Geophys. J. Int. (2000) **142,** 000–000

Global P and PP traveltime tomography: rays versus waves

R. Montelli[1], G. Nolet[1], G. Masters[2], F.A. Dahlen[1], S-H. Hung[3]

[1] *Department of Geosciences, Guyot Hall, Princeton University, Princeton NJ* 08540*, USA*

[2] *Institute of Geophysics and Planetary Physics, University of California San Diego, La Jolla, CA* 92093*, USA*

[3]*Department of Geosciences, National Taiwan University, Taipei, Taiwan*

SUMMARY

This paper presents a comparison of ray-theoretical and finite-frequency traveltime tomography for compressional waves. Our data set consists of 86,405 long-period P and PP-P traveltimes measured by cross-correlation. The traveltime of a finite-frequency wave is sensitive to anomalies in a hollow banana-shaped region surrounding the unperturbed ray path, with the sensitivity being zero on the ray. Because of the minimax nature of the surface reflected PP wave, its sensitivity is more complicated. We compute the 3D traveltime sensitivity efficiently by using the paraxial approximation in conjunction with ray theory and the Born approximation. We compare tomographic models with the same χ^2 fit, for both ray theory and finite-frequency analysis. Depending on the depth and size of the anomaly, the amplitudes of the velocity perturbations in the finite-frequency tomographic images are 30%–50% larger than in the corresponding ray-theoretical images, demonstrating that wavefront healing cannot be neglected when interpreting long-period seismic waves. The images obtained provide clear evidence that a limited number of hotspots are fed by plumes originating in the lower mantle.

Key words: global seismology, tomography, P waves, ray theory, Fréchet derivatives, mantle plumes

1 INTRODUCTION

Global P-wave tomographic models have so far been obtained by applying ray theory. In this paper, we investigate the effects that the ray approximation has on the tomographic images, especially on smaller objects. Waves propagate as rays only in the high-frequency limit of the elastodynamic equations of motion. All scattering interactions of the wave with the heterogeneities in the propagation medium are neglected under the assumption that the velocity field varies slowly on the scale of the wavelength. Rays might bend and be deviated by the velocity structure, but energy is conserved along the ray and is only influenced by the earth's properties along an infinitesimally narrow path that follows Snell's law. The velocity information contained in a P or S wave is reduced to a single number, the time of the first break, which is assumed to correspond to the arrival of the highest-frequency observable wave. This simplifies the mathematics, but it is quite far from physical reality for finite-frequency waves.

If the scale length of 3D velocity heterogeneities is comparable to the width of the Fresnel zone, finite-frequency effects are important. In regions where shadow zones or strong diffractors are present, waves may scatter or diffract, and ray theory also breaks down. In ray theory, waves preserve the time shifts accrued upon passage through an anomaly somewhere along its path. Because of an intrinsic diffraction phenomenon called "wavefront healing", finite-frequency wavefronts do not. Diffraction acts to fill in or heal irregularities in the wavefront. Also, diffracted waves of significant amplitude might interfere with the direct wave and introduce a bias in the traveltime measurements. Consequently, the traveltime of a finite-frequency seismic wave is sensitive to velocity anomalies off the geometrical ray. Simple diffraction theory shows that a region around the ray path affects the traveltimes, whereas structure far from the ray paths play minor roles. Such a volume surrounding the geometrical ray path is called the first Fresnel zone and is loosely defined as the region where significant constructive interference of seismic energy takes place (Wielandt, 1987; Nolet, 1987; Nolet, 1990; Müller et al., 1992; Nolet & Dahlen, 2000). The size of the Fresnel zone depends on the epicentral distance and on the frequency content of the propagated wavefield (Kravtsov, 1988): the higher the frequency, the narrower the Fresnel zone.

The widespread availability of broadband digital data has led to the recent development of accurate techniques for traveltime measurements using cross-correlation of an observed body-wave phase with the corresponding spherical-earth synthetic phase (Bolton & Masters, 2001). Cross-correlation methods have also been used to measure the differential traveltime of two phases at the same station (Kuo et al., 1987; VanDecar & Crosson, 1990; Woodward & Masters, 1991; Su et al., 1994). The measurement obtained in this way provides an integrated arrival-time difference between two waveforms, not simply the difference between onset times. Therefore the ray-theoretical description for a traveltime along the ray may no longer be valid, and finite-frequency waveform modeling may be required. A

major aim of this paper is to investigate the necessity of finite-frequency theory for broadband seismic data.

The idea of using a ray of non-zero width to bridge the gap between rays and waves dates back at least to Hagendoorn (1954). He introduced the concept of beam width defined as the region falling within the first Fresnel zone. Various later attempts have been made to compute Fresnel or influence zones for bandlimited seismic traveltimes in two and three dimensions (Gelchinsky, 1985; Woodward, 1992; Yomogida, 1992; Cardimora & Garmany, 1993; Stark & Nikolayev, 1993; Vasco & Majer, 1993; Li & Tanimoto, 1993; Marquering et al., 1998; Marquering et al., 1999; Dahlen et al., 2000; Hung et al., 2000; Zhao et al., 2000; Ritzwoller et al., 2002; Yoshizawa & Kennett, 2002). These investigations use the single scattering or first-order Born approximation, the Rytov approximation, or the Kirchhoff approximation to compute Fréchet sensitivity kernels that relate traveltime perturbations to velocity anomalies.

Because of the computational difficulties that accompany a three-dimensional formulation of sensitivity kernels for P or S waves, little effort has been put so far into using them for seismic tomography, at least outside the geophysical exploration community. Castle et al. (2000) and Husen & Kissling (2001) use what they call fat rays for the shear wave speed anomalies at the base of the mantle, and a tomographic study of the Antofagasta area (Northern Chile), respectively. However, fat rays account only qualitatively for wavefront healing effects.

Marquering et al. (1998; 1999) and Zhao et al. (2000) present theoretical expressions for the Fréchet kernel for delay times, obtained by summing surface waves and normal modes, respectively. Zhao et al. (2001) uses such expressions in a 3D tomographic study of the Western Pacific region. The normal-mode kernel provides the most general description of the sensitivity kernel around the unperturbed ray. However, since mode summation is computationally expensive, the implementation of the exact kernel in a global inversion of large data sets is not presently feasible. Dahlen et al. (2000) provide an alternative procedure to economically compute the Fréchet kernel of a finite-frequency traveltime measured by cross-correlation of a broad-band waveform with a spherical-earth synthetic seismogram. The Green's function and the response to a point scatterer in Born theory are represented as a sum of rays rather than modes. The paraxial approximation renders the computation of such kernels much more efficient. In this paper we apply the 3D Fréchet kernel formalism of Dahlen et al. (2000) to a global, finite-frequency traveltime dataset, and compare the result with the 3D model obtained by inverting the same data using conventional ray theory.

2 INGREDIENTS OF THE INVERSE PROBLEM

2.1 Data and reference model

We use arrival times of P and PP waves with 20 s dominant period (Bolton & Masters, 2001). We invert 66,238 P traveltimes and 20,167 PP–P differential traveltimes. Systematic and correctable timing errors were introduced by software bugs in various versions of the Quanterra data logger. Measurements have been appropriately corrected for such timing errors, and data obtained from stations with periods of erratic timing have been removed from further analysis.

Absolute times are mainly affected by noise and errors due to source mislocation. They are measured by cross-correlation of an observed pulse with a synthetic which is constructed by convolving the impulse response of the instrument at Albuquerque (ANMO) with a t^* attenuation operator (Bolton & Masters, 2001); the attenuation time t^* is kept constant at 1 s for P waves.

Differential PP–P times are obtained by cross-correlation of the Hilbert-transformed direct P phase with the reflected PP phase (Woodward & Masters, 1991). Here a 1 s t^* operator accounts for the different attenuation histories of the direct and reflected phase, respectively. Differential traveltimes have the advantage of eliminating source and receiver bias, and are most sensitive to shallow structure in the vicinity of the bounce point, thus allowing us to constrain the upper mantle beneath regions of the world where there are no sources or receivers.

Predicted absolute and differential times are computed using the *iasp91* velocity model (Kennett & Engdahl, 1991). We correct for the signal due to ellipticity and for the effect of the crust (including topography). Crustal corrections are computed using the $2^\circ \times 2^\circ$ global crustal model CRUST2.0 (model available through the Reference Earth Model web site: *http://mahi.ucsd.edu/Gabi/rem.html*). The remaining residuals show a baseline shift of about –1 s for PP–P and about +4 s for P (Fig. 1a). Although the origin of these offsets is not very clear, major candidates to explain them are the use of NEIC source locations and/or imperfections in the 1D reference model. Deviations from our assumed constant t* are not able to explain this time shift. Variations in the delays due to attenuation are of the same order as the variations in t*, which are of order of 0.1 s (see Stewart (1984) and eq.(6) in Bolton & Masters (2001)), and therefore far too small to explain the 4 s shift.

Since differential times such as PP–P are insensitive to source mislocations, the –1 s offset is best explained by the inadequacy of the 1D reference velocity model used. We eliminated this offset by making a slight change to the *iasp91* velocity model in the upper mantle transition zone (Fig. 2). This correction is a purely technical remedy and allows us to eliminate the PP–P offset for a range of distances. After centering the PP–P times to zero, the offset in the P data increases to about +5 s (Fig. 1b). The origin of this bias is not fully understood and is the subject of current study. In order to

minimize the effect, we have applied a constant correction to all P traveltimes such that the average $< \delta T_{\rm P} >= 0$. The effect of this baseline shift is to bring the average of the origin time corrections close to zero; it does not introduce discrepancies that cannot be handled by the inversion.

2.2 Model parameterization

We sample the velocity structure by using an irregular distribution of points to form a Delaunay mesh (Watson, 1981; Watson, 1992; Sambridge et al., 1995). In 3D, a Delaunay mesh is a uniquely defined aggregation of space-filling, disjoint, irregular tetrahedra. We build the Delaunay connections by using *qhull*, a package distributed by the Geometry Center of Minneapolis (Barber et al., 1996). Node spacing is adapted to the expected resolving length of our data and ranges from about 200 km in the upper mantle to about 600 km in the lower mantle. The total number of nodes we use to model the mantle is $M = 19,279$. The velocity c at any point $\mathbf{x}$ in the model is defined by linear interpolation within each tetrahedron spanned by this mesh, formally expressed as:

$$\delta c(\mathbf{x}) = \sum_k \delta c_k h_k(\mathbf{x}), \tag{1}$$

where h_k denote the interpolation functions, k being an index over the four nodes of the tetrahedron that contains $\mathbf{x}$.

2.3 Delay times tomography: rays and waves

To investigate the effects of wavefront healing, we compare ray-theoretical tomography with finite-frequency wave tomography. In the following two sections we briefly review the analytical description of both formulations, which in the end both reduce to a discrete system $\mathsf{A}\mathsf{x} = \mathsf{b}$ of N traveltime shifts b_i measured by cross-correlation and M velocity perturbations x_j, which we solve iteratively in a least-squares sense (Paige & Saunders, 1982; Nolet, 1985). The inversion technique that we use is described in detail by Nolet (1987) and Spakman & Nolet (1988).

2.3.1 Rays

In the ray approximation, a measured traveltime residual is given by a 1D line integral along the unperturbed spherical-earth ray:

$$\delta T = -\int_{\rm ray} c(r)^{-2} \delta c(\mathbf{x})\, dl, \tag{2}$$

where dl is the differential arclength along the ray, $c(r)$ is the background wavespeed at radius r (in our case, modified *iasp91*), and $\delta c(\mathbf{x})$ is the 3D heterogeneity that one is seeking to image. Fermat's Principle allows us to use the raypath computed for the background velocity $c(r)$ (e.g. Nolet, 1987). By

virtue of the linear interpolation on the tetrahedron structure (eq. 1), the expression for the traveltime shift δT becomes:

$$\delta T_i = \sum_j A_{ij} \frac{\delta c_j}{c_j}, \tag{3}$$

where $c_j = c(r_j)$ at node j, and the elements of the matrix A are given by:

$$A_{ij} = -\int_{i\text{th path}} dl\, c^{-1}(r) h_j(\mathbf{x}), \tag{4}$$

with i the datum number, and j the node index. In the case of a differential traveltime PP–P, the delay time $\delta T^{\mathrm{PP-P}}$ is given by $\delta(T^{\mathrm{PP}} - T^{\mathrm{P}})$ and the elements of the matrix A are simply the difference between the matrix elements of the two phases individually.

2.3.2 Finite-frequency waves

In finite-frequency tomography the 1D integral along the geometrical ray is replaced by a 3D volume integral:

$$\delta T = \int_{\oplus} K(\mathbf{x}) \frac{\delta c}{c} d^3\mathbf{x} \tag{5}$$

over the entire mantle $\oplus$ in which the wave-speed perturbation is non-zero, $\delta c/c \neq 0$. The quantity $K(\mathbf{x})$ is the 3D Fréchet kernel of a finite-frequency traveltime shift δT that has been measured by cross-correlation of a broadband waveform with a spherical-earth synthetic.

Following Dahlen et al. (2000) the 3D Fréchet kernel $K(\mathbf{x})$ is expressed by a double ray sum over all scattered body waves. This formula reduces to an easily computable expression by invoking the paraxial approximation which eliminates the need to conduct repeated two-point ray tracing. By ignoring all forward scattering rays that are not of the same type as the unperturbed path, the Fréchet kernel reduces to a compact expression given by (Dahlen et al., 2000):

$$K(\mathbf{x}) = -\frac{1}{2\pi c}\sqrt{(|\det(\mathbf{M}' + \mathbf{M}'')|}\ \frac{\displaystyle\int_0^\infty \omega^3 \left|\dot{m}(\omega)\right|^2 \sin\Phi\, d\omega}{\displaystyle\int_0^\infty \omega^2 \left|\dot{m}(\omega)\right|^2 d\omega}, \tag{6}$$

where

$$\Phi = \frac{1}{2}\omega \mathbf{q}^{\mathrm{T}} \cdot (\mathbf{M}' + \mathbf{M}'') \cdot \mathbf{q} - [\mathrm{sig}(\mathbf{M}' + \mathbf{M}'') - 2]\frac{\pi}{4}\ . \tag{7}$$

The matrices $\mathbf{M}'$ and $\mathbf{M}''$ are the forward and backward 2×2 traveltime Hessians along the central ray, and $\mathbf{q}$ is the location vector of a scatterer at $\mathbf{x}$, in ray coordinates (see Fig. 3). The symbols det and sig denote the determinant and the signature, or the number of positive minus the number of negative eigenvalues of $\mathbf{M}' + \mathbf{M}''$, respectively; ω is the angular frequency, and $c = c(r)$ is the background spherical-earth velocity. The kernel for a single, well-isolated seismic phase depends only upon the sum $\mathbf{M}' + \mathbf{M}''$ of forward and backward traveltime Hessians along the central geometrical ray. The

quantity Φ (eq. 7) represents the phase delay of the wave scattered from $\mathbf{x}$. The quantity $|\dot{m}(\omega)|^2$ is the power spectrum of the attenuated synthetic (see section 2.1) and specifies the frequency content of the cross-correlated arrivals. This is a reminder that $K(\mathbf{x})$ is the Fréchet kernel of a finite-frequency traveltime measurement δT. We have ignored a possible bias in dominant frequency caused by the correlation operator emphasizing the early part of the waveform rather than a full period.

