# Optimal observables for multiparameter seismic tomography

Geophysical Journal International

*Geophys. J. Int.* (2014) **198,** 1241–1254 doi: 10.1093/gji/ggu204
GJI Seismology

# Optimal observables for multiparameter seismic tomography

Moritz Bernauer,[1] Andreas Fichtner[2] and Heiner Igel[1]

[1]*Department of Earth and Environmental Sciences, Ludwig-Maximilians-University, Munich, Germany.*
*E-mail: moritz.bernauer@geophysik.uni-muenchen.de*
[2]*Department of Earth Sciences, ETH Zurich, Switzerland*

Accepted 2014 May 28. Received 2014 April 14; in original form 2013 December 22

## SUMMARY

We propose a method for the design of seismic observables with maximum sensitivity to a target model parameter class, and minimum sensitivity to all remaining parameter classes. The resulting optimal observables thereby minimize interparameter trade-offs in multiparameter inverse problems.

Our method is based on the linear combination of fundamental observables that can be any scalar measurement extracted from seismic waveforms. Optimal weights of the fundamental observables are determined with an efficient global search algorithm. While most optimal design methods assume variable source and/or receiver positions, our method has the flexibility to operate with a fixed source–receiver geometry, making it particularly attractive in studies where the mobility of sources and receivers is limited.

In a series of examples we illustrate the construction of optimal observables, and assess the potentials and limitations of the method. The combination of Rayleigh-wave traveltimes in four frequency bands yields an observable with strongly enhanced sensitivity to 3-D density structure. Simultaneously, sensitivity to $S$ velocity is reduced, and sensitivity to $P$ velocity is eliminated. The original three-parameter problem thereby collapses into a simpler two-parameter problem with one dominant parameter.

By defining parameter classes to equal earth model properties within specific regions, our approach mimics the Backus–Gilbert method where data are combined to focus sensitivity in a target region. This concept is illustrated using rotational ground motion measurements as fundamental observables. Forcing dominant sensitivity in the near-receiver region produces an observable that is insensitive to the Earth structure at more than a few wavelengths' distance from the receiver. This observable may be used for local tomography with teleseismic data.

While our test examples use a small number of well-understood fundamental observables, few parameter classes and a radially symmetric earth model, the method itself does not impose such restrictions. It can easily be applied to large numbers of fundamental observables and parameters classes, as well as to 3-D heterogeneous earth models.

**Key words:** Numerical solutions; Inverse theory; Seismic tomography; Computational seismology; Theoretical seismology; Wave propagation.

## 1 INTRODUCTION

### 1.1 Multiparameter seismic inversion

The characterization of the Earth's internal state requires the combined knowledge of a broad spectrum of physical properties that may be inferred from seismic observations. Joint information on $P$-velocity $\alpha$, $S$-velocity $\beta$ and density $\rho$, for instance, is needed to independently constrain temperature and composition (e.g. Trampert *et al.* 2004; Mosca *et al.* 2012). Seismic anisotropy can be related to deformation (e.g. Hess 1964; Gung *et al.* 2003; Debayle & Ricard 2013) and the presence of small-scale heterogeneity (e.g. Backus 1962; Capdeville *et al.* 2010; Fichtner *et al.* 2013). The attenuation of seismic waves indicates high temperatures or the presence of fluids (e.g. Jackson *et al.* 2002; Jackson 2007; Dalton *et al.* 2008), and lateral variations in density provide information on convective flow (e.g. Ishii & Tromp 2001; Mosca *et al.* 2012).

Inferring the state of the Earth from seismic observations is complicated because any observable $d$ is sensitive to a wide range of properties, or parameter classes $m_j$. Arrival times of seismic waves, for example, are sensitive to both velocity and attenuation structure (e.g. Zhou 2009; Ruan & Zhou 2010).

Wave amplitudes generally depend on attenuation and on purely elastic heterogeneities that cause focusing and defocusing (e.g. Dahlen & Baig 2002; Zhou *et al.* 2004). Shear wave splitting is affected by intrinsic anisotropy and small-scale heterogeneity, including fine lamination (e.g. Backus 1962; Babuška & Cara 1991).

All seismic inverse problems are 'multiparameter problems', where trade-offs lead to uncertainties in and differences between seismic earth models. The complete quantification of these trade-offs requires fully probabilistic inversions (e.g. Sambridge & Mosegaard 2002; Bodin & Sambridge 2009; Khan *et al.* 2013) that are not feasible when the number of free parameters is large, or when the solution of the forward problem is computational intensive. While interparameter trade-offs are commonly recognized as a key problem, the design of deterministic (i.e. not probabilistic) multiparameter inversions is mostly guided by subjective choices and technical convenience, instead of being driven by the physics of the problem. The most widely used approaches are (1) ignoring model parameters, (2) scaling, (3) sequential inversion and (4) subspace methods.

Ignoring, for instance, the presence of 3-D heterogeneities in attenuation and *P* velocity is common in surface wave tomography (e.g. Yoshizawa & Kennett 2004; Fishwick *et al.* 2005; Debayle & Ricard 2013). The neglected attenuation and *P*-velocity structures may contaminate *S*-velocity models.

To reduce the effective number of model parameters, scaling relations are also frequently applied. *P*-wave anisotropy, for instance, may be scaled to *S*-wave anisotropy in surface wave tomography (e.g. Panning & Romanowicz 2006; Nettles & Dziewoński 2008) using, for instance, prior information from mineral physics (e.g. Montagner & Anderson 1989). Similarly, density variations have been scaled to seismic velocity variations in joint inversions of gravity and seismic data (e.g. Tondi *et al.* 2000, 2009; Maceira & Ammon 2009). Scaling implements prior assumptions. It therefore precludes the detection of those interesting cases where the assumptions do not hold, for example, when density and velocity heterogeneities are not correlated due to chemical heterogeneities.

Sequential inversion is widely used, for instance, in attenuation tomography. First, velocity variations are constrained using traveltime information. Subsequently, the amplitudes of seismic waves are inverted for 3-D attentuation structure (e.g. Gung & Romanowicz 2004; Kennett & Abdullah 2011). While practical, this approach ignores that the velocity model will to some extent be incorrect because the traveltimes used for its construction have sensitivity to 3-D attenuation as well. Sequential inversion ignores that any observable depends on more than one parameter class, and it relies on the subjective decision that one parameter is more important than another, and therefore should be considered first.

Subspace methods for multiparameter inversion (Kennett 1997; Kennett *et al.* 1988) introduce a scaling between different parameter classes that is determined by the data functional. While being preferable to *a priori* scalings, the use of subspace methods constitutes a subjective choice in itself, and it does not prevent the occurrence of interparameter trade-offs.

### 1.2 The origin of interparameter mappings in tomographic inversion

Interparameter mappings are an artefact of the regularization needed to enforce a unique solution of otherwise ill-posed tomographic systems. This can be seen most easily for the example of a generic linear problem where variations in the observable data vector $\delta\mathbf{d}$ are related to variations in the model parameter vector $\delta\mathbf{m}$ via a sensitivity matrix $\mathbf{G}$,

$$\delta\mathbf{d} = \mathbf{G}\,\delta\mathbf{m}\,. \tag{1}$$

In the unrealistic case where $\mathbf{G}$ is invertible, $\delta\mathbf{m}$ is uniquely determined by $\delta\mathbf{d}$. The resolution matrix $\mathbf{R}$ is then equal to the unit matrix, meaning that variations in one parameter class do not affect the inferred variations in another parameter class. When $\mathbf{G}$ is not invertible, the solution of the tomographic system is often defined as the minimum of the quadratic misfit functional

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

where $\mathbf{C}_d$ and $\mathbf{C}_m$ are the data and model covariance matrices, respectively (e.g. Nolet 2008). The normal equations that result from forcing the derivative of (2) with respect to $\delta\mathbf{m}$ to zero are

$$\delta\mathbf{m} = [\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\,\delta\mathbf{d} = \mathbf{L}\,\delta\mathbf{d}\,, \tag{3}$$

with the pseudo-inverse $\mathbf{L} = [\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}$. For the simplest case of only two parameter classes denoted $\mathbf{m}_1$ and $\mathbf{m}_2$, we have

$$\mathbf{m} = \begin{pmatrix}\mathbf{m}_1\\ \mathbf{m}_2\end{pmatrix}, \qquad \mathbf{G} = (\mathbf{G}_1\;\mathbf{G}_2)\,, \tag{4}$$

and the pseudo-inverse takes the specific form

$$\begin{aligned}\mathbf{L} &= [\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}]^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\\ &= \begin{pmatrix}\mathbf{G}_1^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_1 + \mathbf{C}_{m_1}^{-1} & \mathbf{G}_1^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2\\ \mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_1 & \mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2 + \mathbf{C}_{m_2}^{-1}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{G}_1^{\mathrm{T}}\\ \mathbf{G}_2^{\mathrm{T}}\end{pmatrix}\mathbf{C}_d^{-1}\,.\end{aligned} \tag{5}$$

Note that in eq. (5) the model covariance matrix is assumed block-diagonal, which corresponds to the absence of interparameter trade-offs. Ideally, the prior model covariances $\mathbf{C}_{m_1}$ and $\mathbf{C}_{m_2}$ reflect our—to some degree subjective—*a priori* guess on the plausible variability within the model parameter classes $m_1$ and $m_2$, respectively. More often, however, $\mathbf{C}_{m_1}$ and $\mathbf{C}_{m_2}$ are chosen pragmatically, that is, such that the matrix $[\mathbf{G}^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}]$ in eqs (3) and (5) becomes invertible. This regularization in the form of unavoidably non-zero model covariances—whether idealistic or pragmatic—introduces off-diagonal terms in the resolution matrix $\mathbf{R} = \mathbf{L}\,\mathbf{G}$, explicitly given by

$$\begin{aligned}\mathbf{R} = &\begin{pmatrix}\mathbf{G}_1^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_1 + \mathbf{C}_{m_1}^{-1} & \mathbf{G}_1^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2\\ \mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_1 & \mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2 + \mathbf{C}_{m_2}^{-1}\end{pmatrix}^{-1}\\ &\times\begin{pmatrix}\mathbf{G}_1^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_1 & \mathbf{G}_1^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2\\ \mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_1 & \mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2\end{pmatrix}.\end{aligned} \tag{6}$$

Off-diagonal elements in $\mathbf{R}$ map variations in one parameter class into inferred variations of the other parameter class. Moreover, the regularization imposed on $\delta\mathbf{m}_1$ affects $\delta\mathbf{m}_2$, and vice versa, thereby introducing additional interdependences that have no physical basis.

To avoid unphysical mappings, for instance, from $\mathbf{m}_2$ to $\mathbf{m}_1$, the observables $\mathbf{d}$ should ideally be chosen such that the elements of $\mathbf{C}_d^{-1/2}\mathbf{G}_2$ are much smaller than $\mathbf{C}_{m_2}^{-1/2}$, while keeping $\mathbf{C}_d^{-1/2}\mathbf{G}_1$ at order of $\mathbf{C}_{m_1}^{-1/2}$ or larger. Provided that such observables are available, the resolution matrix condenses to

$$\mathbf{R} = \mathbf{L}\,\mathbf{G} \approx \begin{pmatrix}[\mathbf{G}_1^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_1 + \mathbf{C}_{m_1}^{-1}]^{-1}[\mathbf{G}_1^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_1] & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{pmatrix}, \tag{7}$$

meaning that variations in $\mathbf{m}_2$ have little effect on the inferred variations in $\mathbf{m}_1$. While illustrated here for the specific case of a linear least-squares problem, the phenomenon that necessary regularization induces interparameter trade-offs, is general and methodology-independent. The challenge is therefore to construct observables that have maximum sensitivity with respect to $\mathbf{m}_1$ and minimum sensitivity with respect to $\mathbf{m}_2$.

