Equivalent Heterogeneity Analysis as a Tool for Understanding the Resolving Power of Anisotropic Travel Time Tomography

William Menke
Lamont-Doherty Earth Observatory of Columbia University
(Version 3.0; October 31, 2014)

Abstract.

We investigate whether 2D anisotropic travel time tomography can uniquely determine both the spatially-varying isotropic and anisotropic components of the seismic velocity field. This issue was first studied by Mochizuki (1997) for the special case of Radon's problem (tomography with infinitely long rays), who found it to be non-unique. Our analysis extends this result to all array geometries and demonstrates that *all* such tomographic inversions are non-unique. *Any* travel time dataset can be fit by a model that is either purely isotropic, purely anisotropic, or some combination of the two. However, a pair of purely isotropic and purely anisotropic velocity models that predict the same travel times are very different in other respects, including spatial scale. Thus, prior information can be used to select among equivalent solutions to achieve a "unique" solution embodying a given set of prior expectations about model properties. We extend the notion of a resolution test, used in traditional isotropic tomography, to the anisotropic case. Our *Equivalent Heterogeneity Analysis* focuses on the anisotropic heterogeneity equivalent to a point isotropic heterogeneity, and vice versa. We demonstrate that it provides insights into the structure of an anisotropic tomography problem that facilitates both the selection of appropriate prior information and the interpretation of results. We recommend that it be routinely applied to all surface wave inversions where the presence of anisotropy is suspected, including those based on noise-correlation.

Keywords: travel time tomography, seismic anisotropy, Radon's problem, resolution, non-uniqueness, ambient noise correlation, seismic surface waves

INTRODUCTION

This paper addresses the issue of using 2D tomographic inversion of travel time data (or equivalently, phase delay data) to image seismic velocity in the presence of both heterogeneity (variation with position) and anisotropy (variation with direction of propagation). While a simpler problem than fully three-dimensional tomography, 2D tomography has wide uses in seismology, because several important classes of elastic waves can be viewed, at least approximately, as propagating horizontally across the surface of the earth. 2D tomography has been applied to mantle-refracted body waves such as Pn and Sn (e.g. Hearn, 1996; Pei et al., 2007). However, its widest application has been to Rayleigh and Love waves (surface waves), where a sequence of inversions is used to image the surface wave phase velocity at a suite of periods. In the surface wave case, the earth's material anisotropy leads to azimuthal anisotropy of phase velocity, which in turn causes an azimuthal variability of the travel time (or, equivalently, phase delay) of the surface wave.

Starting in the 1970's and continuing to the present, many authors have used long-period surface waves from large earthquakes observed at teleseismic distances to study the structure of the lithosphere (e.g. Yu and Mitchell, 1979; Tanimoto and Anderson, 1984; Nishimura and Forsyth, 1988; Montagner and Tanimoto, 1991; Ritzwoller and Levshin , 1998; Nettles, M. and A.M. Dziewonski, 2008). Most, but not all, of these authors include azimuthal anisotropy in their inversions; those who omitted it nevertheless recognized its likely presence. These authors are able to achieve impressive global or continental-scale images with spatial resolution of 100-200 km, using surface wave periods as small as about 20s and source-receiver offsets as small as about 1000 km. Finer-scale resolution is difficult to achieve with earthquake sources, owing to the low signal-to-noise ratio at shorter periods and the paucity of shorter source-receiver offsets. However, during the last decade, the development of ambient noise-correlation techniques for reconstructing surface waves propagating between stations has opened up new opportunities for the use of surface waves in high-resolution seismic imaging (Shapiro and Campillo, 2004; Shapiro et al. 2005; Calkins et al., 2011). Surface wave travel times, for periods as short as 8s, can now be routinely calculated by cross-correlating ambient noise observed at two stations, separated by a little as 50 km (Snieder, 2004; Bensen et al., 2007; Ekstrom et al. 2009). The revolutionary aspect of ambient noise correlation is that the number of measurements tends to be larger, and the spatial and azimuthal pattern of paths tends to be better, than traditional earthquake-source methods. The resulting tomographic images often have sufficiently high resolution to permit detailed structural interpretations (e.g. Lin et al., 2007; Yang et al., 2007; Lin et al., 2008; Zha et al., 2014). Owing to the excitement that noise-correlation has generated (both in the community and for this author), revisiting issues associated with 2D tomography is timely and appropriate. In particular, we address here the question of the the degree to which this technique can distinguish anisotropy from heterogeneity. Simply put, can it uniquely determine both?

Seismic velocity is inherently both heterogenous and anisotropic. The latter can be due to intrinsic anisotropy of mineral grains aligned by large-scale ductile deformation (Hess, 1964; Raitt et al., 1969; Silver and Chan, 1988; Nicolas, 1989; Karato et al. 2008) or to the effective anisotropy of materials with fine-scale layering and systems of cracks (Backus, 1962; Menke, 1983) or some combination of the two (Fitchner 2013). This anisotropy needs to be accounted for in a tomographic inversion as it is a source of important information about earth processes. However, an anisotropic earth model is extremely complex and requires 21 functions of position for its complete description (e.g. Aki and Richards, 2002). Notably, for the special case of surface waves propagating in a weakly anisotropic earth, the phase velocity is sensitive to only a few combinations of these functions (Backus, 1965; Smith and Dahlen, 1973). It is possible to formulate a tomographic inversion that includes all 21 functions (e.g. Wu and Lees, 1999). However, most surface wave applications use a simplied form of anisotropy that is described by just the three functions. One of these functions represents the isotropic part of the phase velocity. The other two represent the anisotropic part and encode a $\cos\{2(\theta-\theta_0)\}$ angular dependence (where θ is azimuth of propagation and θ_0 is the azimuth of the slow axis of anisotropy).

The switch from one function in 2D isotropic tomography to three functions in the anisotropic case raises the issue of whether sufficient information is contained in travel time measurements to uniquely determine, even in principle, all three functions. Mochizuki (1997) studied the special case of Radon's problem – tomography with infinitely long rays - and showed that travel time measurements at best can determine only one combination of the three unknown functions. For more realistic experimental geometries, numerical tests succeeded in reconstructing simple patterns of anisotropy (e.g. Wu and Lees, 1999), suggesting that Mochizuki's (1997) result was not applicable to these more realistic cases. As will demonstrate below, the success of these tests was due to the addition of prior information that selected for the simple patterns from among an infinitude of possibilities, and not because Mochizuki's (1997) result was not applicable.

We demonstrate below that *any* travel time dataset can be fit by a model that is either purely isotropic, purely anisotropic, or some combination of the two. However, the spatial patterns of isotropy or anisotropy that are equivalent in the sense of predicting the same travel times are very different in other respects, including spatial scale. Thus, prior information can be used to select among equivalent solutions to achieve a "unique" solution embodying a given set of prior expectations about model properties.

Spatial resolution analysis has proved an extremely powerful tool in understanding non-uniqueness in traditional isotropic tomography problems (Backus and Gilbert, 1968; Wiggins, 1972, see also Menke, 2012; Menke, 2014). We extend ideas of resolution here to anisotropic tomography by focusing on the anisotropic heterogeneity equivalent to a point isotropic heterogeneity, and vice versa. We demonstrate that this *Equivalent Heterogeneity Analysis* provides insights into the structure of an anisotropic tomography problem that facilitates both the selection of appropriate prior information and the interpretation of results.

PRINCIPLES OF 2D ANISOTROPIC TOMOGRAPHY

We limit our study to the case of weak two-dimensional heterogeneity and anisotropy, meaning that the phase velocity, V, can be expressed in terms of a constant background velocity, V_0, and a small perturbation, $\delta V(x, y, \theta)$, which is a function of position in the (x, y) plane and propagation azimuth, θ:

$$V = V_0 + \delta V(x, y, \theta) \tag{1}$$

The phase slowness, $U = 1/V$, can be expressed to first order as:

$$U = [V_0 + \delta V]^{-1} = {V_0}^{-1}\left[1 + \frac{\delta V}{V_0}\right]^{-1} \approx {V_0}^{-1} - \frac{\delta V}{{V_0}^2} \equiv U_0 + \delta U \tag{2}$$

where $U_0 \equiv 1/V_0$ and $\delta U \equiv -\delta V/{V_0}^2$. We will use slowness, and not velocity, as the primary variable, because travel time depends linearly on slowness but nonlinearly on velocity. However, since the perturbations in velocity and slowness are proportional to one another, $\delta U \propto \delta V$, this choice, while convenient, is not fundamental.

The perturbation in phase slowness $\delta U(x, y, \theta)$ of a wave propagating in the (x, y) plane and with azimuth θ (Figure 1a) is modeled as varying with both position and azimuth according to the formula (Smith and Dahlen, 1973):

$$\delta U(x, y, \theta) = A(x, y) + B(x, y)\cos[2\{\theta - \theta_0(x, y)\}]$$

(3)

Here, $A(x,y)$ represents the isotropic part of the model, $B(x,y)$, the anisotropic part and $\theta_0(x,y)$, the azimuth of the axis of anisotropy. The slowest propagation occurs when $\theta = \theta_0$ (that is, θ_0 is the *slow axis* of anisotropy) and the fastest at right angles to it. Note that this model omits $\cos 4\theta$ terms, which though strictly-speaking necessary to fully-represent seismic anisotropy, are usually negligible. The trigonometric identity, $\cos(a-b) = \cos a \cos b + \sin a \sin b$, can be used to rewrite the formula as:

$$\delta U(x,y,\theta) = A(x,y) + B_c(x,y)\cos 2\theta + B_s(x,y)\sin 2\theta \tag{4}$$

with

$$B_c = B\cos 2\theta_0 \;\text{ and }\; B_s = B\sin 2\theta_0$$
$$B = (B_c^2 + B_s^2)^{½} \;\text{ and }\; \theta_0 = ½\tan^{-1}(B_s/B_c) \tag{5}$$

Thus, the anisotropic medium is specified by three spatially-varying *material parameter functions*, $A(x,y)$, $B_c(x,y)$ and $B_s(x,y)$. The function $A(x,y)$ describes the isotropic part of the slowness and the two functions $B_c(x,y)$ and $B_s(x,y)$ describe the anisotropic part. This parameterization avoids explicit reference to the direction of the slow axis of anisotropy.

We rely here on seismic ray theory (e.g. Cerveny, 2005) to link slowness to travel time. Widely used in seismology, it is a high-frequency approximation to the wave equation that is valid when diffraction effects can be ignored; that is, when slowness varies slowly and smoothly with position (when compared to wavelength of the observed seismic waves). We believe that its use here leads to what is in some sense a 'best case' analysis of non-uniqueness; inversion of low-bandwidth data will be more non-unique than our ray-theory based analysis indicates (and as we will demonstrate, below, our ray-theory based analysis points to substantial non-uniqueness).

