Mitigating the ill-posedness of first-arrival traveltime tomography using slopes: application to the eastern Nankai Trough (Japan) OBS data set

S Sambolian, A Gorszczyk, S Operto, Alessandra Ribodetti, B Tavakoli F.

Geophysical Journal International

Geophys. J. Int. (2021) **227,** 898–921
Advance Access publication 2021 July 07
GJI Marine Geosciences and Applied Geophysics

https://doi.org/10.1093/gji/ggab262

Mitigating the ill-posedness of first-arrival traveltime tomography using slopes: application to the eastern Nankai Trough (Japan) OBS data set

S. Sambolian [1], A. Gorszczyk [2,3], S. Operto [1], A. Ribodetti[1] and B. Tavakoli F.[1,*]

[1] *Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, IRD, Géoazur,* 06560 *Valbonne, France. E-mail: sambolian@geoazur.unice.fr*
[2] *Institute for Geophysics, Polish academy of Science,* 01-452 *Warsaw, Poland*
[3] *Université Grenoble Alpes, ISTerre,* 38610 *Gières, France*

Accepted 2021 July 5. Received 2021 June 24; in original form 2021 February 19

SUMMARY

First-arrival traveltime tomography is one of the most used velocity model building techniques especially in sparse wide-angle acquisitions for deep crustal seismic imaging cases. Relying on the inversion of a picked attribute, the absolute traveltimes, the approach is ill-posed in terms of non-uniqueness of the solution. The latter is remedied by proper regularization or the introduction of prior information. Indeed, since traveltime kernels are vulnerable to the velocity–depth ambiguity, the inversion is stabilized by the introduction of complementary data like reflections and explicit reflectors in the velocity models. Here, we propose to supplement first-arrival traveltimes by their slopes, in other words the horizontal component of the slowness vectors at the sources and/or receivers. Slopes are a crucial attribute in state of the art scattering-based or reflection-based tomographic methods like slope tomography or wavefront tomography where the differential information is needed in order to locate the scattering events position or to parametrize the wavefront. The optional but valuable injection of slopes as an objective measure in first-arrival traveltime tomography stabilizes the problem by constraining the emergence angle or in turn implicitly the turning point depth of the rays. We explain why slopes have a tremendous added value in such a tomographic problem and highlight its remedial effect in cases where the medium is unevenly illuminated. We also show that the contribution of slopes become even more significant when the acquisition is sparse as it is generally the case with ocean-bottom seismometer surveys. The inferred models from such an extended time-attributes tomography will be used as initial guesses in a full-waveform inversion workflow context. The proposed strategy is benchmarked in 2-D media against a dip section of the SEG/EAGE overthrust model and then followed by a revisit of ocean bottom seismometers data from the eastern-Nankai subduction margin as a real deep crustal case study.

Key words: Inverse theory; Waveform inversion; Controlled source seismology; Crustal Imaging; Seismic tomography.

1 INTRODUCTION

Seismic traveltime tomography is one of the most widely used techniques due to its applicability to different purposes in passive and controlled-source seismics. Inferring the subsurface properties from seismic recordings, mainly wave speeds, is essential to understand seismogenic processes at the lithospheric scale, evaluating a resource play for exploration purposes or even near-surface characterization for geotechnical assessments.

Since the early works of Aki & Lee (1976) on traveltime tomography and its application using *P*-wave first-arrival traveltimes on a regional scale, inverting traveltimes in a least-squares sense under an infinite frequency approximation of wave propagation (Tarantola &

*Now at: TOTAL, 6400 Pau, France.

Nercessian 1984) was massively developed due to its ease of implementation. In parallel, and due to the limits of ray theory tied to its insensitivity to perturbations outside its infinitesimal travel path, finite-frequency methods accounting for heterogeneity in the first Fresnel zone emerged (e.g. Luo & Schuster 1991; Woodward 1992; Dahlen *et al.* 2000). In the latter, a linearized inversion around a reference model is done through for example the first-order Rytov approximation (Snieder & Lomax 1996) or the first-order Born approximation (Marquering *et al.* 1998). It is also worth noting that Zelt & Chen (2016) introduced a nonlinear variant where traveltimes are frequency dependent.

Regardless of the proven deficiency of ray-based inversion kernels, first-arrival traveltime tomography (FATT) was massively developed in the last decades for controlled-source seismic applications (e.g. Zhu & McMechan 1989; Zelt & Barton 1998). A lot of variants emerged for the method, including the introduction of eikonal solvers as a forward problem solver (Ammon & Vidale 1993) and the use of the adjoint-state method (Taillandier *et al.* 2009) instead of conventional Fréchet-derivative matrix building.

In FATT, the absolute traveltimes of the first-arrivals are inverted solely. The inversion algorithm consists in simulating traveltimes along a specific path that obeys Fermat's principle and then updating subsurface parameters along this trajectory in order to match observations. It is important to remind the reader that in the so-called nonlinear traveltime tomography, even though there is no linearization around the forward problem, it is assumed that the traveltime misfit function in a least-squares sense is locally quadratic with respect to small model perturbations. In this context, the ray-stationarity assumption (Hole 1992) holds in practice even though ray trajectories are implicitly redefined at each iteration through the forward problem.

Another key point to highlight is the fact that the inversion is done in a least-squares framework, mostly suitable for overdetermined problems whereas seismic tomographic problems are generally mixed-determined (Menke 1984). The latter is even more problematic in the context of FATT, depending on the acquisition and subsurface structures, since the medium is only partially illuminated by first-arrivals especially in the case of sparse fixed-spread acquisitions (Zelt 1999). The second problematic point is that the data misfit is supposedly only caused by the inaccuracy of the velocity parameters along the ray trajectory. In practice, the ray trajectory is altered depending on the updates in the first Fresnel zone and not only along it due to the implicit constraint imposed by Fermat's principle. The insensitivity of the data to model perturbations in poorly covered parts of the subsurface, the possibility to converge through numerous fitting solutions due to the first aforementioned point along with the unconstrained non-linearity of the ray trajectory erects an ill-posedness encountered in FATT. In reality, it is actually common to tackle cases with such pathological scenarios especially in crustal scale applications. In Fig. 1, a simplistic model containing two velocity gradients, followed by a sudden high contrast, respectively mimicking the upper crust, the lower crust and the upper mantle, is presented. The partial coverage often encountered in crustal cases is illustrated, the incomplete illumination of the subsurface and a local gap in ray coverage at intermediate depths between 50 and 165 km in distance provoked by the weak change in velocity gradients in the crust and the deep high velocity contrast around the Moho discontinuity (Fig. 1, upper panel). The base of the upper and lower crust are only covered by secondary arrivals, like diving waves beyond the crossover distances (Sheriff & Geldart 1995, p. 96, around 125 and 175 km) and wide-angle reflections in this case (Fig. 1, lower panel). We note that even if the sketch in Fig. 1 represents only what could be recorded as first-arrivals by a single receiver, the redundancy in the data volume related to these low-velocity zones is often insufficient due to a poor angular illumination.

The ill-posedness of FATT has been remedied by introducing additional arrivals like reflections or explicitly parametrized structural priors (Zelt & Smith 1992; Zelt 1999; Korenaga *et al.* 2000) or statistical knowledge on the sought solution during the inversion as a form of regularization (Tikhonov 1963; Delprat-Jannaud & Lailly 1992; Korenaga *et al.* 1997; Ajo-Franklin *et al.* 2007).

In the scope of this study, we do not discuss the different possible regularization recipes or model-driven constraints to mitigate the ill-posedness of FATT but we propose a straightforward data-driven remedy. Zhang & Toksöz (1998) and Zhang *et al.* (1998) inverted the slopes of the traveltime curves instead of absolute traveltimes (i.e. traveltime points) whereas Trinks *et al.* (2003) regularize FATT by introducing the slope information in the data covariance matrix. Along the same line of thought, the objective of this paper is to show how to use the slopes (horizontal component of the slowness vector at the sources and/or receivers) as an additional objective measure in FATT and how it affects the course of the inversion by constraining implicitly the geometry of the ray path and subsequently mitigate the generation of misleading structural artefacts in the models.

In fact, most state of the art scattering-based or reflection-based tomographic methods rely on a differential information as a needed complement to traveltimes. Stereotomography (Billette & Lambaré 1998; Lambaré 2008; Tavakoli F. *et al.* 2017b; Sambolian *et al.* 2019; Tavakoli F. *et al.* 2019), a slope tomography method based on locally coherent events utilizes slopes in order to constrain and define the scattering position associated with reflections or diffractions (Fig. 2a). The latter notion was introduced in automatic migration velocity analysis (Chauris *et al.* 2002) and even was used in the context of hypocentre-velocity problems (Sambolian *et al.* 2021). Prieux *et al.* (2013) used slopes along with traveltimes in order to invert for refracted arrivals but under the original framework, valid only for reflections. In other wavefront-based tomographic methods (Gelchinsky *et al.* 1999; Duveneck 2004; Bauer *et al.* 2017), the second-order derivatives of traveltime are even needed to parametrize the wavefront. On the other hand some methods, like polarization tomography (Hu *et al.* 1994) or double-difference tomography (Monteiller *et al.* 2005; Yuan *et al.* 2016) or seismic gradiometry (Curtis & Robertsson 2002; Langston 2007; de Ridder & Biondi 2015), rather utilize the differential information as a supplement for a higher resolution stabilized inversion.

The remainder of this paper is divided into the following sections. In the first section, we introduce slopes as an objective measure in the framework of FATT based on eikonal solver and the adjoint-state method (Tavakoli F. *et al.* 2018). As a supplement, we compute in the Appendix the Hessian-vector product through the second-order adjoint-state method (Fichtner & Trampert 2011; Métivier *et al.* 2017). We then present a first numerical experiment using analytic expressions where the added value of slopes is assessed in a laterally homogenenous

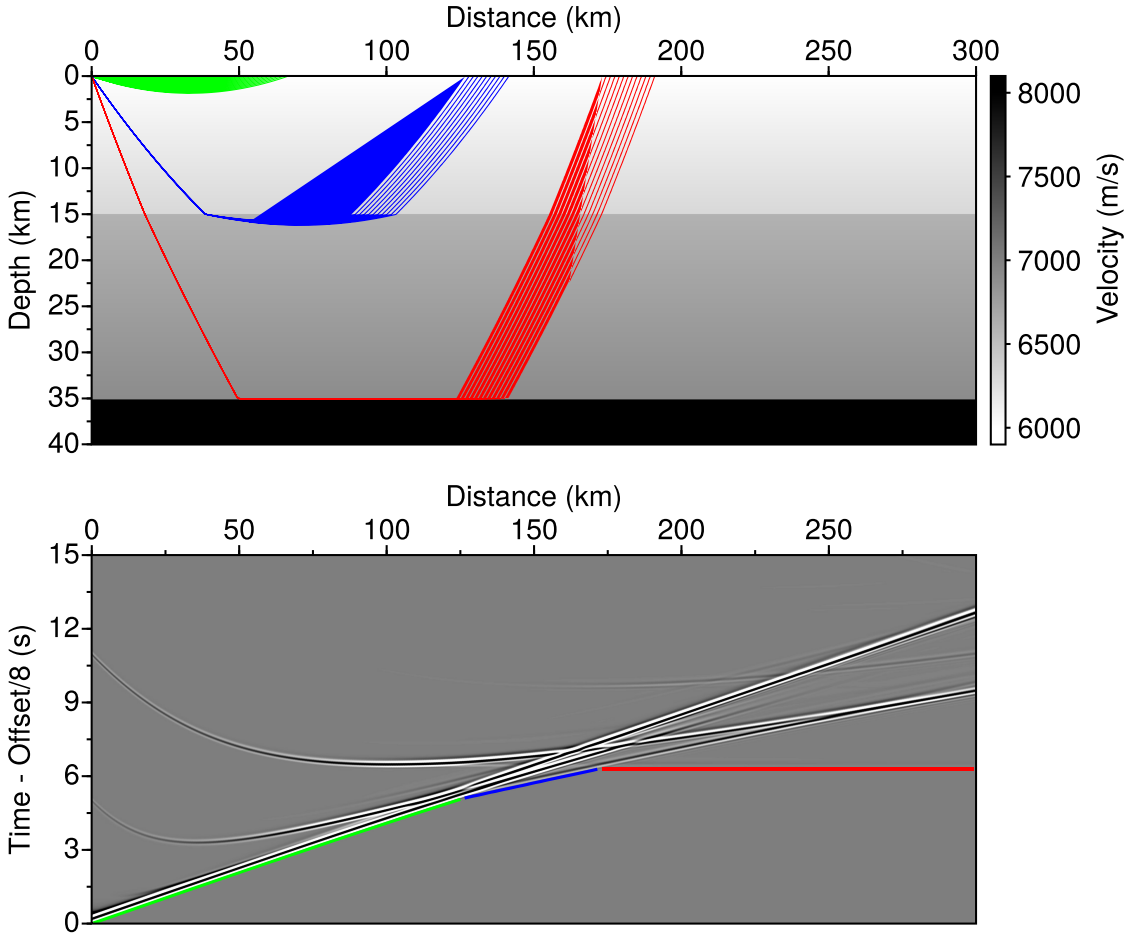

Figure 1. Simple numerical example illustrating the encountered blind zone in crustal imaging case. A gap is visible between 50 and 165 km at intermediate depths. Rays and arrivals in green and red denote the diving waves turning in the upper crust and head-waves propagating along the Moho. Adapted from Zelt (1999, their fig. 11).

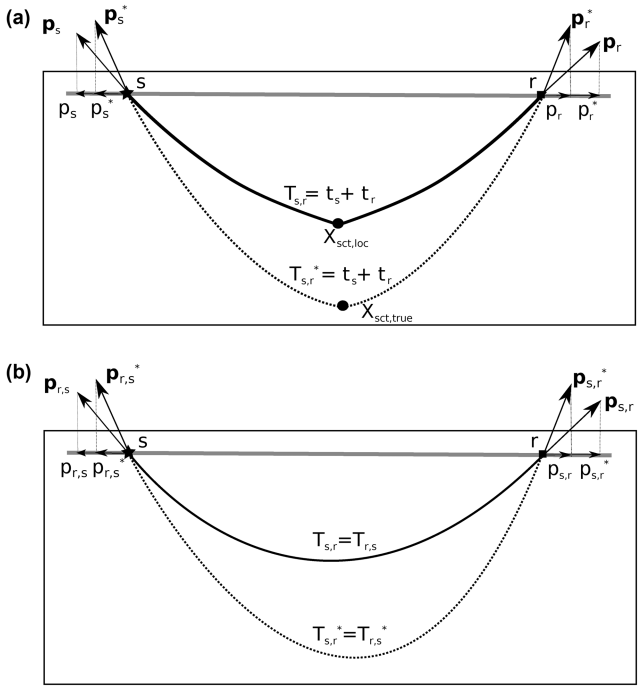

Figure 2. Slopes serving as a objective measure in slope tomography. (a) In the case of reflection slope tomography: the two-way traveltime $T_{s,r}$ (simulated through the computed one way traveltimes t_s and t_r) and the slopes at the source and receiver positions, p_s and p_r are needed to constrain the velocity model and the scattering point position X_{sct}. (b) In the case of FASTT: the slopes at the source and receiver positions, $p_{r,s}$ and $p_{s,r}$ supplement the two-way traveltime $T_{s,r}$ during the velocity model inversion by constraining implictly the geometry of the ray. The star and the square denote the source and receiver positions (with their respective slowness vectors). The recorded data are labelled with the superscript *. The dash line represents the true rays.T

medium where the velocity linearly increases with depth. Using different offset ranges in the acquisition and optimization schemes, we highlight how the slope information mitigates the ambiguity between the top velocity and the velocity gradient. We follow with an application on the SEG/EAGE overthrust model (Aminzadeh *et al.* 1997) where we use first-arrival traveltime/slope tomography (FASTT) as an initial model building tool for full-waveform inversion (FWI, Tarantola 1984; Virieux & Operto 2009). This case study shows how slopes constrain implicitly the geometry of the ray, especially when first-arrivals are guided along refractors. We then asses the method in the framework of crust-scale applications by inverting first-arrival traveltimes and slopes at shot positions extracted from a 2-D ocean bottom seismometer (OBS) data set collected in the eastern Nankai Trough (Tokai area, Japan, Dessa *et al.* 2004b; Operto *et al.* 2006; Górszczyk *et al.* 2017). Through our results, we illustrate the resolution power and robustness of FASTT relative to FATT for different OBS spacings and the impact of the reconstructed tomography models on FWI and prestack-depth migration of aligned towed-streamer data.

The method is developed for 2-D media. In the final conclusion and perspective section, we discuss the extension to 3-D media and to land acquisitions with topography.

2 FIRST-ARRIVAL SLOPE + TRAVELTIME TOMOGRAPHY (FASTT)

2.1 Theory

We define the following nonlinear constrained minimization problem with the aim of retrieving the minimizer $\mathbf{m}$ representing the subsurface parameters,

$$\min_{\mathbf{m}} C(\mathbf{m}) = \min_{\mathbf{m}} \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} (\mathbf{W}_d(\mathbf{d}(\mathbf{m}) - \mathbf{d}^*))^2, \tag{1}$$

where $\mathbf{d}(\mathbf{m})$ and $\mathbf{d}^*$ denote the predicted and observed measurement respectively. In turn, $\mathbf{d}$ groups the different attributes associated with a source-receiver pair (s, r): first-arrival traveltime $T_{s,r}$ and the slopes at the source and receiver positions $p_{r,s}$ and $p_{s,r}$, respectively. In the expression of the slopes, the first and the second subscripts refers to the starting point of the ray and the position where the slope is estimated, respectively (Fig. 2b). Moreover, we define the data space $\mathbf{D}$ as the space defined by the source–receiver pairs (s, r) and the model space $\mathbf{M}$ as the parametrized space of the subsurface domain $\mathbf{\Omega}$ to be imaged. The operator $\mathbf{W}_d$ denotes the inverse of the diagonal covariance matrix associated with each observable class (Tarantola 1987). The latter will serve as a means to render the data dimensionless. For the rest of the paper we develop the formulation as if both the slopes at the source and the receiver are accessible even though the formulation is still valid and advantageous for cases like sparse seabed acquisitions where one of them is only accessible (the slope at the source); as we will show later in this study.

