Geophysical Journal International

Geophys. J. Int. (2016) **207,** 876–888
Advance Access publication 2016 August 23
GJI Seismology

doi: 10.1093/gji/ggw315

Global seismic tomography using Backus–Gilbert inversion

Christophe Zaroli

Institut de Physique du Globe de Strasbourg, UMR 7516*, Université de Strasbourg, EOST/CNRS, France. E-mail: c.zaroli@unistra.fr*

Accepted 2016 August 19. Received 2016 August 18; in original form 2016 May 12

SUMMARY

The appraisal of tomographic models, of fundamental importance towards better understanding the Earth's interior, consists in analysing their resolution and covariance. The discrete theory of Backus–Gilbert, solving all at once the linear problems of model estimation and appraisal, aims at evaluating weighted averages of the true model parameters. Contrary to damped least-squares techniques, one key advantage of Backus–Gilbert inversion is that no subjective regularization is needed to remove the non-uniqueness of the model solution. Indeed, it is often possible to identify unique linear combinations of the parameters even when the parameters themselves are not uniquely defined. In other words, the non-uniqueness can be broken by averaging rather than regularizing. Over the past few decades, many authors have considered that, in addition to a high computational cost, it could be a clumsy affair in the presence of data errors to practically implement the Backus–Gilbert approach to large-scale tomographic applications. In this study, we introduce and adapt to seismic tomography the Subtractive Optimally Localized Averages (SOLA) method, an alternative Backus–Gilbert formulation which retains all its advantages, but is more computationally efficient and versatile in the explicit construction of averaging kernels. As a leitmotiv, we focus on global-scale *S*-wave tomography and show that the SOLA method can successfully be applied to large-scale, linear and discrete tomographic problems.

Key words: Inverse theory; Tomography; Body waves.

1 INTRODUCTION

A detailed knowledge of the magnitude and length scales of seismic attributes of the Earth's mantle is crucial to better understand its physical properties and multiscale dynamics (e.g. Ritsema *et al.* 2007; Simmons *et al.* 2009; Davies *et al.* 2012; Schuberth *et al.* 2012). Though there may be some consensus on long wavelength features of global-scale tomographic models, seismic imaging of shorter scale-length heterogeneities, such as sinking slabs or hot-rising plumes, is usually more controversial. Therefore, a robust analysis of the model resolution and covariance has now become of fundamental importance to push further the physical interpretations of tomographic models (e.g. Trampert 1998).

Damped least-squares (DLS) techniques have been widely used to solve large-scale, linear and discrete tomographic problems (e.g. Aster *et al.* 2012). In the DLS approach, the non-uniqueness of the solution is usually broken by adding to the inverse problem some *subjective* regularization constraints (e.g. model norm damping), often aimed at favouring the model simplicity. This is not to be confused with *objective* regularization constraints added within a Bayesian framework. Thus, the DLS model estimate, resolution and covariance are dependent on these subjective regularization constraints. The DLS model resolution and covariance (or 'uncertainty') can either be fully calculated (e.g. Boschi 2003; Soldati & Boschi 2005; Ritsema *et al.* 2007, 2011), or simply approximated if the tomographic systems are too large (e.g. Nolet *et al.* 1999). Nevertheless, in most of large-scale tomographic studies, resolution analysis has often been reduced to checker-board sensitivity tests, coming with a number of potential drawbacks (e.g. Lévêque *et al.* 1993; Rawlinson & Spakman 2016), and uncertainty assessment has often been ignored, or at best given minimal treatment (e.g. Rawlinson *et al.* 2014).

A fundamentally different approach is that of Backus–Gilbert (B-G; e.g. Backus & Gilbert 1967, 1968, 1970), which belongs to the class of Optimally Localized Averages (OLA) methods. It solves all at once the linear problems of model estimation and appraisal, while a direct control can be operated on the model resolution and covariance. The theory of B–G aims at determining what are the properties of the *continuous* model space that can be estimated given the data and their errors. In principle, there is no need to restrict attention to a particular subspace of the full model space. However, in this study, we shall follow Nolet (1985) and assume some *discrete*, local model parametrization, mainly as a numerical step to perform the calculations and to lessen the computational cost. In the discrete theory of B–G, one aims at evaluating linear combinations of the *true* model parameters. That is, one seeks solutions which are not estimates of the true model parameters themselves, but estimates of weighted averages of the true parameters. The

advantage of seeking to evaluate 'averages' of parameters, rather than 'estimates' of the parameters themselves, is that unique linear combinations of the parameters can often be identified even when the parameters are not uniquely defined (e.g. Menke 1989). In other words, the non-uniqueness of the problem can be removed by the process of averaging rather than by adding subjective regularization constraints. This is a key advantage of B–G compared to DLS. Indeed, the B–G solution refers to the true model parameters, while the DLS solution refers to the 'damped' true model parameters – hence making its physical interpretation with respect to the true model parameters more complicated.

As pointed out by Nolet (2008), '*though the Backus–Gilbert method receives much interest in helioseismology, it is woefully underused in terrestrial applications*'. Over the past few decades, many authors have considered that, in addition to a high computational cost, it could be a clumsy affair in the presence of data errors to practically implement the B–G method to large-scale tomographic applications (e.g. Menke 1989; Parker 1994; Trampert 1998; Aster *et al.* 2012). To the best of our knowledge, the B–G approach has never been applied to large-scale tomographic problems ($M \simeq 10^4$–10^5 parameters, $N \simeq 10^5$–10^6 data).

The goal of this study is to show that a variant of the B–G method can successfully be applied to large-scale, linear and discrete tomographic problems. For this purpose, we shall introduce and adapt to seismic tomography the Subtractive OLA (SOLA) method, an alternative formulation used in helioseismic inversions which retains all the advantages of the B–G approach, but is more computationally efficient and versatile in the explicit construction of averaging kernels (e.g. Pijpers & Thompson 1992, 1994). As a leitmotiv, we shall focus on the case of global-scale *S*-wave tomography and report on the obtained tomographic results (i.e. model estimation and appraisal) in the 400–1710 km depth range, where our data coverage is the most relevant.

2 BACKUS–GILBERT INVERSION

In the following, we aim at applying the B–G approach to large-scale, linear and discrete tomographic problems. Without loss of generality, we shall assume a local model parametrization.

2.1 Preamble

2.1.1 Linear model estimation and appraisal

Using matrix notations, a linear forward problem can be written as

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

where $\mathbf{d} = (d_i)_{1\leq i\leq N}$ denotes the data vector, $\mathbf{n} = (n_i)_{1\leq i\leq N}$ the noise in the data, $\mathbf{m} = (m_j)_{1\leq j\leq M}$ the true model parameters, and $\mathbf{G} = (G_{ij})$ the sensitivity matrix of size $N \times M$. Having in mind the tomographic application in Section 3, we consider in this Section 2 that the model parameters represent 3-D velocity anomalies, with respect to a 1-D reference model, and that the data are time-residuals. In this study, we shall assume that the noise $\mathbf{n}$ has zero mean, and that the data covariance matrix, $\mathbf{C}_d$, can be expressed as

$$\mathbf{C}_d = \mathrm{diag}\left(\sigma_{d_i}^2\right), \qquad 1 \leq i \leq N\,. \tag{2}$$

One aims at estimating a model solution, $\tilde{\mathbf{m}} = (\tilde{m}_j)_{1\leq j\leq M}$, which can be expressed as a linear combination of the data:

$$\tilde{\mathbf{m}} = \mathbf{G}^\dagger\mathbf{d}\,, \tag{3}$$

where the matrix $\mathbf{G}^\dagger$ denotes some generalized inverse operator (e.g. Snieder & Trampert 1999). The model estimate $\tilde{\mathbf{m}}$ can be related to the true model $\mathbf{m}$ as follows:

$$\tilde{\mathbf{m}} = \mathbf{R}\mathbf{m} + \mathbf{G}^\dagger\mathbf{n}\,, \tag{4}$$

where $\mathbf{R}$ denotes the model resolution matrix, defined as

$$\mathbf{R} = \mathbf{G}^\dagger\mathbf{G}\,. \tag{5}$$

In the case where the resolution matrix $\mathbf{R}$ is not the identity $\mathbf{I}_M$, the eq. (4) shows that the model estimate $\tilde{\mathbf{m}}$ results both from the imperfect ability to recover the true model $\mathbf{m}$ even with error-free data (i.e. the term $\mathbf{R}\mathbf{m}$) and from the propagation of data errors into the solution estimate (i.e. the term $\mathbf{G}^\dagger\mathbf{n}$). The appraisal problem consists in estimating and analysing both the resolution matrix $\mathbf{R}$ and the model covariance matrix, $\mathbf{C}_{\tilde{m}}$, defined as

$$\mathbf{C}_{\tilde{m}} = \mathbf{G}^\dagger\mathbf{C}_d(\mathbf{G}^\dagger)^{\mathrm{T}}, \tag{6}$$

where $(\cdot)^{\mathrm{T}}$ is the transpose operator. In the rest of this Section 2, we shall consider that the data vector $\mathbf{d}$ and the sensitivity matrix $\mathbf{G}$ have been scaled by the data errors σ_{d_i}.

2.1.2 Damped least-squares

The least-squares approach consists in finding a model solution that minimizes the data misfit. However, there is an infinity of solutions with the same data misfit. Thus, to remove the non-uniqueness of the solution, the DLS problem consists in finding a model that minimizes both the data misfit *and* the model 'complexity'. This may be achieved by minimizing the quantity:

$$||\mathbf{d} - \mathbf{G}\tilde{\mathbf{m}}||^2 + \Theta^2||\tilde{\mathbf{m}}||^2 \;=\; \min\,, \tag{7}$$

where $||\cdot||$ denotes the L^2-norm, commonly used to measure the model complexity (i.e. zeroth-order Tikhonov regularization), and $\Theta \in \mathbb{R}_+$ is a regularization parameter that controls the damping of the model norm. For simplicity reasons, the model parametrization is assumed to consist in orthonormal basis functions in this Section 2.1.2, such that the norm of the discretized model $\tilde{\mathbf{m}}$, involved in eq. (7), is equal to the norm of the continuous model $\tilde{m}(\mathbf{r})$. The choice of an adequate value for Θ is not an easy one. In the DLS approach, one has to specify the generalized inverse to compute the model resolution and covariance. In the case of zeroth-order Tikhonov regularization, the generalized inverse can be written as

$$\mathbf{G}^\dagger_{\mathrm{DLS}} = \left\{\mathbf{G}^{\mathrm{T}}\mathbf{G} + \Theta^2\mathbf{I}_M\right\}^{-1}\mathbf{G}^{\mathrm{T}}\,, \tag{8}$$

where $\{\cdot\}^{-1}$ denotes the inverse operator. Methods based on Singular Value Decomposition can be used to evaluate $\mathbf{G}^\dagger_{\mathrm{DLS}}$, though powerful computational resources are required for large-scale problems (e.g. Boschi 2003; Soldati & Boschi 2005; Ritsema *et al.* 2007, 2011). Note that the generalized inverse $\mathbf{G}^\dagger_{\mathrm{DLS}}$ depends on Θ, so do the model estimate, the resolution matrix, and the model covariance. Thus, the resolution matrix $\mathbf{R}_{\mathrm{DLS}} = \mathbf{G}^\dagger_{\mathrm{DLS}}\mathbf{G}$ does not represent how resolvable by the data is the true model, but how resolvable is the 'damped' true model. As the effect of damping may differ for different model parameters, this may complicate the physical interpretation of the DLS model estimate based on $\mathbf{R}_{\mathrm{DLS}}$.

