

RESEARCH ARTICLE

Resolution analysis by random probing

10.1002/2015JB012106

Andreas Fichtner¹ and Tristan van Leeuwen²

Key Points:

- Quantitative resolution analysis with low computational requirements
- Applicability to large-scale tomography including full-waveform inversion
- Quantification of spatial resolution and interparameter trade-offs

Correspondence to:

A. Fichtner,
andreas.fichtner@erdw.ethz.ch

Citation:

Fichtner, A., and T. van Leeuwen (2015), Resolution analysis by random probing, *J. Geophys. Res. Solid Earth*, 120, 5549–5573, doi:10.1002/2015JB012106.

Received 9 APR 2015

Accepted 5 JUL 2015

Accepted article online 14 JUL 2015

Published online 6 AUG 2015

¹Department of Earth Sciences, ETH Zurich, Zurich, Switzerland, ²Mathematical Institute, Utrecht University, Utrecht, Netherlands

Abstract We develop and apply methods for resolution analysis in tomography, based on stochastic probing of the Hessian or resolution operators. Key properties of our methods are (i) low algorithmic complexity and easy implementation, (ii) applicability to any tomographic technique, including full-waveform inversion and linearized ray tomography, (iii) applicability in any spatial dimension and to inversions with a large number of model parameters, (iv) low computational costs that are mostly a fraction of those required for synthetic recovery tests, and (v) the ability to quantify both spatial resolution and interparameter trade-offs. Using synthetic full-waveform inversions as benchmarks, we demonstrate that autocorrelations of random-model applications to the Hessian yield various resolution measures, including direction- and position-dependent resolution lengths and the strength of interparameter mappings. We observe that the required number of random test models is around five in one, two, and three dimensions. This means that the proposed resolution analyses are not only more meaningful than recovery tests but also computationally less expensive. We demonstrate the applicability of our method in a 3-D real-data full-waveform inversion for the western Mediterranean. In addition to tomographic problems, resolution analysis by random probing may be used in other inverse methods that constrain continuously distributed properties, including electromagnetic and potential-field inversions, as well as recently emerging geodynamic data assimilation.

1. Introduction

Since its conception and first applications [e.g., *Aki and Lee, 1976; Aki et al., 1976; Dziewoński et al., 1977*], seismic tomography has developed into one of the most powerful tools to study the internal structure of the Earth. Tomographic models play a central role, for instance, in the location of earthquakes, the exploration for natural resources, the modeling of seismically induced ground motion, geodynamic reconstructions, and studies of the Earth's thermochemical state. Despite rapid progress—driven by improved data coverage and methodological advances—tomographic models often differ over length scales that individual studies would claim to be resolved [e.g., *Becker and Boschi, 2002; Fichtner et al., 2012; Schaeffer and Lebedev, 2013*]. While the subjectivity of regularization in deterministic inversions already biases resolution estimates, the problem is aggravated by the fact that uncertainty analyses are often based on synthetic recovery tests, typically of the checkerboard type. In addition to being potentially misleading [*Lévéque et al., 1993*], recovery tests are only qualitative and therefore unable to quantify if two models fall within their mutual uncertainties. The common practice to perform recovery tests seems inconsistent with the wealth of methods for quantitative resolution analysis that has been proposed in the literature.

1.1. Methods for Uncertainty Analysis

Recognizing that the solution of inverse problems for a finite collection of data is synonymous to the description of the infinite dimensional manifold of plausible models, early methods for quantitative resolution analysis were developed by *Backus and Gilbert* [1968, 1970] when data volumes grew too large for manual hypothesis testing.

Subsequent developments in uncertainty quantification for potentially nonlinear problems with a comparatively small number of model parameters include various probabilistic methods to explore the posterior distribution in model space. Examples with geophysical applications are the Metropolis algorithm and its variants [e.g., *Mosegaard and Tarantola, 1995; Sambridge and Mosegaard, 2002; Tarantola, 2005*], the neighborhood algorithm [*Sambridge, 1999a, 1999b; Mosca et al., 2012*], genetic algorithms [*Gallagher et al., 1991; Sambridge and Drijkoningen, 1992*], and neural networks [*Meier et al., 2007a, 2007b; de Wit et al., 2013*].

In the realm of nearly linear problems, the list of frequently used techniques comprises the analysis of posterior covariance and resolution matrices [e.g., *Tarantola, 2005; Nolet, 2008*], efficient methods to compute these matrices for large tomographic systems [*Boschi, 2003; Nolet et al., 1999*], extremal bounds analysis [*Meju and Sakkas, 2007; Meju, 2009*], nullspace shuttles [*Deal and Nolet, 1996*], and position- and direction-dependent point spread functions for ray tomography [*Yanovskaya, 1997, 2000*]. A recent review of widely used methods may be found in *Rawlinson et al. [2014]*. Our inability to explicitly compute and store the resolution matrix for large linear tomographic systems has motivated a new class of stochastic algorithms where specific resolution proxies are estimated from the application of the resolution matrix to an ensemble of random test vectors: *Sacchi et al. [2007]* demonstrate that the diagonal of the migration point spread function can be estimated using random test images. *MacCarthy et al. [2011]* compute stochastic estimates of the trace of the resolution matrix that can be used in generalized cross validation. *An [2012]* proposes to parameterize the resolution matrix and to estimate the parameters from the application to random Earth model perturbations. *Trampert and Fichtner [2013]* estimate horizontal and vertical resolution lengths in the Earth model S40RTS [*Ritsema et al., 2011*] from a single random test model. Methods specifically targeting resolution proxies are members of a larger family of random probing techniques to infer properties of large matrices. These include Hutchinson's method and its variants to estimate the trace [*Hutchinson, 1990; Avron and Toledo, 2011*], the approach of *Bekas et al. [2007]* to approximate the diagonal, different techniques to obtain low-rank approximations [*Frieze et al., 2004; Drineas et al., 2006; Halko et al., 2011*], and to perform principal component analysis [*Rokhlin et al., 2009*].

1.2. Resolution Analysis in Full-Waveform Inversion

New challenges in uncertainty quantification arose with the advent of tomographic techniques based on purely numerical wavefield simulations and adjoint techniques [e.g., *Igel et al., 1996; Pratt, 1999; Dessa et al., 2004; Fichtner et al., 2009a; Virieux and Operto, 2009; Kawai and Geller, 2010; Tape et al., 2010; Rickers et al., 2013; Colli et al., 2013; Afanasiev et al., 2014*]. Commonly referred to as full-waveform inversion, these methods have computationally expensive forward problem solutions, large model spaces, and a nonlinear dependence between measurements and model parameters. Therefore, full-waveform inversion does not fall into the range of applicability of the aforementioned techniques. Alternative approaches include the analysis of wave number coverage in a diffraction tomography context [e.g., *Mora, 1989*] and the comparison with well-log data, when these are available [e.g., *Bleibinhaus et al., 2007*]. Often, however, resolution analyses are not performed, and the solution of the full-waveform inversion problem is reduced to data fitting.

Recently proposed methods for resolution analysis in full-waveform inversion are based on parameterized or low-rank approximations of the Hessian operator, which is equal to the inverse posterior covariance when all errors are Gaussian [*Fichtner and Trampert, 2011a, 2011b; Bui-Thanh et al., 2013*]. While being applicable to large-scale problems, the practical popularity of these methods is, however, limited by their algorithmic complexity and computational requirements. These are main issues that we aim to address.

1.3. Objectives and Outline

The principal objective of this work is to develop and apply methods for uncertainty analysis characterized by (i) applicability to any tomographic problem, including full-waveform inversion, (ii) low algorithmic complexity, (iii) applicability in any spatial dimension and for a potentially large number of model parameters, and (iv) the ability to quantify both spatial resolution and interparameter trade-offs. These methods are based on stochastic probing techniques, already alluded to in section 1.1.

A central role in our approach is played by the point spread function which we briefly review in section 2 in order to set the stage for further developments. Using intuitive analytical examples in one dimension, section 3 is intended to introduce a simple method to infer position-dependent resolution length from the application of the Hessian to random test models and the subsequent autocorrelation of the results. In section 4 we apply and extend this concept using a synthetic full-waveform inversion for S velocity in 2-D. Synthetic multiparameter inversions are then described in section 5, where we compute position- and direction-dependent resolution lengths of the individual parameters, as well as measures that quantify interparameter trade-offs. To demonstrate the applicability of our methods to realistic large-scale problems, we present a 3-D real-data full-waveform inversion example for the western Mediterranean in section 6. Following numerous numerical examples, section 7 is intended to substantiate our empirical observations concerning convergence from a theoretical perspective. Finally, a condensed review of second-order adjoints may be found in the appendix.

2. Point Spread Functions

To set the stage and to introduce basic notation, we start with a brief review of point spread functions, i.e., the response to a point-localized heterogeneity. For this we assume that differences between observed and synthetically computed data, as well as prior information on the Earth model \mathbf{m} , are quantified in terms of a misfit functional $\chi(\mathbf{m})$. In the specific case of a least squares problem, $\chi(\mathbf{m})$ may be defined as

$$\chi(\mathbf{m}) = \frac{1}{2} \mathbf{d}^T \mathbf{C}_d^{-1} \mathbf{d} + \frac{1}{2} \mathbf{m}^T \mathbf{C}_m^{-1} \mathbf{m}, \quad (1)$$

with the data covariance operator \mathbf{C}_d , the model covariance operator \mathbf{C}_m , and a vector of secondary observables \mathbf{d} that may contain differential travel times, differential waveforms, or various other types of observables. We further assume that some optimization method has provided us with an optimal model \mathbf{m} that minimizes $\chi(\mathbf{m})$. Perturbing \mathbf{m} with a point-localized heterogeneity $\delta\mathbf{m}$ in one physical parameter, e.g., P or S velocity, increases the misfit from the minimum $\chi(\mathbf{m})$ to

$$\chi(\mathbf{m} + \delta\mathbf{m}) \approx \chi(\mathbf{m}) + \frac{1}{2} \delta\mathbf{m}^T \mathbf{H}(\mathbf{m}) \delta\mathbf{m}, \quad (2)$$

where $\mathbf{H}(\mathbf{m})$ denotes the Hessian operator of the misfit with respect to \mathbf{m} . Keeping in mind that equation (2) is a second-order approximation in $\delta\mathbf{m}$, we will henceforth replace \approx by $=$ in the interest of a cleaner notation. Recognizing that the misfit gradient is given by $\nabla\chi(\mathbf{m} + \delta\mathbf{m}) = \mathbf{H}(\mathbf{m}) \delta\mathbf{m}$, we can approximately reconstruct \mathbf{m} from its perturbed version $\mathbf{m} + \delta\mathbf{m}$ using a gradient method. Denoting the reconstruction by $\tilde{\mathbf{m}}$, the first iteration of a generic gradient method would yield

$$\tilde{\mathbf{m}} = \mathbf{m} + \delta\mathbf{m} - \mathbf{P}^{-1} \mathbf{H}(\mathbf{m}) \delta\mathbf{m}, \quad (3)$$

with some preconditioner \mathbf{P} that should approximate the Hessian \mathbf{H} [e.g., *Tarantola*, 2005]. In the ideal case where $\mathbf{P} = \mathbf{H}$ is invertible and independent of \mathbf{m} , the iteration converges in one step and the reconstruction is perfect, that is, $\tilde{\mathbf{m}} = \mathbf{m}$. In most applications, \mathbf{H} is neither explicitly available nor invertible. Thus, instead of subtracting the exact point-localized perturbation $\delta\mathbf{m}$ from $\mathbf{m} + \delta\mathbf{m}$, its blurred version

$$\delta\tilde{\mathbf{m}} = \mathbf{P}^{-1} \mathbf{H}(\mathbf{m}) \delta\mathbf{m} \quad (4)$$

will be subtracted. In this context, $\mathbf{P}^{-1} \mathbf{H}(\mathbf{m}) \delta\mathbf{m}$ plays the role of a point spread function and $\mathbf{R} = \mathbf{P}^{-1} \mathbf{H}(\mathbf{m})$ acts as resolution operator. The columns of \mathbf{R} can thus be interpreted as point spread functions. In the worst case, the preconditioner \mathbf{P} is chosen to be proportional to the identity operator \mathbf{I} and the updating scheme (3) simplifies to a steepest-descent method. In this sense, $\mathbf{H}(\mathbf{m}) \delta\mathbf{m}$ can be interpreted as a conservative estimate of the point spread function. In the following paragraphs, we will, in the interest of simplicity, equate the point spread function and its conservative estimate through the Hessian. This convention follows common practice, especially in seismic exploration [e.g., *Berkhout*, 1984; *Nemeth et al.*, 1999; *Valenciano and Biondi*, 2006]. In section 8.2 we discuss the trivial extension to inversion schemes where linear preconditioners are applied.

The power of the point spread function concept lies in its applicability to infinite-dimensional model spaces that we typically encounter in tomographic problems where we seek quantities that are continuously distributed in space. Using the term *averaging kernel*, point spread functions and their dependence on the measurement process were analyzed by *Backus and Gilbert* [1968, 1970]. The notion of directionally dependent point spread functions in tomographic inversions was later discussed by *Yanovskaya* [1997], who also proposed a measure for the width of the point spread function in a specific direction, taking inspiration from *Backus and Gilbert* [1968]. The direction-dependent width of the point spread function quantifies the distance across which a hypothetical point-localized heterogeneity is smeared. It therefore provides useful information on the extent to which specific features in tomographic models can be interpreted.

For most realistic tomographic problems, limited computational resources prevent the explicit calculation and storage of the resolution operator \mathbf{R} and thus the comprehensive quantification of resolution. Motivated by this constraint, numerous methods have been developed in order to estimate specific aspects of matrices that

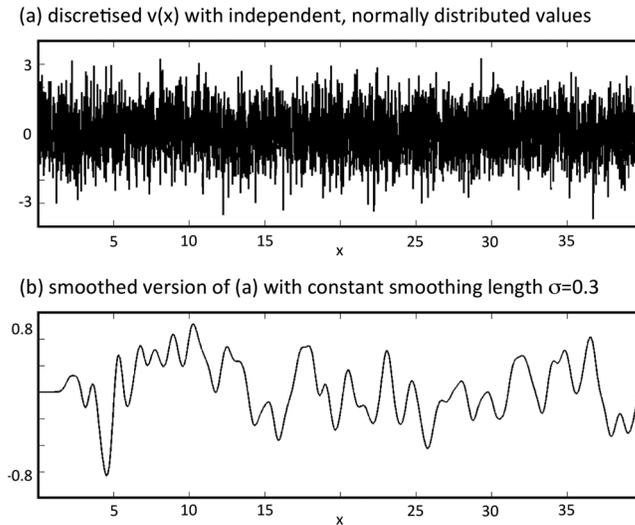


Figure 1. Action of the point spread function on a random function. (a) Discretised random function $v(x)$ where the individual $v_i = v(x_i)$ are independent and normally distributed. (b) Random function low-pass filtered by the Gaussian point spread function from equation (7), $h(y) = \int H(x; y)v(x) dx$. The standard variation in $H(x; y)$ is $\sigma = 0.3$.

are too expensive to store and compute. These include, for instance, Lanczos and Arnoldi iterations to estimate the largest eigenvalues [e.g., Lanczos, 1950; Arnoldi, 1951; Sleijpen and van der Vorst, 2000].

