



HAL
open science

Graph-Space Optimal Transport Concept for Time-Domain Full-Waveform Inversion of Ocean-Bottom Seismometer Data: Nankai Trough Velocity Structure Reconstructed From a 1D Model

Andrzej Górszczyk, Romain Brossier, Ludovic Métivier

► **To cite this version:**

Andrzej Górszczyk, Romain Brossier, Ludovic Métivier. Graph-Space Optimal Transport Concept for Time-Domain Full-Waveform Inversion of Ocean-Bottom Seismometer Data: Nankai Trough Velocity Structure Reconstructed From a 1D Model. *Journal of Geophysical Research: Solid Earth*, 2021, 126 (5), 10.1029/2020JB021504 . hal-03404483

HAL Id: hal-03404483

<https://hal.science/hal-03404483>

Submitted on 26 Oct 2021

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

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

1 **Graph-space optimal transport concept for time-domain**
2 **full-waveform inversion of ocean-bottom seismometer**
3 **data: Nankai Trough velocity structure reconstructed**
4 **from a 1D model**

5 **Andrzej Górszczyk^{1,2}, Romain Brossier¹, Ludovic Métivier^{1,3}**

6 ¹ISTerre, Univ. Grenoble Alpes, F-38000 Grenoble, France

7 ²Institute of Geophysics, Polish Academy of Sciences, Warsaw, Poland

8 ³Univ. Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France

9 **Key Points:**

- 10 • Time-domain FWI
11 • Alternative misfit function
12 • Cycle-skipping mitigation
13 • High-resolution crustal-scale imaging

Corresponding author: Andrzej Górszczyk, agorszczyk@igf.edu.pl

14 Abstract

15 Detailed reconstruction of deep structures with full-waveform inversion (FWI) of wide-angle
 16 ocean-bottom seismometer (OBS) data remains challenging and unconventional. The com-
 17 plexity of the long-offset waveforms increases the nonlinearity of the inverse problem, while
 18 the sparsity of the OBS deployments leads to a poorly constrained model reconstruction.
 19 Consequently, for such a FWI setting it is difficult to derive an initial model that sat-
 20 isfies the cycle-skipping criterion. Searching for a remedy to this issue, we investigate the
 21 graph-space optimal transport (GSOT) technique, which can potentially overcome the cycle-
 22 skipping problem at the initial FWI stage. The key feature of the GSOT cost function is
 23 the convexity with respect to the patterns in the two seismograms, which allows for correct
 24 matching of the arrivals shifted in time for more than half of the wavelet. This in turn
 25 shall allow FWI to handle the large kinematic errors of the starting model. We test this
 26 hypothesis by applying the time-domain acoustic FWI to the synthetic and field data from
 27 the subduction zone environment. We show that despite the complexity of the geological
 28 structure, the GSOT misfit function is able to guide the FWI toward the precise velocity
 29 model reconstruction and data fitting starting from a simple 1D model. The improved con-
 30 vevity of the GSOT misfit function allows FWI to converge even when mismatches between
 31 the observed and synthetic signals reach a few cycles. This ability reduces the constraint on
 32 the kinematic accuracy of the initial model and makes the FWI from the OBS data more
 33 feasible.

34 1 Introduction

35 Full-waveform inversion (FWI) has proven to be a powerful tool able to reconstruct
 36 high-resolution velocity models of the subsurface. Rapid development of this technology
 37 during past decades has led to its successful application in imaging of various geological
 38 targets, from near-surface to global-scale endpoints (see Tromp (2019) for a recent review).
 39 In particular, the oil & gas industry (stimulated by the hydrocarbon market) is currently
 40 routinely applying FWI at the exploration scale for high-resolution imaging of reservoirs
 41 (Plessix & Perkins, 2010; Stopin et al., 2014; Operto et al., 2015; Borisov et al., 2019).
 42 Such an increase in FWI applications has not yet been observed when considering regional-
 43 scale academic imaging case studies based on wide-angle ocean-bottom seismometer (OBS)
 44 acquisitions. Although these kinds of data have a great potential to efficiently constrain
 45 the subsurface velocity at depths beyond the range of typical reflection seismic data, so
 46 far, only a few attempts have been performed to combine them with FWI (Operto et al.,
 47 2006; Kamei et al., 2012; Morgan et al., 2013; Górszczyk et al., 2017; Davy et al., 2017;
 48 Górszczyk et al., 2019). Instead, the routine wide-angle data processing still relies mainly
 49 on ray-tracing and travelttime inversion, utilizing the arrival times of pre-interpreted phases
 50 (Nakanishi et al., 1998; Korenaga et al., 2000; Yu et al., 2016; Czuba, 2016). This fact is
 51 partially due to the limited availability of dense seismic datasets (preferably 1-2 km OBS
 52 spacing) required for the stable application of FWI. Indeed, one of the major challenges for
 53 crustal-scale FWI is related to the large volumes of the models constituting the geological
 54 targets, which, in turn, are recovered from relatively sparse OBS acquisitions. Moreover,
 55 the ultralong-offset data contain a full-range of various arrivals associated with different
 56 propagation regimes and resolution powers, further increasing the nonlinearity of the inverse
 57 problem. The simultaneous inversion of seismic information collected along the long-offset
 58 diving and refraction wavepaths additionally enriched with wide-angle reflection arrivals is
 59 therefore highly nontrivial. This challenge is all the more severe when the standard form
 60 of FWI is considered, namely, FWI relying on a nonlinear least-squares minimization. The
 61 fact that the L^2 misfit function is nonconvex with respect to large kinematic errors between
 62 observed and calculated data (larger than a half-period of the dominant inverted wavelet),
 63 makes the FWI prone to cycle-skipping; that is, due to the inaccurate starting model, the
 64 inversion converges toward a geologically unreliable local minimum.

65 A solution to the cycle-skipping problem can rely either on building more accurate starting
 66 models or looking for more convex misfit functions. Regarding FWI from wide-angle OBS

67 data, the initial velocity model is usually built with first-arrival tomography (FAT, e.g., Zelt
 68 and Barton (1998)). As a consequence, it strongly relies on the ability to precisely pick and
 69 accurately invert the traveltimes of first breaks. However, the picking procedure might be
 70 subjective in the areas of the seismogram where the arrivals are mixed, dominated by noise,
 71 or for far-offset data. In such a case, significant human interaction and subsequent quality
 72 control of the tomographic model is required (Górszcyk et al., 2017). Even with precisely
 73 picked first-breaks, the long-offsets data still imply a long time of wavefield propagation and
 74 therefore a potential accumulation of kinematic error along the wavepath, namely higher
 75 risk of cycle-skipping (Pratt, 2008). Furthermore, the resolution of the final FAT model is
 76 limited to the first Frenel’s zone, which does not guarantee that the model accurately predicts
 77 the later arrivals. The problem can be partially mitigated through the joint refraction and
 78 reflection traveltome tomography (e.g., Gras et al. (2019)) or through the simultaneous
 79 inversion of first-arrival traveltimes and their slopes (horizontal derivatives, e.g., Sambolian,
 80 et al. (2019)). The latter approach has the potential to better constrain the model structure,
 81 although it requires a proper weighting between the update coming from the first-break
 82 traveltimes and from their slopes.

83 Another approach to mitigate the cycle-skipping problem relies on the design of alternative
 84 ways to measure the discrepancy between observed and calculated data to enhance the
 85 convexity of the misfit function with respect to the model parameters. Such methods should
 86 make it possible to start the inversion from cruder initial models while still converging
 87 toward the global minimum. Different families of misfit function with better convexity
 88 have been proposed over the past decade. Some of them still compare the data in the
 89 L^2 sense; however, they incorporate additional constraints (coming, for example, from
 90 the traveltome information (Treister & Haber, 2017)) or modify the data, for example,
 91 taking their envelope (e.g., J. Luo and Wu (2015); Borisov et al. (2017); Bozdağ et al.
 92 (2011)) or instantaneous/unwrapped phase (e.g., Choi and Alkhalifah (2011); Alkhalifah
 93 and Choi (2012); Kamei et al. (2014)). Other misfit functions are based on the cross-
 94 correlation measurement (Y. Luo & Schuster, 1991; van Leeuwen & Mulder, 2010) or a
 95 deconvolution operation (S. Luo & Sava, 2011; Warner & Guasch, 2016; Zhu & Fomel, 2016).
 96 Further promising approaches rely on extending the model space by means of wavefield
 97 reconstruction (van Leeuwen & Herrmann, 2013; Aghamiry et al., 2018), although time-
 98 domain applications of this strategy still seems a challenging problem.

99 Recently, misfit functions relying on the optimal transport (OT) distance have been proposed
 100 for FWI leading to a new family of promising misfit functions (Engquist & Froese, 2014;
 101 Qiu et al., 2017; Y. Yang et al., 2018; Y. Yang & Engquist, 2018; Métivier et al., 2016a,
 102 2019a). In spirit, unlike the L^2 misfit function - which compares the field and the synthetic
 103 seismic data sample by sample - the OT distance searches for the best fitting (mapping)
 104 between the field and the synthetic data. Each such mapping is attributed with a cost,
 105 which increases/decreases if the mapping operation requires larger/smaller modification of
 106 the synthetic data to fit their observed counterpart. The distance between field and synthetic
 107 data is the cost produced by the best mapping operator. The appealing property from OT
 108 distance is its convexity with respect to translation and dilation, which makes it a very good
 109 candidate to produce a convex misfit function for FWI. To date, various OT-based misfit
 110 function implementations have been applied, mainly to different benchmark datasets from
 111 the oil & gas exploration field (e.g. Métivier et al. (2016b); Poncet et al. (2018); Chen and
 112 Peter (2018); He et al. (2019); Sun and Alkhalifah (2019); Pladys et al. (2020); Provenzano
 113 et al. (2020)).

114 In this study, we investigate the potential of the OT distance to mitigate the problem of
 115 the high nonlinearity and nonconvexity of FWI when applied to ultralong-offset OBS data.
 116 As described above, the sparsity of the OBS stations, large size of the models or wide range
 117 of different arrivals in the wavefield, make this acquisition setting fundamentally different
 118 than the one used in exploration-scale seismic imaging. Therefore the robustness of OT-
 119 based crustal-scale FWI has yet to be proven. We follow the graph-space optimal transport
 120 (GSOT) implementation of Métivier et al. (2019a) that compares the discrete graphs of
 121 the signals instead of the signals themselves. Due to the graph-space transformation, the

122 oscillatory seismic data are converted into positive values and can be compared without
 123 losing the convexity of the OT misfit function - even when the data are shifted by more
 124 than half a cycle. The ability of the GSOT to compare data samples within a given time
 125 window shall in principle relax the constraints related to the kinematic accuracy of the
 126 initial velocity model. As such, the GSOT misfit function provides a possibility to reduce
 127 the burden related to the accurate tomographic inversion of the first-arrival traveltimes as
 128 well as their precise picking for the sake of building the initial FWI model. We confront this
 129 hypothesis against synthetic and real 2D OBS datasets from a subduction zone environment.
 130 In both cases, the GSOT misfit function combined with a progressive data selection (from
 131 early to late arrivals) makes it possible to run FWI starting from a crude 1D model. Despite
 132 the significant kinematic error (a few cycles mismatch between the field and synthetic data)
 133 at the initial FWI stage, the GSOT misfit function makes the convergence of the inversion
 134 possible and leads to successful velocity model reconstruction.
 135 In the following section we outline the formulation of our forward and inverse problem and
 136 introduce the GSOT misfit function definition. Then, we illustrate with a synthetic FWI
 137 case study based on the GO_3D_OBS model (Górszczyk & Operto, 2021) how GSOT
 138 makes it possible to apply the FWI starting from a 1D initial velocity model that linearly
 139 increases in depth. We describe our inversion workflow and present the convexity analysis,
 140 followed by the evolution of the velocity model and the data fitting during the inversion.
 141 Further, we apply a similar FWI workflow to the SFJ-OBS field data from the Tokai area of
 142 the Nankai Trough, Japan. We show how we are able to make FWI converge starting from
 143 a 1D initial model and recover the velocity structure consistent with that obtained during
 144 a frequency-domain FWI study performed by Górszczyk et al. (2017) on the same dataset.
 145 Finally, we discuss some aspects of the presented work and summarize the article with a
 146 conclusion.