We recast the minimization problem presented in eq. (1) as the following constrained problem

$$\min_{\mathbf{u},\mathbf{m}} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} J(\mathbf{u}) \quad \text{subject to} \quad \mathbf{F}(\mathbf{u}, \mathbf{m}), \tag{2}$$

where the operator $\mathbf{F}$ gathers the forward problem equations related to the calculation of the data $\mathbf{d}$, $\mathbf{u}$ gathers the state variables and $C(\mathbf{m}) = J(\mathbf{u}^*)$ where $\mathbf{u}^*$ stands for a realization of the physical constraints. We solve the constrained problem, (eq. 1) under a Lagrangian formalism following the adjoint-state method recipe (Haber *et al.* 2000; Plessix 2006). The augmented functional $\mathcal{L}$ in compact form is rewritten as

$$\mathcal{L}(\mathbf{m}, \mathbf{u}, \bar{\mathbf{u}}) = J(\mathbf{u}) - \left\langle \bar{\mathbf{u}} \mid \mathbf{F}(\mathbf{u}, \mathbf{m}) \right\rangle, \tag{3}$$

where $\langle . | . \rangle$ denotes the inner product and $\bar{\mathbf{u}}$ are the adjoint-state variables (or Lagrange multipliers). We proceed with the description of the physical (state) equations gathered by the non-linear forward problem operator $\mathbf{F}$. We infer the predicted traveltime and slopes from traveltime maps $[t_s(\mathbf{x}), t_r(\mathbf{x})]$ computed with a finite-difference factored eikonal solver using the source and receiver positions as injection points (Fomel *et al.* 2009; Tavakoli F. *et al.* 2015).

$$H(\mathbf{x}, \nabla t_s(\mathbf{x})) = 0 \quad \text{with} \quad t_s(\mathbf{s}) = 0, \tag{4}$$

$$H(\mathbf{x}, \nabla t_r(\mathbf{x})) = 0 \quad \text{with} \quad t_r(\mathbf{r}) = 0, \tag{5}$$

where $\mathbf{x} \in \mathbf{\Omega}$ and $\mathbf{s} = (x_s, z_s)$ and $\mathbf{r} = (x_r, z_r)$ denote the source and receiver positions. In the above equations, we impose a Dirichlet boundary condition by zeroing the traveltimes at the source and receiver positions. The operator H stands for the Hamiltonian representation of the Eikonal equation in tilted transversely isotropic (TTI) media (Alkhalifah 1998; Waheed *et al.* 2014) given by

$$H(\mathbf{x}, \nabla t(\mathbf{x})) = A(\mathbf{x})((\mathcal{R}(\mathbf{x})\nabla t(\mathbf{x}))_x)^2 + C(\mathbf{x})((\mathcal{R}(\mathbf{x})\nabla t(\mathbf{x}))_z)^2 + E(\mathbf{x})((\mathcal{R}(\mathbf{x})\nabla t(\mathbf{x}))_x(\mathcal{R}(\mathbf{x})\nabla t(\mathbf{x}))_z)^2 - 1, \tag{6}$$

where $\mathcal{R}$ is a standard rotation matrix and A, C, E are coefficients that embed the model parameters we seek to update depending on the chosen anisotropic parametrization (Alkhalifah & Tsvankin 1995; Plessix & Cao 2011; Gholami *et al.* 2013), for example

$$\begin{aligned} A(\mathbf{x}) &= v_v(\mathbf{x})^2(1+2\epsilon(\mathbf{x})),\\ C(\mathbf{x}) &= v_v(\mathbf{x})^2,\\ E(\mathbf{x}) &= -2v_v(\mathbf{x})^4(\epsilon(\mathbf{x})-\delta(\mathbf{x})), \end{aligned} \tag{7}$$

where v_v, ϵ and δ denote vertical velocity and the Thomsen parameters (Thomsen 1986). We refer the reader to Tavakoli F. *et al.* (2015) and Waheed *et al.* (2015) for a detailed description on the manner of solving eq. (6) in TTI media using the fast sweeping method (Zhao 2005; Luo & Qian 2012) as a global solver and a fixed-point iteration algorithm (Kelley 1995) for handling the quartic term.

In order to extract the traveltime solution at any position $\mathbf{x}_i \in \boldsymbol{\Omega}$ from the traveltime maps $t_s(\mathbf{x})$ and $t_r(\mathbf{x})$, we introduce a sampling operator $Q(\mathbf{x}-\mathbf{x}_i)$ implemented with a Kaiser-windowed sinc function (Hicks 2002). The first-arrival traveltimes $T_{s,r} \in \mathbf{D}$ are obtained straightforwardly by pointing the operator at the receiver position $\mathbf{r}$ in the traveltime map $t_s(\mathbf{x})$ initiated at the source s or vice versa depending on the acquisition and on the computational advantages of using the reciprocity principle

$$T_{s,r} = t_s(\mathbf{r}) = Q(\mathbf{x}-\mathbf{r})t_s(\mathbf{x}) = t_r(\mathbf{s}) = Q(\mathbf{x}-\mathbf{s})t_r(\mathbf{x}) = T_{r,s}, \tag{8}$$

while the slopes $p_r(\mathbf{s}) \in \mathbf{D}$ and $p_s(\mathbf{r}) \in \mathbf{D}$ at source and receiver positions, respectively, are computed in a finite-difference sense as

$$p_r(\mathbf{s}) = p_{r,s} = \frac{\partial T_{s,r}}{\partial x_s} = \frac{\partial t_r(\mathbf{s})}{\partial x_s} \approx \frac{1}{2h}(Q(\mathbf{x}-\mathbf{s}^+) - Q(\mathbf{x}-\mathbf{s}^-))t_r(\mathbf{x}), \tag{9}$$

and

$$p_s(\mathbf{r}) = p_{s,r} = \frac{\partial T_{s,r}}{\partial x_r} = \frac{\partial t_s(\mathbf{r})}{\partial x_r} \approx \frac{1}{2h}(Q(\mathbf{x}-\mathbf{r}^+) - Q(\mathbf{x}-\mathbf{r}^-))t_s(\mathbf{x}), \tag{10}$$

where $\mathbf{s}/\mathbf{r}^{\pm} = \mathbf{s}/\mathbf{r} \pm \mathbf{h}$, $\mathbf{h} = (h, 0)$ and $h > 0$ denotes a horizontal space shift suitable for accurate estimation of the slope, ideally the shot/receiver interval to re-use a precomputed traveltime map. Note that we assume for now that the sources and receivers are deployed on a horizontal line. The reader is referred to the final discussion section where the generalization to land acquisitions with complex topography or dense seabed acquisitions with complex bathymetry is discussed. A more precise strategy for the computation of the slopes would involve solving an additional eikonal-based partial differential equation solving for the emergence angle (Qian & Symes 2002) or tying the traveltime perturbation with respect to the source position (Alkhalifah & Fomel 2010).

Expanding $\mathbf{F}$ in eq. (3) using the state eqs (4), (5), (8), (9), (10) and associating the adjoint-state variables $\bar{\mathbf{u}} = (\xi_{s,r}, \xi_{r,s}, \mu_{s,r}, \lambda_s, \lambda_r)$ to their respective state variables $\mathbf{u} = (p_{s,r}, p_{r,s}, T_{s,r}, t_s, t_r)$, gives

$$\begin{aligned} \mathcal{L}(\mathbf{m},\mathbf{u},\bar{\mathbf{u}}) = J(\mathbf{u}) &- \frac{1}{2}\sum_{s=1}^{N_s}\left\langle \lambda_s(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_s(\mathbf{x}))\right\rangle_{\Omega} - \frac{1}{2}\sum_{r=1}^{N_r}\left\langle \lambda_r(\mathbf{x}) \mid H(\mathbf{x}, \nabla t_r(\mathbf{x}))\right\rangle_{\Omega}\\ &- \sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\xi_{r,s}\left(p_{r,s} - \frac{1}{2h}(Q(\mathbf{x}-\mathbf{s}^+) - Q(\mathbf{x}-\mathbf{s}^-))t_r(\mathbf{x})\right)\\ &- \sum_{r=1}^{N_r}\sum_{s=1}^{N_s}\xi_{s,r}\left(p_{s,r} - \frac{1}{2h}(Q(\mathbf{x}-\mathbf{r}^+) - Q(\mathbf{x}-\mathbf{r}^-))t_s(\mathbf{x})\right)\\ &- \sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\mu_{s,r}\left(T_{s,r} - Q(\mathbf{x}-\mathbf{r})t_s(\mathbf{x})\right), \end{aligned} \tag{11}$$

with the Lagrangian functional $\mathcal{L}$, valid in the subsurface domain $\boldsymbol{\Omega}$, dependent on the subsurface parameter $\mathbf{m}$ through the eikonal equation. According to the first-order optimality conditions, namely the so-called Karush–Kuhn–Tucker (KKT) conditions, a minimizer of a constrained optimization problem is reached at the saddle point of the Lagrangian function (Nocedal & Wright 2006) when the three following equations are satisfied:

$$\begin{cases} \partial\mathcal{L}/\partial\mathbf{u} = 0,\\ \partial\mathcal{L}/\partial\bar{\mathbf{u}} = 0,\\ \partial\mathcal{L}/\partial\mathbf{m} = 0. \end{cases} \tag{12}$$

The joint update of the entire system spanned by $\mathbf{u}$, $\bar{\mathbf{u}}$ and $\mathbf{m}$ is avoided due to computational complexity (Akçelik 2002). We thus resort to the reduced-space approach of the adjoint-state method (Haber *et al.* 2000; Plessix 2006) based on a sequence of variable projections. In other words, the first two KKT conditions of eq. (12) are satisfied by solving the state equations $\partial\mathcal{L}/\partial\bar{\mathbf{u}} = 0$ in a given model $\mathbf{m}_k$ at each iteration k and we then subsequently deduce the Lagrange multipliers by enforcing $\partial\mathcal{L}/\partial\mathbf{u} = 0$ in this manner

$$\frac{\partial\mathcal{L}}{\partial T_{s,r}} = 0 \rightarrow \mu_{s,r} = W_{T_{s,r}}(T_{s,r} - T^*_{s,r}) = W_{T_{s,r}}\Delta T_{s,r}, \tag{13}$$

$$\frac{\partial\mathcal{L}}{\partial p_{r,s}} = 0 \rightarrow \xi_{r,s} = W_{p_{r,s}}(p_{r,s} - p^*_{r,s}) = W_{p_{r,s}}\Delta p_{r,s}, \tag{14}$$

$$\frac{\partial \mathcal{L}}{\partial p_{s,r}} = 0 \rightarrow \xi_{s,r} = W_{p_{s,r}}(p_{s,r} - p^*_{s,r}) = W_{p_{s,r}} \Delta p_{s,r}. \tag{15}$$

The first three adjoint-state variables are the weighted data residuals serving as source terms in the following transport-like equations obtained through the derivation of $\partial\mathcal{L}/\partial t_s = 0$ and $\partial\mathcal{L}/\partial t_r = 0$ in the same manner as Tavakoli F. *et al.* (2017a)

$$\begin{aligned}(\nabla \cdot (\lambda_s(\mathbf{x})\,\mathcal{U}_s))_\Omega &= \frac{1}{2}\sum_{r=1}^{N_r}\Big(\, Q(\mathbf{x}-\mathbf{r})^t\ \mu_{s,r} + (Q(\mathbf{x}-\mathbf{r}^+)^t - Q(\mathbf{x}-\mathbf{r}^-)^t)\frac{\xi_{s,r}}{h}\Big),\\ (\nabla \cdot (\lambda_r(\mathbf{x})\,\mathcal{U}_r))_\Omega &= \frac{1}{2}\sum_{s=1}^{N_s}\Big(\, Q(\mathbf{x}-\mathbf{s})^t\ \mu_{s,r} + (Q(\mathbf{x}-\mathbf{s}^+)^t - Q(\mathbf{x}-\mathbf{s}^-)^t)\frac{\xi_{r,s}}{h}\Big),\end{aligned} \tag{16}$$

where Q^t stands for the adjoint of Q, namely the prolongation operator from **D** in **M**. The above adjoint-state equations satisfied by $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$ are solved with the fast sweeping method similarly to Leung & Qian (2006). The adjoint kernels are formed through the back-projection of the slope and traveltime residuals along two ray tubes following the group vectors $\mathcal{U}_s$ and $\mathcal{U}_r$ connecting the source–receiver pair. Finally, the gradient of the objective function $C(\mathbf{m})$, eq. (1), with respect to any TTI parameter can be inferred by the weighted summation of $\lambda_s(\mathbf{x})$ and $\lambda_r(\mathbf{x})$, which are independent of the chosen parameters for describing the model, as follows:

$$\nabla_{m(\mathbf{x})} C = -\sum_{s=1}^{N_s}\frac{\lambda_s(\mathbf{x})}{2}\frac{\partial H(\mathbf{x},\nabla t_s(\mathbf{x}))}{\partial m(\mathbf{x})} - \sum_{r=1}^{N_r}\frac{\lambda_r(\mathbf{x})}{2}\frac{\partial H(\mathbf{x},\nabla t_r(\mathbf{x}))}{\partial m(\mathbf{x})}. \tag{17}$$

We refer the reader to Tavakoli F. *et al.* (2019, the Appendix) for a detailed development of the gradient with respect to all TTI parameters. Once the gradient is computed and projected through the chain rule of derivatives on a cubic B-spline basis for multiscaling purposes, we proceed with a Newton-based local optimization scheme (Nocedal & Wright 2006)

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k\Big(\frac{\partial^2 C(\mathbf{m}_k)}{\partial \mathbf{m}^2}\Big)^{-1}\Big(\frac{\partial C(\mathbf{m}_k)}{\partial \mathbf{m}}\Big), \tag{18}$$

where the step length $\alpha_k \in \mathbb{R}_+$ satisfies the Armijo rule and the curvature condition of the Wolfe conditions. In practice, for all numerical experiments presented in this study, the inexact line search is managed by the SEISCOPE optimization toolbox (Métivier & Brossier 2016) as well as the inverse Hessian operator approximation through the limited-memory Broyden–Fletcher–Goldfarb–Shannon (L-BFGS) algorithm (Byrd *et al.* 1995). We present in the Appendix, the second-order adjoint-state formulation (Fichtner & Trampert 2011; Métivier *et al.* 2017) of FASTT where the Hessian-vector product, embedding a better approximation of the Hessian, can be used in a truncated Newton scheme (Nash 2000).

2.2 Sensitivity analysis

Before we proceed with the presentation of two full scale applications, we illustrate the advantages of FASTT with respect to FATT in a toy test setup. We consider a two parameter estimation problem of constant gradient medium ($v = v_0 + a \times z$), with the true top layer velocity $v_0 = 2.73\ \mathrm{km\,s^{-1}}$ and the gradient $a = 0.75\ \mathrm{s^{-1}}$ (white cross in Fig. 3) . We use analytical expressions for solving the forward problem and the Fréchet derivatives computation (Udías 2000; Stovas & Alkhalifah 2014). In Fig. 3, we show the inversion results using two different acquisition setups and starting with the same initial guess of $v_0 = 3\ \mathrm{km\,s^{-1}}$ and $a = 0.45\ \mathrm{s^{-1}}$. The first test (Fig. 3a) is done for one shot and mimicks a surface fixed-spread 100-km-long acquisition with 100 m spacing between each receiver; the length of the acquisition ensures a fair illumination at various depths. The results of the first setup show a convergence of both FATT (Fig. 3a) and FASTT (Fig. 3b) towards the sought solution. On the other hand, the optimization path taken by each is clearly different due to the different shape of the attraction basin. In the case of FATT, the basin is flat and lacks curvature therefore the minimum is harder to reach. As a counterpart, the minimum in the case of FASTT is more distinct, hence the convergence in fewer iterations even for this simple case. The introduction of slope observations balances the sensitivity with respect to both parameters while in the case of FATT the elongated ellipse delineating the basin depicts a superior sensitivity to the top velocity parameter. At this point of the discussion, it is rather evident that, in the case of dense long-offset acquisitions where the medium is sufficiently well sampled by multi-incidence crossing rays, fitting traveltimes perfectly ensures a convergence towards a global minimum. Having said that, we now remind the reader that, in practice, flawless data fit is never achieved, long offsets are never long enough and the often contrasted subsurface is not illuminated at all depths by diving waves. The latter case is encountered frequently in subsalt (Shen *et al.* 2018) or deep crustal (Zelt 1999; Korenaga *et al.* 2000) imaging cases, short to intermediate offset first-arrival recordings are trapped by shallow velocity contrast while only very long-offset recordings are actually coming from wave propagation in the deep parts (Fig. 1). The poorly constrained areas at intermediate depth aggravate the non-uniqueness of the completely blind tomography to the velocity updates dictating the turning point depth. The second part (Figs 3c and d) of the toy test looks closely at that kind of phenomena, where we invert for offsets between 5–30 km and 60–100 km as if part of the medium is poorly constrained by the data. The results of the inversion illustrates how FATT struggles in a very flat attraction basin composed of a multitude of parameter combinations (Fig. 3c), while FASTT is still able to converge in a fewer number of iterations (Fig. 3d). The toy test depicted that the introduction of slopes in FATT mitigates its ill-posedness. The effect is logical since the emergence angle and therefore the turning point of the rays are better constrained. In order to

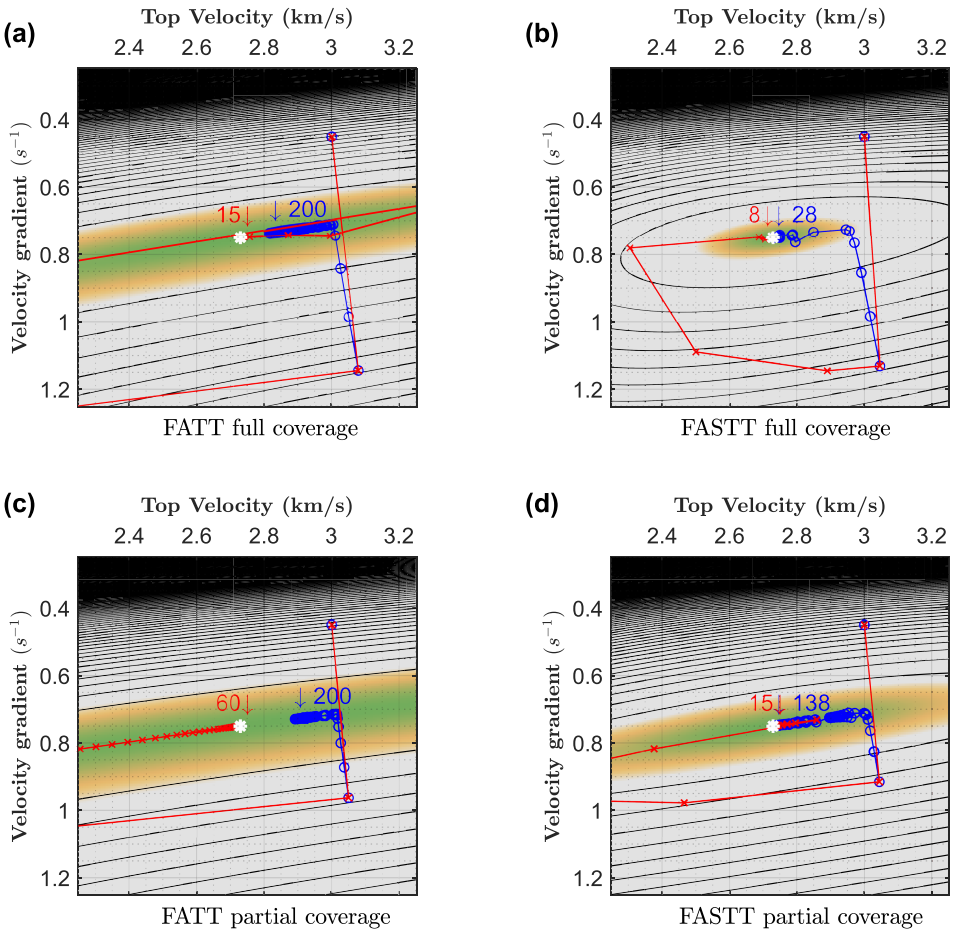

Figure 3. Toy test for a constant vertical velocity-gradient medium. The medium is parametrized by two parameters, the top velocity v_0 and the vertical velocity gradient a. The colored lines denote the optimization paths taken by the FATT (a, c) and FASTT (b, d) using a steepest-descent (blue) and a BFGS (red) scheme. The black curves denote the cost function iso-values. The white diamond denotes the sought minimum. Two acquisition setups: (a and b) a full offset settings; (c and d) a partial acquisition (missing intermediate offsets).

avoid any confusion we remind the reader that we are looking at the slope at the source and receiver and not along the ray, even though in the presented toy test the model is laterally homogeneous, slopes still have a contribution in guiding the rays through the initial condition defined by the emergence angle. In the subsequent sections, we strengthen our claims on full-scale applications.