2.2 Backus–Gilbert and the SOLA method

Without loss of generality, we shall consider that the local model parametrization consists in 'nodes'. The node k will correspond to the kth model parameter, and $\mathbf{r}^{(k)}$ will denote its spatial location.

2.2.1 Usual Backus–Gilbert formulation

Rather than seeking an estimate $\tilde{m}_k$ for each individual true model parameter m_k, the heart of B–G is to seek a weighted average of the true parameters, hereafter denoted as $\hat{m}_k$, localized around the location $\mathbf{r}^{(k)}$. Taking advantage of the linearity of the forward problem and of the averaging process, the model estimate $\hat{m}_k$ can be expressed as a linear combination of the data:

$$\hat{m}_k = \sum_{i=1}^{N} x_i^{(k)} d_i \,, \tag{9}$$

where the N unknown coefficients $x_i^{(k)}$ can be viewed as the analogue of the kth row of the generalized inverse operator in eq. (3). For each node k, the B–G approach consists in directly estimating the N coefficients $x_i^{(k)}$, from which it is straightforward to obtain the model estimate $\hat{m}_k$ with its corresponding variance $\sigma^2_{\hat{m}_k}$ and averaging kernel $A^{(k)}(\mathbf{r})$. The model variance $\sigma^2_{\hat{m}_k}$ is given by:

$$\sigma^2_{\hat{m}_k} = \sum_{i=1}^{N} \left(x_i^{(k)} \sigma_{d_i} \right)^2 . \tag{10}$$

Since the data have been scaled by their errors, the model variance simplifies to $\sum_i (x_i^{(k)})^2$. The estimate $\hat{m}_k$ can be expressed as

$$\hat{m}_k = \sum_{j=1}^{M} \left(\sum_{i=1}^{N} x_i^{(k)} G_{ij} \right) m_j + \sum_{i=1}^{N} x_i^{(k)} n_i \,, \tag{11}$$

where, on the right side, the first term corresponds to a weighted averaging over the true model parameters and the second term to the propagation of data errors. The B–G point-of-view is that for error-free data the estimate $\hat{m}_k$ can be expressed as

$$\hat{m}_k = \int A^{(k)}(\mathbf{r})\, m(\mathbf{r})\, \mathrm{d}^3\mathbf{r} = \sum_{j=1}^{M} \left(\mathbb{V}_j A_j^{(k)} \right) m_j \,, \tag{12}$$

where $A^{(k)}(\mathbf{r})$, hereafter $A^{(k)}$ for short, is the corresponding averaging kernel (also referred to as the resolving kernel). The continuous kernel $A^{(k)}$ is defined from its M discretized values $A_j^{(k)}$, using the same interpolation rule as for the model interpolation (see eq. 25). The term $\mathbb{V}_j$ corresponds to the 'volume' associated to the jth node of the local parametrization (see eq. 26). The M components $A_j^{(k)}$ are related to the N coefficients $x_i^{(k)}$:

$$A_j^{(k)} = \frac{1}{\mathbb{V}_j} \sum_{i=1}^{N} x_i^{(k)} G_{ij} \,, \tag{13}$$

and can be interpreted as the kth row of a resolution matrix $\hat{\mathbf{R}}$:

$$\hat{R}_{kj} = \mathbb{V}_j A_j^{(k)}, \qquad 1 \le j \le M \,. \tag{14}$$

We wish that (12) be an *unbiased* averaging, that is, $\sum_j \hat{R}_{kj}$ must be equal to one, such that $\hat{m}_k = \sum_j \hat{R}_{kj} m_j$ represents a *true physical averaging* over the true model parameters (e.g. Nolet 2008). Thus, the averaging kernel $A^{(k)}$ is imposed to be 'unimodular':

$$\int A^{(k)}(\mathbf{r})\, \mathrm{d}^3\mathbf{r} = \sum_{j=1}^{M} \mathbb{V}_j A_j^{(k)} = 1 \,. \tag{15}$$

For each model estimate $\hat{m}_k$, we wish that all the elements $\hat{R}_{kj}$, that is, the elements $A_j^{(k)}$, be almost zero if the nodes k and j are spatially distant beyond the local resolving length at spatial position $\mathbf{r}^{(k)}$. In the presence of data errors, the B–G method consists in constructing the most peak-shaped resolving kernel $A^{(k)}$, peaked around $\mathbf{r}^{(k)}$, while moderating at most the propagation of data errors into the model estimate. Thus, one aims at minimizing both the spatial spread of $A^{(k)}$ and the model variance $\sigma^2_{\hat{m}_k}$. These are two opposite aims: the well-known trade-off between resolution and variance. Let $W^{(k)} \in \mathbb{R}_+$ be a number that represents the spatial spread of the kernel $A^{(k)}$, and $\eta \in \mathbb{R}_+$ be a trade-off parameter. Thus, for a given value of η, one seeks the set of N coefficients $x_i^{(k)}$ that minimizes the following quantity:

$$W^{(k)} + \eta^2 \sigma^2_{\hat{m}_k} = \min \,, \quad \text{[s.t. unimodular]} \,, \tag{16}$$

where [s.t. unimodular] means 'subject to the constraint of unimodular averaging kernel, that is, eq. (15)'. Usually, $W^{(k)}$ is defined as

$$W^{(k)} = \int J^{(k)}(\mathbf{r}) \left[A^{(k)}(\mathbf{r}) \right]^2 \mathrm{d}^3\mathbf{r} \,, \tag{17}$$

where the function $J^{(k)}$ is designed to favour a peak-shaped kernel $A^{(k)}$ around the spatial position $\mathbf{r}^{(k)}$. This means giving any contribution to the integral of $[A^{(k)}]^2$ away from $\mathbf{r}^{(k)}$ a large weight. For this purpose, it is common practice to use (e.g. Nolet 1985):

$$J^{(k)}(\mathbf{r}) \propto |\mathbf{r} - \mathbf{r}^{(k)}|^4 \,. \tag{18}$$

This minimization problem, that is, eqs (16)–(18), leads to solving for $\mathbf{x}^{(k)}$ the following set of normal equations:

$$\left(\hat{\mathbf{G}}^{(k)} \left(\hat{\mathbf{G}}^{(k)} \right)^{\mathrm{T}} + \eta^2 \mathbf{I}_N \right) \mathbf{x}^{(k)} = \mathbf{0}_N \,, \ \text{[s.t. unimodular]} \,, \tag{19}$$

where the matrix $\hat{\mathbf{G}}^{(k)}$ and the column-vector $\mathbf{x}^{(k)}$ are defined as

$$\hat{\mathbf{G}}^{(k)} = \left(\frac{|\mathbf{r}^{(j)} - \mathbf{r}^{(k)}|^2}{\sqrt{\mathbb{V}_j}} G_{ij} \right) , \quad \mathbf{x}^{(k)} = \left(x_i^{(k)} \right) , \tag{20}$$

with $1 \le i \le N$, and $1 \le j \le M$. The method of Lagrange multipliers can be used to solve eq. (19) for every node k. Since the matrix $\hat{\mathbf{G}}^{(k)}$ depends on the spatial location $\mathbf{r}^{(k)}$ of the considered node k, it has to be recomputed for each one of the M nodes. For large-scale tomographic problems, as we are interested in, this is a significant computational drawback of the usual B–G formulation.

2.2.2 The SOLA method

We introduce and adapt to seismic tomography the SOLA method (e.g. Pijpers & Thompson 1992, 1994), an alternative B–G approach used in helioseismic inversions (e.g. Pijpers 1997; Rabello-Soares *et al.* 1999; Jackiewicz *et al.* 2012). The key idea is to specify an *a priori* 'target form' for the averaging kernel $A^{(k)}$. Let $T^{(k)}$ denotes such a *target kernel*. Rather than minimizing the spread of $A^{(k)}$, as in the usual B–G formulation, one aims at minimizing the integrated squared difference between $A^{(k)}$ and $T^{(k)}$. Pijpers & Thompson (1992) termed this alternative the 'SOLA' method, though Larsen & Hansen (1997) suggest that it may have been rediscovered independently by different authors (e.g. Louis & Maass 1990; Pijpers & Thompson 1992). The SOLA minimization problem can be written as:

$$\int \left[A^{(k)}(\mathbf{r}) - T^{(k)}(\mathbf{r}) \right]^2 \mathrm{d}^3\mathbf{r} + \eta^2 \sigma^2_{\hat{m}_k} = \min, \ \text{[s.t. unimodular]} \tag{21}$$

where the target kernel $T^{(k)}(\mathbf{r})$, hereafter $T^{(k)}$ for short, is defined from its M components $T_j^{(k)}$ using the same interpolation rule as for the model interpolation. The components $T_j^{(k)}$ result from the projection onto the model parametrization of the 'true' target kernel

$T^{(k)}_{\rm true}(\mathbf{r})$, hereafter $T^{(k)}_{\rm true}$ for short. Let $\mathbb{S}^{(k)}$ be the spatial domain where $T^{(k)}_{\rm true}$ is constant inside and zero outside. If $T^{(k)}_{\rm true}$ was a 3-D ball centred in $\mathbf{r}^{(k)}$, its radius would correspond to the minimum resolving length that could, at best, be expected in $\mathbf{r}^{(k)}$. The minimization problem (21) leads to solving the set of normal equations:

$$\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^{\rm T} + \eta^2 \mathbf{I}_N\right)\mathbf{x}^{(k)} = \hat{\mathbf{G}}\mathbf{t}^{(k)}, \qquad [\text{s.t. unimodular}]\,, \tag{22}$$

where the matrix $\hat{\mathbf{G}}$ and the column-vector $\mathbf{t}^{(k)}$ are defined as

$$\hat{\mathbf{G}} = \left(\frac{1}{\sqrt{\mathbb{V}_j}}G_{ij}\right), \qquad \mathbf{t}^{(k)} = \left(T^{(k)}_j\sqrt{\mathbb{V}_j}\right), \tag{23}$$

with $1 \le i \le N$, and $1 \le j \le M$. Though such systems are usually solved using Lagrange multipliers, Nolet (1985) reports that this may lead to an ill-conditioned system of linear equations, and proposes an alternative, more stable approach based on the LSQR algorithm (e.g. Paige & Saunders 1982; Grunberg 2006), that we adopt in this study (see Appendix A1). Computationally speaking, the most important point is that the matrix $\hat{\mathbf{G}}$ does *not* depend on the spatial location $\mathbf{r}^{(k)}$; no need to recompute it for every node k.