The travel time, T (or equivalently the phase delay, $\varphi = \omega T$, where ω is angular frequency), between a source at (x_1, y_1) and a receiver at (x_2, y_2) and separated by a distance, L, is approximated as the ray integrals:

$$T = T_0 + \delta T \;\text{ with }\; T_0 = \int_{s_1}^{s_2} U_0 \, ds \;\text{ and }\; \delta T = \int_{s_1}^{s_2} \delta U[x(s), y(s)] \, ds \tag{6}$$

Here, s is arc-length along the ray connecting source and receiver. In some instances, it may suffice to approximate the ray as a straight line, in which case its azimuth, θ, is constant and (x,y) are linear functions of arc-length, s:

$$x = x_1 + s\cos\theta = a + bs \;\text{ and }\; y = y_1 + s\sin\theta = c + ds \;\text{ with }\; \theta = \tan^{-1}\frac{y_2 - y_1}{x_2 - x_1}$$

$$\text{and } s_1 = 0 \text{ and } s_2 = L \tag{7}$$

Here $a, b, c,$ and d are abbreviations for $x_1, \cos\theta, y_1$ and $\sin\theta$, respectively. In this straight-line case, after inserting Equation 4 into Equation 6 and applying the straight line ray assumption, the travel time becomes:

$$T = T_0 + \delta T \text{ with } T_0 = U_0 L \text{ and}$$
$$\delta T = \int_{s_1}^{s_2} A[x(s), y(s)]\ ds + \cos 2\theta \int_{s_1}^{s_2} B_c[x(s), y(s)]\ ds$$
$$+ \sin 2\theta \int_{s_1}^{s_2} B_c[x(s), y(s)]\ ds \equiv I_A + \cos 2\theta\ I_C + \sin 2\theta\ I_S \tag{8}$$

Here, I_A, I_C and I_S are abbreviations for the three integrals. Note that all three integrals are of the same form; that is, line integrals of their respective integrands over the same straight line segments.

We now focus upon the tomographic imaging problem; that is, what can be learned about the material parameter functions, $A(x, y)$, $B_c(x, y)$ and $B_s(x, y)$ when the travel time function δT has been measured for specific source-receiver geometries. Note that the background slowness, U_0, does not appear explicitly in the formula relating δT to A, B_c and B_s, implying that the results of our analysis will be independent of its value (as long as the assumption of weak heterogeneity and anisotropy holds). Thus, we are free to set $U_0 = 0$, but with the understanding that this choice is made to eliminate the need to carry an irrelevant parameter through the analysis, rather than as a statement about the actual background slowness. Any background slowness can be superimposed, without impacting the results.

ANALYSIS OF A STAR ARRAY

Intuitively, we expect that travel time measurements made along several short ray paths centered on the same point, say (x_0, y_0), but with different azimuths, say $\theta_1, \theta_2, \theta_3 \cdots$ (a "star array", as in the Figure 1b), would be sufficient to determine the average material properties (including the mean direction of the slow axis) near that point. This result can be demonstrated by writing the average of A as $\langle A \rangle = I_A / L$, and similarly for $\langle B_C \rangle$ and $\langle B_S \rangle$. These averages depend upon the ray azimuth, θ, since the line integral depends upon path. However, for smooth models and for sufficiently small L, $A(x, y)$ can be approximated by the first three terms of its Taylor series:

$$A(x, y) \approx A(x_0, y_0) + \left.\frac{\partial A}{\partial x}\right|_{(x_0, y_0)} (x - x_0) + \left.\frac{\partial A}{\partial y}\right|_{(x_0, y_0)} (y - y_0) \tag{9}$$

in a small region of the (x, y) plane that includes the whole ray. Inserting Equation 9 into the formula for I_A in Equation 8, and using the relations $(x - x_0) = s\ \cos\theta$ and $(y - y_0) = s\ \sin\theta$, we find that:

$$\langle A \rangle = I_A/L \approx L^{-1} \int_{-L/2}^{+L/2} \left(A(x_0, y_0) + \left.\frac{\partial A}{\partial x}\right|_{(x_0,y_0)} s\ \cos\theta + \left.\frac{\partial A}{\partial y}\right|_{(x_0,y_0)} s\ \sin\theta \right) ds$$

$$= \frac{A(x_0, y_0)}{L} \int_{-\frac{L}{2}}^{+\frac{L}{2}} ds + \frac{\cos\theta}{L} \left.\frac{\partial A}{\partial x}\right|_{(x_0,y_0)} \int_{-\frac{L}{2}}^{+\frac{L}{2}} s\ ds + \frac{\sin\theta}{L} \left.\frac{\partial A}{\partial y}\right|_{(x_0,y_0)} \int_{-\frac{L}{2}}^{+\frac{L}{2}} s\ ds =$$

$$A(x_0, y_0) + 0 + 0 \tag{10}$$

Note that the first integral equals L and the other two integrals are zero. We conclude that $\langle A \rangle = I_A/L \approx A(x_0, y_0)$ and similarly for $\langle B_C \rangle$ and $\langle B_S \rangle$. Furthermore, these averages are independent of ray direction, as long as the ray is short enough for the linear approximation to be valid. The travel time equation for ray i is then:

$$d_i \equiv \frac{\delta T_i}{L_i} = \langle A \rangle + \cos 2\theta_i\ \langle B_C \rangle + \sin 2\theta_i\ \langle B_S \rangle \tag{11}$$

Here, d_i is an abbreviation for the *path-averaged slowness* $\delta T_i/L_i$. The average material properties, $\langle A \rangle$, $\langle B_C \rangle$ and $\langle B_S \rangle$, can be determined by travel time measurements along three distinct rays. For example, if $(\theta_1, \theta_2, \theta_3) = (0, \pi/4, \pi/2)$:

$$d_1 = \frac{\delta T_1}{L_1} = \langle A \rangle + \langle B_C \rangle + 0$$

$$d_2 = \frac{\delta T_2}{L_2} = \langle A \rangle + 0 + \langle B_S \rangle$$

$$d_3 = \frac{\delta T_3}{L_3} = \langle A \rangle - \langle B_C \rangle + 0 \tag{12}$$

then $\langle A \rangle = ½(d_1 + d_3)$, $\langle B_C \rangle = d_1 - \langle A \rangle$ and $\langle B_S \rangle = d_2 - \langle A \rangle$.

Once $\langle A \rangle$, $\langle B_C \rangle$ and $\langle B_S \rangle$, have been determined, the average slow axis, $\langle \theta_0 \rangle$, and average anisotropy, $\langle B \rangle$, can be computed as:

$$\langle B \rangle \approx (\langle B_c \rangle^2 + \langle B_s \rangle^2)^{½} \text{ and } \langle \theta_0 \rangle \approx ½ \tan^{-1}(\langle B_s \rangle / \langle B_c \rangle) \tag{13}$$

We use approximate signs, because θ_0 and B are non-linear functions of A, B_C and B_S, and so strictly speaking, the average values $\langle\theta_0\rangle$ and $\langle B\rangle$ are not exactly what is obtained by the substitution of average values $\langle A\rangle$, $\langle B_C\rangle$ and $\langle B_S\rangle$ into the functions. Nevertheless, this approximation is usually adequate.

One star array can be used to estimate the material parameters in the vicinity of a single point in a spatially-varying model. A grid of them can be used to estimate these properties on a grid of points, and hence to produce a low-resolution estimate of the model. An example is shown in Figure 2, where a test model is imaged by two grids of star arrays, a fine grid of small star arrays and a coarse grid of large star arrays. As expected, the finer, denser grid does a better job recovering the test model, but in both cases both isotropic and anisotropic features are correctly recovered, or at least those features with a scale length greater than the size L of the star arrays.

The incorporation of star arrays into an experimental design has practical advantage, since it provides data that can discriminate anisotropy from heterogeneity. The caveat is that its success depends on correctly choosing the length L of the arrays, which must be smaller than the spatial scale over which the material parameters vary. This point brings out the role of prior information in achieving a unique solution. From the point of view of uniqueness, very small star arrays are advantageous. However, very small star arrays may not be capable of measuring travel time accurately, since measurement error does not usually scale with array size. Travel time measurements made with small-aperture arrays tend to be very noisy.

RADON'S PROBLEM

Radon's problem is to deduce slowness in a purely isotropic model (that is, the case $A \neq 0$; $B_c = B_s = 0$), using travel time measurements along a complete set of infinitely long straight-line rays; that is, rays corresponding to sources and receivers at $\pm\infty$. By complete, we mean that measurements have been made along rays with all possible orientations and positions. In practice, infinitely long rays are not realizable; a feasible experiment approximating Radon's geometry has the sources and receivers on the boundary of the study region. The non-uniqueness of the anisotropic version of Radon's problem has been investigated in detail by Mochizuki (1997), who concludes that it is substantially non-unique. Mochizuki's (1997) result, which is based on a Fourier representation of slowness, will be discussed later in this section. We first review more general aspects of the problem.

In the traditional formulation of Radon's problem, straight line rays are parameterized by their distance, u, of closest-approach to the origin and the azimuth ϕ of the $\vec{u}$ direction (Figure 1c). The travel time equation (Equation 8) becomes:

$$\delta T(u,\phi) = \int_{ray_{u,\phi}} A[x(s), y(s)]\ ds \tag{14}$$

Since $\phi = \theta + \pi/2$, we can view travel time as a function of either ϕ or θ; that is, as either $\delta T(u,\phi)$ or $\delta T(u,\theta)$. In the discussion below we use the latter form, since it is more compatible with our previous usage.

Radon's problem has been studied extensively. The problem of determining $A(x, y)$ from $\delta T(u, \theta)$ is known to be unique, as long as data from a complete set of rays are available. The *Fourier Slice Theorem* (e.g. Menke 2012; Menke, 2014) shows that exactly enough information is available in $\delta T(u, \theta)$ to construct the Fourier transform $\tilde{\tilde{A}}(k_x, k_y)$ at all wavenumbers (k_x, k_y). Thus, $A(x, y)$ is uniquely determined, since a function is uniquely determined by its Fourier transform. An implication of the Fourier Slice Theorem is that any travel time function, $\delta T(u, \theta)$, can be exactly fit by an isotropic model, irrespective of whether or not the true model from which it was derived was purely isotropic. A tomography experiment that uses infinitely long rays *cannot* prove the existence of anisotropy.

We now inquire whether it is possible to find a purely anisotropic model in which only $B_c(x, y)$ is non-zero and that exactly fits the travel time data. Superficially, this proposition seems possible, since travel time equation (Equation 8 with $A = B_s = 0$) can be manipulated into exactly the same form as Radon's equation, simply by dividing through by $\cos 2\theta$:

$$\delta T^*(u, \theta) \equiv \frac{\delta T(u, \theta)}{\cos 2\theta} = \int_{ray_{u,\theta}} B_c[x(s), y(s)]\ ds \tag{15}$$

However, the new "travel time" function, $\delta T^*(u, \theta)$ is singular at angles where the cosine is zero, making the application of the Fourier Slice Theorem invalid. Physically, these are the ray orientations at which B_c can have no effect on travel time. Therefore, no choice of B_c will fit the travel time along those rays. The same problem would arise if we were to try to fit the travel time with a model in which only $B_s(x, y)$ is non-zero.

