A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging,

S. Operto, D. Pageot, Martin Vallée, J. Virieux, R. Brossier

Geophysical Journal International

Geophys. J. Int. (2013) **193,** 1479–1505
Advance Access publication 2013 March 23

doi: 10.1093/gji/ggs132

A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging

Damien Pageot,[1] Stéphane Operto,[1] Martin Vallée,[1,*] Romain Brossier[2] and Jean Virieux[2]

[1] *Géoazur, Université de Nice Sophia-Antipolis, CNRS, Observatoire de la Côte d'Azur, Sophia Antipolis, France. E-mail: pageot@geoazur.unice.fr*
[2] *ISTerre, Université Joseph Fourier, Grenoble, France*

Accepted 2012 December 30. Received 2012 December 24; in original form 2012 June 19

SUMMARY

The development of dense networks of broad-band seismographs makes teleseismic data amenable to full-waveform inversion (FWI) methods for high-resolution lithospheric imaging. Compared to scattered-field migration, FWI seeks to involve the full seismic wavefield in the inversion. We present a parametric analysis of 2-D frequency-domain FWI in the framework of lithospheric imaging from teleseismic data to identify the main factors that impact on the quality of the reconstructed compressional (*P*)-wave and shear (*S*)-wave speed models. Compared to controlled-source seismology, the main adaptation of FWI to teleseismic configuration consists of the implementation with a scattered-filed formulation of plane-wave sources that impinge on the base of the lithospheric target located below the receiver network at an arbitrary incidence angle. Seismic modelling is performed with a *hp*-adaptive discontinuous Galerkin method on unstructured triangular mesh. A quasi-Newton inversion algorithm provides an approximate accounting for the Hessian operator, which contributes to reduce the footprint of the coarse acquisition geometry in the imaging. A versatile algorithm to compute the gradient of the misfit function with the adjoint-state method allows for abstraction between the forward-problem operators and the meshes that are during seismic modelling and inversion, respectively. An approximate correction for obliquity is derived for future application to real teleseismic data under the two-dimension approximation. Comparisons between the characteristic scales involved in exploration geophysics and in teleseismic seismology suggest that the resolution gain provided by full waveform technologies should be of the same order of magnitude for both applications. We first show the importance of the surface-reflected wavefield to dramatically improve the resolving power of FWI by combining tomography-like and migration-like imaging through the incorporation of the forward-scattered and the backscattered wavefields in the inversion. The resolution of FWI is assessed through checkerboard tests and confirms a resolution of the order of the wavelength for both the *P* and *S* speeds, when the full wavefield is incorporated in the inversion. Secondly, we show that computationally efficient strategies, which consist of decimating the number of frequency components involved in the inversion, do not apply to teleseismic acquisitions, because the scattering-angle bandwidth sampled by plane-wave sources can be narrow and coarsely sampled, compared to that provided by dense profiles of point sources in exploration seismology. The waveform inversion is less sensitive to the band of incidence angles spanned by the plane-wave sources and to the sampling of this band. However, the deficit of vertically propagating plane waves hampers the vertical resolution of planar layers. Aliasing artefacts created by coarse arrays of receivers are illustrated. We show how taking into account the Hessian in the inversion and the suitable management of frequencies in the inversion help to mitigate these artefacts.

*Now at: IPGP, Paris, France.

Acceptable reconstructions are shown for both the *P*- and *S*-wave speeds for a receiver spacing of up to 20 km in the 0.1–0.4 Hz frequency range. Building a reliable initial model for FWI is a highly non-linear problem in exploration seismology. We show how the low-frequency content of the teleseismic sources allow us to build accurate *P*- and *S*-wave speed models starting from simple vertical-gradient velocity models. All of these results are derived using a 2-D realistic synthetic experiment that was performed with noise-free data. Most of conclusions of this parametric study should apply in three dimensions. The impact of noise and the footprint of other experimental parameters such as the estimation of the temporal plane-wave signature and the estimation of the incidence angle of the impinging plane waves, need however to be assessed in the future.

Key words: Inverse theory; Body waves; Seismic tomography; Computational seismology; Wave scattering and diffraction; Wave propagation.

1 INTRODUCTION

With the massive deployment of dense multicomponent broad-band seismological networks in North America, Australia, Japan and Europe, and with the increase in the computational power provided by high-performance computing, there is growing interest in the development of multichannel high-resolution tomographic approaches in earthquake seismology at the regional, continental and global scales. The most popular imaging approaches in earthquake seismology rely on receiver function analysis (Langston 1979; Ammon *et al.* 1990; Ammon 1991) and ray-theoretical traveltime tomography (Dziewonski 1984; Nolet 1987; Fukao *et al.* 2001). The receiver function method attempts to localize discontinuities in the lithosphere beneath the receiver arrays from *P*–*S*-mode conversions. These approaches share some similarities with migration methods in exploration seismology, in the sense that they provide a geometrical picture of the structural discontinuities but do not provide quantitative inferences of the physical properties of the subsurface. In contrast, ray-theoretical traveltime tomography provides low-resolution perturbation models around a starting model. When the sensitivity kernels of the tomography are rays, judicious *ad hoc* smoothing regularization and/or adaptive model parametrization should be found to image the subsurface with a spatial resolution that is as consistent as possible with the theoretical sensitivity of the traveltimes. Moreover, the regularization or the parametrization should ideally be locally adapted to the uneven ray coverage of the subsurface (Wang 1993). These difficulties prompted Dahlen *et al.* (2000) to develop finite-frequency traveltime tomography to more rigorously account for the sensitivity of the traveltimes to the Fresnel volume centred on the ray. Finite-frequency delay times are extracted by cross-correlation of recorded and modelled waveforms of selected phases, and these delay times are linearly related to the model perturbations through the single-scattering Born approximation. In the earlier developments of finite-frequency tomography, computationally efficient approaches were implemented by computing absolute and differential traveltimes by dynamic ray tracing using the paraxial approximation (Dahlen *et al.* 2000). With the advances in high-performance computing, Chen *et al.* (2007a) and Tromp *et al.* (2005) proposed to compute the full seismic wavefields with numerical approaches, such as finite difference or finite element methods, to build the kernel of the tomography. Although the inversion remains limited to phases of selected energetic wave packets, the modelling of the full seismic wavefield allows the non-linear relationship between the data and the subsurface parameters to be accounted for, hence making the tomography amenable to non-linear iterative inversion where the subsurface model is updated at each iteration. This frequency-band tomographic approach based on full wavefield modelling with a robust extraction of frequency-dependent phases as data to be fitted can be implemented with the scattering-integral or adjoint-state methods (Tromp *et al.* 2005; Chen *et al.* 2007a). The main difference between these two approaches relies on the explicit building of the sensitivity or Fréchet derivative matrix in the scattering-integral approach, while the adjoint-state method directly builds the gradient of the misfit function through reverse-time propagation of the residual wavefields. Applications of both the scattering-integral and adjoint-state methods were presented at the regional scale by Chen *et al.* (2007b) and Tape *et al.* (2009). Woodhouse & Dziewonski (1984) developed full-waveform inversion (FWI) of the upper mantle using normal modes summation at the global scale, while partial derivatives for fitting phases and amplitudes are based on the great circle approximation in relation to ray theory. Alternatively, Lekić & Romanowicz (2011) developed a hybrid approach to build global models of the upper–mantle structure by full waveform tomography. Love and Rayleigh waves and long-period body waves are obtained by a spectral element method for the upper mantle, while partial derivatives of the data can be approximately computed with improved ray-based methods (Lui & Romanowicz 1995). Weights are applied to these different synthetic amplitudes in the misfit function estimation. In an attempt to evolve towards the exploitation of the full wavefield, Fichtner *et al.* (2009) and Bozdag *et al.* (2011) proposed new misfit functions where both the phase and the amplitude attributes, such as the envelope, are taken into account in the misfit function.

Another path was followed for high-resolution lithospheric imaging from teleseismic events by Bostock *et al.* (2001), Shragge *et al.* (2001), and Rondenay *et al.* (2001). They proposed to adapt least-squares elastic ray+Born migration/inversion to scattered teleseismic body waves, which was originally developed for controlled-source seismic reflection seismology (Jin *et al.* 1992). In lithospheric imaging from teleseismic data, the sources are external plane waves, which originate from the distant earthquake and which impinge the base of the lithospheric target located beneath the receiver array with a given incidence angle. Ray+Born migration is a local optimization problem, which minimizes the misfit between the recorded and the modelled single-scattered wavefield. The single-scattered wavefield is computed by means of linearization of the forward problem around a smooth background model with the Born approximation. The Green functions in the sensitivity kernel of the linearized forward problem are computed with ray theory. The outputs of the imaging can be perturbation models (i.e. a bandpass filtered version of the subsurface model) of *P*- and *S*-wave speeds, density, and attenuation, which allow the

prediction of the recorded scattered wavefield (Forgues & Lambaré 1997; Thierry *et al.* 1999; Ribodetti *et al.* 2000). A key difference between the finite-frequency tomographic approaches previously introduced and the ray+Born inversion is that the full single-scattered wavefield contained in the coda of the primary wavefields are incorporated within the ray+Born inversion, which makes this kind of method amenable to high-resolution imaging. The potential limits of these approaches are related to the linearization of the forward problem, which prevents all of the wave phenomena from being taken into account during seismic modelling and inversion: the incident wavefield is not involved in the inversion, and must be separated from the scattered wavefield during a pre-processing step, and multiscattering is not taken into account during the linearized modelling. Moreover, the accuracy of the scattered wavefield computed for large scattering angles can be questioned in the framework of the ray+Born approximation (Lambaré 1991, p. 167). This might be an issue in teleseismic geometries, where forward scattering might be the dominant scattering regime. As in any migration techniques, the imaging relies on a scale uncoupling between the large wavelengths contained in the background model, which remain constant over the linear iterations, and the short-wavelength model perturbations that are reconstructed by the migration process. This scale uncoupling requires the background model to match the phase spectrum of the scattered waves within the full frequency band with an error that does not exceed half the period. If this condition is not satisfied, the scattered waves recorded at receiver positions will be backprojected onto the wrong isochrone (i.e. equiphase) surfaces, leading to poor focusing and incorrect positioning. An application of ray+Born migration/inversion to teleseismic data from the Cascadia subduction zone was presented by Rondenay *et al.* (2001) and Rondenay *et al.* (2005). They showed that the key factor for building high-resolution perturbation models of the *P*-wave and *S*-wave speeds is the explicit incorporation of the reflections from the free surface in the ray+Born modelling. This allows them to invert the backscattered wavefield reflected from the lithospheric structures after a first reflection from the free surface. These specific double-scattered events allow a reflection survey to be mimicked and fine-scale discontinuities to be imaged through a backscattering migration process.

In this study, we address some issues related to the feasibility of frequency-domain elastic FWI of teleseismic data for lithospheric imaging. Here, FWI refers to the waveform inversion of the phase and amplitude of all of the arrivals (Tarantola 1984; Pratt *et al.* 1998; Virieux & Operto 2009). In this context, seismic modelling aims to compute the full solution of the wave equation. The approach developed in the present study differs from the above-mentioned finite-frequency tomographic approaches in its ability to invert the full waveform of all of the arrivals without the need to identify a specific phase. On the other hand, it differs from the ray-based migration method developed by Bostock *et al.* (2001) in the sense that we honour the non-linear relationship between the data and the model parameter. In this framework, the misfit between the recorded and modelled data is minimized iteratively in a non-linear sense and the starting model of each iteration is updated with the final model of the previous iteration. As the full wavefield is involved in the inversion, no separation between the direct wavefield and the scattered wavefield is performed. The gradients of the misfit function in non-linear and linear inversions have the same expression and rely on the single-scattering Born approximation: in both cases, the partial derivative of the wavefield with respect to one model parameter represents the wavefield scattered by this model parameter (Pratt *et al.* 1998). However, double scattering is taken into account in the non-linear inversion formula through the second-order term of the Hessian operator (Pratt *et al.* 1998; Fichtner & Trampert 2011a). In non-linear FWI of teleseismic data, the joint inversion of the incident primary wavefield and the wavefield reflected from the free surface implies that tomography-like and migration-like reconstructions are combined, which is amenable to a broad-band reconstruction of the lithospheric target.

We perform 2-D elastic FWI in the frequency domain, which provides a natural framework to design multiscale imaging by successive inversions of increasing frequencies (Pratt 1990). The *P*- and *S*-wave speeds are jointly updated from vertical and radial geophones. Frequency-domain FWI was originally developed for cross-hole geometries, which allow the recording of waves scattered with wide scattering angles (Pratt 1999). This provides a suitable framework to reconstruct the large to intermediate wavelengths of the subsurface, a difficult non-linear issue in reflection seismology. Later on, the potential of using frequency-domain FWI to build high-resolution velocity model from long-offset wide-aperture surface data was demonstrated with realistic synthetic and real data case studies, by for example, Ravaut *et al.* (2004) and Plessix *et al.* (2012) at the oil exploration scale, by for example, Operto *et al.* (2006), Bleibinhaus *et al.* (2007) at the deep crustal scale and by for example, Brenders & Pratt (2007b) at the lithospheric scale. When the acquisition design allows the recording of scattered waves over a broad range of scattering angles, the temporal frequencies and the scattering angle have redundant control on the wavenumber coverage in the model space. This prompted Pratt & Worthington (1990), Pratt (1999), Sirgue & Pratt (2004) and Brenders & Pratt (2007a) to design computationally efficient frequency-domain FWI algorithms by limiting the inversion to a few discrete frequencies, such that the redundancy of the wavenumber coverage was reduced. When a limited number of discrete frequencies need to be modelled for a large number of sources, frequency-domain seismic modelling, which reduces to the resolution of a large and sparse system of linear equations per frequency with multiple right-hand sides (each right-hand side being a source), can be performed efficiently with Gauss elimination techniques at least for 2-D problems because the expensive part related to the lower-upper decomposition of the impedance matrix is independent of the source, and hence is performed only once per frequency (Marfurt 1984; Stekl & Pratt 1998; Brossier *et al.* 2008).

Teleseismic acquisition leads to a quite different scattering-angle illumination than controlled-source seismology, because sources are a sparse set of up-going compressional plane waves, which impinge the base of the lithospheric target located below the array of receivers, instead of a dense line of point sources located near the surface. This raises the following issues that we would like to address hereinafter through a realistic synthetic case study:

(i) What is the resolving power of teleseismic frequency-domain FWI according to the limited scattering-angle illumination provided by a coarse set of incident plane waves? In relation to this, which part of the scattered wavefield carries the most resolving information? Can this information be extracted?

(ii) Can the *P*- and *S*-wave speeds be reliably reconstructed from incident compressional plane waves and multicomponent data?

(iii) Does efficient frequency-domain FWI based on hierarchical inversions of a few discrete frequencies apply to teleseismic configurations?

(iv) Which receiver spacing allows spatial aliasing to be prevented for the teleseismic bandwidth according to the resolving power of the FWI?

(v) According to the low-frequency content of teleseismic plane-wave sources and the acquisition geometry, which state of information do we need in the initial model to converge towards the global minimum of the misfit function?

Our parametric analysis is limited to 2-D geometries. Although the 2-D assumption puts limitations for applications on real teleseismic data because of the obliquity of the incident plane waves, we believe that our exhaustive parametric analysis can yet provide useful guidelines for future applications of 3-D FWI on teleseismic data. Although FWI of teleseismic data should clearly evolve towards 3-D geometries to take advantage of a significant number of useful events, we describe in Appendix A an approximate correction of the obliquity that should allow to apply 2-D elastic FWI of teleseismic data, when a sufficient number of incoming plane waves illuminate the lithospheric target with a moderate obliquity.

In the first part of the present study, we review the key features of the elastic FWI algorithm that we use. In Appendix B, we indicate how the seismic modelling engine implemented with a first-order discontinuous Galerkin frequency-domain method is interfaced with the inversion to compute the gradient of the misfit function with the adjoint-state method. Secondly, we review the key differences between controlled-source and teleseismic acquisitions, and discuss their potential influence on FWI. In the third section, we present a parametric analysis of elastic FWI of teleseismic data with a synthetic case study, for which a complex sedimentary model, that was originally designed to assess seismic imaging methods in exploration seismology, has been scaled to lithospheric dimensions. This parametric analysis aims to address all of the above-mentioned issues. We conclude this study with a discussion of the main results and with a review of some potential difficulties associated with the application of FWI to real teleseismic data.

2 METHOD

2.1 Frequency-domain elastic plane-wave seismic modelling

2.1.1 Discretization of the elastodynamic equations

In this study, seismic modelling is performed by solving the 2-D isotropic *P–SV* velocity–stress equation using a *hp*-adaptive frequency domain discontinuous Galerkin method (Brossier *et al.* 2008, 2010a; Brossier 2011). The 2-D *P–SV* velocity–stress equation in isotropic media is given by

$$
\begin{aligned}
-\mathrm{i}\omega\rho v_x &= \frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xz}}{\partial z} + f_x \\
-\mathrm{i}\omega\rho v_z &= \frac{\partial \sigma_{xz}}{\partial x} + \frac{\partial \sigma_{zz}}{\partial z} + f_z \\
-\mathrm{i}\omega\sigma_{xx} &= (\lambda + 2\mu)\frac{\partial v_x}{\partial x} + \lambda\frac{\partial v_z}{\partial z} \\
-\mathrm{i}\omega\sigma_{zz} &= \lambda\frac{\partial v_x}{\partial x} + (\lambda + 2\mu)\frac{\partial v_z}{\partial z} \\
-\mathrm{i}\omega\sigma_{xz} &= \mu\left\{\frac{\partial v_x}{\partial z} + \frac{\partial v_z}{\partial x}\right\},
\end{aligned}
\tag{1}
$$

where $[v_x(\omega, \mathbf{x}), v_z(\omega, \mathbf{x})]$ and $[\sigma_{xx}(\omega, \mathbf{x}), \sigma_{zz}(\omega, \mathbf{x}), \sigma_{xz}(\omega, \mathbf{x})]$ denote the particle velocities and stresses, respectively. A pure complex imaginary number is denoted by i. The coefficients $\lambda(\mathbf{x})$ and $\mu(\mathbf{x})$ are the Lamé parameters, $\rho(\mathbf{x})$ is the density, ω is the angular frequency, and $f_x(\mathbf{x})$ and $f_z(\mathbf{x})$ are the horizontal and vertical components of external point forces. This equation can be written in matrix form as

$$
\mathbf{A}[\omega, \mathbf{m}(\mathbf{x})]\mathbf{u}(\omega, \mathbf{x}) = \mathbf{s}(\omega, \mathbf{x}), \tag{2}
$$

where $\mathbf{u}$ denotes the velocity–stress wavefield, $\mathbf{s}$ is the source, and $\mathbf{A}$ is the so-called impedance matrix resulting from the discretization of the eq. (1). In this study, we solve the system 2 with the massively parallel sparse direct solver MUMPS (MUMPS-team 2011).

The lithospheric model can be discretized on a unstructured triangular mesh, which implies that the size of the element can be locally adapted to the medium properties (the so-called *h* adaptivity). Furthermore, the interpolation order of the shape functions can vary from one element to the next one, this property being referred to as the *p* adaptivity. The 3-D extension of the discontinuous Galerkin method in the time domain was developed on unstructured tetrahedral meshes by Etienne *et al.* (2010). In the framework of lithospheric imaging, triangular or tetrahedral unstructured meshes are useful to accurately represent complex topographies and the Earth's curvature, which is of the order of 4 km for a 450-long profile. We use a nodal formulation of the discontinuous Galerkin method, based on Lagrange polynomials of order 0, 1 or 2 (referred to as P0, P1 and P2, respectively) and centred fluxes (Hesthaven & Warburton 2008). The medium properties are piecewise constant per element in our current implementation. The number of nodes per element is 1, 3 and 6 for the P0, P1 and P2 interpolations, respectively. The P0 interpolation, which turns out to be equivalent to the finite-volume method, requires to use regular equilateral mesh for accurate simulations with elements of size 1/10 of the minimum wavelength (Brossier *et al.* 2008). Alternatively, P2 interpolation requires elements of size 1/4 of the local wavelength (Brossier 2011). The P1 interpolation can be used on unstructured meshes with elements of size 1/10 of the local wavelength, but does not generally provide distinct advantages relatively to the P0 or P2 interpolations. The P1 interpolation order can, however be combined with the P0 interpolation in specific environments such as shallow-water environments for accurate representation of the sea bottom and accurate positioning of sources and receivers in the near surface (Prieux *et al.* 2011). Optimal meshing strategies for frequency-domain discontinous Galerkin seismic modelling are discussed in Brossier *et al.* (2010a) and Brossier (2011).

Absorbing boundary conditions at the base of the lithospheric model and along its vertical edges are perfectly matched layers (Berenger 1994), while a free-surface boundary condition is used on the top of the model.