Written out explicitly, the 3D integral for the traveltime shift δT is given by:

$$\delta T = -\frac{1}{2\pi}\int_0^L dl \iint_{-\infty}^{\infty} dq_1 dq_2 (1 + q_k \partial_k \ln c) c^{-2} \delta c$$

$$\times \sqrt{(|\det(\mathbf{M}' + \mathbf{M}'')|}\ \frac{\int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin\Phi\, d\omega}{\int_0^\infty \omega^2 |\dot{m}(\omega)|^2\, d\omega}. \tag{8}$$

The limit $\pm\infty$ on the transverse integrals over q_1, q_2 are purely formal; in practice, the kernel $K(\mathbf{x})$ is negligible except within the first one or two Fresnel zones about the central ray.

Again, by virtue of the linear interpolation on the tetrahedron structure (eq. 1), the expression for the traveltime shift δT becomes:

$$\delta T_i = \sum_j A_{ij} \frac{\delta c_j}{c_j}, \tag{9}$$

where the elements of the matrix A are given by:

$$A_{ij} = -\frac{1}{2\pi}\int_{i\text{th path}} dl \iint_{-\infty}^{\infty} dq_1 dq_2 (1 + q_k \partial_k \ln c) c^{-1} h_j(\mathbf{x})$$

$$\times \sqrt{(|\det(\mathbf{M}' + \mathbf{M}'')|}\ \frac{\int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \sin\Phi\ d\omega}{\int_0^\infty \omega^2 |\dot{m}(\omega)|^2\, d\omega}, \tag{10}$$

with $i = 1, 2, ..., N$ the datum number, $j = 1, 2, ..., M$ the node index and $k = 1, 2$ the ray coordinate index. In the case of differential traveltimes $\delta T^{\mathrm{PP-P}} = \delta(T^{\mathrm{PP}} - T^{\mathrm{P}})$, which is related to the difference of the individual Fréchet kernels: $K^{\mathrm{PP-P}}(\mathbf{x}) = K^{\mathrm{PP}}(\mathbf{x}) - K^{\mathrm{P}}(\mathbf{x})$.

The difference between the ray-theoretical and finite-frequency approach resides in the elements of the matrix A. In the ray-theoretical formulation, each row of the matrix A represents the geometrical ray connecting the source $\mathbf{s}$ to the receiver $\mathbf{r}$. The elements A_{ij} are interpolation weights integrated along the arclength of the ray i contained in all tetrahedra having node j as a common vertex. In the finite-frequency modeling, each row of A represents one Fréchet kernel connecting the source $\mathbf{s}$ to the receiver $\mathbf{r}$. Therefore each element of the matrix A_{ij} can be seen as the integrated effect of the kernel i contained in all tetrahedra having node j as a common vertex.

Typical widths of the sensitivity region (i.e., the diameter of the ring around the unperturbed geometrical ray) at the turning point of a direct P wave range from about 1000 km to about 1300 km for a 60° and 80° epicentral distance, respectively. Delay times are relatively insensitive to velocity perturbations close to the geometrical ray. This region of insensitivity is smaller near the source and receiver but can extend to about 400 km near the turning point of a P wave at 80° epicentral distance, giving the characteristic cross-sectional doughnut shape to the banana kernel (Fig. 4a). PP waves show a much more complicated shape of the sensitivity region than direct P (Fig. 4b). The PP wave from a source to a receiver passes through a source-to-receiver caustic where it experiences a non-geometrical $\pi/2$ shift; the backward wave from receiver to source passes through the corresponding receiver-to-source caustic. Upon passage through these caustics the shape of the kernel changes drastically and does not resemble a hollow banana any more. The on-ray PP sensitivity is identically zero between the source and the source-to-receiver caustic, and between the receiver and the receiver-to-source caustic; however, it is nonzero between the two caustics (Fig. 4b). The characteristic zero-to-maximal-to-zero sensitivity variation of the PP waves along the geometrical ray is due to the jumps in the term $\mathrm{sig}(\mathbf{M}' + \mathbf{M}'')\pi/2$ of eq.(7) (Dahlen et al., 2000; Hung et al., 2000). Steps of integration in the computation of elements of the matrix A have been dynamically adapted along the kernel to take into account changes in the size and shape of the sensitivity region.

It is clear from Fig. 4 that a 1D line integral along a ray is an extremely crude approximation of the complex sensitivity region of a PP wave. PP waves are particularly useful in global tomography because they provide upper-mantle constraints in regions where there are no sources or receivers. Fig. 5a shows the sources and stations distribution of the P waves contained in our data set. Significant parts of the globe, particularly in the southern hemisphere, are poorly covered by paths. Because of the bounce points, the introduction of PP waves in the inversion significantly enhances the path coverage in the upper mantle.

Because of the wide span of the sensitivity regions, finite-frequency waves sample a larger volume of the model than do the rays. As result, the matrix A for finite-frequency waves is an order of magnitude less sparse than the one constructed with rays for our parameterization. Nevertheless, the sensitivity is significant only in a limited region around the geometrical ray and many of the matrix elements are small. In Fig. 6 we show a comparison between the column density for the two matrices. We define the density for a given node of the grid to be the sum of all the elements of the matrix in the column correspondent to that node. As expected, the density is larger and broader for the finite-frequency matrix.

3 TECHNICAL ASPECTS OF THE INVERSION

We simultaneously invert for perturbations in velocity $\delta c/c$ and in hypocentral parameters (origin time, longitude, latitude, depth). Our system of inversion becomes $[\mathsf{A}\,\mathsf{H}]\mathsf{x} = \mathsf{b}$ where H is the matrix of (ray-theoretical) partial derivatives with respect to the source parameters and where $\mathsf{x} = [\mathsf{x}_c, \mathsf{x}_h]^{\mathrm{T}}$ now contains both unknown velocity perturbations x_c and source corrections x_h. The quantity b on the right side is the vector of the delay times δT. We have $5,738$ sources and the grid consists of $19,279$ points giving a total of $42,231$ unknowns and $N = 86,405$ observations (see section 2.1). Changes in origin time and source location are computed with respect to NEIC values.

We have many more data than unknowns, yet due to the sparseness of the ray distribution, the problem has a partly underdetermined nature. Also, because of errors, the system of equations is inconsistent. To regularize the inversion, we apply norm damping to the velocity perturbation x_c, which biases to a lower amplitude solution, and to the source corrections x_h, which limits changes in origin times and hypocentral location coordinates: the strength of this $\|\mathsf{x}_c\| \rightarrow 0$ and $\|\mathsf{x}_h\| \rightarrow 0$ norm damping is controlled by two tunable parameter ϵ_c and ϵ_h. To supplement the norm damping, we also apply second derivative damping, $\|\mathsf{S}\mathsf{x}_c\| \rightarrow 0$, governed by a parameter ϵ_s; this biases the solution toward smooth velocity variations in every direction (latitudinal, longitudinal and radial). Due to the irregular nature of the grid, our smoothing operator is not truly a canonical second derivative ∇^2, since it averages over the total number of nearest neighbours to a node. In finite-frequency modeling, the effect of the uneven sampling is reduced by the implementation of kernel volumes.

How are we going to compare two models obtained with two different techniques? The least-squares problems in seismic tomography can be rephrased in a statistical sense. Given a set of data, with known standard deviations, we can ask which is the most likely model from which these data could have arisen. A basic assumption is that data errors are independent and Gaussian distributed, so that the joint likelihood of obtaining the observed data vector b (the quantity we want to maximize) is the product of the individual probabilities. Maximizing the probability is equivalent to minimizing χ^2:

$$\sum_i \left(\frac{\sum_j A_{ij} x_j - b_i}{\sigma_i} \right)^2 = \chi^2, \tag{11}$$

where σ_i are the estimated errors in the data.

Each measurement b_i is characterized by a grade A, B, and C based on the confidence of the pick and how well the waveforms match. Each grade corresponds to a different error σ_i. A priori estimates of the σ_i values given in the traveltime measurement procedure (Bolton & Masters, 2001) were slightly raised on the basis of the misfits obtained in preliminary inversions for P and PP-P

separately. We assigned an error σ of 0.44, 0.53, 0.79s for P residuals, and 0.75, 0.95, 1.15 for PP–P residuals to measurements of quality grades A, B, C, respectively.

To the extent that these errors are normally distributed, and for very large tomographic problems involving a large number N of data, the quantity χ^2 is a sum of N squares of normally distributed quantities, each normalized to unit variance (Press et al., 1992; Parker, 1994). Thus a typical value of χ^2 for a good fit is $\chi^2 = N$. More precisely, the χ^2 statistic has a mean N and a standard deviation $\sqrt{2N}$. In model space this motivates us to look for solutions which lie on the boundary of the allowable misfit region, i.e. where $\chi^2 = N$. Ray-theorerical and finite-frequency tomography will have different regions of allowable misfit. We compare models which have exactly the same χ^2, with $\chi^2 \approx N$.

The least-square system we are minimizing can finally be expressed as:

$$\chi^2 + \epsilon_c \|\mathsf{x}_c\|^2 + \epsilon_h \|\mathsf{x}_h\|^2 + \epsilon_S \|\mathsf{S}\mathsf{x}_c\|^2 = \text{minimum}, \tag{12}$$

where ϵ_s and S are the smoothing factor and smoothing operator, respectively. The three damping factors (ϵ_c for the model norm, ϵ_h for the hypocentral parameters, and ϵ_s for the smoothing) define a three-dimensional space. Each point in this space corresponds to a particular solution, and models with the same χ^2 span 2D surfaces. We experimented with changing both the damping ϵ_c and the smoothing ϵ_s, while keeping the damping factor ϵ_h constant at a value that yields hypocentral corrections of realistic magnitude. In our inversion, the norm damping (ϵ_h) applied to the hypocentral part of the solution is such that 95% of the changes in the origin time lie between ± 1.2 sec and 95% of the hypocentral coordinates between ± 14 km. The actual value of the damping factors has no direct physical meaning, and their effect depends greatly on the eigenvalue distribution of the matrix A. Their function is only to let us navigate between the model norm, smoothness, and the data fit. Though it is possible to obtain damped solutions by halting the iterative matrix solver before full convergence has been obtained, such solutions are less than optimal (in the sense that a lower model norm could be obtained for the same χ^2) and we have always made sure that our solutions iterated to full convergence. Our experience confirms that, because of their differential nature, PP-P data are insensitive to source parameters. We also found that finite-frequency inversion requires much smaller values for damping and smoothing parameters to achieve the same χ^2.

4 INVERSION RESULTS

We first investigate the sensitivity of the misfit, as measured by chi-square divided by the number of data (χ^2/N), to the model norm $|\mathsf{x}_c|^2$ for both inversion formulations, by obtaining a series of models corresponding to different values of ϵ_c and ϵ_s. Fig. 7 shows these solutions in a plane of $|\mathsf{x}_c|^2$ versus χ^2/N. Solid lines correspond to the ray-theoretical solutions, dashed lines denote the finite-

frequency version. It is clear that a very similar trade-off exists for both theories. This graph does not tell the whole story, though, since the smoothness of two models is different even if they occupy the same location in this graph. Therefore, we plot the smoothness versus the model norm in Fig. 8 for a series of solutions obtained at fixed value of $\chi^2/N = 1.18$ (the vertical line in Fig. 7). Here the large difference between ray-theoretical and finite-frequency solutions becomes evident: for the same smoothness (a proxy for spectral content), the finite-frequency models exhibit a significantly larger amplitude anomalies.

In the following, we shall further investigate two models out of the family shown in Fig. 8, denoted by FF (finite frequency) and RT (ray theory), respectively. Fig. 9 shows these two P-wave velocity perturbations with respect to our background model for ray theory (left) and finite-frequency (right) at different depths. The two models are constrained to have the same data misfit, $\chi^2/N = 1.18$, and the same spectral content, $\|\mathsf{S}\mathsf{x}_c\|/\|\mathsf{x}_c\| = 0.16$. One can, by eye, confirm the aggregate result of Fig. 8 for individual anomalies. This is further quantified in Fig. 10, where we show the ratio $(\delta c/c)_{FF}/(\delta c/c)_{RT}$ for those anomalies that are significantly different from zero ($|\delta c/c| > 0.2\%$) Clearly, the average ratio is larger than 1 with a tail in the distribution extending well beyond 2, especially as the depth increases. In the deepest layer (2408–2889 km depth) the average ratio is as large as 1.5, with the larger values belonging to smaller anomalies, as expected. The ratio decreases toward the surface, where it is of the order of 1.3. This decrease is also expected, since the kernels become narrower for shorter raypaths that are mostly located at shallower depth.

The effect of finite-frequency theory is less when we do not restrict our attention to the more significant anomalies. Fig. 11 allows one to inspect how the difference in root-mean-square value of the anomalies (rms) depends on depth. An average discrepancy of about 20% or less (Fig. 11a) is present between the average absolute value of the velocity changes in the two models. A slight depth dependence is visible because the absolute difference between FF and RT remains the same while the rms decreases to a minimum near 2000 km depth. The correlation coefficient (Fig. 11b) between the models is above 0.9 at all depths. We conclude from this analysis that the effects of finite-frequency theory are especially important for significant anomalies of small size.

To confirm that the incorporation of finite-frequency sensitivity kernels makes more of a difference for the smaller significant anomalies, we increased the power of the small anomalies by reducing ϵ_s and compared models for which χ^2 was lower — in this case equal to N exactly. Fig. 12 shows the comparison of the ray-theoretical and finite-frequency tomography as a function of depth for this $\chi^2/N = 1$ case. Again, both solutions have the same roughness or spectral content, in this case $\|\mathsf{S}\mathsf{x}_c\|/\|\mathsf{x}_c\| = 0.64$. Numerous small-scale heterogeneities appear in the model. The inspection of the two solutions shows even more clearly the significantly larger amplitudes of the finite-

frequency model. Because of the presence of small-scale anomalies at all depths, finite-frequency velocity changes are now on average about 1.7 times larger than the corresponding ray-theoretical ones at the base of the mantle (2408–2889 km depth). As in the case of the $\chi^2/N = 1.18$ model, the ratio decreases only mildly toward the surface becoming as high as 1.6 (Fig. 13). A large difference is also visible in the rms (Fig. 14). In this case, the finite-frequency analysis enhances the rms velocity anomalies by about 25% or more near the base of the mantle (Fig. 14a). The two $\chi^2/N = 1$ models are less well correlated, especially near the base of the mantle where the correlation coefficient reduces to 0.6.

The reader might wonder why we did not choose the model with $\chi^2/N = 1$ in the first place. A look at Fig. 7 shows that there is a significant increase in model norm between the two values of the model misfit. This implies that many of the extra details (and the reduced correlation between FF and RT) hinges on our choice of a priori errors in the data, which admittedly are not cast in stone. William of Occam's dictum that the simplest hypothesis is the preferred one, leads us to prefer the much simpler images in Fig. 9, which would be equivalent to $\chi^2 = 1$ if we underestimated the standard errors in the data by only 9%. From Fig. 7 it is however clear that a further reduction in $\|x_c\|$ can only be obtained at the expense of a much larger χ^2/N.

