AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL074996

Key Points:
- Damped least squares tomographic models can be locally biased in poorly sampled regions
- Slow velocity anomalies below Hawaii are biased toward higher amplitudes in model S40RTS
- We show how to efficiently compute unbiased models including their full resolution and covariance

Supporting Information:
- Supporting Information S1

Correspondence to:
C. Zaroli,
c.zaroli@unistra.fr

Citation:
Zaroli, C., Koelemeijer, P., & Lambotte, S. (2017). Toward seeing the Earth's interior through unbiased tomographic lenses. *Geophysical Research Letters*, *44*, 11,399–11,408. https://doi.org/10.1002/2017GL074996

Received 20 JUL 2017
Accepted 27 OCT 2017
Accepted article online 2 NOV 2017
Published online 29 NOV 2017

Toward Seeing the Earth's Interior Through Unbiased Tomographic Lenses

Christophe Zaroli[1], **Paula Koelemeijer**[2], **and Sophie Lambotte**[1]

[1]Institut de Physique du Globe de Strasbourg, UMR 7516, Université de Strasbourg, EOST/CNRS, Strasbourg, France, [2]Department of Earth Sciences, University of Oxford, Oxford, UK

Abstract Geophysical tomographic studies traditionally exploit linear, damped least squares inversion methods. We demonstrate that the resulting models can be locally biased toward lower or higher amplitudes in regions of poor data illumination, potentially causing physical misinterpretations. For example, we show that global model S40RTS is locally biased toward higher amplitudes below isolated receivers where raypaths are quasi-vertical, such as on Hawaii. This leads to questions on the apparent low-velocity structure interpreted as the Hawaii hot spot. We prove that a linear Backus-Gilbert inversion scheme can bring the Earth's interior into focus through unbiased tomographic lenses, as its model estimates are constrained to be averages over the true model. It also efficiently computes the full generalized inverse required to infer both model resolution and its covariance, enabling quantitative interpretations of tomographic models.

1. Introduction

From local to global scale, a continuing challenge in seismology is to build higher-resolution tomographic models, to better constrain both length scales and magnitude of seismic heterogeneities in the Earth's interior, for example, to further improve our knowledge of the physical forces driving plate tectonics and whole-mantle convective processes (e.g., Davies et al., 2012; Romanowicz, 2003). Robust physical interpretations of tomographic images require accurate appraisals of the underlying local model resolution and uncertainty (covariance). However, their quantitative assessment has often been ignored or at best given minimal treatment (e.g., Rawlinson et al., 2014; Rawlinson & Spakman, 2016).

In this context, we aim to focus on inversion methods that can deal with large-scale, linear, discrete tomographic problems ($\gtrsim 10^5$ model parameters, $\gtrsim 10^6$ data) and efficiently compute model estimates with their resolution and covariance. We will not consider probabilistic approaches, which suffer from the curse of model dimensionality (e.g., Burdick & Lekic, 2017; Sambridge et al., 2013), or nonlinear methods such as full-waveform inversion (FWI) (e.g., Fichtner et al., 2009), which are computationally very intensive. FWI also faces more difficulties toward quantifying resolution (e.g., Fichtner & Trampert, 2011), and a full assessment of uncertainties seems out of reach at present.

Geophysical tomographic studies traditionally exploit linear(ized), damped least squares (DLS) inversion methods (e.g., Aster et al., 2012), within ray theoretical or finite frequency (e.g., Dahlen et al., 2000) physical frameworks, using data such as ambient noise (e.g., Zigone et al., 2015), surface wave (e.g., Debayle & Ricard, 2012), body wave (e.g., Sigloch, 2011), normal mode (e.g., Resovsky & Ritzwoller, 1999), or joint data sets (e.g., Koelemeijer et al., 2016, Zaroli et al., 2015). In the DLS approach, the nonuniqueness inherent to the least squares solution is removed by adding ad hoc regularization constraints on the model itself, such as norm damping or smoothing—often aimed to subjectively favor the model simplicity. Such ad hoc regularization schemes are fundamentally different from a priori physical constraints, objectively imposed on the solution within some Bayesian philosophy. Because of such ad hoc regularization constraints, DLS model estimates may be locally *biased* averages over the true-model parameters, thus potentially causing physical misinterpretations (e.g., Nolet, 2008). In this study, we aim to better quantify these "averaging bias" effects and demonstrate that all DLS tomographic studies with uneven data coverage are concerned.

A second focus of this paper will be to show how to efficiently compute bias-free tomographic images; that is, using the "SOLA Backus-Gilbert" inversion method (hereafter SOLA), recently adapted and introduced to

Table 1
A Brief Comparison of Linear Inversion Methods: DLS Versus SOLA

	DLS	SOLA
Minimization problem	$\min_{\tilde{\mathbf{m}}} \underbrace{\|\mathbf{d}-\mathbf{G}\tilde{\mathbf{m}}\|^2}_{\text{data misfit}} + \Theta^2 \underbrace{\|\tilde{\mathbf{m}}\|^2}_{\text{model norm}}$	$\min_{\hat{\mathbf{G}}^\dagger_{k.}} \underbrace{\sum_{j=1}^{M}\left(\hat{R}_{kj}-T_j^{(k)}\right)^2}_{\text{resolution misfit}} + \eta_k^2 \underbrace{\sigma^2_{\tilde{m}_k}}_{\text{model variance}}$ s.t. $\sum_{j=1}^{M}\hat{R}_{kj}=1$
Generalized inverse	$\tilde{\mathbf{G}}^\dagger = \{\mathbf{G}^T\mathbf{G}+\Theta^2\mathbf{I}_M\}^{-1}\mathbf{G}^T$	$\hat{\mathbf{G}}^\dagger \longleftarrow \hat{\mathbf{G}}^\dagger_{k.}$ loop over k (parallel computing)
Averaging bias (kth parameter)	$\sum_{j=1}^{M}\tilde{R}_{kj}$ <1 $=1$ >1	$\sum_{j=1}^{M}\hat{R}_{kj}$ $=1$
Model estimate (kth parameter)	biased[a] unbiased[b] biased[c]	unbiased[b]

[a]The model estimate is biased toward lower amplitudes. [b]To be fully unbiased, the resolving kernels should also be nonnegative. [c]The model estimate is biased toward higher amplitudes.

large-scale, linear, and discrete tomographic problems by Zaroli (2016)—closely related to earlier theoretical developments (e.g., Backus & Gilbert, 1967, 1968, 1970; Nolet, 1985; Pijpers & Thompson, 1992, 1994). In the SOLA approach, model estimates are explicitly constrained to be (*unbiased*) averages over the true-model parameters—no ad hoc model damping or smoothing is needed to find a solution, similarly to probabilistic methods, thus avoiding biasing it. Last, but not least, DLS and SOLA schemes will be compared in terms of their computational efficiency for calculating the full generalized inverse—needed to infer both the model resolution and its covariance.