2.1.2 Implementation of plane-wave sources

In the teleseismic configuration, there is no source excitation within the lithospheric target, and hence the right-hand side in eq. (2) is zero. The source is a plane wave generated by a distant earthquake, which impinges on the base of the target with an arbitrary propagation direction. We implement the plane-wave source with a scattered-field formulation (Taflove & Hagness 2000, p. 220), which has already been used for teleseismic modelling by Pageot *et al.* (2009) and Roecker *et al.* (2010). This study is limited to 2-D modelling for plane-wave incidence. Two-and-a-half modelling was proposed by Takenaka & Kennett (1996) and Roecker *et al.* (2010) to account for the obliquity of the incident plane waves with respect to the vertical plane defined by the receiver network. Their method, which relies on the second-order wave equation for particle velocities, would be too computationally intensive in the frequency-

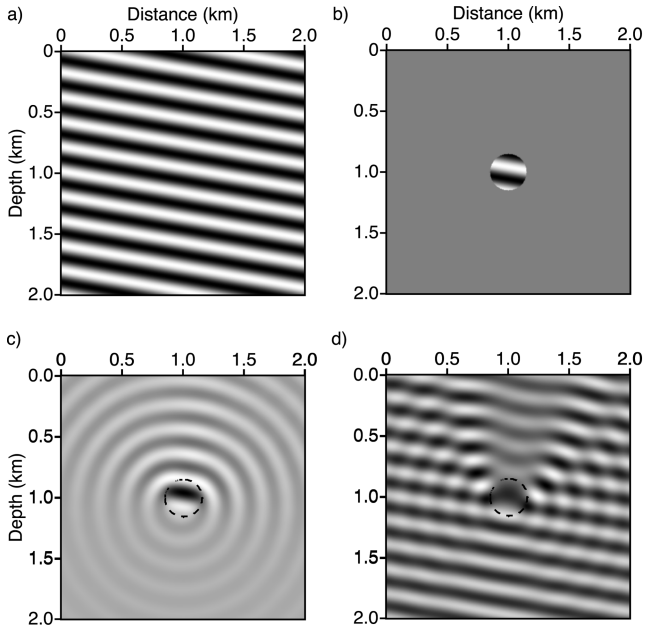

Figure 1. Plane-wave source modelling with the scattered-field formulation. The four main steps of the approach are illustrated for a homogeneous acoustic background model with an inclusion. (a) Analytical monochromatic plane-wave solution in the homogeneous background model. (b) Monochromatic scattering source. (c) Monochromatic scattered wavefield. (d) Total monochromatic wavefield obtained by summing the wavefields shown in (a) and (c).

domain when applied to first-order velocity–stress wave equation, because it would require to compute nine velocity–stress wavefields instead of two (when the second-order wave equation is considered, three particle-velocity wavefields in the 2.5-D approximation are computed instead of two in the 2-D approximation). Instead, we propose to approximately account for the obliquity of the plane waves through a local correction of the wave speed and propagation direction. This approximate correction is described in Appendix A.

The linearity of the wave equation allows us to assume that the total wavefield $\mathbf{u}$ can be written as the sum of an incident wavefield $\mathbf{u}_b$ computed in a background model $\mathbf{m}_b$ and a scattered wavefield $\mathbf{u}_s$. The incident wavefield satisfies the equation

$$\mathbf{A}_b\mathbf{u}_b = \mathbf{0}, \tag{3}$$

where $\mathbf{A}_b(\mathbf{m}_b, \omega)$ is the impedance matrix computed in the background model $\mathbf{m}_b$. Substituting $\mathbf{u}$ by $\mathbf{u}_b + \mathbf{u}_s$ and zeroing $\mathbf{s}$ in eq. (2), and taking the difference with eq. (3), gives the equation that is satisfied by the scattered wavefield

$$\mathbf{A}\mathbf{u}_s = -(\mathbf{A} - \mathbf{A}_b)\,\mathbf{u}_b. \tag{4}$$

By analogy with eq. (2), the scattered wavefield $\mathbf{u}_s$ is computed in the model $\mathbf{m}$ using internal source excitation, which is non-zero at the positions where $\mathbf{m}$ differs from $\mathbf{m}_b$. Once the scattered wavefield has been computed, the total wavefield is simply computed by the summation of $\mathbf{u}_s$ and $\mathbf{u}_b$. Notice that this method is exact, that is, it does not rely on linearization, as the forward-problem operator $\mathbf{A}$ on the left-hand side of eq. (4) is built from the true model and not from the background model. This approach is illustrated in Fig. 1, where the monochromatic full wavefield (Fig. 1d) is computed in a homogeneous background model containing a circular inclusion. The scattering source, the spatial support of which is limited to the inclusion, is shown in Fig. 1(b), and this generates the scattered wavefield shown in Fig. 1(c).

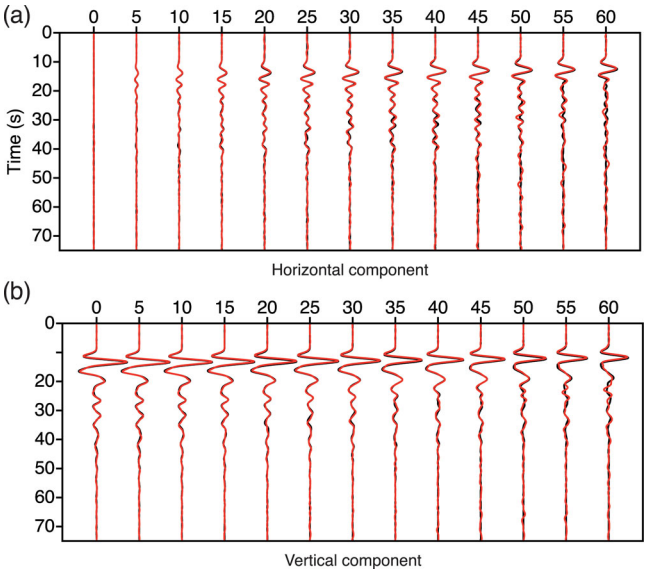

Figure 2. Validation of the scattered-field method against the reflectivity method. Direct comparison between seismograms computed with the reflectivity method (black line) and the scattered-field formulation (red line) in the three-layer model (Table 1) for the radial (a) and vertical (b) components.

Any background model can be used, provided that the incident wavefield can be computed efficiently. In this study, we use a homogeneous half-space with a flat free surface, for which analytical solutions are known. Semi-analytical methods, such as the reflectivity method (Kennett 1983), or the discrete-wavenumber boundary-integral methods (Gaffet & Bouchon 1989; Gaffet 1995) can also be methods of choice if layers, topography (including at the free surface) and the Earth's curvature need to be introduced into the background model.

The implementation of the scattered-wavefield formulation in the discontinuous Galerkin method, eq. (4), is validated against the reflectivity method in Fig. 2, for a three-layer medium with a free surface on top of the model and for different propagation directions (Table 1). The frequency bandwidth ranges between 0 and 0.5 Hz. The source signature is a Ricker wavelet of mean frequency 0.15 Hz. Seismic modelling is performed on an unstructured P2 mesh with elements of size 1/4 of the local wavelength. We show a good agreement between the two simulations for incidence angles ranging between 0° and 60°, although some differences for incidence angles greater than 45° result from artificial reflections at the perfectly matched layer-medium interfaces (mesh refinement does not lead to match improvement, that supports that the above-mentioned mismatch is not related to numerical dispersion). These artefacts should not significantly hamper realistic teleseismic applications, for which the propagation direction of the incident plane waves generally ranges between 20° and 40° at 150 km in depth.

Table 1. Physical parameters of the three-layer velocity model used to compute seismograms with the reflectivity method and the scattered-field formulation developed in this study.

Three-layered model				
Layer	Thickness (km)	V_p (m s^{-1})	V_s (m s^{-1})	ρ (kg m^{-3})
1	5	3000	1800	2000
2	35	6000	3500	2800
3	40	8000	4600	3000

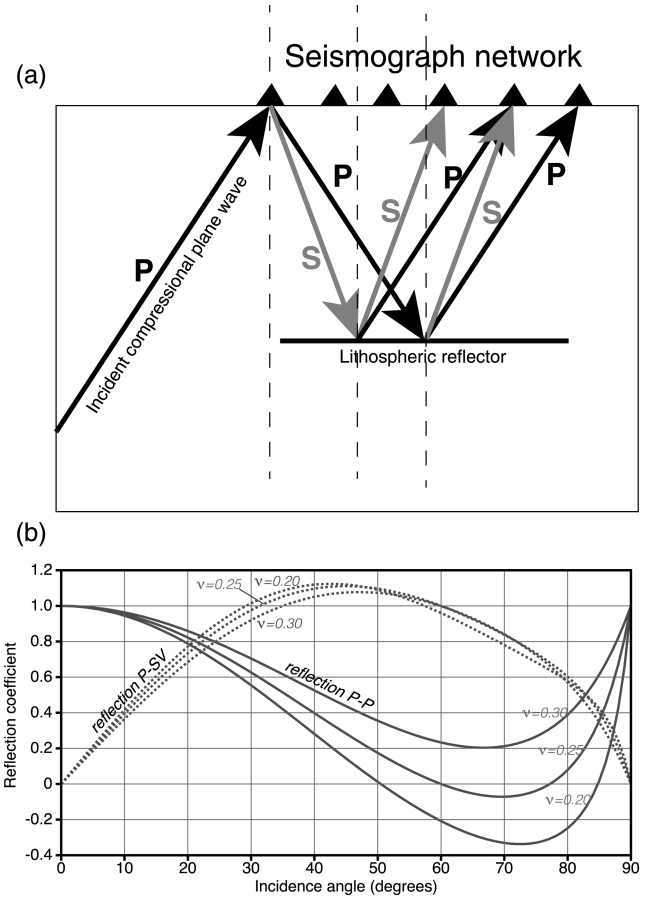

Figure 3. (a) Illustration of a teleseismic experiment, and the corresponding scattering modes. (b) *P–P* and *P–SV* reflection coefficients at the free surface versus the incidence angle of an up-going *P* wave. The curves are plotted for three Poisson ratios, as indicated.

In the present study, we consider only incident *P* waves, which generates *P–P* and *P–SV* reflections at the free surface. Note that no surface waves are generated in a homogeneous half-space for an incident compressional plane wave. The *P–P* and *P–SV* reflections from the free surface can be viewed in the framework of FWI as secondary monodirectional *P*- and *S*-wave sources, which continuously sample the surface of the model. These sources generate wavefields that propagate downwards in the lithospheric model before being reflected by the subsurface discontinuities and recorded at the surface by the receiver arrays (Fig. 3a). The free-surface *P–P* and *P–SV* reflection coefficients, which control the strength of these secondary sources, are shown in Fig. 3(b) for three different Poisson ratio representative of realistic subsurface geological media (Aki & Richards 2002). For incidence angles between 20° and 40°, the *P–P* and *P–SV* reflection coefficients in terms of displacement amplitude ratios are significant, and vary between 0.3 and 0.9 and between 0.7 and 1.1, respectively.

2.2 Frequency-domain elastic FWI

We review here the main features of our FWI algorithm (Brossier 2011) in terms of optimization and regularization that can have a significant impact on the results as we shall show in the parametric analysis. We also present in the Appendix B a versatile approach to compute the gradient of the misfit function with the adjoint-state method from self-adjoint operators, whatever the modelling engine that is used for seismic modelling. FWI is recast as a local optimization where a norm of the data residual vector is minimized iteratively around an initial model (Tarantola 1984; Pratt 1999; Virieux & Operto 2009). We use the least-squares norm of the data misfit augmented with a regularization term that aims to penalize the roughness of the difference between the model $\mathbf{m}$ and a prior model $\mathbf{m}_{prior}$ through the weighting operator $\mathbf{W}_{m_i}$

$$\mathcal{C}(\mathbf{m}) = \frac{1}{2}\Delta\mathbf{d}^{\dagger}\mathbf{W}_d\Delta\mathbf{d} + \frac{1}{2}\sum_{i=1}^{N_p}\lambda_i\left(\mathbf{m}_i - \mathbf{m}_{prior_i}\right)^{\dagger}\mathbf{W}_{m_i}\left(\mathbf{m}_i - \mathbf{m}_{prior_i}\right), \quad (5)$$

where $\Delta\mathbf{d} = \mathbf{R}\mathbf{v}(\mathbf{m}) - \mathbf{d}_{\rm obs}$ denotes the complex-valued monochromatic data residual vector, the difference between the modelled particle-velocity wavefield sampled at the receiver positions through the sampling or detection operator $\mathbf{R}$ and the recorded data $\mathbf{d}_{\rm obs}$. The symbol † denotes the transpose conjugate of a matrix. The multiparameter subsurface model is denoted by $\mathbf{m} = \left(\mathbf{m}_1, ..., \mathbf{m}_{N_p}\right)$, where N_p denotes the number of parameter classes to be updated during FWI. In this study, the parameter classes are the *P*- and *S*-wave speeds. Therefore, $N_p = 2$. Data preconditioning can be applied through the weighting matrix $\mathbf{W}_d$, which weights each component of the data misfit vector. For surface controlled-source acquisition, an amplitude gain with source–receiver offset can be used to strengthen the contribution of long-offset data in the misfit function (Operto *et al.* 2006). This data weighting does not apply to the teleseismic configuration we consider, and we will use the identity for $\mathbf{W}_d$. The scalar hyperparameters λ_i control the respective weights of the data-space and model-spaced misfit functions in eq. (5). Their value can be adapted to each parameter class.

The local minimization of the misfit function at iteration k in the vicinity of the model $\mathbf{m}_k$ gives the Newton descent direction $\mathbf{p}_k$ as

$$\mathbf{p}_k = -\left[\frac{\partial^2 C(\mathbf{m}_k)}{\partial\mathbf{m}^2}\right]^{-1}\frac{\partial C(\mathbf{m}_k)}{\partial\mathbf{m}}. \quad (6)$$

The second and first derivatives of the misfit function in the right-hand side term of eq. (6) are referred to as the Hessian and the gradient of the misfit function, respectively. The updated model $\mathbf{m}_{k+1}$ is related to the initial model $\mathbf{m}_k$ and the descent direction $\mathbf{p}_k$ by the expression

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta\mathbf{m}_k = \mathbf{m}_k + \gamma_k\mathbf{p}_k, \quad (7)$$

where the step length γ_k defines the amount of descent in the direction $\mathbf{p}_k$ and $\Delta\mathbf{m}_k$ is the model perturbation. The estimation of γ_k, which is required by the local quadratic approximation of the non-linear misfit function underlying eq. (6), is performed by line-search through parabolic fitting of the misfit function (e.g. Ravaut *et al.* 2004).

The expression of the Newton descent direction as a function of the sensitivity or Fréchet derivative matrix $\mathbf{J}$ is given by

$$\mathbf{p}_k = \Re\left[\widehat{\mathbf{W}}_m^{-1}\mathbf{J}_k^{\dagger}\mathbf{W}_d\mathbf{J}_k + \widehat{\mathbf{W}}_m^{-1}\left(\frac{\partial\mathbf{J}_k^T}{\partial\mathbf{m}^T}\right)\left(\Delta\mathbf{d}_k^*...\Delta\mathbf{d}_k^*\right) + \Lambda\right]^{-1} \Re\left[\widehat{\mathbf{W}}_m^{-1}\mathbf{J}_k^T\mathbf{W}_d\Delta\mathbf{d}_k^* + \Lambda\left(\mathbf{m}_k - \mathbf{m}_{prior}\right)\right], \quad (8)$$

where Λ is a block diagonal damping matrix, given as

$$\Lambda = \begin{pmatrix} \lambda_1\mathbf{I}_M & ... & \mathbf{0} \\ ... & ... & ... \\ \mathbf{0} & ... & \lambda_{N_p}\mathbf{I}_M \end{pmatrix}, \quad (9)$$

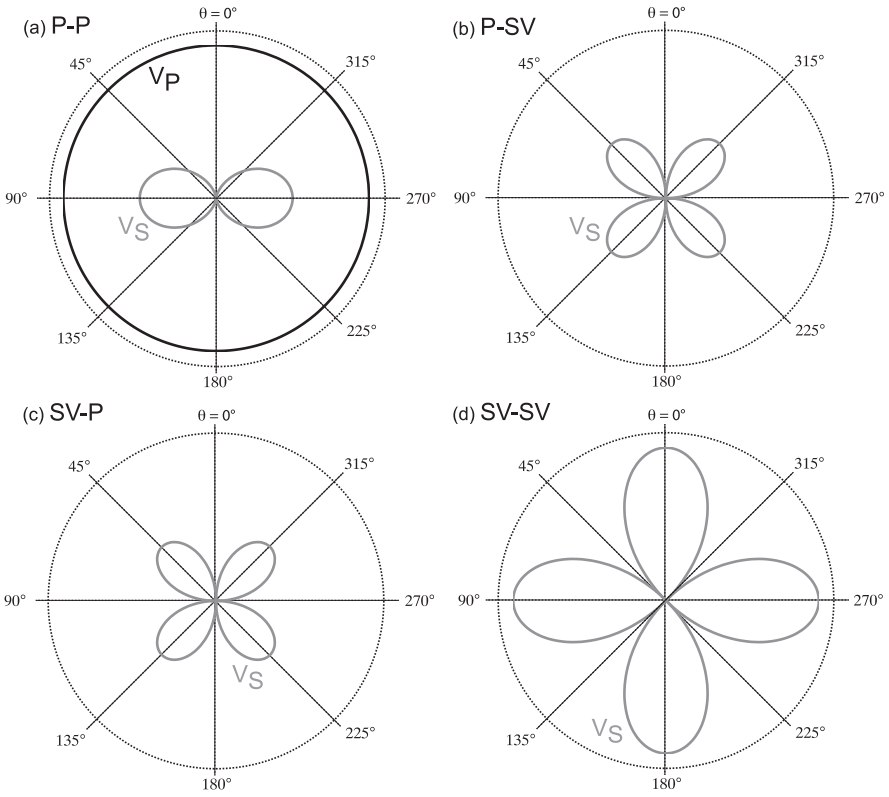

Figure 4. Scattering pattern of V_P (black line) and V_S (grey line) diffractors for the four scattering modes *P–P* (a), *P–SV* (b), *SV–P* (c) and *SV–SV* (d). The scattering patterns are plotted as a function of the scattering angle θ (Fig. 5a). Note that the V_P diffractor generates scattering only for the *P–P* mode. The scattering patterns were derived analytically in the framework of the ray+Born approximation (Forgues & Lambaré 1997).

and $\mathbf{I}_M$ is the identity matrix of dimension M, where M denotes the number of nodes in the computational mesh. The transpose of a matrix, the conjugate of a complex number, and the real part of a complex number are denoted by ^{T}, $*$ and $\Re$, respectively. The matrix $\widehat{\mathbf{W}}_m$ is a $N_p \times N_p$ block diagonal matrix, where each block is formed by the $\mathbf{W}_{m_i}$ matrices. The descent direction is written for one frequency in eq. (8). The simultaneous inversion of multiple frequencies is implemented by summing the contributions of each of the frequencies.

2.2.1 *Sensitivity of multicomponent data to V_P and V_S*

The coefficients of the sensitivity matrix **J** in eq. (8) correspond to the values of the so-called partial derivative wavefields at the geophone positions. These partial derivative wavefields are zero-lag correlated with the data residuals to form the gradient of the misfit function. These partial derivative wavefields satisfy the wave equation for virtual sources $\mathbf{f}_j$ (Pratt *et al.* 1998, their eqs 15 and 16):

$$\mathbf{B}\left(\frac{\partial \mathbf{v}}{\partial m_j}\right)_k = \mathbf{f}_j = -\frac{\partial \mathbf{B}}{\partial m_j}\mathbf{v}_k, \tag{10}$$

where **v** denote the particle-velocity wavefields and **B** the second-order forward modelling operator for particle velocities (Appendix B). These partial derivative wavefields represent the wavefields scattered by the missing heterogeneities in the initial model. The sparse matrix $\frac{\partial \mathbf{B}}{\partial m_j}$ represents the scattering (or radiation) pattern of the virtual secondary source located at position m_j, which generates the partial derivative wavefield $\frac{\partial \mathbf{v}}{\partial m_j}$. As the radiation patterns of the virtual source controls the amplitude of the partial derivative wavefields, they give clear insights into the influence of the selected parameter on the data as a function of the scattering angle, and the trade-off between parameters when multiple classes of parameters are jointly updated. The influence of the parameter on the data as a function of the scattering angle gives, in turn, some insight into the resolution with which the parameter can be imaged: the scattered wavefield associated with large and small scattering angles are sensitive to long and short wavelengths of the subsurface, respectively.

The radiation patterns of the *P* and *S* velocities (denoted by V_P and V_S, respectively) were computed analytically for the *P–P, P–SV, SV–P* and *SV–SV* modes in the framework of the high-frequency ray+Born approximation by Forgues & Lambaré (1997, Fig. 4). A first conclusion is that only the *P–P* mode is sensitive to the V_P perturbation. For this mode, the radiation pattern of V_P is isotropic, and hence a broadband reconstruction of V_P is expected (Fig. 4a, black line). The other modes are not impacted by the V_P perturbations (Figs 4b–d). The influence of the V_S parameter on the *P–P* mode is of smaller amplitude and spans a narrower range of intermediate scattering angles (Fig. 4a, grey line). The union of the radiation patterns of the V_S parameter for the *P–SV, SV–P* and *SV–SV* modes spans the full range of scattering angles (Figs 4b–d, grey lines). Therefore, a broadband reconstruction of V_S is expected provided that the inversion manages to exploit all of the scattering modes. A limited trade-off between V_P and V_S is expected during their reconstruction, because the two classes of parameter have a dominant influence in the data for different scattering modes. Of note, the density is not involved during the inversion in this study.

2.2.2 *Computing the gradient of the misfit function with the adjoint-state method*

Instead of explicitely forming the sensitivity matrix **J** in eq. (8), we compute the gradient of the misfit function with the

adjoint-state method (Lions 1968; Chavent 1974; Tarantola 1984; Plessix 2006; Chavent 2009). We use a second-order forward-modelling operator **B** for particle velocities to derive the gradient of the misfit-function, although seismic modelling is performed from the first-order velocity–stress wave equation. This allows us to manipulate self-adjoint operator and save memory during the gradient computation. Our approach is described in Appendix B in the framework of the scattered-field formulation.

2.2.3 Hessian approximation

The inverse of the full Hessian in eq. (8) contains three terms. The first aims to correct the gradient for linear effects, such as the limited bandwidth of the source, the limited spread of the acquisition geometry, and the geometrical spreading of the data. The second term of the Hessian accounts for double-scattering effects in non-linear inverse problems (Pratt *et al.* 1998). The third regularization term damps the deconvolution action of the first two terms of the Hessian, to improve the conditioning of the Hessian matrix.

