Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM-DSM hybrid method

Vadim Monteiller, Sébastien Chevrot, Dimitri Komatitsch, Yi Wang

Geophysical Journal International

Geophys. J. Int. (2015) **202,** 811–827
GJI Seismology

doi: 10.1093/gji/ggv189

Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method

Vadim Monteiller,[1,2,*] Sébastien Chevrot,[1] Dimitri Komatitsch[2] and Yi Wang[1]

[1] *Géosciences Environnement Toulouse, Observatoire Midi-Pyrénées, CNRS UMR* 5563, *Université Paul Sabatier, Toulouse, France.*
E-mail: vadim.monteiller@geoazur.unice.fr
[2] *Laboratoire de Mécanique et d'Acoustique, CNRS UPR* 7051, *Aix-Marseille University, Centrale Marseille, F*-13402 *Marseille Cedex* 20, *France*

Accepted 2015 April 29. Received 2015 April 29; in original form 2015 January 27

SUMMARY

We present a method for high-resolution imaging of lithospheric structures based on full waveform inversion of teleseismic waveforms. We model the propagation of seismic waves using our recently developed direct solution method/spectral-element method hybrid technique, which allows us to simulate the propagation of short-period teleseismic waves through a regional 3-D model. We implement an iterative quasi-Newton method based upon the L-BFGS algorithm, where the gradient of the misfit function is computed using the adjoint-state method. Compared to gradient or conjugate-gradient methods, the L-BFGS algorithm has a much faster convergence rate. We illustrate the potential of this method on a synthetic test case that consists of a crustal model with a crustal discontinuity at 25 km depth and a sharp Moho jump. This model contains short- and long-wavelength heterogeneities along the lateral and vertical directions. The iterative inversion starts from a smooth 1-D model derived from the IASP91 reference Earth model. We invert both radial and vertical component waveforms, starting from long-period signals filtered at 10 s and gradually decreasing the cut-off period down to 1.25 s. This multiscale algorithm quickly converges towards a model that is very close to the true model, in contrast to inversions involving short-period waveforms only, which always get trapped into a local minimum of the cost function.

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

1 INTRODUCTION

Thanks to the densification of both permanent and temporary seismic recording arrays, resolution in tomographic images of the continental lithosphere has been considerably refined in the last decade. For example, the lateral resolution in the regional tomographic model of the Pyrenees obtained from the use of both PYROPE and IBERARRAY dense temporary deployments, quantified from the analysis of the resolution matrix, is about 25 km (Chevrot *et al.* 2014). While such a resolution is sufficient to map major structural units, it is insufficient to provide sharp images of lithospheric architecture, mainly because of strong vertical smearing of the velocity anomalies. Indeed, that same study found that vertical resolution is of the order of 80–100 km at lithospheric depths, and that it quickly deteriorates at greater depth. Such poor vertical resolution in regional tomography stems from the subvertical incidence of teleseismic body waves and from the fact that the sensitivity of traveltimes to seismic velocities is broadly distributed along and around the geometrical ray. A direct consequence of this poor vertical resolution is that crustal structures can also significantly contaminate tomographic models down to about 200 km depth (Chevrot *et al.* 2014). These fundamental limitations of traveltime tomography are unavoidable, even if finite-frequency effects are taken into account.

Alternative imaging approaches that make a more thorough use of the information contained in seismic records have been introduced. For example, ray-based migration techniques have been developed to exploit converted and multiply reflected phases that arrive in the coda of teleseismic *P* waves in order to map seismic discontinuities (Bostock & Rondenay 1999; Bostock *et al.* 2001; Rondenay 2009). These approaches led to improved images of interfaces in subduction environments (e.g. Bostock 1998; Kawakatsu & Watada 2008; Rondenay *et al.* 2008; Audet *et al.* 2009), providing crucial constraints on the dehydration of subducted materials and water transportation in the mantle. These images are very efficient at mapping small-scale structural variations of the subsurface, such as velocity discontinuities, but in contrast to classical traveltime tomography they lack the sensitivity to longer-wavelength structures. In order to obtain tomographic images of the continental lithosphere, prone to detailed geological interpretation, it will thus be necessary to turn to methods that can constrain the full heterogeneity spectrum, from a few tens of kilometres down to a few kilometres. Such a

[*] Now at: Géoazur, CNRS UMR 7329, Nice, France.

breakthrough implies moving towards full waveform inversion (see, e.g. Liu & Gu 2012, for a more detailed discussion), which has a theoretical resolution limit of the order of half of the shortest wavelength in the seismic wavefield (Virieux & Operto 2009). Imaging the lithosphere with a resolution of the order of a few kilometres will thus require to exploit records of teleseismic body waves in the period range 1–10 s.

Nowadays, efficient numerical methods to simulate the propagation of elastic waves based on full waveform modelling in a 3-D Earth model are available, for instance the finite-difference method (see, e.g. Virieux & Operto 2009, for a review), the spectral-element method (SEM; e.g. Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999; Vai *et al.* 1999; Komatitsch & Tromp 2002) or a standard finite-element method (e.g. Kallivokas *et al.* 2013), but they still suffer from a high computational cost in the context of solving inverse problems, which currently limits their application to periods larger than a few seconds in the case of the full Earth, even on the largest supercomputers (Tsuboi *et al.* 2003; Peter *et al.* 2011). Another important aspect of waveform inversion is the non-linear dependence of the waveforms on the velocity model. Waveform inversion problems are thus solved based on iterative methods, which requires solving 3-D wave propagation problems a very large number of times, making such imaging approaches even more prohibitive. Various simplifications of the problem have been proposed in order to reduce the computing power required to invert short-period teleseismic *P* waveforms. Apart from simplifying the modelling by using asymptotic approaches (e.g. Bostock *et al.* 2001) or considering 2-D geometries only (e.g. Pageot *et al.* 2013; Nissen-Meyer *et al.* 2014; Tong *et al.* 2014a), a possibility is to drastically reduce the size of the 3-D computational domain (e.g. Monteiller *et al.* 2013; Tong *et al.* 2014b). The forward problem then reduces to injecting an incident wavefield, produced by a distant teleseismic source, inside a regional 3-D domain. In other words, solving the forward problem involves a hybrid method in which modelling of short-period teleseismic waves is performed in two steps. The first step involves propagating the seismic waves from the source to the regional domain, and the second involves simulating the wavefield in the regional domain located below the array of receivers. These hybrid methods have received considerable attention over the last decade, and several approaches have been used to model wave propagation inside the regional domain: the SEM with the incidence of plane waves or a normal-mode summation, direct solution method (DSM) or frequency–wavenumber (F-K) wavefield (Vai *et al.* 1999; Capdeville *et al.* 2003a,b; Chevrot *et al.* 2004; Godinho *et al.* 2009; Monteiller *et al.* 2013; Tong *et al.* 2014a,b), the finite-difference method (Chen *et al.* 2005; Zhao *et al.* 2008; Roecker *et al.* 2010) or the frequency-domain discontinuous Galerkin method (Pageot *et al.* 2013). This is also true in other fields involving acoustic waves, for instance ultrasonic non-destructive testing (Gengembre *et al.* 2004). Many of these articles considered an incoming plane wave injected into a Cartesian 2-D domain. While reducing the dimension of the problem certainly allows for a drastic reduction of the computational cost, a 2-D geometry is of limited interest in practice because earthquakes almost never occur along the azimuth of acquisition profiles, which are also seldom rectilinear. In addition, as pointed out by Rondenay (2009), assuming a planar teleseismic wavefront is safe if the size of the domain is not too large (say smaller than 100 km), but this limits its application to rather local studies of the shallow Earth. Applying full waveform inversion to teleseismic wave records at the regional scale thus requires improving the numerical solvers to address these shortcomings.

In Monteiller *et al.* (2013), we introduced the first high-frequency 3-D hybrid modelling method by matching an incident wavefield computed in a spherically symmetric Earth model based on the DSM (Geller & Ohminato 1994; Geller & Takeuchi 1995; Takeuchi *et al.* 1996) with a regional wavefield computed based on the SEM (Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999; Vai *et al.* 1999) using the SPECFEM3D open-source software package. DSM can provide accurate synthetic seismograms up to frequencies as high as 2 Hz (Kawai *et al.* 2006) but at a cost that is significantly larger than that of the F-K method used for instance in Tong *et al.* (2014a,b). However, computing the incident wavefield in a spherical Earth based on DSM allows us to consider larger and deeper regional domains, which more than compensates for the additional computational cost compared to FK. Another motivation for doing so is to later be able to consider a regional domain that can be located anywhere inside the Earth and not only at its surface, as envisioned by Masson *et al.* (2014). Let us mention that other options such as a calculation of the Green's function of the Earth by Minor Integration (GEMINI; Friederich & Dalkolmo 1995) or a combination of 2-D axisymmetric numerical simulations for elementary sources (Nissen-Meyer *et al.* 2014) could be used instead of DSM for the incident field calculation, but normal-mode summation could not because it cannot accurately go below seismic periods of 5 s or so (e.g. Komatitsch *et al.* 2010). Eventually, 3-D computation of the incident wavefield with SPECFEM will also become a viable alternative. In that case, our hybrid approach would still present the advantage of drastically reducing the cost of full waveform inversion inside the regional 3-D domain.

It is important to mention that by definition of our one-way coupling strategy, scattered waves generated by 3-D heterogeneities superimposed to the 1-D background model cannot be further modelled by this hybrid method when they travel out of the SEM domain and are reflected back into that SEM domain off 1-D interfaces located inside the DSM domain. This is by construction the main limitation of the approach that we have chosen. However, the contributions of such scattered waves is of second order compared to the incident wavefield, and in most cases can safely be neglected. For a more complete but also more involved, fully-coupled strategy one can refer for instance to Capdeville *et al.* (2003a,b). It is also worth mentioning that such code-coupling or method-coupling strategies are successfully used in other fields (see, e.g. Formaggia *et al.* 2001; Han *et al.* 2014; Mariotti *et al.* 2015).

In this paper, we therefore present a 3-D full waveform inversion method that relies on such a hybrid method based on the DSM/SEM hybrid technique of Monteiller *et al.* (2013). We give a brief overview of that hybrid method in Section 2. We then recall the waveform inversion problem and the computation of the gradient kernel based on the adjoint method (Chavent 1974; Lailly 1983; Tarantola 1984; Tromp *et al.* 2005; Fichtner *et al.* 2006; Plessix 2006; Tromp *et al.* 2008; Virieux & Operto 2009; Fichtner 2010) in Section 3. Different iterative algorithms can be used to solve nonlinear waveform inversion problems, from the simple gradient method to the far more efficient but more involved Gauss–Newton method. We discuss them in Section 4 in which we also recall the principles of the Limited-Memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, which allows for a practical implementation of the quasi-Newton algorithm without the need to compute and store the inverse Hessian kernel. Finally, we apply our full waveform inversion method to a synthetic test chequerboard case in Section 5 and a sharp Moho jump in Section 6.

2 FORWARD PROBLEM: THE DSM/SEM HYBRID METHOD

To simulate the propagation of teleseismic waves at the regional scale, we use the hybrid DSM/SEM method introduced in Monteiller *et al.* (2013). Here we will only give a brief overview of the method and refer the reader to that paper for further details. The main advantage of the hybrid method is to restrict the costly 3-D computations to a regional domain of limited size, which is a crucial point in terms of calculation cost for iterative waveform inversion. Each of the DSM and SEM methods provides very accurate solutions of the wave equation in its own domain, and the problem is thus to match the global and regional wavefields on the boundaries of the regional mesh.

DSM is a Galerkin method that solves the weak form of the equation of motion in the frequency domain (e.g. Geller & Ohminato 1994; Geller & Takeuchi 1995). By carefully tuning the vertical grid spacing, maximum angular order and cut-off depth, the calculations can be made efficient while keeping high accuracy, even at frequencies as high as 2 Hz (Kawai *et al.* 2006). This method is thus well suited to model full short-period teleseismic wavefields. In the 3-D regional domain, we use the SEM (e.g. Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999; Vai *et al.* 1999; Komatitsch *et al.* 2005; Tromp *et al.* 2008), which is a highly accurate continuous Galerkin technique to model seismic wave propagation in elastic or anelastic (viscoelastic) media. The SEM is based upon the weak form of the seismic wave equation solved in the time domain. Because it uses high-degree polynomial basis functions, can handle very distorted meshes and does not necessitate interpolation of material properties, it is highly accurate and allows one to include all the complexity that may affect the seismic wavefields: topography of the free surface and of internal discontinuities, anelasticity, anisotropy and lateral variations of elastic parameters and density. In Monteiller *et al.* (2013), for instance, we showed that topography produces very significant body-to-surface wave conversions, which are predominant in the coda of teleseismic *P* waves. Using the DSM code, we compute the tractions and velocities produced by each source at all the SEM grid points located on the edges of the regional SEM mesh and store them to disk. We then read them back at the beginning of each SEM simulation. In the regional mesh, it is not necessary to use a discontinuous (i.e. geometrically non-conforming) mesh and thus a discontinuous Galerkin formulation because material property contrasts are not drastic, therefore, resorting to a continuous Galerkin formulation is sufficient.

3 FULL WAVEFORM INVERSION

Full waveform inversion means that one considers the observed seismograms (possibly filtered) as the basic observables that one wants to fit. One thus searches for the model that minimizes the mean squared difference between observed and synthetic seismograms. In other words, the goal is to find a structural model that can explain a larger portion of seismological records, and not simply the phase of a few seismic arrivals.

