



HAL
open science

On the sensitivity of teleseismic full-waveform inversion to earth parametrization, initial model and acquisition design

S Beller, V. Monteiller, L. Combe, S. Operto, G. Nolet

► To cite this version:

S Beller, V. Monteiller, L. Combe, S. Operto, G. Nolet. On the sensitivity of teleseismic full-waveform inversion to earth parametrization, initial model and acquisition design. *Geophysical Journal International*, 2018, 212 (2), pp.1344 - 1368. 10.1093/gji/ggx480 . hal-01726968

HAL Id: hal-01726968

<https://hal.science/hal-01726968>

Submitted on 16 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

On the sensitivity of teleseismic full-waveform inversion to earth parametrization, initial model and acquisition design

S. Beller, V. Monteiller,* L. Combe, S. Operto and G. Nolet

Université Côte d'Azur, CNRS, OCA, IRD, Géoazur, Valbonne, France. E-mail: beller@geoazur.unice.fr

Accepted 2017 November 2. Received 2017 August 27; in original form 2017 March 21

SUMMARY

Full-waveform inversion (FWI) is not yet a mature imaging technology for lithospheric imaging from teleseismic data. Therefore, its promise and pitfalls need to be assessed more accurately according to the specifications of teleseismic experiments. Three important issues are related to (1) the choice of the lithospheric parametrization for optimization and visualization, (2) the initial model and (3) the acquisition design, in particular in terms of receiver spread and sampling. These three issues are investigated with a realistic synthetic example inspired by the CIFALPS experiment in the Western Alps. Isotropic elastic FWI is implemented with an adjoint-state formalism and aims to update three parameter classes by minimization of a classical least-squares difference-based misfit function. Three different subsurface parametrizations, combining density (ρ) with P and S wave speeds (V_p and V_s), P and S impedances (I_p and I_s), or elastic moduli (λ and μ) are first discussed based on their radiation patterns before their assessment by FWI. We conclude that the (ρ , λ , μ) parametrization provides the FWI models that best correlate with the true ones after recombining *a posteriori* the (ρ , λ , μ) optimization parameters into I_p and I_s . Owing to the low frequency content of teleseismic data, 1-D reference global models as PREM provide sufficiently accurate initial models for FWI after smoothing that is necessary to remove the imprint of the layering. Two kinds of station deployments are assessed: coarse areal geometry versus dense linear one. We unambiguously conclude that a coarse areal geometry should be favoured as it dramatically increases the penetration in depth of the imaging as well as the horizontal resolution. This results because the areal geometry significantly increases local wavenumber coverage, through a broader sampling of the scattering and dip angles, compared to a linear deployment.

Key words: Inverse theory; Waveform inversion; Body waves; Computational seismology; Crustal imaging; Seismic tomography.

1 INTRODUCTION

Building high-resolution and quantitative images of the lithosphere from body waves is a key challenge in earthquake seismology. The methodological challenge results from the heterogeneous nature of the crust and the complex interaction of the wavefields with both these heterogeneities and the free surface. From the geodynamical viewpoint, building images of the lithosphere with the best possible resolution according to the available frequency band (namely, of the order of the wavelength) is crucial to correlate tectonic deformation in the crust with deeper mantle processes arising in the asthenosphere. Lithospheric images can be built either from regional earthquakes (earthquakes nucleated in the lithospheric target) or from distant earthquakes, located several thousands of kilometres

away from the target and commonly referred to as teleseisms. One key advantage of the teleseismic configuration is to potentially provide a large catalogue of events that are suitable for lithospheric imaging by either traveltimes tomography, receiver function analysis or waveform inversion techniques. The teleseismic wavefield can fairly be approximated by an incident planar wavefield incoming from outside of the lithospheric target with different incidence angles and backazimuths. This planar configuration is conducive to a rather uniform illumination of the lithospheric target but might be a limiting resolution factor due to the limited angular illumination provided by plane-wave sources as opposed to point sources.

Teleseismic data sets are conventionally processed by either traveltimes tomography or receiver function analysis. Traveltimes tomography, when limited to the first-arrival P -wave, provides subsurface model with a resolution of the order of the first Fresnel zone width (Williamson 1991). In the teleseismic setting where the incident planar P -wave wavefields impinge the base of the lithospheric target with incidence angles ranging between 20° and 60° , the

* Now at: Laboratoire de Mécanique et Acoustique, CNRS, Marseille, France.

tomographic models tend to exhibit significant vertical smearing of lithospheric structures along a banana-shape sensitivity kernel (Marquering *et al.* 1999), hence attesting to the limited resolution of these approaches. This vertical resolution issue has prompted some authors to enrich teleseismic data sets with either regional (Bavalia *et al.* 2016) or global (Weidle & Widiyantoro 2005) data sets to perform *P*-wave traveltimes tomography. Conversely, migration of receiver functions (Vinnik 1977; Langston 1979) can provide sharper images of lithospheric discontinuities (Ryberg & Weber 2000; Zhu & Kanamori 2000). However, stacking of *P*-wave receiver functions are inherently contaminated by secondary arrivals, that limit their use to first order PS converted waves generated by strong impedance contrast at major lithospheric discontinuities (Moho, 410 km and, 660 km boundaries).

The current densification of permanent and temporary arrays in conjunction with the continuous growth of computing facilities permit the consideration of waveform inversion methods that are expected to provide quantitative images of the subsurface with a resolution close to the wavelength. Among teleseismic waveform inversion techniques, least-squares migration of scattered teleseismic body waves was first proposed by Bostock *et al.* (2001), Shragge *et al.* (2001), Rondenay *et al.* (2001) and Rondenay *et al.* (2005). In their approach, the forward problem is linearized with the single-scattering Born approximation, Green functions are computed by dynamic ray tracing and the migration is recast as a linear waveform inversion accordingly, where the misfit between the recorded and modelled scattered wavefields is minimized in a least-squares sense. Bostock *et al.* (2001) emphasized the key role of the free surface as a secondary source of down-going *P* and *S* waves. Compared to the migration of receiver functions, which are often focused on the forward-scattered converted *P*-*S* waves, migration of reflected waves back-scattered by lithospheric reflectors after a first reflection from the free surface produces sharp *P*- and *S*-wave velocity perturbation models of the lithosphere.

A natural extension of the former linear waveform inversion is nonlinear full-waveform inversion (FWI), a technique also borrowed from exploration geophysics (Tarantola 1984; Mora 1987; Pratt 1999; Virieux & Operto 2009). The main difference with the above-mentioned linear waveform inversion is that the full wavefield residuals are minimized in FWI as opposed to the single-scattered wavefield residuals in scattering migration. This implies that the background subsurface model, in which seismic modelling is performed, is updated at each FWI iteration, while the background model is kept the same over iterations in scattering migration. A key advantage of the nonlinear inversion is that, as short-scale features are injected in the subsurface model over several inversion iterations, they can help to more fully account for converted waves and multiscattering during modelling. Conversely, a potential advantage of the linear inversion is that the explicit separation between the background wavefield and the scattered wavefield during a pre-processing stage and the use of ray tracing to select specific phases can be helpful to drive the inversion towards the best resolving information in the data (back-scattered *P*-*S* converted waves). Alternatively to FWI, Shang *et al.* (2012) proposed to image lithospheric reflectors by reverse time migration, while the background velocity models are built by reflection tomography from the free surface multiples (Burdick *et al.* 2014). This purely reflection approach discards the contribution of the incident wavefields and uncouples the update of the long wavelengths of the lithosphere by migration-based velocity analysis from the shorter ones by migration through an explicit scale separation as conventionally performed in multichannel seismic reflection processing.

In its conventional form, FWI seeks to minimize in a least-squares sense the sample-to-sample misfit between the recorded and modelled seismograms. This implies that both traveltimes and amplitudes of all of the arrivals are involved in the misfit function. One question which arises in relation to the use of the amplitudes is whether secondary parameters such as density and attenuation can be reliably updated during FWI. While multi-parameter reconstruction for density and attenuation is an active field of research in controlled-source seismology (e.g. Hicks & Pratt 2001; Malinowski *et al.* 2011; Kamei & Pratt 2013; Kurzmann *et al.* 2013; Prioux *et al.* 2013; Groos *et al.* 2014), the path taken by the earthquake seismology community has been rather to modify the sensitivity kernels of the FWI to remove amplitudes effects and recast the waveform inversion as a finite-frequency traveltimes inversion of the first arrival (Marquering *et al.* 1999) or selected wave packets (Maggi *et al.* 2009). This approach is generally referred to as adjoint tomography by the earthquake seismological community when seismic modelling is performed with full-wave numerical techniques such as the spectral element method (Komatitsch & Tromp 1999). Here, by adjoint is meant a mathematical technique that allows for the efficient computation of the gradient of a functional without forming the sensitivity matrix (see Tromp *et al.* 2005; Fichtner *et al.* 2006a,b; Plessix 2006 for a review). An alternative to the adjoint approach is the scattering-integral approach which relies on the explicit computation of the sensitivity matrix (Chen *et al.* 2007). Several applications of adjoint tomography at regional, continental and global scales have been presented by Tape *et al.* (2010), Fichtner *et al.* (2009, 2010, 2013), Zhu *et al.* (2012, 2015) and Bozdağ *et al.* (2016).

During these last years, there have been a few attempts to apply FWI to teleseismic data for lithospheric imaging. The first attempts were performed in 2-D and 2.5-D using a frequency-domain implementation (Roecker *et al.* 2010; Pageot *et al.* 2013; Baker & Roecker 2014). A frequency domain formulation was used because 2-D seismic modelling can be performed efficiently with Gauss elimination techniques when a large number of sources are processed for a few discrete frequencies (Pratt 1999). However, Pageot *et al.* (2013) showed that a fine sampling of the frequencies was required during the inversion to balance the narrow and coarse scattering angle illumination provided by a few plane wave sources. This, along with the computational cost of large matrix factorization, lead them to the conclusion that a time-domain implementation was more suitable than the frequency-domain counterpart to perform 3-D teleseismic FWI.

3-D time-domain FWI of teleseismic data has been first investigated by Tong *et al.* (2014a), who assessed the ability of FWI to reconstruct reflector geometry and density, V_p and V_s models of simple volumetric structures such as cubic inclusion or slab through a careful analysis of sensitivity kernels of specific phases. Monteiller *et al.* (2015b) assessed the resolution power of FWI to image a simple crustal model with a sharp Moho from four teleseismic events. The reliability of FWI for teleseismic application was further demonstrated with an application to real data collected in the framework of the Pyrope experiment across the Pyrénées range (Wang *et al.* 2016). FWI of five events recorded by 29 stations provided tomographic *P* and *S* wave velocity (V_p and V_s) models of the Pyrenean lithospheric structure with an unprecedented resolution. Moreover, Wang *et al.* (2016) showed the consistency between the FWI V_p and V_s models and migrated images inferred from receiver functions. In parallel with this, Beller *et al.* (2017) applied FWI on nine teleseismic events collected during the CIFALPS experiment (Zhao *et al.* 2015) in the Western Alps. They use a FWI approach

similar to Monteiller *et al.* (2015b) except that the density (ρ) was processed as an optimization parameter and jointly updated with V_p and V_s during the inversion. The FWI models show convincing images of the continental subduction and the Ivrea body as well as a low V_s zone in the lower lithosphere supporting the hypothesis of a slab detachment. The inversion was pushed up to a maximum frequency of 0.2 Hz. The ρ model shows structures that are consistent with those shown in the V_s model in the upper lithosphere. However, the maximum depth at which ρ can be imaged is smaller than for V_s suggesting that ρ has been mostly reconstructed from free-surface multiples. Another striking result is the good resolution of the imaging in the cross-line direction, perpendicular to the main receiver line of the CIFALPS experiment, although a quite sparse distribution of stations in the cross-line direction. In particular, Beller *et al.* (2017) successfully recovered the arc shape of the Ivrea body, whose ρ model allows it to match the Bouguer anomaly.

Despite this recent success, there is still a need to appraise the promises and pitfalls of 3-D teleseismic FWI. First, the plane-wave configuration raises the issue of the resolution with which lithospheric structures can be imaged from a few teleseismic events. Although Bostock *et al.* (2001) have emphasised the key role of the free surface multiples, the extent to which second-order back-scattered waves can be exploited by FWI when the full wavefield is processed in one go, that is, without explicit separation between the incident and scattered wavefields, is not yet clear. A second issue is related to the choice of the best subsurface parametrization (the parameter set that fully describes the subsurface) and the optimization parameters (the subset of parameters that are updated during the inversion as opposed to the passive parameters that are kept fixed) for teleseismic FWI. This issue is far from being neutral as long as one seeks to update secondary parameters such as density, attenuation or anisotropic parameters (see Operto *et al.* 2013, for a tutorial). The choice of the subsurface parametrization and the optimization parameters is mainly driven by the need to mitigate parameter cross-talk, while preserving a sufficient resolution in the imaging. The most common tools to define a suitable parametrization for FWI relies on the analysis of the so-called radiation patterns (Tarantola 1986; Forgues & Lambaré 1997; Gholami *et al.* 2013; Alkhalifah & Plessix 2014), the principal component analysis of the sensitivity matrix (Sieminski *et al.* 2009) or the eigenvector analysis of the Hessian operator (Plessix & Cao 2011).

A third issue is related to the role of the initial model in teleseismic FWI. It is well acknowledged that this initial model should allow the prediction of traveltimes with an error that does not exceed half the period to prevent cycle skipping artefacts when a conventional least-squares difference-based misfit function is used (e.g. Virieux & Operto 2009). This condition is challenging to fulfil when a large number of wavelengths are propagated. This arises when seismic data lack low frequencies as in controlled-source seismology and the acquisition geometry involve long propagation distances. The low frequency content of teleseismic data should make FWI reasonably immune to cycle skipping. However, it is unclear if other factors than the kinematic accuracy of the initial model, for example the smoothness versus blocky character of the velocity model or site effects near the free surface, can drive the FWI towards a local minimum. A fourth obvious factor is related to the footprint of the station spread and sampling in teleseismic FWI. Clarifying this issue is crucial to design reliable surveys amenable to high resolution imaging methods such as FWI. Using too coarse stations network can generate aliasing artefacts of different nature (Shragge *et al.* 2001; Rondenay *et al.* 2005; Pageot *et al.* 2010). This has prompted the design of several teleseismic acquisitions

in the form of dense profile of stations spaced 5 km apart at the expense of coarser areal deployments (Rondenay *et al.* (2001, Cascadia experiment), Iglesias *et al.* (2010, MASE experiment), Zhao *et al.* (2015, CIFALPS experiment). While such a dense linear set-up might be suitable to sample teleseismic events arriving with an azimuth aligned with the receiver line, it may not be suitable to record waves that are scattered off the receiver line or process events reaching the array with a significant obliquity.

Among other parameters that may impact the reconstruction of lithospheric models by teleseismic FWI, we do not consider the effects of the heterogeneities that are located outside the lithospheric box. This issue has been recently investigated by Masson & Romanowicz (2017) for both traveltime and full-waveform tomography. For full-waveform tomography, they have shown that acceptable results can be obtained even when no *a priori* information other than a 1-D reference model is available outside the lithospheric target (Masson & Romanowicz 2017, their fig. 11). However, the FWI results can be significantly improved when a smooth traveltime tomography model is used rather than a 1-D reference model as a background model outside the lithospheric target (Masson & Romanowicz 2017, their fig. 14). We do not consider either the effects of the source wavelet estimation and noisy data. One issue with the source signature estimation is the potential trade-off with the subsurface parameter updates, in particular when these two optimization processes are performed in an alternating manner in the inversion iterations. This prompts us to estimate the source signature prior to the FWI and keep it fixed afterwards during the CIFALPS real data case study (Beller *et al.* 2017). This prior source estimation can be useful to mitigate the effects of inaccurate source location and moment tensor estimation, and also absorb the effects of remote large-scale heterogeneities located outside the lithospheric target as these heterogeneities will mainly generate a traveltime perturbation of the nearly planar incident teleseismic wavefields (Monteiller *et al.* 2015a). Since we aim to keep the conclusions of this study as general as possible, we do not add noise to the data during the synthetic experiments since the amount of noise as well as the ability of the acquisition to mitigate its effect by stacking is really case dependent. Also, for a real data application, noisy data can thoroughly be rejected or pre-processed as it was illustrated in Beller *et al.* (2017).

The aim of this study is to gain new insights on the three above-mentioned issues through the application of 3-D time-domain FWI on a realistic synthetic case study inspired from the CIFALPS experiment.

In the first part of this study, we review the main theoretical ingredients of the FWI. Then, we present the synthetic lithospheric model of the western Alps and describe the anatomy of the data computed in this model. We interpret the dominant arrivals recorded by the three geophone components and identify part of the signal that underwent aliasing. This is followed by a parametric analysis of the FWI. We first assess three different subsurface parametrizations: (ρ, λ, μ) , (ρ, v_p, v_s) and (ρ, I_p, I_s) , where λ and μ are the Lamé parameters and I_p and I_s the P and S impedances. Our numerical examples suggest that the (ρ, λ, μ) parametrization provides the best FWI results after recombining *a posteriori* the ρ , λ and μ optimization-parameter models into I_p and I_s . Second, we assess the sensitivity of the FWI to the initial models. We show that a smooth tomographic model is the most suitable one for FWI, although 1-D reference models as PREM are enough to achieve acceptable FWI results. Third, we assess the sensitivity of the teleseismic FWI to two different kinds of acquisition geometry: coarse areal configuration versus dense linear acquisition. Our results strongly support that

areal deployment should be first designed at the expense of a dense linear deployment for the application of teleseismic FWI. For such a configuration, areal deployments allows to increase the range of usable data, specifically, those with a high signal-to-noise ratio emanating from earthquakes not aligned with a linear array. Besides, 2-D arrays can sample a broader wavenumber spectrum than linear arrays, and hence improve the expected resolution of FWI images at the expense of data redundancy. This redundancy can nonetheless be achieved by including events occurring in a common tectonic area.

2 TELESEISMIC FULL-WAVEFORM INVERSION

This section reviews some basic principles of FWI that will be useful to interpret the results of the numerical experiments carried out in this study. For sake of compactness, we review these principles with a matrix formalism suitable for the frequency-domain formulation of FWI (Pratt *et al.* 1998), although our teleseismic FWI is fully implemented in the time domain.

2.1 Theoretical background

Let's assume that a modelling engine such as those based upon hybrid methods is available to compute synthetic teleseismic seismograms (Monteiller *et al.* 2013; Tong *et al.* 2014b; Beller *et al.* 2017). The elastodynamic equations can be written in matrix form as

$$\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{s}, \quad (1)$$

where the vectors \mathbf{s} and \mathbf{u} denote the seismic source and the synthetic wavefield, respectively, and the matrix \mathbf{A} is the wave modelling operator, the coefficients of which depend on frequency (or time) and subsurface properties gathered in the vector \mathbf{m} . The most natural parametrization of the subsurface involve the density ρ and elastic moduli $\mathbf{c} = c_{ijkl}$ since this parametrization directly results from the equation of motion and the Hooke's constitutive law.