147 2 Methods

148 2.1 Forward modeling

149 In this study, we use a 2D finite-difference (2^{nd} order in time and 4^{th} order in space)
 150 visco-acoustic isotropic time-domain modeling. It is derived from the 3D visco-acoustic
 151 VTI modeling and inversion formulation proposed by P. Yang et al. (2018). The intrinsic
 152 attenuation mechanism that we use is based on the generalized Maxwell body including three
 153 standard linear solid attenuation mechanisms (P. Yang et al., 2016a). As it is common for
 154 wide-angle studies, we process the shots in a reciprocal manner, taking advantage of the
 155 spatial reciprocity of the Green's functions. In such a configuration, we treat the vertical
 156 geophone at the OBS location as a vertical force source (vertical derivative of the pressure),
 157 and we extract the pressure wavefield at the air-gun shot position. Each OBS is processed
 158 independently, taking advantage of MPI parallelism over the sources, which translates to
 159 good scalability of the modeling step, especially for large-scale applications. We impose a
 160 free-surface boundary condition at the sea/air interface and absorbing sponge layers on the
 161 lateral sides and at the bottom of the model (Cerjan et al., 1985). Sources and receivers
 162 are accurately positioned in the finite-difference grid using the windowed sinc interpolation
 163 (Hicks, 2002).

164 2.2 Inversion

In the standard FWI formulation, the mismatch between observed and synthetic seis-
 mograms is measured as an L^2 distance. This L^2 distance between two datasets (\mathbf{d}_1) and
 (\mathbf{d}_2) with N_s sources and N_r receivers is expressed as:

$$J(\mathbf{d}_1, \mathbf{d}_2) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int_0^T (\mathbf{d}_1(\mathbf{x}_r, t; \mathbf{x}_s) - \mathbf{d}_2(\mathbf{x}_r, t; \mathbf{x}_s))^2 dt, \quad (1)$$

and the associated L^2 FWI misfit function is:

$$\mathcal{C}(\mathbf{m}) = J(\mathbf{d}_{cal}[\mathbf{m}], \mathbf{d}_{obs}). \quad (2)$$

165 Here, the model properties are denoted by the vector \mathbf{m} . The total recording time for the
 166 seismogram is denoted by T . The observed and synthetic trace computed in the model \mathbf{m} , as-
 167 sociated with source \mathbf{x}_s and receiver \mathbf{x}_r , are denoted by $\mathbf{d}_{obs}(\mathbf{x}_r, t; \mathbf{x}_s)$ and $\mathbf{d}_{cal}[\mathbf{m}](\mathbf{x}_r, t; \mathbf{x}_s)$,
 168 respectively.

169 The associated inversion is formulated as the minimization of the misfit function \mathcal{C} . Due
 170 to the size of the corresponding discrete problem, the inversion is performed through the
 171 local optimization techniques. Starting from an initial model \mathbf{m}_0 , those are based on the
 172 following iteration:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta\mathbf{m}_k, \quad (3)$$

174 where k is the iteration number and the model perturbation $\Delta\mathbf{m}_k$ is given by:

$$\Delta\mathbf{m}_k = -\alpha_k Q_k \nabla_{\mathbf{m}} \mathcal{C}_k, \quad (4)$$

176 In equation (4), $-\nabla_{\mathbf{m}} \mathcal{C}$ is the steepest-descent direction and α_k is the step length magnitude
 177 at iteration k estimated through a line search process (Nocedal & Wright, 2006). The
 178 gradient $\nabla_{\mathbf{m}} \mathcal{C}$ is determined with the adjoint-state method (Plessix, 2006). The operator
 179 Q_k is an approximation of the inverse Hessian operator at iteration k . It is computed
 180 following the l -BFGS approximation (Nocedal, 1980; Byrd et al., 1995) and combined with
 181 a diagonal preconditioning operator based on a pseudo-Hessian strategy (Shin et al., 2001;
 182 Choi & Shin, 2008). This diagonal preconditioning compensates for the amplitude decay at
 183 depth in the gradient due to the geometrical spreading associated with the surface acquisition
 184 configuration we are using. Our implementation uses the SEISCOPE optimization toolbox
 185 (Métivier & Brossier, 2016). A Gaussian smoothing operator is also applied to the gradient,
 186 with a correlation length defined as a fraction of the local wavelength. The latter is estimated
 187 through a reference frequency and the current local velocity.

188 2.3 GSOT misfit function

The FWI methodology introduced in the previous section suffers from a severe limita-
 tion, namely, the nonconvexity of the L^2 misfit function with respect to the velocity model.
 If the starting model m_0 generates synthetic data shifted in time by more than half the dom-
 inant period from the observed signal, then the L^2 misfit function will guide the inversion
 toward a wrong solution. This key issue is known as the cycle-skipping criterion (Virieux &
 Operto, 2009), and to overcome it in this study we consider a GSOT-based misfit function.
 The concept of OT dates back to the French engineer Monge (1781), who first formulated the
 OT problem in an attempt to minimize the effort of workers transporting sand on a bridge
 building site. This originally ill-posed OT problem was later re-explored by Kantorovich
 (1942) who provided a well-posed relaxation of the original Monge’s problem. Based on
 the Kantorovich OT problem, it is possible to properly define an OT distance in the space
 of probability distributions. An important property of the proposed OT distance is the
 convexity with respect to translation and dilation. This means that for seismic signals, the
 OT could be convex with respect to time shifts, which is a very good proxy for convexity
 with respect to velocity perturbations.

However, the oscillatory nature of the seismic signals makes them impossible to be repre-
 sented as a probability distribution. Several strategies have been proposed so far to overcome
 this issue. Some approaches rely on a prior nonlinear transformation of the signal to make
 it positive (square, exponential, softmax) followed by a normalization (Engquist & Froese,
 2014; Qiu et al., 2017; Y. Yang et al., 2018; Y. Yang & Engquist, 2018). Another method
 consists of considering a specific instance of OT distance (1-Wasserstein distance), which
 can be extended to oscillatory signals (Métivier et al., 2016b, 2016a).

However, the nonlinear transformation of the data does not appear to us as a satisfactory
 option. This is because the chase for a robust misfit function shall not only focus on the im-
 proved convexity, but also on the ability to exploit all of the signal attributes for the sake of

better model reconstruction. Nonlinear signal transformation can cause loss of information, or might induce uncontrolled emphasis on particular phases. This is especially true in the case of long-offset OBS data that contain a wide variety of phases with different amplitudes and time signatures. The alternative approach based on the 1-Wasserstein distance has shown interesting properties, especially because it is possible to apply it to multi-D shot gathers and account for the lateral coherency of the data in this shot-gather representation (see, for instance, successful applications to the exploration-scale field data by Poncet et al. (2018); Messud and Sedova (2019); Sedova et al. (2019); Carotti et al. (2020); Shutova et al. (2020)). However, one identified drawback of this approach is the loss of convexity with respect to large time shifts (Métivier et al., 2018): the valley of attraction of the misfit function is only slightly larger than that of the L^2 norm. In other words, the convergence of FWI towards global minimum still strongly relies on the kinematic accuracy of the initial model. The interest of this approach thus relies more on its ability to better extract low-frequency information from noisy data by exploiting the lateral continuity of events in shot-gather representations rather than being resilient to large kinematic errors.

In this study, we employ an alternative OT approach based on the comparison of the discrete graph of seismic traces, following the original idea of (Thorpe et al., 2016). Mathematically, this GSOT variant of OT represents each seismic trace as a sum of Dirac delta functions in a 2D space with the time and amplitude dimensions. Comparison of the discrete graphs of the observed and synthetic seismic data (instead of the data themselves) ensures the positivity of the considered quantities and maintains the convexity of the distance with respect to time and amplitude shifts (Métivier et al., 2019a). Compared with 1-Wasserstein distance, this is a trace-by-trace approach and therefore does not exploit the lateral coherency of seismic event.

The formulation for GSOT we use has been introduced in Métivier et al. (2019a). First, we define a seismic trace $d(t)$ discretized as (d_1, \dots, d_n) . We denote its discrete graph by $(t, d) = ((t_1, d_1), \dots, (t_N, d_N)) \in \mathcal{R}^{2N}$. Let d_{cal} and d_{obs} be a calculated and observed trace respectively. With (t, d_{cal}) and (t, d_{obs}) we denote their discrete graphs consisting of N delta Dirac functions in a 2D space. The GSOT distance between d_{cal} and d_{obs} is given by solving the linear sum assignment problem:

$$h_2(d_{cal}, d_{obs}) = \min_{\sigma \in S(N)} \sum_{i=1}^N c_{i, \sigma(i)}(d_{cal}, d_{obs}), \quad (5)$$

where $S(N)$ denotes the space of permutation of $\{1, \dots, N\}$, and c_{ij} is the distance between the points i and j of the discrete graph of d_{cal} and d_{obs} :

$$c_{ij}(d_{cal}, d_{obs}) = |t_i - t_j|^2 + \left| \frac{\tau}{A} (d_{cal,i} - d_{obs,j}) \right|^2. \quad (6)$$

In Equation 6, A is the maximum peak amplitude difference between observed and calculated data, while τ is the maximum estimated time shift between d_{obs} and d_{cal} . This ensures the convexity of the GSOT distance for time up to approximately τ . The assignment problem is efficiently solved using the auction algorithm (Bertsekas & Castanon, 1989; Métivier et al., 2019b). The final cost function we use for the purpose of FWI application with N_s shots containing N_r receivers is defined as:

$$\min_m f[m] = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} w^{s,r} h_2(d_{cal}^{s,r}[m], d_{obs}^{s,r}), \quad (7)$$

where $d_{obs}^{s,r}$ and $d_{cal}^{s,r}[m]$ are the observed and synthetic (calculated in model m) traces respectively, which are associated with source s and receiver r . The $w^{s,r}$ is a trace-by-trace weighting factor, typically used to restore the AVO trend in the data. This trend is removed from the trace-by-trace GSOT approach, and the amplitude of each trace is treated separately through the normalization factor $\frac{\tau}{A}$.

Regarding the gradient computation, we refer to (Métivier et al., 2019a). Let us however state the main result. Due to the adjoint-state strategy, the only modification in the gradient computation compared with the L^2 misfit function is related to the computation of the

adjoint source. The latter can be expressed, in the GSOT approach, as a generalized residual between synthetic and field data. For a given trace, this generalized residual is equal to the difference between synthetic and field data at time samples connected by the optimal permutation σ solution of the problem (5). Mathematically, this generalized residual is defined by:

$$r_i = d_{cal,i} - d_{obs,\sigma(i)} \quad (8)$$

189 with σ being the solution of (5). We also see immediately that when σ is equal to the identity,
 190 we recover the conventional L^2 misfit function with the conventional residual (difference between synthetic and observed data).

