



HAL
open science

Slope tomography based on eikonal solvers and the adjoint-state method

B. Tavakoli F., S. Operto, A. Ribodetti, J. Virieux

► **To cite this version:**

B. Tavakoli F., S. Operto, A. Ribodetti, J. Virieux. Slope tomography based on eikonal solvers and the adjoint-state method. *Geophysical Journal International*, 2017, 209 (3), pp.1629 - 1647. 10.1093/gji/ggx111 . hal-01524296

HAL Id: hal-01524296

<https://hal.science/hal-01524296>

Submitted on 19 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

Slope tomography based on eikonal solvers and the adjoint-state method

B. Tavakoli F.,¹ S. Operto,¹ A. Ribodetti¹ and J. Virieux²

¹*Geoazur, Université Côte d'Azur, CNRS, IRD, OCA, Valbonne, France. E-mail: tavakoli@geoazur.unice.fr*

²*ISTerre, Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR, F-38000 Grenoble, France*

Accepted 2017 March 14. Received 2017 March 2; in original form 2016 October 31

SUMMARY

Velocity macromodel building is a crucial step in the seismic imaging workflow as it provides the necessary background model for migration or full waveform inversion. In this study, we present a new formulation of stereotomography that can handle more efficiently long-offset acquisition, complex geological structures and large-scale data sets. Stereotomography is a slope tomographic method based upon a semi-automatic picking of local coherent events. Each local coherent event, characterized by its two-way traveltime and two slopes in common-shot and common-receiver gathers, is tied to a scatterer or a reflector segment in the subsurface. Ray tracing provides a natural forward engine to compute traveltime and slopes but can suffer from non-uniform ray sampling in presence of complex media and long-offset acquisitions. Moreover, most implementations of stereotomography explicitly build a sensitivity matrix, leading to the resolution of large systems of linear equations, which can be cumbersome when large-scale data sets are considered. Overcoming these issues comes with a new matrix-free formulation of stereotomography: a factored eikonal solver based on the fast sweeping method to compute first-arrival traveltimes and an adjoint-state formulation to compute the gradient of the misfit function. By solving eikonal equation from sources and receivers, we make the computational cost proportional to the number of sources and receivers while it is independent of picked events density in each shot and receiver gather. The model space involves the subsurface velocities and the scatterer coordinates, while the dips of the reflector segments are implicitly represented by the spatial support of the adjoint sources and are updated through the joint localization of nearby scatterers. We present an application on the complex Marmousi model for a towed-streamer acquisition and a realistic distribution of local events. We show that the estimated model, built without any prior knowledge of the velocities, provides a reliable initial model for frequency-domain FWI of long-offset data for a starting frequency of 4 Hz, although some artefacts at the reservoir level result from a deficit of illumination. This formulation of slope tomography provides a computationally efficient alternative to waveform inversion method such as reflection waveform inversion or differential-semblance optimization to build an initial model for pre-stack depth migration and conventional FWI.

Key words: Inverse theory; Numerical modelling; Tomography; Seismic tomography.

1 INTRODUCTION

Building a velocity macromodel from reflection data remains one of the most crucial and challenging issues in seismic imaging. The difficulty arises from the nonlinearity of the inverse problem associated with the long-wavelength reconstruction of the subsurface; by contrast the imaging of reflectivity by migration is a far more linear problem (Claerbout 1985). Velocity model building provides the necessary background or starting model to perform pre-stack-depth migration (Etgen *et al.* 2009) or full waveform inversion (FWI; Tarantola 1984).

Among the most popular methods for velocity macromodel building, reflection traveltime tomography (Bishop *et al.* 1985; Farra & Madariaga 1988) and migration-based velocity analysis (MVA; Gardner *et al.* 1974; Al-Yahya 1989) were specifically designed for seismic reflection data. Reflection traveltime tomography, which builds a velocity model through the minimization of the traveltime residuals, is computationally efficient but relies on tedious picking of continuous horizons. MVA methods rely on an explicit scale separation between the background velocities and the reflectivity to iteratively alternate the velocity update and the migration. The velocity update is driven by flattening the reflectors in the

common image gathers or by minimizing reflection data residuals in the time domain after demigration. Among the MVA methods, some rely on picking in the pre-stack migration volume (Al-Yahya 1989) or on more automatic waveform-based misfit criteria as in differential semblance optimization (DSO; Symes 1998; Chauris & Noble 2001) or migration-based traveltimes inversion (MBTT; Clément *et al.* 2001). Recently, the governing ideas of the MBTT method have been recast in the framework of FWI, leading to the so-called reflection waveform inversion (RWI; Xu *et al.* 2012; Brossier *et al.* 2015; Wu & Alkhalifah 2015; Wang *et al.* 2016) with an extension to the joint inversion of diving waves and reflections proposed by Zhou *et al.* (2015). One drawback of MVA and RWI is the computational cost resulting from the migration performed at each cycle of the alternating optimization.

The computational cost of MVA and the above-mentioned variants of the FWI make traveltimes tomography an attractive technique for velocity model building, where we do not need repeated migrations. However, picking either in the data domain or in the migrated domain is a burden in traveltimes tomography. As a remedy for the issue of horizon-based picking in reflection tomography, use of local coherent events in pre-stack data sets has been a breakthrough, first introduced by Rieber (1936) and followed by Riabinin (1957) and Sword (1987) in the framework of the controlled directional reception (CDR) method. These local events are interpreted as arrivals reflected from small reflecting facets. Main motivation of CDR has been to add, in the data domain, the slopes of the local coherent events to the traveltimes to better constrain the velocity model building process.

Billette & Lambaré (1998) extended the CDR method to stereotomography, which relies on a semi-automatic picking of traveltimes and slopes of local coherent events in both shot and receiver gathers. Each of these local events is tied to a reflecting/diffracting facet in the subsurface. The misfit between recorded and modelled two-way traveltimes and slopes at the source and receiver positions is minimized to update the subsurface velocities and the coordinates of the facets in depth. Moreover, take-off angles and the one-way traveltimes of the rays connecting the facet to the source and the receiver can be added as optimization parameters with the aim of making the forward problem more efficient by avoiding two-point ray tracing or making the inverse problem less prone to local minima by expanding the search space. For this, the source and receiver positions need also to be fitted since there is no guarantee that the rays associated with the initial guess of take-off angles and one-way traveltimes will reach the shot and receiver positions.

Through different applications of stereotomography (Billette *et al.* 2003; Lambaré *et al.* 2004; Nag *et al.* 2006; Alerini *et al.* 2007; Prieux *et al.* 2013), some practical workflows have been designed to mitigate the nonlinearity of the inversion. These strategies rely on a suitable scaling of the different optimization parameter classes, suitable initial localization of the facets and a multiscale approach during which the grid on which the velocity model is parametrized is progressively refined. In spite of all these efforts, stereotomography still suffers from the limitations of ray-based techniques to handle complex media (Lambaré 2008). A second limitation of computational nature is related to the fact that the inversion algorithm relies on the explicit building from paraxial quantities of the sensitivity matrix which has a (data size) \times (model size) complexity. Although this matrix-based implementation is suitable for a sensitivity analysis of the multiparameter inversion, it represents a significant computational burden when huge 3-D data sets are tackled. A third limitation is related to the significant number of optimization parameters of different nature in ray-based stereotomography (e.g.

velocity and ray parameters) that can make the inversion poorly conditioned and hence difficult to scale.

In order to overcome these difficulties, we present a new formulation of stereotomography. The first key ingredient is the computation of traveltimes by solving the factored eikonal solver with the fast sweeping method (FSM; Zhao 2005; Fomel *et al.* 2009). Our main motivation behind the use of eikonal solver at the expense of ray tracing is to avoid the issue of non-uniform ray sampling that can arise in presence of low velocity zone or area of rapidly varying wave speeds. Moreover, a factored form of the eikonal equation improves the accuracy of the traveltimes computation nearby the sources. Exploiting source–receiver reciprocity, we compute the synthetic slopes at the source and receiver positions by finite differences from traveltimes maps computed from neighbour source and receiver positions. By doing so, we limit the computational burden of our approach by solving the eikonal equations from the source and receiver positions rather than from the facets. The second key ingredient is the computation of the gradient of the stereotomography misfit function with the matrix-free adjoint-state method. Plessix (2006) developed the gradient of the stereotomography misfit function from the ray equations and showed how solving the adjoint-state system is less computationally intensive than solving the propagator system required to compute the sensitivity matrix with paraxial ray equations. In our formulation, the state equations resulting from the use of a finite-difference eikonal solver as forward-modelling engine (eikonal equation and slope estimation from finite differences of traveltimes maps) drive us towards a model space formed by the subsurface velocities and the coordinates of the scatterers. This model space may be easier to manage than those used in ray-based stereotomography (Billette & Lambaré 1998) because it involves a smaller number of parameter classes. Therefore, this choice may mitigate issues of parameter cross-talk and ill-conditioned inversion. In this study, we support the legitimacy of this statement with an application to the complex Marmousi model. Other applications of classic stereotomography on the Marmousi model with different model parametrizations are presented by Chauris *et al.* (2002b) and Billette *et al.* (2003).

In the following, we first review the principles of the classical ray-based stereotomography. Then we introduce the forward-problem (state) equations that are used to compute traveltimes and slopes in our formulation referred to as adjoint slope tomography. From these equations, we introduce the data and model spaces, and compute the gradient with a formulation of a semi-discretized adjoint-state method. We then discuss some practical aspects of the implementation of the method before showing its relevance with three synthetic examples. The third example shows an application to the Marmousi model performed with a realistic synthetic data set inferred from a sparse picking of reflectors in the Marmousi model. The accuracy of the estimated model is assessed as a starting model for frequency-domain FWI with a realistic starting frequency of 4 Hz. We manage to build a reliable initial velocity model for FWI without taking advantage of any prior knowledge of the velocity structure. However, our results also show the sensitivity of the method to sparse distribution of picks in deep complex zone or near the ends of the acquisition layout.

2 2-D CLASSICAL STEREOTOMOGRAPHY

Slope tomography methods allow for the estimation of subsurface velocity models from slopes and traveltimes of local coherent events

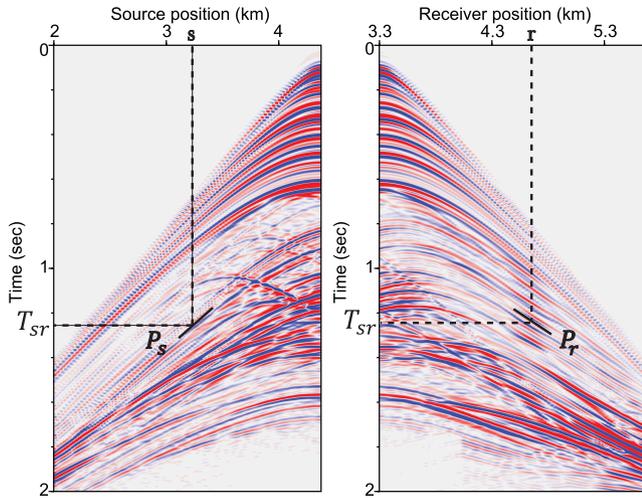


Figure 1. Local coherent events in stereotomography. Each local coherent event is described by the source and receiver positions (s, r), the slopes (p_s, p_r) picked in common-shot and common-receiver gathers and two-way traveltime T_{SR} .

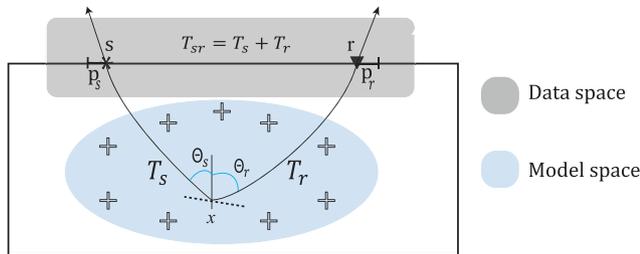


Figure 2. Data space and model space in classical stereotomography. The symbols $+$ denote the velocity nodes. Two rays are shot towards the source ‘s’ and receiver ‘r’ from the scatterer x with take-off angles Θ_s and Θ_r . Corresponding one-way traveltimes are T_s and T_r . The velocities, the scatterer coordinates x and the ray attributes (Θ_s, Θ_r, T_s and T_r) form the model space of classical stereotomography. The scatterer position and the two take-off angles (Θ_s, Θ_r) define a dip bar (or migration facet) shown by the segment running through the scatterer. The horizontal component of the slowness vectors at the source and receiver positions, p_s and p_r , the two-way traveltime T_{SR} and the source and receiver positions form the data space of the classical stereotomography (adapted from Billette & Lambaré 1998).

on unmigrated data (Fig. 1). These local coherent events correspond most of the time to a reflection from a small reflector segment (Billette *et al.* 2003). However, they can also represent diffractions or diving waves (Plessix 2006; Prioux *et al.* 2013). Since the structures that generate local coherent events are parametrized by point diffractor in slope tomography, we will refer to them as scatterers in the following. For two dimensional acquisitions, a stereotomographic data set consists of N locally coherent events that can be parametrized by

$$\mathbf{d} = [(s, r, T_{s,r}, p_s, p_r)]_{n=1}^N, \quad (1)$$

where s and r denote the position of the source and receiver, p_s and p_r are the slopes picked in the common-receiver and common-shot gathers for source s and receiver r , respectively (equivalently, the horizontal component of slowness vector at the position of source and receiver r), and $T_{s,r}$ is the two-way traveltime (Fig. 2).

The scatterer related to each pair of local events is located at a position denoted by x from which a pair of rays are propagated in a background velocity model towards the source and receiver. The take-off angles of these two rays define the dip component

of the scatterer (dip bar in Fig. 2) that is mapped by the source–receiver pair. The scatterer position and the take-off angles can be represented by dip bars as shown by Billette *et al.* (2003, their figs 4 and 9) the assemblage of which can be viewed as a skeleton of a migrated image. This ray-based description leads to the following definition of the stereotomographic parameters

$$\mathbf{m} = \{(\mathbf{x}, \theta_s, \theta_r, T_s, T_r)_{n=1}^N, [c_m]_{m=1}^M\}, \quad (2)$$

where fields T_s and T_r stand for one-way traveltimes of the rays shot with take-off angles θ_s and θ_r from the scatterer towards the source and receiver (Fig. 2). Quantities T_s , T_r , θ_s and θ_r are introduced as optimization parameters mainly for sake of efficient ray tracing with end points near the actual source and receiver locations. The fact that the end points do not match the true shot and receiver positions requires involving these later as observables in the data space (eq. 1). Alternatively, two-point ray tracing can be performed at the expense of computational efficiency. This option allows one to remove T_s, T_r, θ_s and θ_r from the model space and s and r from the data space. The velocity macromodel can be parametrized by the coefficients c_m of cardinal cubic B-spline functions, which ensure the second-order continuity of velocity model required by paraxial estimations (Billette *et al.* 2003).