3. Intuitive Examples

We start the development of resolution analysis methods with intuitive 1-D examples to illustrate basic concepts. In one dimension, the Hessian operator is a scalar with two space dependencies, x and y . The action of $H(x; y)$ on a continuously distributed model parameter $v(y)$ takes the form of an integral, that is,

$$Hv = \int H(x; y)v(y) dy. \quad (5)$$

The semicolon in $H(x; y)$ separates the two spatial dependencies of H . The advantage of this notation will become more apparent in later examples in two

and three dimensions where x and y are vectors, the components of which we will separate by colons. Our goal is to infer properties of $H(x; y)$ from its applications to test functions $v(x)$, where for each x_i and $x_j \neq x_i$, the random values $v(x_i)$ and $v(x_j)$ have identical and random distributions with zero mean. Furthermore, we request that the pointwise covariance satisfies the δ property

$$\text{cov}[v(x), v(x')] = \mathbb{E}[v(x)v(x')] = \delta(x - x'), \quad (6)$$

where \mathbb{E} denotes the expectation, that is, the theoretically expected mean of a random variable.

3.1. Recovering the Width of the Point Spread Function

For the purpose of illustration, we first assume that the point spread function $H(x; y)$ is exactly Gaussian, i.e.,

$$H(x; y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-y)^2}, \quad (7)$$

with standard deviation σ . While H itself is hardly ever known explicitly, we can efficiently compute its action on a random function $v(x)$ using second-order adjoints [Santosa and Symes, 1988; Fichtner and Trampert, 2011a]. We shall denote the action of H on v by

$$h(y) = \int H(x; y)v(x) dx = \frac{1}{\sqrt{2\pi}\sigma} \int e^{-\frac{1}{2\sigma^2}(x-y)^2} v(x) dx, \quad (8)$$

In equation (8), the point spread function $H(x; y)$ acts as a convolutional low-pass filter on $v(x)$, meaning that $h(y)$ is a smoothed version of $v(x)$ with an average correlation length approximately equal to σ . An example of a random $v(x)$ and its low-pass-filtered version $h(y)$ is shown in Figure (1).

We can recover information on $H(x; y)$ from $h(y)$ by computing, for instance, the autocorrelation of $h(y)$:

$$c(x) = \int h(x+y)h(y) dy = \iint H(z; x+y)H(z'; y)v(z)v(z') dz dz'. \quad (9)$$

Taking the expectation of c and using equation (6) gives

$$\begin{aligned} \mathbb{E}[c(x)] &= \iint H(z; x+y)H(z'; y)\mathbb{E}[v(z)v(z')] dz dz' \\ &= \iint H(z; x+y)H(z; y) dz dy = \text{const.} e^{-\frac{1}{4\sigma^2}x^2}. \end{aligned} \quad (10)$$

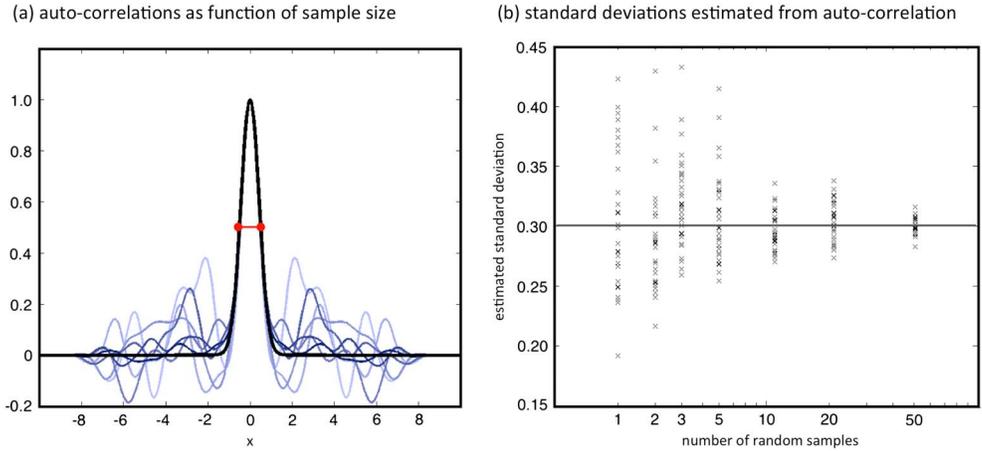


Figure 2. Autocorrelation example. (a) Autocorrelations of $h(y) = \int H(x; y)v(x) dx$ with the point spread function $H(x; y)$ taken to be the Gaussian from equation (7) with $\sigma = 0.3$. Increasing numbers of samples (1, 2, 3, 5, 10, 20, and 50) are plotted with increasingly darker tones of blue. The theoretical asymptotic for infinitely many samples, $e^{-\frac{1}{4\sigma^2}x^2}$, is shown in black. The half width of the central peak in the autocorrelations, indicated by the red bar, is nearly independent of the number of samples. (b) Standard deviations of the point spread function estimated from the half width of the autocorrelation for 30 realizations with 1, 2, 3, 5, 10, 20, and 50 samples.

The expected correlation is also Gaussian, with standard deviation equal to $\sqrt{2}\sigma$. It follows that the width of the point spread function can be inferred by autocorrelating the action of the point spread function on a sufficiently large set of random functions. In practice, we approximate the theoretical expectation \mathbb{E} by the empirical expectation, i.e., the sample average

$$\mathbb{E}_n = \frac{1}{n} [c_1(x) + c_2(x) + \dots + c_n(x)], \quad (11)$$

where each c_i is obtained from the application of $H(x; y)$ to a realization (sample) $v_i(x)$ of the random function $v(x)$. Figure 2a shows averaged autocorrelations of $h(y)$ for the cases of 1, 2, 3, 5, 10, 20, and 50 samples of $v(x)$. For $v(x)$ we choose a pointwise normal distribution, which leads to nearly optimal convergence, as demonstrated in section 7. With increasing number of samples, the autocorrelations approach $e^{-\frac{1}{4\sigma^2}x^2}$, as predicted by equation (10). While numerous sidelobes still appear for large numbers of samples, the central peak of the Gaussian is well recovered when few samples are used. This empirical observation can be underpinned theoretically (see section 7.2). The rapid convergence around $x = 0$ suggests that a small number of samples may suffice to obtain useful estimates of the asymptotic width of the Gaussian and thus of the width of the point spread function. This conjecture is confirmed by Figure 2b, which shows the width of the point spread function estimated from the width of the autocorrelation function at half its maximum height. When only one sample, i.e., one random function $v(x)$, is used, 60% of the estimates for σ range within $\pm 15\%$ of its true value.

3.2. Spatially Variable Resolution Length

As a result of uneven coverage, resolution in realistic applications is spatially variable. Assuming that the point spread function takes the form of the modified Gaussian with spatial dependence of σ

$$H(x; y) = \frac{1}{\sqrt{2\pi\sigma(x)\sigma(y)}} e^{-\frac{(x-y)^2}{2\sigma(x)\sigma(y)}}, \quad (12)$$

we can attempt to recover the space-dependent variance $\sigma(x)$ using an adapted version of the algorithm proposed in section 3.1: First, we apply $v(x)$ to $H(x; y)$, to form the smoothed version $h(y)$ of $v(x)$, as shown in equation (8) and Figures 3a and 3b. Second, we autocorrelate $h(y)$ over an interval $[x - \Delta x, x + \Delta x]$ that is larger than the expected width of the point spread function, but small enough to capture its spatial variability:

$$c(\bar{x}, x) = \int_{x-\Delta x}^{x+\Delta x} h(\bar{x} + y)h(y) dy. \quad (13)$$

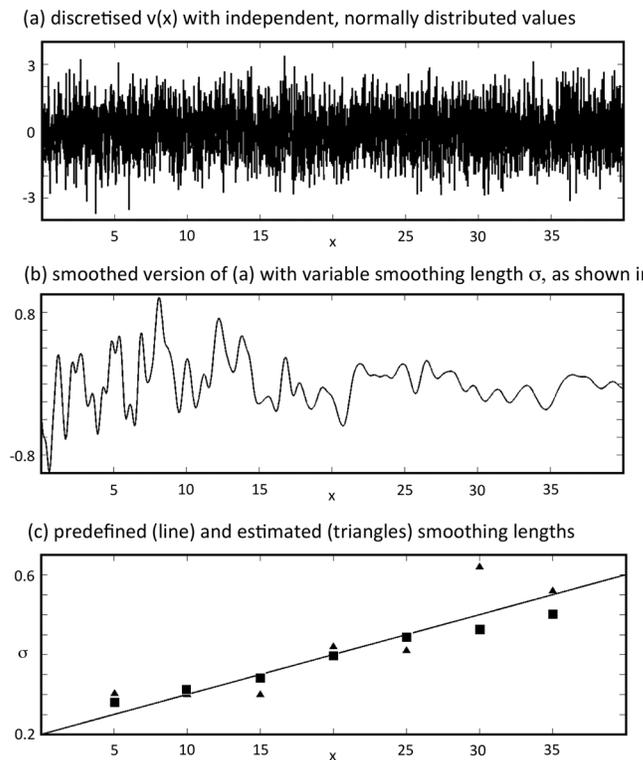


Figure 3. Action of the point spread function with space variable smoothing length on a random function. (a) Discretised random function $v(x)$ where the individual $v_i = v(x_i)$ is independent and normally distributed. (b) Smoothed version of $v(x)$ with linearly increasing smoothing length $\sigma(x)$, as shown below. (c) The predefined smoothing length $\sigma(x)$ is shown as solid line. Estimated smoothing lengths are shown as triangles (one sample: black, triangles, five samples: squares).

To approach more realistic scenarios, we continue with synthetic full-waveform inversions where point spread functions may deviate from an idealized Gaussian. This will allow us to refine the concepts introduced in section 3 while being able to compare the derived resolution proxies with the quality of the reconstructed target model.

4.1. Synthetic Inversion as Reference

For simplicity, we consider the 2-D scalar wave equation

$$\rho \partial_t^2 u - \partial_{x_1} (\mu \partial_{x_1} u) - \partial_{x_2} (\mu \partial_{x_2} u) = f, \quad (14)$$

within a domain $G \subset \mathbb{R}^2$. In equation (14), u , ρ , μ , and f denote the displacement field, density, shear modulus, and an external force density. We solve equation (14) numerically using a time-domain fourth-order finite difference scheme [Virieux, 1984] with absorbing boundaries [Cerjan et al., 1985].

As illustrated in Figure 4, the computational domain is split into two parts, with average velocities approximating the phase velocity of Rayleigh waves at 10 s period in oceanic ($\beta = \sqrt{\mu/\rho} = 3.5$ km/s) and continental ($\beta = 3.0$ km/s) regions, respectively. The source-receiver geometry, also shown in Figure 4, is intended to mimic realistic setups with clusters of events and with receivers deployed irregularly along lines or within structured arrays. In the synthetic inversions we measure L_2 waveform misfit

$$\chi = \sum_{\text{receivers}} \int [u - u_{\text{data}}]^2 dt + \mathbf{m}^T \mathbf{C}_m^{-1} \mathbf{m}, \quad (15)$$

acknowledging that misfit functionals with higher robustness or emphasis on phase differences are more useful in real-data applications [e.g., Luo and Schuster, 1991; Gee and Jordan, 1992; Fichtner et al., 2008; Brossier et al., 2009; van Leeuwen and Mulder, 2010; Rickers et al., 2012]. The model covariance operator \mathbf{C}_m in

Averaging over realizations of $v(x)$ provides correlations as a function of the central position x of the correlation interval $[x - \Delta x, x + \Delta x]$. We may thus obtain an estimate for the width of the point spread function at position x by measuring the width of the position-dependent correlation function $c(\bar{x}, x)$.

An example is shown in Figure 3c where we tried to recover a linearly increasing σ using one and five realizations of $v(x)$. With few exceptions, the estimates fall within 10% of the target value in both cases.

3.3. Preliminary Conclusions

At this point, we can draw some preliminary conclusions that will guide the developments of the forthcoming sections: (i) The Hessian acts as a smoother of random functions. The smoothed functions carry information on resolution that can be extracted through different types of correlations. (ii) While the correlations themselves may require large sample sizes to converge, useful resolution proxies may already be obtained with very few, and possibly just one, sample.

4. Theory I: Single-Parameter Inversion

In section 3 we used point spread functions with a predefined functional shape.

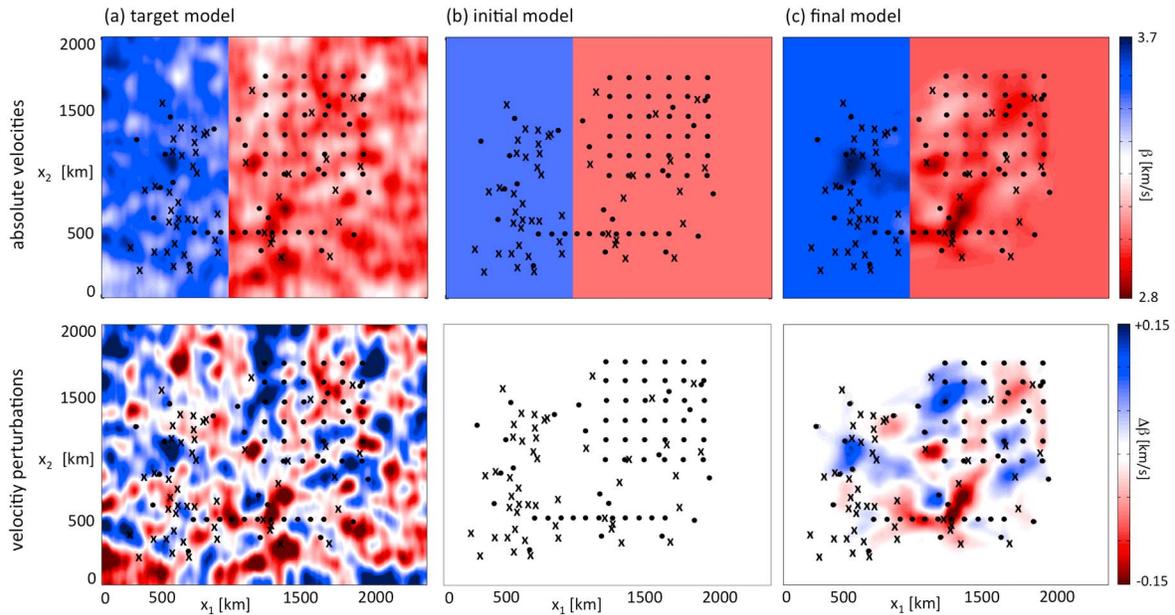


Figure 4. Synthetic single-parameter inversion to recover S velocity β . (top row) Absolute velocities and (bottom row) velocity perturbations $\Delta\beta$. (a) Input v_S structure used to compute artificial data. Sources are marked by crosses; receivers by dots. (b) Initial model. (c) Final model after 10 conjugate-gradient iterations.

equation (15) imposes a smoothing of the model over length scales of the minimum wavelength in order to suppress unwanted artifacts. This smoothing is automatically included in the Hessian through the definition of the misfit, and it acts to increase resolution length. The optimization is based on a conjugate-gradient iteration, where gradients are computed via the adjoint method [e.g., Lailly, 1983; Tarantola, 1988; Tromp et al., 2005; Fichtner et al., 2006].