A purely anisotropic model that includes both $B_c(x, y)$ and $B_s(x, y)$ *can* be made to work. We first define:

$$\delta T^C(u, \theta) \equiv \cos^2 2\theta\ \ \delta T(u, \theta) \quad \text{and} \quad \delta T^S(u, \theta) \equiv \sin^2 2\theta\ \ \delta T(u, \theta) \tag{16}$$

Note that $\delta T = \delta T^C + \delta T^S$. The travel time integrals analogous to Equation 15 are:

$$\delta T^{C*}(u, \theta) = \frac{\delta T^C(u, \theta)}{\cos 2\theta} = \cos 2\theta\ \ \delta T(u, \theta) = \int_{ray_{u,\theta}} B_c[x(s), y(s)]\ ds$$
$$\delta T^{S*}(u, \theta) = \frac{\delta T^S(u, \theta)}{\sin 2\theta} = \sin 2\theta\ \ \delta T(u, \theta) = \int_{ray_{u,\theta}} B_s[x(s), y(s)]\ ds \tag{17}$$

The quantities $\delta T^{C*}$and δT^{S*} have no singularities, so we can construct a $B_c(x, y)$ and a $B_s(x, y)$ that fits them exactly. Finally, we note that

$$\cos 2\theta \int_{ray_{u,\theta}} B_c[x(s), y(s)]\ ds + \sin 2\theta \int_{ray_{u,\theta}} B_s[x(s), y(s)]\ ds = \ \delta T^C + \delta T^S = \ \delta T \tag{18}$$

We have constructed a purely anisotropic expression that fits the travel time data exactly. Note that the linear combination of isotropic and anisotropic models, $(1-\alpha)A(x,y)$, $\alpha B_c(x,y)$ and $\alpha B_s(x,y)$, satisfy the travel time data exactly for any value of the parameter, α. A whole family of models with different mixes of heterogeneity and anisotropy can be constructed. If we define:

$$\delta T^C(u,\theta) \equiv c(\theta)\delta T(u,\theta) \quad \text{and} \quad \delta T^S(u,\theta) \equiv s(\theta)\delta T(u,\theta)$$
$$\text{and} \ \ \delta T^A(u,\theta) \equiv \{1 - c(\theta) - s(\theta)\}\, \delta T(u,\theta) \tag{19}$$

where $c(\theta)$ and $s(\theta)$ are chosen to have appropriately-placed zeros that removed the singularities but are otherwise arbitrary, then δT^A, $\delta T^C / \cos 2\theta$, and $\delta T^S / \sin 2\theta$ can be separately inverted to a set of A, B_c and B_s that, taken together, fit the travel time data exactly. Evidently, many such functions $c(\theta)$ and $s(\theta)$ exist, since one set of acceptable choices is:

$$c(\theta) = \sum_{n=0}^{\infty} c_n \cos^{\mathrm{n}} 2\theta \quad \text{and} \quad s(\theta) = \sum_{n=1}^{\infty} s_n \sin^{\mathrm{n}} 2\theta \tag{20}$$

where c_n and s_n are arbitrary (up to a convergence requirement).

MOCHIZUKI (1977) ANALYSIS OF RADON'S PROBLEM

We now return to Mochizuki's (1997) analysis of non-uniqueness. Mochizuki's (1997) considers a very general form of slowness:

$$\delta U(x,y,\theta) = \sum_{n=0}^{\infty} f_n^c(x,y) \cos n\theta + f_n^s(x,y) \sin n\theta \tag{21}$$

Note that all possible angular behaviors are considered, including those with odd n. The contribution of the even-n terms is unchanged when source and receiver are interchanged; that is, when θ is replaced with $\theta + \pi$. This behavior is characteristic of anisotropy. The contribution of the odd-n terms switches sign when the source and receiver are interchanged. This behavior is *not* characteristic of anisotropy, but can be used to model other wave propagation effects, such as those arising from dipping layers. The parameterization used in this paper (Equation 4) includes only the $n = 0$ isotropic term and the two $n = 2$ anisotropic terms.

Mochizuki's (1997) first result shows that the even-n terms can be determined independently of the odd-n terms. The former depends only upon the sum of $\delta U(x, y, \theta)$and $\delta U(x, y, \theta+\pi)$ and the latter depends only upon the difference. Provided that measurements made in both directions are averaged, the odd-n terms, arising say from dipping layers, will not bias the estimate of anisotropy.

Mochizuki's (1997) second result addresses the issue of non-uniqueness. It is an adaptation of the Fourier Slice Theorem and uses as primary variables the 2D Fourier transforms $\hat{\hat{f}}_n^c(k_r, m)$ and $\hat{\hat{f}}_n^s(k_r, m)$ of the spatially-varying f's in Equation 21. Here (k_r, m) are radial and azimuthal wavenumbers, respectively. The travel time data are shown to be sufficient to constrain exactly one linear combination of $\hat{f}_n^c$'s and exactly one linear combination of $\hat{f}_n^s$'s, rather than all of the $\hat{f}_n^c$'s and $\hat{f}_n^s$'s, individually. This result implies that the $n = 0$ isotropic terms and the $n = 2$ anisotropic terms (the focus of this paper) cannot be separately determined. This is the same behavior investigated earlier in this section through Equation 19.

EQUIVALENT HETEROGENEITIES FOR RADON'S PROBLEM

While a range of isotropic and anisotropic models can fit a given travel time data set, not all of them may be sensible when judged against prior information about the study region. It may be possible to rule out some models because they contain features that are physically implausible, such as very small-scale isotropic heterogeneity or rapidly fluctuating directions of the slow axis of anisotropy.

Some insight on this issue can be gained by studying the types of solutions that are possible when the true model contains a single point-like heterogeneity that is either purely isotropic or purely anisotropic. As shown in Appendix B, these solutions can be derived analytically for Radon's problem. However, from the perspective of anisotropic tomography, Radon's problem is just one of many source-receiver configurations – and not the most commonly encountered, either. Hence, we will focus on universally-applicable inversion techniques based on generalized least squares (e.g. Menke, 2012; Menke, 2014; see also Appendix A), rather than on methods applicable only to Radon's problem. Almost all seismic tomography suffers from non-uniqueness due to under-sampling. The same regularization (damping) schemes that are used to handle this type of non-uniqueness also have application to non-uniqueness associated with anisotropy.

We consider a sequence of experiments in which an exact travel time dataset is computed from the true model and then inverted for an estimated model, using the inverse method described in Appendix A and a regularization (damping) scheme that alternately forces the estimated model to be purely isotropic or purely anisotropic. This process, which we call *Equivalent Heterogeneity Analysis*, results in four estimated models:

(A) The purely isotropic model equivalent to a point-like isotropic heterogeneity

(B) The purely anisotropic model equivalent to a point-like isotropic heterogeneity

(C) The purely isotropic estimated model equivalent to a point-like anisotropic heterogeneity.

(D) The purely anisotropic estimated model equivalent to a point-like anisotropic heterogeneity

Note that we have included (A) in this tabulation, even though a perfect experiment (such as Radon's problem) would determine that the estimated and true models are identical. In real experiments, both the inherent non-uniqueness associated with anisotropy and the practical non-uniqueness caused by a poor distribution of sources and receivers are present. Cases (B) and (C) explore how isotropy and anisotropy trade off; and cases (A) and (D) function as traditional resolution tests. Taken as a group, the structure of these four estimated models can help in the interpretation of inversions of real data.

Figure 3 shows equivalent heterogeneities for Radon's problem (or actually the closest feasible approximation with sources and receivers on the boundary of the study region). An isotropic heterogeneity (Figure 3a) can be more-or-less exactly recovered by a purely isotropic inversion (Figure 3b), except for a little smoothing resulting from the regularization (even so, the travel time error is less than 1%). The purely anisotropic estimated model (Figure 3c) is radially-symmetric (as is expected, since the true heterogeneity and the ray pattern both have exact rotational symmetry) and is spatially-diffuse. Its effective diameter is at least twice the diameter as the true isotropic heterogeneity. An analytic calculation (Appendix B) indicates that the strength of the anisotropy falls off as $(\text{distance})^{-2}$. The equivalence of a point-like isotropic heterogeneity and a spatially-distributed radial anisotropic heterogeneity could possibly be problematic in some geodynamical contexts. For instance, a mantle plume might be expected to cause both a thermal anomaly on the earth's surface, which would be expressed as a point-like isotropic anomaly, and a radially-diverging flow pattern, which would be expressed as a radial pattern of fast axes. Unfortunately, the two features cannot be distinguished by Radon's problem (or, as we will show below, by any other experimental configuration, either).

The anisotropic heterogeneity (Figure 3d) is not exactly recovered by the purely-anisotropic inversion (Figure 3f). The estimated model has a much wider anomaly, with a more complicated pattern of slow axes, although with some correspondence with the true model in its central region. Yet this result is not a mistake; it fits the travel times of the much simpler true model to within a percent. It is a consequence of the extreme non-uniqueness of anisotropic inversions. The purely-isotropic estimated model (Part E) is dipolar in shape with slow lobes parallel to the slow axis of the true heterogeneity, as is predicted by Mochizuki (1997) and as discussed in Appendix B. The amplitude of the heterogeneity falls of as $(\text{distance})^{-2}$. The dipolar shape might be construed as good news in the geodynamical context, since geodynamical situations in which isotropic dipoles arise are rare; an interpretation in terms of anisotropy will often be preferable.

An extended region of spatially-constant anisotropy (Figure 4a) can be thought of as a grid of many point-line anisotropic heterogeneities (as in Figure 3d) that covers the extended region. The equivalent isotropic heterogeneity is constructed by replacing each point-like anisotropic heterogeneity with an isotropic dipole and summing (Figure 4b). Within the interior of the region, the positive and negative lobes of adjacent dipoles overlap and cancel, causing the interior to be homogeneous or nearly so. The dipoles on the boundary will not cancel, so the homogenous region will be surrounded by a thin zone of strong and very rapidly fluctuating isotropic heterogeneities. This pattern is very easily recognized. In many cases, the interpretation of the region as one of spatially-constant anisotropy will be geodynamically more plausible than that of a homogenous isotropic region with an extremely complicated boundary.

EQUIVALENT HETEROGENEITIES FOR MORE REALISTIC ARRAYS

A few experimental geometries in seismic imaging, such as imaging an ocean basin with sources and receivers located on its coastlines, correspond closely to Radon's problem. However, stations more commonly are placed within the study region, for example, on a regular grid (Figure 5).

Intuitively, one might expect this array geometry to be a significant improvement over Radon's, as the stations in the interior of the study region provide short ray paths like those of the star-array discussed earlier. Unfortunately, this is not the case, at least for the sparse station spacing used in the example (Figure 6). The scale lengths over which one can form star-arrays is just too large to be relevant to the imaging of the point-like heterogeneities used here. The equivalent heterogeneities are quite similar in shape, but arguably worse than those of Radon's problem, since they exhibit a strong rectilinear bias which is due to rows and columns of the array. Switching to a hexagonal array with the same station spacing (not shown) removes the rectilinear bias, but still results in equivalent heterogeneities very similar in shape to those of Radon's problem.