3 GLOBAL TOMOGRAPHY WITH THE SOLA METHOD

Using the SOLA method, we aim at estimating isotropic 3–D variations of shear-velocities in the mantle, with respect to the radial reference model iasp91 (Kennett & Engdahl 1991), and are interested in quantitatively analysing the model resolution and uncertainty.

3.1 Statement of the tomographic problem

Innovative theoretical developments on the finite-frequency behaviour of the propagation of body waves (e.g. Dahlen *et al.* 2000; Tromp *et al.* 2005) have recently received increasing attention in seismic tomography (e.g. Montelli *et al.* 2004a; Sigloch *et al.* 2008; Zaroli *et al.* 2010; Tian *et al.* 2011; Mercerat *et al.* 2014; Schuberth *et al.* 2015; Zaroli *et al.* 2015). In this study, we shall exploit a globally distributed set of finite-frequency shear-wave delay times, which consists in *S* and *SS* seismic phases measured at 22 s central period (Zaroli *et al.* 2010); that is a subset of the data used in Zaroli *et al.* (2015). The corresponding forward problem is linear:

$$d_i = \int K_i(\mathbf{r})\, m(\mathbf{r})\, \mathrm{d}^3\mathbf{r} + n_i\,, \tag{24}$$

with d_i the ith delay-time, n_i the noise in d_i, m the 'true' model, and K_i the ith finite-frequency sensitivity kernel (Dahlen *et al.* 2000); K_i is numerically computed as in Zaroli *et al.* (2013). In this work, the total number of data is $N = 79\,765$, and the whole mantle is parametrized by $M = 38\,125$ parameters. We shall refer to $\tilde{m}_{\rm DLS}$ as the tomographic model obtained from the DLS inversion of the same data set. For the purposes of this study (i.e. estimating model uncertainties), the data errors σ_{d_i} that were used in Zaroli *et al.* (2015) are increased by 30 per cent such that the reduced chi-square (e.g. Nolet 2008) is one for $\tilde{m}_{\rm DLS}$.

In this study, we use the same local model parametrization as detailed in Zaroli *et al.* (2015). To make this paper self-contained, it is briefly summarized in the following. First, the mantle is radially divided into 18 spherical layers. The minimum radial resolving length that may potentially be achieved, at best, corresponds to the width of each layer. Here, it gradually varies from 100 km at the surface, 130 km in the transition-zone, and up to 200 km in the lower-mantle. Laterally, each layer is spanned with Delaunay triangulations whose spatial distributions are optimized according to ray density (see Fig. 1). The parametrization is then made up of spherical triangular prisms; let $\mathcal{T}_p$ denotes the pth prism. The three vertices on the top of each prism were referred to as 'nodes' in Zaroli *et al.* (2015). The minimum lateral resolving length that may potentially be achieved, at best, corresponds to the nodes spacing: ∼200–1000 km (see Fig. 1). From here on, the spatial location of the jth node is defined as: $\mathbf{r}^{(j)} = (\bar{r}^{(j)}, \theta^{(j)}, \phi^{(j)})$, where $\theta^{(j)}$ and $\phi^{(j)}$ are the co-latitude and longitude of node j, respectively, and $\bar{r}^{(j)}$ is the 'middle radius' of the layer to which node j belongs to. The spatial locations of the three nodes of a given prism correspond to three model parameters. It is convenient to index the parameters with respect to the prism $\mathcal{T}_p$, that is, $m_q^{\mathcal{T}_p}$, with $q = \{1, 2, 3\}$. Each parameter $m_q^{\mathcal{T}_p}$ can be mapped back to its globally indexed, unique, parameter m_j, such as $(\mathcal{T}_p, q) \to j$. The continuous model is linearly related to the model parameters as follows:

$$m(\mathbf{r}) = \sum_{q=1}^{3} b_q^{\mathcal{T}_p}(\mathbf{r})\, m_q^{\mathcal{T}_p}\,, \quad \forall \mathbf{r} \in \mathcal{T}_p\,, \tag{25}$$

where $b_q^{\mathcal{T}_p}$ denote barycentric coordinates. With our irregular grid, the volumetric term $\mathbb{V}_j$ associated to the j-th node is given by:

$$\mathbb{V}_j = \sum_{(\mathcal{T}_p, q)\to j} \int_{\mathbf{r}\in\mathcal{T}_p} b_q^{\mathcal{T}_p}(\mathbf{r})\, d^3\mathbf{r}\,, \tag{26}$$

and the elements of the sensitivity matrix $\mathbf{G}$ can be expressed as

$$G_{ij} = \sum_{(\mathcal{T}_p, q)\to j} \left(\int_{\mathbf{r}\in\mathcal{T}_p} K_i(\mathbf{r})\, b_q^{\mathcal{T}_p}(\mathbf{r})\, d^3\mathbf{r}\right), \tag{27}$$

where $\sum$ is a sum over all pairs $(\mathcal{T}_p, q)$ satisfying to $(\mathcal{T}_p, q) \to j$.

3.2 Target kernels and trade-off parameters

Our model parametrization has been tuned to the heterogeneous body-wave data coverage. Indeed, ray density may often be considered as a proxy for the local resolving length (e.g. Vasco *et al.* 2003). In this study, we report that some suitable form for a true target kernel $T^{(k)}_{\rm true}$ is that of a 3-D spheroid corresponding to the minimum lateral and radial resolving lengths that could potentially be expected in $\mathbf{r}^{(k)}$ with our irregular parametrization. This choice ensures that, for every node k, the estimation of the model ($\hat{m}_k$) and of its appraisal ($A^{(k)}$ and $\sigma_{\hat{m}_k}$) be constrained by several data. Such a spheroid-shape target kernel $T^{(k)}_{\rm true}$ is centred in $\mathbf{r}^{(k)}$; its horizontal and vertical radii are respectively defined as

$$\mathcal{L}^{(k)}_H = \frac{1}{\#\mathbb{N}_k}\sum_{j\in\mathbb{N}_k}\mathcal{D}_{jk}\,, \qquad \mathcal{L}^{(k)}_V = \left|\mathcal{Z}^{(k)}_+ - \mathcal{Z}^{(k)}_-\right|\,, \tag{28}$$

with $\mathbb{N}_k$ the set of natural-neighbour nodes with respect to k, $\mathcal{D}_{jk}$ the distance between j and k, $\#\mathbb{N}_k$ the cardinality of $\mathbb{N}_k$, and $\mathcal{Z}^{(k)}_+$ and $\mathcal{Z}^{(k)}_-$ the top and bottom depth of the layer to which node k belongs to, respectively. The spatial domain $\mathbb{S}^{(k)}$ over which such a target kernel $T^{(k)}_{\rm true}$ is defined can be expressed as

$$\left\{(x, y, z) \in \mathbb{R}^3,\ \frac{x^2}{(\mathcal{L}^{(k)}_H)^2} + \frac{y^2}{(\mathcal{L}^{(k)}_H)^2} + \frac{z^2}{(\mathcal{L}^{(k)}_V)^2} \le 1\right\}, \tag{29}$$

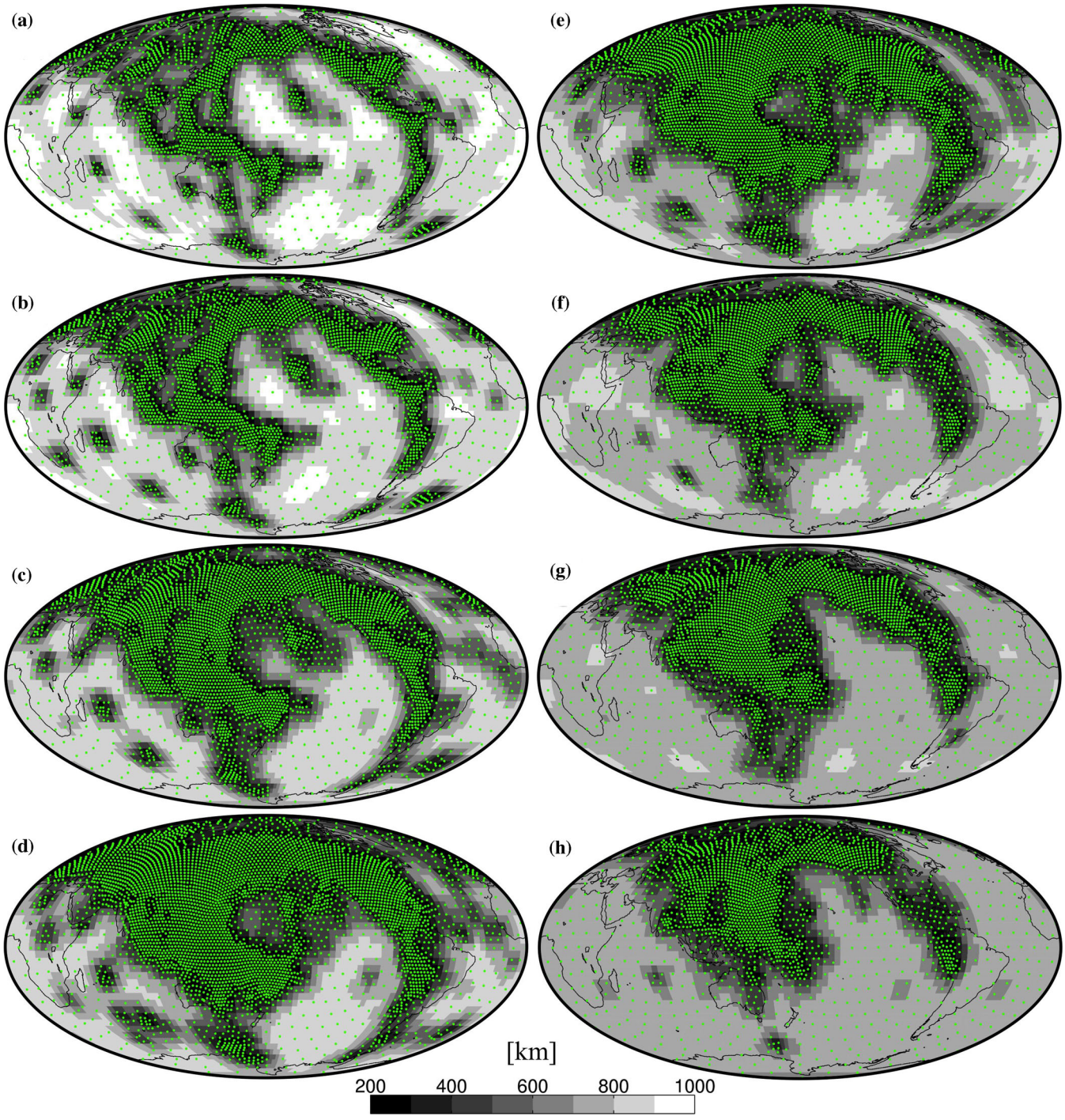

Figure 1. Spatial distribution of all the nodes (green dots) in the 400–1710 km depth range. The depths are (in km): (a) 400–530; (b) 530–660; (c) 660–810; (d) 810–960; (e) 960–1110; (f) 1110–1310; (g) 1310–1510; (h) 1510–1710. The black-to-white colourscale shows the corresponding variations of $\mathcal{L}_H^{(k)}$ (see Section 3.2).