We will illustrate and quantitatively argue for these points using a simple tomographic "toy" problem and a "real-data-based" global tomographic model, S40RTS (Ritsema et al., 2011).

2. Material and Methods

2.1. Preamble

We are interested in linear(ized), discrete forward problems of the form

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{n}\,, \tag{1}$$

where $\mathbf{d} = (d_i)_{1\le i\le N}$ denotes the data, $\mathbf{G} = (G_{ij})_{1\le i,j\le N,M}$ the sensitivity matrix, $\mathbf{m} = (m_j)_{1\le j\le M}$ the true-model parameters, and $\mathbf{n} = (n_i)_{1\le i\le N}$ the noise. Let us consider, without loss of generality, that the data are time residuals, the model parameters are velocity anomalies, the model parametrization is local and orthonormal, the noise is randomly drawn from a normal distribution with zero mean and standard deviation σ_n, and the data covariance matrix is $\mathbf{C}_d = \sigma_n^2\mathbf{I}_N$. From hereon, the data and sensitivity matrix are normalized by the data errors (so that $\mathbf{C}_d = \mathbf{I}_N$).

2.2. Model Estimate, Appraisal, and Averaging Bias

One aims at finding a model estimate, $\tilde{\mathbf{m}}$, as a linear combination of the data

$$\tilde{\mathbf{m}} = \tilde{\mathbf{G}}^\dagger\mathbf{d}\,, \tag{2}$$

where the matrix $\tilde{\mathbf{G}}^\dagger$ denotes *some* generalized inverse; $\tilde{\mathbf{m}}$ can then be decomposed as

$$\tilde{\mathbf{m}} = \tilde{\mathbf{R}}\mathbf{m} + \tilde{\mathbf{G}}^\dagger\mathbf{n}\,, \tag{3}$$

where $\tilde{\mathbf{R}} = \tilde{\mathbf{G}}^\dagger\mathbf{G}$ is the model resolution matrix. The first term, $\tilde{\mathbf{R}}\mathbf{m}$, represents the filtered true model ($\mathbf{m}$ filtered by $\tilde{\mathbf{R}}$). It shows our inability, if $\tilde{\mathbf{R}} \neq \mathbf{I}_M$, to perfectly recover the true model (even with error-free data). The second term, $\tilde{\mathbf{G}}^\dagger\mathbf{n}$, is the propagation of data errors into the model estimate. The appraisal problem consists in computing and analyzing both $\tilde{\mathbf{R}}$ and the model covariance matrix $\mathbf{C}_{\tilde{m}} = \tilde{\mathbf{G}}^\dagger\mathbf{C}_d(\tilde{\mathbf{G}}^\dagger)^T$. Note that the model estimate $\tilde{\mathbf{m}}$, resolution $\tilde{\mathbf{R}}$, and covariance $\mathbf{C}_{\tilde{m}}$ can directly be inferred from the generalized inverse $\tilde{\mathbf{G}}^\dagger$. Therefore, computing the full generalized inverse is the cornerstone of any linear inversion method; we will see how it differs for DLS and SOLA (see section 2.3, Table 1, and supporting information).

We wish that $\tilde{\mathbf{R}}\mathbf{m}$ represents an unbiased averaging over the true model $\mathbf{m}$. Here we refer to the kth row of the resolution matrix, $\tilde{\mathbf{R}}_{k.} = (\tilde{R}_{kj})_{1\le j\le M}$, as the resolving kernel linearly relating the kth parameter estimate, $\tilde{m}_k$, to the

true-model parameters ($\tilde{m}_k = \sum_j \tilde{R}_{kj} m_j$, ignoring the term of propagated data errors). Let us now define the averaging bias quantity $\tilde{U}_k$ related to the parameter estimate $\tilde{m}_k$ as follows:

$$\tilde{U}_k = \sum_{j=1}^{M} \tilde{R}_{kj} \,. \tag{4}$$

Since the resolving kernel $\tilde{\mathbf{R}}_{k.}$ is expected to be nonnegative (see section 3.1.3), then if the averaging bias quantity $\tilde{U}_k$ differs from unity, the parameter estimate $\tilde{m}_k$ will represent some *biased* averaging over the true-model parameters — biased toward lower or higher amplitudes, depending on whether $\tilde{U}_k$ is lower or higher than 1, respectively.

A primary goal of the toy problem will consist of quantifying how much this averaging bias effect affects both the DLS and SOLA model solutions.

2.3. DLS Versus SOLA

From hereon, the superscripts ˜ (tilde) and ˆ (hat) will refer to DLS and SOLA, respectively. We now aim to briefly point out how the DLS and SOLA methods differ (summarized in Table 1); further details on the SOLA tomographic scheme are given in the supporting information.

2.3.1. Biased or Not

The simplest DLS method consists of finding a model estimate $\tilde{\mathbf{m}}$ that minimizes both the data misfit and the model L_2 norm, that is, $\|\mathbf{d} - \mathbf{G}\tilde{\mathbf{m}}\|^2 + \Theta^2\|\tilde{\mathbf{m}}\|^2 = \min$, where Θ is some ad hoc damping parameter. The DLS model $\tilde{\mathbf{m}}$ refers to the *damped* true model (e.g., Menke, 1989; Nolet, 2008) and thus may represent locally biased averages over the true-model parameters (i.e., $\tilde{U}_k$ may differ locally from 1).