A subsurface model is iteratively updated by minimization of the least-squares misfit between picked and modelled data. The inverse problem is solved through the explicit building of the sensitivity matrix, which can be inferred from paraxial ray quantities (Billette & Lambaré 1998). The resulting tomographic system augmented with smoothing constraints is solved with a linear conjugate gradient such as LSQR (Paige & Saunders 1982). In practice, different strategies have been implemented in stereotomography to cope with potential nonlinearities and ill-posedness resulting from the multiparameter (i.e. velocity, ray parameters and scatterer positions) nature of the reconstruction. Billette (1998) proposed a workflow subdivided in three steps: initialization of ray parameters in a homogeneous background velocity model, estimation of a constant gradient velocity model that best fits the stereotomographic observed data, and joint inversion of the ray parameters, scatterer positions and spline coefficients modifying the background constant gradient velocity model. This procedure faced difficulties in case of complex geological structures where the second step cannot converge to a sufficiently accurate velocity model (Billette *et al.* 2003). Time stripping (Alerini *et al.* 2007) is another technique to incorporate the stereotomographic picks in a progressive manner, according to their traveltimes. For very smooth and layering velocity models, this approach seems efficient but, since the position of the scatterers and the one-way traveltimes are updated independently, this method may fail to converge in presence of strong lateral velocity variations. Target-oriented strategies are also possible to reduce the nonlinearity of the inversion (Billette 1998).

To establish a powerful, automatic and comprehensive approach, Billette *et al.* (2003) introduced a multiscale approach. They proposed an initialization of the model parameters by considering a sparse distribution of cubic cardinal B-spline nodes in each direction and a simple geometrical consideration for the ray parameters in a homogeneous background velocity model. The next step is the localization, that includes updating independently the ray parameters and positions of scatterers while the velocity is fixed at its initial value. In the third step, they perform a joint inversion of all model parameters through a multiscale approach where the grid of spline nodes is progressively refined over the different multiscale steps. In this study, we design a workflow along these lines, with different model parametrization and inversion implementation.

3 ADJOINT SLOPE TOMOGRAPHY

Compared to classical stereotomography, our implementation of slope tomography relies, on the one hand, on solving the forward problem by numerical resolution of the eikonal equation and, on the other hand, on using the adjoint-state method to compute the gradient of the slope tomography misfit function.

Although eikonal solvers do not provide explicit information on rays attributes such as slopes, they avoid the issue of non-uniform ray sampling caused by low or rapidly varying velocities in particular near grazing angles. It is worth reminding that the aim of the slope tomography is to build smooth velocity models. Therefore, our motivation behind using eikonal solvers is not to use diffractions or head waves generated by low or high velocity contrasts, these arrivals being challenging to pick in the data, but rather to take advantage of robust and accurate computation of uniformly sampled traveltimes maps in smooth media.

Current implementations of eikonal solvers do not identify multiple arrivals generated by multipathing which remains a computational challenge (Qian & Leung 2004). Therefore, we consider only the fastest solution in the present formulation. Operto *et al.* (2000, their figs 10 and 11) showed that ray+Born migration/inversion using first-arrival traveltimes provide acceptable quantitative migrated images of the Marmousi model as long as the background model is sufficiently smooth. In this case, the amplitudes of the velocity perturbations are reasonably estimated where the ray field is folded. This suggests that the information carried out by multiple arrivals in the data is accounted for in an average sense. Indeed, this amplitude recovery from first-traveltime arrivals is achieved at the expense of the quality of the focusing, that is hampered by the smoothness of the background model (compare figs 7 and 9 in Operto *et al.* 2000). We can transpose this reasoning from the migration task to the velocity model building counterpart by assuming that the velocity models estimated by slope tomography are smooth enough such that first-arrival eikonal solver predicts in average sense traveltimes of multiple arrivals that would have been picked in the data. In summary, our choice of eikonal solver rather than ray tracing is pragmatical. It provides a robust tool to compute uniformly sampled traveltimes maps that accurately predict picked traveltimes as long as the velocity model is forced to be smooth enough.

For solving the inverse problem, we compute explicitly the gradient of the misfit function with respect to the model parameters using an adjoint-state formulation. This allows us to avoid building the sensitivity matrix and performing large-scale matrix resolution. The matrix-free nature of this approach makes the inversion scheme more capable of handling large data sets. In the following section, we review the model and data spaces used in our forward-modelling engine and the adjoint formulation of slope tomography. All the mathematical symbols used in our formulation are presented in Table 1.

3.1 Data and model space definition

In adjoint slope tomography, we consider descending wave propagation from the source and receiver positions towards the scatterers (Fig. 3), unlike ray-based approaches where rays are shot from the scatterer towards the source and receiver positions with prescribed shooting angles (Billette 1998). These shooting angles explicitly define the dip of the scatterer (i.e. reflector segment if the scatterer lies on a reflector) sampled by the source–receiver pair. In our approach, we shall show that the dip component of a scatterer is

Table 1. Mathematical symbols.

Symbol	Description
N	Total number of scatterers.
M	Total number of cubic B-spline nodes.
N_s	Number of distinct shot gathers.
N_r	Number of distinct receiver gathers.
N_r^s	Number of distinct receivers in the shot-gather s .
N_n^r	Number of distinct sources in the receiver-gather r .
$N_n^{s,r}$	Number of scatterers for source–receiver pair (s, r) .
$n_{s,r}$	n th scatterer associated with source and receiver pair (s, r) .
$\mathbf{x}_{n_{s,r}}$	Position of n th scatterer associated with source–receiver pair (s, r) .
$\mathbf{x}_s, \mathbf{x}_r$	Position of source s and receiver r .
$T_{s,r,n_{s,r}}^*$	Observed two-way traveltimes for scatterer $n_{s,r}$.
$T_{s,r,n_{s,r}}$	Calculated two-way traveltimes for scatterer $n_{s,r}$.
$p_{s,n_{s,r}}^*$	Observed local slope for scatterer $n_{s,r}$ at the position of s th source of receiver-gather r .
$p_{s,n_{s,r}}$	Calculated local slope for scatterer $n_{s,r}$ at the position of s th source of receiver-gather r .
$p_{r,n_{s,r}}^*$	Observed local slope for scatterer $n_{s,r}$ at the position of r th receiver of shot-gather s .
$p_{r,n_{s,r}}$	Calculated local slope for scatterer $n_{s,r}$ at the position of r th receiver of shot-gather s .
$(\dots)^t$	Transpose operator.
\mathbf{v}	Velocity model on the Cartesian grid.
\mathbf{c}	B-spline velocity coefficients: $\mathbf{v} = \mathbf{B}\mathbf{c}$
$\mathbf{m} = \{\mathbf{c}, \mathbf{x}_{n_{s,r}}\}$	Adjoint slope tomography parameters.
$C(\mathbf{m})$	Adjoint slope tomography misfit function.

implicitly embedded in the adjoint sources. This implicit representation results from the way we estimate the slopes at the source and receiver positions by finite differences.

Since one eikonal resolution provides a traveltimes map in the whole target and each shot and receiver gather can be associated with several picked coherent events, descending propagations from the surface to the scatterers mitigates the number of forward modellings as the number of shots and receivers is expected to be one to two order of magnitude smaller than the number of scatterers. The local-coherent events, picked on common-shot and common-receiver gathers, are parametrized by their two-way traveltimes and the two slopes (in 2-D) at the source and receiver positions. These quantities define the data space of the adjoint slope tomography. To each picked event in the data space is associated a scatterer in the subsurface. Accordingly, we label each scatterer with the subscripts of the source and receiver to which it is related. This prompts the following notations for the data space of the adjoint slope tomography:

$$\mathbf{d} = (T_{s,r,n_{s,r}}, p_{s,n_{s,r}}, p_{r,n_{s,r}}) \Big|_{s=1}^{N_s} \Big|_{r=1}^{N_r} \Big|_{n_{s,r}=1}^{N_n^{s,r}}. \quad (3)$$

where s and r denote a source and receiver, $n_{s,r}$ a scatterer associated with the pair (s, r) , $T_{s,r,n_{s,r}}$ the two-way traveltimes along the paths connecting the source s and the receiver r to the scatterer $n_{s,r}$, $p_{s,n_{s,r}}$ and $p_{r,n_{s,r}}$ the horizontal component of slowness vector at the source and receiver position, respectively. Unlike in classical stereotomography, the source and receiver positions are not part of the data space since we perform wave propagation from these exact positions. In classical stereotomography, one needs to minimize distances between source or receiver position and the end points of the rays emitted with the prescribed shooting angle from the scatterer: some specificities of the ray-tracing engine has led to a particular extension of both data space (source/receiver positions) and model space (shooting angles and one-way traveltimes).

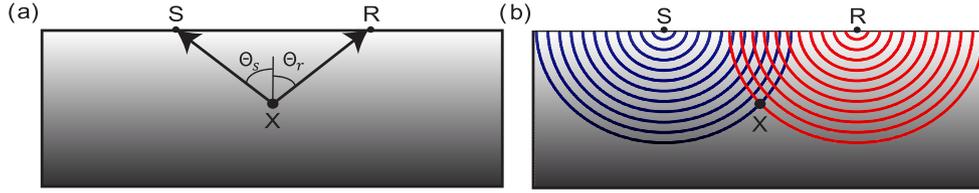


Figure 3. Layout of forward modelling in (a) classic stereotomography: two rays are back propagated from scatterer towards the associated source and receiver, and in (b) adjoint slope tomography: two traveltime maps are generated from associated source and receiver towards the scatterer.

Consequently, the model space involves two categories of parameter classes: coordinates of the scatterers $\mathbf{x}_{n_s,r} = (x_{n_s,r}, z_{n_s,r})$ and subsurface velocities:

$$\mathbf{m} = \left([c_m]_{m=1}^M, \{ \mathbf{x}_{n_s,r} \}_{s=1}^{N_s} \mid \{ \mathbf{x}_{n_s,r} \}_{r=1}^{N_r} \right)^t. \quad (4)$$

Here, c_m denote the coefficients of the cubic cardinal B-splines, that are used to parametrize the subsurface velocity model (Billette 1998). This sampling of the velocity model operates as an implicit regularization of the stereotomography. While the cubic cardinal B-spline parametrization is generally a natural choice for ordinary differential equations such as (paraxial) ray tracing, eikonal solvers as partial differential equations are often formulated for efficiency on a Cartesian grid. Nevertheless, we consider that B-splines will describe in a compact way the smooth velocity structure we are looking for. We may consider this sampling choice as a hard constraint on the model space. We shall also promote a soft constraint through the Gaussian filtering of the misfit gradient.

3.2 Forward problem

According to the data space defined in the eq. (3), the forward-modelling engine requires calculation of the slopes at the source and receiver and the two-way traveltimes for each triplet (s, r, n_s, r) . We first compute the traveltime maps $t_s(\mathbf{x})$ and $t_r(\mathbf{x})$ initiated from each source and each receiver, respectively, in the velocity model using zero traveltimes at the source and receiver positions as Dirichlet boundary conditions in the eikonal equation

$$H(\mathbf{x}, \nabla t_s(\mathbf{x})) = |\nabla t_s(\mathbf{x})|^2 - \frac{1}{v^2(\mathbf{x})} = 0, \quad (5)$$

$$t_s(\mathbf{x}_s) = 0 \quad (6)$$

and

$$H(\mathbf{x}, \nabla t_r(\mathbf{x})) = |\nabla t_r(\mathbf{x})|^2 - \frac{1}{v^2(\mathbf{x})} = 0, \quad (7)$$

$$t_r(\mathbf{x}_r) = 0 \quad (8)$$

where the velocity model $v(\mathbf{x})$ is parametrized on the Cartesian grid that is used for building the discrete field solution.

The two-way traveltimes $T_{s,r,n_s,r}$ for the scatterer n_s, r are given by

$$T_{s,r,n_s,r} = Q_{n_s,r} \mathbf{t}_s + Q_{n_s,r} \mathbf{t}_r, \quad (9)$$

where $Q_{n_s,r}$ is a sampling operator, which extracts the traveltime at the scatterers position, $\mathbf{x}_{n_s,r}$. In this study, we implement this sampling operator with a windowed sinc function (Hicks 2002) although other options can be considered.

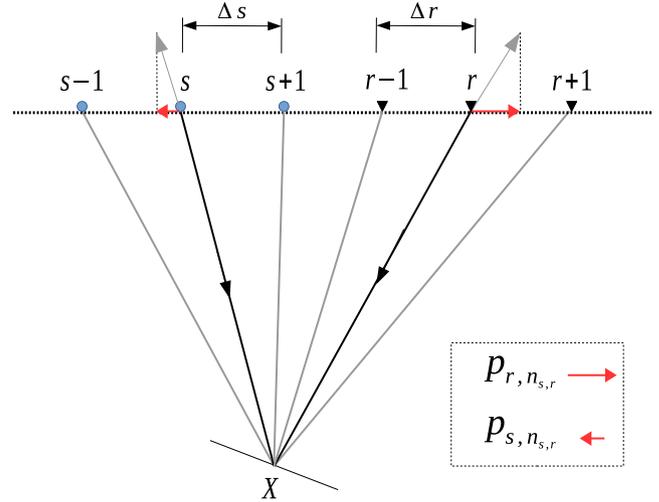


Figure 4. Horizontal component of slowness vectors at the source/receiver positions, $(p_{s,n_s,r}, p_{r,n_s,r})$, are inferred from the traveltime fields emitted from neighbour sources/receivers. Δs and Δr are the source and receivers interval, respectively.

We compute the traveltime maps with a factored eikonal solver based on FSM (Fomel *et al.* 2009). Although the local finite-difference stencil considers only outward directions when computing traveltimes for external nodes of the grid, the sweeping scheme guarantees the coverage of all the wave propagation directions. Moreover, the factorization technique allows one to remove the singularity at the point source position. The extension of this method to TTI anisotropic media as proposed by Waheed *et al.* (2015) and Tavakoli F. *et al.* (2015) may provide the necessary forward-modelling engine for adjoint slope tomography in anisotropic media.

As previously mentioned, two slopes that are picked in a shot and receiver gathers and tied to a scatterer in the subsurface correspond to the horizontal component of the slowness vectors at the receiver and source positions (Fig. 2). Exploiting the reciprocity of the wave propagation between source and scatterer and between receiver and scatterer, and assuming a dense sampling of shots and receivers, we estimate in a finite-difference sense these two slopes from the traveltime maps generated by the left and right neighbour receivers/sources (Fig. 4). This leads to the following expression of the slopes

$$p_{s,n_s,r} = \frac{\partial t_s(\mathbf{x}_s)}{\partial x_s} \approx (Q_{n_s,r} \mathbf{t}_{s+1} - Q_{n_s,r} \mathbf{t}_{s-1}) / 2\Delta s, \quad (10)$$

$$p_{r,n_s,r} = \frac{\partial t_r(\mathbf{x}_r)}{\partial x_r} \approx (Q_{n_s,r} \mathbf{t}_{r+1} - Q_{n_s,r} \mathbf{t}_{r-1}) / 2\Delta r, \quad (11)$$

where Δs and Δr denote the source and receiver intervals, respectively.