where (x, y, z) denote coordinates with respect to the local Cartesian frame naturally associated to the node k. The spatial variations of $\mathcal{L}_H^{(k)}$ within the 400–1710 km depth range are shown in Fig. 1. Following Pijpers (1997), we choose to impose that the projected target kernel $T^{(k)}$ be unimodular, that is:

$$\int T^{(k)}(\mathbf{r})\,\mathrm{d}^3\mathbf{r} = \sum_{j=1}^{M} \mathbb{V}_j\, T_j^{(k)} = 1\,, \tag{30}$$

though this constraint is not strictly necessary. Therefore, with our irregular grid, the M components $T_j^{(k)}$ can simply be expressed as:

$$T_j^{(k)} = \frac{\mathbb{V}_j^{-1} \sum_{(\mathcal{T}_p,q)\to j} \int_{\mathbf{r}\in\mathcal{T}_p\cap\mathbb{S}^{(k)}} b_q^{\mathcal{T}_p}(\mathbf{r})\,\mathrm{d}^3\mathbf{r}}{\sum_{j=1}^{M}\left(\sum_{(\mathcal{T}_p,q)\to j} \int_{\mathbf{r}\in\mathcal{T}_p\cap\mathbb{S}^{(k)}} b_q^{\mathcal{T}_p}(\mathbf{r})\,\mathrm{d}^3\mathbf{r}\right)}\,, \tag{31}$$

where integrals are computed using Riemann sums.

To solve the SOLA minimization problem (21), one has to define some relevant value for the trade-off parameter η, such that a suitable

compromise can be found between minimizing both $\int [A^{(k)} - T^{(k)}]^2$ (i.e. good model resolution) and $\sigma^2_{\hat{m}_k}$ (i.e. small model variance). Since the value of η is free to differ for every node k, one should write $\eta^{(k)}$. In the presence of data noise, a potentially serious drawback of OLA methods may be the necessity to find at most M relevant values for the M trade-off parameters $\eta^{(k)}$ (e.g. Parker 1994). An option could be to estimate $\eta^{(k)}$ from the analysis of the monotonically decreasing curve of trade-off between $\int [A^{(k)} - T^{(k)}]^2$ and $\sigma^2_{\hat{m}_k}$ (e.g. Backus & Gilbert 1968; Pijpers & Thompson 1994). Ideally, this trade-off curve could be L-shaped, making almost straightforward to design an automatic criterion (e.g. Hansen & O'leary 1993) to find some relevant compromise between resolution and variance. After having computed and visualized several of those trade-off curves, we report that some of them may be far from being L-shaped (i.e. no sharp corner). Moreover, computing all these trade-off curves would computationally be very expensive, since for every node k one has to solve the system (22) as many times as the number of tested values for $\eta^{(k)}$. Last, but not least, though in the DLS approach one can see if some value for the damping parameter does lead to a globally coherent tomographic model, in OLA methods since only one node k is treated at a time one cannot see if a particular value for $\eta^{(k)}$ is globally coherent. Trampert (1998) summarizes the previous issues by recognizing that: '*being practically difficult to implement in the presence of data errors, resulted in the Backus–Gilbert method not finding many applications in seismic tomography*'.

An important finding of this study is to report that *globally coherent* tomographic images can be obtained from SOLA inversions (see Fig. 2), by simply using for all the nodes in the zth layer the same 'constant' value $C^{(z)}$ for the corresponding trade-off parameters $\eta^{(k)}$. In practice, our strategy has been to consider all the nodes in the zth layer, to test a few different constant values $C^{(z)}$, and to retain the one giving for us the most appealing tomographic solution at that depth. Finally, the same constant value C^{LM} has been retained for all the nodes in the lower-mantle (660–1710 km depth), and another constant value $C^{\mathrm{TZ}} = \frac{3}{4} \times C^{\mathrm{LM}}$ for those in the transition zone (400–660 km depth).

A more comprehensive study on how to choose appropriate target kernels and trade-off parameters will be the subject of future work. For instance, one could wonder what was the actual influence of having used such spheroid-shape target kernels, explicitly tuned to the irregular data coverage, on the very reduced number of different trade-off parameters finally used here? Though there is some 'subjectivity' in the SOLA method, note that different choices would result in different solutions, all of them corresponding to physical interpretations with respect to the *true* model parameters.

3.3 Tomographic results

We aim at presenting the obtained SOLA model and its appraisal; we shall focus on the 400–1710 km depth range, where the spatial coverage of our data is the most relevant.

3.3.1 Model

Fig. 2 presents our global-scale S-wave tomographic model, $\hat{m}_{\mathrm{sola}}$, which has been built using the alternative SOLA formulation of the B–G method. A detailed analysis or physical interpretation of this model is beyond the scope of this study; that will be the subject of future work. In the author's opinion, this first SOLA model is 'visually appealing'. To share this opinion with the reader, Fig. 3 shows a visual comparison between this SOLA model ($\hat{m}_{\mathrm{sola}}$) and an equivalent DLS model ($\tilde{m}_{\mathrm{DLS}}$, see Section 3.1); only the inversion approach differs. First, the two models are shown in Figs 3(a) and (b) within the transition zone, at 530–660 km depth. Though the two models look similar, significative differences can easily be spotted. As an example, the reader may focus on the complex system of subducted slabs (positive velocity anomalies) related to the subduction regions of New-Hebrides-Fiji, Tonga-Kermadec and Hikurangi. Similarly, the models are shown in Figs 3(c) and (d) at 1310–1510 km depth, focussing on the Farallon subduction system. The slab fragments F1 and F2, pointed out in Fig. 3(d), will be discussed in Section 3.3.3. In general, seismic features appear to be 'smoother or better focalized', and with 'higher amplitudes', in the SOLA model than in the DLS model. A quantitative comparison between these two models, $\hat{m}_{\mathrm{sola}}$ and $\tilde{m}_{\mathrm{DLS}}$, is beyond the scope of this study. For instance, one could also add some subjective smoothing constraints to the DLS solution. However, note that the inherent smoothness of the SOLA model naturally results from the averaging kernels themselves. In the same line, note that the overall smaller amplitudes in the DLS model, compared to the SOLA one, are probably due to the imposed constraint of model norm damping. Finally, it is our view that the SOLA inversion flowchart, as described in Sections 2.2 and 3.2, does lead to a 'globally coherent' tomographic model, the first one to be based on the B–G approach.

3.3.2 Appraisal

As mentioned in Section 2.2, in the B–G approach a reduction in the model uncertainty $\sigma_{\hat{m}_k}$ is obtained by spatially enlarging the resolving kernel $A^{(k)}$. Fig. 4 shows the model uncertainty $\sigma_{\hat{m}_k}$ obtained with the SOLA method, at four depths ranging from the transition zone to the mid lower-mantle. This model uncertainty is comprised within the interval 0.05–0.25 per cent in the whole 400–1710 km depth range. As expected, the spatial variations of the model uncertainty are almost similar to those of the ray-density driven target kernels, see Figs 1(a,c,e,g) and 4. As a remark, the uncertainties $\sigma_{\hat{m}_k}$ merely represent the propagation of data errors. They do not take into account the error due to the fact that, in practice, the averaging kernel $A^{(k)}$ cannot reproduce the form of the target kernel $T^{(k)}$ 'perfectly'. Following Backus & Gilbert (1970) and Pijpers & Thompson (1994), future work could aim at estimating an upper bound to the systematic error due to this effect.

We report that the unimodular condition is fully respected for the resolving kernels $A^{(k)}$ obtained with the SOLA method, since their volumetric integral, theoretically equal to one, is always comprised within the interval $1 \pm 2 \times 10^{-8}$. As a remark, this unimodular constraint cannot be imposed on the rows of the DLS resolution matrix, $\mathbf{R}_{\mathrm{DLS}}$, which may also complicate physical interpretations based on $\mathbf{R}_{\mathrm{DLS}}$ (e.g. Nolet 2008). Moreover, we report that the resolving kernels are not significantly oscillating (i.e. no prominent side lobes, see Fig. 5), which confirms that suitable forms for the target kernels have been used (e.g. Pijpers & Thompson 1994). Keeping in mind the trade-off between model variance and resolution, one may look at Fig. 4 and intuitively guess what are at first order the spatial variations of the model resolution. However, as warned by Parker (1994): '*we should always plot [the resolving kernels]*'. Indeed, as will be discussed in Sections 3.3.3 and 4.2, some of the resolving kernels may potentially be smeared or leaked in complex ways, requiring us to analyse them through 3-D visual plots to comprehensively apprehend the spatial model resolution.