To initially avoid the complication of interparameter trade-offs, we start with a single-parameter problem where we invert for the target distribution of the S velocity β , shown in Figure 4a. Taking a model without lateral variations in the “oceanic” and “continental” parts as initial model (Figure 4b), we obtain our final model (Figure 4c) after 10 conjugate-gradient iterations that reduced the initial misfit by 92%.

4.2. Point Spread Functions

Using a comparison of target model and final model as a measure of resolution is tempting but potentially misleading. Features that seem well reconstructed may be accidental artifacts or the result of an optimization scheme tuned toward a specific target model. Point spread functions offer an objective measure of resolution that is independent of the inversion scheme and unavoidably subjective regularization.

Figure 5 shows a collection of point spread functions $H(\mathbf{x}; \mathbf{y})$ for point-localized shear velocity heterogeneities $\delta\beta$ at positions $\mathbf{x} = (x_1, x_2)$. The height, width, and lateral smearing of the point spread functions vary widely. While clearly peaked point spread functions can be found in the center of the domain, point spread functions outside the source-receiver configuration are practically zero, indicating poor resolution. The computation of point spread functions for all grid points—and thus the recovery of the full Hessian—would be ideal, but computationally too expensive, especially in three dimensions.

A resolution proxy that can be computed with only one application of the Hessian to a test model is the volume of the point spread functions $V(\mathbf{x})$ as a function of position \mathbf{x} . This volume is a proxy for the impact of a point-localized heterogeneity on the final model. Choosing a constant shear velocity perturbation $\delta\beta = 1$, we find

$$V(\mathbf{x}) = \int H(\mathbf{x}; \mathbf{y}) \delta\beta \, d\mathbf{y} = \int H(\mathbf{x}; \mathbf{y}) \, d\mathbf{y}, \quad (16)$$

which is the volume of the point spread function for a point-localized shear velocity perturbation at position \mathbf{x} . The volumes of the point spread functions for the previously described synthetic inversion are displayed

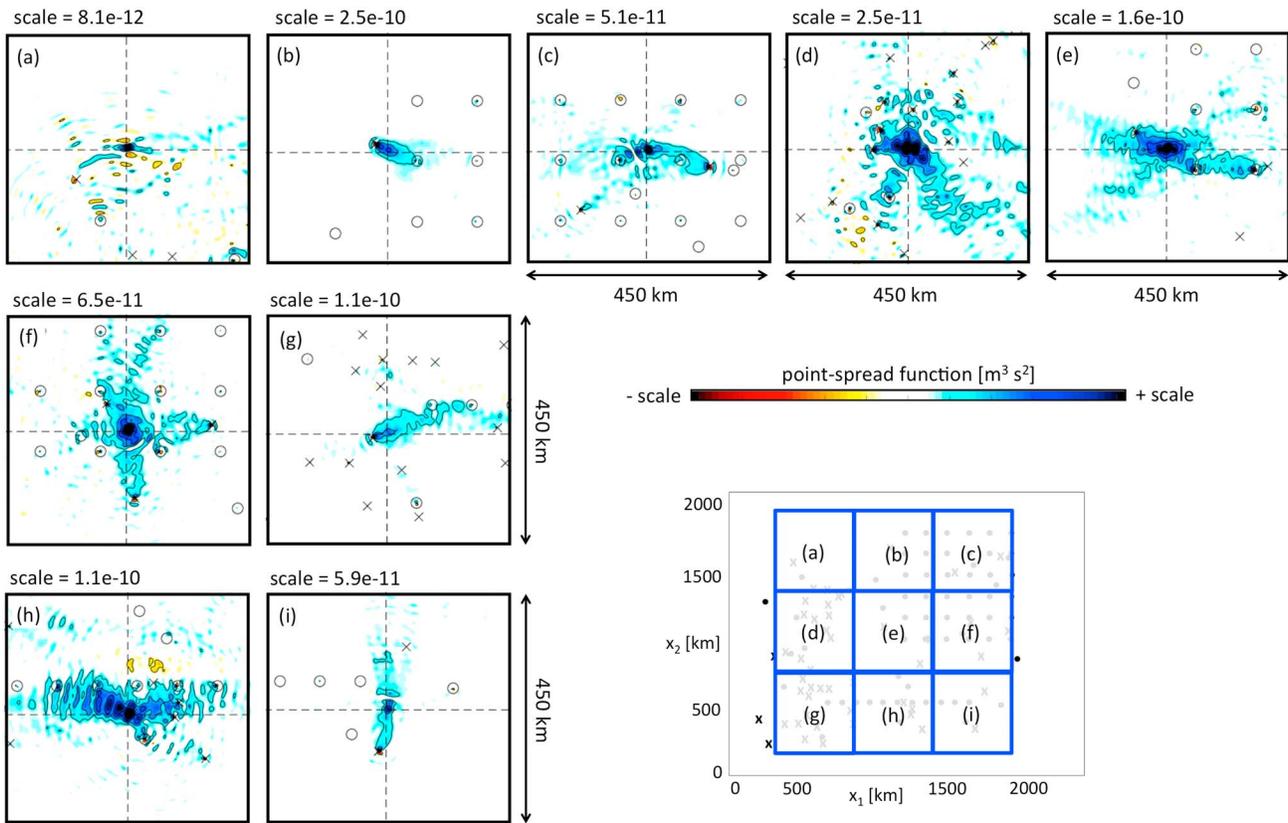


Figure 5. (a–i) Collection of point spread functions for different locations in the model. Dashed cross hairs indicate sample location of point spread function (PSF). The position of the individual point spread function figure is indicated in the bottom right panel. Note the different scales, ranging from $8.1 \cdot 10^{-12}$ in the top left corner of the model (Figure 5a) to $2.5 \cdot 10^{-10}$ in Figure 5b.

in Figure 6. Heterogeneities in regions with small point spread function volumes have small effects on the observed misfit, regardless of their shape and lateral extent, meaning that resolution is poor. In regions where point spread function volumes are large, heterogeneities can have significant impact on the misfit. However, in the case of very wide point spread functions, these heterogeneities must have either a large amplitude or a large lateral extent. The latter can be estimated from the resolution length, considered in the following paragraph.

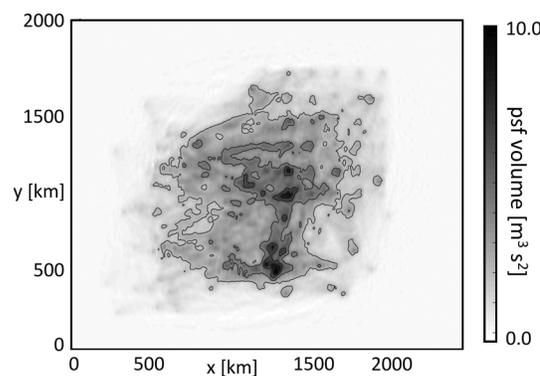


Figure 6. Volume of point spread functions as a function of position \mathbf{x} in the computational domain. The volumes are computed by the application of the Hessian $H(\mathbf{x}; \mathbf{y})$ to a constant shear velocity perturbation $\delta\beta = 1$, that is, $\int H(\mathbf{x}; \mathbf{y}) \delta\beta \, d\mathbf{y} = \int H(\mathbf{x}; \mathbf{y}) \, d\mathbf{y}$.

4.3. Direction-Dependent Resolution Length

To compute resolution length as a function of position, we generalize the ideas introduced in section 3.2. Denoting the action of the Hessian $H(\mathbf{x}; \mathbf{y})$ on a random test function $v(\mathbf{y})$ by

$$h(\mathbf{x}) = \int H(\mathbf{x}; \mathbf{y})v(\mathbf{y}) \, d\mathbf{y}, \quad (17)$$

we autocorrelate $h(\mathbf{x})$ in the x_1 direction over an interval $[\bar{x}_1 - \Delta x_1, \bar{x}_1 + \Delta x_1]$ that is larger than the expected width of the point spread function but small enough to capture its spatial variability:

$$c(\bar{x}_1, \mathbf{x}) = \int_{x_1 - \Delta x_1}^{x_1 + \Delta x_1} h(\bar{x}_1 + x'_1, x_2) h(x'_1, x_2) \, dx'_1. \quad (18)$$

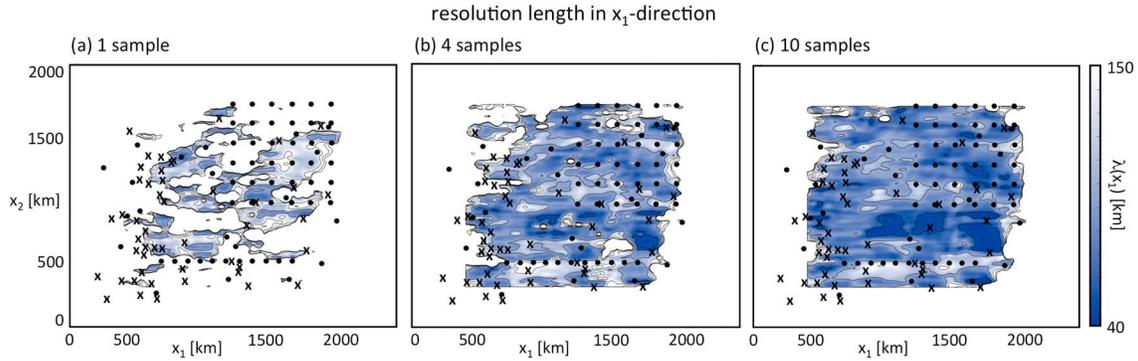


Figure 7. Resolution length in x_1 direction computed with (a) 1, (b) 4, and (c) 10 random samples.

The quantity $c(\bar{x}_1, \mathbf{x})$ is a local correlation in x_1 direction at position $\mathbf{x} = (x_1, x_2)$ with offset variable \bar{x}_1 . Introducing the definition of $h(\mathbf{x})$ from equation (17) into equation (18), we obtain

$$c(\bar{x}_1, \mathbf{x}) = \int_{x_1 - \Delta x_1}^{x_1 + \Delta x_1} \int \int H(\bar{x}_1 + x'_1, x_2; \mathbf{y}) H(x'_1, x_2; \mathbf{y}') v(\mathbf{y}) v(\mathbf{y}') dx'_1 d\mathbf{y} d\mathbf{y}' . \quad (19)$$

Imposing that realizations of $v(\mathbf{y})$ are uncorrelated with zero mean and unit variance, i.e.,

$$\mathbb{E}[v(\mathbf{y})v(\mathbf{y}')] = \delta(\mathbf{y} - \mathbf{y}') , \quad (20)$$

the expected value of the local autocorrelation (19) condenses to

$$\mathbb{E}[c(\bar{x}_1, \mathbf{x})] = \int_{x_1 - \Delta x_1}^{x_1 + \Delta x_1} \int H(\bar{x}_1 + x'_1, x_2; \mathbf{y}) H(x'_1, x_2; \mathbf{y}) dx'_1 d\mathbf{y} . \quad (21)$$

Following the intuitive example from section 3.1, we define the width of $\mathbb{E}[c(\bar{x}_1, \mathbf{x})]$ divided by $\sqrt{2}$ as the resolution length at position \mathbf{x} in x_1 direction, denoted by $\lambda(x_1; \mathbf{x})$. To measure the width of the local autocorrelation $\mathbb{E}[c(\bar{x}_1, \mathbf{x})]$, we again determine the distance between the two points where it reaches half of its maximum value. Alternatives to this measurement are discussed in section 8.3.

Resolution length in x_1 direction for the synthetic inversion described in section 4.1 is shown in Figure 7. To approximate the expectation of the autocorrelation we use 1, 4, and 10 random samples. While the use of only one sample mostly leads to an overestimation of resolution length, the result stabilizes when four or more samples are used.

The comparison of Figures 5 and 7 demonstrates that the resolution lengths computed via equation (18) are indeed good proxies for the widths of the point spread functions. Figure 5h shows a point spread function that is smeared over approximately 150 km in x_1 direction, which is in good agreement with $\lambda(x_1) \approx 130$ km in Figure 7. In contrast, the point spread function in Figure 5i is almost perfectly peaked in x_1 direction, in agreement with $\lambda(x_1) \approx 50$ km for that position.

The definition of resolution length in x_1 direction based on equation (18) is easily adapted to other directions by changing the orientation of the autocorrelation integral. Correlating, for instance, in x_2 direction yields the resolution length in x_2 direction—which is displayed for our synthetic inversion in Figure 8. The difference between Figures 7 and 8 reveals a strong directional dependence of resolution. This situation is typical for tomographic problems where source–receiver lines often have a preferential orientation, that is, the azimuthal coverage is anisotropic [e.g., Yanovskaya, 1997; Fichtner and Trampert, 2011a]. A similar situation occurs in teleseismic body wave tomography where predominantly vertically oriented rays result in good horizontal coverage combined with vertical smearing.

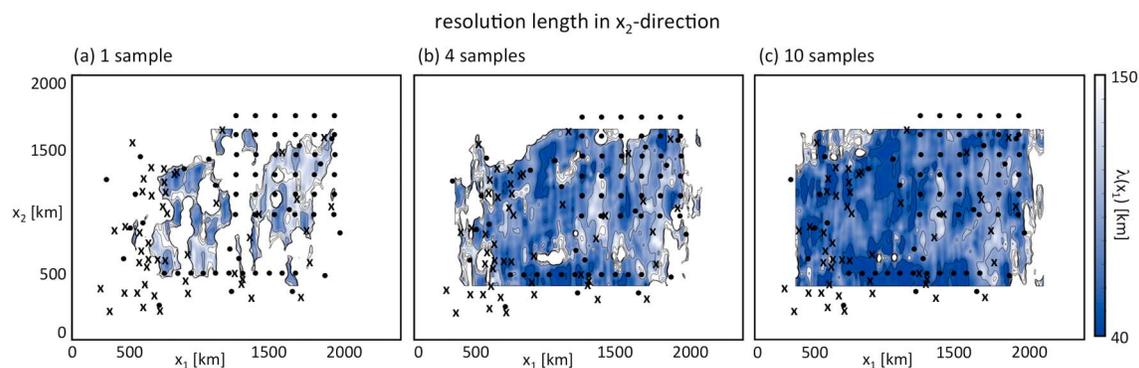


Figure 8. Resolution length in x_2 direction computed with (a) 1, (b) 4, and (c) 10 random samples.

5. Theory II: Multiparameter Inversion

5.1. Reference Synthetic Inversion

As an extension of the single-parameter inversion in the previous section, we consider a two-parameter inversion where we seek to constrain the two-dimensional distributions of shear velocity β and density ρ . The synthetic inversion experiment is summarized in Figure 9. The background and initial model for β is identical to the one used for the single-parameter inversion and shown in Figure 4. The initial and background model for ρ is homogeneous at 3000 kg/m^3 . The misfit functional is the same as the one used in section 4.1, the only difference being that the covariance operator now imposes smoothing of both model parameters, though without forcing correlations between them. To explore the extent to which v_s and ρ can be constrained independently, the target heterogeneities for both parameters are uncorrelated.