With this reciprocity-based strategy, slopes at the source and receiver positions are inferred from traveltimes sampled at the scatterer positions, namely far away from the sources or the receivers where traveltimes have singular values. Therefore, we expect these traveltimes to be accurate enough for a reliable finite-difference estimation of the slopes at the sources or at the receivers, thanks to the smoothness of the velocity model. Alternatively, slopes can be estimated through a partial differential equation by solving the angle equation $\nabla\theta \cdot \nabla\mathbf{t} = \mathbf{0}$ in isotropic media. This means that the gradient of the take-off angle θ is constant along the ray (Noble *et al.* 2012; Belayouni 2013) which can be recast into the more general expression $\partial H/\partial \mathbf{p} \cdot d\mathbf{p}$ for anisotropic media with point source condition. We prefer the semi-discretized formulation because, in this case, synthetic fields to be found are only the two traveltime maps. We are confident that the finite-difference estimation of slopes is accurate enough in smooth media we are looking for, thanks to the use of the reciprocity.

Eqs (5)–(11) define state equations that will be used to estimate the gradient of the adjoint slope tomography misfit function. With our formulation, the number of forward problems scales to the number of sources and receivers.

3.3 Multiparameter inverse problem

From the definition of the data space (eq. 3), the misfit function of the adjoint slope tomography is given by

$$\begin{aligned} C(\mathbf{m}) = & \frac{1}{2\sigma_{T_{s,r}}^2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{n_{s,r}=1}^{N_n^{s,r}} (T_{s,r,n_{s,r}}(\mathbf{m}) - T_{s,r,n_{s,r}}^*)^2 \\ & + \frac{1}{2\sigma_{p_s}^2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{n_{s,r}=1}^{N_n^{s,r}} (p_{s,n_{s,r}}(\mathbf{m}) - p_{s,n_{s,r}}^*)^2 \\ & + \frac{1}{2\sigma_{p_r}^2} \sum_{r=1}^{N_r} \sum_{s=1}^{N_s} \sum_{n_{s,r}=1}^{N_n^{s,r}} (p_{r,n_{s,r}}(\mathbf{m}) - p_{r,n_{s,r}}^*)^2, \end{aligned} \quad (12)$$

where the symbol * denotes the picked data and quantities $\sigma_{T_{s,r}}^2$, $\sigma_{p_s}^2$, and $\sigma_{p_r}^2$, are elements of a diagonal covariance matrix containing the variance of each data class. The inverse of this matrix plays the role of a weighting operator which balances the relative contribution of each residual class during the inversion (Tarantola 1987).

Minimizing the cost function $C(\mathbf{m})$ (eq. 12) is a nonlinear problem which can be tackled either with global or local optimization techniques. Although global optimization can be used for 2-D tomographic problems (Datta & Sen 2016; Sajeve *et al.* 2016), we focus this study on Newton-based local optimization scheme that updates iteratively the model parameters \mathbf{m} as

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \left(\frac{\partial^2 C}{\partial \mathbf{m}^2}(\mathbf{m}_k) \right)^{-1} \frac{\partial C}{\partial \mathbf{m}}(\mathbf{m}_k). \quad (13)$$

Here the positive real step length α_k is estimated by line search while satisfying the Wolfe conditions (Nocedal & Wright 2006). We use the l-BFGS method (Byrd *et al.* 1995) implemented in the SEISCOPE optimization toolbox (Métivier & Brossier 2016) to account for the Hessian operator in the inversion. Considering the contribution of the Hessian operator is crucial to manage the potential leakage between the two different parameter classes (velocity and positions of scatterers). Because this is often not enough, a common practice makes the different parameter classes

dimensionless in multiparameter inversion to better balance the relative contribution of each parameter class. This will improve the condition number of the inversion. In our case, with the units for velocity (m s^{-1}) and for scattered positions (m), we have not found the need to do so.

3.4 Gradient computation with the adjoint-state method

We compute the gradient of the misfit function (eq. 12) with the adjoint-state method implemented through the Lagrangian formalism (Chavent 1974; Plessix 2006).

We build the Lagrangian misfit function by augmenting the misfit function C with equality constraints requiring that the state variables are realizations of the state equations (eqs 5–11):

$$\begin{aligned} \mathcal{L}(\mathbf{m}, \mathbf{u}, \bar{\mathbf{u}}) &= \mathcal{H}(\mathbf{u}, \mathbf{m}) \\ &- \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{n_{s,r}=1}^{N_n^{s,r}} \mu_{s,r,n_{s,r}} (T_{s,r,n_{s,r}} - Q_{n_{s,r}} \mathbf{t}_s - Q_{n_{s,r}} \mathbf{t}_r) \\ &- \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \sum_{n_{s,r}=1}^{N_n^{s,r}} \xi_{s,n_{s,r}} (p_{s,n_{s,r}} - (Q_{n_{s,r}} \mathbf{t}_{s+1} - Q_{n_{s,r}} \mathbf{t}_{s-1})/2\Delta s) \\ &- \sum_{r=1}^{N_r} \sum_{s=1}^{N_s} \sum_{n_{s,r}=1}^{N_n^{s,r}} \xi_{r,n_{s,r}} (p_{r,n_{s,r}} - (Q_{n_{s,r}} \mathbf{t}_{r+1} - Q_{n_{s,r}} \mathbf{t}_{r-1})/2\Delta r) \\ &- \frac{1}{2} \sum_{s=1}^{N_s} \left\langle \lambda_s(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_s(\mathbf{x})) \right\rangle_{\Omega} - \frac{1}{2} \sum_{r=1}^{N_r} \left\langle \lambda_r(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_r(\mathbf{x})) \right\rangle_{\Omega} \\ &- \sum_{s=1}^{N_s} \psi_s(\mathbf{x}_s) - \sum_{r=1}^{N_r} \psi_r(\mathbf{x}_r), \end{aligned} \quad (14)$$

where scalar products, $\langle \cdot, \cdot \rangle$, are defined on the targeted subsurface domain Ω . The arguments of the Lagrangian $\mathbf{u} = (T_{s,r,n_{s,r}}, p_{s,n_{s,r}}, p_{r,n_{s,r}}, \mathbf{t}_s, \mathbf{t}_r, t_s(\mathbf{x}_s), t_r(\mathbf{x}_r))$ and $\bar{\mathbf{u}} = (\mu_{s,r,n_{s,r}}, \xi_{s,n_{s,r}}, \xi_{r,n_{s,r}}, \lambda_s, \lambda_r, \psi_s, \psi_r)$ gather the state and adjoint-state variables, respectively. Since the state variables, the adjoint-state variables and the model parameters are processed as independent quantities in the eq. (14), we rewrite $C(\mathbf{m})$ as $\mathcal{H}(\mathbf{u}, \mathbf{m})$. Finally, we substitute in the state equation satisfied by $p_{s,n_{s,r}}$ and $p_{r,n_{s,r}}$ the formal expression of the slopes, namely $\partial t_s/\partial x_s$ and $\partial t_r/\partial x_r$, by their finite-difference approximation as provided in eqs (10) and (11), eliminating the need of considering slope maps.

The adjoint-state equations are obtained by zeroing the partial derivative of the Lagrangian with respect to the state variables. It is straightforward to derive the expression of the adjoint-state variables $\mu_{s,r,n_{s,r}}$, $\xi_{s,n_{s,r}}$ and $\xi_{r,n_{s,r}}$ which correspond to the traveltime and slope residuals weighted by the coefficients of the covariance matrices.

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial T_{s,r,n_{s,r}}} = 0 &\rightarrow \mu_{s,r,n_{s,r}} = \frac{1}{\sigma_{T_{s,r}}^2} (T_{s,r,n_{s,r}} - T_{s,r,n_{s,r}}^*) = \frac{\Delta T_{s,r,n_{s,r}}}{\sigma_{T_{s,r}}^2}, \\ \frac{\partial \mathcal{L}}{\partial p_{s,n_{s,r}}} = 0 &\rightarrow \xi_{s,n_{s,r}} = \frac{1}{\sigma_{p_s}^2} (p_{s,n_{s,r}} - p_{s,n_{s,r}}^*) = \frac{\Delta p_{s,n_{s,r}}}{\sigma_{p_s}^2}, \\ \frac{\partial \mathcal{L}}{\partial p_{r,n_{s,r}}} = 0 &\rightarrow \xi_{r,n_{s,r}} = \frac{1}{\sigma_{p_r}^2} (p_{r,n_{s,r}} - p_{r,n_{s,r}}^*) = \frac{\Delta p_{r,n_{s,r}}}{\sigma_{p_r}^2}. \end{aligned}$$

Zeroing the partial derivative of the Lagrangian with respect to $t_s(\mathbf{x}_s)$ and $t_r(\mathbf{x}_r)$ indicates the adjoint-state variables ψ_s and ψ_r are zero, therefore, the zero traveltime boundary conditions

on the sources and receivers do not insert any information to our formalism.

To develop the adjoint-state equations satisfied by $\lambda_s(\mathbf{x})$ (namely, $\partial\mathcal{L}/\partial\mathbf{t}_s = 0$), we have to identify all of the terms containing \mathbf{t}_s in the summation over sources of the state equations satisfied by $p_{s,n_s,r}$ (third line in eq. (14)). We find two contributions coming from the neighbour source $s+1$ and $s-1$. We finally obtain for $\partial\mathcal{L}/\partial\mathbf{t}_s$

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial\mathbf{t}_s} = & \sum_{r=1}^{N_s^t} \sum_{n_s,r=1}^{N_n^{s,r}} \mathcal{Q}_{n_s,r}^t \mu_{s,r,n_s,r} \\ & + \frac{1}{2\Delta s} \sum_{r=1}^{N_r^{s-1}} \sum_{n_{s-1},r=1}^{N_n^{s-1,r}} \mathcal{Q}_{n_{s-1},r}^t \xi_{s-1,n_{s-1},r} \\ & - \frac{1}{2\Delta s} \sum_{r=1}^{N_r^{s+1}} \sum_{n_{s+1},r=1}^{N_n^{s+1,r}} \mathcal{Q}_{n_{s+1},r}^t \xi_{s+1,n_{s+1},r} \\ & + \sum_{s=1}^{N_s} \left(\left(\nabla \cdot (\lambda_s(\mathbf{x}) \nabla \mathbf{t}_s(\mathbf{x})) \right)_{\Omega} - \left(\lambda_s(\mathbf{x}) \nabla \mathbf{t}_s(\mathbf{x}) \cdot \mathbf{n} \right)_{\Gamma} \right), \end{aligned} \quad (15)$$

where the last line has been obtained by integration by parts (Taillandier *et al.* 2009). Here, Γ denotes the boundaries of Ω and \mathbf{n} denotes the outward unit normal vectors to them. Without loss of generality, we shall impose Dirichlet boundary conditions for the adjoint variables, so that $\lambda_s(\mathbf{x}) \nabla \mathbf{t}_s(\mathbf{x}) \cdot \mathbf{n} = 0$ over Γ : this will be a non-restrictive hypothesis as long as all sources and receivers are inside the numerical domain. In a similar way we calculate $\partial\mathcal{L}/\partial\mathbf{t}_r$.

Zeroing eq. (15) gives the adjoint-state equation satisfied by $\lambda_s(\mathbf{x})$

$$\begin{aligned} \left(\nabla \cdot (\lambda_s(\mathbf{x}) \nabla \mathbf{t}_s(\mathbf{x})) \right)_{\Omega} = & - \sum_{r=1}^{N_s^t} \sum_{n_s,r=1}^{N_n^{s,r}} \mathcal{Q}_{n_s,r}^t \mu_{s,r,n_s,r} \\ & + \frac{1}{2\Delta s} \sum_{r=1}^{N_r^{s+1}} \sum_{n_{s+1},r=1}^{N_n^{s+1,r}} \mathcal{Q}_{n_{s+1},r}^t \xi_{s+1,n_{s+1},r} \\ & - \frac{1}{2\Delta s} \sum_{r=1}^{N_r^{s-1}} \sum_{n_{s-1},r=1}^{N_n^{s-1,r}} \mathcal{Q}_{n_{s-1},r}^t \xi_{s-1,n_{s-1},r}. \end{aligned} \quad (16)$$

In a similar way, the adjoint-state variables $\lambda_r(\mathbf{x})$ satisfy

$$\begin{aligned} \left(\nabla \cdot (\lambda_r(\mathbf{x}) \nabla \mathbf{t}_r(\mathbf{x})) \right)_{\Omega} = & - \sum_{s=1}^{N_s^t} \sum_{n_s,r=1}^{N_n^{s,r}} \mathcal{Q}_{n_s,r}^t \mu_{s,r,n_s,r} \\ & + \frac{1}{2\Delta r} \sum_{s=1}^{N_s^{r+1}} \sum_{n_{s,r+1}=1}^{N_n^{s,r+1}} \mathcal{Q}_{n_{s,r+1}}^t \xi_{r+1,n_{s,r+1}} \\ & - \frac{1}{2\Delta r} \sum_{s=1}^{N_s^{r-1}} \sum_{n_{s,r-1}=1}^{N_n^{s,r-1}} \mathcal{Q}_{n_{s,r-1}}^t \xi_{r-1,n_{s,r-1}}. \end{aligned} \quad (17)$$

We use the same approach as Taillandier *et al.* (2009) based on the FSM method (Zhao 2005) to solve the linear eqs (16) and (17) for $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$. However, we implement the traveltime and slope residuals as source terms at the scatterer positions in the right-hand sides of the adjoint equations rather than as a boundary condition in Taillandier *et al.* (2009, their eq. 12). This results because we assume that the shots and receivers are inside the domain Ω . These adjoint sources back-propagate the traveltime and slopes

residuals along two ray tubes connecting the scatterers to the shots and receivers. Traveltime residuals $\Delta T_{s,r,n_s,r}$ are back-propagated from the scatterer position $n_{s,r}$ towards the shot s and receiver r , as it would be in reflection traveltime tomography. Together with these traveltime residuals, the slope residuals, that are tied to the neighbour sources $s+1, s-1$ (eq. 16), and receivers $r+1, r-1$ (eq. 17), are back-propagated from the scatterers $n_{s-1,r}, n_{s+1,r}, n_{s,r-1}$ and $n_{s,r+1}$ towards the shot s and receiver r . Therefore, the spatial support spanned by these neighbour scatterers constrain the dip angle at the scatterer $n_{s,r}$ and, hence can be viewed as a discrete approximation of the two take-off angles that are conventionally estimated in ray-based stereotomography. In the example section, we illustrate the shape of the adjoint-state variables $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ and the role of residuals in scatterers repositioning. The contribution of all of the scatterers associated with one source or receiver and its neighbours are gathered in the right-hand side by summation. Therefore, the number of adjoint-state problems to be solved is equal to the number of sources and receivers and hence is independent of the number of scatterers, an important property which makes the computational cost independent of the number of picked events.

The gradient of the misfit function (12) with respect to the model parameters is obtained by taking the partial derivative of the Lagrangian (14) with respect to the model parameters.