3 APPLICATION TO THE SEG/EAGE OVERTHRUST MODEL

In the following section, we benchmark the performance of FATT and FASTT against a 2-D section of the SEG/EAGE overthrust model (Aminzadeh *et al.* 1997) in a FWI workflow context. The 20-km-long overthrust model (Fig. 4d) contains some challenging features for tomography and in particular ray-based approaches. The main target is a dipping thrust structure surrounded by alternating positive and negative velocity contrasts layers, other features include small scale fractures and channels but are retrievable only at the FWI step. The thrust structure and the non-deformed high velocity basement are the major cause of ray trapping and shadow zones making the problem challenging for ray-based methods.

For the sake of avoiding incomplete coverage of the main target situated at the edge we extended the model laterally 25 km from each side using the boundary values. This extension, only done at the tomography stage of the workflow, will also ensure a proper undershooting of the target through long-offset arrivals that are refracted on top of the basement. The observed data are simulated in a slightly smoothed version, using the eikonal solver described earlier, guaranteeing the validity of ray theory while preserving the kinematics of the model.

3.1 Dense acquisition results

In the first test, we consider a dense 70 km fixed-spread acquisition (100 m source–receiver spacing) and an initial constant gradient model defined with $v_0 = 2000\,\mathrm{m\,s^{-1}}$ and $a = 0.8\ \mathrm{s^{-1}}$ (Fig. 4a). We use only one slope in order to draw a clear comparison with a subsequent sparse acquisition setting test. Same strategy is used during both FATT and FASTT, the inversion is regularized through gradient smoothing using a

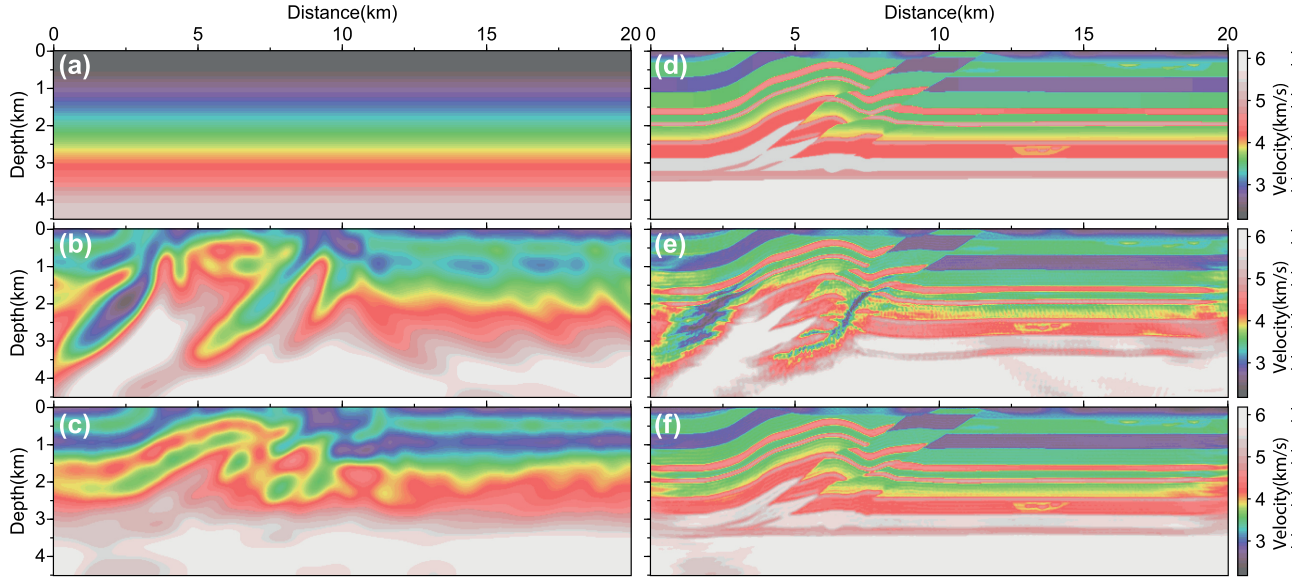

Figure 4. Overthrust case study. Dense acquisition case. (a) Initial velocity model. (b and c) Velocity models inferred from FATT and FASTT, respectively. (d) True velocity model. (e and f) Final velocity model inferred from FWI using (b) and (c) as initial guesses, respectively.

300 m correlation length Gaussian filter and we use a multi-scale reconstruction approach through progressive B-spline refinement. We use line-search failure as a stopping criterion, in other words when a step length satisfying the Wolfe conditions cannot be found anymore. The results for FATT and FASTT (Figs 4b and c) exhibit different features. In the FATT model (Fig. 4b), the main velocity trend is coherent with what would exist in a smooth version of the true target (Fig. 4d). However, some artefacts appear along the main dipping structure of the thrust (2.5−7.5 km). On other hand, the FASTT model (Fig. 4c) contains well delineated features even in the main target. Before we elaborate on the causes behind the discrepancies between the two inverted models, we use them as initial guesses for FWI and asses their impact on the final result. We note that any practitioner would not use the obtained FATT model as initial guess for FWI and opt for a suitably regularized inversion result. However, in order to draw a clear comparison between the FATT and FASTT using the same workflow, both will be used as initial guesses for FWI.

We proceed with a frequency-domain FWI in the frequency band of 3–20 Hz through a hierarchical frequency continuation scheme (Pratt 1999) subdivided in eight groups. We use the same acquisition setup as the one used during tomography, but now the original 20 km profile is considered for FWI. The latter taking advantage of the improved subsurface illumination provided both by finite-frequency wavepath kernels and late arrivals (by opposition to ray path kernels and first-arrivals). No regularization was used at this stage but we precondition the descent direction with a diagonal pseudo-Hessian (Shin *et al.* 2001). The FWI result (Fig. 4e) using the FATT model as initial guess shows artefacts in the main target but a good recovery of the layered structures, clearly incriminating the poor result at the tomography stage. On the contrary, FWI (Fig. 4f) was able to retrieve perfectly all the structures using FASTT result as an initial model.

We investigate the source of artefacts in the FATT model by taking a look at some rays traced in the smooth version of the model used for the data simulation and the inverted tomographic models (Fig. 5). For the rest of the section, we refer to the smooth version of the true model as the target model since it represents the best case scenario outcome of our tomography if perfect resolution was attainable. By first examining the rays traced in the target model (Fig. 5a), we notice a predictable channeling effect occurring along the thrust but more interestingly the channeling is also along the velocity contrast interfaces at 1.5 and 3 km depth. The rays distribution in the FATT model (Fig. 5b) shows a lack of ray path along the aforementioned interfaces and a significantly less rich angular coverage, whereas in the case of FASTT (Fig. 5c), the rays follow the same paths as the ones seen in the target model. This observation implies that FATT was not as sensitive as FASTT to the small contrasted layer and took a different direction during the inversion. Indeed, since FATT was insensitive to some velocity variations, the rays went deeper and in turn the inversion went in a direction that forced the fitting of the traveltime by generating some compensating artefacts. The latter is a typical result of the ill-posedness of FATT, a usual remedy for this deficiency is of course a proper regularization at the cost of resolution, that is not however the purpose of our discussion as mentioned earlier. We reiterate that FASTT is more robust to this effect since the emergence angle and in turn the ray path is better constrained, hence leading to a more stable inversion.

As an additional support to our claim, we show the gradient computed on the Cartesian grid (not the one used in the optimization scheme since we project it on a cubic B-spline basis) at the first iteration. We can clearly notice in Fig. 6(a) how the amplitudes of the gradient are dominated by the long-offset rays density in the deep part while the variations in the shallow part are somewhat monotonic and laterally homogeneous. On the other hand, in the case of FASTT (Fig. 6b) the gradient is very contrasted and in particular along the thrusts where the channeling occurs since slopes are more sensitive to velocity variations surrounding the ray path.

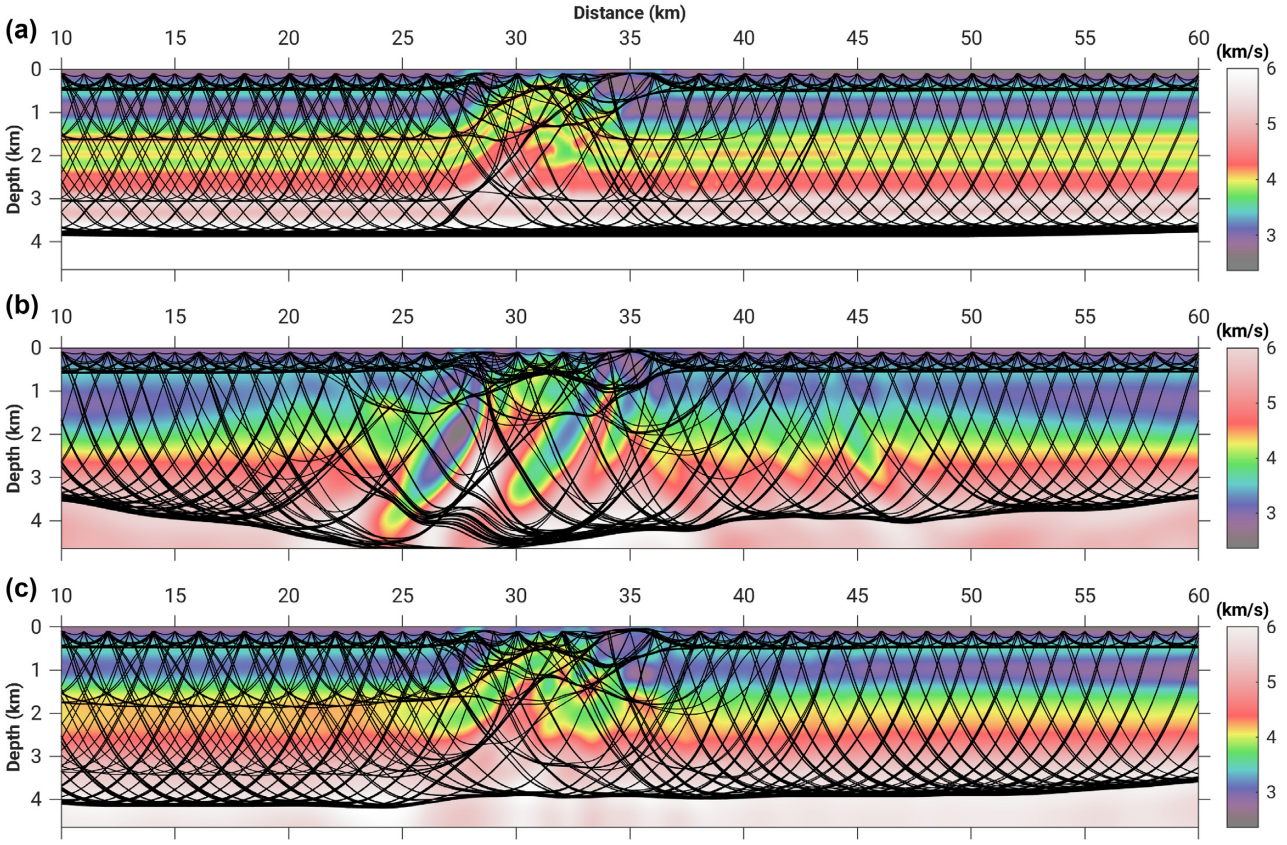

Figure 5. Overthrust case study. Rays traced in the extended target model (a), the tomographic models (b and c) inferred from FATT and FASTT, respectively.

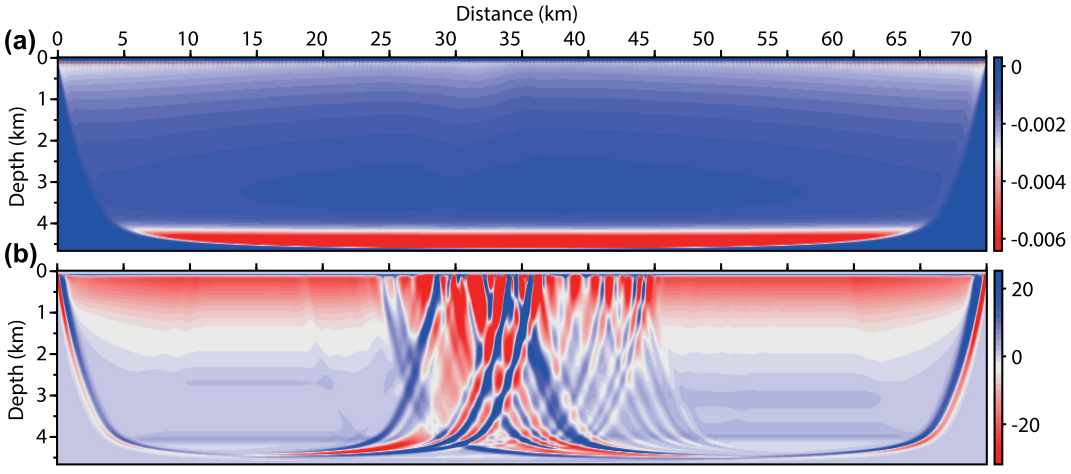

Figure 6. Overthrust case study. Gradient computed at the first iteration in the case of FATT (a) and FASTT (b).

3.2 Sparse acquisition results

In order to check for the effect of a more realistic ocean bottom seismometers (OBS) spacing and push the tomography to its limits, we repeat the same test but with a 2 km receiver spacing while using the same workflow across all stages even the gradient smoothing. The inverted models (Fig. 7) using the sparse acquisition illustrates a bigger discrepancy between FATT and FASTT results in this unfavorable setup. The velocity model inferred from FATT (Fig. 7a) is contaminated by kernel imprints clearly revealing that a more aggressive regularization at the cost of resolution is needed to compensate for the sparsity of the acquisition. Using the same regularization, FASTT was able to retrieve a solution that is fairly comparable to the one of the dense acquisition setup. Both methods fitted their corresponding data (Fig. 8, red and blue dashes). However, FATT reaching a local minimum was not able to implicitly fit the slopes. This test proves that in the case of unfavorable acquisition setups and a lack of proper regularization during the inversion, fitting traveltimes is not enough since it is a deficient attribute.

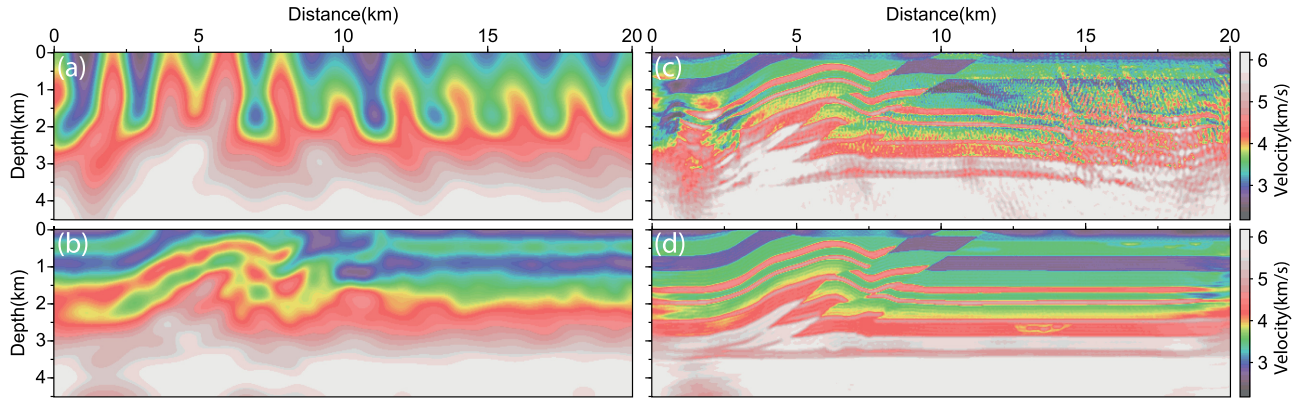

Figure 7. Overthrust case study. Sparse acquisition case. (a and b) Velocity models inferred from FATT and FASTT, respectively. (c and d) Final velocity model inferred from FWI using (a) and (b) as initial guesses, respectively.

