, $\mathbf{Gm} = \mathbf{d}$ (Appendix, Section L). Each row of this equation presents a so-called sensitivity kernel; that is, the relationship between a travel time or amplitude measurement and the values of all of the control points. Because the data depend only on control points near the ray path, the sensitivity kernel (row of $\mathbf{G}$), when plotted

in (x,y), has the form of a narrow band centered on the ray path. The equation, $\mathbf{Gm} = \mathbf{d}$, is solved in the least squares sense. We use the damped least squares solution (Appendix, Section L), because it can handle the case where some control points are not constrained by data by forcing them to the background value. Construction of the matrix equation is facilitated by the simple form of the cubic interpolation algorithm; the relationship between the value of a function and its derivatives and the values of the control points are simple linear algebraic formula of known form. The solution of the least squares equation is facilitated by $\mathbf{G}$ being sparse (most elements are zero).

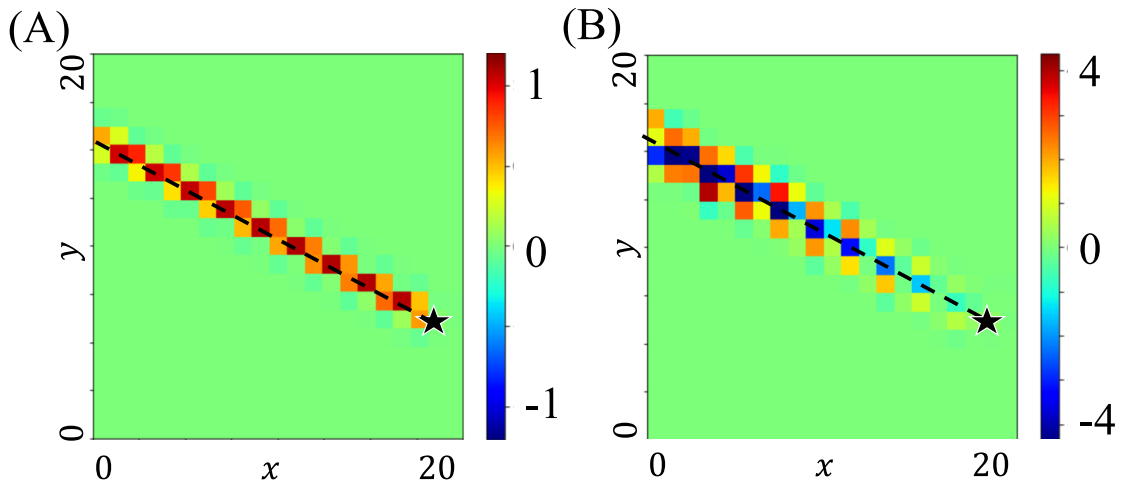

Fig. 3.3: Exemplary sensitivity kernels (colors) for a single ray path (dotted line) to a receiver (star). (A) Sensitivity of travel time to slowness. (B) Sensitivity of amplitude to slowness. The sensitivity of amplitude to attenuation is identical to (A), except its sign is reversed.

Experiments Set A, concerning inversions for velocity (or slowness), only. This set of experiments addresses the advisability of supplementing travel time inversions with amplitudes, when the unknowns are the spatially-varying pattern of phase slownesses. In the best-case scenario of a large number ($2\times$ overdetermined) of low-noise (1 %) data, travel time data alone, amplitude data alone, and both data types used together all lead to high-quality slowness images (not shown). Note that the local conservation energy effect regularizes the amplitude-only inversion. Without it, the data would only depend on the 2^{nd} derivative of the slowness, and a linear term could not be resolved in the image.

Experiment A.1, to assess the effect of amplitude noise on the quality of the solution. In this experiment, we hold the travel time noise constant, at 0.1 % and the perform a sequence of geotomographic inversions in which the amplitude noise is increased from 0.1 % to 1000 %. Ten travel time only inversions and ten travel time and amplitude inversions are performed at each noise level, with randomly-drawn earthquake source azimuths, receiver locations and noise values, and the median improvement of the estimated phase velocity models is tabulated. The median improvement is a near-constant 90 % for noise levels below 20 % and slowly falls off at higher noise levels, becoming negligibly small beyond about 300 % (Fig. 3.4A). This behavior indicates that even moderately noisy amplitude data, with noise levels in the 10 % – 100 % range can substantially improve the

results of an inversion.

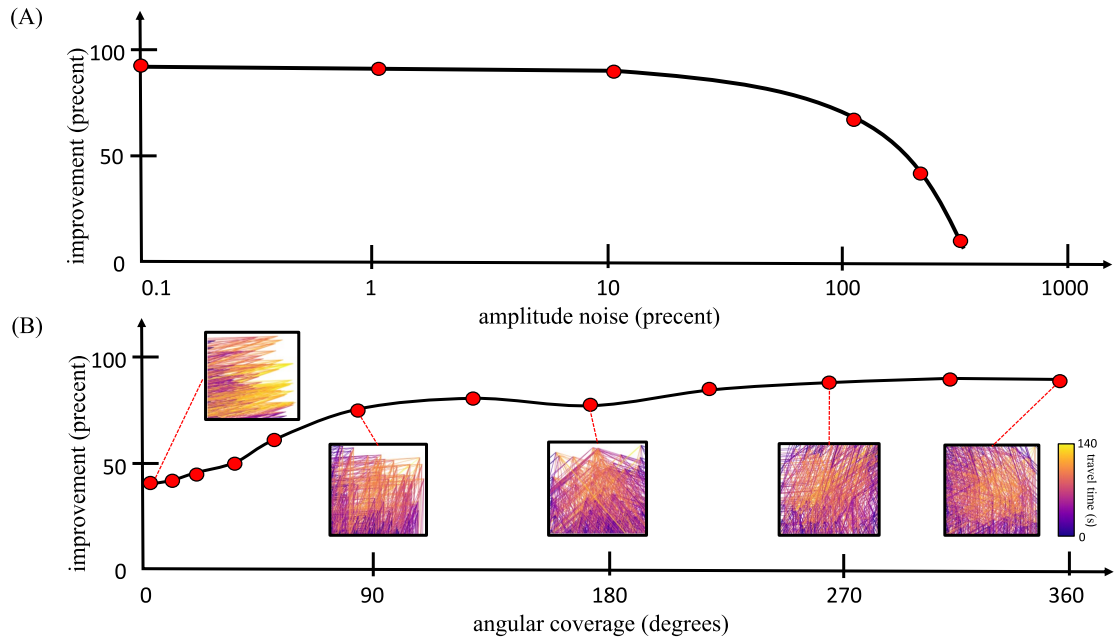

Fig. 3.4: Slowness-only inversion. (A) Percent improvement in slowness image recovery, as a function of noise in the amplitude data. The image contains 21×21 pixels and the number of stations and the number of earthquakes both are equal to $L = 30$. The travel time noise level is held fixed at 0.1 %. The improvement is relatively insensitive to noise level, for levels less 100 %. (B) Percent improvement in slowness image recovery, as a function of angular coverage of earthquake sources. The image contains 21×21 pixels, the number of stations is 81 and the number of earthquakes is 10. The travel time and amplitude noise levels are held fixed at 1 %.

Experiment A.2, to assess the ability of amplitude data to improve an inversion when the azimuthal spread of earthquake sources is poor. In this experiment, we hold the travel time and amplitude noise constant at 1 % and the perform a sequence of geotomographic inversions in which the azimuth interval in increased from ten degrees to 360 degrees. Ten travel time only inversions and ten travel time and amplitude inversions are performed for each azimuth range, with randomly-drawn earthquake source azimuths, receiver locations and noise values, and the median improvement of the estimated phase velocity models is tabulated. The improvement increases from about 40 % for small azimuthal intervals to about 90 % at the larger (Fig. 3.4B). This behavior indicates that even amplitude data can substantially improve the results of an inversion, irrespective of whether the azimuthal coverage of sources is poor or excellent. In particular, the 40 % improvement in the poor-range coverage is especially important, for the travel time only inversion (not shown) is especially poor in this case.

Experiment A.3, to assess the ability of amplitude data to improve an inversion when the number of data is low. In this experiment, we hold the travel time and amplitude noise constant at 1 % and 0.1 %, respectively, and the perform a sequence of geotomographic inversions where the number of data are slowly increases, from a low number to a high one, bracketing the number, $L^2 = 441$, at which the travel time only inversion is even determined. Ten inversions are performed for each case, with randomly-drawn earth-

quake source azimuths, receiver locations and noise values, and the median improvement of the estimated phase velocity model is tabulated. The median improvement rises from about 10 %, when the inversion is very underdetermined, to about 90 % when it is very overdetermined (Fig. 3.5). Of special interest are cases where the travel time inversion is close to even-determined, for then the travel time and amplitude inversion is visibly much better than the travel time only one. In practice, most inversions are close to being even-determined, because those designing them often seek the highest resolution (and hence the most model parameters) that their dataset possibly can support. The addition of amplitude data leads to marked improvement in these cases.

Overall, this set of experiments illustrates the advisability of supplementing travel time inversions with amplitudes. It leads to markedly improved phase velocity estimates, even when the amplitude noise level is moderately high, and when the number of data or their azimuthal coverage is poor.

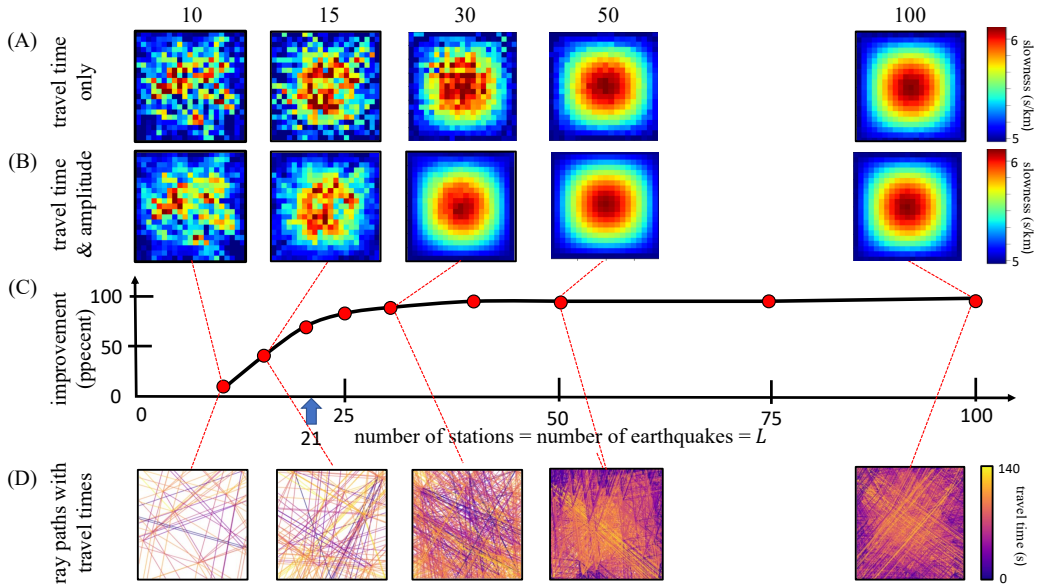

Fig. 3.5: Slowness-only inversion. Slowness image improvement when amplitude data is added to a travel time inversion, as a function of the number of data added. The image contains $M = 21 \times 21$ model parameters and the number of stations and the number of earthquakes both are equal to L. The travel time and amplitude noise levels are held fixed at 1 % and 0.1 %, respectively. The number of travel time data equals the number of model parameters when $L = 21$ (blue arrow). (A) Inversion using travel time data, only, for selected values of L. The true image (not shown) is indistinguishable from the $L = 100$ case. (B) Inversion using both travel time and amplitude data, only, for selected values of L. Note improvement compared to (A). (C) Percent improvement in slowness image recovery as a function of L. (D) Ray paths corresponding to the images in (A) and (B), colored according to their travel times.