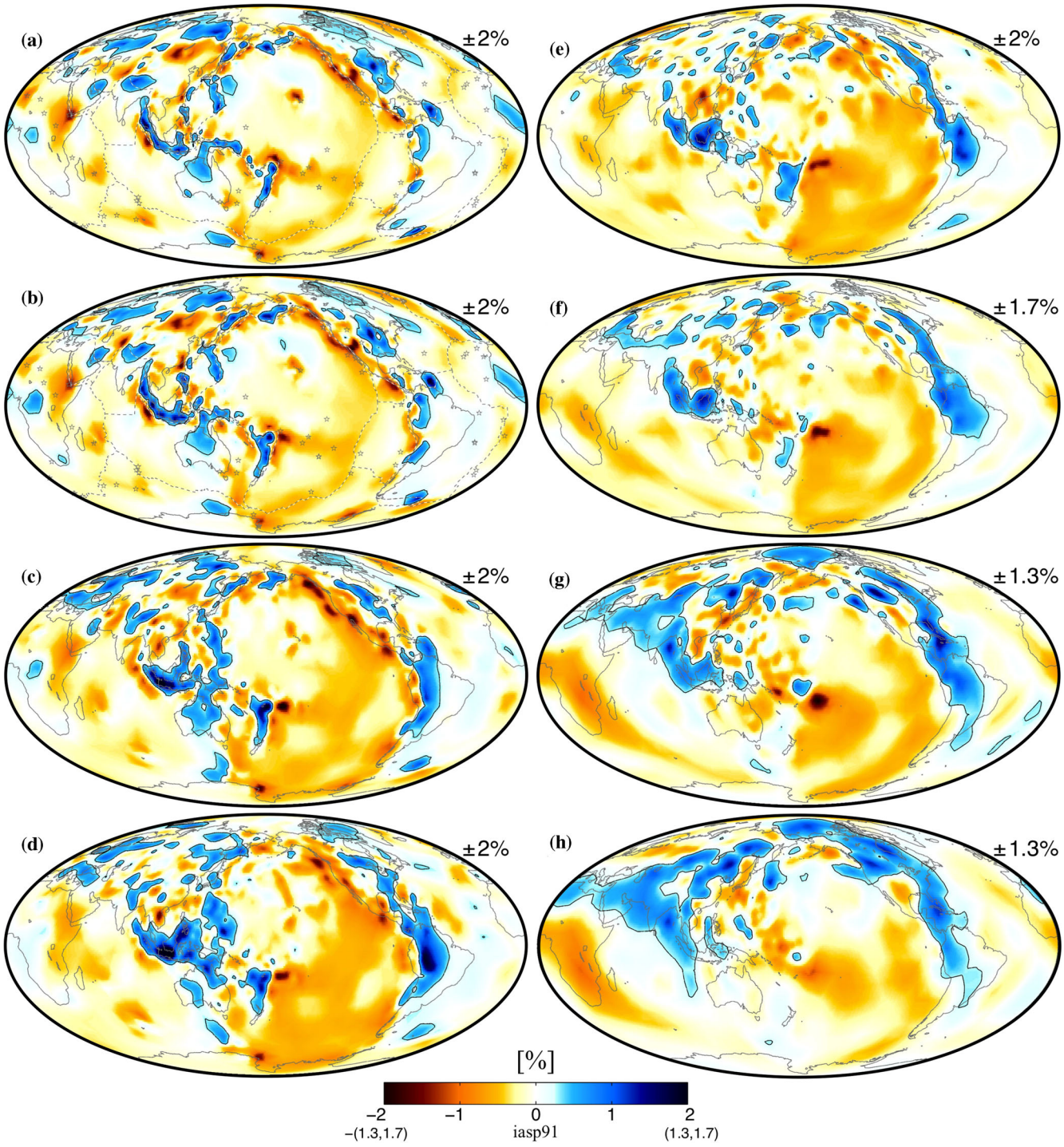

Figure 2. Global *S*-wave tomographic model ($\hat{m}_{\mathrm{sola}}$) obtained from using the alternative SOLA formulation of the Backus–Gilbert method. The depths are (in km): (a) 400–530; (b) 530–660; (c) 660–810; (d) 810–960; (e) 960–1110; (f) 1110–1310; (g) 1310–1510; (h) 1510–1710. The colourscale bounds are indicated in each panel. In panels (a) and (b), the dashed lines denote the tectonic plate boundaries and the grey stars represent some hotspots (e.g. Anderson & Schramm 2005). Positive velocity anomalies are emphasized with black contour lines if larger than some threshold: 0.3 per cent in panels (a)–(e), 0.25 per cent in panel (f) and 0.2 per cent in panels (g)–(h).

3.3.3 Examples

In the following, we aim at illustrating the importance of the model appraisal when it comes to physical interpretations of global-scale tomographic models. We focus on two seismic features present in our SOLA model, and discuss them in the light of their estimated resolution and uncertainty.

First, we aim at demonstrating whether or not is resolved a prominent slow-velocity feature, located South-West from Hawaii in the 810–960 km depth layer. A zoom-in of our SOLA model is shown in Fig. 5(left), on which the negative velocity anomaly is contoured with a black solid line. Inside this contour is depicted with a green dot a particular node k, whose model estimate and uncertainty

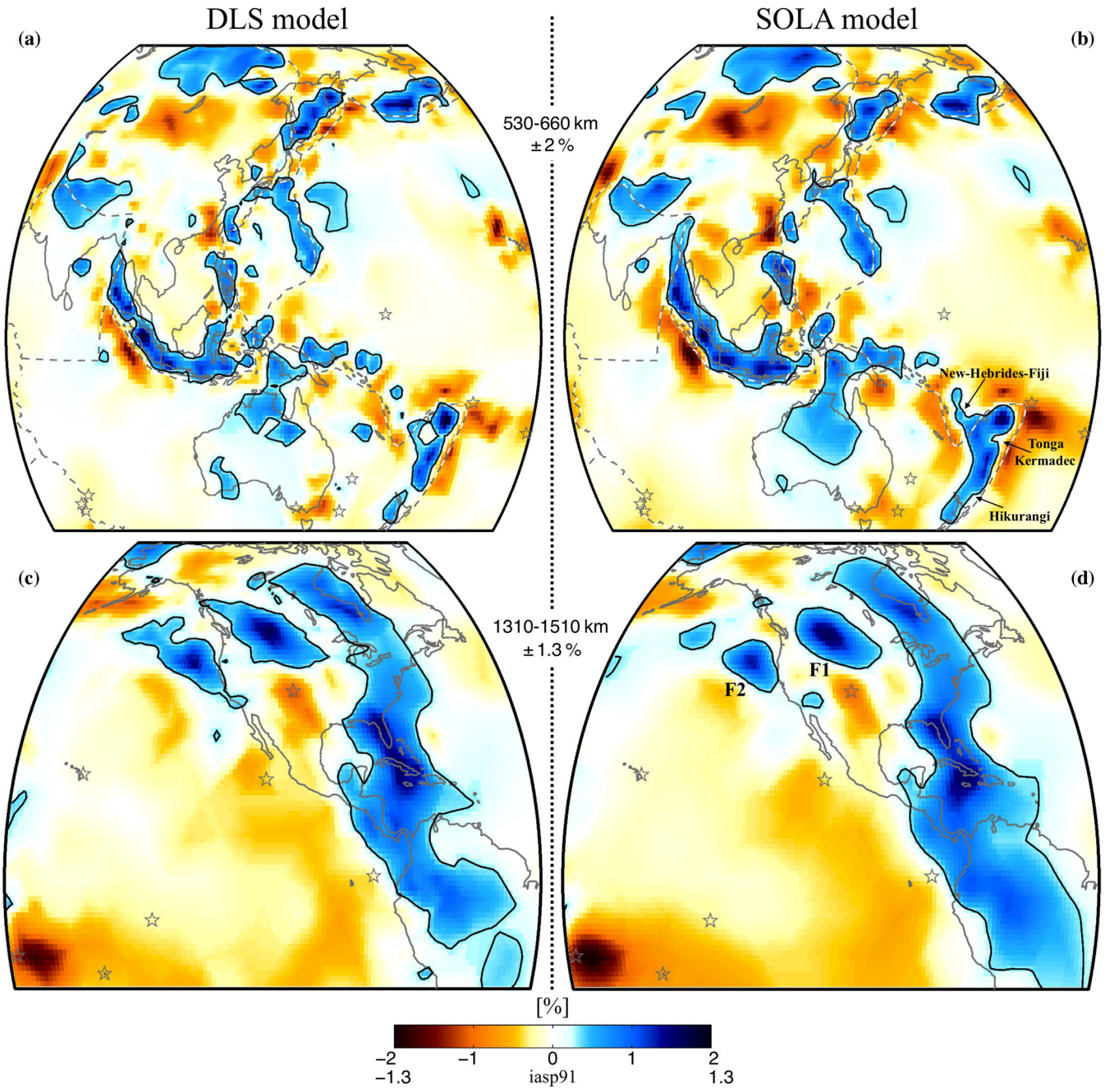

Figure 3. Visual comparison of tomographic models derived from: (a, c) Damped least-squares (DLS) inversion ($\tilde{m}_{\mathrm{DLS}}$); (b, d) SOLA inversion ($\hat{m}_{\mathrm{sola}}$). The comparison is focussed on: (a, b) several active subduction zones in the transition zone, at 530–660 km depth; (c, d) the ancient and complex Farallon subduction system taking place in the mid lower mantle, at 1310–1510 km depth. Black contour lines and grey stars/dashed-lines are the same as in Fig. 2.

are: $\hat{m}_k = -0.87$ per cent, and $\sigma_{\hat{m}_k} = 0.10$ per cent. The target kernel $T^{(k)}$ is shown in Fig. 5(left). Its form is that of a spheroid centred in $\mathbf{r}^{(k)}$, at 885 km depth, with horizontal and vertical radii: $\mathcal{L}_H^{(k)} = 300$ km, and $\mathcal{L}_V^{(k)} = 150$ km. The resolving kernel $A^{(k)}$ is also shown in Fig. 5(left) at different depths ranging from the near surface to the mid lower-mantle. Clearly, the shape of $A^{(k)}$ is complex and not confined into a small region near $\mathbf{r}^{(k)}$. That is, $A^{(k)}$ is obliquely spread out from the deep lower-mantle to the near surface, below the Hawaii island. Since $A^{(k)}$ defines a volume where different seismic heterogeneities cannot be distinguished, the contoured slow-velocity feature imaged at 885 km depth is not resolvable by the data. The most likely scenario to explain this contoured feature is that of a negative velocity anomaly located right below the Hawaii island, which would have been spread out down into the lower-mantle. The best candidate surely is the shallow part of the Hawaii hotspot, whose the seismic signature is visible on all the tomographic images within the 400–1110 km depth range, see Figs 2(a)–(e). Though an experienced geophysicist would probably argue, solely based on the tomographic images, that this contoured feature is an artefact due to the strongly anisotropic ray path geometry in the middle of the Pacific, the B–G approach has allowed us to more quantitatively point out this artefact. Note that resolving kernels with such a complex shape clearly require some 3-D visual inspection to fully apprehend them.