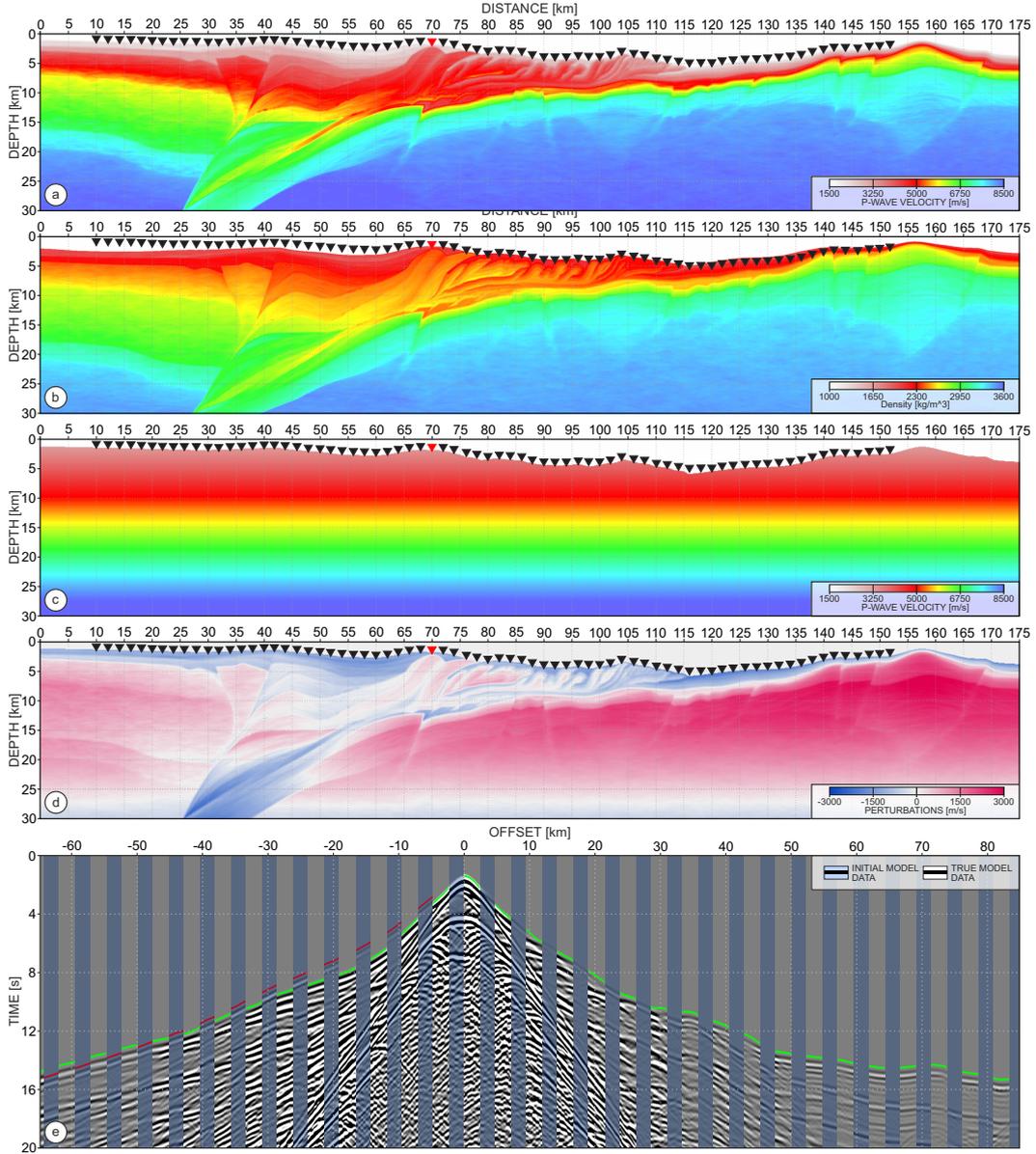


Figure 1. (a)-(b) True V_p and ρ models; (c) Initial V_p model; (d) Velocity perturbations - difference between (a) and (c); (d) OBS gather recorded at 70 km of model distance (red triangle in (a)-(d)). Every 20 traces generated in the true model are interleaved with the following 20 traces generated in the initial model. Green/red lines mark the first breaks in true/initial data.

191 3 Synthetic test

192 In this section, we illustrate how the GSOT misfit function can lead FWI from long-
 193 offset OBS data into the correct velocity reconstruction of deep crustal targets starting from
 194 a crude 1D model. We use the 2D P-wave velocity and density (V_p and ρ) models extracted
 195 from the GO_3D_OBS (Górszczyk & Operto, 2021) crustal-scale model of a subduction
 196 zone (see Figure 1a-b). This model has been designed to benchmark different tomographic
 197 and inversion approaches with a special emphasis on FWI of long-offset stationary-receiver
 198 data. Our test assumes modeling of the acoustic data within the target model and subse-
 199 quent inversion of these data starting from the kinematically wrong initial model.

200 3.1 Experiment setup

201 The acquisition settings that we design comprises 72 receivers distributed along the
 202 seabed with 2 km spacing intervals, as presented in Figure 1a, and 1500 shots distributed
 203 every 100 m (10 m depth) between 5 km and 155 km of model distance. To generate the
 204 synthetic dataset, we use a 2D time-domain acoustic isotropic modeling with a 1.5 Hz Ricker
 205 source wavelet. We consider a 20 s propagation time.

206 During the inversion we focus mainly on the ability of GSOT to overcome the problem of
 207 the kinematic inaccuracy of the initial velocity model. For this purpose, we consistently
 208 keep the true ρ model that was also used for synthetic data generation. Alternatively, we
 209 could use constant density model for both forward modeling and FWI steps. This could
 210 theoretically make the problem simpler for GSOT-based FWI, since in our setting the ve-
 211 locity update has to take into account spatially varying density parameter. Moreover, the
 212 presence of density contrasts can create the arrivals (especially short-spread reflections and
 213 less produced wide-angle reflections) originating at the interfaces, that are not existing in
 214 the velocity model at the initial stage. As a starting V_p model, we use the 1D model that
 215 linearly increases in depth, presented in Figure 1c. The difference between the true and
 216 initial models is presented in Figure 1d. Note the significant velocity perturbations reaching
 217 ~ 3000 m/s. In Figure 1e, we present in an interleaved manner the corresponding data gener-
 218 ated in the true and initial models. The complexity of the structure from the true model
 219 is clearly reflected by the various interfering arrivals. Large velocity differences from Figure
 220 1d are reflected by a significant kinematic mismatch between the first-arrival traveltimes
 221 marked by the green (true) and red (initial) lines in Figure 1e. This clear cycle-skipping
 222 revealed in the OBS gather example makes it impossible to run FWI with the standard L^2
 223 misfit function which is stuck from the first iteration.

We perform time-domain acoustic FWI with the l -BFGS optimization scheme and the

	STAGE1	STAGE2	STAGE3	STAGE4
τ value	4.0 s	4.0 s	4.0 s	4.0 s
Time window	0.2 s + 0.5 s taper	0.2 s + 0.5 s taper	0.2 s + 10 s taper	Full data
Amplitude	NO AVO	AVO	AVO	AVO
Smoothing	2.0 \times 2.0	1.6 \times 0.8	0.8 \times 0.4	0.8 \times 0.4
Iterations	50	20	50	150

224 **Table 1.** Summary of the inversion steps in the synthetic test

225 preconditioning strategy described previously. We invert for the synthetic data without ap-
 226 plying any band-pass filter, and we follow the 4-stage multiscale strategy presented in Table
 227 1. The workflow is based on progressively extended time windows and reduced gradient
 228 smoothing realised by the means of the Gaussian filter with correlation lengths matching to
 229 the local wavelength. The time-window of length of 0.2 s starts at the first-arrival time and
 230 is extended from one FWI stage to another with progressively longer taper. This makes it
 231 possible to limit the volume of the data used at the beginning of the inversion. In Table 1,

232 the time window that we use at the initial FWI stages is narrow and focused mainly around
 233 the first arrivals. Consequently, the volume of the data that must be compared by the GSOT
 234 misfit function is reduced. Muting of the secondary arrivals and wide-angle reflections from
 235 the observed data reduces the risk of matching them with the first arrivals from the synthetic
 236 data. Such a mismatch may likely occur taking into account the large time shifts between
 237 the observed and initial data in Figure 1e. Moreover, progressive introduction of the later
 238 arrivals translates to the narrowing of the scattering angles associated with smaller scale of
 239 perturbations. Combined with the tuned correlation lengths of the smoothing operator, this
 240 approach makes it possible to keep control of the spatial scale of the model update at each
 241 FWI stage. In addition, to re-estimate the maximum peak amplitude difference between
 242 the observed and calculated data (A in Equation 6), we restart the inversion after every 10
 243 FWI iterations.

244 At the initial FWI stage, we use constant $w^{s,r}$ equal to 1; namely, we compare trace-
 245 normalized seismograms to increase the contribution of the far-offset data in the gradient.
 246 Those data correspond to the waves penetrating in the deeper subsurface and have typically
 247 much smaller amplitudes than the near-offset data. Trace normalization at the initial FWI
 248 stage mitigates this issue at the price of the loss of the amplitude trend with the offset.
 249 For the consecutive FWI stages the $w^{s,r}$ is calculated as a trace-by-trace RMS value of the
 250 observed seismogram to restore the amplitude vs offset (AVO) trend in our misfit function.

251 3.2 Convexity analysis

To analyze the convexity of the GSOT at stage 1 of our inversion, we generate a population of 10 V_p models V_α according to the formula:

$$V_\alpha = (1 - |\alpha|)V_{true} + |\alpha|V_{init} \quad (9)$$

252 where $\alpha = [-1.0, -0.9, \dots, 0.9, 1.0]$. For each of the models we calculate the initial misfit function
 253 values using $\tau = 0.005$ s and $\tau = 4$ s. The first value of τ mimics the L^2 norm, since the data
 254 sampling in our test equals 0.005 s. For this scenario the convexity of the GSOT-misfit
 255 function is reduced to the L^2 norm convexity. The second value of τ is empirically selected
 256 to account for significant time shifts between the early arrivals in the true and initial data.
 257 Consequently, the convexity of the GSOT misfit function is increased. The normalized values
 258 of the misfit functions for both scenarios are plotted in Figure 2. One can observe that for
 259 $\tau = 0.005$ s, the global minimum valley is in between $\alpha \sim -0.3$, and $\alpha \sim 0.3$. Starting from an
 260 initial model that is farther from the true model would result in converging toward a local
 261 minimum. Conversely, for $\tau = 4$ s, we obtain a convex misfit function with a single global
 262 minimum for the whole range of α values, including the case $|\alpha| = 1.0$ which corresponds to
 263 taking the 1D model as an initial model. Note, that the absolute value of α in Equation 9
 264 is used only to create the apparent symmetry of the misfit function with respect to $\alpha = 0.0$ in
 265 Figure 2. Namely, the initial models for a given $\pm\alpha$ values are exactly the same and so are
 266 the values of the misfit functions.

267 Notably, with the broadening of the global minimum valley, it also becomes less sharp,
 268 which is an indication of potential resolution loss. As a consequence, keeping large values of
 269 τ can lead to less resolved velocity model reconstructions. For the synthetic test presented
 270 here, we however keep a constant $\tau = 4$ s at each stage of the inversion. For the field data
 271 application, various factors such as presence of noise, acoustic approximation of the wave
 272 propagation and simplified subsurface parametrization, or imperfect source wavelet, can
 273 further broaden the null space of the misfit function. Therefore, in such a case, we shall
 274 keep the τ value according to the kinematic errors of the model and as small as possible to
 275 narrow the global minimum valley.

276 3.3 Results

The final recovered V_p model is presented in Figure 3a. Figure 3b-e show the velocity perturbations recovered at each of inversion stage. Each of the panels is overlaid with

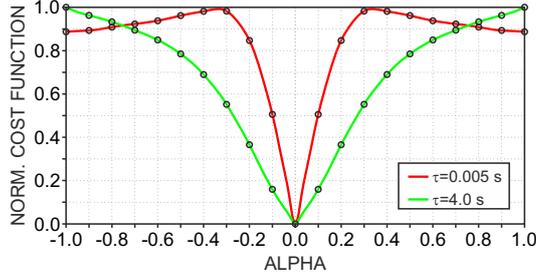


Figure 2. GSOT cost-function convexity for $\tau = 0.005$ s and $\tau = 4$ s at the first stage of our test.

three transparent masks that change the saturation of the underlying color according to the relative error ϵ of the resolved velocity model calculated as follow:

$$\epsilon = \left| \frac{m_{true} - m_{fwi}}{m_{true}} \right| * 100\%. \quad (10)$$

277 In Equation 10 m_{true} and m_{fwi} are the true model and the model recovered at given FWI
 278 stage respectively. We introduce three ϵ intervals: (1) $\epsilon \leq 2.5\%$ (light saturation); (2)
 279 $2.5\% < \epsilon \leq 5.0\%$ (intermediate saturation); (3) $5.0\% < \epsilon$ (strong saturation, see arrows in
 280 Figure 3b). Note how from one stage to another the resolution of the recovered structure is
 281 increased, and how the area of light-saturated colors ($\epsilon \leq 2.5\%$) is progressively expanding.
 282 Despite, the fact that we invert for low frequency band (below 4.5 Hz) (although assuming
 283 no noise and known source signature), the complex velocity perturbations in Figure 3e were
 284 recovered in the details (compare with true perturbations in Figure 1d). In particular, on
 285 the landward part the steeply dipping faults marking the old backstop at approximately 30-
 286 40 km distance as well as the thin subduction channel on top of the oceanic crust between
 287 40 and 55 km distance are clearly imaged. In the central part, the accretionary wedge
 288 containing a sequence of deformed thrusts is clearly reconstructed and also includes the
 289 low-velocity zone at 12 km depth between 60 and 80 km distance. In the deep water
 290 part, the rapid increase in velocity is properly recovered, and the sharp contrast of velocity
 291 between sediments deposited in the trench and the top of the oceanic crust is imaged.
 292 Furthermore, the signatures of the faults cutting through the subducting crust and the upper
 293 mantle are also imaged. In the deeper parts of the mantle, we start observing some velocity
 294 mismatches that possibly result from the limited and mainly subhorizontal illumination of
 295 those segments. This is confirmed by Figure 3f where we show the perturbation logs (marked
 296 in Figure 3e with black-dashed lines) representing true (red line) and reconstructed (green
 297 line) V_p perturbations (difference between true/initial and reconstructed/initial V_p
 298 models). One can observe that the logs remain in very good agreement even for the most complex
 299 structures down to approximately 20 km of the model depth. In Figure 4a-b, we show the
 300 initial and final data-fitting comparison for the OBS marked with a red triangle in Figure
 301 3a. The red-blue color scale corresponds to the calculated data, while the black-colored
 302 phases denote the observed data. The presence of large time shifts between the observed
 303 and initially predicted data in Figure 4a are a clear indication of the cycle-skipping. Despite
 304 it, the fact that there is no red color in Figure 4b indicates that the GSOT-based FWI
 305 approach we have designed leads to precise data and model reconstruction.