We use the quasi-Newton limited-memory Broyden–Flechter–Goldfarb–Shanno (L-BFGS) optimization algorithm to solve eq. (8) (Nocedal 1980; Nocedal & Wright 1999). The L-BFGS algorithm recursively computes an approximation of the product of the inverse of the Hessian with the gradient, from a few gradients and a few solution vectors from the previous iterations. In what follows in the present study, we will highlight the improvements of the teleseismic FWI results that are achieved through the estimation of the Hessian performed using the L-BFGS optimization. As an initial guess of the inverse of the Hessian, we use a diagonal approximation of the approximate Hessian (the linear term) damped by the Λ matrix,

$$\mathbf{H}_0 = \mathrm{diag}\left(\widehat{\mathbf{W}}_m^{-1}\mathrm{diag}\left\{\mathbf{J}^{(k)^\dagger}\mathbf{W}_d\mathbf{J}^{(k)}\right\} + \Lambda\right)^{-1}. \tag{11}$$

In the matrix Λ, we use the same value of the damping factors λ_i for V_P and V_S, eq. (5). This value is defined as a fixed percentage (1 per cent) of the maximum coefficient of the diagonal Hessian.

2.2.4 Regularization

In the present study, we use $\mathbf{m}_{prior} = \mathbf{m}_k$, which allows us to cancel out the regularization term in the gradient of the misfit function in eq. (8). The operator $\widehat{\mathbf{W}}_m^{-1}$ is a smoothing operator that is implemented with a 2-D Gaussian function, the vertical and horizontal correlation lengths are defined as fractions of the P and S wavelengths. The same value of the fraction is used for V_P and V_S. eq. (8) shows that the Tikhonov regularization reduces to a smoothing of the Hessian and of the gradient when $\mathbf{m}_{prior} = \mathbf{m}_k$. This form of regularization can be viewed as preconditioning of the descent direction in an attempt to steer the inversion towards smooth solutions (Sirgue & Pratt 2004; Guitton *et al.* 2012).

3 SAMPLING, RESOLUTION AND NON-LINEARITY ISSUES IN TELESEISMIC FWI

Before showing the application of FWI to synthetic teleseismic data, it is worth reviewing some key differences between controlled-source and teleseismic acquisitions.

Resolving power of FWI and related strategies in exploration seismics

Crustal-scale and lithospheric-scale FWI was originally developed for wide-aperture controlled-source seismic exploration, where dense arrays of point sources are generally deployed near the surface (e.g. Pratt *et al.* 1996; Brenders & Pratt 2007a,b). The broad radiation pattern of the point sources allows incident waves to be generated with a broad range of incidence angles. These waves are scattered by the subsurface heterogeneities, with a wide range of scattering angles leading to a wide variety of arrivals in the wavefields, such as pre- and post-critical reflections, diving waves, head waves and diffractions. Long-offset multifold surface acquisition geometries are required to record all of these waves, and hence to really exploit the resolving power of FWI (Operto *et al.* 2006).

In the framework of diffraction tomography, Wu & Toksöz (1987) showed that the gradient of the misfit function with respect to a model parameter located at the position **x** can be viewed as a truncated inverse Fourier summation where the arguments of the basis functions are the wavenumber components injected into the subsurface model at the position **x** (see also Sirgue & Pratt (2004, their eq. 11) for a discussion in the framework of FWI). The truncation of the Fourier series, which limits the resolving power of the imaging, is controlled by the source bandwidth, the scattering-angle bandwidth sampled by the source–receiver acquisition geometry, and the scattering pattern of the model parameters. In the more general framework of asymptotic inverse-scattering theory (Miller *et al.* 1987; Wu & Toksöz 1987; Lambaré *et al.* 2003), these wavenumber vectors are formed by the sum of the two slowness vectors $\mathbf{p}_s$ and $\mathbf{p}_r$ associated with the rays connecting the source and the receiver to the diffractor point **x**: $\mathbf{k} = (k_x, k_z) = k(\cos\phi, \sin\phi) = \omega(\mathbf{p}_s + \mathbf{p}_r)$ (Fig. 5a). These wavenumbers are related to the local wave speed c_0, the angular frequency ω, the scattering angle θ and the local direction of propagation of the incident wave ϕ_S by the expressions

$$k = \frac{2\omega}{c_0}\cos(\theta/2), \tag{12}$$

and

$$\phi = \phi_S + \frac{\theta}{2}. \tag{13}$$

The expression of the modulus of the wavenumber vector, eq. (12), shows that there is double control of the frequency and the scattering angle on the wavenumber coverage. This wavenumber coverage can be strongly redundant in the directions ϕ, for which the acquisition geometry continuously samples a broadband of scattering angles. In other words, two close frequencies contribute to image two bands of wavenumbers that significantly overlap. In reflection seismics, this typically occurs for vertical wavenumbers (Sirgue & Pratt 2004). This prompted Pratt & Worthington (1990) and Pratt (1999) to reduce this vertical wavenumber redundancy by limiting the inversion to a few discrete frequencies, and hence design computationally efficient algorithms. Sirgue & Pratt (2004) defined a frequency interval in homogeneous media that eliminates this vertical wavenumber redundancy: the highest vertical wavenumber mapped by one frequency should be equal to the smallest vertical wavenumber constrained by the next frequency (Fig. 5b, Sirgue & Pratt 2004). This condition leads to an increasing frequency interval as the frequency increases.

The second conclusion that can be drawn from eq. (12), is that the wide scattering angles associated with diving waves and wide-spread reflections and low frequencies control the reconstruction of the long wavelengths of the subsurface, while the short scattering angles associated with short-spread reflections and high frequencies control the reconstruction of the short wavelengths. This prompted Pratt & Worthington (1990) and Pratt (1999) to design multiscale

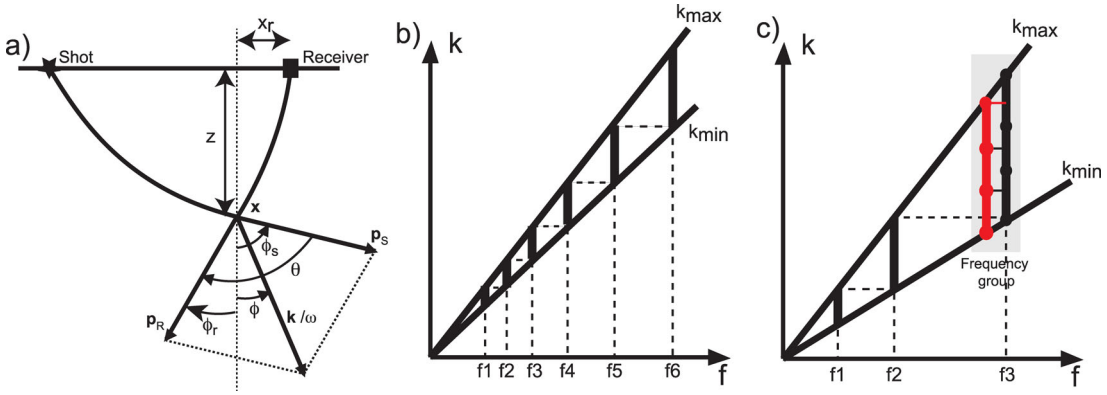

Figure 5. Sampling issues in sparse diffraction tomography. (a) Relationship between local wavenumber vector **k** at diffractor point **x** and angular frequency ω, scattering angle θ and local incidence angle ϕ. The slowness vectors associated with the rays connecting the source **s** and the receiver **r** to the diffractor **x** are denoted by $\mathbf{p}_s$ and $\mathbf{p}_r$, respectively. (b, c) Vertical wavenumber coverage versus frequency for a reflection from a horizontal interface [adapted from Sirgue & Pratt (2004)]. The solid black lines represent the minimum and maximum vertical wavenumbers spanned by one frequency. (b) Case of a narrow illumination of incidence angles. This scenario corresponds to a set of earthquakes with similar back-azimuth. A small frequency interval is required to guarantee continuous sampling of the wavenumber spectrum. (c) The case of a wide but coarse illumination of incidence angles. This scenario corresponds to the wide-spread distribution of a few earthquakes with distinct backazimuths. The circles represent wavenumbers, which are coarsely sampled within the wavenumber band spanned by one frequency. Simultaneous inversions of close multiple frequencies should help to prevent undersampling of the wavenumber band (red circles).

FWI that proceeds hierarchically from the low frequencies to the higher frequencies, to reduce the non-linearity of the FWI. A second nested level of data preconditioning can be viewed by injecting progressively shorter scattering angles into the inversion, through time windowing or time damping, which can be combined with offset windowing (Shipp & Singh 2002; Brossier *et al.* 2009a; Wang & Rao 2009).

All these manipulations of the data are designed to mitigate the non-linearity of the inverse problem, to have a better chance of converging on the global minimum.

Sampling issues in teleseismic FWI

Teleseismic acquisitions lead to relatively different subsurface illumination: the sources are a limited number of up-going plane waves instead of a dense line of point sources on the surface. This can strongly impact on the resolving power of FWI and lead to spatial aliasing.

As the gradient of the misfit function is discretized in the frequency domain, spatial heterogeneities are periodized in the spatial domain with a period of $1/\Delta k$, where Δk denotes the local sampling interval of the wavenumber in a given spatial direction ϕ. The sampling interval Δk should be sufficiently small to prevent periodization of the subsurface heterogeneities within the lithospheric domain of interest. Some examples of such periodization artefacts were shown in frequency-domain pre-stack-depth migration by Mulder & Plessix (2004), and for lithospheric-scale FWI by Pageot *et al.* (2010, their fig. 3). As it is not possible to control the distribution of sources in teleseismic imaging, it is likely that the frequency sampling that is used during teleseismic FWI will be a key issue to prevent spatial aliasing. On this basis, and according to eq. (12), it is worth remembering that wavenumbers that are mapped in the subsurface model are linearly related to frequency.

A first possible teleseismic configuration corresponds to a few earthquakes, that arrive at the base of the lithospheric target with similar incidence angles. This will provide a narrow scattering-angle illumination, which will require the refining of the frequency sampling in FWI, to prevent gaps in the wavenumber spectrum each time the FWI proceeds with the next frequency, and to satisfy the criterion of Sirgue & Pratt (2004, Fig. 5b). This refining of the frequency sampling has a strong impact on the computational efficiency of frequency-domain modelling, which scales linearly with the number of frequencies.

A second possible scenario corresponds to a few widely spread teleseismic events that arrive at the base of the lithospheric target with distinct incidence angles. This will provide wide, but coarsely-sampled, scattering-angle illumination (Fig. 5c). In this case, the wavenumber band reconstructed by single-frequency inversion can be under-sampled, because of the limited number of sources. One strategy to refine the wavenumber sampling and to prevent these aliasing effects might be to simultaneously invert a subset of close frequencies, rather than a single frequency, during one step of the multiscale inversion.

In both cases, it is likely that the teleseismic configuration requires finer sampling of frequencies than in controlled-source seismology to perform reliable FWI.

Improving the resolving power of lithospheric FWI with second-order scattering

The incident plane-wave sources impinge upon the target from beneath and are recorded on the surface, leading to low-resolution tomography-like reconstruction associated with large scattering angles (Pratt *et al.* 1996). To overcome this resolution limitation, Bostock *et al.* (2001) proposed to use the surface reflections as secondary *P* and *SV* plane-wave sources. These secondary plane-wave sources can be viewed as a continuous line of monodirectional sources at the surface. They generate downgoing waves, which reflect from the lithospheric discontinuities for all of the scattering modes (*P–P, P–SV, SV–P, SV–SV*), before being recorded at the surface (Fig. 3a). Migration-like inversion of reflection arrivals increases the resolution of the imaging of both the *P* and *S* velocity structures, because shorter scattering angles are considered, compared to tomography-like reconstruction. In this context, the maximum resolution of the FWI achievable at a scattering point is half a wavelength, according to eq. (12), and it is reached only for a scattering angle of zero.

What receiver spacing do we need to prevent spatial aliasing?

The receiver sampling in teleseismic acquisitions is generally sparser than for controlled-source experiments, and hence it needs

to be assessed according to the frequency content of the source, to prevent spatial aliasing in all of the spatial directions.

A first condition for the receiver spacing can be derived from the radius of the Fresnel zone for a reflection of an incident downgoing plane wave source from a reflector at a depth z: $R = \sqrt{\lambda\, z}$ (Sheriff & Geldart 1995, p. 155). One receiver will contribute to image a segment of the reflector, the length of which is the diameter of the Fresnel zone. A continuous image of the reflector will be obtained if the segments imaged by two close receivers overlap. This gives the condition

$$\Delta x_r < 2\sqrt{\lambda\, z}. \tag{14}$$

As an illustrative example, for a shear-wave velocity of 3 km s^{-1}, a frequency of 0.5 Hz and a depth of 20 km, the minimum receiver interval should be of the order of 20 km. If this condition is not satisfied, acquisition footprints hamper the shallow part of the FWI model. Acquisition footprint in the framework of exploration geophysics are shown in Ben Hadj Ali *et al.* (2008) and Sirgue *et al.* (2010).

Horizontal aliasing will occur if the sampling interval of the horizontal wavenumbers resulting from the receiver spacing does not satisfy the Nyquist-Shannon sampling criterion.

If we assume a theoretical resolution of half a wavelength (this resolution would be achieved for the reflection of a horizontally propagating plane wave from a vertical reflector), the sampling theorem requires that

$$\frac{\lambda}{2} < 1/\left(2\,\Delta k_x\right). \tag{15}$$

According to eqs (12) and (13), we have

$$k_x = \frac{\omega}{c}\left[\sin(\phi_S) + \sin(\phi_r)\right]. \tag{16}$$

For a downgoing plane wave source in a homogeneous medium, we can easily infer the variation of k_x that would result from perturbation of the position of the receiver by differentiation of k_x with respect to ϕ_r and from the relationship $\tan(\phi_r) = x_r/z$,

$$\Delta k_x = \frac{2\pi}{\lambda z} cos^3(\phi_r)\Delta x_r, \tag{17}$$

where z denotes the depth of the scatterer, and x_r the horizontal offset between the scatterer and the receiver (Fig. 5a). The sampling theorem, eq. (15), is satisfied if

$$\frac{\lambda}{2} < \frac{\lambda\, z}{2\cos^3(\phi_r)\Delta x_r}, \tag{18}$$

which leads to,

$$\Delta x_r < \frac{z}{\cos^3(\phi_r)}. \tag{19}$$

The condition to be satisfied by the receiver sampling is close to that derived by Rondenay *et al.* (2005), who concluded from an analysis of the migration operator that aliasing becomes significant at depths smaller than twice the receiver spacing. For completeness, Rondenay *et al.* (2005) and Brenders & Pratt (2007a) also proposed that the incoming wavefield should be properly sampled by the receiver array at the surface to avoid spatial aliasing. This leads to the condition that the receiver spacing for teleseismic experiment should be smaller than half the apparent wavelength of the incoming wavefield, while only the minimum of the source and receiver spacings should satisfy this condition for a controlled-source experiment.

Which initial model do we need?

Two other key differences between teleseismic and controlled-source experiments are the frequency content of the sources and the dimension of the target. In the Born approximation, the initial model should allow the prediction of the traveltimes of the scattered wavefields with an error lower than half the period of the monochromatic signal. If this condition is not satisfied, cycle skipping artefacts will arise (e.g. Virieux & Operto 2009, their fig. 7). To prevent cycle-skipping artefacts, the accuracy of the starting model must be improved or the number of propagated wavelengths must be reduced, because the relative traveltime error is inversely proportional to the number of propagated wavelengths (Pratt 2008; Virieux & Operto 2009),

$$\frac{\Delta t}{T} \geq \frac{1}{2N_\lambda}. \tag{20}$$

The most obvious remedy to reduce the number of propagated wavelengths is the introduction of low frequencies in the inversion. As an illustrative example, a recent application of FWI to low-frequency land data has shown that a simple vertical-velocity gradient model at the oil-exploration scale provides a sufficiently accurate initial model to perform FWI when the starting frequency is as low as 1 Hz (Plessix *et al.* 2012). Therefore, a key issue of the present study is the assessment of the kind of initial model needed by FWI to prevent cycle skipping artefacts according to the low frequency content of teleseismic sources and the size of the lithospheric target. In particular, we want to determine whether the available starting velocity models as 1-D velocity models extracted from global 1-D Earth models like the Preliminary Reference Earth Model (Dziewonski & Anderson 1981), AKI135 or IASP91 are sufficiently accurate to perform reliable FWI in the teleseismic bandwidth.

Low frequencies of teleseismic earthquakes are useful to relax the condition of accurate initial model. Another benefit of the teleseismic geometry compared to surface controlled-source geometries is to allow for the recording of transmitted waves, which travel from the bottom of the target to the surface where receivers are deployed. This setting is similar to cross-hole geometries, except that the source and receiver arrays are horizontal rather than vertical. This geometry together with the low frequencies of the sources provide a suitable framework to continuously sample the wavenumber of the lithospheric target as the FWI workflow evolves from the forward-scattered waves and the low frequencies to the backscattered waves and the high frequencies. This full scattering-angle illumination provided by teleseismic geometry contrasts with the narrower illumination provided by surface short-spread reflection experiments for which the intermediate wavelengths belongs to the null space of the FWI (Jannane *et al.* 1989).

As in controlled-source experiments, multiscale strategies can be viewed to hierarchically invert subdata sets with an increasing resolving power.

4 PARAMETRIC ANALYSIS OF TELESEISMIC FWI

In this section, we attempt to answer the issues raised in the previous section with a realistic synthetic example.

4.1 The lithospheric SEG/EAGE overthrust model

To perform our parametric analysis, we chose a vertical section of the 3-D complex onshore Society of Exploration Geophysics (SEG)/European Association of Geoscientists and Engineers (EAGE) velocity model (Aminzadeh *et al.* 1997), which was

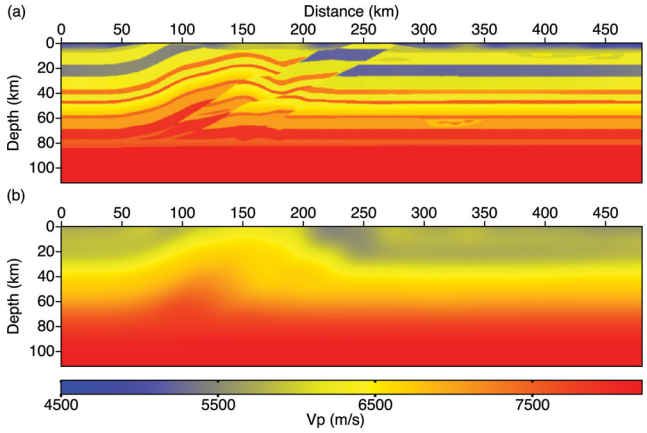

Figure 6. Lithospheric SEG/EAGE overthrust model (a) True and (b) starting *P*-wave velocity models. The *S*-wave velocity model is derived from the *P*-wave velocity model using a constant Poisson ratio of 0.25.

originally designed to assess seismic imaging methods in exploration seismology (Fig. 6a). The dimensions of the original model are 20 km in distance and 4.5 km in depth. The main features of the overthrust model are a weathered layer in the near surface, several buried palaeochannels, and two main thrust faults, which cross-cut the sedimentary cover. This sedimentary cover overlays a decollement level at 4 km in depth. An *S*-wave velocity model is built from the *P*-wave velocity model, assuming a constant Poisson ratio of 0.25. This velocity model was imaged by elastic FWI of surface wide-aperture data by Brossier *et al.* (2009a). The imaging was performed with and without free-surface effects in the data (i.e. surface waves). Accurate reconstructions of the *P* and *S* velocities were obtained from body waves. The inversion of surface waves in addition to body waves significantly increases the non-linearity of the inversion, which requires two levels of data preconditioning to be considered, by frequency selection and time damping. Since surface waves should not have a strong impact on teleseismic FWI of incident body waves, we will use the results of the elastic FWI of body waves of Brossier *et al.* (2009a, their fig. 9) as a reference to assess the results of the imaging at the lithospheric scale.

Table 2 outlines the characteristic scales involved in seismic exploration and teleseismic seismology for the overthrust case study. The model dimensions have been scaled by a factor of around 24. The *P*-wave velocities range from 2.36 to 6 km s^{-1} in the original overthrust model, compared to 4.7–8.1 km s^{-1} in the lithospheric model. The frequency bandwidth, which was used during FWI by Brossier *et al.* (2009a), is [1.7–7] Hz, against [0.05–0.4] Hz in the present study. Considering the mean wave speeds and the lowest frequency, in the lithospheric model, the *P* and *S* wavelengths are of the order of 12.8 and 7.4 km, respectively, and in the original model, of the order of 0.6 and 0.34 km, respectively. These wavelengths give rough estimates of the spatial resolution that should be achieved by reflection FWI. The ratio between the characteristic wavelengths propagated in the original and lithospheric models are of the same order of magnitude as the ratio between the characteristic dimensions of the two models, which prompts us to conclude than the resolution gain provided by FWI relative to conventional approaches like traveltime tomography should be equivalent at the seismic exploration and teleseismic scales.

It is also worth estimating the minimum number of propagated wavelengths in both models, to determine the potential risk of cycle skipping in FWI. This number of propagated wavelengths is computed by considering the lowest frequency, the mean wave speeds, and a propagation distance that corresponds to wide-angle reflections with an offset of 20 and 200 km for the original and lithospheric models, respectively. In the lithospheric case, we augment the surface-to-surface reflection path with a one-way path from the bottom of the model to the surface, to account for the reflection from the free surface. We use a propagation distance for the wide-angle reflection of 22 and 445 km for the sedimentary and the lithospheric cases, respectively. In this case, the number of propagated wavelengths is almost three times lower in the lithospheric case than in the seismic exploration case. Therefore, the low frequency content of the teleseismic sources should help to relax the requirement of an accurate model for FWI. Notice that this ratio of three was estimated by considering a sedimentary model as a seismic exploration target. Other scales involves the deep crustal scale and wide-angle surveys (Operto *et al.* 2006; Bleibinhaus *et al.* 2007, 2009; Kamei *et al.* 2012) and the near surface scale (Romdhane *et al.* 2011; Bleibinhaus & Hilberg 2012). At the deep crustal scale, the target size is of the order of one-to-two hundred kilometers in length and few tens of kilometres in depth and the frequency bandwidth typically ranges from 3 to 15 Hz. In this case, the number of propagated wavelengths will be significantly higher than at the teleseismic scale and, hence represent a challenging experimental setup to perform reliable FWI. In contrast, the number of propagated wavelengths at the near-surface scale can be smaller by one order of magnitude compared to the teleseismic scale, because of the limited dimensions of the target. Therefore, the challenging issues of the near surface scale are more related to the extreme heterogeneity of the near surface and the exploitation of the high-frequency content of the source.