While the procedure set forward in Equation 17 for fitting travel time with either purely isotropic or purely anisotropic models was developed in the context of Radon's problem, it is equally applicable to all other array configurations, since no part of its derivation requires that the rays be infinitely long (though they do have to be straight). Fundamentally, all anisotropic tomography – even the star array - suffers from the same non-uniqueness. The appearance of uniqueness in the star array is created by the addition of prior information that the model varies smoothly (no faster than linearly) across the array. Smoothness constraints can resolve non-uniqueness in other settings, as well. For instance, it would allow the selection of a large-anisotropic-domain solution (Figure 4a) over a more highly spatially-fluctuating isotropic solution (Figure 4b). Such considerations allowed Wu and Lees (1999) to successfully recover a model containing just a few large anisotropic domains.

Irregular arrays, and especially arrays with shapes tuned to linear tectonic features such as spreading centers, are common in seismology. The array (Figure 7) we consider here has a shape similar to the Eastern Lau Spreading Center (ELSC) array, a temporary deployment of ocean-bottom seismometers that took place in 2010-2011 (Zha et al., 2013). It consists of two linear sub-arrays that are perpendicular to the spreading center, a more scattered grouping of stations parallel to the spreading center and between the linear sub-arrays, and a few outlying stations. While the central stations are closely spaced, we simulate the high noise level often encountered in ocean-bottom seismometers by omitting rays shorter than one fifth the overall array diameter.

Because of the irregularity of the array, the Equivalent Point heterogeneities are a strong function of the position of point-like heterogeneity. Results for several positions of the point-like heterogeneity must be analyzed in order to develop a good understanding of the behavior of the array. We start with a point-like heterogeneity at the center of the array, where the station density is the highest (Figure 7). The array resolves both a true isotropic heterogeneity (compare

Figure 7a and b) and a true anisotropic heterogeneity (compare Figure 7d and f) very well. The anisotropic heterogeneity that is equivalent to the true isotropic heterogeneity (compare Figure 7a and c) has a large size and a very disorganized pattern of slow directions. If encountered when interpreting real-world data, it is arguably legitimate to use Occam's Razor to reject this extremely complex anisotropic heterogeneity in favor of the much simpler isotropic one. As in all previous cases, the isotropic heterogeneity equivalent to the true point-like anisotropic heterogeneity is dipolar in character, though owing to the irregularity of the array, a little more irregular in shape than the cases considered previously.

When the true point-like heterogeneity is placed at the margin of the array, the Equivalent Heterogeneities take on more complicated shapes (Figure 9) but retain some of the same features discussed previously. Note, for instance, that the anisotropic heterogeneity equivalent to the point-like isotropic heterogeneity (Figure 8c) is much more linear in character than in previous examples. This linear pattern could be problematical for geodynamic interpretations in a spreading center environment, where linear mantle flow patterns are plausible. This result is a reminder that imaging results from the periphery of an array should always be interpreted cautiously.

DISCUSSION AND CONCLUSIONS

All 2D anisotropic tomography problems suffer from the same non-uniqueness first identified by Mochizuki (1997) for Radon's problem. *Any* travel time dataset can be fit by a model that is either purely isotropic, purely anisotropic, or some combination of the two, if heterogeneities of all shapes and spatial scales are permitted. However, the spatial patterns of equivalent isotropic and anisotropic heterogeneities are substantially different. When one is point-like, the other is spatially-extended. Thus, prior information can be used to select among equivalent solutions to achieve a "unique" solution embodying a given set of prior expectations about model properties.

We extend ideas of resolution analysis, first developed by Backus and Gilbert (1968) and Wiggins (1972) to understand non-uniqueness in a spatial context, to the anisotropic tomography problem. The resulting *Equivalent Heterogeneity Analysis* provides insights into the structure of an anisotropic tomography problem that facilitates both the selection of appropriate prior information and the interpretation of results. We recommend that it be routinely applied to all surface wave inversions where the presence of anisotropy is suspected, including those based on ambient noise correlation.

Data and Resources. Station locations for the Eastern Lau Spreading Center array are freely available and accessed through Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC) as Array YL.

Acknowledgements. This work was supported by the National Science Foundation under grants OCE-0426369 and EAR 11-47742.

References

Aki, K. and P.G. Richards, Quantitative Seismology, Second Edition, University Science Books, 702pp.

Backus, G.E. (1962), Long-wave elastic anisotropy produced by horizontal layering, J. Geophys. Res., 67(11), 4427–4440, doi:10.1029/JZ067i011p04427.

Backus, G.E. (1965), Possible form of seismic anisotropy of the upper mantle under the oceans, J. Geophys. Res 70, 3429-3439.

Backus GE, Gilbert, JF (1968), The resolving power of gross earth data, *Geophys. J. Roy. Astron. Soc*. **16**, 169–205.

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*(3), 1239-1260, doi: 10.1111/j.1365-246X.2007.03374.x.

Calkins, J. A., G. A. Abers, G. Ekstrom, K. C. Creager, and S. Rondenay (2011), Shallow structure of the Cascadia subduction zone beneath western Washington from spectral ambient noise correlation, *J Geophys Res-Sol Ea*, *116*, doi: 10.1029/2010jb007657.

Cerveny, V. (2005), Seismic Ray Theory, Cambridge University Press, 724pp.

Ekström, G., G. A. Abers, and S. C. Webb (2009), Determination of surface-wave phase velocities across USArray from noise and Aki's spectral formulation, *Geophysical Research Letters*, *36*(18), doi: 10.1029/2009gl039131.

Fitchner, A., B.L.N. Kennett and J. Trampert (2013), Separating intrinsic and apparent anisotropy, Phys. Earth Planet Int., doi:10.1016/j.pepi.2013.1003.1000.

Gradshteyn I.S. and I.M. Ryzhik (1980), Tables of Integrals, Series and Products, Corrected and Enlarged Edition, Academic Press, New York, 1160pp.

Hearn, T. M. (1996), Anisotropic Pn tomography in the western United States, J. Geophys. Res., 101(B4), 8403–8414, doi:10.1029/96JB00114.

Hess, H. (1964), Seismic anisotropy of the uppermost mantle under the oceans, Nature 203, 629-631, 1964.

Karato, S., H. Jung, I. Katayam, P Skemer (2008), Geodynamic significance of seismic anisotropy of the upper mantle: New insights from laboratory studies, Ann. Rev. Earth Planet. Sci. 36, 59=95. doi: 10.1146annrev.earth.36.031207.124120.

Lin, F.-C., Moschetti, M. P. and Ritzwoller, M. H. (2008), Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps. Geophysical Journal International, 173: 281–298. doi: 10.1111/j.1365-246X.2008.03720.x

Lin, F.-C., Ritzwoller, M. H., Townend, J., Bannister, S. and Savage, M. K. (2007), Ambient noise Rayleigh wave tomography of New Zealand. Geophysical Journal International, 170: 649–666. doi: 10.1111/j.1365-246X.2007.03414.x

Menke, W., 1983. On the effect of P-S coupling on the apparent attenuation of elastic waves in randomly layered media, Geophys. Res. Lett. 10, 1145-114.

Menke, W. (2012), Geophysical Data Analysis: Discrete Inverse Theory, MATLAB Edition, Elsevier, 293 pp.

Menke, W., Review of the Generalized Least Squares Method, in press in Surveys in Geophysics, 2014.

Mochizuki, E. (1997), Nonuniqueness of two-dimensional anisotropic tomography, Seism. Soc. Am. Bull., 87, 261-264.

Montagner, J.-P and T. Tanimoto (1991), Global upper mantle tomography of seismic velocities and anisotropies, J. Geophys. Res. 96, 20,337-20,351.

Nettles, M. and A.M. Dziewonski (2008), Radially anisotropic shear-velocity structure of the upper mantle beneath North America, J. Geophys. Res. 113, B02303, 10.1029/2006JB004819

Nicolas, A. (1989), Structures of Ophiolites and Dynamics of Oceanic Lithosphere, Springer, 367pp.

Nishimura, C. and D.W. Forsyth (1988), Rayleigh wave phase velocities in the Pacific with implications for azimuthal anisotropy and lateral heterogeneities, Geophys. J. 94, 479-501.

Pei, S., J. Zhao, Y. Sun, Z. Xu, S. Wang, H. Liu, C. A. Rowe, M. N. Toksöz, and X. Gao (2007), Upper mantle seismic velocities and anisotropy in China determined through Pn and Sn tomography, J. Geophys. Res., 112, B05312, doi:10.1029/2006JB004409.

Raitt, R. W., G. G. Shor Jr., T. J. G. Francis, and G. B. Morris (1969), Anisotropy of the Pacific upper mantle, J. Geophys. Res., 74(12), 3095–3109, doi:10.1029/JB074i012p03095.

Ritzwoller, M. H., and A. L. Levshin (1998), Eurasian surface wave tomography: Group velocities, J. Geophys. Res., 103(B3), 4839–4878, doi:10.1029/97JB02622.

Shapiro, N. M., and M. Campillo (2004), Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophysical Research Letters*, *31*(7), doi: 10.1029/2004gl019491.

Shapiro, N. M., M. Campillo, L. Stehly, and M. H. Ritzwoller (2005), High-resolution surface-wave tomography from ambient seismic noise, Science, 307 (5715), 1615–1618.

Silver, P.G. and W.W. Chan (1988), Implications for continental scale structure and evolution from seismic anisotropy, Nature 335, 34-39.

Smith, M. L., and F. A. Dahlen (1973), Azimuthal Dependence of Love and Rayleigh-Wave Propagation in a Slightly Anisotropic Medium, *Journal of Geophysical Research*, *78*(17), 3321-3333, doi: Doi 10.1029/Jb078i017p03321.

Snieder, R. (2004), Extracting the Green’s function from the correlation of coda waves: A derivation based on stationary phase, *Physical Review E*, *69*(4), doi: 10.1103/PhysRevE.69.046610.

Tanimoto, T. and D.L. Anderson (1984), Mapping convection in the mantle, Geophys. Res. Lett. 11, 327-336.

Wiggins RA (1972), The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure. *Rev. Geophys. Space Phys*. **10**, 251–285.

Wu, H. and J.M. Lees, J. M. (1999), Cartesian Parameterization of Anisotropic Traveltime Tomography, Geophys. J. Int. 137, 64-80.

Yang, Y., Ritzwoller, M. H., Levshin, A. L. and Shapiro, N. M. (2007), Ambient noise Rayleigh wave tomography across Europe. Geophysical Journal International, 168: 259–274. doi: 10.1111/j.1365-246X.2006.03203.x

Yu., G.K. and B.J. Mitchell, Regionalized shear velocity moels of the Pacific upper mantle from observed Love and Rayleigh wave dispersion, Geophys. J. R. Astr. Soc. 57, 311-341, 1979.

Zha, Y., S. C. Webb, and W. Menke (2013), Determining the orientations of ocean bottom seismometers using ambient noise correlation, Geophysical Research Letters, 40(14), 3585-3590, doi: 10.1002/Grl.50698.

