Geophysical Journal International

Geophys. J. Int. (2019) **218,** 619–630
Advance Access publication 2019 April 13
GJI Seismology

doi: 10.1093/gji/ggz175

Seismic tomography using parameter-free Backus–Gilbert inversion

Christophe Zaroli

Institut de Physique du Globe de Strasbourg, UMR 7516, Université de Strasbourg, EOST/CNRS, F-67084 Strasbourg cedex, France.
E-mail: *c.zaroli@unistra.fr*

Accepted 2019 April 11. Received 2019 March 8; in original form 2018 December 3

SUMMARY

This proof-of-concept study presents a parameter-free, linear Backus–Gilbert inversion scheme, tractable for seismic tomography problems. It leads to efficient computations of unbiased tomographic images, accompanied by meaningful resolution and uncertainty informations. Moreover, as there is no need to parametrize the model space in this parameter-free approach, it enables numerically accurate data sensitivity kernels to be effectively exploited in tomographic inversions. This is a major benefit over discrete tomographic methods, for which data sensitivity kernels are often inaccurate, as they are projected on a given model parametrization prior to be exploited in the inversion, and these parametrizations are usually coarse to limit the number of parameters and keep tractable the problems of model estimation and/or appraisal. Therefore, this new tomographic scheme fuels great hopes on better constraining multiscale seismic heterogeneities in the Earth's interior by exploiting accurate data sensitivity kernels, that is, taking full advantage of known wave-propagation physics, and enabling quantitative appraisals of tomographic features. As a remark, since its computational cost grows as a function of the total number of data squared, it may be better suited to handle moderate-size data sets, typically encountered in regional-scale tomography. Theoretical developments are illustrated within a finite-frequency physical framework. A set of 27 070 teleseismic *S*-wave time residuals is inverted, with focus on imaging and appraising shear-wave velocity anomalies lying in the mantle below Southeast Asia, in the 350–1410 km depth range.

Key words: Inverse theory; Tomography; Body waves.

1 INTRODUCTION

Tomographic images inferred from seismic data can be exploited to provide constraints within which to frame and answer fundamental questions on the Earth's present-day internal structure, composition and dynamics (e.g. Kennett & Bunge 2008; Nolet 2008; Romanowicz 2008). Recently, in an effort to build higher resolution tomographic models, finite-frequency effects (e.g. wave-front-healing) present in seismic data have started to be accounted for through the use of finite-frequency data sensitivity kernels (e.g. Dahlen *et al.* 2000; Hung *et al.* 2001; Montelli *et al.* 2004a; Tromp *et al.* 2005; Mercerat *et al.* 2014; Zaroli *et al.* 2015). Interest for finite-frequency tomography has been fuelled by continued evidence for structure-related dispersion exhibited in local to global scale sets of *P*- and *S*-wave cross-correlation time residuals (e.g. Hung *et al.* 2004; Yang *et al.* 2006; Sigloch & Nolet 2006; Zaroli *et al.* 2010; Hosseini & Sigloch 2015; Schuberth *et al.* 2015). However, despite theoretical improvement upon the infinite-frequency approximation of ray theory (e.g. Nolet *et al.* 2005), several studies have questioned on the actual ability to better constrain small-scale seismic heterogeneities in finite-frequency tomographic models (e.g. Van der Hilst & de Hoop 2005; Dahlen & Nolet 2005; Boschi *et al.* 2006; Chevrot *et al.* 2012; Maceira *et al.* 2015; Maguire *et al.* 2018). Indeed, benefits from using a finite-frequency wave-propagation physical approach may be hampered by several factors, such as the data quality and spatial coverage, and, as further discussed, the inversion scheme.

Most of linear or linearized tomographic inversions to date, within ray-theory or finite-frequency frameworks, have been carried out through two main technical steps (e.g. Aster *et al.* 2012): (1) Parametrizing the model space with a finite number of parameters; (2) Applying damped least-squares (DLS) methods to estimate these parameters (model estimation), and, though often ignored because of prohibitive computational costs (e.g. Rawlinson *et al.* 2010, 2014), their resolution and uncertainty (model appraisal). The first step implies that data sensitivity kernels have to be projected on a given parametrization of the model space prior to be effectively exploited in the inversion (e.g. Nolet 2008). Consequently, in order to fully exploit them for modelling finite-frequency effects in the data, it is of crucial importance that finite-frequency data sensitivity kernels remain numerically accurate after such projection. Although there are various ways for parametrizing models (e.g. Sambridge *et al.* 1995; Montelli *et al.* 2004b; Ritsema *et al.* 2011; Chevrot *et al.* 2012; Zaroli *et al.* 2015; Maguire *et al.* 2018), the number of parameters often has to be limited to keep computationally tractable the problems of model estimation and/or appraisal. Therefore, finite-frequency sensitivity kernels typically are projected on coarse parametrizations, and thus look like 'fat' ray-theoretical sensitivity kernels (i.e. no sensitivity variation all around the ray path). Those projected sensitivity kernels are then no more suitable for finite-frequency imaging purposes—a major '*reason why finite-frequency theory gave so far results similar to ray theory*' (Chevrot *et al.* 2012). Another concern, which links to the second step, is that amplitudes of DLS models may represent locally biased averages over the true-model parameters in regions of poor data illumination (Zaroli *et al.* 2017). This averaging bias effect is related to adding *ad hoc* regularization constraints (different from *a priori* physical constraints) on the model, such as L_2-norm damping, to remove the non-uniqueness inherent to the least-squares solution (e.g. Menke 1989; Nolet 2008; Voronin & Zaroli 2018). Since uneven data coverage prevails in local to global scale tomographic experiments, most DLS models are prone to be locally biased—what may lead to model misinterpretations.

In this study, we aim to present a new tomographic scheme that overcomes all the drawbacks related to the previous two technical steps. First of all, a fundamental insight from the pioneering works by Backus & Gilbert (1967, 1968, 1970) is that tomographic problems are invariably, at least partly, underdetermined: '*the collection of Earth models which yield the physically observed values of any independent set of gross Earth data is either empty or infinite dimensional*' (Backus & Gilbert 1967). Recognizing this fact, the linear Backus–Gilbert (B–G) inversion scheme, which belongs to the class of Optimally Localized Averages (OLA) methods, seeks not to construct a particular model solution, that is, to estimate model parameters, but instead to determine spatially localized, unbiased averages over the continuous true-model properties. Thus, the B–G approach seems relevant to move toward parameter-free and unbiased tomography, while enabling to solve all at once the problems of model estimation and appraisal. However, many authors subsequently found it to be too computationally intensive, as well as impractical in the presence of data errors (e.g. Menke 1989; Parker 1994; Trampert 1998; Aster *et al.* 2012). Recently, following the discrete B–G framework suggested by Nolet (1985), for which a local parametrization of the model space is assumed, Zaroli (2016) uncovered an efficient way of adapting a variant of B–G, namely the SOLA method (Subtractive OLA, proposed by Pijpers & Thompson (1992) for helio-seismic inversions), to large-scale, linear and discrete seismic tomography problems, even in the presence of data errors. The reader is referred to Zaroli *et al.* (2017) for a formal comparison of the discrete SOLA and DLS inversion schemes in terms of model estimation and appraisal, as well as for a quantitative illustration of averaging bias effects that may occur in DLS models— both based on synthetic tomographic experiments.

The goal of this study is then to extend the method of discrete SOLA tomography to the parameter-free case, so that the model space can retain its infinite dimensional nature and a specific model parametrization never be introduced—and accurate data sensitivity kernels be effectively exploited in the inversion. Section 2 presents theoretical developments on this new method, named *parameter-free SOLA tomography*, within a finite-frequency physical framework (Dahlen *et al.* 2000). For illustrative purposes, the parameter-free SOLA approach is applied to a finite-frequency inversion of 27 070 teleseismic *S*-wave time residuals, with focus on imaging and appraising shear-wave velocity anomalies lying in the mantle below Southeast Asia, in the 350–1410 km depth range. In Section 3, parameter-free SOLA tomography is discussed against discrete SOLA and DLS, and several perspectives and future applications are highlighted. Computational aspects involved in parameter-free SOLA tomography are discussed in Appendix A, including how to reduce the costs and memory requirements.

Finally, this proof-of-concept paper aims to open a new way for solving linear tomographic problems, that: (1) allows accurate data sensitivity kernels to be effectively exploited in tomographic inversions; (2) leads to efficient, embarrassingly parallel, computations of unbiased tomographic images accompanied by meaningful resolution and uncertainty informations, enabling quantitative appraisals of tomographic features; (3) is tractable even with limited computational resources, provided moderate-size data sets—as frequently encountered in regional-scale experiments.

2 PARAMETER-FREE SOLA TOMOGRAPHY

2.1 Preamble

We are interested in linear tomographic problems of the form

$$d_i = \int K_i\,(\mathbf{r})\,m\,(\mathbf{r})\,\mathrm{d}^3\mathbf{r} + n_i\,, \quad 1 \le i \le N\,, \tag{1}$$

where d_i is the ith datum, K_i the sensitivity kernel, n_i the noise, and m the 'true' model. As a leitmotiv, we consider the case of finite-frequency *S*-wave time-residual tomography, aimed at imaging 3-D shear-wave velocity anomalies in the Earth's mantle. Thus, $m(\mathbf{r})$ denotes the shear-wave velocity perturbation in $\mathbf{r}$ with respect to a radial velocity model, d_i represents an *S*-wave time residual measured by cross-correlating a pair of observed and synthetic waveforms filtered around a given central period, and $K_i(\mathbf{r})$ is a volumetric, *S*-wave time-residual sensitivity kernel which depends on the filtering period (Marquering *et al.* 1998; Dahlen *et al.* 2000). In this study, we assume that the noise $(n_i)_{1\le i\le N}$ has zero mean and the data covariance matrix is diagonal: $\mathbf{C}_d = \mathrm{diag}(\sigma_{d_i}^2)_{1\le i\le N}$. From hereon, both the data d_i and sensitivity kernels K_i are scaled by the data errors σ_{d_i}.