After 24 conjugate-gradient iterations, the misfit dropped by 2 orders of magnitude. While v_s heterogeneities are well recovered within the source–receiver array, density variations can hardly be constrained. This result is expected for various reasons: (i) The density model is smooth, lacking sharp impedance contrast that may be reconstructed from reflected waves [e.g., *Igel et al.*, 1996; *Prieux et al.*, 2013]. (ii) The scattering of body waves off density heterogeneities is predominantly backward, meaning that waveform effects appear only very late in the seismograms [*Wu and Aki*, 1985; *Tarantola*, 1986]. (iii) The inversion suffers from the absence of surface waves that have significant sensitivity to short-wavelength density variations [*Takeuchi and Saito*, 1972;

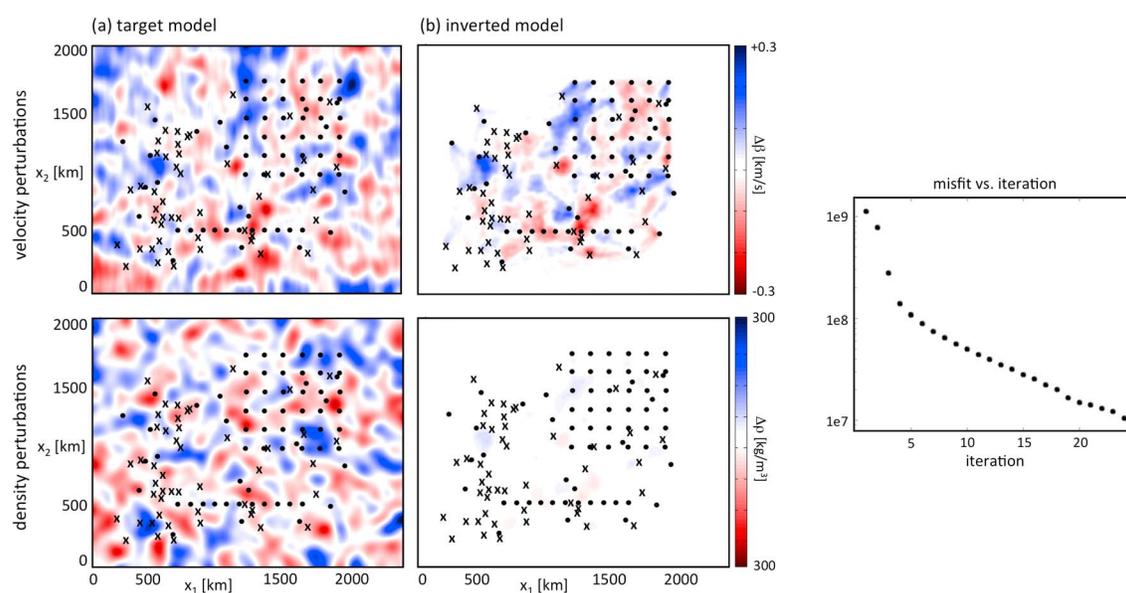


Figure 9. Synthetic two-parameter inversion for shear velocity (v_s) and density (ρ). (a) The target models for v_s and ρ are uncorrelated in order to reveal the extent to which both parameters can be constrained independently. (b) The inverted models after 24 conjugate-gradient iterations during which the misfit was reduced by 2 orders of magnitude. The misfit evolution is shown to the right.

Tanimoto, 1991]. Furthermore, we observe that the inverted density model seems to correlate more with the target β model than with the target ρ model.

5.2. The Multiparameter Hessian

Spatial resolution and trade-offs in the multiparameter scenario are described by a Hessian operator matrix $\mathbf{H}(\mathbf{x}; \mathbf{y})$ which, in the case of our numerical example, takes the form

$$\mathbf{H}(\mathbf{x}; \mathbf{y}) = \begin{pmatrix} H_{\beta\beta}(\mathbf{x}; \mathbf{y}) & H_{\beta\rho}(\mathbf{x}; \mathbf{y}) \\ H_{\rho\beta}(\mathbf{x}; \mathbf{y}) & H_{\rho\rho}(\mathbf{x}; \mathbf{y}) \end{pmatrix} \quad (22)$$

The diagonal elements $H_{\beta\beta}(\mathbf{x}; \mathbf{y})$ and $H_{\rho\rho}(\mathbf{x}; \mathbf{y})$ describe the spatial resolution of their respective parameters, that is, S velocity β and density ρ . As in the single-parameter case, $H_{\beta\beta}(\mathbf{x}; \mathbf{y})$ can be interpreted as the \mathbf{x} -dependent point spread function corresponding to an S velocity perturbation point localized at position \mathbf{y} . A similar interpretation holds for $H_{\rho\rho}(\mathbf{x}; \mathbf{y})$. Interparameter trade-offs are encoded in the off-diagonal element $H_{\beta\rho}(\mathbf{x}; \mathbf{y})$. A density perturbation point localized at position \mathbf{y} will map into a blurred S velocity heterogeneity with the shape of $H_{\beta\rho}(\mathbf{x}; \mathbf{y})$; and vice versa.

5.3. Probing the Multiparameter Hessian

There are two different aspects of interparameter trade-offs: (i) The volume of the interparameter point spread function $H_{\beta\rho}(\mathbf{x}; \mathbf{y})$ describes the strength of a heterogeneity that maps from β to ρ or vice versa. (ii) The direction-dependent width of $H_{\beta\rho}(\mathbf{x}; \mathbf{y})$ captures the extent to which a point-localized heterogeneity in one parameter smears the image of the other parameter. In the following, we will consider both aspects separately.

5.3.1. The Strength of Interparameter Mapping

To estimate the strength of interparameter mappings, we start with a random test vector

$$\mathbf{v}(\mathbf{y}) = \begin{pmatrix} v_{\beta}(\mathbf{y}) \\ v_{\rho}(\mathbf{y}) \end{pmatrix}. \quad (23)$$

The application of \mathbf{H} to \mathbf{v} yields the vector

$$\begin{pmatrix} h_{\beta}(\mathbf{x}) \\ h_{\rho}(\mathbf{x}) \end{pmatrix} = \int \mathbf{H}(\mathbf{x}; \mathbf{y}) \mathbf{v}(\mathbf{y}) \, d\mathbf{y}. \quad (24)$$

For the expectation of h_{β} we find

$$\mathbb{E}[h_{\beta}(\mathbf{x})] = \int H_{\beta\beta}(\mathbf{x}; \mathbf{y}) \mathbb{E}[v_{\beta}(\mathbf{y})] \, d\mathbf{y} + \int H_{\beta\rho}(\mathbf{x}; \mathbf{y}) \mathbb{E}[v_{\rho}(\mathbf{x})] \, d\mathbf{y}. \quad (25)$$

Choosing the ensemble of random test vectors such that $\mathbf{E}[v_{\beta}(\mathbf{y})] = 0$ and $\mathbf{E}[v_{\rho}(\mathbf{x})] = 1$, yields

$$\mathbb{E}[h_{\beta}(\mathbf{x})] = V_{\beta\rho} = \int H_{\beta\rho}(\mathbf{x}; \mathbf{y}) \, d\mathbf{y}, \quad (26)$$

which is the volume of the interparameter point spread function that describes the strength of apparent heterogeneities in ρ that result from trade-offs with β structure, and vice versa. An expected value of 1 for v_{ρ} can be achieved most easily by choosing all ensemble members to be constant over the whole domain. This is similar to the computation of the point spread function volume in the single-parameter case described in equation (16) of section 4.2.

Going back to equation (25), we may alternatively choose test vectors with $\mathbb{E}[v_{\beta}(\mathbf{y})] = 1$ and $\mathbb{E}[v_{\rho}(\mathbf{x})] = 0$, which causes $\mathbb{E}[h_{\beta}(\mathbf{x})]$ to equal the point spread function volume for S velocity:

$$\mathbb{E}[h_{\beta}(\mathbf{x})] = V_{\beta\beta} = \int H_{\beta\beta}(\mathbf{x}; \mathbf{y}) \, d\mathbf{y}. \quad (27)$$

Similar arguments can be made for the expectation of h_{ρ} . It follows that by choosing different expectations of the test vectors, we can obtain the position-dependent point spread function volumes either for specific parameters themselves ($V_{\beta\beta}$, $V_{\rho\rho}$) or for the interparameter trade-offs ($V_{\beta\rho}$).

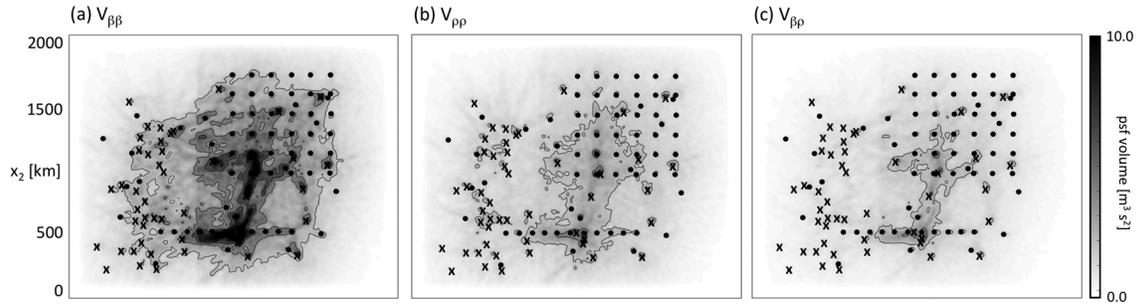


Figure 10. Point spread function volumes (a) $V_{\beta\beta}$, (b) $V_{\rho\rho}$, and (c) $V_{\beta\rho}$ computed with 10 test vectors. The volume of point spread functions for density, $V_{\rho\rho}$, is small compared to the one for shear velocity, $V_{\beta\beta}$, indicating poor recovery of density heterogeneities. The comparable amplitudes of $V_{\rho\rho}$ and $V_{\beta\rho}$ suggest that the artifacts introduced into the density image by S velocity heterogeneities are roughly as large as the recovered density heterogeneities themselves.

The point spread function volumes $V_{\beta\beta}$, $V_{\rho\rho}$, and $V_{\beta\rho}$ for the synthetic inversion described in section 5.1 are presented in Figure 10. An ensemble of 10 test vectors was used for their computation. Point spread function volumes for S velocity, $V_{\beta\beta}$ are similar to those for the single-parameter case shown in Figure 6, suggesting that the recovery β does not suffer much from the introduction of a second physical parameter. However, $V_{\beta\beta}$ varies by a factor of around 5 throughout the source-receiver array, indicating that the quality of the S velocity reconstruction is strongly position dependent. For density, point spread function volumes $V_{\rho\rho}$ are generally more than a factor 3 smaller than $V_{\beta\beta}$, in accord with the result of the synthetic inversion where density is poorly recovered. Furthermore, the nearly equal amplitudes of $V_{\rho\rho}$ and $V_{\beta\rho}$ suggest that the mapping from β to ρ is approximately as significant as the recovery of ρ itself.

5.3.2. Interparameter Smearing

To extract information on interparameter smearing by random probing of the Hessian, we consider two different random test vectors

$$\mathbf{v}(\mathbf{y}) = \begin{pmatrix} v_\beta(\mathbf{y}) \\ v_\rho(\mathbf{y}) \end{pmatrix}, \quad \mathbf{v}'(\mathbf{y}) = \begin{pmatrix} v'_\beta(\mathbf{y}) \\ v'_\rho(\mathbf{y}) \end{pmatrix} \quad (28)$$

where v_β , v_ρ , and v'_ρ are random perturbations in β and ρ . Note that the two vectors have identical S velocity perturbations but different density perturbations. The application of \mathbf{H} to \mathbf{v} and \mathbf{v}' yields the vectors $(h_\beta, h_\rho)^T$ and $(h'_\beta, h'_\rho)^T$, respectively. First, we correlate h_β and h'_β :

$$\begin{aligned} c_{\beta\beta}(\mathbf{x}) &= \int h_\beta(x_1 + x'_1, x_2) h'_\beta(x'_1, x_2) dx'_1 \\ &= \int \left[\int H_{\beta\beta}(x_1 + x'_1, x_2; \mathbf{y}) v_\beta(\mathbf{y}) d\mathbf{y} + \int H_{\beta\rho}(x_1 + x'_1, x_2; \mathbf{y}) v_\rho(\mathbf{y}) d\mathbf{y} \right] \\ &\quad \left[\int H_{\beta\beta}(x'_1, x_2; \mathbf{y}') v_\beta(\mathbf{y}') d\mathbf{y}' + \int H_{\beta\rho}(x'_1, x_2; \mathbf{y}') v'_\rho(\mathbf{y}') d\mathbf{y}' \right] dx'_1 \end{aligned} \quad (29)$$

In the interest of a lighter notation, we omitted in equation (29) the option to compute local correlations that would capture the possible spatial variability of the point spread function. It may, however, be introduced with ease. Before computing the expectation of $c_{\beta\beta}$, we note the expectations of the four different products between the random test functions v_β , v_ρ , and v'_ρ :

$$\mathbb{E}[v_\beta(\mathbf{y})v_\beta(\mathbf{y}')] = \delta(\mathbf{y} - \mathbf{y}'), \mathbb{E}[v_\beta(\mathbf{y})v_\rho(\mathbf{y}')] = 0, \mathbb{E}[v_\beta(\mathbf{y})v'_\rho(\mathbf{y}')] = 0, \mathbb{E}[v_\rho(\mathbf{y})v'_\rho(\mathbf{y}')] = 0. \quad (30)$$

With the help of relations (30), the expectation of the correlation $c_{\beta\beta}$ takes the simple form

$$\mathbb{E}[c_{\beta\beta}(\mathbf{x})] = \iint H_{\beta\beta}(x_1 + x'_1, x_2; \mathbf{y}) H_{\beta\beta}(x'_1, x_2; \mathbf{y}) dx'_1 d\mathbf{y}. \quad (31)$$

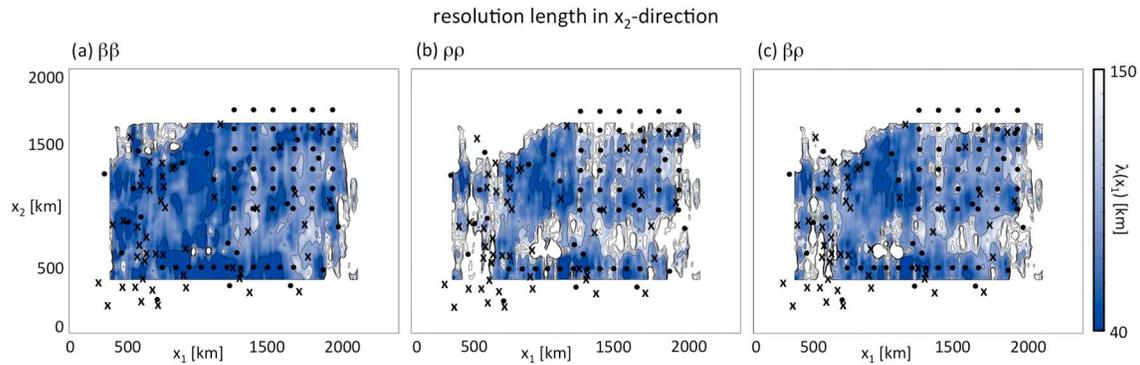


Figure 11. Resolution lengths in x_2 direction for (a) S velocity, (b) density, and (c) the interparameter mapping. An ensemble of 10 test vectors was used. Resolution for S velocity is comparable to the single-parameter case shown in Figure 8c. For density, resolution lengths are significantly larger than those for S velocity. Furthermore, the smearing of the interparameter mapping in x_2 direction is similar to the smearing of density heterogeneities.