Experiments Set B, concerning joint inversions for spatially-varying phase slowness and attenuation factor. This set of experiments addresses the performance of such inversions in the presence of noise and the degree to which phase slowness and attenuation factor trade off.

Experiment B.1, to assess the effect of amplitude noise on the quality of the solution. In this experiment, we hold the travel time noise constant, at 0.1 % and the perform

a sequence of geotomographic inversions in which the amplitude noise is increased from 0.1 % to 1.3 %. One inversion is performed at each noise level, with randomly-drawn earthquake source azimuths, receiver locations and noise values, and the median improvement of the estimated attenuation factor model is tabulated. In all cases, the phase slowness model is well-recovered (Fig. 3.6). However, at the higher noise levels, the error in the attenuation model can be very significant. This error grows with amplitude noise level at roughly a 1:1 rate, but with occasional much worse outliers (Fig. 3.6C). This behavior indicates that noisy amplitude data is of very little utility in a joint inversion.

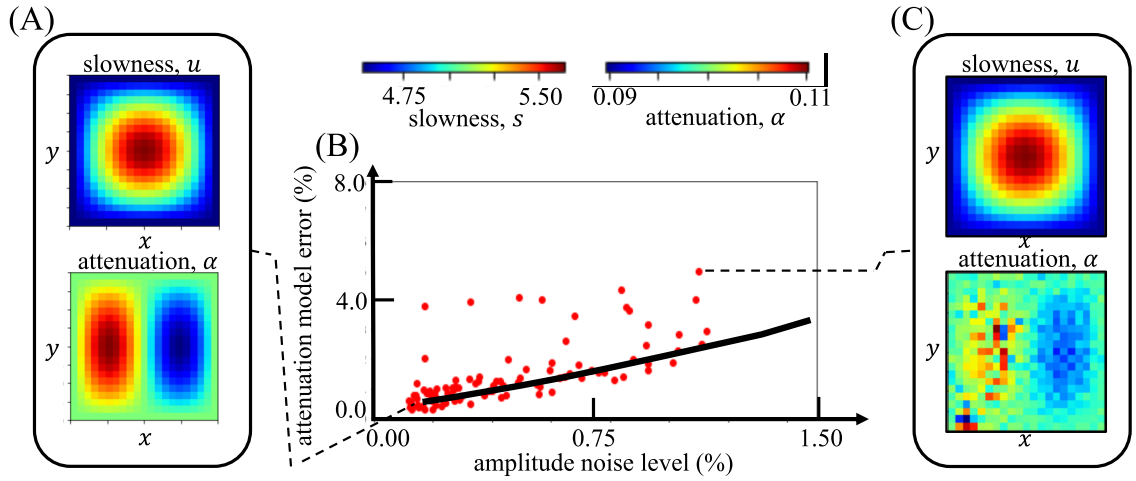

Fig. 3.6: Robustness of joint inversion for slowness and attenuation in the presence of noise. (A) Result of inversion, when the noise level of the amplitude data is 0.1 %. These images are indistinguishable from the true images (not shown). (B) Model error (red circles) as a function of amplitude noise level, with median curve (black). (C) Result of inversion, when the noise level of the amplitude data is 1.3 %.

Experiment B.2, to assess the ability of amplitude data to improve an inversion when the number of data is low. In this experiment, we hold the travel time and amplitude noise constant at 1 % and 0.1 %, respectively, and the perform a sequence of geotomographic inversions where the number of data, N, are slowly increased but the number of model parameters, M, is held constant, bracketing the $N/M = 1$ even-determined case. One inversion is performed for each case, with randomly-drawn earthquake source azimuths, receiver locations and noise values, and the error of the attenuation factor model is tabulated. The error is large (< 50 %) for cases with low number of data ($N/M < 3$), and then dramatically falls off to < 5 % above that value (Fig. 3.7). This result suggests that joint inversion is only viable for very overdetermined problems, that is, when the number of travel time and amplitude data greatly exceed the number of phase slowness and attenuation factor model parameters.

Experiment B.3, to assess how phase slowness and attenuation factor trade off in joint inversions. We employ the 'squeezing procedure' (*Lerner-Lam and Jordan*, 1987; W. *Menke*, 2018) to address this issue. A damping factor is introduced into the inversion that penalizes deviations of the attenuation factor from the background level, and this factor is successively increased, to 'squeeze' spatially-varying structure into the phase slowness (Fig. 3.8). Squeezed solution typically have higher prediction error than un-

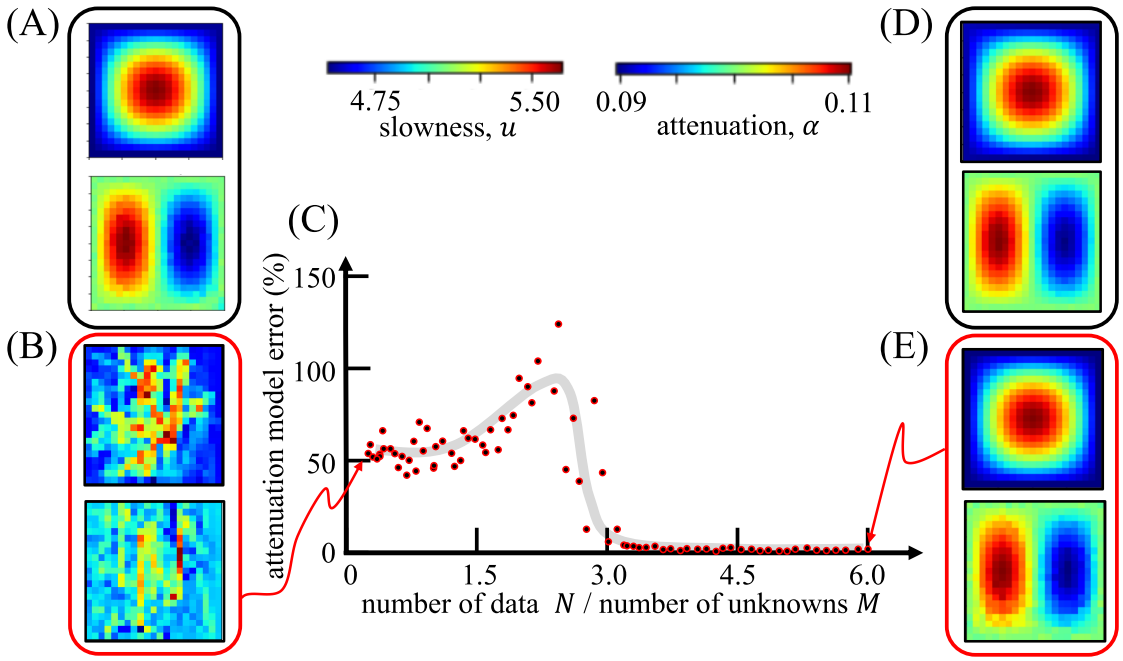

Fig. 3.7: Comparison of slowness-only and joint slowness-attenuation inversions. (A) Slowness inversion (top), using travel times only, and attenuation inversion (bottom), using amplitudes only, with N data and M model parameters, and $N/M = 0.2$. The results are almost indistinguishable from the true images (Fig. 2.1). (B) Joint slowness-attenuation inversion for $N/M = 0.2$. Note that both the slowness and attenuation images are poorly recovered. (C) Attenuation model error (red circles) for a suite if randomly generated models, as a function of N/M, with median curve (black). (D) Slowness inversion, using travel times only, and attenuation inversion, using amplitude data only, for a large number of data ($N/M = 6.0$). (E) Joint slowness-attenuation inversion for $N/M = 6.0$. Both slowness and attenuation are well-recovered.

squeezed ones. We find that almost all the spatial variability of the attenuation factor can be squeezed away, with a corresponding $3\times$ increase in travel time prediction error and $10\times$ increase in amplitude prediction error. The resulting phase slowness image is incorrect, but still spatially coherent and similar in some respects to the true one. In a real-world application, damping factors needs to be introduced to 'normalize' for the relative size of the slowness and attenuation model parameters, and their choice is typically somewhat subjective. The increase in prediction error associated with over-damping the attenuation factor is significant, but still small enough that it might be mistaken for observational error in a real-world application, leading to poor solution. Consequently, a suite of damping factors should always be investigated when performing joint inversions.

Overall, this set of experiments illustrates that joint inversions are not viable except when problem is very overdetermined (by a factor of > 3; see Fig. 3.7C) and when the amplitude noise level is very low (say < 10 %). Even then, significant trading off of the phase slowness and attenuation factor parts of the solution may occur.

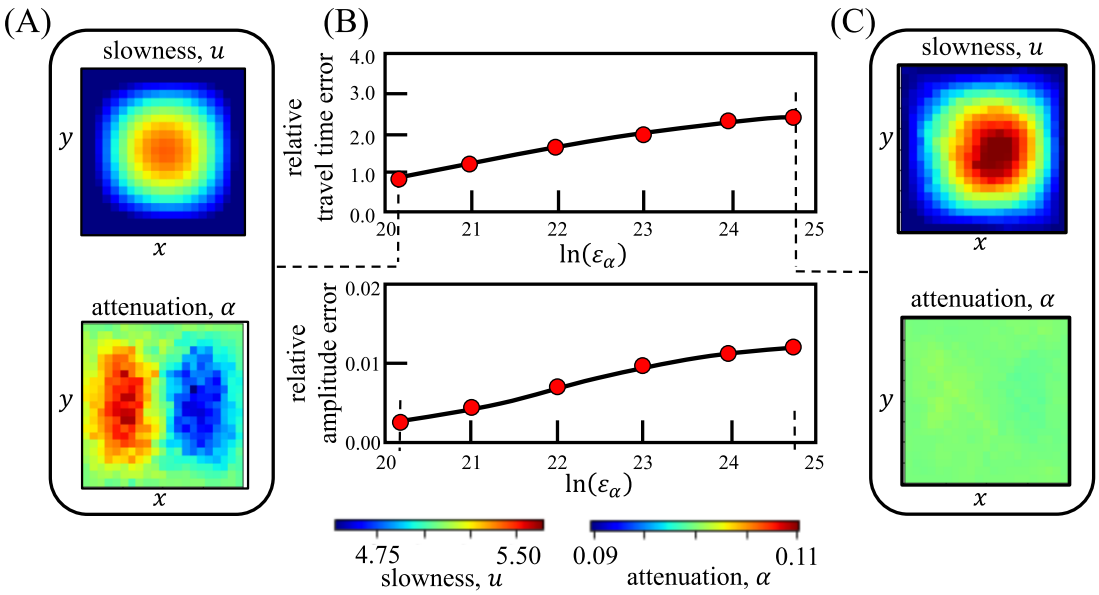

Fig. 3.8: Squeezing of attenuation in joint slowness-attenuation inversions. (A) Reference case for minimal squeezing (amplitude damping factor of $\ln \varepsilon_\alpha = 20.1$). The number of stations and the number of earthquakes both are equal to $L = 40$ ($N/M \approx 5$). The travel time and amplitude noise levels are 0.1 % and 1 %, respectively. Estimated 21×21 pixels images of slowness (top) and attenuation (bottom) are visually identical to the true images (Fig. 3.3). (B) Increase in travel time error (top) and amplitude error (bottom) as the amplitude damping factor, ε_α, is increased. (C) Estimated slowness (top) and attenuation (bottom) for strong amplitude damping of $\ln \varepsilon_\alpha = 24.7$, which results in the attenuation image having no spatial variability. Note that the slowness image is still reasonably well-recovered.