FWI is a data fitting procedure that aims to find the subsurface parameters \mathbf{m}^* that minimize a certain measure \mathcal{C} of the data misfit, in its most conventional form, the least-squares norm of the sample-to-sample difference between the observed and the modelled data

$$\mathbf{m}^* = \arg \min_{\mathbf{m}} \mathcal{C}(\mathbf{m}), \quad (2)$$

where $\mathcal{C}(\mathbf{m}) = \frac{1}{2} \|\Delta \mathbf{d}\|_2^2$ and $\Delta \mathbf{d} = \mathbf{R}\mathbf{u}(\mathbf{m}) - \mathbf{d}$ is the difference between the recorded data \mathbf{d} and the modelled wavefield sampled at the receiver positions by the operator \mathbf{R} . Because of the computational burden of the seismic modelling and the huge dimension of the model space, FWI is recast as an iterative local optimization problem that seeks the minimum of the misfit function in the vicinity of a starting model \mathbf{m}_k at iteration k . Zeroing the gradient of the misfit function after a second-order Taylor expansion of the misfit function (2) around the starting model provides the Newton descent direction. Multiplying this descent direction with a step length γ_k , that can be found by a line search procedure (Nocedal & Wright 2006), provides the model perturbation $\Delta \mathbf{m}_{k+1}$ at iteration k

$$\Delta \mathbf{m}_{k+1} = -\gamma_k \mathbf{H}^{-1}(\mathbf{m}_k) \nabla_{\mathbf{m}} \mathcal{C}(\mathbf{m}_k), \quad (3)$$

where \mathbf{H} and $\nabla_{\mathbf{m}} \mathcal{C}$ are the Hessian (the second derivative of the misfit function) and the gradient of the misfit function with respect to \mathbf{m} , respectively.

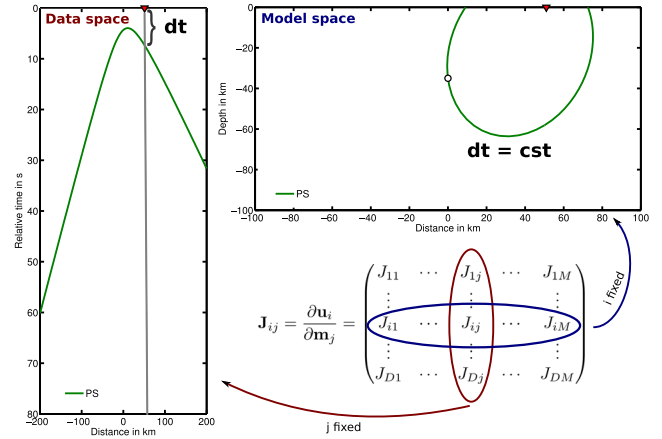


Figure 1. Sensitivity matrix and FWI imaging principle. The column j of the sensitivity matrix describes the synthetic wavefield that would be scattered by a diffractor located at m_j and sampled at the stations positions. In this figure, we show the case of a P - S diffraction. The support of this scattered wavefield in the time-distance domain follows a diffraction hyperbola. The gradient of the misfit function is formed by a weighted stack of the data residuals along these trajectories. Conversely, the line i of the sensitivity matrix describes the positions in space along which a sample of the data residuals $\partial \mathbf{u}_i$ is projected. These positions describe an isochronal ellipse as in scattering migration. Summation over all possible ellipses focuses by constructive interference the image of the scatterer with a resolution controlled by the source bandwidth and the angular illumination provided by the acquisition geometry (adapted from Rondenay *et al.* 2005).

2.2 Gradient of the misfit function

The gradient of the misfit function is given by

$$\nabla_{\mathbf{m}} \mathcal{C} = \left(\frac{\partial \mathbf{R} \mathbf{u}(\mathbf{m})}{\partial \mathbf{m}} \right)^T \Delta \mathbf{d} = \mathbf{J}^T \Delta \mathbf{d}, \quad (4)$$

where \mathbf{J} denotes the Fréchet derivatives matrix or sensitivity matrix (Pratt *et al.* 1998). Differentiation of eq. (1) with respect to a model parameter m gives

$$\mathbf{A} \frac{\partial \mathbf{u}}{\partial m} = -\frac{\partial \mathbf{A}}{\partial m} \mathbf{u}. \quad (5)$$

Analogy between eqs (1) and (5) shows that the partial derivative wavefield $\partial \mathbf{u} / \partial m$ is the solution of the elastodynamic equations for a secondary virtual source $-(\partial \mathbf{A} / \partial m) \mathbf{u}$. The sparse operator $\partial \mathbf{A} / \partial m$ defines the spatial and temporal supports of the virtual source at the position of m in space and at the arrival time of \mathbf{u} at m in time, respectively as well as its radiation pattern. Therefore, the partial derivative wavefield can be interpreted as the wavefield scattered by a point diffractor located at m . The traveltime curve followed by this partial derivative wavefield sampled at the receiver positions (this sampling gives one column of the sensitivity matrix) describes a diffraction hyperbola (Fig. 1). The gradient of the misfit function is simply formed by the zero-lag correlation between the sampled partial derivative wavefield and the data residuals or in other words by the summation along the diffraction hyperbola of the product between the partial derivative wavefield and the data residuals. The aim of this correlation is to pick among all of the residuals those which result from a missing heterogeneity located at m .

Although eqs (4) and (5) draw a clear connection between FWI and diffraction tomography, the gradient of the misfit function can be more efficiently computed with the adjoint-state method at the cost of two forward simulations per source (Plessix 2006, for

a review). Following Pratt *et al.* (1998), injecting the expression of the partial derivative wavefield (5) into (4) gives

$$\nabla_m \mathcal{C} = -\mathbf{u}^T \left(\frac{\partial \mathbf{A}}{\partial m} \right)^T (\mathbf{A}^{-1})^T \Delta \mathbf{d}^* = - \left(\frac{\partial \mathbf{A}}{\partial m} \mathbf{u} \right)^T \mathbf{u}^{\dagger*}, \quad (6)$$

where the adjoint wavefield \mathbf{u}^\dagger satisfies

$$\mathbf{A}^T \mathbf{u}^{\dagger*} = \mathbf{R}^T \Delta \mathbf{d}^*. \quad (7)$$

This expression shows that the gradient is obtained by a zero-lag correlation of the virtual source $-(\partial \mathbf{A} / \partial m) \mathbf{u}$ with the adjoint wavefield \mathbf{u}^\dagger , that is back-propagated in the subsurface using the residuals as a composite source term. Considering only one sample of the data residuals associated with a source s , a receiver r and a traveltime T , its energy will be smeared in the subsurface along positions x that satisfy the imaging condition $T_{sx} + T_{rx} = T$. Here, T_{sx} and T_{rx} denote the traveltimes of a wave connecting the source and the receiver to the scatterer x , respectively. These positions describe an isochronal surface, which is an ellipse in a 2-D homogeneous background medium (Fig. 1).

All of these scattering concepts and imaging condition are typically met in migration techniques (such as ray+Kirchhoff or ray+Born migration) (McMechan & Fuis 1987; Miller *et al.* 1987; Lumley *et al.* 1994; Rondenay *et al.* 2005) and highlight the connections that can be drawn between FWI and generalized diffraction tomography (Devaney 1984; Wu & Toksöz 1987) as reviewed by Pratt *et al.* (1998).

For a general elastic medium parametrized by the density ρ and the coefficients $\mathbf{c} = c_{ijkl}$ of the stiffness tensor, the gradient is simply given by:

$$\begin{aligned} \nabla_\rho \mathcal{C}(\mathbf{x}) &= \int_0^T \mathbf{u}^\dagger(\mathbf{x}, t) \cdot \partial_t^2 \mathbf{u}(\mathbf{x}, t) dt, \\ \nabla_c \mathcal{C}(\mathbf{x}) &= \int_0^T \epsilon^\dagger(\mathbf{x}, t) :: \epsilon(\mathbf{x}, t) dt, \end{aligned} \quad (8)$$

where \mathbf{c} denotes the stiffness tensor, \mathbf{u} and \mathbf{u}^\dagger the displacement and adjoint displacement wavefields, and ϵ and ϵ^\dagger the strain tensor and the adjoint strain tensor, respectively (Tromp *et al.* 2005). Other parametrizations may be considered and deduced from the former using the chain-rule of derivatives (Köhn *et al.* 2012), three of them are inspected in Section 2.4.

2.3 Resolution analysis

We review here the main experimental factors that control the spatial resolution of the FWI. In the framework of diffraction tomography and FWI respectively, Wu & Toksöz (1987) and Sirgue & Pratt (2004) showed that the gradient of the misfit function with respect to one parameter has the form of a truncated inverse Fourier series in which the arguments of the basis functions are the wavenumber components locally injected in the targeted model at the position of the parameter. The truncation of this Fourier series, which results from the limited bandwidth of the source and the limited spread of the acquisition, limits the resolving power of FWI.

In the high-frequency approximation (or local plane wave approximation), the local wavenumber vectors \mathbf{k} injected at a given position in a 2-D subsurface model (Fig. 2a) can be inferred from the ray paths connecting the sources and the receivers to the scatterer leading to the following expression

$$\mathbf{k} = \mathbf{k}_s + \mathbf{k}_r = (k_x, k_y, k_z) = k(\sin \phi \cos \psi, \sin \phi \sin \psi, \cos \phi) \quad (9)$$

where

$$k = \frac{2\omega}{c} \cos\left(\frac{\theta}{2}\right) \quad (10)$$

and c is the local wave speed, ω the angular frequency, ϕ , ψ and θ are the dip, azimuth and scattering angles, respectively (Miller *et al.* 1987; Wu & Toksöz 1987; Lambaré *et al.* 2003).

The expression of \mathbf{k} shows that the local wavelength and the angles ϕ , ψ and θ are the four parameters that control the spatial resolution of the FWI. In teleseismic applications when a few up-going wavefields can lead to a coarse and narrow sampling of θ , the use of a broad frequency band is useful to guarantee a proper sampling of the wavenumbers in the spatial directions spanned by the dip ϕ and azimuth ψ angles. Moreover, free-surface multiples may enrich the range of dip and short scattering angles sampled by the acquisition, hence increasing the vertical resolution of the imaging.

Beyond the resolution limitations imposed by the limited frequency bandwidth of the source and the limited spread of the acquisition, the radiation pattern of the virtual sources further controls the amplitude versus θ variations of the partial derivative wavefields and hence potentially narrow the real range of wavenumbers mapped in the FWI gradient. In the next section, we analyse the radiation patterns of the virtual sources for the three subsurface parametrizations that are used in this study for a theoretical assessment of their impact on the FWI resolution and the trade-off between parameters.

2.4 Radiation patterns for different subsurface parametrization

Inferences on the sensitivity of the wavefield to different parameter classes can be drawn from the radiation patterns of elastic waves scattered by small elastic heterogeneities (Wu & Aki 1985; Tarantola 1986). Such patterns give clear insights on the influence of a single parameter onto the data for different scattering regime (forwards versus backwards) (Fig. 2a). Moreover, the overlap of the radiation patterns of different parameters with θ indicates potential trade-off between these parameters. On the one hand, one may want to mitigate these trade-offs by choosing a parametrization that minimizes these overlaps. On the other hand, this strategy might narrow the influence of the parameters with θ , hence degrading the resolution with which parameters are reconstructed accordingly. Many different parametrizations can be considered for FWI. However, in the case of isotropic elasticity, the most common ones combine ρ with either wave speeds (V_p , V_s), impedances (I_p , I_s) or elastic moduli (λ , μ). The radiation patterns of these three parametrizations are shown in Figs 2(b)–(d) for a 2-D homogeneous medium. The presentation of the radiation patterns is complemented with multi-parameter FWI gradients computed for an incident P wave with an incidence angle of 45° and one station (Fig. 3). The background medium is a homogeneous half space to which is added a small circular inclusion at 50 km depth to generate the data residuals.

A first noticeable feature, that is shared by the three considered parametrizations, is that λ (resp. V_p and I_p) generates only P – P scattering with an isotropic radiation pattern (Figs 2b–d, (4–6), blue curve). Therefore, the radiation pattern of λ (resp. V_p and I_p) has no impact on the resolution with which this parameter is reconstructed: the local resolution is only controlled by the local P wavelength and the range of P – P scattering angles spanned by the acquisition. The FWI gradient with respect to λ (resp. V_p and I_p) highlights the dominant contribution of the forward-scattered P – P mode delineated by the half-ellipse isochronal curve with one focus

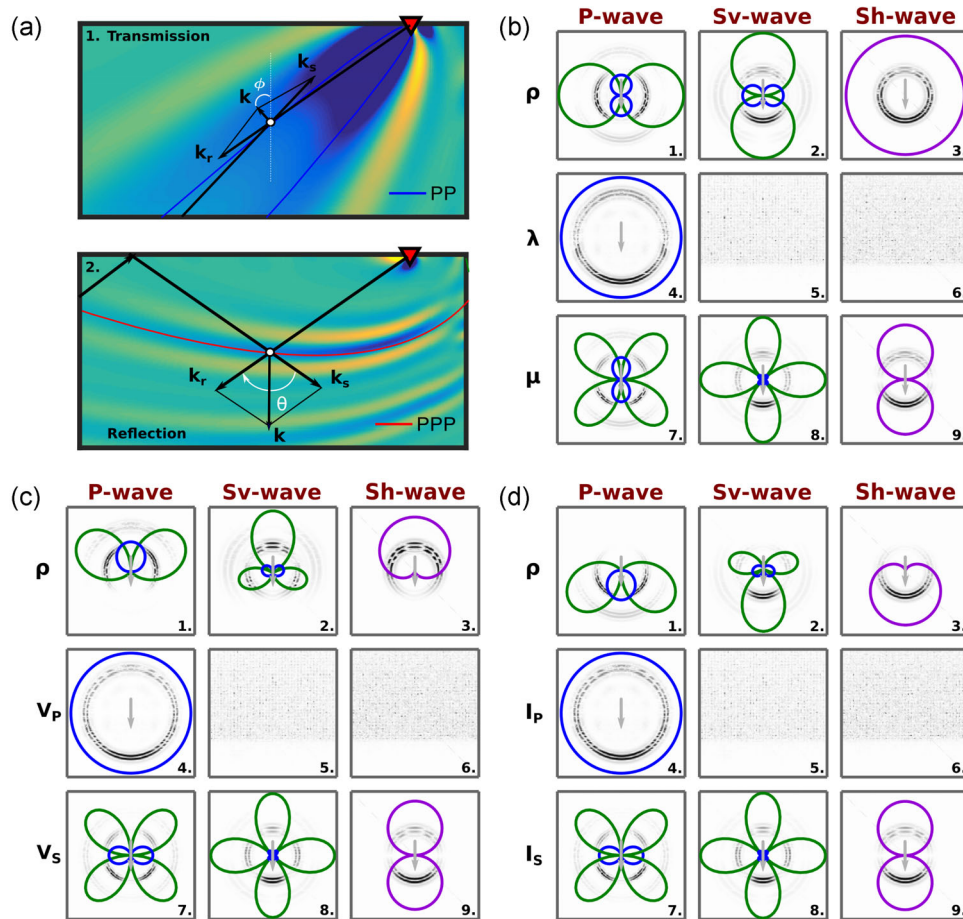


Figure 2. Radiation patterns for elastic isotropic media. (a) Schematic illustration of the relationship between the wavenumber components mapped in the subsurface at a given position with the angular illumination provided by the acquisition and the reflectors of the background subsurface model (here, the free surface). In a teleseismic setting, the forward scattering generated during the upgoing propagation of the incident wavefield towards the surface will contribute to preferentially map the low-to-intermediate horizontal wavenumbers, while the backward scattering generated after the reflection of the incident wavefield from the free surface will contribute to preferentially map the high-to-intermediate vertical wavenumbers. Radiation patterns of elastic scatterers for the (b) (ρ , λ , μ), (c) (ρ , V_p , V_s) and (d) (ρ , I_p , I_s) parameterizations. From left to right, each column presents the interaction of a P (1,4,7), S_v (2,5,8) and Sh -wave (3,6,9) incident plane wave with, from top to bottom, a ρ (1–3), λ (4–6) (resp. V_p and I_p) and μ (7–9) (resp. V_s and I_s) scatterer. All the analytical radiation patterns (Wu & Aki 1985) are superimposed with the scattered wavefields that are computed numerically. Both patterns are scaled with respect to their respective maximum scattered energy. The anisotropic behaviour of the amplitudes of numerical patterns translates a slight deviation from Born theory (the heterogeneity is too wide comparatively to the propagated wavelength). Arrows indicate the direction of the incident plane wave (black for P -wave, red for S_v wave and red-black for Sh wave). Radiation-pattern curves are plotted in blue, green and purple for the P , S_v and Sh scattered modes, respectively. See the text for the interpretation.

at the station (Figs 3b,e and h, blue curve). This forward-scattering regime is associated with wide θ and will contribute to update the long wavelengths of λ (resp. V_p and I_p) (Fig. 2a). Superimposed on the forward-scattered P - P sensitivity kernel, we also see, with much smaller amplitudes, the migration isochrone associated with the doubly P - P back-scattered wave from the free-surface and the inclusion (Figs 3b, e and h, red curve). This back-scattering regime is associated with smaller θ and hence will contribute to update shorter wavelengths of λ (resp. V_p and I_p) (Fig. 2a). Since the forward-scattering regime is dominant, the FWI will tend to update the long wavelengths first before updating the shorter ones from the back-scattered waves once the residuals of the incident wave will have been reduced. This hierarchy between forward and backward scattering naturally honours a multi-scale approach which is useful to mitigate the nonlinearity of the FWI.

In contrast, μ (resp. V_s and I_s) gives rise to P - P , P - S , S - S and S - P scattering, with stronger diffracted S waves relative to the P counterparts (Figs 2b–d, 7–9). The P - S , S - S and S - P radiation

patterns are identical for the μ , V_s and I_s and the union of these three radiation patterns covers a broad range of θ that is conducive to a broad-band reconstruction of these parameters if the imprint of all of these scattering modes is significant in the data. However, remembering that only incident P waves are considered in this study, the two lobes of the P - S_v radiation pattern pointing forward in the direction of the arrow with an angle of 45° in Figs 2(b)–(d) (7, green curves) show that the long wavelengths of μ (resp. V_s and I_s) cannot be updated from the P - S_v mode since there is no radiation at very wide θ . This is supported by the gradient with respect to μ (resp. V_s and I_s) in Figs 3c, f and i, dark green curve, which shows the more focused spatial domain spanned by the P - S_v forward scattering mode relative to the P - P one (Figs 3b, e and h, blue curve). Note also that the amplitudes of the P - S_v forward scattering mode are small compared to those of the doubly back-scattered modes in Figs 3(c), (f) and (i) due to the low values of the P - S transmission coefficient relative to those of the P - S_v reflection coefficient at the free surface (Pageot *et al.* 2013, their fig. 3).