### 1.3 Optimal experimental design

Optimizing sensitivity to and resolution of specific model subspaces falls into the field of experimental design, for which an extensive body of literature exists. In geophysics, most of the research in experimental design focused on small-scale, exploration or engineering applications where the positions of sources and/or receivers is variable. Pioneering work was concerned with finding receiver configurations for locating earthquakes (e.g. Rabinowitz & Steinberg 1990; Hardt & Scherbaum 1994) and for ocean tomography (e.g. Barth & Wunsch 1990) such that resolution is maximized while keeping the financial and logistic effort as small as possible. Under the assumption that the relation between model parameters and data is nearly linear, tools from linear algebra can be used to find a suitable balance between benefits and costs, expressed in the form of a design objective function (e.g. Maurer & Boerner 1998; Curtis 1999a; Curtis *et al.* 2004; Coles & Morgan 2009). Methods applicable to general non-linear problems were developed, for instance, by van den Berg *et al.* (2003) and Winterfors & Curtis (2008). For a comprehensive review on optimal experimental design in geophysics, the reader is referred to Curtis (2004a,b) and Maurer *et al.* (2010).

A common characteristic of most experimental design methods is that the type of measurement is fixed while sources and/or receivers are assumed mobile. This is the case, for instance, in traveltime cross-hole tomography where traveltime measurements are made on seismograms recorded by seismometers at variable depths (e.g. Curtis 1999b). While being successful in exploration and engineering geophysics, such methods are not directly applicable to most regional- and global-scale problems where the ability to move sources or receivers is limited. For instance, the location of earthquakes that act as sources in tomographic studies is largely controlled by tectonic processes on which we have no influence. Furthermore, the positioning of seismic stations is often dictated by geographic or political circumstances.

The relative immobility of sources and receivers in regional- and global-scale studies requires a different approach to experimental design. Instead of moving sources and receivers while keeping the type of measurement fixed, we propose to design optimal measurements or observables based on a fixed source–receiver geometry.

### 1.4 Outline

In the following paragraphs, we develop and apply a method for the design of seismic observables that reduce interparameter trade-offs and mappings by having maximum sensitivity with respect to the parameter class of interest while being as insensitive as possible to all other parameter classes. These 'optimal observables' are defined as a linear combination of 'fundamental observables' that are generally sensitive to multiple parameter classes. Possible choices of fundamental observables include, but are not limited to, finite-frequency traveltimes and amplitudes of body and surface waves, different types of waveform misfit in selected time intervals or various measurements made on interstation correlations of deterministic or quasi-random wavefields. The optimal weights in the linear combination are determined in a two-stage optimization process that requires the sensitivity kernels of the fundamental observables.

This paper is organized as follows: In Section 2, we develop a general method that allows us to design optimal observables on the basis of a collection of fundamental observables with less favourable properties. The theory will be followed by a series of examples where we design observables with maximum sensitivity to density (Section 3.1), and maximum sensitivity in the near-receiver structure (3.2). These examples are primarily intended to illustrate the methodology, and to reveal its potentials and limitations. Finally, in Section 4, we discuss various details of our method, including pre-requisites for the successful construction of optimal observables, the role of prior model covariance and parameter scaling, as well as the use of optimal observables in iterative multiparameter inversions.

In order to be independent of a specific inversion setup or discretization, we will mostly work in the continuous space domain. The discrete versions of model parameters and sensitivity kernels can be obtained by projection onto an application-specific set of basis functions.

## 2 THEORY

### 2.1 Constructing optimal observables

We start our development with a set of fundamental observables $d_1, \ldots, d_n$, and a set of continuously distributed parameter classes $m_1(\mathbf{x}), \ldots, m_p(\mathbf{x})$, where $\mathbf{x}$ denotes the position vector. Fundamental observables can be, for instance, traveltimes in various frequency bands; and possible parameter classes include $P$ velocity, $S$ velocity and attenuation. Without loss of generality, we assume that the parameter class of interest corresponds to $m_1$. Our goal is to find the linear combination of fundamental observables

$$d = \sum_{i=1}^{n} w_i d_i , \tag{8}$$

that has maximum sensitivity to parameter class $m_1$, while having minimum sensitivity to all other parameter classes $m_2, \ldots, m_p$. To ensure well-posedness of this optimization problem, we constrain the squared sum of the weights to equal 1, that is,

$$\sum_{i=1}^{n} w_i^2 = 1 . \tag{9}$$

Our task is to identify the appropriate vector of weighting coefficients $\mathbf{w} = (w_1, \ldots, w_n)^{\mathrm{T}}$. For this, we write the variation of the combined observable $d$,

$$\delta d = \sum_{i=1}^{n} w_i \delta d_i , \tag{10}$$

in terms of the Fréchet or sensitivity kernels $K_{ij}$ of the fundamental observables, defined as

$$\delta d_i = \sum_{j=1}^{p} \int_V K_{ij}(\mathbf{x})\, \delta m_j(\mathbf{x})\, d^3\mathbf{x} , \tag{11}$$

with the earth model volume denoted by $V$. Eq. (11) is the space-continuous version of eq. (1), with the sensitivity kernel $K_{ij}$ playing

the role of the sensitivity matrix **G**. Inserting (11) into (10) gives

$$\delta d = \sum_{j=1}^{p} \int_V K_j(\mathbf{x})\, \delta m_j(\mathbf{x})\, d^3\mathbf{x}\,, \tag{12}$$

where the sensitivity $K_j$ of the combined observable with respect to parameter class $m_j$ is simply the weighted sum of the fundamental sensitivities $K_{ij}$,

$$K_j(\mathbf{x}) = \sum_{i=1}^{n} w_i K_{ij}(\mathbf{x})\,. \tag{13}$$

For $d$ to be optimal, we require that the sensitivity power (Sieminski *et al.* 2009)

$$P_j(\mathbf{w}) = \int_V K_j^2(\mathbf{x})\, d^3\mathbf{x} \tag{14}$$

is large for $j = 1$ and small for $j > 1$. We can cast this simultaneous maximization and minimization of sensitivity powers, into a single maximization problem by defining the objective functional

$$\mathbb{P}(\mathbf{w}) = \sum_{j=1}^{p} b_j P_j(\mathbf{w})\,, \tag{15}$$

with balancing coefficients $b_1 > 0$ and $b_{j>0} < 0$. The balancing coefficients control the extent to which the sensitivity power $P_1$ is maximized while the sensitivity powers $P_{j>1}$ are minimized. Choosing, for instance, $b_1$ large relative to the $b_{j>1}$ will result in a combined observable $d$ that has large sensitivity with respect to $m_1$, but the sensitivities with respect to the remaining parameter classes $m_{j>1}$ may also still be large. On the other hand, choosing $b_1$ small relative to the $b_{j>1}$ will only minimize sensitivity with respect to the parameter classes $m_{j>1}$ without necessarily maximizing sensitivity with respect to $m_1$. Thus, appropriate balancing is needed, as we will explain in the following paragraphs.

Determining optimal weights $w_i$ and balancing coefficients $b_j$ simultaneously, is a non-linear problem that can be computationally expensive. To keep the computational burden at a manageable level, we developed a two-stage procedure for the solution of the optimization problem: Stage 1: We consider the balancing coefficients $b_j$ constant, and we try to find weights $w_i$ that maximize (15) subject to the constraint (9). The Lagrange function for this constrained optimization problem is

$$\mathcal{L}(\mathbf{w}, \lambda) = \sum_{j=1}^{p} b_j P_j(\mathbf{w}) - \lambda \left( \sum_{i=1}^{n} w_i^2 - 1 \right)\,. \tag{16}$$

Differentiating eq. (16) with respect to **w**, and using (13) and (14), yields

$$\frac{\partial \mathcal{L}}{\partial \mathrm{w}_l} = \sum_{j=1}^{p} b_j \int_V K_{lj}(\mathbf{x}) \left[ \sum_{i=1}^{n} \mathrm{w}_i K_{ij}(\mathbf{x}) \right] d^3\mathbf{x} - \lambda \mathrm{w}_l\,, \tag{17}$$

with $l = 1, \ldots, n$. Setting the left-hand side of (17) to zero, and defining the matrix

$$M_{li} := \sum_{j=1}^{p} b_j \int_V K_{lj}(\mathbf{x}) K_{ij}(\mathbf{x})\, d^3\mathbf{x} \tag{18}$$

with $l, i = 1, \ldots, n$, condenses the optimization problem into the eigenvalue equation

$$\mathbf{M}\mathbf{w} = \lambda \mathbf{w}. \tag{19}$$

All eigenvalues $\lambda$ are real because **M** is symmetric. Furthermore, we have $\mathbb{P}(\mathbf{w}) = \lambda$ for $\frac{\partial \mathcal{L}}{\partial \mathrm{w}_l} = 0$ (see Appendix A). Consequently, the solution of the maximization problem is the weight eigenvector **w** that corresponds to the maximum eigenvalue $\lambda_{\max}$ of **M**. The size of **M** is controlled by the number of fundamental observables, which is typically small (in fact, around five in the examples shown later). It follows that the computational costs for solving the eigenvalue problem are generally low. Stage 2: So far, the solution vector **w** of the maximization problem still depends on the balancing vector $\mathbf{b} = (b_1, \ldots, b_p)^{\mathrm{T}}$ that we assumed fixed in eq. (16), that is, $\mathbf{w} = \mathbf{w}(\mathbf{b})$. As an additional optimization step, the balancing coefficients may be further tuned to improve the properties of the combined observable $d$. For this, we draw test balancing vectors $\mathbf{b}_{\mathrm{test}}$ and evaluate the criterion

$$\frac{P_1[\mathbf{w}(\mathbf{b}_{\mathrm{test}})]}{\prod_{j=2}^{p} Pj[\mathbf{w}(\mathbf{b}_{\mathrm{test}})]} = \max\,, \tag{20}$$

which only involves the solution of the small eigenvalue problem (19). We then choose the balancing vector for which (20) attains its global maximum. In the applications of Section 3, we generate test balancing vectors using a regular grid search.

We note that other criteria for finding optimal balancing coefficients may be used, depending on specific requirements and applications. Eq. (20) proved useful for our purposes.

### 2.2 Parameter classes

The method developed in Section 2.1 is very general, and does not pose any restrictions on what we define as a parameter class. In addition to physical properties such as $P$ velocity and $S$ velocity, parameter classes can also be defined as physical properties in specific regions of the Earth. In this sense, $m_1$ may, for instance, be the $S$ velocity in the D" region, and $m_2$ the $S$ velocity in the remaining mantle and the crust. Using a suitable set of fundamental observables, our method may then be used to design observables with primary sensitivity to D" structure. An example where sensitivity is focused into a specific region, is shown in Section 3.2.

## 3 EXAMPLES

To illustrate the construction of optimal observables, we provide various examples in a regional tomography setting. In Section 3.1, we focus on traveltime measurements used to construct combined observables with primary sensitivity to 3-D density structure. This will be followed, in Section 3.2, by the combination of rotational and translational ground motion measurements into an observable that is sensitive only to near-receiver structure.

For our experiments we compute synthetic wavefields for a point source at 13.9 km depth beneath southern Greece. These are recorded at 20.98° epicentral distance in Malaga, Spain, as illustrated in Fig. 1 (left-hand panel). In the interest of reproducibility, we use the 1-D earth model ak135 (Kennett *et al.* 1995), and the source parameters correspond to the $M_{\mathrm{w}}5.5$ event in the southern Greece region on 2010 January 18 taken from the Global CMT catalogue (www.globalcmt.org). Velocity seismograms for the vertical component and frequency bands from 60 to 90 s (solid line) and 90 to 130 s (dashed line) are plotted in the right-hand panel of Fig. 1.