3.1 Principle

We want to minimize the classical waveform misfit function:

$$\chi(\mathbf{m}) = \sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T \frac{1}{2}\|\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)-\mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)\|^2\,\mathrm{d}t. \quad (1)$$

This functional quantifies the L^2 difference between the observed waveforms $\mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)$ at receivers $\mathbf{x}_r$, $r=1,\ldots,M$ produced by sources at $\mathbf{x}_s$, $s=1,\ldots,N$, and the corresponding synthetic seismograms $\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)$ computed in model $\mathbf{m}$. While this misfit function is indeed classical, it is worth mentioning that in the case of noisy real data other norms could be used, since in the oil industry for instance it is known that the L^1 norm (Crase *et al.* 1990; Brossier *et al.* 2010), hybrid L^1-L^2 norms (Bube & Langan 1997), Hubert norm (Ha *et al.* 2009), Student-*t* distribution (Aravkin *et al.* 2011; Jeong *et al.* 2015), etc., can be more robust than the L^2 norm used here in the context of synthetic data with no noise. In the vicinity of $\mathbf{m}$, the misfit function can be expanded into a Taylor series:

$$\chi(\mathbf{m}+\delta\mathbf{m}) \approx \chi(\mathbf{m}) + \mathbf{g}(\mathbf{m})\cdot\delta\mathbf{m} + \delta\mathbf{m}\cdot\mathbf{H}(\mathbf{m})\cdot\delta\mathbf{m}\,, \quad (2)$$

where $\mathbf{g}(\mathbf{m})$ is the gradient of the waveform misfit function:

$$\mathbf{g}(\mathbf{m}) = \frac{\partial\chi(\mathbf{m})}{\partial\mathbf{m}}\,, \quad (3)$$

and $\mathbf{H}(\mathbf{m})$ the Hessian:

$$\mathbf{H}(\mathbf{m}) = \frac{\partial^2\chi(\mathbf{m})}{\partial\mathbf{m}^2}\,. \quad (4)$$

In the following, for simplicity the dependence of the gradient and Hessian on the model will be implicitly assumed and omitted in the notations. The nearest minimum of χ in eq. (2) with respect to the model perturbation $\delta\mathbf{m}$ is reached for

$$\delta\mathbf{m} = -\mathbf{H}^{-1}\cdot\mathbf{g}\,. \quad (5)$$

The local minimum of eq. (1) is thus given by perturbing the model in the direction of the gradient preconditioned by the inverse Hessian.

3.2 Computation of the gradient based on the adjoint method

A direct method to compute the gradient is to take the derivative of eq. (1) with respect to model parameters:

$$\frac{\partial\chi(\mathbf{m})}{\partial\mathbf{m}} = -\sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T \frac{\partial\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)}{\partial\mathbf{m}} \cdot[\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)-\mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)]\,\mathrm{d}t\,. \quad (6)$$

This equation can be reformulated as the matrix–vector product:

$$\mathbf{g} = -\mathbf{J}^*\cdot\delta\mathbf{d}\,, \quad (7)$$

where $\mathbf{J}^*$ is the adjoint of the Jacobian matrix of the forward problem that contains the Fréchet derivatives of the data with respect to model parameters and $\delta\mathbf{d}$ is the vector that contains the data residuals. The determination of $\mathbf{J}$ would require computing the Fréchet derivatives for each time step in the time window considered and for all the source-station pairs, which is completely prohibitive on current supercomputers (let us note that this situation may change one day). However, it is possible to obtain this gradient without computing the Jacobian matrix explicitly. The approach to determine the gradient without computing the Fréchet derivatives was introduced in nonlinear optimization by Chavent (1974) working with J. L. Lions, and later applied to seismic exploration problems by Bamberger *et al.* (1977), Bamberger *et al.* (1982), Lailly (1983) and Tarantola (1984). The idea is to resort to the adjoint state, which corresponds to the wavefield emitted and back-propagated from the receivers (e.g. Tromp *et al.* 2005; Fichtner *et al.* 2006; Plessix 2006; Tromp *et al.* 2008).

Let us give an outline of the theory to compute the gradient based on the adjoint method and refer the reader to, for example, Tromp *et al.* (2005) and Tromp *et al.* (2008) for further details. The perturbation of the misfit function can be expressed as:

$$\delta\chi(\mathbf{m}) = \sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T [\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t) - \mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)]\cdot\delta\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t)\,\mathrm{d}t, \tag{8}$$

where $\delta\mathbf{u}$ is the perturbation of displacement given by the first-order Born approximation (e.g. Hudson 1977):

$$\delta\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t) = -\int_0^t\int_V \big[\delta\rho(\mathbf{x})\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t')\cdot\partial_{t'}^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t') + \nabla\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t'):\delta\mathbf{c}(\mathbf{x}):\nabla\mathbf{u}(\mathbf{x};t')\big]\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t'. \tag{9}$$

In this expression, $\mathbf{G}$ is the Green's tensor, $\delta\rho$ the perturbation of density, δc the perturbation of the fourth-order elasticity tensor, and a colon denotes a double tensor contraction operation. Inserting eq. (9) into eq. (8) we obtain

$$\delta\chi(\mathbf{m}) = -\sum_{s=1}^{N}\sum_{r=1}^{M}\int_0^T [\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;t) - \mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;t)] \times\int_0^t\int_V \big[\delta\rho(\mathbf{x})\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t')\cdot\partial_{t'}^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t') + \nabla\mathbf{G}(\mathbf{x}_r,\mathbf{x};t-t'):\delta\mathbf{c}(\mathbf{x}):\nabla\mathbf{u}(\mathbf{x},t')\big]\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t'\,\mathrm{d}t\,. \tag{10}$$

Defining the waveform adjoint source for each source $\mathbf{x}_s$

$$\mathbf{f}^\dagger(\mathbf{x},\mathbf{x}_s;t) = \sum_{r=1}^{M}[\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;T-t) - \mathbf{d}(\mathbf{x}_r,\mathbf{x}_s;T-t)]\,\delta(\mathbf{x}-\mathbf{x}_r)\,, \tag{11}$$

and the corresponding adjoint wavefield

$$\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_s;t) = \int_0^{t'}\int_V \mathbf{G}(\mathbf{x},\mathbf{x}';t'-t)\cdot\mathbf{f}^\dagger(\mathbf{x}',\mathbf{x}_s;t)\,\mathrm{d}^3\mathbf{x}'\,\mathrm{d}t\,, \tag{12}$$

the perturbation of the misfit function may be expressed as:

$$\delta\chi(\mathbf{m}) = -\sum_{s=1}^{N}\int_V\int_0^T \big[\delta\rho\,\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_s;T-t)\cdot\partial_t^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t) + \nabla\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_s;T-t):\delta\mathbf{c}:\nabla\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\big]\,\mathrm{d}^3\mathbf{x}\,\mathrm{d}t\,. \tag{13}$$

At this point, we make some assumptions on the nature of the elasticity tensor. A general fourth-order elasticity tensor is described by 21 elastic parameters, a very large number that makes its complete characterization way beyond the reach of any tomographic approach. For the time being, let us thus consider isotropic elasticity tensors, described by the two Lamé parameters λ and μ:

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\,. \tag{14}$$

In this case, eq. (13) can be written as:

$$\delta\chi(\mathbf{m}) = -\sum_{s=1}^{N}\int_V \big[K_\rho(\mathbf{x},\mathbf{x}_s)\delta\ln\rho(\mathbf{x}) + K_\lambda(\mathbf{x},\mathbf{x}_s)\delta\ln\lambda(\mathbf{x}) + K_\mu(\mathbf{x},\mathbf{x}_s)\delta\ln\mu(\mathbf{x})\big]\,\mathrm{d}^3\mathbf{x}\,, \tag{15}$$

where ln() is the natural logarithm and where the Fréchet derivatives with respect to the density and Lamé parameters are given by:

$$K_\rho(\mathbf{x},\mathbf{x}_s) = -\int_0^T \rho(\mathbf{x})\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_s;T-t)\cdot\partial_t^2\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\,\mathrm{d}t, \tag{16}$$

$$K_\lambda(\mathbf{x},\mathbf{x}_s) = -\int_0^T \lambda(\mathbf{x})\nabla\cdot\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_s;T-t)\nabla\cdot\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\,\mathrm{d}t, \tag{17}$$

$$K_\mu(\mathbf{x},\mathbf{x}_s) = -2\int_0^T \mu(\mathbf{x})(\mathbf{x})\nabla\mathbf{u}^\dagger(\mathbf{x},\mathbf{x}_s;T-t):\nabla\mathbf{u}(\mathbf{x},\mathbf{x}_s;t)\,\mathrm{d}t\,. \tag{18}$$

Since the propagation of seismic waves mainly depends on compressional wave speed α and shear wave speed β, but also because these seismic velocities are easier to interpret, tomographic models are usually described based on these two parameters. With this new parametrization, the perturbation of the misfit function may be written as:

$$\delta\chi(\mathbf{m}) = -\sum_{s=1}^{N}\int_V \big[K'_\rho(\mathbf{x},\mathbf{x}_s)\delta\ln\rho(\mathbf{x}) + K'_\alpha(\mathbf{x},\mathbf{x}_s)\delta\ln\alpha(\mathbf{x}) + K'_\beta(\mathbf{x},\mathbf{x}_s)\delta\ln\beta(\mathbf{x})\big]\,\mathrm{d}^3\mathbf{x}\,, \tag{19}$$

where

$$K'_\rho(\mathbf{x},\mathbf{x}_s) = K_\rho(\mathbf{x},\mathbf{x}_s) + K_\lambda(\mathbf{x},\mathbf{x}_s) + K_\mu(\mathbf{x},\mathbf{x}_s) \tag{20}$$

$$K'_\alpha(\mathbf{x},\mathbf{x}_s) = 2\left(\frac{\lambda+2\mu}{\lambda}\right)K_\lambda(\mathbf{x},\mathbf{x}_s) \tag{21}$$

$$K'_\beta(\mathbf{x},\mathbf{x}_s) = 2K_\mu - \frac{4\mu}{\lambda}K_\lambda(\mathbf{x},\mathbf{x}_s)\,. \tag{22}$$

As can be seen from these expressions, the principle of the adjoint-state method is to correlate two wavefields: the direct (i.e. forward) field that propagates from the source to the receivers, and the adjoint field that propagates from all the receivers backward in time. The same approach can be followed for any type of seismic observable (phase, amplitude, envelope, time series, etc.), provided the appropriate adjoint source is used (Tromp *et al.* 2005, 2008). For example, for the cross-correlated traveltime of a seismic phase, the adjoint source is defined as the velocity of that synthetic phase weighted by the traveltime residual.

Computing the gradient based on the adjoint-state method requires performing two simulations per source (forward and adjoint fields) regardless of the type of observable. However, to define the adjoint field one must know the adjoint source, and that source is computed from the results of the forward simulation. One must therefore perform the forward simulation before the adjoint simulation. A straightforward solution for time-domain methods would be to store the whole forward field to disk at each time step during the forward run and then read it back during the adjoint simulation to calculate the interaction of these two fields. In 2-D this is feasible but in the 3-D case for very short seismic periods and without lossy compression, downsampling or a large amount of disk or memory checkpointing (e.g. Fichtner *et al.* 2009; Rubio Dalmau *et al.* 2014; Cyr *et al.* 2015) the amount of disk storage required would currently be too large. However, let us note again that this situation may change in the future. In the mean time, a standard possible solution is to perform three simulations per source (Tromp *et al.* 2008; Peter *et al.* 2011), that is, perform the forward calculation twice, once to compute the adjoint sources and once again at the same time as the adjoint simulation to correlate the two fields and sum their interaction on the fly over all the time steps. Doing so for an elastic Earth, one only needs a small amount of disk storage to store the last time step of the forward run, which is then used as an

initial condition to redo the forward run backwards, as well as the field on the outer edges of the mesh for each time steps in order to be able to undo the absorbing boundary conditions.

3.3 The SEM and tomographic (inversion) grids

Because it is more convenient to use a topologically regular tomographic grid than a non-structured finite-element mesh, we distinguish the grid used to solve the wave propagation problem (the SEM grid) and the tomographic grid, which we choose to be regular. The former is subdivided into a number of non-overlapping hexahedral elements controlled by their 27 corners, mid-edge, mid-face and central control points. Inside each spectral element, the model and the wave field are defined on a basis of Lagrange polynomials of degree $N = 4$ (Komatitsch & Tromp 1999) and thus based on $N + 1 = 5$ interpolation points along each of the three spatial directions. Each element thus contains $(N + 1)^3 = 125$ points, at which the elastic parameters as well as the wave field are sampled. The mesh is a 'chunk' of the Earth, that is, a portion of the so-called 'cubed sphere' (Sadourny 1972; Ronchi *et al.* 1996; Chaljub *et al.* 2003) that is used to mesh part of the Earth based on hexahedra. We use the conventions given in Chevrot *et al.* (2012) to map the geographical coordinates to the cubed sphere coordinates and conversely. The chunk is defined by the position of the centre of its upper surface, referenced in the geographic coordinate system by its north–south and east–west extensions expressed in degrees, and by the azimuth of the X axis in the regional mesh.

To project the model defined on the SEM mesh onto the topologically regular tomographic grid, we search for the SEM elements that contain the vertices of each given tomographic cell. We then use the products of Lagrange polynomials (i.e. the SEM interpolation matrix) to interpolate the model parameters at the eight corners of this tomographic cell and assign to it the average of the values at these eight corners. We similarly compute the gradient from the correlation of the forward and adjoint wavefields, and for accuracy reasons we compute the volume integrals needed based on the SEM Gauss–Lobatto–Legendre quadrature formula. This requires interpolating the values of the correlations between the forward and adjoint wavefields not only at the vertices of each tomographic cell but also at all the collocation points located inside that cell. Once the model and the gradients for all the sources are projected onto the tomographic grid, we use the summed gradient to update the current tomographic model, which we then project back to the SEM mesh before proceeding to the next iteration. For efficiency, we once and for all define a mapping function that matches the SEM elements and their collocation point identifiers with the identifiers of the cell in the tomographic grid. Based on this mapping, the inverse projection becomes straightforward.