The Backus-Gilbert (discrete) approach aims to directly identify *averages* over the true-model parameters (e.g., Backus & Gilbert, 1967, 1968, 1970; Nolet, 1985). The SOLA tomographic method (Zaroli, 2016) retains all the advantages of the original Backus-Gilbert scheme but is more computationally efficient and versatile in the construction of resolving (averaging) kernels (e.g., Pijpers & Thompson, 1992, 1994).

Each row of the SOLA generalized inverse is individually computed by solving a specific minimization problem. The kth row $\hat{\mathbf{G}}^{\dagger}_{k.} = (\hat{G}^{\dagger}_{ki})_{1 \le i \le N}$ is found such that $\sum_j \left(\hat{R}_{kj} - T_j^{(k)}\right)^2 + \eta_k^2 \sigma^2_{\hat{m}_k} = \min$, subject to $\sum_j \hat{R}_{kj} = 1$, where $T_j^{(k)}$ denote all M components of the kth target resolving kernel and η_k is the kth trade-off parameter (resolution misfit versus model variance). Therefore, since $\hat{U}_k$ is constrained to be unity, the kth parameter estimate, $\hat{m}_k = \sum_i \hat{G}^{\dagger}_{ki} d_i$, is expected to be unbiased — so for the full model estimate $\hat{\mathbf{m}}$.

2.3.2. Computational Efficiency

We compare the theoretical, computational efficiency of both DLS and SOLA methods to calculate the full generalized inverse matrix — required to infer the model estimate, resolution and covariance. First, the DLS generalized inverse can be expressed as $\tilde{\mathbf{G}}^{\dagger} = \{\mathbf{G}^T\mathbf{G} + \Theta^2\mathbf{I}_M\}^{-1}\mathbf{G}^T$ and can be calculated from the eigen decomposition of $\mathbf{G}^T\mathbf{G}$. However, since $\mathbf{G}^T\mathbf{G}$ is much less sparse than the sensitivity matrix $\mathbf{G}$, it may be too large to be diagonalized or even to fit in computer memory for large-scale tomographic applications. Thus, in practice, $\tilde{\mathbf{G}}^{\dagger}$ is often not computed (e.g., Nolet, 2008; Rawlinson et al., 2014). On the other hand, the SOLA scheme is extremely parallel to compute the generalized inverse $\hat{\mathbf{G}}^{\dagger}$, since each row is computed independently from the others (section 2.3.1). In addition, as detailed in the supporting information, computing the kth row only involves one LSQR inversion of a unique (i.e., independent from index k) matrix $\mathbf{Q}^{(\eta)}$, of size $(M+1) \times (N-1)$, that is almost as sparse as $\mathbf{G}$ — so that $\mathbf{Q}^{(\eta)}$ can easily fit in computer memory and its sparsity be exploited by the LSQR algorithm (Paige & Saunders, 1982). Therefore, depending on the computational facilities (e.g., number of processors), SOLA may be more efficient than DLS for computing generalized inverses.

2.4. Toy Problem: Settings

We aim to carry out a synthetic, 2-D tomographic experiment to better quantify the averaging bias effect in DLS models and formally compare DLS and SOLA tomographic results.

2.4.1. Tomographic Experiment

Here the model parameters are shear wave velocity perturbations (dlnVs), and the parametrization consists of $M = 1{,}024$ square pixels of unit area each (see Figure 1a). In the framework of ray theory, data represent onset delay times of direct S waves, whose raypaths are straight lines from one black dot to another. All pairs of black dots with an interdistance larger than 8 pixel units are considered as suitable raypaths. Total number of data is $N = 9{,}778$. Each element G_{ij} of the sensitivity matrix represents the length of the ith ray inside the