### 3.1 Combining surface wave traveltimes to enhance sensitivity to 3-D density variations

Trade-offs between parameter classes characterize inversions for 3-D density structure. While the Earth's mass, moment of inertia and

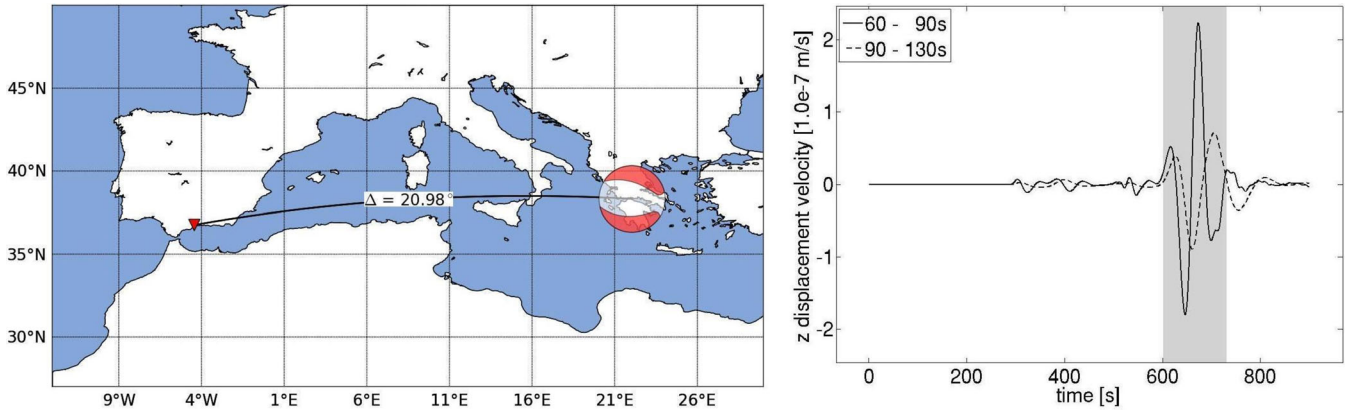


**Figure 1.** Left-hand side: source–receiver geometry for an event beneath southern Greece (longitude: 22.05°E, latitude: 38.32°N, depth: 13.9 km) and a receiver located in Malaga, Spain (longitude: 4.43°W, latitude: 36.76°N, depth: 0.0 km). The epicentral distance is 20.98°. The source mechanism is plotted at the epicentre. Right-hand side: vertical-component seismograms of the displacement velocity. The frequency bands are 60–90 s (solid line) and 90–130 s (dashed line). The Rayleigh-wave window is marked by a grey bar.

gravest normal modes constrain radial density structure to within ∼1 per cent over 200-km-depth intervals (Kennett 1998), 3-D density remains poorly known. Deterministic inversions are frequently biased by convenience-driven assumptions, including the neglect of compositional heterogeneities (e.g. Simmons *et al.* 2010) and imposed correlations between velocities and density (Tondi *et al.* 2000, 2009). Joint deterministic inversions of gravity and normal-mode data yield geodynamically plausible density heterogeneities for the lower mantle (Ishii & Tromp 1999, 2001), but their reliability remains debated mostly because sensitivity to density is small and resolution estimates are biased by regularization (e.g. Resovsky & Ritzwoller 1999; Romanowicz 2001; Kuo & Romanowicz 2002). Probabilistic inversions for 3-D density avoid regularization (e.g. Resovsky & Trampert 2003; Mosca *et al.* 2012) but are limited to smaller numbers of model parameters.

A central problem in density tomography is that the sensitivity of traveltime measurements to density variations tends to be small, oscillatory and parametrization-dependent. In isotropic models parametrized in terms of *P*-velocity $\alpha$, *S*-velocity $\beta$ and density $\rho$, the traveltimes of body waves are nearly insensitive to density due to backward scattering (Wu & Aki 1985; Tarantola 1986). While not being close to zero, the sensitivities of surface wave traveltimes are vertically oscillating, meaning that positive and negative contributions of large-scale perturbations tend to cancel (e.g. Takeuchi & Saito 1972). The alternative parametrization in terms of bulk modulus $\kappa$, shear modulus $\mu$ and density, leads to large sensitvities of body- and surface wave traveltimes to density, but also to strong interparameter trade-offs.

In the following paragraphs, we try to construct combinations of surface wave traveltime measurements with dominant sensitivity to 3-D density variations. As fundamental observables we use cross-correlation time-shifts of vertical-component surface waves in the period bands 30–40 s, 40–60 s, 60–90 s and 90–130 s, that we denote by $d_1 = d_{30-40}$, $d_2 = d_{40-60}$, $d_3 = d_{60-90}$ and $d_4 = d_{90-130}$. Thus, the number of fundamental observables is $n = 4$. The measurement time window is indicated by grey shading in the synthetic seismograms of Fig. 1. Sensitivity kernels needed in the optimization scheme are computed with the help of adjoint techniques (e.g. Tarantola 1988; Tromp *et al.* 2005; Fichtner *et al.* 2006; Liu & Tromp 2008; Chen 2011). Under the assumption that observed and synthetic waveforms are sufficiently similar, the sensitivity kernels for cross-correlation time-shift measurements are independent of observed data (Luo & Schuster 1991), meaning that we can construct optimal observables without requiring actual data.

Knowing that the inversion for 3-D density is a major challenge in seismic tomography, our goals are modest and mostly centred around a proof of the algorithmic concept introduced in Section 2. Density is a difficult parameter that is well suited to illustrate the functioning of our method, as well as potential pitfalls and limitations. In a first numerical experiment, we construct an optimal observable with minimum sensitivity to $\alpha$ and $\beta$, and maximum sensitivity to the vertical density gradient $\partial_r \rho$ that can be related to the pressure derivative of density, observable in mineral physics experiments. This will be followed by a similar, though more difficult, attempt to find a optimal observable for density when the earth model is parametrized in terms of the bulk modulus $\kappa$, the shear modulus $\mu$ and density. In Section 3.1.3 we extend the analysis of a single source–receiver pair to a realistic configuration with multiple sources and receivers.

### *3.1.1 Optimal observable for the vertical density gradient $\partial_r \rho$*

To construct optimal observables for the vertical density gradient $\partial_r \rho$, we start with the computation of sensitivity kernels for the relative perturbations $\delta \ln \alpha$, $\delta \ln \beta$ and $\delta \ln \rho$, where the reference values are taken from the 1-D earth model ak135 (Kennett *et al.* 1995). The kernels are displayed in the upper three rows of Fig. 2 for the frequency band from 60 to 90 s. The left-hand panels show vertical slices through the source–receiver plane at 37.5°N. The right-hand panel contains vertical slices at 9.0°E. While sensitivity with respect to $\delta \ln \alpha$ is non-zero only within the upper ∼30 km, sensitivity to $\delta \ln \beta$ reaches its maximum around 80 km depth. The density kernel is characterized by a sign change around 50 km depth. It follows that density variations extending much more than 50 km in depth hardly affect 60 to 90 s surface wave traveltimes because positive and negative contributions tend to cancel. This characteristic is nearly period-independent.

As we demonstrate in Appendix B, the kernel for the vertical density gradient, $K_{\partial_r \rho}$, can be computed from $K_\rho$ via a simple integral over radius from the centre of the Earth ($r = 0$) to the surface ($r = R$):

$$K_{\partial_r \rho}(\theta, \phi, r) = - \int_{r'=0}^{r} K_\rho(\theta, \phi, r') \, \mathrm{d}r' \, . \tag{21}$$

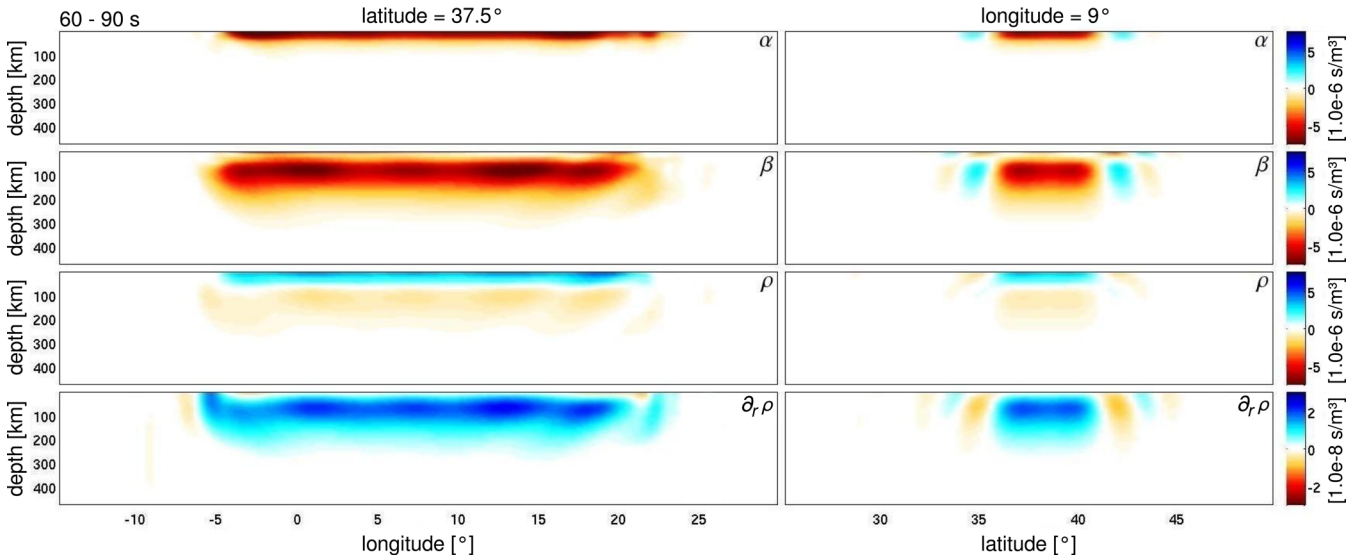


**Figure 2.** Vertical slices through the sensitivity kernels of a 60–90 s surface wave traveltime with respect to the relative variations $\delta\ln\alpha$ (1st row), $\delta\ln\beta$ (2nd row) and $\delta\ln\rho$ (3rd row). The 4th row displays slices through the relative density gradient kernel, that is, the kernel corresponding to $\delta\ln\partial_r\rho$.

Kernels for relative perturbations in the vertical density gradient, $\delta\ln\partial_r\rho$, are obtained from $K_{\partial_r\rho}$ by multiplication with a reference value for $\partial_r\rho$. In the case of the 1-D reference earth model ak135 (Kennett *et al.* 1995), this reference takes the nearly constant value of $6.05 \cdot 10^{-4}$ kg m$^{-4}$ from 35 to 410 km depth. Vertical slices through the sensitivity kernel for $\delta\ln\partial_r\rho$ are shown in the fourth row of Fig. 2. Unlike $K_\rho$, the density gradient kernel $K_{\partial_r\rho}$ does not change sign with depth. It follows that variations of $\partial_r\rho$ extending over more than ∼50 km in depth will affect the traveltime observations. Equipped with sensitivity kernels for $\delta\ln\alpha$, $\delta\ln\beta$ and $\delta\ln\partial_r\rho$ in the individual period bands, we can now use the machinery developed in Section 2 to construct a composite observable that is optimally sensitive to the vertical density gradient. The resulting optimal weights are $w_1 = -0.11$, $w_2 = 0.41$, $w_3 = -0.73$ and $w_4 = 0.54$, meaning that, according to eq. (8), the optimal observable takes the form

$$d = -0.11\, d_{30-40} + 0.41\, d_{40-60} - 0.73\, d_{60-90} + 0.54\, d_{90-130}. \tag{22}$$

Sensitivity kernels corresponding to this optimal observable are displayed in Fig. 3. Compared to the 60–90 s period band kernels shown in Fig. 2, *P*-velocity sensitivity is almost completely eliminated, and the maxima of *S*-velocity sensitivity is reduced by ∼50 per cent. Furthermore, the *S*-velocity kernel of the optimal observable is oscillatory, meaning that large-scale variations in *S* velocity have an even smaller effect. At the same time, the density gradient kernel maintains the same amplitude as within the individual period bands.

A condensed version of the sensitivity kernels for the optimal observable is shown in Fig. 4 in the form of horizontally averaged kernels that emphasize their radial structure. While the sensitivity of the optimal observable relative to $\partial_r\rho$ consists of a single lobe with amplitude similar to the individual period band kernels, the sensitivities to $\alpha$ and $\beta$ differ significantly from their individual contributions. Most importantly, and as seen in the vertical profiles of Fig. 3, sensitivity to $\alpha$ is eliminated almost completely, meaning that the original three-parameter inverse problem has reduced to a two-parameter problem that does not suffer from trade-offs with shallow *P*-velocity structure.