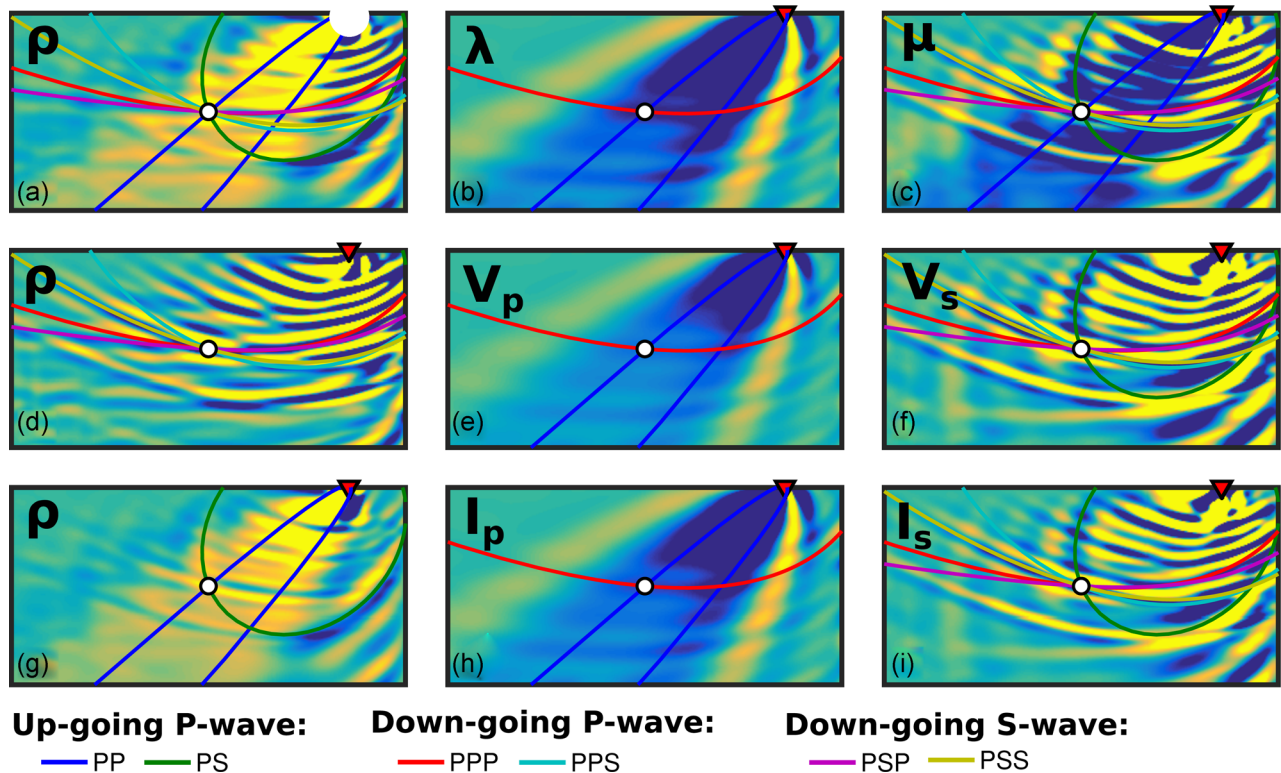


Figure 3. Anatomy of a teleseismic gradient for one source–receiver pair and different parametrizations. FWI gradients computed for (a–c) (ρ , λ , μ), (d–f) (ρ , V_p , V_s) and (g–i) (ρ , I_p , I_s) parametrizations. The source is a 45° incident plane-wave travelling through an homogeneous half-space with a small heterogeneity located on the centre of the figure (white circle). The gradient is computed for a three-component sensor located at the position of the red triangle. The isochronal curves of the scattered-waves generated through the interaction of the upgoing P -wave and downgoing P and S -wave with the small heterogeneity are superimposed on the gradients. See the text for interpretation.

Interestingly, the only difference between the radiation patterns of μ , V_s and I_s is that μ gives rise to two lobes of P – P scattering at wide and small θ (Fig. 2b,(7), blue curve), while V_s and I_s give rise to P – P scattering at intermediate θ (Figs 2c and d,(7), blue curve). This difference manifests in Fig. 3(c), blue curve, by the clear imprint of the P – P forward-scattering mode in the μ gradient within a half-ellipse sensitivity kernel, unlike in the V_s and I_s gradients. Owing that the long wavelengths of μ (resp. V_s and I_s) cannot be updated from the P – S_v scattering mode as above-mentioned, the ability to update the long wavelengths of μ from the P – P mode might be an argument in favour of the (ρ , λ , μ) parametrization in particular if the inversion is started from a crude initial μ model.

Potential trade-off between λ (resp. V_p and I_p) and μ (resp. V_s and I_s) would result from the overlap between the P – P radiation patterns of μ (resp. V_s and I_s) and λ (resp. V_p and I_p). These trade-offs might be more significant during the update of the long wavelengths of λ and μ because the P – P forward-scattering sensitivity kernel of these two parameters are very similar (Figs 3b and c). This comment highlights the fact that it is often challenging to update multiple parameter classes with a good resolution, while minimizing the trade-offs between them.

The radiation patterns of ρ show significant differences for the three parametrizations, although ρ generates scattering for the P – P , P – S , S – S and S – P scattering modes in the three cases (Figs 2b–d,(1–3)). For (ρ , λ , μ), the radiation patterns are symmetric with respect to the incident plane wave front (Fig. 2b, (1–3)). In other words, forward and backward scattering are equally generated by ρ with this parametrization. Moreover, combining the contribution of the P and S scattered waves provides a sensitivity over the full range

of scattering angles. Conversely, ρ scatters waves in a preferential direction with the two other parametrizations: backward for (ρ , V_p , V_s) (Fig. 2c,(1–3)) and forward for (ρ , I_p , I_s) (Fig. 2d,(1–3)). This implies that (ρ , I_p , I_s) is suitable to update the long to intermediate wavelengths of ρ , while (ρ , V_p , V_s) is more suitable to update the short to intermediate wavelengths. We might conclude that (ρ , λ , μ) might be the most suitable one to update ρ because the first-order P – P and P – S forward-scattering modes can yet generate models of good resolution, which can be further improved by the contribution of the doubly S_v – P and S_v – S_v back-scattered waves. Note also that the Sh – Sh radiation pattern is isotropic and hence can contribute to significantly increase the resolution of ρ .

This theoretical analysis is supported by the gradients with respect to ρ (Figs 3a, d and g). For (ρ , λ , μ), we show well-balanced contributions of the forward- and backward-scattering regimes (Fig. 3a), while the broad P – P elliptical sensitivity kernel of the forward-scattering (delineated by the blue curve) is dominant in the (ρ , I_p , I_s) parametrization (Fig. 3g). Conversely, the more spatially focused migration-like isochronal surfaces of the backward scattering are dominant for the (ρ , V_p , V_s) parametrization (Fig. 3d).

Significant trade-offs are expected between λ (resp. V_p , I_p) and ρ and between μ (resp. V_s , I_s) and ρ according to their respective radiation patterns. However, it is worth remembering that λ (resp. V_p , I_p) and μ (resp. V_s , I_s) control both the kinematic and the dynamic attributes of seismic waves, while density mostly influences amplitudes. Heuristically, we may think that the long-wavelength reconstructions of λ (resp. V_p , I_p) and μ (resp. V_s , I_s) will be primarily tied to the need to fit traveltimes and that amplitudes will play a second-order role in their update, hence limiting the imprint

of these parameter trade-offs. Controlling the trade-off between the update of the small to intermediate wavelengths of ρ and λ (resp. V_p , I_p) or ρ and μ (resp. V_s , I_s) from reflections at short to intermediate scattering angles is probably more challenging.

A last comment concerns the presence of ghost isochronal surfaces in the gradients that have been focused off the position of the inclusion (Fig. 3). These isochrones result from the misinterpretation by the gradient of either (1) multiple-scattered waves or (2) different modes of converted waves generated by unmodelled heterogeneity. These artefacts can be partially cancelled out during the gradient computation by the destructive interferences of the contributions of several receivers and events. Moreover, the Gauss-Newton Hessian can further contribute to remove these artefacts by a deconvolution-like processing since the zero-lag correlation between the partial derivative wavefields will mimic the same undesired correlations as the zero-lag correlation between the partial derivative wavefields and the data residuals. Note that the artefacts related to multiple-scattered arrivals generated by the discontinuities of the background model should not be confused with those generated by multiple scattering from unmodelled heterogeneities. These multiple-scattered arrivals are present in the recorded data but are not modelled during the FWI gradient calculation, which relies on the single-scattering approximation. In this case, the artefacts generated by the recorded multiple-scattered arrivals can be removed from the FWI gradient by the second-order (nonlinear) term of the Hessian.

3 THE WESTERN ALPS LITHOSPHERIC MODEL

3.1 Model description

Beller *et al.* (2017) applied FWI on nine teleseismic events recorded by the CIFALPS temporary array to image the lithospheric structure of the South-Western Alps (Fig. 4a). In the light of this study and for the purpose of the current study, we design a lithospheric model that is representative of the retrieved structure of the Alpine lithosphere.

The synthetic model (Fig. 4b) represents the subduction of the European continental crust below the Adriatic plate in the Western Alps. This model is 400 km \times 200 km long and 200 km deep. In its shallowest part, the model exhibits slow velocities/low density sedimentary basins, that are 10 km thick to mimic the presence of the SE-France and Po plain sedimentary basins (Fig. 4a, red). The upper European crust is indented by a piece of Adriatic lithospheric mantle (Fig. 4b, light blue) referred to as the Ivrea body mantle wedge. From the West to the East, the model displays two distinct lower crusts (Fig. 4a, green), the first dips towards the east up to a maximum depth of 100 km and characterizes the subduction of the European continental crust that under-thrusts the Ivrea body, whilst the other depicts the more conventional lower crust of the Adriatic plate. Beneath that, we designed a lithospheric mantle that gently varies from 100 to 125 km depth beneath the Ivrea body. The asthenospheric mantle located below also contains a piece of detached slab at the bottom end of the model (Fig. 4b).

Overall velocities of the synthetic model are based on the IASP91 model (Kennett & Engdahl 1991) and densities are taken from the PEM-C model (Dziewonski *et al.* 1975). Seismic modelling in the lithospheric target is performed with a grid injection technique described in Beller *et al.* (2017): we first perform a global-earth simulation in the PREM model (Dziewonski & Anderson 1981) with the AxiSEM software (Nissen-Meyer *et al.* 2014) and store the 3-D teleseismic wavefield at the boundaries of the lithospheric target.

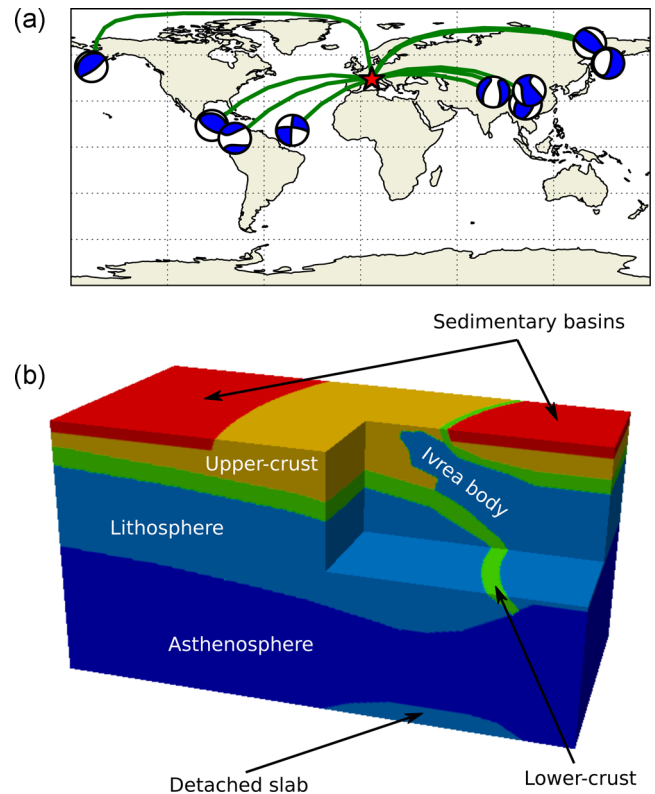


Figure 4. Synthetic example design. (a) The nine teleseismic events selected from the CIFALPS data set for FWI. (b) The 3-D Alps lithospheric model representative of a continental collision.

Second, the teleseismic wavefield is propagated in the lithospheric target with the spectral element method (Komatitsch *et al.* 1998) through a grid injection technique (Monteiller *et al.* 2013). To avoid numerical artefacts during the grid injection, we force a smooth transition between our synthetic model and the outer PREM model with a cosine taper in a 20 km thick layer along the boundaries of the lithospheric target. For SEM modelling, the lithospheric target is discretised with a hexahedral mesh of 10 km wide elements. This leads to 2 000 000 degrees of freedom in the mesh. We perform SEM modelling in this lithospheric model using 9 computer nodes of 24 cores each. With this computational resources, one FWI gradient computations takes around 3 min of wall time.

3.2 Teleseismic data anatomy

Since FWI is a data-fitting procedure, it is relevant to interpret the potential information content of the teleseismic wavefield and correlate ‘with one’s own eyes’ the geological features of the synthetic model with their seismic response. For that purpose, we show in Fig. 5 synthetic seismograms recorded by a dense profile of stations sampled 5 km apart for an event, the characteristics of which are given in Table 1.

The relatively shallow depth (12 km) of this event together with the dominant period of the source signature (≈ 5 s) prevents the identification of source-related multiples such as pP and sP phases, which are hidden within the coda of the direct P -wave arrival (their theoretical arrival times from the onset of the direct P arrival are 3.95 and 5.6 s, respectively). In contrast, the intermediate epicentral distance of the event (63°) allows the PcP phase to be well-separated from the incident P -wave within the displayed time window. This phase has a theoretical arrival time of 37.6 s, according to PREM

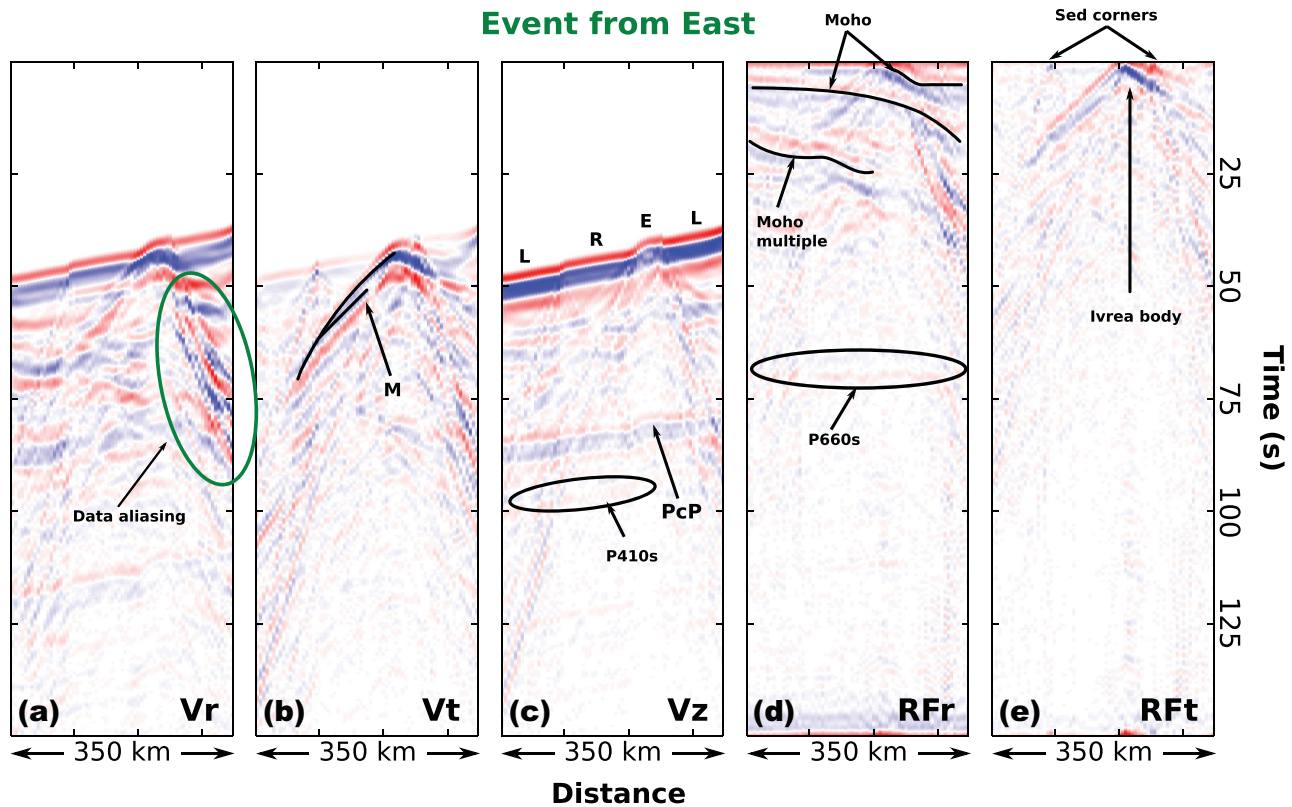



Figure 5. Synthetic data computed in the Alps lithospheric model. From left to right: radial, transverse and vertical components of particle velocities followed by radial and transverse receiver functions corresponding to the event 22 in Beller *et al.* (2017). E: early arrivals; L: late arrivals; R: reference arrivals; M: Moho reflection. Black ellipses indicate P -to- S wave conversions at upper-mantle reflectors (P410s and P660s), while the green ellipse highlights data aliasing of steep diffractions generated by the Ivrea body.

Table 1. Selected event for teleseismic waveform modelling indicating origin time, location and moment tensor solution from Global CMT project (Dziewonski *et al.* 1981; Ekström *et al.* 2012). Half duration (Hdur) and time shift (Tshift) are given in seconds.

Id	Event	Lat (°) Lon (°)	Dist (°) Baz (°)	Hdur Tshift	Depth (km)	Mw CMT
22	2013-02-14	67.65 142.512	62.98 17.37	5.30 6.26	12	6.7 

and the TauP toolkit (Crotwell *et al.* 1999), and is characterized by a sub-planar move-out (Fig. 5c).