In our view, discretizing a finite-frequency sensitivity kernel K_i on a grid made of 50 km edge-length cubic cells is sufficient to fully capture its form (i.e. all its characterizing sensitivity variations), as illustrated in Fig. 1(a) for an *SS* phase. Discretizing sensitivity kernels is necessary to perform numerical calculations, but is fundamentally different from discretizing the model space. That is, discretizing the model space leads sensitivity kernels to be projected onto the model parametrization prior to be exploited in the inversion. Fig. 1(b) shows the projection of the same *SS* sensitivity kernel on a tomographic grid made of 200 km edge-length cubic cells. One sees that the projected kernel looks like a 'fat' ray-theoretical sensitivity kernel, that is, there is almost no sensitivity variation all around the ray path. Such a projected kernel is no more suitable for finite-frequency imaging purposes.

Most tomographic experiments rely on worldwide distributed earthquakes and/or receivers, for example, see Fig. 2(a), so that sensitivity kernels may sample any part of the mantle, and, therefore, the *entire* mantle needs to be parametrized. To fully capture the form of every finite-frequency sensitivity kernels, one could parametrize the whole mantle using 50 km edge-length cubic cells, which would lead to 7.2 million parameters and then make intractable the problems of model estimation and/or appraisal.

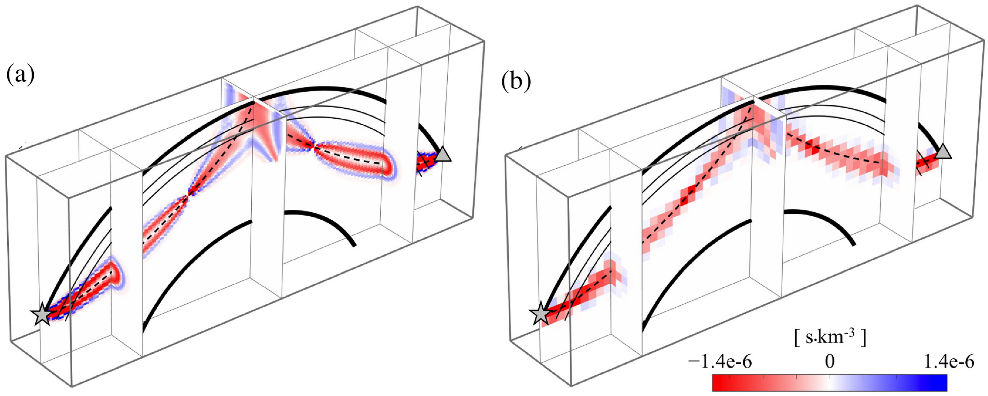

Figure 1. (a) 3-D, finite-frequency, *SS*-phase time-residual sensitivity kernel (120° epicentral distance, 22 s central period of a passband Gaussian filter). It is discretized on a local Cartesian grid made of regularly spaced 50 km edge-length cubic cells, spanning a rectangular parallelepiped region (thick grey solid line). The Earth's surface and core–mantle boundary (transition zone) are depicted with thick (thin) black solid lines, respectively, and the earthquake and receiver with a star and triangle. (b) Same *SS* sensitivity kernel but after projection on a tomographic grid that consists in regularly spaced 200 km edge-length cubic cells; note that the projected sensitivity kernel looks like a 'fat' ray-theoretical kernel—that is, almost no sensitivity variation all around the geometrical ray path (black dashed line).

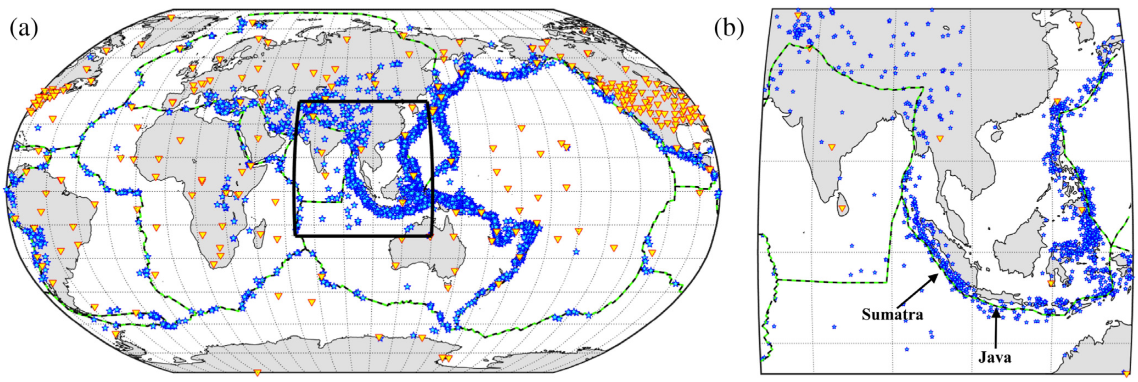

Figure 2. (a) Globally distributed earthquakes (∼4000 stars) and receivers (∼250 triangles) corresponding to the set of teleseismic *S*-wave data used in this study. Tectonic plates are drawn in black–green dashed lines. (b) Zoom-in on the 'Southeast Asia' region (black frame).

In the following, we aim at showing from theory to practice that parameter-free SOLA tomography: (1) provides an efficient way to circumvent the need for parametrizing the model space, enabling numerically accurate sensitivity kernels to be exploited in tomographic inversions; (2) leads to efficient, all-at-once, computations of unbiased tomographic images, accompanied by resolution and uncertainty informations.

2.2 Theory

In the B–G approach, one explicitly seeks an estimate, $\hat{m}^{(k)}$, that represents a *weighted average* over the continuous true-model properties, $m(\mathbf{r})$. This averaging process takes place through an averaging kernel, $A^{(k)}(\mathbf{r})$, that we wish to be spatially localized around a given query point, $\mathbf{r}^{(k)}$. This leads to writing:

$$\hat{m}^{(k)} = \int A^{(k)}(\mathbf{r})\, m(\mathbf{r})\, \mathrm{d}^3\mathbf{r} \quad (+\,\text{propagated noise})\,. \tag{2}$$

We wish that the integral $\int A^{(k)} m$ yields *unbiased* averages over the true model m. The averaging kernel $A^{(k)}$ should then satisfy to the following 'unimodular condition':

$$\int A^{(k)}(\mathbf{r})\, \mathrm{d}^3\mathbf{r} = 1\,; \tag{3}$$

and also be non-negative. As a remark, the model estimate $\hat{m}^{(k)}$ is said to be *biased* (Nolet 2008) if the averaging kernel $A^{(k)}$ does not meet (3). Zaroli *et al.* (2017) show (though in a discrete framework) that if $\int A^{(k)}$ is larger (smaller) than one, then $\hat{m}^{(k)}$ may be biased toward higher (lower) amplitude values, respectively, and thus not represent anymore a true averaging over the true model. They demonstrate that this *averaging bias* effect may occur in DLS models, especially in regions with poor data coverage. Since B–G (or SOLA) estimates are explicitly constrained to meet (3), they are expected to be unbiased. Averaging kernels, also referred to as resolving kernels, inform us on the local resolving length in tomographic images, that is, the minimum size of velocity anomalies that could be locally detected. For example, if a resolving kernel $A^{(k)}$ was constant-valued inside a 3–D ball centred on a query point $\mathbf{r}^{(k)}$ and zero elsewhere, then the ball's radius would correspond to the resolving length that one could, at best, expect to reach in $\mathbf{r}^{(k)}$. Since the forward problem (1) is linear, one can seek the estimate $\hat{m}^{(k)}$ as a linear combination of the data:

$$\hat{m}^{(k)} = \sum_{i=1}^{N} x_i^{(k)} d_i\,, \tag{4}$$

where the N unknown, real-valued coefficients

$$\mathbf{x}^{(k)} = \left(x_i^{(k)}\right)_{1\le i\le N} \tag{5}$$

represent a generalized inverse operator that maps the data to the estimate. From eqs (1), (2) and (4), one can write the estimate as:

$$\hat{m}^{(k)} = \int \underbrace{\left(\sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r}) \right)}_{A^{(k)}} m(\mathbf{r})\, \mathrm{d}^3\mathbf{r} + \underbrace{\sum_{i=1}^{N} x_i^{(k)} n_i}_{\substack{\text{propagated}\\ \text{noise}}} , \tag{6}$$

and the resolving kernel $A^{(k)}$ can be formally expressed as a linear combination of the sensitivity kernels:

$$A^{(k)}(\mathbf{r}) = \sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r}) . \tag{7}$$