3.4 Regularization

To stabilize the inversion it is necessary to regularize the problem. To do so, we add a classical penalty condition on the norm of the Laplacian of the model to the waveform misfit function and define a new misfit function:

$$\chi'(\mathbf{m}) = \chi(\mathbf{m}) + \frac{\eta}{2}\|\Delta m\|^2, \tag{23}$$

where η is a smoothing coefficient that weights the contribution of the norm of the Laplacian with respect to the waveform misfit. Denoting $L(\mathbf{m})$ the norm of the Laplacian:

$$L(\mathbf{m}) = \frac{1}{2}\|\Delta m\|^2\,, \tag{24}$$

the gradient of the new misfit function is then given by:

$$\nabla\chi'(\mathbf{m}) = \mathbf{g}(\mathbf{m}) + \eta\nabla L(\mathbf{m})\,. \tag{25}$$

We have recalled in the previous section how to compute $\mathbf{g}(\mathbf{m})$ based on the adjoint method. The Laplacian of the model can be computed using a finite-difference operator in the regular 3-D Cartesian inversion grid:

$$\begin{aligned}\Delta m_{(i,j,k)} = &-8m_{(i,j,k)} + m_{(i-1,j,k)} + m_{(i+1,j,k)} + m_{(i,j-1,k)}\\ &+m_{(i,j+1,k)} + m_{(i,j,k-1)} + m_{(i,j,k+1)}.\end{aligned} \tag{26}$$

This formula can easily be adapted to the case of different smoothing values along the horizontal and vertical directions:

$$\begin{aligned}\Delta m_{(i,j,k)} = &(-4l_h - 2l_v)m_{(i,j,k)} + l_h(m_{(i-1,j,k)} + m_{(i+1,j,k)}\\ &+ m_{(i,j-1,k)} + m_{(i,j+1,k)}) + l_v(m_{(i,j,k-1)} + m_{(i,j,k+1)}),\end{aligned} \tag{27}$$

where l_h and l_v are the smoothing coefficients along the horizontal and vertical directions, respectively. Let us denote Δ_d the discrete Laplacian operator in the inversion grid, which is described by a symmetric band-diagonal matrix. The norm of the discrete Laplacian can then be written as:

$$L(\mathbf{m}) = \frac{1}{2}\|\Delta_d\mathbf{m}\|^2 = \frac{1}{2}\langle\Delta_d\mathbf{m}|\Delta_d\mathbf{m}\rangle\,, \tag{28}$$

where $\langle\rangle$ denotes the scalar product. Using these definitions, the product

$$\nabla L(\mathbf{m}).\delta\mathbf{m} = \langle\Delta_d\mathbf{m}|\Delta_d\delta\mathbf{m}\rangle \tag{29}$$

can be written as:

$$\nabla L(\mathbf{m}).\delta\mathbf{m} = \langle\Delta_d^*\Delta_d\mathbf{m}|\delta\mathbf{m}\rangle = \langle\Delta_d\Delta_d\mathbf{m}|\delta\mathbf{m}\rangle\,, \tag{30}$$

since the discrete Laplacian is self-adjoint. From this last expression we thus see that the gradient of the penalty function $L(\mathbf{m})$ is simply obtained by applying the discrete Laplacian to the model vector $\mathbf{m}$ twice:

$$\nabla L(\mathbf{m}) = \Delta_d\Delta_d\mathbf{m}. \tag{31}$$

4 ITERATIVE OPTIMIZATION ALGORITHMS

Different classes of nonlinear optimization methods can be used to find the minimum of a cost function. Stochastic methods randomly explore the model space with no other information than the value of the cost function (e.g. Mosegaard & Tarantola 1995; Sambridge 1999). Their main advantage is that they converge towards the global minimum of the cost function, but at the price of a much larger computational cost. In practice, these methods are thus efficient only if the dimension of the parameter space to explore is small, which is not the case in 3-D full waveform inversion problems.

Steepest descent methods search for a perturbation of the model along the direction defined by the gradient of the misfit function. Their main two drawbacks are that the inversion may converge towards a local minimum and remain trapped there, and that it may converge very slowly. They are therefore not recommended for very large problems in which the cost to estimate the misfit function and the gradient is important.

The Gauss–Newton method is the most efficient technique in terms of convergence speed. In this method, the search direction is obtained by preconditioning the gradient by the inverse Hessian

(e.g. Pratt *et al.* 1998; Pratt 1999). However, this method requires to first compute and store the Hessian, which requires a huge amount of computation and storage space, and then to solve a very large system of linear equations. In the case of 3-D full waveform inversion, such an approach is currently completely prohibitive.

For all these reasons, we thus choose to use the L-BFGS method (Nocedal & Wright 2006) to handle the full waveform inversion problem. The main appeal of this method is that while it is affiliated to Gauss–Newton methods, it does not require computing and storing the Hessian (or its inverse) explicitly.

4.1 The L-BFGS method

Using iterative methods, it is possible to compute an estimate of the inverse Hessian based only on the knowledge of the gradient at the previous iterations, the quality of the approximation improving with the number of previous iterations used. The method generates a series of models that gradually converge towards a minimum of the misfit function (which may be local) and a series of matrices that converge towards the inverse Hessian.

The BFGS formula to compute $\mathbf{H}_k$, the approximate inverse Hessian at iteration k, is given by (Nocedal & Wright 2006):

$$\mathbf{H}_k^{-1} \simeq \mathbf{H}_{k-1}^{-1} - \frac{\mathbf{H}_{k-1}^{-1} \cdot \mathbf{s}_{k-1} \otimes \mathbf{s}_{k-1} \cdot \mathbf{H}_{k-1}^{-1}}{\mathbf{s}_{k-1} \cdot \mathbf{H}_{k-1} \cdot \mathbf{s}_{k-1}} + \frac{\mathbf{y}_{k-1} \otimes \mathbf{y}_{k-1}}{\mathbf{y}_{k-1} \cdot \mathbf{s}_{k-1}}, \qquad (32)$$

where $\otimes$ is the tensor product, $\mathbf{s}_k = \mathbf{m}_k - \mathbf{m}_{k-1}$ is the difference between the current model and the model at the previous iteration, and $\mathbf{y}_k = \nabla\chi'_k - \nabla\chi'_{k-1}$ is the gradient change. Using eq. (32) one can iteratively calculate an estimate of the inverse Hessian $\mathbf{H}^{-1}$ based on the knowledge of the approximation of the inverse Hessian at the previous iteration, the difference $\mathbf{s}$ between the current model and its value at the previous iteration, and the difference $\mathbf{y}$ between the current gradient and the gradient at the previous iteration. Compared to the gradient method, convergence of BFGS is much faster (for a mathematical proof see, e.g. Nocedal & Wright 2006) for the same numerical cost. Convergence is about a factor of four faster in the examples that we will show below. Compared to the classical Gauss–Newton method, BFGS is also easier to implement because it does not require to compute and store the inverse Hessian. The reader is referred to chapter 7 of Nocedal & Wright (2006) for a more detailed presentation of the BFGS algorithm.

To compute the search direction at iteration k

$$\mathbf{p}_k = -\mathbf{H}_k^{-1} \cdot \nabla\chi'_k, \qquad (33)$$

one only needs to perform a matrix–vector multiplication. However, in the case of large problems it is currently impossible to compute and store even the approximate inverse Hessian matrix. Since in eq. (33) one does not need to explicitly store it but only be able to compute its effect on a vector (the gradient), a modified method called the L-BFGS algorithm (for 'limited-memory BFGS') has been developed (Nocedal & Wright 2006) in order to compute the matrix–vector product in eq. (33) without having to store the inverse Hessian. The principle of L-BFGS is to use eq. (32) iteratively to compute the product of the inverse Hessian, using the gradient from the initial inverse Hessian and the history of models and gradients accumulated in the iterations of the algorithm. In this case, one only needs to store a set of models and gradients, which represents only a fraction of the storage space required to store the complete inverse Hessian. The number of previous models and gradients that are kept in memory is a parameter chosen by the user. An initial inverse Hessian $\mathbf{H}_0^{-1}$ also needs to be specified; we will present below a simple approach to define it.

4.2 Calculation of the step length

Once the descent direction $\mathbf{p}_k$ at iteration k has been obtained, it is necessary to determine the step length, or in other words to decide how far to move along that direction. This problem can be formulated as finding the step s that minimizes

$$\phi(s) = \chi'(\mathbf{m}_k + s\mathbf{p}_k). \qquad (34)$$

In practice, determining that optimal step precisely may require to test a large number of step lengths, which can thus be very expensive. However, one should keep in mind that $\chi'(\mathbf{m})$ rather than $\phi(s)$ is the quantity that we need to minimize. It is thus sufficient to find an approximate step at minimal cost that honours certain conditions in order to make the optimization method converge. In practice, the step length variations between two iterations must be sufficiently large so that the algorithm requires a moderate number of iterations to converge, and sufficiently small to avoid the divergence of the algorithm. A good compromise is to use the so-called Wolfe conditions to select the step length (Nocedal & Wright 2006). These rules test if the current step provides a sufficient decrease of both the cost function and the gradient. Introducing parameters $0 < c_1 < c_2 < 1$, and $\phi'(s)$ the derivative of ϕ with respect to s, the step length is kept if:

$$\phi(s) \leq \phi(0) + c_1 s\phi'(0) \quad \text{and} \quad |\phi'(s)| \leq c_2|\phi'(0)|. \qquad (35)$$

If these two conditions are not satisfied, a new step is tested. If

$$\phi(s) > \phi(0) + c_1 s\phi'(0) \qquad (36)$$

the step is too long, and one then tests a smaller step length. On the other hand, if:

$$\phi(s) \leq \phi(0) + c_1 s\phi'(0) \quad \text{and} \quad |\phi'(s)| > c_2|\phi'(0)|, \qquad (37)$$

the step is too short, and one then tests a longer step. When it is no longer possible to find a step that satisfies these relations, convergence is reached and one then stops the algorithm. Tuning parameters c_1 and c_2 makes the selection rules more or less restrictive in terms of accepting the step length. For example, if c_1 is chosen close to 0, it is easier to honour the first inequality. In our implementation, we select $c_1 = 0.1$ and $c_2 = 0.9$, the standard values recommended by Nocedal & Wright (2006).

5 TEST OF THE METHOD ON A CHEQUERBOARD MODEL

In order to test the inversion method described in the previous sections, we consider a simple ideal (in terms of ray coverage) synthetic chequerboard test. The chequerboard model is composed of 20 km cube-shaped velocity anomalies of ± 0.5 km s^{-1} embedded in a homogeneous model in which $V_p = 8$ km s^{-1}, $V_s = 4.5$ km s^{-1} and $\rho = 3000$ kg m^{-3}. We perform the computations in a 100 km $\times$ 100 km $\times$ 60 km regular Cartesian grid discretized with 2 km cubic elements. The model is illuminated by 20 different plane-wave sources coming from the south, east, north and west directions and with incidence angles of 20, 30, 45, 60 and 80. These plane waves are recorded by an array of 2116 stations uniformly distributed at the free surface with an inter-station spacing of 2 km. The source wavelet is a Gaussian with a dominant period of 1.25 s, and we invert the vertical component of P waveforms in time windows that start 5 s before the P-wave arrival and end 10 s after. The starting model in the iterative algorithm is the homogeneous background model. In the beginning of the inversion, we

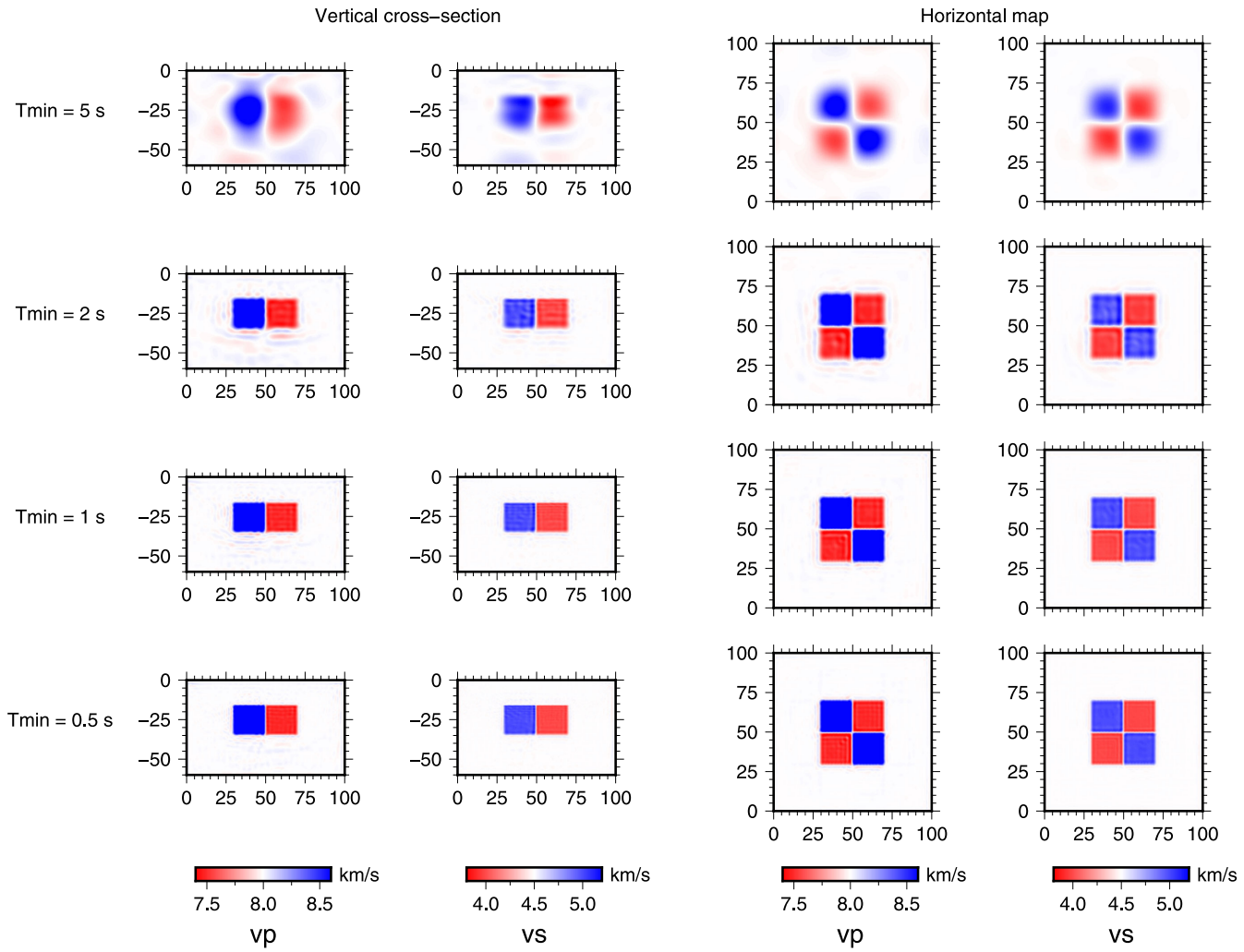

Figure 1. Results of full waveform inversion for a chequerboard test. The plots show vertical cross-sections (left) and map views (right) of the V_p and V_s models obtained with a hierarchical algorithm starting with data low-pass filtered at 5 s (top) and then gradually decreasing the corner period of the filter to 0.5 s (bottom). We used 20 plane waves with 5 coming from the north, 5 from the south, 5 from the east and 5 from the west, with incidence angles of 20, 30, 45 and 80° for each of the four waves in each case. The misfit function decreases from 13.16 to 0.14 in 18 iterations for the data filtered at 5 s, from 6.45 to 0.17 in 62 iterations for the data filtered at 2 s, from 1.99 to 0.54 in 76 iterations for the data filtered at 1 s and from 1.8 to 1.05 in 63 iterations for the data filtered at 0.5 s.