Zha, Y., S. C. Webb, S. S. Wei, D. A. Wiens, D. K. Blackman, W. H. Menke, R. A. Dunn, J. A. Conders (2014), Upper mantle shear velocity structure beneath the Eastern Lau Spreading Center from OBS ambient noise tomography, in press in Earth Planet. Sci. Lett., 2014

Mailing Address: William Menke, LDEO, 61 Route 9W, Palisades NY 10964 USA, MENKE@LDEO.COLUMBIA.EDU

APPENDIX A: FOURIER DATA KERNELS FOR THE 2D TOMOGRAPHY INVERSE PROBLEM

Here we formulate the 2D tomography problem using a Fourier (sines and cosines) representation of slowness. A Fourier basis has two advantages over the usual pixilated basis: the ray integrals can be performed analytically; and smoothness regularization can be implemented simply by suppressing higher wavenumber components. The ray integrals that appear in the formula for travel time (Equation 8) all have the form:

$$C_p \int_{ray_p} f[x(s), y(s)]\ ds \tag{A1}$$

where $f(x, y)$ is a smooth function of two spatial variables, (x, y). Ray p is a straight line connecting a source at (x_1, y_1) to a receiver at (x_2, y_2). The function $f(x, y)$ is meant to represent any of the material property functions, so $C_p = 1$ when $f = A$, $C_p = \cos 2\theta_p$ when $f = B_c$ and $C_p = \sin 2\theta_p$ when $f = B_s$.

We approximate the function $f(x, y)$ using a two-dimensional Fourier series:

$$f(x, y) = \sum_{i=0}^{N_x} \sum_{j=0}^{N_y} \left\{ A_{ij}\, f_{ij}^{cc}(x, y) + \ B_{ij}\, f_{ij}^{cs}(x, y) + \ C_{ij}\, f_{ij}^{sc}(x, y) + \ D_{ij}\, f_{ij}^{ss}(x, y) \right\} \tag{A2}$$

with basis functions:

$$\begin{gathered} f_{ij}^{cc}(x, y) = \cos\left(k_x^{(i)} x\right) \cos\left(k_y^{(j)} y\right) \ \ and \ \ f_{ij}^{cs}(x, y) = \cos\left(k_x^{(i)} x\right) \sin\left(k_y^{(j)} y\right) \\ f_{ij}^{sc}(x, y) = \sin\left(k_x^{(i)} x\right) \cos\left(k_y^{(j)} y\right) \ \ and \ \ f_{ij}^{ss}(x, y) = \sin\left(k_x^{(i)} x\right) \sin\left(k_y^{(j)} y\right) \end{gathered} \tag{A3}$$

These basis functions contain the spatial wavenumbers:

$$k_x^{(i)} = i\Delta k_x \ \text{ and } \ k_y^{(j)} = j\Delta k_y \tag{A4}$$

All coefficients multiplying sines of zero wavenumber are constrained to be equal to zero:

$$B_{i0} = C_{0j} = D_{i0} = D_{j0} = 0 \tag{A5}$$

The spatial wavenumbers have uniform spacing Δk_x and Δk_y along the wavenumber axes. Thus, the function $f(x, y)$ is represented by $K = 4N_x N_y - 2N_x - 2N_y + 1$ real coefficients (or model parameters), A_{ij}, B_{ij}, C_{ij} and D_{ij}. The motivation for using a Fourier basis is that smoothness constraints easily can be implemented by preferentially damping the higher wavenumber coefficients. We use here a sine and cosine basis, as contrasted to a complex exponential basis, because the latter would require complicated constraints on the symmetry of the complex coefficients in order to guarantee that $f(x, y)$ is purely real.

We now insert the Fourier series into the line integral and rearrange:

$$\sum_{i=0}^{N_x} \sum_{j=0}^{N_y} \{ G_{pij}^{cc} A_{ij} + G_{pij}^{cs} B_{ij} + G_{pij}^{sc} C_{ij} + G_{pij}^{ss} D_{ij} \}$$

$$G_{pij}^{cc} = C_p \int_{ray_p} f_{ij}^{cc}[x(s), y(s)]\, ds \;\text{ and }\; G_{pij}^{cs} = C_p \int_{ray_p} f_{ij}^{cs}[x(s), y(s)]\, ds$$

$$G_{pij}^{sc} = C_p \int_{ray_p} f_{ij}^{sc}[x(s), y(s)]\, ds \;\text{ and }\; G_{pij}^{ss} = C_p \int_{ray_p} f_{ij}^{ss}[x(s), y(s)]\, ds$$

(A6)

Here G_{pij}^{cc}, G_{pij}^{cs}, G_{pij}^{sc} and G_{pij}^{ss} are data kernels that relate the model parameters to the travel time data via a linear algebraic equation. The line integrals can be performed analytically, since the integrands are elementary trigonometric functions and since x and y are linear functions of arc-length, s ($x = a + bs$ and $y = c + ds$, as in Equation 7). The result is:

$$\begin{aligned} G_{pij}^{cc} = C_p\Big\{ & \cos\big(k_x^{(i)}a\big)\cos\big(k_y^{(j)}c\big)\; I_{cc}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & - \cos\big(k_x^{(i)}a\big)\sin\big(k_y^{(j)}c\big)\; I_{cs}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & - \sin\big(k_x^{(i)}a\big)\cos\big(k_y^{(j)}c\big)\; I_{sc}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & + \sin\big(k_x^{(i)}a\big)\sin\big(k_y^{(j)}c\big)\; I_{ss}\big(k_x^{(i)}b, k_y^{(j)}d, L\big)\Big\} \end{aligned}$$

$$\begin{aligned} G_{pij}^{cs} = C_p\Big\{ & \sin\big(k_x^{(i)}a\big)\cos\big(k_y^{(j)}c\big)\; I_{cc}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & - \sin\big(k_x^{(i)}a\big)\sin\big(k_y^{(j)}c\big)\; I_{cs}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & + \cos\big(k_x^{(i)}a\big)\cos\big(k_y^{(j)}c\big)\; I_{sc}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & - \cos\big(k_x^{(i)}a\big)\sin\big(k_y^{(j)}c\big)\; I_{ss}\big(k_x^{(i)}b, k_y^{(j)}d, L\big)\Big\} \end{aligned}$$

$$\begin{aligned} G_{pij}^{sc} = C_p\Big\{ & \cos\big(k_x^{(i)}a\big)\sin\big(k_y^{(j)}c\big)\; I_{cc}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & + \cos\big(k_x^{(i)}a\big)\cos\big(k_y^{(j)}c\big)\; I_{cs}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & - \sin\big(k_x^{(i)}a\big)\sin\big(k_y^{(j)}c\big)\; I_{sc}\big(k_x^{(i)}b, k_y^{(j)}d, L\big) \\ & - \sin\big(k_x^{(i)}a\big)\cos\big(k_y^{(j)}c\big)\; I_{ss}\big(k_x^{(i)}b, k_y^{(j)}d, L\big)\Big\} \end{aligned}$$

$$G_{pij}^{ss} = C_p\Big\{\sin\big(k_x^{(i)}a\big)\sin\big(k_y^{(j)}c\big)\; I_{cc}\big(k_x^{(i)}b, k_y^{(j)}d, L\big)$$

$$+\sin\left(k_x^{(i)}a\right)\cos\left(k_y^{(j)}c\right)\ I_{cs}\left(k_x^{(i)}b,k_y^{(j)}d,L\right)$$

$$+\cos\left(k_x^{(i)}a\right)\sin\left(k_y^{(j)}c\right)\ I_{sc}\left(k_x^{(i)}b,k_y^{(j)}d,L\right)$$

$$+\cos\left(k_x^{(i)}a\right)\cos\left(k_y^{(j)}c\right)\ I_{ss}\left(k_x^{(i)}b,k_y^{(j)}d,L\right)\Big\}$$

(A7)

Here the *I's* are the integrals:

$$I_{cc}(a_1,a_2,s_0)=\int_0^{s_0}\cos(a_1 s)\cos(a_2 s)ds=\frac{\sin\{(a_2-a_1)s_0\}}{2(a_2-a_1)}+\frac{\sin\{(a_2+a_1)s_0\}}{2(a_2+a_1)}$$

$$I_{cs}(a_1,a_2,s_0)=-\frac{\cos\{(a_2-a_1)s_0\}}{2(a_2-a_1)}-\frac{\cos\{(a_2+a_1)s_0\}}{2(a_2+a_1)}+\frac{a_2}{{a_2}^2-{a_1}^2}$$

$$I_{sc}(a_1,a_2,s_0)=\frac{\cos\{(a_2-a_1)s_0\}}{2(a_2-a_1)}-\frac{\cos\{(a_2+a_1)s_0\}}{2(a_2+a_1)}-\frac{a_1}{{a_2}^2-{a_1}^2}$$

$$I_{ss}(a_1,a_2,s_0)=\frac{\sin\{(a_2-a_1)s_0\}}{2(a_2-a_1)}-\frac{\sin\{(a_2+a_1)s_0\}}{2(a_2+a_1)}$$

(A8)

Note that in the limit $s_0\to 0$, these integrals all approach zero. Note also that the integrals have removable singularities as $|a_2|\to|a_1|$, In the $a_2\to a_1$, case we find:

$$I_{cc}(a_1,a_2,s_0)\approx\frac{s_0}{2}-\frac{(a_2-a_1)^2{s_0}^3}{12}+\frac{\sin\{(a_2+a_1)s_0\}}{2(a_2+a_1)}$$

$$I_{cs}(a_1,a_2,s_0)\approx\frac{1}{4}(a_2-a_1)\,{s_0}^2-\frac{[\cos\{(a_2+a_1)s_0\}-1]}{2(a_2+a_1)}$$

$$I_{sc}(a_1,a_2,s_0)\approx-\frac{1}{4}(a_2-a_1)\,{s_0}^2-\frac{[\cos\{(a_2+a_1)s_0\}-1]}{2(a_2+a_1)}$$

$$I_{ss}(a_1,a_2,s_0)\approx\frac{s_0}{2}-\frac{(a_2-a_1)^2{s_0}^3}{12}-\frac{\sin\{(a_2+a_1)s_0\}}{2(a_2+a_1)}$$

(A9)

And in the $a_2\to -a_1$ case, we find:

$$I_{cc}(a_1,a_2,s_0)\approx\frac{\sin\{(a_2-a_1)s_0\}}{2(a_2-a_1)}+\frac{s_0}{2}-\frac{(a_2+a_1)^2{s_0}^3}{12}$$

$$I_{cs}(a_1,a_2,s_0)\approx-\frac{[\cos\{(a_2-a_1)s_0\}-1]}{2(a_2-a_1)}+\frac{1}{4}(a_2+a_1)\,{s_0}^2$$

$$I_{sc}(a_1,a_2,s_0) \approx \frac{[\cos\{(a_2-a_1)s_0\}-1]}{2(a_2-a_1)} + \frac{1}{4}(a_2+a_1)\,{s_0}^2$$

$$I_{ss}(a_1,a_2,s_0) \approx \frac{\sin\{(a_2-a_1)s_0\}}{2(a_2-a_1)} - \frac{s_0}{2} + \frac{(a_2+a_1)^2{s_0}^3}{12}$$

(A10)

A typical tomography problem has many thousands of rays, so in all likelihood a few of them will correspond to these exceptional cases. Software that implements the tomographic inversion must therefore detect and deal with them.