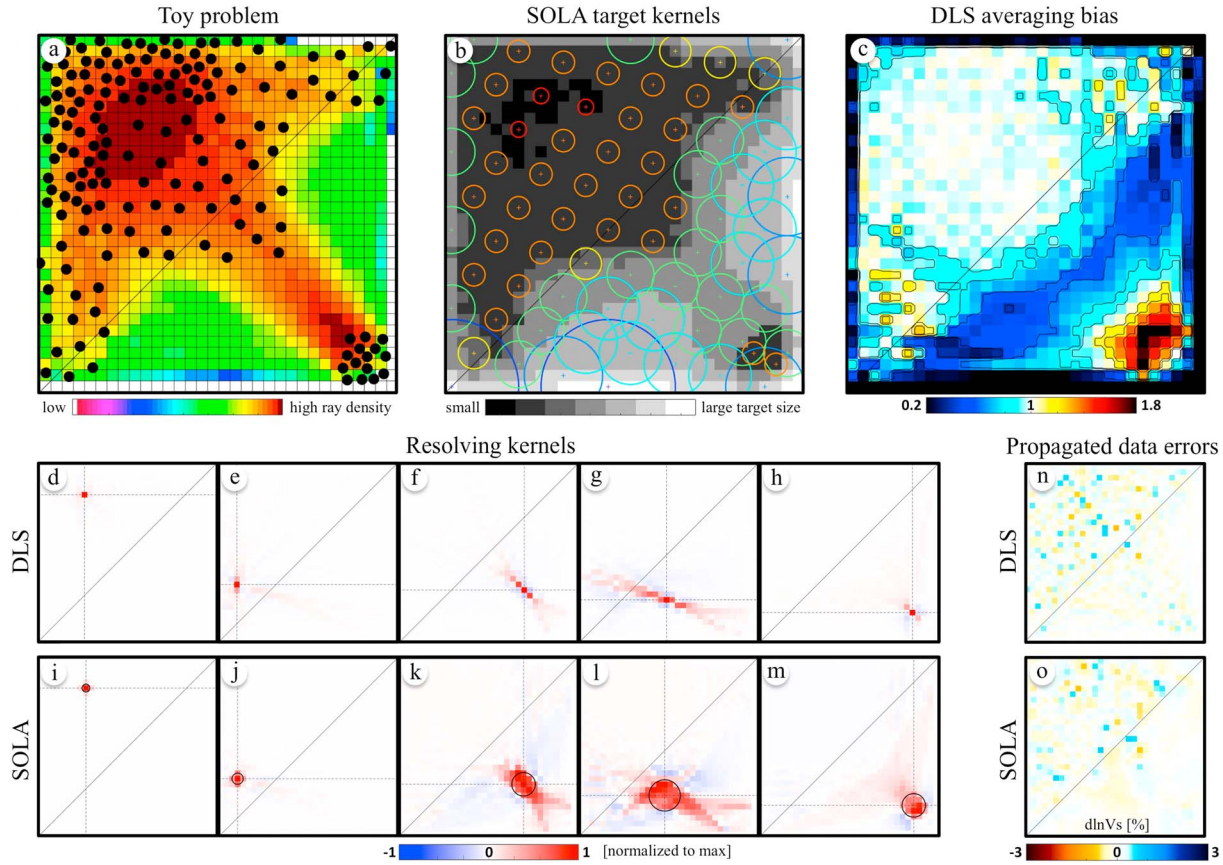

Figure 1. Toy problem. (a) Data coverage (ray density) and model parametrization (pixels); (b) circular forms of some SOLA target resolving kernels (each target kernel is constant inside such a circle and zero outside); (c) DLS averaging bias quantity ($\tilde{U}_k$); (d–h) DLS resolving kernels; (i–m) SOLA resolving kernels (black circles denote target forms); and (n–o) DLS and SOLA propagated data errors, respectively.

jth pixel. Normalized ray density is shown in Figure 1a, where the quantity $\log_{10} \frac{\sum_i |G_{ij}|}{\max_j \sum_i |G_{ij}|}$ is color-plotted for every pixel j. For a given true model $\mathbf{m}$, the data are computed as $\mathbf{d} \leftarrow \mathbf{Gm} + \mathbf{n}$, for random noise $\mathbf{n}$ with realistic σ_n (e.g., Bolton & Masters, 2001).

2.4.2. Tunable Inversion Parameters

We select the damping Θ such that the DLS model estimate fits the data at the level of "reduced chi-square" equal to 1 (e.g., Nolet, 2008) and mostly follow Zaroli (2016) to specify suitable SOLA target resolving kernels and trade-off parameters; three steps are considered in this process. In the first step, we aim to determine all M target resolving kernels such that their spatial extent represents some relevant a priori estimate of the local resolving length around each pixel location; here their form is chosen to be circular, see Figure 1b. We use the ray density as a first-order proxy for the spatial variations of the local resolution and make an educated guess about the resolving-length bounds. In the second step, we aim to find a single, constant value η suited for all M trade-off parameters η_k such that it leads to a globally coherent model solution. Finally, in the third step, we aim to ensure that all M resolving kernels are mostly well localized (close to the target kernels) and non-negative (nonoscillating), while the model uncertainties $\sigma_{\hat{m}_k}$ remain limited (for more details, see supporting information). The variability of the DLS and SOLA solutions, as a function of some of these tunable inversion parameters, is discussed and illustrated in the supporting information (see Figures S1–S4).

3. Results

3.1. Toy Problem: DLS Versus SOLA Tomographic Lenses

3.1.1. Subexperiments

We consider five subexperiments to better illustrate the toy problem results. Five true models (see Figure 2, first column) are used to generate five data sets, keeping the noise unchanged, to be inverted with both the DLS and SOLA methods. Since specifying the SOLA target kernels and trade-off parameters only depends on