We performed one final test to investigate the influence of the model parameterization on the inversion. To verify that our previous observations are independent of the chosen grid, we performed an inversion with a much finer parameterization of 39,048 points (i.e., approximately twice as many). The distribution of the nodes is, as before, roughly proportional to the expected resolving length of our data. The distance among the more closely spaced nodes ranges from about 100 km at the surface to about 600 km near the core-mantle boundary. None of the conclusions reached earlier were affected by this test. We do see small differences in the velocity anomalies, mainly at shallow depths. The original parameterization with $19,279$ nodes was evidently fine enough to capture the smaller structure that is affected by finite-frequency effects.

We personally do not believe that "variance reduction" is a particularly useful parameter to evaluate a tomographic inversion. It is as much a measure of the adequacy of the starting model as it is a measure of goodness of fit of the final solution. However, since other global tomographic models are often specified in terms of their variance reduction, we give our values here. Both inversions with the $19,279$-point grid have a variance reduction of about 84% for $\chi^2/N = 1.18$, and of about 87% for $\chi^2/N = 1$. In both cases, roughly half of the variance reduction is due to the velocity anomalies and half is due to the hypocenter corrections.

5 DISCUSSION

Even though this tomographic study is primarily meant to study the difference between ray-theoretical and finite-frequency inversions, the results obtained are of enough interest that we also offer a few speculations. A more extensive discussion of the velocity structure, including a rigorous resolution analysis, is given in an already published tomographic study, in which we combined the long-period data with short-period data from the ISC (Montelli et al., 2004).

There is a remarkable agreement between the low-velocity anomalies in the finite-frequency model (FF) and the locations of well known hotspots visible both in the maps (Fig. 9) and in the cross sections (Fig. 15). A list of major hotspots with their location is given in Table 1. We distinctly see Easter Island, Tahiti, Hawaii, Bouvet, Kerguelen, Azores, Canary Island, Cape Verde, Tibesti, Kilimanjaro, Galapagos and Ascension.

Fast velocity anomalies with amplitudes above 1% are observed beneath Tonga-Kermadec, Tasmania, Java and below Asia, the latter presumably identifiable as the Tethys slab (Van der Hilst et al., 1997; Grand et al., 1997; Grand, 1994; Bijwaard et al., 1998; Van der Voo et al., 1999; Gu et al., 2001). Also clear is the familiar signature of the Farallon plate migrating eastward with depth (Grand, 1994; Grand et al., 1997; Van der Hilst et al., 1997; Mégnin & Romanowicz, 2000). At 1800 km depth the high-velocity anomalies below North and South America begin to disappear while becoming more pronounced below Central America in the lowermost mantle.

In the lowermost mantle the pattern of heterogeneity is dominated by two large-scale slow velocity anomalies, one in the eastern Atlantic Ocean and one under the South Pacific (Dziewonski et al., 1991; Dziewonski et al., 1993; Grand, 1994; Su et al., 1994; Liu & Dziewonski, 1998; Masters et al., 1996; Ritsema et al., 1999; Mégnin & Romanowicz, 2000; Romanowicz & Gung, 2002), as well as by fast velocity under the circum-Pacific common to many other tomographic models (Grand et al., 1997). Both slow regions are loci of major hotspots. The South Pacific "superswell", which is considered to be an exhausted remnant of the Mid-Cretaceous upwelling beneath the Pacific Basin (Larson, 1991; McNutt, 1998), underlies Easter Island, Tahiti and Samoa (Fig. 15b). The Atlantic "superswell" contains Kerguelen, the African superplume, the African hotspots, Cape Verde, Canary Island, and extends all the way to the North Sea; with a clear signature of the shallow Jan Mayen seamount connected with a deeper anomaly below Greenland (Figs. 9 and 15a,c,d).

Crough & Jurdy (1980) removed subduction-related geoid highs from the observed geoid and found a residual field which has a simple form of two large, elliptical highs surrounded by lows. Broad residual geoid highs are in the central Pacific and the Africa/eastern Atlantic region, in perfect correlation with the regions of highest hotspot concentration. Because hotspots are regions of mantle upwelling, they can contribute significantly to geoid anomalies. The two low-velocity anomalies in

the finite-frequency velocity map (FF)are very well correlated with the geoid highs of Crough & Jurdy (1980).

Many features are visible both in the ray-theoretical (RT) and finite-frequency (FF) inversions. However, the continuity of anomalies is generally greater for the finite-frequency images. Also, the finite-frequency tomographic images provide compelling evidence that many hotspots are fed from the lower mantle (Fig. 15). The Pacific superplume seems to feed the spreading of the South Pacific, whereas the Atlantic megaplume feeds the spreading not only of the North and South Atlantic but also of the Indian Ocean, through a clearly visible conduit "leaning" toward Kerguelen (Fig. 15d). The interaction of the African superplume with both the Mid-Atlantic Ridge and the Mid-Indian Ridge is present in the shear-velocity models obtained previously by Ritsema et al. (1999), Mégnin & Romanowicz (2000) and Romanowicz & Gung (2002).

Major hotspots which do not seem connected to a lower mantle plume include Afar, Ascension, Etna, Galapagos, Iceland, Kilimanjaro, Madeira, Reunion, Tristan. These all seem to originate in the mid mantle. Indications that Iceland is not a deep-seated anomaly were already presented by Ritsema et al. (1999), and a shallow origin was argued from indirect evidence by Foulger & Pearson (2001), Foulger et al. (2001) and Foulger (2003). The result of our inversion confirms these observations and clearly contradict the finding of Bijwaard & Spakman (1999), who proposed an Iceland plume extending all the way to the core-mantle boundary.

6 RESOLUTION

To determine the reliability of our tomographic images and to compare the resolving power of finite-frequency and ray-theoretical inversions for relatively small-scale structures, we have performed a limited number of resolution tests using synthetic plumes. We introduce velocity anomalies having the form of vertical cylinders at location of known hotspots: Iceland, Ascension, Kerguelen, Hawaii, Tahiti and Easter Island. The input velocity perturbation is a circularly symmetric Gaussian centered on the cylindrical axis of the hotspot. The highest velocity perturbation in the center of the cylinder is defined by following the pattern for temperature derivatives of P-wave velocities in the mantle as a function of depth given by Karato (1993). Assuming a temperature $T = 300^\circ$ K at the center of the plume, the maximum input velocity perturbation is -2.4% above 600 km depth, -1.2% between 600 and 1000 km depth, and -1% below 1000 km depth. The corresponding velocity perturbation is assigned to each point of the grid lying within the cylinder. Because of the tetrahedral nature of our parameterization, the input velocity model deviates slightly from a smooth cylinder, depending on the distribution of model nodes within and near the synthetic plume.

The radius of the input cylinders is roughly the same as the quasi-cylindrical anomalies in the

smoother tomographic model (Fig. 9), i.e., the standard width of the Gaussian is about 500 km. To verify the vertical resolution we performed two kinds of tests. In the first test, the synthetic hotspots reach the core-mantle boundary (Fig. 16a); in a second test, they are confined to depths above 660 km, to simulate plumes originating in the upper-mantle (Fig. 16b). These tests allow us to examine whether the absence of a plume tail at great depth is due to a lack of resolution, and whether a deep feature is due to the leakage of a shallow anomaly to larger depth in the mantle. Synthetic delay times are computed using finite-frequency theory, and inverted using both ray theory and finite-frequency kernels (Fig. 16 "exact data"). The same tests are then repeated adding normally distributed random noises to the synthetic residuals (Fig. 16 "noise added").

Overall we have adequate resolution at almost all synthetic hotspot locations to recover the shape and depth extent of the anomalies, no matter if they are shallow or deep. A comparison by eye of the results shows that finite-frequency analysis tends to more faithfully recover the amplitude of the anomalies. This is further quantified in Fig. 17, where we show the finite-frequency and ray-theoretical maximum values of the reconstructed velocity anomalies $\delta c/c$ for each synthetic hotspot as a function of depth for plumes of 500 km radius (Fig. 17a) and 300 km (Fig. 17b), respectively. In general, the amplitudes recovered using finite-frequency theory are greater than those recovered using ray theory, as we observed in the tomographic models. In absolute terms, the difference is larger in the upper mantle. In relative terms, the effect of finite-frequency is still larger in the lower mantle, provided the damping does not prevail, as it does for many synthetic plumes in the southern hemisphere. For Iceland, the best resolved synthetic plume, the finite-frequency perturbations are a factor of 1.2 times larger than the ray-theoretical ones in the upper mantle, and 1.3 in the lower mantle. The lack of resolution in the lower mantle is more evident with smaller scale plumes, such as those of 300 km radius used in Fig. 17b. Deep anomalies beneath Kerguelen, Tahiti and Ascension seem to be particularly poorly resolved; the same is true for a shallow anomaly beneath Kerguelen (Fig. 16b). We attribute this poor resolution to a lack of ray path coverage in these particular regions, visible also from Figs. 5 and 6. Montelli et al. (2004) have further improved the resolution with the introduction of the high-frequency data (ISC delays).

7 CONCLUSION

We have introduced 3D finite-frequency Fréchet sensitivity kernels into global P-wave tomography, and compared the results with the corresponding tomographic images obtained using conventional seismic ray theory. We inverted P and PP-P cross-correlation traveltimes of 20 s dominant period. The results demonstrate that finite-frequency analysis of such long-period waves significantly affects the final images. The amplitudes of the velocity perturbations in our finite-frequency model are 30%–50%

higher than those obtained with ray theory, depending upon the depth and size of the heterogeneity. This demonstrates that the neglect of wavefront healing effects is a major shortcoming of ray theory. Finally, the results in this paper show that it is possible, even with Fresnel zones seismic waves with wavelengths of the order of one thousand km, to image structures as narrow as mantle plumes. While some plumes are still visible in the ray-theoretical inversion, properly accounting for finite-frequency wavefront healing effects leads to an increase in anomaly amplitudes, which raises the signal-to-noise ratio to significant levels. It is likely that other improvements also contributed to our success in imaging plumes. The correction for timing errors may have repaired the data somewhat, but we do not suspect these errors to have been disastrous. Also, the use of the unstructured grid, adapted to the decreased resolution at depth, plays an additional important role in delineating anomalies at the limit of resolution.

ACKNOWLEDGMENTS

We are extremely grateful to Jason Morgan for providing his newest list of hotspot locations and for sharing with us his valuable insights. This research was made possible through financial support from the National Science Foundation under grants EAR-9814570 and EAR-0105387. We thank Malcolm Sambridge for providing us the software for the *walking algorithm* in Delaunay meshes, and Brian Schlottmann, Frederik Simons and Sergio Speziale for very helpful discussions. The original manuscript was significantly improved by the constructive reviews of Jeroen Ritsema and an anonymous reviewer. Figures have been plotted using GMT (Wessel & Smith, 1995).

REFERENCES

Barber, C. B., Dobkin, D. P., & Huhdanpaa, H. T., 1996. The Quickhull algorithm for convex hulls, *ACM Trans. on Mathematical Software*, **22**, 469–483.

Bijwaard, H. & Spakman, W., 1999. Tomographic evidence for a narrow whole mantle plume below Iceland, *Earth Planet. Sci. Lett.*, **166**, 121–126.

Bijwaard, H., Spakman, W., & Engdahl, E. R., 1998. Closing the gap between regional and global travel time tomography, *J. Geophys. Res.*, **103**, 30055–30078.

Bolton, H. & Masters, G., 2001. Travel times of P and S from global digital seismic networks: Implication for the relative variation of P and S velocity in the mantle, *J. Geophys. Res.*, **106**, 13527–13540.

Cardimora, S. & Garmany, J., 1993. Smoothing operators for waveform tomographic imaging, *Geophysics*, **58**, 1646–1654.

Castle, J. C., Creager, K. C., Winchester, J. P., & Van der Hilst, R. D., 2000. Shear wave speeds at the base of the mantle, *J. Geophys. Res.*, **105**, 21543–21557.

Crough, S. T. & Jurdy, D. M., 1980. Subducted lithosphere, hot-spots and the geoid, *Earth Planet. Sci. Lett.*, **48**, 15–22.

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

Dziewonski, A. M., Su, W.-J., & Woodward, R. L., 1991. Grand structures of the Earth's Interior, *Eos Trans. AGU*, **72**, 451.

Dziewonski, A. M., Forte, A. M., Su, W.-J., & Woodward, R., 1993. Seismic tomography and geodynamics, in *Relating Geophysical Structures and Processes: The Jeffreys Volume*, edited by K. Aki & R. Dmowska, vol. 76, pp. 67–105, AGU Geophys. Monograph.

Foulger, G. R., 2003. Plumes, or plate tectonic processes?, *Astronomy and Geophysics*, **43**, 6.19–6.23.

Foulger, G. R. & Pearson, D. G., 2001. Is Iceland underlain by a plume in the lower mantle? Seismology and helium isotopes, *Geophys. J. Int.*, **145**, F1–F5.

Foulger, G. R., Pritchard, M. J., Julian, B. R., Evans, J. R., Allen, R. M., Nolet, G., Morgan, W. J., Bergsson, B. H., Erlendsson, P., Jakobsdottir, S., Ragnarsson, S., Stefansson, R., & Vogfjörd, K., 2001. Seismic tomography shows that upwelling beneath Iceland is confined to the upper mantle, *Geophys. J. Int.*, **146**, 504–530.

Gelchinsky, B., 1985. The formulae for the calculation of the Fresnel zones or volumes, *J. Geophys.*, **57**, 33–41.

Grand, S. P., 1994. Mantle shear structure beneath the americas and surrounding oceans, *J. Geophys. Res.*, **99**, 11591–11621.

Grand, S. P., Van der Hilst, R. D., & Widiyantoro, S., 1997. Global Seismic Tomography: a snapshot of convection in the Earth, *GSA Today*, **7**, 1–3.

Gu, Y. J., Dziewonski, A. M., Su, W., & Ekström, G., 2001. Models of the mantle shear velocity and discontinuities in the pattern of lateral heterogeneities, *J. Geophys. Res.*, **106**, 11169–11199.

Hagendoorn, O., 1954. A process of seismic reflection interpretation, *Geophys. Prospecting*, **2**, 85–127.

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

Husen, S. & Kissling, E., 2001. Local earthquake tomography between rays and waves: fat ray tomography, *Phys. Earth Planet. Inter.*, **123**, 129–149.

Karato, S.-I., 1993. Importance of anelasticity in the interpretation of seismic tomography, *Geophys. Res. Lett.*, **20**, 1623–1626.

Kennett, B. L. N. & Engdahl, E. R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.*, **105**, 429–465.

Kravtsov, Y. A., 1988. Rays and caustics as physical objects, in *Progress in optics XXVI*, edited by E. Wolf, pp. 229–348, Elsevier Science Publisher.