In an anisotropic tomography problem, each of the three material property functions is represented by its own Fourier series. The series for $A(x,y)$ has coefficients, say, $\left(A_{ij}^A, B_{ij}^A, C_{ij}^A, D_{ij}^A\right)$, the series for $B_c(x,y)$, $\left(A_{ij}^C, B_{ij}^C, C_{ij}^C, D_{ij}^C\right)$ and the series for $B_s(x,y)$, $\left(A_{ij}^S, B_{ij}^S, C_{ij}^S, D_{ij}^S\right)$. All of these coefficients can be grouped into a single model parameter vector, **m**, of length $M = 3K \approx 12N_xN_y$. The travel time measurements can be arranged in a vector, **d**, of length, say, N. Data and model parameters are connected by the linear matrix equation $\mathbf{d} = \mathbf{Gm}$, where the elements of the matrix, **G**, are the data kernels derived above. This equation can be solved by a standard method, such as generalized least squares.

In our implementation, we add a second equation, $\mathbf{0} = \mathbf{Hm}$, the effect of which is to suppress (or *damp*) the higher wavenumber components of the model. The matrix, **H**, is an $M \times M$ diagonal matrix whose elements depend upon the wavenumbers of the corresponding model parameter and whether it belongs to the Fourier series of the isotropic function A or the anisotropic functions B_c and B_s.

$$A\text{: } H_{ii} = w^{iso}({k_x}^2 + {k_y}^2)^{P^{iso}}$$

$$B_c \text{ and } B_s\text{: } H_{ii} = w^{ani}({k_x}^2 + {k_y}^2)^{P^{ani}}$$

(A11)

The relative smoothness of the isotropic and anisotropic parts of the estimated model can be controlled by judicious choice of the constants w^{iso}, w^{ani}, P^{iso} and P^{ani}.

APPENDIX B: EQUIVALENT POINT HETEROGENEITIES FOR RADON'S PROBLEM

Anisotropic Heterogeneity Equivalent to a Point Isotropic Heterogeneity. Our goal is to design a pattern of anisotropy (B_s, B_c) that is equivalent to a point isotropic heterogeneity at the origin, in the sense that both lead to travel time $\delta T = \delta T_0$ for rays passing through the origin, and zero travel time for rays that miss the origin. The problem has radial symmetry, so we work in polar coordinates (r,θ). Because of the symmetry, the slow axis of anisotropy θ_0 everywhere must point away from the origin (that is, $\theta_0 = \theta$), so:

$$\begin{bmatrix} B_s \\ B_c \end{bmatrix} = \begin{bmatrix} \sin 2\theta \\ \cos 2\theta \end{bmatrix} R(r) \quad \text{since} \quad \theta_0 = ½\tan^{-1}(B_s/B_c) = ½\tan^{-1}(\tan 2\theta) = \theta$$

(B1)

Here $R(r)$ is an as yet undetermined function that depends only upon radius, r. Note that $B = \left(B_s{}^2 + B_c{}^2\right)^{1/2} = R(r)$. Now consider an indefinitely long straight-line ray that passes a distance r_0 from the origin (Figure 10). Since the problem has radial symmetry, we may consider this ray to be parallel to the x-axis without loss of generality. A point (x, r), with $r = (x^2 + r_0^2)^{1/2}$, on the ray makes an angle θ with respect to the slow axis of anisotropy (that is, the radial direction). The travel time δT is the integral of $B \cos 2\theta$ along this ray. Note that:

$$B \cos 2\theta = R(r) \cos 2\theta = R(r) \cos^2 \theta - R(r) \sin^2 \theta = \frac{x^2}{x^2 + r_0^2} R(r) - \frac{r_0^2}{x^2 + r_0^2} R(r) \tag{B2}$$

The function $R(r)$ must be chosen so that:

$$\delta T(r_0) = \int_{-\infty}^{+\infty} B \cos 2\theta \, dx = 0 \quad \text{for} \quad r_0 > 0 \tag{B3}$$

The reader may verify that the correct choice is $R(r) = Cr^{-2}$, where C is an arbitraty constant, by using integrals 2.173.1 and 2.175.4 of Gradshteyn and Ryzhik (1980) (a result that we have also checked numerically). The travel time along the $r_0 = 0$ ray is infinite, since the function Cr^{-2} has a non-integrable singularity at the origin and the ray passes through it. However, the radial symmetry of the problem actually implies zero – not infinite - anisotropy at the origin. We resolve this inconsistency by defining a scale length ε over which the anisotropy falls to zero:

$$R(r) = C \frac{r^2}{(r^4 + \varepsilon^4)} \tag{B4}$$

This function behaves as Cr^{-2} when $r \gg \varepsilon$ and as $C\varepsilon^{-4}r^2$ when $r \ll \varepsilon$. It is integrable because it has no singularity at the origin. The reader may verify that the choice $C = \sqrt{2}\, \varepsilon \delta T_0 / \pi$ leads to a ray with travel time $\delta T = \delta T_0$, by using integral 2.132.3 of Gradshteyn and Ryzhik (1980) (a result that we have also checked numerically). The equivalent anomaly is then:

$$\begin{bmatrix} B_s \\ B_c \end{bmatrix} = \frac{\sqrt{2}\, \delta T_0}{\pi} \frac{\varepsilon r^2}{(r^4 + \varepsilon^4)} \begin{bmatrix} \sin 2\theta \\ \cos 2\theta \end{bmatrix} \tag{B5}$$

This result indicates that the anisotropic heterogeneity equivalent to a point isotropic heterogeneity is not point-like, but rather is spatially-distributed. Furthermore, while its intensity falls off with distance, it does so relatively slowly, as (distance)$^{-2}$.

The sum of the spatially-distributed anisotropic anomaly and the negative of the point-like isotropic anomaly is a null solution, meaning that it has no travel time anomaly. Any number of these null solutions can be added to the estimated model without changing the degree to which it fits the data.

Isotropic Heterogeneity Equivalent to a Point Anisotropic Heterogeneity. Our goal is to design an isotropic heterogeneity $A(r,\theta)$ (where (r,θ) are polar coordinates) that is equivalent to a point anisotropic heterogeneity at the origin, in the sense that both lead to travel time $\delta T = \delta T_0\ \cos 2(\theta-\theta_0)$ for rays passing through the origin, and zero travel time for rays that miss the origin. Here θ_0 is the azimuth of the slow axis of anisotropy. Inspired by the previous result, we try the function:

$$A(r,\theta) \propto \frac{\cos 2(\theta-\theta_0)}{r^2} \tag{B6}$$

As before, we must demonstrate that the ray integral is zero for any ray passing a distance $r_0 > 0$ away from the origin. Since θ_0 is arbitrary, we can choose the ray to be parallel to the x-axis without loss of generality (Figure 9). We now manipulate (A2.6) using standard trigonometric identities:

$$\begin{aligned} \frac{\cos 2(\theta-\theta_0)}{r^2} &= \cos(2\theta_0)\frac{\cos(2\theta)}{r^2} + \sin(2\theta_0)\frac{\sin(2\theta)}{r^2} \\ &= \cos(2\theta_0)\frac{[\cos^2(\theta)-\sin^2(\theta)]}{r^2} + \sin(2\theta_0)\frac{\sin(2\theta)}{r^2} \\ &= \cos(2\theta_0)\frac{x^2-r_0^2}{(x^2+r_0^2)^2} + \sin(2\theta_0)\frac{r_0 x}{(x^2+r_0^2)^2} \end{aligned} \tag{B7}$$

The ray integral of the first term has already been shown to be zero. The ray integral of the second term is zero because the second term is an odd function of x. Thus, the travel time of all rays with $r_0 > 0$ is zero.

As in the previous section, the travel time along the $r_0 = 0$ ray is infinite, since the function r^{-2} has a non-integrable singularity at the origin and the ray passes through it. However, depending upon the ray orientation, (A2.5) implies that the point at the origin has both negative and positive A – a contradiction. As before, we resolve this inconsistency by requiring that the heterogeneity falls to zero within a small distance ε of the origin. The heterogeneity is then:

$$A(r,\theta) = \frac{\sqrt{2}\,\delta T_0}{\pi}\frac{\varepsilon r^2}{(r^4+\varepsilon^4)}\ \cos 2(\theta-\theta_0) \tag{B8}$$

This anomaly is similar in form to the one given in Equation 45 of Mochizuki (1997) for the isotropic anomaly equivalent to a spatially-compact anisotropic heterogeneity with a Gaussian spatial pattern. The sum of the spatially-distributed isotropic anomaly and the negative of the point-like anisotropic anomaly is another null solution.

Figures and Captions

Figure 1

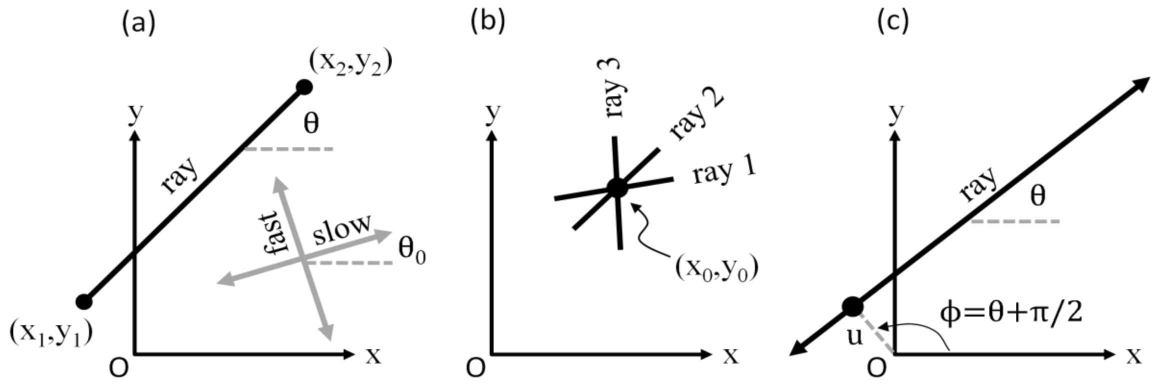

Fig. 1. (a) Coordinate system used in this paper. Ray (black line) has with azimuth θ and endpoints at (x_1, y_1) and (x_2, y_2). Slow and fast directions of anisotropy (grey lines) have azimuth θ_0 and $\theta_0 + \pi/2$, respectively. (b) Star array consisting of three short rays, centered at point (x_0, y_0). (c) In Radon's problem, the position an orientation of a ray is parameterized by its distance u of closest approach to the origin and by the azimuth ϕ of the ray-perpendicular direction. Note that $\phi = \theta + \pi/2$.