306 4 Field data application

307 In this section we present the GSOT-based FWI application to the TKY-21 field OBS
 308 data from the subduction zone located in the eastern Nankai Trough (Tokai segment, Fig-
 309 ure 5a). This region was historically a subject of numerous studies investigating its geody-
 310 namical background in the context of devastating earthquakes frequently occurring in the

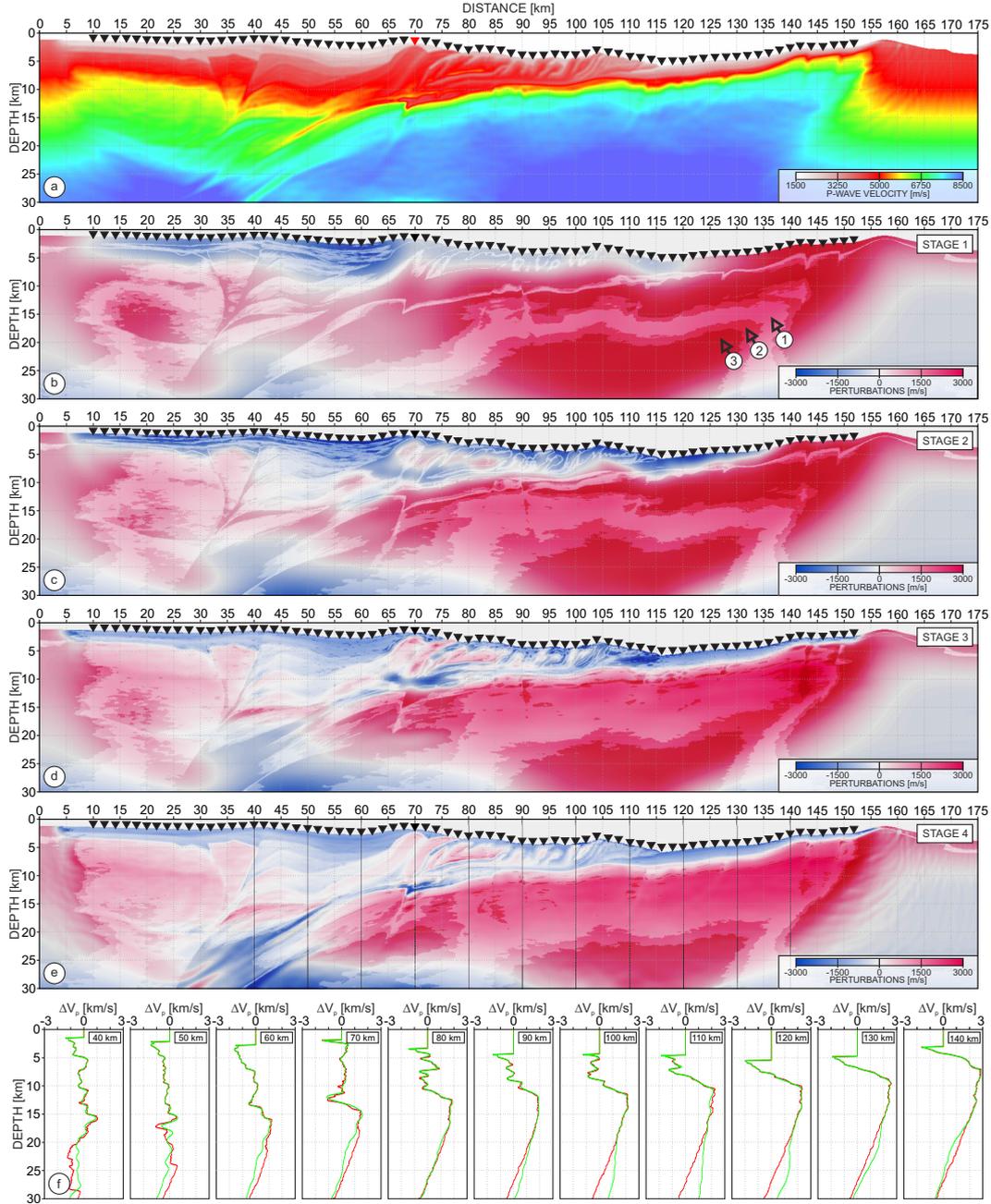


Figure 3. (a) Final V_p model after FWI stage 4; Red triangle marks the position of the OBS from Figure 4; (b)-(e) Recovered perturbations - difference between the initial and the final model after each FWI stage; Each panel is overlaid with three transparent masks that modify the color saturation and correspond to three intervals of model reconstruction error ϵ : (1) $\epsilon \leq 2.5\%$; (2) $2.5\% < \epsilon \leq 5.0\%$; (3) $5.0\% < \epsilon$. See the arrows in (b), which mark the respective saturation levels. (f) True (red) and reconstructed (green) V_p perturbation logs (see vertical black-dashed lines in (e) for locations);

311 Nankai Trough. Among different research efforts, the detailed velocity model reconstruction
 312 via frequency-domain FWI of the TKY-21 OBS line was first performed by Operto et
 313 al. (2006) and later revisited by Górszczyk et al. (2017) with Laplace-Fourier FWI. These

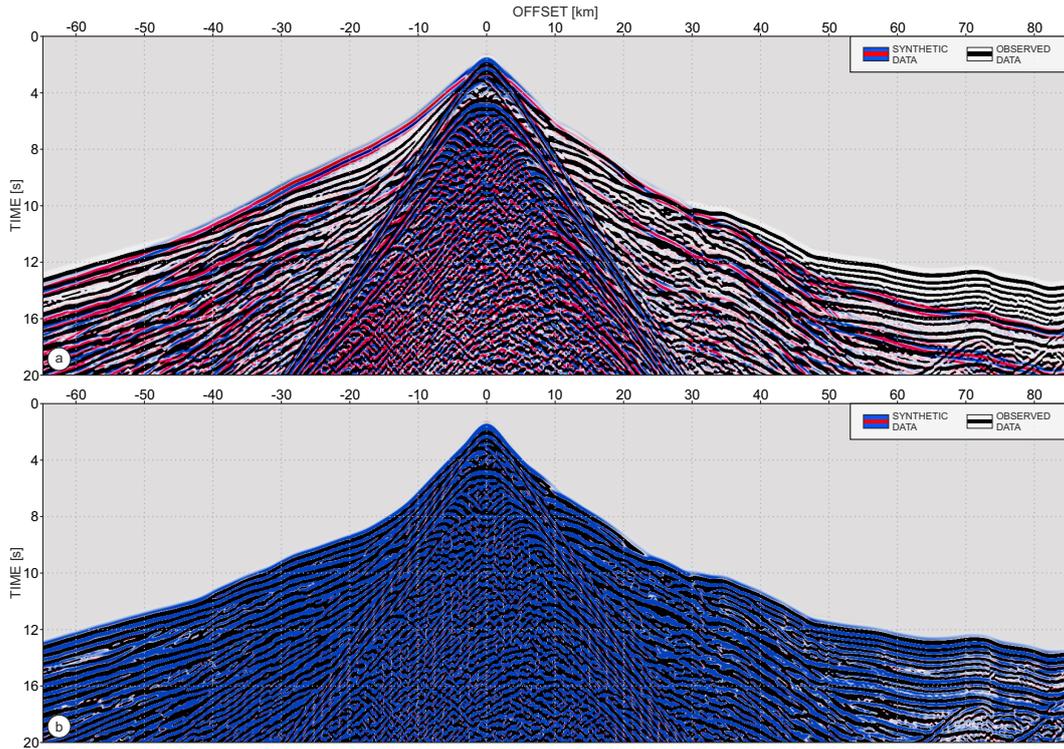


Figure 4. Comparison of the data fitting at the initial (a) and the final (b) FWI stage. Observed

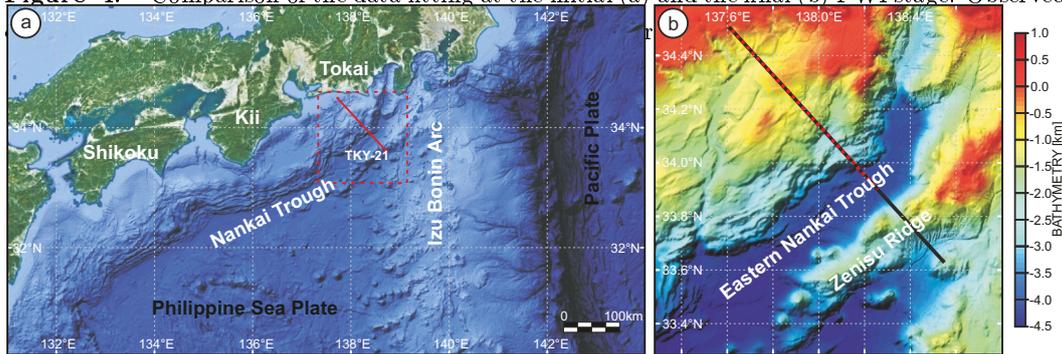


Figure 5. (a) The tectonic settings of the Nankai Trough area. The solid red line represents the seismic profile of the TKY-21 experiment. (b) Zoomed view of the TKY-21 survey area, with bathymetry variations. The black and dashed red lines represent the shot and the receiver lines respectively.

314 frequency-domain results gave us a good reference point for the time-domain GSOT-based
 315 FWI application we present here.

316 4.1 Data preprocessing

317 The OBS TKY-21 dataset was acquired by JAMSTEC in 2001 during the KY0106 cruise
 318 of R/V Kaiyo within the framework of the French-Japanese project SFJ-OBS. The acquisition
 319 consisted of 100 OBS stations (facilitated with 4.5-Hz three-component geophones and
 320 hydrophones) deployed within 1 km intervals and 1404 air-gun shots (total array volume
 321 of 196 l) spaced 100 m apart. This acquisition geometry leads to a relatively dense spatial
 322 sampling of the target by the propagating wavefield which is at least sufficiently redundant
 323 for FWI to be applied.

324 Before running the inversion we perform an elementary data preprocessing procedure. We

325 remove the DC component before applying 3D to 2D amplitude-spreading correction (scal-
 326 ing factor equal to the square root of time). We then apply a spectral whitening to partially
 327 mitigate the bubble effect, and we bandpass filter the data between 1.5 and 3.5 Hz. Finally,
 328 we normalize the OBS gathers with their amplitude RMS value to compensate for differences
 329 in the instrument response and seabed coupling.

330 4.2 Workflow

331 Our field data FWI application relies on the workflow designed during the synthetic
 332 study previously presented. The analogous 4-stage FWI scheme is summarized in Table 2.
 333 The parameters are tuned to meet the real data conditions, namely, possible elastic effects,
 334 rapid amplitude decay with offset, inaccuracy of source wavelet, and presence of noise. In
 335 particular, we reduce the value of τ as the model accuracy increases from one stage to
 336 another. We set the time windows to proceed from first arrivals to the later ones. We use a
 337 maximum time window of ~ 2 s after the first arrival to limit the amount of energetic elastic
 338 arrivals (mainly from near and intermediate offsets) in our acoustic inversion. In stage 2, we
 339 introduce an intermediate misfit function weighting based on the square root of the AVO.
 340 This is due to the large amplitude variations between near- and far-offset traces (several
 341 orders of magnitude) observed in this dataset. We run the inversion until no significant
 342 model updates are observed. We pre-estimate one source wavelet for all OBS gathers from
 the short-offset water wave (Pratt, 1999), and we keep it the same for all FWI stages. Figure

	STAGE1	STAGE2	STAGE3	STAGE4
τ value	4.0 s	1.0 s	1.0 s	0.2 s
Time window	0.1 s + 0.5 s taper	0.1 s + 1.0 s taper	0.1 s + 2.0 s taper	0.1 s + 2.0 s taper
Amplitude	NO AVO	SQRT(AVO)	AVO	AVO
Smoothing	2.0 \times 2.0	1.6 \times 0.8	1.6 \times 0.8	0.8 \times 0.4
Iterations	80	40	40	140

Table 2. Summary of the inversion steps in the field data application

343 6a presents the initial velocity model (1D velocity profile) derived as a horizontal average
 344 of the initial FWI model used by Górszczyk et al. (2017) which was obtained with FAT. To
 345 illustrate the kinematic mismatch between the field and the synthetic data generated in this
 346 initial model, we show in Figure 6b the gather from the OBS marked by the red triangle
 347 in Figure 6a. Every 20 traces of the field data are interleaved with subsequent 20 traces
 348 of the synthetic data (blue shading). It is clear that with the increasing offset, the time
 349 shifts between trace intervals become larger and locally reach few cycles. Moreover, one can
 350 observe the local change in the time shifts sign depending on the offset. This indicates that
 351 the initial model lacks large-scale low and high velocity perturbations.