Equation (31) is the analogue of equation (21) for the single-parameter case, from which the width of the point spread function for S velocity can be inferred. Again making use of relations (30), we can compute the expected correlation of h_ρ and h'_ρ , denoted by $\mathbb{E}[c_{\rho\rho}(\mathbf{x})]$:

$$\mathbb{E}[c_{\rho\rho}(\mathbf{x})] = \iint H_{\beta\rho}(x_1 + x'_1, x_2; \mathbf{y}) H_{\beta\rho}(x'_1, x_2; \mathbf{y}) dx'_1 dy. \quad (32)$$

It follows by analogy that $\mathbb{E}[c_{\rho\rho}(\mathbf{x})]$ allows us to estimate the direction- and position-dependent width of the perceived density heterogeneity induced by a point-localized S velocity perturbation. This is, by virtue of the symmetry of \mathbf{H} , identical to the apparent S velocity heterogeneity caused by a point-localized density perturbation.

Figure 11 displays resolution lengths in x_2 direction for our synthetic inversion, computed from 10 test vectors. The resolution length for S velocity is similar to the one for the single-parameter case shown in Figure 8, thus indicating that S velocity resolution does not suffer significantly from the introduction of density as a second model parameter class. For density, the resolution length in x_2 direction is roughly twice as large as for S velocity, suggesting more pronounced smearing. However, this result must be interpreted with caution because the amplitude recovery for density, encoded in the point spread function volume $V_{\rho\rho}$, is comparatively poor and to some extent overwhelmed by trade-offs with S velocity structure (see Figures 10b and 10c). The smearing of the interparameter mapping between β and ρ , shown in Figure 11c, follows almost the same pattern as the smearing of the individual parameters.

6. Realistic Example in 3-D

We illustrate the proposed methods for resolution analysis with a real-data full-waveform inversion for the western Mediterranean. For this inversion, we used 13,089 three-component recordings from 52 regional earthquakes. For broadband stations belonging to permanent national/regional networks we obtained waveform data from global (IRIS, www.iris.edu) and European (ORFEUS, www.orfeus-eu.org) data centers. This was complemented by data from the temporary IberArray [Díaz *et al.*, 2009], the PASSCAL experiment PICASSO (90 stations in southern Spain and northern Morocco), and smaller arrays deployed by the University of Münster, Germany (15 stations) and Bristol, UK (6 stations). The source-receiver configuration is shown in Figure 12.

As initial model we used a full-waveform tomography of Eurasia [Fichtner *et al.*, 2013] that already contains the large-scale features of the western Mediterranean. To solve the forward problem of viscoelastic wave propagation we employ a GPU-accelerated spectral-element code [Fichtner *et al.*, 2009b]. Following a semiautomatic selection of measurement time windows [Krischer *et al.*, 2015], we quantify time- and frequency-dependent phase differences between observed and synthetic seismograms [Fichtner *et al.*, 2008]. In order to suppress artifacts that we expect to be unresolved, we add a regularization term to the time-frequency phase misfit. The regularization term is the same as in equation (15), and it effectively acts as a smoother over distances of the minimum wavelength. Including the regularization term into the misfit definition automatically guarantees that it is consistently incorporated into the Hessian and the estimates of resolution length. Being

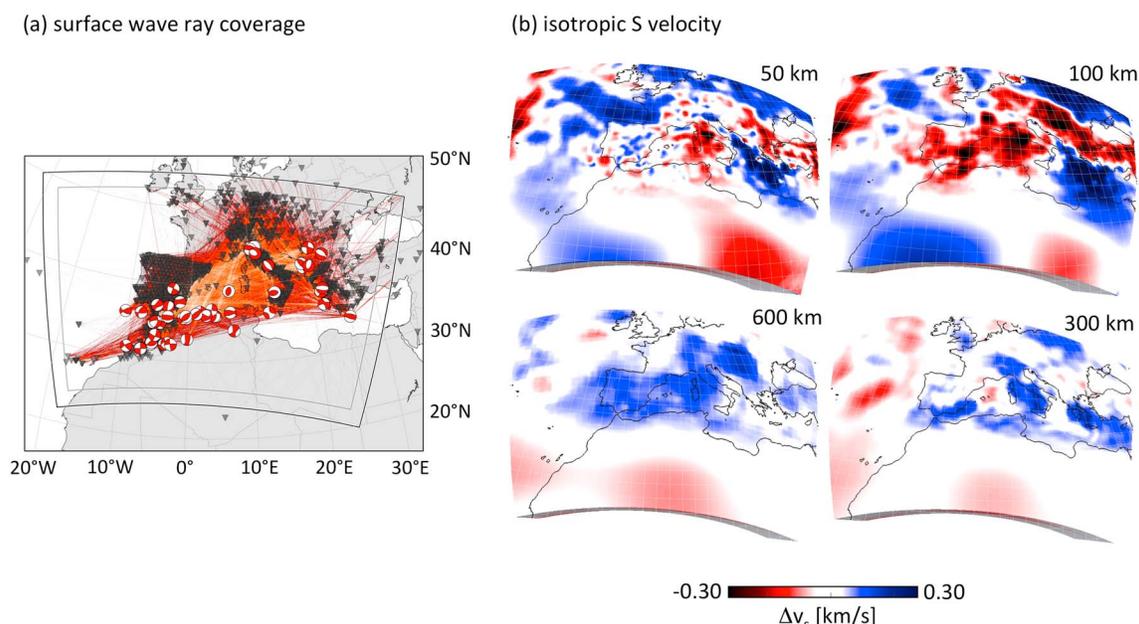


Figure 12. Full-waveform inversion for the western Mediterranean. (a) Source-receiver distribution in the western Mediterranean region. Beach balls indicate the source mechanisms and locations of the 52 earthquakes used in our inversion. Great circles connecting sources and receivers are plotted in lighter color when coverage is high. The black and gray boxes mark the boundary of the computational domain and the inner boundary of the absorbing layer, respectively. Recordings are only chosen from within the inner gray box to avoid contamination by artificial reflections from imperfect absorbing boundaries. (b) Absolute variations of the isotropic S velocity, v_s , relative to the lateral average in the region.

insensitive to an amplitude scaling, the phase differences avoid trade-offs with the event magnitudes and the station calibration that may not be well constrained. These differences are then minimized iteratively using a conjugate-gradient method that is based on the computation of Fréchet kernels with the help of adjoint techniques. Figure 12 shows horizontal slices through the final distribution of the isotropic S velocity v_s , defined as $v_s^2 = \frac{2}{3}v_{sv}^2 + \frac{1}{3}v_{sh}^2$ [e.g., Babuška and Cara, 1991; Panning and Romanowicz, 2006].

Using the previously described methods, we estimate 3-D distributions of resolution lengths using a total of five samples. Resolution length in N-S, E-W, and radial directions are summarized in Figure 13. As a result of the heterogeneous coverage, the distributions of resolution lengths exhibit strong spatial variability, as well as directional dependence. This observation is not specific to full-waveform inversion and also typically found in ray tomography [e.g., Yanovskaya, 1997, 2000]. In the horizontal directions, resolution length around 50 km depth varies between 30 and nearly 400 km. Local minima of horizontal resolution lengths appear beneath regions of particularly dense coverage, such as the Aegean and the Iberian peninsula. However, horizontal resolution is poor beneath the south-central Mediterranean where few surface ray paths cross.

Resolution length in radial direction is mostly shorter than that in the horizontal directions because of fundamental- and higher-mode surface wave data that are naturally incorporated in the full-waveform inversion. Regions with poor horizontal resolution may in fact be well resolved in radial direction when few surface wave paths spread heterogeneities over wide areas while still constraining them to roughly the right depth range.

With increasing depth, the resolution lengths are less dominated by surface waves and more a function of body wave coverage. Consequently, the resolved volume shrinks with increasing depth. Despite the smaller volume, horizontal resolution lengths of ~ 50 km are possible to depths of around 600 km in regions where sufficiently many wave paths cross. As a result of predominantly E-W oriented source-receiver paths (Figure 12a), resolution in N-S direction is mostly better than in E-W direction.

7. Convergence and the Choice of Sampling Distributions

In the following paragraphs we provide theoretical considerations on the choice of the random sampling distribution and the convergence of the autocorrelation functions. Key results of this paragraph are as follows: (i) Normally distributed random functions ensure nearly optimal convergence of the empirical covariance

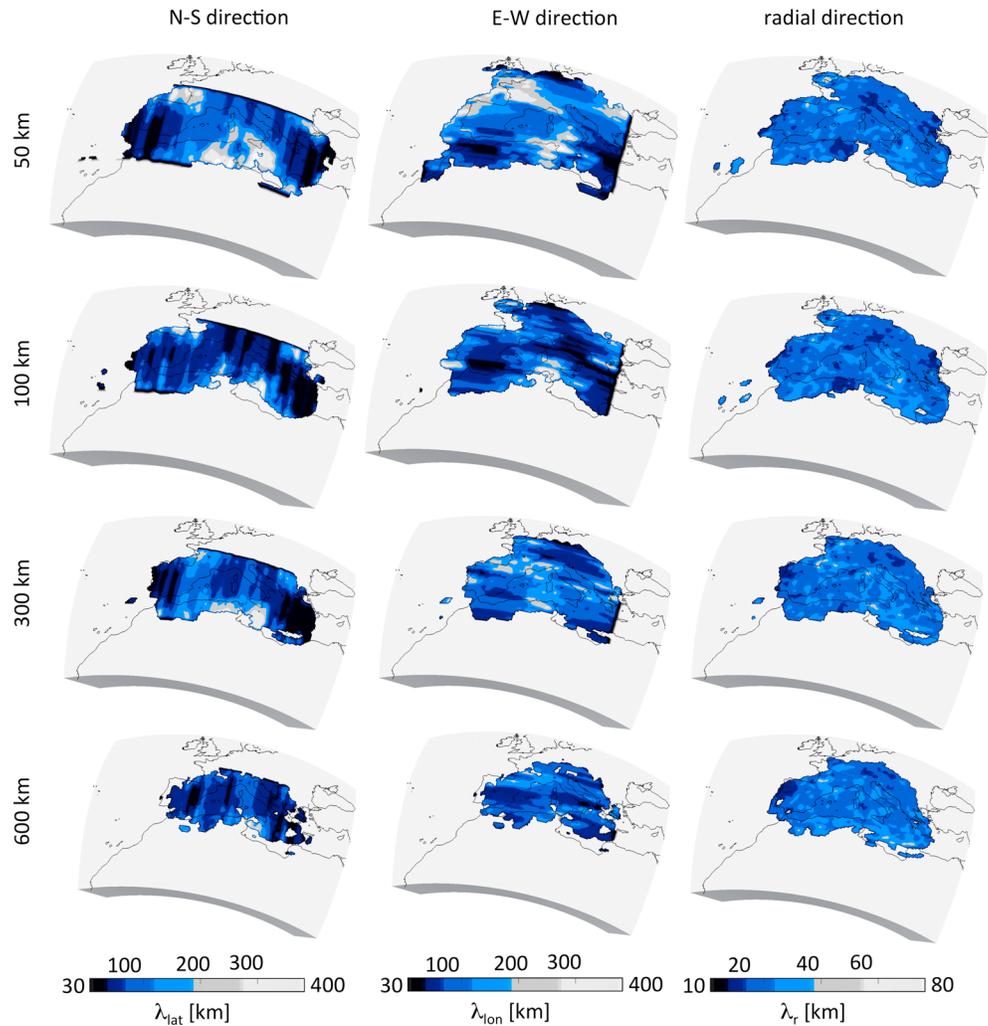


Figure 13. Slices through the 3-D distributions of resolution lengths. Resolution length in (left column) N-S direction, (middle column) E-W direction, and (right column) radial direction. Regions where the volume of the local point spread functions is below 1% of the maximum point spread function volume are blanked.

$\mathbb{E}[v(x)v(x')]$ toward $\delta(x - x')$. (ii) The autocorrelation $c(x)$ converges most rapidly near $x = 0$, meaning that its variance can be estimated from values around $x = 0$ based on a small number of random samples. (iii) Autocorrelating over a larger interval improves convergence and can compensate for a small number of samples.

7.1. Optimal Properties of the Sampling Distribution

To ensure a rapid convergence of the resolution proxies, the empirical expectation $\mathbb{E}_n[v(x)v(x')]$ should converge quickly toward $\delta(x - x')$, as noted in equation (6). To infer properties of sampling distributions that are suitable in this sense, we consider two fixed points x and $x' \neq x$. In the interest of a light notation, we define

$$v(x)v(x') = v v' , \tag{33}$$

where v and v' are two identically distributed and independent random variables with mean 0. Our goal is to estimate the convergence of the empirical expectation

$$\mathbb{E}_n = \frac{1}{n}(v_1 v'_1 + v_2 v'_2 + \dots + v_n v'_n), \tag{34}$$

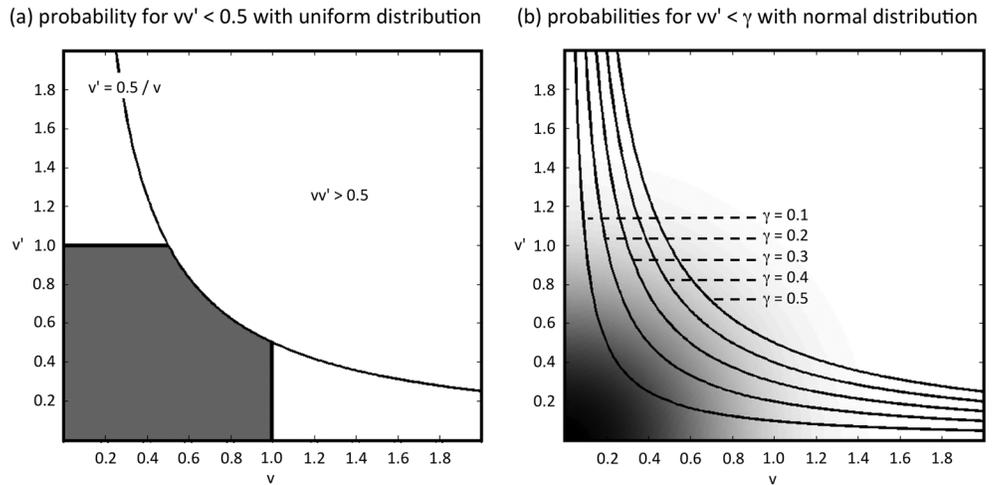


Figure 14. Probabilities for vv' being smaller than some value γ . (a) When v and v' are uniformly distributed between 0 and 1, the probability of $vv' < 0.5$ is equal to the gray shaded area, which is bounded by the coordinate axes, $v = 1$, $v' = 1$, and the curve $v' = 0.5/v$. (b) For an arbitrary distribution of v and v' the probability of $vv' < \gamma$ is equal to the integral over the distribution bounded by the curve $v' = \gamma/v$. Shown are, for illustration, a Gaussian distribution and the curves $v' = \gamma/v$ with $\gamma = 0.1, 0.2, 0.3, 0.4, 0.5$.

toward its theoretical limit 0 for increasing sample size n . To keep the development instructive, we temporarily assume that v is uniformly distributed between 0 and 1. As shown in Figure 14a, the probability that the product vv' is smaller than some value γ is then given by

$$\mathbb{P}(vv' < \gamma) = \int_{v=0}^1 \int_{v'=0}^{\min(1, \gamma/v)} dv' dv. \tag{35}$$

When v has an arbitrary probability density $p(v)$, equation (35) is generalized to

$$\mathbb{P}(|vv'| < \gamma) = 4 \int_0^\infty \int_0^{\gamma/v} p(v)p(v') dv' dv, \tag{36}$$

where we made the reasonable assumption that p is symmetric. This is illustrated in Figure 14b.