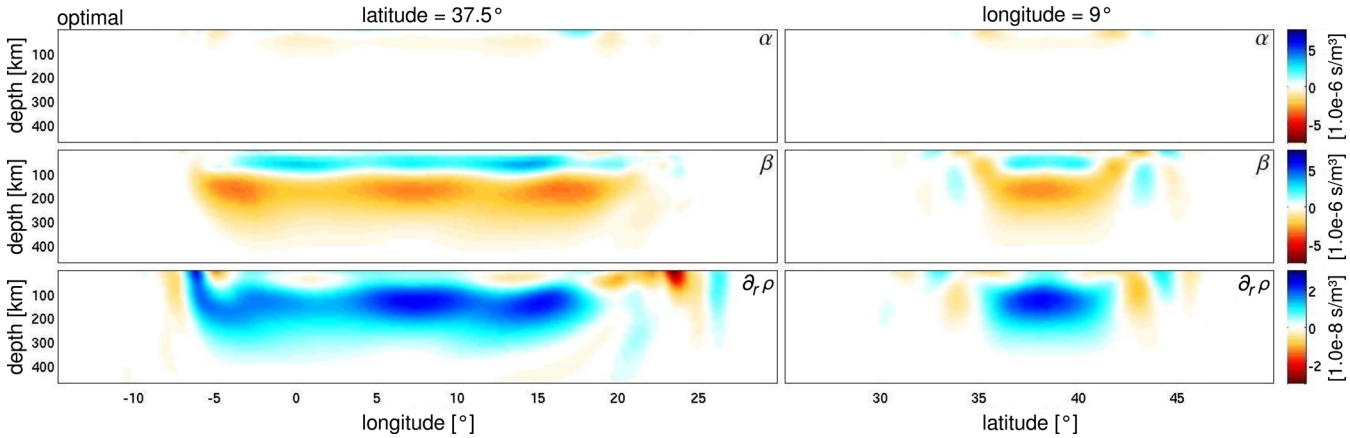


**Figure 3.** Vertical slices through the sensitivity kernels of the optimal observable $d$ (eq. 22) with respect to the relative variations $\delta\ln\alpha$ (1st row), $\delta\ln\beta$ (2nd row) and $\delta\ln\partial_r\rho$ (3rd row). Sensitivity to *P* velocity is nearly eliminated, and sensitivity to *S* velocity is significantly reduced, relative to kernels for the individual period bands.

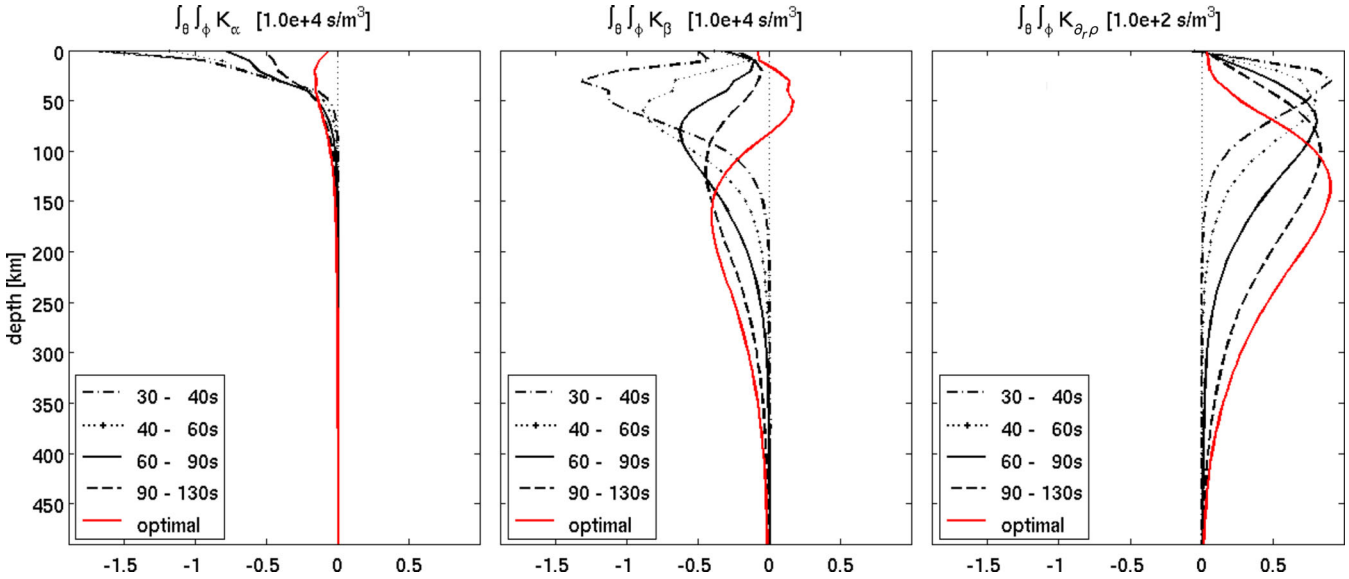


**Figure 4.** Horizontally averaged sensitivity kernels with respect to $\alpha$ (left panel), $\beta$ (centre panel) and $\partial_r \rho$ (right panel). Kernels for the individual period bands are plotted in black. Red curves represent the kernels corresponding to the optimal observable. Sensitivity of the optimal observable to $\alpha$ is forced to almost zero, meaning that the original three-parameter problem has reduced to a two-parameter problem that is nearly independent of $P$ velocity.

This first experiment indicates that the optimization algorithm proposed in Section 2 indeed produces observables that have increased sensitivity with respect to the target parameter, while having reduced sensitivity with respect to the remaining parameters. Remarkable aspects of this experiment are that relative sensitivity to density structure can be enhanced, and sensitivity to $P$-velocity structure can be nearly eliminated by combining only a small number of Rayleigh-wave traveltime measurements that by themselves have weak sensitivity to density and strong sensitivity to $P$ velocity within the crust.

### *3.1.2 Changing parametrization: ($\alpha$, $\beta$, $\rho$) → ($\kappa$, $\mu$, $\rho$)*

A complicating aspect of density tomography is the strong dependence of the density kernel on the choice of free parameters that can have profound effects on the result of deterministic regularized inversions (Cara *et al.* 1984). The effect of changing the parametrization from $\alpha$, $\beta$ and $\rho$ to the bulk modulus $\kappa$, shear modulus $\mu$ and density, is visualized in Fig. 5 for the period band 60–90 s. While sensitivity with respect to $\kappa$ is non-zero only in the upper ∼30 km below the surface, sensitivity with respect to $\mu$ reaches its maximum at depths between 50 and 100 km and tends to zero towards the surface. The density kernel is characterized by an opposite sign relative to the $\kappa$ and $\mu$ kernels, but there is no sign change with depth as in the $\alpha$, $\beta$, $\rho$ parametrization. Solving the optimization problem produces the weights $w_1 = -0.28$, $w_2 = 0.6$, $w_3 = -0.65$ and $w_4 = 0.37$ that we use to construct the optimal composite observable. Vertical slices through and horizontal averages of the resulting sensitivity kernels are displayed in Figs 6 and 7. Similar to the $P$-velocity $\alpha$ in the previous example, sensitivity to $\kappa$ is strongly reduced, meaning that the optimal observable is effectively sensitive to two parameters only. Sensitivity to $\mu$, however, remains comparatively large. The failure to strongly reduce shear modulus sensitivity results from the nearly period-independent geometry of the density kernels, the horizontal averages of which

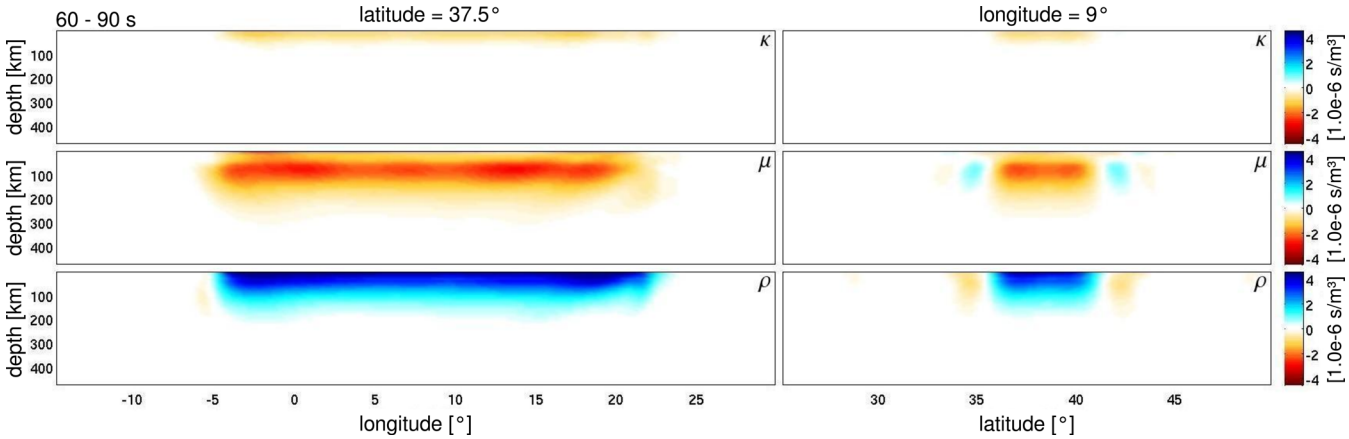


**Figure 5.** Cross-correlation time-shift kernels with respect to the bulk modulus $\kappa$ (first row), the shear modulus $\mu$ (second row) and density $\rho$ (third row) for the frequency band 60–90 s. The vertical slices are shown at 37.5°N (left-hand column) and 9.0°E (right-hand column). The similarly shaped sensitivity kernels for $\mu$ and $\rho$ indicate that strong trade-offs exist between both parameters.

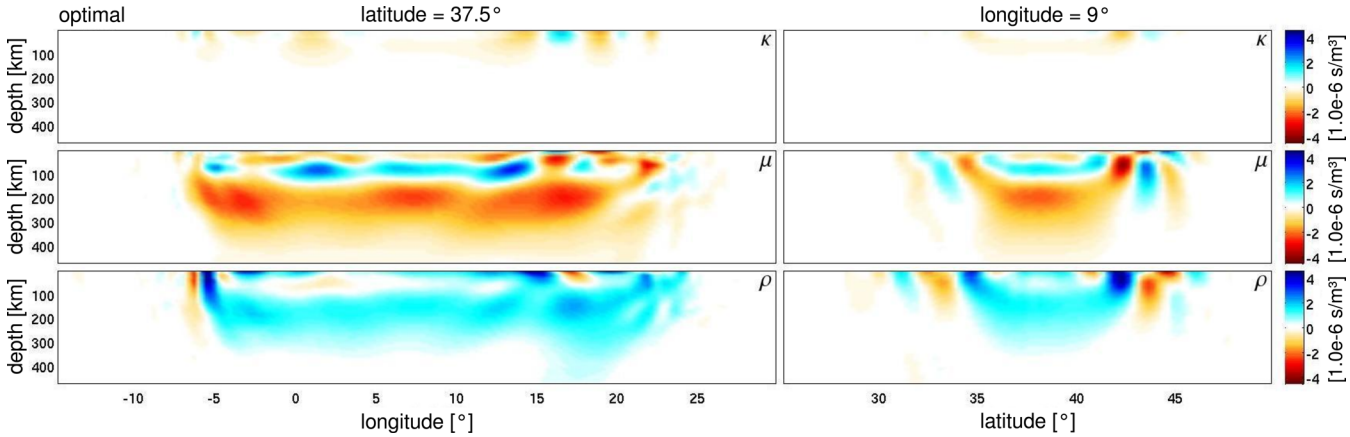


**Figure 6.** Vertical slices through sensitivity kernels for the optimal observable in the $\kappa$, $\mu$, $\rho$ parametrization. While the sensitivity to $\kappa$ is negligibly small, sensitivity to $\mu$ remains large. The comparatively small reduction of sensitivity to $\mu$ results from the small variability in the geometry of the individual density kernels for different period bands, shown in the right-hand panel of Fig. 7. The main effect of the optimization is to introduce additional sign changes with depth in the sensitivity to $\mu$.