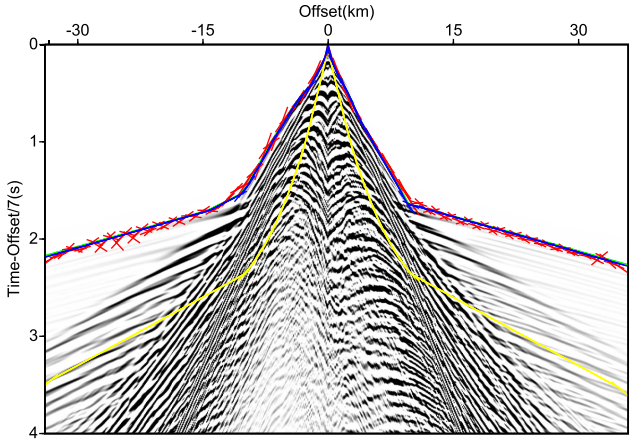

Figure 8. Overthrust case study. Common-receiver gather simulated at the top of the thrust structures in the extended target model superimposed by traveltimes and slopes calculated in the initial (yellow dashes), exact model (green dashes), FATT (red dashes) and FASTT models (blue dashes).

In the same manner as the first test, we proceed with FWI using both FATT and FASTT models as initial models. Starting from FATT model, the unsurprising results show that FWI could recover some parts of the overthrust model (Fig. 7c) since the first-arrival is fitted but is still contaminated by some unavoidable artefacts while starting from the FASTT model the FWI result is very satisfying (Fig. 7d). We proceed with an examination of the seismograms match post-FWI in both cases to check for cycle-skipping (Fig. 9). We clarify to the reader that in instance of good match, the waveforms should be represented by a blue/black color scale opposed to the red/black for mismatched waveforms. The FATT+FWI seismograms (Fig. 9a) do not exhibit a flagrant cycle-skipping pattern, however we can see a series of mismatches in late arrivals at short to intermediate offsets and early arrivals at long-offset due to the artefacts present in the model. The kinematics is as expected very well recovered in the FASTT+FWI case (Fig. 9b).

4 APPLICATION TO THE EASTERN NANKAI TROUGH (JAPAN)

We revisit a 2-D real data crustal case study in the eastern Nankai Trough, offshore Japan (Dessa *et al.* 2004a,b; Operto *et al.* 2006; Górszczyk *et al.* 2017, 2019). This offshore subduction zone (Fig. 10) is of interest due to its seismicity induced by the N–S convergence of the Philippine Sea plate and the Eurasia plate at the Nankai Trough (Le Pichon *et al.* 1996). The easternmost segment referred to as the Tokai area, delineated by the colliding Izu–Bonin–Mariana arc, entices a lot of research studies since it remains unruptured and accumulating stress for over one and half centuries (Ando 1975). During this study, we use multichannel seismic (MCS) and OBS data acquired in the frame of the Seize France Japan (SFJ) project (Dessa *et al.* 2004b). We first perform FATT and FASTT using first breaks picked on OBS data followed with the FWI workflow designed by Górszczyk *et al.* (2017). We redo the experiment using a decimated data set keeping one instrument out of ten in order to assess the impact of the acquisition sampling on our results. We finally evaluate the kinematic sounds of the inverted tomographic models through a preserved-amplitude ray+Born depth migration/inversion (Thierry *et al.* 1999) of an aligned MCS profile carried out with a 4.5-km-long streamer (Górszczyk *et al.* 2019).

The OBS survey was carried out with 100 instruments spaced 1 km apart, 93 of them are usable (Fig. 10, red line) and an aligned 140-km-shot profile with a 100 m shot interval (Fig. 10, black line). The shot profile is 40-km longer than the OBS line leading to a deficit of

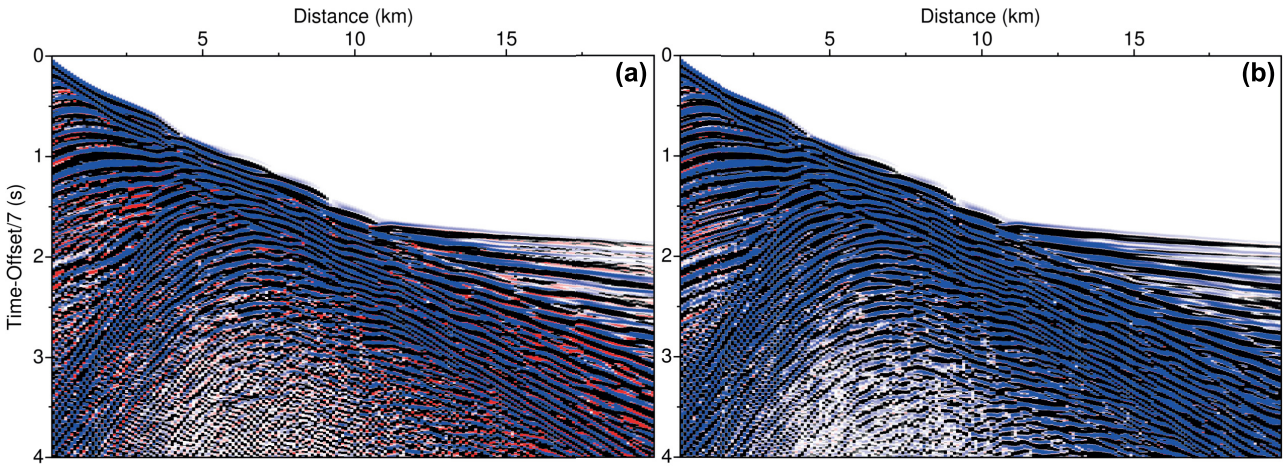

Figure 9. Overthrust case study. Common-receiver gather simulated at the start of the exact model in blue/red superimposed by a common-receiver gather in black/transparent simulated at the same position in the FATT+FWI (a) and FASTT+FWI (b) models of Figs 7(c) and (d).

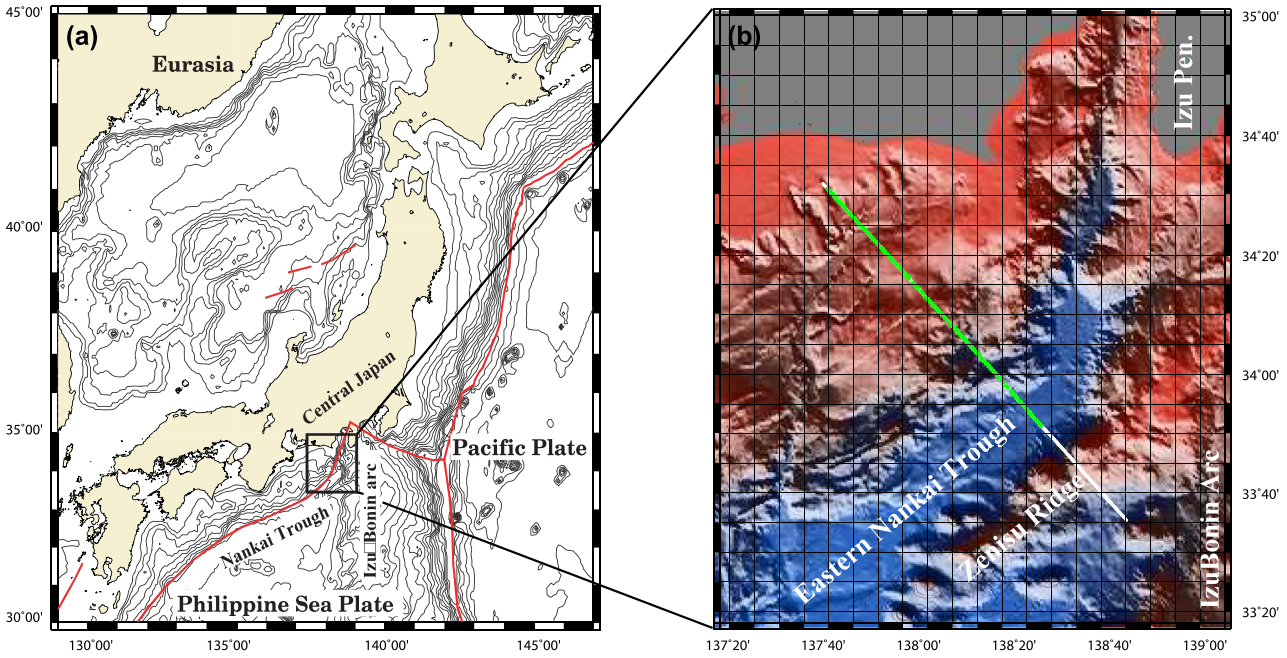

Figure 10. Nankai case study. (a) Geodynamical context. (b) SFJ acquisition map, the white line and greens stars delineates the shot profile and OBS positions respectively. Adapted from (Operto *et al.* 2006).

short offsets at the seaward end of the survey above the Zenisu ridge, a compressive structure running parallel to the trench and located just seaward to it (Fig. 10b). This lack of short offsets at the end of the model aggravates the ill-posedness of FATT and can generate artefacts when the starting model of the tomography does not contain some priors about the structure of the Zenisu ridge (Dessa *et al.* 2004a). For tomography, we use 124 248 first breaks previously inverted by Górszczyk *et al.* (2017). We approximate the slopes at the sources in a finite-difference sense after a spline interpolation of the traveltime curves (Fig. 11). We do not endorse this suboptimal strategy versus explicit slope picking but we opted for it in order to have a comparable result with respect to previous publications using the same picks.

To introduce the main structural units of the survey area, we show in Fig. 12 the FWI model and its detrended version developed by Górszczyk *et al.* (2017). From southeast to northwest, the main structural domains involve the trench axis between 105 and 88 km distance with weakly deformed sedimentary fill (WDU), the active accretionary wedge (MDU) bounded on the west by the Tokai thrust, the Miocene wedge (HDU) bounded on the west by the Kodaiba thrust and the backstop (BST) which undergoes an important compressive tectonic regime highlighted by several presently inactivated major thrusts and underplated crustal sheets (Le Pichon *et al.* 1996; Dessa *et al.* 2004b; Henry *et al.* 2004). A still ambiguous area on top of the subducting oceanic crust (SOC) is located between 55 and 60 km distance at the position

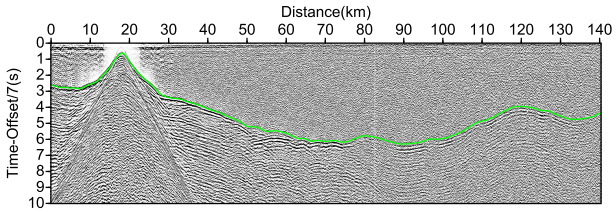

Figure 11. Nankai case study. OBS-17 seismograms overlain by their corresponding first-break picks (green line). The seismograms have been processed by spectral whitening, bandpass filtering and automatic gain control. The Pn wave is recorded as a first-arrival beyond 70 km offset. Note the footprint of the Zenisu ridge on the traveltime curve at long offset.

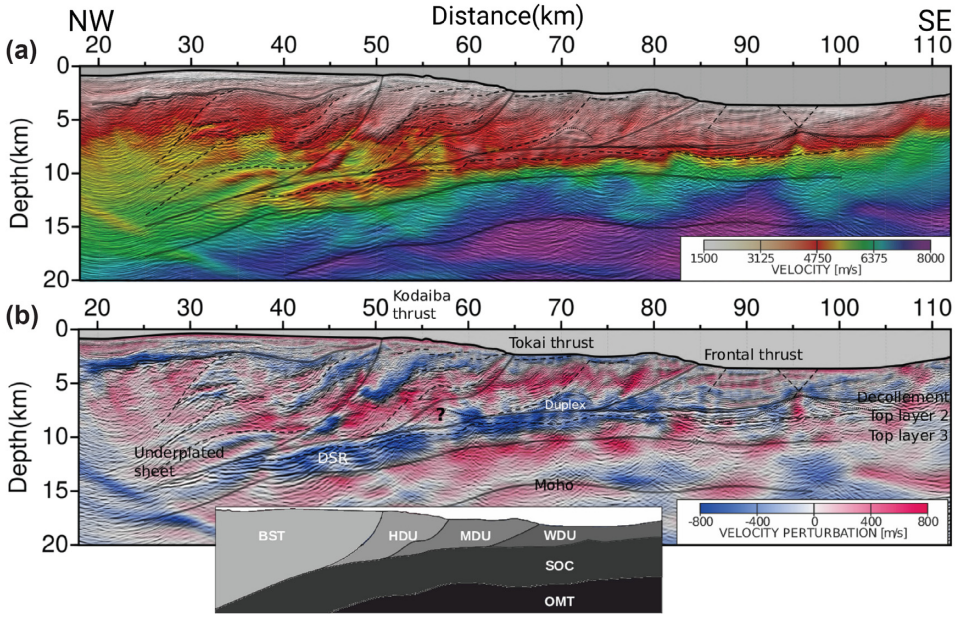

Figure 12. FWI model of the eastern Nankai through (Górszczyk *et al.* 2017, 2019). (a) FWI velocity model. (b) Detrended FWI velocity model. A depth migrated section inferred from the MCS data and a gross structural line drawing delineating the main structural units and tectonic features such as the Tokai and Kodaiba thrusts is superimposed on the models. The inset delineates the main structural domains as interpreted by Henry *et al.* (2004). SOC: subducting oceanic crust. OMT: oceanic mantle. WDU: weakly deformed unit (trench fill); MDU: moderately deformed unit (active wedge); HDU: heavily deformed unit (Miocene wedge); BST: backstop. DSR: deep strong reflector. The question mark in (b) points the possible location of the Palaeo-Zenisu ridge (Le Pichon *et al.* 1996). Note that the figure shows only the section of the model that is well covered by the survey (i.e. by both the OBS array and the shot profile). Accordingly, the Zenisu ridge at the seaward end of the model is not shown. Adapted from Górszczyk *et al.* (2019).

of the presumed subducting palaeo-Zenisu ridge (Le Pichon *et al.* 1996, Fig. 12, question mark). The decollement on top of the subducting oceanic crust is identified by a continuous reflector between 60 and 90 km distance and 7.5 km depth. A striking feature is a Deep Strong Reflector (DSR) correlated with a low-velocity zone on top of the subducting oceanic crust where it pinches out the backstop at 40 km distance and 11.6 km depth.

4.1 Complete acquisition results

We run FATT and FASTT starting with an initial constant gradient velocity model ($v_0 = 4000\,\mathrm{m\,s^{-1}}$ and $a = 0.2\,\mathrm{s^{-1}}$) following the bathymetry level (Fig. 13a). Unlike the starting velocity model used for FATT by Dessa *et al.* (2004a) and Górszczyk *et al.* (2017), our choice of the initial guess is not based on any prior such as the dip of the subducting slab and the crustal root at the Zenisu ridge. The workflow is similar to the one presented in the previous section: we use a multiscale reconstruction approach through progressive B-spline refinement and gradient smoothing (500 m × 800 m correlation length) as a regularization. Following a line-search failure after 50 and 85 iterations for FATT and FASTT respectively, the data misfit is significantly reduced (Fig. 14). Both inversions reached a very close traveltime misfit with approximately 98 per cent of the traveltime residuals falling below the FWI cycle-skipping threshold for a starting frequency of 1.5 Hz (Pratt 2008, their eq. 1). In Fig. 14, we notice that the slope residuals in the case of FASTT are significantly reduced opposed to the less implicitly fitted slopes in the case of FATT. The inferred tomographic models (Figs 13b and c) exhibit a traveltime RMS of 28 ms noting that the poor illuminated areas at the end of the shot profile contribute to most of it, particularly in the case of FASTT. Even though the traveltime data fit and the RMS error are similar, the structures seen in the models are different. The FATT model (Fig. 13b) contains some traveltime sensitivity kernel imprints inherited from ray paths initiated along the Zenisu ridge and emerging later on throughout the middle. Indeed, an unlikely wavy pattern affects

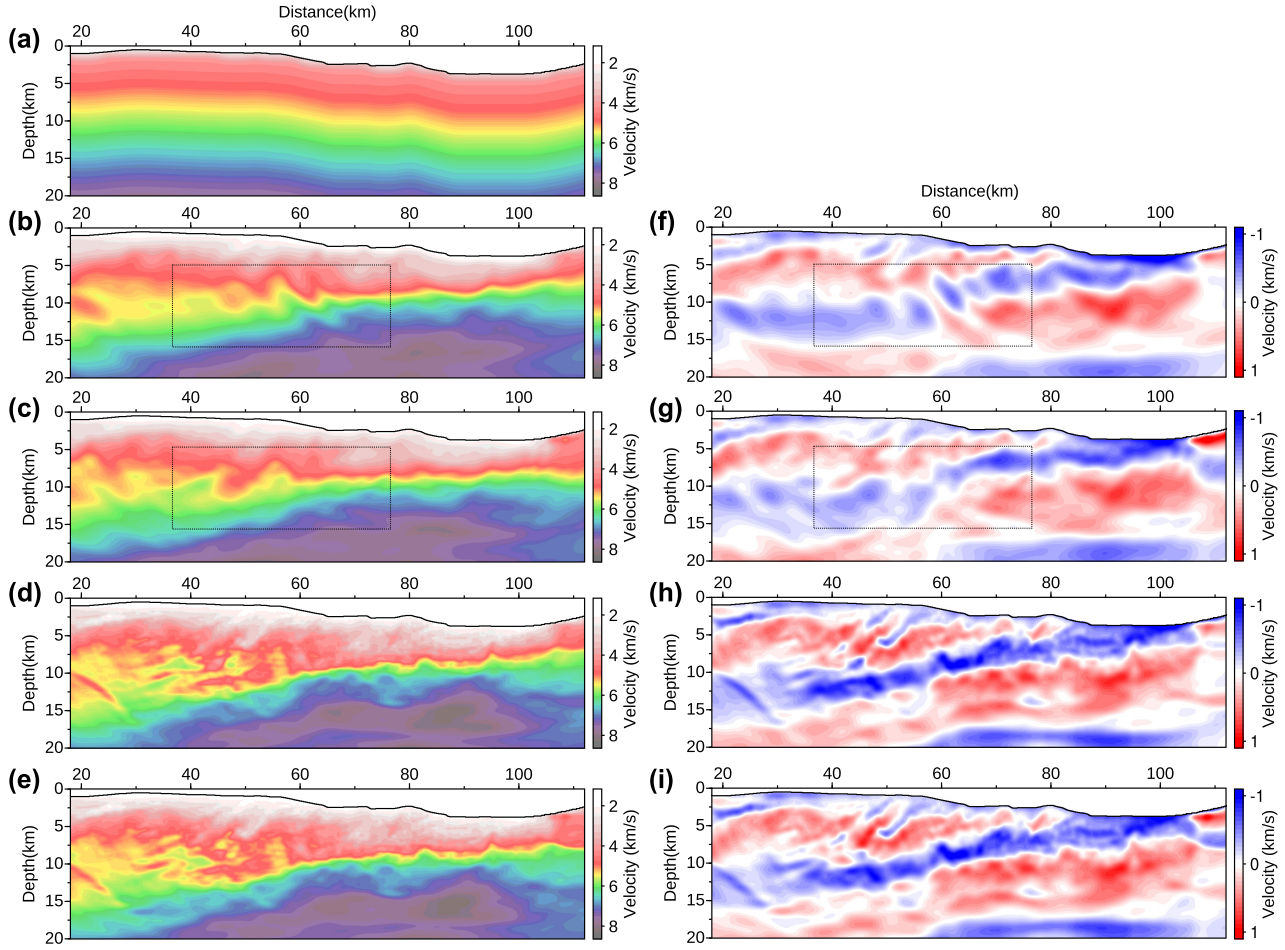

Figure 13. Nankai case study. Dense acquisition results. (a) Initial velocity model. (b and c) Velocity models inferred from FATT and FASTT, respectively. (d and e) Final velocity model inferred from FWI using (b) and (c) as initial guesses, respectively. (f–i) The detrended version of the inverted velocity models (b-e). The dash box delineates the most striking differences between the FATT and the FASTT models. The black line in all of the panels delineates the bathymetry.