At first glance, owing to the low frequency content of the data, it is challenging to identify reflections from major lithospheric boundaries in the recorded data, which are dominated by the incident P wavefield and energetic diffraction from the Ivrea body. However, the footprint of several lithospheric structures, such as sedimentary basins and Ivrea body, is clearly visible in the onset time of the direct P arrival. In particular, the slow/buoyant sedimentary basins delay the travelttime of the direct P arrival and increase its amplitudes (Fig. 5, L), while their pinch-outs generate energetic diffractions with sharp dips (Fig. 5, Sed corners). In contrast, the fast Ivrea body advances the direct P arrival onset and decreases its amplitudes (Fig. 5, E). Moreover, it behaves as a strong secondary point source near the surface, that scatters waves downwards in the lithosphere. With respect to this comment, we interpret a late arrival with a hyperbolic moveout as the reflection from the Moho generated from this diffracting source (Fig. 5, M). The signature of the upper-mantle discontinuities corresponding to the top and

bottom of the transition zone at 410 and 660 km depth are also visible at around 40 and 70 s after the first arrival (Fig. 5, P410s and P660s). These discontinuities are not part of the lithospheric target and hence will not be imaged by FWI. However, their signature in the data will not impact upon the imaging since they will be injected in the modelled synthetic seismograms by the grid injection technique.

The computation of receiver functions by frequency-domain deconvolution (Langston 1979) further helps us to analyse the contribution of late reflections from crustal heterogeneities by removing the imprint of the incident wavefield (Fig. 5d and e). The radial receiver function shows the geometry of the Moho of the European and Adriatic plate, with a pronounced dip for the former. On the transverse component, receiver functions are dominated by three diffraction hyperbola created by the pinch-out of the sedimentary basins and the Ivrea body. These arrivals are spatially aliased due to their significant dips. Later arrivals might indicate the location of the lithosphere–asthenosphere boundary (LAB). However, as is frequently the case in P -wave receiver function analysis, the signature of the LAB is hampered by the Moho free-surface multiples, which makes its identification challenging.

4 PARAMETRIC ANALYSIS OF TELESEISMIC FWI

4.1 Experimental setup

For all the experiments involved in the parametric study, we perform FWI for the same nine events as those used for the CICALPS

experiment (Fig. 4a). We refer the reader to Beller *et al.* (2017, their table 1) for the precise location of the nine events used for FWI. We performed a multi-scale FWI (Bunks *et al.* 1995) by successive inversion of data sets of increasingly high-frequency content. We have generated three data sets by low-pass filtering the raw data with cut-off periods of 20, 10 and 5 s, respectively. For each frequency band, the optimization is stopped when: (1) the misfit reaches 1 per cent, (2) 47 iterations or the dedicated wall-time is reached and (3) when no suitable step-length is found according to Wolfe criteria (Nocedal & Wright 2006). We do not add noise to the data sets during the FWI experiments to minimise the role of the regularization and hence maximise the footprint of the experimental factors we want to assess (subsurface parametrization, initial model, acquisition geometry).

Unlike the real CIFALPS data application, the temporal source signature is assumed to be known and hence is not estimated during FWI. We do not apply any time-dependent preconditioning (namely, time windowing or amplitude balancing) on the data, but we normalize each observed and computed common-event gathers by the maximal amplitude of the vertical component of the observed data to equally balance their contribution in the gradient of the misfit function. We use the L-BFGS optimization algorithm (Nocedal 1980), a quasi-Newton approach that recursively estimates the action of the inverse Hessian on the gradient, without other additional preconditioning on the gradient. The optimization parameters are the logarithm of the physical parameters to further balance the parameters in depth. A mild regularization is applied by smoothing the gradient with a Gaussian function with a correlation length of 2.4 km in all three Cartesian directions.

In the following numerical examples, each experimental configuration deviates from the one we consider as the reference one in only one respect: subsurface parametrization, starting model or acquisition geometry. In the reference case, FWI is applied using:

- (i) a (ρ, V_p, V_s) parametrization,
- (ii) a starting model that consists in a smoothed version of the true model (this model will be referred to as the Tomographic model),
- (iii) a 2-D regular grid of stations with a receiver-sampling interval of 5 km in the two horizontal Cartesian directions,

as shown in the second line of Table 2.

For a quantitative insight on the relevance of the FWI models, we computed a correlation-based misfit between a vertical section of the FWI models and the Alps lithospheric model, this vertical section being aligned with the CIFALPS profile (Beller *et al.* 2017, their fig. 3). Even though such a norm may not be the most suitable one (in the sense that it cannot discriminate the footprint of resolution and amplitude errors), it complements in a more quantitative way the qualitative visual assessment of the subsurface models. The statistics of the results of the parametric analysis are outlined in Table 2.

4.2 Which earth parametrization for teleseismic FWI?

We first provide some insights on the most suitable subsurface parametrization for teleseismic FWI. The conclusions we draw in this section, do depend on the acquisition geometry considered for FWI applications. While they are arguably valuable for teleseismic application where few incident plane waves impinges the lithospheric target from beneath, these conclusions cannot be generalized for other kind of seismic acquisitions such as reflection ones for which the reader is referred to Tarantola (1986). We

Table 2. Statistics of the FWI applications. Setup: experimental setup in terms of initial model (m_0), optimization parameters ($param$) and acquisition geometry ($acqui$). Misfit function (5 s): initial/final data misfit function and data misfit function reduction in per cent for the last multi-scale step (cut-off period: 5 s). Misfit reduction: data misfit reduction in per cent for each frequency band (20 s, 10 s and 5 s). Model misfit: initial/final correlation-based model misfit and misfit reduction in per cent. Iterations: number of FWI iteration for each frequency band (20, 10, 5 s).

m_0	Setup	param	acqui	Misfit function (5 s)			Misfit reduction (per cent)			Model misfit			Iterations		
				f_0	f_{end}	per cent	20 s	10 s	5 s	$m_o - m_{true}$	m- m_{true}	per cent	20 s	10 s	5 s
Tom		(ρ, λ, μ)	N05	7.57×10^3	4.00×10^2	5.28	1.88	7.87	12.86	3.89×10^7	2.73×10^7	70.33	44	47	45
Tom		(ρ, V_p, V_s)	N05	7.57×10^3	8.95×10^2	11.82	1.51	29.38	23.31	3.89×10^7	2.51×10^7	64.76	42	14	36
Tom		(ρ, V_p, V_s)	N05	7.57×10^3	1.17×10^3	15.46	3.32	8.22	26.29	3.89×10^7	3.22×10^7	82.98	44	42	19
Smooth		(ρ, V_p, V_s)	N05	1.64×10^4	1.07×10^3	6.41	1.72	19.4	18.84	4.59×10^7	2.82×10^7	61.37	34	41	31
PREM		(ρ, V_p, V_s)	N05	1.01×10^4	1.01×10^3	6.27	1.56	19.04	21.58	4.66×10^7	3.17×10^7	68	35	11	40
Gradient		(ρ, V_p, V_s)	N05	8.45×10^4	7.73×10^3	2.96	1.58	15.08	18.44	1.33×10^8	5.76×10^7	43.15	35	39	37
Tom		(ρ, V_p, V_s)	N10	2.04×10^3	2.39×10^2	11.72	1.7	18.12	16.04	3.89×10^7	2.90×10^7	74.7	36	44	29
Tom		(ρ, V_p, V_s)	N25	4.00×10^2	1.68×10^2	42	1.16	14.97	75.34	3.89×10^7	2.92×10^7	75.24	38	14	2
Tom		(ρ, V_p, V_s)	N50	9.13×10^1	3.79×10^1	41.51	2.16	12.14	48.97	3.89×10^7	4.33×10^7	111.49	35	13	4
Tom		(ρ, V_p, V_s)	L05	3.21×10^2	2.58×10^2	67.07	2.36	24.16	100	3.89×10^7	4.87×10^7	125.19	44	11	0
Tom		(ρ, V_p, V_s)	L10	1.64×10^2	1.10×10^2	80.37	2.35	7.8	93.22	3.89×10^7	4.80×10^7	123.39	44	42	2
Tom		(ρ, V_p, V_s)	CIF	6.43×10^2	1.37×10^2	21.31	7.78	11.34	20.79	3.89×10^7	5.11×10^7	131.36	30	35	44

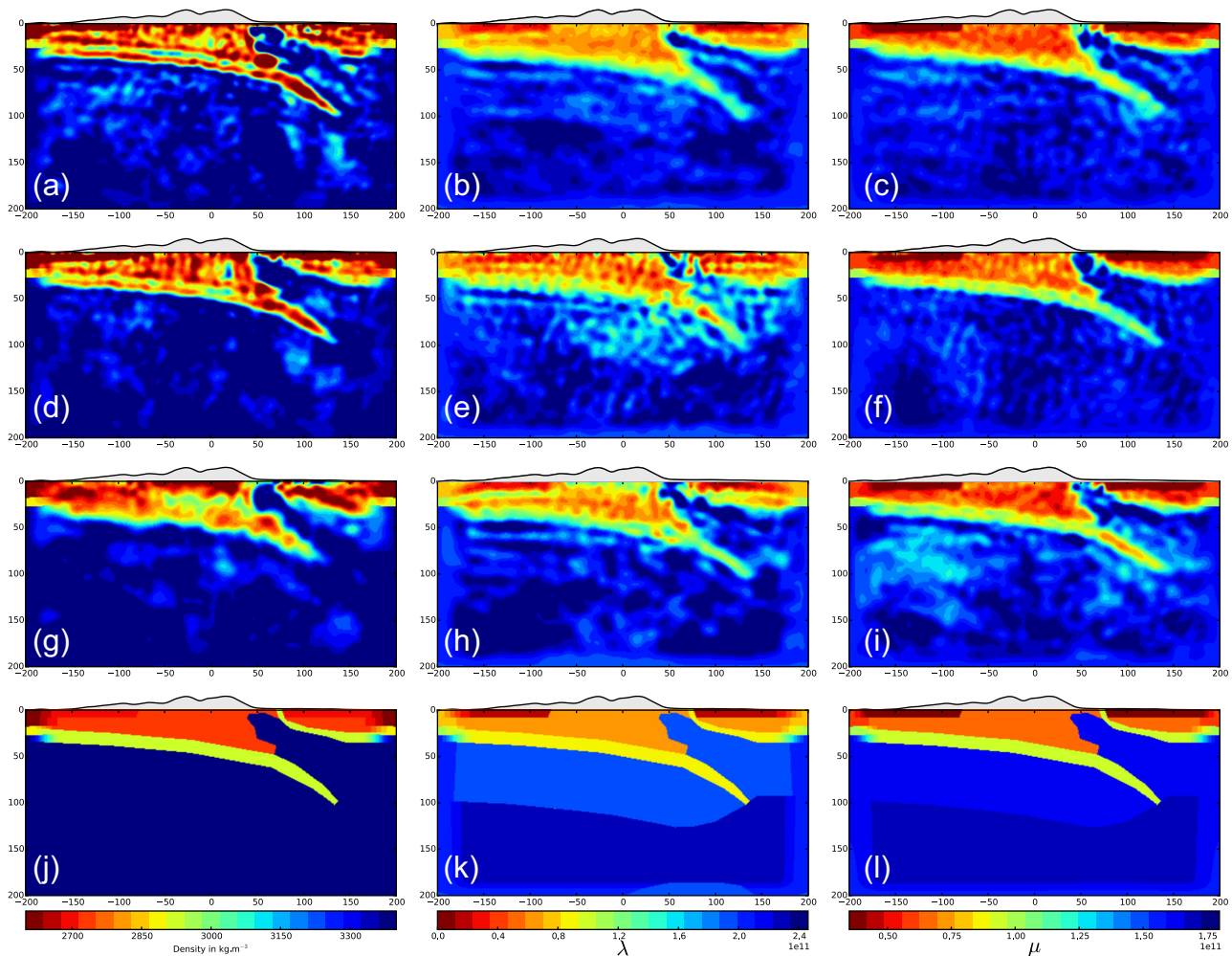


Figure 6. Subsurface parametrization assessment. From left to right are shown the ρ , λ and μ models that are built *a posteriori* by the nonlinear recombination of the optimization parameters. These optimization parameters are (a–c) (ρ , λ , μ), (d–f) (ρ , V_p , V_s) and (g–i) (ρ , I_p , I_s). The first row directly shows the optimization parameters (ρ , λ , μ) accordingly. The last row (j–l) shows the true (ρ , λ , μ) models for comparison.

perform FWI for three different elastic parametrizations, (ρ , λ , μ), (ρ , V_p , V_s) and (ρ , I_p , I_s), and we jointly update the three parameter classes involved in the chosen parametrization. For each of the three inversions, we show the parameters that have been updated by the inversion, referred to as optimization parameters, as well as the parameters of the two other parametrizations, these parameters being built by the *a posteriori* nonlinear recombination of the optimization parameters. These parameters will be referred to as visualization parameters.

Three (ρ , λ , μ), (ρ , V_p , V_s) and (ρ , I_p , I_s) models are shown in Figs 6(a)–(i), 7(a)–(i) and 8(a)–(i), respectively. For each of these figures, two multi-parameter models (visualization parameters) have been built *a posteriori* by the nonlinear recombination of the optimization parameters, while the third one directly shows the optimization parameters. On each figure (Figs 6a–i, 7a–i and 8a–i), the optimization parameters are (ρ , λ , μ), (ρ , V_p , V_s), (ρ , I_p , I_s), from top to bottom. The presentation of the subsurface models are complemented by Fig. 9, whose histograms show the level of correlation between the reconstructed models generated by each FWI and the true subsurface models.

As a guideline for the visual assessment of the subsurface models, we focus on the reconstruction of the sedimentary basins, the Ivrea body and the subducting slab. According to these 3 criteria, we con-

clude that the best subsurface models for the kinematic parameters, namely, V_p , λ and I_p and V_s , μ and I_s , are the I_p and I_s (visualization) models inferred from the (ρ , λ , μ) optimization models (Fig. 8b and c). This qualitative assessment is supported by the metric outlined in Fig. 9, that confirms that these I_p and I_s models show the best correlation with the true ones.

The ρ model inferred from the (ρ , λ , μ) parametrization seems slightly better resolved than the one inferred from the (ρ , V_p , V_s) parametrization, at the expense of the noise level. In contrast, the ρ model inferred from the (ρ , I_p , I_s) parametrization is much smoother, that is consistent with the former analysis of the radiation pattern. This visual assessment is supported by Fig. 9 in the sense that the ρ model inferred from the (ρ , V_p , V_s) parametrization shows the best correlation with the true model, while the one inferred from the (ρ , λ , μ) parametrization shows the poorest one, which might result from the noise impacting this model. However, this noise did not prevent a reliable recombination of the I_p and I_s models from the (ρ , λ , μ) optimization parameters.

One may wonder why the (ρ , λ , μ) parametrization provided the best subsurface models after recombination of I_p and I_s . If we limit our comparative analysis to the optimization parameters, μ is the one which shows the best correlation with the true model among the kinematic parameter more closely related to shear waves

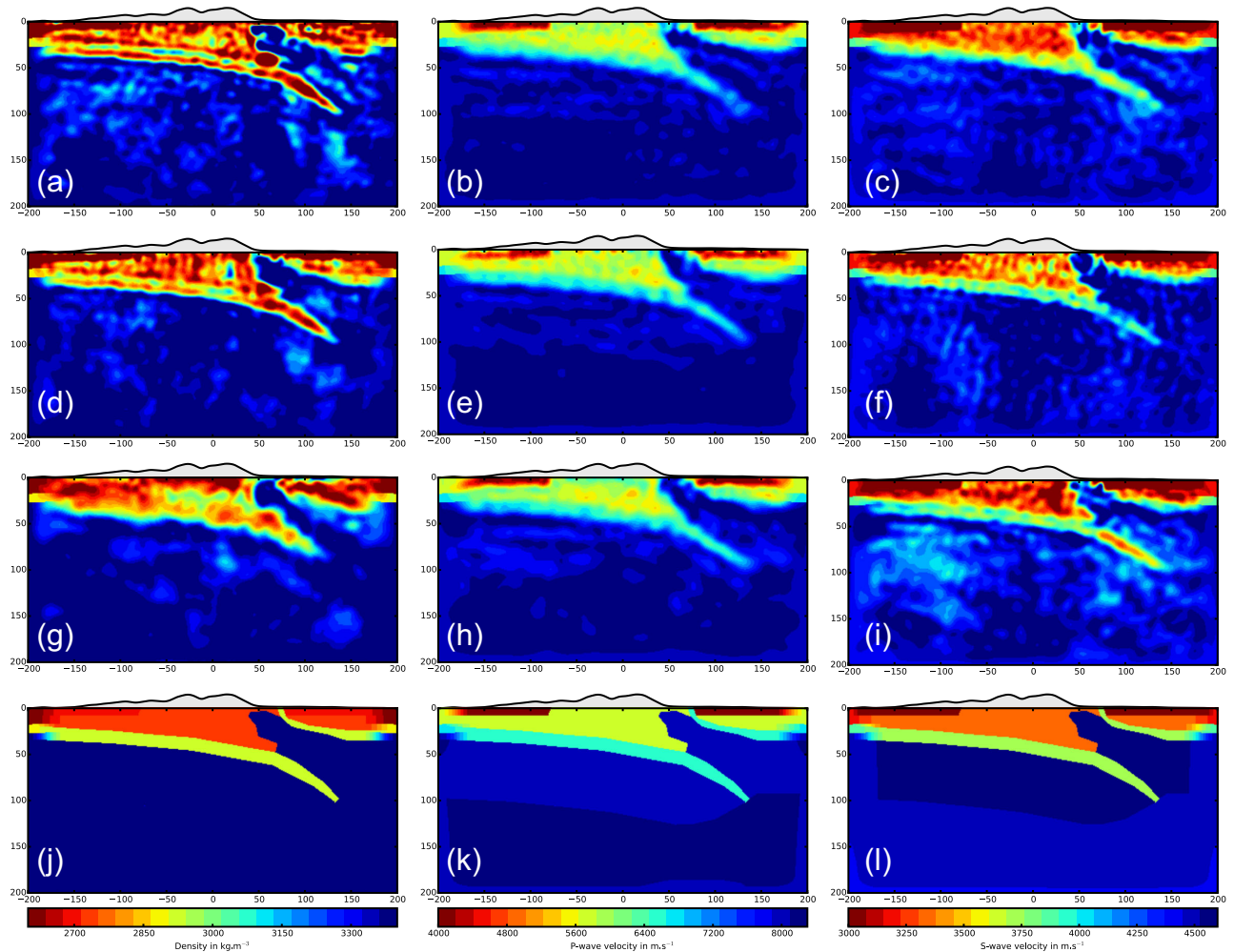


Figure 7. Subsurface parametrization assessment. From left to right are shown the ρ , V_p and V_s models that are built *a posteriori* by nonlinear recombination of the optimization parameters. These optimization parameters are (a–c) (ρ , λ , μ), (d–f) (ρ , V_p , V_s) and (g–i) (ρ , I_p , I_s). The second row directly shows the optimization parameters (ρ , V_p , V_s) accordingly. The last row (j–l) shows the true (ρ , V_p , V_s) models for comparison.