The term $\sum_i x_i^{(k)} n_i$ in (6) represents the amount of data noise that propagates into $\hat{m}^{(k)}$. The variance in the model estimate $\hat{m}^{(k)}$ can be expressed as:

$$\sigma^2_{\hat{m}^{(k)}} = \sum_{i=1}^{N} \left(x_i^{(k)} \right)^2 \left(\sigma_{d_i} \right)^2 = \sum_{i=1}^{N} \left(x_i^{(k)} \right)^2 , \tag{8}$$

since the data were scaled by their errors. The uncertainty $\sigma_{\hat{m}^{(k)}}$ informs us on the level of propagated noise in the ‘weighted average’ estimate $\hat{m}^{(k)}$. As a remark, $\sigma_{\hat{m}^{(k)}}$ cannot inform us on how much $\hat{m}^{(k)}$ may differ from the true-model value $m(\mathbf{r}^{(k)})$—at least when the spatial variations of m are non-smooth and/or the spatial extent of $A^{(k)}$ is far from a Dirac delta function. Both the resolving kernels and uncertainties are needed for quantitative model appraisals, to apprehend whether emerging structures in tomographic images are resolved given the data and their errors (see Section 2.4). Once the generalized inverse $\mathbf{x}^{(k)}$ is known, one can directly infer the estimate $\hat{m}^{(k)}$, resolving kernel $A^{(k)}$ and uncertainty $\sigma_{\hat{m}^{(k)}}$:

$$\mathbf{x}^{(k)} \Longrightarrow \begin{cases} \sum_{i=1}^{N} x_i^{(k)} d_i & \longrightarrow \hat{m}^{(k)} \\ \sum_{i=1}^{N} x_i^{(k)} K_i(\mathbf{r}) & \longrightarrow A^{(k)}(\mathbf{r}) \\ (\sum_{i=1}^{N} (x_i^{(k)})^2)^{1/2} & \longrightarrow \sigma_{\hat{m}^{(k)}} . \end{cases} \tag{9}$$

The B–G approach consists in directly solving for the generalized inverse $\mathbf{x}^{(k)}$, such that $\mathbf{x}^{(k)}$ leads to the most peak-shaped resolving kernel $A^{(k)}$ around the query point $\mathbf{r}^{(k)}$, while moderating at most the propagated noise, that is, minimizing the variance $\sigma^2_{\hat{m}^{(k)}}$. Zaroli (2016) has introduced and adapted to large-scale, linear and discrete seismic tomography problems the SOLA method, an alternative B–G formulation which retains all its advantages but is more efficient and versatile in the explicit construction of resolving kernels (Pijpers & Thompson 1992). We aim at extending the discrete SOLA tomographic method to the parameter-free case, named *parameter-free SOLA tomography*. The key idea of SOLA is to specify an *a priori* target form $T^{(k)}$ for each resolving kernel $A^{(k)}$. Those ‘target’ resolving kernels, $T^{(k)}$, are referred as target kernels for short. In the parameter-free case, a target kernel $T^{(k)}$ is formally defined as follows:

$$T^{(k)}(\mathbf{r}) = \begin{cases} 1/\int_{\mathbf{r}\in\mathbb{S}^{(k)}} \mathrm{d}^3\mathbf{r} & \text{if} \quad \mathbf{r} \in \mathbb{S}^{(k)} \\ 0 & \quad\text{elsewhere} \end{cases} , \tag{10}$$

where $\mathbb{S}^{(k)}$ is a volumetric region well localized in the model space (e.g. a ball or a spheroid), which is centred on $\mathbf{r}^{(k)}$ and whose size represents an *a priori* estimate of the local resolution. Note that (10) implies that target kernels also satisfy to the unimodular condition:

$$\int T^{(k)}(\mathbf{r})\, \mathrm{d}^3\mathbf{r} = 1 . \tag{11}$$

Rather than minimizing the spread of each resolving kernel, SOLA aims at minimizing the integrated squared difference between each resolving kernel and its associated target kernel. That is, for every query point, $\mathbf{r}^{(k)}$, the parameter-free SOLA minimization problem consists in finding the coefficients $\mathbf{x}^{(k)} \in \mathbb{R}^N$ such that:

$$\begin{cases} \underbrace{\int \left[A^{(k)}(\mathbf{r}) - T^{(k)}(\mathbf{r}) \right]^2 d^3\mathbf{r}}_{\substack{\text{resolution}\\ \text{misfit}}} + \eta^2 \underbrace{\sigma^2_{\hat{m}^{(k)}}}_{\substack{\text{model}\\ \text{variance}}} = \min \\ \text{s.t.} \quad \underbrace{\int A^{(k)}(\mathbf{r}) d^3\mathbf{r} = 1}_{\substack{\text{unimodular}\\ \text{condition}}} . \end{cases} \tag{12}$$

Since the value of the trade-off (resolution versus uncertainty) parameter η is free to differ for every query point $\mathbf{r}^{(k)}$, one should rather write it as $\eta^{(k)}$. In this study, we choose it to be constant-valued and then drop the k subscript. Indeed, as suggested by Zaroli (2016) and Zaroli *et al.* (2017), a constant trade-off parameter η may lead to ‘*globally coherent*’ tomographic images when the size of target kernels is set to spatially vary as ray density, that is, a proxy for the *a priori* local resolution (see Section 2.4). The parameter-free SOLA minimization problem (12) can be written in the matrix form:

$$\begin{cases} \mathbf{F}^{(\eta)}\mathbf{x}^{(k)} & = \mathbf{u}^{(k)} \\ \text{s.t.} \quad \mathbf{c}^{\mathrm{T}}\mathbf{x}^{(k)} & = 1 , \end{cases} \tag{13}$$

where elements of the symmetric matrix $\mathbf{F}^{(\eta)} = (F^{(\eta)}_{ii'})_{1\le i,i' \le N}$, and vectors $\mathbf{c} = (c_i)_{1\le i\le N}$ and $\mathbf{u}^{(k)} = (u_i^{(k)})_{1\le i\le N}$ are given by

$$\begin{cases} F^{(\eta)}_{ii'} & = \int K_i(\mathbf{r}) K_{i'}(\mathbf{r})\, \mathrm{d}^3\mathbf{r} + \eta^2 \delta_{ii'} \\ c_i & = \int K_i(\mathbf{r})\, \mathrm{d}^3\mathbf{r} \\ u_i^{(k)} & = \int T^{(k)}(\mathbf{r})\, K_i(\mathbf{r}) \mathrm{d}^3\mathbf{r} , \end{cases} \tag{14}$$

with δ the Kronecker symbol. Though systems as (13) are usually solved using Lagrange multipliers, we rather follow Nolet (1985) and Zaroli (2016) and use an LSQR-based approach. Let us consider the three column vectors (assuming $c_1 \neq 0$):

$$\hat{\mathbf{x}}^{(k)} = \left(x_i^{(k)} \right)_{2\le i\le N} , \quad \hat{\mathbf{c}} = \left(\frac{c_i}{c_1} \right)_{2\le i\le N} , \quad \mathbf{e}_1 = (\delta_{i1})_{1\le i\le N} . \tag{15}$$

We wish the resolving kernel $A^{(k)}$ to satisfy to the unimodular condition (3), which may also be written as follows:

$$\mathbf{c}^{\mathrm{T}}\mathbf{x}^{(k)} = 1 . \tag{16}$$

The first element of $\mathbf{x}^{(k)}$ can be expressed in terms of the others:

$$x_1^{(k)} = c_1^{-1} - \hat{\mathbf{c}}^{\mathrm{T}}\hat{\mathbf{x}}^{(k)} , \tag{17}$$

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

$$\mathbf{x}^{(k)} = \mathbf{B}\hat{\mathbf{x}}^{(k)} + c_1^{-1}\mathbf{e}_1 , \tag{18}$$

where the matrix $\mathbf{B}$ is defined as:

$$\mathbf{B} = \begin{pmatrix} -\hat{\mathbf{c}}^{\mathrm{T}} \\ \mathbf{I}_{N-1} \end{pmatrix} , \tag{19}$$

with $\mathbf{I}_{N-1}$ the identity matrix of order $N-1$. The parameter-free SOLA problem (13) consists in solving for $\hat{\mathbf{x}}^{(k)}$ the system:

$$\mathbf{H}^{(\eta)}\hat{\mathbf{x}}^{(k)} = \mathbf{v}^{(k,\eta)} , \tag{20}$$

and then inferring the generalized inverse solution $\mathbf{x}^{(k)}$ from $\hat{\mathbf{x}}^{(k)}$, where the matrix $\mathbf{H}^{(\eta)}$ and vector $\mathbf{v}^{(k,\eta)}$ are defined as:

$$\begin{cases} \mathbf{H}^{(\eta)} & = \mathbf{F}^{(\eta)}\mathbf{B} \\ \mathbf{v}^{(k,\eta)} & = \mathbf{u}^{(k)} - \mathbf{F}^{(\eta)} c_1^{-1}\mathbf{e}_1 . \end{cases} \tag{21}$$

We use LSQR (Paige & Saunders 1982) to numerically solve (20); for a given η, it iteratively converges to the solution:

$$\hat{\mathbf{x}}^{(k,\eta)} = \underset{\hat{\mathbf{x}}^{(k)} \in \mathbb{R}^{N-1}}{\arg\min} : ||\mathbf{v}^{(k,\eta)} - \mathbf{H}^{(\eta)}\hat{\mathbf{x}}^{(k)}||^2 , \tag{22}$$

where $|| \cdot ||$ denotes the L_2-norm.