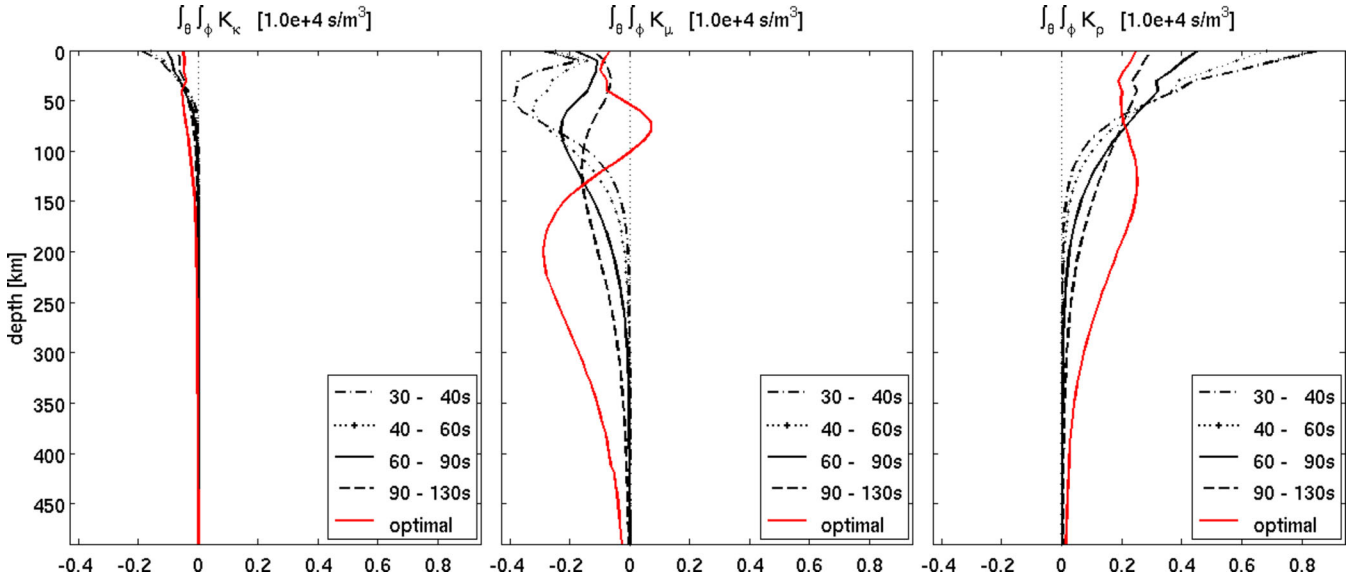


**Figure 7.** Horizontally averaged sensitivity kernels with respect to $\kappa$ (left panel), $\mu$ (centre panel) and $\rho$ (right panel). Kernels for the individual period bands are plotted in black. Red curves represent the kernels corresponding to the optimal observable. Sensitivity of the optimal observable to $\alpha$ is forced to almost zero. The comparatively small variability in the geometry of the individual density kernels for different period bands limits the ability to find an optimal observable with a density sensitivity that is significantly increased relative to the shear modulus sensitivity.

are shown in the right-hand panel of Fig. 7. The small geometric variability reduces the ability to find linear combinations with fundamentally new properties, thereby illustrating a limitation of our method that we further discuss in Section 4.1.

One of the main effects of the optimization algorithm is to introduce additional sign changes with depth in the sensitivity to $\mu$. These sign changes reduce the impact of large-scale variations in $\mu$ on the optimal observable. In the following section we will further investigate this effect in a scenario with multiple sources and receivers.

### *3.1.3 A synthetic tomography scenario*

In this section, we apply the construction of optimal observables to a configuration with multiple sources and receivers, as shown in Fig. 8. This experiment is intended to serve several purposes: (1) test the applicability of the optimization algorithm to multiple source–receiver pairs, (2) check the reproducibility of the weighting coefficients $w_i$ found for a single source–receiver pair in the previous example and (3) test with actual measurements if density truly becomes the dominant parameter.

Our scenario includes 10 shallow events in the Mediterranean region, with epicentres taken from the Global CMT catalogue (www.globalcmt.org). The positions of the 43 stations in the experiment correspond to station locations in the seismic networks IberArray (http://iberarray.ictja.csic.es) and ISIDe (http://iside.rm.ingv.it/iside). This distribution of sources and receivers ensures a dense coverage of the Mediterranean region except for the southeastern part.

For the construction of optimal observables, we again consider cross-correlation time-shifts of vertical-component Rayleigh waves for the frequency bands 30–40 s, 40–60 s and 90–130 s. We furthermore impose that the optimal weights $w_i$ be the same for each

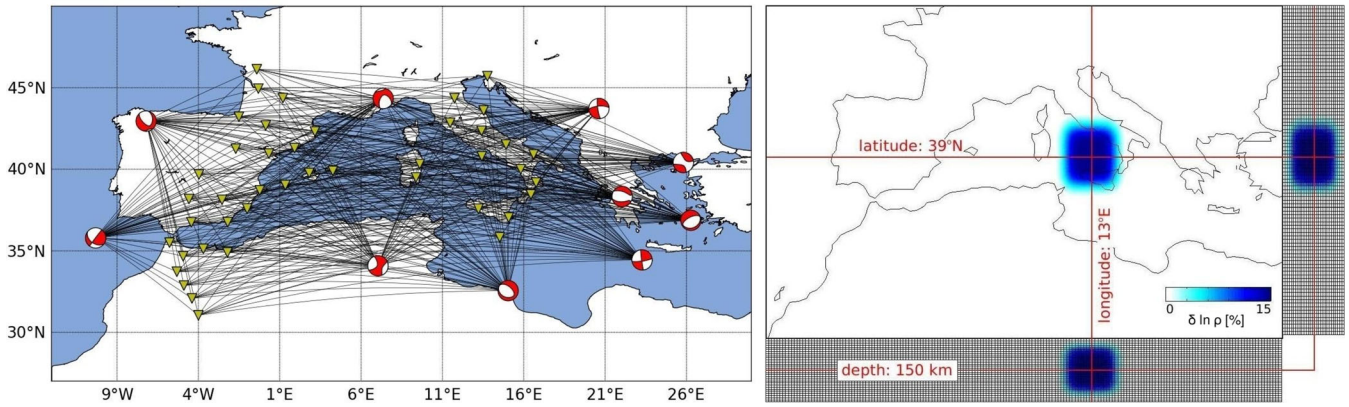


**Figure 8.** Left-hand side: source/receiver geometry of the synthetic tomography scenario. 10 seismic events visualized by their corresponding source mechanisms encircle the Mediterranean Sea. The 43 receiver locations are marked by yellow triangles. Great circle paths are represented by black lines between each source/receiver pair. Right-hand side: density perturbation of +15 per cent centred in the Tyrrhenian Sea at 13°E, 39°N and 150 km depth.

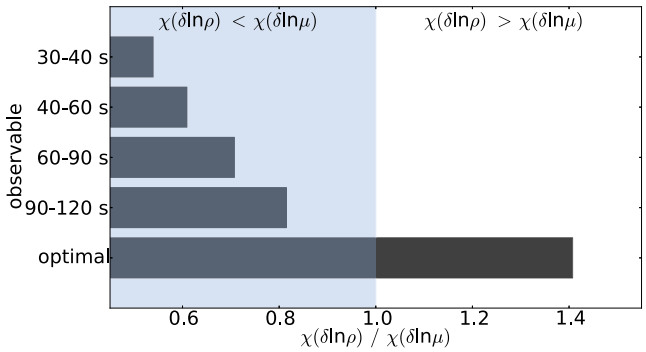


**Figure 9.** Comparison of cumulative time delays resulting from the density heterogeneity ($\chi(\delta\ln\rho)$) and the shear modulus heterogeneity ($\chi(\delta\ln\mu)$) shown in Fig. 8. The effect of the density heterogeneity is clearly dominant for the optimal observable.

receiver and event. For this scenario, the optimization algorithm reproduces exactly the same weights as in the single source–receiver case of Section 3.1.2. This result is plausible, at least for a radially symmetric background model where surface wave sensitivity kernels have translational invariance and depth profiles that are nearly independent of epicentral distance. Differences in the source mechanisms play a minor role. To test if the optimal observable indeed reacts primarily to density perturbations, we add heterogeneities to the 1-D model, centred at 13°E, 39°N and 150 km depth (see right-hand panel of Fig. 8). In a first simulation we computed vertical-component seismograms for the individual frequency bands and a +15 per cent density perturbation. For realistic earth models, this value may be exaggerated, but it ensures that the traveltime differences between the 1-D model and its perturbed version are large and free of discretization errors. In a second simulation we replaced the density perturbation by a −15 per cent shear modulus perturbation. Following these simulations, we compute the cumulative rms time delays $\chi$ observed for all source–receiver pairs for both types of perturbations; and for the fundamental, as well as for the optimal observable.

The result is displayed in Fig. 9 in the form of relative time delays for density and shear modulus perturbations, $\chi(\delta\ln\rho)/\chi(\delta\ln\mu)$. Within the individual period bands, the time delays produced by the shear modulus heterogeneity is generally larger than the time delays produced by the density heterogeneity of comparable size. For the optimal observable, however, time delays due to the density heterogeneity are more than 40 per cent larger, supporting the results of the sensitivity analysis of Section 3.1.2 that already suggested the dominance of density structure. This result furthermore confirms that the optimization algorithm produces the desired results, and indeed provides observables that mostly react to changes in the target parameter class.

### 3.2 Sensitivity optimization involving translational and rotational ground motion measurements

So far, we were concerned with the optimization of sensitivities with respect to a set of structural parameters. In the following application we demonstrate that our algorithm remains applicable when the parameter classes are not structural parameters themselves, but structural parameters within certain target regions of the Earth. We consider the apparent *S*-velocity $\beta_a$, defined at the receiver location $\mathbf{x}^r$ as the ratio of rms translational velocity and rotation amplitudes of a seismic recording (e.g. Fichtner & Igel 2009):

$$\beta_a(\mathbf{x}^r) := \frac{1}{2}\frac{\|\mathbf{v}(\mathbf{x}^r)\|_2}{\|\boldsymbol{\omega}(\mathbf{x}^r)\|_2}\,. \tag{23}$$

Quantities on the right-hand side of eq. (23) denote the velocity amplitude $\|\mathbf{v}(\mathbf{x}^r)\|_2^2 = \int \mathbf{v}^2(\mathbf{x}^r, t)\,\mathrm{d}t$ and the rotation amplitude $\|\boldsymbol{\omega}(\mathbf{x}^r)\|_2^2 = \int (\frac{1}{2}\nabla\times\mathbf{u})^2(\mathbf{x}^r, t)\,\mathrm{d}t$. The quotient rule of differentiation implies that the sensitivity of $\beta_a$ is equal to the difference between the sensitivities of $\|\mathbf{v}(\mathbf{x}^r)\|_2$ and $\|\boldsymbol{\omega}(\mathbf{x}^r)\|_2$ :

$$K_\beta(\beta_a) = K_\beta(\|\mathbf{v}(\mathbf{x}^r)\|_2) - K_\beta(\|\boldsymbol{\omega}(\mathbf{x}^r)\|_2). \tag{24}$$

The key property of $K_\beta(\beta_a)$ is that sensitivity is concentrated in the immediate vicinity of the receiver, and tends to zero towards the source (Fichtner & Igel 2009; Bernauer *et al.* 2012). This property suggests a variant of local tomography with teleseismic data that is fully independent of both traveltime measurements and deep Earth structure far from the local target region (Bernauer *et al.* 2009).