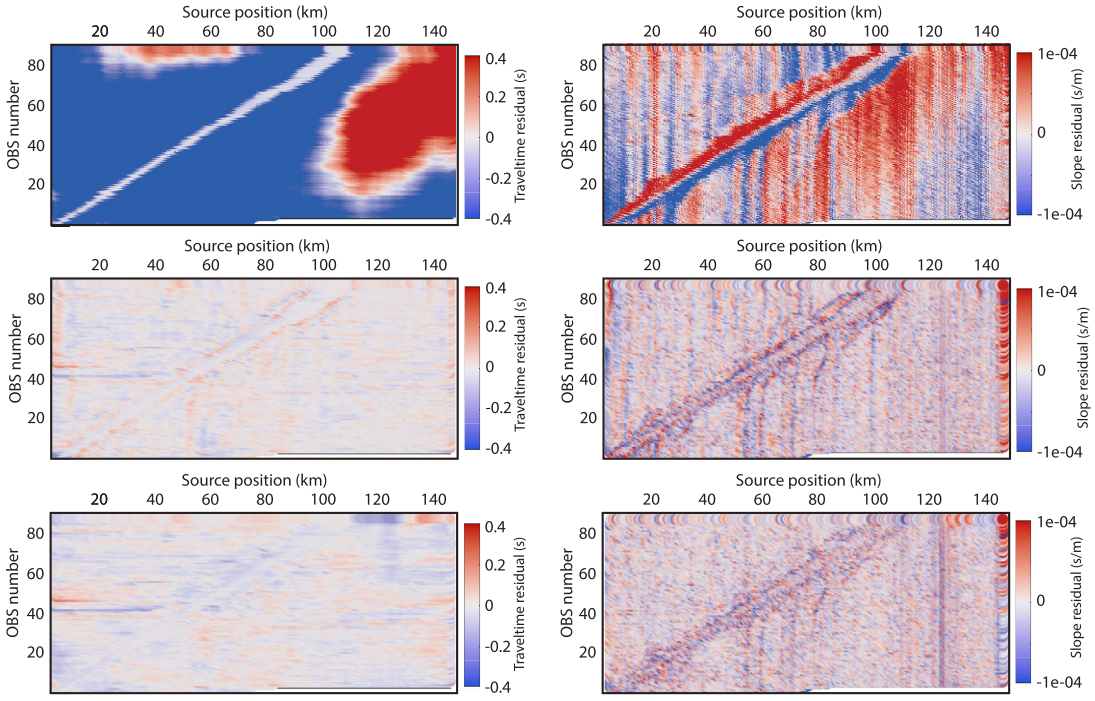

Figure 14. Nankai case study. Traveltime (left-hand panels) and slope (right-hand panels) misfit at the initial stage (top panels), post-FATT (middle panels) and post-FASTT (bottom panels).

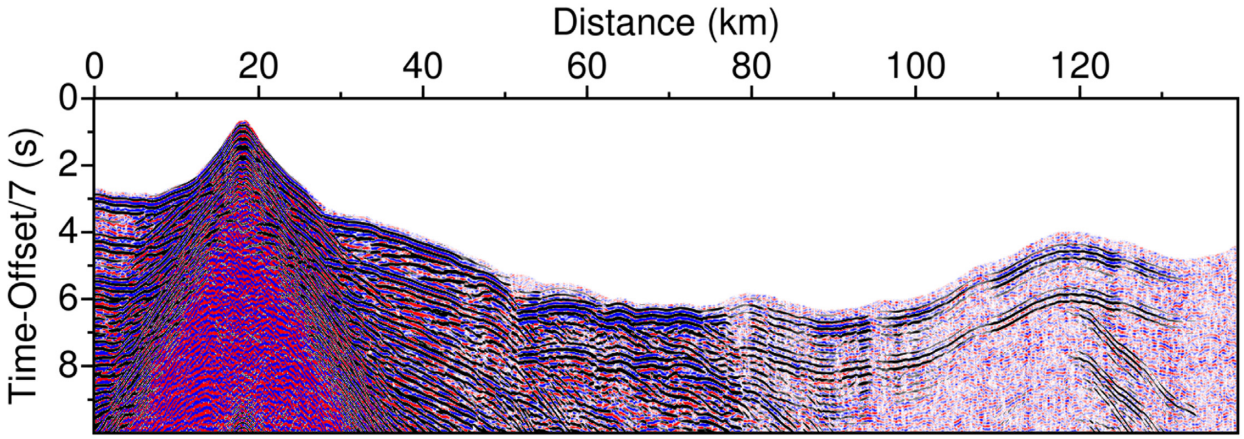

Figure 15. Nankai case study. OBS-17 seismogram in blue/red superimposed by a seismogram in black/transparent simulated at the same position in the FASTT+FWI model.

the top of the oceanic crust at around 10 km depth and dips ocean-ward in the opposite direction to the main tectonic trend highlighted by the landward-dipping faults affecting the backstop and the accretionary wedges (Fig. 12). We remind the reader that using a more adequate regularization at the cost of resolution is a suitable remedy as shown by the FATT results of Górszczyk *et al.* (2017, their fig. 5d). In contrast, the velocity trend in the FASTT model (Fig. 13c) looks more consistent with the structural dips interpreted in Fig. 12.

Without concluding on the validity of these models based upon prior geological knowledge and previous studies, we proceed with a frequency-domain acoustic Laplace-Fourier FWI (Shin & Cha 2009; Brossier *et al.* 2009) in the frequency band 1.5−8 Hz. We use the same workflow as the one used by Górszczyk *et al.* (2017), except that we use a constant density model for all inversions in order to associate the discrepancies in the results to wave speeds solely. We refer the reader to Górszczyk *et al.* (2017) for a detailed description on the triple-nested hierarchical management of frequencies, offsets and the Laplace constant. The FATT+FWI and FASTT+FWI inferred models are very similar (Figs 13d and e). Interestingly, the structures evoked previously exhibit a dip similar to the one seen in FASTT, hence validating the claim of the dip inconsistency in the FATT models. In fact, the layer-stripping approach along with the artificial low frequencies generated by aggressive time damping in the Laplace–Fourier domain render the FWI workflow more robust to initial guesses, hence recovering the sought structures even in the case of the initial FATT model. This experiment alludes that at the tomography stage, the use of the FATT model for migration or direct geological interpretation of the accretionary prism would be misleading. The detrended models (Figs 13f–i) support the interpretation made in this section. The kernel imprint in the case of FATT is even more pronounced (Fig. 13f) and the model was indeed corrected at the FWI level (Fig. 13h). The early arrivals waveform match between the real and the modeled seismograms at the end of the FWI process proves that the retrieved velocity model is reliable (Fig. 15).

4.2 Coarse acquisition results

We repeat the whole workflow presented above while using part of the data (11 OBS, approximately 10 km spacing) since in most crustal case studies the acquisition is coarser than the one used in the Nankai case study. The tomographic problem becomes more challenging due to the insufficient redundancy in the data and in turn the deficient illumination of the subsurface. We restart FATT and FASTT with a larger correlation length of the Gaussian filter on the gradient (1000 m × 1600 m), ensuring the validity of our asymptotic kernel. The tomography inferred models (Figs 16a and b) exhibit an aggravated version of the pathology seen in the full data case. The FATT model is polluted by strong kernel imprints below the backstop area and the accretionary wedge between 20 and 80 km distances (Fig. 16a), while the FASTT model is less affected (Fig. 16b). The trench fill beyond 85 km distance is not that affected in both cases since it is shallow, hence well illuminated (Fig. 1). In this case of sparse acquisition, the FWI results are more affected by the artefacts introduced through the initial models (Figs 16c and d). These differences are clearly highlighted by the detrended models shown in Figs 16(h) and (g). The most obvious artefacts take the form of patchy velocity updates in the FATT+FWI model that cross over the entire crust between 40 and 80 km distances. These artefacts impact dramatically the structure of the subducting slab as well as the geometry of the megathrust on top of it. In particular, significant differences between the megatrust reconstructed by the two FWI models are pointed by the arrows in Fig. 16 between 40 and 60 km distances. The area delineated by the dot rectangle in Fig. 16 highlights other discrepancies at shallower depths in the geometry of the thrusts and crustal sheets affecting the backstop. All these differences can indeed affect significantly the interpretation of the structural factors controlling rupture process of megathrust earthquakes.

4.2.1 Seismogram modelling

We examine the data of OBS-20 (Fig. 17a) and simulate time-domain seismograms in the four models of Fig. 16. The seismograms simulated in the tomography models (Figs 17b–c) embed a different level of complexity, hence highlighting different resolution of these models. Indeed,

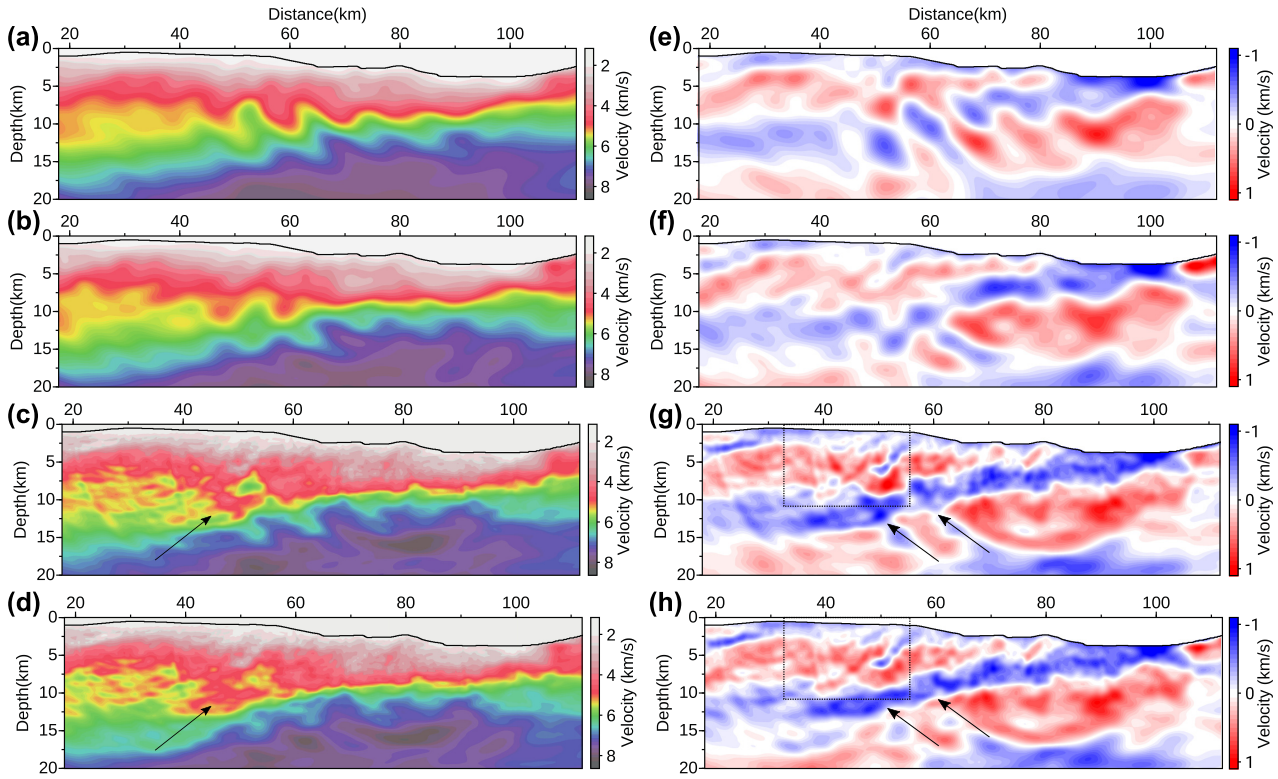

Figure 16. Nankai case study. Partial acquisition results. (a and b) Velocity models inferred from FATT and FASTT, respectively. (c and d) Final velocity model inferred from FWI using (a) and (b) as initial guesses, respectively. The black line in all of the panels delineates the bathymetry. (e–h) The detrended version of the inverted velocity models (a–d). The arrows and the dot rectangle point major differences between the FWI models inferred from FATT and FASTT.

the first-arrival in the FASTT model at 60 km is more complex than in the FATT model. In particular, it shows what is interpreted as a low-amplitude wave guided by the dipping structure in the accretionary prism (Figs 17b–c, black arrow). In fact, this wave may mimic a head wave trapped along the Tokai thrust. Other than the latter, a later reflection, possibly from the top of the subducting oceanic crust, missing in the case of FATT, is also detected in the seismograms computed in the FASTT model (Figs 17 b–c, grey arrow). This low-amplitude first-arrival followed by an energetic post-critical reflections are observed in the real OBS gather of Fig. 11 in the offset range 50 km-60 km. The reader is also referred to fig. S3d of Górszczyk *et al.* (2017) where ray tracing performed in the FWI velocity model highlights the trapping of the first-arrival rays along several thrusts slicing the backstop and the accretionary wedges. Looking at the post-FWI seismograms (Figs 17d–e), we notice that the suspected head wave is recorded in both cases, meaning that the inversion was going in the right direction and the FWI tried to rectify the shortcomings of the FATT model. Even though both FWI waveform look somehow similar from a kinematic point a view that is not the case from a dynamic view point. The amplitudes in the seismograms computed in the FASTT+FWI velocity model are mildly sharper than those of the FATT+FWI counterpart for some part of the early arrivals and more significantly for late reflections arrivals. This phenomenon attests that the FWI starting from the FATT initial model was converging in the right direction but was however late in comparison to its counterpart starting from the FASTT model (Fig. 17, white arrow).