The gradient of the misfit function with respect to the velocities \mathbf{v} on the Cartesian grid is the sum of the adjoint-state variables $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ weighted by the velocity raised to the power 3.

$$\nabla_{\mathbf{v}(\mathbf{x})} C = - \sum_{s=1}^{N_s} \frac{\lambda_s(\mathbf{x})}{v(\mathbf{x})^3} - \sum_{r=1}^{N_r} \frac{\lambda_r(\mathbf{x})}{v(\mathbf{x})^3}. \quad (18)$$

The gradient with respect to the cubic B-spline coefficients $\mathbf{c} = \mathbf{c}_m |_{m=1}^M$ can be inferred from the eq. (18) by applying the chain rule of derivatives

$$\nabla_{\mathbf{c}} C = \mathbf{B}^t \nabla_{\mathbf{v}} C, \quad (19)$$

where \mathbf{B} stands for the cubic cardinal B-spline operator ($\mathbf{v} = \mathbf{B}\mathbf{c}$) and t denotes the adjoint operator.

The gradient with respect to the scatterer coordinates is given by

$$\begin{aligned} \nabla_{\mathbf{x}_{n_s,r}} C = & \mu_{s,r,n_s,r} \left(\frac{\partial \mathcal{Q}_{n_s,r}}{\partial \mathbf{x}_{n_s,r}} \mathbf{t}_s + \frac{\partial \mathcal{Q}_{n_s,r}}{\partial \mathbf{x}_{n_s,r}} \mathbf{t}_r \right) \\ & + \frac{\xi_{s,n_s,r}}{2\Delta s} \left(\frac{\partial \mathcal{Q}_{n_s,r}}{\partial \mathbf{x}_{n_s,r}} \mathbf{t}_{s+1} - \frac{\partial \mathcal{Q}_{n_s,r}}{\partial \mathbf{x}_{n_s,r}} \mathbf{t}_{s-1} \right) \\ & + \frac{\xi_{r,n_s,r}}{2\Delta r} \left(\frac{\partial \mathcal{Q}_{n_s,r}}{\partial \mathbf{x}_{n_s,r}} \mathbf{t}_{r+1} - \frac{\partial \mathcal{Q}_{n_s,r}}{\partial \mathbf{x}_{n_s,r}} \mathbf{t}_{r-1} \right), \end{aligned} \quad (20)$$

where the terms $\partial \mathcal{Q}_{n_s,r} / \partial \mathbf{x}_{n_s,r}$ can be unambiguously obtained through the derivative of the windowed sinc function (Hicks 2002). The coordinates of the scatterers are updated from the traveltime maps computed from the sources and receivers to which they are related as well as from those computed from the neighbouring sources and receivers. Again, this highlights the additional constraints provided by slopes to locate the scatterers. The different steps required to compute the gradient are outlined in Algorithm 1.

3.5 Implementation of adjoint slope tomography in practice

We subdivide the adjoint slope tomography into three main steps following the strategy proposed by Billette *et al.* (2003). Three steps, (1) initialization of scatterer positions, (2) preliminary localization

Algorithm 1 Gradient algorithm.

```

1:  $\mathbf{v} = \mathbf{B} \mathbf{c}$ 
2: for  $s / r = 1$  to  $N_s / N_r$  do
3:   Compute  $\mathbf{t}_s / \mathbf{t}_r$ , eqs (5)–(8)
4: end for
5: for  $s / r = 1$  to  $N_s / N_r$  do
6:   for  $r / s = 1$  to  $N_r^s / N_s^r$  do
7:     for  $n_{s,r} = 1$  to  $N_n^{s,r}$  do
8:       Compute  $T_{s,r,n_{s,r}}, p_{s,n_{s,r}} / p_{r,n_{s,r}}$ , eqs (9)–(11)
9:     end for
10:  end for
11: end for
12: Compute  $C$ , eq. (12)
13: for  $s / r = 1$  to  $N_s / N_r$  do
14:   Compute  $\lambda_s / \lambda_r$ , eqs (16) and (17)
15: end for
16: Compute  $\nabla_{\mathbf{v}} C$ , eq. (18)
17: Compute  $\nabla_{\mathbf{c}} C = \mathbf{B}' \nabla_{\mathbf{v}} C$ , eq. (19)
18: Compute  $\nabla_{\mathbf{x}} C$ , eq. (20)

```

of the scatterers and (3) multiscale inversion for joint velocity and scatterer updates, aim to mitigate the nonlinearity and ill-posedness of the inversion (Algorithm 2).

Initialization of scatterer positions. An initial guess of each scatterer position can be computed analytically from the observed traveltimes and slopes by assuming straight rays (Billette *et al.* 2003, their appendix A). At this stage, no initial velocity model needs to be defined: each picked data related to a scatterer defines an effective homogeneous velocity. As underlined by Billette *et al.* (2003), closed-form solutions have always been found in our numerical examples and, the computational cost is negligible.

Localization of scatterer positions. In order to remove noisy picks and make the initial positions of the scatterers as close as possible to their true positions, we update the scatterer positions keeping a fixed background velocity model. One may use either a homogeneous velocity model or a constant gradient velocity model to compute the traveltimes and slopes analytically and hence make this re-localization step computationally efficient. This background velocity model might be chosen to make the distribution of the scatterers as even as possible in the subsurface. For example, a too slow homogeneous model might tend to squeeze the scatterers in the shallow part of the subsurface medium. In contrast, a constant gradient velocity model can help to equally distribute the scatterers in depth. If no *prior* information coming from well logs are available, finding an appropriate constant gradient model by trial and error or grid search is not challenging due to the computational efficiency of the localization process. Since the velocity model is not updated during the localization step, the resolution of the eikonal equation needs to be performed only during the first iteration, where we store the traveltimes maps.

Multi-scaling joint inversion. The last step consists of simultaneously updating all of the model parameters through a multiscale approach. Multi-scale imaging is performed through a progressive refinement of the spline parametrization, that is implemented taking advantage of the subdivision property of the B-spline surfaces (see de Boor (1978) or Virieux & Farra (1991, their appendix)). We complement the regularization performed by the coarse B-spline parametrization by an additional Gaussian smoothing of the gradient computed on the Cartesian grid, similar to the one used by Taillandier *et al.* (2009) for first-arrival traveltimes tomography. This

smoothing acts as a soft constraint to reduce the null space, as opposed to the sampling strategy of the model space which is a hard constraint.

With this smoothing regularization, the gradient of the misfit function with respect to the B-spline coefficients can be written as

$$\nabla_{\mathbf{c}} C = \mathbf{B}' \mathbf{G} \nabla_{\mathbf{v}} C, \quad (21)$$

where \mathbf{G} is the Gaussian smoothing operator and the 3-D operator $\mathbf{B}' \mathbf{G}$ can be written as the tensorial product of three 1-D operators (Operto *et al.* 2003, appendix B). The main aim of the Gaussian smoothing is to filter out the footprint of the sources, receivers and scatterers in the velocity gradient computed in the Cartesian grid.

Algorithm 2 Adjoint slope tomography workflow. N_{ms} : number of multiscale step; \mathcal{S} : B-spline subdivision operator. For sake of clarity, the model parameters \mathbf{m} , the B-spline velocity coefficients \mathbf{c} , the misfit function C and its gradient ∇C are indexed by the scale step i and the iteration number k .

```

1: Initialization of scatterer positions (Billette et al. 2013, their appendix A)
2: Set initial B-spline velocity model  $\mathbf{c}_0$ 
3: Preliminary re-localization of scatterers in  $\mathbf{c}_0$ 
4: for  $i = 1$  to  $N_{ms}$  do
5:    $\mathbf{c}_{0,i} = \mathcal{S} \mathbf{c}_{N_{it},i-1}$ 
6:   for  $k = 1$  to  $N_{it}$  do
7:     call LFBFGS( $\mathbf{m}_{k-1,i}, C_{k-1,i}, \nabla C_{k-1,i}$ )
8:      $\mathbf{m}_{k,i} = \mathbf{m}_{k-1,i} + \Delta \mathbf{m}_{k-1,i}$ 
9:   end for
10: end for
11:  $\mathbf{v}_{final} = \mathbf{B}_{N_{ms}} \mathbf{c}_{N_{ms},N_{it}}$ 

```

4 SYNTHETIC EXAMPLES

In this section, we assess the adjoint slope tomography with synthetic examples of increasing complexity. In the first and second examples, we design an ideal experimental set-up with a dense distribution of scatterers and a wide-aperture acquisition geometry to prevent velocity-versus-depth ambiguity (Bube *et al.* 2005) during the joint update of velocities and scatterers. The third example aims to reconstruct a smooth version of the Marmousi model from realistic synthetic picks that would be generated from a towed-streamer acquisition. We assess the accuracy of the stereotomography velocity model by using it as an initial model for frequency-domain FWI. For all of the following tests, we compute the observed data set, that is, $\{T_{s,r,n_{s,r}}, p_{s,n_{s,r}}, p_{r,n_{s,r}}\}$, in the true velocity model and for the true scatterer positions with our forward-modelling engine. Assessment of adjoint slope tomography in more realistic settings, where two-way traveltimes and slopes are picked in the common shot/receiver gathers or in the common-image gathers, is left to future studies.

4.1 Example 1: circular anomaly

In this example, the true subsurface model consists of a constant-gradient velocity background model ($v(x, z) = v_0 + a \times z$) inside a model of 20 km \times 5 km, that contains a smooth circular velocity anomaly of radius 750 m (Fig. 5). We use $v_0 = 1000 \text{ m s}^{-1}$ and $a = 0.9 \text{ s}^{-1}$ in the background model, while the velocity reaches 3820 m s^{-1} in the centre of the inclusion. A line of sources, spaced 200 m apart, is set at 500 m depth. Each shot is recorded by five receivers at 500 m depth with offsets of 0.8, 1.6, 2.4, 3.2 and 4 km.

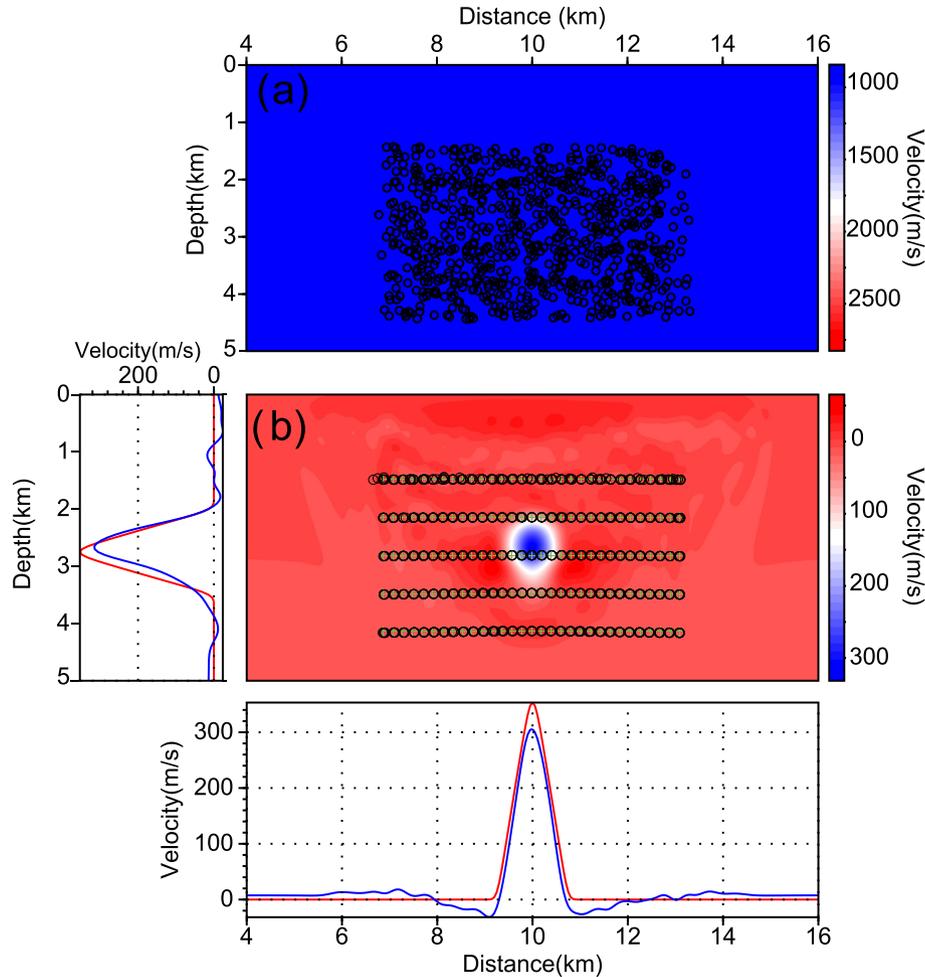


Figure 5. Circular inclusion example. (a) Initial position of scatterers, randomly distributed. (b) Final adjoint slope tomography velocity perturbation model (inverted velocity model minus the true constant gradient background velocity) with superimposed exact (cross) and calculated (circle) scatterers positions. Diagrams show the direct comparison between calculated (blue) and exact (red) velocity perturbations across horizontal and vertical profiles cross-cutting the centre of the anomaly.

The scatterer layout is composed of 155 scatterers with a spacing of 200 and 700 m in the horizontal and vertical directions, respectively. Each scatterer is located midway between a source and receiver positions. This setting provides a reasonable angular illumination of the target. The same Cartesian discretization of the velocity model is used for the forward and inverse problems using a grid interval of 50 m. We use a low-pass Gaussian filtering of the misfit function gradient for regularization. The correlation length for this filtering is 200 m. Setting $(\sigma_{T_{s,r}}, \sigma_{p_s}, \sigma_{p_r}) = (1 \text{ ms}, 0.01 \text{ ms m}^{-1}, 0.01 \text{ ms m}^{-1})$ balances the three terms in the misfit function (eq. 12).

The initial velocity model is homogeneous with a wave-speed of 1000 m s^{-1} . The initial estimates for positions of the scatterers are randomly distributed around their exact positions with a maximum deviation of 400 m (Fig. 5a). Since the main goal of this test is to assess the capability of the method in finding the circular anomaly and the scatterer positions, we design the optimization process as follows. First, we find value 0.4 for a that allows for the best data fit using the initial positions of the scatterers. Then, starting from the resulting approximated background velocity model, we jointly update the velocities and the scatterer positions. The final velocity model and scatterer positions after 100 iterations suggest no leakage between the two parameter classes, since the scatterers

were moved to their correct positions (Fig. 5b). The decrease of the misfit function over iterations is shown in Fig. 6. The reconstruction of the inclusion shows some vertical smearing of the bottom end of the inclusion that may result from the narrowing of the angular illumination with depth (Fig. 5). In contrast to the reconstruction in the vertical direction affected by smearing, some low-velocity perturbations on both sides of the inclusion in the horizontal direction probably reflect a small deficit of low wavenumbers (Fig. 5, horizontal profiles). These limited bandwidth effects can be related to the shape of the adjoint slope tomography sensitivity kernels, which connect the scatterers in depth to the sources and receivers at surface (Fig. 7). The elongated shape of these sensitivity kernels in subvertical directions favours a smooth reconstruction of the vertical wavenumbers as in transmission tomography, while the lateral deviation of the kernels generated by heterogeneities located between the surface and the scatterer may favour the update of intermediate horizontal wavenumbers at the expense of the low components as in reflection tomography.