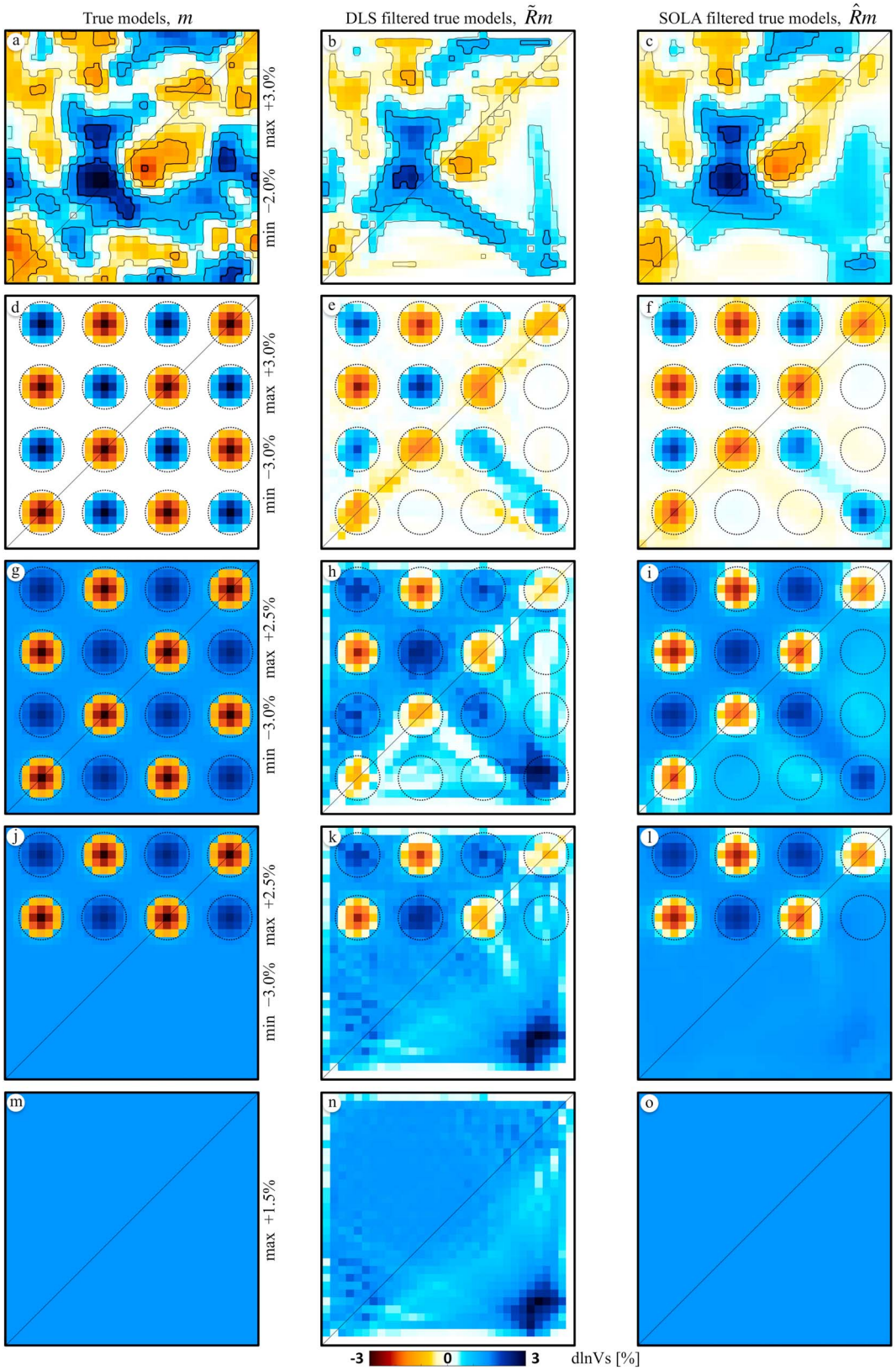

Figure 2. Toy problem. (a, d, g, j, m) True models, **m** (minimum and maximum values are indicated); (b, e, h, k, n) DLS filtered true models, $\tilde{\mathbf{R}}\mathbf{m}$; and (c, f, i, l, o) SOLA filtered true models, $\hat{\mathbf{R}}\mathbf{m}$. Anomalies of the first true model (Figure 2a) are spatially distributed like a 2-D Gaussian random field.

the fixed data geometry and errors, they are kept the same in all five SOLA inversions. We also make sure that the same damping value is selected in all five DLS inversions. Therefore, the DLS and SOLA generalized inverses remain the same through the subexperiments — so that the filtered true models do vary (see Figure 2 second and third columns) but the resolving kernels and propagated data errors (displayed in Figure 1) remain unchanged.

3.1.2. Averaging Bias

In Figure 1c, the DLS averaging bias quantity $\tilde{U}_k$ is plotted for every pixel k. Though $\tilde{U}_k$ is almost unity in regions with "good" ray coverage (spatially dense and isotropic, mainly in the upper left triangle), it varies within the range 0.2–1.8 where the coverage is "poor" (sparse or anisotropic, mainly in the lower right triangle).

We observe that $\tilde{U}_k \gg 1$ in the lower right corner, around where the isolated cluster of black dots is located. This peculiar, dense, and anisotropic local ray coverage is similar to the case of isolated receivers located on oceanic islands, such as on Hawaii, for which most of the arriving teleseismic body wave rays are almost unidirectional (vertical) in the few hundreds kilometers below the receivers (see section 3.2).

That region, where $\tilde{U}_k$ is much greater than 1, is surrounded by another region where $\tilde{U}_k \ll 1$ (in the lower right triangle). Thus, the spatial variations of $\tilde{U}_k$ appear to be complex and to reflect the overall heterogenous raypaths distribution.

Note that $\tilde{U}_k$ can not only be smaller than 1, as suggested by Nolet (2008), but can also be *larger* than 1. In other words, DLS models will represent averages over the true models that can be locally biased toward lower or higher amplitudes.

3.1.3. Resolving Kernels

Meaningful model interpretations require the resolving kernels to be both unbiased (no averaging bias) and spatially well localized around the considered pixels.

In Figures 1d–1m we display some normalized DLS and SOLA resolving kernels. Note that most are mainly nonnegative. Their spatial extent is similar in regions of good ray coverage (upper left triangle) but differs where the coverage becomes poorer (lower right triangle). For instance, DLS kernels tend to be more stretched along the preferential ray directions (see Figures 1f–1h) compared to SOLA kernels (see Figures 1k–1m).

Some DLS resolving kernels are misleading. For example, we show in Figure 1h a kernel $\tilde{\mathbf{R}}_{k'.}$ that is located in the lower right corner region (dense, anisotropic coverage); index k' refers to the considered pixel. Its spatial extent is closely restricted around the k'th pixel, and its value $\tilde{R}_{k'k'}$ at the k'th pixel location is relatively high (0.46). Therefore, at first glance, one could be tempted to conclude that the k'th parameter estimate $\tilde{m}_{k'}$ represents a true-model averaging over a small-size area around the k'th pixel (i.e, high resolution), with a moderate uncertainty $\sigma_{\tilde{m}_{k'}}$ (0.23%). However, since the averaging bias quantity $\tilde{U}_{k'}$ is larger than 1 (1.8), the estimate $\tilde{m}_{k'}$ actually represents a biased average — toward 80% higher amplitude. Note that the SOLA kernel $\hat{\mathbf{R}}_{k'.}$ (see Figure 1m) expands over a broader region (i.e., poorer resolution), as expected from the local, anisotropic raypath distribution.

This illustrates how important it is to take the averaging bias effect into account in the appraisal of DLS models, to avoid misinterpretations, and that SOLA resolving kernels are often more reliable and better suited for robust model interpretations.

3.1.4. Model Estimates

We now aim to compare DLS and SOLA tomographic lenses in terms of model estimates. In the toy problem, both the DLS and SOLA propagated data errors are moderate, see Figures 1n and 1o, with respect to the amplitudes of the true models (see Figure 2, first column). Note that the propagated noise is almost randomly distributed although overall larger in the upper left triangle where the local resolution is better. Thus, we focus on directly comparing the DLS and SOLA "filtered true models."

All five true models, $\mathbf{m}$, corresponding to the aforementioned subexperiments (section 3.1.1), are displayed in Figure 2, first column. The DLS and SOLA filtered true models, $\tilde{\mathbf{R}}\mathbf{m}$ and $\hat{\mathbf{R}}\mathbf{m}$, are shown in Figure 2 (second and third columns, respectively). Differences between the DLS and SOLA filtered models are mainly related to differences of resolving kernels (averaging bias effect and spatial extent).

In regions where the ray coverage is good (upper left triangle), the DLS and SOLA filtered models look similar, and the input true models are well recovered. That is, the DLS averaging bias quantity ($\tilde{U}_k$) is close to 1 (see Figure 1c), and the DLS and SOLA resolving kernels have similar, mostly short, spatial extents (see Figures 1d, 1e, and 1i, and 1j).