4 Discussion and conclusions

Our results indicate that the addition of amplitude data to travel time data can substantially improve determinations of Earth structure in Rayleigh wave geotomography, when phase slowness (or, equivalently phase velocity) is the parameter that is being imaged. The underlying reason is that amplitudes are sensitive to the behavior of the phase slowness perpendicular to the ray path, whereas travel times are sensitive to it behavior along the ray path. The two are complimentary. Significant improvement occurs even when the amplitude data a very noisy (1:1 signal to noise ratio), when the azimuthal pattern of ray paths is poor, and when the overall amount of data is low and the problem underdetermined.

The prospect for joint imaging of travel time and attenuation factor is not as favorable. Our experiments produce faithful images of the two only when problem is overdetermined (by a factor of 3) and at high signal-to-noise ratios ($> 10:1$). Squeezing experiments indicate that travel time and attenuation factor partially trade off, in the sense that it is possible to erroneously 'map' attenuation structure into the slowness image with only a modest increase in data misfit. These findings validate the *Dalton and Ekström* (2006b) approach to attenuation factor geotomography, in which a slowness image based exclusively on travel times is used to 'correct' an amplitude dataset for geometric spreading and local conservation of every effects, before it is used in an attenuation factor geotomography.

Acknowledgements

We thank Dallas and Michael Kaplan for their leadership for the Lamont Summer Internship program, which one of us (C.R.) attended. This work was partially supported by National Science Foundation Site: Research Experience for Undergraduates Grant OCE-2050923.

References

Aki, K. and P. G. Richards, 2009. *Quantitative Seismology*. Second Edition. Mill Valley, California, USA: University Science Books, 742. ISBN: 978-1-891389-63-4.

Bensen, G. D., M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro and Y. Yang, 2007. Processing Seismic Ambient Noise Data to Obtain Reliable Broad-band Surface Wave Dispersion Measurements. *Geophysical Journal International*, **169**, 1239–1260. DOI: 10.1111/j.1365-246X.2007.03374.x.

Bowden, D. C., V. C. Tsai and F.-C. Lin, 2017. Amplification and Attenuation Across USArray Using Ambient Noise Wavefront Tracking. *Journal of Geophysical Research: Solid Earth*. DOI: 10.1002/2017JB014804.

Cafferky, S. and B. Schmandt, 2015. Teleseismic P Wave Spectra from USArray and Implications for Upper Mantle Attenuation and Scattering. *Geochemistry, Geophysics, Geosystems*, **16**, 3343–3361. URL: https://doi.org/10.1002/2015GC005993.

Cammarano, F., S. Goes, P. Vacher and D. Giardini, 2003. Inferring Upper-Mantle Temperatures from Seismic Velocities. *Physics of the Earth and Planetary Interiors*, **139**, 147–163. DOI: 10.1016/0012-821X(95)00238-8.

Cerveny, V., 2001. *Seismic Ray Theory*. Cambridge, UK: Cambridge University Press, 722. ISBN: 978-0-511529-39-9.

Dahlen, F. A. and J. Tromp, eds., 1998. *Theoretical Global Seismology*. Princeton, NJ, USA: Princeton University Press. ISBN: 978-0-691001-24-1.

Dalton, C. A., X. Bao and Z. Ma, 2017. The Thermal Structure of Cratonic Lithosphere from Global Rayleigh Wave Attenuation. *Earth and Planetary Science Letters*, **457**, 250–262. DOI: 10.1016/j.epsl.2016.10.020.

Dalton, C. A. and G. Ekström, 2006a. Constraints on Global Maps of Phase Velocity from Surface-Wave Amplitudes. *Geophysical Journal International*, **167**, 820–826. DOI: 10.1111/j.1365-246X.2006.03118.x.

Dalton, C. A. and G. Ekström, 2006b. Global Models of Surface Wave Attenuation. *Journal of Geophysical Research*, **111**, B05317. DOI: 10.1029/2005JB003997.

Ekström, G., 2017. Short-Period Surface-Wave Phase Velocities across the Conterminous United States. *Physics of the Earth and Planetary Interiors*, **270**, 168–175. DOI: 10.1016/j.pepi.2017.07.003.

Hildebrand, F., 1962. *Advanced Calculus for Applications*. Englewood Cliffs, New Jersey, USA: Prentice Hall International, 646. ISBN: 0130111716.

Keys, R., 1981. Cubic Convolution Interpolation for Digital Image Processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **29**, 1153–1160. DOI: 10.1109/TASSP.1981.1163711.

Lee, S. van der and G. Nolet, 1997. Upper Mantle S Velocity Structure of North America. *Journal of Geophysical Research: Solid Earth*, **102**, 22815–22838. DOI: 10.1029/97JB01168.

Lerner-Lam, A. and T. H. Jordan, 1987. How Thick Are the Continents? *Journal of Geophysical Research: Solid Earth*, **92**, 14007–14026. DOI: 10.1029/JB092iB13p14007.

Li, A. and R. S. Detrick, 2006. Seismic Structure of Iceland from Rayleigh Wave Inversions and Geodynamic Implications. *Earth and Planetary Science Letters*, **241**, 901–912. DOI: 10.1016/j.epsl.2005.11.048.

Menke, W., 2018. *Geophysical Data Analysis: Discrete Inverse Theory*. Fourth. Elsevier, 350. ISBN: 978-0-128135-55-6.

Menke, W., 2020. A Connection between Geometrical Spreading and the Adjoint Field in Travel Time Tomography. *Applied Mathematics*, **11**, 84–96. DOI: 10.4236/am.2020.112009.

Menke, W. and J. Menke, 2016. *Environmental Data Analysis with MATLAB*. Second Edition. Academic Press (Elsevier), 342. ISBN: 978-0-128045-50-3.

Nettles, M. and A. M. Dziewonski, 2008. Radially Anisotropic Shear Velocity Structure of the Upper Mantle Globally and beneath North America. *Journal of Geophysical Research*, **113**, B02303. DOI: 10.1029/2006JB004819.

Nishimura, C. E. and D. W. Forsyth, 1989. The Anisotropic Structure of the Upper Mantle in the Pacific. *Geophysical Journal International*, **96**, 203–229. DOI: 10.1111/j.1365-246X.1989.tb00566.x.

Richards, P. G. and W. Menke, 1983. The Apparent Attenuation of a Scattering Medium. *Bulletin of the Seismological Society of America*, **73**, 1005–1021.

Shen, W. and M. H. Ritzwoller, 2016. Crustal and Uppermost Mantle Structure beneath the United States. *Journal of Geophysical Research: Solid Earth*. DOI: 10.1002/2016JB012887.

Synge, J. L., 1956-1957. Flux of Energy for Elastic Waves in Anisotropic Media. *Proceedings of the Royal Irish Academy: Section A, Mathematical and Physical Sciences*, **58**, 12–21.

Waldhauser, F. and W. Ellsworth, 2000. A Double Difference Earthquake Location Algorithm: Method and Application to the Northern Hayward Fault, California. *Bulletin of the Seismological Society of America*, **90**, 1353–1368.

Wang, Z. and F. A. Dahlen, 1995. Validity of Surface-Wave Ray Theory on a Laterally Heterogeneous Earth. *Geophysical Journal International*, **123**, 757–773. DOI: 10.1111/j.1365-246X.1995.tb06888.x.

West, M. and W. Menke, 2000. "Fluid-induced Changes in Shear Velocity from Surface Waves". In: *Proceedings of the Symposium on the Application of Geophysics to Engineering and Environmental Problems*. Ed. by Michael H. Powers, Ahmed B. Ibrahim and Lorraine Cramer. Environmental and Engineering Geophysical Society, 21–28.

Zha, Y., S. C. Webb, S. Wei, D. A. Wiens, D. K. Blackman, W. H. Menke, R. A. Dunn and J. A. Conder, 2014. Upper Mantle Shear Velocity Structure beneath the Eastern Lau Spreading Center from OBS Ambient Noise Tomography. *Earth and Planetary Science Letters*, **408**, 194–206. DOI: 10.1016/j.epsl.2014.09.035.

Appendix A. *Sensitivity kernels*

In ray theory, the sensitivity kernel, $g(s,s')$, relates a small change, δD in a datum, D, to a small change, δw, in a material property, w, via the ray integral

$$\delta D(s) = \int_0^s g(s,s')\delta w(s')\, \mathrm{d}s' \; . \tag{A.1}$$

Here, s is arc-length along the ray and the integral represents a path integral along the ray from its start at 0 to its end at s. Sensitivity integrals are commonly used in linearized inversions to reconstruct w from observations of D.

Appendix B. *Sensitivity kernel for travel time*

The sensitivity integral that relates travel time, t, to slowness (reciprocal velocity), u, is (e.g., *Cerveny*, 2001; *Aki and Richards*, 2009; W. *Menke*, 2020)

$$\delta t(s) = \int_0^s 1 \delta u(s') \, \mathrm{d}s' \, . \tag{B.1}$$

In this case, the sensitivity kernel is unity.

Appendix C. *Sensitivity kernel for geometrical spreading amplitude*

Here, we derive the sensitivity kernel that relates a small change in ray theoretical amplitude, A, to a small change in slowness, u. This derivation is similar to one provided by *Dahlen and Tromp* (1998), but makes different assumptions about the geometry of the measurements. In particular, slowness is considered to be a two-dimensional Cartesian field and the incident wave is assumed to be planar.