2.3 Numerical considerations

Since a constant-valued trade-off parameter η is assumed, there are P parameter-free SOLA tomographic systems (20) to be set up, and solved, where P is the total number of query points $\mathbf{r}^{(k)}$. First, one needs to compute P vectors $\mathbf{v}^{(k,\eta)} = (v_i^{(k,\eta)})_{1\le i\le N}$, with

$$v_i^{(k,\eta)} = \underbrace{\int T^{(k)} K_i}_{u_i^{(k)}} - \underbrace{\left[\int K_i K_1 + \eta^2 \delta_{i1}\right]}_{F_{i1}^{(\eta)}} c_1^{-1} . \tag{23}$$

This task mainly consists in calculating, at most, $P \times N$ integrals $\int T^{(k)} K_i$; it is cheap to compute the small fraction of non-zero integrals $\int K_i K_1$ (see Appendix A5). In this study, we assume that there are much more data than query points, that is, $P/N \ll 1$ (e.g. $P/N \simeq 16$ per cent in Section 2.4). Our view is that there is no need to consider too many query points, that is, not more than required to fit the spatial variations of the *a priori* local resolution; the same argument holds for data-driven, irregular tomographic grids (e.g. Sambridge *et al.* 1995; Nolet & Montelli 2005; Zaroli *et al.* 2015). Concerning the application in Section 2.4, we report that it is not costly to compute, in parallel, all those, at most, $P \times N \ll N^2$ integrals $\int T^{(k)} K_i$. Numerical details for calculating them are discussed in Appendix A. Second, one has to compute the matrix $\mathbf{H}^{(\eta)}$, of size $N \times (N-1)$, whose elements are given by

$$H_{\mu\nu}^{(\eta)} = \underbrace{\int K_\mu K_{\nu+1} + \eta^2 \delta_{\mu,\nu+1}}_{F_{\mu,\nu+1}^{(\eta)}} - \underbrace{\left[\int K_\mu K_1 + \eta^2 \delta_{\mu 1}\right]}_{F_{\mu 1}^{(\eta)}} \underbrace{\frac{c_{\nu+1}}{c_1}}_{\hat{c}_\nu} , \tag{24}$$

where $1 \le \mu \le N$ and $1 \le \nu \le N-1$. Since the matrix $\mathbf{H}^{(\eta)}$ does *not* depend on the query point, it does not need to be recomputed P times. This nice property is due to the SOLA formulation itself—a crucial advantage compared to the B–G approach (e.g. Pijpers 1997). The matrix $\mathbf{H}^{(\eta)}$ can be easily derived from the symmetric matrix $\mathbf{F}^{(\eta)}$ of order N. The main computational difficulty of parameter-free SOLA tomography, compared to discrete SOLA, arises from the calculation of $\mathbf{F}^{(\eta)}$. Though it is not costly to compute its N diagonal elements, $F_{ii}^{(\eta)} = \int K_i^2 + \eta^2$, it may be expensive to compute its $N(N-1)/2$ off-diagonal elements, $F_{ii'}^{(\eta)} = \int K_i K_{i'}$. Indeed, as the number of integrals $\int K_i K_{i'}$ grows as a function of N^2, this may be a computational burden when facing large data sets (see Section 3), especially if one aims at fully capturing the form of finite-frequency sensitivity kernels (see Section 2.1). Numerical details for calculating $\mathbf{F}^{(\eta)}$ are given in Appendix A. It is straightforward to calculate $\mathbf{H}^{(\eta)}$ and $\mathbf{v}^{(k,\eta)}$ for different η values. Last, but not least, it is worth of noting that parameter-free SOLA tomography is well suited for parallel computations, since the problem can be easily separated into a number of independent computational tasks, for example, to calculate all the P vectors $\mathbf{v}^{(k,\eta)}$, the $\sim N^2/2$ integrals $\int K_i K_{i'}$, and the P LSQR-solution vectors $\hat{\mathbf{x}}^{(k,\eta)}$. Finally, we show in Appendix A how to reduce the computational costs and memory requirements involved in setting up and solving parameter-free SOLA tomographic systems, tailored to teleseismic finite-frequency body-wave mantle tomography.

2.4 Application

We aim at applying the parameter-free SOLA tomographic method, within a finite-frequency physical framework (Dahlen *et al.* 2000), to the problem of imaging and appraising isotropic variations of shear-wave velocities in the mantle region below Southeast Asia, with respect to the reference radial velocity model iasp91 (Kennett & Engdahl 1991). The Southeast-Asia region, depicted in Fig. 2(b), is chosen without any particular purpose, except that it is characterized by several deep subduction systems, and because of a large amount of data available in that region, making it worth for first time testing parameter-free SOLA tomography.

Our data set consists in $N = 27\,070$ teleseismic S and SS time residuals, measured by cross-correlation technique at 22 s central period (passband Gaussian filter) (Zaroli *et al.* 2010). The associated, globally distributed earthquakes and receivers are shown in Fig. 2(a). These data are a subset of those used in previous, global-scale, discrete SOLA and DLS tomographic experiments (Zaroli 2016). To generate it, a selection criterion was applied to ensure that every data sensitivity kernels sample the mantle region of interest. Estimates of data errors include earthquake-location and measurement-process errors (Zaroli *et al.* 2010, 2013); these original errors were additionally increased by 30 per cent by Zaroli (2016) to have unit reduced chi-square for a global DLS model. Each finite-frequency sensitivity kernel is calculated at 22 s period on a grid made of 50 km edge-length cubic cells (see Fig. 1a and Appendix A1), using analytical formulae (Zaroli *et al.* 2013) for which is assumed a Gaussian source power spectrum (Hung *et al.* 2001).

Each target kernel $T^{(k)}$ is a spheroid centred on a query point. We follow Zaroli (2016) to specify the locations and sizes of target kernels. We use the ray density as a first-order proxy for the spatial variations of the *a priori* local resolution, make an educated guess about the *a priori* resolving-length bounds, and then interpolate to determine the sizes of target kernels at given locations. To limit the number of query points, their locations are chosen such that they fit the spatial distribution of the *a priori* local resolving length. Fig. 3 displays lateral views of all the $P = 4310$ target kernels. They are spanning the region of interest at eight depths in the 350–1410 km range, where our data coverage is the most relevant. The lateral radius of target kernels is driven by ray density; it ranges from 200 to 1000 km and represents the *a priori*, isotropic, local, lateral resolving length. Their radial radius gradually varies from 130 to 200 km in the 350–1410 km depth range, respectively, and represents the *a priori*, local, radial resolving length. A constant value is chosen for the trade-off parameter η, after having tested a few different values. Note that SOLA solutions depend on tunable inversion parameters (target kernels and trade-off parameter), so that different choices would result in different, unbiased model estimates and appraisals—leading to different, fully quantitative and thus meaningful (at least in a mathematical sense) model interpretations. As a remark, in the case of a synthetic tomographic problem, Zaroli *et al.* (2017) illustrate the variability of SOLA solutions (i.e. model, resolution, uncertainty) as a function of the trade-off parameter η.

Conventional tomographic images can be built, for plotting purposes, from linear interpolations among the model estimates, $\hat{m}^{(k)}$. Fig. 4 shows the resulting images in the upper mantle (350 km depth), transition zone (465, 595 km depth) and mid lower mantle (735, 885, 1035, 1210, 1410 km depth). Some features seem to be worth of further investigations. For example, those characterized by strongly positive velocity anomalies (bluish); some of them

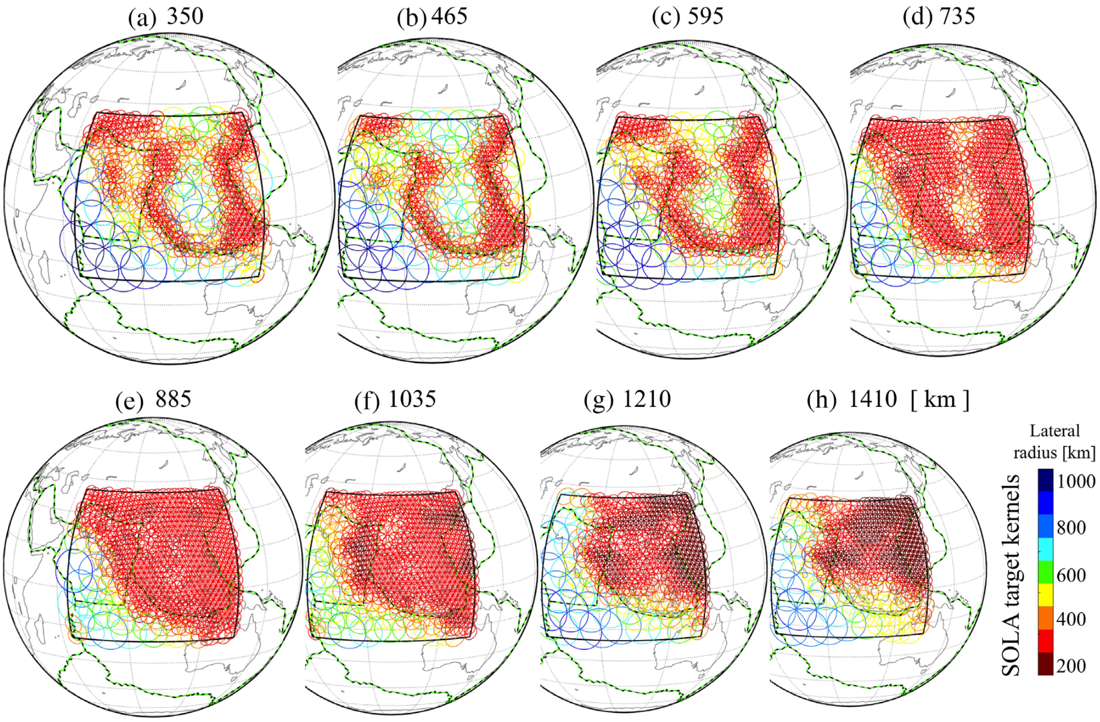

Figure 3. Drawn circles represent lateral, 2–D views of all the 4310 parameter-free SOLA target kernels spanning Southeast Asia at eight different depths in the mantle (350 to 1410 km). Note that a query point lies at the centre of each circle, whose radius is colour coded and ray-density driven (see Section 2.4).