In regions of poor ray coverage (lower right triangle), the DLS averaging bias quantity ($\tilde{U}_k$) significantly deviates from 1 (see Figure 1c), and the spatial extent of the DLS and SOLA resolving kernels highly differs (see Figures 1f–1h and 1k–1m). Therefore, the DLS and SOLA filtered models often strongly differ—more or less depending on the input models and their spatial relationships with the resolving kernels.

In some DLS filtered models (see Figures 2h, 2k, and 2n), several apparent structural features can be directly related to the averaging bias effect (see Figure 1c). For example, in the lower right corner the bias effect is to locally increase—up to 80%—the amplitudes in the filtered models, thus artificially giving rise to the observed *darker* blue anomaly. In section 3.2, in the context of global tomography, we will quantify whether a similar bias effect could happen below some isolated, oceanic receivers (e.g., Hawaii), where the body wave ray coverage is expected to be pathologically similar.

In addition, in those poorly covered regions, the input anomalies tend to be more stretched (smeared and elongated) in the DLS filtered models than in the SOLA ones (see Figures 2b, 2e, and 2h and Figures 2c, 2f, and 2i, respectively). This is expected, since the shape of DLS kernels is often more stretched (see section 3.1.3). Finally, in these poorly sampled regions, the DLS models suffer from both the averaging bias effect and some enhanced stretching effect, so that the input structures are better retrieved in the SOLA models.

3.1.5. Remarks

One wonders whether the averaging bias effect could be diminished by using an irregular, data-driven, local parametrization, so that the model parameters are sampled more uniformly by the data. Despite using such an adaptive grid, Zaroli (2016) reports significant differences between some global, real-data based DLS and SOLA tomographic models, indicating that the parametrization may not make up for all the bias.

The bias effect cannot simply be undone by scaling DLS model estimates by the averaging bias quantities. In general, such scaled solutions (i.e., $\tilde{m}_k/\tilde{U}_k$) would not be meaningful in terms of model interpretations (e.g., Nolet, 2008).

Though the SOLA method is not specifically aimed at minimizing the data misfit (see Table 1), we report that for the toy problem all five SOLA model estimates ($\hat{\mathbf{m}}$) do fit their corresponding data sets at the level of reduced chi-square equal to 1, as do the DLS models ($\tilde{\mathbf{m}}$).

Finally, in terms of model estimates, the SOLA tomographic recipe (Zaroli, 2016) performs much better than (as well as) the DLS approach in the regions of poor (good) data coverage, respectively.

3.2. Global Mantle Tomography: Uncovering the Bias Effect

We are interested in quantifying the averaging bias effect in real-data-based DLS models and focus on global mantle tomography. When considering a large number of data and model parameters, computing the full model resolution matrix, required to evaluate the bias effect, can become computationally challenging (see section 2.3.2). Thus, in most global tomographic studies, the resolution matrix has often been ignored, or at best its diagonal elements have been approximated (e.g., Rawlinson & Spakman, 2016; Trampert et al., 2013). We consider model S40RTS, one of the few models for which the resolution matrix, $\tilde{\mathbf{R}}_{S40}$, was fully calculated (Ritsema et al., 2011). S40RTS consists of isotropic, 3-D shear wave velocity variations in the whole mantle, resulting from a joint DLS inversion of surface wave, body wave, and normal mode data.

In particular, we aim to investigate how much and where S40RTS may be locally biased toward *higher* amplitudes. As already mentioned (sections 3.1.2 and 3.1.4), such bias effects are expected to happen in mantle regions below isolated receivers (e.g., located on oceanic islands), where body wave raypaths should predominantly be unidirectional (vertical). Indeed, such anisotropic ray coverages are similar to that encountered in the toy problem (lower right corner). As a remark, one could also consider other mantle regions such as below subduction zones or oceanic ridges, where most body waves are primarily vertically radiating from earthquakes.

Let us consider the input model $\mathbf{m}$ shown in Figures 3a, 3d, and 3g, which consists of 3-D shear wave velocity perturbations, spatially distributed like a Gaussian random field within the whole mantle, except in the Pacific and Antarctica regions, where four very long wavelength low-velocity anomalies are superimposed (much longer than the local resolving lengths), centered in Hawaii, Tahiti, Samoa, and Mountt Erebus. These broad-extent, slow features are radially extending throughout the entire mantle and laterally varying from dlnVs $\simeq -1.3\%$ at their center to 0 in a linear fashion over 40° great-circle distance (see Figure 3d). Some isolated (groups of) receivers, for which there are many recorded arrival times in the S40RTS data set, appear to be located at these four locations (see large-size white triangles in Figure 3a).