Table 2. Characteristic scales involved in exploration seismology (e.g. Brossier *et al.* 2009a) and in teleseismic imaging. L_x, L_z (km): horizontal and vertical dimensions of the target. $V_{P\min}$, $V_{P\max}$ (km s^{-1}): minimum and maximum *P*-wave speeds. $V_{S\min}$ $V_{S\max}$ (km s^{-1}): minimum and maximum *S*-wave speeds. $f_{\min}$, $f_{\max}$ (Hz): frequency bandwidth. $\lambda_{P\min}$, $\lambda_{P\max}$ (km): minimum and maximum compressional wavelengths. $\lambda_{S\min}$, $\lambda_{S\max}$ (km): minimum and maximum shear wavelengths. $N_{\lambda_{P\min}}$, $N_{\lambda_{S\min}}$: number of propagated compressional and shear wavelengths for the starting frequency $f_{\min}$.

	Exploration scale	Lithospheric scale
L_x L_z (km)	20.0, 4.65	473, 110
$V_{P\min}$ $V_{P\max}$ (km s^{-1})	2.36–6	4.7–8.1
$V_{S\min}$ $V_{S\max}$ (km s^{-1})	1.36–3.45	2.71–4.68
$f_{\min}$ $f_{\max}$ (Hz)	1.7–7.0	0.05–0.5
$\lambda_{P\min} = V_{P\mathrm{mean}}/\omega_{\min}$ $\lambda_{P\max} = V_{P\mathrm{mean}}/f_{\max}$ (km)	0.6–2.46	12.8–128
$\lambda_{S\min} = V_{S\mathrm{mean}}/\omega_{\min}$ $\lambda_{S\max} = V_{S\mathrm{mean}}/f_{\max}$ (km)	0.34–1.41	7.39–73.9
$N_{\lambda_{P\min}} = L_z/\lambda_{P\max}$ $N_{\lambda_{S\min}} = L_z/\lambda_{S\max}$	9–16	3.5–6

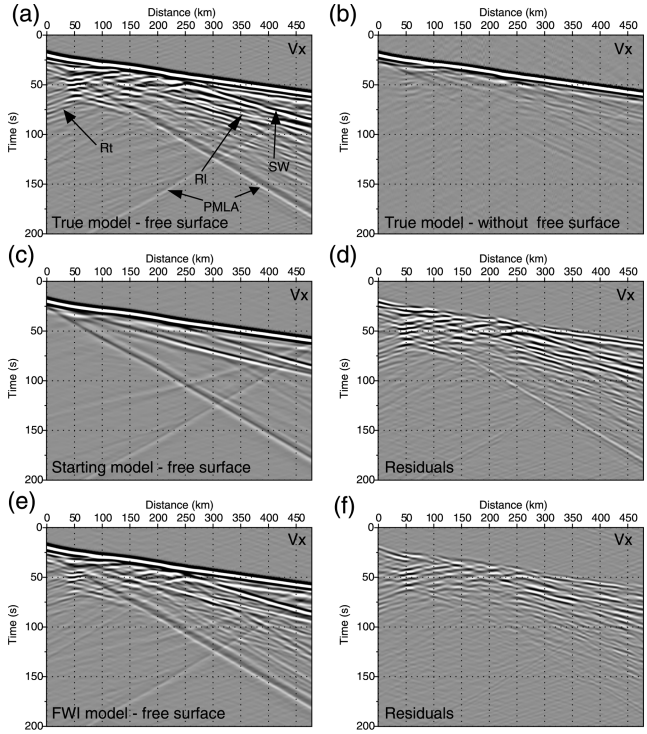

Figure 7. (a–b) Seismograms computed in the true V_P and V_S models: the radial component. The incidence angle of the up-going plane wave is –40°. A free-surface boundary condition (a) and an absorbing boundary condition (b) are used on top of the model. R_T: Doubly reflected phases from the free surface and from the thrusts. R_P: Doubly reflected phases from the free surface and from the horizontal layers in the right part of the model. *PMLA*: parasite reflections from the perfectly matched layer-medium interfaces. *SW*: a *P*–*S* diffraction in the near surface towards 250 km of distance. (c) As for (a), for seismograms computed in the initial model (Fig. 6b). (d) Residuals between seismograms shown in (a) and (c). (e) Same as (a) for seismograms computed in the FWI model of Figs 14(g) and (h). (f) Residuals between the seismograms shown in (a) and (e).

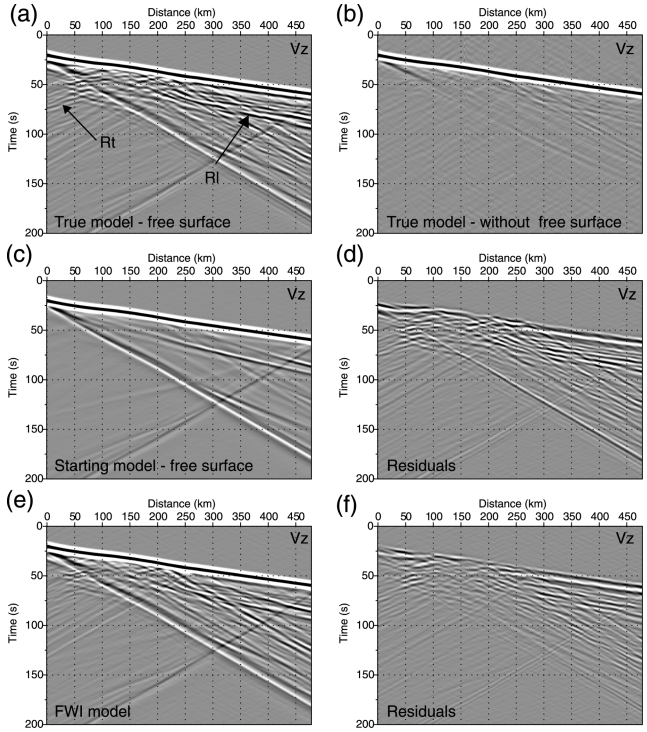

Figure 8. As for Fig. 7, for the vertical component.

4.2 Transmission versus reflection FWI of teleseismic data

Bostock *et al.* (2001) showed the need to consider all of the reflections and conversions from the lithospheric reflectors after a first reflection from the free surface. Seismograms computed in the true velocity models when free surface and absorbing boundary conditions are implemented on top of the model can be compared in Figs 7(a, b) and 8(a, b), for the radial and vertical geophones, respectively. The incidence angle of the plane-wave source is –40°. In seismograms computed with the free surface boundary condition, we show the doubly reflected phases from the free surface and the thrusts (Figs 7a, b and 8a, b, phase Rt). The traveltimes of these reflections decrease from the left-hand side to the right-hand side of Figs 7(a) and 8(a), as the thrust plane becomes closer to the surface. On the right-hand side of Figs 7(a) and 8(a), we show the doubly reflected phases from the free surface and the horizontal layers of the model (Figs 7a and 8a, phase Rl). We also interpret a phase on the horizontal component that originates at 250 km in distance and that propagates with a slower apparent velocity than the first-arrival as a *P*–*S* diffracted waves (Fig. 7a, SW). We apply FWI to the data sets computed without and with the free surface boundary condition on top of the model. The initial models are obtained by smoothing the true model with 2-D Gaussian functions of horizontal and vertical correlation lengths of 10 km (Fig. 6b). This resolution should mimic the resolution of first-arrival traveltime tomography, which is of the order of the first Fresnel zone width (Williamson 1991). The data set consists of 17 plane-wave sources with incidence angles ranging from −40° to 40° with a sampling rate of 5° (Table 3). The resulting phases are recorded by a dense network of 399 vertical and radial geophones deployed on the surface with a spacing of 1.2 km. Thirty-seven frequencies between 0.1 and 0.4 Hz were inverted sequentially, using the final model of one frequency inversion as the initial model for the next frequency inversion. We discretize the overthrust model with a regular equilateral mesh of size 1/10 of the minimum *S* wavelength and we perform seismic modelling with P0 interpolation. Unlike the simulation performed for the three-layer model (Fig. 2), we perform seismic modelling with a regular P0 mesh instead of a unstructured P2 mesh, because the building of the P0 impedance matrix is two-time faster than the P2 one, while the LU factorization and the resolution steps take a similar amount of time for the two meshes. This probably results because the mesher has more difficulties to exploit the *h* adaptivity when the subsurface models contain a large numbers of thin layers, as opposed to blocky models composed of few homogeneous layers (Brossier *et al.* 2010a). The same mesh was used for each hierarchical frequency inversions, although a mesh refinement could have been performed as the inversion was evolving towards higher frequencies. Computations were performed on 12 quadri-core processors with 8-Gygabytes of memory each. A LU factorization of the P0 impedance matrix typically takes 20 s and the substitution step to get the solution 8 s. Fifteen non-linear iterations were performed per frequency with L-BFGS optimization. The comparison between the final FWI V_P and V_S models inferred from the two data sets shows the dramatic resolution improvement that is achieved when free-surface reflections are involved in the inversion (Fig. 9). Figs 10(a–d) and 11(a–d) show a more detailed view of this resolution improvement along two vertical profiles extracted from the

Table 3. FWI experimental set-up. Nomenclature: Fig: figure number. f_0 (Hz), f_{max} (Hz): initial and final frequencies used in FWI. N_f: number of frequencies used in FWI. $Delta_R$(km): receiver spacing. N_s: number of plane-wave sources. Angles (°): range of incidence angles spanned by the plane-wave sources. *Op*: Optimization algorithm [L-BFGS (L) versus conjugate gradient (C)]. *BC*: Boundary condition on top of the model (A: absorbing; FS: free surface). *FG*: single-frequency (S) versus multi-frequency (M) hierarchical inversions. By multi-frequency inversion is meant successive inversions of overlapping frequency groups. E_P, E_S: error in percentage in the V_P and V_S FWI models.

Fig	f_0	f_{max}	N_f	Δ_R	N_s	Angles	*Op*	*BC*	*G*	E_P (%)	E_S (%)
9(a, b)	0.1	0.4	37	1.2	17	−40,+40	L	A	S	4.40	4.30
9(c, d)	0.1	0.4	37	1.2	17	−40,+40	L	FS	S	2.78	2.43
12(a, b)	0.1	0.4	37	1.2	17	−40,+40	C	FS	S	2.92	2.62
14(a, b)	0.1	0.4	7	1.2	17	−40,+40	L	FS	S	3.00	3.02
14(c, d)	0.1	0.4	19	1.2	17	−40,+40	L	FS	S	2.80	2.37
14(e, f)	0.1	0.4	19	1.2	17	−40,+40	L	FS	M	2.80	2.32
14(g, h)	0.1	0.4	37	1.2	17	−40,+40	L	FS	M	2.58	2.16
17(a, b)	0.1	0.4	19	1.2	5	−40,+40	L	FS	S	3.00	2.58
17(c, d)	0.1	0.4	37	1.2	9	−40,+40	L	FS	S	2.87	2.56
17(e, f)	0.1	0.4	37	1.2	17	−40,+40	L	FS	S	2.78	2.43
17(g, h)	0.1	0.4	37	1.2	8	−40,−20;+20,+40	L	FS	S	2.98	2.43
17 (i, j)	0.1	0.4	37	1.2	4	+20,+40	L	FS	S	3.42	3.30
17(k, l)	0.1	0.4	37	1.2	4	−40,−20	L	FS	S	3.41	3.21
18(a, b)	0.1	0.4	37	2.4	17	−40,+40	L	FS	S	2.80	2.43
18(c, d)	0.1	0.4	37	4.8	17	−40,+40	L	FS	S	2.82	2.51
18(e, f)	0.1	0.4	37	9.6	17	−40,+40	L	FS	S	2.85	2.57
18(g, h)	0.1	0.4	37	19.2	17	−40,+40	L	FS	S	3.29	3.06
18(i, j)	0.1	0.4	37	19.2	17	−40,+40	L	FS	S	2.73	2.46
18(k, l)	0.1	0.4	37	19.2	17	−40,+40	L	FS	S	3.61	3.81
19(a, b)	0.1	0.4	37	1.2	17	−40,+40	L	FS	S	2.86	3.61
19(c, d)	0.05	0.4	37	1.2	17	−40,+40	L	FS	S	2.75	3.61
19(e, f)	0.05	0.4	37	1.2	17	−40,+40	L	FS	S	10.0	10.1
19(g, h)	0.015	0.4	37	1.2	17	−40,+40	L	FS	S	8.03	5.56

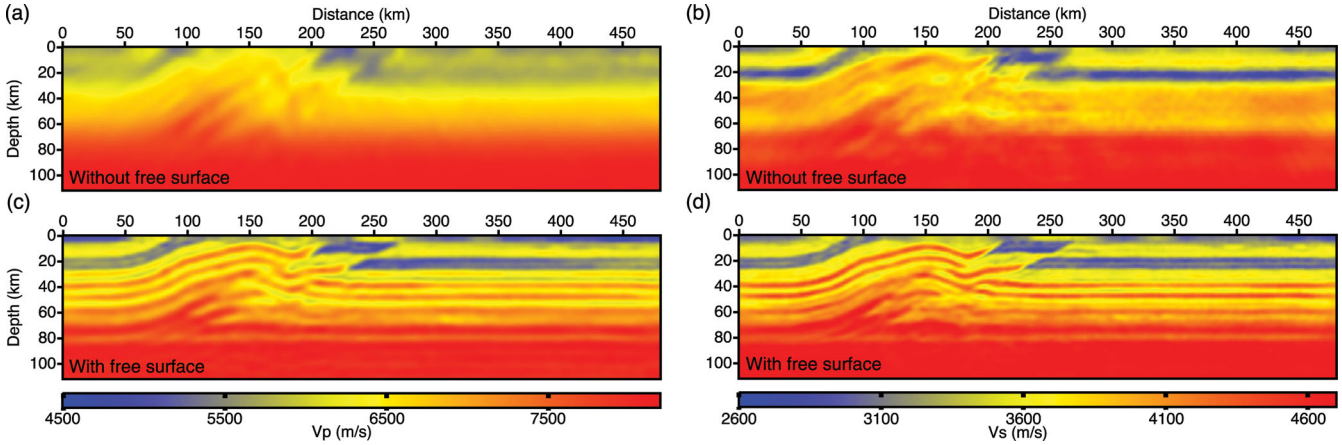

Figure 9. Importance of the free surface in teleseismic FWI. V_P (a) and V_S (b) FWI velocity models when an absorbing boundary condition on top of the model (infinite model) is used to compute the seismic wavefield in the true model and in the reconstructed models. (c, d) As for (a, b), but a free-surface boundary condition is implemented on top of the models. See Table 3 for the FWI set-up.

true models, the initial models and the final FWI models at 125 km (Figs 10a–d) and 325 km (Figs 11a–d) in distance.

4.3 Conjugate-gradient versus L-BFGS quasi-Newton optimization

The conjugate-gradient method is one of the most popular optimization algorithm to perform FWI (Mora 1987). Another optimization algorithm that is suitable for FWI is the BFGS algorithm, recast in a limited memory storage version by Nocedal (1980): L-BFGS. Brossier *et al.* (2009b) have shown how the L-BFGS optimization improves the reconstruction of the elastic overthrust model from land surface seismic data compared to the conjugate gradient algorithm. Comparison between the lithospheric FWI models inferred from the conjugate gradient (Figs 12a and b) and the L-BFGS (Figs 12c and d) algorithms confirms the improvements that are achieved when the action of the Hessian on the gradient is taken into account. The inversion set-up is the same as that used in the previous section with a free-surface boundary condition on top of

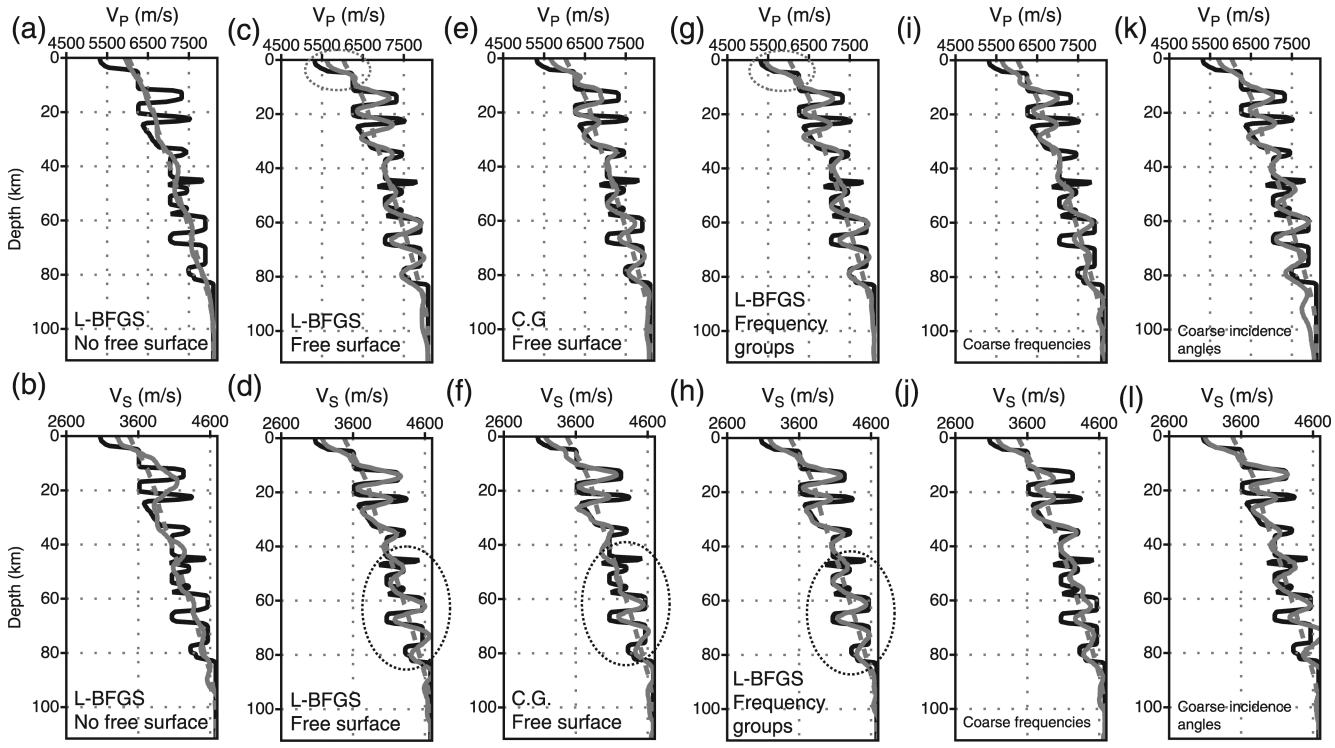

Figure 10. Vertical V_P (top panel) and V_S (bottom panel) profiles extracted at 125 km in distance from the true (black lines), the FWI (solid grey line) and the starting (dashed grey lines) models. (a, b), (c, d), (e, f), (g, h), (i, j), (k, l) V_P and V_S profiles extracted from the FWI models of Figs 9(a, b), 9(c, d), 12(a, b), 14(g, h), 14(a, b) and 17(a, b), respectively. The dashed ellipses in (c–g) delineate the shallow velocities, where the use of frequency groups provides a more reliable reconstruction compared to single-frequency inversion. The dashed ellipses in (d–f) delineate part of the velocity profile where the L-BFGS optimization provides more reliable focusing than the conjugate gradient algorithm.

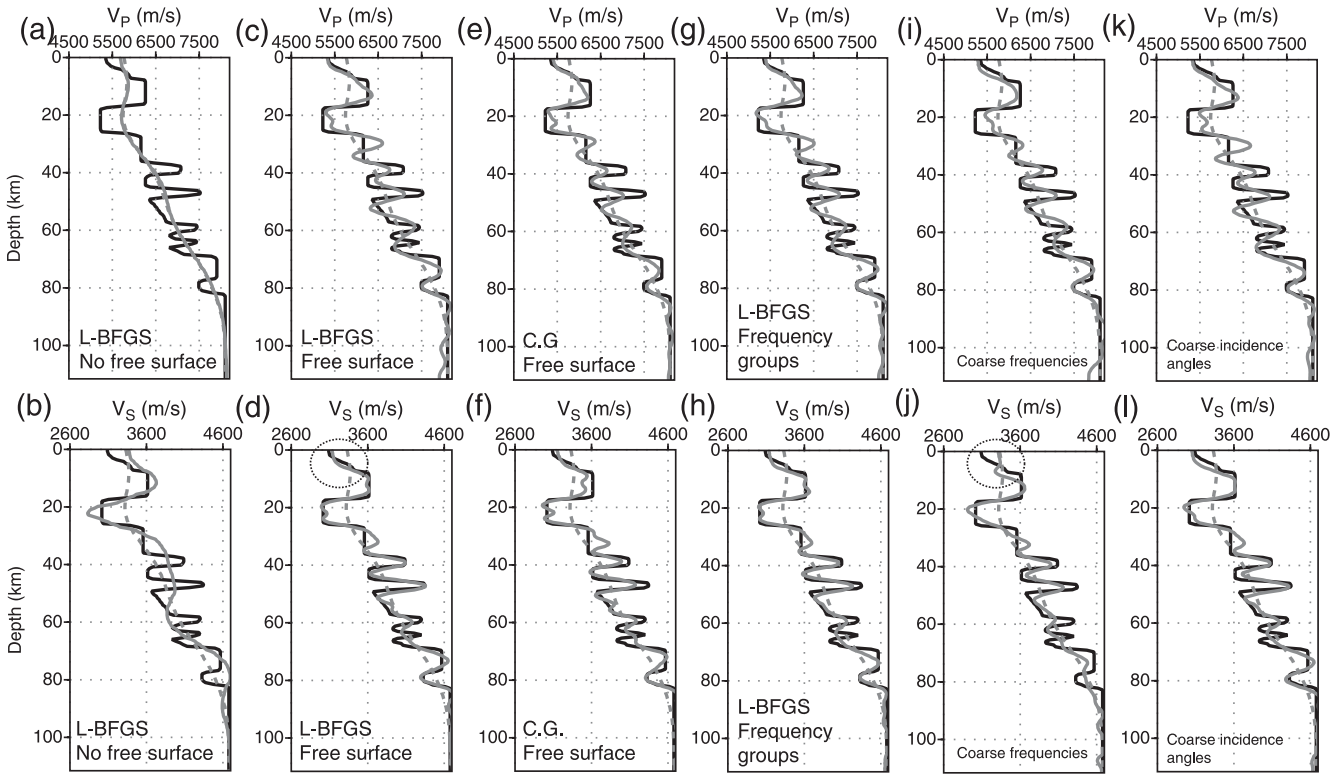

Figure 11. As for Fig. 10 for the vertical profiles located at 325 km in distance. The dashed ellipses in (d–j) delineate the shallow parts of the model, where the *S*-wave speed is not reconstructed correctly from a coarse subset of frequencies.