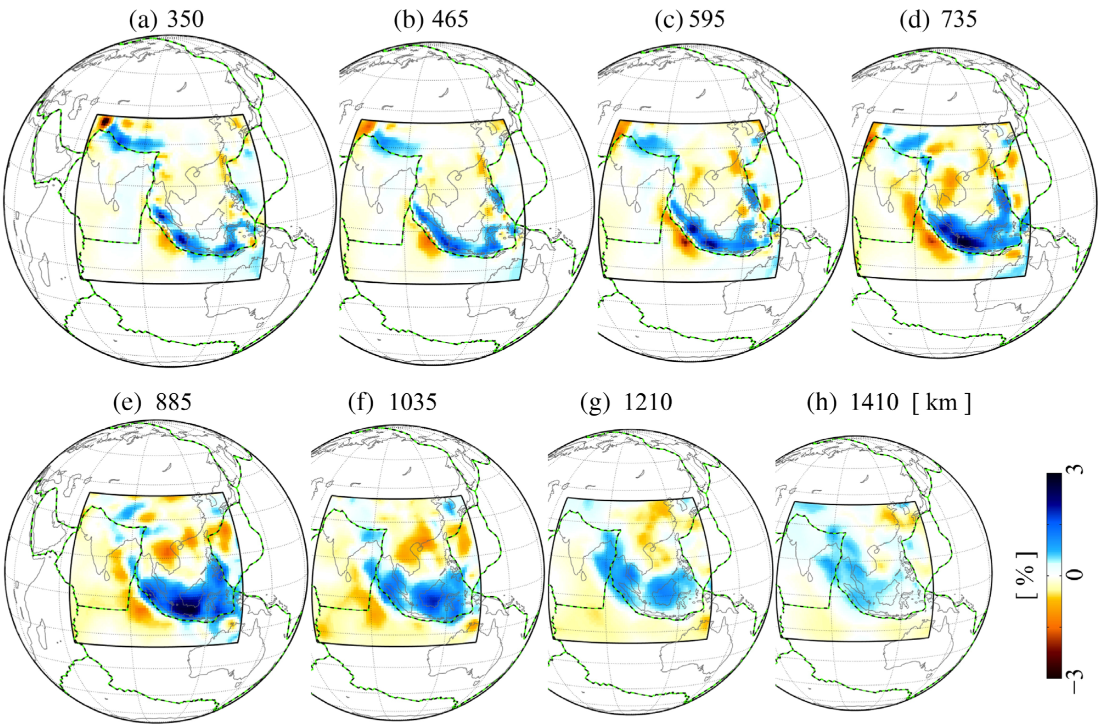

Figure 4. Parameter-free SOLA tomographic images (see Section 2.4).

correspond to major deep subducted slabs, as in the Sumatra and Java regions (indicated in Fig. 2b). Even more interesting are the negative anomalies (reddish) appearing on west–southwest side of the Sumatra slab in the 350–1035 km depth range (see Figs 4a–f), while none are showing up nearby on the south side of the Java slab. Quantitatively interpreting these complex structural features would require to analyse them in the light of resolving kernels and uncertainties, which is beyond the scope of this work (see Section 3). As a remark, Zaroli *et al.* (2017) report that for their synthetic tomographic experiments, the discrete SOLA models do fit the data at the same level as the DLS models—while the SOLA method is not specifically aimed at minimizing the data misfit. In this study, one

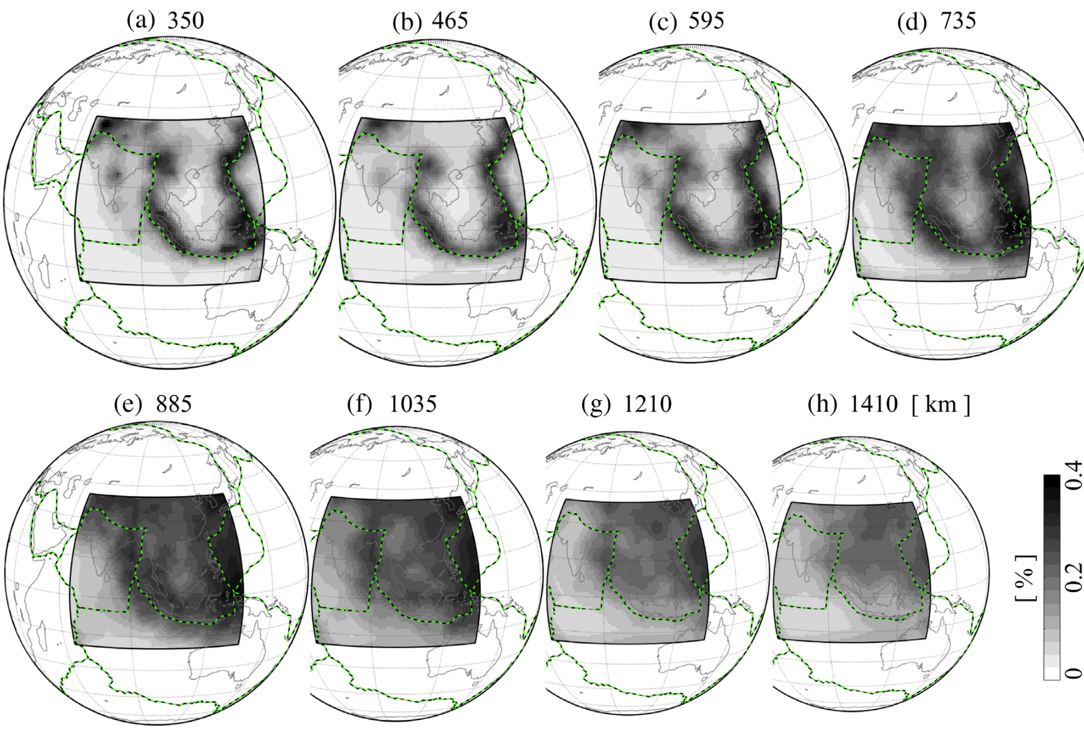

Figure 5. Parameter-free SOLA uncertainties (see Section 2.4).

cannot compute the data misfit because the parameter-free SOLA model is calculated in a limited region (Southeast Asia), while the data coverage spans almost the whole mantle.

Fig. 5 shows interpolated maps of uncertainties, $\sigma_{\hat{m}^{(k)}}$. They merely represent the amount of data noise that propagates into the model estimates, and come with some underlying assumptions on the noise itself (which is assumed to follow uncorrelated, zero-mean, Gaussian statistics). While the amplitudes of the model estimates are within ± 3 per cent, one sees that their uncertainties may reach at most 0.4 per cent. As expected, the spatial variations of uncertainties are similar to those of the ray-density driven target kernels (see Fig. 3). In the regions where the size of target kernels is large (small), that is, the *a priori* local resolution is poor (good), the uncertainty is low (high), respectively—the unavoidable trade-off between resolution and uncertainty (e.g. Menke 1989).

Resolving kernels, $A^{(k)}$, have to be calculated in a consistent way with respect to the discretization of data sensitivity kernels. In this study, each resolving kernel is then computed on a grid which consists in 50 km edge-length cubic cells surrounding the considered query point (see Appendix A4), enabling us to fully capture its form. Fig. 6 shows three examples of resolving kernels, associated to three query points located below Sulawesi Island at 350, 595 and 1035 km depth. Horizontal and vertical cross-sections through those resolving kernels are shown. As expected for teleseismic *S*-wave tomography, their lateral (radial) extent is smaller (larger) in the upper than lower mantle, respectively. That is, vertical smearing (horizontal leaking) mainly occurs in the upper (lower) mantle, respectively. As a remark, one possible artefact that SOLA could be prone to arises when resolving kernels are significantly negative, since they cannot be considered as truly averaging kernels. This does not appear to be the case in this study; for example, only weak negative values are observed in Fig. 6. Pijpers & Thompson (1994) and Zaroli *et al.* (2017) discuss how to avoid such artefacts by enlarging the size of target kernels. Note that the target kernels may differ from the actual resolving kernels (see Fig. 6). This simply means that the *a priori* local resolving length was chosen too optimistically. However, as long as the resolving kernels are mostly non-negative and spatially well localized, they can be exploited for appraising the actual local resolution in tomographic images.