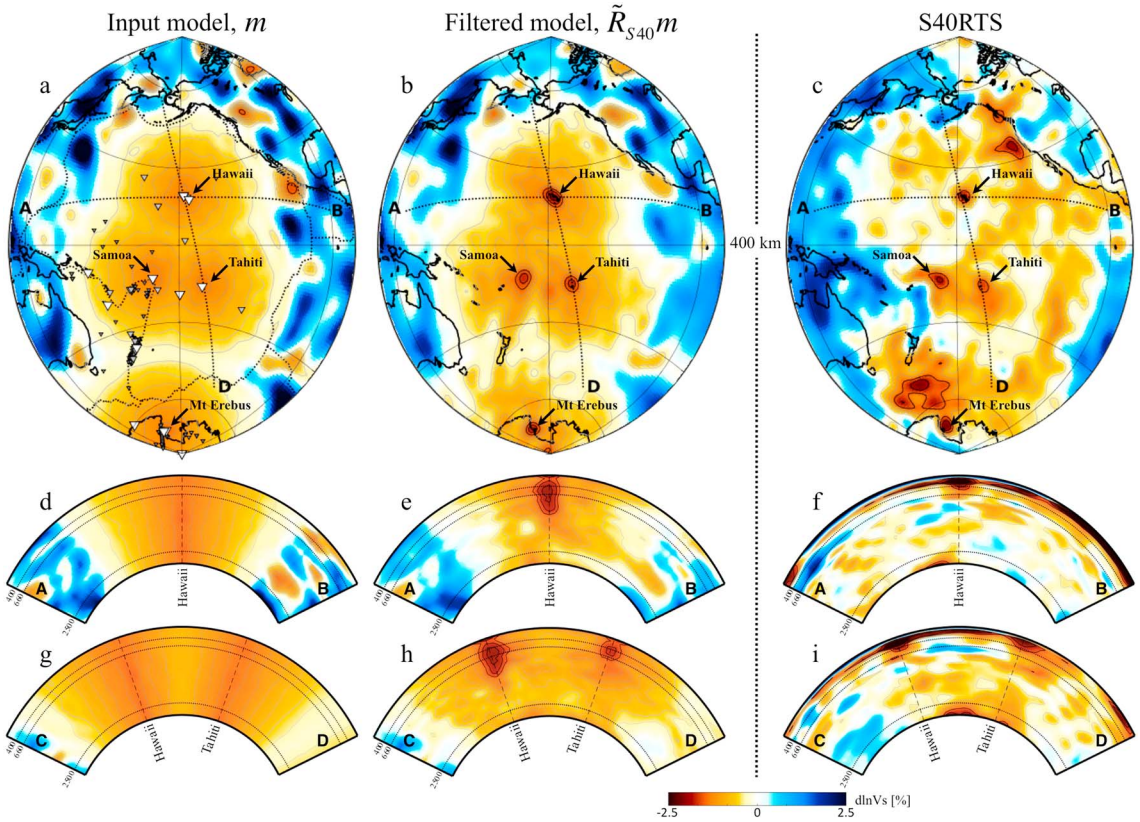

Figure 3. Quantitative illustration of the local averaging bias effect in model S40RTS. (a, d, g) Input model, **m**; (b, e, h) filtered model, $\tilde{\mathbf{R}}_{S40}\mathbf{m}$; (c, f, i) model S40RTS. In Figure 3a large-size white (medium-size gray, small-size black) triangles denote all receivers located within the Pacific or Antarctica regions for which there is a large (moderate, small) number of recorded body wave arrival times in S40RTS data set, respectively.

The filtered model $\tilde{\mathbf{R}}_{S40}\mathbf{m}$ is shown in Figures 3b, 3e, and 3h. Quite remarkably, note the four *darker* red anomalies showing up at all four locations, for instance, at 400 km depth (see Figure 3b). Amplitudes have been locally enhanced, up to 50% higher values (dlnVs $\simeq -2\%$), thus creating some artificial, "hot spot-like" features (see Figures 3e and 3h). Below Hawaii (Tahiti), this bias effect is significant from the near surface down to 1200 (700) km depth — amplitudes are increased by 10–50% (10–25%), respectively.

Model S40RTS is shown in Figures 3c, 3f, and 3i to be compared with filtered model $\tilde{\mathbf{R}}_{S40}\mathbf{m}$. One should bear in mind that S40RTS is the sum of the filtered (unknown) true-mantle model and the propagated (poorly known) data errors. Four low velocity features are visible in S40RTS, for instance at 400 km depth, below Hawaii, Tahiti, Samoa, and Mount Erebus (see Figure 3c) — often referred to as "hot spots" (e.g., Montelli et al., 2006).

The apparent Hawaii and Tahiti hot spots (see Figures 3f and 3i) overlap with the mantle regions where the bias effect is significant (see Figures 3e and 3h), which then raises the following question: What is the actual imprint of the bias effect on these apparent hot spot features? Recall that the bias signature also depends on what the input true model is and how it spatially relates to the resolving kernels (see Figure 2, second column and section 3.1.4). Most importantly, could this bias imprint be misleading in terms of physical interpretations? In other words, could the apparent low shear wave velocity structures, interpreted as the Hawaii and Tahiti hot spots, to some extent be questioned?

Although there are other robust geophysical, geochemical, or geological evidences favoring such hot spot features in the mantle (e.g., Courtillot et al., 2003), our results should at least be a "quantitative reminder" that the appraisal of DLS tomographic images should definitely be more quantitative, including quantifying these bias effects.

Finally, we have shown how crucial it could be to quantify the averaging bias effect in real-data-based DLS models to avoid physical misinterpretations (if any). As a remark, this bias effect should at least be accounted

for by "tomographically filtering" input models, for example, when comparing geodynamic and tomographic structures (e.g., Davies et al., 2012; Ritsema et al., 2007; Schuberth et al., 2009).

4. Conclusion

In this study, we have demonstrated that the SOLA Backus-Gilbert inversion approach (Zaroli, 2016) is better suited than damped least squares (DLS) methods for solving large-scale, linear(ized), discrete tomographic problems, that is,

1. DLS models may be locally biased toward lower or higher amplitudes in regions of poor data illumination, potentially causing physical misinterpretations.
2. SOLA models are explicitly constrained to be (unbiased) averages over the true-model parameters.
3. The extremely parallel SOLA scheme is more efficient for computing the generalized inverse—required to infer the model estimate, resolution and covariance.

Since uneven data coverage has been, and will remain, a serious issue in seismic tomography, from local to global scale, such averaging bias effect should systematically be investigated in DLS models. For example, we have shown that global model S40RTS is locally biased toward up to 50% higher amplitudes below isolated receivers where raypaths are quasi-vertical, such as on Hawaii—which leads to questions on the apparent low shear wave velocity structure interpreted as the Hawaii hot spot.

Finally, the way is now clearly open for seeing the Earth's interior through *unbiased*, SOLA tomographic lenses—including quantitative model interpretations.

Acknowledgments
The authors are grateful to the Editor Andrew Newman and to Nick Rawlinson and two anonymous reviewers whose constructive comments have helped to improve this paper. No new data or models are presented in this theoretical contribution. The S40RTS filtering code and model are available from Jeroen Ritsema/Paula Koelemeijer, and the computer codes for the SOLA inversions are available from the corresponding author—upon reasonable request. C. Z. is grateful to D. Al-Attar, A. Maggi, G. Nolet, B. S. A. Schuberth, and S. Voronin for stimulating discussions at various stages of this research. Most of this manuscript was written while C. Z. was on sabbatical leave (CRCT, CNU Section 35) at the University of Oxford, United Kingdom.