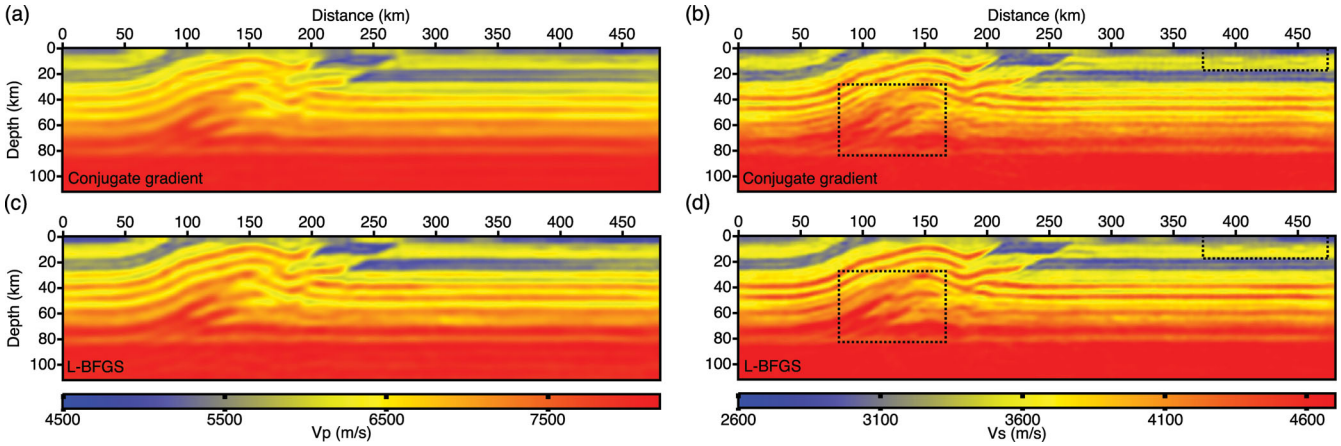

Figure 12. Optimization algorithm in FWI. (a, b) V_P (a) and V_S (b) FWI velocity models computed with a conjugate gradient algorithm. (c, d) As for (a, b) for the L-BFGS algorithm. Note the improved reconstruction of the thrusts and the shallow channels in the V_S model (dashed rectangles). See Table 3 for the FWI set-up.

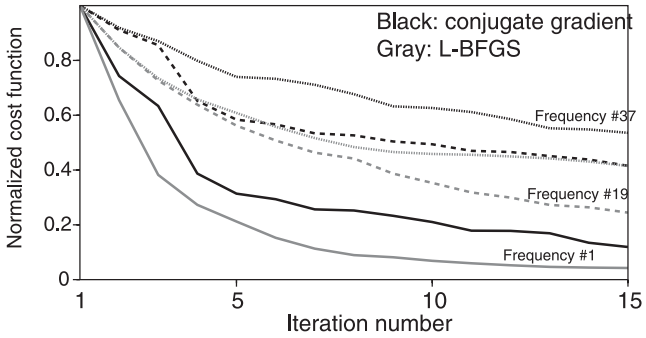

Figure 13. Optimization algorithm in FWI. Normalized misfit function versus iteration number for frequencies of 0.1, 0.25 and 0.4 Hz. The misfit function obtained with the conjugate gradient and L-BFGS are plotted in grey and black, respectively.

the model (Table 3). Fifteen non-linear iterations are performed with both the conjugate gradient and the L-BFGS optimization. Of note, the descent direction provided by the conjugate-gradient algorithm has been preconditioned by the diagonal terms of the approximate Hessian, eq. (13). The better focusing of the V_P model obtained with L-BFGS suggests an improved convergence rate, which is further confirmed by the misfit function plotted against the iteration number for three frequencies (Fig. 13). Moreover, the V_S model obtained with L-BFGS shows an improved signal-to-noise ratio in the thrust area and in the shallow part of the model, where horizontal aliasing effects have been efficiently reduced. The improved convergence rate achieved with L-BFGS is confirmed by vertical profiles extracted from the FWI models obtained with L-BFGS and the conjugate gradient algorithms: amplitudes of the velocity perturbations are reconstructed much better with the L-BFGS algorithm (Figs 10c–f and 11c–f).

In what follows for the present study, all of the results were obtained with L-BFGS, except when explicitly stated to the contrary.

4.4 Influence of frequency sampling

Designing a suitable subset of frequencies to prevent aliasing artefacts in FWI of teleseismic data is a key issue, according to the sparse and/or narrow illumination of scattering angles provided by a limited number of plane wave sources (Fig. 5). The results of FWI applied sequentially to an increasing number of discrete frequencies are shown in Figs 14(a–f). Seventeen plane-wave sources with incidence angles ranging from -40° to 40° with a sampling rate of 5° are considered for FWI, and lead to a broadband sampling of the incidence angles (Table 3). The V_P and V_S models obtained with 19 and 37 frequencies are of similar quality. A significant degradation of the results, which is more obvious in the V_S model because of shorter-propagated wavelengths, is shown when only seven frequencies are used. This suggests that a notch in the wavenumber spectrum is created each time the inversion progresses to the next frequency (Fig. 5b). Comparison with vertical profiles extracted from FWI models suggests that large errors can occur in the near surface when the frequency interval is too coarse (Figs 11d and j, dashed ellipse). We also gather the 37 frequencies into three overlapping frequency groups, [0.1–0.2] Hz, [0.2–0.3] Hz and [0.3–0.4] Hz, and apply successive inversions to these three frequency groups. The resulting V_S model (Fig. 14h) shows a slightly improved signal-to-noise ratio in the deep part of the model compared to that inferred from sequential inversion of single frequencies (Fig. 14f), while the V_P models inferred from the two inversions does not show significant differences (Figs 14e and g). This improvement is highlighted in vertical profiles extracted from the FWI V_S models obtained without and with frequency groups in Figs 10(d) and (h), dashed ellipse. We will see later that the benefit provided by frequency groups is more obvious when the receiver interval is increased.

We conclude that even if the plane-wave sources span over a broad range of incidence angles, the frequency interval used in teleseismic FWI should be significantly refined compared to those commonly used in efficient FWI of controlled-source data. As a comparative example, Brenders & Pratt (2007a) concluded that only four frequencies between 0.8 and 7 Hz are needed to build a lithospheric model of the *P*-wave speed from a dense surface acquisition.

We compute time-domain synthetic seismograms in the initial model and in the final FWI models inferred from the frequency-group inversion (Fig. 14g, h) to determine which part of the wavefield was matched during FWI (Figs 7c–f and 8c–f). These seismograms can be qualitatively compared with those computed in the true model (Figs 7a, b and 8a, b), while direct comparisons between the seismograms computed in the true model and in the

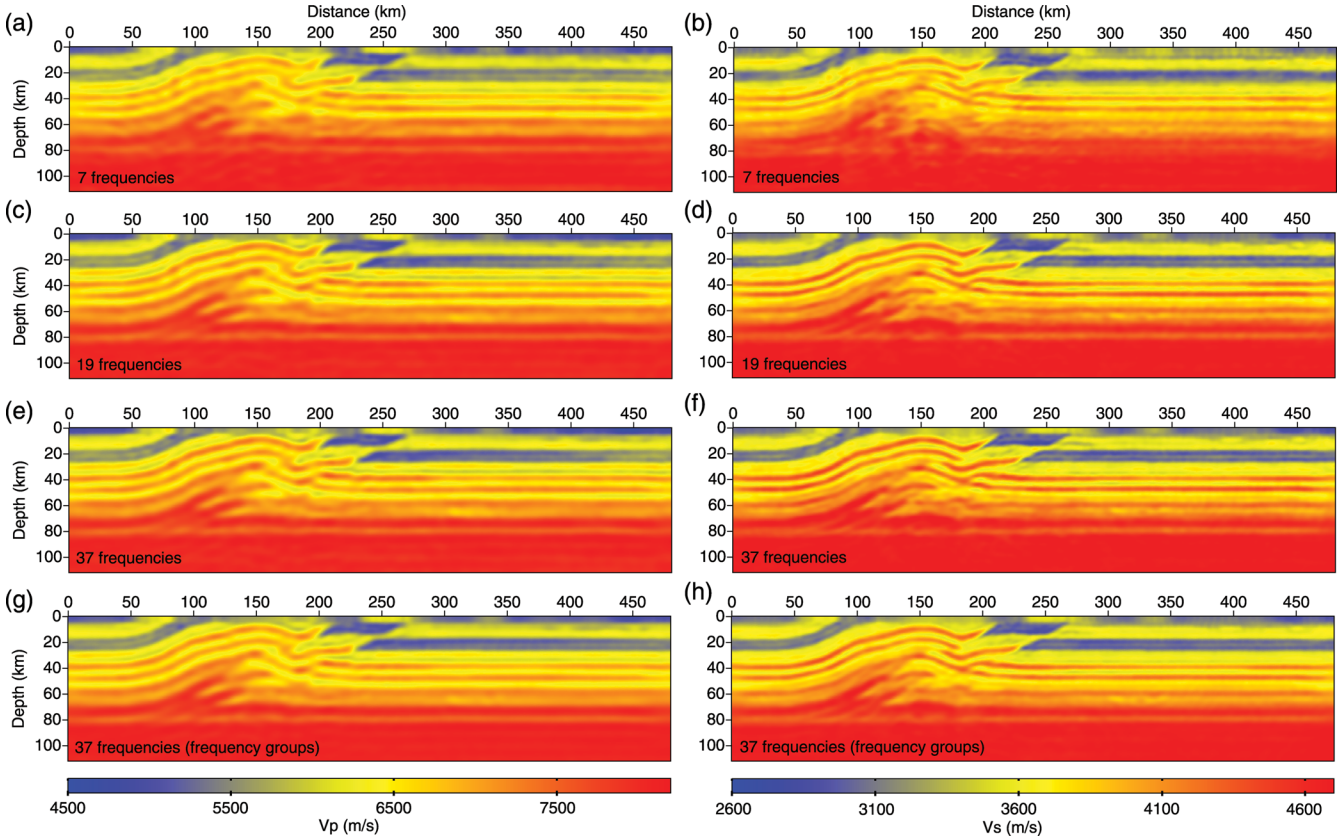

Figure 14. Impact of frequency sampling on teleseismic FWI. (a–f) Final V_P and V_S FWI models obtained by sequential inversions of single frequencies: 7(a, b), 19(c, d), 37(e, f) frequencies are inverted. (g, h) Successive inversions of three overlapping frequency groups, for the total of 37 frequencies involved in the inversion. Note the improved signal-to-noise ratio in the V_S model in (h) relatively to (f). See Table 3 for the FWI set-up.

initial and final FWI models are shown in Figs 15 and 16 for the horizontal and vertical geophones, respectively. We show that most of the residuals associated with the high-amplitude primary wavefield were cancelled out by FWI. Significant residuals associated with the low-amplitude doubly scattered wavefield remain. However, direct comparisons between the seismograms computed in the true model and in the final FWI models show that the phase of almost all of the arrivals are matched (Figs 15b and 16b), which suggests that most of the remaining residuals result from underestimated wavefield amplitudes due to an insufficient number of iterations.

4.5 Influence of incidence-angle sampling and bandwidth

We now determine the impact of the sampling of the incidence angles of the plane-wave sources (see Table 3 for the FWI set-up). Thirty-seven frequencies were inverted sequentially with the L-BFGS algorithm. The receiver spacing was 1.2 km. The incidence angles were uniformly sampled between -40° and $+40^\circ$. We show progressive degradation of the FWI models when the number of plane-wave sources decreases from 17 to five (Figs 17a–f, 10k, l and 11k, l). However, the FWI is clearly less sensitive to the incidence-angle sampling relative to the frequency sampling. This is consistent with the relationship between the wavenumber, the frequency and the scattering angle, eq. (12): the wavenumber is linearly related to the frequency, while it is related to a more slowly varying cosine function of the scattering angle.

We now consider a more realistic setting in terms of plane-wave coverage. First, we still consider a symmetric illumination from the right and from the left of the target, but we remove plane waves that propagate nearly vertically, as these arrivals would correspond to earthquakes located on the other side of the earth. The FWI V_P and V_S models obtained with 10 plane-wave sources, the incidence angles of which range between -40° and -20° and between $+20^\circ$ and $+40^\circ$, are shown in Figs 17(g) and (h), and they can be compared with those of Figs 17(e) and (f), for which 17 plane waves with incidence angles that continuously range between -40° and $+40^\circ$ are considered. We show overall degradation of the vertical resolution associated with a lack of vertical-wavenumber coverage. For example, the shallow channels on the right-hand side of the V_S models are not visible anymore, as well as some thrusts in the left-hand part of the model (Fig. 17h, dashed rectangle). The V_P model is affected to a lesser extent than the V_S counterpart, because the compressional wavelengths are higher than the shear counterparts. However, the degradation in resolution of the horizontal layers is clear on the right-hand side of the model, where the vertically propagating plane waves contribute efficiently to inject short vertical wavenumbers in the model (compare Figs 17e and g).

Secondly, we consider an even less favourable setting, where the plane waves are not symmetrically distributed any more: five plane waves arrive either from the left (Figs 17i and j) or the right (Figs 17k and l) of the lithospheric target, and cover a range of incidence angles between -40° and -20° and between 20° and 40°, respectively. As expected, the imaging of the thrusts is more significantly impacted

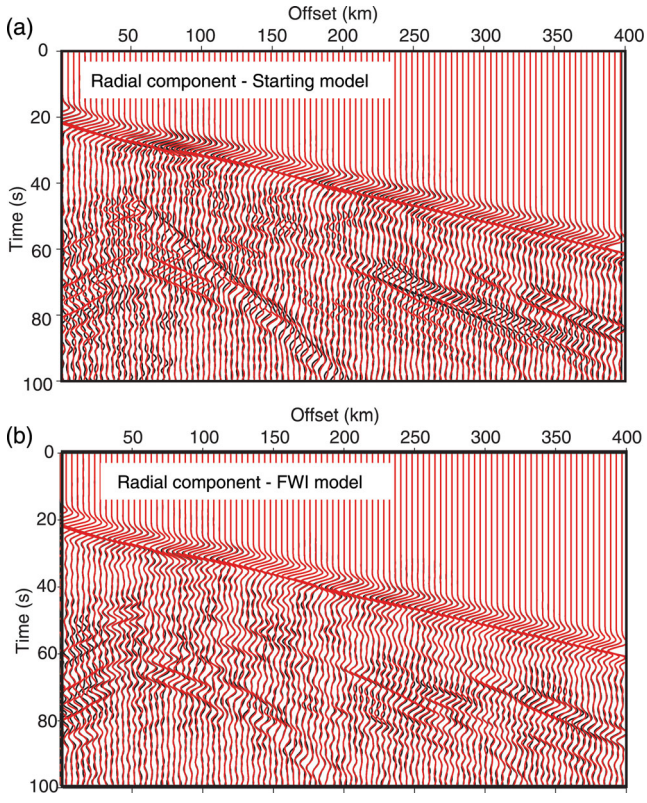

Figure 15. Radial component. (a) Direct comparison between seismograms computed in the true model (black) and in the initial model (red). (b) As for (a), but the red seismograms are computed in the FWI models of Figs 14(g) and (h). The seismograms are plotted with an automatic gain control to assess more easily the match of the phase of the secondary scattered wavefield.

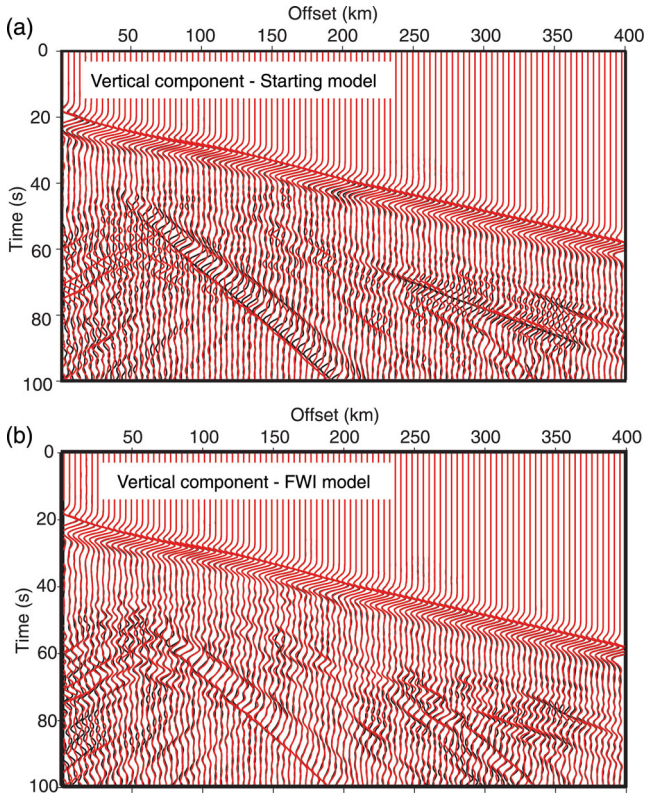

Figure 16. Vertical component. As for Fig. 15 for the vertical component.

when the plane waves come from the right, because the limited extent of the receiver array prevents the recording of reflections from the dipping thrusts propagating from right to left (compare Figs 17i–j and 17k–l). To a lesser extent, plane waves coming from the left are less favourable than those coming from the right to image the deep part of the horizontal layers on the right of the model. This might be related to the decreasing amplitudes of the wavefields as the number of propagated wavelengths increases, which makes the inversion less sensitive to these arrivals. It is worth noting that the shallow part of the model on the right-hand side appears to be less affected by aliasing artefacts and shows better reconstruction of a channel when the plane waves come from the left. This might arise because a *P*–*S* wave scattered near the middle of the model (Fig. 5a, SW) could have contributed to filling the gap in terms of wavenumber coverage, to improve the imaging of the first 10 km of the target.

4.6 Impact of receiver sampling

Even if modern dense networks of broadband stations allow data sets to be recorded that are potentially amenable to multichannel high-resolution seismic imaging methods such as FWI, receiver spacing remains a critical issue to prevent spatial aliasing according to the theoretical resolution of FWI. We applied FWI to several data sets computed with increasing receiver spacing, ranging from 2.4 km to 19.2 km (Fig. 18). Thirty-seven frequencies were inverted sequentially, and 17 plane-wave sources with incidence angles that ranged between -40° and $+40^\circ$ were involved in the inversion (Table 3). Horizontal artefacts started appearing in the shallow part for a receiver spacing of 9.6 km (Figs 18e and f), and became significant for a receiver spacing of 19.2 km in the full model (Figs 18g and h). Replacing the sequential inversion of single frequencies by the sequential inversion of overlapping frequency groups efficiently reduced the aliasing artefacts in the deep part of the model, without significant extra computational cost (Figs 18i and j). This supports the idea that these aliasing effects mainly results from the coarse sampling of the horizontal-wavenumber band that is injected into the subsurface model during each single-frequency inversion, and that the simultaneous inversion of several close frequencies can contribute to fill in this wavenumber band (Fig. 5c). Shallow artefacts with a periodic horizontal pattern of around 20 km however remains down to 20-km in depth in the V_S model of Fig. 18(j). These artefacts can be related to the radius of the Fresnel zone and are consistent with the condition of eq. (15) for a wave speed of 3 km s^{-1}, a frequency of 0.5 Hz and a depth of 20 km, that gives a radius of the Fresnel zone of 10 km and $\Delta x_r < 20$ km. For completeness, in Figs 18(k) and (l), we also show the results of FWI for a receiver spacing of 19.2 km when the frequencies are inverted sequentially and when the conjugate gradient optimization is used. The V_S model is significantly noisier than the V_S model obtained when L-BFGS optimization is combined with suitable management of frequencies during waveform inversion. This highlights on the one hand the importance of the Hessian in FWI, as a deconvolution operator, and on the other hand the importance of frequency management in multiscale FWI of coarse teleseismic data.