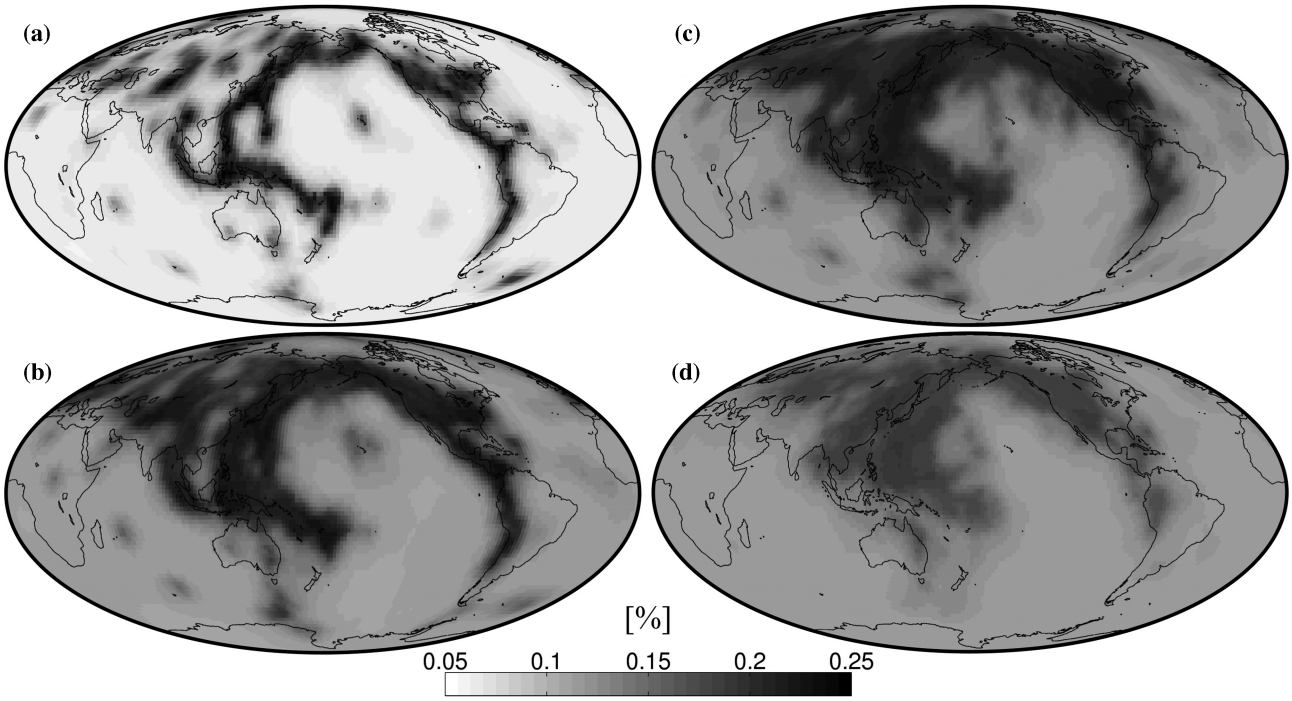

Figure 4. Examples of the model uncertainties $\sigma_{\hat{m}_k}$ obtained in this study using the alternative SOLA formulation of the Backus–Gilbert method. Results are shown for several layers from the transition zone to the mid lower mantle. The depth ranges are (in km): (a) 400–530; (b) 660–810; (c) 960–1110; (d) 1310–1510.

In this second example, we are interested in the Farallon subduction system (positive velocity anomalies). A zoom-in within the 1310–1510 km depth layer is shown in Fig. 3(d). In addition to the main subducted slab beneath eastern North-America, one may identify at least two detached slab fragments located further to the West, denoted F1 and F2 as in Zaroli *et al.* (2015). The Farallon system also takes place at shallower and deeper depths. In particular when looking at Figs 2(f)–(h), note that: (1) F1 could potentially be connected with some fast-velocity material upwards (at least at 1110–1310 km depth); (2) F2 seems to 'suddenly' appear at 1310–1510 km depth offshore California; (3) F1, F2, and the main slab all seem to collapse together at 1510–1710 km depth. Such a complex Farallon system in the mid lower-mantle has not been so sharply imaged in previous 'global' scale tomographic models (e.g. Masters *et al.* 1996; Van der Hilst *et al.* 1997; Montelli *et al.* 2004b; Houser *et al.* 2008; Ritsema *et al.* 2011; French & Romanowicz 2014; Moulik & Ekström 2014). However, the features F1 and F2 have recently been imaged, and even physically interpreted, in some 'regional' scale (multifrequency) *P*-wave tomographic studies (Sigloch *et al.* 2008; Sigloch 2011; Sigloch & Mihalynuk 2013). It has therefore been proposed that F1 could correspond to the deep root in lower mantle of the Cascadia subduction system, and F2 be related to some intra-oceanic subduction. In the aforementioned regional tomographic studies, only checker-board tests were used to assess the model resolution, and no quantification of the model uncertainty was performed. As the reader may have guessed, we are now interested in demonstrating whether for instance F2 is or not resolved, in our case. In Fig. 5(right), a zoom-in of our SOLA model is shown; F2 is contoured with a black solid line. Inside this contour is depicted with a green dot a particular node k; the model estimate and uncertainty are: $\hat{m}_k = 0.85$ per cent, and $\sigma_{\hat{m}_k} = 0.18$ per cent. The target kernel $T^{(k)}$ is shown in Fig. 5(right). Its form is that of a spheroid centred in $\mathbf{r}^{(k)}$, at 1410 km depth, with horizontal and vertical radii: $\mathcal{L}_H^{(k)} = 200$ km, and $\mathcal{L}_V^{(k)} = 200$ km. The resolving kernel $A^{(k)}$ is shown in Fig. 5(right) within the 1110–1710 km depth range. Note that $A^{(k)}$ is confined into a small enough volume near $\mathbf{r}^{(k)}$, with respect to the size of F2. Thus, the local resolving length, both laterally and radially, is small enough to claim that F2 is 'resolvable' by the data. Furthermore, since the uncertainty $\sigma_{\hat{m}_k}$ is almost five times smaller than the model estimate $\hat{m}_k$, one can conclude that F2 is 'resolved' given the data and their errors.

4 PERSPECTIVES

We aim at discussing some future work perspectives related to the use of the SOLA method in seismic tomography.

4.1 New tomographic applications

Future work would naturally consist in applying the SOLA method to other linear or linearized tomographic problems, such as, but not limited to: ambient noise tomography (e.g. Zigone *et al.* 2015), surface-wave tomography (e.g. Debayle & Ricard 2012), body-wave tomography (e.g. Sigloch 2011), normal-mode tomography (e.g. Resovsky & Ritzwoller 1999) and joint tomography (e.g. Ritsema *et al.* 2011; Zaroli *et al.* 2015; Koelemeijer *et al.* 2016). Significantly increasing the number of data and model parameters will require more processors and computer memory than the limited computational resources used in this work (see Appendix A3). However, with modern computational facilities (e.g. High Performance Computing Clusters, Graphics Processing Units, memory prices decreasing with time), the SOLA method should be applicable to linear and discrete tomographic problems of much larger size than faced in this study. As a final remark, though the B–G theory has been developed for linear problems, note that Snieder (1991) extended the theory to *weakly non-linear* problems.

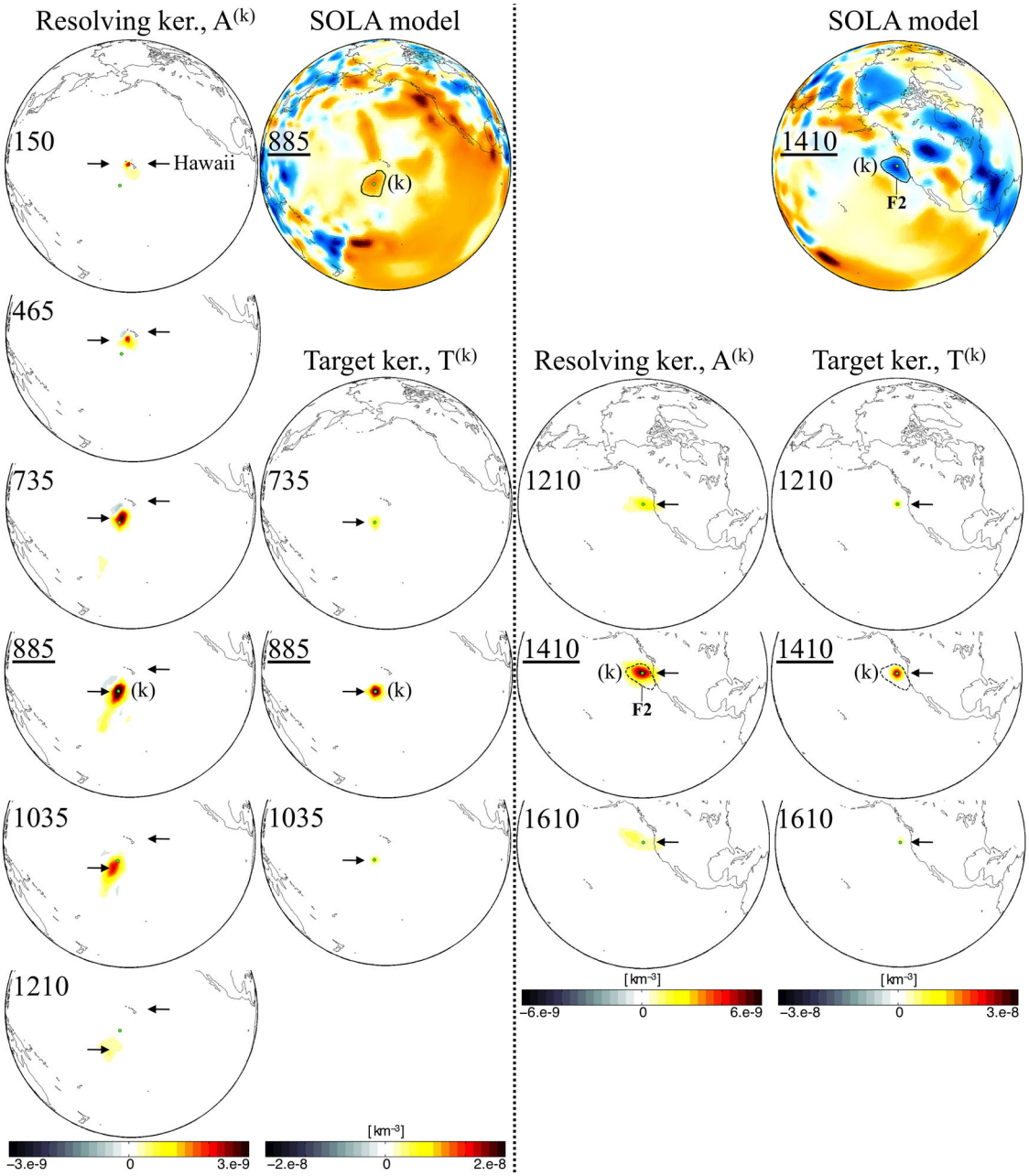

Figure 5. Examples of two resolving (averaging) kernels $A^{(k)}$ obtained in this study using the alternative SOLA formulation of the Backus–Gilbert method (see Section 3.3.3). Left: resolution analysis of the contoured slow-velocity feature at 885 km depth. Right: Resolution analysis of the contoured fast-velocity feature F2 at 1410 km depth. Are also shown the corresponding zoom-ins of the SOLA model $\hat{m}_{\mathrm{sola}}$, same colourscales as in Fig. 2 and the target kernels $T^{(k)}$. The spatial locations of the two nodes (k) are depicted with green dots. Each number represents the middle depth, in km, of one layer of the parametrization.