Our aim is to reproduce the localization of sensitivity near the receiver by constructing an optimal observable, instead of using the definition of the apparent *S*-velocity $\beta_a$ from eq. (23). For this, we define $d_v = \|\mathbf{v}\|_2$ and $d_\omega = \|\boldsymbol{\omega}\|_2$ as our fundamental observables. As first parameter class, we choose the *S*-velocity $\beta$ within a region $R_1$ that extends 500 km around the receiver. The second parameter class is the *S* velocity in the remaining part of the model, denoted by $R_2$. With the optimization scheme developed in Section 2, we

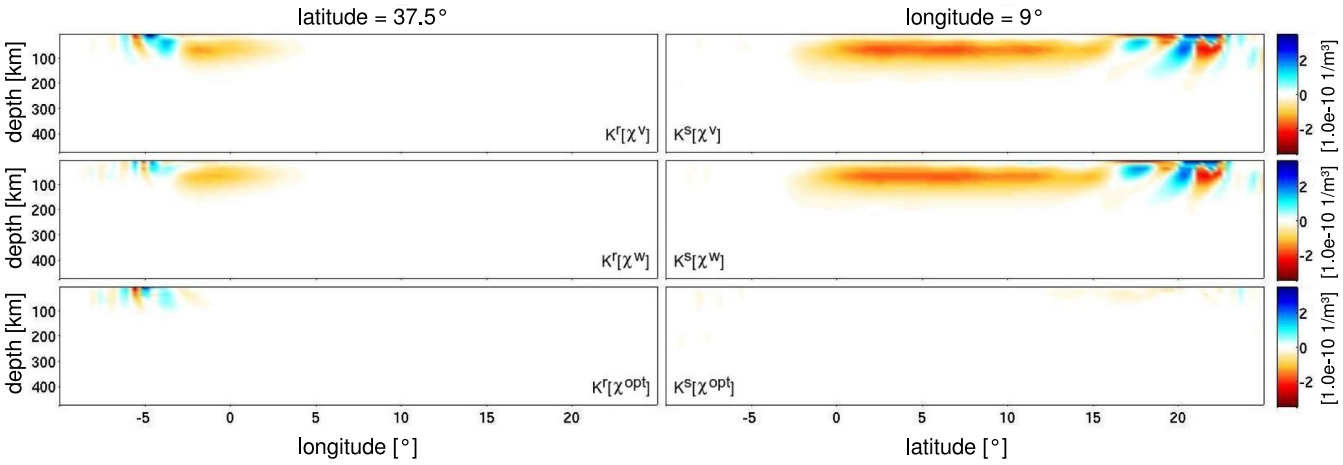


**Figure 10.** Vertical slices at 37.5°N through the sensitivity kernels $G\,K_\beta(\|\mathbf{v}\|)$ (top left-hand side), $G\,K_\beta(\|\boldsymbol{\omega}\|)$ (centre left-hand side), $(1-G)\,K_\beta(\|\mathbf{v}\|)$ (top right-hand side) and $(1-G)\,K_\beta(\|\boldsymbol{\omega}\|)$ (centre right-hand side). The function $G$ is a Gaussian function that is centred at the receiver position and has a half-width of 500 km. The source–receiver geometry and the source characteristics are the same as in Fig. 1. The third row displays the kernels for the optimal observable $d = 0.71\,(d_v - d_\omega)$. Shear wave sensitivity is entirely concentrated in the receiver region, and zero at distances of more than a few hundred kilometres from the receiver.

construct an optimal observable $d = w_v d_v + w_\omega d_\omega$ with maximum sensitivity in $R_1$ and minimum sensitivity in $R_2$.

In the practical implementation, we write the variation of $d_v$ as

$$\delta d_v = \int_V G\,K_\beta(\|\mathbf{v}\|)\,\delta\beta\,d^3\mathbf{x} + \int_V (1-G)\,K_\beta(\|\mathbf{v}\|)\,\delta\beta\,d^3\mathbf{x}\,, \quad (25)$$

where $G$ is a Gaussian function that is centred at the receiver position and has a half-width of 500 km. Similarly, for the variation of $d_\omega$, we have

$$\delta d_\omega = \int_V G\,K_\beta(\|\boldsymbol{\omega}\|)\,\delta\beta\,d^3\mathbf{x} + \int_V (1-G)\,K_\beta(\|\boldsymbol{\omega}\|)\,\delta\beta\,d^3\mathbf{x}\,. \quad (26)$$

Eqs (25) and (26) are special cases of eq. (11) in the theoretical developments, and they imply that the four kernels entering the optimization scheme are $G\,K_\beta(\|\mathbf{v}\|)$, $G\,K_\beta(\|\boldsymbol{\omega}\|)$, $(1-G)\,K_\beta(\|\mathbf{v}\|)$ and $(1-G)\,K_\beta(\|\boldsymbol{\omega}\|)$. These kernels are displayed in the first two rows of Fig. 10. All kernels were computed with the same setup as in Section 3.1 but for a frequency band of 40–100 s.

Inserting the sensitivity kernels in the optimization algorithm provides the weighting coefficients $w_v = 0.71$ and $w_\omega = -0.71$, meaning that the observable with maximum sensitivity in the receiver region and minimum sensitivity elsewhere is just the difference of $d_v$ and $d_\omega$, that is,

$$d = 0.71\,(d_v - d_\omega)\,. \quad (27)$$

The factor $0.71 \approx \sqrt{1/2}$ results from the constraint that the squared sum of all weights be equal to 1 (see eq. 9). The sensitivity kernel corresponding to $d$ is given by

$$K_\beta(d) = 0.71\,\left[K_\beta(\|\mathbf{v}(\mathbf{x}^r)\|_2) - K_\beta(\|\boldsymbol{\omega}(\mathbf{x}^r)\|_2)\right] = 0.71\,K_\beta(\beta_a), \quad (28)$$

and is shown in the third column of Fig. 10. This result is remarkable for various reasons: First, the optimization scheme indeed succeeds to focus sensitivity in the receiver region while completely erasing sensitivity in any other part of the earth model. It follows that any shear velocity perturbation more than a few hundred kilometres from the receiver has no first-order effect on the optimal observable $d$. Second, we fully reproduce the sensitivity of the apparent $S$-velocity $\beta_a$, up to a constant that results from the construction procedure of the optimal observable. Third, the result illustrates that kernels cannot be linked uniquely to an observable. Both $\beta_a = d_v/d_\omega$ and $d = d_v - d_\omega$ have, as a consequence of the quotient rule, identical sensitivity kernels. Finally, we note that separating the complete kernels into low- and high-frequency components, instead of concentrating on different areas as in eqs (25) and (26), also leads to exactly the same linear combination as in eq. (27). The variations in $d_v$ and $d_\omega$ are expressed as

$$\delta d_v = \int_V F * K_\beta(\|\mathbf{v}\|)\,\delta\beta\,d^3\mathbf{x} + \int_V (1-F) * K_\beta(\|\mathbf{v}\|)\,\delta\beta\,d^3\mathbf{x}, \quad (29)$$

and

$$\delta d_\omega = \int_V F * K_\beta(\|\boldsymbol{\omega}\|)\,\delta\beta\,d^3\mathbf{x} + \int_V (1-F) * K_\beta(\|\boldsymbol{\omega}\|)\,\delta\beta\,d^3\mathbf{x}, \quad (30)$$

where $F$ is a spatial high-pass filter and $(1-F)$ the corresponding low-pass filter. This result shows that the sensitivity close to the receiver in fact corresponds to the low-wavenumber contributions of the sensitivity kernels for $d_v$ and $d_\omega$.

## 4 DISCUSSION

The series of examples shown in Section 3 illustrates the applicability of the proposed method, but also raises various questions that deserve a more detailed discussion. These include the circumstances under which useful optimal observables can be designed, the role of quasi-subjective prior covariances, actual multiparameter inversion schemes based on optimal observables and the relation of our method to experimental design and the Backus–Gilbert method.

### 4.1 Successfully designing optimal observables

The extent to which sensitivity with respect to the target parameter class can be maximized while minimizing sensitivity with respect to the remaining parameter classes depends critically on the shape of the sensitivity kernels for the fundamental observables. A large variability in the shape of the kernels generally favours the successful construction of an observable that is truly optimal in the sense of being sensitive to one parameter class only.

In the example of Section 3.1.1, where we considered the sensitivity of Rayleigh-wave traveltimes to the vertical density gradient $\partial_r\rho$, the shape of the individual kernels was strongly period-dependent. This dependence allowed us to reduce the sensitivity to $\alpha$ and $\beta$ while keeping the sensitivity to $\partial_r\rho$ at a high level.

We encountered a less favourable situation when changing the parametrization to $\kappa$, $\mu$ and $\rho$ in Section 3.1.2. In this parametrization, the geometry of the density kernels is nearly period-independent, which limits the ability to find linear combinations with fundamentally different properties.

Along these lines, it is clear that the success of our methodology relies on the choice of fundamental observables. For the purpose of illustration, we deliberately chose a small number of fundamental observables, the physics of which are well understood. These observables allowed us to display both the functioning and limitations of our method. For future applications, the range of fundamental observables should clearly be extended, to include, for instance, time- and frequency-dependent amplitudes or various array-like measurements like interstation correlations.

### 4.2 The role of prior model covariance and parameter scaling

The problem of interparameter trade-offs in regularized inversions has two components—one related to our prior assumptions on model covariance, and one related to the structural sensitivity of specific observables. The interplay of both components becomes most apparent in the generalized inverse, given in eq. (5) for the case of two parameter classes.

Normalizing, for instance, parameter class $\mathbf{m}_1$ by a very small prior standard variation $\sigma_1$, will scale the corresponding $\mathbf{G}_1$ to $\sigma_1\mathbf{G}_1$, meaning that the generalized inverse $\mathbf{L}$ effectively becomes

$$\mathbf{L} \approx \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & [\mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2 + \mathbf{I}] \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{G}_1^{\mathrm{T}} \\ \mathbf{G}_2^{\mathrm{T}} \end{pmatrix} \mathbf{C}_d^{-1}. \tag{31}$$

The corresponding resolution matrix would be

$$\mathbf{R} \approx \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & [\mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2 + \mathbf{I}]^{-1}[\mathbf{G}_2^{\mathrm{T}}\mathbf{C}_d^{-1}\mathbf{G}_2] \end{pmatrix}, \tag{32}$$

meaning that the prior assumption on model covariance causes $\delta\mathbf{m}_1$ to be completely unresolved. This illustrates the effect of the prior on perceived resolution in deterministic inversions; and there is no generic solution other than probabilistic inversions that are, however, not feasible for high-dimensional model spaces.

Our development focuses on the second component of interparameter trade-offs that is related to the choice of observable. Independent of any prior assumptions, we ask how much sensitivity with respect to one parameter can be increased relative to sensitivity with respect to another parameter by designing optimal observables. The usefulness of these optimal observables in an actual regularized deterministic inversion also depends on the choices of prior covariances for the various parameter classes. Prior knowledge on the variability of model parameters can be naturally incorporated into our approach by scaling model parameters by their prior standard deviations.