For a sufficiently large critical γ , denoted γ_c , the joint probability density $p(v)p(v')$ will be concentrated almost entirely underneath the curve $v' = \gamma_c/v$, meaning that $\mathbb{P}(|vv'| < \gamma_c) \approx 1$. At this point we can invoke Hoeffding's inequality [Hoeffding, 1963] to estimate the probability with which the empirical expectation with sample size n , \mathbb{E}_n , deviates from its expected value 0:

$$\mathbb{P}(|\mathbb{E}_n| > \alpha) \lesssim 2 e^{-\frac{n\alpha^2}{2\gamma_c^2}}. \tag{37}$$

Equation (37) indicates that the probability of \mathbb{E}_n exceeding some threshold α decreases exponentially with increasing n . Furthermore, the smaller the critical γ_c can be chosen, the more quickly the right-hand side of (37) will approach 0. It follows that convergence speed increases the more $p(v)$ is concentrated around $v = 0$. However, in addition to being concentrated around $v = 0$, we require the variance of v to equal 1, i.e., $\mathbb{E}(v^2) = 1$, in order to ensure the δ property of equation (6).

It follows that the sampling distribution for the random numbers v should have both concentration near the origin and variance 1. A normal distribution with unit variance—used for all the numerical examples presented in previous sections—is thus expected to provide useful results. We refrain from trying to find the optimal sampling distribution as the result will anyway depend on the unavoidably subjective way of measuring concentration.

In addition to providing guidance for the choice of the sampling distribution, equation (37) also provides a theoretical justification for the use of a small number of samples. The exponential in n on the right-hand side of (37) decays rapidly for small n and is nearly flat for large n , meaning that the addition of new samples has an exponentially decreasing effect.

7.2. Convergence of the Autocorrelation

A key observation already made in the introductory section 3 is the rapid convergence of the autocorrelation $c(x)$ in the vicinity of $x = 0$. In the following we underpin this observation theoretically, thereby demonstrating its general validity. For this, we consider a Gaussian random function $v(x)$, $x \in \mathcal{D} \subset \mathbb{R}$, with mean zero,

$$\mathbb{E}[v(x)] = 0, \quad (38)$$

and covariance

$$\mathbb{E}[v(x)v(x')] = \delta(x - x'), \quad (39)$$

for all $x, x' \in \mathcal{D}$. It is readily verified that the random variable

$$h(x) = \int_{\mathcal{D}} dx' H(x, x')v(x'), \quad (40)$$

is also Gaussian with mean 0 and covariance

$$\mathbb{E}[h(x)h(x')] = \int_{\mathcal{D}} H(x, y)H(x', y) dy. \quad (41)$$

If the Hessian is stationary, that is, $H(x, y) \equiv H(x - y)$, we can view $h(x)$ as a stationary stochastic process with covariance

$$C(x) = \mathbb{E}[h(y)h(y + x)] = \int_{\mathcal{D}} H(y - x')H(y - x' + x) dx'. \quad (42)$$

If $H(x)$ is a decaying function, $C(x)$ will decay as well, albeit more slowly. For $H(x) = e^{-ax^2}$, for example, we have $C(x) \propto e^{-(\alpha/2)x^2}$.

7.2.1. Estimating the Covariance of Autocorrelations From the Sample Average

We can estimate the covariance of a stochastic process via the autocorrelation—even without assuming a specific functional for H , as done in section 3. To prove this statement, we let $\mathcal{D} = [0, L]$ and define the autocorrelation as

$$c(x) = \frac{1}{L - x} \int_0^{L-x} h(y)h(y + x) dy. \quad (43)$$

It follows that

$$\mathbb{E}[c(x)] = C(x). \quad (44)$$

The second moment of c is given by

$$\mathbb{E}[c(x)^2] = \frac{1}{(L - x)^2} \int_0^{L-x} \int_0^{L-x} \mathbb{E}[h(y)h(y + x)h(y')h(y' + x)], dy dy', \quad (45)$$

which, using Isserlis' theorem [e.g., *Isserlis, 1918; Lindgren et al., 2013*], can be written as

$$\mathbb{E}[c(x)^2] = \frac{1}{(L - x)^2} \int_0^{L-x} \int_0^{L-x} [C(x)^2 + C(y - y')^2 + C(y - y' + x)C(y - y' - x)] dy dy'. \quad (46)$$

The variance, then, is given by

$$\mathbb{E}[c(x)^2] - \mathbb{E}[c(x)]^2 = \frac{1}{(L - x)^2} \int_0^{L-x} \int_0^{L-x} [C(y - y')^2 + C(y - y' + x)C(y - y' - x)] dy dy', \quad (47)$$

which can be rewritten as

$$\mathbb{E}[c(x)^2] - \mathbb{E}[c(x)]^2 = \frac{1}{(L - x)} \int_{-(L-x)}^{(L-x)} [C(y)^2 + C(y + x)C(y - x)] dy. \quad (48)$$

For large L and small x (relative to the decay of $C(x)$), the integral is constant and the variance can be expressed as

$$\mathbb{E}[c(x)^2] - \mathbb{E}[c(x)]^2 \propto \frac{1}{(L-x)}. \quad (49)$$

In practice, we would estimate the expectation using a number of independent samples of the random function v , yielding independent samples $h_i(x)$ for $i = 1, 2, \dots, n$. The estimator is given by

$$c_n(x) = \frac{1}{n(L-x)} \sum_{i=1}^n \int_0^{L-x} h_i(y)h_i(y+x) dy. \quad (50)$$

It follows that the expectation of $c_n(x)$ is given by

$$\mathbb{E}[c_n(x)] = C(x). \quad (51)$$

The expected variance of the sample average autocorrelation $c_n(x)$ is therefore given by

$$\mathbb{E}[(c_n(x) - C(x))^2] \propto \frac{1}{n(L-x)}. \quad (52)$$

Relation (52) implies that a small sample size n can be compensated by a large L , that is, integration over a wider interval. Moreover, the error is smallest for small x , which explains why the width of the autocorrelation can be estimated reliably from values around $x = 0$ even when the sample size is small.

8. Discussion

In the following we discuss extensions and technical details of our method, as well as its potentials and limitations.

8.1. Advantages and Limitations of the Proposed Method

8.1.1. Convergence and Scalability

In addition to its algorithmic simplicity, the small number of samples needed to obtain useful resolution proxies is a key advantage of our method. Empirically, we observe that approximately five random samples are sufficient to ensure that the resolution length estimates have converged sufficiently to serve as a useful aid in the interpretation of tomographic models. In section 7.2 we underpin this observation theoretically by showing that the autocorrelation converges rapidly near the origin, thereby allowing for a reliable estimation of its width even when only a small number of samples is used.

Two factors contribute to the rapid convergence: (i) The sampling distribution of the random test functions $v(x)$ for fixed x is Gaussian with expectation 0 and variance 1. As shown theoretically in section 7, this choice ensures exponential convergence with increasing sample size. (ii) At least as important is the measurement process, i.e., the way in which the width of the autocorrelations is estimated. As illustrated in Figure 2, the autocorrelations may contain significant sidelobes even when 10 samples or more are used. Nevertheless, the half width of the autocorrelations generally provides reliable estimates of their width for few samples. We elaborate more on this aspect in section 8.3. In our numerical experiments we observe that the required number of samples is approximately the same in one, two, and three dimensions. Currently, we cannot offer a rigorous explanation for this perfect scaling property.

8.1.2. Computational Requirements

Owing to its fast convergence, our method is computationally inexpensive. The computation of N Hessian model application requires $N + 1$ forward and $2N$ adjoint simulations [Santosa and Symes, 1988; Fichtner and Trampert, 2011a], meaning that a total of 16 simulations is needed to estimate resolution with five random samples. To put this number in perspective, we note that one iteration in a full-waveform inversion requires at least three simulations to solve the forward problem, compute Fréchet kernels, and find a step length. It follows that the computational requirements of the proposed resolution analysis are roughly equivalent to a full-waveform inversion with only five iterations. However, between 10 and 100 iterations are typically needed in real-data applications [e.g., Igel et al., 1996; Pratt, 1999; Dessa et al., 2004; Fichtner et al., 2009a; Virieux and Operto, 2009; Kawai and Geller, 2010; Tape et al., 2010; Rickers et al., 2013; Afanasiev et al., 2014].

It follows that the proposed method is not only more quantitative than simple recovery tests, it is also computationally less expensive.

8.1.3. Local Resolution Analysis

A disadvantage of our method is its local nature, meaning the fact that resolution is quantified in the vicinity of the optimal model. To provide meaningful results, the optimization scheme must have arrived at a model where the curvature of the misfit surface—quantified in terms of the Hessian—is similar to the one at the actual optimum. While the optimal model can usually only be approached but not reached, convergence toward it can be greatly accelerated through the use of Newton-type methods [Pratt *et al.*, 1998], which are particularly beneficial in multiparameter inversions [Métivier *et al.*, 2013, 2014].

The presence of potentially meaningful local minima cannot be detected by our method. This would require the use of Monte Carlo methods [e.g., Mosegaard and Tarantola, 1995; Sambridge and Mosegaard, 2002; Tarantola, 2005] or other techniques that explore the model space more comprehensively.

Reaching the global minimum instead of a local one can be challenging, depending on the nature of the data and the complexity of the Earth. Methods to almost ensure convergence toward the global optimum include the progression from low to high frequencies during the inversion [Bunks *et al.*, 1995; Fichtner, 2010; Sirgue and Pratt, 2004] and the extension of the search space to higher dimensions [Symes, 2008; Almomin and Biondi, 2012; van Leeuwen and Herrmann, 2013; Biondi and Almomin, 2014].

8.2. Extension to the Application of (Iterative) Preconditioners

Throughout most of our developments, we used the Hessian as a conservative approximation of the point spread function. This simplification was based on the argument from section 2 that the reconstruction of a point-localized heterogeneity $\delta\mathbf{m}$ with one steepest descent iteration is proportional to $\mathbf{H}(\mathbf{m})\delta\mathbf{m}$. Methods that are more elaborate than steepest descent apply a preconditioner \mathbf{P}^{-1} , meaning that the reconstruction would be proportional to $\mathbf{P}^{-1}\mathbf{H}(\mathbf{m})\delta\mathbf{m}$ (see equation (3)).

We note that the proposed resolution analysis tools are applicable to this more general case where a preconditioner is applied. All proofs of the autocorrelation properties remain valid when \mathbf{H} is replaced by $\tilde{\mathbf{H}} = \mathbf{P}^{-1}\mathbf{H}$. This includes methods such as truncated Newton [e.g., Métivier *et al.*, 2013, 2014] where the application of \mathbf{P} is computed iteratively.

8.3. Measuring the Width of Autocorrelations and Point Spread Functions

Measuring the width of correlation functions is a key element in the proposed set of methods. In the absence of a universal measure of width, subjective and pragmatic choices are required that will affect the resolution length estimates to some extent. Common measures, including the variance computed over a finite interval and the various deltaness measures proposed by Backus and Gilbert [1968, 1970], require unimodal functions. As illustrated in Figures 2 and 5, autocorrelation functions are mostly not unimodal. The typical presence of sidelobes in the autocorrelations can result in meaningless estimates of their width when the measurement requires unimodality.

For this reason, we decided to estimate the variance σ of an autocorrelation from its half-width $\ell_{1/2}$, indicated by the red bar in Figure 2a. Both are related by $2\sigma^2 \log \frac{1}{2} = -\ell_{1/2}^2$. In addition to being comparatively robust against sidelobes, the estimated half width converges quickly, meaning that only a small number of samples is needed to approximate the true value sufficiently well. This, in effect, is shown also in Figure 2a. In this sense, the choice of a measure of width also affects the convergence speed of the resolution length estimates.

8.4. The Size of Random Perturbations

The computation of resolution lengths rests on the application of the Hessian to and the propagation of wavefields through random test models (see Appendix A). These simulations pose two types of potential problems: (i) Numerical methods used to solve the wave equation pose restrictions on the admissible types of heterogeneities that can be treated accurately. (ii) Some model perturbations may be too weak to produce wavefield perturbations above the numerical noise; or they may be so strong that meaningful misfit measurements become impossible. In the following paragraphs we discuss suitable choices for the spatial extent and the amplitude of random perturbations that produce reliable results while honoring restrictions imposed by numerical wave propagation.

8.4.1. Spatial Extent

Valuable guidance concerning the spatial extent of the random heterogeneities comes from the one-dimensional experiments in section 3, where we used analytically defined point spread functions. Choosing the width of the individual perturbations to be less than half the width of the point spread function limits the errors in the resolution length estimates to below $\sim 10\%$, with variations resulting from the random nature

of the estimation procedure. In practice, the width of the spatially variable point spread function is unknown. However, the minimum expected resolution length can be estimated from physical arguments and will usually be on the order of 1 wavelength. Realizing that most numerical wave equation solvers require on the order of 10 grid points per wavelength to produce accurate solutions, we deduce that the spatial extent of the random perturbations should be around five numerical grid points or less.