(V_s , I_s , μ), (Fig. 9, arrows). This might result from the key contribution of the forward-scattered P – P mode during the long-wavelength reconstruction of μ (Fig. 3c). A second reason may be related to the high-wavenumber content of the ρ model obtained with the (ρ , λ , μ) parametrization (Fig. 3a). When I_s is inferred from ρ and μ by multiplication ($I_s = \sqrt{\rho\mu}$), it is likely that the high wavenumber content of ρ has complemented the small-to-intermediate wavenumber content of μ to focus a broadband I_s model. This improved resolution of the visualization I_p and I_s models relative to the λ and μ optimization models is quite visible in the reconstruction of the Ivrea body and subducting slab (compare Figs 8b and c with 6b and c). A similar line of thought would explain why the sedimentary basins in the visualization I_p and I_s models inferred from the (ρ , λ , μ) optimization parameters (Figs 8b and c) are better focused than those of the I_p and I_s optimization models (Figs 8h and i). The lower high-wavenumber content of the optimization μ model, in particular at great depths (Fig. 6c, subducting slab), relative to the optimization I_s and V_s models might result from the dominant imprint of the P – P forward scattering in the μ gradient relative to that of the back-scattering components. This might have dominantly driven the deep update of μ towards a low-to-intermediate wavenumber reconstruction at the expense of the high wavenumber components. Again, multiplication of μ with ρ might have contributed to build

a posteriori an I_s model, whose quality is higher than that of the optimization I_s and V_s model. The same comment could apply to the I_p reconstruction from λ and μ . Another issue concerns potential cross-talk between ρ and the two kinematic parameters. The multiplication of ρ with a kinematic parameter to build the impedance might help to remove cross-talk as suggested by Prioux *et al.* (2013).

A common practice in FWI studies to appraise the quality of output models consists in comparing FWI models vertical profiles to their respecting counterparts in the initial and targeted models. Such a quality assessment is presented in Fig. 10 for impedance models of the three above-mentioned parametrization. The localization of interest is $x = 100$ km. FWI clearly succeed in recovering the absolute values of I_p and I_s models as well as those of the ρ model. Biases, appearing as vertical undulations, are attributed to limited-bandwidth effects resulting from the limited bandwidth of signals and the limited angular illumination or by trade-off effects with the wave speed. This is particularly true for the density that is characterized by a more selective radiation pattern than the S wave speed). Note that, whatever the choice of parametrization used during optimization, I_s and I_p models (obtained directly by inversion or after recombination of the other optimization parameters) fairly well reproduce the amplitudes of the targeted I_p and I_s structures

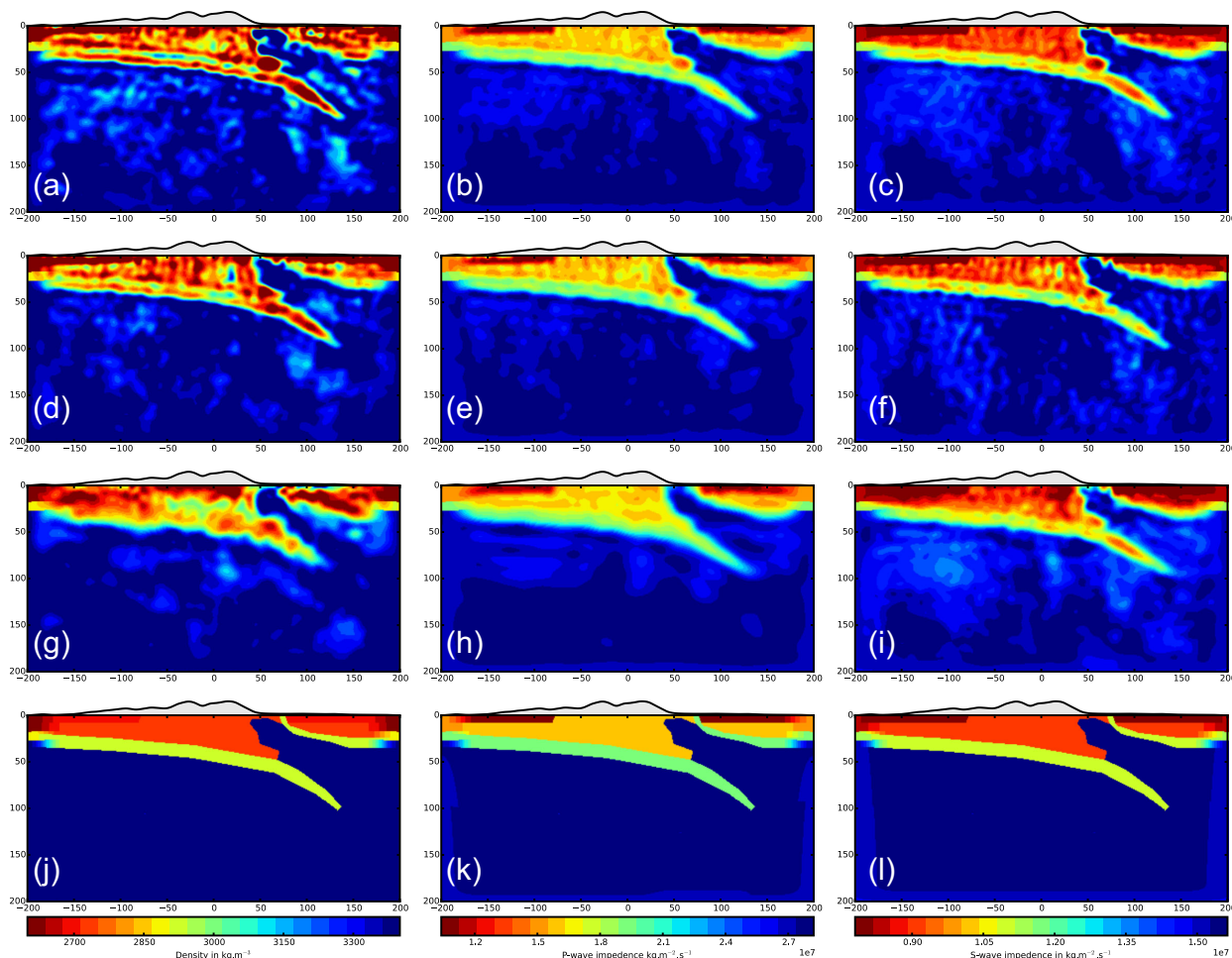


Figure 8. Subsurface parametrization assessment. From left to right are shown the ρ , I_p and I_s models that are built *a posteriori* by nonlinear recombination of the optimization parameters. These optimization parameters are (a–c) (ρ , λ , μ), (d–f) (ρ , V_p , V_s) and (g–i) ρ , I_p , I_s). The third row directly shows the optimization parameters (ρ , I_p , I_s) accordingly. The last row (j–l) shows the true (ρ , I_p , I_s) models for comparison.

aside from the previously mentioned filtering effects. Therefore, the P and S impedances might be robust parameters for structural interpretations.

4.3 Which initial model for teleseismic FWI ?

FWI is a nonlinear and ill-posed inverse problem that is generally tackled with local optimization techniques. In this framework, a requirement for the optimization algorithm to converge towards a reliable subsurface model is that the initial model should be accurate enough to position the starting point within the attraction basin of the global minimum. Forwardly, this issue is closely related to the cycle skipping phenomenon that occurs in FWI whenever a seismic wiggle obtained from an inaccurate predictive model is shifted by more than half the period with respect to the corresponding observed wiggle (Virieux & Operto 2009). If cycle-skipping occurs, data residuals will be back-projected onto a wrong isochrone surface hence conducting FWI towards a local minimum.

A conventional way of mitigating cycle skipping effects is to progressively increase the number of propagated wavelengths during seismic modelling, since the relative time error is inversely proportional to this number (Pratt 2008). A first data-driven strategy to achieve this goal relies on a frequency continuation approach where data sets of increasing high-frequency content are hierarchi-

cally processed (Bunks *et al.* 1995). This frequency continuation can be complemented by a time continuation that can be designed by gently increasing the selected time window over iterations. Both data continuation strategies contribute to implement a multi-scale imaging proceeding from the low wavenumber updates to the higher ones.

The frequency continuation approach requires that the recorded data have a sufficiently low frequency content, which is not always the case in exploration seismology although the design of broadband technology has become an active field of research (e.g. Baeten *et al.* 2013). To overcome this issue, this community has focused its efforts on the upstream building of reliable initial model by traveltime and slope tomography and migration-based velocity analysis, the design of elaborated misfit functions with better convexity properties or by expanding the search space (e.g. Symes 2008; Lambaré 2008; van Leeuwen & Herrmann 2013; Métivier *et al.* 2016). In the case of teleseismic data, the cycle skipping issue should be prevented by the availability of long period (low frequency) data. Furthermore, reference 1-D global earth model such as PREM, IASP91 or AK135 are generally accurate enough to predict the arrivals of global seismic phases with a confidence interval in between one and two seconds therefore preventing the cycle skipping issue for representative teleseismic periods (20 s to 1 s).

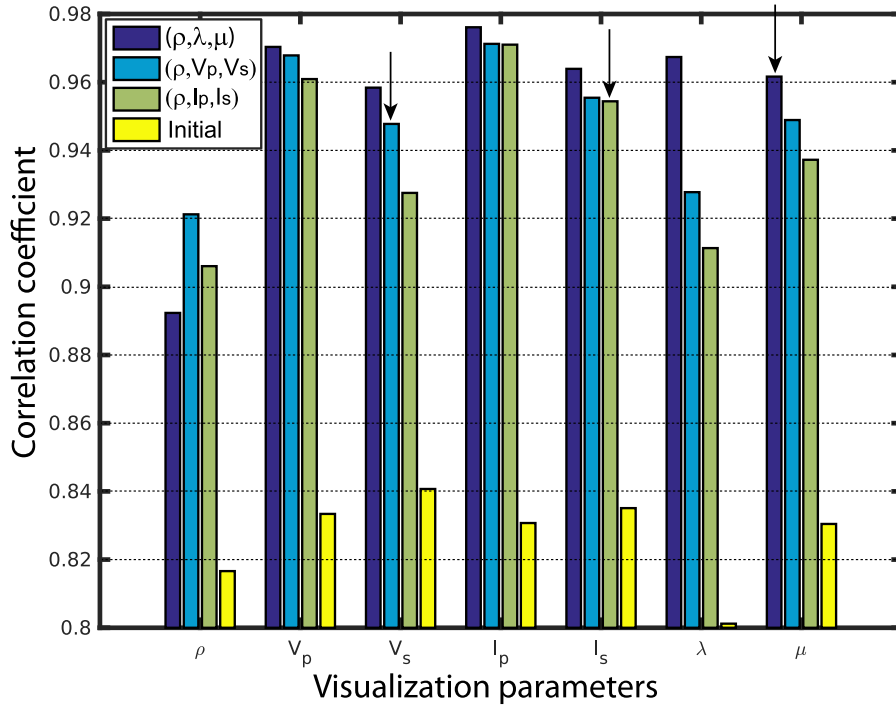


Figure 9. FWI model quality. The colours label the subsurface parametrization used for FWI or in other words the FWI optimization parameters. *Initial* refers to the initial models for each parameter class. The horizontal axis labels the visualization parameter that has been built *a posteriori* by nonlinear recombination of the optimization parameters. The vertical axis labels the correlation-based misfit function between the visualization and the true models. The three arrows point the quality of the optimization parameters μ , V_s and I_s (see the text for explanations).

Vertical logs at position x=100km

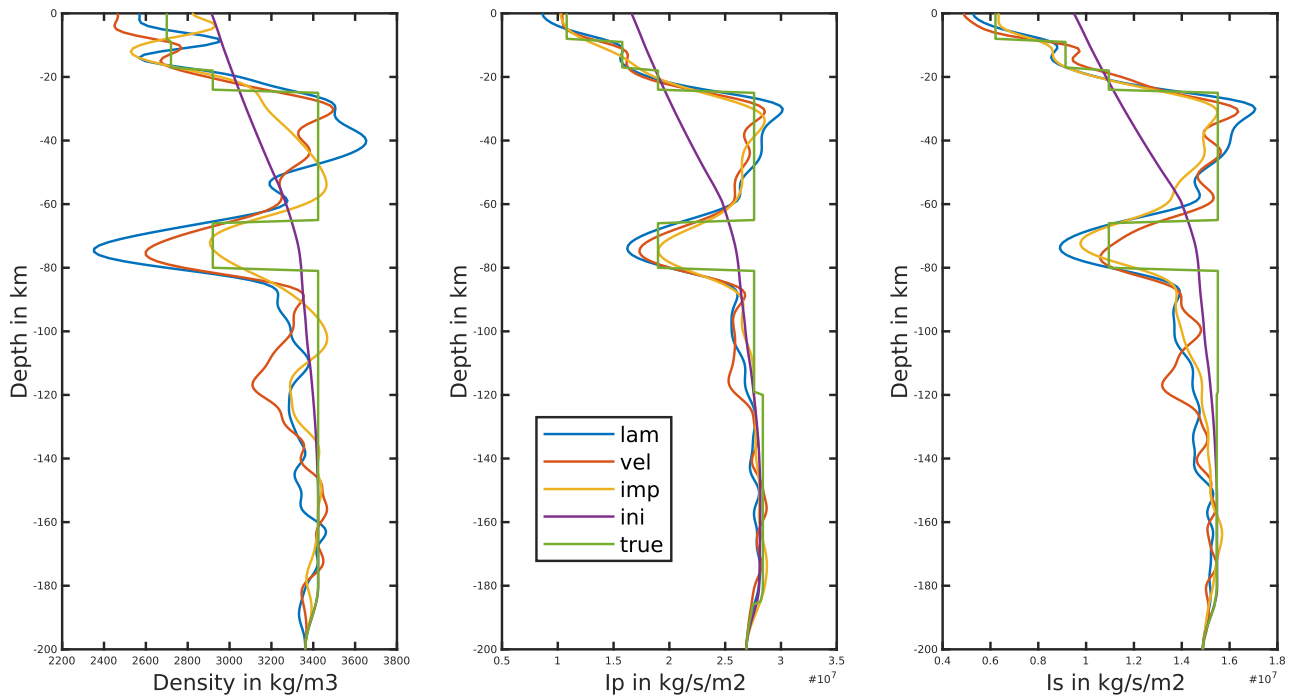


Figure 10. FWI vertical logs. From left to right are presented vertical profiles (at $x = 100$ km) of ρ , I_p and I_s visualization parameters generated from FWI models obtained from the optimization of Lamé parameters (*lam*), seismic velocities (*vel*) and seismic impedances (*imp*). Their counterparts in the initial (*ini*) and targeted models (*true*) are also given for comparison.

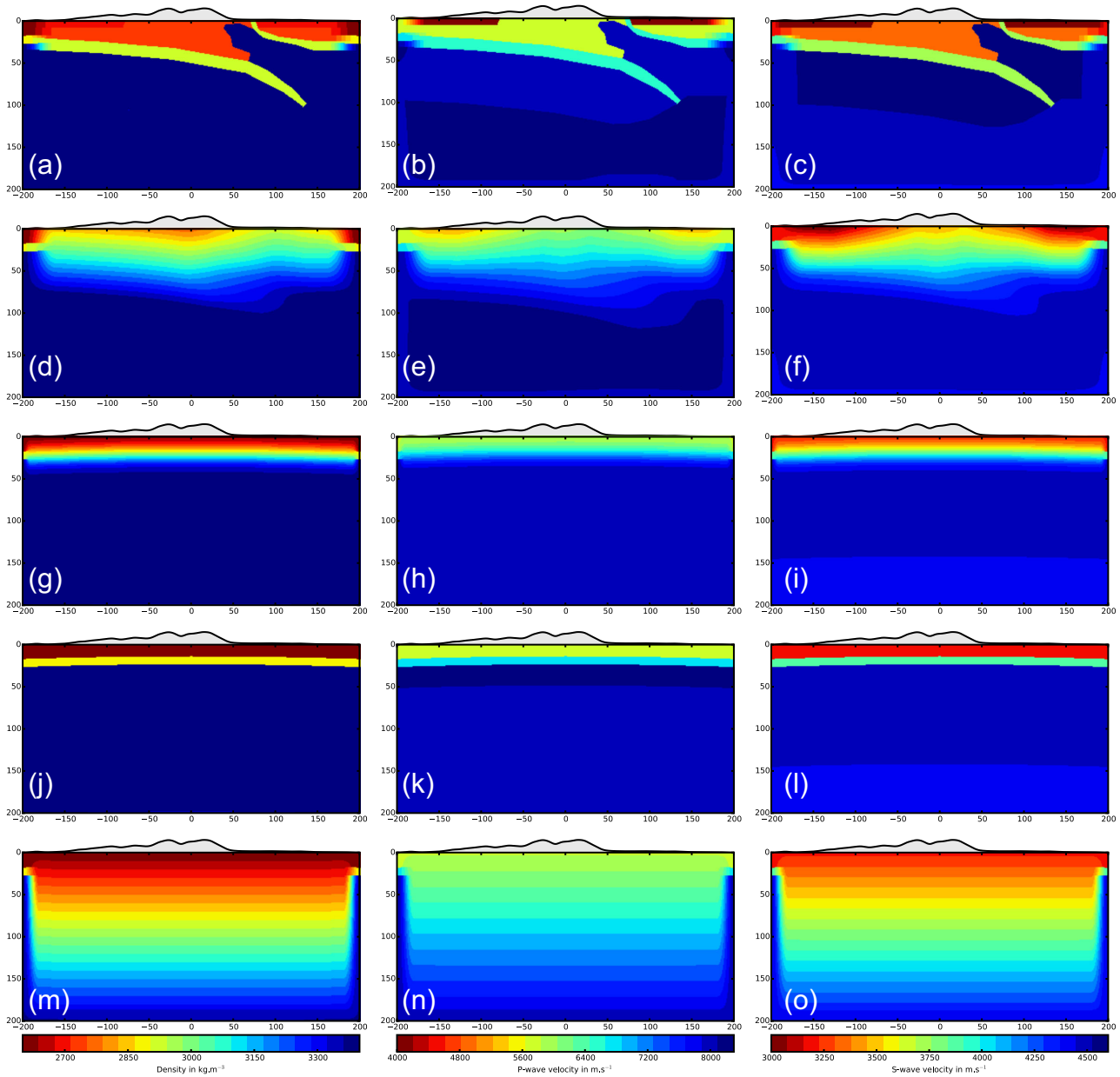


Figure 11. Starting models for FWI. From left to right ρ , v_p and v_s models for: (a–c) the input model used to generate synthetic data, (d–f) tomographic model computed by smoothing of the input model, (g–i) smoothed version of PREM model, (j–l) PREM model and (m–o) vertical gradient.

To verify the previous statement, we analyse the sensitivity of the FWI to four different starting models. A first initial model (Figs 11d–f) is built by Gaussian smoothing of the true Alps lithospheric model with a correlation length of 50 km in the three spatial directions. Despite being smooth, this model still captures some of the 3-D complexity of the target and can be considered as analogous to an initial model built by first-arrival traveltimes tomography. In absence of such a knowledge about the lithospheric velocity and density structure, another option is to resort to available 1-D reference earth models. We investigate here two possible initial models based on the PREM reference model: the layered original one and a smoothed version obtained by Gaussian filtering with a 10 km correlation length (Figs 11g–i) and (Figs 11j–l). A last starting model is a simple vertical velocity gradient (Figs 11m–o).