4.2 Visualizing the spatial model resolution

The shape of some resolving kernels $A^{(k)}$ may be complex and thus require 3-D visual inspections (see Section 3.3.3 and Fig. 5(left)). It may not be very convenient to show in a single paper all the M resolving kernels plotted in a 3-D fashion. From our experience, the shape of many resolving kernels approximately resembles that of an ellipsoid (see Fig. 5(right)). Let $E^{(k)}$ denotes such an ellipsoid, corresponding to node k. For simplicity reasons, let us assume that $E^{(k)}$ be centred in $\mathbf{r}^{(k)}$, though its centre of mass could be a free parameter. Thus, $E^{(k)}$ may simply be characterized by four parameters: its major and minor horizontal radius, $a^{(k)}$ and $b^{(k)}$, respectively; its vertical radius $c^{(k)}$; the azimuth $\psi^{(k)}$ of its major horizontal axis. Future work could consist in automatically estimating these four parameters for all the resolving kernels whose the shape is ellipsoidal; one could then visualize through global-scale 2-D plots the radial resolving length, $c^{(k)}$, and the direction-dependent lateral resolving length, given by ($a^{(k)}$, $b^{(k)}$, $\psi^{(k)}$). Some criterion would be needed to automatically discriminate non-ellipsoidal kernels; only those complex ones would remain to be 3-D visualized.

4.3 SOLA tomography using a parameter-free approach

To go beyond this ‘discrete’ study, one could aim at applying the SOLA method to tomographic problems in a ‘parameter-free’

manner, closer to the philosophy of Backus & Gilbert (1967), that is without introducing any model parametrization. Thus, the averaging kernel is expressed as $A^{(k)}(\mathbf{r}) = \sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r})$, and one has to solve the 'continuous' system of N equations ($1 \leq i' \leq N$):

$$\sum_{i=1}^{N} x_i^{(k)} \int K_i(\mathbf{r}) K_{i'}(\mathbf{r}) \mathrm{d}^3\mathbf{r} + \eta^2 x_{i'}^{(k)} = \int T^{(k)}(\mathbf{r}) K_{i'}(\mathbf{r}) \mathrm{d}^3\mathbf{r}\,, \qquad (32)$$

subject to the unimodular condition $\sum_{i=1}^{N} x_i^{(k)} \int K_i(\mathbf{r}) d^3\mathbf{r} = 1$. Pijpers & Thompson (1994) show that (32) corresponds to solving a linear system $\mathbf{H}\mathbf{x}^{(k)} = \mathbf{v}^{(k)}$, where the symmetric matrix $\mathbf{H}$ is of order $N+1$ and mainly contains a large number of integrals of pairs of sensitivity kernels, $\int K_i(\mathbf{r}) K_{i'}(\mathbf{r}) d^3\mathbf{r}$, which now have to be calculated. The vector $\mathbf{v}^{(k)}$ mainly contains the integrals $\int T^{(k)}(\mathbf{r}) K_{i'}(\mathbf{r}) \mathrm{d}^3\mathbf{r}$, which also need to be computed. In our view, one should keep the target kernel $T^{(k)}$ tuned to some proxy for the local resolving length (e.g. ray density). Though the computational cost to solve (32) could be challenging, when considering a large number of data and of points $\mathbf{r}^{(k)}$, the advantages would be that: (1) A given, subjective finite-dimensional parametrization would not have to be introduced; (2) The shape of the resolving kernel $A^{(k)}$ would be fully captured, thus improving the model estimation and the quantitative assessment of its resolution and uncertainty.

5 CONCLUSIONS

To push further the physical interpretations of tomographic models, the appraisal problem has now become as important as the model estimation. As we have seen, the B–G method can solve 'all at once' the linear problems of model estimation and appraisal. Furthermore, one key advantage of B–G is that its solution refers to the true model, while the DLS solution, often used in tomographic studies, refers to the 'damped' true model – hence making its physical interpretation with respect to the true model more complicated.

In this study, we have introduced and adapted to linear and discrete tomographic problems the SOLA method, an alternative B–G formulation which retains all its advantages, but is more computationally efficient and versatile in the explicit construction of averaging kernels. We have focused on the case of global-scale *S*-wave tomography, and have presented the first SOLA tomographic model of the Earth's mantle, globally coherent within the 400–1710 km depth range. In particular, when compared to an equivalent DLS model, the seismic features in our SOLA model generally appear to be 'smoother or better focalized', and with 'higher amplitudes'. To illustrate the appraisal of our SOLA model, we have shown that a small-scale slab fragment (F2), related to the complex Farallon subduction system, is resolved given the data and their errors.

Finally, since over the past few decades '*the Backus–Gilbert theory has received more citation than actual application*' (Parker 1994), the main result of this study is to have shown that the SOLA variant of the B–G method can successfully be applied to large-scale, linear and discrete tomographic problems.

ACKNOWLEDGEMENTS

The author is grateful to the editor, Lapo Boschi, and to David Al-Attar and an anonymous reviewer whose detailed comments have helped to improve this paper. The author wishes to sincerely thank Sophie Lambotte, for fruitful discussions, and Franco Zaroli, for invaluable help on some program optimization.

REFERENCES

Anderson, D.L. & Schramm, K.A., 2005. Global hotspot maps, in *Plates, Plumes and Paradigm,* GSA Special Paper, 388, pp. 19–29, eds Foulger, G.R., Natland, J.H., Presnall, D.C. & Anderson, D.L., Geological Society of America.

Aster, R.C., Borchers, B. & Thurber, C., 2012. *Parameter Estimation and Inverse Problems,* revised edn, Elsevier.

Backus, G. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross Earth data, *Phil. Trans. R. Soc. A.,* **266**(1173), doi:10.1098/rsta.1970.0005 .

Backus, G. & Gilbert, J.F., 1967. Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. astr. Soc.,* **13,** 247–276.

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

Boschi, L., 2003. Measures of resolution in global body wave tomography, *Geophys. Res. Lett.,* **30**(19), doi:10.1029/2003GL018222.

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

Davies, D.R., Goes, S., Davies, J., 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 planet. Sci. Lett.,* **353–354,** 253–269.

Debayle, E. & Ricard, Y., 2012. A global shear velocity model of the upper mantle from fundamental and higher rayleigh mode measurements, *J. geophys. Res.,* **117,** B10308, doi:10.1029/2012JB009288.

French, S.W. & Romanowicz, B., 2014. Whole-mantle radially anisotropic shear-velocity structure from spectral-element waveform tomography, *Geophys. J. Int.,* **199**(3), 1303–1327.

Grunberg, M., 2006. Conception d'une méthode de maillage 3D parallèle pour la construction d'un modèle de Terre réaliste par la tomographie sismique, *PhD thesis,* Strasbourg University, France.

Hansen, C. & O'leary, D., 1993. The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.,* **14**(6), 1487–1503.

Houser, C., Masters, G., Shearer, P.M. & Laske, G., 2008. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophys. J. Int.,* **174,** 195–212.

Jackiewicz, J., Birch, A.C., Gizon, L., Hanasoge, S.M., Hohage, T., Ruffio, J.-B. & Svanda, M., 2012. Multichannel three-dimensional SOLA inversion for local helioseismology, *Sol. Phys.,* **276,** 19–33.

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

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, *Geophys. J. Int.,* **204**(2), 1024–1039.

Larsen, R.M. & Hansen, P.C., 1997. Efficient implementations of the SOLA mollifier method, *Astron. Astrophys. Suppl. Ser.,* **121,** 587–598.

Lévêque, J.-J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic inversion, *Geophys. J. Int.,* **115,** 313–318.

Louis, A.K. & Maass, P., 1990. A mollifier method for linear operator equations of the first kind, *Inverse Probl.,* **6,** 427–490.

Masters, G., Johnson, S., Laske, G., Bolton, H. & Davies, J.H., 1996. A shear-velocity model of the mantle, *Phil. Trans. R. Soc. A,* **354,** 1385–1410.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory,* revised edn, Academic Press.

Mercerat, E.D., Nolet, G. & Zaroli, C., 2014. Cross-borehole tomography with correlation delay times, *Geophysics,* **79**(1), R1–R12.

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

Montelli, R., Nolet, G., Masters, G., Dahlen, F.A. & Hung, S.-H., 2004b. Global P and PP traveltime tomography: rays versus waves, *Geophys. J. Int.,* **158,** 636–654.

Moulik, P. & Ekström, G., 2014. An anisotropic shear velocity model of the Earth's mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophys. J. Int.,* **199,** 1713–1738.