For the analytically defined point spread functions in section 3 there is no lower bound on the spatial extent of the perturbations. In practice, however, the discretization of space derivatives in numerical solvers does impose such a lower bound. In spectral-element methods, for instance, Earth models must be represented as low-order polynomials within each element [e.g., *Faccioli et al.*, 1996; *Komatitsch and Vilotte*, 1998; *Fichtner et al.*, 2009b; *Peter et al.*, 2011]. Similar smoothness constraints also apply to the Direct Solution Method in order to ensure accurate approximations of integrals in the computation of mass and stiffness matrices [e.g., *Geller and Ohminato*, 1994; *Cummins et al.*, 1997]. Some variants of the discontinuous Galerkin method impose constant material properties within elements [e.g., *Dumbser et al.*, 2007a, 2007b; *de la Puente et al.*, 2007]. Finite difference methods require smooth heterogeneities to ensure that spatial derivatives are computed accurately [e.g., *Igel et al.*, 1995; *Moczo et al.*, 2000; *Bohlen*, 2002]. Most of the requirements imposed by numerical methods can be met by choosing the width of the random perturbations to be around half a wavelength. This is in accord with the previously mentioned upper bound that results from the requirement to estimate accurate resolution lengths. Wider perturbations, for instance, in response to numerical restrictions, will generally increase the estimated resolution lengths, thus making them more conservative.

8.4.2. Amplitude

Propagating wavefields through perturbed Earth models in order to evaluate Hessian model applications can cause complications in the misfit measurement process. Small point-localized perturbations needed for the computation of point spread functions may result in wavefield perturbations below the numerical error. Conversely, large model perturbations can lead to changes of the wavefield that are so substantial that measurements of phase differences or time shifts suffer from cycle skip problems [*Luo and Schuster*, 1991; *Gee and Jordan*, 1992; *Fichtner et al.*, 2008; *van Leeuwen and Mulder*, 2010]. To avoid such complications, the model perturbations should be scaled to produce wavefield changes that are large enough to be above the numerical error and small enough to produce meaningful measurements. Subsequently, the Hessian model application can be scaled back to its proper value. For example, the volume of the point spread functions as a function of position (equation (16)) may be computed as

$$\int H(\mathbf{x}_i; \mathbf{y}) d\mathbf{y} = \frac{1}{\epsilon} \int H(\mathbf{x}_i; \mathbf{y}) \epsilon d\mathbf{y}, \quad (53)$$

with a suitably selected ϵ . In practice, the choice of ϵ should be guided by the range of linearity of the measurement. For instance, travel times measured by cross correlation are linearly related to seismic velocity perturbations of up to 10% [*Mercerat and Nolet*, 2013], meaning that random velocity perturbations in the percent range would be a meaningful choice. Velocity perturbations on the percent level are also within the validity range of the Born approximation, that is, the regime where seismic waveforms are linearly dependent on 3-D Earth structure [e.g., *Parisi et al.*, 2015]. In the case of physical parameters for which the range of linearity is less well known, it should be explored through numerical perturbation tests prior to the application of the proposed resolution analysis tools.

8.5. Potentials

We illustrate our method using examples from full-waveform inversion in two and three dimensions. This choice is motivated by the fact that full-waveform inversion currently poses the largest computational challenges of all methods in the tomographic toolbox. In this sense, the applicability to full-waveform inversion ensures that other tomographic techniques can also benefit from the proposed resolution analyses. Depending on the specifics of the method, small modifications may be necessary. In linearized travel time tomography, for instance, random test models would be applied to the resolution matrix and not to the Hessian operator, as in full-waveform inversion.

In addition to tomographic problems, resolution analysis by random probing may be used in other inverse methods that try to constrain continuously distributed properties. This includes traditional electromagnetic and potential-field inversions but also emerging geodynamic data assimilation [e.g., *Bunge et al.*, 2003; *Liu and Gurnis*, 2008; *Worthen et al.*, 2014].

9. Conclusions

We develop and apply random probing techniques for resolution analysis in tomographic inverse problems. These are based on applications of the Hessian or resolution operators to random test models and the subsequent manipulation of the results. Depending on the properties of the test vectors, different measures of resolution can be estimated. Examples include direction- and position-dependent resolution lengths, the position-dependent volumes of point spread functions, and the strength of interparameter mappings in multiparameter inversion.

The most important properties of the proposed resolution analysis are the following:

1. Low algorithmic complexity and ease of implementation. No modification of preexisting forward modeling and inversion tools is needed.
2. The required number of test models is around five for inversions in two and three dimensions. This ensures that resolution analysis is much less expensive than the actual inversion.
3. Ability to quantify both spatial resolution and interparameter trade-offs.

Our developments are general in the sense that they are applicable to any inverse problem that seeks to constrain quantities that are continuously distributed in space. Electromagnetic and potential-field inversions, as well as geodynamic data assimilation fall into this class of problems.

Appendix A: A Review of Second-Order Adjoints

We present a brief review on the computation of Hessian model applications based on second-order adjoints. For more details and examples, the reader is referred to *Fichtner and Trampert* [2011a].

A1. General Development

The first variation of the misfit functional χ with respect to an infinitesimal model perturbation $\delta\mathbf{m}_1$ can be written in terms of the sensitivity kernel $\mathbf{K}(\mathbf{x})$:

$$\delta_1\chi = \int \delta\mathbf{m}_1^T(\mathbf{x}) \mathbf{K}(\mathbf{x}) d\mathbf{x}. \quad (\text{A1})$$

Using equation (A1), we express the second variation of χ in terms of both the sensitivity kernel \mathbf{K} and the Hessian \mathbf{H} :

$$\begin{aligned} \delta_1\delta_2\chi &= \iint \delta\mathbf{m}_1^T(\mathbf{x}) \mathbf{H}(\mathbf{x}; \mathbf{y}) \delta\mathbf{m}_2(\mathbf{y}) d\mathbf{x} d\mathbf{y} \\ &= \int \delta\mathbf{m}_1^T(\mathbf{x}) \delta_2\mathbf{K}(\mathbf{x}) d\mathbf{x}, \end{aligned} \quad (\text{A2})$$

where $\delta_2\mathbf{K}(\mathbf{x})$ denotes the variation of \mathbf{K} with respect to $\delta\mathbf{m}_2$. Comparing the second and third terms of equation (A2) provides an expression for the application of the Hessian to a model perturbation $\delta\mathbf{m}$ in terms of the variation of the sensitivity kernel:

$$\int \mathbf{H}(\mathbf{x}; \mathbf{y}) \delta\mathbf{m}(\mathbf{y}) d\mathbf{y} = \delta\mathbf{K}(\mathbf{x}). \quad (\text{A3})$$

The kernel \mathbf{K} is an explicit function of the Earth model \mathbf{m} , and also depends linearly on the forward wavefield \mathbf{u} and the adjoint wavefield \mathbf{u}^\dagger [e.g., *Lailly*, 1983; *Tarantola*, 1988; *Tromp et al.*, 2005; *Fichtner et al.*, 2006]. Taking these dependencies into account and omitting the dependence of \mathbf{K} on \mathbf{x} , we find

$$\delta\mathbf{K} = \mathbf{K}'(\mathbf{m}, \mathbf{u}, \mathbf{u}^\dagger) + \mathbf{K}(\mathbf{m}, \delta\mathbf{u}, \mathbf{u}^\dagger) + \mathbf{K}(\mathbf{m}, \mathbf{u}, \delta\mathbf{u}^\dagger). \quad (\text{A4})$$

The symbol \mathbf{K}' denotes the partial derivative of \mathbf{K} with respect to the model parameters, i.e., the derivative with respect to the explicit dependence of \mathbf{K} on \mathbf{m} .

A2. Example

We consider, as an example, the case where the Earth model is parameterized in terms of the P velocity α , the S velocity β , and density ρ , i.e., $\mathbf{m} = (\alpha, \beta, \rho)^T$ and $\mathbf{K} = (K_\alpha, K_\beta, K_\rho)$. Equation (A1) then takes the special form

$$\delta_1\chi = \int K_\alpha \delta\alpha d\mathbf{x} + K_\beta \delta\beta d\mathbf{x} + K_\rho \delta\rho d\mathbf{x}. \quad (\text{A5})$$

The S velocity kernel, K_β , can be expressed in terms of the forward strain field $\epsilon_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i)$ and the adjoint strain field $\epsilon_{ij}^\dagger = \frac{1}{2}(\partial_i u_j^\dagger + \partial_j u_i^\dagger)$ [e.g., *Tromp et al.*, 2005; *Fichtner et al.*, 2006; *Liu and Tromp*, 2008]:

$$K_\beta = 4\rho\beta \int \epsilon_{ij}\epsilon_{ij}^\dagger dt - 4\rho\alpha \int \epsilon_{ii}\epsilon_{jj}^\dagger dt. \quad (\text{A6})$$

Using the general formula from equation (A4), we find that the application of the diagonal Hessian element $H_{\beta\beta}$ to a shear velocity perturbation $\delta\beta$ is given by

$$\begin{aligned} \int H_{\beta\beta} \delta\beta d\mathbf{x} &= 4\rho \delta\beta \int \underbrace{\epsilon_{ij}\epsilon_{ij}^\dagger}_{\mathbf{K}(\mathbf{m},\mathbf{u},\mathbf{u}^\dagger)} dt \\ &+ 4\rho\beta \int \underbrace{\delta_\beta \epsilon_{ij}\epsilon_{ij}^\dagger}_{\mathbf{K}(\mathbf{m},\delta\mathbf{u},\mathbf{u}^\dagger)} dt - 4\rho\alpha \int \delta_\beta \epsilon_{ii}\epsilon_{jj}^\dagger dt \\ &+ 4\rho\beta \int \underbrace{\epsilon_{ij}\delta_\beta \epsilon_{ij}^\dagger}_{\mathbf{K}(\mathbf{m},\mathbf{u},\delta\mathbf{u}^\dagger)} dt - 4\rho\alpha \int \epsilon_{ii}\delta_\beta \epsilon_{jj}^\dagger dt. \end{aligned} \quad (\text{A7})$$

Applications of other components of the Hessian to the different types of model perturbations can be computed following the same recipe. The perturbations of the forward and adjoint wavefields can be computed most easily with finite differences, that is, by taking the difference of wavefields in the original and in the perturbed Earth model [*Fichtner and Trampert*, 2011a, 2011b].

Acknowledgments

The authors would like to thank Mike Afanasiev, Anatoly Baumstein, Nienke Blom, Christian Böhm, Shравan Hanasoge, Mauricio Sacchi, William Symes, Jeannot Trampert, Paul Williamson, Tatiana Yanovskaya, and Saulė Žukauskaitė for useful discussions and brainstorming sessions. The comments of two anonymous reviewers and the Associate Editor helped to improve this manuscript. This research was supported by the Swiss National Supercomputing Center (CSCS) in the form of the GeoScale and CH1 projects, by the Swiss National Science Foundation (SNF) under grant 200021_149143 and by the Netherlands Organization for Scientific Research (VIDI grant 864.11.008). All seismic waveform data used in this study are available via the data centers of ORFEUS (www.orfeus-eu.org) and IRIS (www.iris.edu).

References

- Afanasiev, M. V., R. G. Pratt, R. Kamei, and G. McDowell (2014), Waveform-based simulated annealing of crosshole transmission data: A semi-global method for estimating seismic anisotropy, *Geophys. J. Int.*, *199*, 1586–1607.
- Aki, K., and W. H. K. Lee (1976), Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model, *J. Geophys. Res.*, *81*, 4381–4399.
- Aki, K., A. Christofferson, and E. S. Husebye (1976), Determination of three-dimensional seismic structure of the lithosphere, *J. Geophys. Res.*, *81*, 277–296.
- Almomin, A., and B. Biondi (2012), Tomographic full waveform inversion: Practical and computationally feasible approach, Expanded Abstracts SEG– 2012-0976 presented at 2012 SEG Annual Meeting, pp. 500–505, Las Vegas, Nevada, 4–9 Nov.
- An, M. (2012), A simple method for determining the spatial resolution of a general inverse problem, *Geophys. J. Int.*, *191*, 849–864.
- Arnoldi, W. E. (1951), The principle of minimized iteration in the solution of the matrix eigenvalue problem, *Q. Appl. Math.*, *9*, 17–29.
- Avron, H., and S. Toledo (2011), Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix, *J. ACM*, *58*, 8, doi:10.1145/1944345.1944349.
- Babuška, V., and M. Cara (1991), *Seismic Anisotropy in the Earth*, Kluwer Acad., Dordrecht, Netherlands.
- Backus, G. E., and F. Gilbert (1968), The resolving power of gross Earth data, *Geophys. J. R. Astron. Soc.*, *16*, 169–205.
- Backus, G. E., and F. Gilbert (1970), Uniqueness in the inversion of inaccurate gross Earth data, *Philos. Trans. R. Soc. London, Ser. A*, *266*, 123–192.
- Becker, T. W., and L. Boschi (2002), A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosys.*, *3*, 1003, doi:10.1029/2001GC000168.
- Bekas, C., E. Kokopoulou, and Y. Saad (2007), An estimator for the diagonal of a matrix, *Appl. Numer. Math.*, *57*, 1214–1229.
- Berkhout, A. J. (1984), *Handbook of Geophysical Exploration: Seismic Resolution*, Geophys. Press Ltd., London, Amsterdam.
- Biondi, B., and A. Almomin (2014), Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion, *Geophysics*, *79*, WA129–WA140.
- Bleibinhaus, F., J. A. Hole, T. Ryberg, and G. S. Fuis (2007), Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging, *J. Geophys. Res.*, *112*, B06315, doi:10.1029/2006JB004611.
- Bohlen, T. (2002), Parallel 3-D viscoelastic finite difference modelling, *Comput. Geosci.*, *28*, 887–899.
- Boschi, L. (2003), Measures of resolution in global body wave tomography, *Geophys. Res. Lett.*, *30*(19), 1978, doi:10.1029/2003GL018222.
- Brossier, R., S. Operto, and J. Virieux (2009), Robust elastic frequency-domain full waveform inversion using the L_1 norm, *Geophys. Res. Lett.*, *36*, L20310, doi:10.1029/2009GL039458.
- Bui-Thanh, T., O. Ghattas, J. Martin, and G. Stadler (2013), A computational framework for infinite-dimensional Bayesian inverse problems part I: The linearized case, with application to global seismic inversion, *SIAM J. Sci. Comput.*, *35*, A2494–A2523, doi:10.1137/12089586X.
- Bunge, H.-P., C. R. Hagelberg, and B. J. Travis (2003), Mantle circulation models with variational data assimilation: Inferring past mantle flow and structure from plate motion histories and seismic tomography, *Geophys. J. Int.*, *152*, 280–301.
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent (1995), Multiscale seismic waveform inversion, *Geophysics*, *60*, 1457–1473.
- Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef (1985), A nonreflecting boundary condition for discrete acoustic and elastic wave equations, *Geophysics*, *50*, 705–708.
- Colli, L., A. Fichtner, and H.-P. Bunge (2013), Full waveform tomography of the upper mantle in the South Atlantic region: Imaging westward fluxing shallow asthenosphere?, *Tectonophysics*, *604*, 26–40, doi:10.1016/j.tecto.2013.06.015.
- Cummins, P. R., N. Takeuchi, and R. J. Geller (1997), Computation of complete synthetic seismograms for laterally heterogeneous models using the Direct Solution Method, *Geophys. J. Int.*, *130*, 1–16.