For illustrative purposes, Fig. 7 shows the kernel of the misfit function gradient with respect to the velocity in a homogeneous background velocity model. Here we consider one 4 km offset source–receiver pair and one scatterer located midway between

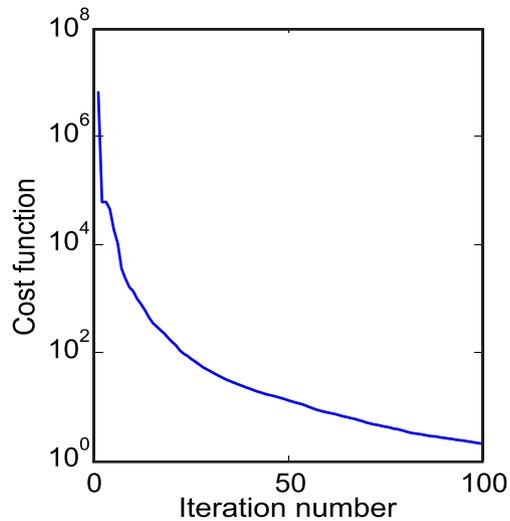


Figure 6. Circular inclusion example. Convergence diagram in logarithmic scale for the test in Fig. 5.

the source and receiver positions at 4.3 km depth. This kernel is formed by the superposition of the adjoint-state variables $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ weighed by the velocity raised to a power three (eq. 18). Since the background model is homogeneous, the two neighbouring sources and the two neighbouring receivers that are required to build the right-hand side terms of the adjoint-state equations (eqs 16 and 17) involve two scatterers located on both sides of the scatterer in question. The central and the two neighbour scatterers define the spatial support of the source of the adjoint-state equation, that back-propagate the time and slope residuals towards the source (left ray tube) and receiver (right ray tube) positions. The spatial support spanned by the two neighbouring scatterers define the so-called dip bar (here, a horizontal bar) (Billette *et al.* 2003), from which two rays are shot with appropriate take-off angles in ray-based stereotomography so that they honour the Snell's law at the scatterer. While these two take-off angles are explicitly part of the optimization parameters in ray-based stereotomography, they are implicitly estimated by adjoint slope tomography through the joint localization of the nearby scatterers, hence defining a model space with a more limited number of parameter classes.

4.2 Example 2: curved layer anomaly

The target of the second example is a layer with a synclinal geometry that is embedded in a 20 km \times 5 km homogeneous velocity background model (Fig. 8a). The aim of this test is to validate the adjoint slope tomography in a classical reflection configuration and

assess the contribution of the slope information to overcome potential velocity–depth cross-talk during the velocity estimation. The observables are generated from two series of scatterers that follow the top and the bottom of the layer and a horizontal line of scatterers located below the layer at a depth of 4200 m. This latter line of scatterers provides the necessary illumination in depth to constrain the bottom reflector of the layer.

The velocity in the homogeneous background model is 3000 m s⁻¹ and the velocity within the smoothed layer anomaly reaches a maximum value of 4000 m s⁻¹. A 2400 m single-offset acquisition is designed with sources and receivers on the surface meaning that each scatterer is reached by only one source–receiver pair. This implies that, without slope information, the inversion may not cope with the velocity–depth ambiguity. In this example, we use the multiscale approach during which the spline interval decreases throughout five steps from 1000 to 60 m and from 2500 to 150 m in the vertical and horizontal directions, respectively. The initial velocity model is homogeneous with an overestimated velocity of $v_0 = 3300$ m s⁻¹. The initial scatterers are located midway between the source and the receiver at a depth of $v_0 \times t/2$ where t denotes the true two-way traveltimes (Fig. 8a). Therefore, the initial dips at the scatterers as predicted by the initial slopes at the source and receiver positions are horizontal and do not embed any prior information on the geometry of the layer. Moreover, these horizontal dips do not match those that would be inferred from the relative positions of the initial scatterers since these latter are not horizontally distributed (Fig. 8a).

The velocity model and the position of scatterers that have been reconstructed after 210 iterations are shown in Fig. 8(b). Here the correlation length of Gaussian filter is 200 m. Overall, the shape of the syncline is well recovered as shown by the good agreement between the true and the reconstructed lines of scatterers aligned along the top and bottom reflectors with some small mismatches towards the ends of the layer (Fig. 8b). The bottom line of the scatterers shows more obvious footprint of velocity–depth ambiguity towards the ends of the line in the form of underestimated depths balanced by two circular blobs of low velocities. Overall, these mismatches remain acceptable owing the limited offset coverage considered in this example. The tomography reconstructs a smooth representation of the layer as shown by the horizontal and vertical profiles in Fig. 8. In the vertical direction, the smoothing of the layer generates overestimated velocities just above the top reflector that are balanced by underestimated background velocities in the first 1.5 km in depth. Below the bottom reflector, the velocities tend to be overestimated to balance the underestimated velocities in the smooth reconstructed layer. Towards the middle of the last scatterer line, a slight downward shift of the reconstructed line also contributes to partially balance these overestimated velocities below the layer.

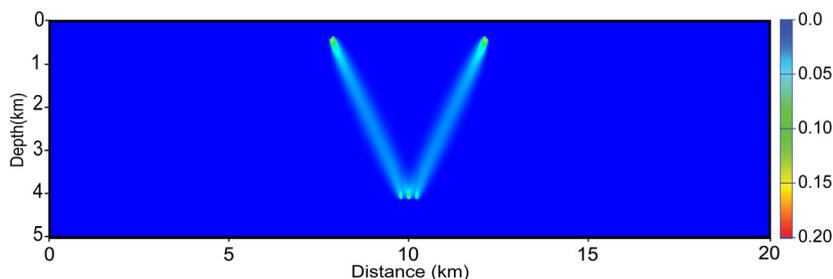


Figure 7. Circular inclusion example. The kernel of misfit function gradient with respect to the velocity in a homogeneous background velocity model for one 4 km offset source–receiver pair. The kernel consists of adjoint-state variables $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ which are weighted by the velocity raised power three. Here the source–receiver includes one scatterer at 4.3 km depth. Migration facets can be constructed by neighbour scatterers.

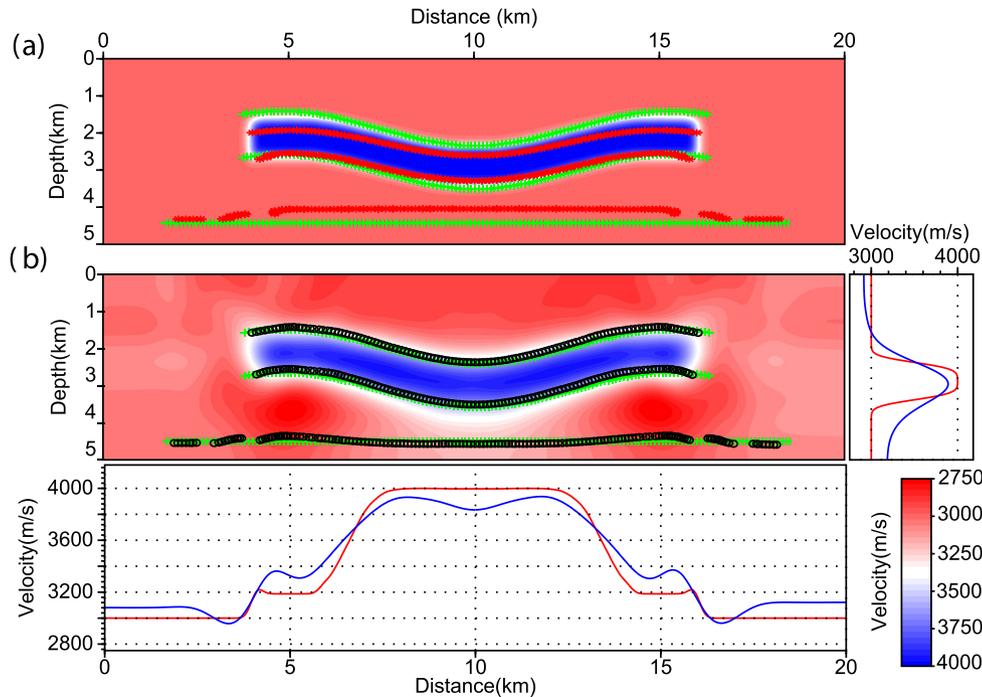


Figure 8. Layer example. (a) Exact velocity model with superimposed exact (green '+') and initial (red '*') scatterer positions. (b) Reconstructed velocity model by adjoint slope tomography with superimposed exact (green '+'), initial (red '*') and calculated (black 'o') scatterer positions. Diagrams show the direct comparison between calculated (blue) and exact velocities (red) across horizontal and vertical profiles cross-cutting the velocity model at 2.5 km depth and 10 km distance.

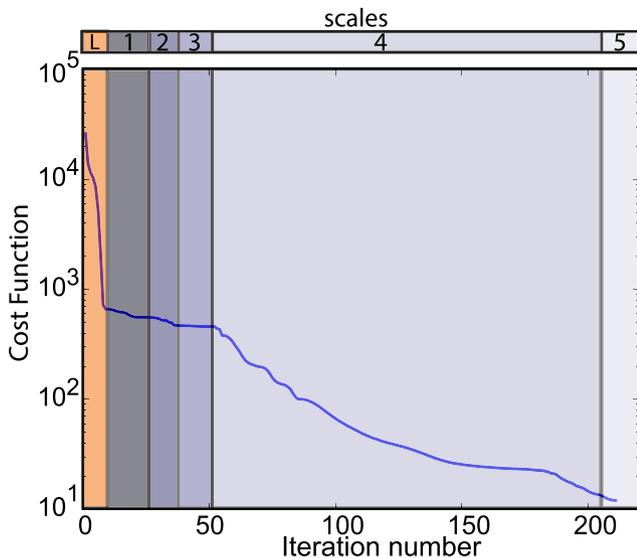


Figure 9. Layer example. Convergence diagram in logarithmic scale for the example in Fig. 8. The colours show the localization step ('L') and scales number in multiscaling approach scheme.

Fig. 9 shows the convergence diagram for multiscaling optimization including different five scales.

In order to assess more precisely the contribution of the source and receiver slopes in the inversion, we perform three inversions that rely only on slope information (Fig. 10d), traveltimes information (Fig. 10c) or both (Fig. 10b). The Fig. 10(a) shows a close-up of the initial model where the layer exhibits some dips as shown by the superimposed true and initial scatter positions. Note that there are twice more scatterers along an interface than shot-receiver pairs. We initialize the scatterer positions midway between shot and

receiver positions, which implies that pairs of scatterers share the almost same position (Fig. 10a, red asterisk symbol), while the true scatterers are uniformly distributed along the reflectors (Fig. 10a, green plus symbol). We design this configuration in order to assess the ability of the slope tomography to move the scatterers both horizontally and vertically.

Traveltime inversion fails to position correctly the scatterers (Fig. 10c). First, the depths of the scatterers on the top interface and bottom interfaces are overestimated and underestimated, respectively. This inaccurate vertical positioning is balanced by erroneous velocity update, manifested by overestimated velocities in the background model and underestimated velocities in the layer, in order to fit traveltimes. This highlights velocity–depth cross-talk during the multiparameter reconstruction. Second, the traveltime inversion fails to move the scatterer horizontally, that highlights the missing slope constraints as suggested by the results of the slope inversion. In other words, in absence of the slopes information, there is not any constraint to locate the scatterer along the isochrone which is defined by observed traveltimes (Chauris *et al.* 2002a).

Fig. 10(d) shows that the slope-only inversion fails to fit traveltimes, that is highlighted by the fact that the depth of the scatterers associated with the upper interface are underestimated and the velocities in the background model are overestimated. Unlike the traveltime inversion, the vertical mispositioning of the scatterers does not balance the erroneous velocities to honour traveltimes (these latter being not involved in the inversion). However, slope inversion manages to horizontally move the scatterers to their true horizontal coordinates, unlike traveltime inversion. This sensitivity to the horizontal positions can be intuitively understood from the gradient of the misfit function with respect to scatterer coordinates, which involves the difference between sampled neighbouring traveltimes maps at the scatterer position, eq. (20).

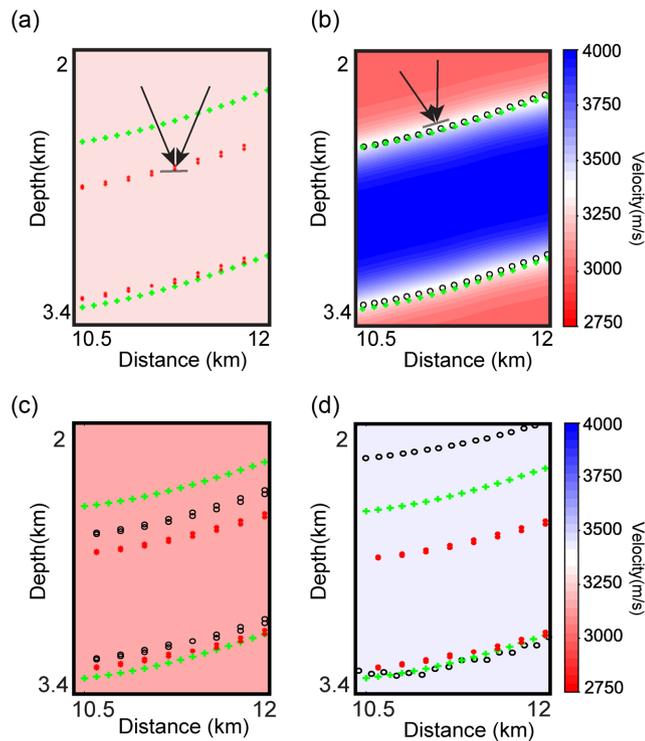


Figure 10. Layer example. Comparison between joint traveltimes + slope, traveltimes only and slope only inversions. (a) Close-up of the initial velocity model with superimposed true (green '+') and initial (red '*') scatterer positions. The initial positions of the scatterers are midway the source and receiver positions meaning that the local dip predicted by the shot and receiver slopes would be horizontal. Note that pairs of scatterers share the same initial position. Close-up of the velocity model after (b) joint traveltimes and slope inversion, (c) traveltimes-only inversion, and (d) slope-only inversion. The retrieved scatterer positions are shown by black circles. See the text for detailed interpretation.

Fig. 10(b) shows how the joint inversion of traveltimes and slopes allows to overcome these artefacts and retrieve both the correct velocities and the horizontal and vertical coordinates of the scatterers.