Kuo, B.-Y., Forsyth, D. W., & Wysession, M., 1987. Lateral heterogeneity and azimuthal anisotropy in the North Atlantic determined from SS-S differential travel times, *J. Geophys. Res.*, **92**, 6421–6436.

Larson, R. L., 1991. Latest pulse of Earth: evidence for a mid-Cretaceous superplume, *Geology*, **19**, 547–550.

Li, X.-D. & Tanimoto, T., 1993. Waveforms of long-period body waves in a slightly aspherical Earth model, *Geophys. J. Int.*, **121**, 695–709.

Liu, X.-F. & Dziewonski, A. M., 1998. Global analysis of shear wave velocity anomalies in the lower-most mantle, in *The Core-Mantle Boundary Region, Geodynamics Serie*, edited by M. Gurnis, M. Wysession, E. Knittle, & B. Buffett, vol. 28, pp. 21–36, American Geophysical Union.

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

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

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

McNutt, M. K., 1998. Superswells, *Rev. Geophysics*, **36**, 211–244.

Mégnin, C. & Romanowicz, B., 2000. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophys. J. Int.*, **143**, 709–728.

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

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

Nolet, G., 1985. Solving or resolving inadequate and noisy tomographic systems, *J. Comp.Phys.*, **61**, 463–482.

Nolet, G., 1987. Seismic wave propagation and seismic tomography, in *Seismic Tomography*, edited by G. Nolet, pp. 1–23, D. Reidel Publishing Company.

Nolet, G., 1990. Partitioned wave-form inversion and 2D structure under the NARS array, *J. Geophys. Res.*, **95**, 8513–8526.

Nolet, G. & Dahlen, F. A., 2000. Wavefront healing and the evolution of seismic delay times, *J. Geophys. Res.*,

105, 19043–19054.

Paige, C. C. & Saunders, M. A., 1982. LSQR: An algorithm for sparse linear equations and sparse least squares, *ACM Trans.Math.Software*, **8**, 43–71.

Parker, R. L., 1994. *Geophysical Inverse Theory*, Princeton University Press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P., 1992. *Numerical Recipes in Fortran 77, Second Edition, The Art of Scientific Computing*, vol. 1, Cambridge University Press.

Ritsema, J., Van Heijst, H. J., & Woodhouse, J. H., 1999. Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science*, **286**, 1925–1928.

Ritzwoller, M. H., Shapiro, N. M., Barmin, M. P., & Levshin, A. L., 2002. Global surface wave diffraction tomography, *J. Geophys. Res.*, **107**, 10.1029/2002JB001777.

Romanowicz, B. & Gung, Y. C., 2002. Superplumes from the core-mantle boundary to the lithosphere: implications for heat flux, *Science*, **296**, 513–516.

Sambridge, M., Braun, J., & McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours, *Geophys. J. Int.*, **122**, 837–857.

Spakman, W. & Nolet, G., 1988. Imaging algorithms, accuracy and resolution in delay time tomography, in *Mathematical Geophysics*, edited by N. J. Vlaar, G. Nolet, M. J. R. Wortel, & S. A. P. L. Cloetingh, pp. 155–188, Reidel.

Stark, P. B. & Nikolayev, D. I., 1993. Toward tubular tomography, *J. Geophys. Res.*, **98**, 8095–8106.

Stewart, R. C., 1984. Q and the rise and fall of a seismic pulse, *Geophys. J. Roy. Astron. Soc.*, **76**, 793–805.

Su, W.-J., Woodward, R. L., & Dziewonski, A. M., 1994. Degree 12 model of shear velocity heterogeneity in the mantle, *J. Geophys. Res.*, **99**, 6945–6980.

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

Van der Voo, R., Spakman, W., & Bijwaard, H., 1999. Mesozoic subducted slabs under Siberia, *Nature*, **397**, 246–249.

VanDecar, J. C. & Crosson, R. S., 1990. Determination of teleseismic relative phase arrival times using multi-channel cross correlation and least squares, *Bull. Seismol. Soc. Am.*, **80**, 150–169.

Vasco, D. W. & Majer, E. L., 1993. Wavepath travel time tomography, *Geophys. J. Int.*, **115**, 1055–1069.

Watson, D. F., 1981. Computing the n-dimensional Delaunay tesselation with application to Voronoi polytopes, *Comput. J.*, **24**, 167–172.

Watson, D. F., 1992. *Contouring: A guide to the analysis and display of spatial data*, Pergamon, Oxford.

Wessel, P. & Smith, W. H. F., 1995. New version of the Generic Mapping Tools released, *Eos Trans. AGU*, **76**, 329.

Wielandt, E., 1987. On the validity of the ray approximation for interpreting delay times, in *Seismic Tomography*, edited by G. Nolet, pp. 85–98, D. Reidel Publishing Company.

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

Woodward, R. L. & Masters, G., 1991. Global upper mantle structure from long-period differential travel times,

J. Geophys. Res., **96**, 6351–6377.

Yomogida, K., 1992. Fresnel zone inversion for lateral heterogeneities in the Earth, *Pageoph*, **138**, 391–406.

Yoshizawa, K. & Kennett, B. L. N., 2002. Three-dimensional fréchet differential kernels for seismic delay times, *Geophys. J. Int.*, **149**, 440–453.

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

Zhao, L., Chen, L., & Jordan, T. H., 2001. Full Three-dimensional tomography experiments in the Western Pacific Region, *Eos Trans. AGU*, **82**(S52F-0698), 47.

Figure 1. Top: Histograms of PP–P residuals of data used in this study. Bottom: P residuals computed with (left) the *iasp91* velocity model and (right) a modified version of the *iasp91* velocity model shown in Fig. 2. $<\delta T>$ indicates the average delay time. The offset of $\sim +5$ sec is discussed in the text.

Figure 2. Comparison of the *iasp91* velocity model (solid line) with a model obtained by applying a slight perturbation (0.7%) to the velocity of the *iasp91* velocity model (dotted line) between 400 and 660 km. The modified *iasp91* model removes the bias in the PP-P data.

Figure 3. Perpendicular projection of a scatterer $\mathbf{x}$ onto the paraxial point ξ, situated on the central geometrical ray from the source $\mathbf{s}$ to the receiver $\mathbf{r}$. The off-path difference vector is expressed in terms of two orthogonal unit vectors: $q = q_1\hat{\mathbf{q}}_\mathbf{1} + q_2\hat{\mathbf{q}}_\mathbf{2}$. The ray centered coordinates of the scatterer are $\mathbf{x} = (q_1, q_2, l)$ where l is the arclength along the central ray (Dahlen et al., 2000).

Figure 4. Ray-perpendicular cross sections of the Fréchet kernel (a) for a P wave between the source and the turning point and (b) for a PP wave between the source and the bounce point. The shape of the PP kernel changes drastically upon passage of a caustic.

Figure 5. (a) Distribution of sources (stars) and receivers (triangle) for P data. (b) Distribution of sources (stars), stations (triangles) and bounce points (dots) for PP data.

Figure 6. Sections at different depth of the density of the matrix A for ray theory (left) and finite-frequency waves (right), expressed as the sum of the absolute values of the elements of each column of the matrix A. Note that the maps have been "wrapped around" to aid in the visualization of patterns in the vicinity of the Greenwich meridian.

Figure 7. Model norm versus χ^2/N for a combination of values of norm damping ϵ_c and smoothing ϵ_S. Solutions for ray-theory (solid line) and finite-frequency (dotted line) tomographies are compared. Symbols – listed in the legend – correspond to different norm damping parameters. Smoothing increases from upper right to lower left along the curves. The two white dots indicate the FF and RT solutions, respectively, which are discussed in the text and in Figs. 9– 11.

Figure 8. Model norm versus roughness of the solution for finite-frequency (stars) and ray-theoretical (dots) tomography with $\chi^2/N = 1.18$. The two white dots indicate locations of finite-frequency and ray theory models compared in Figs. 9– 11.

Figure 9. Comparison between velocity maps of the smooth, $\chi^2/N = 1.18$ model for ray theory (left) and finite-frequency theory (right) at different depths. The quantity c is the velocity in the reference model shown in Fig. 2. Maps have been "wrapped around" to aid in visualization of patterns both in the Atlantic and the Pacific Oceans.

Figure 10. Histograms showing the ratio between finite-frequency and ray-theoretical velocity changes at different depths, for the model with $\chi^2/N = 1.18$. Only changes with absolute value larger than 0.2% are binned. Depth in the histograms is representative of the depth at the bottom of the shell considered ($z_{\rm bott}$). There are six shells from the surface of the Earth down to the core-mantle boundary. The number in the corner represents the bottom depth of the shell in kilometers. For each layer the total number of points of the grid present in that layer ($\#{\rm pts}$) and their average spacing ($d_{\rm avg}$) are indicated.

Figure 11. (Left) Root-mean-square velocity perturbation $\delta c/c$ versus depth for the finite-frequency (FF – dashed line) and ray-theoretical tomography (RT – solid line). (Right) Correlation coefficient between the finite-frequency and ray-theoretical models versus depth. The data fit criterion in both inversions is $\chi^2/N = 1.18$.

Figure 12. Comparison between velocity maps of the rough, $\chi^2/N = 1$ model for ray theory (left) and finite-frequency (right) at different depths. Maps have been "wrapped around" to aid visualization of patterns both in the Atlantic and the Pacific Oceans.

Figure 13. Histograms showing the ratio between finite-frequency and ray-theoretical velocity changes at different depths, for the models with $\chi^2/N = 1$. Only changes with absolute value larger than 0.2% are binned. There are six shells from the surface of the Earth down to the core-mantle boundary. The number in the corner represents the bottom depth of the shell in kilometers.

Figure 14. (Left) Root-mean-square velocity perturbation $\delta c/c$ versus depth for the finite-frequency (FF–dashed line) and ray theoretical tomography (RT – solid line). (Right) Correlation coefficient between the finite-frequency and ray-theoretical models versus depth. The data fit criterion in both inversions is $\chi^2/N = 1$.

Figure 15. Cross sections of the finite-frequency model with $\chi^2/N = 1.18$. The top figure shows the four great circle paths. Letters (a)–(d) on the paths match plots below. (a) cross section across Greenland and Iceland (pole location $94.82°$W, $11.57°$N), (b) cross section through the Pacific superwell (pole location $96.50°$E, $62.44°$N), (c) cross section across La Reunion and the African hotspots (pole location $125.39°$W, $45.09°$N), (d) cross section across the Atlantic superwell and Hawaii (pole location $90.94°$W, $41.16°$N). Two-letter hotspot identifiers are listed in Table 1.

Figure 16. Resolution tests for six of the major hotspots observed in the tomographic images. (Left) Recovered velocity model (Actual) for $\chi^2/N = 1.18$; (Right) Resolution tests: from left to right we present the input model (Input), the recovered model obtained by inverting the synthetic delay times δT using the ray-theoretical inverse (RT out) and the finite-frequency inverse (FF out), respectively. The rightmost two columns show the corresponding recovered models in the case we invert the synthetic residuals after the addition of normally distributed random noise. Panel (a) shows the results with the synthetic hotspots reaching the core-mantle boundary; panel (b) shows the results with the hotspots originating in the upper mantle (around 660 km depth). Two-letter hotspot identifiers are listed in Table 1.

Figure 17. Difference between the finite-frequency and ray theory maximum velocity perturbation $\delta c/c$ as a function of depth for the 5 plumes showed in Fig. 16. Synthetic times are obtained by using finite-frequency modeling. No random errors were added in this case. Two-letter hotspot identifiers are listed in Table 1. Panel (a) shows the results of the test done with synthetic plumes with 500 km radius, (b) shows the results with plumes of 300 km radius

Table 1. List of major hotspots clearly seen in our tomographic images (locations are from W. Jason Morgan, personal communication 2003)

Hotspot	Latitude	Longitude	Label
Amsterdam	38.7°S	77.5°E	AM
Ascension	7.9°S	14.3°W	AS
Azores	37.9°N	26°W	AZ
Bouvet	54.4°S	3.4°E	BV
Canary	28.2°N	18°W	CA
Cape Verde	14.9°N	24.3°W	CV
Easter	26.8°S	107.6°W	ES
Kerguelen	49.6°S	69°E	KG
Hawaii	19.1°N	155.1°W	HW
Guadalupe	26.8°N	112.4°W	BC
Iceland	64.4°N	17.3°W	IC
Kilimanjaro	3°S	37.5°E	KL
Madeira	32.7°N	17°W	MA
Reunion	21.2°S	55.7°E	RE
Tahiti	18.1°S	148.3°W	TH
Tasmania	40.8°S	146°E	TA
Tibesti	20.8°N	17.5°E	TI
Yellowstone	44.5°N	110.4°W	YW

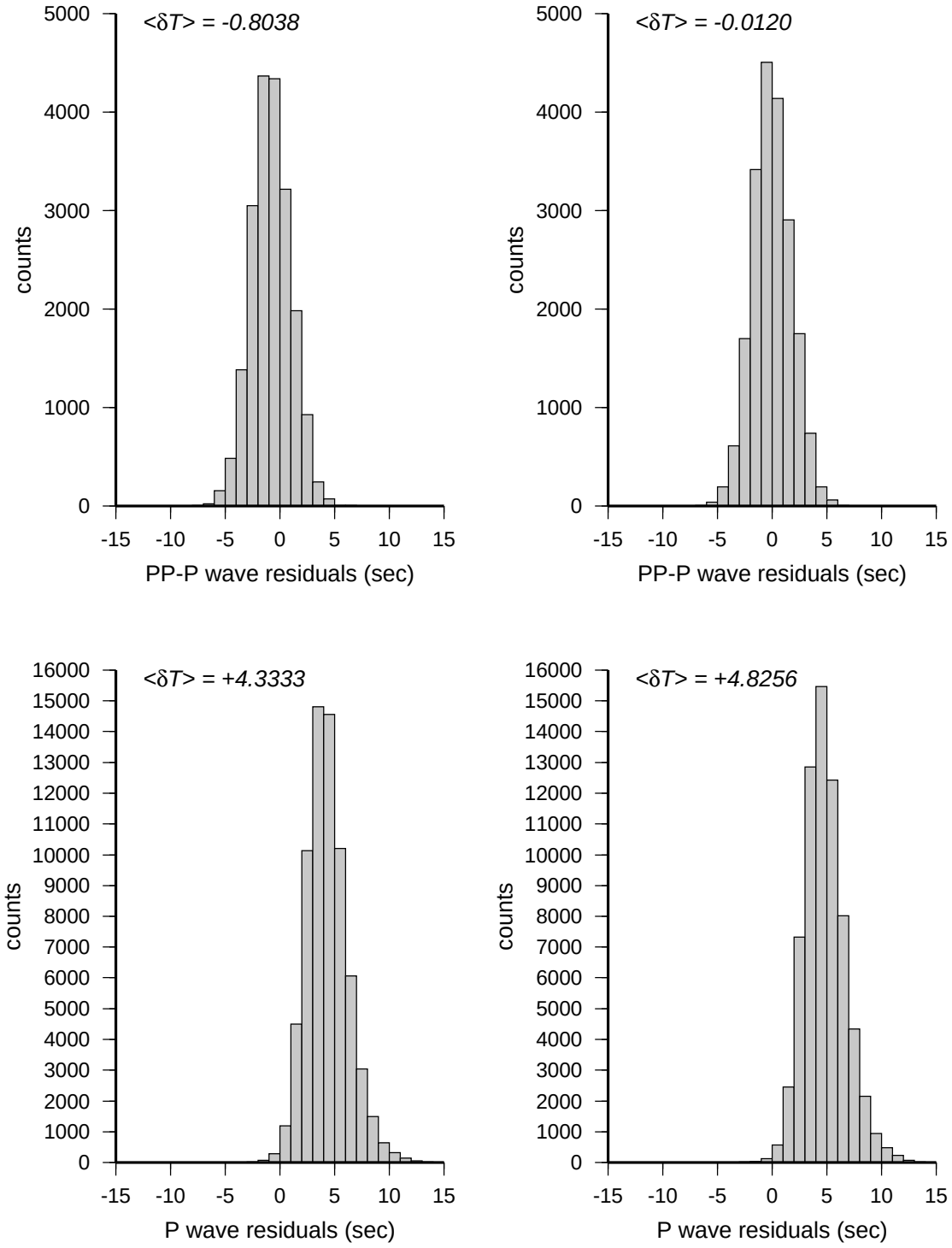


Figure 1. Top: Histograms of PP–P residuals of data used in this study. Bottom: P residuals computed with (left) the *iasp91* velocity model and (right) a modified version of the *iasp91* velocity model shown in Fig. 2. $<\delta T>$ indicates the average delay time. The offset of $\sim +5$ sec is discussed in the text.