- Deal, M. M., and G. Nolet (1996), Nullspace shuttles, *Geophys. J. Int.*, *124*, 372–380.
- de la Puente, J., M. Dumbser, M. Käser, and H. Igel (2007), An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes - IV. Anisotropy, *Geophys. J. Int.*, *169*, 1210–1228.
- de Wit, R. W. L., A. Valentine, and J. Trampert (2013), Bayesian inference of Earth's radial seismic structure from body-wave travel times using neural networks, *Geophys. J. Int.*, *195*, 408–422.
- Dessa, J. X., S. Operto, S. Kodaira, A. Nakanishi, G. Pascal, J. Virieux, and Y. Kaneda (2004), Multiscale seismic imaging of the Eastern Nankai trough by full waveform inversion, *Geophys. Res. Lett.*, *31*, L18606, doi:10.1029/2004GL020453.
- Díaz, J., A. Villaseñor, J. Gallart, J. Morales, A. Pazos, D. Córdoba, J. Pulgar, J. L. García-Lobón, M. Harnañi, and Topolberia Seismic Working Group (2009), The IBERARRAY broadband seismic network: A new tool to investigate the deep structure beneath Iberia, *ORFEUS Newsl.*, *8*, 1–6.
- Drineas, P., R. Kannan, and M. W. Mahoney (2006), Fast Monte Carlo algorithms for matrices II: Computing a low-rank approximation to a matrix, *SIAM J. Comput.*, *36*, 158–183.
- Dumbser, M., M. Käser, and J. de la Puente (2007a), Arbitrary high-order finite volume schemes for seismic wave propagation on unstructured meshes in 2D and 3D, *Geophys. J. Int.*, *171*, 665–694.
- Dumbser, M., M. Käser, and E. Toro (2007b), An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes, Part V: Local time stepping and p -adaptivity, *Geophys. J. Int.*, *171*, 695–717.
- Dziewoński, A. M., B. H. Hager, and R. J. O'Connell (1977), Large-scale heterogeneities in the lower mantle, *J. Geophys. Res.*, *82*, 239–255.
- Faccioli, E., F. Maggio, A. Quarteroni, and A. Tagliani (1996), Spectral-domain decomposition methods for the solution of acoustic and elastic wave equations, *Geophysics*, *61*(4), 1160–1174.
- Fichtner, A. (2010), *Full Seismic Waveform Modelling and Inversion*, Springer, Germany.
- Fichtner, A., and J. Trampert (2011a), Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.*, *185*, 775–798.
- Fichtner, A., and J. Trampert (2011b), Resolution analysis in full waveform inversion, *Geophys. J. Int.*, *187*, 1604–1624.
- Fichtner, A., H.-P. Bunge, and H. Igel (2006), The adjoint method in seismology - I. Theory, *Phys. Earth Planet. Inter.*, *157*, 86–104.
- Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge (2008), Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *Geophys. J. Int.*, *175*, 665–685.
- Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge (2009a), Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, *179*, 1703–1725.
- Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge (2009b), Spectral-element simulation and inversion of seismic waves in a spherical section of the Earth, *J. Numer. Anal. Ind. Appl. Math.*, *4*, 11–22.
- Fichtner, A., S. Fishwick, K. Yoshizawa, and B. L. N. Kennett (2012), Optimal spherical spline filters for the analysis and comparison of regional-scale tomographic models, *Phys. Earth Planet. Inter.*, *190–191*, 44–50.
- Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. Villasenor (2013), Multi-scale full waveform inversion, *Geophys. J. Int.*, *194*, 534–556.
- Frieze, A., R. Kannan, and S. Vempala (2004), Fast Monte Carlo algorithms for finding low-rank approximations, *J. ACM*, *51*, 1025–1041.
- Gallagher, K., M. S. Sambridge, and G. G. Drijkoningen (1991), Genetic algorithms: An evolution of Monte Carlo methods for strongly non-linear geophysical optimization problems, *Geophys. Res. Lett.*, *18*, 2177–2180.
- Gee, L. S., and T. H. Jordan (1992), Generalized seismological data functionals, *Geophys. J. Int.*, *111*, 363–390.
- Geller, R., and T. Ohminato (1994), Computation of synthetic seismograms and their partial derivatives for heterogeneous media with arbitrary natural boundary conditions using the Direct Solution Method, *Geophys. J. Int.*, *116*, 421–446.
- Halko, N., P. G. Martinsson, and J. A. Tropp (2011), Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, *SIAM Rev.*, *53*, 217–288.
- Hoeffding, W. (1963), Probability inequalities for sums of bounded random variables, *J. Am. Stat. Assoc.*, *58*, 13–30.
- Hutchinson, M. F. (1990), A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines, *Commun. Stat.*, *19*, 433–450.
- Igel, H., P. Mora, and B. Rioulet (1995), Anisotropic wave propagation through FD grids, *Geophysics*, *60*, 1203–1216.
- Igel, H., H. Djikpesse, and A. Tarantola (1996), Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio, *Geophys. J. Int.*, *124*, 363–371.
- Isserlis, L. (1918), On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables, *Biometrika*, *12*, 134–139.
- Kawai, K., and R. J. Geller (2010), Waveform inversion for localized seismic structure and an application to D'' structure beneath the Pacific, *J. Geophys. Res.*, *115*, B01305, doi:10.1029/2009JB006503.
- Komatitsch, D., and J. P. Vilotte (1998), The spectral element method: An effective tool to simulate the seismic response of 2D and 3D geological structures, *Bull. Seismol. Soc. Am.*, *88*, 368–392.
- Krischer, L., A. Fichtner, S. Žukauskaitė, and H. Igel (2015), Large-scale seismic inversion framework, *Seismol. Res. Lett.*, *86*, 1198–1207, doi:10.1785/0220140248.
- Lailly, P. (1983), The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering: Theory and Application*, edited by J. Bednar et al., pp. 206–220, Soc. Ind. Appl. Math., Philadelphia, Pa.
- Lanczos, C. (1950), An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bur. Stand.*, *45*, 255–282.
- Lévêque, J. J., L. Rivera, and G. Wittlinger (1993), On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, *115*, 313–318.
- Lindgren, G., H. Rootzen, and M. Sandsten (2013), *Stationary Stochastic Processes for Scientists and Engineers*, CRC Press, Boca Raton, Fla.
- Liu, L., and M. Gurnis (2008), Simultaneous inversion of mantle properties and initial conditions using an adjoint of mantle convection, *J. Geophys. Res.*, *113*, B08405, doi:10.1029/2008JB005594.
- Liu, Q., and J. Tromp (2008), Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.*, *174*, 265–286.
- Luo, Y., and G. T. Schuster (1991), Wave-equation traveltimes inversion, *Geophysics*, *56*, 645–653.
- MacCarthy, J. K., B. Borchers, and R. C. Aster (2011), Efficient stochastic estimation of the model resolution matrix diagonal and generalized cross-validation for large geophysical inverse problems, *J. Geophys. Res.*, *116*, B10304, doi:10.1029/2011JB008234.
- Meier, U., A. Curtis, and J. Trampert (2007a), Global crustal thickness from neural network inversion of surface wave data, *Geophys. J. Int.*, *169*, 706–722.

- Meier, U., A. Curtis, and J. Trampert (2007b), Fully nonlinear inversion of fundamental mode surface waves for a global crustal model, *Geophys. Res. Lett.*, *34*, L16304, doi:10.1029/2007GL030989.
- Meju, M. A. (2009), Regularized extremal bounds analysis (REBA): An approach to quantifying uncertainty in nonlinear geophysical inverse problems, *Geophys. Res. Lett.*, *36*, L03304, doi:10.1029/2008GL036407.
- Meju, M. A., and V. Sakkas (2007), Heterogeneous crust and upper mantle across southern Kenya and the relationship to surface deformation as inferred from magnetotelluric imaging, *J. Geophys. Res.*, *112*, B04103, doi:10.1029/2005JB004028.
- Mercerat, D., and G. Nolet (2013), On the linearity of cross-correlation delay times in finite-frequency tomography, *Geophys. J. Int.*, *192*, 681–687.
- Métivier, L., F. Breteau, R. Brossier, S. Operto, and J. Virieux (2014), Full waveform inversion and the truncated Newton method: Quantitative imaging of complex subsurface structures, *Geophys. Prospect.*, *62*, 1353–1375.
- Métivier, L., R. Brossier, J. Virieux, and S. Operto (2013), Full waveform inversion and the truncated Newton method, *SIAM J. Sci. Comput.*, *35*, B401–B437.
- Moczo, P., J. Kristek, and L. Halada (2000), 3D fourth-order staggered-grid finite-difference schemes: Stability and grid dispersion, *Bull. Seismol. Soc. Am.*, *90*, 587–603.
- Mora, P. (1989), Inversion = migration + tomography, *Geophysics*, *54*, 1575–1586.
- Mosca, I., L. Cobden, A. Deuss, J. Ritsema, and J. Trampert (2012), Seismic and mineralogical structures of the lower mantle from probabilistic tomography, *J. Geophys. Res.*, *117*, B06304, doi:10.1029/2011JB008851.
- Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*, 12,431–12,447.
- Nemeth, T., C. Wu, and T. Schuster (1999), Least-squares migration of incomplete reflection data, *Geophysics*, *64*, 208–221.
- Nolet, G. (2008), *A Breviary of Seismic Tomography*, Cambridge Univ. Press, Cambridge, U.K.
- Nolet, G., R. Montelli, and J. Virieux (1999), Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.*, *138*, 36–44.
- Panning, M., and B. Romanowicz (2006), A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.*, *167*, 361–379.
- Parisi, L., A. M. G. Ferreira, and Y. Capdeville (2015), Validity domain of the Born approximation for seismic waveform modelling in realistic 3-D Earth structure, *Geophys. J. Int.*, *200*, 908–914.
- Peter, D., et al. (2011), Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.*, *186*, 721–739.
- Pratt, R. G. (1999), Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model, *Geophysics*, *64*, 888–901.
- Pratt, R. G., C. Shin, and G. J. Hicks (1998), Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.*, *133*, 341–362.
- Prieux, V., R. Brossier, S. Operto, and J. Virieux (2013), Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: Imaging compressional wave speed, density and attenuation, *Geophys. J. Int.*, *194*, 1640–1664.
- Rawlinson, N., A. Fichtner, M. Sambridge, and M. K. Young (2014), Seismic tomography and the assessment of uncertainty, *Adv. Geophys.*, *55*, 1–76.
- Rickers, F., A. Fichtner, and J. Trampert (2012), Imaging mantle plumes with instantaneous phase measurements of diffracted waves, *Geophys. J. Int.*, *190*, 650–664.
- Rickers, F., A. Fichtner, and J. Trampert (2013), The Iceland-Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: Evidence from full-waveform inversion, *Earth Planet. Sci. Lett.*, *367*, 39–51.
- Ritsema, J., A. Deuss, H. J. van Heijst, and J. H. Woodhouse (2011), S40RTS: A degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophys. J. Int.*, *184*, 1223–1236.
- Rokhlin, V., A. Szlam, and M. Tygert (2009), A randomized algorithm for principal component analysis, *SIAM J. Matrix Anal. Appl.*, *31*, 1100–1124.
- Sacchi, M. D., J. Wang, and H. Kuehl (2007), Estimation of the diagonal of the migration blurring kernel through a stochastic approximation, Extended Abstract SEG–2007-2437 presented at 2007 SEG Annual Meeting, 23–28 September, San Antonio, Tex., pp. 2437–2441.
- Sambridge, M. S. (1999a), Geophysical inversion with the Neighbourhood Algorithm: 1. Searching a parameter space, *Geophys. J. Int.*, *138*, 479–494.
- Sambridge, M. S. (1999b), Geophysical inversion with the Neighbourhood Algorithm: 2. Appraising the ensemble, *Geophys. J. Int.*, *138*, 727–746.
- Sambridge, M. S., and G. G. Drijkoningen (1992), Genetic algorithms in seismic waveform inversion, *Geophys. J. Int.*, *109*, 323–342.
- Sambridge, M. S., and K. Mosegaard (2002), Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, *40*(3), 1009, doi:10.1029/2000RG000089.
- Santosa, F., and W. W. Symes (1988), Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology, *Inverse Prob.*, *4*, 211–233.
- Schaeffer, A. J., and S. Lebedev (2013), Global shear speed structure of the upper mantle and transition zone, *Geophys. J. Int.*, *194*, 417–449.
- Sirgue, L., and R. G. Pratt (2004), Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, *Geophysics*, *69*, 231–248.
- Sleijpen, G. L. G., and H. van der Vorst (2000), A Jacobi–Davidson iteration method for linear eigenvalue problems, *SIAM Rev.*, *42*, 267–293.
- Symes, W. W. (2008), Migration velocity analysis and waveform inversion, *Geophys. Prospect.*, *56*, 765–790.
- Takeuchi, H., and M. Saito (1972), Seismic surface waves, in *Methods in Computational Physics*, vol. 11, edited by B. A. Bolt, pp. 217–295, Academic Press, New York.
- Tanimoto, T. (1991), Waveform inversion for three-dimensional density and S wave structure, *J. Geophys. Res.*, *96*, 8167–8189.
- Tape, C., Q. Liu, A. Maggi, and J. Tromp (2010), Seismic tomography of the southern California crust based upon spectral-element and adjoint methods, *Geophys. J. Int.*, *180*, 433–462.
- Tarantola, A. (1986), A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, *51*, 1893–1903.
- Tarantola, A. (1988), Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure Appl. Geophys.*, *128*, 365–399.
- Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, 2nd ed., Soc. for Ind. and Appl. Math., Philadelphia, Pa.
- Trampert, J., and A. Fichtner (2013), Resolution tests revisited: The power of random numbers, *Geophys. J. Int.*, *192*, 676–680.
- Tromp, J., C. Tape, and Q. Liu (2005), Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, *160*, 195–216.

- Valenciano, A., and B. Biondi (2006), Target-oriented wave-equation inversion, *Geophysics*, *71*, A35–A38.
- van Leeuwen, T., and F. J. Herrmann (2013), Mitigating local minima in full-waveform inversion by expanding the search space, *Geophys. J. Int.*, *195*, 661–667.
- van Leeuwen, T., and W. A. Mulder (2010), A correlation-based misfit criterion for wave-equation traveltome tomography, *Geophys. J. Int.*, *182*, 1383–1394.
- Virieux, J. (1984), SH wave propagation in heterogeneous media: Velocity-stress finite difference method, *Geophysics*, *49*, 1933–1942.
- Virieux, J., and S. Operto (2009), An overview of full waveform inversion in exploration geophysics, *Geophysics*, *74*, WCC127–WCC152.
- Worthen, J., G. Stadler, N. Petra, M. Gurnis, and O. Ghattas (2014), Towards adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow, *Phys. Earth Planet. Inter.*, *234*, 23–34.
- Wu, R., and K. Aki (1985), Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics*, *50*, 582–595.
- Yanovskaya, T. B. (1997), Resolution estimation in the problems of seismic ray tomography, *Izv. Phys. Solid Earth*, *33*, 76–80.
- Yanovskaya, T. B. (2000), Development of methods for surface wave tomography based on the Backus-Gilbert approach, in *Computational Seismology and Geodynamics*, vol. 7, edited by V. I. Keilis-Borok and G. M. Molchan, pp. 5–16, AGU, Washington, D. C.