4.7 Which initial model and starting frequency do we need?

Building a reliable starting model for FWI is one of the most topical issue in exploration geophysics at present (Virieux & Operto 2009), even if the current trend is to design new acquisition devices that allow for the emission of low frequencies to satisfy the cycle-skipping

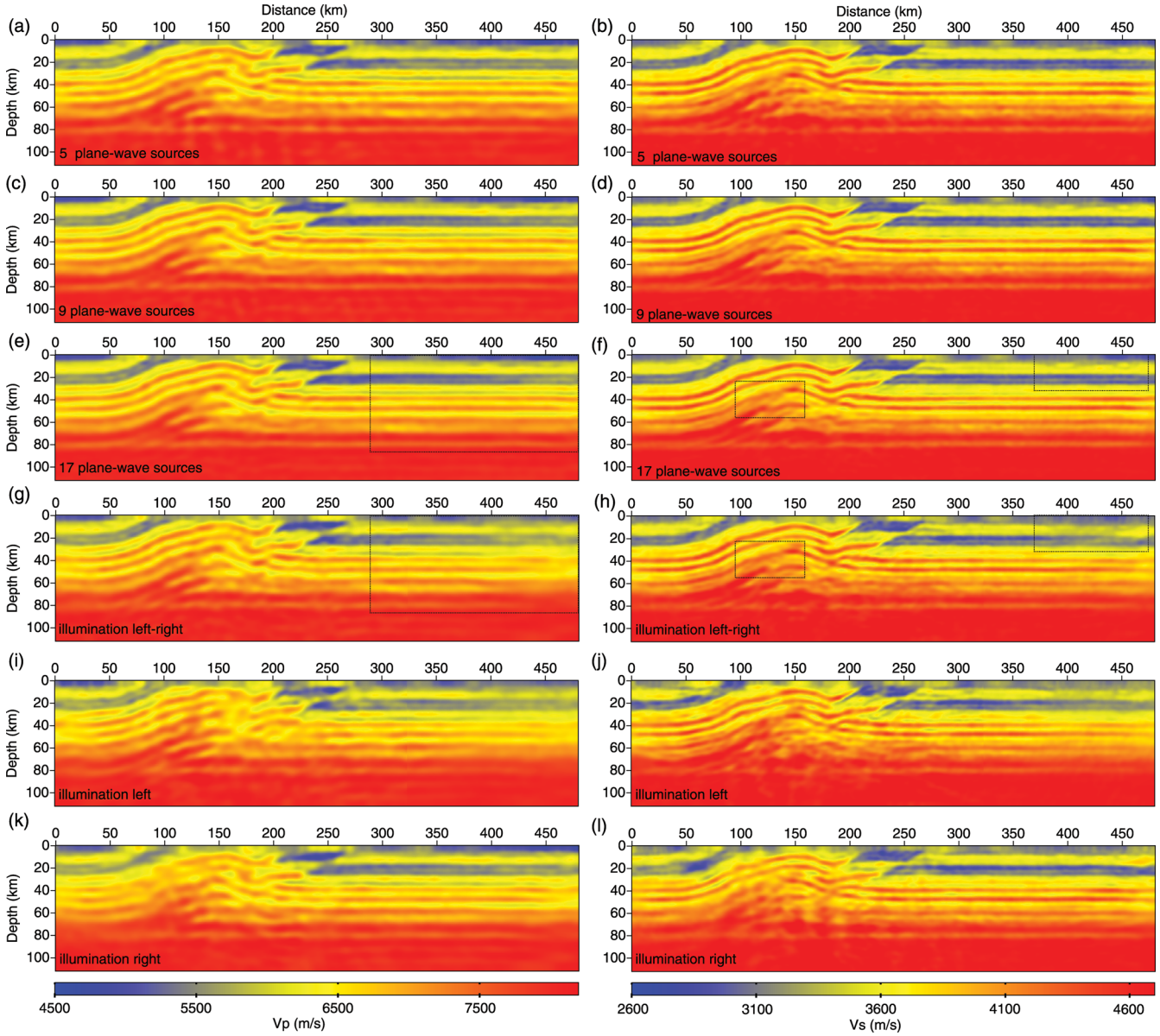

Figure 17. Impact of incidence-angle sampling and coverage. (a, b) V_P (a) and V_S (b) FWI models obtained for five plane-wave sources. Incidence angles range between −40° to +40° with an interval of 20°. (c, d) As for (a, b), for nine plane-wave sources (angle interval: 10°). (e, f) As for (a, b), for 17 plane-wave sources (angle interval: 5°). (g, h) As for (a, b), for 10 plane-wave sources, with incidence angles ranging between −40° and –20° and between +20° and +40°. (g, h) As for (a, b), for five plane-wave sources, with incidence angles ranging between −40° and −20°. (k, l) As for (a, b), for five plane-wave sources, with incidence angles ranging between +20° and +40°. See Table 3 for the complete FWI set-up.

criterion, eq. (20) (Soubaras & Whiting 2011; Plessix *et al.* 2012). In the teleseismic configuration, low frequencies are naturally provided by large-magnitude earthquakes. Therefore, it is worth investigating which kind of initial model guarantees reliable lithospheric imaging from teleseismic data by FWI. For this purpose, we consider now as the initial model a laterally homogeneous velocity gradient model, instead of the smooth background model of Fig. 6(b). We performed the sequential inversions of 37 and 46 frequencies associated with starting frequencies of 0.1 Hz (Figs 19a and b) and 0.05 Hz (Figs 19c and d), respectively (Table 3). We show a lack of small wavenumbers in the thrust area of the V_S model inferred from the velocity-gradient starting model and the initial frequency of 0.1 Hz: these small wavenumbers lacks in the velocity-gradient starting model and belong to the null-space when the initial frequency is 0.1 Hz. These small wavenumbers are injected in the FWI models when the starting frequency is decreased to 0.05 Hz. For this starting frequency of 0.05 Hz, the FWI V_P and V_S models inferred from the starting velocity-gradient model do not show significant differences with the FWI models inferred from the smooth background models and a starting frequency of 0.1 Hz (Figs 9c and d). This highlights the trade-off between the need for an accurate initial model and the need for low frequencies. Of note, the FWI models inferred from the smooth background model with starting frequencies of 0.05 Hz and 0.1 Hz do not show significant differences (not shown here). We conclude that reliable FWI of teleseismic data should be possible starting from a crude vertical-velocity gradient model. We also apply FWI for homogeneous starting models of *P*-wave and *S*-wave speeds of 6 km s^{-1} and 3.4 km s^{-1}, respectively

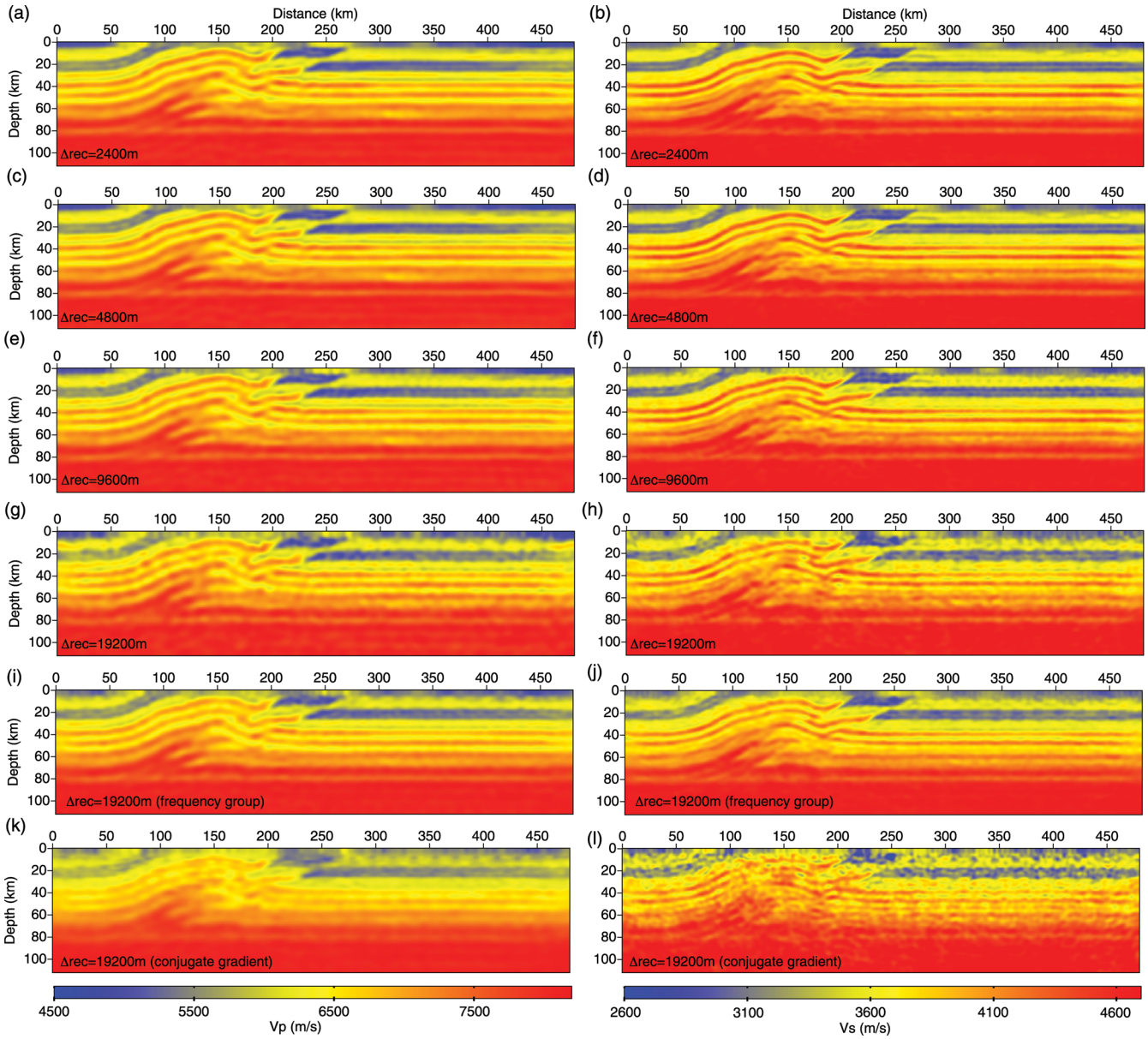

Figure 18. Impact of receiver spacing. V_P (left-hand panel) and V_S (right-hand panel) FWI velocity models obtained with a receiver interval of 2.4 km (a, b), 4.8 km (c, d), 9.6 km (e, f) and 19.2 km (g, h). Sequential inversions of 37 frequencies are performed (Table 3). (i, j) As for (g, h), but sequential inversions of three overlapping frequency groups are performed instead of sequential inversions of single frequencies. Note the significant reduction in the spatial aliasing. (k, l) As for (g, h), but the conjugate gradient algorithm is used instead of the L-BFGS.

(Figs 19e–h). In this case, the inversion fails to converge towards the correct solution, even for a starting frequency as low as 0.015 Hz (Figs 19g and h). The reason for this might be related to inaccurate amplitude modelling at the free surface, rather than to kinematic inaccuracies.

4.8 Resolution analysis by checkerboard tests

We conclude the present study by resolution analysis of FWI performed through checkerboard tests: the true lithospheric model to be reconstructed is formed by the initial background model (Fig. 6b), on which we superimpose a checkerboard perturbation model. Other possible approaches for resolution analysis of FWI consists of computing point-spread functions through an approximate estimation of the Hessian (Fichtner & Trampert 2011b). We seek to reconstruct the checkerboard, starting from the smooth background model, and following the same hierarchical inversion procedure as for the overthrust model imaging. We seek to highlight the intrinsic resolving power of FWI for the teleseismic frequency bandwidth. Therefore, we consider an ideal acquisition device with a receiver spacing of 1.2 km and a broad range of incident-angle illumination between -40° and $+40^\circ$.

The final FWI V_P and V_S models obtained for elements of dimension 19.2 km, 9.6 km and 4.8 km are shown in Fig. 20. Thirty-seven frequencies gathered in three overlapping frequency groups and 17 plane-wave sources were involved in the inversion. The inversion succeeds in reconstructing the 19.2 km elements for both the V_P and V_S models (Figs 20a and b), which is consistent because the size of the elements is greater than the minimum P and

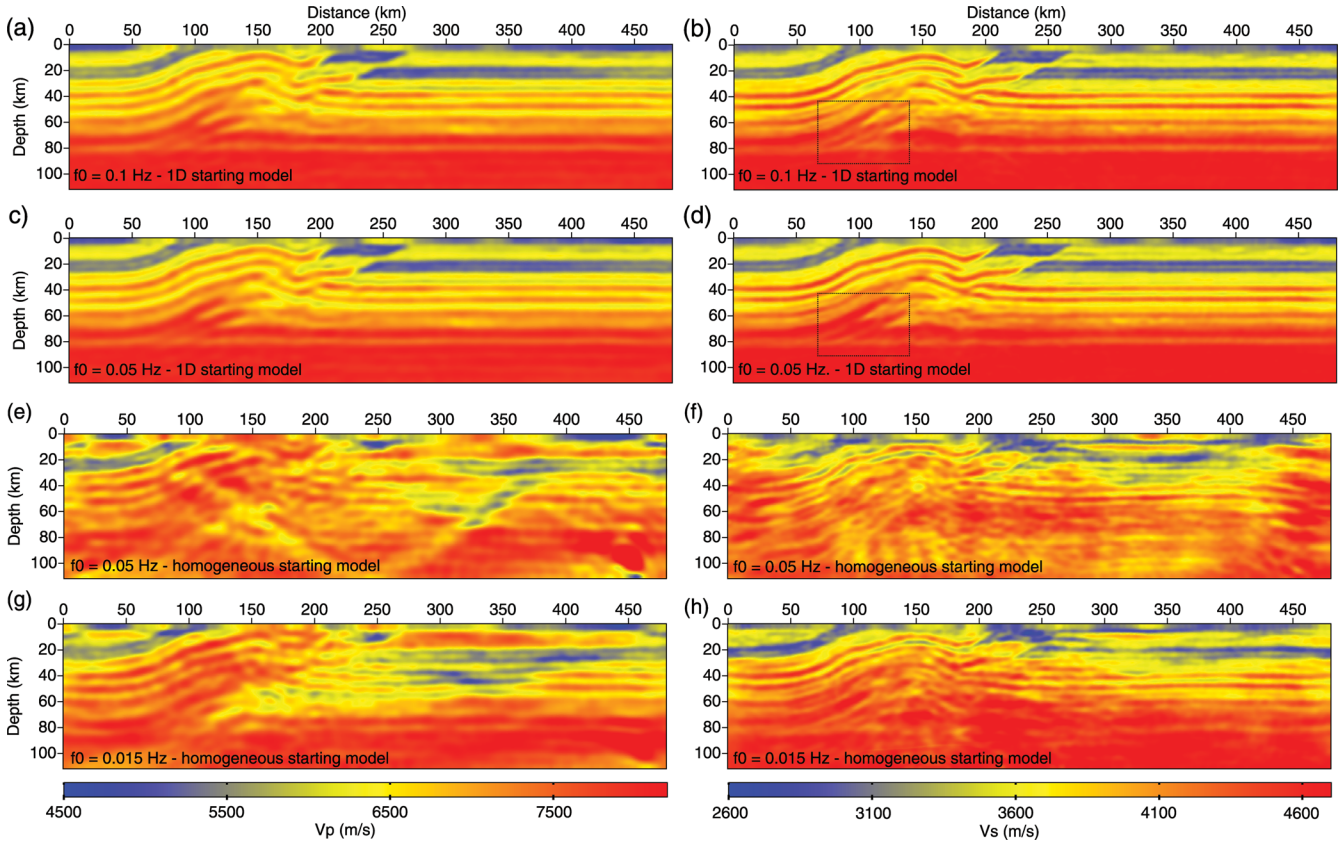

Figure 19. Impact of the initial model. (a, b) V_P (a) and V_S (b) FWI models inferred from vertical velocity-gradient starting models. The starting frequency was 0.1 Hz. Note the deficit of small wavenumbers in the thrust area (dashed rectangle). (c, d) As for (a, b), for a starting frequency of 0.05 Hz. Note how the gap of low wavenumbers is filled in. (e, f) The starting models are homogeneous with $V_P = 6$ km s^{-1} and $V_S = 3.4$ km s^{-1}. The starting frequency was 0.05 Hz. (g, h) As for (e, f), for a starting frequency of 0.015 Hz. In the last two cases, the inversion fails to converge towards the global minimum. See Table 3 for the FWI set-up.

S wavelengths (12.8 km and 7.4 km, respectively), as outlined in Table 2. The 9.6 km elements were reconstructed with underestimated amplitudes in the deep part of the V_P model (Fig. 20c). The *S*-wave velocity model reconstruction shows overall more consistent amplitudes (Fig. 20d). However, the shape of the elements is poorly reconstructed in several parts of the model, where the background model shows significant lateral variations (i.e. in the thrust area). This might result from the more limited *S*-wave illumination compared to the *P*-wave counterpart, which arises because the incident wavefield is compressional. The FWI fails to reconstruct the 4.8 km elements in the V_P model, except in the first 5 km of the subsurface (Fig. 20e). This is consistent, because the size of the elements is below half the minimum *P*-wavelength. The elements are successfully reconstructed in the V_S model down to 60 km in depth, although the footprint of the initial model is visible near the left-hand side of the model (Fig. 20f). The successful reconstruction of the 4.8 km elements down to 60 km in depth is consistent with the minimum shear wavelength, the dimension of which is of the same order of magnitude as the size of the elements.

For completeness, we show the results of the checkerboard tests for the 4.8 km elements when the conjugate gradient algorithm was used instead of L-BFGS (Figs 20g–j). The frequency groups are inverted in Figs 20(g) and (h), while single frequencies are inverted in Figs 20(i) and (j). Comparisons between the FWI results shown in Figs 20(e, f) and 20(g, j) highlight again the importance of the Hessian and of the frequency management in FWI of coarse data.

4.9 Full waveform inversion and migration

FWI and depth migration rely on similar imaging principles: the perturbation model or the migrated image are built by zero-lag correlation of the incident wavefield and the adjoint wavefield, which is backpropagated from the receiver positions. Therefore, a depth migrated image is provided by the gradient of the FWI misfit function at the first iteration, this gradient being stacked over the full seismic bandwidth (Lailly 1983, 1984). These depth migrations based on the full wave equation are referred to as reverse time migration and can be performed in the frequency domain (Mulder & Plessix 2004; Kim *et al.* 2011; Prieux *et al.* 2011). The data residuals, which are backpropagated from the receiver positions in FWI, are replaced by the recorded wavefield in migration. Typically, this recorded wavefield corresponds to the reflection wavefield in surface seismic exploration. A key difference between migration and FWI is that the former relies on scale uncoupling between the background model and the migrated image as the background model is only used for wave propagation from the source and from the receivers to perform the zero-lag correlation of the incident and adjoint wavefields. For FWI, aside the correlation, the data residuals, which are the source of the adjoint equation, depend also on the velocity model leading

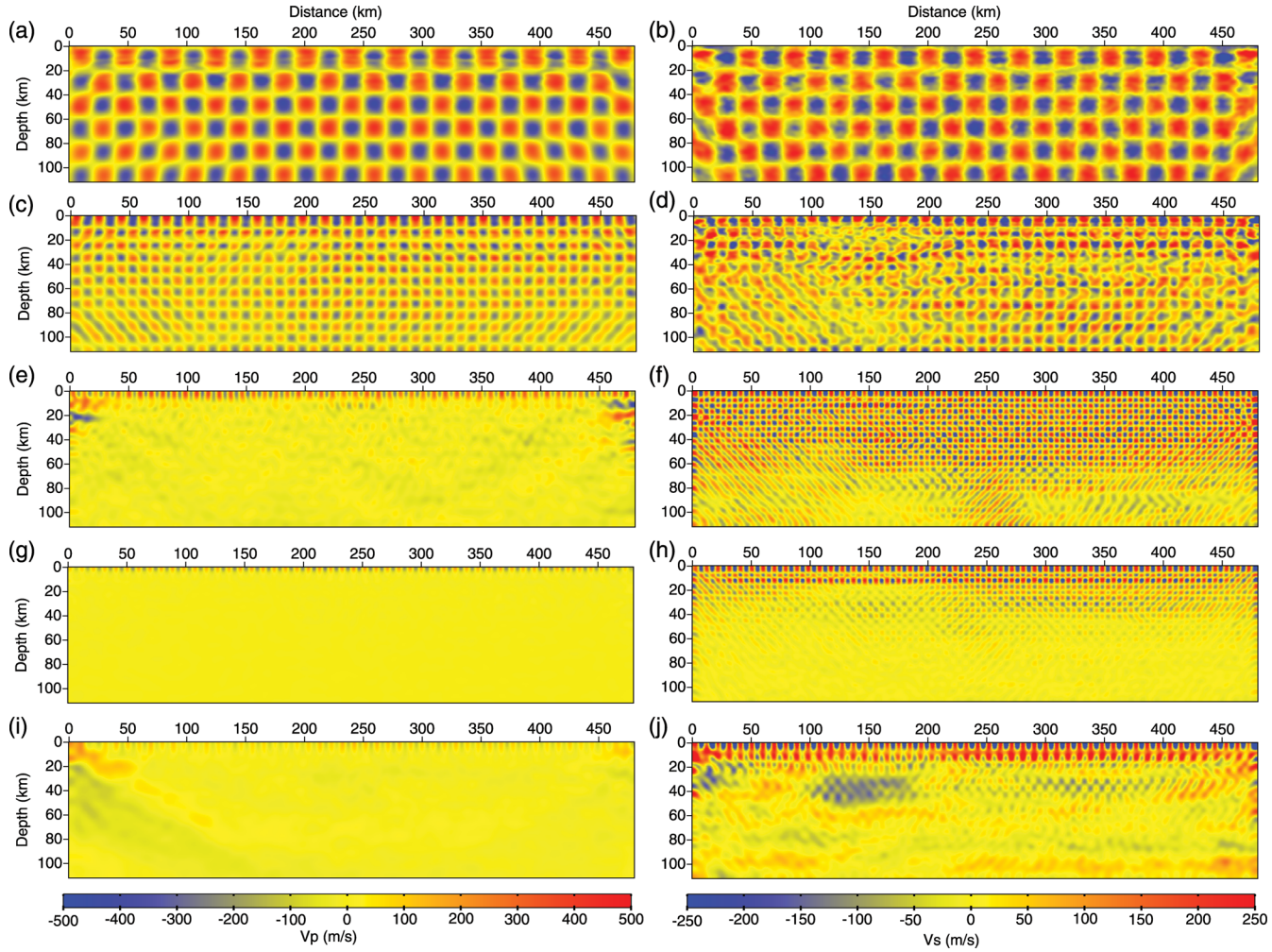

Figure 20. Resolution analysis by checkerboard tests. V_P (left-hand panel) and V_S (right-hand panel) FWI perturbation models for elements of dimension (a, b) 19.2 km, (c, d) 9.6 km and (e, f) 4.8 km. In (a–f), the inversions were applied to three overlapping frequency groups using the L-BFGS optimization. (g, h) As for (e, f), except that the conjugate gradient algorithm is used instead of L-BFGS. (i, j) As for (g, h), except that single frequencies were inverted rather than frequency groups.

to a non-linear iterative process during which the model is updated by model perturbations at each iteration.