4.2.2 Depth migration

As a further quality control of the tomographic model, we perform a pre-stack depth ray+Born inversion/migration (Thierry *et al.* 1999) of an aligned MCS profile using the FATT and FASTT models of Figs 16(a) and (b) as background models (Fig. 18). Comparing the migrated sections superimposed on the FATT and FASTT velocity models show unambiguously the improved reflectivity imaging obtained with the FASTT model. For example, the decollement highlighted by the black arrows in Fig. 18 is more continuous and can be followed over a larger range of distances in the migrated image inferred from the FASTT model. Moreover, this almost flat reflector complies more accurately with the smooth velocity variations of the FASTT model below the decollement, while this reflector crosses the unlikely wavy velocity variations of the FATT model. The almost seismically transparent high-velocity patch above the decollement of the FASTT model and the migrated reflectors delineating this velocity patch better match the interpreted duplex in Fig. 12 than the FATT counterpart. Several reflectors such as thrusts and ramp affecting the accretionary wedge clearly better comply with the velocity variations of the FASTT model (Fig. 18, grey arrows). Also, the reflectors in the sedimentary fill of the trench axis are generally more focused in the migrated section inferred from the FASTT model and better comply with the velocity variations of the latter. This is further supported by having a look at the angle-domain common image gathers (CIG) extracted at the trench fill position since that part is the most layered and is sufficiently illuminated using a

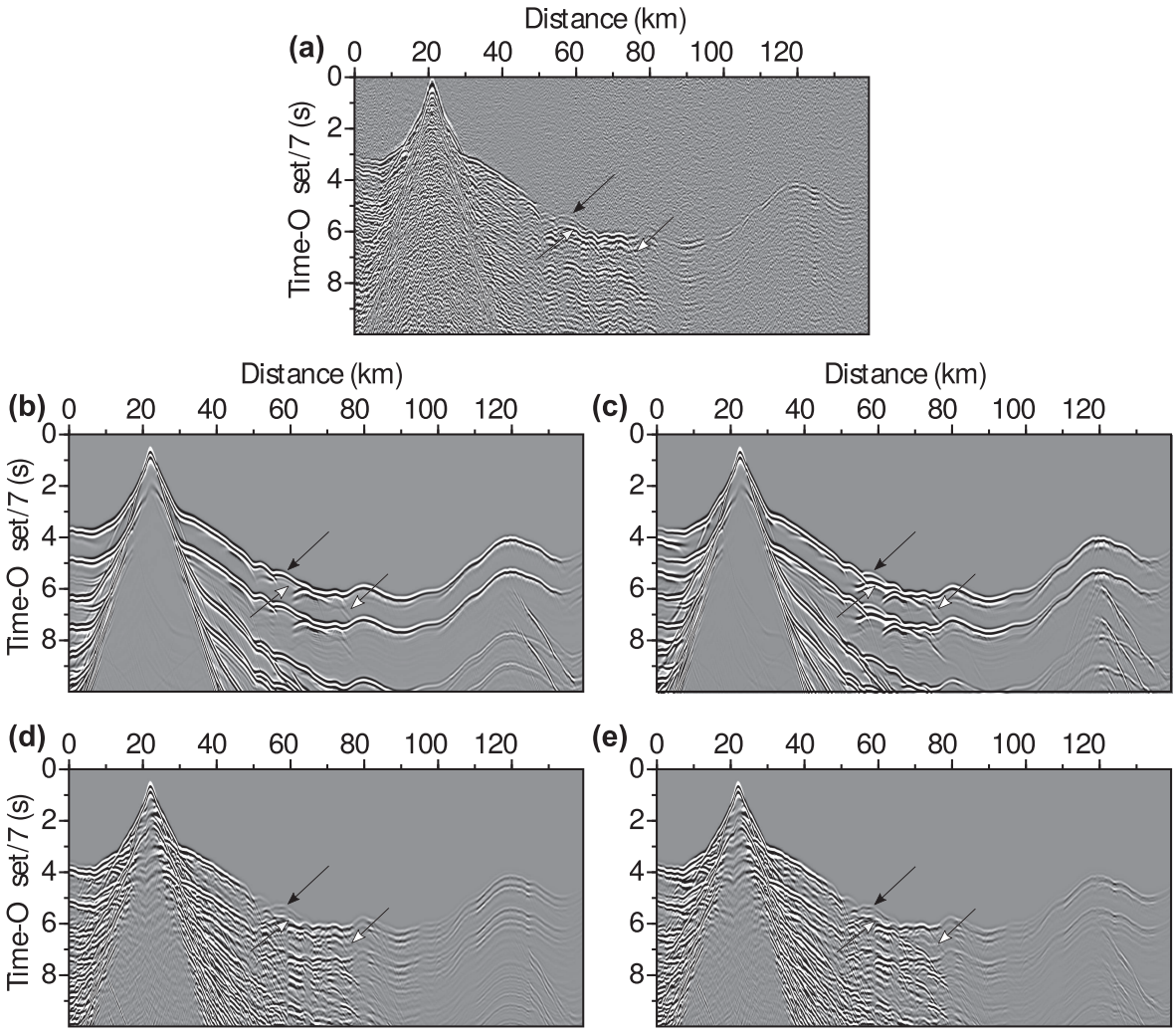

Figure 17. Nankai case study. Recorded OBS-20 seismogram (a). Simulated OBS-20 seismograms in the FATT (b), FASTT (c), FATT+FWI (d) and FASTT+FWI (e) models. The black arrow points a wave channeled along dipping structure in the accretionary prism while the grey arrow points a post-critical reflection from below (probably the top of the oceanic crust). The white arrow points contrasted amplitudes and focusing of a post-critical reflection in the seismograms computed in the FATT+FWI and FASTT+FWI models.

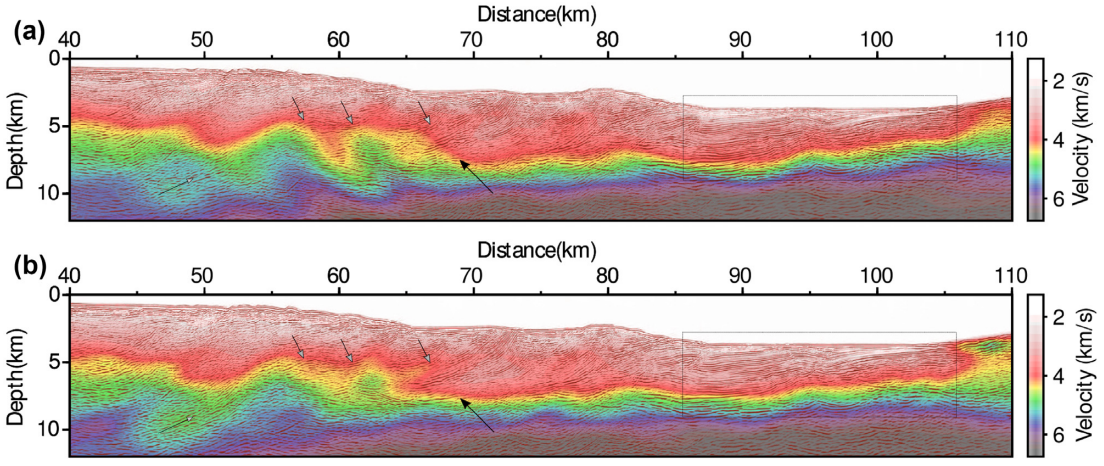

Figure 18. Nankai case study. Depth migrated images using the (a) FATT and (b) FASTT models of Figs 16(a) and (b) as background velocity models. The black arrows point the decollement on top of the subducting oceanic crust. The almost horizontal decollement intersects unlikely macrovelocity variations in the FATT model, while the velocities founded by FASTT comply more accurately with the geometry of the decollement. The grey arrows point the thrust and ramp that better comply with the velocity variations of the FASTT model in the accretionary wedge. The rectangle delineates the sedimentary fill of the trench axis where the reflectors are better focused in the migrated section computed with the FASTT model.

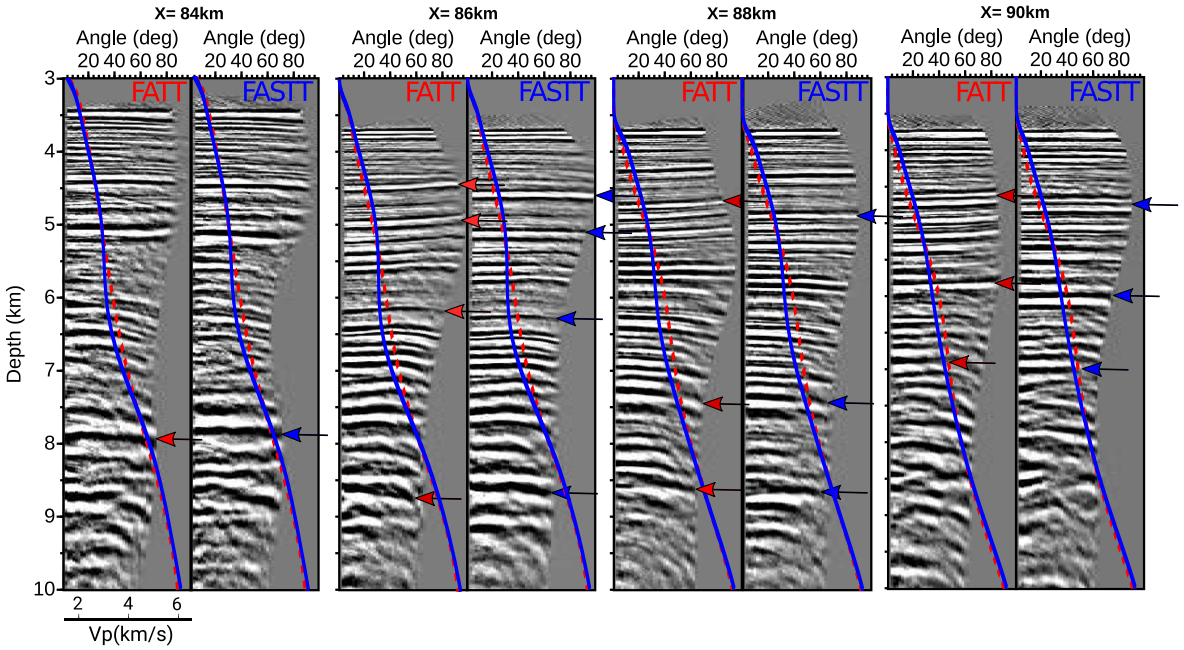

Figure 19. Nankai case study. Angle-domain common image gathers inferred by a pre-stack ray+Born inversion/migration using the FATT and FASTT models seen in Fig. 16, extracted at different positions in the trench fill. The solid red and dashed blue curves are vertical velocity profiles extracted at the same position as the CIGs from the FATT and FASTT, respectively. The arrows point at the main differences.

vintage 4.5 km streamer. The direct comparison of the CIGs (Fig. 19) flatness shows that the FASTT result is indeed more reliable in terms of kinematics. In order to relate the differences in the CIGs to the inverted velocity models (Figs 16a and b), vertical velocity profiles extracted at the same position as the CIGs (Fig. 19). The velocity profiles (Fig. 19) show an underestimation of velocity in the FATT model (Fig. 16, dashed red curve) compared to the FASTT model (Fig. 16, solid blue curve) from 3 to 5 km depth. The latter point is consistent with the upward curving and flat events seen in the CIGs at these depth in the case of FATT and FASTT, respectively. Looking at deeper sections than the trench fill, another difference is observed down to the top of the oceanic crust. The downward dipping events in the CIGs obtained using the FATT model reveal that the velocities are overestimated, contrarily to the FASTT case where the events are flat. In the deeper section down to the middle to lower oceanic crust, the events in the CIGs appear at slightly different depths and are not focused in the same manner. At these depths and in deeper parts, even though in the FASTT case the events seem more sharp, it become harder to interpret the results looking at the CIGs and the velocity profiles since wide-angle data were inverted at the velocity model building step while a limited-offset reflection data were used in the migration.

Finally, as a closure, we migrate the MCS data using the best FWI model presented in this study (Fig. 13e). We note that the use of a FWI derived model is the best option since the latter benefit from reflections and late arrivals. The final integrated imaging results are shown in Fig. 20 with three complementary styles of representation that highlight the different scales contained in the FWI model and the migrated image. Fig. 20(a) shows the FWI model together with the migrated section superimposed in transparency. Fig. 20(b) adds in transparency the velocity gradient of the FWI model (namely, the sum of the horizontal and vertical derivatives) to highlight the short-scale (migrated) components reconstructed by FWI. This style of representation delineates for example fairly well the top of the subducting oceanic crust and the Moho and hence ideally supplements at crustal depths the migrated image inferred from the MCS data. Compared to Fig. 20(a), Fig. 20(c) replaces the FWI model by its detrended version to highlight the structural units at intermediate scale reconstructed by FWI. This detrended representation style highlights the crustal sheets in the backstop as well as the sedimentary units in the accretionary wedges. These structural units comply fairly well with the short-scale reflectivity imaged by the migration of the MCS data (Figs 20b and c). Among the main features that can be easily interpreted: In the shallow part, the geometry of the forearc basin to the east (30 km distance) and the slope basin to the west (65–75 km distances) are fairly well delineated in both the migrated section and the FWI velocity model (Fig. 20a). In the backstop, albeit the penetration limitations induced by the vintage 4.5-km-long streamer, several migrated reflectors delineate the underplated crustal sheets down to 10 km depth (Fig. 20c). The ramps and flats characterizing the complex geometry of the Tokai thrust (Fig. 12) can be interpreted in both the MCS migrated section and the FWI velocity model (from 50 km distance where the Tokai thrust seems to branch upward from the megathrust to 65 km distance where it outcrops on the seafloor, Fig. 20a). This is however a quite complex area where we cannot preclude some significant 3-D effects as suggested by the enigmatic high-velocity perturbation shown at 60-km distance and 10-km depth in the detrended velocity model at the presumed location of the Palaeo-Zenisu ridge (Fig. 20c). Overall, the match between the reflectivity and the FWI velocities in the active wedge and the sedimentary trench fill is spectacular as highlighted by the imaging of a seamount-like structure draped by some layers at 95 km distance and overhung by a graben (Fig. 20). It should be also noted that the top of the slab in the western-most part of the model (40–50 km in distance to 10–15 km in depth) is probably not positioned in depth as accurately as in the results of Górszczyk *et al.*

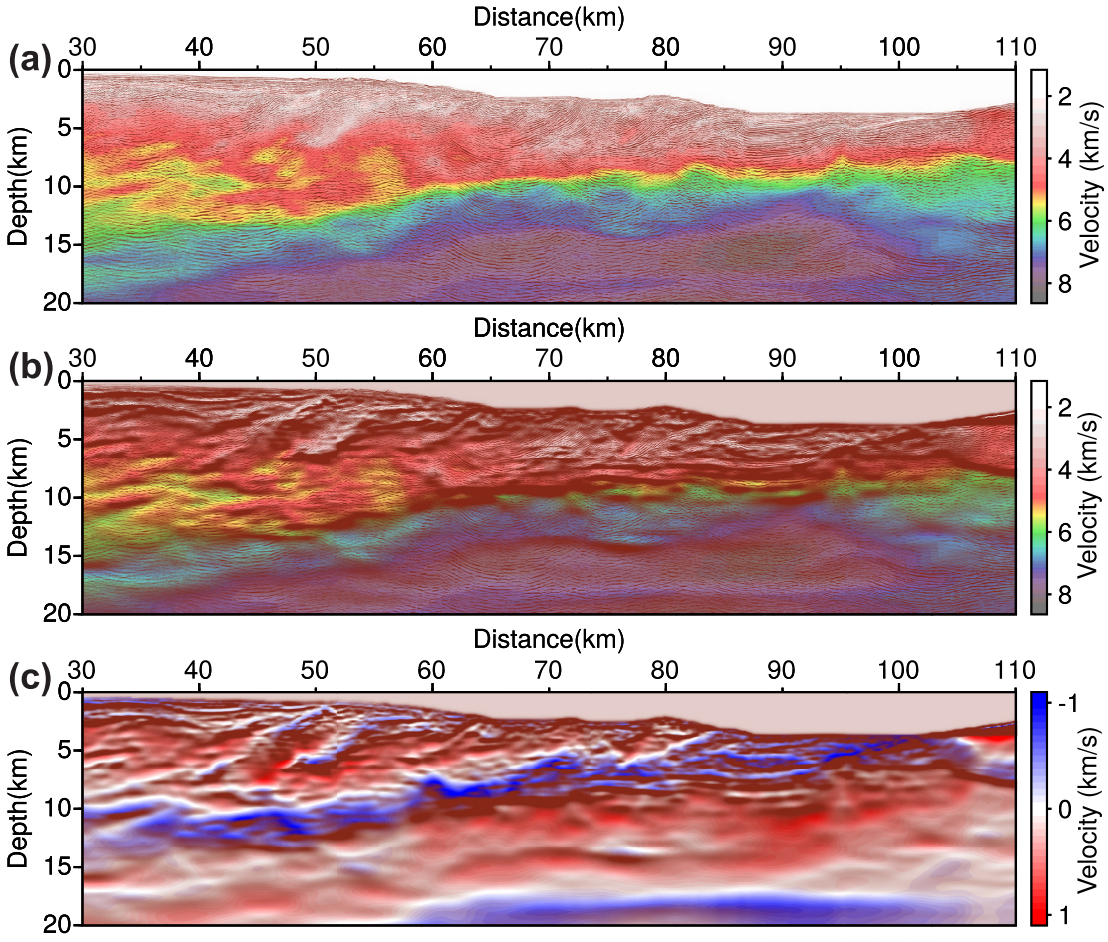

Figure 20. Nankai case study. (a) FASTT+FWI model superimposed by its corresponding ray+Born migrated image. (b) Same as (a) but the velocity gradient of the FWI model (the sum of the horizontal and vertical derivative) is also superimposed in transparency. This representation style highlights the short-scale (migrated) components of the FWI model. (c) The detrended version of the FWI model, superimposed by the velocity gradient of the FWI model, is shown to highlight the intermediate-scale structural units reconstructed by FWI.

(2017) with pull-up effects (For example, note the different depths (11 km and 11.6 km) of the low velocity zone at 42 km distance in Figs 12 and 20a). This mispositioning is inherited from the starting velocity model that we used for tomography (Fig. 13a), which does not contain any prior about the dip of the slab. In this deep poorly illuminated area close to the ends of the acquisition layout, FASTT was not able to fully solve the velocity–depth ambiguity in the spite of the added-value provided by slopes.

5 DISCUSSION

We propose a simple approach to mitigate the ill-posedness of first-arrival traveltime tomography by adding slopes at sources and/or receivers to traveltimes as optimization measurements. Improved results are obtained by better implicitly constraining the starting and ending incidence angles of the rays connecting a source to a receiver, and hence the turning points of the diving rays. This is useful to mitigate the non-uniqueness of the solution of FATT when the first-arrival rays are blind to the lower part of crustal layers as illustrated in Fig. 1. We also show that using slopes leads to velocity models which have a higher resolution than the counterpart built by FATT. This improved resolution results from the fact that differential traveltimes are more sensitive to the velocity gradients. We also show that the added-value provided by slopes in first-arrival tomography increases as the acquisition is coarser without the need of aggressive regularization.

Slopes measurements can be easily inferred by finite differences once the first breaks have been picked. Alternatively, automatic picking tools classically used in reflection slope tomography to pick locally coherent events can be used. These tools rely on local slant stacks to automatically track locally coherent events in common-shot or common-receiver gathers (Taner *et al.* 1979; Billette *et al.* 2003; Lambaré 2008). While slopes at sources and receivers are necessary to implement reflection slope tomography, 2-D first-arrival slope tomography can be performed with only one slope since a first-arrival ray is unambiguously defined in a given velocity model by the source and receiver positions. This allows one to readily apply FASTT to sparse stationary-recording surveys such as OBS surveys. When considering sparse 3-D OBS surveys, the source dimensions may be downsampled in the cross direction preventing the picking of the azimuth angle. However, it is

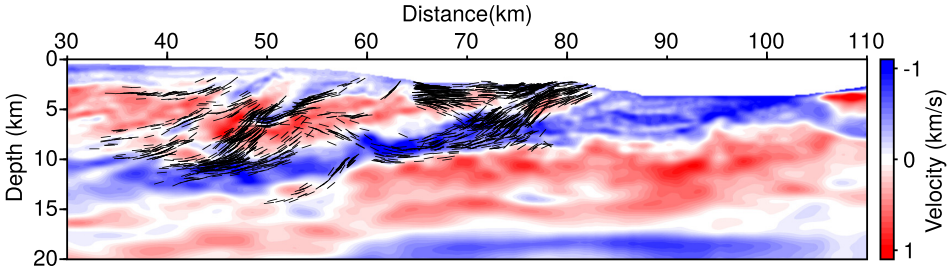

Figure 21. An illustration of kinematic migration with the Nankai case study. A small set of secondary arrivals were picked and migrated kinematically by looking at the intersection between the isochrone defined by the two-way traveltime and the ray leaving the shot position with the picked slope. The located scatterer is plotted as a migration facet the dip of which is tangent to the isochrone at the scatterer location.

likely that using only one slope at shot positions will still be beneficial to perform 3-D FASTT. An open issue is the measurement of slopes at OBS positions when the acquisition is sparse. This is beneficial to involve the receiver slope in FASTT or to combine FASTT with reflection slope tomography. One possible solution is to estimate the incidence and azimuth angles at OBS positions from the three components of the OBS by polarization analysis, with the open question of the measurement accuracy (Hu & Menke 1992; Hu *et al.* 1994). Alternatively, the reciprocity principle can be used to estimate the slope at a given OBS position using the records of the other OBSs triggered by the shots located at the vertex of the targeted instrument (Alerini 2006; Alerini *et al.* 2009). It should also be noted that in the case of rugged topographies often encountered in onshore case studies, slopes are corrected according to the undulant surface locally by performing an analysis using the horizontal and vertical component of the slowness vectors (Jin & Zhang 2018).