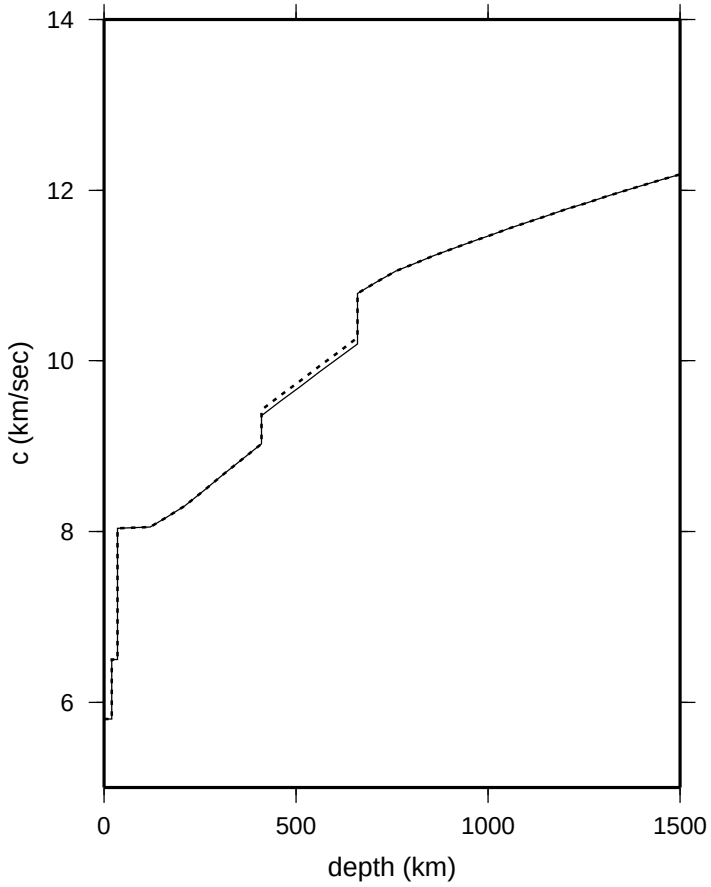


Figure 2. Comparison of the *iasp91* velocity model (solid line) with a model obtained by applying a slight perturbation (0.7%) to the velocity of the *iasp91* velocity model (dotted line) between 400 and 660 km. The modified *iasp91* model removes the bias in the PP-P data.

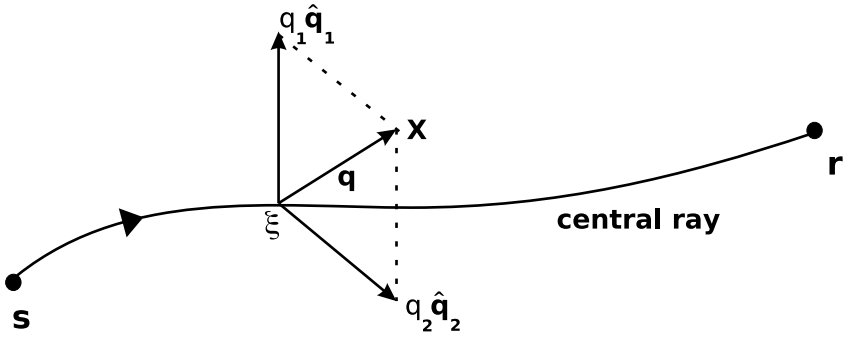


Figure 3. Perpendicular projection of a scatterer $\mathbf{x}$ onto the paraxial point ξ, situated on the central geometrical ray from the source $\mathbf{s}$ to the receiver $\mathbf{r}$. The off-path difference vector is expressed in terms of two orthogonal unit vectors: $q = q_1\hat{\mathbf{q}}_\mathbf{1} + q_2\hat{\mathbf{q}}_\mathbf{2}$. The ray centered coordinates of the scatterer are $\mathbf{x} = (q_1, q_2, l)$ where l is the arclength along the central ray (Dahlen et al., 2000).

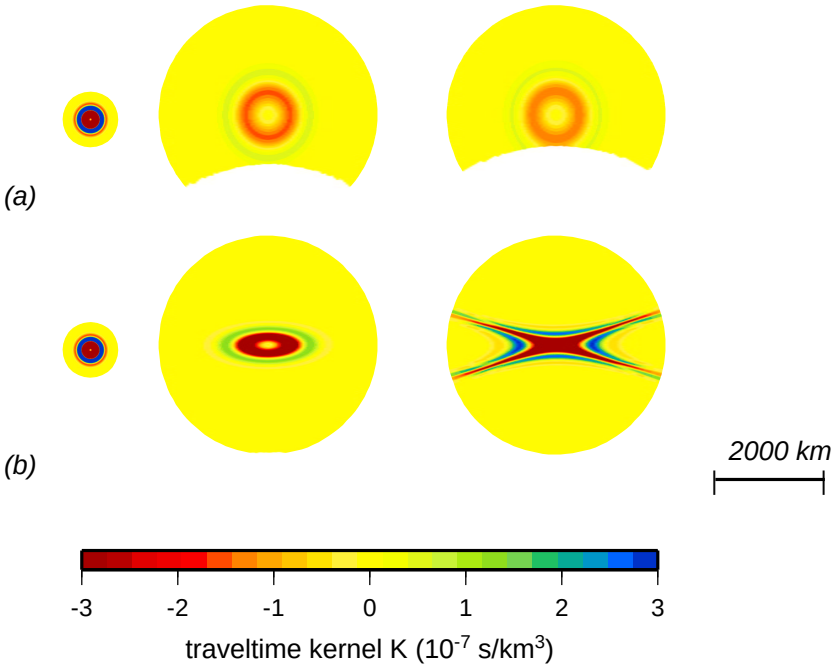
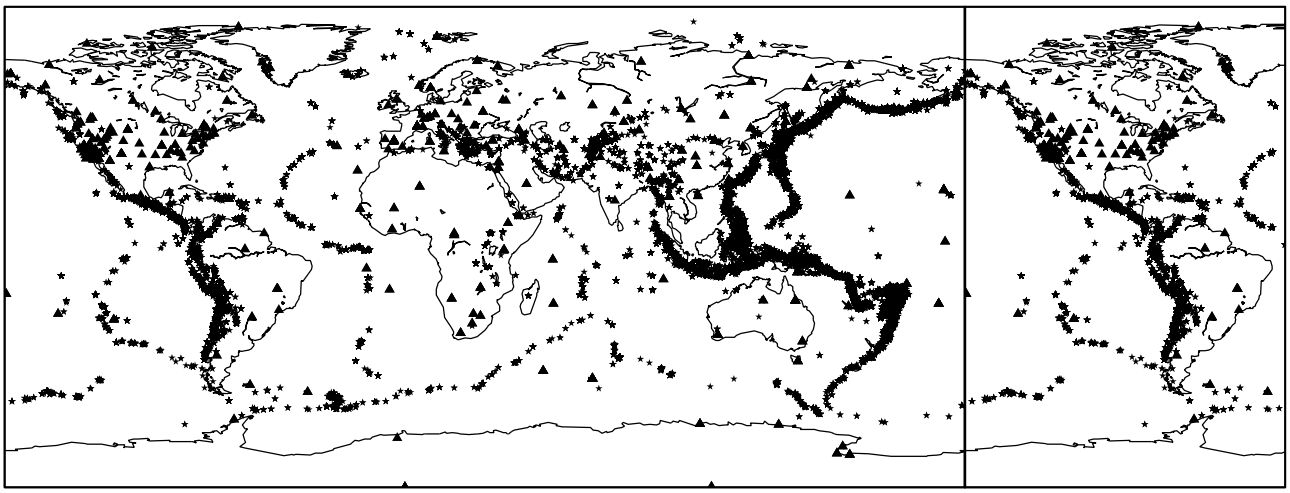
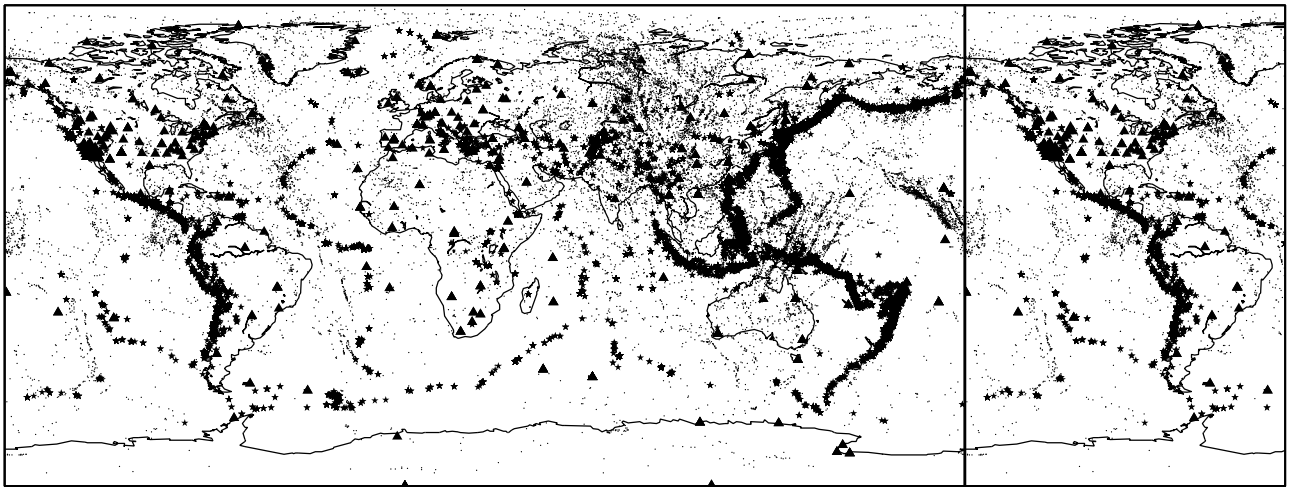


Figure 4. Ray-perpendicular cross sections of the Fréchet kernel (a) for a P wave between the source and the turning point and (b) for a PP wave between the source and the bounce point. The shape of the PP kernel changes drastically upon passage of a caustic.



(a) P data



(b) PP-P data

Figure 5. (a) Distribution of sources (stars) and receivers (triangle) for P data. (b) Distribution of sources (stars), stations (triangles) and bounce points (dots) for PP data.

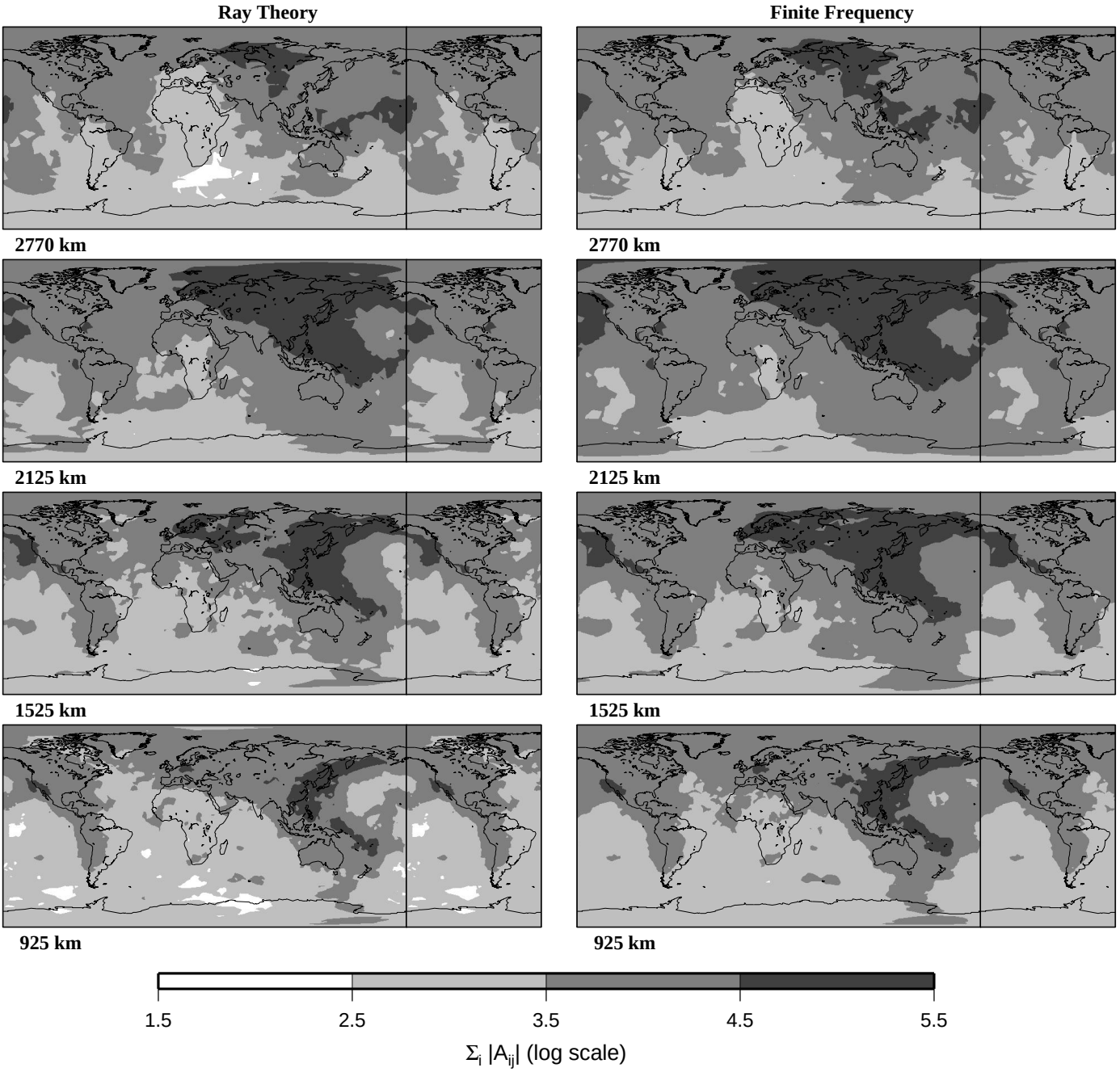


Figure 6. Sections at different depth of the density of the matrix A for ray theory (left) and finite-frequency waves (right), expressed as the sum of the absolute values of the elements of each column of the matrix A. Note that the maps have been "wrapped around" to aid in the visualization of patterns in the vicinity of the Greenwich meridian.