Finally, to illustrate the appraisal of tomographic features with the parameter-free SOLA approach, we discuss whether is resolved the lateral flattening of the Sumatra–Java slabs in mid lower mantle. Indeed, the lateral extent of these slabs appears to be much smaller in the upper mantle (see Fig. 4a) and transition zone (see Figs 4b and c), compared to mid lower mantle (see Figs 4d–g). To have a clue on the variations with depth of the lateral resolving length in this region, let us consider the three resolving kernels shown in Fig. 6 and estimate their lateral extent, that is, the local lateral resolving length. We report that, below Sulawesi Island, the lateral resolving length is, at most, 200 km (300, 500) at a depth of 350 km (595, 1035), respectively. In particular, Fig. 6(c) indicates that the lateral extent of these slabs around Sulawesi Island at 1035 km depth is much larger than the local lateral resolving length. Moreover, model estimates are 0.95, 0.92 and 1.03 per cent for the corresponding query points below Sulawesi Island at 350, 595 and 1035 km depth, while their uncertainties are three times smaller, that is, 0.33, 0.30 and 0.29, respectively. Thus, one can argue that the slab lateral flattening that takes place in mid lower mantle, at least below Sulawesi Island, is resolved given our data and their errors.

3 DISCUSSION AND PERSPECTIVES

First, we aim to discuss when parameter-free SOLA tomography should be preferred to discrete SOLA tomography, and vice versa. Let M^{∞} be the minimum number of parameters required to parametrize the entire model space, so that every projected sensitivity kernels are accurate. In the context of teleseismic, finite-frequency, *S*-wave mantle tomography, it leads to $M^{\infty} \simeq 10^6$–10^7

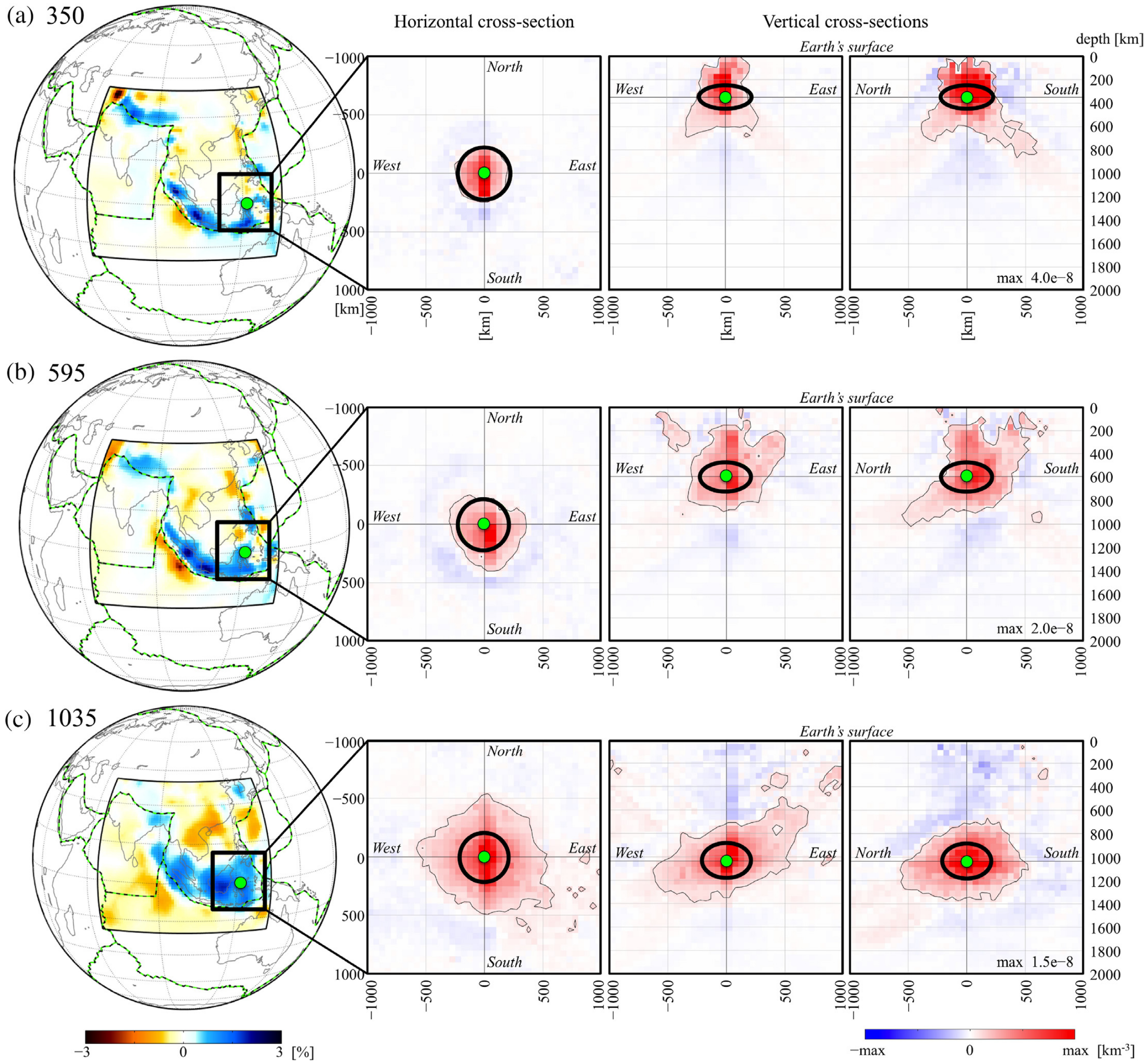

Figure 6. Visualization of horizontal and vertical cross-sections across three parameter-free SOLA resolving kernels (see Section 2.4). The associated three query points are highlighted with green dots; they are located at three selected depths (350, 595 and 1035 km) below the Sulawesi island. Tomographic images are also displayed. Each drawn black circle (ellipse) represents the horizontal (vertical) spatial extent of the corresponding spheroid-shape target kernel, respectively.

(see Section 2.1). If one aims at fully exploiting finite-frequency theory, but cannot handle discrete SOLA inversions with M^∞ parameters, then one should definitely use the parameter-free SOLA approach. However, if the total number of data is too high, for example, $N \gg 10^5$, it may not be tractable to compute the $\sim N^2/2$ elements of the matrix $\mathbf{F}^{(\eta)}$ (see Section 2.3, Appendix A). Hence, one may have no choice but to move back to discrete SOLA with a total number of parameters $M \ll M^\infty$ (and thus simply have to project N sensitivity kernels on a given, coarse tomographic grid). Consequences would be that some, if not all, projected sensitivity kernels would become unsuitable for finite-frequency imaging purposes (see Fig. 1b). Note that parameter-free SOLA tomography is particularly well suited for regional-scale experiments, for which moderate-size data sets are typically encountered, enabling to take full advantage of finite-frequency theory even with modest computational resources (see Appendix A6).

As an additional comparison of parameter-free SOLA versus other tomographic schemes, let us reconsider the standard, discrete DLS approach, and focus on the problem of imaging and appraising deep mantle plumes, a topic of high interest, recently revisited by Maguire *et al.* (2018). In their study, various plume models and earthquakes–receivers settings are considered to generate synthetic sets of teleseismic body-wave time residuals, inverted using DLS and finite-frequency sensitivity kernels. Relying on powerful computational facilities, they are able to parametrize the entire mantle using a Cartesian cubed sphere approach (e.g. Ronchi *et al.* 1996), which consists in ∼3.5 million roughly cubic cells (∼65 km edge length), enabling projected finite-frequency sensitivity kernels to be accurate. As a remark, other recent studies (e.g. Charléty *et al.* 2013; Nolet *et al.* 2019) were able to derive teleseismic, DLS-based tomographic images when using so many parameters. However, handling millions parameters makes prohibitive to compute the full

DLS generalized inverse (e.g. Bogiatzis *et al.* 2016). Hence, resolution and uncertainty informations cannot be fully taken into account to quantitatively analyse, for example, plume-like, features in DLS images. Note that the typical size of Maguire *et al.* (2018)'s data sets is $N \simeq 5.\,10^4$, what parameter-free SOLA can handle with relatively modest resources and while solving all at once both the imaging and appraising problems. Moreover, DLS images may be locally biased in regions with poor data illumination, due to *ad hoc* regularization, such as below isolated receivers where ray paths are quasi-vertical (Zaroli *et al.* 2017)—while SOLA solutions are explicitly constrained to be unbiased. Maguire *et al.* (2018) identify part of this bias effect. Assuming an hypothetical vertical conduit of 'slow' anomalies in the mantle below Hawaii, they show that the recovered plume may contain prominent 'fast' anomalies—what is a reminder that DLS-based tomographic images could lead to physical misinterpretations.

In the following, we aim to highlight some perspectives related to parameter-free SOLA tomography. First, since it yields unbiased images with resolution and uncertainty informations, one could naturally aim at evaluating in a fully quantitative way whether specific features of interest (e.g. mantle plumes, slabs) are resolved, or not. Future work could consist in designing algorithms aimed at better apprehending and visualizing this new wealth of available informations (model, resolution, uncertainty), enabling to quickly identify resolved features. Since reliable estimates of model uncertainties $\sigma_{\hat{m}^{(k)}}$ require reliable estimates of data errors σ_{d_i}, one could aim at better evaluating the noise contributions in various data sets, and also to investigate the impact of assuming a wrong noise model on the SOLA results.