352 To obtain a better insight into the kinematic inaccuracy of the initial model, we estimate
 353 the order of the cycle-skipping for the whole dataset. Figure 6c presents the map of the
 354 skipped cycles between the observed and synthetic first-arrivals for each trace used during
 355 inversion. The number of skipped cycles was calculated as a difference between the first-
 356 break traveltimes picked from the observed data and the synthetic traveltimes calculated in
 357 the initial FWI model from Figure 6a. This difference was further divided by the dominant
 358 period of the data (assuming a 3.5 Hz dominant frequency). One can observe, that there are
 359 just a few regions where the observed first-arrival phases match the synthetic ones (white
 360 color), while the majority of the synthetic traces are shifted with respect to their field
 361 equivalent by at least two cycles. Not surprisingly, with such an initial model, FWI with
 362 the L^2 misfit function is unable to converge from the very first iteration of stage 1. In
 363 contrast, the FWI with the GSOT-based misfit function provides a consistent model and
 364 data update along the different stages of the inversion.
 365

366

4.3 Results

367

368

369

370

371

372

373

374

375

376

377

378

379

380

381

382

383

384

385

386

387

388

389

390

391

Figure 7a-d and Figure 8 illustrate, respectively, the model and the data evolution after each FWI stage. One can observe that after stage 1 we recover long-wavelength positive (red) and negative (blue) perturbations (see the inset in Figure 7a) which leads to a smooth model exhibiting the general trend of the subduction zone. However, the synthetic data in Figure 8a (blue-shaded traces) are still locally significantly cycle-skipped. This mismatch effects most likely from the strong smoothing of the gradient at this stage, which hampers the intermediate- and small-scale perturbations needed to explain the data more precisely. Therefore, in stage 2, we use a smaller τ value to improve the convergence of inversion, and simultaneously, we increase the resolution of the introduced perturbation by means of reduced smoothing and a slightly extended time window. Importantly, we switch to a data weighting according to the square root of the AVO. In this way, we put more weight into the energetic short- and intermediate-offset data that carry rich information about the underlying geological features. It is clear that after stage 2, the reconstructed model (Figure 7b) has much higher resolution. One can observe the signature of the complex structures building the accretionary prism. The shape of the oceanic crust and the Moho is also reconstructed. This increase in the model resolution is reflected by the improved data fitting presented in Figure 8b. There is no clear evidence of cycle-skipping within the first arrivals which indicates a significant improvement of the model with respect to the previous stage. In the inset, one can observe that the synthetic data exhibit a complex package of wide-angle reflections - although their kinematics and dynamics are not precisely reconstructed yet. The model after stage 3 (Figure 7c) shows mostly improvement of the resolution of the shallow and intermediate structures. This result is due to the weighting of the misfit function applied according to the true AVO. At this stage, we further improve the continuity of the phases between the synthetic and field data panels, as seen in Figure 8c, as well as for later arrivals. The final model is presented in Figure 7d. A smaller τ value and a smaller

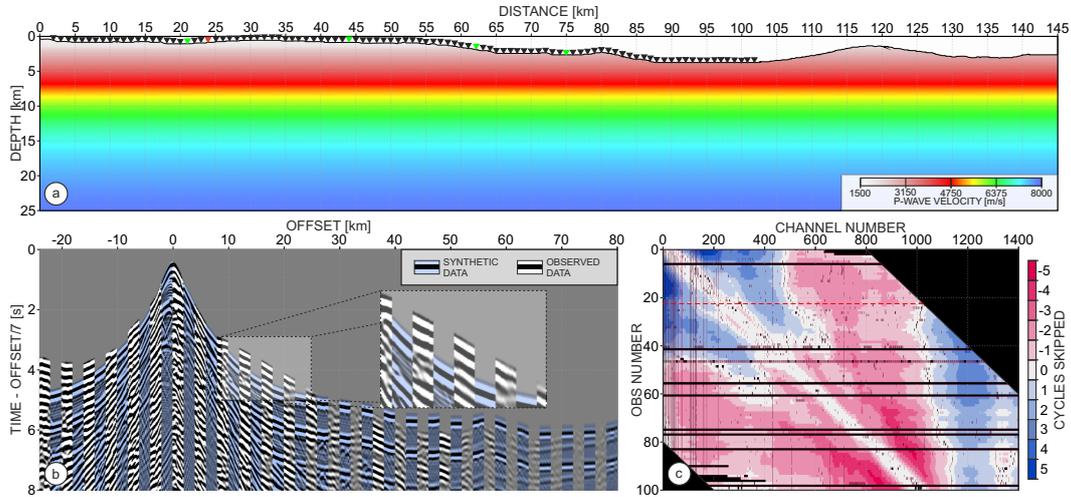


Figure 6. (a) 1D initial FWI model. Triangles mark the OBS positions. Red triangle mark the position of the OBS in (b). Green triangles mark the position of the OBS in Figure 9. (b) Gather extracted from the OBS 23. Every 20 traces of the observed data (white and black phases) are interleaved with the following 20 traces of the synthetic data (light blue and black phases) generated in the initial FWI model from (a). Inset present the zoom on the heavily cycle-skipped data. (c) Map showing the number of skipped cycles for each trace in the dataset - arranged in the source-receiver coordinates. Blue/red color indicates that the first breaks of the synthetic data arrive later/earlier than their observed counterparts. Black color marks the traces excluded from the inversion.

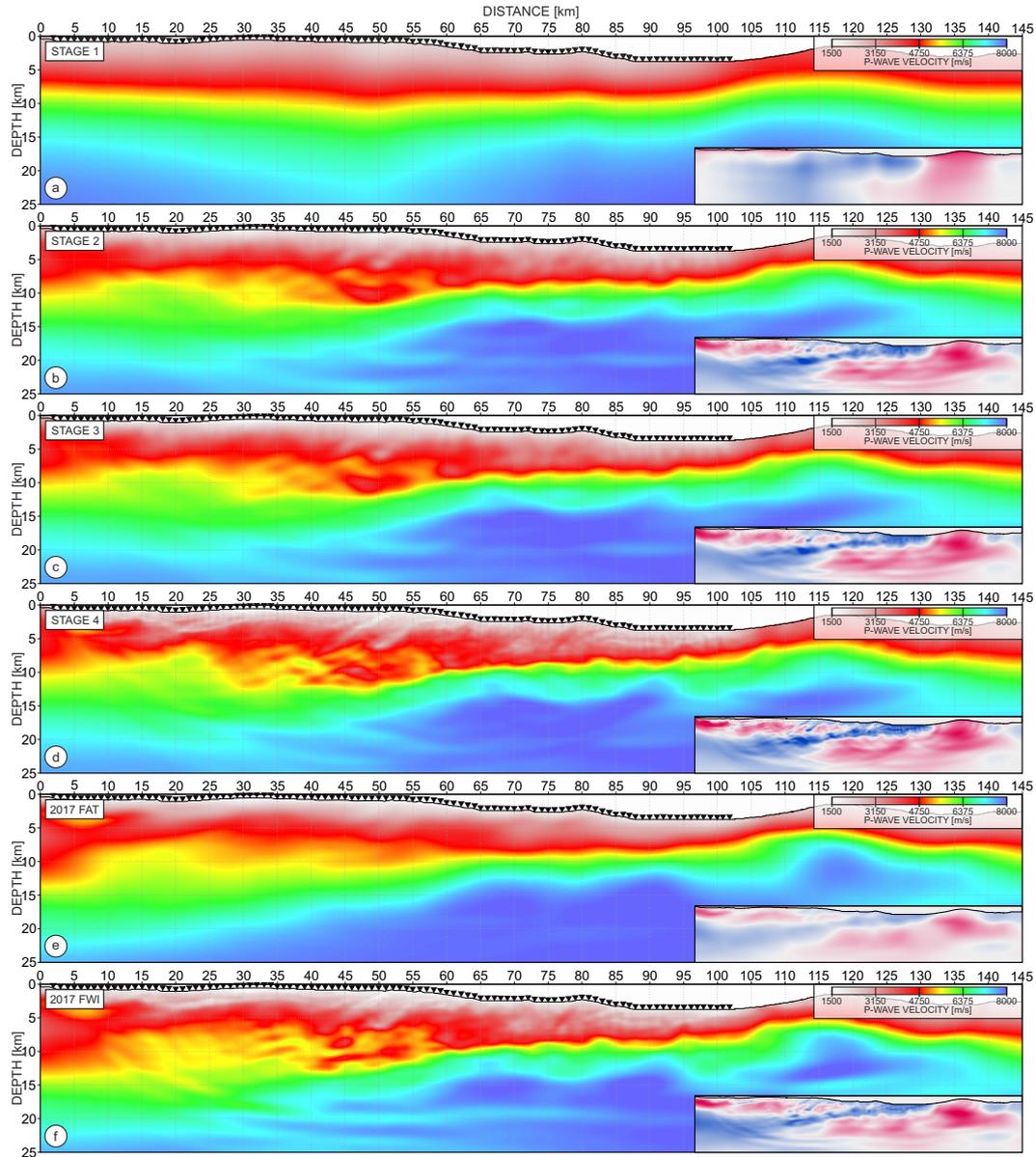


Figure 7. (a)-(d) Model evolution after each FWI stage. (e)-(f) Results from Górszczyk et al. (2017). Initial model derived with FAT and the model derived with frequency-domain FWI (frequency range up to 4 Hz). Insets show the difference between the presented and the initial model from Figure 6a. The color-scale is ± 2000 m/s.

392 smoothing make it possible to sharpen the structure within the accretionary prism. Shallow
 393 sedimentary basins, as well as sequences of thrusts of various scales are clearly visible now.
 394 The characteristic undulations of the subducting oceanic crust in the Tokai region, coupled
 395 with a wavy nature of the underlying Moho are also made evident. The final velocity per-
 396 turbations presented in the inset exhibit a wide range of introduced structures varying both
 397 in terms of spatial scale and magnitude. In Figure 8d one can observe further improvement
 398 of the continuity of phases and amplitude trends between synthetic and field data traces.