Figure 2

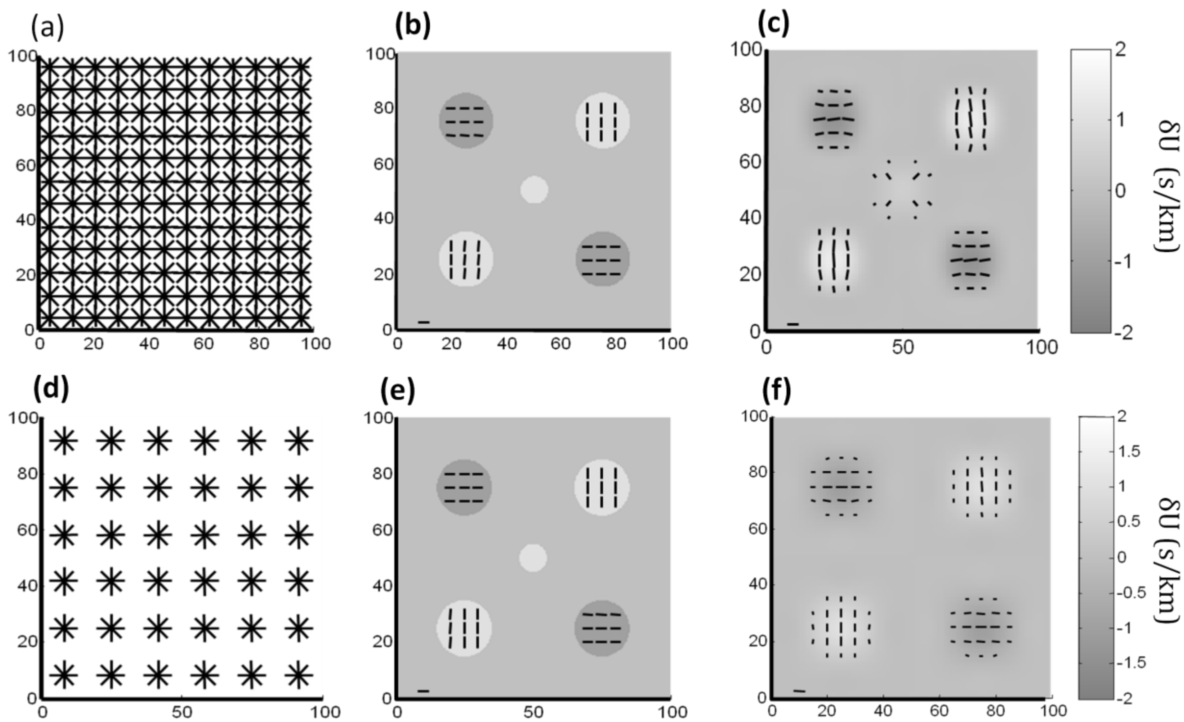

Fig. 2. Model estimated by a grid of star arrays. (a) Cartesian grid of star arrays. (b) The true model consists of circular heterogeneities. Each heterogeneity has a constant isotropic part, A, (depicted in grey shades), anisotropic part, B, and slow axis, θ_0 (depicted with black bars, whose length scales with B and whose orientation reflects θ_0). The bold bar in the lower left corresponds to $B = 0.2$. (c) Estimated model. (d)-(f) Same as (a)-(c), except for a sparser grid of larger star arrays.

Figure 3

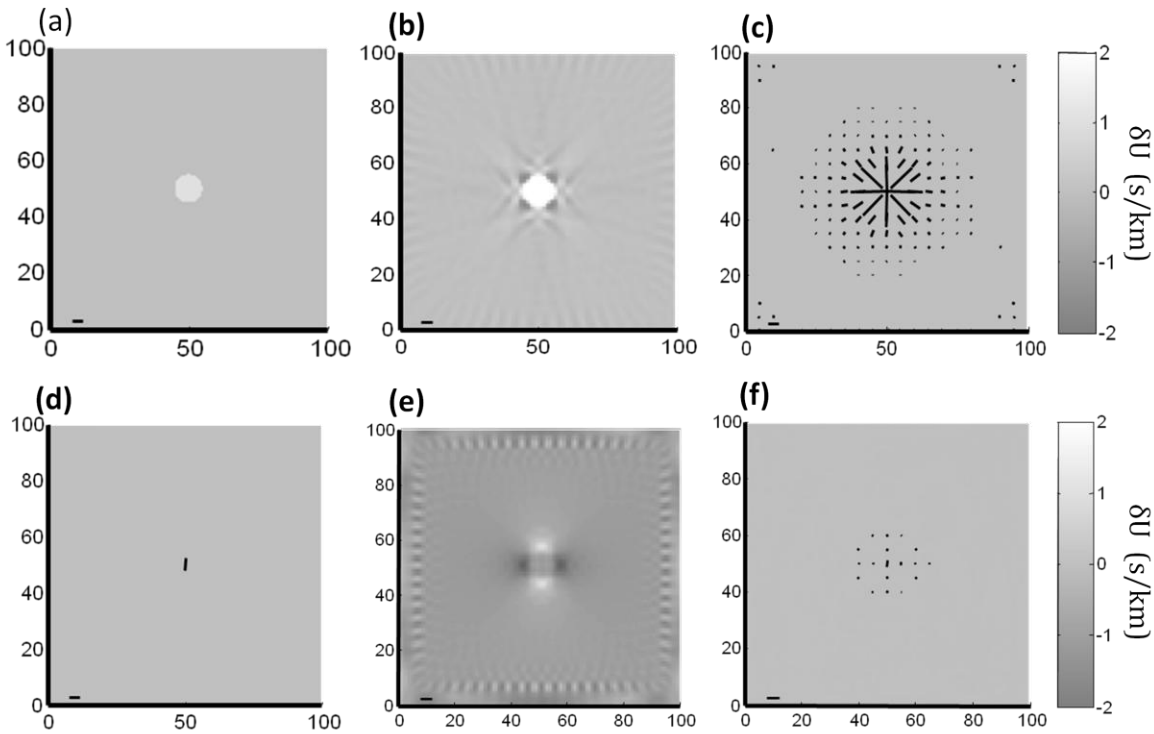

Fig. 3. Equivalent Heterogeneities for Radon's Problem. (a) True model has purely isotropic circular heterogeneity ($A = 1$) at its center. (b) Purely isotropic estimated model. (c) Purely anisotropic estimated model. (d) True model has purely anisotropic circular heterogeneity ($B = 0.2, \theta_0 = \pi/2$) at its center. (e) Purely isotropic estimated model. (f) Purely anisotropic estimated model. All estimated models have less than 1% travel time error.

Figure 4

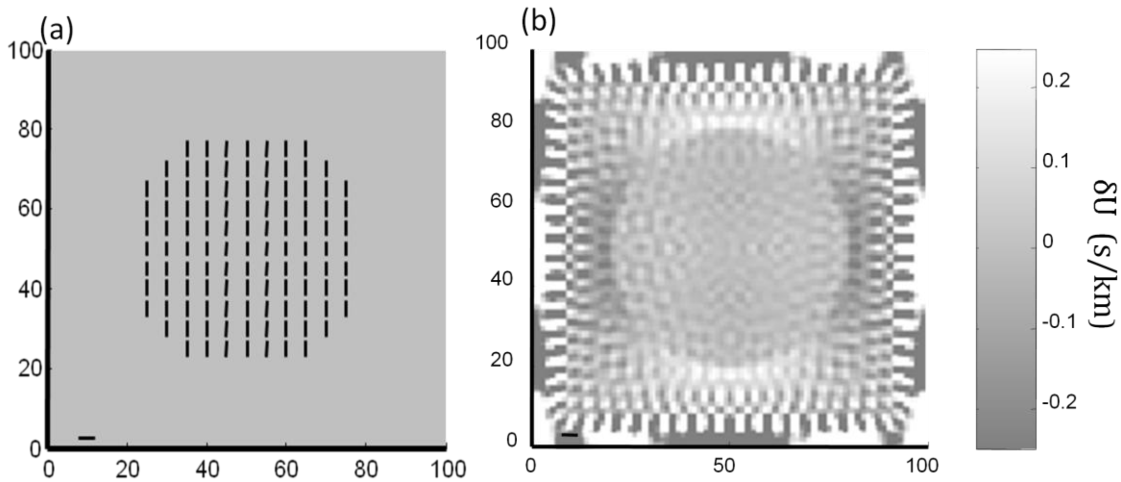

Fig. 4. Equivalent Heterogeneities for Radon's Problem. (a) True model has a large, circular, purely anisotropic heterogeneity ($B = 0.2, \theta_0 = \pi/2$)) at its center. (b) Purely isotropic estimated model, which has less than 1% error, has strongest heterogeneity around its edges.

Figure 5

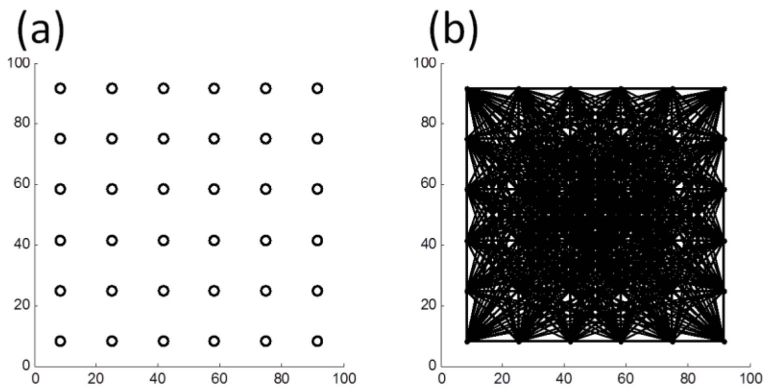

Fig. 5. (a) Regular grid of stations. (b) Rays connecting all pairs of stations are used in the tomographic inversion.