### 4.3 Iterative inversion for multiple parameter classes

To use optimally designed observables in a tomographic inversion, we propose to proceed as follows: Again using the case of two parameter classes for illustration, we first construct an optimal observable $\mathbf{d}_1$ for $\mathbf{m}_1$ and a second optimal observable $\mathbf{d}_2$ for $\mathbf{m}_2$. The linearized relation between observations and model parameters then takes the form

$$\begin{pmatrix} \delta\mathbf{d}_1 \\ \delta\mathbf{d}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{pmatrix} \begin{pmatrix} \delta\mathbf{m}_1 \\ \delta\mathbf{m}_2 \end{pmatrix}. \tag{33}$$

In the ideal, but hardly achievable, scenario where $\mathbf{d}_1$ is insensitive to $\mathbf{m}_2$ and $\mathbf{d}_2$ is insensitive to $\mathbf{m}_1$, the off-diagonal contributions $\mathbf{G}_{12}$ and $\mathbf{G}_{21}$ in eq. (33) cancel, and the two parameter classes are decoupled. Under more realistic circumstances, $\mathbf{G}_{12}$ and $\mathbf{G}_{21}$ will be non-zero but by construction smaller than the diagonal elements $\mathbf{G}_{11}$ and $\mathbf{G}_{22}$. To emphasize the dominance of the diagonal terms, we rewrite (33) as

$$\begin{aligned} \mathbf{G}_{11}\delta\mathbf{m}_1 &= \delta\mathbf{d}_1 - \mathbf{G}_{12}\delta\mathbf{m}_2, \\ \mathbf{G}_{22}\delta\mathbf{m}_2 &= \delta\mathbf{d}_2 - \mathbf{G}_{21}\delta\mathbf{m}_1. \end{aligned} \tag{34}$$

Following (Kennett & Sambridge 1998), the first iteration consists in solving a regularized version of

$$\begin{aligned} \mathbf{G}_{11}\delta\mathbf{m}_1^{(1)} &= \delta\mathbf{d}_1, \\ \mathbf{G}_{22}\delta\mathbf{m}_2^{(1)} &= \delta\mathbf{d}_2. \end{aligned} \tag{35}$$

As a result of the enforced diagonal dominance, the first iterates $\delta\mathbf{m}_1^{(1)}$ and $\delta\mathbf{m}_2^{(1)}$ will already be useful approximations to $\delta\mathbf{m}_1$ and $\delta\mathbf{m}_2$. Subsequently, the off-diagonal terms are incorporated iteratively by solving regularized versions of

$$\begin{aligned} \mathbf{G}_{11}\delta\mathbf{m}_1^{(i+1)} &= \delta\mathbf{d}_1 - \mathbf{G}_{12}\delta\mathbf{m}_2^{(i)}, \\ \mathbf{G}_{22}\delta\mathbf{m}_2^{(i+1)} &= \delta\mathbf{d}_2 - \mathbf{G}_{21}\delta\mathbf{m}_1^{(i)}, \end{aligned} \tag{36}$$

with $i = 1, \ldots$ . During the inversion, only the right-hand side of the equations changes, meaning that it can be solved efficiently once generalized inverses of $\mathbf{G}_{11}$ and $\mathbf{G}_{22}$ are constructed.

### 4.4 The influence of different reference models and the effects of non-linearity on the inversion process

As outlined in Section 2, the construction of optimal observables depends critically on the Fréchet or sensitivity kernels for the fundamental observables. Since the sensitivity kernels depend on the properties of the earth model, the weights used to construct optimal observables depend on the earth model, too. This implies that sensitivity kernels for fundamental observables, as well as optimal weights, should in principle be re-computed in each iteration of an iterative non-linear inversion.

While iterative re-computations would be required in order to obtain exactly the optimal observables, a more relaxed approach can be taken in regional- to global-scale tomographies where the variations from a 1-D background model are usually small. As shown by Zhou *et al.* (2011), the geometry and amplitudes of surface wave sensitivity kernels are only insignificantly affected by plausible lateral heterogeneities in global earth models. A similar result for body waves was presented by Mercerat & Nolet (2012). The weak dependence of sensitivity kernels on percent-level heterogeneities justifies the use of a 1-D earth model in our examples, and it also reduces the computational cost of the proposed optimization scheme because a re-computation of optimal weights in each iteration is unlikely to be necessary.

Nevertheless, the proposed optimization algorithm is also applicable to local studies where the successive introduction of 3-D

heterogeneities throughout the inversion may significantly affect the characteristics of the sensitivity kernels of optimal observables.

### 4.5 Relation to the Backus–Gilbert method and experimental design

The linear combination of data for optimal-design purposes, proposed in eq. (8), is reminiscent of the Backus–Gilbert approach for linear inverse problems with a single model parameter class (Backus & Gilbert 1968). Backus and Gilbert suggested to combine data as $d = \sum_{i=1}^{n} w_i d_i$ such that the composite sensitivity kernel $K(\mathbf{x}) = \sum_{i=1}^{n} K_i(\mathbf{x})$ is close to $\delta(\mathbf{x} - \mathbf{x}_0)$, thereby producing a direct estimate of earth model properties via $\delta d = \int K(\mathbf{x})\, \delta m(\mathbf{x})\, d^3\mathbf{x} \approx \delta m(\mathbf{x}_0)$. The concept is therefore similar, but the design goal and optimization scheme differ significantly.

As indicated already in Section 3.2, our method can be adapted to mimic the Backus–Gilbert method. Defining parameter class 1 to equal, for instance, $S$ velocity at location $\mathbf{x} = \mathbf{x}_0$, and parameter class 2 to equal $S$ velocity at all other positions $\mathbf{x} \neq \mathbf{x}_0$, would produce an optimal observable with an associated kernel that is as much $\delta$-like as allowed by the data. While theoretically appealing, this approach would suffer from the same problems as the Backus–Gilbert method itself: To compute a whole-earth model, the optimization procedure needs to be repeated for all positions $\mathbf{x}$; and the resulting earth model may then not explain the data because the union of optimal point estimates is not necessarily an optimal collective estimate. This explains why applications of the Backus–Gilbert method are relatively few in number (e.g. Chou & Booker 1979; Trampert & van Heijst 2002).

As mentioned in the Introduction further links exist to experimental design frequently used to optimize the source–receiver geometry in geophysical exploration (e.g. Curtis 1999b). Our method can be used for a similar purpose by defining the fundamental observables to be measurements for a collection of potential source–receiver pairs. In the final experiment, only those source–receiver pairs with a weight above a pre-defined threshold would then be used.

## 5 CONCLUSIONS

We developed a method for the construction of seismic observables that have maximum sensitivity with respect to a target model parameter class, while having minimum sensitivity with respect to all remaining parameter classes. This is achieved through the optimal linear combination of fundamental observables that can be any scalar measurement extracted from seismic recordings. The resulting optimal observables minimize interparameter trade-offs that result from regularization in ill-posed multiparameter inverse problems.

In a series of examples we have shown that surface wave traveltime observations in different frequency bands can be combined such that sensitivity to 3-D density structure increases substantially. Simultaneously, sensitivity to $S$ velocity (or shear modulus) is reduced, and sensitivity to $P$ velocity (or bulk modulus) is practically eliminated, thereby reducing a three-parameter problem into a simpler two-parameter problem.

Using rotational ground motion measurements, allows us to construct an observable with dominant sensitivity in the near-receiver region, and zero sensitivity at more than a few wavelengths distance from the receiver. This observable is identical to the apparent shear velocity $\beta_a$ (eq. 23), originally defined on the basis of purely physical arguments (Fichtner & Igel 2009; Bernauer *et al.* 2012).

In the interest of reproducibility and clarity, our examples used a small number of well-understood fundamental observables, few parameter classes and a radially symmetric earth model. The method itself, however, does not impose such restrictions, and it can easily be applied to large numbers of fundamental observables and parameters classes, as well as to 3-D heterogeneous earth models.

## ACKNOWLEDGEMENTS

The authors would like to thank Malcolm Sambridge and Jeannot Trampert for inspiring discussions and the provision of numerous useful references. This work has benefited from thoughtful and constructive comments of Andrew Curtis and an anonymous reviewer. The research presented in this paper was supported by the International Graduate School THESIS within the Bavarian Elite Network, the European Commission (Marie Curie Actions, ITN QUEST), the German Research Foundation (DFG, IG 16/15-1), and the Netherlands Organisation for Scientific Research (NWO) under VIDI grant 864.11.008. The numerical computations were performed on the National Supercomputer SuperMUC maintained by the Leibniz-Rechenzentrum.

## REFERENCES

Babuška, V. & Cara, M., 1991. *Seismic Anisotropy in the Earth,* Kluwer.

Backus, G.E., 1962. Long-wave elastic anisotropy produced by horizontal layering, *J. geophys. Res.,* **67,** 4427–4440.

Backus, G.E. & Gilbert, F., 1968. The resolving power of gross Earth data, *Geophys. J. R. astr. Soc.,* **16,** 169–205.

Barth, N. & Wunsch, C., 1990. Oceanographic experiment design by simulated annealing, *J. Phys. Oceanogr.,* **20,** 1249–1263.

Bernauer, M., Fichtner, A. & Igel, H., 2009. Inferring earth structure from combined measurements of rotational and translational ground motions, *Geophysics,* **74,** WCD41–WCD47.

Bernauer, M., Fichtner, A. & Igel, H., 2012. Measurements of translation, rotation and strain: new approaches to seismic processing and inversion, *J. Seismol.,* **16,** 669–681.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178,** 1411–1436.

Capdeville, Y., Guillot, L. & Marigo, J.J., 2010. 2-D nonperiodic homogenization to upscale elastic media for P-SV waves, *Geophys. J. Int.,* **182,** 903–922.

Cara, M., Lévèque, J.J. & Maupin, V., 1984. Density-versus-depth models from multimode surface waves, *Geophys. Res. Lett.,* **11,** 633–636.

Chen, P., 2011. Full-wave seismic data assimilation: theoretical background and recent advances, *Geophys. J. Int.,* **168,** 1527–1552.

Chou, C.W. & Booker, J.R., 1979. A Backus-Gilbert approach to inversion of travel-time data for three-dimensional velocity structure, *Geophys. J. R. astr. Soc.,* **59,** 325–344.

Coles, D.A. & Morgan, F.D., 2009. A method of fast, sequential experimental design for linearized geophysical inverse problems, *Geophys. J. Int.,* **178,** 145–158.

Curtis, A., 1999a. Optimal design of focused experiments and surveys, *Geophys. J. Int.,* **139,** 205–215.

Curtis, A., 1999b. Optimal experiment design: cross-borehole tomographic examples, *Geophys. J. Int.,* **136,** 637–650.

Curtis, A., 2004a. Theory of model-based geophysical survey and experimental design Part A: linear problems, *Leading Edge,* **23,** 997–1004.

Curtis, A., 2004b. Theory of model-based geophysical survey and experimental design Part B: nonlinear problems, *Leading Edge,* **23,** 1112–1117.

Curtis, A., Michelini, A., Leslie, D. & Lomax, A., 2004. A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys, *Geophys. J. Int.,* **157,** 595–606.

Dahlen, F.A. & Baig, F.A., 2002. Fréchet kernels for body wave amplitudes, *Geophys. J. Int.,* **150,** 440–466.

Dalton, C.A., Ekström, G. & Dziewonski, A.M., 2008. The global attenuation structure of the upper mantle, *J. geophys. Res.,* **113,** doi:10.1029/2007JB005429.

Debayle, E. & Ricard, Y., 2013. Seismic observations of large-scale deformation at the bottom of fast-moving plates, *Earth planet. Sci. Lett.,* **376,** 165–177.

Fichtner, A. & Igel, H., 2009. Sensitivity densities for rotational ground motion measurements, *Bull. seism. Soc. Am.,* **99,** 1302–1314.

Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology: I. Theory, *Phys. Earth planet. Inter.,* **157,** 86–104.

Fichtner, A., Kennett, B.L.N. & Trampert, J., 2013. Separating intrinsic and apparent anisotropy, *Phys. Earth planet. Inter.,* **219,** 11–20.

Fishwick, S., Kennett, B.L.N. & Reading, A.M., 2005. Contrasts in lithospheric structure within the Australian Craton, *Earth planet. Sci. Lett.,* **231,** 163–176.

Gung, Y.C. & Romanowicz, B., 2004. Q tomography of the upper mantle using three component long period waveforms, *Geophys. J. Int.,* **157,** 813–830.

Gung, Y.C., Panning, M. & Romanowicz, B., 2003. Global anisotropy and the thickness of continents, *Geophys. J. Int.,* **422,** 707–711.

Hardt, M. & Scherbaum, F., 1994. The design of optimum networks for aftershock recordings, *Geophys. J. Int.,* **117,** 716–726.

Hess, H.H., 1964. Seismic anisotropy of the uppermost mantle under oceans, *Nature,* **203,** 629–631.

Ishii, M. & Tromp, J., 1999. Normal-mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle, *Science,* **285,** 1231–1236.

Ishii, M. & Tromp, J., 2001. Even-degree lateral variations in the Earth's mantle constrained by free oscillations and the free-air gravity anomaly, *Geophys. J. Int.,* **145,** 77–96.

Jackson, I., 2007. Physical origins of anelasticity and attenuation in rock, in *Treatise on Geophysics,* pp. 493–525, ed. Schubert, G., Elsevier.

