Appendix

A Inversion for azimuthal anisotropy

This first appendix summarizes a few key points treated in more detail in [37].

For each of the paths $p = 1, ..., M$ a waveform fit results in a set of $K(p)$ depth-dependent constraints. These individual constraints are uncorrelated: this is achieved explicitly by the diagonalization of their error covariance matrix [68]. By themselves modal Fréchet kernels are not orthogonal, so phase velocity measurements in a particular frequency band [69] cannot be converted into depth-dependent shear velocity models without this extra step [37]. In the formalism of [68] every seismogram yields a set of uncorrelated depth-distributed averages of the three-dimensional wavespeed structure:

$$q_k^p = \int_{P_p} \langle G_k^p(r) | \, \delta\beta(\ , \theta, \varphi) \rangle \; d\Delta. \qquad \text{(A.1)}$$

Here, $k = 1, ..., K(p)$ indexes the number of independent constraints that have been retained after applying an error cut-off criterion (discussed by [70]), and $d\Delta$ indicates the incremental epicentral distance along the great circle path P_p. The "data" q_k^p have been scaled to unit variance. We've used the notation: $\langle f | g \rangle \equiv \int_0^a f(r) g(r) dr$, where the integration goes from the center ($r = 0$) to the surface ($r = a$) of the Earth. The 3-D shear wave speed anomaly is denoted as $\delta\beta(\mathbf{r}) = \delta\beta(\ , \theta, \varphi)$ and $G_k^p(r)$ is a kernel function which indicates how the data constraint q_k^p is related to the model parameters $\delta\beta$.

The precise form of $G_k^p(r)$ is determined by the error structure of the parameters determined in the nonlinear waveform inversion; practically, as more (and better) information is available on the velocity structure, the total number k increases to some number $K(p)$ for a particular path. Insofar as this information is constrained by higher modes or lower-frequency fundamental modes of the seismogram, the sensitivity of the kernel $G_k^p(r)$ will shift to deeper structure as k increases. The value of q_k^p is influenced by the local path reference model that was used in obtaining the fit [71]; the velocity anomaly is with respect to a regional 1-D background model common to all paths.

The velocity anomalies are expanded onto a set of equal-area surface cells parameterized as $l_t(\theta, \varphi)$ for $t = 1, ..., T$ where $l_t(\theta, \varphi) = 1$ if the horizontal coordinates (θ, φ) are inside the cell, and 0 elsewhere. The radial basis functions are denoted as $h_j(r), j = 1, ..., J$; they can be discrete layers or high-order splines, but are here taken to be boxcar (at the crustal, 400 and 670 km discontinuities) and tent functions (interpolating between nodes at 15, 30, 80, 140, 200 and 300 km). The surface and depth parameterization describe a set of basis functions $s_i(\mathbf{r}) = l_t(\mathbf{r}/r) \, h_j(r)$, with $i = 1, ..., N$ and $N = TJ$ upon which to expand $\delta\beta(\mathbf{r})$ as $\delta\beta(\mathbf{r}) = \sum_i^N b_i s_i(\mathbf{r})$. Inserting these definitions into Eq. A.1 we obtain [70]:

$$q_k^p = \sum_i^N b_i \, L_t^p \, \langle G_k^p(r) | h_j(r) \rangle \; . \qquad \text{(A.2)}$$

We've introduced L_t^p, the path length of path p in surface cell t. All k constraints of one path sample the same cells on the surface, but the depth-averaging functions differ.

Defining $\mathcal{G}$ to be dependent on the path lengths in the surface cells and on the sensitivity to depth structure which is influenced by the mode structure of the seismogram and the diagonalized error covariance matrix of the waveform inversion, we may write:

$$q_k^p = \sum_i^N \mathcal{G}_{ki}^p b_i \quad \text{or} \quad \mathbf{q} = \mathcal{G} \cdot \mathbf{b}. \qquad \text{(A.3)}$$

The dimensions of the matrices involved are $[\mathbf{q}] = \sum_{p=1}^M K(p) \times 1$, $[\mathbf{G}] = \sum_{p=1}^M K(p) \times N$, and $[\mathbf{b}] = N \times 1$.