We illustrate the relationship between FWI and migration in Fig. 21. A theoretical migrated image can be generated by applying a depth-to-time conversion of the true perturbation model (i.e. the difference between the true model and the initial model), bandpass filter each time profile of the resulting model, and apply the reciprocal time-to-depth conversion to the bandpass filtered model (e.g. Thierry *et al.* 1999). The low-pass frequency of the bandpass filter corresponds to the maximum frequency of the source. The high-pass frequency of the bandpass filter would correspond to the smallest frequency of the source for zero-offset recordings. However, non-zero scattering angles push the imaging towards wavenumbers that are smaller than the one associated with the lowest temporal frequency in virtue of eq. (12). The theoretical migrated image is shown in Fig. 21(a) for the V_S model and for a filter bandwidth [0.05 Hz; 4.5 Hz], which was chosen slightly broader than the source bandwidth [0.1 Hz−0.4 Hz]. The same bandpass filtering procedure is applied when the final FWI V_S model (Fig. 14h) is used instead of the true model (Fig. 21b). The resulting perturbation model closely matches the true migrated image, that highlights that FWI embeds a migration task if the inversion can be pushed towards sufficiently-high frequencies. The migrated image for V_S perturbations obtained by frequency-domain elastic reverse time migration is shown in Fig. 21(c). We show a poor-quality migrated section with a significant deficit of small-wavenumber content (compare the wavenumber content of the migrated images of Fig. 21).

Although the high wavenumbers of the migrated image of the Fig. 21(c) are positioned at the correct depths (suggesting a sufficiently-accurate background model for migration of the reflection wavefield), the small-to-intermediate wavenumbers of the V_S lithospheric model are lacking in the migrated image. This might result because the incident source is compressional and the migration background model is smooth. Therefore, no *P*–*S* conversion occur during the propagation from the bottom of the target to the surface, until the reflection from the free surface. Without forward-scattered PS waves, the long-to-intermediate wavelengths of the V_S models cannot be reconstructed: only the short wavelengths of V_S are imaged from the back-scattered PS waves generated by the free surface and the lithospheric reflectors.

One may wonder why FWI succeeds in reconstructing these long-to-intermediate wavelengths of the V_S model (Fig. 21b), unlike

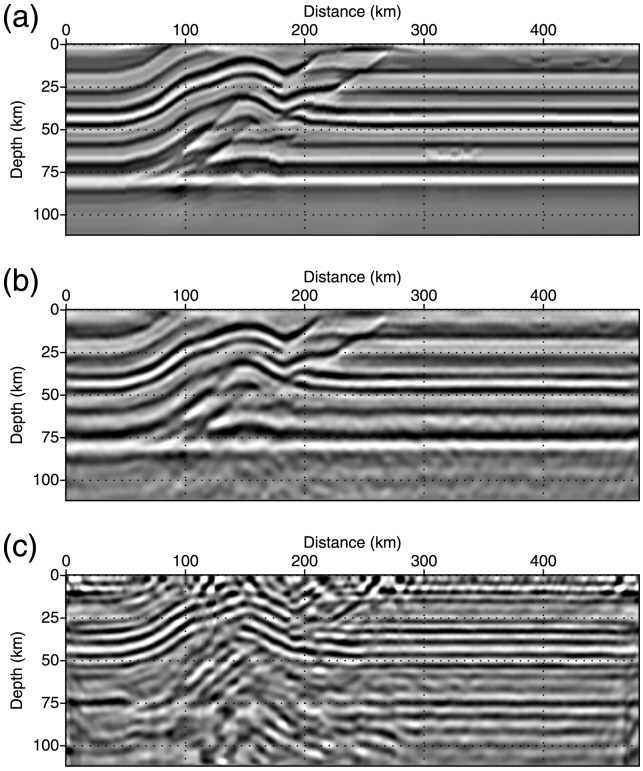

Figure 21. Relationship between FWI and migration. (a) True migrated image, which corresponds to the difference between the true model and the starting model after bandpass filtering in the time domain (see text for more details). (b) Same as (a) but the final FWI V_S model replaces the true model. (c) Reverse time migrated image computed in the frequency domain using the starting model.

migration (Fig. 21c). The reason might be that, as the FWI hierarchically progresses towards high frequencies, shorter wavelengths are injected in the lithospheric model during each non-linear iteration of the FWI. The discontinuities that are progressively injected at each non-linear iteration of the FWI lead to forward-scattered PS waves which are amenable to the building of the long-to-intermediate wavelengths of V_S during the late stages of FWI. Notice that, it this statement is correct, the imaging process follows a different path than the conventional multiscale approach which reconstructs first the long wavelengths before the short wavelengths. This results from the particular geometry and the source type of teleseismic acquisition. The reconstruction of short wavelengths before the long wavelengths is made possible by the redundant control of frequencies and scattering angles on the wavenumber coverage.

These results highlight the fundamental difference between the linear imaging process underlying migration and the non-linear one underlying FWI and the improved imaging that is expected from FWI by progressively updating the background model in which seismic modelling is performed.

5 DISCUSSION AND CONCLUSION

In the present study, we have discussed the feasibility of efficient frequency-domain FWI for lithospheric imaging from teleseismic data through a synthetic case study. We focused our analysis on the sensitivity of the FWI to several factors.

First, we have shown that the surface-reflected wavefields carry the most resolving information on the lithospheric structure for both the *P*-wave and *S*-wave wave velocities. In this context, the free surface can be seen as a source on the surface, which propagates downgoing *P* and *S* plane waves that are amenable to reflection from the main discontinuities before their recording on the surface. The incident primary wavefields and the reflection wavefields can be combined in a joint tomography-like and migration-like waveform inversion to broaden the range of the scattering angles involved in the inversion, and hence, to improve the resolution of the lithospheric models. This double illumination provides a distinct advantage over conventional surface acquisitions in exploration geophysics, for which building the large wavelengths of the subsurface is challenging. The ability of FWI to account for forward scattering and backward scattering through full wave modelling is a distinct advantage over conventional ray-theoretical teleseismic tomographic approaches that lack resolving power. Our numerical experiments show that a theoretical resolution of the order of one wavelength is achievable for both the *P* and *S* velocity models if the full *P* and *P*–*S* converted wavefields can be taken into account.

Considering the low-frequency content of teleseismic sources and the sampling of wide scattering angles during the up-going propagation of the primary wavefield, we can conclude that teleseismic acquisitions should provide a suitable framework to reduce the risk of cycle skipping, and to build the large wavelengths of the structure. We have shown that a velocity-gradient model provides a suitable initial model to perform FWI for a starting frequency of 0.05 Hz. If the starting frequency is increased to 0.1 Hz, imaging of deep dipping structures starts to be hampered by a deficit of small wavenumbers.

In the general framework of diffraction tomography, the temporal frequency and the scattering angle have double control on the resolving power of the waveform inversion. Efficient frequency-domain FWI of wide-aperture data in exploration geophysics (e.g. cross-hole data, long-offset refraction experiments) is conventionally applied to a few discrete frequencies to design computationally efficient algorithms when seismic modelling is performed in the frequency domain. This frequency decimation is possible because point sources have a broad directivity pattern, and hence allow a broad range of scattering angles to be finely sampled. We have shown that such aggressive frequency decimation is generally not permitted in teleseismic FWI, because the limited number of incident plane-wave sources leads to a coarser and narrower sampling of the scattering angles. In this context, refining the frequency interval is required to prevent notches in the wavenumber spectra each time a new frequency component is inverted. When a few plane-wave sources sample a broad band of incidence angles, the wavenumber spectrum of the subsurface that is constrained during a single-frequency inversion can be under-sampled, which leads to spatial aliasing. To prevent these artefacts, it is worth performing simultaneous inversion of multiple frequencies rather than inversion of single frequencies. This does not add prohibitive extra computational cost if the frequency bandwidths of two consecutive groups do not significantly overlap.

Even if the advent of modern seismic instrumentation allows for the deployment of denser and denser station networks, the maximum receiver spacing that allows reliable FWI to be performed should be estimated according to the high resolving power of this imaging. First, we have shown the importance of the L-BFGS optimization algorithm to improve the focusing of the imaging, and to reduce the footprint of the aliasing artefacts through accounting for the Hessian operator. We have shown that horizontal aliasing starts appearing in the shallow part of the shear-wave velocity models

for a receiver spacing of the order of 10 km. This receiver spacing is of the order of the shear wavelength in the shallow part of the lithospheric model. Reliable *P*-wave and *S*-wave velocity models are obtained when the receiver spacing is up to 20 km, as long as L-BFGS optimization and frequency groups are used during the inversion.

Although we perform FWI in the frequency domain, the conclusions of this study also apply to time-domain FWI. As the frequency interval is reduced in frequency-domain inversion, frequency-domain and time-domain inversions become equivalent, as a similar amount of data is injected in both inversions. We have shown that teleseismic geometries require inversion of a large number of frequencies for a limited number of sources. From a numerical viewpoint, this result questions the relevance of frequency-domain approaches based on Gauss-elimination techniques, as a new LU decomposition of the impedance matrix needs to be performed for each frequency. Therefore, it is likely that time-domain inversion will be the approach of choice to perform 3-D FWI of teleseismic data. Alternatively, frequency-domain inversion based on iterative solvers can compete with the time-domain approach if efficient preconditioner allows for the iteration number to be independent of frequency (Plessix 2009).

Our parametric analysis was limited to the 2-D assumption. One may wonder whether the conclusions inferred from this study can be generalized in three dimensions. Some of the factors that were analysed in the present study should have similar influence in two and three dimensions. First, accounting for free surface effects will be a key to improve the spatial resolution of 3-D FWI of teleseismic data, as it is in 2-D. We have shown that a fine frequency interval is required to prevent spatial aliasing to overcome the narrow and coarse scattering angle illumination provided by a sparse set of plane-wave sources. This requirement will likely remain in 3-D because, even if the number of available teleseisms will dramatically increase in 3-D experiments relatively to 2-D ones, the range of incidence angles of the plane wave sources will remain bounded by the useful range of epicentral distances (30–60°), and the scattering-angle illumination will remain impacted on by the uneven teleseismic coverage. The conditions that must be satisfied by the receiver interval to prevent shallow artefacts in relation with the radius of the Fresnel zone will apply equally well in 3-D. Such artefacts referred to as acquisition footprint were already noticed in 3-D FWI results obtained with coarse land or seabottom data sets (e.g. Ben Hadj Ali *et al.* 2008; Sirgue *et al.* 2010).

Indeed, this study does not pretend to cover all of the factors that can impact on FWI of teleseismic data. We have used the same modelling engine to compute the data in the true model and in the FWI models (the so-called inverse crime), and noise-free data to illustrate the footprint of selected factors and to quantify the intrinsic resolving power of FWI.

One difficulty with real data application will probably be related to noise in the data. The footprint of this noise will be difficult to remove during imaging because the data redundancy provided by teleseismic acquisition is not as high as in controlled-source seismology. This is a significant issue in the sense that noise will dominantly hamper the inversion of the low-amplitude secondary scattered waves, which have the most resolving power. Evolving towards 3-D FWI seems necessary to strengthen the data redundancy and exploit these low-amplitude back-scattered waves. We have not investigated some data-preconditioning strategies in the present study that might help to strengthen the contribution of the secondary scattered wavefield in FWI. One possible strategy would consist of proceeding sequentially from the early arrivals to the later-arriving phases in the data space, to reconstruct the long wavelengths of the lithospheric model before the shorter wavelengths. This multiscale approach can be easily implemented with time damping applied from the first arrival, which can be easily picked on teleseismic data (Brossier *et al.* 2009b). With this hierarchical approach, the misfit reduction of the secondary scattered wavefield should be facilitated during the late stages of the FWI, once the misfit of the primary wavefields has been efficiently reduced during the early stage of the FWI. Moreover, these multiscale approaches should also help to start the FWI from crude initial models without cycle skipping artefacts, as a limited number of propagated wavelengths will be propagated during the early stages of the FWI, where only the early arrivals are involved in the inversion.

FWI requires the estimation of the source signature (here, by source signature is meant a temporal wavelet representative of the incident plane wave). The source estimation is generally nested with the estimation of the subsurface parameters in frequency-domain FWI (Pratt 1999). The source estimation can also be used for quality control of FWI models (Brenders & Pratt 2007b; Malinowski *et al.* 2011; Prieux *et al.* 2011). In the teleseismic configuration, there might be a significant trade-off between the timing of the source wavelet, the velocity models, and the incidence angle of the incoming plane-wave source, to match the first-arrival traveltimes. Moreover, the teleseismic source signature can be much more complex than that of controlled sources. For example, the *P* and *S* ghost reflections on the free surface (the so-called pP and sP arrivals) and the footprint of the crustal heterogeneities from the source side can interfere with useful signals in the time window considered for FWI, and prevent the implementation of multiscale approaches by application of time-windowing to the data set.

Finally, the reliability of the amplitude information can be questioned in teleseismic FWI. The influence of the amplitudes in FWI can be reduced by using the L1 norm of data residuals instead of the L2 norm, to compute the data misfit function. In frequency-domain FWI, this amounts to back-propagation of the residuals normalized by their modulus, instead of the residuals themselves (Brossier *et al.* 2010b). Alternatively, the influence of the amplitudes can be reduced (although not removed) by only inverting the phase spectrum, which can be easily implemented with a logarithmic norm of the data residuals (Shin & Min 2006; Bednar *et al.* 2007; Shin *et al.* 2007). Empirical data preconditioning were also proposed to reduce the influence of the amplitudes on the inversion such as trace normalization by spectral whitening (Operto *et al.* 2004; Malinowski & Operto 2008; Malinowski *et al.* 2011; Bleibinhaus & Hilberg 2012) or amplitude-versus-offset scaling (Brenders & Pratt 2007b).

Application of elastic frequency-domain FWI to real teleseismic data will be the aim of future studies to assess the feasibility of this technology for seismological applications.

ACKNOWLEDGEMENTS

Access to the high-performance computing facilities of MESOCENTRE SIGAMM computer centres provided the required computer resources and we gratefully acknowledge this facility and the support of the staff. Finally, this study was carried out within the framework of the SEISCOPE consortium (http://seiscope.oca.eu) sponsored by BP, CGG-Veritas, ENI, EXXON-Mobil, PETROBRAS, SAUDI ARAMCO, SHELL, STATOIL and TOTAL. We would like to thank the Editor Saskia Goes, Florian Bleibinhaus (University Salzburg) and an anonymous reviewer for their helpful comments on the manuscripts.

REFERENCES

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology, Theory and Methods,* 2nd edn, University Science Books, Sausalito, California.

Aminzadeh, F., Brac, J. & Kunz, T., 1997. 3-D Salt and Overthrust models: SEG/EAGE 3-D Modeling Series No. 1.

Ammon, C.J., 1991. The isolation of receiver effects from teleseismic p waveform, *Bull. seism. Soc. Am.,* **81,** 2504–2510.

Ammon, C.J., Randall, G.E. & Zandt, G., 1990. On the nonuniqueness of receiver function inversions, *J. geophys. Res.,* **95,** 15 303–15 318.

Bednar, J.B., Shin, C. & Pyun, S., 2007. Comparison of waveform inversion, part 2: phase approach, *Geophys. Prospect.,* **55,** 465–475.

Ben Hadj Ali, H., Operto, S. & Virieux, J., 2008. Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data, *Geophysics,* **73**(5), VE101–VE117.

Berenger, J.-P., 1994. A perfectly matched layer for absorption of electromagnetic waves, *J. Comput. Phys.,* **114,** 185–200.

Bleibinhaus, F. & Hilberg, S., 2012. Shape and structure of the Salzach Valley, Austria, from seismic traveltime tomography and full waveform inversion, *Geophys. J. Int.,* **189**(3), 1701–1716.

Bleibinhaus, F., Hole, J.A., Ryberg, T. & Fuis, G.S., 2007. Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging, *J. geophys. Res.,* **112,** doi:10.1029/2006JB004611.

Bleibinhaus, F., Lester, R.W. & Hole, J.A., 2009. Applying waveform inversion to wide-angle seismic surveys, *Tectonophysics,* **472,** 238–248.

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.

Bozdag, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.,* **185,** 845–870.

Brenders, A.J. & Pratt, R.G., 2007a. Efficient waveform tomography for lithospheric imaging: implications for realistic 2D acquisition geometries and low frequency data, *Geophys. J. Int.,* **168,** 152–170.

Brenders, A.J. & Pratt, R.G., 2007b. Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model, *Geophys. J. Int.,* **168,** 133–151.

Brossier, R., 2011. Two-dimensional frequency-domain visco-elastic full waveform inversion: parallel algorithms, optimization and performance, *Comput. Geosci.,* **37,** 444–455.

Brossier, R., Etienne, V., Operto, S. & Virieux, J., 2010a. Frequency-domain numerical modelling of visco-acoustic waves based on finite-difference and finite-element discontinuous galerkin methods, in *Acoustic Waves,* pp. 125–158, ed. Dissanayake, D.W., SCIYO, Rijeka, Croatia.

Brossier, R., Operto, S. & Virieux, J., 2009a. 2D elastic frequency-domain full-waveform inversion for imaging complex onshore structures, in *Proceedings of the 71th Annual EAGE Conference & Exhibition,* Amsterdam, Expanded Abstracts, U019, EAGE.

Brossier, R., Operto, S. & Virieux, J., 2009b. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics,* **74,** WCC63–WCC76.

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

Brossier, R., Virieux, J. & Operto, S., 2008. Parsimonious finite-volume frequency-domain method for 2-D P-SV-wave modelling, *Geophys. J. Int.,* **175,** 541–559.

Chavent, G., 1974. Identification of parameter distributed systems, in *Identification of Function Parameters in Partial Differential Equations,* pp. 31–48, eds Goodson, R. & Polis, M., American Society of Mechanical Engineers, New York.

Chavent, G., 2009. *Nonlinear Least Squares for Inverse Problems,* Springer, Dordrecht/Heidelberg/London/New York.

Chen, P., Jordan, T. & Zhao, L., 2007a. Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods, *Geophys. J. Int.,* **170,** 175–181.

Chen, P., Zhao, L. & Jordan, T.H., 2007b. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.,* **97,** 1094–1120.

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

Dziewonski, A.M., 1984. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.,* **89,** 5929–5952.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Etienne, V., Chaljub, E., Virieux, J. & Glinsky, N., 2010. An hp-adaptive discontinuous Galerkin finite-element method for 3D elastic wave modelling, *Geophys. J. Int.,* **183,** 941–962.

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

Fichtner, A. & Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.,* **185,** 775–798.

Fichtner, A. & Trampert, J., 2011b. Resolution analysis in full waveform inversion, *Geophys. J. Int.,* **187**(3), 1604–1624.

Forgues, E. & Lambaré, G., 1997. Parameterization study for acoustic and elastic ray+born inversion, *J. Seism. Explor.,* **6,** 253–278.

Fukao, Y., Widiyantoro, S. & Obayashi, M., 2001. Stagnant slabs in the upper and lower mantle transition region, *Rev. Geophys.,* **39,** 291–323.

Gaffet, S., 1995. Teleseismic waveform modeling including geometrical effects of superficial geological structures near to seismic sources, *Bull. seism. Soc. Am.,* **85,** 1068–1079.

Gaffet, S. & Bouchon, M., 1989. Effects of two-dimensional topographies using the discrete wavenumber-boundary integral equation method in P-SV cases, *J. acoust. Soc. Am.,* **85,** 2277–2283.

Guitton, A., Ayeni, G. & Dìaz, E., 2012. Constrained full-waveform inversion by model reparameterization, *Geophysics,* **77,** R117–R127.

Hesthaven, J.S. & Warburton, T., 2008. *Nodal Discontinuous Galerkin Method: Algorithms, Analysis, and Application,* Springer, New York.

Jannane, M. *et al.*, 1989. Wavelengths of Earth structures that can be resolved from seismic reflection data, *Geophysics,* **54,** 906–910.

Jin, S., Madariaga, R., Virieux, J. & Lambaré, G., 1992. Two-dimensional asymptotic iterative elastic inversion, *Geophys. J. Int.,* **108,** 575–588.

Kamei, R., Pratt, R.G. & Tsuji, T., 2012. Waveform tomography imaging of a megasplay fault system in the seismogenic Nankai subduction zone, *Earth planet. Sci. Lett.,* **317-318,** 343–353.

Kennett, B. L.N., 1983. *Seismic Wave Propagation in Stratified Media,* Cambridge University Press, Cambridge, UK.

Kim, Y., Min, D.-J. & Shin, C., 2011. Frequency-domain reverse-time migration with source estimation, *Geophysics,* **76,** S41–S49.

Lailly, P., 1983. The seismic problem as a sequence of before-stack migrations, *Presented at the Conference on Inverse Scattering: Theory and Applications,* SIAM, Philadelphia.

Lailly, P., 1984. The seismic inverse problem as a sequence of before stack migrations, in *Proceedigns of the Conference on Inverse Scattering: Theory and Application, Expanded Abstracts,* SIAM, Philadelphia, pp. 206–220.

Lambaré, G., 1991. Inversion linearisée de données de sismique réflexion par une méthode quasi-newtonienne, *PhD thesis,* Université de Paris VII.

Lambaré, G., Operto, S., Podvin, P., Thierry, P. & Noble, M., 2003. 3-D ray+Born migration/inversion—part 1: theory, *Geophysics,* **68,** 1348–1356.

Langston, C.A., 1979. Structure under mount rainier, washington, inferred from teleseismic body waves, *J. geophys. Res.,* **84,** 4749–4762.

Lekić, V. & Romanowicz, B., 2011. Inferring upper-mantle structure by full waveform tomography with the spectral element method, *Geophys. J. Int.,* **185,** 799–831.