The derivation starts with the equation for ray with position, $\mathbf{x}(s)$, arc-length, s, propagating through a medium with slowness, $u(\mathbf{x})$ (*Aki and Richards*, 2009) (Fig. C.1):

$$\frac{\mathrm{d}}{\mathrm{d}s} u \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} = \nabla u, \text{ or, equivalently, } \frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}s^2} + \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} \left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} \cdot \frac{1}{u} \nabla u \right) = \frac{1}{u} \nabla u \tag{C.1}$$

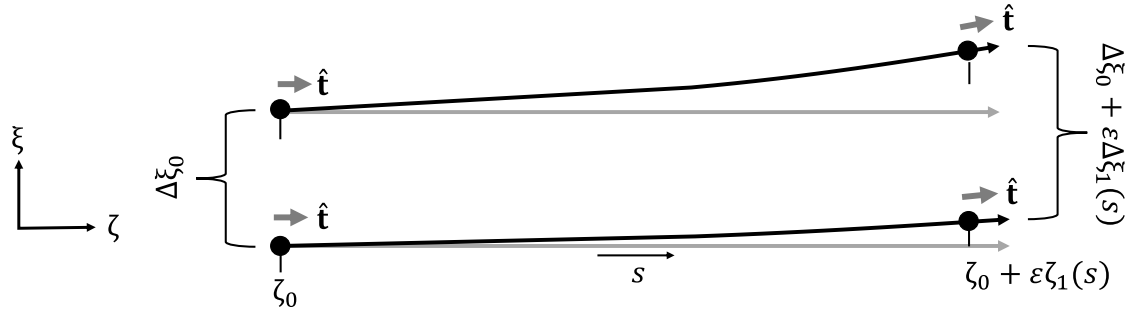

Fig. C.1: Ray geometry used to quantify the lensing effect. Two parallel rays, propagating in the ζ-direction, have an initial separation, $\Delta\xi_0$. Their ray tangents, $\hat{\mathbf{t}}$, are parallel. After propagating an arc-length, s, through a medium with a spatially-varying slowness perturbation, εu_1, where ε is a small number, the rays are no longer parallel and their separation has become $\Delta\xi_0 + \varepsilon\Delta\xi_1$.

A quantity, $f \equiv \ln u$, is defined, so that $\nabla f = u^{-1} \nabla u$. The ray equation becomes

$$\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}s^2} + \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} \left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} \cdot \nabla f \right) = \nabla f \, . \tag{C.2}$$

Later, we will assume that the plane wave is propagating in the ζ-direction, with the ray-perpendicular coordinate, ξ. The quantity $f(\zeta, \xi)$ is written as a constant background part, f_0, plus a small spatially-varying part, εf_1, where ε is a small parameter. Expanding

f_1 in a Taylor series and keeping the first three terms yields

$$f = f_0 + \varepsilon f_1 = f_0 + \varepsilon b_\xi(\zeta)\xi + \frac{1}{2}\varepsilon C_{\xi\xi}(\zeta)\xi^2 \text{ and } \nabla f = \varepsilon \begin{bmatrix} \mathrm{d}b_\zeta/\mathrm{d}\zeta + (\mathrm{d}C_{\xi\xi}/\mathrm{d}\zeta)\xi \\ b_\xi + C_{\xi\xi}\xi \end{bmatrix} . \tag{C.3}$$

Here, $b_\xi \equiv \partial f/\partial\xi$ and $C_{\xi\xi} \equiv \partial^2 f/\partial\xi^2$ are the first and second derivatives, respectively. We now treat the case of initially-parallel rays; that is, a plane wave. The position vector for a ray that is initially parallel to the ζ-direction is written as the sum of an unperturbed part (with subscript 0) and a perturbed part (with subscript 1). Because $\mathrm{d}\mathbf{x}/\mathrm{d}s$ is a unit vector that initially is $[\zeta,\xi]^\mathsf{T} = [1,0]^\mathsf{T}$, only its ξ-component can have a first order perturbation. Position along the ray, and its derivatives, are

$$\mathbf{x} = \begin{bmatrix} \zeta_0(s) \\ \xi_0 + \varepsilon\xi_1(s) \end{bmatrix} , \text{ and } \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} = \begin{bmatrix} \mathrm{d}\zeta_0/\mathrm{d}s \\ \varepsilon(\mathrm{d}\xi_1/\mathrm{d}s) \end{bmatrix} , \text{ and } \frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}s^2} = \begin{bmatrix} \mathrm{d}^2\zeta_0/\mathrm{d}s^2 \\ \varepsilon(\mathrm{d}^2\xi_1/\mathrm{d}s^2) \end{bmatrix} . \tag{C.4}$$

The second term on left-hand side of ray equation becomes

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s}\cdot\nabla f &= \begin{bmatrix} \mathrm{d}\zeta_0/\mathrm{d}s \\ \varepsilon(\mathrm{d}\xi_1/\mathrm{d}s) \end{bmatrix} \cdot \varepsilon \begin{bmatrix} \mathrm{d}b_\zeta/\mathrm{d}\zeta + (\mathrm{d}C_{\xi\xi}/\mathrm{d}\zeta)y_0 + \varepsilon(\mathrm{d}C_{\xi\xi}/\mathrm{d}x) \\ b_\xi + C_{\xi\xi}\xi_0 + \varepsilon C_{\xi\xi}\xi_1 \end{bmatrix} \\ &= \varepsilon\left(\frac{\mathrm{d}\zeta_0}{\mathrm{d}s}\right)\left(\frac{\partial b_\zeta}{\partial\zeta} + \left(\frac{\partial C_{\xi\xi}}{\partial\zeta}\right)\xi_0\right) + \mathscr{O}(\varepsilon^2) , \end{aligned} \tag{C.5}$$

from whence we find

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s}\left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s}\cdot\nabla f\right) = \begin{bmatrix} \varepsilon(\mathrm{d}\zeta_0/\mathrm{d}s)^2 \\ \mathscr{O}(\varepsilon^2) \end{bmatrix}\left(\frac{\partial b_\zeta}{\partial\zeta} + \left(\frac{\partial C_{\xi\xi}}{\partial\zeta}\xi_0\right)\right) + \mathscr{O}(\varepsilon^2) . \tag{C.6}$$

Then, the ray equation, written to first order in ε, is

$$\begin{aligned} &\begin{bmatrix} \mathrm{d}^2\zeta_0/\mathrm{d}s^2 \\ \varepsilon\left(\mathrm{d}^2\xi_1/\mathrm{d}s^2\right) \end{bmatrix} + \begin{bmatrix} \varepsilon(\mathrm{d}\zeta_0/\mathrm{d}s)^2 \\ \mathscr{O}(\varepsilon^2) \end{bmatrix}\left(\partial b_\zeta/\partial\zeta + (\partial C_{\xi\xi}/\partial\zeta)\xi_0\right) \\ &\qquad = \varepsilon \begin{bmatrix} \partial b_\zeta/\partial\zeta + (\partial C_{\xi\xi}/\partial\zeta)\xi_0 \\ b_\xi + C_{\xi\xi}\xi_0 \end{bmatrix} . \end{aligned} \tag{C.7}$$

The zeroth order ray equation,

$$\frac{\mathrm{d}^2\zeta_0}{\mathrm{d}s^2} = 0 \text{ , has solution } \zeta_0(s) = s \text{ and } \frac{\mathrm{d}\zeta_0}{\mathrm{d}s} = 1 . \tag{C.8}$$

This solution represents a plane wave propagating in the ζ-direction. First order ray equations for two rays initially at ξ_0^A and ξ_0^B, respectively, are:

$$\begin{aligned} \frac{\mathrm{d}^2\xi_1^\mathrm{A}}{\mathrm{d}s^2} &= b_\xi + C_{\xi\xi}\xi_0^\mathrm{A} \\ \frac{\mathrm{d}^2\xi_1^\mathrm{B}}{\mathrm{d}s^2} &= b_\xi + C_{\xi\xi}\xi_0^\mathrm{B} \end{aligned} \tag{C.9}$$

Subtracting Equation (C.9) for the difference, $\Delta\xi$, yields

$$\frac{\mathrm{d}^2\Delta\xi_1}{\mathrm{d}s^2} = C_{yy}\Delta\xi_0 \ , \tag{C.10}$$

with $\Delta\xi \equiv \xi^{\mathrm{B}} - \xi^{\mathrm{A}} = (\xi_0^{\mathrm{B}} - \xi_0^{\mathrm{A}}) + \varepsilon(\xi_1^{\mathrm{B}} - \xi_1^{\mathrm{A}}) \equiv \Delta\xi_0 + \varepsilon\Delta\xi_1$.

Integrating this equation twice yields:

$$\frac{\mathrm{d}\Delta\xi_1}{\mathrm{d}s} = \Delta\xi_0 \int_0^s C_{\xi\xi}(s')\ \mathrm{d}s' \tag{C.11}$$

$$\Delta\xi_1 = \Delta\xi_0 \int_0^s \left[\int_0^{s'} C_{\xi\xi}(s'')\ \mathrm{d}s''\right]\ \mathrm{d}s' \text{ or } \frac{\Delta\xi_1}{\Delta\xi_0} = \int_0^s \left[\int_0^{s'} C_{\xi\xi}(s'')\ \mathrm{d}s''\right]\ \mathrm{d}s' \tag{C.12}$$

In the next section, the iterated integral is shown to be equivalent to the single integral

$$\frac{\Delta\xi_1}{\Delta\xi_0} = \int_0^s [s-s']C_{\xi\xi}(s')\ \mathrm{d}s' \ . \tag{C.13}$$

For constant $C_{\xi\xi}$, the quantity, $\Delta\xi_1/\Delta\xi_0$, grows or shrinks with distance, depending upon the sign of $C_{\xi\xi}$. This effect corresponds to ray divergence and convergence. Because energy is confined to the ray tube, for propagation in two dimensions, the energy density, E, of the wave is proportional to the reciprocal of the cross-sectional of the ray tube:

$$\frac{E}{E_0} = \frac{\Delta\xi_0}{\Delta\xi_0 + \Delta\xi_1} = \left[1 + \frac{\Delta\xi_1}{\Delta\xi_0}\right]^{-1} \tag{C.14}$$

Because the amplitude, A, of a wave is proportional to the square root of its energy density we have

$$\frac{A}{A_0} \propto \left[1 + \frac{\Delta\xi_1}{\Delta\xi_0}\right]^{-1/2} \approx 1 - \frac{1}{2}\frac{\Delta\xi_1}{\Delta\xi_0} \ . \tag{C.15}$$

Writing $A = A_0 + \Delta A$, we find:

$$\frac{\Delta A}{A_0} \approx -\frac{1}{2}\frac{\Delta\xi_1}{\Delta\xi_0} \tag{C.16}$$

Equations (C.13) and (C.16) can be combined into a sensitivity integral relating a perturbation, δu, in slowness to a perturbation, $\delta(\Delta A/A_0)$, in amplitude:

$$\delta a(s) \equiv \delta\left(\frac{\Delta A(s)}{A_0}\right) = \int_0^s \left\{-\frac{1}{2}[s-s']\frac{\partial^2}{\partial\xi^2}\right\}\delta u(s')\ \mathrm{d}s' \tag{C.17}$$

Here we have abbreviated $\Delta A(s)/A_0$ as $a(s)$. Consequently, the kernel is

$$g(s,s') = -\frac{1}{2}[s-s']\frac{\partial^2}{\partial\xi^2} \ . \tag{C.18}$$

Appendix D. *Identity involving an iterated integral*

The proof that

$$\int_0^s \left[\int_0^{s'} f(s'')\, \mathrm{d}s'' \right] \mathrm{d}s' = \int_0^s [s - s'] f(s')\, \mathrm{d}s' \tag{D.1}$$

makes use of the Heaviside step function $H(s_2 - s_1)$, defined to be unity when $s_2 > s_1$ and zero otherwise. This allows the top limit in the inner integral to be increased to s:

$$I(s) \equiv \int_0^s \left[\int_0^{s'} f(s'')\, \mathrm{d}s'' \right] \mathrm{d}s' = \int_0^s \left[\int_0^s H(s' - s'') f(s'')\, \mathrm{d}s'' \right] \mathrm{d}s' \tag{D.2}$$