invert the waveforms low-pass filtered at 0.2 Hz and gradually increase the corner frequency of the filter up to 2 Hz. Moving from low to high frequencies helps the inversion to converge towards a model close to the true model (Fig. 1) without getting trapped in a secondary minimum of the misfit function (e.g. Pratt 1999; Pageot *et al.* 2013).

6 A MORE REALISTIC SYNTHETIC CASE

To illustrate the feasibility and potential of our approach, we now apply our waveform inversion method to a synthetic model with a crustal root. While this model represents a simple analogue of the crust beneath a narrow mountain range, at this point we did not introduce a topography profile as in Monteiller *et al.* (2013). It is worth mentioning that our approach only requires computational resources that are nowadays accessible: all the simulations are run on a moderate-size cluster with a number of processor cores used typically smaller than 500.

6.1 The computational grid and the synthetic experiment settings

The computational grid is a chunk of the sphere with a size of 2° in the north–south direction, 1° in the east–west direction and 75 km in the vertical direction (Fig. 2). This chunk is discretized using 2.5 km × 2.5 km × 2.5 km hexahedral elements. The minimum shear wave velocity is 3.36 km s^{-1} and the calculations are performed for a minimum seismic period of 1 s. We consider four different explosive sources located at 600 km depth. Sources 1 and 2 are located at an epicentral distance of 60° with respect to the centre of the chunk, while sources 3 and 4 are located at 50°. The backazimuths for the four sources are respectively 0, 180, 20 and 60°.

We construct a lithospheric model (Fig. 3) starting from the IASP91 reference Earth model (Kennett & Engdahl 1991). In this model, the crust consists of two homogeneous layers separated by an interface located at 20 km depth. The *P*- and *S*-wave velocities are respectively 5.8 and 3.36 km s^{-1} in the top layer, and 6.5 and 3.75 km s^{-1} in the bottom layer. The Moho is located at 35 km

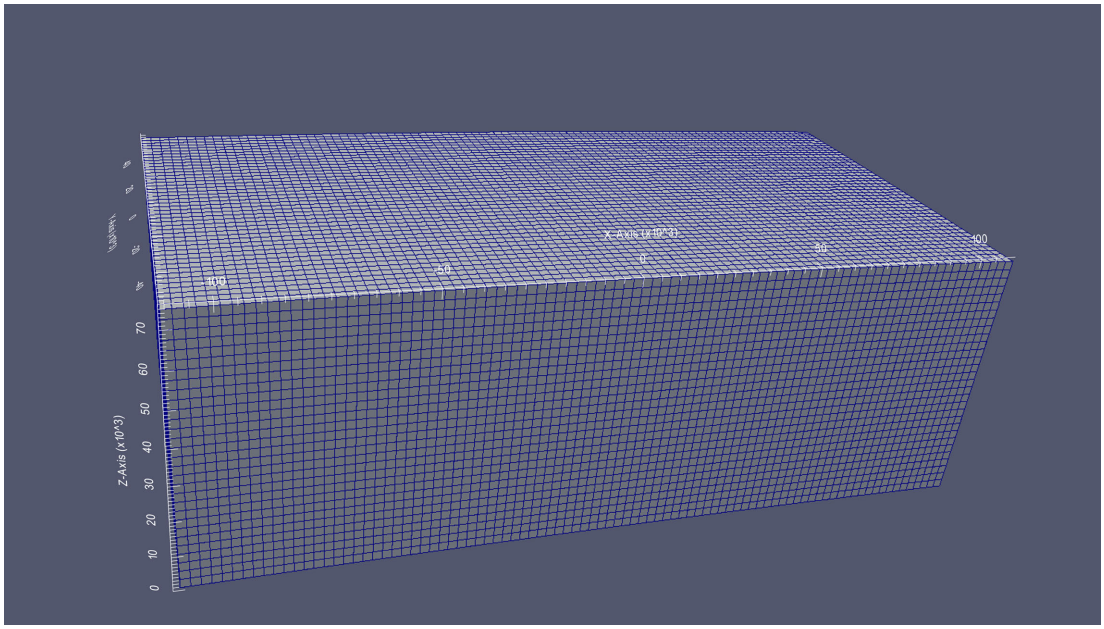

Figure 2. The spectral element mesh that we use in the SPECFEM3D software package. The X-axis is oriented along the south–north direction, the Y-axis along the west–east direction and the Z-axis from bottom to top. A linear profile of 40 recording stations is placed on the free surface along the X-axis at $y = 0$. This mesh is composed of 121 500 spectral elements having a total of 8 910 835 degrees of freedom. The minimum distance between two mesh points is 431 m and the maximum distance is 1406 m.

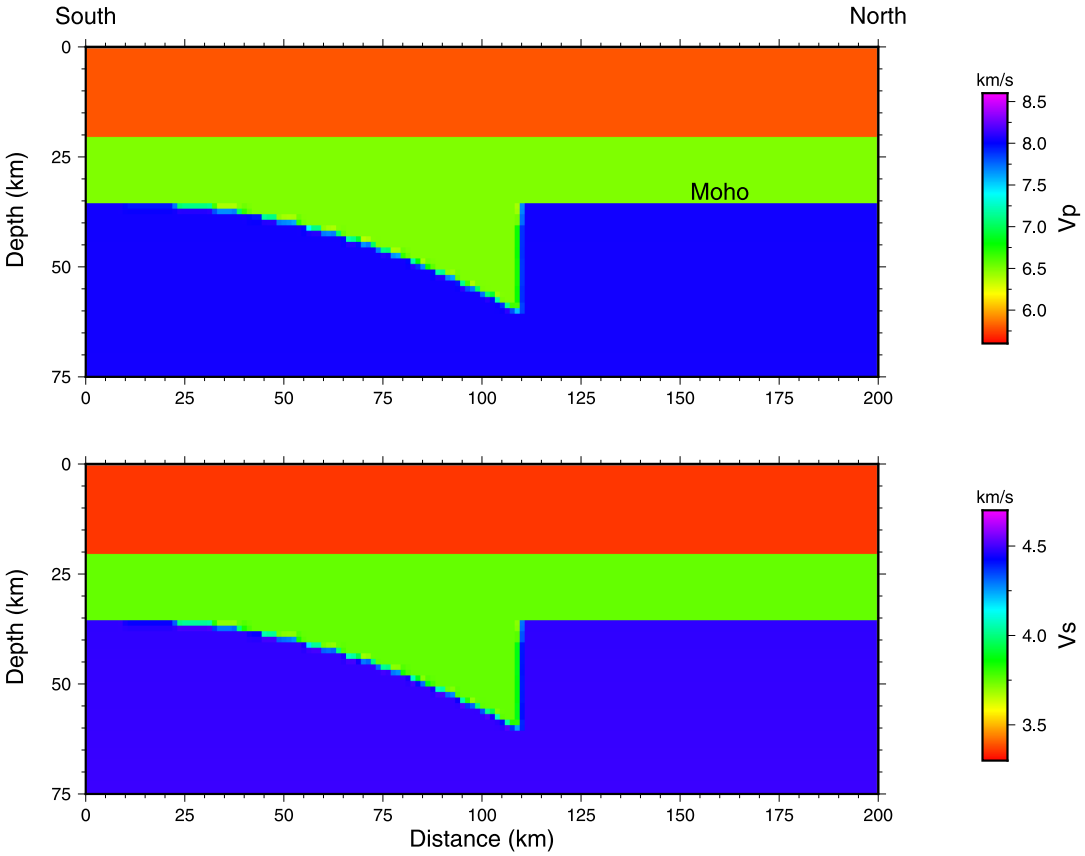

Figure 3. Synthetic model of P (top) and S (bottom) velocities.

depth. We modify this 1-D model by changing the geometry of the Moho. To do so, we introduce a 25 km Moho jump at $X = 11$ km, the X-axis following a meridian. To the south of this Moho jump, the Moho has a polynomial variation of degree 2, from 35 km depth at the southern edge of the grid to a maximum depth of 60 km. The negative velocity anomaly related to this lower crust thickening is about 20 per cent for the P-wave speed and 15 per cent for the S-wave speed. We place 40 stations along a north–south profile

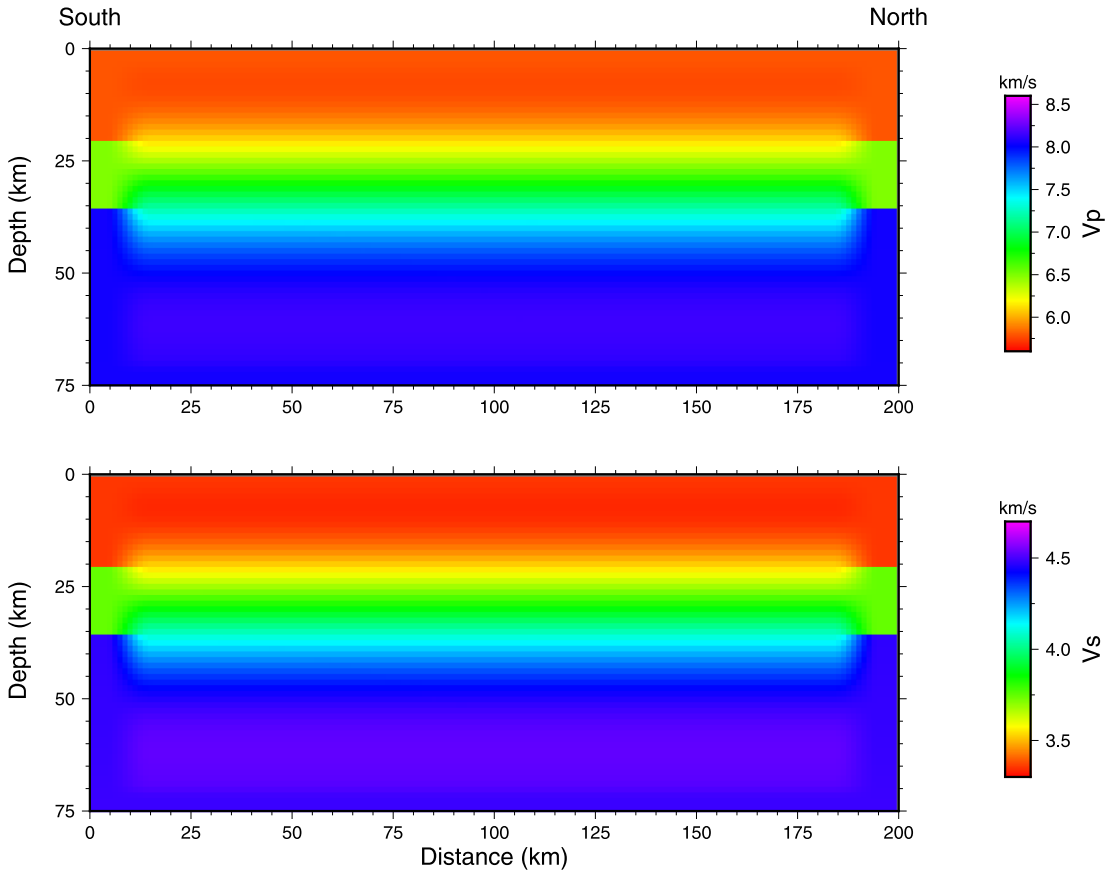

Figure 4. The smooth 1-D *P* and *S* velocity models that we use as starting models for waveform inversion.

located in the middle of the mesh chunk, with a station spacing of 5.5 km.

The choice of the starting model is important in nonlinear iterative methods because it can significantly influence the convergence speed as well as the results of the inversion. Since we use the IASP91 reference Earth model to compute the incident teleseismic wavefields based on DSM, it would seem natural to also use this model as the initial model. However, this choice would be far from optimal because the gradients at the first iterations would be strongly contaminated by the seismic signature of the interfaces in the starting model. Unless one has a good *a priori* knowledge of the geometry of the interfaces, it is thus better to start from a smooth model having no interface. This smooth initial model must preserve the arrival times of the direct waves computed in the reference Earth model. This constraint is met by searching for the smooth 1-D polynomial model that conserves the traveltime of vertically propagating *P* and *S* waves in the IASP91 model. We match this smooth model, shown in Fig. 4, to the IASP91 model on the edges of the SEM grid using a cosine taper function with a width of 15 km.