References

Aster, R. C., Borchers, B., & Thurber, C. (2012). *Parameter estimation and inverse problems* (2nd ed.). Oxford, UK: Elsevier.
Backus, G., & Gilbert, J. F. (1967). Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal of Royal Astronomical Society*, *13*, 247–276.
Backus, G., & Gilbert, J. F. (1968). The resolving power of gross Earth data. *Geophysical Journal of Royal Astronomical Society*, *16*, 169–205.
Backus, G., & Gilbert, J. F. (1970). Uniqueness in the inversion of inaccurate gross Earth data. *Philosophical Transactions of the Royal Society A*, *266*(1173), 123–192.
Bolton, H., & Masters, G. (2001). Traveltimes of *P* and *S* from global digital seismic networks: Implication for the relative variation of *P* and *S* velocity in the mantle. *Journal of Geophysical Research*, *106*, 13,527–13,540.
Burdick, S., & Lekic, V. (2017). Velocity variations and uncertainty from transdimensional *P*-wave tomography of North America. *Geophysical Journal International*, *209*, 1337–1351.
Courtillot, V., Davaille, A., Besse, J., & Stock, J. (2003). Three distinct types of hotspots in the Earth's mantle. *Earth and Planetary Science Letters*, *205*, 295–308.
Dahlen, F. A., Hung, S.-H., & Nolet, G. (2000). Fréchet kernels for finite-frequency traveltimes—I. Theory. *Geophysical Journal International*, *141*, 157–174.
Davies, D. R., Goes, S., Davies, J. H., Schuberth, B. S. A., Bunge, H.-P., & Ritsema, J. (2012). Reconciling dynamic and seismic models of Earth's lower mantle: The dominant role of thermal heterogeneity. *Earth and Planetary Science Letters*, *353-354*, 253–269.
Debayle, E., & Ricard, Y. (2012). A global shear velocity model of the upper mantle from fundamental and higher Rayleigh modeeter estimation and inverse prob measurements. *Journal of Geophysical Research*, *117*, B10308. https://doi.org/10.1029/2012JB009288
Fichtner, A., Kennett, B. L. N., & Igel, H. (2009). Full waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. *Geophysical Journal International*, *179*, 1703–1725.
Fichtner, A., & Trampert, J. (2011). Resolution analysis in full waveform inversion. *Geophysical Journal International*, *187*, 1604–1624. https://doi.org/10.1111/j.1365-246X.2011.05218.x
Koelemeijer, P., Ritsema, R., Deuss, A., & Van Heijst, H.-J. (2016). SP12RTS: A degree-12 model of shear- and compressional-wave velocity for Earth's mantle. *Geophysical Journal International*, *204*(2), 1024–1039.
Menke, W. (1989). *Geophysical data analysis: Discrete inverse theory* (3rd ed.). San Diego, CA: Academic Press.
Montelli, R., Nolet, G., Dahlen, F. A., & Masters, G. (2006). A catalogue of deep mantle plumes: New results from finite-frequency tomography. *Geochemistry, Geophysics, Geosystems*, *7*, Q11007. https://doi.org/10.1029/2006GC001248
Nolet, G. (1985). Solving or resolving inadequate and noisy tomographic systems. *Journal of Computational Physics*, *61*, 463–482.
Nolet, G (2008). *A breviary of seismic tomography*. Cambridge, UK: Cambridge University Press.
Paige, C., & Saunders, M. (1982). LSQR: An algorithm for sparse, linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, *8*, 43–71.
Pijpers, F., & Thompson, M. (1992). Faster formulations of the optimally localized averages method for helioseismic inversions. *Astronomy and Astrophysics*, *262*, L33–L36.
Pijpers, F., & Thompson, M. (1994). The SOLA method for helioseismic inversion. *Astronomy and Astrophysics*, *281*, 231–240.
Rawlinson, N., Fichtner, A., Sambridge, M., & Young, M. K. (2014). Seismic tomography and the assessment of uncertainty. *Advances in Geophysics*, *55*, 1–76.
Rawlinson, N., & Spakman, W. (2016). On the use of sensitivity tests in seismic tomography. *Geophysical Journal International*, *205*, 1221–1243.
Resovsky, J., & Ritzwoller, M. H. (1999). A degree 8 mantle shear velocity model from normal mode observations below 3mhz. *Journal of Geophysical Research*, *104*(B1), 993–1014.
Ritsema, J., Deuss, A., van Heijst, H.-J., & Woodhouse, J. H. (2011). S40RTS: A degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. *Geophysical Journal International*, *184*, 1223–1236.

Ritsema, J., McNamara, A. K., & Bull, A. L. (2007). Tomographic filtering of geodynamic models: Implications for model interpretation and large-scale mantle structure. *Journal of Geophysical Research*, *112*, B01303. https://doi.org/10.1029/2006JB004566

Romanowicz, B. (2003). Global mantle tomography: Progress status in the past 10 years. *Annual Review of Earth and Planetary Sciences*, *31*, 303–328.

Sambridge, M., Bodin, T., Gallagher, K., & Tkalcic (2013). Transdimensional inference in the geosciences. *Philosophical Transactions of the Royal Society A*, *371*(1984), 20110547.

Schuberth, B. S. A., Bunge, H.-P., & Ritsema, J. (2009). Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone? *Geochemistry, Geophysics, Geosystems*, *10*, Q05W03. https://doi.org/10.1029/2009GC002401

Sigloch, K. (2011). Mantle provinces under North America from multifrequency *P* wave tomography. *Geochemistry, Geophysics, Geosystems*, *12*, Q02W08. https://doi.org/10.1029/2010GC003421

Trampert, J., Fichtner, A., & Ritsema, J. (2013). Resolution tests revisited: The power of random numbers. *Geophysical Journal International*, *192*(2), 676–680.

Zaroli, C. (2016). Global seismic tomography using Backus–Gilbert inversion. *Geophysical Journal International*, *207*(2), 876–888.

Zaroli, C., Lambotte, S., & Lévêque, J.-J. (2015). Joint inversion of normal-mode and finite-frequency *S*-wave data using an irregular tomographic grid. *Geophysical Journal International*, *203*(3), 1665–1681.

Zigone, D., Ben-Zion, Y., Campillo, M., & Roux, P. (2015). Seismic tomography of the Southern California plate boundary region from noise-based Rayleigh and Love waves. *Pure and Applied Mathematics*, *172*, 1007–1032.