Nolet, G., 1985. Solving or resolving inadequate and noisy tomographic systems, *J. Comput. Phys.,* **61,** 463–482.
Nolet, G., 2008. *A Breviary of Seismic Tomography,* Cambridge Univ. Press.
Nolet, G., Montelli, R. & Virieux, J., 1999. Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.,* **138,** 36–44.
Paige, C.C. & Saunders, M., 1982. LSQR: an algorithm for sparse, linear equations and sparse least squares, *ACM Trans. Math. Softw.,* **8,** 43–71.
Parker, R.L., 1994. *Geophysical Inverse Theory,* Princeton Univ. Press.
Pijpers, F.P., 1997. Inversions in astronomy and the SOLA method, in *Inverse Problems in Wave Propagation,* pp. 419–442, eds Chavent, G. *et al.*, Springer Science + Business Media.
Pijpers, F.P. & Thompson, M.J., 1992. Faster formulations of the Optimally Localized Averages method for helioseismic inversions, *Astron. Astrophys.,* **262,** L33–L36.
Pijpers, F.P. & Thompson, M.J., 1994. The SOLA method for helioseismic inversion, *Astron. Astrophys.,* **281,** 231–240.
Rabello-Soares, M.C., Basu, S. & Christensen-Dalsgaard, J., 1999. On the choice of parameters in solar-structure inversion, *Mon. Not. R. astr. Soc.,* **309,** 35–47.
Rawlinson, N. & Spakman, W., 2016. On the use of sensitivity tests in seismic tomography, *Geophys. J. Int.,* **205**(2), 1221–1243.
Rawlinson, N., Fichtner, A., Sambridge, M. & Young, M.K., 2014. Seismic tomography and the assessment of uncertainty, *Adv. Geophys.,* **55,** 1–76.
Resovsky, J. & Ritzwoller, M.H., 1999. A degree 8 mantle shear velocity model from normal mode observations below 3 mHz, *J. geophys. Res.,* **104**(B1), 993–1014.
Ritsema, J., McNamara, A.K. & Bull, A.L., 2007. Tomographic filtering of geodynamic models: implications for model interpretation and large-scale mantle structure, *J. geophys. Res.,* **112,** B01303, doi:10.1029/2006JB004566.
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, *Geophys. J. Int.,* **184,** 1223–1236.
Schuberth, B.S.A., Zaroli, C. & Nolet, G., 2012. Synthetic seismograms for a synthetic Earth: long-period *P*- and *S*-wave traveltime variations can be explained by temperature alone, *Geophys. J. Int.*, **188**(3), 1393–1412.
Schuberth, B.S.A., Zaroli, C. & Nolet, G., 2015. Traveltime dispersion in an isotropic elastic mantle: strong lower mantle signal in differential-frequency residuals, *Geophys. J. Int.,* **203**(3), 2099–2118.
Sigloch, K., 2011. Mantle provinces under North America from multi-frequency P wave tomography, *Geochem. Geophys. Geosyst.,* **12,** Q02W08, doi:10.1029/2010GC003421.
Sigloch, K. & Mihalynuk, M., 2013. Intra-oceanic subduction shaped the assembly of cordilleran North America, *Nature,* **496**(7443), 50–56.
Sigloch, K., McQuarrie, M. & Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nature Geosci.,* **1,** 458–462.
Simmons, N.A., Forte, A.M. & Grand, S.P., 2009. Joint seismic, geodynamic and mineral physical constraints on three-dimensional mantle heterogeneity: implications for the relative importance of thermal versus compositional heterogeneity, *Geophys. J. Int.,* **177**(3), 1284–1304.
Snieder, R., 1991. An extension of Backus-Gilbert theory to nonlinear inverse problems, *Inverse Probl.,* **7,** 409–433.
Snieder, R. & Trampert, J., 1999. Inverse problems in geophysics, in *Wavefield Inversion,* pp. 119–190, ed. Wirgin, A., Springer Verlag.
Soldati, G. & Boschi, L., 2005. The resolution of whole Earth seismic tomographic models, *Geophys. J. Int.,* **161**(1), 143–153.
Tian, Y., Zhou, Y., Sigloch, K., Nolet, G. & Laske, G., 2011. Structure of North American mantle constrained by simultaneous inversion of multiple-frequency SH, SS, and Love waves, *J. geophys. Res.,* **116,** B02307, doi:10.1029/2010JB007704.
Trampert, J., 1998. Global seismic tomography: the inverse problem and beyond, *Inverse Probl.,* **14,** 371–385.
Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160,** 195–216.
Van der Hilst, R.D., Widiyantoro, S. & Engdahl, E.R., 1997. Evidence for deep mantle circulation from global tomography, *Nature,* **386,** 578–584.
Vasco, D.W., Johnson, L.R. & Marques, O., 2003. Resolution, uncertainty, and whole Earth tomography, *J. geophys. Res.,* **108,** doi:10.1029/2001JB000412.
Zaroli, C., Debayle, E. & Sambridge, M., 2010. Frequency-dependent effects on global S-wave traveltimes: wavefront-healing, scattering and attenuation, *Geophys. J. Int.,* **182,** 1025–1042.
Zaroli, C., Sambridge, M., Lévêque, J.-J., Debayle, E. & Nolet, G., 2013. An objective rationale for the choice of regularisation parameter with application to global multiple-frequency S-wave tomography, *Solid Earth,* **4,** 357–371.
Zaroli, C., Lambotte, S. & Lévêque, J.-J., 2015. Joint inversion of normal-mode and finite-frequency S-wave data using an irregular tomographic grid, *Geophys. J. Int.,* **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 appl. Geophys.,* **172,** 1007–1032.

APPENDIX A: NUMERICAL CONSIDERATIONS

A1 SOLA using LSQR

Though systems such as (22) are usually solved using Lagrange multipliers, in this study we rather use the alternative LSQR approach proposed by Nolet (1985), as detailed in the following.

First, let the column-vector $\mathbf{c}$ be:

$$\mathbf{c} = (c_i)_{1\le i\le N}\,, \qquad c_i = \sum_{j=1}^{M} G_{ij}\,. \tag{A1}$$

Thus, the unimodular constraint, that is, eq. (15), can be rewritten as:

$$\mathbf{c}^{\mathrm{T}}\mathbf{x}^{(k)} = 1\,. \tag{A2}$$

One has to solve for $\mathbf{x}^{(k)}$ the following SOLA system of equations:

$$\begin{pmatrix} \hat{\mathbf{G}}^{\mathrm{T}} \\ \eta\mathbf{I}_N \end{pmatrix}\mathbf{x}^{(k)} = \begin{pmatrix} \mathbf{t}^{(k)} \\ \mathbf{0}_N \end{pmatrix}, \qquad [\text{s.t.} \quad \mathbf{c}^{\mathrm{T}}\mathbf{x}^{(k)} = 1]\,. \tag{A3}$$

Using eq. (A2), one element of the vector $\mathbf{x}^{(k)}$—for example, its first element $x_1^{(k)}$—can be expressed in terms of the others. Let us consider the three column-vectors:

$$\hat{\mathbf{x}}^{(k)} = \left(x_i^{(k)}\right)_{2\le i\le N}\,, \quad \hat{\mathbf{c}} = \left(\frac{c_i}{c_1}\right)_{2\le i\le N}\,, \quad \mathbf{e}_1 = (\delta_{i1})_{1\le i\le N}\,, \tag{A4}$$

assuming $c_1 \neq 0$ and δ being the Kronecker symbol, such that

$$x_1^{(k)} = c_1^{-1} - \hat{\mathbf{c}}^{\mathrm{T}}\hat{\mathbf{x}}^{(k)}\,. \tag{A5}$$

Let us also consider the matrix $\mathbf{B}$ defined as:

$$\mathbf{B} = \begin{pmatrix} -\hat{\mathbf{c}}^{\mathrm{T}} \\ \mathbf{I}_{N-1} \end{pmatrix}, \tag{A6}$$

so that the vector $\mathbf{x}^{(k)}$ can be written in function of $\hat{\mathbf{x}}^{(k)}$, that is,

$$\mathbf{x}^{(k)} = \mathbf{B}\hat{\mathbf{x}}^{(k)} + c_1^{-1}\mathbf{e}_1\,. \tag{A7}$$

Finally, the SOLA system (A3) can be rewritten into a system of normal equations to be solved for $\hat{\mathbf{x}}^{(k)}$ using the LSQR

algorithm:

$$\begin{pmatrix} \mathbf{Q}^{(\eta)} \\ \eta \mathbf{I}_{N-1} \end{pmatrix} \hat{\mathbf{x}}^{(k)} = \begin{pmatrix} \mathbf{y}^{(k,\eta)} \\ \mathbf{0}_{N-1} \end{pmatrix}, \tag{A8}$$

where $\mathbf{Q}^{(\eta)}$ is a matrix, of size $(M+1) \times (N-1)$, and $\mathbf{y}^{(k,\eta)}$ is a column-vector, which are defined as

$$\mathbf{Q}^{(\eta)} = \begin{pmatrix} \hat{\mathbf{G}}^{\mathrm{T}} \mathbf{B} \\ -\eta \hat{\mathbf{c}}^{\mathrm{T}} \end{pmatrix}, \qquad \mathbf{y}^{(k,\eta)} = \begin{pmatrix} \mathbf{t}^{(k)} - c_1^{-1} \hat{\mathbf{G}}^{\mathrm{T}} \mathbf{e}_1 \\ -c_1^{-1} \eta \end{pmatrix}. \tag{A9}$$

Thus, for a given trade-off parameter η, the LSQR algorithm iteratively converges to the solution $\hat{\mathbf{x}}^{(k,\eta)}$, such that

$$\hat{\mathbf{x}}^{(k,\eta)} = \underset{\hat{\mathbf{x}}^{(k)} \in \mathbb{R}^{N-1}}{\arg\min} : ||\mathbf{y}^{(k,\eta)} - \mathbf{Q}^{(\eta)} \hat{\mathbf{x}}^{(k)}||^2 + \eta^2 ||\hat{\mathbf{x}}^{(k)}||^2 , \tag{A10}$$

where $|| \cdot ||$ denotes the L^2-norm. Computationally speaking, the most important point is that the matrix $\mathbf{Q}^{(\eta)}$ does *not* depend on the spatial location $\mathbf{r}^{(k)}$; thus, no need to recompute it for every node k. Since the vector $\mathbf{y}^{(k,\eta)}$ depends on $\mathbf{r}^{(k)}$, it has to be recomputed for each node k; this task is computationally cheap. The last row of $\mathbf{Q}^{(\eta)}$ and last element of $\mathbf{y}^{(k,\eta)}$ both depend on η.

A2 Matrix sparsity

In global body-wave tomography, the matrix $\hat{\mathbf{G}}$ is highly sparse. Matrix sparsity may be a very useful property in terms of storage, efficiency of algorithms, memory footprint, etc. Though only 2.2 per cent of all the elements of $\hat{\mathbf{G}}$ are non-zero, the density varies from 0.2 per cent and up to 6.4 per cent for individual rows of $\hat{\mathbf{G}}$. The system (A8) involves computing the matrix $\mathbf{Q}^{(\eta)}$; except for its last row, its elements are

$$Q^{(\eta)}_{\mu,\nu} = -\hat{c}_\nu \hat{G}_{1,\mu} + \hat{G}_{\nu+1,\mu} , \tag{A11}$$

with $1 \leq \mu \leq M$, and $1 \leq \nu \leq N-1$. The M elements $\hat{G}_{1,\mu}$ represent the first row of $\hat{\mathbf{G}}$. All the elements of $\hat{\mathbf{c}}$ are non-zero. Thus, the total density of $\mathbf{Q}^{(\eta)}$ may vary from 2.4 per cent to 8.6 per cent depending upon the density of the first row of $\hat{\mathbf{G}}$. Preserving at most the sparsity of $\mathbf{Q}^{(\eta)}$ is crucial, especially as we plan to increase the number of data and parameters in future inversions (see Section 4.1). To optimize the sparsity of $\mathbf{Q}^{(\eta)}$, we find the sparsest row of $\hat{\mathbf{G}}$ and switch it with the original first row of $\hat{\mathbf{G}}$, such that the total density of $\mathbf{Q}^{(\eta)}$ is 2.4 per cent. To be consistent, the analogous operation is performed on the data vector $\mathbf{d}$ (i.e. switch of d_1).

A3 Computational facilities

Our computational facilities consist in $P = 20$ processors, Intel Xeon E5-4657L 2.40 GHz. As mentioned in Section 3.2, our strategy is to consider all the $M^{(z)}$ nodes belonging to the zth layer, and to use the 'same' constant value $C^{(z)}$ for the corresponding trade-off parameters. Since the LSQR algorithm can independently be applied to every node k, we partition the $M^{(z)}$ nodes into P subsets, each one containing $M^{(z)}/P$ nodes. For algorithms needs, and to reduce the disk input-output utilization per process, the matrix $\mathbf{Q}^{(\eta)}$ is stored twice on every processor, corresponding to a total of $6.4 \times P = 128$ Go memory. LSQR is used in a sequential fashion for each subset of nodes. The P subsets are treated in parallel on all the P processors. In our case, it takes about $t \simeq 3$ min (CPU time) to run 100 LSQR iterations for a single node k. We report that 100 iterations were sufficient for convergence needs; further iterating did not significantly change the model estimation, resolution and variance. Thus, computing the M solution vectors, $\hat{\mathbf{x}}^{(k,\eta)}$, takes about: $t \times M/P \simeq 4$ d (CPU time).