The FWI results for the different initial models are shown in Fig. 12. As expected, the best FWI model has been inferred from

the initial tomography-like model (Figs 12d–f and Table 2). Using the smooth PREM model as an initial model leads to acceptable FWI results, that may be improved by a more aggressive regularization (Figs 12g–i and Table 2). The FWI v_p model is very similar to the one inferred from the tomography-like model, while the ρ and v_s models are slightly noisier. Overall, this result confirms that smoothed reference earth models are accurate enough for teleseismic FWI. When the original PREM model is used as initial model, the footprint of the sharp crustal layering of the initial model overlap the smoother structures updated by FWI (Figs 12j–l). Indeed, this high-wavenumber footprint cannot be removed or updated by FWI because its spectral content is beyond the resolution limits of FWI. Moreover, this prior high-wavenumber content in the initial model, which can be mispositioned in depth, can drive the inversion towards a local minimum by generating, during seismic modelling, transmitted and reflected waves that are badly positioned in time.

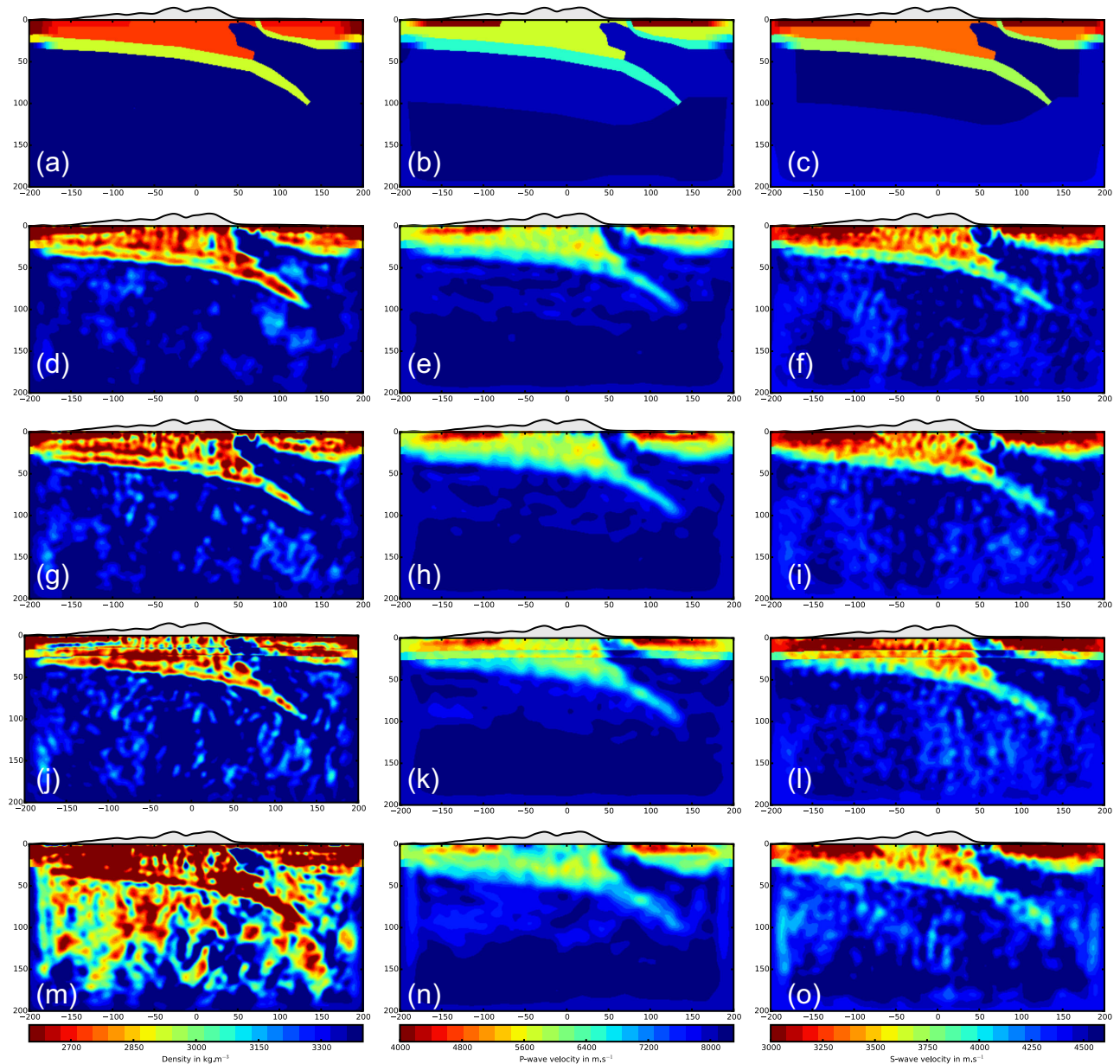


Figure 12. Assessment of initial models. From left to right, ρ , V_p and V_s models. (a–c) True models. (d–o) FWI models inferred from (d–f) the smooth tomography-like model, (g–i) the smoothed version of the PREM model, (j–l) the original PREM model and (m–o) the vertical gradient model.

Interestingly the vertical gradient model is accurate enough to produce acceptable velocity models (Figs 12m–o and Table 2). However, the ρ model shows saturated amplitudes suggesting a higher sensitivity of the inversion to this parameter. A possible reason is that, among all of the tested initial models, the vertical gradient model is the one which does not generate any wave partitioning except at the free surface, hence leading to significant amplitude residuals. In this case, ρ has been over-saturated, that is to say more important than for the other density models, to remove these amplitude residuals during the early iterations. This over-saturation of ρ might also have favoured some leakage with the velocity updates. This leakage from the velocities to the ρ updates is suggested by the less contrasted reconstruction of the V_p parameter in Fig. 12(n).

In summary, a good initial model for teleseismic FWI should be smooth enough to avoid involving wavenumber components in the initial models that cannot be easily updated by FWI or which

are beyond its resolution limit. On the other hand, our numerical experiments suggest that the initial model should be accurate enough to predict well enough amplitude phenomena in particular if second-order parameter such as density is jointly updated with the wave speeds. If this condition is not satisfied, the update of the density might become exaggeratedly sensitive to these amplitude residuals and might prevent the update of the high wavenumber components of the wave speeds in the upper part of the lithospheric model.

4.4 Sensitivity to acquisition design

4.4.1 Linear versus areal acquisitions

High-resolution seismic imaging techniques such as receiver function migration or waveform tomography require dense array of seismic sensors to adequately sample the teleseismic wavefield as

well as the imaging kernel. On the one hand, seismic experiments have been designed for high-resolution receiver function analysis or scattering migration with dense profiles of stations spaced 5 to 10 km apart. Cascadia, Hi-Climb, MASE, CIFALPS or Pyrope experiments are examples of such dense linear acquisitions. On the other hand, areal geometries designed with a coarse grid of stations (30 to 70 km interval) have been carried out to perform tomography at a larger scale. USArray, Iberarray or the incoming AlpArray are examples of such experiments, where the receiver grid is either fixed or moving.

To the best of our knowledge, no synthetic experiment has been performed to assess which acquisition geometry is the most suitable one for teleseismic FWI at the lithospheric scale. To fill this gap, we performed FWI for seven station layouts, that have been designed based upon the two above-mentioned acquisition trends. Four areal acquisitions consist of 2-D receiver grids with a station-sampling interval of 5, 10, 25 and 50 km, leading to pools of 2001, 525, 105 and 21 stations, respectively (Figs 13d–o). Two acquisitions consist of one line of stations spaced 5 and 10 km apart (Figs 13p–u), leading to pools of 69 and 35 stations, respectively (Figs 13p–u). Finally, for sake of comparison with the real data case study presented in Beller *et al.* (2017), we also assess the CIFALPS acquisition, that combines a dense profile of stations spaced 5 km apart with a sparse irregular spread of permanent stations (Figs 13v–x).

The FWI models for each experiment are shown in Fig. 13. As expected, the denser areal geometry gives the best resolved models for each of the three parameter classes. The coarsening from 5 to 10 km of the receiver grid does not significantly degrade the reconstruction of the synthetic model neither in terms of spatial resolution nor in terms of noise (Figs 13d–i). Furthermore, the 25 km acquisition still provides reliable reconstruction of the synthetic model even though its resolution is clearly more limited. Here, this poorer resolution results because the FWI stopped after two iterations during the processing of the last frequency band (see Table 2). For a 50 km acquisition, despite a quite low resolution, the main structural heterogeneities of the synthetic model are still recognizable. However, a strong acquisition footprint in the form of energetic velocity anomalies at the station positions every 50 km is quite visible in the *P*-wave velocity model (Fig. 13n). Overall, we show a progressive degradation of the spatial resolution as the receiver sampling becomes coarser. This results because the range of scattering and dip angles sampled by the source-receiver layout becomes narrower and sparser as the station interval increases. The narrowing of the scattering angle illumination will limit the resolution of the reconstruction perpendicularly to the dip, while the coarsening of the scattering-angle illumination can generate wraparound of the structures if narrow-frequency bands are processed. The scattering-angle illumination as a function of the dip angle (noted ϕ in the current study, eq. 9), is illustrated in Rondenay *et al.* (2005, their fig. 3) for a linear teleseismic acquisition and two different station samplings.

Compared to areal geometries, linear acquisitions drastically affect the ability of FWI in recovering well-resolved lithospheric models. Although six of the nine earthquakes have backazimuths approximately aligned with the receiver line (Fig. 4a), the FWI models show a degraded resolution of the V_p and V_s models as well as incomplete reconstruction of density at great depths. This results because the linear deployment illuminates more incompletely (in terms of bandwidth and sampling) the scattering and the two dip angles than an areal geometry, even for the image points located in the vertical plane of the receiver line. As an illustrative example, any non-cylindrical structure cross-cutting the vertical plane of the

profile such as the Ivrea body will generate out-of-plane scattering which cannot be recorded without an areal geometry. The limited number of off-line stations available in the real CIFALPS acquisition helps increasing the spatial resolution at the expense of the signal-to-noise ratio, which might have been degraded due to spatial aliasing (Figs 13v–x). It is worth remembering that, for sake of consistency between the different experiments, we did not apply significant regularization during FWI. When dealing with such sparse acquisition, more aggressive smoothing constraints along the spatial directions that are not sampled well by the acquisition geometry would improve the results as suggested by the results of Beller *et al.* (2017) on the CIFALPS real data.

4.4.2 Acquisition footprint and aliasing

The noise in the recovered FWI model might be due to the conjunction of two factors. First, FWI models are contaminated by acquisition footprint in the form of high-amplitude perturbations at the station positions, this acquisition footprint being particularly obvious for the 50 km station grid. These localized perturbations result from the high amplitudes of the adjoint wavefields at the station positions and the arrival time of the incident planar wavefields at the station. Therefore, the product of these two wavefields generate some bright spots at the station positions if no efficient correction for these amplitude effects are applied. The Hessian, in particular its diagonal coefficients, is expected to correct for them. However, the poor approximation of the Hessian provided by l-BFGS during the first FWI iterations might have contributed to leave a significant imprint of this amplitude singularities in the FWI models. These high-amplitude perturbations can behave as strong diffractors that propagate noise in the FWI models over iterations.

Second, aliasing effects may contaminate the recovered image at high-frequencies therefore adding noise in the reconstructed image. In the framework of Kirchhoff migration, Lumley *et al.* (1994) defined three kinds of aliasing: data, image and operator aliasing.

Data aliasing occurs when the recorded wavefield is not properly sampled in one or more dimension according to the Nyquist theorem. Spatial aliasing can occur when the apparent wavelength of the incoming wavefield at the surface is not sampled adequately by the receiver grid. Even though the 5 s-period incident *P* wavefield satisfies the sampling criterion, sharp small-scale structures near the surface can behave as energetic diffractors that scatter energy along steep diffraction hyperbola in the time–distance domain. In this study, the edges of the Ivrea body and the pinch-out of the sedimentary basins generate such energetic diffractions that undergo aliasing even for a 5 km station-sampling interval (Fig. 5, green ellipse). An obvious means to mitigate data aliasing is to limit the inversion to low-frequency data at the expense of the spatial resolution of the imaging. Another strategy would consist in re-sampling the acquisition by interpolation of seismic traces (Wilson & Guitton 2007).

Image aliasing occurs when the mesh, that parametrizes the subsurface target, is too coarse to properly sample the output of the imaging. In FWI, the maximum resolution is half a wavelength and the subsurface should be meshed accordingly, that is with at least four grid points per wavelengths in a finite-difference framework. In our case, we design the hexahedral mesh with at least five degrees of freedom per wavelength according to the above-mentioned requirement.

Operator aliasing occurs when the integral operator that transforms the data in model updates is not properly sampled or, in other

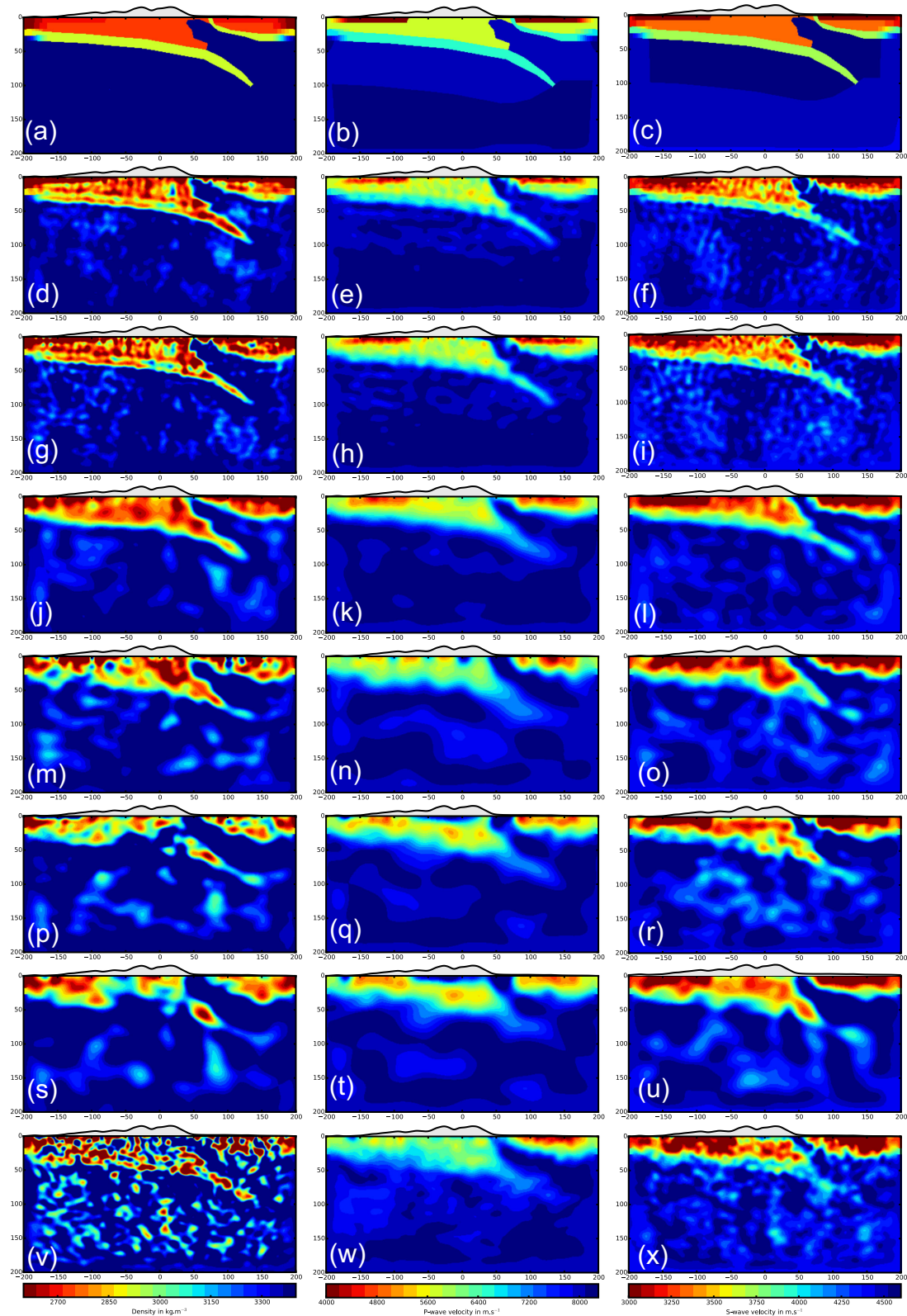


Figure 13. Sensitivity of FWI to the station layout. From left to right, ρ , V_p and V_s models. (a–c) True (ρ , V_p , V_s) models used to generate the recorded data; FWI models inferred from 2-D grids of station with inter-station spacings of 5 km (d–f), 10 km (g–i), 25 km (j–l) and 50 km (m–o). FWI models inferred from linear acquisition with inter-station spacings of 10 km (p–r) and 5 km (s–u). (v–x) FWI models inferred from the CIFALPS acquisition as used in Beller *et al.* (2017).

words, when the dip angle of the wavenumber vectors, eq. (9), are undersampled, leading to wraparound in space. This form of aliasing in the model space can occur even if the data are not aliased. In the framework of inverse scattering techniques, Rondenay *et al.* (2005, their Fig. 4) illustrated how the sampling of the stations impacts upon the sampling of the weighting function applied to the scattered data during their weighted stack along diffraction hyperbolas. This migration operator becomes increasingly smooth with depth, which means that operator aliasing is mostly generated at shallow depths. Undersampling this weighting function leads to wraparound artefacts in the form of high wavenumber ‘speckle’ in the image. Considering that the FWI gradient and the migration kernel obey the same imaging principle, it is likely that this kinds of aliasing has affected our FWI results.

Other sources of noise can result from the poor fold of teleseismic acquisitions. The subsurface imaging in FWI relies on the Huygen’s principle: any continuous structure in the subsurface is imaged by the interference of the isochrone surfaces generated by back-projection of data residual samples. These interferences are constructive at the locus of the missing subsurface structures and destructive off this locus. If the stations sampling is too coarse, the destructive interferences off the locus of the missing heterogeneities is not efficient enough to cancel out the undesired contributions of the isochrones.

Beyond the receiver sampling, the source distribution in teleseismic imaging can potentially generate aliasing noise and degrade spatial resolution. In a parametric study of 2.5-D frequency domain teleseismic FWI, Pageot *et al.* (2013) have shown the footprint of the source distribution and sampling on the quality of the imaging. Narrowing the range of incidence angles sampled by the events contribute to narrow the wavenumber coverage accordingly and hence, degrade the spatial resolution of the imaging. However, the source sampling for a fixed range of incidence angles has little effects on the imaging quality provided that the temporal frequency dimension is finely sampled. This results because a planar wavefield leads to a uniform spatial sampling of the subsurface at the expense of the angular illumination. More serious aliasing artefacts occurred when a 2-D lithospheric target is illuminated from only one side, leading to significant gaps in the dip angle coverage. In 3-D applications, this aliasing may be more prominent since two dip angles need to be sampled.