Then, the order of the arguments in the Heaviside function are reversed, using the identity $H(s' - s'') = 1 - H(s'' - s')$.

$$I(s) = \int_0^s \left[\int_0^s \left\{ 1 - H(s'' - s') \right\} f(s'')\, \mathrm{d}s'' \right] \mathrm{d}s' \tag{D.3}$$

Then, the order of integration is swapped:

$$I(s) = \int_0^s \left[\int_0^s \left\{ 1 - H(s'' - s') \right\} \mathrm{d}s' \right] f(s'')\, \mathrm{d}s'' \tag{D.4}$$

The inner integral can now be performed analytically, yielding

$$I(s) = \int_0^s (s - s'') f(s'')\, \mathrm{d}s'' \ . \tag{D.5}$$

Appendix E. *Relationship between second derivatives of f and u*

Tomography requires the derivative $C_{yy} = \partial^2 f / \partial y^2$, where $f = \ln(u)$ is approximated as a linear operator acting on slowness, u. We use the approximation $\ln(1+t) \approx t$, valid when $|t| \ll 1$ to develop such a relationship. Suppose that $u = u_0 + \Delta u$, where u_0 is a constant, average value, and Δu is spatially varying and small compared to u_0. Then:

$$f = \ln(u_0 + \Delta u) = \ln \left[u_0 \left(1 + \frac{\Delta u}{u_0} \right) \right] = \ln(u) + \ln \left(1 + \frac{\Delta u}{u_0} \right) \approx \ln(u) + \frac{\Delta u}{u_0} \tag{E.1}$$

Here we have used the identity $\ln(ab) = \ln(a) + \ln(b)$. Because only Δu is spatially variable,

$$\frac{\partial^2 f}{\partial \xi^2} = \frac{1}{u_0} \frac{\partial^2 \Delta u}{\partial \xi^2} = \frac{1}{u_0} \frac{\partial^2 u}{\partial \xi^2} \ . \tag{E.2}$$

A numerical test (not shown) indicates that this formula is reasonably accurate for the tomographic application developed here.

Appendix F. Rotation of second derivative to ray coordinates

In practice, the symmetric matrix, $\mathbf{D}$, of second derivates is first computed:

$$\mathbf{D} = \begin{bmatrix} \dfrac{\partial^2 u}{\partial x^2} & \dfrac{\partial^2 u}{\partial x \partial y} \\ \dfrac{\partial^2 u}{\partial x \partial y} & \dfrac{\partial^2 u}{\partial y^2} \end{bmatrix} \tag{F.1}$$

Then, the derivative, $\mathrm{d}^2u/\mathrm{d}\xi^2$, that appears in the sensitivity integral is inferred from $\mathbf{D}$. Suppose that the ray direction, ζ, is oriented an angle, θ, counter-clockwise with respect to the x-axis. In the special case of $\theta = 0$, we find $\partial^2 u/\partial \xi^2 = \partial^2 u/\partial y^2 = D_{22}$. In the general case of $\theta \neq 0$, the matrix $\mathbf{D}$ needs to be rotated to a new primed coordinate system:

$$\mathbf{D}' = \mathbf{R}\mathbf{D}\mathbf{R}^{\mathsf{T}} \text{ with } \mathbf{R} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{F.2}$$

The second derivative in the ξ direction is identified as D'_{22}, or:

$$\frac{\mathrm{d}^2 u}{\mathrm{d}\xi^2} = \sin^2\theta \frac{\mathrm{d}^2 u}{\mathrm{d}x^2} + 2\sin\theta\cos\theta \frac{\mathrm{d}^2 u}{\mathrm{d}x\,\mathrm{d}y} + \cos^2\theta \frac{\mathrm{d}^2 u}{\mathrm{d}y^2} \tag{F.3}$$

Appendix G. Sensitivity kernel for intrinsic attenuation

Rock friction causes intrinsic attenuation; that is, the loss amplitude of a wave with distance travelled. The decrease in amplitude is calculated as (*Aki and Richards*, 2009)

$$\ln\left(\frac{A(s)}{A_0}\right) = \int_0^s (-\alpha)\, \mathrm{d}s' \,. \tag{G.1}$$

The attenuation factor, $\alpha(x,y)$, is a material property that quantifies the rate of loss of amplitude with distance traveled. It is related to the quality factor, Q, and angular frequency, ω, of the wave, by $\alpha = (\frac{1}{2}\omega u_0)/Q$. In many instances, the energy loss is small, so that

$$\ln\left(\frac{A}{A_0}\right) = \ln\left(\frac{A_0 + \Delta A}{A_0}\right) = \ln\left(1 + \frac{\Delta A}{A_0}\right) \approx \frac{\Delta A}{A_0} \,. \tag{G.2}$$

Here, we used the approximation $\ln(1+t) \approx t$, valid when $|t| \ll 1$. The sensitivity equation is then

$$\delta a(s) = \int_0^s (-1)\delta a\, \mathrm{d}s' \,, \tag{G.3}$$

and the sensitivity kernel is (-1). Although they act on different material properties, the sensitivity kernels for travel time and amplitude loss by rock friction differ only by a sign.

Appendix H. *Interpolation scheme and sensitivity derivatives for control points*

Slowness, u, and attenuation, α, are represented on $M_x \times M_y$ Cartesian grids of control points, $u_{ij} = u(x_i, y_j)$ and $\alpha_{ij} = \alpha(x_i, y_j)$, with uniform spacing, Δx and Δy. A convolutional form of cubic interpolation, due to *Keys* (1981), is used to estimate the slowness and attenuation, u^{est} and α^{est}, at an arbitrary point (x, y) within the grid. Slowness interpolation is treated here; attenuation interpolation is analogous. The process begins with the identification of a 4×4 grid of control points, such that (x, y) lies within its central square. The control points in this 4×4 grid are denoted $u_{-1,-1}$ through $u_{2,2}$, and comprise the 4×4 matrix, $\mathbf{U}$. The coordinate system is then shifted and rescaled, so that the central square covers the intervals $0 \le x \le 1$ and $0 \le y \le 1$ and so that $\Delta\hat{x} = \Delta\hat{y} = 1$.

A vector, $\mathbf{b}$, is then formed by applying an interpolant to the rows of $\mathbf{U}$:

$$\mathbf{b}^T = \mathbf{f}^T(\hat{x})\mathbf{M}\mathbf{U} \text{ with } \mathbf{f}(\hat{x}) = \frac{1}{2}[1, \hat{x}, \hat{x}^2, \hat{x}^3]^T \tag{H.1}$$

$$\text{and } \mathbf{M} = \begin{bmatrix} 0 & 2 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 2 & -5 & 4 & -1 \\ -1 & 3 & -3 & 1 \end{bmatrix} \text{ and } \mathbf{U} = \begin{bmatrix} u_{-1,-1} & u_{-1,0} & u_{-1,1} & u_{-1,2} \\ u_{0,-1} & u_{0,0} & u_{0,1} & u_{0,2} \\ u_{1,-1} & u_{1,0} & u_{1,1} & u_{1,2} \\ u_{2,-1} & u_{2,0} & u_{2,1} & u_{2,2} \end{bmatrix} .$$

This process creates estimates of slowness at $(\hat{x}, \hat{y}_j)$, where $\hat{y}_j$ are the $\hat{y}$-coordinates of the control points. Then, a similar interpolant is applied to the elements of $\mathbf{b}$:

$$u^{est}(\hat{x}, \hat{y}) = \mathbf{f}^T(\hat{y})\mathbf{M}\mathbf{b} = \mathbf{f}^T(\hat{x})\mathbf{M}\mathbf{U}\mathbf{M}^T\mathbf{f}(\hat{y}) \text{ with } \mathbf{f}(\hat{y}) = \frac{1}{2}[1, \hat{y}, \hat{y}^2, \hat{y}^3]^T \tag{H.2}$$

This process creates an estimate of slowness, $u^{est}(\hat{x}, \hat{y})$. Issues arise when $\mathbf{U}$ must be extracted from near the edges of the overall u_{ij} grid, because then one of its rows or columns (or both) is undefined. We fill them in via linearly interpolation, using the two adjacent rows or columns (or both) of $\mathbf{U}$. This procedure forces the second derivative of $u^{est}(\hat{x}, \hat{y})$ to be close to zero at the boundaries of the overall grid.

Spatial derivatives can be formed by analytic differentiation of the interpolation formula:

$$\frac{\partial u^{est}}{\partial x} = (\Delta x)\dot{\mathbf{f}}^T(\hat{x})\mathbf{M}\mathbf{U}\mathbf{M}^T\mathbf{f}(\hat{y}) \text{ and } \frac{\partial u^{est}}{\partial y} = (\Delta y)\mathbf{f}^T(\hat{x})\mathbf{M}\mathbf{U}\mathbf{M}^T\dot{\mathbf{f}}(\hat{y}) \tag{H.3}$$

$$\frac{\partial^2 u^{est}}{\partial x^2} = (\Delta x)^2\ddot{\mathbf{f}}^T(\hat{x})\mathbf{M}\mathbf{U}\mathbf{M}^T\mathbf{f}(\hat{y}) \text{ and } \frac{\partial^2 u^{est}}{\partial y^2} = (\Delta y)^2\mathbf{f}^T(\hat{x})\mathbf{M}\mathbf{U}\mathbf{M}^T\ddot{\mathbf{f}}(\hat{y})$$

$$\text{and } \frac{\partial^2 u^{est}}{\partial x \partial y} = (\Delta x \Delta y)\dot{\mathbf{f}}^T(\hat{x})\mathbf{M}\mathbf{U}\mathbf{M}^T\dot{\mathbf{f}}(\hat{y})$$

$$\text{with } \dot{\mathbf{f}}(\hat{x}) = \frac{1}{2}[0, 1, 2\hat{x}, 3\hat{x}^2]^T \text{ and } \dot{\mathbf{f}}(\hat{y}) = \frac{1}{2}[0, 1, 2\hat{y}, 3\hat{y}^2]^T$$

$$\text{and } \ddot{\mathbf{f}}(\hat{x}) = \frac{1}{2}[0, 0, 2, 6\hat{x}]^T \text{ and } \ddot{\mathbf{f}}(\hat{y}) = \frac{1}{2}[0, 0, 2, 6\hat{y}]^T .$$

The factors of Δx and Δy are introduced to account for the scaling. The interpolation of attenuation is completely analogous; the matrix, $\mathbf{U}$, of slowness control points merely is replaced by an analogous matrix, say $\mathbf{A}$, of attenuation control points.