4.3 Example 3: Marmousi model

4.3.1 Building velocity macromodel

We now consider a more realistic example with the complex Marmousi model (Fig. 11a; Bourgeois *et al.* 1991). Because of its structural complexity, the reconstruction of this velocity model by tomographic methods is challenging. As mentioned by Lambaré (2008) and shown by Billette *et al.* (2003), ray-based stereotomography has not fully succeeded in reconstructing the smooth components of this model, for three possible reasons: nonlinearity of the inverse problem, intrinsic limitations of ray theory, and unreliability of stereotomographic picking. We generate the targeted velocity model of the stereotomographic inversion by Gaussian smoothing of the original Marmousi model with vertical and horizontal correlation lengths of 100 m (Fig. 11b). This smoothing is sufficiently mild to guarantee that the resulting model provides a good background model for pre-stack depth migration (e.g. Thierry *et al.* 1999) or a good initial model for FWI considering a realistic starting frequency of the order of 4 Hz. Note that we use a less aggressive smoothing (100 m correlation length instead of 240 m) than Billette *et al.* (2003) and Chauris *et al.* (2002b) to generate the true

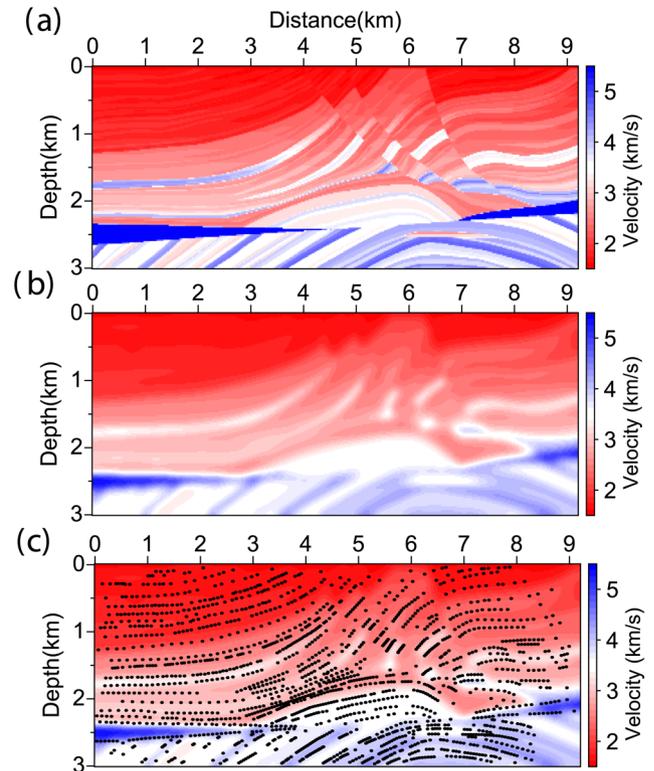


Figure 11. Marmousi example. (a) True velocity model. (b) Smoothed Marmousi velocity model used as the targeted model of the adjoint slope tomography. (c) Scatterers that have been used to generate the traveltimes and slope data set from the velocity model shown in (b).

background velocity model to be reconstructed by stereotomography. In our application this can contribute to make the distribution of scatterers in depth more non-uniform and generate multipathing. The grid interval in the smoothed Marmousi model is 20 m.

To build the data set for inversion, we pick manually the main reflectors in the Marmousi model (Fig. 11c) and compute the corresponding reflection traveltimes and slopes numerically with our forward engine. To generate a realistic data set, we assign a source–receiver pair to each scatterer according to the Fermat’s principle by searching the scatterer that has the minimum reflection traveltime along a reflector. Moreover, we check that the found scatterer corresponds to a specular reflection point; if not, we remove its associated traveltime and slopes from the data set. This condition is fulfilled if the sum of the slowness vectors estimated at the scatterer position from the source and receiver traveltime-map gradients is aligned with the normal to the reflector. We consider a towed-streamer acquisition with offsets ranging between 100 m and 3425 m. Ninety-one shots every 100 m are recorded by 134 receivers spaced 25 m apart. The first shot is located at a distance of 7400 m in Fig. 11 and the acquisition layout is moving from right to left, implying a deficit of scatterers in the bottom-right part of the target. According to this source–receiver geometry, our reflector picking in the Marmousi model (Fig. 11c) results in more than 6000 observables in the data domain. Two shot gathers located at $x = 800$ m and $x = 4800$ m are shown in Fig. 12 with superimposed traveltimes and slopes generated from the reflector picks shown in Fig. 11(c). Both of these gathers include the deepest scatterers; the shot gather at $x = 4800$ m, unlike the other one, includes the reflection from complex geological structures. Our picking of the Marmousi reflectors should be reasonably representative of the one that would

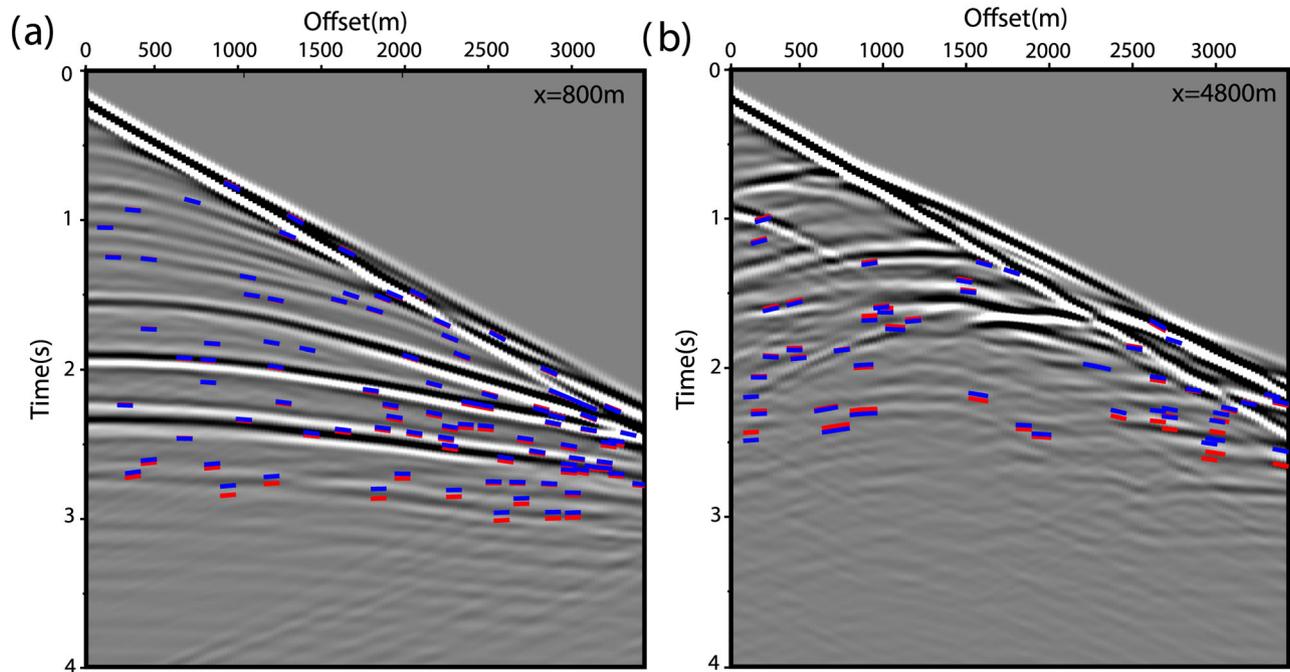


Figure 12. Marmousi example. Examples of traveltime and slope observables (red segments) generated from the velocity model of Fig. 11(b) and the scatterers of Fig. 11(c). The shot positions are $x = 800$ m (a) and $x = 4800$ m (b). Each red segment shows the traveltime and the slope of each picked event. The blue segments represent the traveltimes and slopes computed in the final adjoint slope tomographic model (Fig. 13f).

have been performed in the pre-stack migrated domain (namely, in common-image gathers). Indeed, traveltime and slope observables can be generated from image-domain picks by demigration (Chauris *et al.* 2002a; Guillaume *et al.* 2008), in a manner comparable to our ad-hoc approach except that we took advantage of our knowledge of the true model to measure the dips instead of picking them. Our reflector picking leads to a sparse distribution of locally coherent events in the gathers as illustrated in Fig. 12. Note that some picks in the complex zone have not satisfied the specular-point condition. Therefore, our inversion can suffer from a deficit of illumination in this complex part of the model.

In order to reconstruct the smooth Marmousi velocity model, we follow the stereotomographic workflow introduced in Section 3.5. During the localization step and the following slope tomographic inversion, we use $(\sigma_{T_s,r}, \sigma_{p_s}, \sigma_{p_r}) = (1 \text{ ms}, 0.01 \text{ ms m}^{-1}, 0.01 \text{ ms m}^{-1})$. Also, we estimate an appropriate correlation length of 200 m for Gaussian smoothing by trial and error. After the initialization step (Fig. 13a), we performed a localization of the scatterers from a homogeneous velocity model with a 2000 m s^{-1} velocity (Fig. 13b). Then, we removed outliers associated with scatterers that fall outside the limits of the subsurface target. After localizing the scatterers, we performed four successive slope tomographic inversions by refining the velocity grid by a factor 2 in both directions at each step. During these four multiresolution steps, the horizontal and vertical B-spline node spacings decrease from 1300 to 160 m and from 400 to 50 m, respectively. During these four steps, the inversion has performed 3, 147, 50 and 165 iterations to generate the velocity models and the scatterer positions shown in Figs 13(c)–(f). Overall, the positions of the scatterers tend to align with the reflectors of the Marmousi model as the resolution of the velocity model improves over the four multiscale steps with however, a certain hierarchy driven by the local structural complexity of the velocity model. During the first-scale inversion, the scatterer positions and the background velocities were

not significantly updated because of the B-spline grid intervals are too coarse (Fig. 13c). Large-scale variations of the background velocities are introduced in the whole model during the second step; however, the scatterers start being aligned with the reflectors only in the left part of the model where the structure is simpler (Fig. 13d). During the third step, the scatterer positions are moved to their correct positions in the upper-right part of the model but the inversion fails to significantly update their positions at the reservoir level (Fig. 13e). During the fourth step, the scatterers are aligned along the reflectors with improved accuracy as the resolution of the velocity model is improved with the most significant updates in the complex zone at the reservoir level (Fig. 13f). Through these iterative updates, the most obvious inaccuracies occur at the reservoir depths and near the bottom-right end of the model where the overburden exhibits significant dips and the acquisition geometry provides a limited illumination. The misfit function plotted as a function of the iteration number over the localization and the four multiscale steps shows a regular decrease of the data misfit, suggesting a reasonable tuning of the inversion (Fig. 14). If we would have used a constant-gradient velocity model during the localization instead of a homogeneous model, we could have started the slope tomography inversions on a finer spline grid to converge in a smaller number of iterations towards a final model almost identical to the one shown here. In this test, by considering a homogeneous background velocity model, we intended to assess the capability of the model to image smooth components of complex structures almost from scratch.

As a quality control of the final results, we have superimposed the dip bars onto the original Marmousi velocity model (Fig. 15b). The dip associated to each scatterer is computed a posteriori from the traveltime gradient vectors ∇t_s and ∇t_r , computed in the slope tomography velocity model (Fig. 15a). We also show the direct comparison between the dip bars (black segments) and the true

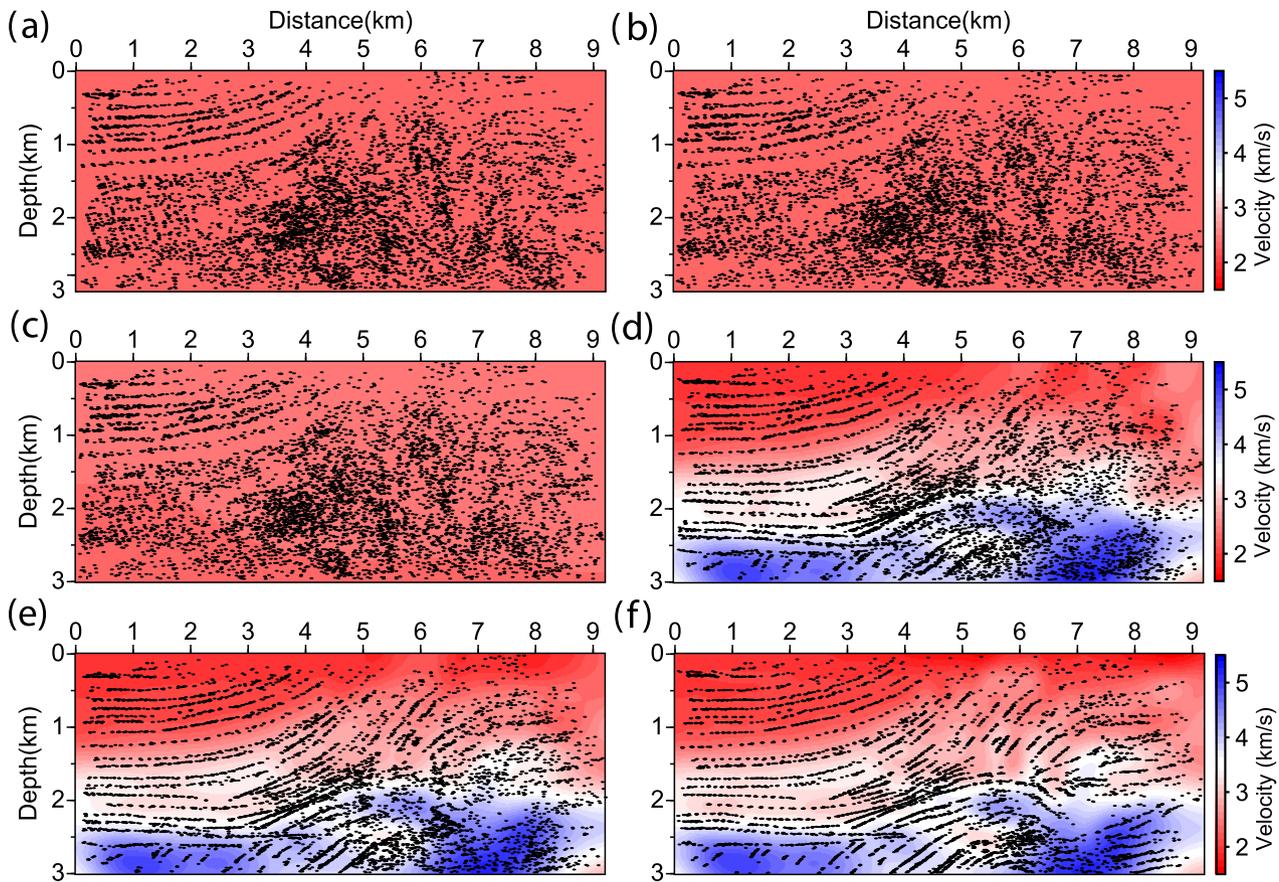


Figure 13. Marmousi example. Adjoint slope tomography results. Velocity model and scatterer positions after (a) the initialization step, (b) the localization step and (c–f) four multiscale adjoint slope tomographic inversions with spline-grid refinement. See the text for details.