In a similar line to earlier tomographic filtering studies (e.g. Ritsema *et al.* 2007; Schuberth *et al.* 2009; Davies *et al.* 2012; Zaroli *et al.* 2017), one could investigate how user-defined input features (e.g. mantle plumes, subducted slabs, whole-mantle geodynamical models) are seen through SOLA resolving kernels, that is, analysing the term 'filtered input model' in eq. (25). Furthermore, with the explicit knowledge of the SOLA generalized inverse, one could also investigate the amount of data noise which is expected to propagate into the recovered tomographic features. That is, the output model estimate $\hat{m}^{(k)}_{\text{out}}$ for a specific query point can be related to a given input model m_{in} as follows:

$$\begin{aligned}\hat{m}^{(k)}_{\text{out}} &= \textstyle\sum_i x_i^{(k)} \underbrace{\left(\int K_i m_{\text{in}} + n_i^{\text{synth}}\right)}_{\text{synthetic data, } d_i^{\text{synth}}} \\ &= \underbrace{\int A^{(k)} m_{\text{in}}}_{\text{filtered input model}} + \underbrace{\sum_i x_i^{(k)} n_i^{\text{synth}}}_{\text{propagated synthetic noise}}\end{aligned} \tag{25}$$

where the ith synthetic noise component n_i^{synth} is randomly drawn from a zero-mean normal distribution with unit standard deviation [since the original tomographic system (1) was scaled by the data errors], and the generalized inverse components $x_i^{(k)}$ and the resolving kernel $A^{(k)}$ are those from the actual SOLA tomographic experiment. Therefore, as in the discrete SOLA case (Zaroli *et al.* 2017), the parameter-free SOLA framework provides an efficient and fully quantitative way for comparing input and output features, by means of analysing the filtered input model *and* the propagation of synthetic noise into the tomographic model solution. Note that the parameter-free SOLA 'filtered input model' can also be indirectly computed as $\int A^{(k)} m_{\text{in}} = \sum_i x_i^{(k)} \int K_i m_{\text{in}}$. This may be useful if it is cheaper to compute the N integrals $\int K_i m_{\text{in}}$ compared to the P integrals $\int A^{(k)} m_{\text{in}}$ (requiring all the resolving kernels to be explicitly calculated).

The parameter-free SOLA approach could also be exploited to invert onset-time data for which sensitivity kernels are infinitesimally narrow rays. Since one expects less pairs of crossing rays, compared to crossing finite-frequency kernels, it should be cheaper to compute the integrals $\int K_i^{\text{ray}} K_{i'}^{\text{ray}}$. Onset-time and correlation-time data, modelled by rays and finite-frequency sensitivity kernels (e.g. Montelli *et al.* 2004b; Obayashi *et al.* 2013; Hosseini 2016), could also be jointly inverted with SOLA. In addition, since fully accurate data sensitivity kernels are exploited in parameter-free SOLA inversions, one could revisit comparisons of ray-theory versus finite-frequency versus multifrequency tomography, while comparing differences in terms of the images and their appraisals. Finally, parameter-free SOLA tomography could be applied to other linear problems, such as, for example, finite-frequency surface-wave tomography (e.g. Zhou *et al.* 2005; Nolet 2008).

4 CONCLUSION

We have presented a parameter-free, linear B–G inversion scheme, tractable for seismic tomography problems—named *parameter-free SOLA tomography*. Theoretical and numerical developments have been illustrated for teleseismic body-wave mantle tomography, in a finite-frequency physical framework. This new tomographic scheme leads to efficient, embarrassingly parallel, computations of unbiased images, accompanied by meaningful resolution and uncertainty informations. Furthermore, since it does not assume any parametrization of the model space, it enables numerically accurate data sensitivity kernels to be effectively exploited in tomographic inversions. This is a key advantage over discrete tomographic methods for which data sensitivity kernels are often inaccurate, as they are projected on coarse parametrizations. The most costly task of parameter-free SOLA tomography is the calculation of $\sim N^2/2$ volumetric integrals of the form $\int K_i K_{i'}$, where $(K_i, K_{i'})$ is a pair of data sensitivity kernels and N the total number of data, what could make intractable very large data sets (e.g. $N \gg 10^5$). Nevertheless, using modest computational facilities, we have successfully inverted a set of 27 070 teleseismic, finite-frequency, S-wave time residuals, with focus on imaging and appraising shear wave velocity anomalies lying in the mantle below Southeast Asia, in the 350–1410 km depth range.

To conclude, parameter-free SOLA tomography is particularly well suited for regional-scale experiments, for which moderate-size data sets are frequently encountered, so that limited computational resources are sufficient—while enabling quantitative appraisals of tomographic features, and to take full advantage of finite-frequency data sensitivity kernels.

ACKNOWLEDGEMENTS

The author is grateful to Lapo Boschi (editor), Kasra Hosseini and an anonymous reviewer whose constructive comments have helped to improve this paper, and wishes to thank David Al-Attar for stimulating discussions on Backus–Gilbert theory, Sophie Lambotte for reading an earlier version of this manuscript and Franco Zaroli for invaluable help on some program optimization related to B-trees.

REFERENCES

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

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.

Backus, G. & Gilbert, J.F., 1970. Uniqueness in the inversion of inaccurate gross Earth data, *Phil. Trans. R. Soc. A.,* **266**(1173), 123–192.

Bayer, R. & McCreight, E., 1972. Organization and maintenance of large ordered indexes, *Acta Informatica,* **1**(3), 173–189.

Bogiatzis, P., Ishii, M. & Davis, T., 2016. Towards using direct methods in seismic tomography: computation of the full resolution matrix using high-performance computing and sparse QR factorization, *Geophys. J. Int.,* **205,** 830–836.

Boschi, L., Becker, T.W., Soldati, G. & Dziewonski, A.M., 2006. On the relevance of Born theory in global seismic tomography, *Geophys. Res. Lett.,* **33,** L06302, doi:10.1029/2005GL025063.

Charléty, J., Voronin, S., Nolet, G., Loris, I., Simons, F., Sigloch, K. & Daubechies, I.C., 2013. Global seismic tomography with sparsity constraints: comparison with smoothing and damping regularization, *J. geophys. Res,* **118,** 1–13.

Chevrot, S., Martin, R. & Komatitsch, D., 2012. Optimized discrete wavelet transforms in the cubed sphere with the lifting scheme—implications for global finite-frequency tomography, *Geophys. J. Int.,* **191,** 1391–1402.

Dahlen, F.A. & Nolet, G., 2005. Comment on the paper "On sensitivity kernels for wave-equation transmission tomography" by de Hoop and Van der Hilst, *Geophys. J. Int.,* **163,** 949–951.

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.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 planet. Sci. Lett.,* **353-354,** 253–269.

Hosseini, K., 2016. Global multiple-frequency seismic tomography using teleseismic and core-diffracted body waves, *PhD thesis*, University of Munich (LMU).

Hosseini, K. & Sigloch, K., 2015. Multifrequency measurements of core-diffracted P waves (Pdiff) for global waveform tomography, *Geophys. J. Int.,* **203**(1), 506–521.

Hung, S.-H., Dahlen, F.A. & Nolet, G., 2001. Wavefront-healing: a banana-doughnut perspective, *Geophys. J. Int.,* **146,** 289–312.

Hung, S.-H., Shen, Y. & Chiao, L.-Y., 2004. Imaging seismic velocity structure beneath the Iceland hotspot: a finite-frequency approach, *J. geophys. Res,* **109,** B08305, doi:10.1029/2003JB002889.

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

Kennett, B.L.N. & Bunge, H.-P., 2008. *Geophysical Continua,* Cambridge Univ. Press.

Maceira, M., Larmat, C., Porritt, R.W., Higdon, D.M., Rowe, C.A. & Allen, R.M., 2015. On the validation of seismic imaging methods: finite-frequency or ray theory? *Geophys. Res. Lett.,* **42**(2), 323–330.

Maguire, R., Ritsema, J., Bonnin, M., van Keken, P. & Goes, S., 2018. Evaluating the resolution of deep mantle plumes in teleseismic traveltime tomography, *J. geophys. Res.,* **123**(1), 384–400.

Marquering, H., Nolet, G. & Dahlen, F.A., 1998. Three-dimensional waveform sensitivity kernels, *Geophys. J. Int.,* **132,** 521–534.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory,* revisededn, 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.

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., 2005. Optimum parameterization of tomographic models, *Geophys. J. Int.,* **161,** 365–372.

Nolet, G., Dahlen, F.A. & Montelli, R., 2005. Traveltimes and amplitudes of seismic waves: a re-assessment, in *Seismic Earth: Analysis of broadband seismograms,* Vol. **157,** pp. 37–48, eds Levander, A. & Nolet, G., AGU Monograph Series.

Nolet, G. *et al.*, 2019. Imaging the Galápagos mantle plume with an unconventional application of floating seismometers, *Sci. Rep.,* **9**(1), 1326, doi:10.1038/s41598-018-36835-w.

Obayashi, M., Yoshimitsu, J., Nolet, G., Fukao, Y., Shiobara, H., Sugioka, H., Miyamachi, H. & Gao, Y., 2013. Finite frequency whole mantle *P* wave tomography: improvement of subducted slab images, *Geophys. Res. Lett.,* **40**(21), 5652–5657.

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 University Press.

Pijpers, F.P., 1997. Inversions in astronomy and the SOLA method, in *Inverse Problems in Wave Propagation,* pp. 419–442, eds Chavent, G., Sacks, P., Papanicolaou, G. & Symes, W.W., 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.