Fig. 5 compares the vertical and radial seismograms computed in the 1-D smooth starting model of Fig. 4 to those computed in the 3-D model of Fig. 3. The seismograms are filtered in different frequency bands to show the effect of the filter on the seismic waveforms that are used at different stages of the hierarchical iterative waveform inversion algorithm. The most salient feature observed on the vertical component seismograms is that the direct *P* wave (the first arrival in the record section) recorded at stations 12 to 20, located above the strong low-velocity anomaly related to the thickened crust, is both delayed and amplified. This effect is mostly visible at short period. The complex Moho topography has a clear signature on the traveltime of the PpPmP phase (*P* wave reflected beneath the free surface and reflected back by the Moho), which arrives about 10 s after the *P* wave at the stations located at the two extremities of the profile. In particular, the curvature of the Moho in the southern domain can be directly observed through its influence on the onset of the PpPmP phase. Other notable effects on the vertical record section are the diffracted waves that are scattered at the extremities of the vertical segment of the crust-mantle interface. Again, these effects are mostly visible at short period. At long period, crustal multiples and diffracted waves interfere, producing smooth amplitude modulations of the direct *P* wave and its coda.

6.2 Gradient of the cost function and pre-conditioning

In the first iteration, the gradients are computed in the smooth 1-D model (Fig. 4). Total gradients for V_p and V_s obtained by summing the gradients computed for the four individual sources are shown in Fig. 6. These gradients are computed by considering different time window lengths around the direct *P* wave and different filterings. The V_p and V_s gradients show very different behaviours. The V_p gradient is remarkably stable when the length of the time window is increased, meaning that it is dominated by the contribution of the direct *P* wave. Nevertheless, a closer look reveals that for window lengths extending beyond 20 s after the *P* wave, a sharp band of enhanced sensitivity is observed in the V_p gradient. This band, which is more clearly seen at short period, approximately follows the Moho geometry in the true model. This sensitivity to sharp velocity gradients is carried by the PpPmp phase, which arrives around

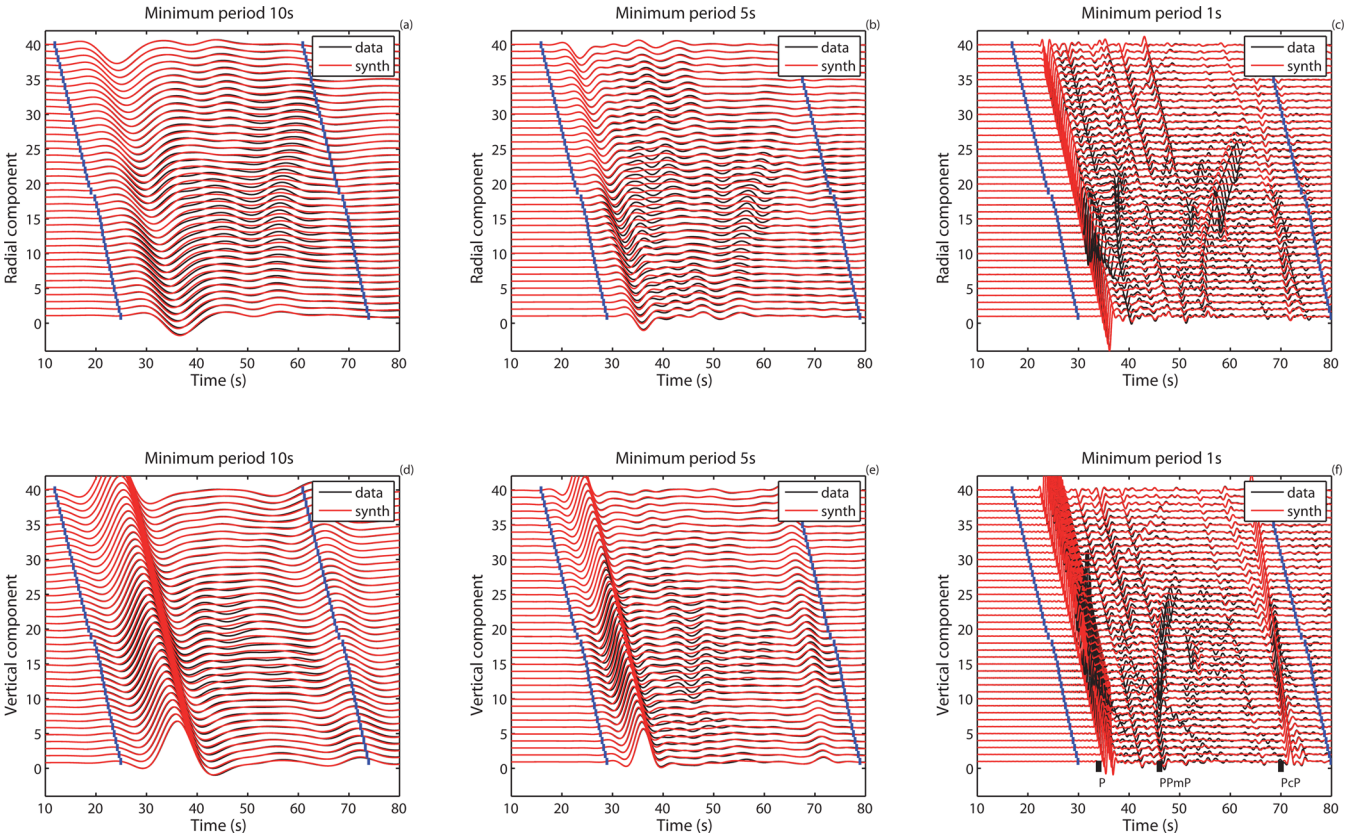

Figure 5. Radial and vertical synthetic seismograms for the first source (distance 60°, backazimuth 0°) computed in the 3-D model of Fig. 3 (black lines) and in the smooth starting model of Fig. 4 (red lines), low-pass filtered at 10 s (a, d), 5 s (b, e) and 1 s (c, f). The blue lines show the time windows used in the inversions.

10 s after the P wave, as can be seen in Fig. 5. For periods longer than 5 s, the V_p gradient is dominated by the contribution of the low-velocity anomaly associated with the thick crustal root. Note that the geometry of this crustal root is better retrieved with longer time windows that include the reverberations on the Moho. Again, this is consistent with a V_p gradient dominated by the contribution of the direct P wave, that is, to long wavelength structures, the later reverberations being mainly sensitive to sharp velocity interfaces, that is, to short wavelength structures. At shorter period (1–2 s), the gradient exhibits a large central V-shape anomaly that takes its roots along the vertical edge of the crust-mantle boundary. It is probably produced by the contribution of waves diffracted on the vertical edge of the Moho, illuminated from the north or from the south by the different teleseismic sources.

In contrast, the V_s gradient shows strong sensitivity to seismic velocity jumps at the mid-crust and crust-mantle interfaces at all the periods, regardless of the time window length considered. This is not surprising, since PmS, the phase converted from P to S at the Moho, arrives 4–6 s after the P wave and is thus always present in the parts of the seismograms that are inverted. Other later-arriving phases that are multiply reflected off crustal discontinuities such as the PpPms or the PpSms are only present in the longest time window, which ends 25 s after the P wave (see Fig. 5). Previous studies have shown that including these backscattered waves in the inversion leads to drastic resolution improvement (e.g. Bostock *et al.* 2001; Rondenay 2009; Pageot *et al.* 2013). This is clearly seen in the last column of the top of Fig. 6 where the Moho is more sharply defined than in the case of gradients computed for shorter periods. At periods longer than 5 s, the V_s gradient starts to bring information on the long-wavelength part of the V_s model.

As pointed out by Virieux & Operto (2009), there is a close connection between full waveform inversion and diffraction tomography (Wu & Toksoz 1987). The wavenumbers in the model space that can be resolved depend on the source–receiver configuration through the relation (Sirgue & Pratt 2004; Virieux & Operto 2009; Pageot *et al.* 2013):

$$k = \frac{2\omega}{c} \cos \frac{\theta}{2}, \tag{38}$$

where θ is the scattering angle and c the local velocity. From eq. (38), it can be seen that low-frequency data and forward scattered waves ($\theta \sim 180°$) resolve the long wavelengths of the medium. On the other hand, high-frequency data and backscattered waves ($\theta \sim 0°$) resolve the small wavelengths of the medium, with a theoretical resolution limit of half a wavelength. Finally, since V_s is significantly smaller than V_p (by a factor of typically about 1.7–2), this relation also predicts that for the same frequency content, shear waves will constrain significantly shorter wavelengths than compressional waves. Applied to a teleseismic source–receiver configuration, eq. (38) thus qualitatively explains the main observations that can be made on the V_p and V_s gradients.

Another notable feature in the V_p and V_s gradients is the strong amplification of sensitivity close to the surface. This effect is only barely visible in the gradients plots because of the amplitude cut-off in the colour scales. The near-surface amplification mainly results from the geometrical spreading of seismic waves, which tends to

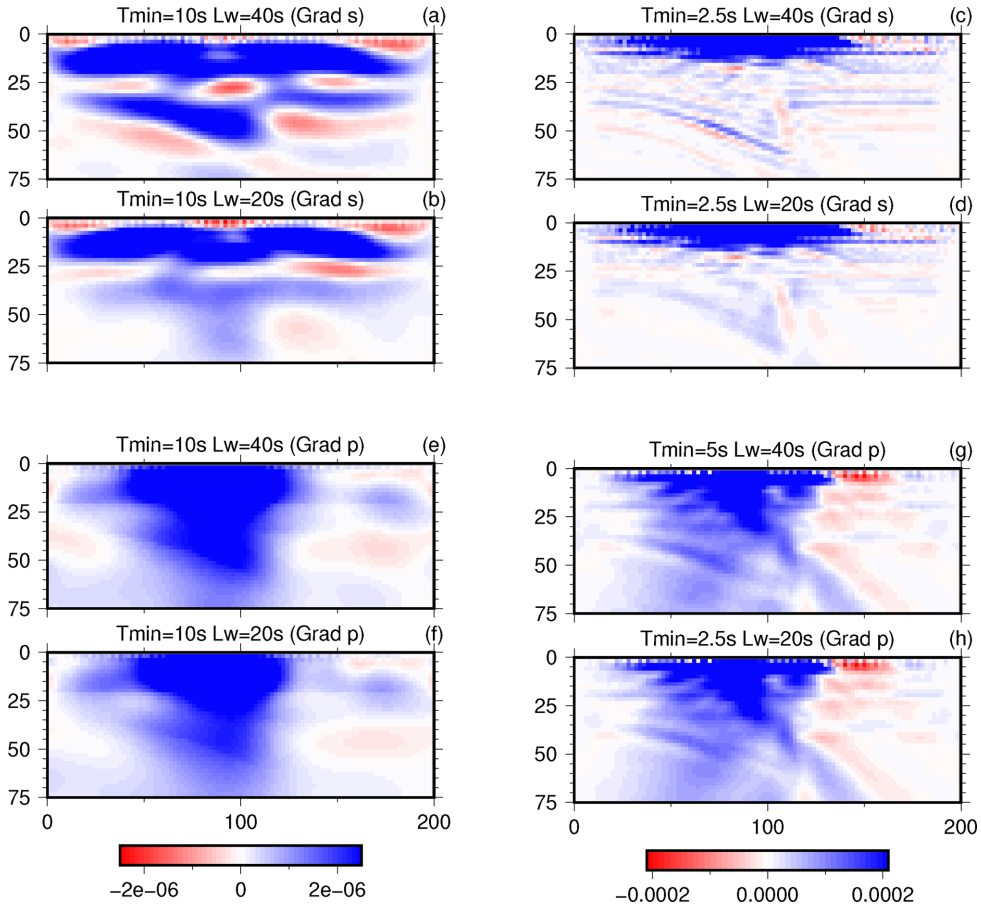

Figure 6. Gradient of the misfit function for V_s (a, b, c, d) and V_p (e, f, g, h) computed for a time window that starts 10 s before the direct P wave and ends 20 s (b, d, f, h) or 40 s (a, b, e, g) after. We filtered the waveforms with a 10 s (a, b, e, f) or 2.5 s (c, d, g, h) low-pass Butterworth filter.

diminish their amplitude at depth. These contrasted sensitivities in the V_p and V_s gradients are usually accounted for and balanced by a pre-conditioning operator. In the L-BFGS algorithm, one needs to define an approximate inverse Hessian that is later refined in the different iterations of the algorithm; for that purpose in the following we will define the initial inverse Hessian based on the square root of depth at each point in the tomographic grid.

6.3 A first inversion test

We run a first test by considering a time window starting 10 s before and ending 50 s after the P wave, for periods down to 2.5 s, inverting the broadband seismograms simultaneously for the four sources.

6.3.1 Results

Fig. 7 shows the convergence history for V_p and V_s models obtained during the iterations of the L-BFGS algorithm. Compared to the input model, the final V_p model fails to resolve the sharp interfaces of the input crustal model but the thick crustal root is well retrieved, with a geometry close to the original and with very little artefacts. In fact, the results of this first attempt to use the P waveforms are already very good for the V_p model. However, the results for V_s are less accurate: in the V_s model the seismic interfaces are more sharply defined, but very strong artefacts are visible in the lower crust. Such very contrasted behaviours directly result from the different sensitivities of teleseismic P wavefields to V_p and V_s, as pointed out above.