4.4.3 Resolution analysis

We assess now more precisely the footprint of the receiver sampling on the spatial resolution of the FWI. We performed two spike tests, the aim of which is to image a grid of small spherical inclusions spaced 20 km apart. These inclusions are superimposed on the smooth tomographic lithospheric model, which is used as a starting model for FWI. The diameter of the anomalies is 20 km and the amplitude of the perturbations are $\pm 100 \text{ m s}^{-1}$ and $\pm 100 \text{ kg m}^{-3}$ for wave speeds and ρ , respectively, with an alternating sign from one inclusion to the next. We perform this resolution analysis for two station deployments: the first consists of a grid of stations sampled 25 km apart (Fig. 15), while the second one involves a line of stations spaced 5 km apart (Fig. 16). The number of stations involved in the two acquisitions is roughly equivalent. In Figs 15 and 16, we extract a vertical section from the FWI models along the receiver line of the linear acquisition and several horizontal slices at increasing depths.

To assist the interpretation of these FWI results, we show the local wavenumber coverage, eqs (9) and (10), that is provided by the nine

events and the two station layouts for a homogeneous background model, at four diffractor points located at $(x[\text{km}], y[\text{km}], z[\text{km}]) = (0,0,10)$ (Fig. 14a), $(0,0,50)$ (Fig. 14b), $(0,50,10)$ (Fig. 14c) and $(0,50,50)$ (Fig. 14d). Note that the (2π) factor in the expression of the wavenumber modulus, eq. (10), has not been considered to generate Fig. 14 for a more direct access to the sampled wavenumber. The wavenumber coverage was computed considering the forward-scattered P - P and P - S modes and all the mode conversions of the doubly back-scattered waves from the free surface and the scatterers discretizing the homogeneous model. The P and S wave speeds in the homogeneous medium are 7 and 4 km s^{-1} , respectively. These figures give some idea of the bandwidth of the wavenumber spectrum illuminated by the acquisitions as well as the sampling of this bandwidth, keeping in mind that the weighting performed by the radiation patterns and the reflection coefficients at the free-surface are not taken into account. As a guideline, a circle of radius 0.04 km^{-1} is superimposed on each wavenumber spectrum to assess the sampling of the dip angles for a fixed wavenumber modulus. This angular sampling is directly controlled by the station spacing. Any undersampling of these dip angles will manifest in the space domain by spatial wraparound. A wavenumber of 0.04 km^{-1} roughly corresponds to the upper bound of the spectrum of a Gaussian inclusion of correlation length 10 km. Therefore, a full reconstruction of the inclusions shown in Figs 15 and 16 would be achieved if the area covered by the red circle was filled up in Fig. 14.

For the areal geometry, FWI succeeds in reconstructing the three parameters at shallow depths (Fig. 15). The resolution and the amplitudes of the V_p updates degrade with depth faster than those of V_s and ρ . This might result because all of the scattering modes contribute to the update of V_s and ρ unlike V_p hence leading to a higher fold during the V_s and ρ updates. Moreover, the P - S reflection coefficient at the free surface is higher than the P - P one (Pageot *et al.* 2013, their fig. 3), making the doubly P - S - S reflections from the lithospheric reflectors more energetic than the P - P - P counterparts. Increasing the fold is quite important when the imaging mostly rely on the contribution of second-order back-scattered waves which is likely the case at these depths. Moreover, the signature of the forward scattered wavefield diffracted from the deep inclusions might have been healed when arriving at the surface, hence contributing to degrade the V_p resolution at depth. We also show alternating positive and negative perturbations in the ρ vertical section at large depths (Fig. 15a, dash ellipse). This distortion of the inclusion shape highlights limited bandwidth effects, which result because ρ is mainly updated from back-scattered waves according to its radiation pattern. The range of scattering angles sampled by the acquisition decreases with depth hence narrowing the vertical wavenumber bandwidth accordingly. More surprisingly, the horizontal resolution is not impacted significantly by the limited backazimuthal coverage provided by the nine events (Figs 15b-f). In other words, the reconstruction of the inclusions in the horizontal planes shows only a moderate smearing along the y direction, which becomes more obvious below 130 km depth. This moderate smearing in the y direction is consistent with the elliptic shape of the (k_x, k_y) coverage provided by the areal acquisition with a major axis along the k_x axis (Figs 14a-d-4). This good horizontal wavenumber coverage can be explained by the fact that spherical inclusions behave as point diffractors that scatter waves in all the directions. These scattered waves are recorded by the areal geometry along all of the horizontal directions hence providing a significant coverage of the horizontal components of the wavenumber vectors.

Noise is also shown at shallow depths (Figs 15a and b) and decrease with depth. This noise has a higher frequency content on

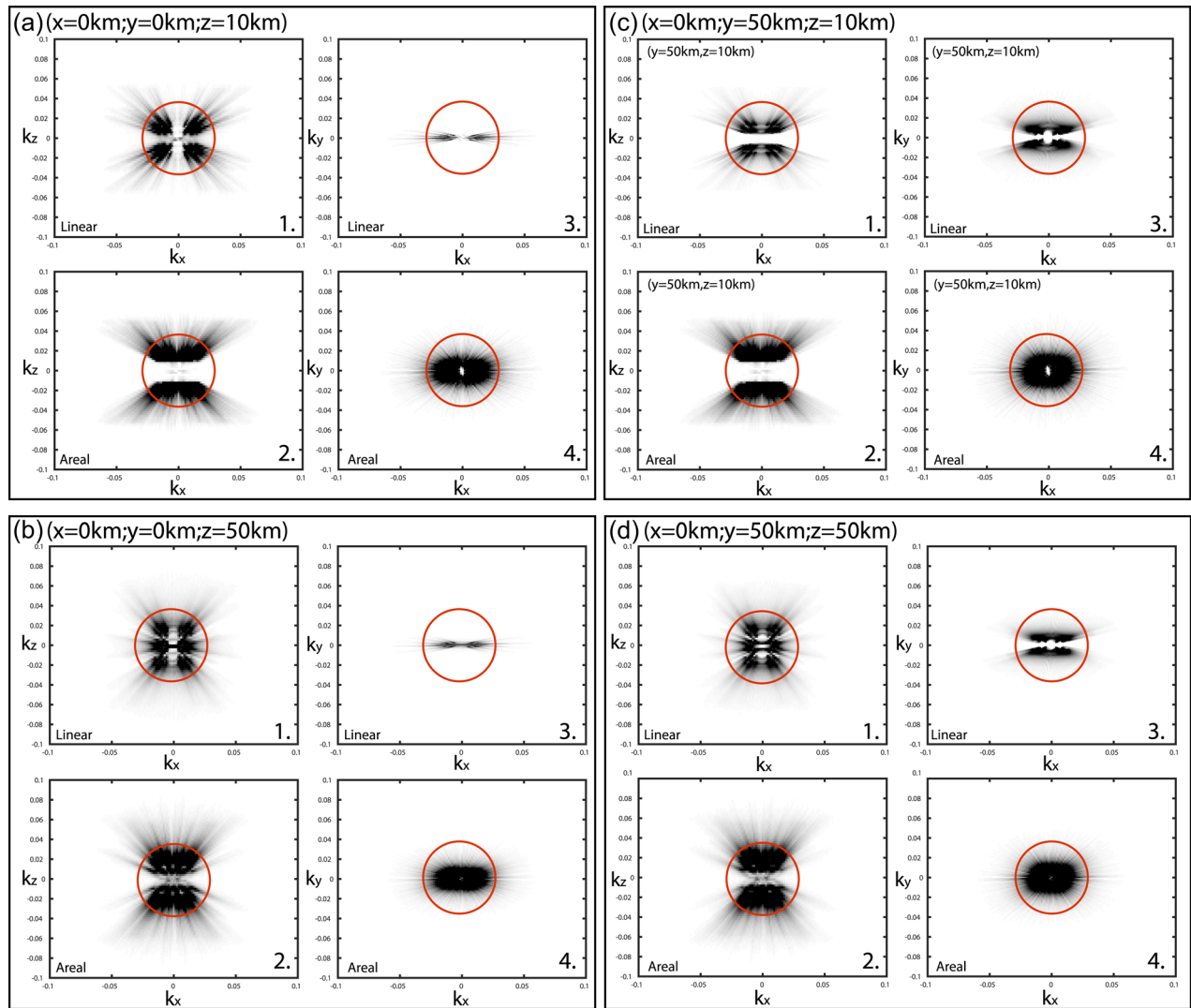


Figure 14. Wavenumber illumination provided by the linear (top panels) and areal (bottom panels) acquisitions for a scatterer located at $(x[\text{km}], y[\text{km}], z[\text{km}]) =$ (a) $(0,0,10)$, (b) $(0,0,50)$, (c) $(0,50,10)$ and (d) $(0,50,50)$. The first two scatterers are aligned with the station profile of the linear acquisition, while the last two ones are out of plane. In (a–d), the left panels show the (k_x, k_z) spectrum, while the right panels show the (k_x, k_y) spectrum. The red circle is superimposed as a guideline to assess the sampling of the dip angle along this circle.

the V_s and ρ models than on the V_p model because S waves have a leading role in the V_s and ρ updates, unlike in the V_p one. This noise can result from the acquisition footprint from the receiver side and spatial aliasing due to dip angle undersampling. The decrease of the noise with depth is consistent with the fact that (k_x, k_z) bandwidth becomes wider towards low wavenumbers with depth as highlighted in Figs 14(a)–(d)-2. This contributes to make the (k_x, k_z) coverage more uniform with depth, and hence less prone to aliasing, in the area delineated by the red circle (Fig. 14a–d-2).

For the line geometry, the reconstruction of the three parameters has been significantly degraded hence confirming the previous FWI results on the Alps model (Fig. 16). In the vertical section, only the shallow V_p , V_s and ρ inclusions are reconstructed at 10 km depth, while the inversion failed to reconstruct them at greater depths. Comparing these results with those obtained with the areal geometry highlights the key contribution of off-line stations to reconstruct the inclusions at great depths in the vertical plane defined by the central receiver line. Again, this results because the spherical inclusions behave as point diffractor that scatter waves in all spatial directions. Therefore, the wavefield scattered by the in-line inclu-

sions can be recorded by the full station layout leading to a broad scattering and dip angle illumination and a high fold. The broader and more-uniform (k_x, k_z) coverage generated by the areal acquisition relative to the linear one supports this statement (Figs 14a–d (1 versus 2)). Note however that it is not guaranteed that more cylindrical geological structures would emphasise the same add-value of the areal geometry relative to the linear one. A second possible factor that explains the rapidly decreasing sensitivity in depth of the imaging is that the surface waves that are generated by conversion of the incident P wave onto Rayleigh waves near the surface might have a significant contribution in both the vertical and horizontal resolution. The wavelength of the Rayleigh wave at 0.1 Hz (30 km) is consistent with the depth resolution shown in Fig. 16(a).

Concerning horizontal resolution, the depth slices at 10 km depth show quite noisy reconstructions of the ρ and V_s inclusions, while only the inclusions along the receiver line are reconstructed in the V_p slices. Compared to the results obtained with the areal geometry, the smearing along the cross-line (y) direction is more significant, which reflects a deficit of wavenumber coverage along this

2D array - 25 km

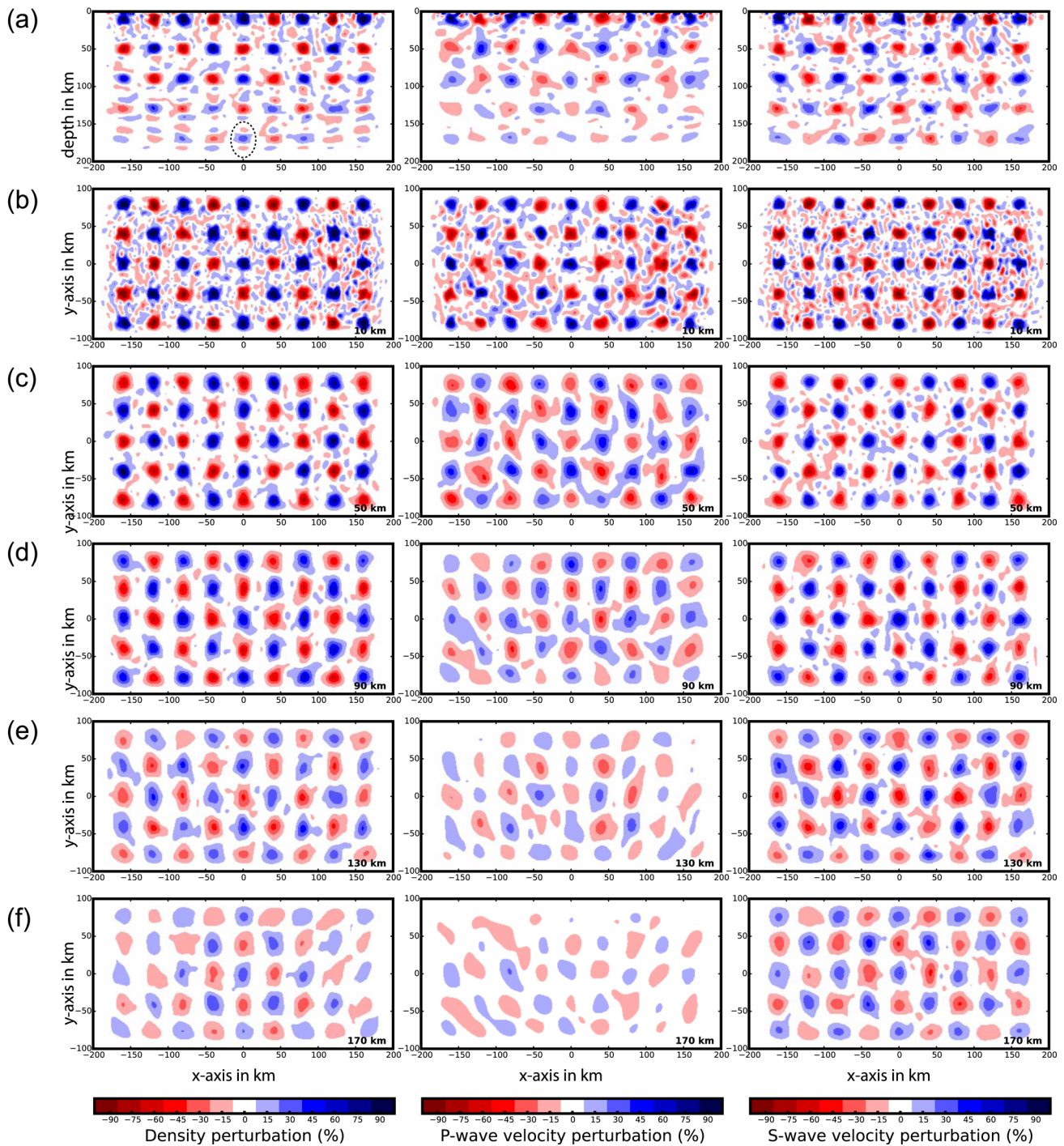


Figure 15. Resolution test for the areal acquisition. (a) Vertical section in the middle of the receiver grid. (b–f) Depth slices at 10, 50, 90, 130 and 170 km depth. From left to right, reconstructed ρ , V_p and V_s inclusions. The dash ellipse in (a) (density perturbations) highlights the inclusion distortion with alternated positive and negative vertical perturbations resulting from limited bandwidth effects (see the text for details).

dimension, while the significant amount of noise in the ρ and V_s models is likely due to aliasing. The poor illumination of the k_y components by the linear acquisition is well illustrated by the narrow (k_x, k_y) coverage in Figs 14(a)–(d-3). The failure of the V_p reconstruction off the receiver line might result from the small amplitudes of the P – P waves that are back-scattered towards the receiver line (back-scattered waves are those which dominantly sample the

horizontal components of the wavenumber vectors), while V_s and ρ reconstructions might have benefited from the various off-line back-scattering modes generated by the inclusions and the free surface. As for the vertical resolution, a second factor which can explain the contrasted horizontal resolution in the shallow updates of V_p compared to V_s and ρ is related to the contribution of Rayleigh waves, which are more sensitive to V_s perturbations.

1D array - 5 km

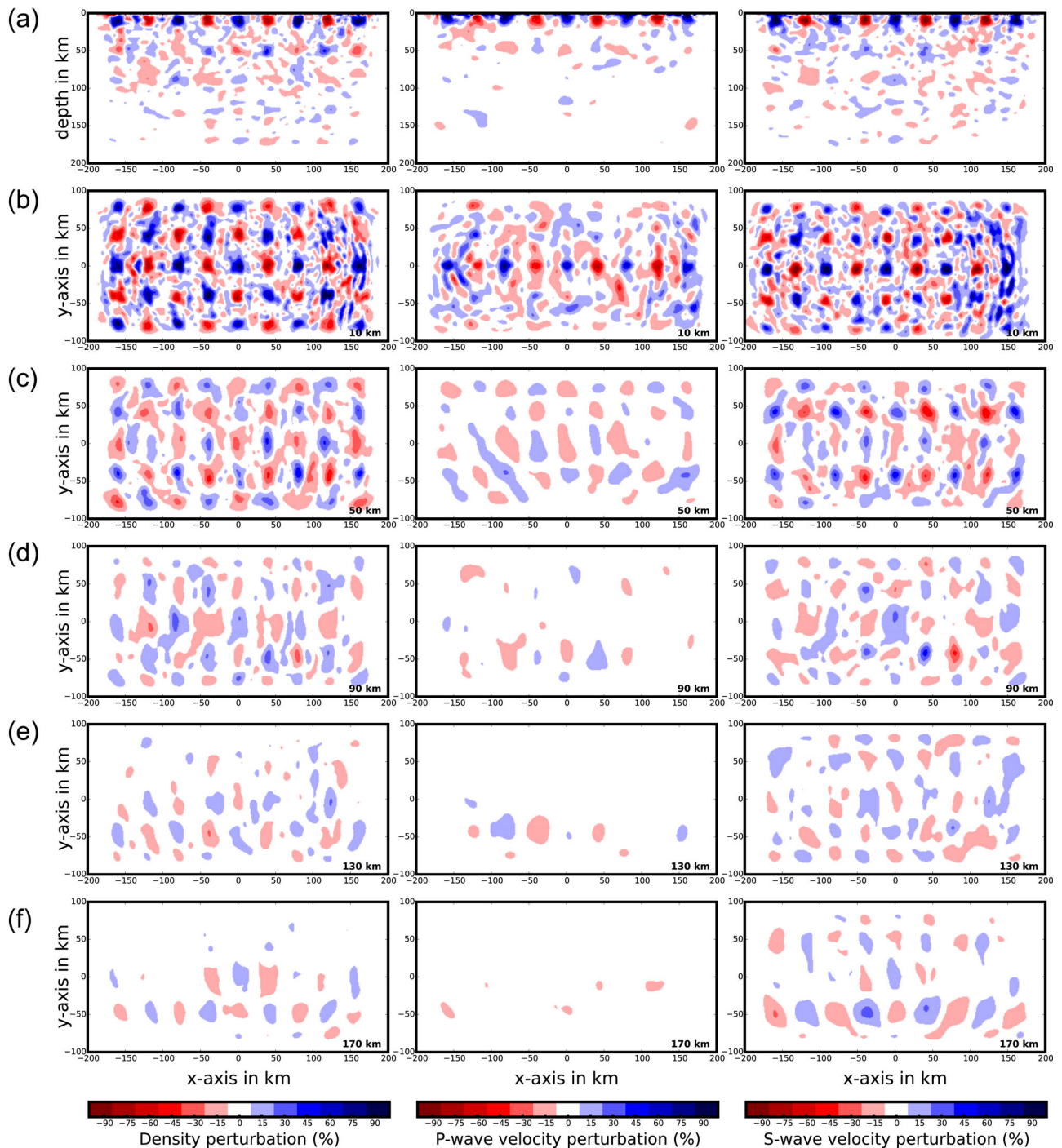


Figure 16. Resolution test for the linear acquisition. (a) Vertical section along the receiver line. (b–f) Depth slices at 10, 50, 90, 130 and 170 km depth. From left to right, reconstructed ρ , V_p and V_s inclusions.