399 We can also compare the model evolution (Figure 7a-d) with the results obtained by
 400 (Górszczyk et al., 2017). Figure 7e-f shows the initial model (derived with FAT) and the
 401 model from Laplace-Fourier FWI after inverting frequency groups between 1.5 Hz and 4.0
 402 Hz. Direct and detailed judgement of the final models (Figure 7d and f) is not straightfor-

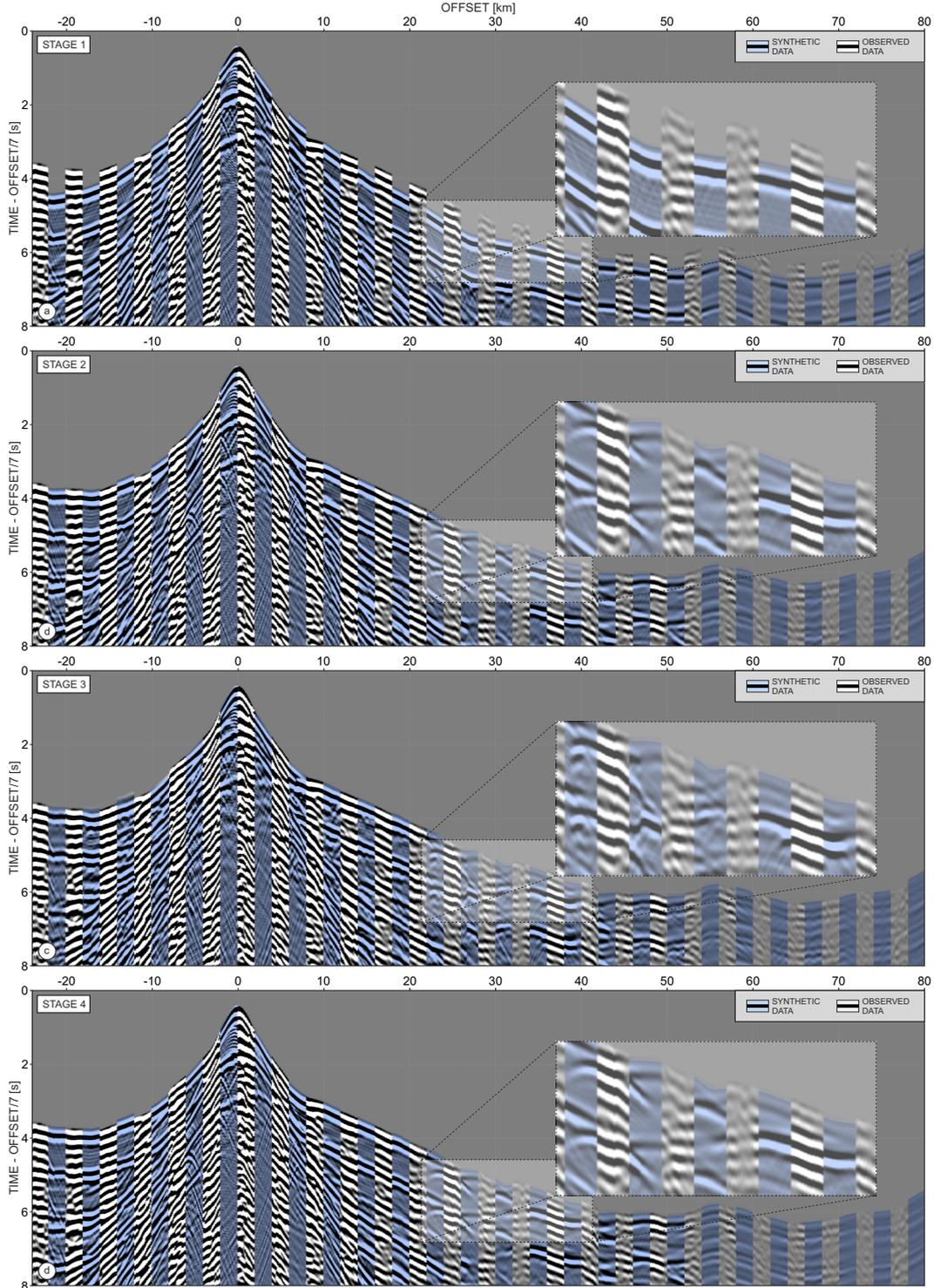


Figure 8. (a)-(d) Comparison of data fitting (OBS 23, Figure 6b) after FWI stages 1-4 respectively. Every 20 traces of the observed data are interleaved with the following 20 traces of the synthetic data (blue-shaded traces). Insets show the zoom on the complex waveform package.

403 ward due to the different parametrizations used in both workflows. However, we can observe
 404 a good agreement between the two velocity structures despite the fact that the time-domain
 405 FWI results were obtained starting from a simple 1D model. One can observe that the tomographic
 406 model in Figure 7e has higher resolution than the model after FWI stage 1 (Figure

7a, still producing the cycle-skipped data) and lower resolution than the model after FWI stage 2 (Figure 7b). We may therefore argue, that using possibly inaccurate tomographic model (e.g. equivalent of the model in Figure 7a) as a starting point for GSOT-based FWI, shall allow us to start the inversion from stage 2 and still converge towards global minimum. Together with a notable data fitting obtained with the GSOT-based FWI, this justifies the accuracy of the model reconstruction. We will come back to the time- versus frequency-domain FWI in the Discussion section.

To obtain a better insight into the final data fitting in Figure 9, we present another 4 observed/synthetic OBS gathers recorded at different locations along the profile (green triangles in Figure 7). The presented waveforms cover a broad range of arrivals that might be recorded in the subduction zone environment. One can observe the good continuity of the phases not only in terms of the first arrivals but also in terms of later wide-angle reflections and free-surface multiples appearing at various offsets. The amplitude trends of the synthetic data also follow those from field data - although the acoustic approximation of our FWI causes limited amplitude reconstruction in particular for later arrivals. It is noteworthy again here that for our inversion, we use only the data from the 2 s time window starting at the first arrival. Such a time window makes it possible to incorporate the majority of the P-wave arrivals constraining the velocity model, and at the same time, it limits the amount of waveforms related to elastic effects that we cannot handle properly due to the simplified approximation of physics we are using here.

5 Discussion

5.1 Application of FWI to OBS data

FWI proved to be a powerful technique able to recover velocity models of the subsurface with the theoretical resolution limit reaching half of the wavelet. This resolution can be indeed obtained assuming that the target is illuminated by the wavefield propagating at different angles and spanning wide range of wave-vectors at this target. Such a illumination can be obtained with the dense long-offset OBS deployments, which record the data that contain various waveforms - including those travelling in the deep subsurface and undershooting the target structure. Additionally, the OBS experiments provide 4C data, opening the perspective for decoupling the crosstalk between physical parameters during multiparameter FWI (Operto et al., 2013). In practice, however, the wavefield associated with a surface seismic acquisition, provides only limited sampling of the model-space by the wave-vectors. The issue becomes even more severe in case of the academic 2D OBS deployments, where the OBS spacing is often larger than 10 km and the low-frequency content of the data is missing. Application of FWI to such a dataset leaves major doubts about the geological correctness of the reconstructed model - even if one is able to fit the data. Unfortunately, the number of the OBS stations available for a given seismic experiment is always limited. It is therefore necessary to optimize the acquisition setting, if the aim is to process the resulting data using FWI. Below we mention some of the acquisition-related factors and dependencies, which from the practical point of view, can have a crucial impact on the successful application of FWI to the OBS data.

5.1.1 Complexity and depth of the target

With decreasing number of the OBS stations the spatial aliasing in the shallow part of the model increases (Brenders & Pratt, 2007a). However, the sampling of the intermediate and deeper parts of the model (e.g. crust or upper mantle) can be sufficient to retrieve a meaningful structural information with the resolution higher than in case of FAT. Of course this might be possible assuming that there is no large kinematic errors in the shallow part of the model (e.g. resulting from the shallow high-velocity layers, which are difficult to reconstruct without dense receiver coverage). On the other hand, the sampling of the deeper parts of the model will most likely rely on the sub-horizontal wavefield propagation, which

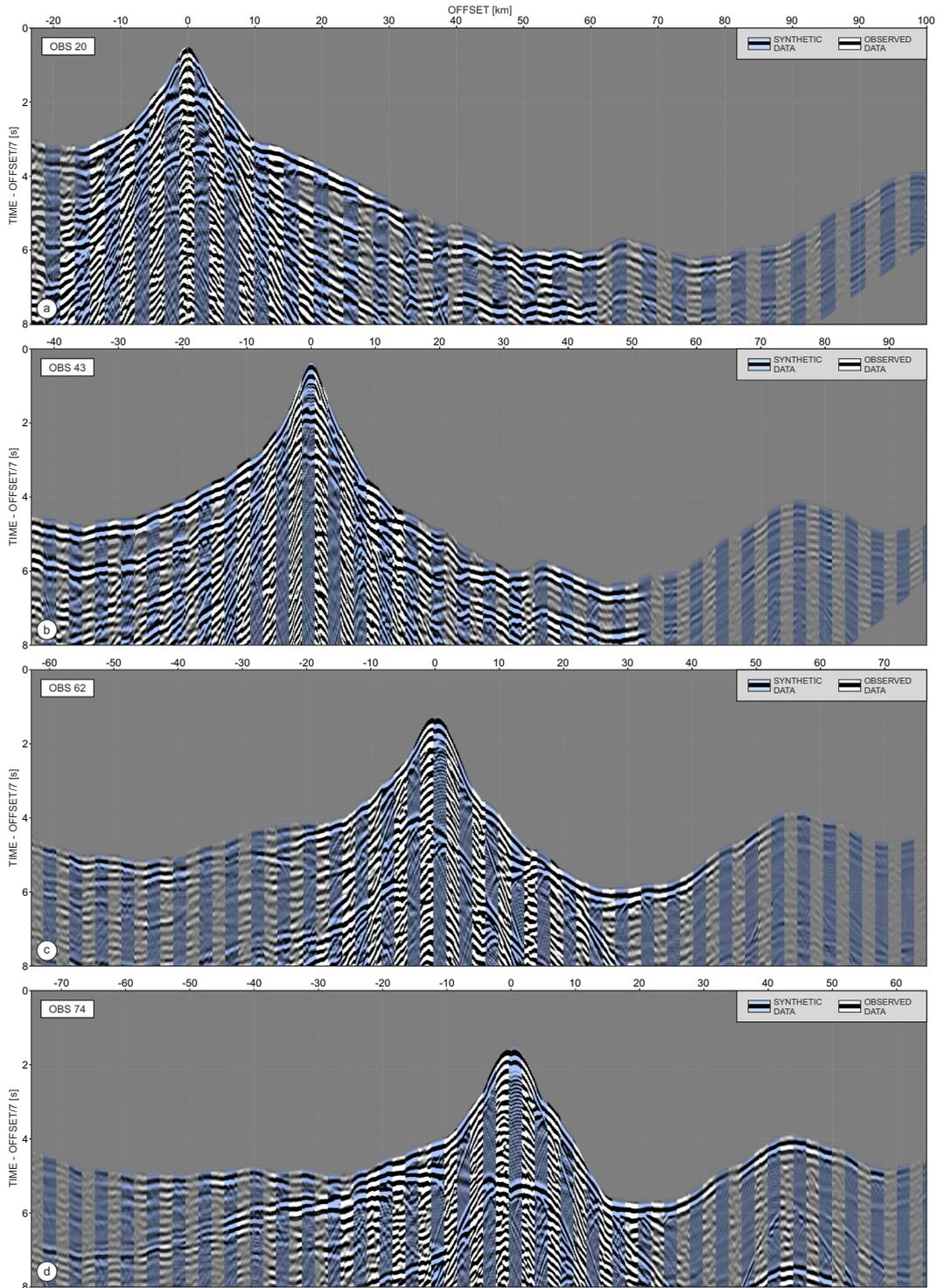


Figure 9. (a)-(d) Comparison of data fitting after stage 4 of FWI - the OBS positions are marked by the green triangles in Figure 6a. Every 20 traces of the observed data are interleaved with the following 20 traces of the synthetic data (blue-shaded traces).

457 will translate to the reconstruction based only on the limited range of wave-vectors. This
 458 issue could be partially stabilised if the various types of arrivals (i.e. reflections, wide-angle
 459 reflections, diving waves, refraction) originate at the same target and feed the FWI with

460 more seismic information. In other words, if the recorded wavefield contains clear arrivals
 461 that represent different propagation regimes and sample the same subsurface area, then
 462 the precise reconstruction of such a waveforms can potentially mitigate the sparsity of the
 463 OBS acquisition and stabilise the inversion. The drawback in turn, will be the increased
 464 nonlinearity of the inverse problem related to the more complex input to FWI.

465 **5.1.2 Availability of the multichannel streamer (MCS) data**

466 Incorporating MCS data into the model building process can significantly constrain the
 467 velocities down to the depths approximately equal to the streamer-length. This makes it
 468 possible to partially compensate for the undersampling of the shallow subsurface related
 469 to the sparse OBS deployment. The constrain from the MCS data can rely either on the
 470 building more accurate initial model, either on updating the final FWI model from the OBS
 471 data. Depending on the length of the streamer and frequency content of the data, one
 472 can aim on updating the velocity model with tomography (based on reflections, slopes or
 473 waveforms (Sambolian, Operto, et al., 2019; Takougang & Calvert, 2011)) or FWI (Qin &
 474 Singh, 2017). In case of deep-water environment, re-datuming of the data to the seafloor
 475 with downward continuation approach (Gras et al., 2019) can also be considered to extract
 476 the diving waves for FWI. Moreover MCS data can be used for depth-migration with the
 477 model derived from the OBS data as background velocity field (Gorszczyk et al., 2019). This
 478 can bring additional high-resolution geological information, but also validate the correctness
 479 of the velocity model, which shall flatten reflections in the common image gathers.