In travel time tomography, the grids, u_{ij} and α_{ij}, are reconstructed from measurements that depend upon u^{est} and α^{est}, a process that requires the sensitivity derivative $\partial u^{est}/\partial u_{ij}$ and $\partial \alpha^{est}/\partial \alpha_{ij}$. These derivatives are zero for control points outside the 4×4 grid. For control points within the 4×4 grid, they can be determined by direct differentiation of the interpolation formula

$$\frac{\partial u^{est}}{\partial u_{ij}} = \frac{\partial \alpha^{est}}{\partial \alpha_{ij}} = \mathbf{f}^{\mathsf{T}}(\hat{x})\mathbf{M}\mathbf{P}^{(ij)}\mathbf{M}^{\mathsf{T}}\mathbf{f}(\hat{y}) \ , \text{ with } \mathbf{P}^{(ij)} \equiv \frac{\partial \mathbf{U}}{\partial u_{ij}} \ . \tag{H.4}$$

The 4×4 matrix $\mathbf{P}^{(ij)}$ is zero, except for its (i,j) element, which is unity; that is $[\mathbf{P}^{(ij)}]_{pq} = \delta_{ip}\delta_{jq}$, where δ_{ij} is the Kronecker delta function. Thus, the sensitivity derivative is just the interpolant evaluated at $(\hat{x},\hat{y})$, but with $\mathbf{U}$ (or $\mathbf{A}$) replaced with $\mathbf{P}^{(ij)}$. However, as $\partial u^{est}/\partial u_{ij}$ and $\partial \alpha^{est}/\partial \alpha_{ij}$ are 44 matrices, the interpolant must be evaluated sixteen times, once for each combination of i and j. Furthermore, cases where $\mathbf{U}$ and $\mathbf{A}$ are at the edge of the overall grid must be handled specially, because these matrices then contain fewer than 16 independent elements.

The sensitivity of the second derivatives, which are needed for amplitude tomography, are computed in an analogous fashion:

$$\begin{aligned}
\frac{\partial}{\partial u_{ij}}\left(\frac{\partial^2 u^{est}}{\partial x^2}\right) = \frac{\partial}{\partial \alpha_{ij}}\left(\frac{\partial^2 \alpha^{est}}{\partial x^2}\right) &= (\Delta x)^2 \ddot{\mathbf{f}}^{\mathsf{T}}(\hat{x})\mathbf{M}\mathbf{P}^{(ij)}\mathbf{M}^{\mathsf{T}}\mathbf{f}(\hat{y}) \\
\frac{\partial}{\partial u_{ij}}\left(\frac{\partial^2 u^{est}}{\partial y^2}\right) = \frac{\partial}{\partial \alpha_{ij}}\left(\frac{\partial^2 \alpha^{est}}{\partial y^2}\right) &= (\Delta y)^2 \mathbf{f}^{\mathsf{T}}(\hat{x})\mathbf{M}\mathbf{P}^{(ij)}\mathbf{M}^{\mathsf{T}}\ddot{\mathbf{f}}(\hat{y}) \\
\frac{\partial}{\partial u_{ij}}\left(\frac{\partial^2 u^{est}}{\partial x \partial y}\right) = \frac{\partial}{\partial \alpha_{ij}}\left(\frac{\partial^2 \alpha^{est}}{\partial x \partial y}\right) &= (\Delta x \Delta y)\dot{\mathbf{f}}^{\mathsf{T}}(\hat{x})\mathbf{M}\mathbf{P}^{(ij)}\mathbf{M}^{\mathsf{T}}\dot{\mathbf{f}}(\hat{y})
\end{aligned} \tag{H.5}$$

Appendix I. *Changes in amplitude due to conservation of energy*

The energy flux, F_i, (rate of energy transport per unit area) of an elastic wave in the direction, i, is (*Synge*, 1956-1957)

$$F_i = -\tau_{ij}\dot{u}_j \ . \tag{I.1}$$

The flux conserves the local energy density, E, in the sense that $\dot{E} = -F_{i,\,i}$.

Although our goal is to study the Rayleigh surface wave, we first consider a horizontally propagating shear body wave, in order to provide a point of comparison. This shear wave has displacement, $u = [0,0,A]^{\mathsf{T}}\cos(\omega x/b - \omega t)$, where ω is angular frequency and

b is shear velocity. The non-zero components of stress and velocity are

$$\begin{aligned} \tau_{xz} &= \rho b^2 \left(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{z} \right) = -\omega \rho b A \sin(\omega x/b - \omega t) \\ \dot{u}_z &= \omega A \sin(\omega x/b - \omega t) \ , \end{aligned} \tag{I.2}$$

from whence the energy flux, F_x, and its time average, $\langle F_x \rangle$, are

$$\begin{aligned} F_x &= -\tau_{xz} \dot{u}_x = \omega^2 \rho b A^2 \sin^2(\omega x/b - \omega t) \\ \langle F_x \rangle &= -\tau_{xz} \dot{u}_x = \frac{1}{2} \omega^2 \rho b A^2 \ . \end{aligned} \tag{I.3}$$

The time-averaged flux is $\langle F_x \rangle = -\tau_{xz} \dot{u}_x = \frac{1}{2} \omega^2 \rho b A^2$. Unlike a surface wave, the shear wave as a planar wavefront with no depth, z, dependence. Nevertheless, we are free to consider the total flux over some depth interval, $(0, L)$, as

$$\langle F_x \rangle_{\mathrm{T}} = \int_0^L \langle F_x \rangle \, \mathrm{d}z = \frac{1}{2} \omega^2 \rho b A^2 L \ . \tag{I.4}$$

When the shear wave propagates from a medium with one shear velocity of a medium with another, and neglecting reflections at the boundary, conservation of energy implies that $\langle F_x \rangle_{\mathrm{T}}$ is constant:

$$\mathrm{d}\langle F_x \rangle_{\mathrm{T}} = 0 = \frac{1}{2} \omega^2 \rho A^2 L \, \mathrm{d}b + \frac{1}{2} \omega^2 \rho b L \cdot 2A \, \mathrm{d}A \ , \text{ or, } \frac{\mathrm{d}A}{A} = -\frac{1}{2} \frac{\mathrm{d}b}{b} \ . \tag{I.5}$$

Consequently, the amplitude decreases as the shear velocity increases, with a proportionality factor of $1/2$.

We now perform a similar calculation for the Rayleigh wave, restricting our analysis to the case of a homogeneous halfspace with shear velocity, b, and compressional velocity, $a = rb$, where r is a ratio. The calculation is more involved, because the way in which the depth distribution of energy changes with velocity needs to be taken into account. According to *Aki and Richards* (2009, Equation 5.56), the phase velocity of the Rayleigh wave, $V_r = 1/u$, where the slowness, u, is given by the zero of the function, $R(u)$:

$$R(u) = \left(\frac{1}{b^2} - 2u^2 \right)^2 - 4u^2 \xi \eta \ , \text{ with } \xi \equiv \left(u^2 - \frac{1}{a^2} \right)^{1/2} \text{ and } \eta = \left(u^2 - \frac{1}{b^2} \right)^{1/2} \tag{I.6}$$

Note that we can rewrite this condition as $f(z) = 0$, with $z = bu$ and

$$f(z) = \left(\frac{1}{(bu)^2} - 2 \right)^2 - 4 \left(1 - \frac{1}{(bu)^2} \right)^{1/2} \left(1 - \frac{1}{r^2 (bu)^2} \right)^{1/2} \ . \tag{I.7}$$

The derivative, $\mathrm{d}u/\mathrm{d}b$, can be calculated using the implicit function theorem (e.g., *Hilde-*

brand, 1962):

$$\frac{\mathrm{d}f}{\mathrm{d}u} = \frac{\mathrm{d}f}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}u} = \frac{\mathrm{d}f}{\mathrm{d}z}b \text{ and } \frac{\mathrm{d}f}{\mathrm{d}b} = \frac{\mathrm{d}f}{\mathrm{d}z}\frac{\mathrm{d}z}{\mathrm{d}b} = \frac{\mathrm{d}f}{\mathrm{d}z}u, \text{ so } \frac{\mathrm{d}u}{\mathrm{d}b} = -\frac{\mathrm{d}f}{\mathrm{d}b}\Big/\frac{\mathrm{d}f}{\mathrm{d}u} = -\frac{u}{b} \tag{I.8}$$

Furthermore, the derivative of any function of the form, $g(bu)$, is zero:

$$\frac{\mathrm{d}}{\mathrm{d}b}g(bu) = g'(bu)\left[u + b\frac{\mathrm{d}u}{\mathrm{d}b}\right] = 0 \tag{I.9}$$

Consequently,

$$\frac{\mathrm{d}}{\mathrm{d}b}(au) = \frac{\mathrm{d}}{\mathrm{d}b}(bu) = \frac{\mathrm{d}}{\mathrm{d}b}\left(\frac{u}{\eta}\right) = \frac{\mathrm{d}}{\mathrm{d}b}\left(\frac{u}{\xi}\right) = 0\ . \tag{I.10}$$

The equation, $R(u) = 0$, can be solved numerically for u via Newton's method, using the analytic derivative:

$$\frac{\mathrm{d}R}{\mathrm{d}u} = -8p\left(\frac{1}{b^2} - 2u^2\right) - 8u\eta\xi - 4u^3\eta^{-1}\xi - 4u^3\eta\xi^{-1} \tag{I.11}$$

This expression has been checked numerically, and gives $V_r = \gamma b$ with $\gamma \approx 0.92$ (for $r = 1.8$).

One of the consequences of Eqn. (I.10) is that the depth decay rates, say, $r_a = \omega\xi = \omega u(\xi/u)$ and $r_b = \eta\omega = \omega u(\eta/u)$ have derivatives $\partial r_a/\partial u = \omega(\xi/u)$ and $\partial r_b/\partial u = \omega(\eta/u)$ that are positive constants. The depth decay rates increase with phase slowness (and decreases with phase velocity). Consequently, as the shear velocity is decreased, Rayleigh wave energy is squeezed towards the Earth's surface.