Figure 6

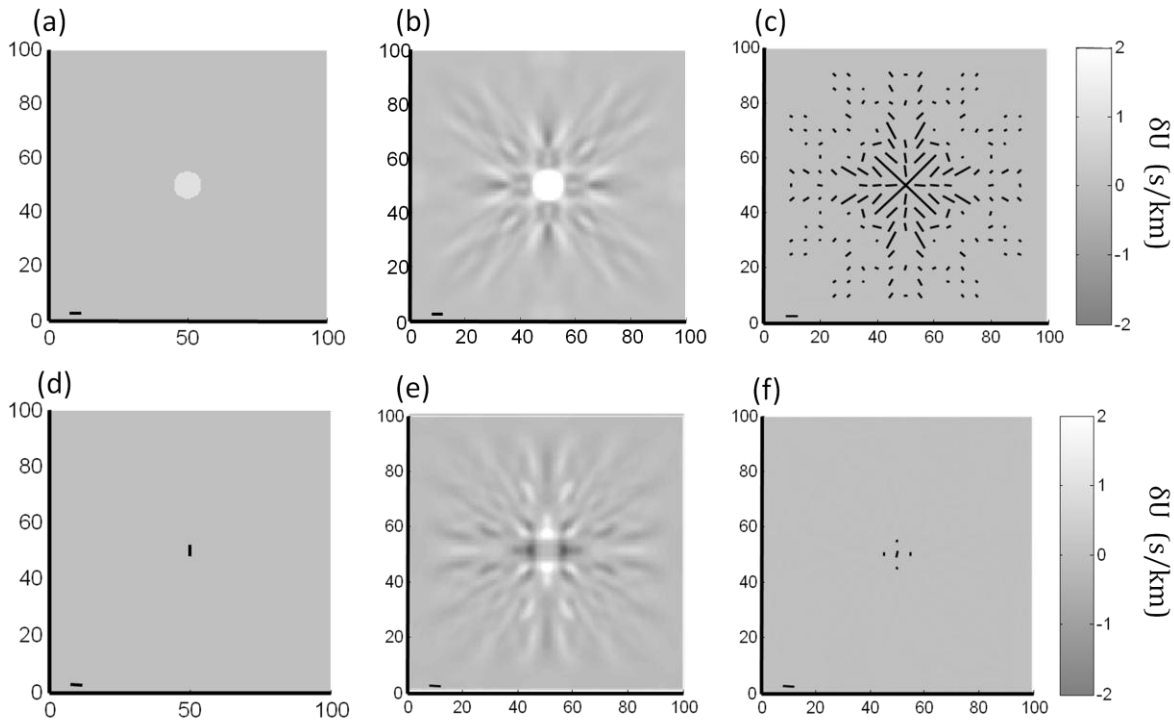

Fig. 6. Equivalent Heterogeneities for a regular grid of stations. (a) True model has purely isotropic circular heterogeneity ($A = 1$) at its center. (b) Purely isotropic estimated model. (c) Purely anisotropic estimated model. (d) True model has purely anisotropic circular heterogeneity ($B = 0.2, \theta_0 = \pi/2$) at its center. (e) Purely isotropic estimated model. (f) Purely isotropic estimated model. All estimated models have less than 1% travel time error.

Figure 7

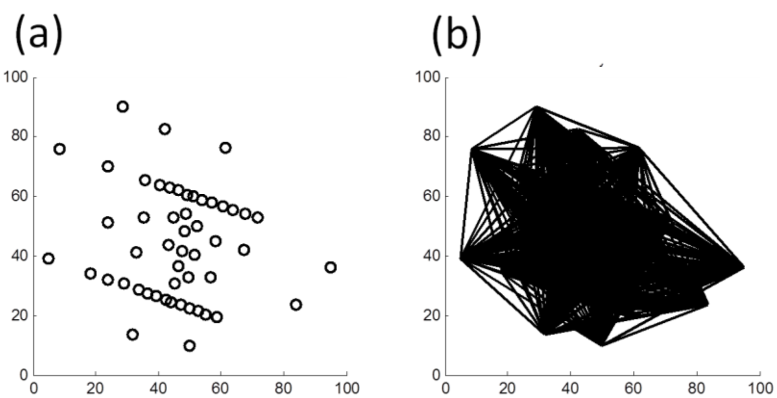

Fig. 7. (a) Irregular array of stations, with a shape similar to the 2009-2010 Eastern Lau Spreading Center array. (b) Rays between all stations separated by at least 20 km.

Figure 8

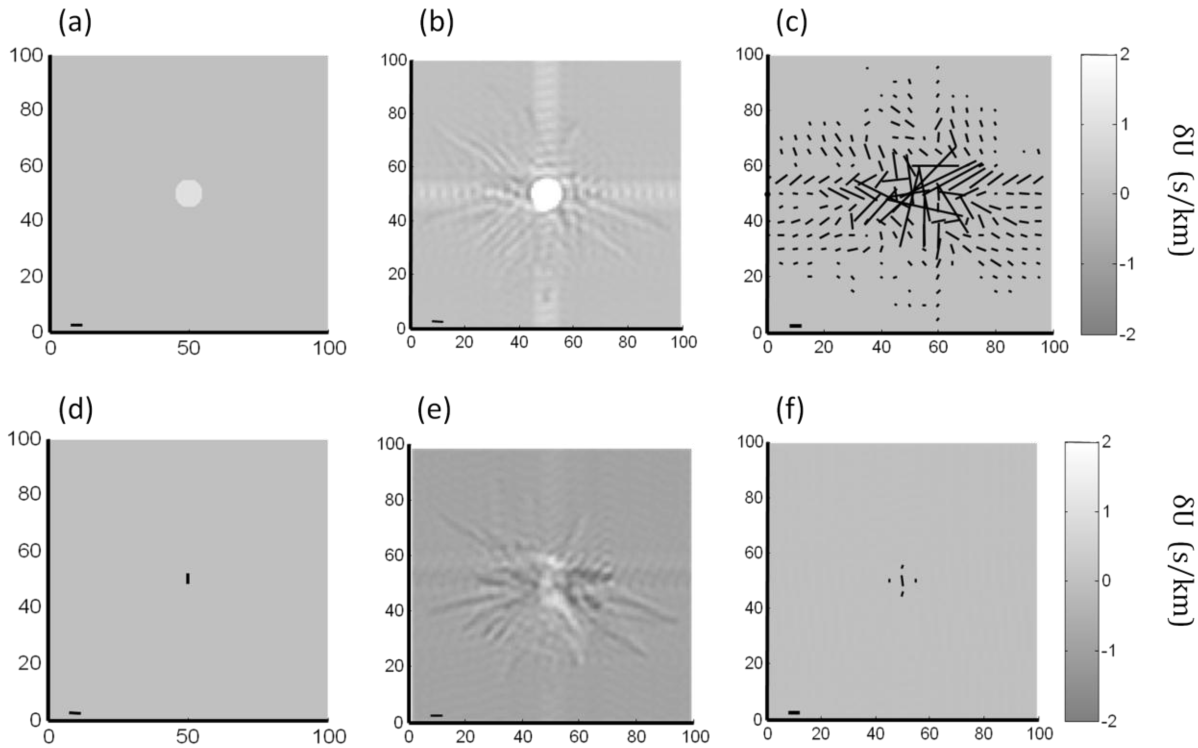

Fig. 8. Equivalent Heterogeneities for irregular array. (a) True model has purely isotropic circular heterogeneity ($A = 1$) at its center. (b) Purely isotropic estimated model. (c) Purely anisotropic estimated model. (d) True model has purely anisotropic circular heterogeneity ($B = 0.2, \theta_0 = \pi/2$) at its center. (e) Purely isotropic estimated model. (f) Purely anisotropic estimated model. All estimated models have less than 1% travel time error.

Figure 9

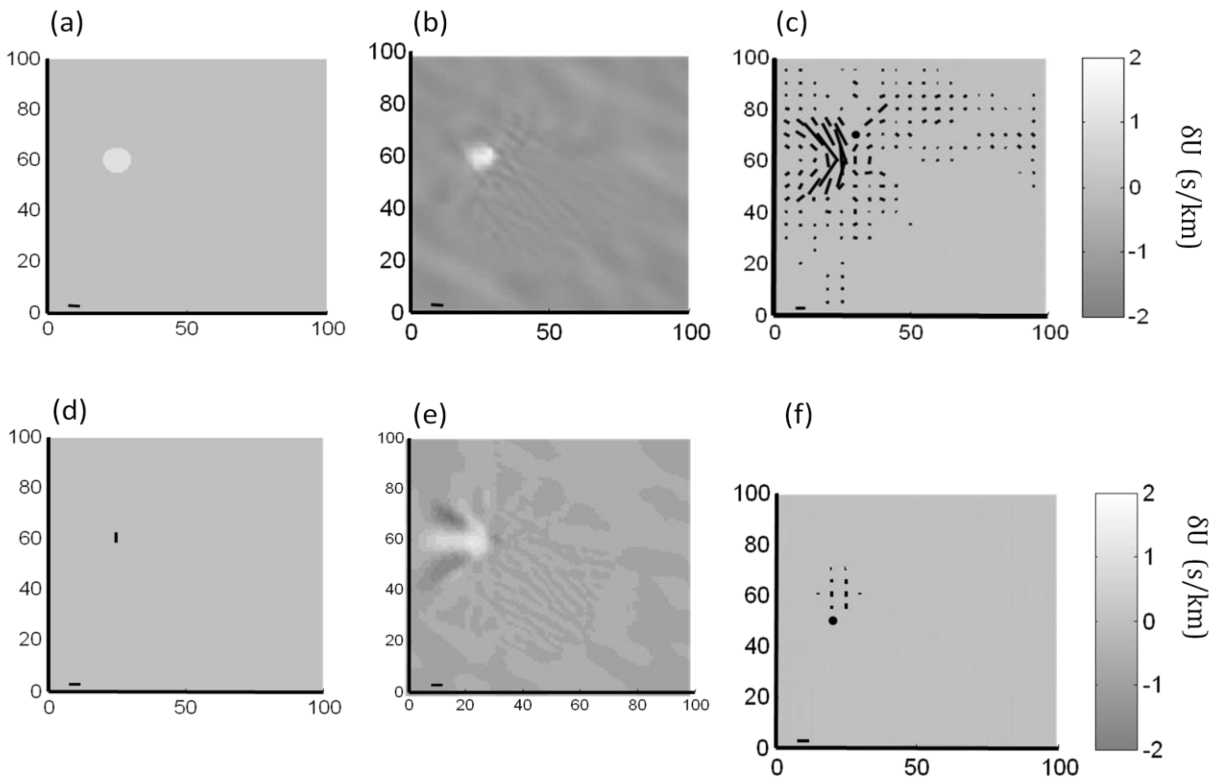

Fig. 9. Equivalent Heterogeneities for irregular array. Same as Figure 8, except with the heterogeneity moved to the edge of the array.

Figure 10

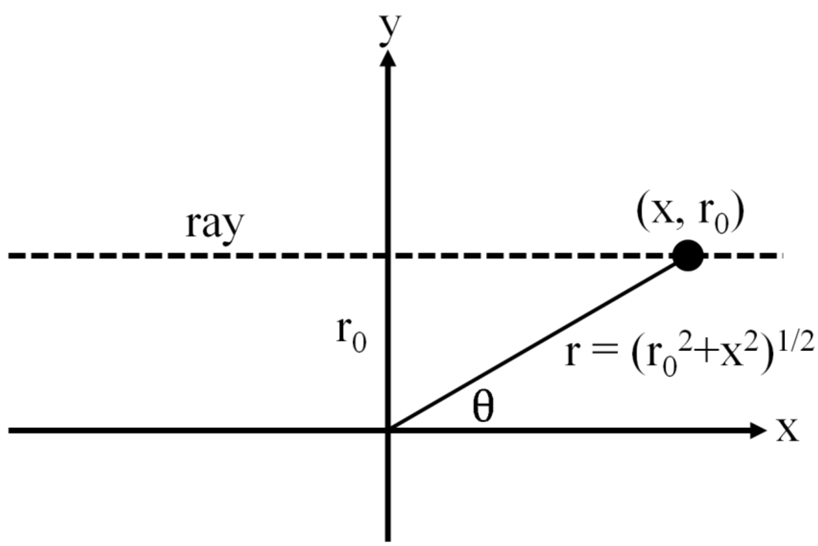

Fig. 10. Geometry of ray used in travel time integral.