The slopes of late arrivals (namely, any single-scattered arrival and surface multiples) at shot positions can be used also to perform a kinematic migration using the FASTT or FWI models as background model. The aim of kinematic migration is to locate a scatterer in the subsurface at the intersection between the isochrone defined by the two-way traveltime and the ray leaving the shot (or receiver) position with the measured slope (Chauris *et al.* 2002; Sambolian *et al.* 2019). This kinematic migration builds a skeleton of the structure that may be useful to guide a line drawing for structural interpretation or as a quality control of the background velocity model. This can provide also a useful tool to clarify the origin of ambiguous arrivals recorded in complex geological environment and better understand all the arrivals that are involved in FWI. To illustrate how this might work, Fig. 21 shows a set of migration facets located in the final FWI model. It should be noted that the picking was performed in synthetic seismograms computed in the FWI velocity model rather than in the real data. One can see that the dip of the migrated facets comply with the dip of several thrusts in the backstop and the accretionary wedges, a splay fault branching upward from the plate boundary, and several horizons in the sedimentary slope basin. Finally, several facets with antithetic dips are shown in the enigmatic area at the position of the presumed palaeo-Zenisu ridge (60 km distance) supporting the presence of a major tectonic feature at this position. It is also worth noting a fault in the subducting slab located seaward to this feature, which might be similar to those observed seaward to the Zenisu ridge (Mazzotti *et al.* 2002).

6 CONCLUSIONS

We propose the use of slopes as an additional objective measure in the context of first-arrival traveltime tomography as a remedy for the ill-posedness of the problem. We present the formulation of the problem under the framework based on eikonal solvers and the adjoint-state method. We bring forward the added value of slopes and explain how they constrain better the ray path and thus lead to a more stable inversion. In addition to a toy test based on analytical expressions, we defend our claims on two full-scale applications. We elaborate on the deficiency of the traveltimes and illustrate its impact on the resulting structurally misleading models opposed to the reliable ones inferred through our proposed strategy. In future works, the proposed scheme is easily embedded in our reflection slope tomography method (Sambolian *et al.* 2019). We will investigate cases where both reflection and first-arrival are available especially in a challenging OBS context where reflection slopes are not straightforwardly accessible on the receiver side. Evidently, our joint approach is perfectly applicable in cases where OBS and streamer data are coupled and will be presented in upcoming studies.

ACKNOWLEDGEMENTS

This study was granted access to the HPC resources of SIGAMM infrastructure (http://crimson.oca.eu), hosted by Observatoire de la Côte d'Azur and which is supported by the Région Provence-Alpes-Côte d'Azur, and the HPC resources of CINES/IDRIS/TGCC under the allocation A0050410596 made by GENCI. We thank JAMSTEC for providing us with the SFJ-OBS data. Many thanks to Prof. Jean Virieux (ISTerre, Grenoble) for fruitful discussions. The Editor Prof. Frederik Simons and Prof. Jun Korenaga are gratefully acknowledged for their valuable feedback during the review process.

DATA AVAILABILITY

The access to the SFJ-OBS data set can be negotiated with JAMSTEC and is accessible through their database (http://www.jamstec.go.jp/obsmcs_db/e/index.html).

REFERENCES

Ajo-Franklin, J.B., Minsley, B.J. & Daley, T.M., 2007. Applying compactness constraints to differential traveltime tomography, *Geophysics,* **72**(4), R67–R75.

Akçelik, V., 2002. Multiscale Newton-Krylov methods for inverse acoustic wave propagation, *PhD thesis*, Carnegie Mellon University, Pittsburgh, PA.

Aki, K. & Lee, W., 1976. Determination of three-dimentional velocity anomalies under a seismic array using first p-arrival times from local earthquakes, *J. geophys. Res.,* **81**(4), 4381–4399.

Alerini, M., 2006. Stereotomography for nodes data, in *Proceedings of the 68nd Annual Meeting,* Extended Abstracts B043, 12–15 June 2006, Vienna, Eur. Ass. Geosc. Eng.

Alerini, M., Traub, B., Ravaut, C. & Duveneck, E., 2009. Prestack depth imaging of ocean-bottom node data, *Geophysics,* **74**(6), WCA57–WCA63.

Alkhalifah, T., 1998. Acoustic approximations for processing in transversely isotropic media, *Geophysics,* **63,** 623–631.

Alkhalifah, T. & Fomel, S., 2010. An eikonal based formulation for traveltime perturbation with respect to the source location, *Geophysics,* **75**(6), T175–T183.

Alkhalifah, T. & Tsvankin, I., 1995. Velocity analysis for transversely isotropic media, *Geophysics,* **60,** 1550–1566.

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

Ammon, C. & Vidale, J., 1993. Tomography without rays, *Bull. seism. Soc. Am.,* **83**(2), 509–528.

Ando, M., 1975. Possibility of a major earthquake in the Tokai district, Japan and its pre-estimated seismotectonic effects, *Tectonophysics,* **27,** 69–85.

Bauer, A., Schwarz, B. & Gajewski, D., 2017. Utilizing diffractions in wavefront tomography, *Geophysics,* **82**(2), R65–R73.

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.

Brossier, R., Operto, S. & Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics,* **74**(6), WCC105–WCC118.

Brown, P.N., 1987. A local convergence theory for combined inexact-newton/finite-difference projection methods, *SIAM J. Numer. Anal.,* **24**(2), 407–434.

Byrd, R.H., Lu, P. & Nocedal, J., 1995. A limited memory algorithm for bound constrained optimization, *SIAM J. Scient. Stat. Comput.,* **16,** 1190–1208.

Chauris, H., Noble, M., Lambaré, G. & Podvin, P., 2002. Migration velocity analysis from locally coherent events in 2-D laterally heterogeneous media, Part I: theoretical aspects, *Geophysics,* **67**(4), 1202–1212.

Curtis, A. & Robertsson, J. O.A., 2002. Volumetric wavefield recording and wave equation inversion for near-surface material properties, *Geophysics,* **67**(5), 1602–1611.

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

de Ridder, S.A.L. & Biondi, B.L., 2015. Near-surface Scholte wave velocities at Ekofisk from short noise recordings by seismic noise gradiometry, *Geophys. Res. Lett.,* **42**(17), 7031–7038.

Delprat-Jannaud, F. & Lailly, P., 1992. What information on the earth model do reflection travel times provide?, *J. geophys. Res.,* **97**(B13), 19 827–19 844.

Dessa, J.X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Uhira, K. & Kaneda, Y., 2004a. Deep seismic imaging of the eastern Nankai trough (Japan) from multifold ocean bottom seismometer data by combined traveltime tomography and prestack depth migration, *J. geophys. Res.,* **109**(B02111), doi:10.1029/2003JB002689.

Dessa, J.X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Virieux, J. & Kaneda, Y., 2004b. Multiscale seismic imaging of the eastern nankai trough by full waveform inversion, *Geophys. Res. Lett.,* **31**(L18606), doi:10.1029/2004GL020453.

Duveneck, E., 2004. Velocity model estimation with data-derived wavefront attributes, *Geophysics,* **69**(1), 265–274.

Fichtner, A. & Trampert, J., 2011. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.,* **185**(2), 775–798.

Fomel, S., Luo, S. & Zhao, H.-K., 2009. Fast sweeping method for the factored eikonal equation, *J. Comput. Phys.,* **228,** 6440–6455.

Gelchinsky, B., Berkovitch, A. & Keydar, S., 1999. Multifocusing homeomorphic imaging Part 1. Basic concepts and formulas, *J. appl. Geophys.,* **42,** 229–242.

Gholami, Y., Brossier, R., Operto, S., Ribodetti, A. & Virieux, J., 2013. Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 1: sensitivity and trade-off analysis, *Geophysics,* **78**(2), R81–R105.

Górszczyk, A., Operto, S. & Malinowski, M., 2017. Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern nankai trough revisited, *J. geophys. Res.,* **122**(6), 4601–4630.

Górszczyk, A., Operto, S., Schenini, L. & Yamada, Y., 2019. Crustal-scale depth imaging via joint FWI of OBS data and PSDM of MCS data: a case study from the eastern nankai trough, *Solid Earth,* **10,** 765–784.

Haber, E., Ascher, U.M. & Oldenburg, D., 2000. On optimization techniques for solving nonlinear inverse problems, *Inverse Problems,* **16**(5), 1263.

Henry, P. *et al.*, 2004. Deformation processes and earthquakes in Nankai, in *Proceedings of 2004 GeoHazards Workshop*.Shizuoka, Japan

Hicks, G.J., 2002. Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions, *Geophysics,* **67,** 156–166.

Hole, J.A., 1992. Nonlinear high-resolution three-dimensional seismic travel time tomography, *J. geophys. Res.,* **97,** 6553–6562.

Hu, G. & Menke, W., 1992. Formal inversion of laterally heterogeneous velocity structure from P-wave polarization data, *Geophys. J. Int.,* **110,** 63–69.

Hu, G., Menke, W. & Powell, C., 1994. Polarization tomography for P wave velocity structure in southern california, *J. geophys. Res.,* **99,** 15 245–15 256.

Jin, C. & Zhang, J., 2018. Stereotomography of seismic data acquired on undulant topography, *Geophysics,* **83**(4), U35–U41.

Kelley, C., 1995. *Iterative Methods for Linear and Nonlinear Equations,* SIAM.

Knoll, D. & Keyes, D., 2004. Jacobian-free Newton–Krylov methods: a survey of approaches and applications, *J. Comput. Phys.,* **193**(2), 357–397.

Korenaga, J., Holbrook, W.S., Singh, S.C. & Minshull, T.A., 1997. Natural gas hydrates on the southeast U.S. margin: Constraints from full waveform and traveltime inversions of wide-angle seismic data, *J. geophys. Res.,* **102**(B7), 15 345–15 365.

Korenaga, J., Holbrook, W.S., Kent, G.M., Kelemen, P.B., Detrick, R.S., Larsen, H.C., Hopper, J.R. & Dahl-Jensen, T., 2000. Crustal structure of the southeast Greenland margin from joint refraction and reflection seismic tomography, *J. geophys. Res.,* **105,** 21 591–21 614.

Lambaré, G., 2008. Stereotomography, *Geophysics,* **73**(5), VE25–VE34.

Langston, C.A., 2007. Spatial gradient analysis for linear seismic arrays, *Bull. seism. Soc. Am.,* **97**(1B), 265–280.

Le Pichon, X., Lallemant, S., Tokuyama, H., Thoué, F., Huchon, P. & Henry, P., 1996. Structure and evolution of the backstop in the eastern Nankai trough area (Japan): implications for the soon-to-come Tokai earthquake, *Island Arc,* **5,** 440–454.

Leung, S. & Qian, J., 2006. An adjoint state method for three-dimensional transmission traveltime tomography using first-arrivals, *Commun. Math. Sci.,* **4**(1), 249–266.

Luo, S. & Qian, J., 2012. Fast sweeping method for factored anisotropic eikonal equations: multiplicative and additive factors, *J. Scient. Comput.,* **52,** 360–382.

Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics,* **56**(5), 645–653.

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

Mazzotti, S., Lallemant, S., Henry, P., Pichon, X.L., Tokuyama, H. & Takahashi, N., 2002. Intraplate shortening and underthrusting of a large basement ridge in the eastern Nankai subduction zone, *Mar. Geol.,* **187,** 63–88.

Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press, Inc.

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., Operto, S. & Virieux, J., 2017. Full waveform inversion and the truncated Newton method, *SIAM Rev.,* **59**(1), 153–195.

Monteiller, V., Got, J.-L., Virieux, J. & Okubo, P., 2005. An efficient algorithm for double-difference tomography and location in heterogeneous media, with an application to the Kilauea volcano, *J. geophys. Res.,* **110**(B12306), doi:10.1029/2004JB00346.

Nash, S.G., 2000. A survey of truncated Newton methods, *J. Comput. Appl. Math.,* **124,** 45–59.

Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization,* 2nd edn, Springer.

Operto, S., Virieux, J., Dessa, J.X. & Pascal, G., 2006. Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough, *J. geophys. Res.,* **111**(B09306), doi:10.1029/2005JB003835.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**(2), 495–503.

Plessix, R.E. & Cao, Q., 2011. A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium, *Geophys. J. Int.,* **185,** 539–556.

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

Pratt, R.G., 2008. Waveform tomography - successes, cautionary tales, and future directions, in *Presented at the 70th Annual EAGE Conference & Exhibition,* Roma, WO11 – Full–Waveform Inversion: current status and perspectives.

Prieux, V., Lambaré, G., Operto, S. & Virieux, J., 2013. Building starting model for full waveform inversion from wide-aperture data by stereotomography, *Geophys. Prospect.,* **61**(Issue supplement: 60 year anniversary issue), 109–137.

Qian, J. & Symes, W., 2002. An adaptive finite-difference method for traveltimes and amplitudes, *Geophysics,* **67,** 167–176.

Sambolian, S., Operto, S., Ribodetti, A., Tavakoli, B. & Virieux, J., 2019. Parsimonious slope tomography based on eikonal solvers and the adjoint-state method, *Geophys. J. Int.,* **218**(1), 456–478.

Sambolian, S., Operto, S., Ribodetti, A. & Virieux, J., 2021. Consistent seismic event location and subsurface parameters inversion through slope tomography: a variable-projection approach, *Geophys. J. Int.,* **224**(3), 1956–1979.

Shen, L., Ahmed, I., Brenders, A., Dellinger, J., Etgen, J. & Michell, S., 2018. Full-waveform inversion: the next leap forward in subsalt imaging, *Leading Edge,* **37**(1), 2–80.

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

Shin, C. & Cha, Y.H., 2009. Waveform inversion in the Laplace-Fourier domain, *Geophys. J. Int.,* **177,** 1067–1079.

Shin, C., Jang, S. & Min, D.J., 2001. Improved amplitude preservation for prestack depth migration by inverse scattering theory, *Geophys. Prospect.,* **49,** 592–606.

Snieder, R. & Lomax, A., 1996. Wavefield smoothing and the effect of rough velocity perturbations on arrival times and amplitudes, *Geophys. J. Int.,* **125,** 796–812.

Stovas, A. & Alkhalifah, T., 2014. Analytical approximations of diving-wave imaging in constant-gradient medium, *Geophysics,* **79**(4), S131–S140.

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.

Taner, M.T., Koehler, F. & Sheriff, R.E., 1979. Complex seismic trace analysis, *Geophysics,* **44,** 1041–1063.

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,

Tarantola, A. & Nercessian, A., 1984. Three-dimensional inversion without blocks, *Geophys. J. R. astr. Soc.,* **769,** 299–306, doi.org/10.1111/j.1365-246X.1984.tb05047.x.

Tavakoli F., B., Operto, S., Ribodetti, A., Sambolian, S. & Virieux, J., 2018. Anisotropic first-arrival slope and traveltime tomography (FASTT), in *Proceedings of the 80th EAGE Conference and Exhibition 2018,* June 2018, Vol. 2018, pp.1–5

Tavakoli F., B., Operto, S., Ribodetti, A. & Virieux, J., 2019. Matrix-free anisotropic slope tomography: theory and application, *Geophysics,* **84**(1), R35–R57.

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,* **687,** 3576–3581.

Tavakoli F., B., Operto, S., Ribodetti, A. & Virieux, J., 2017a. Anisotropic slope tomography based on Eikonal solver and adjoint-state method, in *Proceedings of the 79th EAGE Conference and Exhibition 2017,* Paris.

Tavakoli F., B., Operto, S., Ribodetti, A. & Virieux, J., 2017b. Slope tomography based on eikonal solvers and the adjoint-state method, *Geophys. J. Int.,* **209**(3), 1629–1647.

Thierry, P., Lambaré, G., Podvin, P. & Noble, M., 1999. 3-D preserved amplitude prestack depth migration on a workstation, *Geophysics,* **64**(1), 222–229.

Thomsen, L.A., 1986. Weak elastic anisotropy, *Geophysics,* **51,** 1954–1966.

Tikhonov, A.N., 1963. Resolution of ill-posed problems and the regularization method (in Russian, French translation, Mir, Moscow, 1976), *Dokl. Akad. Nauk SSSR,* **151,** 501–504.

Trinks, I., Singh, S., Chapman, C., Barton, P. & Cherett, A., 2003. High-resolution traveltime and slowness tomography, in *Expanded Abstracts,* pp. 742–745.

Udías, A., 2000. *Principles of Seismology,* Cambridge Univ. Press.

Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics,* **74**(6), WCC1–WCC26.

Waheed, U.B., Yarman, C.E. & Flagg, G., 2014. An iterative fast sweeping based eikonal solver for tilted orthorhombic media, in *Expanded Abstracts,* pp. 480–485, Society of Exploration Geophysics.

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.

Woodward, M.J., 1992. Wave-equation tomography, *Geophysics,* **57,** 15–26.

Yuan, Y.O., Simons, F.J. & Tromp, J., 2016. Double-difference adjoint tomography, *Geophys. J. Int.,* **206**(3), 1599–1618.