6.3.2 Comparison between the gradient and the L-BFGS methods

We applied the gradient and the L-BFGS methods to the complete set of seismograms, iterating upon convergence of the algorithm. As mentioned above, we define convergence as the moment at which the algorithm cannot honour the Wolfe conditions any more. With the gradient method, a 99.79 per cent reduction of the misfit is obtained after convergence, which is reached after 732 iterations. The final model obtained with the L-BFGS algorithm gives the same reduction of the misfit, but more importantly convergence is reached much faster, after only 137 iterations. Fig. 8 compares the evolution of the misfit function for the gradient and L-BFGS algorithms. Since each L-BFGS iteration requires the same amount of computations as the gradient method, this comparison demonstrates that owing to its faster convergence rate the L-BFGS method is superior because it requires fewer iterations and thus a smaller total number of calculations than the gradient method to reach convergence.

6.3.3 Comparison between adjoint tomography and full waveform inversion

The seismological community has started to compute synthetic seismograms in 3-D global and regional models and incorporate them in a minimization procedure of phase misfits measured in different

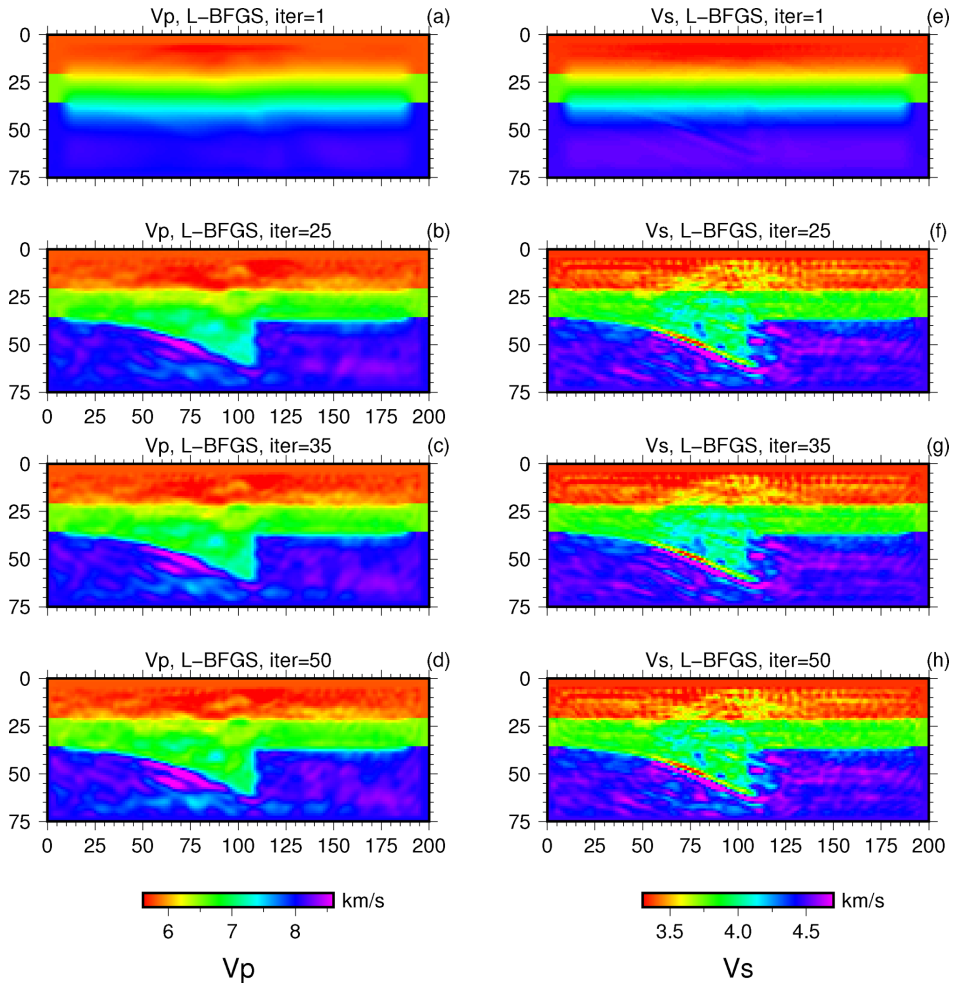

Figure 7. V_p (left) and V_s models obtained after 10 (a, e), 25 (b, f), 35 (c, g) and 50 (d, h) iterations of the L-BFGS algorithm.

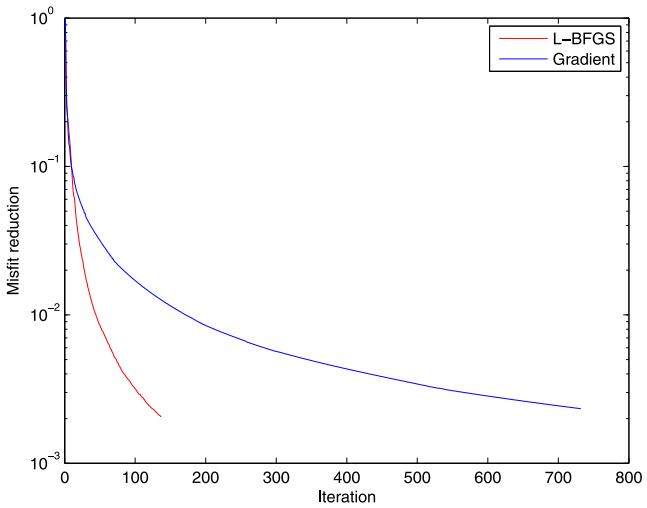

Figure 8. Decay of the misfit function as a function of the number of iterations in the gradient (blue) and L-BFGS (red) algorithms.

frequency bands (Fichtner *et al.* 2009; Tape *et al.* 2009). This can also be related to full waveform inversion with transmission data as developed for instance by Gee & Jordan (1992) and Luo & Schuster (1991). This minimization procedure is also iterative, and requires computing gradient kernels based on the adjoint method, but in this case using differential traveltime adjoint sources instead of waveform adjoint sources as in the full waveform inversion approach. This tomographic approach, sometimes called adjoint tomography in the recent literature, has gained popularity owing to its superior resolution potential compared to classical ray tomography. Since this approach relies on the phase of the seismic wavefield, it is robust and quasi-linearly related to Earth's structures. At this point, it is thus interesting to investigate whether resorting to waveform adjoint sources instead of differential traveltime adjoint sources, that is, using full waveforms instead of phase information only, can offer significant improvement.

To address this issue, we ran a set of inversions considering vertical components of the direct *P*-wave train in a time window that starts 5 s before the *P*-wave onset and ends 7 s afterwards. We filtered the seismograms with a 5 s low-pass Butterworth filter. We first implemented adjoint tomography, in which we minimized the phase misfit between the direct *P* wave in the real and synthetic vertical components of the seismograms, starting from the initial smooth 1-D model described above. After 15 iterations, the algorithm converged towards a model that provides a misfit reduction of 99 per cent. The final model, shown at the bottom of Fig. 9, thus provides an almost perfect fit of phase anomalies. We then performed full waveform inversion in the same data set, following the approach described above. The algorithm converged after five iterations, with a misfit reduction of 96 per cent. The model obtained by full waveform inversion is shown at the top of Fig. 9.

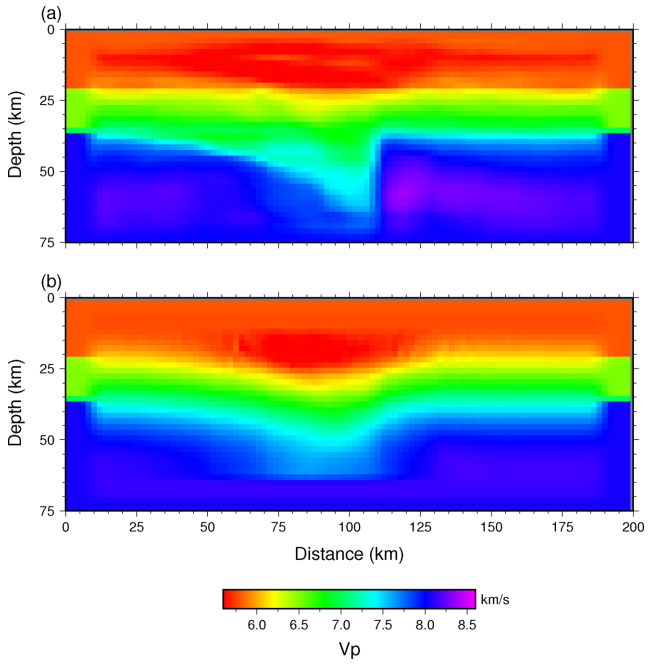

Figure 9. Models obtained by full waveform inversion (a) and adjoint tomography (b) from vertical-component direct *P* waves, after 15 iterations of the L-BFGS in the case of adjoint tomography and 5 iterations in the case of full waveform inversion.

Both methods are able to retrieve the thick crustal root beneath the central part of the profile, with very little vertical smearing. This is remarkable since, as in the previous inversion tests, these models were obtained with only four teleseismic sources. On the other hand, in the full waveform inversion model, the geometry of the crustal root is much more sharply defined, which clearly points to significantly finer resolution, both laterally and vertically. In particular, the shape and the velocity contrasts of the main seismic interfaces are both very well retrieved. This strongly suggests that there is rich and useful information contained in the amplitude of teleseismic waves, even at relatively long period (∼5 s) and that for the same numerical cost, full waveform inversion can lead to more accurate results than adjoint tomography.

6.4 Hierarchical multiscale full waveform inversion

The first inversion tests that we have performed above have shown that the iterative full waveform algorithm may converge towards a model that can be far from the true model, especially for the V_s model. Indeed, the V_s model in Fig. 7 presents very strong high- and low-velocity anomalies in the lower crust. In this section, let us thus propose and explore some possible improvements of the inversion algorithm.

A well-known limitation of full waveform inversion is that the traveltimes predicted by the starting model must match the true traveltimes with an error less than half the dominant period of the signals. If not, cycle skipping occurs and inversion converges towards a local minimum (e.g. Virieux & Operto 2009). To avoid this problem, a simple solution is to adopt a multiscale or hierarchical algorithm, starting from the low frequencies and gradually moving to including higher frequencies as well (e.g. Pratt 1999; Pageot *et al.* 2013). Such a multiscale approach has been used in a number of studies to obtain detailed velocity models of the crust and lithosphere (e.g. Tape *et al.* 2009, 2010; Fichtner *et al.* 2009, 2013; Zhu & Tromp 2013). We thus performed inversions of *P* waveforms over longer time windows, starting 10 s before the *P* arrival and ending 50 s after. In the first inversion we filtered the data between 0.01 and 0.1 Hz using a Butterworth bandpass filter. The starting model is the smooth model shown in Fig. 4. The L-BFGS algorithm is iterated upon convergence, usually reached after five or six iterations, and then repeated with an upper corner frequency gradually increased to 0.8 Hz by steps of 0.1 Hz. Each new inversion starts from the 3-D model obtained after convergence of the previous inversion. Smoothing constraints are imposed to the model by damping its Laplacian. Without them, the models would be contaminated by strong artefacts, which tend to be stronger at shorter period. Some models obtained at different stages of the inversion are shown in Fig. 11. The comparisons shown in Fig. 10 of the waveforms obtained in the final 3-D model (Figs 11c–d) with the waveforms obtained in the true model (Fig. 3) demonstrate that the final model provides a very good match of the waveforms in all the frequency bands considered.

In a second test, we considered shorter time windows, which this time ended 25 s after the *P* wave, and left all the other parameters unchanged. The models obtained in this new inversion test are shown in Figs 11(b) and (f). Comparison of the models clearly suggests that shortening the time window has a detrimental effect on the results of waveform inversion. Closer inspection of Fig. 11 reveals that the V_s model is always more poorly retrieved than the V_p model, with a quality that deteriorates at frequencies above 0.2 Hz. This comes from the fact that at these frequencies the shear wavelength is smaller than station spacing and thus aliasing appears. We thus tried a new inversion with a station spacing of 500 m. The results are shown in Figs 11(d) and (h). The small artefacts that were present with a coarser array of recording stations are considerably reduced. The density of sensors is thus a crucial parameter that controls spatial resolution (see also Pageot *et al.* 2013, for a more detailed discussion of the relation between resolution and sampling of the seismic wavefield), especially for shear wave velocities. In all the above we have used the same penalty coefficients to smooth the V_p and V_s models, while the *P* wavelength is approximately twice the *S* wavelength. This would suggest that stronger smoothing constraints should be applied to the shear velocity models. There is thus still room for improvement of the full waveform inversion algorithm used. Such improvements will be investigated in more details in future work when we start applying our waveform inversion approach to real data.

7 DISCUSSION

Using a sequential, multiscale waveform inversion of seismograms recorded by a rectilinear array, we have shown that it is possible to recover V_p and V_s over a broad range of scales, including the geometry of sharp seismic discontinuities such as the mid-crust and crust-mantle interfaces. These velocity models have a much finer resolution than what could be obtained based on classical ray-based traveltime tomography or even adjoint tomography. In particular, the depth resolution is remarkable, with almost no visible vertical smearing of the low-velocity crustal anomaly into the mantle, which is a typical shortcoming of regional body wave tomography (e.g. Chevrot *et al.* 2014).