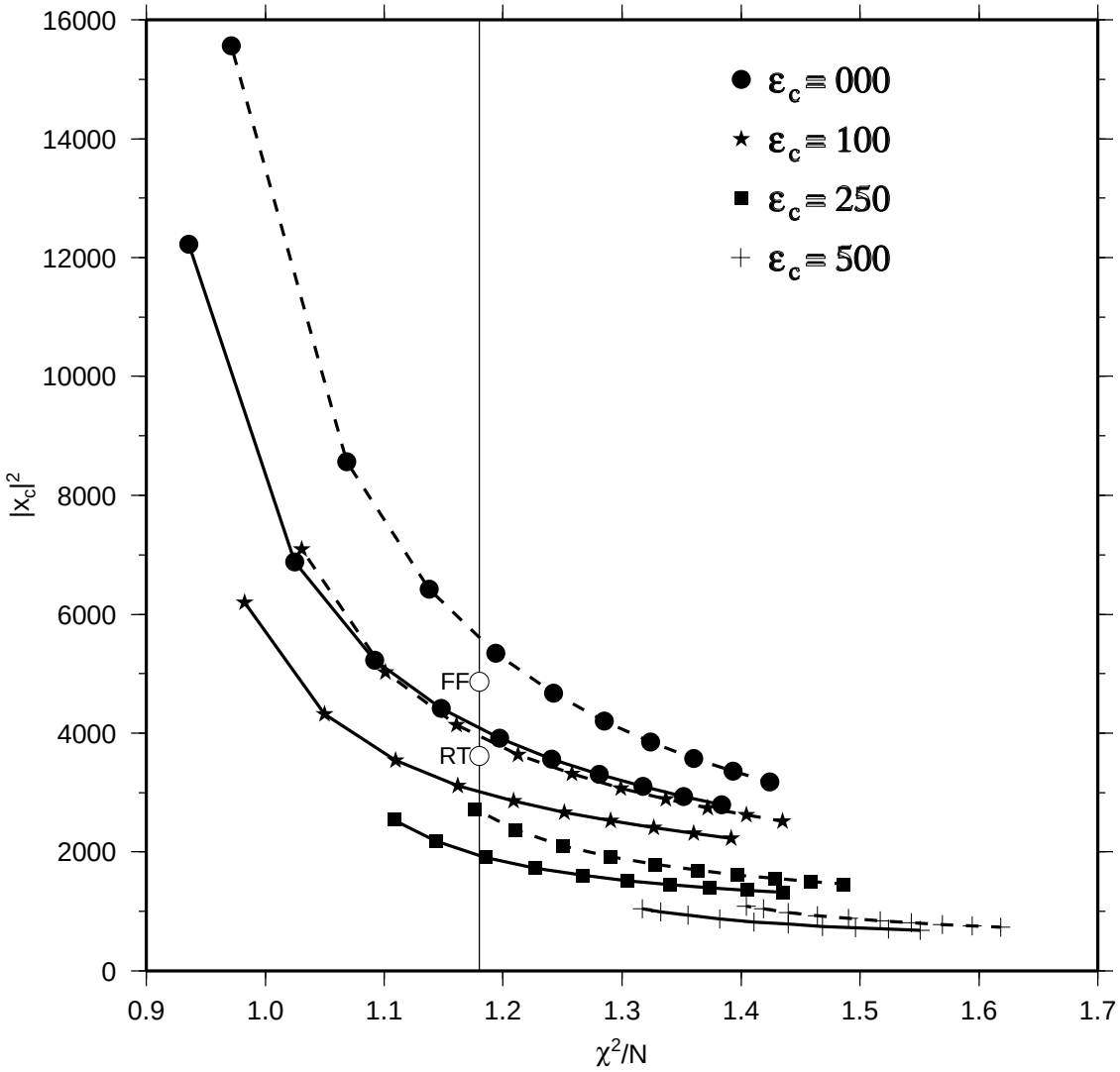


Figure 7. Model norm versus χ^2/N for a combination of values of norm damping ϵ_c and smoothing ϵ_S. Solutions for ray-theory (solid line) and finite-frequency (dotted line) tomographies are compared. Symbols – listed in the legend – correspond to different norm damping parameters. Smoothing increases from upper right to lower left along the curves. The two white dots indicate the FF and RT solutions, respectively, which are discussed in the text and in Figs. 9– 11.

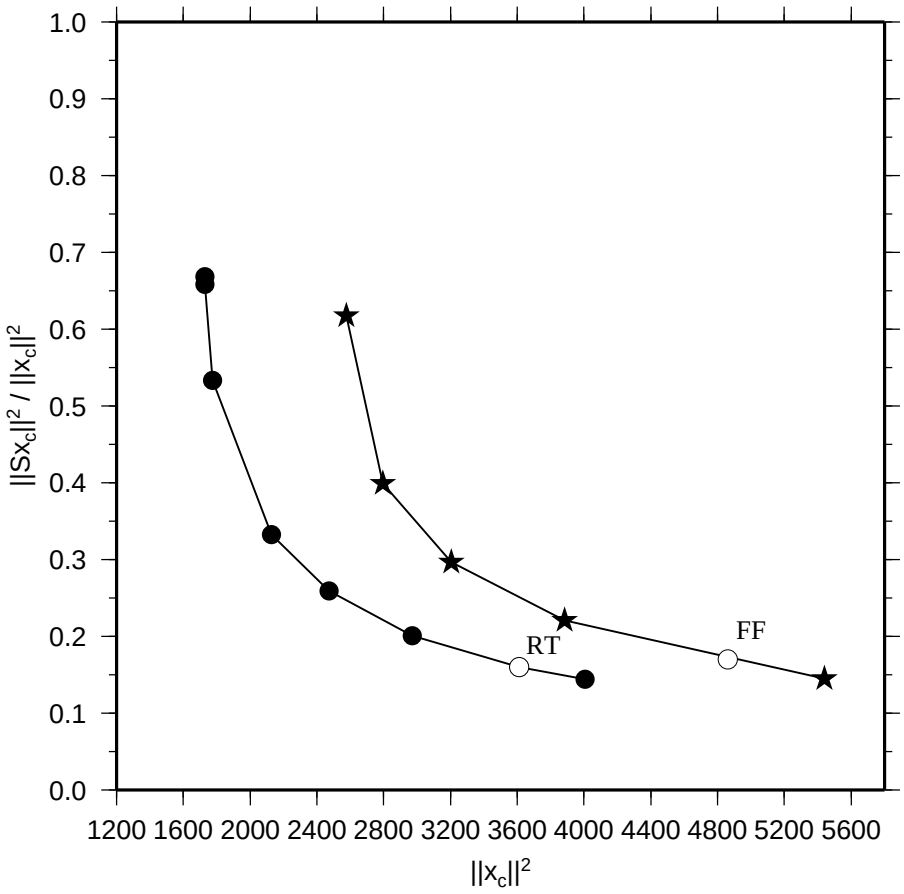


Figure 8. Model norm versus roughness of the solution for finite-frequency (stars) and ray-theoretical (dots) tomography with $\chi^2/N = 1.18$. The two white dots indicate locations of finite-frequency and ray theory models compared in Figs. 9– 11.

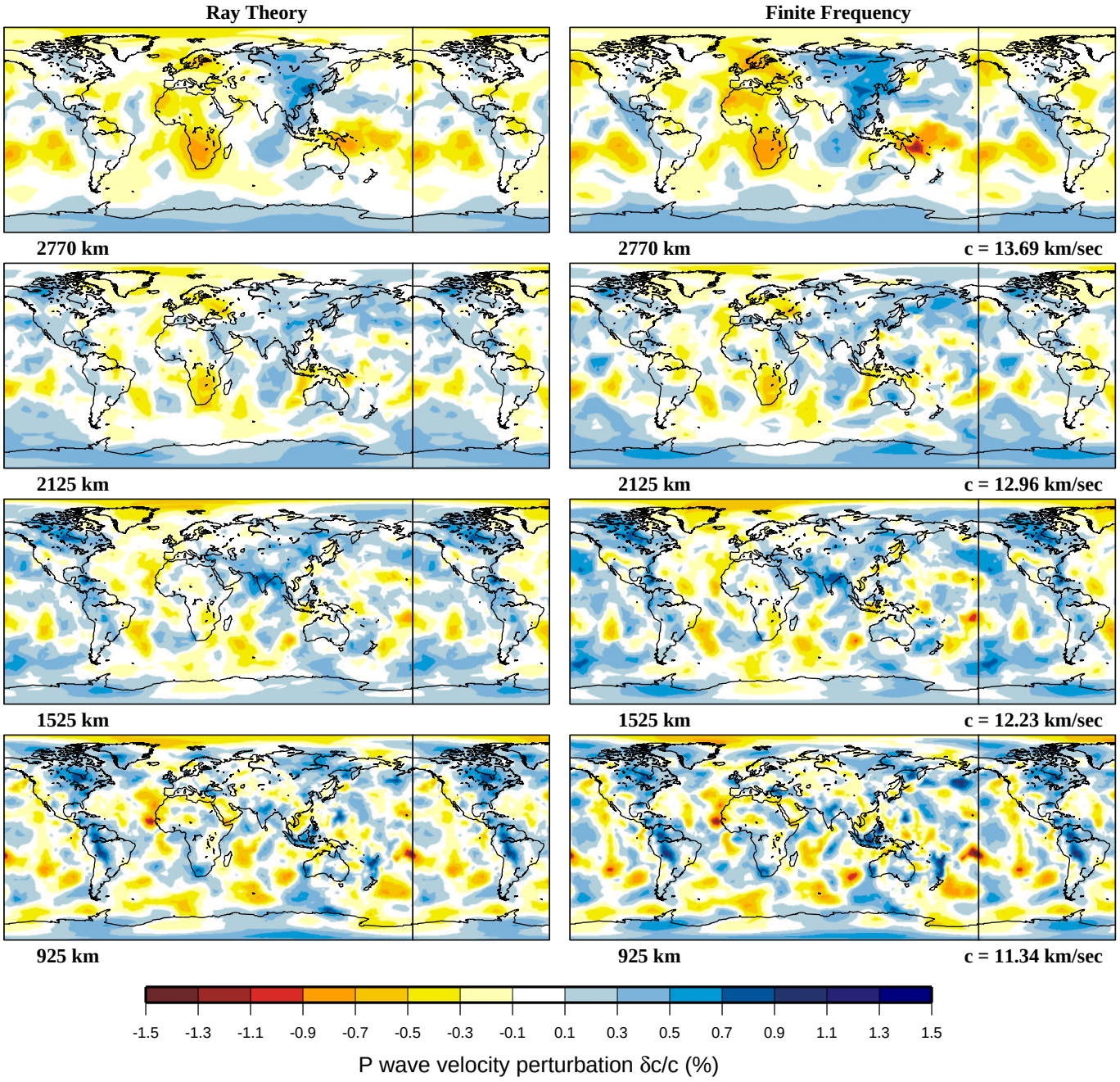


Figure 9. Comparison between velocity maps of the smooth, $\chi^2/N = 1.18$ model for ray theory (left) and finite-frequency theory (right) at different depths. The quantity c is the velocity in the reference model shown in Fig. 2. Maps have been "wrapped around" to aid in visualization of patterns both in the Atlantic and the Pacific Oceans.

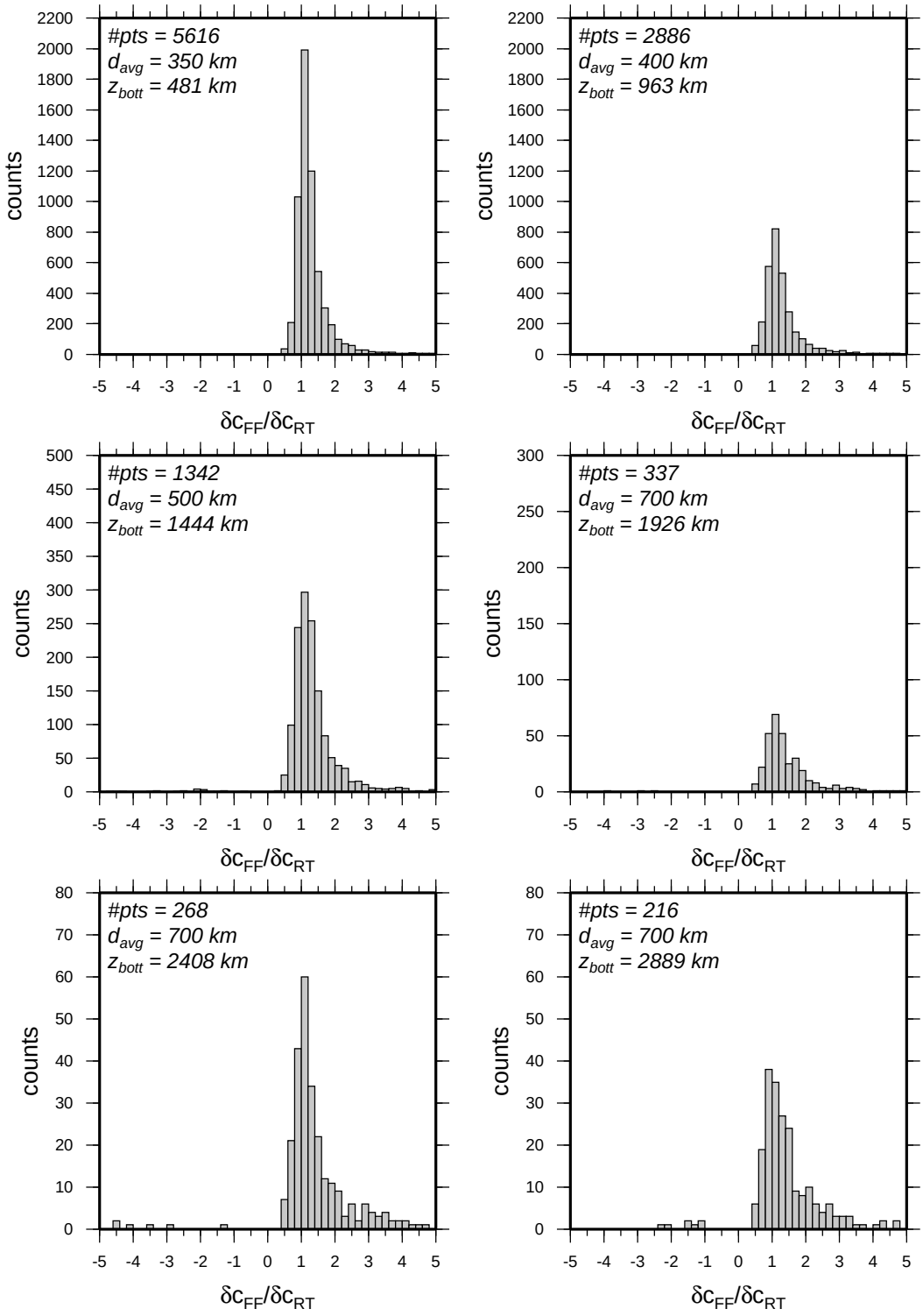


Figure 10. Histograms showing the ratio between finite-frequency and ray-theoretical velocity changes at different depths, for the model with $\chi^2/N = 1.18$. Only changes with absolute value larger than 0.2% are binned. Depth in the histograms is representative of the depth at the bottom of the shell considered (z_{bott}). There are six shells from the surface of the Earth down to the core-mantle boundary. The number in the corner represents the bottom depth of the shell in kilometers. For each layer the total number of points of the grid present in that layer ($\#\mathrm{pts}$) and their average spacing (d_{avg}) are indicated.