Zelt, C. & Barton, P.J., 1998. Three-dimensional seismic refraction tomography: a comparison of two methods applied to data from the Faeroe basin, *J. geophys. Res.,* **103**(B4), 7187–7210.
Zelt, C. & Smith, R.B., 1992. Seismic traveltime inversion for 2-D crustal velocity structure, *Geophys. J. Int.,* **108,** 16–34.
Zelt, C.A., 1999. Modelling strategies and model assessment for wide-angle seismic traveltime data, *Geophys. J. Int.,* **139**(1), 183–204.
Zelt, C.A. & Chen, J., 2016. Frequency-dependent traveltime tomography for near-surface seismic refraction data, *Geophys. J. Int.,* **207,** 72–88.
Zhang, J. & Toksöz, M.N., 1998. Nonlinear refraction traveltime tomography, *Geophysics,* **63**(5), 1726–1737.
Zhang, J., ten Brink, U.S. & Toksöz, M.N., 1998. Nonlinear refraction and reflection travel time tomography, *J. geophys. Res.,* **103**(B12), 29 743–29 757.
Zhao, H., 2005. A fast sweeping method for eikonal equations, *Math. Comput.,* **74,** 603–627.
Zhu, X. & McMechan, G.A., 1989. Estimation of a two-dimensional seismic compressional-wave velocity distribution by iterative tomographic imaging, *Int. J. Imag. Syst. Technol.,* **1**(1), 13–17.

APPENDIX: SECOND-ORDER ADJOINT-STATE FORMULATION

In the following, we develop the formulation of FASTT using the second-order recipe of the adjoint-state method (Fichtner & Trampert 2011; Métivier *et al.* 2017). In a truncated newton scheme (Nash 2000), we seek the solution of the Hessian-vector product, in order to solve the Newton system with the linear conjugate gradient method. We define the functional $h_w(\mathbf{m})$ with w being an arbitrary vector as follows:

$$h_w(\mathbf{m}) = \Big\langle \nabla C(\mathbf{m}) \mid w \Big\rangle. \tag{A1}$$

By definition the gradient of the functional expressed above is

$$\nabla h_w(\mathbf{m}) = H(\mathbf{m})w. \tag{A2}$$

Accordingly, computing the Hessian-vector product amounts to computing the gradient of the functional $h_w(\mathbf{m})$. We proceed with the computation of the Hessian-vector product $H(\mathbf{m})w$ under a Lagrangian formalism. The Lagrangian operator $\mathcal{L}_w$ is associated with eleven state variables (the five state variables and their associated adjoint-state variables of the first-order adjoint and the gradient g) and their respective adjoint-state variables denoted by α. We express the gradient for velocity explicitly in the formulation since it's the sole parameter (isotropic cases). The augmented functional is expressed as

$$\begin{aligned}
&\mathcal{L}_w(\mathbf{v}, \mathbf{g}, \mathbf{p_{s_r}}, \mathbf{p_{r_s}}, \mathbf{T_{s,r}}, \mathbf{t_s}, \mathbf{t_r}, \xi_{\mathbf{s_r}}, \xi_{\mathbf{r_s}}, \mu_{\mathbf{s,r}}, \lambda_{\mathbf{s}}, \lambda_{\mathbf{r}}, \alpha_{\mathbf{1}}, \alpha_{\mathbf{2_{s_r}}}, \alpha_{\mathbf{3_{r_s}}}, \alpha_{\mathbf{4_{s,r}}}, \alpha_{\mathbf{5_s}}, \alpha_{\mathbf{6_r}}, \alpha_{\mathbf{7_{s_r}}}, \alpha_{\mathbf{8_{r_s}}}, \alpha_{\mathbf{9_{s,r}}}, \alpha_{\mathbf{10_s}}, \alpha_{\mathbf{11_r}}) = \\
&\Big\langle g \mid w \Big\rangle - \sum_{s=1}^{N_s}\sum_{r=1}^{N_r} \alpha_1(\boldsymbol{x}) \left(g + \frac{\lambda_s(\boldsymbol{x})}{v(\boldsymbol{x})^3} + \frac{\lambda_r(\boldsymbol{x})}{v(\boldsymbol{x})^3} \right) \\
&- \sum_{s=1}^{N_s}\sum_{r=1}^{N_r} \alpha_{2_{s_r}} \left(p_{s_r} - \frac{1}{2h}(Q(\mathbf{x}-\mathbf{s}^+) - Q(\mathbf{x}-\mathbf{s}^-))t_r(\boldsymbol{x}) \right) \\
&- \sum_{r=1}^{N_r}\sum_{s=1}^{N_s} \alpha_{3_{r_s}} \left(p_{r_s} - \frac{1}{2h}(Q(\mathbf{x}-\mathbf{r}^+) - Q(\mathbf{x}-\mathbf{r}^-))t_s(\boldsymbol{x}) \right) \\
&- \sum_{s=1}^{N_s}\sum_{r=1}^{N_r} \alpha_{4_{s,r}} \left(T_{s,r} - \frac{1}{2}(Q(\mathbf{x}-\mathbf{r})t_s(\boldsymbol{x}) + Q(\mathbf{x}-\mathbf{s})t_r(\boldsymbol{x})) \right) \\
&- \frac{1}{2}\sum_{s=1}^{N_s} \Big\langle \alpha_{5_s}(\boldsymbol{x}) \mid H(\boldsymbol{x}, \nabla t_s(\boldsymbol{x})) \Big\rangle_\Omega - \frac{1}{2}\sum_{r=1}^{N_r} \Big\langle \alpha_{6_r}(\boldsymbol{x}) \mid H(\boldsymbol{x}, \nabla t_r(\boldsymbol{x})) \Big\rangle_\Omega \\
&- \sum_{s=1}^{N_s}\sum_{r=1}^{N_r} \alpha_{7_{s_r}} \left(\xi_{s_r} - \frac{\Delta p_{s_r}}{\sigma^2_{p_s}} \right) \\
&- \sum_{r=1}^{N_r}\sum_{s=1}^{N_s} \alpha_{8_{r_s}} \left(\xi_{r_s} - \frac{\Delta p_{r_s}}{\sigma^2_{p_r}} \right) \\
&- \sum_{s=1}^{N_s}\sum_{r=1}^{N_r} \alpha_{9_{s,r}} \left(\mu_{s,r} - \frac{\Delta T_{s,r}}{\sigma^2_{T_{s,r}}} \right) \\
&- \sum_{s=1}^{N_\mathbf{s}} \alpha_{10_s}(\boldsymbol{x}) \left(\Big(\nabla \cdot (\lambda_s(\boldsymbol{x}) \nabla t_s(\boldsymbol{x})) \Big)_\Omega - \frac{1}{2}\sum_{r=1}^{N_r} \Big(Q(\mathbf{x}-\mathbf{r})^t \ \mu_{s,r} + (Q(\mathbf{x}-\mathbf{r}^+)^t - (Q(\mathbf{x}-\mathbf{r}^-)^t) \frac{\xi_{r_s}}{h} \Big) \right) \\
&- \sum_{r=1}^{N_\mathbf{r}} \alpha_{11_r}(\boldsymbol{x}) \left(\Big(\nabla \cdot (\lambda_r(\boldsymbol{x}) \nabla t_r(\boldsymbol{x})) \Big)_\Omega - \frac{1}{2}\sum_{s=1}^{N_s} \Big(Q(\mathbf{x}-\mathbf{s})^t \ \mu_{s,r} + (Q(\mathbf{x}-\mathbf{s}^+)^t - (Q(\mathbf{x}-\mathbf{s}^-)^t) \frac{\xi_{s_r}}{h} \Big) \right).
\end{aligned} \tag{A3}$$

In the same manner as the first-order formulation with the exception of the traveltimes being written in function of both traveltime maps for the sake of symmetry/clarity in the terms, the partial derivative of the new augmented functional with respect to every new state variable are

zeroed. Starting by the gradient g:

$$\frac{\partial \mathcal{L}_w}{\partial g} = 0 \;\rightarrow\; \boxed{\alpha_1 = w}. \tag{A4}$$

Followed by the objective parameters:

$$\frac{\partial \mathcal{L}_w}{\partial p_{s_r}} = 0 \;\rightarrow\; \boxed{\alpha_{2_{s_r}} = \frac{\alpha_{7_{s_r}}}{\sigma^2_{p_s}}}, \tag{A5}$$

$$\frac{\partial \mathcal{L}_w}{\partial p_{r_s}} = 0 \;\rightarrow\; \boxed{\alpha_{3_{r_s}} = \frac{\alpha_{8_{r_s}}}{\sigma^2_{p_r}}}, \tag{A6}$$

$$\frac{\partial \mathcal{L}_w}{\partial T_{s,r}} = 0 \;\rightarrow\; \boxed{\alpha_{4_{s,r}} = \frac{\alpha_{9_{s,r}}}{\sigma^2_{T_{s,r}}}}, \tag{A7}$$

then with respect to the first-order adjoint-state variables:

$$\frac{\partial \mathcal{L}_w}{\partial \xi_{s_r}} = 0 \;\rightarrow\; \boxed{\alpha_{7_{s_r}} = -\frac{1}{2h}\Big((Q(\mathbf{x}-\mathbf{s}^+) - Q(\mathbf{x}-\mathbf{s}^-))\alpha_{11_r}(\boldsymbol{x})\Big)}, \tag{A8}$$

$$\frac{\partial \mathcal{L}_w}{\partial \xi_{r_s}} = 0 \;\rightarrow\; \boxed{\alpha_{8_{r_s}} = -\frac{1}{2h}\Big((Q(\mathbf{x}-\mathbf{r}^+) - Q(\mathbf{x}-\mathbf{r}^-))\alpha_{10_s}(\boldsymbol{x})\Big)}, \tag{A9}$$

$$\frac{\partial \mathcal{L}_w}{\partial \mu_{s,r}} = 0 \;\rightarrow\; \boxed{\alpha_{9_{s,r}} = -\frac{1}{2}Q(\mathbf{x}-\mathbf{r})\alpha_{10_s}(\boldsymbol{x}) - \frac{1}{2}Q(\mathbf{x}-\mathbf{s})\alpha_{11_r}(\boldsymbol{x})}. \tag{A10}$$

For the rest of the development we will recall the same Dirichlet boundary conditions as the first-order adjoint. Before proceeding with the expression of the adjoint-state variable α_{10_s}, for the sake of clarity, we express the solution of the following term that undergoes three integration by parts:

$$\begin{aligned}\frac{\partial}{\partial \lambda_s}\Big(\alpha_{10_s}(\boldsymbol{x})\nabla\cdot(\lambda_s(\boldsymbol{x})\nabla t_s(\boldsymbol{x}))\Big)_\Omega &= -\frac{\partial}{\partial \lambda_s}\Big((\lambda_s(\boldsymbol{x})\nabla t_s(\boldsymbol{x}))\cdot\nabla\alpha_{10_s}(\boldsymbol{x})\Big)_\Omega + \frac{\partial}{\partial \lambda_s}\Big(\alpha_{10_s}(\boldsymbol{x})\lambda_s(\boldsymbol{x})\nabla t_s(\boldsymbol{x})\cdot\vec{n}\Big)_\Gamma \\ &= -\Big(\nabla\alpha_{10_s}(\boldsymbol{x})\cdot\nabla t_s(\boldsymbol{x})\Big)_\Omega + \Big(\alpha_{10_s}(\boldsymbol{x})\nabla t_s(\boldsymbol{x})\cdot\vec{n}\Big)_\Gamma \\ &= \Big(\alpha_{10_s}(\boldsymbol{x})\Delta t_s(\boldsymbol{x})\Big)_\Omega - \Big(\alpha_{10_s}(\boldsymbol{x})\nabla t_s(\boldsymbol{x})\cdot\vec{n}\Big)_\Gamma + \Big(\alpha_{10_s}(\boldsymbol{x})\nabla t_s(\boldsymbol{x})\cdot\vec{n}\Big)_\Gamma.\end{aligned} \tag{A11}$$

We proceed in the same manner for the receiver's equivalent contribution. The following expressions of α_{10_s} and α_{11_r} will invoke the Laplacian of their corresponding traveltime maps

$$\frac{\partial \mathcal{L}_w}{\partial \lambda_s} = -\frac{\alpha_1}{v(\boldsymbol{x})^3} - \alpha_{10_s}(\boldsymbol{x})\Delta\mathbf{t}_s(\boldsymbol{x}) \;\rightarrow\; \frac{\partial \mathcal{L}_w}{\partial \lambda_s} = 0 \;\rightarrow\; \boxed{\alpha_{10_s}(\boldsymbol{x}) = \frac{-w}{\Delta\mathbf{t}_s(\boldsymbol{x})v(\boldsymbol{x})^3}}, \tag{A12}$$

$$\frac{\partial \mathcal{L}_w}{\partial \lambda_r} = -\frac{\alpha_1}{v(\boldsymbol{x})^3} - \alpha_{11_r}(\boldsymbol{x})\Delta\mathbf{t}_r(\boldsymbol{x}) \;\rightarrow\; \frac{\partial \mathcal{L}_w}{\partial \lambda_r} = 0 \;\rightarrow\; \boxed{\alpha_{11_r}(\boldsymbol{x}) = \frac{-w}{\Delta\mathbf{t}_r(\boldsymbol{x})v(\boldsymbol{x})^3}}. \tag{A13}$$

Before proceeding with the expression of the next adjoint-state variable α_{5_s}, we develop the following term through a series of integration by parts

$$\begin{aligned}\frac{\partial}{\partial t_s}\Big(\alpha_{10_s}(\boldsymbol{x})\nabla\cdot(\lambda_s(\boldsymbol{x})\nabla t_s(\boldsymbol{x}))\Big)_\Omega &= -\frac{\partial}{\partial t_s}\Big((\lambda_s(\boldsymbol{x})\nabla t_s(\boldsymbol{x}))\cdot\nabla\alpha_{10_s}(\boldsymbol{x})\Big)_\Omega + \frac{\partial}{\partial t_s}\Big(\alpha_{10_s}(\boldsymbol{x})\lambda_s(\boldsymbol{x})\nabla t_s(\boldsymbol{x})\cdot\vec{n}\Big)_\Gamma \\ &= \frac{\partial}{\partial t_s}\Big[\Big(\lambda_s(\boldsymbol{x})t_s(\boldsymbol{x})\cdot\Delta\alpha_{10_s}(\boldsymbol{x})\Big)_\Omega - \Big(\lambda_s(\boldsymbol{x})t_s(\boldsymbol{x})\nabla\alpha_{10_s}(\boldsymbol{x})\cdot\vec{n}\Big)_\Gamma\Big] \\ &= \Big(\lambda_s(\boldsymbol{x})\Delta\alpha_{10_s}(\boldsymbol{x})\Big)_\Omega - \Big(\nabla\alpha_{10_s}(\boldsymbol{x})\cdot\vec{n}\Big)_\Gamma \\ &= -\Big(\nabla\alpha_{10_s}(\boldsymbol{x})\nabla\lambda_s(\boldsymbol{x})\Big)_\Omega.\end{aligned} \tag{A14}$$

Making use of the previous expression, we can derive the following equations satisfied by α_{5_s} and α_{6_r}

$$\frac{\partial \mathcal{L}_w}{\partial t_s} = 0 \;\rightarrow\; \boxed{\Big(\nabla\cdot(\alpha_{5_s}(\boldsymbol{x})\nabla t_s(\boldsymbol{x}))\Big)_\Omega = \frac{1}{2}\sum_{r=1}^{N_r}\Big(\; Q(\mathbf{x}-\mathbf{r})^t\;\alpha_{4_{s,r}} + (Q(\mathbf{x}-\mathbf{r}^+)^t - Q(\mathbf{x}-\mathbf{r}^-)^t)\frac{\alpha_{3_{r_s}}}{h}\Big) + \nabla\alpha_{10_s}(\boldsymbol{x})\nabla\lambda_s(\boldsymbol{x})}, \tag{A15}$$

$$\frac{\partial \mathcal{L}_w}{\partial t_r} = 0 \;\rightarrow\; \boxed{\Big(\nabla\cdot(\alpha_{6_r}(\boldsymbol{x})\nabla t_r(\boldsymbol{x}))\Big)_\Omega = \frac{1}{2}\sum_{r=1}^{N_r}\Big(\; Q(\mathbf{x}-\mathbf{s})^t\;\alpha_{4_{s,r}} + (Q(\mathbf{x}-\mathbf{s}^+)^t - Q(\mathbf{x}-\mathbf{s}^-)^t)\frac{\alpha_{2_{s_r}}}{h}\Big) + \nabla\alpha_{11_r}(\boldsymbol{x})\nabla\lambda_r(\boldsymbol{x})}. \tag{A16}$$

We can see that the resultant second order adjoint kernel satisfied by α_{5_s} and α_{6_s} is solved in the same manner as the first order adjoint kernel. The first two terms of the right-hand side have similar structures to the ones defining the first order adjoint-field however in this case they are not directly linked to the residuals of the data class. Another difference is the additional term representing the contribution

of the second order term of the Hessian, eliminating this term reduces the formulation to a Gauss–Newton equivalent contribution. The computational overhead associated with the presented second-order formulation revolves around solving for α_{5_s} and α_{6_s} using the fast sweeping method and their entailed derivatives terms solved straightforwardly by finite-difference. Finally, the Hessian-vector product would take this shape

$$H(\mathbf{v})w = \sum_{s=1}^{N_s}\left(3w\frac{\lambda_s(\boldsymbol{x})}{v(\boldsymbol{x})^4} - \frac{\alpha_{5_s}(\boldsymbol{x})}{v(\boldsymbol{x})^3}\right) + \sum_{r=1}^{N_r}\left(3w\frac{\lambda_r(\boldsymbol{x})}{v(\boldsymbol{x})^4} - \frac{\alpha_{6_r}(\boldsymbol{x})}{v(\boldsymbol{x})^3}\right). \tag{A17}$$

We also note that the Hessian-vector product can be solved in a finite-difference sense (Brown 1987) in the following manner:

$$Hw = \frac{g(m+\varepsilon w) - g(m)}{\varepsilon}, \tag{A18}$$

where ε is a parameter perturbation. As with any finite-difference based methods the choice of the step, in this case the perturbation, is critical since the error is proportional to the chosen perturbation which is in turn not straightforward to tune. The choice of ε depends on the tackled problem, however some generalized strategies (Brown 1987; Knoll & Keyes 2004; Nocedal & Wright 2006) have been already proposed. Once a suitable strategy for the order of magnitude of the perturbation is chosen the scheme is pretty straightforward and does not exhibit any extra cost compared to the second-order adjoint-state implementation.