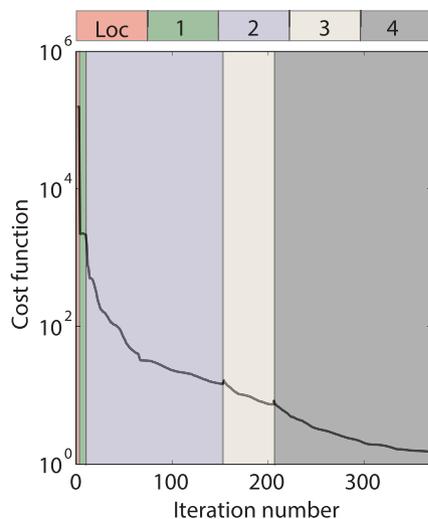


Figure 14. Marmousi example. Misfit function versus iteration number plotted with a logarithmic scale. The colours delineate the localization step ('Loc') and the scales number the multiscale inversion.

position of the scatterers resulting from the picking of the Marmousi reflectors in Fig. 15(c). Figs 15(a) and (c) show a fairly good agreement at all depths in the left part of the Marmousi model where the dips are reasonably mild. In the more complex central and right parts, the alignment remains acceptable in the shallow part, while the adjoint slope tomography failed to provide a reliable reconstruc-

tion at depths greater than 2.5 km beyond 6 km of distance. This mispositioning of the dip bars in the complex deep part manifest as significant late traveltime misfit in the shot gathers of Fig. 12. An insufficient illumination aperture at the reservoir depths, resulting from the strong lateral variations in the overburden and the position of the target near the right end of the acquisition, might explain this failure.

The final estimated velocity model by adjoint slope tomography is complex enough to generate multivalued ray fields. This is illustrated by calculating rays + wave fronts (Lambaré *et al.* 1996) for this velocity model (Fig. 16a). In Fig. 16(a), the rays+wave fronts are superimposed with our eikonal solver solution, first-arrival wave fronts, for a source at $x = 6$ km. The complexity of the slope tomographic velocity model generates caustics which are singular points for the first-arrival wave fronts (Fig. 16b). Our finite difference approach for the slope estimations near these caustic points can be erroneous; another possible reason for mispositioning of the dip bars in the complex zone (Fig. 15c). However, the calculated first-arrival wave fronts with the eikonal solver are completely matched with the solutions of wavefront construction method.

4.3.2 FWI with initial model from adjoint slope tomography

We now assess the quality of the adjoint slope tomographic velocity model (Fig. 17a) as a starting model for frequency-domain FWI of long-offset data. As such, we consider a stationary-receiver acquisition geometry with a 9 km maximum offset. The aim of this

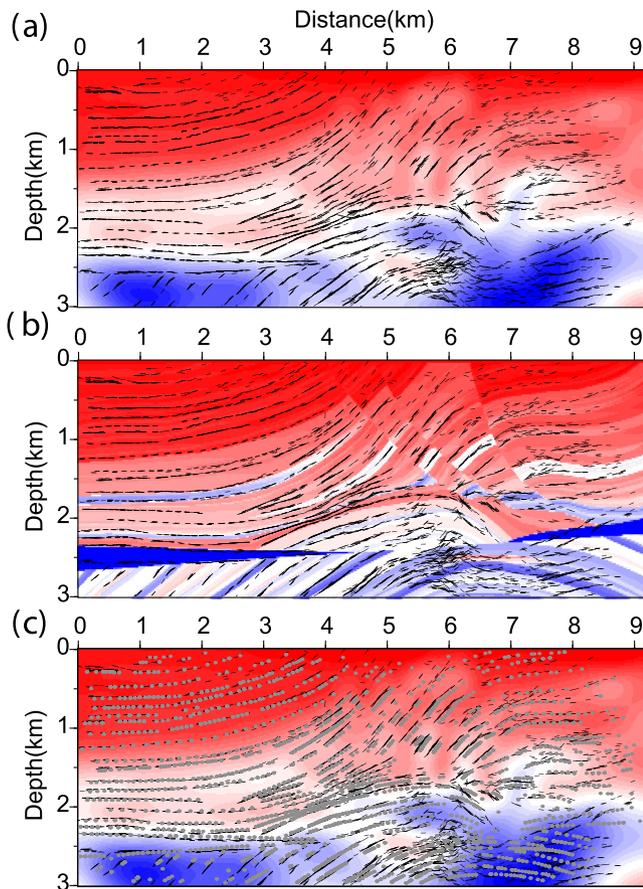


Figure 15. Marmousi example. Scatterer positions with dip bars estimated by adjoint slope tomography superimposed on (a) the final inverted velocity model and (b) the true Marmousi model. (c) Direct comparison between the reconstructed dip bars (black segments) and the scatterers (grey circle) that have been used to generate the data set (Fig. 11c).

increased offset range is twofold. First, it increases the nonlinearity of FWI associated with cycle skipping of both diving waves and wide-angle reflections and hence provides a suitable framework to assess the adjoint slope tomographic model as a good initial model for FWI. Second, this long-offset acquisition provides a sufficient wide scattering-angle illumination to prevent some notches between the wavenumber content of the slope tomographic velocity model and that of the perturbation model generated by FWI when a starting frequency as high than 4 Hz is used. The acquisition consists of 41 sources every 200 m that are recorded by 231 receivers spaced 40 m apart. We invert sequentially five frequency components of the wavefields (4, 5, 7, 9, 12 and 16 Hz) with the l-BFGS algorithm. The final inverted velocity model by FWI (Fig. 17b) shows a good qualitative agreement with the true Marmousi velocity model except near the bottom-right of the model (Fig. 11b). Fig. 18 shows a more quantitative assessment of the accuracy of the FWI model by the direct comparison between a series of vertical profiles extracted from the final adjoint slope tomographic model, the final FWI model and the true Marmousi model. The agreement between the FWI and the true profiles is good down to 2 km depth both in terms of resolution content, positioning in depth and velocity estimation, except near the right-end of the model. The resolution degrades near the bottom of the model as the number of available events becomes more limited.

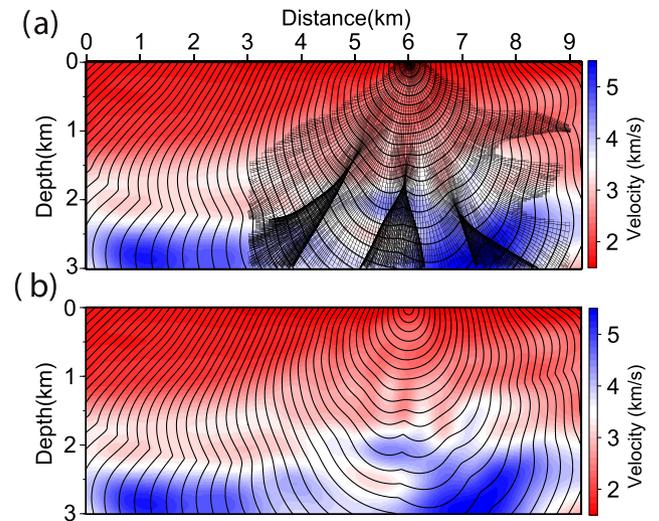


Figure 16. Marmousi example. (a) Smooth Marmousi model built by slope tomography with superimposed ray+wave fronts, computed with the wavefront construction method of Lambaré *et al.* (1996), and wave fronts computed with the factored eikonal solver. (b) Same as (a) without the ray field.

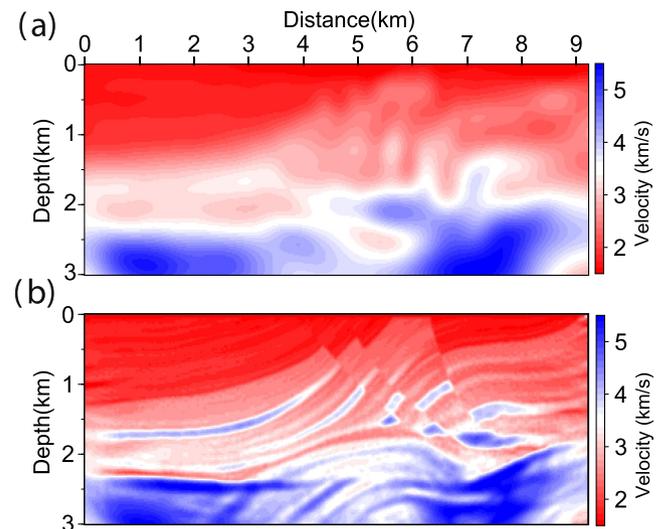


Figure 17. Marmousi example. (a) Final inverted velocity model by adjoint slope tomography. (b) Final FWI velocity model using the adjoint slope tomography model (a) as initial model.

5 DISCUSSION

We have presented a new formulation of stereotomography. The key ingredients are the computation of the traveltimes with an eikonal solver, the inference of the slopes at the source and receiver positions from these traveltimes by finite differences and the computation of the misfit function gradient with the matrix-free adjoint state method. Our formalism relies on a limited number of parameter classes (subsurface velocities and scatterer coordinates) compared to the ray-based stereotomography developed by Billette & Lambaré (1998). This directly results from the use of an eikonal solver instead of a ray-tracing algorithm to compute traveltimes and slopes. The eikonal equation using the shots and receivers as sources and the finite-difference estimation of the slopes from the traveltime solutions of the two eikonal resolutions define the

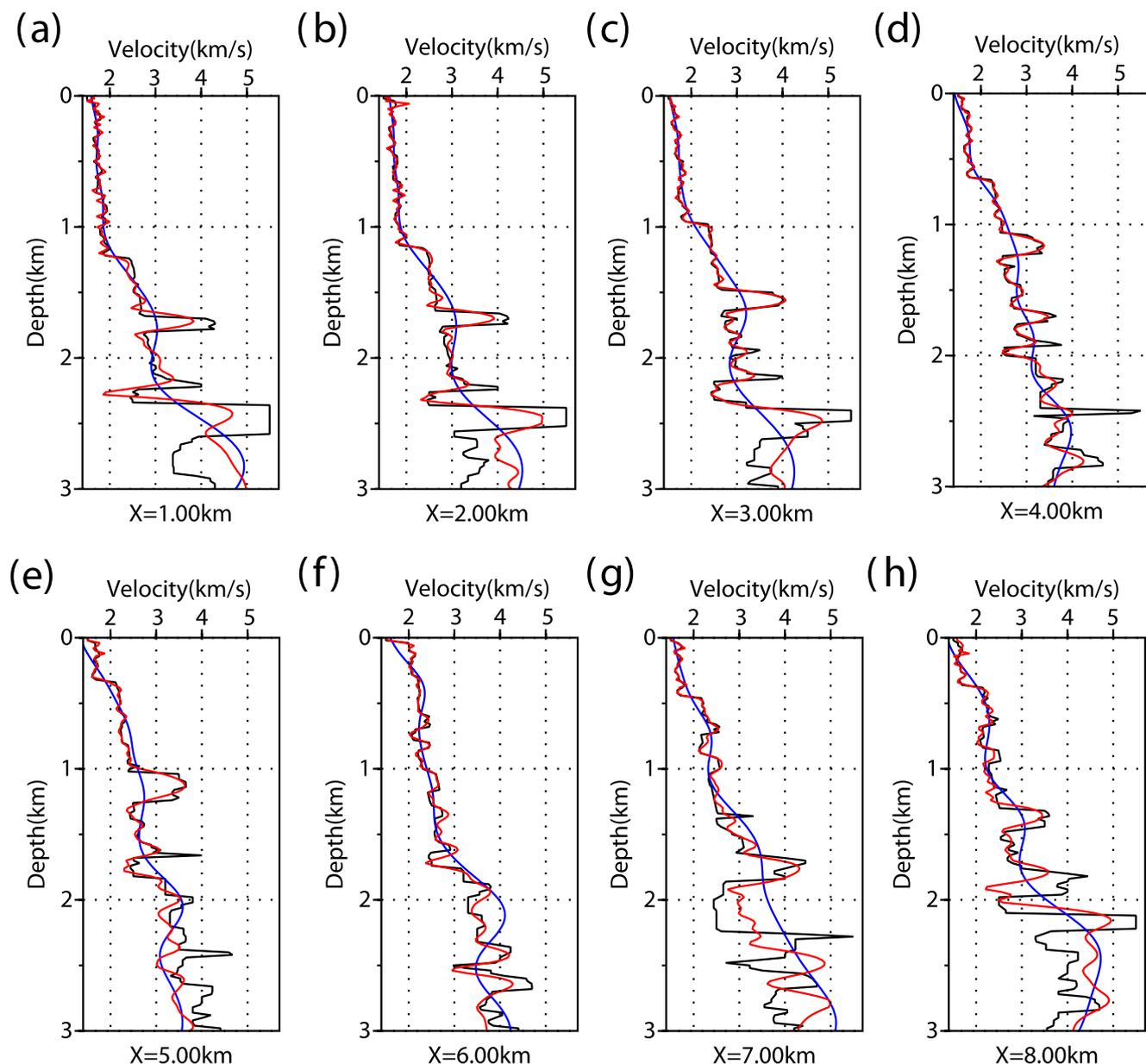


Figure 18. Marmousi example. Direct comparison of logs between exact velocity model in Fig. 11(a), black line, and the estimated velocity by adjoint slope tomography in Fig. 17(a), blue line, and inverted velocity model by frequency domain FWI in Fig. 17(b), red line, while using adjoint slope tomography solution as initial model.

state equation of the adjoint problem. The adjoint-state variables associated with the source-side and receiver-side traveltimes are solution of a transport-like equation, which builds the sensitivity kernels of the velocity update along two ray tubes connecting the scatterer to the sources and receivers. The right-hand side of these adjoint-equations describe a source term whose spatial support, weighted by the two-way traveltime and slope residuals, defines a migration facet or dip bar at the scatterer position. The position of a scatterer is updated from the local gradients of the traveltime maps generated from the sources and receivers to which it is related and also from the gradient of the traveltimes generated by neighbouring sources and receivers. Again, the contribution of neighbour sources and receivers highlight the additional constraints provided by slopes to localize the scatterers. Our parsimonious parametrization may

contribute to make the inverse problem underlying adjoint slope tomography more robust and easier to tune than the ray-based counterpart by avoiding over-parametrization that can lead to significant parameter cross-talk and ill-conditioned inversion (Billéte & Lambaré 1998). Moreover, the computational burden resulting from the use of an eikonal solver instead of ray tracing is efficiently balanced by the fact that traveltime computations are performed from the source and receiver positions rather than from the scatterers. The second advantage is related to our ability to process large-scale data set without any matrix resolution. In return, we do not have access to the sensitivity matrix that is useful to perform local a posteriori resolution analysis. However, the computational efficiency of our approach can be used to build the resolution matrix numerically through spike tests. Alternatively, second-order

adjoint-state method can be used to implement truncated Newton optimization as an alternative to the LBFGS optimization. The efficient computation of Hessian-vector product with second-order adjoint state method (Métivier *et al.* 2013, 2014) can be also used to perform this resolution analysis (Fichtner & Trampert 2011). Implementation of the truncated Newton method in our adjoint slope tomography is ongoing.

One of the main limitations of our method compared to more automatic waveform inversion techniques such as reflection waveform inversion (Xu *et al.* 2012; Brossier *et al.* 2015; Zhou *et al.* 2015) is related to their ability to handle complex structures (Lambaré 2008). The main bottleneck remains the picking of local coherent events, which has not been investigated in this study. Beyond the picking issue, any fast solver providing the traveltime of the most-energetic arrival (Kim 2001) or of all the arrivals (Qian & Leung 2004) instead of the first traveltime might improve the imaging in complex area where ray folding occurs, as has been shown in a pre-stack-depth migration context (Operto *et al.* 2000).