So far we have only investigated a significantly heterogeneous crustal model, containing both short- and long-wavelength heterogeneities. For the same 5 km receiver spacing as the one considered here, the overlap of the Fresnel zones of adjacent receivers will

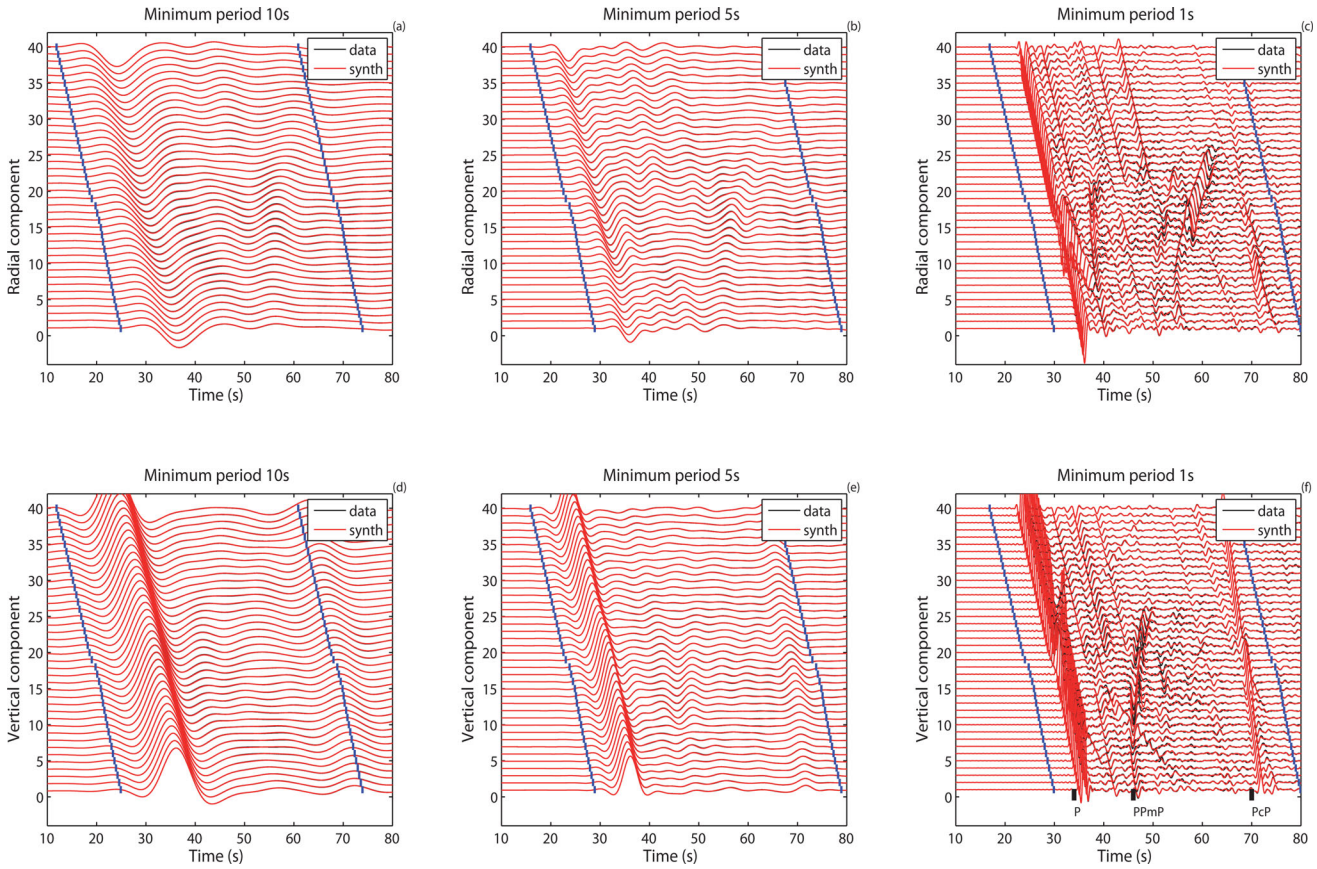

Figure 10. Radial and vertical synthetic seismograms for the first source (distance 60°, backazimuth 0°) computed in the 3-D model of Fig. 3 (black lines) and in the final model of Figs 11(c) and (g) (red lines). The seismograms are low-pass filtered at 10 s (a, d), 5 s (b, e) and 1 s (c, f).

increase with depth and we thus expect even finer resolution for imaging lithospheric mantle structures. This opens important and new perspectives for instance for detailed imaging of subduction zones or of deep crustal and lithospheric roots of high mountain ranges (e.g. the Himalayas).

Our synthetic experiment tried to reproduce the characteristics of temporary experiments as closely as possible. For example, the 5 km receiver spacing is similar to that in many recent temporary deployments (Chevrot *et al.* 2014). The ability to obtain well-resolved models with an extremely limited number of sources (here we used only four different teleseismic sources) also suggests that this approach can be used even for data obtained in short-duration temporary experiments (typically a few months) with reasonable hope of success. For a much larger number of teleseismic sources, source encoding may be a viable option to accelerate the computations. For example, Castellanos *et al.* (2015) have shown that the L-BFGS method can be implemented through a source-encoding approach, which leads to significant reduction of the computational cost.

In this study we have only considered the direct *P* waves, but since the DSM–SEM hybrid method can model any type of incoming wave, it would be straightforward to include any other later-arriving phase in the inversion. Considering seismic phases with different slownesses (or incidence angles) would improve spatial coverage and consequently resolution in the final tomographic models. For example, including the direct *S* waves should allow for better reconstruction of the V_s model.

Before applying our full waveform inversion approach to real waveform data, a number of issues will need to be addressed. First, in our synthetic experiments, we considered simple impulsive wavelets with short temporal support. As a result, the different seismic phases are well separated in time, at least at short period. In contrast, natural earthquakes will have more complex and longer source wavelets. These source wavelets and source locations will need to be estimated before the inversion or included into the inversion problem (e.g. Liu *et al.* 2004). Another potential source of complexity may come from topographic relief. In addition to perturbing the arrival times and amplitudes of seismic waves, topography of the free surface (e.g. Lee *et al.* 2008, 2009; Köhler *et al.* 2012) generates strong body-to-surface wave conversions that contaminate the coda of the *P* wave (Monteiller *et al.* 2013). This signal-generated noise will decrease the signal-to-noise ratio of later-arriving phases such as the *P*-to-*S* converted phases or the crustal multiples that are key to image the sharp velocity interfaces. However, since simple migration approaches of *P*-to-*S* converted waves have largely demonstrated their potential to image seismic discontinuities, there should be no reason for full waveform inversion, which more accurately describes all the wave propagation effects, to fail. Finally, site effects may affect the amplitude of teleseismic *P* waves, in particular at short periods (e.g. Zhou *et al.* 2003). It will thus be necessary to separate local site amplifications from the effects of focusing/defocusing of seismic waves produced by propagation through an heterogeneous medium.

8 CONCLUSIONS AND FUTURE WORK

Building on the recently developed DSM/SEM hybrid method to compute synthetic seismograms of short-period teleseismic *P*

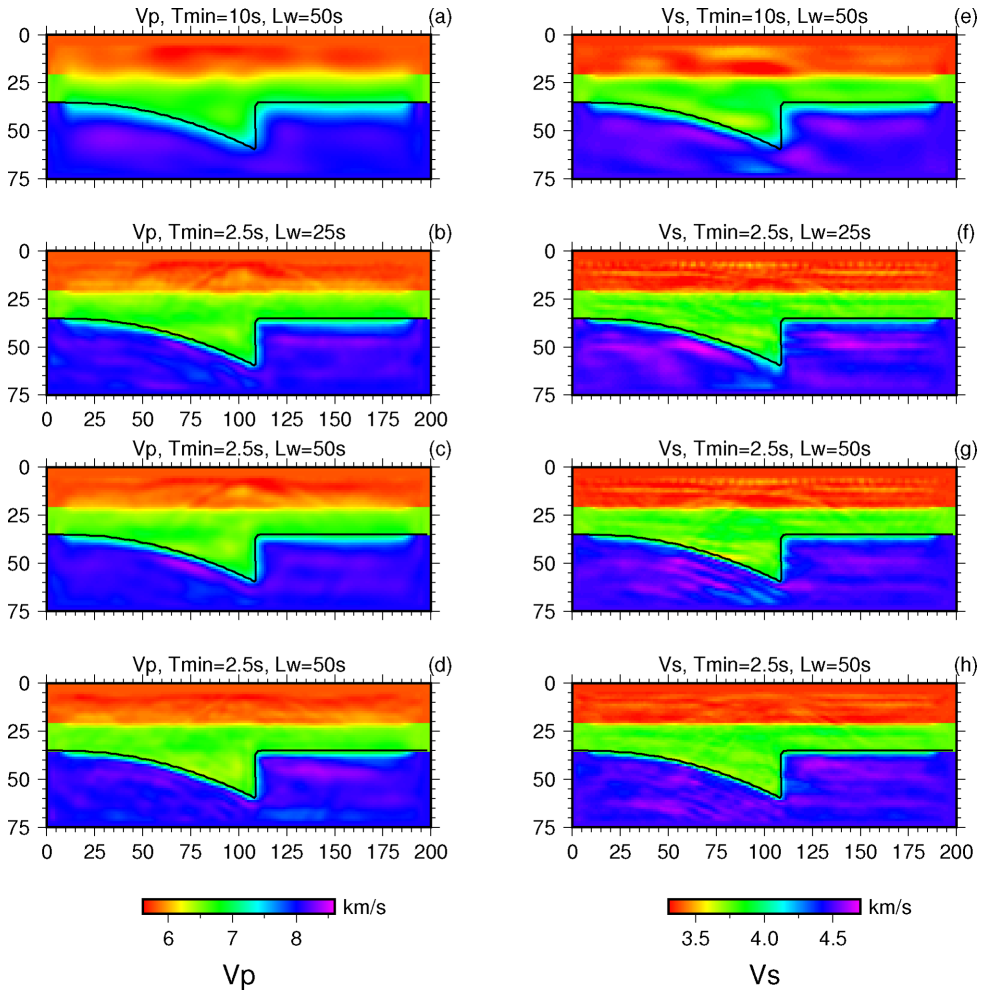

Figure 11. Full waveform inversion with the hierarchical multiscale approach. The panels show the (a) V_p and (e) V_s models obtained in the first inversion, which uses a 0.001–0.1 Hz frequency range. Panels (b) and (f) show the V_p and V_s models obtained using a 0.001–0.4 Hz frequency range with a 35 s time window going from 10 s before to 25 s after the *P* wave, panels (c) and (g) using a 60 s time window going from 10 s before to 50 s after the *P* wave, and panels (d) and (h) for a station spacing of 500 m instead of 5.5 km.

waves, we have implemented an adjoint approach to compute the gradient of traveltime or waveform misfit functions in 3-D regional models. This key ingredient for full waveform inversion of seismic records opens new perspectives for high-resolution imaging beneath dense permanent or temporary arrays of seismic sensors. We have demonstrated that this nonlinear full waveform inversion problem can be efficiently handled based on the L-BFGS algorithm, a memory-efficient implementation of the Gauss–Newton algorithm that does not require computing and storing the inverse Hessian matrix. Compared to the gradient method, the L-BFGS method converges faster (in about four times fewer iterations in the examples that we have shown) and also leads to models that provide stronger waveform misfit reductions. This points to avoiding gradient-based methods to solve nonlinear waveform inversion problem. We have shown that a sequential multiscale inversion, in which a V_p model is first inverted from the vertical component of the *P* wave and then refined in a subsequent inversion for both V_p and V_s using both vertical and radial components, leads to dramatic improvement of the V_s model retrieved compared to the direct joint inversion of V_p and V_s. Convergence is achieved after a few tens of iterations (typically less than 20), providing high-resolution models of V_p and V_s in which both the short- and long-wavelength structures are very well retrieved. This key advantage over conventional reflection seismic imaging comes from the presence of both direct transmitted *P* waves, which constrain the 3-D background velocity model, and reflected/converted waves in the coda of the *P* waves, which mainly constrain the geometry of seismic interfaces. Given the ease of access to moderate-size computing clusters and the increasing development of dense regional seismic recording arrays worldwide, we anticipate rapidly growing interest in regional full waveform imaging from both seismologists and geologists alike, who will, perhaps for the first time, have the possibility to analyse high-resolution tomographic images prone to detailed interpretations in varied tectonic environments. Application of this new full waveform imaging approach to real data is under way.

ACKNOWLEDGEMENTS

We thank Nozomu Takeuchi, Qinya Liu, Ping Tong, Roland Martin, Yann Capdeville, Stéphane Operto and Clara Castellanos for fruitful discussion, and the Computational Infrastructure for Geodynamics (CIG) for support. We also thank Andreas Fichtner and an anonymous reviewer for useful comments that improved the manuscript. Our full waveform inversion software package SPECFEM3D, including the inversion routines and the coupling with DSM

presented above, is available open source at geodynamics.org. This study was supported in part by the French Research Agency 'ANR blanc' programme (PYROPE project, #ANR-09-BLAN-0229) and by the ANR 'Interdisciplinary Program on Application Software towards Exascale Computing for Global Scale Issues' (SEISMIC_IMAGING project, #ANR-10-G8EX-002). Part of this work was performed in the context of the European 'Mont-Blanc: European scalable and power efficient HPC platform based on low-power embedded technology' #288777 project of call FP7-ICT-2011-7, and with support from the Simone and Cino del Duca/Institut de France / French Academy of Sciences Foundation under grant #095164. This work was also granted access to the French HPC resources of TGCC under allocation #2014-gen7165 made by GENCI and of the Aix-Marseille Supercomputing Mesocentre under allocation #14b013.

REFERENCES

Aravkin, A., van Leeuwen, T. & Herrmann, F., 2011. Robust full-waveform inversion using the Student's *t*-distribution, in *SEG Technical Program Expanded Abstracts 2011,* pp. 2669–2673, doi:10.1190/1.3627747.

Audet, P., Bostock, M.G., Christensen, N.I. & Peacock, S.M., 2009. Seismic evidence for overpressured subducted oceanic crust and megathrust fault sealing, *Nature,* **457,** 76–78.

Bamberger, A., Chavent, G. & Lailly, P., 1977. Une application de la théorie du contrôle à un problème inverse sismique, *Ann. Geophys.,* **33,** 183–200.

Bamberger, A., Chavent, G., Hemons, C. & Lailly, P., 1982. Inversion of normal incidence seismograms, *Geophysics,* **47**(5), 757–770.