480 **5.1.3 Frequency content of the data**

481 One of the main issue for robust FWI is deficiency of the low-frequencies in the active
 482 seismic data. This results form the difficulty of generating and recording of the signal energy
 483 below ~ 2 Hz. Although some enhancement of the low-frequency content in the data can be
 484 done at the preprocessing stage (e.g. spiking deconvolution, Yilmaz (2001)), the presence of
 485 noise (typical for the OBS data) can hamper the processing and lead to the FWI application
 486 without access to the low-frequency data. In the first line, this translates to the cycle-
 487 skipping issue, which can be solved with building an accurate starting model - difficult to
 488 derive from the sparse data. Alternatively, one can use more convex misfit functions, which
 489 are able to introduce a tomographic-like model update and fill the resolution gap between
 490 the long-wavelength structure of the initial model and the short wavelength perturbations
 491 resulting from the FWI of high-frequency data. Secondly, even when the cycle-skipping
 492 problem is solved, the spatial aliasing resulting not only from the coarse OBS sampling, but
 493 also from the lack of low-frequency content, can lead to the artefacts in the model update -
 494 in particular in the shallow subsurface. MCS data can be utilised to mitigate this issue in
 495 the manner mentioned above. One can also consider more robust gradient regularisations,
 496 which are able to produce a broadband wavenumber update of the model, or which can
 497 utilise a priori information about the underlying structure (Aghamiry et al., 2019; Peters &
 498 Herrmann, 2017; Trinh et al., 2017).

499 **5.1.4 2D or 3D OBS deployment**

500 Most of the active seismic OBS experiments conducted up to now were performed along
 501 the 2D profiles. The SFJ-OBS 2001 dataset, which we processed here was acquired with 100
 502 OBS stations deployed with 1 km spacing. Such a setting can be consider as a very dense
 503 deployment for the academic surveys. However, with 100 receivers one can perform a 3D
 504 OBS acquisition, covering $100 \text{ km} \times 20 \text{ km}$ area assuming $\sim 5 \text{ km}$ OBS spacing. This will
 505 still be considered a dense academic acquisition setting. Which of those two is more optimal
 506 form the point of view of FWI application? Can the 3D deployment with few shooting
 507 lines compensate for the sparser OBS spacing - taking advantage form narrow-azimuthal
 508 coverage? Was SFJ-OBS experiment oversampling the underlying structure? Or maybe the

509 3D setup will undersample the model space? Perhaps some of those issues might be clarified,
 510 if we will start performing more routinely the realistic numerical studies, to optimize the
 511 active seismic OBS experiments for the purpose of further FWI processing. This might
 512 lead to the development of the leading-edge acquisition approaches which will allow to fully
 513 exploit the potential of the crustal-scale FWI from the OBS data.

514 **5.2 Initial model-building**

515 We have presented how the GSOT misfit function can guide FWI into a correct solution
 516 starting from a crude 1D model. This raises the question of whether we still need to build
 517 accurate initial models, for example, with FAT. From a practical point of view, there are a
 518 few aspects to consider. In terms of computational time, traveltimes tomography is cheap.
 519 However, precise picking of the first breaks might be uncertain and time consuming when the
 520 interpretation of the first-arrival phases is ambiguous. This, in turn, might require iterative
 521 refinement of the picked traveltimes and repetitive FAT inversion (Górszczyk et al., 2017).
 522 On the other hand, time-windowing of the data during our FWI still requires the first-arrival
 523 traveltimes to be provided. There is, however, a fundamental difference here. The precision
 524 of the traveltimes defining the time window that we are using can be significantly lower
 525 compared to what would be usually required to design a correct (in the sense of satisfying
 526 the cycle-skipping) velocity model through FAT (even when uncertainty of picks is taken
 527 into account). From these premises, we could envision a workflow that would consist of
 528 first running a computationally cheap FAT with potentially inaccurate first-arrival picks to
 529 obtain a first smooth velocity model. This smooth velocity model, even if it were not to
 530 satisfy the cycle-skipping criteria could serve as an initial velocity model for a GSOT-based
 531 FWI. In such a way, one could start FWI with a better initial model than a simple 1D
 532 profile and still benefit from the improved convexity of the GSOT misfit function and less
 533 expanded FWI workflow comparing to what we presented. Such an approach might be
 534 especially important from the perspective of large-scale 3D FWI.

535 **5.3 Computational cost of the GSOT approach**

536 The computation of the GSOT misfit requires the solution of an optimal assignment
 537 problem for each trace. By virtue of the auction algorithm, the solution of such a problem
 538 can be relatively fast (less than 1 s on a single core), for instances of a problem with a
 539 number of discrete point N inferior to 1000. This is approximately the order of magnitude
 540 we attain for FWI applications at crustal scale, considering a resampling of the data in
 541 time. For our field data application, we decimate the data in time by a factor of 3. This
 542 leads to N equal to 1334 discrete points per trace and the time step equal to 0.015 s.
 543 Such a data-sampling is still much denser than that required by the Nyquist law (for the
 544 frequency range we consider here) and introduces negligible differences into the gradient
 545 compared to the original sampling. In addition, to speed up the solution of the optimal
 546 assignment problem, we have implemented a localization of the algorithm that depends on
 547 the maximum expected time shift (defined by the parameter τ). We thus observe that the
 548 computation time for a gradient using the GSOT misfit function is between 93 s and 66 s
 549 depending on whether τ is taken equal to 4 s or 0.2 s. With a reference time of 49 s in the L^2
 550 case, this represents an additional cost between 89% and 34%. Let us also mention that a
 551 complexity analysis reveals that the computational cost of GSOT scales with \mathcal{O}^3 , where \mathcal{O} is
 552 the reference frequency of the data, while the incident and adjoint modeling steps required
 553 for the FWI gradient building scale with \mathcal{O}^4 . This indicates a favorable trend for the higher
 554 costs associated with the GSOT strategy for larger-scale applications.

555 **5.4 Acoustic vs visco-acoustic modeling**

556 Unlike frequency domain modeling, for which the implementation of the attenuation is
 557 done through the introduction of a complex-valued velocity, the time-domain visco-acoustic

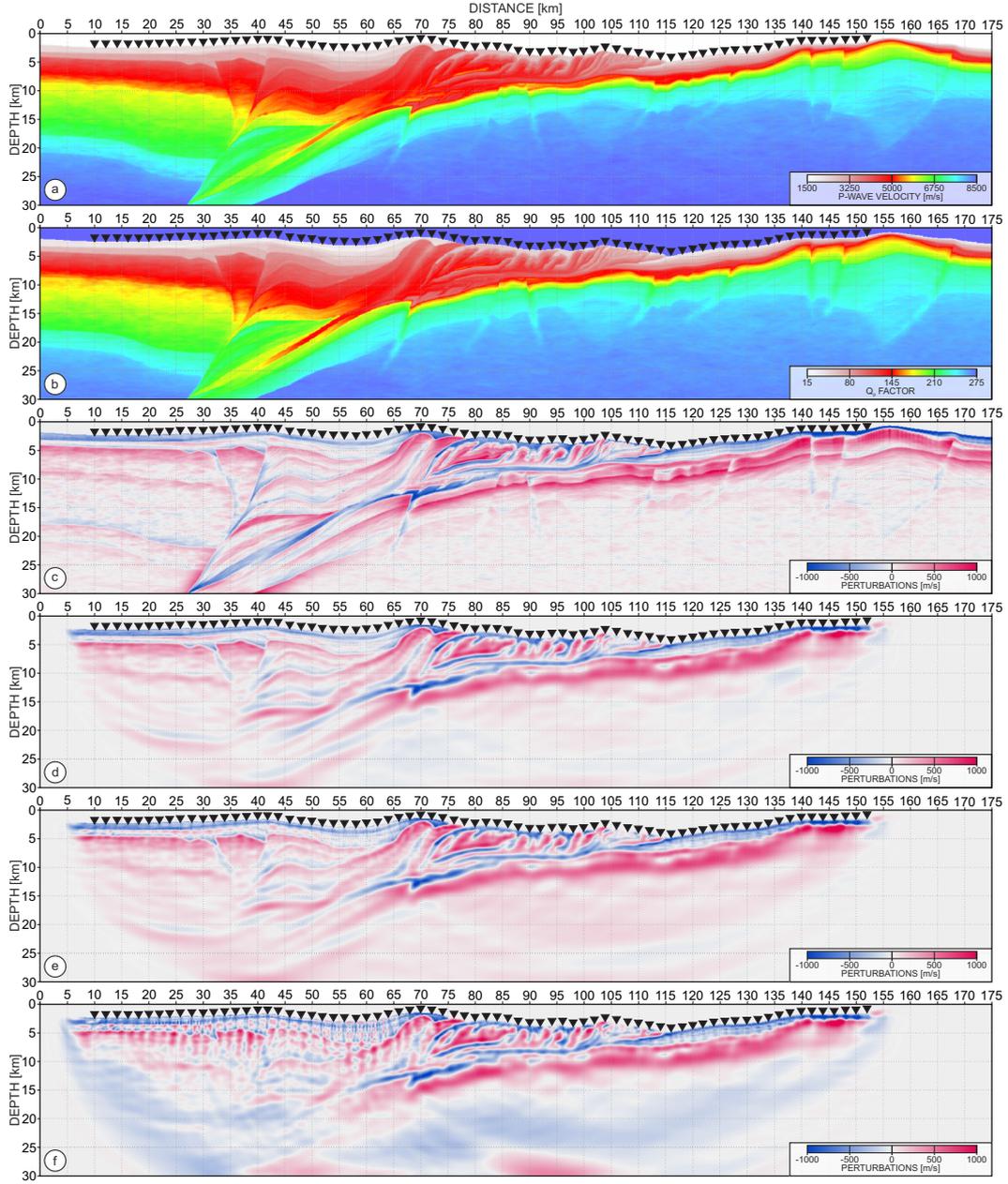


Figure 10. (a)-(b) True V_p and Q_p models; (c) Velocity perturbations - difference between true and initial V_p model; Reconstructed velocity perturbations after FWI with visco-acoustic modeling using (d) true Q_p model; (e) constant $Q_p=100$ in solid part and $Q_p=10000$ in the water column; (f) constant $Q_p=10000$.

558 modeling comes at an extra computing cost. In our implementation we use the checkpoint-
 559 assisted reverse forward simulation (CARFS) strategy proposed by P. Yang et al. (2016b).
 560 Despite the fact that this approach is more computationally efficient than conventional
 561 checkpointing our visco-acoustic modeling is still approximately 3 times more expensive
 562 than its acoustic counterpart. On the other hand, previous studies report that better FWI
 563 results can be obtained, even when using a crude attenuation model for field data FWI
 564 applications (e.g. Kurzmann et al. (2013); Górszczyk et al. (2017); Operto and Miniussi
 565 (2018)). This might be especially true for the ultralong-offset data inversion where number

566 of propagated wavelets is large.
567 Since the GO_3D_OBS model (Górszczyk & Operto, 2021) also contains the definition
568 of the Q_p parameter, we can easily run a synthetic analysis of the FWI with different
569 attenuation scenarios. We consider the same acquisition parameters as for the synthetic
570 study presented before. Figure 10a-b shows the true V_p and Q_p models used to generate the
571 data. Note that the Q_p values span between 15 and 275 in the solid part and are constant
572 ($Q_p=10000$) for the water column. Figure 10c shows the difference between the true and
573 the initial V_p model (smooth version of the true model) used in this exercise. We run 3
574 different GSOT-based FWI tests (200 iterations) where we aim at the V_p reconstruction
575 using visco-acoustic modeling with different Q_p models. We also keep the true ρ model (not
576 presented here) both for the modeling and the inversion step. In the first scenario we use
577 the true Q_p model (Figure 10b), and not surprisingly, the reconstructed perturbations in
578 Figure 10d match well with those from Figure 10c. In the second test, we use a simple
579 Q_p model where we set the $Q_p=100$ in the solid part and $Q_p=10000$ in the water column.
580 The corresponding perturbation model is presented in Figure 10e. We start observing an
581 indication of an acquisition footprint in the shallow sedimentary part for which low Q_p
582 values are present in the true model. Moreover generally faster perturbations are observed
583 in the deeper model part. Despite those differences, the reconstructed structure is still close
584 to the one presented in Figure 10d. In the third test we use the constant $Q_p=10000$ model
585 (approximately no attenuation). The resulting velocity structure in Figure 10f is clearly
586 damaged due to the lack of accounting for the viscous effect during the waveform modeling.
587 In our field data study, even though we invert the data within a relatively low frequency band
588 we use a constant Q_p model in the solid part ($Q_p=200$). The final results from FWI with
589 visco-acoustic modeling show less artifacts in the reconstructed velocity model and provide
590 better data fitting (from 5% to 15% depending on the OBS) than the FWI with acoustic
591 modeling. One can expect that this difference might be even more significant when moving
592 to higher frequency ranges and/or strongly attenuating media. Therefore, considering a
593 simple attenuation model during FWI - even as a passive parameter - clearly improves the
594 final inversion results.