We remind that our synthetic tests were performed for noise-free synthetic data thus limiting our acquisition assessment to ideal cases where only the scattering, dip and azimuth angle illumination provided by the source-receiver geometry controls the resolution of the FWI images. In this setting, we have shown that, compared to a linear acquisition, an areal acquisition will contribute to improve the resolution of the imaging by sampling a broader wavenumber spectrum of the lithospheric target.

For a real data application, noise will indeed impact upon this resolution. The impact of this noise can be mitigated by involving more redundant information in the imaging process either by injecting more close events in the inversion or by decreasing the station interval. The issue with the second option is that, for a given number of stations, a refinement of the station interval along a preferential direction (linear-type acquisition) will be performed at the expense of the angular illumination and hence the imaging

resolution (Fig. 14). Moreover, depending on the geodynamical context (that will guide the orientation of the receiver line along the main structural dips) and the distribution of the teleseismic events, there might be a limited number of events reasonably aligned in the azimuth of the receiver line, hence further degrading the useful angular illumination.

This is the reason why we would favour the first option, that is combining a coarse areal acquisition to cover a broad angular bandwidth with a large number of events to guarantee a redundant sampling of this bandwidth. In practice, moderate to strong earthquakes tend to be very localized on the Earth since they mostly occur along plate boundaries. Therefore, it is reasonable to assume that large catalogues of close events can be constituted to build redundant teleseismic data sets for FWI. In summary, combining coarse areal station deployments with large catalogues of redundant events is from our viewpoint the best strategy to conciliate the resolution and the signal-to-noise specifications.

5 CONCLUSIONS

We have analysed several theoretical and experimental factors that have a significant influence on teleseismic FWI for lithospheric imaging. We have designed a realistic lithospheric model of the Western Alps and used nine events collected during the CIFALPS experiment to assess teleseismic FWI with a realistic experimental set-up.

The first open question that has been addressed deals with the best subsurface parametrization for teleseismic FWI in the elastic isotropic approximation. We have concluded that (ρ, λ, μ) were the most suitable optimization parameters based on the analysis of the radiation patterns. Recombining *a posteriori* these three parameters into P and S impedances provide the lithospheric reconstruction that correlates best with the true model. During this recombination, sharp density reconstruction through the multiplication with the Lamé parameters, contribute to better focus some key structures such as shallow sedimentary basin, the Ivrea body and the subducting slab by injecting high wavenumber components. However, (ρ, V_s, V_p) or (ρ, I_p, I_s) optimization parameters also provide reliable and close results without recombination. The conclusions drawn from this synthetic experiment will need to be validated against real data applications where the presence of noise might damage the reconstruction of the density and hence the quality of the I_p and I_s built by recombination of the (ρ, λ, μ) optimization parameters.

A second important conclusion is that 1-D global reference models such as PREM provide reliable initial models for teleseismic FWI. Such reference models need to be smoothed before being used for FWI in order to the imprint of the layering. The short wavelengths associated with this layering cannot be updated by FWI as they are beyond the resolution limits of FWI and may drive the inversion to a local minimum if their geometries is inaccurate. The low frequency content of the teleseismic data makes the inversion almost immune to cycle skipping. However, too strong amplitude residuals generated by too simple initial models, such as vertical gradient models, can make the inversion to overupdate the density parameter at the expense of the wave-speed updates.

Designing reliable acquisition for teleseismic FWI is a key issue since the best trade-off before station sampling and layout spread should be found for a given pool of stations. We have tested two main acquisition trends corresponding to areal geometries designed with coarse regular grids of stations and dense linear acquisitions. Our results clearly indicate that areal geometry should be designed

to improve both the penetration in depth of the imaging and the horizontal resolution. The main reason is that, compared to a linear acquisition, an areal geometry will tremendously increase the fold and the horizontal and vertical wavenumber coverages with which a diffractor point will be imaged. This statement applies both to diffractor points located in the vertical section defined by the linear acquisition and out of this vertical section.

ACKNOWLEDGEMENTS

This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by CGG, CHEVRON, EXXON-MOBIL, JGI, SHELL, SINOPEC, STATOIL, TOTAL and WOODSIDE. This study was granted access to the HPC resources of SIGAMM infrastructure (<http://crimson.oca.eu>), hosted by Observatoire de la Côte d'Azur and which is supported by the Provence-Alpes Côte d'Azur region, and the HPC resources of CINES/IDRIS under the allocation 046091 made by GENCI.

REFERENCES

- Alkhalifah, T. & Plessix, R., 2014. A recipe for practical full-waveform inversion in anisotropic media: an analytical parameter resolution study, *Geophysics*, **79**(3), R91–R101.
- Baeten, G., de Maag, J.W., Plessix, R.-E., Klaassen, R., Qureshi, T., Kleemeyer, M., ten Kroode, F. & Rujie, Z., 2013. The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study, *Geophys. Prospect.*, **61**(4), 701–711.
- Baker, B. & Roecker, S., 2014. A full waveform tomography algorithm for teleseismic body and surface waves in 2.5 dimensions, *Geophys. J. Int.*, **198**, 1775–1794.
- Bavalia, K., Motaghia, K., Soboutia, F., Ghodsa, A., Abbasib, M., Priestley, K., Mortezaejada, G. & Rezaeiana, M., 2016. Lithospheric structure beneath NW Iran using regional and teleseismic travel-time tomography, *Phys. Earth planet. Inter.*, **253**, 97–107.
- Beller, S., Monteiller, V., Operto, S., Nolet, G., Paul, A. & Zhao, L., 2017. Lithospheric architecture of the South-Western Alps revealed by multi-parameter teleseismic full-waveform inversion, *Geophys. J. Int.*, in press, doi:10.1093/gji/ggx216.
- Bostock, M., Rondenay, S. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves I. theory for oblique incidence, *J. geophys. Res.*, **106**(12), 30 771–30 782.
- Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.*, **207**(3), 1739–1766.
- Bunks, C., Salek, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473.
- Burdick, S., de Hoop, M.V., Wang, S. & van der Hilst, R.D., 2014. Reverse-time migration-based reflection tomography using teleseismic free surface multiples, *Geophys. J. Int.*, **16**(2), 996–1017.
- Chen, P., Jordan, T. & Zhao, L., 2007. Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods, *Geophys. J. Int.*, **170**, 175–181.
- Crotwell, H.P., Owens, J.T. & Ritsema, J., 1999. The TauP Toolkit: flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.*, **70**, 154–160.
- Devaney, A., 1984. Geophysical diffraction tomography, *IEEE Trans. Geosci. Remote Sens.*, **GE-22**(1), 3–13.
- Dziewonski, A., Hales, A. & Lapwood, E., 1975. Parametrically simple earth models consistent with geophysical data, *Phys. Earth planet. Inter.*, **10**(1), 12–48.
- Dziewonski, A., Chou, T.-A. & Woodhouse, J., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. geophys. Res.*, **86**(B4), 2825–2852.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**, 297–356.

- Ekström, G., Nettles, M. & Dziewoński, A., 2012. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes, *Phys. Earth planet. Inter.*, **200**, 1–9.
- Fichtner, A., Bunge, H.P. & Igel, H., 2006a. The adjoint method in seismology: I - Theory, *Phys. Earth planet. Inter.*, **157**(1–2), 86–104.
- Fichtner, A., Bunge, H.P. & Igel, H., 2006b. The adjoint method in seismology: II - Applications, *Phys. Earth planet. Inter.*, **157**(1–2), 105–123.
- Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, **179**(3), 1703–1725.
- Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.P., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and past states of the Australasian upper mantle, *Earth planet. Sci. Lett.*, **290**(3–4), 270–280.
- Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & Villaseñor, A.V., 2013. Multiscale full waveform inversion, *Geophys. J. Int.*, **194**, 534–556.
- Forgues, E. & Lambaré, G., 1997. Parameterization study for acoustic and elastic ray+born inversion, *J. Seismic Explor.*, **6**, 253–278.
- Gholami, Y., Brossier, R., Operto, S., Ribodetti, A. & Virieux, J., 2013. Which parametrization is suitable for acoustic VTI full waveform inversion? - Part I: Sensitivity and trade-off analysis, *Geophysics*, **78**(2), R81–R105.
- Groos, L., Schäfer, M., Forbriger, T. & Bohlen, T., 2014. The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves, *Geophysics*, **79**(6), R247–R261.
- Hicks, G.J. & Pratt, R.G., 2001. Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit, *Geophysics*, **66**(2), 598–612.
- Iglesias, A., Clayton, R.W., Perez-Campos, X., Singh, S.K., Pacheco, J.F., Garcia, D. & Valdes-Gonzalez, C., 2010. S wave velocity structure below central Mexico using high-resolution surface wave tomography, *J. geophys. Res.*, **115**, B06307, doi:10.1029/2009JB006332.
- Kamei, R. & Pratt, R.G., 2013. Inversion strategies for visco-acoustic waveform inversion, *Geophys. J. Int.*, **194**, 859–894.
- Kennett, B.L.N. & Engdahl, E.R., 1991. Travel times for global earthquake location and phase association, *Geophys. J. Int.*, **105**, 429–465.
- Köhn, D., De Nil, D., Kurzman, A., Przebindowska, A. & Bohlen, T., 2012. On the influence of model parametrization in elastic full waveform tomography, *Geophys. J. Int.*, **191**, 325–345.
- Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for 3D seismic wave propagation, *Geophys. J. Int.*, **139**, 806–822.
- Komatitsch, D., Tromp, J. & Vilotte, J.P., 1998. The spectral element method for elastic wave equations: application to 2D and 3D seismic problems, in *Expanded Abstracts*, vol. II, pp. 1460–1463, Bulletin of the Seismological Society of America.
- Kurzman, A., Przebindowska, A., Kohn, D. & Bohlen, T., 2013. Acoustic full waveform tomography in the presence of attenuation: a sensitivity analysis, *Geophys. J. Int.*, **195**(2), 985–1000.
- Lambaré, G., 2008. Stereotomography, *Geophysics*, **73**(5), VE25–VE34.
- Lambaré, G., Operto, S., Podvin, P., Thierry, P. & Noble, M., 2003. 3-D ray+Born migration/inversion - Part I: 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.
- Lumley, D.E., Claerbout, J.F. & Bevc, D., 1994. Anti-aliased Kirchhoff 3-D migration, in *Expanded Abstracts*, 64th Annual SEG Meeting and Exposition, pp. 1282–1285, Society of Exploration Geophysics.
- Maggi, A., Tape, C., Chen, M., Chao, D. & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.*, **178**, 257–281.
- 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**(3), 1179–1204.
- Marquering, H., Dahlen, F.A. & Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.*, **137**, 805–815.
- Masson, Y. & Romanowicz, B., 2017. Box tomography: localised imaging of remote targets buried in an unknown medium, a step forward for understanding key structures in the deep earth, *Geophys. J. Int.*, **211**, 141–163.
- McMechan, G.A. & Fuis, G.S., 1987. Ray equation migration of wide-angle reflections from southern Alaska, *J. geophys. Res.*, **92**(1), 407–420.
- Métivier, L., Brossier, R., Méridot, Q., Oudet, E. & Virieux, J., 2016. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion, *Geophys. J. Int.*, **205**, 345–377.
- Miller, D., Oristaglio, M. & Beylkin, G., 1987. A new slant on seismic imaging: Migration and integral geometry, *Geophysics*, **52**(7), 943–964.
- Monteiller, V., Chevrot, S., Komatitsch, D. & Tromp, J., 2013. A hybrid method to compute short period synthetic seismograms of teleseismic body waves in a 3-D regional model, *Geophys. J. Int.*, **192**(1), 230–247.
- Monteiller, V., Beller, S., Operto, S. & Virieux, J., 2015a. Influence of global heterogeneities on regional imaging based upon full waveform inversion of teleseismic wavefield, in *EGU General Assembly Conference Abstracts*, vol. 17.
- Monteiller, V., Chevrot, S., Komatitsch, D. & Wang, Y., 2015b. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method, *Geophys. J. Int.*, **202**(2), 811–827.
- Mora, P.R., 1987. Nonlinear two-dimensional elastic inversion of multi-offset seismic data, *Geophysics*, **52**, 1211–1228.
- Nissen-Meyer, T., van Driel, M., Stähler, S., Hosseini, K., Hempel, S., Auer, L., Colombi, A. & Fournier, A., 2014. Axisem: broadband 3-D seismic wavefields in axisymmetric media, *Solid Earth*, **5**(1), 425–445.
- Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage, *Math. Comput.*, **35**(151), 773–782.
- Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization*, 2nd edn, Springer.
- Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prieux, V., Ribodetti, A. & Virieux, J., 2013. A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice, *Leading Edge*, Special section Full Waveform Inversion (September), 1040–1054.
- Pageot, D., Operto, S., Vallée, M., Brossier, R. & Virieux, J., 2010. Lithospheric imaging from teleseismic data by frequency-domain elastic full-waveform tomography, in *Expanded Abstracts*, p. WS6, EAGE.
- Pageot, D., Operto, S., Vallée, M., Brossier, R. & Virieux, J., 2013. A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging, *Geophys. J. Int.*, **193**(3), 1479–1505.
- 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**(2), 495–503.
- Plessix, R.E. & Cao, Q., 2011. A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium, *Geophys. J. Int.*, **185**, 539–556.
- Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, part I: theory and verification in a physical scale model, *Geophysics*, **64**, 888–901.
- Pratt, R.G., 2008. Waveform tomography - successes, cautionary tales, and future directions, in *Presented at the 70th Annual EAGE Conference & Exhibition*, Roma, pp. 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.
- Prieux, V., Brossier, R., Operto, S. & Virieux, J., 2013. Multiparameter full waveform inversion of multicomponent OBC data from valhall. Part I: Imaging compressional wavespeed, density and attenuation, *Geophys. J. Int.*, **194**(3), 1640–1664.
- Roecker, S., Baker, B. & McLaughlin, J., 2010. A finite-difference algorithm for full waveform teleseismic tomography, *Geophys. J. Int.*, **181**, 1017–1040.
- 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**(12), 30 795–30 807.

- 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*, *Geophysical Monograph 157*, pp. 187–203, ed. Alan Levander, G.N., American Geophysical Union, Washington, DC.
- Ryberg, T. & Weber, M., 2000. Receiver function arrays: a reflection seismic approach, *Geophys. J. Int.*, **141**(1), 1–11.
- Shang, X., de Hoop, M.V. & van der Hilst, R.D., 2012. Beyond receiver functions: passive source reverse time migration and inverse scattering of converted waves, *Geophys. Res. Lett.*, **39**, L15308, doi:10.1029/2012GL052289.
- Shragge, J., Bostock, M.G. & Rondenay, S., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves 2. Numerical examples, *J. geophys. Res.*, **106**(12), 30 783–30 793.
- Sieminski, A., Trampert, J. & Tromp, J., 2009. Principal component analysis of anisotropic finite-frequency sensitivity kernels, *Geophys. J. Int.*, **179**(2), 1186–1198.
- Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics*, **69**(1), 231–248.
- Symes, W.W., 2008. Migration velocity analysis and waveform inversion, *Geophys. Prospect.*, **56**, 765–790.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.*, **180**, 433–462.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- Tarantola, A., 1986. A strategy for non linear inversion of seismic reflection data, *Geophysics*, **51**(10), 1893–1903.
- Tong, P., Chen, C.-W., Komatitsch, D., Basini, P. & Liu, Q., 2014a. High-resolution seismic array imaging based on an SEM-FK hybrid method, *Geophys. J. Int.*, **197**, 369–395.
- Tong, P., Komatitsch, D., Tseng, T.-L., Hung, S.-H., Chen, C.-W., Basini, P. & Liu, Q., 2014b. A 3-D spectral-element and frequency-wave number hybrid method for high-resolution seismic array imaging, *Geophys. Res. Lett.*, **41**, 7025–7034.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- van Leeuwen, T. & Herrmann, F.J., 2013. Mitigating local minima in full-waveform inversion by expanding the search space, *Geophys. J. Int.*, **195**, 661–667.
- Vinnik, L., 1977. Detection of waves converted from p to sv in the mantle, *Phys. Earth planet. inter.*, **15**(1), 39–45.
- Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Wang, Y. et al., 2016. The deep roots of the western Pyrenees revealed by full waveform inversion of teleseismic p waves, *Geology*, **44**(6), 475–478.
- Weidle, C. & Widiyantoro, S., 2005. Improving depth resolution of teleseismic tomography by simultaneous inversion of teleseismic and global P-wave traveltimes data—application to the Vrancea region in southeastern Europe, *Geophys. J. Int.*, **162**(3), 811–823.
- Williamson, P., 1991. A guide to the limits of resolution imposed by scattering in ray tomography, *Geophysics*, **56**, 202–207.
- Wilson, C. & Guitton, A., 2007. Teleseismic wavefield interpolation and signal extraction using high-resolution linear radon transforms, *Geophys. J. Int.*, **168**(1), 171–181.
- Wu, R.S. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics*, **50**(4), 582–595.
- Wu, R.S. & Toksöz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics*, **52**, 11–25.
- Zhu, L. & Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions, *J. geophys. Res.*, **105**(B2), 2969–2980.
- Zhu, H., Bozdağ, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nat. Geosci.*, **5**, 493–498.
- Zhu, H., Bozdağ, E. & Tromp, J., 2015. Seismic structure of the European upper mantle based on adjoint tomography, *Geophys. J. Int.*, **201**(1), 18–52.
- Zhao, L. et al., 2015. First seismic evidence for continental subduction beneath the western alps, *Geology*, **43**(9), 815–818.