Bostock, M.G., 1998. Mantle stratigraphy and evolution of the Slave province, *J. geophys. Res.,* **103,** 21 183–21 200.

Bostock, M.G. & Rondenay, S., 1999. Migration of scattered teleseismic body waves, *Geophys. J. Int.,* **137,** 732–746.

Bostock, M.G., Rondenay, S. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves - 1. Theory for oblique incidence, *J. geophys. Res.,* **106,** 30 771–30 782.

Brossier, R., Operto, S. & Virieux, J., 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion?, *Geophysics,* **75**(3), R37–R46.

Bube, K.P. & Langan, R.T., 1997. Hybrid L_1/L_2 minimization with applications to tomography, *Geophysics,* **62**(4), 1183–1195.

Capdeville, Y., Chaljub, E., Vilotte, J.P. & Montagner, J.P., 2003a. Coupling the spectral element method with a modal solution for elastic wave propagation in global Earth models, *Geophys. J. Int.,* **152,** 34–67.

Capdeville, Y., To, A. & Romanowicz, B., 2003b. Coupling spectral elements and modes in a spherical Earth: an extension to the 'sandwich' case, *Geophys. J. Int.,* **154,** 44–57.

Castellanos, C., Métivier, L., Operto, S., Brossier, R. & Virieux, J., 2015. Fast full waveform inversion with source encoding and second-order optimization methods, *Geophys. J. Int.,* **200,** 720–744.

Chaljub, E., Capdeville, Y. & Vilotte, J.P., 2003. Solving elastodynamics in a fluid-solid heterogeneous sphere: a parallel spectral element approximation on non-conforming grids, *J. Comput. Phys.,* **187**(2), 457–491.

Chavent, G., 1974. Identification of function parameters in partial differential equations, in *Identification of Parameter Distributed Systems,* pp. 31–48, eds. Goodson, R.E. & Polis, M. American Society of Mechanical Engineers.

Chen, L., Wen, L. & Zheng, T., 2005. A wave equation migration method for receiver function imaging: 1. Theory, *J. geophys. Res.,* **110,** B11309, doi:10.1029/2005JB003665.

Chevrot, S., Favier, N. & Komatitsch, D., 2004. Shear wave splitting in three-dimensional anisotropic media, *Geophys. J. Int.,* **159,** 711–720.

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**(3), 1391–1402.

Chevrot, S. *et al.*, 2014. High resolution imaging of the Pyrenees and Massif Central from the data of the PYROPE and IBERARRAY portable array deployments, *J. geophys. Res.,* **119**(8), 6399–6420.

Crase, E., Pica, A., Noble, M., McDonald, J. & Tarantola, A., 1990. Robust elastic non-linear waveform inversion: application to real data, *Geophysics,* **55,** 527–538.

Cyr, E.C., Shadid, J.N. & Wildey, T., 2015. Towards efficient backward-in-time adjoint computations using data compression techniques, *Comput. Methods Appl. Mech. Eng.,* **288,** 24–44.

Fichtner, A., 2010. *Full Seismic Waveform Modelling and Inversion,* Advances in Geophysical and Environmental Mechanics and Mathematics, Springer Verlag.

Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology: I. Theory, *Phys. Earth planet. Inter.,* **157**(1–2), 86–104.

Fichtner, A., Kennett, B. L.N., Igel, H. & Bunge, H.P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.,* **179**(3), 1703–1725.

Fichtner, A., Saygin, E., Taymaz, T., Cupillard, P., Capdeville, Y. & Trampert, J., 2013. The deep structure of the North Anatolian Fault Zone, *Earth planet. Sci. Lett.,* **373,** 109–117.

Formaggia, L., Gerbeau, J.F., Nobile, F. & Quarteroni, A., 2001. On the coupling of 3D and 1D Navier-Stokes equations for flow problems in compliant vessels, *Comput. Methods Appl. Mech. Eng.,* **191**(6–7), 561–582.

Friederich, W. & Dalkolmo, J., 1995. Complete synthetic seismograms for a spherically-symmetric Earth by a numerical computation of the Green's function in the frequency domain, *Geophys. J. Int.,* **122**(2), 537–550.

Gee, L.S. & Jordan, T.H., 1992. Generalised seismological data functionals, *Geophys. J. Int.,* **111,** 363–390.

Geller, R.J. & Ohminato, T., 1994. Computation of synthetic seismograms and their partial derivatives for heterogeneous media with arbitrary natural boundary-conditions using the Direct Solution Method, *Geophys. J. Int.,* **116,** 421–446.

Geller, R.J. & Takeuchi, N., 1995. A new method for computing highly accurate DSM synthetic seismograms, *Geophys. J. Int.,* **123,** 449–470.

Gengembre, N., Lhémery, A., Omote, R., Fouquet, T. & Schumm, A., 2004. A semi-analytic-FEM hybrid model for simulating UT configurations involving complicated interactions of waves with defects, in *Review of Progress in Quantitative Nondestructive Evaluation,* Vol. 23, pp. 74–80, eds Thompson, D.O. & Chimenti, D.E. American Institute of Physics.

Godinho, L., Mendes, P.A., Tadeu, A., Cadena-Isaza, A., Smerzini, C., Sánchez-Sesma, F.J., Madec, R. & Komatitsch, D., 2009. Numerical simulation of ground rotations along 2D topographical profiles under the incidence of elastic plane waves, *Bull. seism. Soc. Am.,* **99**(2B), 1147–1161.

Ha, T., Chung, W. & Shin, C., 2009. Waveform inversion using a back-propagation algorithm and a Huber function norm, *Geophysics,* **74**(3), R15–R24.

Han, J.-H., Kim, Y.-J. & Karkoub, M., 2014. Wave propagation modeling of fluid-filled pipes using hybrid analytical/two-dimensional finite element method, *Wave Motion,* **51**(7), 1193–1208.

Hudson, J.A., 1977. Scattered waves in the coda of P, *J. geophys. Res.,* **43,** 359–374.

Jeong, W., Kang, M., Kim, S., Min, D.-J. & Kim, W.-K., 2015. Full waveform inversion using Student's *t* distribution: a numerical study for elastic waveform inversion and simultaneous-source method, *Pure appl. Geophys.,* pp. 1–19.

Kallivokas, L.F., Fathi, A., Kucukcoban, S., Stokoe, K.H. II, Bielak, J. & Ghattas, O., 2013. Site characterization using full waveform inversion, *Soil Dyn. Earthq. Eng.,* **47,** 62–82.

Kawai, K., Takeuchi, N. & Geller, R.J., 2006. Complete synthetic seismograms up to 2 Hz for transversely isotropic spherically symmetric media, *Geophys. J. Int.,* **164,** 411–424.

Kawakatsu, H. & Watada, S., 2008. Seismic evidence for deep-water transportation in the mantle, *Science,* **316,** 1468–1471.

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

Köhler, A., Weidle, C. & Maupin, V., 2012. On the effect of topography on surface wave propagation in the ambient noise frequency range, *J. Seismol.,* **16**(2), 221–231.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation, *Geophys. J. Int.,* **139**(3), 806–822.

Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation-I. Validation, *Geophys. J. Int.,* **149**(2), 390–412.

Komatitsch, D. & Vilotte, J.P., 1998. The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.,* **88**(2), 368–392.

Komatitsch, D., Tsuboi, S. & Tromp, J., 2005. The spectral-element method in seismology, in *Seismic Earth: Array Analysis of Broadband Seismograms, Vol. 157 of Geophysical Monograph,* pp. 205–228, eds Levander, A. & Nolet, G. American Geophysical Union.

Komatitsch, D., Vinnik, L.P. & Chevrot, S., 2010. SHdiff/SVdiff splitting in an isotropic Earth, *J. geophys. Res.,* **115**(B7), B07312, doi:10.1029/2009JB006795.

Lailly, P., 1983. The seismic inverse problem as a sequence of before-stack migrations, in *Proceedings of the Conference on Inverse Scattering, Theory and Application Expanded Abstracts,* pp. 206–220, eds Bednar, J.B., Redner, R., Robinson, E. & Weglein, A. Society of Industrial and Applied Mathematics, Philadelphia.

Lee, S.J., Chen, H.W., Liu, Q., Komatitsch, D., Huang, B.S. & Tromp, J., 2008. Three-dimensional simulations of seismic wave propagation in the Taipei basin with realistic topography based upon the spectral-element method, *Bull. seism. Soc. Am.,* **98**(1), 253–264.

Lee, S.J., Komatitsch, D., Huang, B.S. & Tromp, J., 2009. Effects of topography on seismic wave propagation: An example from northern Taiwan, *Bull. seism. Soc. Am.,* **99**(1), 314–325.

Liu, Q. & Gu, Y.J., 2012. Seismic imaging: From classical to adjoint tomography, *Tectonophysics,* **566–567,** 31–66.

Liu, Q., Polet, J., Komatitsch, D. & Tromp, J., 2004. Spectral-element moment tensor inversions for earthquakes in Southern California, *Bull. seism. Soc. Am.,* **94**(5), 1748–1761.

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

Mariotti, C., Le Piver, F. & Aubry, L., 2015. A least-squares coupling method between a finite element code and a discrete element code, *Int. J. Numer. Methods Eng.,* **101**(10), 731–743.

Masson, Y., Cupillard, P., Capdeville, Y. & Romanowicz, B., 2014. On the numerical implementation of time-reversal mirrors for tomographic imaging, *Geophys. J. Int.,* **196**(3), 1580–1599.

Monteiller, V., Chevrot, S., Komatitsch, D. & Fuji, N., 2013. A hybrid method to compute short-period synthetic seismograms of teleseismic body waves in a 3-D regional model, *Geophys. J. Int.,* **192,** 230–247.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100,** 12 431–12 447.

Nissen-Meyer, T., van Driel, M., Stähler, S.C., Hosseini, K., Hempel, S., Auer, L., Colombi, A. & Fournier, A., 2014. AxiSEM: Broadband 3-D seismic wavefields in axisymmetric media, *Solid Earth,* **5**(1), 425–445.

Nocedal, J. & Wright, S., 2006. *Numerical Optimization,* 2nd edn, Springer Series in Operations Research and Financial Engineering, Springer Verlag.

Pageot, D., Operto, S., Vallée, M., Brossier, R. & Virieux, J., 2013. A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging, *Geophys. J. Int.,* **193,** 1479–1505.

Peter, D. *et al.*, 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.,* **186,** 721–739.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167,** 495–503.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model, *Geophysics,* **64,** 888–901.

Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.,* **133,** 341–362.

Roecker, S., Baker, B. & McLaughlin, J., 2010. A finite-difference algorithm for full waveform teleseismic tomography, *Geophys. J. Int.,* **181,** 1017–1040.

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

Rondenay, S., 2009. Upper mantle imaging with array recordings of converted and scattered teleseismic waves, *Surv. Geophys.,* **30,** 377–405.

Rondenay, S., Abers, G.A. & van Keken, P.E., 2008. Seismic imaging of subduction zone metamorphism, *Geology,* **36,** 275–278.

Rubio Dalmau, F., Hanzich, M., de la Puente, J. & Gutiérrez, N., 2014. Lossy data compression with DCT transforms, in *Proceedings of the EAGE Workshop on High Performance Computing for Upstream,* p. HPC30, Chania, Crete, Greece.

Sadourny, R., 1972. Conservative finite-difference approximations of the primitive equations on quasi-uniform spherical grids, *Mon. Weather Rev.,* **100,** 136–144.

Sambridge, M., 1999. Geophysical inversion with a neighbourhood algorithm - I. Searching a parameter space, *Geophys. J. Int.,* **138,** 479–494.

Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, *Geophysics,* **69,** 231–248.

Takeuchi, N., Geller, R. & Cummins, P., 1996. Highly accurate P-SV complete synthetic seismograms using modified DSM operators, *Geophys. Res. Lett.,* **23,** 1175–1178.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science,* **325,** 988–992.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.,* **180,** 433–462.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49,** 1259–1266.

Tong, P., Chen, C.-W., Komatitsch, D., Basini, P. & Liu, Q., 2014a. High-resolution seismic array imaging based on a SEM-FK hybrid method, *Geophys. J. Int.,* **197**(1), 369–395.

Tong, P., Komatitsch, D., Tseng, T.-L., Hung, S.-H., Chen, C.-W., Basini, P. & Liu, Q., 2014b. A 3-D spectral-element and frequency-wave number hybrid method for high-resolution seismic array imaging, *Geophys. Res. Lett.,* **41**(20), 7025–7034.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**(1), 195–216.

Tromp, J., Komatitsch, D. & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Commun. Comput. Phys.,* **3**(1), 1–32.

Tsuboi, S., Komatitsch, D., Ji, C. & Tromp, J., 2003. Broadband modeling of the 2002 Denali fault earthquake on the Earth Simulator, *Phys. Earth planet. Inter.,* **139**(3–4), 305–313.

Vai, R., Castillo-Covarrubias, J.M., Sánchez-Sesma, F.J., Komatitsch, D. & Vilotte, J.P., 1999. Elastic wave propagation in an irregularly layered medium, *Soil Dyn. Earthq. Eng.,* **18**(1), 11–18.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics,* **74,** WCC1–WCC26.

Wu, R.S. & Toksoz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics,* **52,** 11–25.

Zhao, L., Wen, L., Chen, L. & Zheng, T., 2008. A two-dimensional hybrid method for modeling seismic wave propagation in anisotropic media, *J. geophys. Res.,* **113,** B12307, doi:10.1029/2008JB005733.

Zhou, Y., Nolet, G. & Dahlen, F.A., 2003. Surface sediment effects on teleseismic P wave amplitude, *J. geophys. Res.,* **108,** 2417, doi:10.1029/2002JB002331.

Zhu, H. & Tromp, J., 2013. Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean, *Science,* **341,** 871–875.