B Hermite multispectrogram analysis

This second appendix summarizes a few key points treated in more detail in [44].

For two spatially variable (non-stationary) random processes $\{G\}$ (gravity) and $\{H\}$ (topography), defined on $\mathbf{r} = (x, y)$ in the spatial domain and on $\mathbf{k} = (k_x, k_y)$ in the Fourier domain, the coherence-square function relating both fields, γ^2_{GH}, is defined as the ratio of their cross-spectral density, S_{GH}, normalized by the individual power spectral densities, S_{GG} and S_{HH}:

$$\gamma^2_{GH}(\mathbf{r}, \mathbf{k}) = \frac{|S_{GH}(\mathbf{r}, \mathbf{k})|^2}{S_{GG}(\mathbf{r}, \mathbf{k}) S_{HH}(\mathbf{r}, \mathbf{k})}. \quad \text{(B.1)}$$

An estimate $\hat{S}$ of the spectral density function is obtained by taking a weighted average of N windowed periodograms:

$$\hat{S}_{GG}(\mathbf{r}, \mathbf{k}) = \left(\sum_{n=0}^{N-1} \lambda_n \right)^{-1} \times \sum_{n=0}^{N-1} \lambda_n \left| \int h_n(\boldsymbol{\tau} - \mathbf{r}) G(\boldsymbol{\tau}) e^{-i2\pi \mathbf{k} \cdot \boldsymbol{\tau}} \, d\boldsymbol{\tau} \right|^2. \quad \text{(B.2)}$$

with data windows h_n and weights λ_n. The integration is implemented discretely by the Fast Fourier Transform algorithm. The properties of the windows and weights determine the resolution and variance of the estimate. Maximal, symmetric concentration in the space-wavenumber domain [72,73] is obtained by using data windows that are Hermite polynomials (calculated by recursion), modulated by the Gaussian:

$$h_n(x) = \frac{H_n(x) e^{-x^2/2}}{\pi^{1/4} \sqrt{2^n n!}}. \quad \text{(B.3)}$$

The weights are given by:

$$\lambda_n(R) = \frac{1}{n!} \gamma(R^2/2, n+1), \quad \text{(B.4)}$$

where γ denotes the incomplete gamma function. We used $R = 3$ and $N = R^2$ windows in each dimension. The parameter R regulates the resolution of the spectral estimate. As it is the radius of a spherical domain in the space-wavenumber phase plane, spatial resolution is not traded off with spectral resolution, and no spurious anisotropy is introduced by combining dimensions. This is the main advantage over conventional multi-taper methods with Slepian functions as windows, which are not designed to capture spatial variations and must assume local stationary instead. We sampled the spatial domain at points about 500 km apart [44].

The data windows h_n are eigenfunctions of a spatio-spectral concentration operator, with eigenvalues λ_n. To study multi-dimensional processes, windows are formed by taking the outer product of all pairs of 1-D functions, and the eigenvalues by multiplication. Eq. B.4 shows the latter can be calculated cheaply without actually solving the concentration eigenvalue problem. This is responsible for the speed advantage of the method compared to the Slepian multitaper method.

With orthonormal data windows the individual spectral estimates are approximately uncorrelated and the variance of the coherence estimate $\hat{\gamma}^2$ decreases with the number of windows. In function of the true coherence γ^2, this variance is given by

$$\sigma^2\{\hat{\gamma}^2\} = 2\gamma^2 \frac{(1-\gamma^2)^2}{N}. \quad \text{(B.5)}$$

However, the $\sim 1/N$ decrease in the estimation variance is achieved at the expense of widening the concentration region $R = \sqrt{N}$, which degrades the spectral resolution. The errors quoted will be the square root of the average estimation variance divided by the square root of the number of points estimated on a line with constant azimuth through the center of the spectrum. In [44] we assess the significance of the minima by comparing the separation of azimuthal profiles of coherence in “weak” and “strong” directions with their errors.