Rawlinson, N., Pozgay, S. & Fishwick, S., 2010. Seismic tomography: a window into deep Earth, *Phys. Earth planet. Inter.,* **178**(3-4), 101–135.

Rawlinson, N., Fichtner, A., Sambridge, M. & Young, M.K., 2014. Seismic tomography and the assessment of uncertainty, *Adv. Geophys.,* **55,** 1–76.

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.

Romanowicz, B., 2008. Using seismic waves to image Earth's internal structure, *Nature,* **451,** 266–268.

Ronchi, C., Iacono, R. & Paolucci, P., 1996. The cubed sphere: a new method for the solution of partial differential equations in spherical geometry, *J. Comput. Phys.,* **124,** 93–114.

Sambridge, M., Braun, J. & McQueen, H., 1995. Geophysical parameterisation and interpolation of irregular data using natural neighbours, *Geophys. J. Int.,* **122,** 837–857.

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? *Geochem. Geophys. Geosyst.,* **10**(5), Q05W03, doi:10.1029/2009GC002401.

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. & Nolet, G., 2006. Measuring finite-frequency body-wave amplitudes and traveltimes, *Geophys. J. Int.,* **167,** 271–287.

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. & de Hoop, M.V., 2005. Banana-doughnut kernels and mantle tomography, *Geophys. J. Int.,* **163,** 956–961.

Voronin, S. & Zaroli, C., 2018. Survey of computational methods for inverse problems, *Recent Trends in Computational Science and Engineering,* doi:10.5772/intechopen.73332.

Yang, T., Shen, Y., van der Lee, S., Solomon, S. & Hung, S.-H., 2006. Upper mantle beneath the Azores hotspot from finite-frequency seismic tomography, *Earth planet. Sci. Lett.,* **250,** 11–26.

Zaroli, C., 2010. Global multiple-frequency S-wave tomography of the Earth's mantle, *PhD thesis*, Strasbourg University.

Zaroli, C., 2016. Global seismic tomography using Backus–Gilbert inversion, *Geophys. J. Int.,* **207**(2), 876–888.

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.

Zaroli, C., Koelemeijer, P. & Lambotte, S., 2017. Toward seeing the Earth's interior through unbiased tomographic lenses, *Geophys. Res. Lett.,* **44,** 11399–11408.

Zhou, Y., Dahlen, F., Nolet, G. & Laske, G., 2005. Finite-frequency effects in global surface wave tomography, *Geophys. J. Int.,* **182,** 1025–1042.

APPENDIX A: COMPUTATIONAL ASPECTS

We show how to reduce the computational costs and storage requirements involved in setting up and solving the parameter-free SOLA systems (see Section 2.2), tailored to teleseismic body-wave mantle tomography within a finite-frequency physical framework.

A1 Sensitivity kernel discretization

Each finite-frequency data sensitivity kernel K_i is discretized on a local Cartesian grid that consists in regularly spaced, 50 km edge-length, cubic cells spanning a rectangular parallelepiped region surrounding K_i as illustrated in Fig. 1(a). This avoids using a global grid spanning the whole mantle, which would lead to consider and store much more cells, that is, several millions versus a few hundreds thousands (see Section 2.1), with most cells zero-valued. It is straightforward to transform these local Cartesian coordinates to the global Cartesian coordinates (Zaroli 2010), which is useful, for example, when evaluating whether two sensitivity kernels K_i and $K_{i'}$ are simultaneously non-zero valued at the same location (see Appendix A2).

A2 Parameter-free SOLA systems

In this study, each sensitivity kernel K_i is stored in a self-balancing B-tree structure (Bayer & McCreight 1972). Stored informations are non-zero kernel values $K_i(x, y, z)$ and associated local Cartesian coordinates (x, y, z). We report that this B-tree approach enables us to efficiently, and elegantly, numerically compute all the integrals encountered in our parameter-free SOLA tomographic problem (see Section 2.2), that is: $\int T^{(k)}K_i$, $\int K_i$, $\int K_i^2$, $\int K_i K_{i'}$.

Computing the symmetric matrix $\mathbf{F}^{(\eta)}$ is by far the most costly task in parameter-free SOLA tomography (see Section 2.3, Appendix A6). Thus, we aim to explain how are evaluated all the integrals of the form $\int K_i K_{i'=i+1\cdots N}$, where i is fixed. These $(N-i)$ integrals correspond to all the elements of the ith row of the upper-right half part of $\mathbf{F}^{(\eta)}$. We proceed as follows: 1) All the non-zero values of K_i and their locations (with respect to the local Cartesian grid, tailored to K_i), that is, $\{K_i(x, y, z), x, y, z\}$, are computed and stored in a B-tree structure referred as $Bt[K_i]$; 2) All the non-zero values of $K_{i'}$ and their locations, that is, $\{K_{i'}(x', y', z'), x', y', z'\}$, are computed on the fly; 3) Each triplet (x', y', z') is searched for in $Bt[K_i]$; if it is found, that is, if there is a triplet (x, y, z) such that $(x, y, z) \equiv (x', y', z')$ (meaning that (x, y, z) and (x', y', z') stand for the same global Cartesian coordinates, and that both $K_i(x, y, z)$ and $K_{i'}(x', y', z')$ are non-zero), then the integral value is updated: $\int K_i K_{i'} \leftarrow \int K_i K_{i'} + K_i(x, y, z)K_{i'}(x', y', z')\Delta V$, where $\Delta V = 50^3$ km^3; 4) Repeat 2) and 3) with $i' = i + 1 \cdots N$.

A few additional remarks. The search time of (x', y', z') in $Bt[K_i]$ is independent of whether (x', y', z') is found or not. This nice property is due to the self-balancing structure itself of B-trees. Since there is at most $Bt[K_i]$ and $K_{i'}$ to be stored at the time, our approach is not costly in terms of memory footprint. We choose to separate the calculation of all the rows of the upper-right half part of $\mathbf{F}^{(\eta)}$ into parallel tasks, such that the first N_1 contiguous rows are computed on a first processor, the next N_2 rows are computed on a second processor, etc. When using the previous algorithm to compute the first N_1 rows, one actually has to compute once K_1, twice K_2, three times K_3, etc. Therefore, to further speed-up the calculation of the first N_1 rows (and so on) of the upper-right half part of $\mathbf{F}^{(\eta)}$, all the sensitivity kernels K_i are i-index sorted such that their computational costs (or a proxy for it, e.g. the total ray path distance) are in decreasing order. The numbers of contiguous rows (N_1, N_2, etc.) may be chosen such that the work load on every processors is almost identical. Finally, note that new data sets can be easily embedded in existing matrix $\mathbf{F}^{(\eta)}$ and vectors $\mathbf{u}^{(k)}$.

A3 Non-crossing sensitivity kernels

To alleviate the computational burden of building the matrix $\mathbf{F}^{(\eta)}$, one may try to reduce the number of integrals $\int K_i K_{i'}$ to be effectively calculated. That is, to identify *a priori* some pairs of body-wave sensitivity kernels $(K_i, K_{i'})$ that do not cross each other; in which case $\int K_i K_{i'} = 0$. Different strategies could be designed for that purpose. For example, let assume that each sensitivity kernel spans a planar region within the mantle, geometrically defined by the earthquake–receiver great-circle arc. Then, from our ability to identify non-crossing great-circle arcs, one could infer at least some pairs of non-crossing sensitivity kernels. We have adapted this geometrical criterion to take into account volumetric finite-frequency sensitivity kernels, by simply considering 'fat' great-circle arcs.

A4 Resolving kernels

In parameter-free SOLA tomography, each resolving kernel $A^{(k)}$ has to be computed from eq. (7) in a consistent way with respect to the discretization of data sensitivity kernels (see Appendix A1). Thus, $A^{(k)}$ is calculated on a grid which consists in regularly spaced, 50 km edge-length, cubic cells spanning a volumetric region surrounding the query point (see Fig. 6). In practice, to speed-up the computation of $A^{(k)}$, we proceed as follows: (1) Storing $A^{(k)}$ itself in a B-tree; (2) Identifying *a priori* whether a given sensitivity kernel K_i will not cross the volumetric region where we aim at calculating $A^{(k)}$, in which case K_i would not contribute to $A^{(k)}$. Step (2) is based on a geometrical criterion similar to Appendix A3.

A5 Matrix sparsity

Preserving the sparsity of the matrix $\mathbf{H}^{(\eta)}$ is crucial in terms of storage, efficiency of LSQR solver, memory footprint, etc. From eq. (24) one sees that the sparsity of $\mathbf{H}^{(\eta)}$ can be optimized, by selecting the first-indexed sensitivity kernel K_1 such that it leads to minimizing the number of non-zero integrals $\int K_\mu K_1$, that is, maximizing the number of pairs of non-crossing kernels (K_μ, K_1). The kernel K_1 is found using a brute-force (computationally cheap) search to count the number of crossing kernels for each individual kernel using the geometrical criterion mentioned in Appendix A3. We report that, in the case of our application (see Section 2.4), $\mathbf{H}^{(\eta)}$ is ∼2 per cent dense.

A6 Computational cost

Concerning our experiment (see Section 2.4), for which the number of data is $N = 27\,070$, it takes ∼1 wk (CPU time) to compute in parallel using 70 processors (Intel Xeon E5-4657L 2.40 GHz) all the $\sim N^2/2$ elements of the symmetric matrix $\mathbf{F}^{(\eta)}$ – by far the most costly task faced in parameter-free SOLA tomography.