According to *Aki and Richards* (2009, Equations 5.52 and 5.53), the displacement of the Rayleigh wave is the sum of contributions from evanescent compressional and shear waves:

$$P\begin{bmatrix} au \\ i\xi a \end{bmatrix}\exp(-\omega\xi z)\exp(i\omega ux) + S\begin{bmatrix} i\eta b \\ -bu \end{bmatrix}\exp(-\omega\eta z)\exp(i\omega ux) \tag{I.12}$$

Here, ω is angular frequency and P and S are the amplitudes of the compressional and shear evanescent waves, respectively. We now eliminate S from this equation. According to *Aki and Richards* (2009, Equation 5.55 and 5.56), the conditions that the shear traction and normal tractions be zero on the free surface at $z = 0$ imply:

$$-2iuab\xi P = (1 - 2b^2u^2)S \text{ and } (1 - 2b^2u^2)P = 2i\left(\frac{b^3u}{a}\right)\eta S \tag{I.13}$$

Algebraic manipulation of the second equation yields an equation for $ib\eta S$:

$$ib\eta S = \frac{1}{2}\left(\frac{a}{b^2u}\right)(1 - 2b^2u^2)P = \frac{1}{2}\left(\frac{a}{u}\right)\left(1/b^2 - 2u^2\right) \tag{I.14}$$

Another algebraic manipulation of the second equation yields an equation for $-buS$:

$$-buS = \frac{i}{2}\left(\frac{a}{b^2}\right)\frac{(1-2b^2u^2)}{\eta}P = \frac{ia}{2}\frac{\left(1/b^2-2u^2\right)}{\eta} \tag{I.15}$$

Hence, P and S in the displacement equation can be eliminated in favor of a single amplitude, A:

$$\begin{bmatrix} U_1 \\ U_3 \end{bmatrix} \exp(i\omega ux) = A\left\{ \begin{bmatrix} u_x^{\mathrm{P}} \\ u_z^{\mathrm{P}} \end{bmatrix} Z_{\mathrm{P}} + \begin{bmatrix} u_x^{\mathrm{S}} \\ u_z^{\mathrm{S}} \end{bmatrix} Z_{\mathrm{S}} \right\} \exp(i\omega px) \ , \tag{I.16}$$

with

$$\begin{aligned} Z_{\mathrm{P}} = \exp(-\omega\xi z) \text{ and } Z_{\mathrm{S}} = \exp(-\omega\eta z) \\ u_x^{\mathrm{P}} = u_x^{\mathrm{PR}} = au \text{ and } u_x^{\mathrm{S}} = u_x^{\mathrm{SR}} = \frac{1}{2}\left(\frac{a}{u}\right)\left(1/b^2-2u^2\right) = \frac{1}{2}(au)\left(\frac{1}{u^2b^2}-2\right) \\ u_z^{\mathrm{P}} = iu_z^{\mathrm{PI}} = i\xi a = i\left(\frac{\xi}{u}\right)au \text{ and } u_z^{\mathrm{S}} = iu_z^{\mathrm{SI}} = \frac{ia}{2}\frac{\left(1/b^2-2u^2\right)}{\eta} = \frac{iau}{2}\left(\frac{1}{u^2b^2}-2\right)\frac{u}{\eta} \ . \end{aligned} \tag{I.17}$$

These formulas have been checked numerically by verifying that they imply zero stress on the free surface. Note that the derivatives, $\mathrm{d}u_x^{\mathrm{P}}/\mathrm{d}b = \mathrm{d}u_z^{\mathrm{P}}/\mathrm{d}b = \mathrm{d}u_x^{\mathrm{S}}/\mathrm{d}b = \mathrm{d}u_z^{\mathrm{S}}/\mathrm{d}b = 0$. Consequently, for an observer at $z = 0$, no distinction needs to me made between $A^{-1}\,\mathrm{d}A/\mathrm{d}b$, $U_x^{-1}\,\mathrm{d}U_x/\mathrm{d}b$ and $U_z^{-1}\,\mathrm{d}U_z/\mathrm{d}b$; they are equal.

According to *Synge* (1956-1957), for a real displacement, the positive and negative frequency components of an elastic wave are for from the complex conjugate pairs:

$$u_i = U_i\exp(-i\omega t) + \overline{U_i}\exp(+i\omega t) = 2U_i^{\mathrm{R}}\cos(\omega t) + 2U_i^{\mathrm{I}}\sin(\omega t) \tag{I.18}$$

Then, the time-averaged energy flux, $\langle F_1\rangle$, of an elastic wave in the x-direction is

$$\begin{aligned} \frac{\langle F_1\rangle}{-2\omega} &= \rho a^2\left(U_{1,1}^{\mathrm{R}}U_1^{\mathrm{I}} - U_{1,1}^{\mathrm{I}}U_1^{\mathrm{R}}\right) \\ &\quad + \rho b^2\left(U_{3,3}^{\mathrm{R}}U_1^{\mathrm{I}} + U_{1,3}^{\mathrm{R}}U_3^{\mathrm{I}} + U_{3,1}^{\mathrm{R}}U_3^{\mathrm{I}} - U_{,3}^{\mathrm{I}}U_1^{\mathrm{R}} - U_{1,3}^{\mathrm{I}}U_3^{\mathrm{R}} - U_{3,1}^{\mathrm{I}}U_3^{\mathrm{R}}\right) \ . \end{aligned} \tag{I.19}$$

Here, the superscripts R and I refer to real and imaginary parts. Now suppose

$$\begin{aligned} U_1 &= Au_x^{\mathrm{P}}Z_{\mathrm{P}}\exp(i\omega px) + Au_x^{\mathrm{S}}Z_{\mathrm{S}}\exp(i\omega px) \\ U_3 &= Au_z^{\mathrm{P}}Z_{\mathrm{P}}\exp(i\omega px) + Au_z^{\mathrm{S}}Z_{\mathrm{S}}\exp(i\omega px) \ . \end{aligned} \tag{I.20}$$

Noting that

$$\frac{\partial}{\partial x}\exp(i\omega ux) = i\omega u\exp(i\omega ux) \text{ and } \frac{\partial}{\partial z}Z_{\mathrm{P}} = -\omega\xi Z_{\mathrm{P}} \text{ and } \frac{\partial}{\partial z}Z_{\mathrm{S}} = -\omega\eta Z_{\mathrm{S}} \ , \tag{I.21}$$

and evaluating the functions at $x = 0$, we find:

$$\begin{aligned}
U_1\big|_{x=0} &= Au_x^{\mathrm{PR}} Z_{\mathrm{P}} + Au_x^{\mathrm{SR}} Z_{\mathrm{S}} \text{ (which is real)} \\
U_3\big|_{x=0} &= iAu_z^{\mathrm{PI}} Z_{\mathrm{P}} + iAu_z^{\mathrm{SI}} Z_{\mathrm{S}} \text{ (which is imaginary)} \\
U_{1,1}\big|_{x=0} &= i\omega uAu_x^{\mathrm{PR}} Z_{\mathrm{P}} + i\omega uAu_x^{\mathrm{SR}} Z_{\mathrm{S}} \text{ (which is imaginary)} \\
U_{3,1}\big|_{x=0} &= -\omega uAu_z^{\mathrm{PI}} Z_{\mathrm{P}} - \omega uAu_z^{\mathrm{SI}} Z_{\mathrm{S}} \text{ (which is real)} \\
U_{1,3}\big|_{x=0} &= -\omega \xi Au_x^{\mathrm{PR}} Z_{\mathrm{P}} - \omega \eta Au_x^{\mathrm{SR}} Z_{\mathrm{S}} \text{ (which is real)} \\
U_{3,3}\big|_{x=0} &= -i\omega \xi Au_z^{\mathrm{PI}} Z_{\mathrm{P}} - i\omega \eta Au_z^{\mathrm{SI}} Z_{\mathrm{S}} \text{ (which is imaginary)}
\end{aligned} \tag{I.22}$$

These formulas have been checked numerically. The flux simplifies to

$$\langle F_x \rangle \equiv \langle F_x \rangle_{\mathrm{A}} + \langle F_x \rangle_{\mathrm{B}} \text{ , with} \tag{I.23}$$

$$\begin{aligned}
\langle F_x \rangle_{\mathrm{A}} &= -2\omega\rho a^2 \left(-U_{1,1}^{\mathrm{I}} U_1^{\mathrm{R}}\right) \\
&= 2\omega^2 \rho a^2 A^2 u \left[\left(u_x^{\mathrm{PR}}\right)^2 Z_{\mathrm{P}}^2 + \left(u_x^{\mathrm{SR}}\right)^2 Z_{\mathrm{S}}^2 + 2u_x^{\mathrm{PR}} u_x^{\mathrm{SR}} Z_{\mathrm{P}} Z_{\mathrm{S}} \right] \\
\langle F_x \rangle_{\mathrm{B}} &= -2\omega\rho b^2 \left(U_{1,3}^{\mathrm{R}} U_3^{\mathrm{I}} + U_{3,1}^{\mathrm{R}} U_3^{\mathrm{I}} - U_{3,3}^{\mathrm{I}} U_1^{\mathrm{R}}\right) \\
&= 2\omega^2 \rho b^2 A^2 \Big\{ (\xi - \eta) u_x^{\mathrm{PR}} u_z^{\mathrm{SI}} Z_{\mathrm{P}} Z_{\mathrm{S}} + (\eta - \xi) u_z^{\mathrm{PI}} u_x^{\mathrm{SR}} Z_{\mathrm{P}} Z_{\mathrm{S}} \\
&\quad + pu_z^{\mathrm{PI}} u_z^{\mathrm{PI}} Z_{\mathrm{P}} Z_{\mathrm{P}} + 2uu_z^{\mathrm{PI}} u_z^{\mathrm{SI}} Z_{\mathrm{P}} Z_{\mathrm{S}} + pu_z^{\mathrm{SI}} u_z^{\mathrm{SI}} Z_{\mathrm{S}} Z_{\mathrm{S}} \Big\} .
\end{aligned}$$

These formulas have been checked numerically. The vertically integrated horizontal flux, $\langle F_x \rangle_{\mathrm{T}}$, is now defined in the limit, $L \to \infty$. As only the Z's are functions of z, we find:

$$\begin{aligned}
\int_0^\infty Z_{\mathrm{P}}^2 \, \mathrm{d}z &= \int_0^\infty \exp(-2\omega\xi z) \, \mathrm{d}z = \frac{\exp(-2\omega\xi z)}{-2\omega\xi}\bigg|_0^\infty = \frac{1}{2\omega\xi} \\
\int_0^\infty Z_{\mathrm{S}}^2 \, \mathrm{d}z &= \frac{1}{2\omega\eta} \text{ and} \\
\int_0^\infty Z_{\mathrm{P}} Z_{\mathrm{S}} \, \mathrm{d}z &= \frac{1}{\omega(\xi + \eta)}
\end{aligned} \tag{I.24}$$

Defining $\langle F_x \rangle_{\mathrm{T}} \equiv \omega\rho b^2 A^2 C$ and $C \equiv (C_{\mathrm{A}} + C_{\mathrm{B}})$, we find:

$$\begin{aligned}
C_{\mathrm{A}} &= (au)^2 \left[\frac{u}{\xi} + \frac{1}{4} \left(\frac{1}{u^2 b^2} - 2 \right)^2 \frac{u}{\eta} + 2 \left(\frac{1}{u^2 b^2} - 2 \right) \frac{u}{(\xi + \eta)} \right] \\
C_B &= (au)^2 \Bigg\{ \left(\frac{1}{u^2 b^2} - 2 \right) \left(\frac{u}{\eta} \right) \frac{(\xi - \eta)}{(\xi + \eta)} + \left(\frac{\eta - \xi}{\eta + \xi} \right) \left(\frac{1}{u^2 b^2} - 2 \right) \left(\frac{\xi}{u} \right) \\
&\quad + \left(\frac{\xi}{u} \right) + 2 \left(\frac{1}{u^2 b^2} - 2 \right) \frac{u}{\eta} \frac{\xi}{(\xi + \eta)} + \frac{1}{4} \left(\frac{1}{u^2 b^2} - 2 \right)^2 \left(\frac{u}{\eta} \right)^3 \Bigg\}
\end{aligned} \tag{I.25}$$

These formulas have been checked numerically. By inspection, C, is a function of (bu), only, so $\mathrm{d}C/\mathrm{d}b = 0$. Numerical calculations indicate that $C \approx 9.0$ (for $r = 1.8$). Conservation of energy implies

$$\mathrm{d}\langle F_x \rangle_\mathrm{T} = 0 = \mathrm{d}(\omega \rho b^2 A^2 C) = \rho C (2A^2 b \, \mathrm{d}b + 2b^2 A \, \mathrm{d}A) \text{ or } \frac{\mathrm{d}A}{A} = -\frac{\mathrm{d}b}{b} \; . \tag{I.26}$$

This result differs from the shear wave case by a factor of two. Unlike the shear wave, the depth distribution of energy in the Rayleigh wave changes with b, and influences the sensitivity of amplitude to velocity changes.