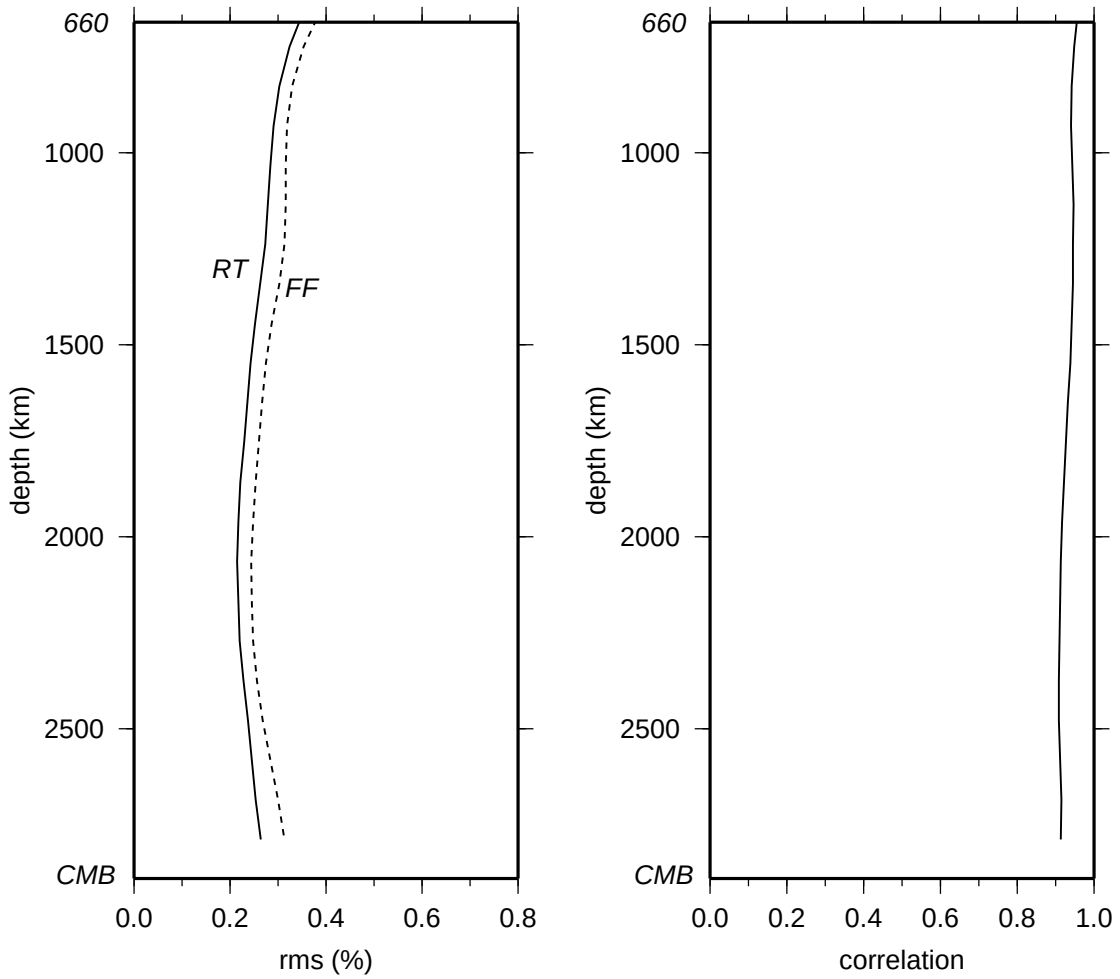


Figure 11. (Left) Root-mean-square velocity perturbation $\delta c/c$ versus depth for the finite-frequency (FF – dashed line) and ray-theoretical tomography (RT – solid line). (Right) Correlation coefficient between the finite-frequency and ray-theoretical models versus depth. The data fit criterion in both inversions is $\chi^2/N = 1.18$.

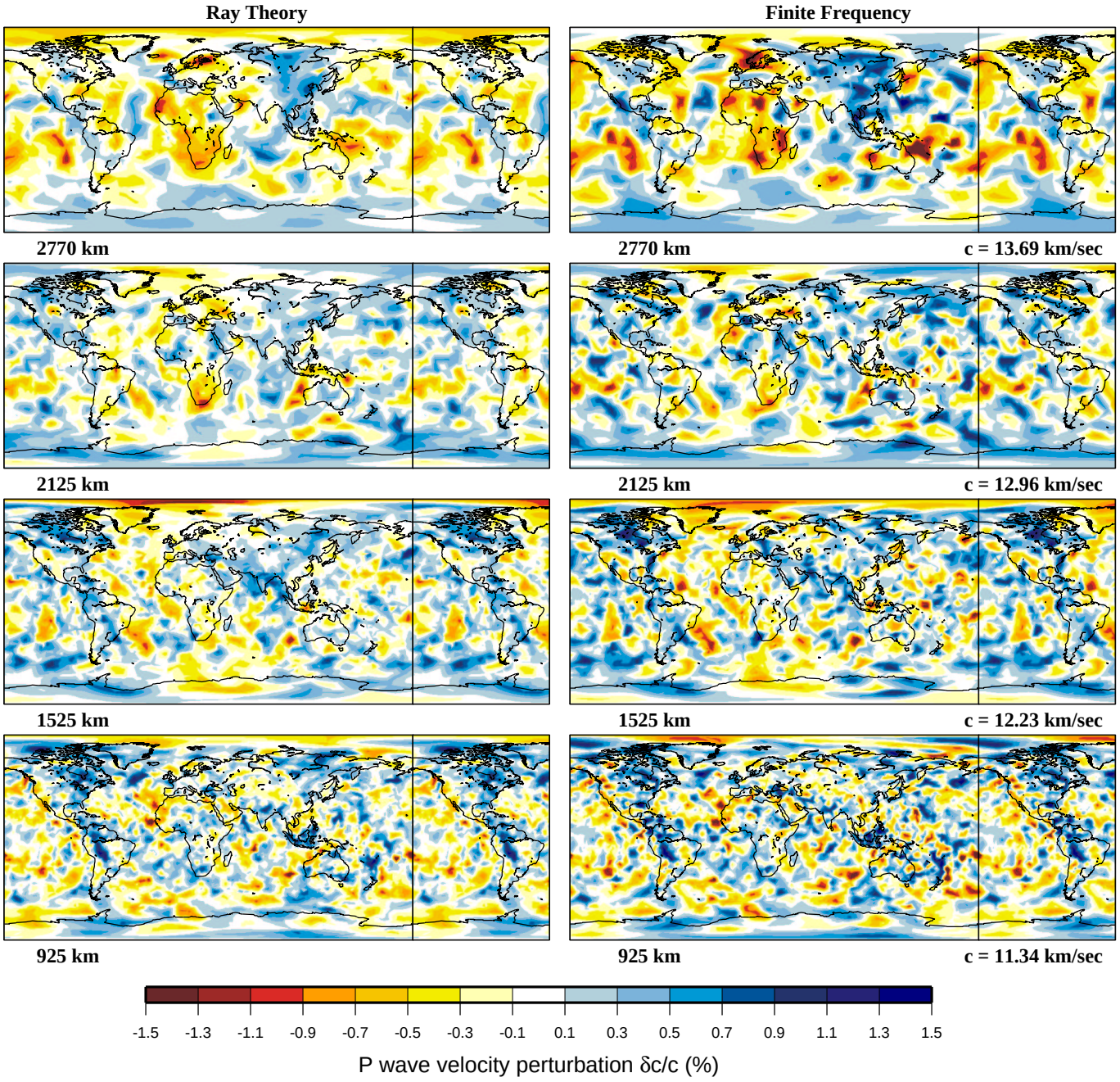


Figure 12. Comparison between velocity maps of the rough, $\chi^2/N = 1$ model for ray theory (left) and finite-frequency (right) at different depths. Maps have been ”wrapped around“ to aid visualization of patterns both in the Atlantic and the Pacific Oceans.

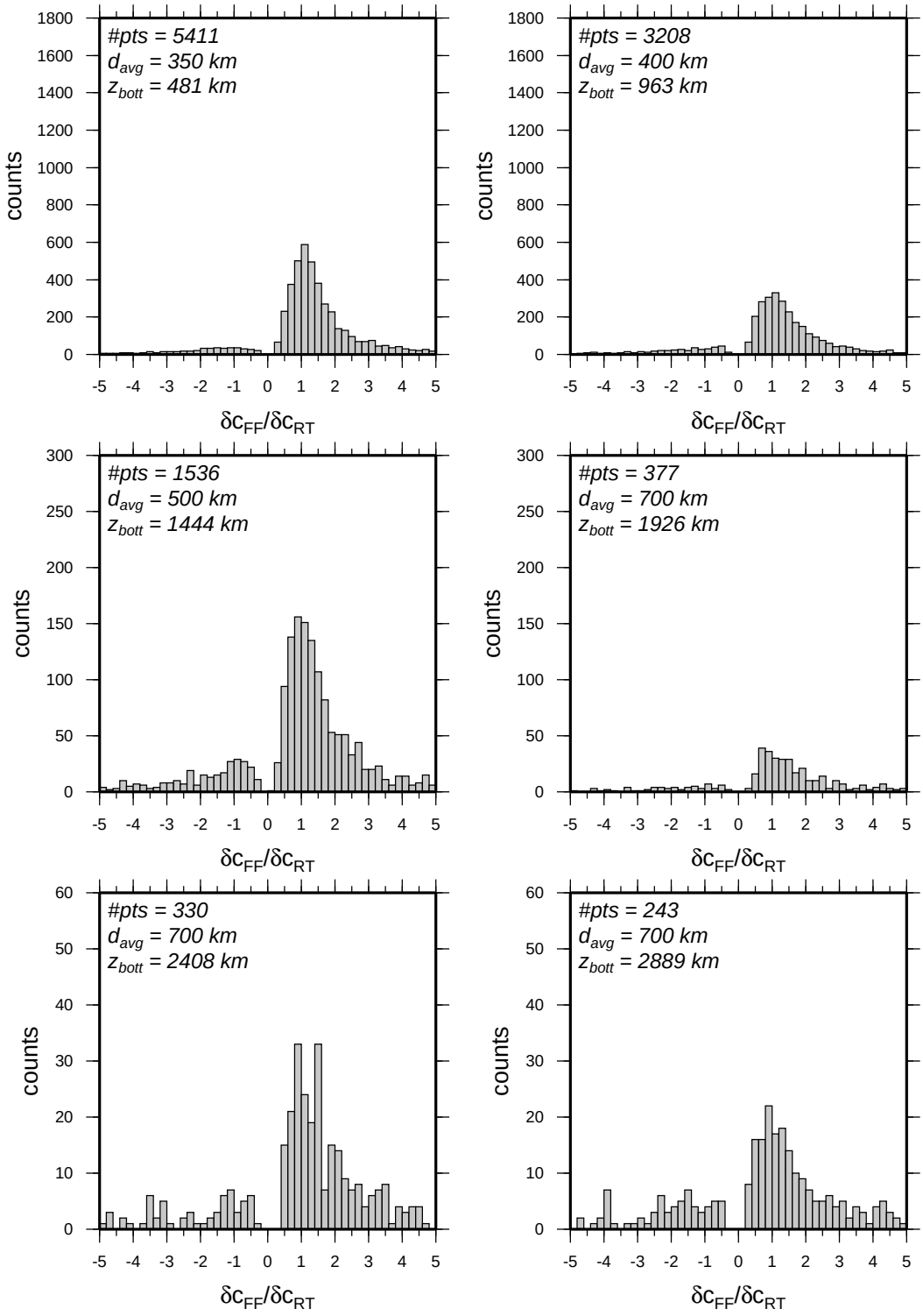


Figure 13. Histograms showing the ratio between finite-frequency and ray-theoretical velocity changes at different depths, for the models with $\chi^2/N = 1$. Only changes with absolute value larger than 0.2% are binned. There are six shells from the surface of the Earth down to the core-mantle boundary. The number in the corner represents the bottom depth of the shell in kilometers.