Chauris *et al.* (2002a,b) have shown how the picking and stereotomographic inversion can be implemented in the pre-stack depth migration domain. Indeed, our approach can be also used when picking is performed in the migrated domain after demigration (modelling) of seismic invariants (Guillaume *et al.* 2008). The benefit of this strategy is that picking can be performed more easily in the migrated domain after suitable pre-processing of the common-image gathers, while the demigration allows one to avoid repeating migration at each iteration of the slope tomography. Chauris *et al.* (2002a,b) also show how migration-based velocity analysis (or slope tomography after demigration) can be performed with a reduced data-space and model-space parametrization involving one slope in the data space and the velocity model in the model space. This reduced parametrization can be considered in the framework of our adjoint formulation, although its robustness should be assessed through numerical investigation. Do we converge to the true minimum, when considering implicitly scatterer positions? Do we mitigate nonlinearities issues when adding these positions to the model space?

In slope tomographic methods, one scatterer is related to one shot-receiver pair and all of the scatterers are processed independently during the inversion. However, due to the intrinsic redundancy of seismic multioffset acquisition, one position in the subsurface is sampled by many shot-receiver pairs. Any optimization constraints on the relative position (i.e. proximity) of scatterers and the alignment of facets that are related to neighbouring sources and receivers should make the inversion more robust.

In this study, we assess adjoint slope tomography for short-spread reflection acquisitions. The use of an eikonal solver provides a robust modelling engine to extend adjoint slope tomography to long-offset acquisitions where traveltimes and slopes of both diving waves and reflected waves can be involved in the inversion as shown by Prieux *et al.* (2013). Use of long-offset data raises the issue of anisotropy, which can be taken into account with a factored eikonal solver proposed by Waheed *et al.* (2015) and Tavakoli F. *et al.* (2015) or more sophisticated approaches based on the resolution of the Hamilton-Jacobi equation with level-set techniques (Qian *et al.* 2003). Extension of adjoint slope tomography to TTI anisotropy is part of our future work. In this long-offset framework, adjoint slope tomography can provide reliable initial model for FWI in the sense that both approaches share the idea of using the full information content in the data through a dense picking of high-frequency scattered waves in slope tomography and the use of the waveforms of all of the arrivals in FWI.

6 CONCLUSIONS

We develop an adjoint formulation of the slope tomography as a new formulation of stereotomography to avoid large-scale matrix resolution. Moreover, a finite-difference eikonal solver is proposed as an alternative to ray tracing to avoid the issue of non-uniform ray sampling that can arise at long offsets and in complex models. Compared to the classical ray-based implementation of stereotomography, the model space involves subsurface velocities and scatterer positions, but not shooting angles and one-way traveltimes. The local subsurface dips, that are parametrized by the two shooting angles in ray-based stereotomography, are implicitly represented in the spatial support of the adjoint source and reconstructed through the update of close scatterers. The eikonal resolutions are performed from the shot and receiver positions to mitigate the number of the forward modelling required by the misfit function gradient: the number of state and adjoint problems to be computed scales to the number of sources and receivers but is independent of the number of scatterers, hence leading to an efficient computational framework. Exploiting the reciprocity principle of wave propagation, the slopes at the source and receiver positions are estimated in a finite-difference sense from traveltimes extracted in the subsurface far away from the source singularities. In this framework, source and receiver positions need not to be considered in the data space as traveltimes can be sampled accurately at arbitrary positions in the finite-difference grid with suitable interpolation schemes. The chosen smaller model space, compared to the model space in classic ray-based stereotomography, mitigates potential leakage between parameters and balances possible nonlinear features of the optimization scheme. The *l*-BFGS optimization scheme, which accounts for the Hessian in a cheap way, mitigates the leakage between wave speeds and scatterer positions. Other second-order optimization algorithms, such as the Truncated Newton method, can be also implemented in the adjoint slope tomography. Our approach takes advantage of a multiscale strategy during which the velocity grid is progressively refined. Through the estimation of velocity model for Marmousi model, we show that adjoint slope tomography is an appropriate candidate for building background velocity model for pre-stack depth migration and FWI. We plan to extend our formulation to 3-D TTI media as well as to consider applications to real case studies.

ACKNOWLEDGEMENTS

We thank René-Edouard Plessix, Samuel Gray and an anonymous reviewer for useful comments which have improved the writing of this paper. This study is partially funded by the SEISCOPE consortium sponsored by CGG, CHEVRON, EXXON-MOBIL, JGI, SHELL, SINOPEC, STATOIL, TOTAL and WOODSIDE (see website <http://seiscope2.osug.fr>). We thank L. Métivier (LJK/CNRS) for providing us the SEISCOPE optimization toolbox, R. Brossier for the discussions on the inverse scheme and Gilles Lambaré and Thibaut Allemand for fruitful discussions on stereotomography.

REFERENCES

- Alerini, M., Lambaré, G., Baina, R., Podvin, P. & Bégat, S.L., 2007. Two-dimensional PP/PS-stereotomography: *P*- and *S*-waves velocities estimation from OBC data, *Geophys. J. Int.*, **170**, 725–736.
- Al-Yahya, K., 1989. Velocity analysis by iterative profile migration, *Geophysics*, **54**(6), 718–729.

- Belayouni, N., 2013. Nouveaux algorithmes efficaces de modélisation 2D et 3D: Temps des premières arrivées, angles à la source et amplitudes, *PhD thesis*, Ecole nationale supérieure des mines de Paris.
- Billette, F., 1998. Estimation de macro-modèles de vitesse en sismique réflexion par stéréotomographie, *PhD thesis*, Université Paris VII.
- Billette, F. & Lambaré, G., 1998. Velocity macro-model estimation from seismic reflection data by stereotomography, *Geophys. J. Int.*, **135**(2), 671–680.
- Billette, F., Le Bégat, S., Podvin, P. & Lambaré, G., 2003. Practical aspects and applications of 2D stereotomography, *Geophysics*, **68**, 1008–1021.
- Bishop, T.N., Bube, K.P., Cutler, R.T., Langan, R.T., Love, P.L., Resnick, J.R., Shuey, R.T. & Spinder, D.A., 1985. Tomographic determination of velocity and depth in laterally varying media, *Geophysics*, **50**, 903–923.
- Bourgeois, A., Bourget, M., Lailly, P., Poulet, M., Ricarte, P. & Versteeg, R., 1991. Marmousi, model and data, in *The Marmousi Experience, Proceedings of the 1990 EAGE workshop on Practical Aspects of Seismic Data Inversion*, pp. 5–16, eds Versteeg, R. & Grau, G., Eur. Ass. Expl. Geophys.
- Brossier, R., Operto, S. & Virieux, J., 2015. Velocity model building from seismic reflection data by full waveform inversion, *Geophys. Prospect.*, **63**, 354–367.
- Bube, K.P., Langan, R.T. & Nemeth, T., 2005. Analysis of the spectral hole in velocity versus depth resolution for reflection traveltimes with limited aperture, *Geophysics*, **70**(3), U37–U45.
- Byrd, R.H., Lu, P. & Nocedal, J., 1995. A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Stat. Comput.*, **16**, 1190–1208.
- Chauris, H. & Noble, M., 2001. 2D velocity macro model estimation from seismic reflection data by local differential semblance optimization: applications on synthetic and real data, *Geophys. J. Int.*, **144**, 14–26.
- Chauris, H., Noble, M., Lambaré, G. & Podvin, P., 2002a. Migration velocity analysis from locally coherent events in 2-D laterally heterogeneous media, Part I: Theoretical aspects, *Geophysics*, **67**(4), 1202–1212.
- Chauris, H., Noble, M., Lambaré, G. & Podvin, P., 2002b. Migration velocity analysis from locally coherent events in 2-D laterally heterogeneous media, Part II: Applications on synthetic and real data, *Geophysics*, **67**(4), 1213–1224.
- Chavent, G., 1974. Identification of parameter distributed systems, in *Identification of Function Parameters in Partial Differential Equations*, pp. 31–48, eds Goodson, R. & Polis, M., American Society of Mechanical Engineers.
- Claerbout, J., 1985. *Imaging the Earth's Interior*, Blackwell Scientific Publication.
- Clément, F., Chavent, G. & Gómez, S., 2001. Migration-based travel-time waveform inversion of 2-D simple structures: a synthetic example, *Geophysics*, **66**, 845–860.
- Datta, D. & Sen, M.K., 2016. Estimating starting models for full waveform inversion using a global optimization method, *Geophysics*, **81**(4), R211–R223.
- de Boor, C., 1978. *A Practical Guide to Splines*, Springer-Verlag.
- Etgen, J., Gray, S.H. & Zhang, Y., 2009. An overview of depth imaging in exploration geophysics, *Geophysics*, **74**(6), WCA5–WCA17.
- Farra, V. & Madariaga, R., 1988. Non-linear reflection tomography, *Geophys. J.*, **95**, 135–147.
- Fichtner, A. & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophys. J. Int.*, **187**, 1604–1624.
- Fomel, S., Luo, S. & Zhao, H.-K., 2009. Fast sweeping method for the factored eikonal equation, *J. Comput. Phys.*, **228**, 6440–6455.
- Gardner, G.H., French, W.S. & Matzok, T., 1974. Elements of migration and velocity analysis, *Geophysics*, **39**, 811–825.
- Guillaume, P. et al., 2008. Kinematic invariants: an efficient and flexible approach for velocity model building, in *Advanced Velocity Model Building Techniques for Depth Imaging*, SEG Workshop.
- Hicks, G.J., 2002. Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions, *Geophysics*, **67**, 156–166.
- Kim, S., 2001. The most-energetic traveltime of seismic waves, *Appl. Math. Lett.*, **14**, 313–319.
- Lambaré, G., 2008. Stereotomography, *Geophysics*, **73**(5), VE25–VE34.
- Lambaré, G., Lucio, P.S. & Hanyga, A., 1996. Two-dimensional multivalued traveltime and amplitude maps by uniform sampling of ray field, *Geophys. J. Int.*, **125**, 584–598.
- Lambaré, G., Alerini, M., Baina, R. & Podvin, P., 2004. Stereotomography: a semi-automatic approach for velocity macromodel estimation, *Geophys. Prospect.*, **52**, 671–681.
- Métivier, L. & Brossier, R., 2016. The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication, *Geophysics*, **81**(2), F11–F25.
- Métivier, L., Brossier, R., Virieux, J. & Operto, S., 2013. Full Waveform Inversion and the truncated Newton method, *SIAM J. Sci. Comput.*, **35**(2), B401–B437.
- Métivier, L., Bretaudeau, F., Brossier, R., Operto, S. & Virieux, J., 2014. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures, *Geophys. Prospect.*, **62**, 1353–1375.
- Nag, S., Alerini, M., Duveneck, E. & Ursin, B., 2006. 2-D stereotomography for anisotropic media, in *Expanded Abstracts*, pp. 3305–3308.
- Noble, M., Gesret, A. & Belayouni, N., 2012. A new 3D eikonal solver for accurate traveltimes, takeoff angles and amplitudes, in *74th EAGE Conference & Exhibition incorporating SPE EUROPEC 2012*, Extended abstract.
- Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization*, 2nd edn, Springer.
- Operto, S., Xu, S. & Lambaré, G., 2000. Can we image quantitatively complex models with rays?, *Geophysics*, **65**(4), 1223–1238.
- Operto, S., Lambaré, G., Podvin, P. & Thierry, P., 2003. 3-D ray-Born migration/inversion. Part 2: Application to the SEG/EAGE overthrust experiment, *Geophysics*, **68**(4), 1357–1370.
- Paige, C.C. & Saunders, M.A., 1982. LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw.*, **8**(1), 43–71.
- 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.
- Prieux, V., Lambaré, G., Operto, S. & Virieux, J., 2013. Building starting model for full waveform inversion from wide-aperture data by stereotomography, *Geophysical Prospecting*, **61**(Issue supplement: 60 year anniversary issue), 109–137.
- Qian, J. & Leung, S., 2004. A level set based Eulerian method for paraxial multivalued traveltimes, *J. Comput. Phys.*, **197**, 711–736.
- Qian, J., Cheng, L.-T. & Osher, S., 2003. A level set based eulerian approach for anisotropic wave propagations, *Wave Motion*, **37**(4), 365–379.
- Riabinkin, L.A., 1957. *Fundamentals of Resolving Power of Controlled Directional Reception (CDR) of Seismic Waves*, vol. 14, pp. 36–60, Society of Exploration Geophysics, Translated and paraphrased from *Prikladnaya*, **16**, 3–36.
- Rieber, F., 1936. A new reflection system with controlled directional sensitivity, *Geophysics*, **1**, 97–106.
- Sajeva, A., M.Aleardi, Stucchi, E., Bienati & N., MazzottiA., 2016. Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the Marmousi model, *Geophysics*, **81**(4), R173–R184.
- Sword, C.H., 1987. Tomographic determination of interval velocities from reflection seismic data: The method of controlled directional reception, *PhD thesis*, Stanford University.
- Symes, W.W., 1998. High frequency asymptotics, differential semblance and velocity estimation, in *Expanded Abstracts*, pp. 1616–1619, Society of Exploration Geophysics.
- Taillandier, C., Noble, M., Chauris, H. & Calandra, H., 2009. First-arrival travel time tomography based on the adjoint state method, *Geophysics*, **74**(6), WCB1–WCB10.

- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier.
- Tavakoli F., B., Ribodetti, A., Virieux, J. & Operto, S., 2015. An iterative factored eikonal solver for TTI media, in *SEG Technical Program Expanded Abstracts 2015*, vol. 687, pp. 3576–3581.
- Thierry, P., Operto, S. & Lambaré, G., 1999. Fast 2D ray-Born inversion/migration in complex media, *Geophysics*, **64**(1), 162–181.
- Virieux, J. & Farra, V., 1991. Ray tracing in 3D complex isotropic media: an analysis of the problem, *Geophysics*, **16**, 2057–2069.
- Waheed, U.B., Yarman, C.E. & Flagg, G., 2015. An iterative, fast-sweeping-based eikonal solver for 3D tilted anisotropic media, *Geophysics*, **80**, C49–C58.
- Wang, F., Donno, D., Chauris, H., Calandra, H. & Audebert, F., 2016. Waveform inversion based on wavefield decomposition, *Geophysics*, **81**(6), R457–R470.
- Wu, Z. & Alkhalifah, T., 2015. Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion, *Geophysics*, **80**(6), R317–R329.
- Xu, S., Wang, D., Chen, F., Lambaré, G. & Zhang, Y., 2012. Inversion on reflected seismic wave, in *SEG Technical Program Expanded Abstracts 2012*, pp. 1–7.
- Zhao, H., 2005. A fast sweeping method for eikonal equations, *Math. Comput.*, **74**, 603–627.
- Zhou, W., Brossier, R., Operto, S. & Virieux, J., 2015. Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation, *Geophys. J. Int.*, **202**(3), 1535–1554.