Appendix J. *Amplitude sensitivity due to conservation of energy*

Equation (I.26) can be written in terms of the Rayleigh wave velocity, $V_r = \gamma b$:

$$\frac{\mathrm{d}A}{\mathrm{d}V_r} = -\frac{A}{V_r} \tag{J.1}$$

We would not expect this formula to hold exactly in an Earth structure that is more complicated than a halfspace, because the way in which changes in velocity redistribute energy with depth are more complicated. In the layered-earth case, we expect $\mathrm{d}A/\mathrm{d}V_r = -\varphi A/V_r$, where the factor, φ, needs to be determined by numerical simulation. In the common case of a shallow, low-velocity layer atop a halfspace, we believe that at high-frequencies, $\varphi \approx 1$, because at those frequencies the Rayleigh wave energy is mostly within the very top of the layer, which is effectively a halfspace. Similarly, a low-frequencies, most of the Rayleigh wave energy is mostly within the lower halfspace, and again $\varphi \approx 1$. At intermediate frequencies, the Rayleigh wave will behave more like a propagating wave confined to layer, for which Eqn. (I.5) gives $\varphi = 1/2$.

Appendix K. *Sensitivity kernel associated with energy conservation*

The amplitude equation can be rewritten in terms of the Rayleigh wave slowness, $u = V_r^{-1}$, by noting that $\mathrm{d}V_r/V_r = -\mathrm{d}u/u$, yielding $\mathrm{d}A/A = \varphi \, \mathrm{d}u/u$. The change in amplitude, A, associated with energy conservation is then

$$\delta a(x_\mathrm{i}, y_\mathrm{i}) \equiv \delta \left(\frac{\Delta A(x_\mathrm{i}, y_\mathrm{i})}{A_0} \right) = \varphi \frac{\delta u(x_\mathrm{i}, y_\mathrm{i})}{u_0(x_\mathrm{i}, y_\mathrm{i})} \; . \tag{K.1}$$

Here, $u_0(x_\mathrm{i}, y_\mathrm{i})$ is the reference Rayleigh wave phase velocity at horizontal position, $(x_\mathrm{i}, y_\mathrm{i})$. Although purely local, Eqn. (K.1) can be converted to a sensitivity kernel by introducing a Dirac function

$$\delta a(s) = \varphi \int_0^{s+\varepsilon} \frac{\delta(s - s')}{u_0(s)} \delta u \, \mathrm{d}s' \; . \tag{K.2}$$

Here, $\varepsilon \ll s$ is a small positive number.

Appendix L. *Solution of the inverse problem*

We use Riemann quadrature to approximate a ray integral as a sum. For example, for travel time:

$$\delta t(s_{\mathrm{i}}) = \int_0^{s_i} \delta u(s')\, \mathrm{d}s' \approx \sum_{\mathrm{j}=1}^{N_{\mathrm{i}}} \Delta s \cdot \delta u^{\mathrm{est}}(s_{\mathrm{j}}) \tag{L.1}$$

Here, $s_{\mathrm{j}} = \mathrm{j} \cdot \Delta s$ and the number, N_{i} of points along the ray path i has been chosen so that Δs is smaller than the grid spacing. Inserting the sensitivity of the interpolated slowness (Equation H.4) and reversing the order of summation yields

$$\delta t(s_{\mathrm{i}}) = \sum_{\mathrm{j}=1}^{N_{\mathrm{i}}} \Delta s \sum_{\mathrm{p}=1}^{M} \sum_{\mathrm{q}=1}^{M} \frac{\partial \delta u^{\mathrm{est}}(s_{\mathrm{j}})}{\partial u_{\mathrm{pq}}} \delta u_{\mathrm{pq}} = \sum_{\mathrm{p}=1}^{M_x} \sum_{\mathrm{q}=1}^{M_y} \left(\sum_{\mathrm{j}=1}^{N_{\mathrm{i}}} \Delta s \frac{\partial u^{\mathrm{est}}(s_{\mathrm{j}})}{\partial u_{\mathrm{pq}}} \right) \delta u_{\mathrm{pq}} \,. \tag{L.2}$$

The sign of the right-hand-side of Eqn. (L.2) a positive, consistent with the well-known behavior of a slowness increase leading to an increase in travel time. Sensitivity integrals for the geometrical spreading, intrinsic attenuation and energy conservation contributions to amplitude are handled similarly:

$$\begin{aligned} \delta a(s_{\mathrm{i}}) = & \sum_{\mathrm{p}=1}^{M_x} \sum_{\mathrm{q}=1}^{M_y} \left(\sum_{\mathrm{j}=1}^{N_{\mathrm{i}}} (-\frac{1}{2}[s - s']\Delta s) \frac{\partial}{\partial u_{\mathrm{pq}}} \left(\frac{\partial^2 u^{\mathrm{est}}}{\partial \xi^2} \right) \right) \delta u_{\mathrm{pq}} \\ & + \sum_{\mathrm{p}=1}^{M_x} \sum_{\mathrm{q}=1}^{M_y} \left(\frac{\varphi}{u_0(s_{\mathrm{i}})} \frac{\partial u^{\mathrm{est}}(s_{\mathrm{i}})}{\partial u_{\mathrm{pq}}} \right) \delta u_{\mathrm{pq}} \\ & + \sum_{\mathrm{p}=1}^{M_x} \sum_{\mathrm{q}=1}^{M_y} \left(\sum_{\mathrm{j}=1}^{N_{\mathrm{i}}} (-\Delta s) \frac{\partial \alpha^{\mathrm{est}}(s_{\mathrm{j}})}{\partial \alpha_{\mathrm{pq}}} \right) \delta \alpha_{\mathrm{pq}} \,, \end{aligned} \tag{L.3}$$

with

$$\begin{aligned} & \frac{\partial}{\partial u_{\mathrm{pq}}} \left(\frac{\partial^2 u^{\mathrm{est}}}{\partial \xi^2} \right) \\ & = \sin^2\theta \frac{\partial}{\partial u_{\mathrm{pq}}} \left(\frac{\partial^2 u^{\mathrm{est}}}{\partial x^2} \right) + 2 \sin\theta \cos\theta \frac{\partial}{\partial u_{\mathrm{pq}}} \left(\frac{\partial^2 u^{\mathrm{est}}}{\partial x \partial y} \right) \\ & + \cos^2\theta \frac{\partial}{\partial u_{\mathrm{pq}}} \left(\frac{\partial^2 u^{\mathrm{est}}}{\partial y^2} \right) \,. \end{aligned}$$

As is well-known, a positive perturbation in slowness causes rays to converge, leading to a positive perturbation in amplitude. Such a slowness perturbation is associated with a negative $\partial^2 u / \partial \xi^2$, consistent with the minus sign in the first term of Equation (L.3). As was discussed in Section 10, a positive perturbation in slowness leads to a positive perturbation in amplitude, consistent with the positive sign in the second term of Equation (L.3). As is well-known, a positive perturbation in attenuation factor leads to a negative perturbation in amplitude, consistent with the negative sign in the third term of Equation (L.3). Equations L.2 and (L.3) are linear algebraic equations relating observations of travel time and amplitude to the unknowns control points, δu_{ij} and $\delta \alpha_{\mathrm{ij}}$. The unknowns can be 'un-

raveled' into vectors, $\mathbf{m}^u$ and $\mathbf{m}^\alpha$, each of length $M_x \times M_y$, and can be grouped into a single model parameter vector, $\mathbf{m} = [\mathbf{m}^u, \mathbf{m}^\alpha]^\mathsf{T}$ of length, $M = 2M_x \times M_y$. The data vectors, δt and δa, say of length, N_t and N_a, respectively, can be grouped into a single data vector, $\Delta\mathbf{d} = [\delta\mathbf{t}, \delta\mathbf{a}]^\mathsf{T}$ of length, $N = N_t + N_a$.

Then, Equations (I.2) and (I.3) can be written as the $N \times M$ matrix equation

$$\begin{bmatrix} \delta\mathbf{t} \\ \delta\mathbf{a} \end{bmatrix} = \begin{bmatrix} \mathbf{G}^{t,u} & \mathbf{0} \\ \mathbf{G}^{a,u} & \mathbf{G}^{a,\alpha} \end{bmatrix} \begin{bmatrix} \mathbf{m}^u \\ \mathbf{m}^\alpha \end{bmatrix} \text{ , or } \Delta\mathbf{d} = \mathbf{G}\Delta\mathbf{m} \text{ .} \tag{L.4}$$

Here, the sub-matrices, $\mathbf{G}^{t,u}$, $\mathbf{G}^{a,u}$ and $\mathbf{G}^{a,\alpha}$ are unraveled versions of the ones defined in Equations (L.2) and (L.3). Their sparseness can be estimated by examining a ray that crosses the model in the x-direction. It interacts with a $4 \times M_x$ swath of control points, out of a total of $M_x M_y$. Thus, the fraction of non-zero elements on a row of is $4M_x/(M_x M_y) = 4/M_y$. Rays with other orientations interact with a similar number of control points. Averaging the x and y-directions yields an overall estimate of $2/M_x + 2/M_y$ non-zero elements. For a 1000×1000 grid, the sub-matrices are estimated to be about 99.6 % sparse. The combined matrix, $\mathbf{G}$, is even sparser, for it contains a block of zeroes. In all our calculation, we assume $\varphi = 1$, the value for a Rayleigh wave in a halfspace.

The damped least squares solution used in the paper is (W. *Menke*, 2018)

$$\mathbf{m}^{\text{est}} = \{\mathbf{G}^\mathsf{T}\mathbf{G} + \mathbf{W}\}^{-1}\mathbf{G}^\mathsf{T}\{\mathbf{d} - \mathbf{G}\mathbf{m}_0\} + \mathbf{m}_0 \text{ .} \tag{L.5}$$

Here, $\mathbf{m}_0 = [\mathbf{m}_0^u, \mathbf{m}_0^\alpha]^\mathsf{T}$ is a spatially-constant background solution and $\mathbf{W}$ is a diagonal matrix of damping factors, with the upper half of the diagonal having a constant value, ε_u, and the lower half, the constant value, ε_α. Small values of the damping constants have little effect on well-resolved control points, but drive unresolved ones toward the background. Large values of one of the damping constants, say ε_α, will force the solution to have most of its spatial variability in u and not α (at the expense of fitting the data poorly).

The $M \times M$ Gram matrix, $\mathbf{G}^\mathsf{T}\mathbf{G}$, encountered in the damped least squares solution (Equation I.5) is typically very large. The examples in this paper, with $M = 882$, were constructed by explicitly calculating the Gram matrix, and then solving the damped least squares system with Python's `scipy.linalg.solve` method. Additionally, we have tested and found practical a sparse matrix implementation that uses the biconjugate gradient method to solve Equation (I.5) without explicitly constructing the Gram matrix (see W. *Menke and* J. *Menke*, 2016, Section 5.8).