Lions, J.L., 1968. *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles,* Dunod, Paris.

Lui, X.-D. & Romanowicz, B., 1995. Comparison of global waveform inversion with and without considering cross-branch modal coupling, *Geophys. J. Int.,* **121,** 695–709.

Malinowski, M. & Operto, S., 2008. Quantitative imaging of the Permo-Mesozoic complex and its basement by frequency domain waveform

tomography of wide-aperture seismic data from the Polish basin, *Geophys. Prospect.,* **56,** 805–825.

Malinowski, M., Operto, S. & Ribodetti, A., 2011. High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full waveform inversion, *Geophys. J. Int.,* **186,** 1179–1204.

Marfurt, K., 1984. Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations, *Geophysics,* **49,** 533–549.

Miller, D., Oristaglio, M. & Beylkin, G., 1987. A new slant on seismic imaging: migration and integral geometry, *Geophysics,* **52,** 943–964.

Mora, P.R., 1987. Nonlinear two-dimensional elastic inversion of multi-offset seismic data, *Geophysics,* **52,** 1211–1228.

Mulder, W.A. & Plessix, R.E., 2004. How to choose a subset of frequencies in frequency-domain finite-difference migration, *Geophys. J. Int.,* **158,** 801–812.

MUMPS-team, 2011. *MUMPS — MUltifrontal Massively Parallel Solver users' guide — version 4.10.0 (2011 May 10). ENSEEIHT-ENS Lyon,* http://www.enseeiht.fr/apo/MUMPS/ or http://graal.ens-lyon.fr/MUMPS.

Nocedal, J., 1980. Updating quasi-newton matrices with limited storage, *Math. Comput.,* **35,** 773–782.

Nocedal, J. & Wright, S.J., 1999. *Numerical Optimization,* Springer, New York, USA.

Nolet, G., 1987. *Seismic Tomography with Applications in Global Seismology and Exploration Geophysics,* D. Reidel Publishing Company, Dordrecht, Holland.

Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A. & Dell'Aversana, P., 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study, *Geophys. Prospect.,* **52,** 625–651.

Operto, S., Virieux, J., Dessa, J.X. & Pascal, G., 2006. Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough, *J. geophys. Res.,* **111,** doi:10.1029/2005JB003835.

Pageot, D., Operto, S., Vallée, M., Brossier, R. & Virieux, J., 2010. Lithospheric imaging from teleseismic data by frequency-domain elastic full-waveform tomography, in *Proceedings of the 72th Annual EAGE Conference & Exhibition,* Barcelona, Expanded Abstracts, WS6, EAGE.

Pageot, D., Operto, S., Vallée, M. & Virieux, J., 2009. Lithospheric imaging from teleseismic data by frequency-domain full-waveform tomography, in *Proceedings of the AGU 2009 Fall Meeting,* San Francisco, Expanded Abstracts, S31D–07 POSTER, Society of Exploration Geophysicists.

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.

Plessix, R.E., 2009. Three-dimensional frequency-domain full-waveform inversion with an iterative solver, *Geophysics,* **74,** WCC53–WCC61.

Plessix, R.-E., Baeten, G., de Maag, J.W. & ten Kroode, F., 2012. Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set, *Geophys. Prospect.,* **60,** 733–747.

Pratt, R.G., 1990. Inverse theory applied to multi-source cross-hole tomography. Part II: elastic wave-equation method, *Geophys. Prospect.,* **38,** 311–330.

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

Pratt, R.G., 2008. Waveform tomography — successes, cautionary tales, and future directions, in *Proceedings of the 70th Annual International Meeting,* EAGE, Expanded Abstracts, WO11–Full–Waveform Inversion: current status and perspectives.

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.

Pratt, R.G., Song, Z.M., Williamson, P.R. & Warner, M., 1996. Two-dimensional velocity models from wide-angle seismic data by wavefield inversion, *Geophys. J. Int.,* **124,** 323–340.

Pratt, R.G. & Worthington, M.H., 1990. Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method, *Geophys. Prospect.,* **38,** 287–310.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O. & Kommedal, J., 2011. On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study, *Geophys. J. Int.,* **187,** 1495–1515.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A. & dell'Aversana, P., 2004. Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt, *Geophys. J. Int.,* **159,** 1032–1056.

Ribodetti, A., Operto, S., Virieux, J., Lambaré, G., Valéro, H.-P. & Gibert, D., 2000. Asymptotic viscoacoustic diffraction tomography of ultrasonic laboratory data: a tool for rock properties analysis, *Geophys. J. Int.,* **140,** 324–340.

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

Romdhane, A., Grandjean, G., Brossier, R., Réjiba, F., Operto, S. & Virieux, J., 2011. Shallow structures characterization by 2d elastic waveform inversion, *Geophysics,* **76,** R81–R93.

Rondenay, S., Bostock, M.G. & Fischer, K.M., 2005. Multichannel inversion of scattered teleseismic body waves: practical considerations and applicability, in *Seismic Earth: Array Analysis of Broadband Seismograms,* ed. Alan Levander, G.N., Geophysical Monograph 157, 187–203. American Geophysical Union Washington, DC.

Rondenay, S., Bostock, M.G. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves. 3. Application to the Cascadia 1993 data set, *J. geophys. Res.,* **106,** 30 795–30 807.

Sheriff, R.E. & Geldart, L.P., 1995. *Exploration Seismology,* 2nd edn, Cambridge University Press, Cambridge, UK.

Shin, C.D. & Min, D.-J., 2006. Waveform inversion using a logarithmic wavefield, *Geophysics,* **71,** R31–R42.

Shin, C., Pyun, S. & Bednar, J.B., 2007. Comparison of waveform inversion. Part 1: conventional wavefield vs logarithmic wavefield, *Geophys. Prospect.,* **55,** 449–464.

Shipp, R.M. & Singh, S.C., 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophys. J. Int.,* **151,** 325–344.

Shragge, J., Bostock, M.G. & Rondenay, S., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves 2. Numerical examples, *J. geophys. Res.,* **106,** 30 783–30 793.

Sirgue, L., Barkved, O.I., Dellinger, J., Etgen, J., Albertin, U. & Kommedal, J.H., 2010. Full waveform inversion: the next leap forward in imaging at Valhall, *First Break,* **28,** 65–70.

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

Soubaras, R. & Whiting, P., 2011. Variable depth streamer—the new broadband acquisition system, *SEG Tech. Prog., Expanded Abstracts,* **30,** 4349–4353.

Stekl, I. & Pratt, R.G., 1998. Accurate viscoelastic modeling by frequency-domain finite difference using rotated operators, *Geophysics,* **63,** 1779–1794.

Taflove, A. & Hagness, S.C., 2000. *Computational Electrodynamics: The Finite-Difference Time-Domaine Method,* Artech House, Boston, London.

Takenaka, H.B. & Kennett, L.N., 1996. A 2.5-D time-domain elastodynamic equation for plane-wave incidence, *Geophys. J. Int.,* **125,** F5–F9.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. 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.

Thierry, P., Operto, S. & Lambaré, G., 1999. Fast 2D ray-Born inversion/migration in complex media, *Geophysics,* **64,** 162–181.

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

Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics,* **74,** WCC127–WCC152.

Wang, B., 1993. Improvement of seismic travel-time inversion methods and application to observed data, *PhD thesis,* Purdue University.
Wang, Y. & Rao, Y., 2009. Reflection seismic waveform tomography, *J. geophys. Res.,* **114,** doi:10.1029/2008JB005916.
Williamson, P., 1991. A guide to the limits of resolution imposed by scattering in ray tomography, *Geophysics,* **56,** 202–207.
Woodhouse, J. & Dziewonski, A., 1984. Mapping the upper mantle: three dimensional modelling of earth structure by inversion of seismic waveforms, *J. geophys. Res.,* **89,** 5953–5986.
Wu, R.S. & Toksöz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics,* **52,** 11–25.

APPENDIX A: CORRECTION FOR OBLIQUITY

In this appendix, we propose an approximate correction for obliquity to perform 2-D FWI of teleseseismic data. The method relies on the assumption of a horizontally stratified medium, for which the horizontal component of the slowness vector does not vary.

Let us denote by (x, z) the Cartesian coordinate system defined by the receiver line and depth.

Let us consider a plane wave that propagates with a wave speed c in the direction of the axis $\tilde{x}$ with an obliquity angle ϕ relative to the axis y (Fig. A1), and an incidence angle ξ. Since the source and the medium are invariant along $\tilde{y}$, the wavefield satisfies a 2-D wave equation in the coordinate system $(\tilde{x}, z)$.

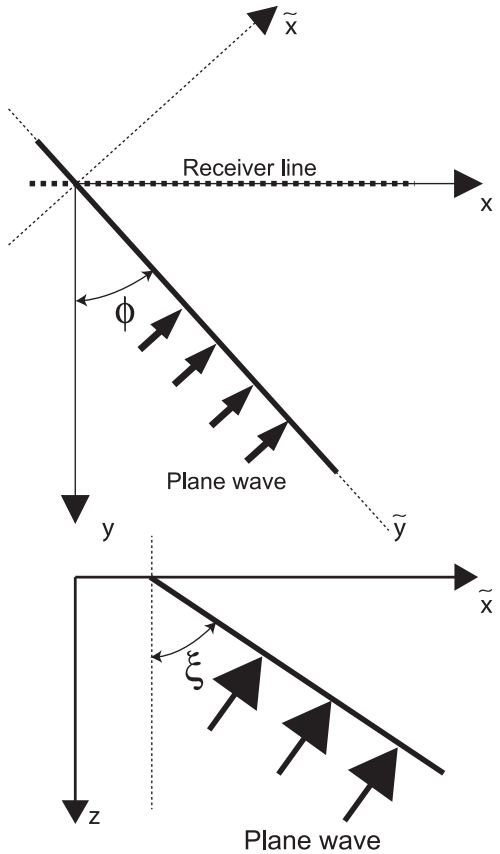

Figure A1. Correction for obliquity. The plane-wave source propagates in a horizontal direction with an angle ϕ with respect the axis perpendicular to the receiver line. The incidence angle is denoted by the symbol ξ in the vertical plane.

Under these assumptions, the solution of the 2-D wave equation for particle velocities in the coordinate system $(\tilde{x}, z)$ is of the form

$$\begin{aligned} V_{\tilde{x}} &= A\,\sin\phi\,\exp\left[-\mathrm{i}\omega\left(\tfrac{\tilde{x}\,\sin\phi}{c}+\tfrac{z\,\cos\phi}{c}-t\right)\right],\\ V_z &= A\,\cos\phi\,\exp\left[-\mathrm{i}\omega\left(\tfrac{\tilde{x}\,\sin\phi}{c}+\tfrac{z\,\cos\phi}{c}-t\right)\right], \end{aligned} \tag{A1}$$

where ϕ is the incidence angle and $\sin\phi/c$ is the apparent horizontal slowness in the vertical plane defined by the receiver line.

If we substitute $\tilde{x}$ by its expression as a function of x and y, $\tilde{x} = x\,\cos\theta + y\,\sin\theta$, we get

$$\begin{aligned} V_{\tilde{x}} &= A\,\sin\phi\,\exp\left[-\mathrm{i}\omega\left(\tfrac{x\,\sin\phi\,\cos\theta}{c}+\tfrac{y\,\sin\phi\,\sin\theta}{c}+\tfrac{z\,\cos\phi}{c}-t\right)\right],\\ V_z &= A\,\cos\phi\,\exp\left[-\mathrm{i}\omega\left(\tfrac{x\,\sin\phi\,\cos\theta}{c}+\tfrac{y\,\sin\phi\,\sin\theta}{c}+\tfrac{z\,\cos\phi}{c}-t\right)\right]. \end{aligned} \tag{A2}$$

In the vertical plane of the receivers (defined by $y = 0$), the expression reduces to the expressions

$$\begin{aligned} V_x &= A\,\sin\phi\,\cos\theta\,\exp\left[-\mathrm{i}\omega\left(\tfrac{x\,\sin\phi\,\cos\theta}{c}+\tfrac{z\,\cos\phi}{c}-t\right)\right],\\ V_z &= A\,\cos\phi\,\exp\left[-\mathrm{i}\omega\left(\tfrac{x\,\sin\phi\,\cos\theta}{c}+\tfrac{z\,\cos\phi}{c}-t\right)\right]. \end{aligned} \tag{A3}$$

We seek a plane wave defined by an apparent wave speed c_a and an apparent incidence angle ϕ_a, which is equivalent to the plane wave of eq. (A3). This leads to the following system of two equations for two unknowns, as

$$\begin{aligned} \tfrac{\sin\phi_a}{c_a} &= \tfrac{\sin\phi_a\,\cos\theta}{c_a},\\ \tfrac{\cos\phi_a}{c_a} &= \tfrac{\cos\phi_a}{c_a}. \end{aligned} \tag{A4}$$

This gives for ϕ_a and c_a the relationships

$$\tan\,\phi_a = \tan\phi\,\cos\theta, \tag{A5}$$

and

$$\begin{aligned} c_a(\mathbf{x}) = c(\mathbf{x})\frac{1}{\sqrt{1-\sin^2\theta\,\sin^2\phi(\mathbf{x})}} &= c\frac{1}{\sqrt{1-p^2\,c^2(\mathbf{x})\,\sin^2\theta}}\\ &= \xi(\mathbf{x})c(\mathbf{x}), \end{aligned} \tag{A6}$$

where the slowness is denoted by p.

The wave speed c_a is the apparent velocity of the plane wave in the vertical plane defined by the receiver line, and angle ϕ_a is the angle between the z-axis and the line of intersection between the plane wave and the (x, z)-plane. The explicit dependency of the apparent wave speed c_a on the spatial coordinates $\mathbf{x}$ is explicitly written to note that the correction factor ξ is local, and hence that it should be applied at each subsurface position. This correction is exact as long as the medium is horizontally stratified. The approximation should be acceptable for moderate obliquity when the geological structures are reasonably cylindrical (such as subduction zones) and the receiver line is oriented perpendicular to the dips.

These corrections can be implemented in the scattered-field formulation of the seismic modelling. The apparent incident angles are taken into account during the computation of the background wavefield $\mathbf{u}_b$. Since we choose a homogeneous half space for $\mathbf{m}_b$, the analytical solution for $\mathbf{u}_b$ is computed by replacing the incidence angle of the plane wave at the base of the lithospheric target by the apparent incidence angle, eq. (A5). The impedance matrices $\mathbf{B}$ and $\mathbf{B}_b$ are built for the apparent velocity models given by eq. (A6). Of note, an apparent impedance matrix needs to be computed for each incidence angle, as the correction ξ depends on this angle. This is a clear drawback when a frequency-domain modelling approach based on Gauss elimination is used, as the LU decomposition needs

to be performed for each teleseism (i.e. source), while this decomposition is independent of source in conventional applications.

During the inversion, the gradient of the misfit function is computed with respect to the apparent wave speed for each teleseism, as these wave speeds are implemented in the forward-problem wave equation. Then, the gradients of the misfit function with respect to the true wave speeds are inferred from the gradients computed from the apparent wave speeds using the chain rule, before being stacked and smoothed with the 2-D Gaussian smoothing operator

$$\nabla C_c = \nabla C_{c_a}\, \xi. \tag{A7}$$

APPENDIX B: COMPUTING THE GRADIENT WITH THE ADJOINT-STATE METHOD

In this appendix, we derive the gradient of the misfit function with the adjoint-state method (Lions 1968; Chavent 1974; Tarantola 1984; Plessix 2006; Chavent 2009) using the second-order wave equation for particle velocities as the state equation, while seismic modelling is performed with a first-order velocity–stress wave equation, eq. (1). This allows us to manipulate self-adjoint operator and save memory during the gradient computation.

The second-order wave equation for particle velocities can be inferred from the velocity–stress system, eq. (1), by eliminating the stress wavefields which gives us the system

$$\begin{aligned}\omega^2 \rho v_x &= \frac{\partial}{\partial x}(\lambda + 2\mu)\frac{\partial v_x}{\partial x} + \frac{\partial}{\partial z}\lambda\frac{\partial v_z}{\partial z} - \mathrm{i}\omega f_x,\\ \omega^2 \rho v_z &= \frac{\partial}{\partial x}\lambda\frac{\partial v_w}{\partial x} + \frac{\partial}{\partial z}(\lambda + 2\mu)\frac{\partial v_z}{\partial z} - \mathrm{i}\omega f_z.\end{aligned} \tag{B1}$$

After spatial discretization, the second-order order wave equation, eq. (B1), can be recast in matrix form as

$$\mathbf{B}\mathbf{v} = \mathbf{s}_v. \tag{B2}$$

where $\mathbf{B}$ denotes the impedance matrices associated with the second-order system. To build the matrix $\mathbf{B}$, we first discretize the velocity–stress system with a simple finite volume scheme, which is equivalent to the P0 discontinuous Galerkin scheme, and eliminate the auxiliary stress wavefields following the parsimonious approach developed by Brossier *et al.* (2008). However, we do not use the second-order wave equation for seismic modelling because implementation of the higher-order P1–P2 discontinuous Galerkin method in the second-order wave equation is quite complex. We rely on velocities of the first-order formulation for computing synthetics used in the misfit function evaluation taking advantage of the *hp*-adaptivity to perform accurate wavefield modelling. The matrix $\mathbf{B}$ is symmetric, and hence it defines self-adjoint operator, unlike the velocity–stress forward-problem operator, eq. (1). The velocity wavefields are denoted by $\mathbf{v}(\mathbf{x}, \omega) = [v_x(\mathbf{x}, \omega), v_z(\mathbf{x}, \omega)]$. The right-hand side $\mathbf{s}_v(\mathbf{x}, \omega) = (-\mathrm{i}\omega f_x, -\mathrm{i}\omega f_z)$ is the sources. The first-order and the second-order wave equations, eqs (2) and (B2), give the same solutions for the particle velocities, apart from numerical errors that result from the discretization. This occurs as long as the source coefficients of the second-order wave equation are the time derivatives of those of the first-order wave equation (check the expression of $\mathbf{s}_v$ and $\mathbf{s}_u$).

The scattered-field formulation, eq. (4), indifferently applies to the second-order wave equation with the equation

$$\mathbf{B}\mathbf{v}_s = -(\mathbf{B} - \mathbf{B}_b)\,\mathbf{v}_b. \tag{B3}$$

In the following, we shall assume that $\mathbf{v}_s = \mathbf{P}\mathbf{R}_v\mathbf{u}_s$ and $\mathbf{v}_b = \mathbf{P}\mathbf{R}_v\mathbf{u}_b$. The restriction operator $\mathbf{R}_v$ extracts the particle velocity components from the velocity–stress vector computed with the discontinuous Galerkin method, and the interpolation operator $\mathbf{P}$ projects the particle-velocity wavefield solutions computed at the nodes of the P0, P1, P2 elements with the discontinuous Galerkin method onto the barycenter of these elements, for consistency with the piecewise constant discretization of the finite-volume forward-problem operator $\mathbf{B}$ (Brossier 2011).

When $\mathbf{m}_{\text{prior}} = \mathbf{m}_k$ and $\mathbf{W}_d = \mathbf{I}$ in eq. (8), the adjoint-state method gives the following expression of the gradient of the misfit function with respect to the model parameter m_j Pratt *et al.* (1998, their eq. 25):

$$\nabla \mathcal{C}_{m_j} = \Re\left\{\mathbf{v}^T\left(\frac{\partial \mathbf{B}}{\partial m_j}\right)^T \lambda^*\right\}, \tag{B4}$$

where the state variable $\mathbf{v}$ and the adjoint-state variable λ satisfy

$$\mathbf{v} = \mathbf{v}_b - \mathbf{B}^{-1}(\mathbf{B} - \mathbf{B}_b)\,\mathbf{v}_b, \tag{B5}$$

and

$$\mathbf{B}\lambda^* = \mathbf{R}^T(\mathbf{R}\mathbf{v} - \mathbf{d}_{\text{obs}})^*, \tag{B6}$$

respectively. The adjoint-state variable corresponds to the back-propagated wavefield using the assemblage of the residuals as a composite source. In eq. (B6), we exploit the symmetry of the matrix $\mathbf{B}$ to remove the transpose operator. This has limited implications in frequency-domain modelling where a matrix or its transpose can indifferently be applied to a vector. However, the symmetry of $\mathbf{B}$ can greatly simplify the implementation of the adjoint-state method in the time domain because the same discrete forward modelling operator is used to compute the incident and adjoint wavefields.

As the second-order modelling operator $\mathbf{B}$ is difficult to discretize with the discontinuous Galerkin method, we compute the state and the adjoint variables from the first-order velocity–stress forward modelling operator, eqs (4), using the following sequences:

For the state variable,

$$\begin{aligned}\mathbf{A}\mathbf{u}_s &= -(\mathbf{A} - \mathbf{A}_b)\,\mathbf{u}_b,\\ \mathbf{v} &= \mathbf{P}\mathbf{R}_v(\mathbf{u}_b + \mathbf{u}_s),\end{aligned} \tag{B7}$$

and, for the adjoint-state variable,

$$\begin{aligned}\mathbf{A}\beta^* &= \frac{-1}{\iota\omega}\mathbf{R}_v^T\mathbf{R}^T(\mathbf{R}\mathbf{v} - \mathbf{d}_{\text{obs}})^*,\\ \lambda &= \mathbf{P}\mathbf{R}_v\beta.\end{aligned} \tag{B8}$$

Combination of the first-order and second-order forward problem operators $\mathbf{A}$ and $\mathbf{B}$ allows us, on the one hand to perform seismic modelling with P0, P1, and P2 discontinuous Galerkin method on unstructured triangular meshes, and on the other hand to derive the expression of the gradient of the misfit function from the P0 second-order forward-problem operator $\mathbf{B}$. As $\mathbf{B}$ is self-adjoint, the same forward-modelling operator is used to compute the state and adjoint-state wavefields. Moreover, only the storage of the particle-velocity wavefields is required during the gradient computation, which allows significant memory saving considering 3-D FWI applications.