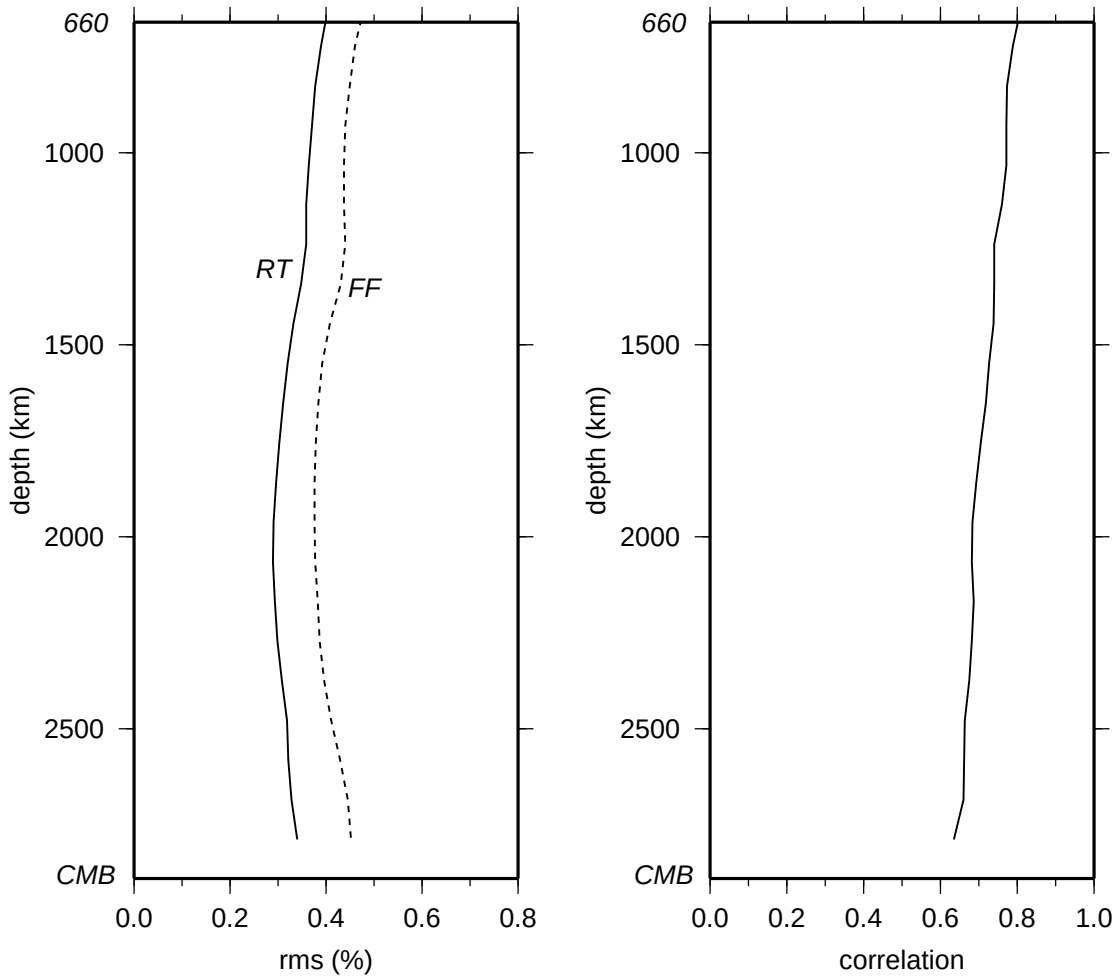


Figure 14. (Left) Root-mean-square velocity perturbation $\delta c/c$ versus depth for the finite-frequency (FF–dashed line) and ray theoretical tomography (RT – solid line). (Right) Correlation coefficient between the finite-frequency and ray-theoretical models versus depth. The data fit criterion in both inversions is $\chi^2/N = 1$.

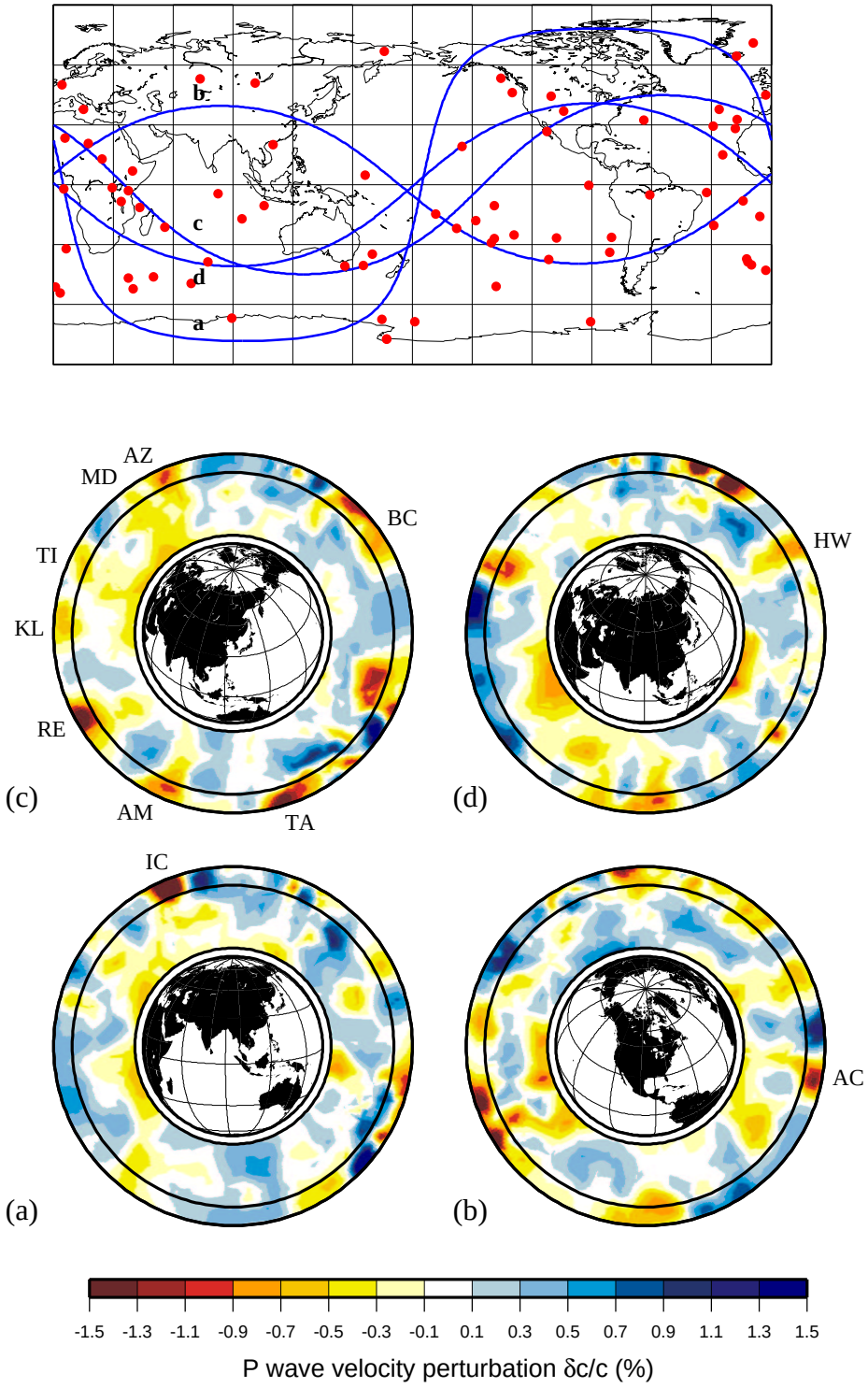


Figure 15. Cross sections of the finite-frequency model with $\chi^2/N = 1.18$. The top figure shows the four great circle paths. Letters (a)–(d) on the paths match plots below. (a) cross section across Greenland and Iceland (pole location $94.82°$W, $11.57°$N), (b) cross section through the Pacific superwell (pole location $96.50°$E, $62.44°$N), (c) cross section across La Reunion and the African hotspots (pole location $125.39°$W, $45.09°$N), (d) cross section across the Atlantic superwell and Hawaii (pole location $90.94°$W, $41.16°$N). Two-letter hotspot identifiers are listed in Table 1.

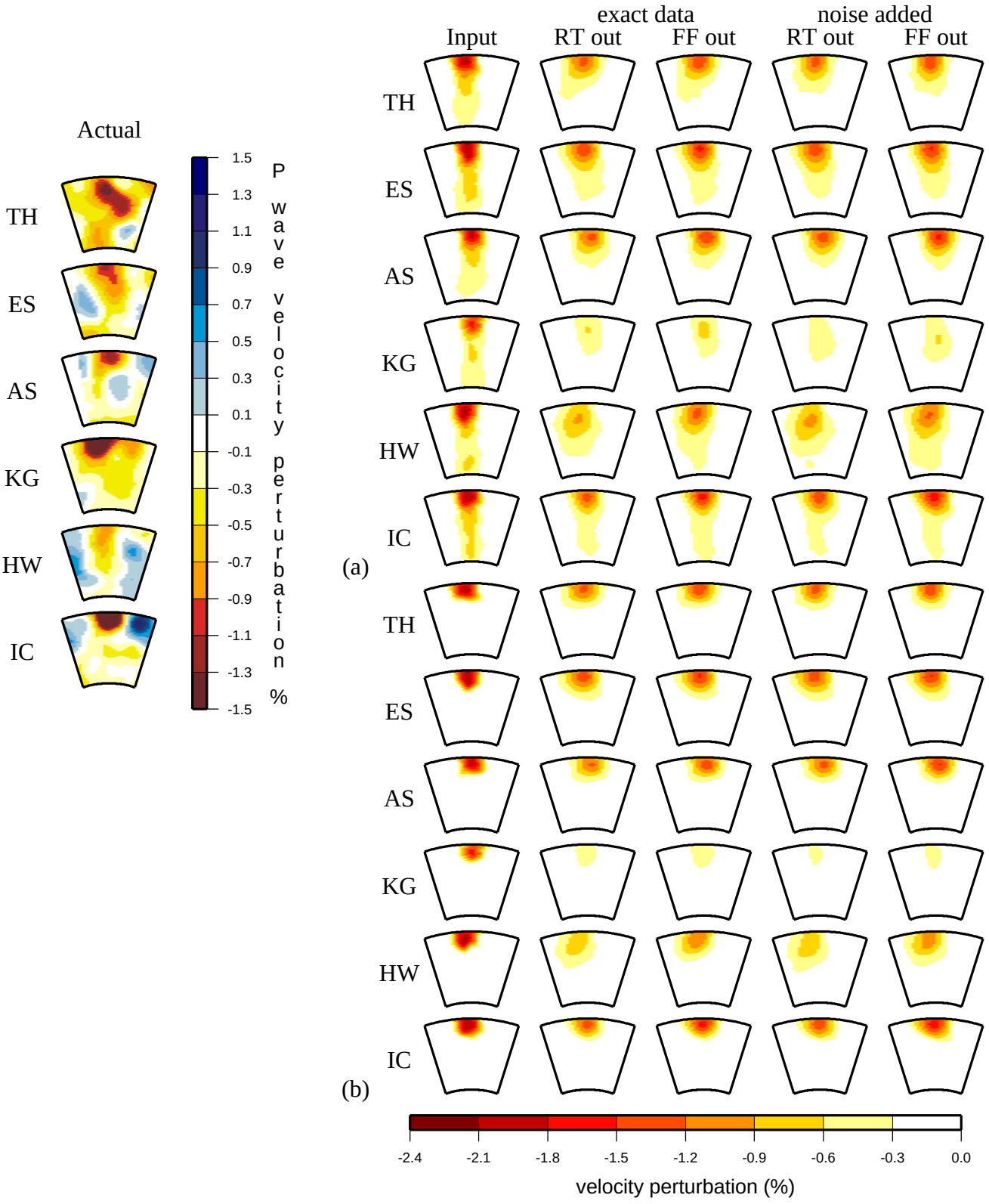


Figure 16. Resolution tests for six of the major hotspots observed in the tomographic images. (Left) Recovered velocity model (Actual) for $\chi^2/N = 1.18$; (Right) Resolution tests: from left to right we present the input model (Input), the recovered model obtained by inverting the synthetic delay times δT using the ray-theoretical inverse (RT out) and the finite-frequency inverse (FF out), respectively. The rightmost two columns show the corresponding recovered models in the case we invert the synthetic residuals after the addition of normally distributed random noise. Panel (a) shows the results with the synthetic hotspots reaching the core-mantle boundary; panel (b) shows the results with the hotspots originating in the upper mantle (around 660 km depth). Two-letter hotspot identifiers are listed in Table 1.

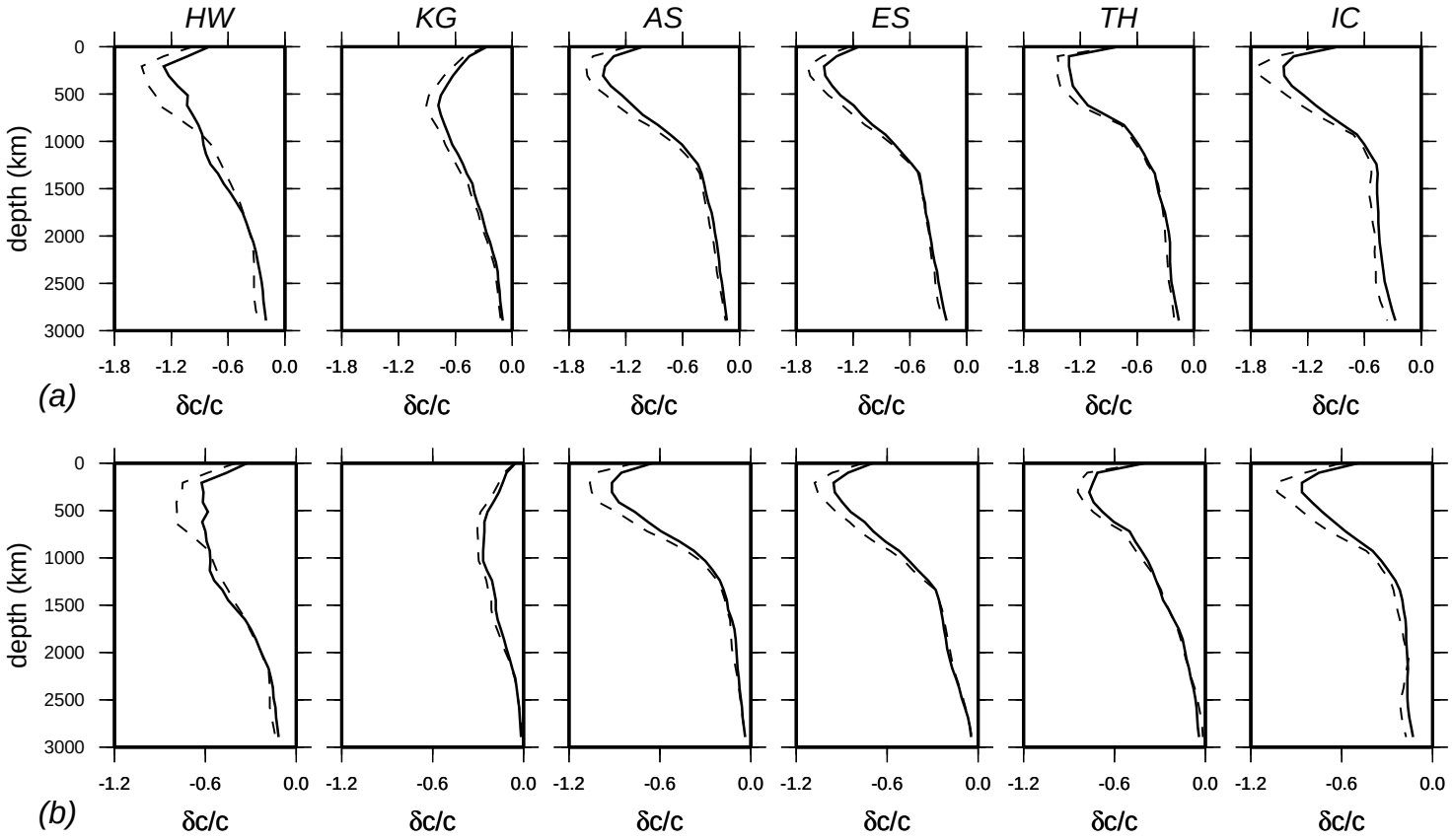


Figure 17. Difference between the finite-frequency and ray theory maximum velocity perturbation $\delta c/c$ as a function of depth for the 5 plumes showed in Fig. 16. Synthetic times are obtained by using finite-frequency modeling. No random errors were added in this case. Two-letter hotspot identifiers are listed in Table 1. Panel (a) shows the results of the test done with synthetic plumes with 500 km radius, (b) shows the results with plumes of 300 km radius