Jackson, I., Fitz Gerald, J.D., Faul, U.H. & Tan, B.H., 2002. Grain-size-sensitive seismic wave attenuation in polycrystalline olivine, *J. geophys. Res.,* **107,** doi:10.1029/2001JB001225.

Kennett, B.L.N., 1997. Observational and theoretical constraints on crustal and upper mantle, *Phys. Earth. planet. Inter.,* **47,** 319–332.

Kennett, B.L.N., 1998. On the density distribution within the Earth, *Geophys. J. Int.,* **132,** 374–382.

Kennett, B.L.N. & Abdullah, A., 2011. Seismic wave attenuation beneath the Australasian region, *Austr. J. Earth Sci.,* **58,** 285–295.

Kennett, B.L.N. & Sambridge, M.S., 1998. Inversion for multiple parameter classes, *Geophys. J. Int.,* **135,** 304–306.

Kennett, B.L.N., Sambridge, M.S. & Williamson, P.R., 1988. Subspace methods for large inverse problems with multiple parameter classes, *Geophys. J. Int.,* **94,** 237–247.

Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.,* **122,** 108–124.

Khan, A., Zunino, A. & Deschamps, F., 2013. Upper mantle compositional variations and discontinuity topography imaged beneath australia from Bayesian inversion of surface-wave phase velocities and thermochemical modeling, *J. geophys. Res.,* **118,** doi:10.1002/jgrb.50304.

Kuo, C. & Romanowicz, B., 2002. On the resolution of density anomalies in the Earth's mantle using spectral fitting of normal mode data, *Geophys. J. Int.,* **150,** 162–179.

Liu, Q. & Tromp, J., 2008. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.,* **174,** 265–286.

Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics,* **56,** 645–653.

Maceira, M. & Ammon, C.J., 2009. Joint inversion of surface wave velocity and gravity observations and its application to central asian basins shear velocity structure, *J. geophys. Res.,* **114,** doi:1029/2007JB005157.

Maurer, H. & Boerner, D.E., 1998. Optimized and robust experimental design: a non-linear application to EM sounding, *Geophys. J. Int.,* **132,** 458–468.

Maurer, H., Curtis, A. & Boerner, D.E., 2010. Recent advances in optimized geophysical survey design, *Geophysics,* **75,** A177–A195.

Mercerat, E.D. & Nolet, G., 2012. Comparison of ray- and adjoint-based sensitivity kernels for body wave seismic tomography, *Geophys. Res. Lett.,* **39,** doi:10.1029/2012GL052002.

Montagner, J.P. & Anderson, D.L., 1989. Petrological constraints on seismic anisotropy, *Phys. Earth planet. Inter.,* **54,** 82–105.

Mosca, I., Cobden, L., Deuss, A., Ritsema, J. & Trampert, J., 2012. Seismic and mineralogical structures of the lower mantle from probabilistic tomography, *J. geophys. Res.,* **117,** doi:10.1029/2011JB008851.

Nettles, M. & Dziewoński, A.M., 2008. Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America, *J. geophys. Res.,* **113,** doi:10.1029/2006JB004819.

Nolet, G., 2008. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun,* Cambridge Univ. Press.

Panning, M. & Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.,* **167,** 361–379.

Rabinowitz, N. & Steinberg, D.M., 1990. Optimal configuration of a seismographic network: a statistical approach, *Bull. seism. Soc. Am.,* **80,** 187–196.

Resovsky, J. & Ritzwoller, M., 1999. Regularisation uncertainty in density models estimated from normal mode data, *Geophys. Res. Lett.,* **26,** 2319–2322.

Resovsky, J. & Trampert, J., 2003. Using probabilistic seismic tomography to test mantle velocity-density relationships, *Earth planet. Sci. Lett.,* **215,** 121–134.

Romanowicz, B., 2001. Can we resolve 3D density heterogeneity in the lower mantle? *Geophys. Res. Lett.,* **28,** 1107–1110.

Ruan, Y. & Zhou, Y., 2010. The effects of 3-D anelasticity, Q structure on surface wave phase delays, *Geophys. J. Int.,* **181,** 479–492.

Sambridge, M.S. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40,** doi:10.1029/2000RG000089.

Sieminski, A., Trampert, J. & Tromp, J., 2009. Principal component analysis of anisotropic finite-frequency kernels, *Geophys. J. Int.,* **179,** 1186–1198.

Simmons, N.A., Forte, A.M., Boschi, L. & Grand, S.P., 2010. GyPSuM: a joint tomography model of mantle density and seismic wave speeds, *J. geophys. Res.,* **115,** doi:10.1029/2010JB007631.

Takeuchi, H. & Saito, M., 1972. Seismic surface waves, in *Methods in Computational Physics,* Vol. **11,** pp. 217–295, ed. Bolt, B.A., Academic Press, London & New York.

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.

Tondi, R., de Franco, R. & Barzaghi, R., 2000. Sequential inversion of refraction and wide-angle reflection traveltimes and gravity data for two-dimensional velocity structures, *Geophys. J. Int.,* **141,** 679–698.

Tondi, R., Achauer, U., Landes, M., Davi, R. & Besutiu, L., 2009. Unveiling seismic and density structure beneath the Vrancea seismogenic zone, Romania, *J. geophys. Res.,* **141,** doi:10.1029/2008JB005992.

Trampert, J. & van Heijst, H.J., 2002. Global azimuthal anisotropy in the transition zone, *Science,* **296,** 1297–1299.

Trampert, J., Deschamps, F., Resovsky, J. & Yuen, D., 2004. Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science,* **306,** 853–856.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160,** 195–216.

van den Berg, J., Curtis, A. & Trampert, J., 2003. Optimal nonlinear Bayesian experimental design: an application to amplitude versus offset experiments, *Geophys. J. Int.,* **155,** 411–421.

Winterfors, E. & Curtis, A., 2008. Numerical detection and reduction of non-uniqueness in nonlinear inverse problems, *Inverse Probl.,* **24,** doi:10.1088/0266-5611/24/2/025016.

Wu, R. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics,* **50,** 582–595.

Yoshizawa, K. & Kennett, B.L.N., 2004. Multi-mode surface wave tomography for the Australian region using a 3-stage approach incorporating finite-frequency effects, *J. geophys. Res.,* **109,** doi:10.1029/2002JB002254.

Zhou, Y., 2009. Surface-wave sensitivity to 3-D anelasticity, *Geophys. J. Int.,* **178,** 1403–1410.

Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.,* **158,** 142–168.

Zhou, Y., Lui, Q. & Tromp, J., 2011. Surface-wave sensitivity: mode summation versus adjoint SEM, *Geophys. J. Int.,* **187,** 1560–1576.

## APPENDIX A: OPTIMAL EIGENVALUE

Given $\sum_{i=1}^{n} w_i^2 = 1$ and $\frac{\partial \mathcal{L}}{\partial w_l} = 0$ we have $\mathcal{L}(\mathbf{w}, \lambda) = \lambda$.

Proof:

$$\begin{aligned}
\mathcal{L}(\mathbf{w}, \lambda) &= \sum_{j=1}^{p} b_j \int_V \left[\sum_{i=1}^{n} w_i K_{ij}(\mathbf{x})\right]^2 d^3\mathbf{x} - \lambda\left(\sum_{i=1}^{n} w_i^2 - 1\right) \\
&= \sum_{j=1}^{p} b_j \int_V \left[\sum_{i=1}^{n} w_i K_{ij}(\mathbf{x})\right]^2 d^3\mathbf{x} \\
&= \sum_{j=1}^{p} b_j \int_V [w_1 w_1 K_{1j} K_{1j} + \cdots + w_1 w_n K_{1j} K_{nj} + \\
&\quad \vdots \qquad\qquad \vdots \\
&\quad + w_n w_1 K_{nj} K_{1j} + \cdots + w_n w_n K_{nj} K_{nj}]\, d^3\mathbf{x} \\
&= w_1 M_{11} w_1 + \cdots + w_1 M_{1n} w_n + \\
&\quad \vdots \qquad\qquad \vdots \\
&\quad + w_n M_{n1} w_1 + \cdots + w_n M_{nn} w_n \\
&= \sum_{l,i=1}^{n} w_l M_{li} w_i = \mathbf{w}^{\mathrm{T}} \mathbf{M} \mathbf{w} = \lambda \mathbf{w}^{\mathrm{T}} \mathbf{w} = \lambda.
\end{aligned}$$

## APPENDIX B: SENSITIVITY KERNELS FOR VERTICAL DENSITY GRADIENTS

To derive expressions for sensitivity kernels with respect to vertical density gradients, $\partial_r \rho$, we start with the general first-order expression that relates a change in the data $\delta d$ to a change in density, $\delta\rho$:

$$\delta d = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \int_{r=0}^{R} K_\rho(\theta, \phi, r)\, \delta\rho(\theta, \phi, z)\, \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}r, \tag{B1}$$

where $R$ denotes the radius of the Earth. In eq. (B1), we incorporated the term $r^2 \sin\theta$ in the spherical volume element in the definition of the density kernel $K_\rho$, thereby leaving a simple integral of $K_\rho \delta\rho$ over colatitude $\theta$, longitude $\phi$ and radius $r$. Substituting the identity

$$K_\rho(\theta, \phi, r) = \partial_r \int_{r'=0}^{r} K_\rho(\theta, \phi, r')\, \mathrm{d}r' \tag{B2}$$

into eq. (B1), yields

$$\delta d = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \int_{r=0}^{R} \left[\partial_r \int_{r'=0}^{r} K_\rho(\theta, \phi, r')\, \mathrm{d}r'\right] \delta\rho(\theta, \phi, z)\, \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}r. \tag{B3}$$

Integrating by parts with respect to $r$, transforms (B3) into

$$\begin{aligned}
\delta d = &\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \left[\int_{r'=0}^{r} K_\rho(\theta, \phi, r')\, \mathrm{d}r'\right] \delta\rho(\theta, \phi, z)\, \mathrm{d}\theta\, \mathrm{d}\phi \Big|_{r=0}^{R} \\
&- \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \int_{r=0}^{R} \left[\int_{r'=0}^{r} K_\rho(\theta, \phi, r')\, \mathrm{d}r'\right] \partial_r \delta\rho(\theta, \phi, z)\, \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}r.
\end{aligned} \tag{B4}$$

Recognizing that the vertical integral over the density kernel $K_\rho$ in $\alpha - \beta - \rho$ parametrization is nearly zero (e.g. Takeuchi & Saito 1972), the first term on the right-hand side of (B4) can be ignored. Using the identity $\partial_r \delta\rho = \partial_r(\rho - \rho_0) = \partial_r \rho - \partial_r \rho_0 = \delta \partial_r \rho$, eq. (B4) simplifies to

$$\delta d = -\int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \int_{r=0}^{R} \left[\int_{r'=0}^{r} K_\rho(\theta, \phi, r')\, \mathrm{d}r'\right] \delta \partial_r \rho(\theta, \phi, z)\, \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}r. \tag{B5}$$

From eq. (B5) we see that the sensitivity kernel with respect to the vertical density gradient $\partial_r \rho$ is given by the negative integral of $K_\rho$ with respect to $r$:

$$K_{\partial_r \rho}(\theta, \phi, r) = -\int_{r'=0}^{r} K_\rho(\theta, \phi, r')\, \mathrm{d}r'\,. \tag{B6}$$

With $K_{\partial_r \rho}$ we retrieve the generic expression relating a change in data to a change in model parameters:

$$\delta d = \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \int_{r=0}^{R} K_{\partial_r \rho}(\theta, \phi, r)\, \delta \partial_r \rho(\theta, \phi, z)\, \mathrm{d}\theta\, \mathrm{d}\phi\, \mathrm{d}r\,. \tag{B7}$$

Kernels for relative perturbations $\delta \ln \partial_r \rho$ are obtained from $K_{\partial_r \rho}$ by multiplication with a reference value for $\partial_r \rho$. In the case of the 1-D reference earth model ak135 (Kennett *et al.* 1995), used in the examples of Section 3, $\partial_r \rho$ takes the nearly constant value of $6.05 \cdot 10^{-4}$ kg m$^{-4}$ from 35 to 410 km depth.