595 5.5 Time- versus frequency-domain crustal-scale FWI

596 Since our GSOT misfit function relies on the comparison of seismograms in the time
597 domain, we consequently apply the time-domain implementation of FWI. Indeed, the com-
598 putational efficiency of frequency-domain crustal-scale FWI for processing the 2D stationary
599 receiver data was exploited in previous studies (e.g. Sirgue and Pratt (2004); Brenders and
600 Pratt (2007b); Ravaut et al. (2004); Operto et al. (2006); Kamei et al. (2012); Malinowski
601 et al. (2011); Górszczyk et al. (2017) . Furthermore, the development of more robust direct
602 solvers coupled with low-rank techniques currently allows undertaking 3D imaging problems
603 at the exploration scale with frequency-domain FWI (Operto et al., 2015; Li et al., 2019).
604 However, their extension to the large 3D imaging problems is still prohibitive in terms
605 of memory demand, volume of computation and communication. Moreover, the typically
606 sparse OBS deployments translate to a limited number of reciprocal sources that have to
607 be processed within the large computing domain. Consequently, efficient parallelism over
608 the sources and over the subdomains can be implemented to speed up the processing in the
609 time domain. Therefore, while the frequency-domain FWI is still contemporaneously eas-
610 ily utilized for 2D crustal-scale OBS data, the perspective of changing the seismic imaging
611 paradigm toward the 3D high-resolution model reconstruction requires robust FWI schemes
612 designed in the time domain.

613 6 Conclusions

614 We have illustrated how the GSOT misfit function can relax the initial model design
615 for crustal-scale FWI from OBS data. In both cases of synthetic and field studies, we were
616 able to guide GSOT-based FWI toward a correct solution starting from a simple 1D model.

617 Significantly better convexity of the GSOT than that of the L^2 misfit function makes it
618 possible to handle the kinematic errors in the initial model that are of the order of a few
619 cycles. Consequently, this approach allows for saving efforts related to the precise picking
620 and inversion of the first-arrival traveltimes from the ultralong-offset data, when the initial
621 FWI model is obtained with FAT.

622 Prospectively, the development of the misfit functions with improved convexity, such as the
623 GSOT, can be expected to stimulate the more routine FWI applications to the wide-angle
624 OBS data. Indeed, with the mitigation of the cycle-skipping problem, the feasibility of the
625 regional FWI would rely more on the ability to use a more realistic physics approximation,
626 more robust regularization techniques, as well as the availability of the seismic data recorded
627 by the sufficiently dense OBS deployments. In particular, the 3D areal OBS acquisitions
628 would essentially improve the illumination of the deep structures, thus making the whole
629 inversion process better constrained. Therefore, the step forward regarding the way that we
630 image the lithospheric targets entails investigating further challenges within the framework
631 of the 3D crustal-scale FWI.

632 **Acknowledgments**

633 This study was partially funded by: (i) the SEISCOPE consortium
634 (<http://seiscope2.osug.fr>), sponsored by AKERBP, CGG, CHEVRON, EQUINOR,
635 EXXON-MOBIL, JGI, SHELL, SINOPEC, SISPROBE, and TOTAL, and (ii) the
636 Polish National Science Center (grant no: 2019/33/B/ST10/01014). The study was
637 granted access to the HPC PL-Grid Infrastructure (grant id: 3dwind2). The access
638 to SFJ-OBS field dataset can be negotiated with JAMSTEC agency and is accessi-
639 ble through their database (http://www.jamstec.go.jp/obsmcs_db/e/index.html). The
640 GO_3D_OBS synthetic model can be accessed through the official data portal of IG PAS
641 (https://dataportal.igf.edu.pl/dataset/go_3d_obs).

642 **References**

- 643 Aghamiry, H., Gholami, A., & Operto, S. (2018). Improving full-waveform inversion based
644 on wavefield reconstruction via Bregman iterations. In *Expanded abstracts, 80th annual*
645 *eage meeting (copenhagen)*.
- 646 Aghamiry, H., Gholami, A., & Operto, S. (2019). Admm-based multi-parameter wavefield
647 reconstruction inversion in VTI acoustic media with TV regularization. *Geophysical*
648 *Journal International*, 219(2), 1316–1333. doi: <https://doi.org/10.1093/gji/ggz369>
- 649 Alkhalifah, T., & Choi, Y. (2012). Taming waveform inversion non-linearity through phase
650 unwrapping of the model and objective functions. *Geophysical Journal International*,
651 191, 1171–1178.
- 652 Bertsekas, D. P., & Castanon, D. (1989). The auction algorithm for the transportation
653 problem. *Annals of Operations Research*, 20(1), 67–96. doi: 10.1007/BF02216923
- 654 Borisov, D., Gao, F., Williamson, P., & Tromp, J. (2019). Application of 2d full-waveform
655 inversion on exploration land data. *Geophysics*, 85(2), 1–57.
- 656 Borisov, D., Modrak, R., Gao, F., & Tromp, J. (2017). Spectral-element based 3D elastic
657 full-waveform inversion of surface waves in the presence of irregular topography using
658 an envelope-based misfit function. *Geophysics*.
- 659 Bozdağ, E., Trampert, J., & Tromp, J. (2011). Misfit functions for full waveform inver-
660 sion based on instantaneous phase and envelope measurements. *Geophysical Journal*
661 *International*, 185(2), 845–870.
- 662 Brenders, A. J., & Pratt, R. G. (2007a). Efficient waveform tomography for lithospheric
663 imaging: implications for realistic 2D acquisition geometries and low frequency data.
664 *Geophysical Journal International*, 168, 152–170.
- 665 Brenders, A. J., & Pratt, R. G. (2007b). Full waveform tomography for lithospheric imaging:
666 results from a blind test in a realistic crustal model. *Geophysical Journal International*,
667 168, 133–151.
- 668 Byrd, R. H., Lu, P., & Nocedal, J. (1995). A limited memory algorithm for bound con-
669 strained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16,
670 1190–1208.
- 671 Carotti, D., Hermant, O., Masclet, S., Reinier, M., Messud, J., Sedova, A., & Lam-
672 baré, G. (2020). Optimal transport full waveform inversion - applications. In
673 *82th Annual EAGE Meeting (Amsterdam)*. European Association of Geoscientists &
674 Engineers. Retrieved from [https://www.earthdoc.org/content/papers/10.3997/](https://www.earthdoc.org/content/papers/10.3997/2214-4609.202011288)
675 [2214-4609.202011288](https://www.earthdoc.org/content/papers/10.3997/2214-4609.202011288) doi: <https://doi.org/10.3997/2214-4609.202011288>
- 676 Cerjan, C., Kosloff, D., Kosloff, R., & Reshef, M. (1985). A nonreflecting boundary condition
677 for discrete acoustic and elastic wave equations. *Geophysics*, 50(4), 2117–2131.
- 678 Chen, F., & Peter, D. (2018, June). Constructing Misfit Function for Full Waveform Inver-
679 sion Based on Sliced Wasserstein Distance. In *80th EAGE conference and exhibition*
680 *2018*. EAGE Publications BV. doi: 10.3997/2214-4609.201801030
- 681 Choi, Y., & Alkhalifah, T. (2011). Frequency-domain waveform inversion using the un-
682 wrapped phase. *SEG Technical Program Expanded Abstracts*, 30(1), 2576–2580.
- 683 Choi, Y., & Shin, C. (2008). Frequency-Domain Elastic Full Waveform Inversion Using
684 the New Pseudo-Hessian Matrix: Experience Of Elastic Marmousi 2 Synthetic Data.
685 *Bulletin of the Seismological Society of America*, 98(5), 2402–2415.
- 686 Czuba, W. (2016, November). 3-d seismic tomographic modelling of the crustal struc-
687 ture of northwestern svalbard based on deep seismic soundings. *Geophysical Journal*
688 *International*, 208(1), 508–520. doi: 10.1093/gji/ggw418
- 689 Davy, R. G., Morgan, J. V., Minshull, T. A., Bayrakci, G., Bull, J. M., Klaeschen, D., ...
690 Cresswell, D. (2017, October). Resolving the fine-scale velocity structure of continental
691 hyperextension at the Deep Galicia Margin using full-waveform inversion. *Geophysical*
692 *Journal International*, 212(1), 244–263. Retrieved from [https://doi.org/10.1093/](https://doi.org/10.1093/gji/ggx415)
693 [gji/ggx415](https://doi.org/10.1093/gji/ggx415) doi: 10.1093/gji/ggx415
- 694 Engquist, B., & Froese, B. D. (2014). Application of the Wasserstein metric to seismic
695 signals. *Communications in Mathematical Science*, 12(5), 979–988.

- 696 Górszczyk, A., & Operto, S. (2021). GO_3D_OBS: The multi-parameter benchmark
697 geomodel for seismic imaging method assessment and next-generation 3D survey
698 design (version 1.0). *Geoscientific Model Development*, *14*(3), 1773–1799. doi:
699 <https://doi.org/10.5194/gmd-14-1773-2021>
- 700 Górszczyk, A., Operto, S., & Malinowski, M. (2017, June). Toward a robust workflow
701 for deep crustal imaging by FWI of OBS data: The eastern nankai trough revisited.
702 *Journal of Geophysical Research: Solid Earth*, *122*(6), 4601–4630. doi: 10.1002/
703 2016jb013891
- 704 Górszczyk, A., Operto, S., Schenini, L., & Yamada, Y. (2019). Crustal-scale depth imaging
705 via joint FWI of OBS data and PSDM of MCS data: a case study from the eastern
706 nankai trough. *Solid Earth*, *10*, 765–784. doi: <https://doi.org/10.5194/se-10-765-2019>
- 707 Gras, C., Dagnino, D., Jiménez-Tejero, C. E., Meléndez, A., Sallarès, V., & Ranero, C. R.
708 (2019, October). Full-waveform inversion of short-offset, band-limited seismic data in
709 the alboran basin (SE iberia). *Solid Earth*, *10*(6), 1833–1855. doi: 10.5194/se-10-1833
710 -2019
- 711 He, W., Brossier, R., Métivier, L., & Plessix, R.-É. (2019). Land seismic multi-parameter full
712 waveform inversion in elastic VTI media by simultaneously interpreting body waves
713 and surface waves with an optimal transport based objective function. *Geophysical*
714 *Journal International*, *219*(3), 1970–1988. doi: 10.1093/gji/ggz414
- 715 Hicks, G. J. (2002). Arbitrary source and receiver positioning in finite-difference schemes
716 using Kaiser windowed sinc functions. *Geophysics*, *67*, 156–166.
- 717 Kamei, R., Pratt, R. G., & Tsuji, T. (2012). Waveform tomography imaging of a megasplay
718 fault system in the seismogenic Nankai subduction zone. *Earth and Planetary Science*
719 *Letters*, *317–318*, 343–353.
- 720 Kamei, R., Pratt, R. G., & Tsuji, T. (2014, April). Misfit functionals in Laplace-Fourier
721 domain waveform inversion, with application to wide-angle ocean bottom seismograph
722 data. *Geophysical Prospecting*, *62*(5), 1054–1074. doi: 10.1111/1365-2478.12127
- 723 Kantorovich, L. (1942). On the transfer of masses. *Dokl. Acad. Nauk. USSR*, *37*, 7–8.
- 724 Korenaga, J., Holbrook, W. S., Kent, G. M., Kelemen, P. B., Detrick, R. S., Larsen, H. C.,
725 ... Dahl-Jensen, T. (2000). Crustal structure of the southeast greenland margin from
726 joint refraction and reflection seismic tomography. *Journal of Geophysical Research*,
727 *105*, 21591–21614.
- 728 Kurzmann, A., Przewindowska, A., Kohn, D., & Bohlen, T. (2013). Acoustic full waveform
729 tomography in the presence of attenuation: a sensitivity analysis. *Geophysical Journal*
730 *International*, *195*(2), 985–1000.
- 731 Li, Y., Brossier, R., & Métivier, L. (2019). 3D frequency-domain elastic wave modeling using
732 spectral element method with a parallel direct linear solver. In *Expanded abstracts,*
733 *81th annual EAGE conference & exhibition, london* (p. Th R06 05). EAGE. doi:
734 10.3997/2214-4609.201901534
- 735 Luo, J., & Wu, R.-S. (2015). Seismic envelope inversion: reduction of local minima and
736 noise resistance. *Geophysical Prospecting*, *63*(3), 597–614.
- 737 Luo, S., & Sava, P. (2011). A deconvolution-based objective function for wave-equation
738 inversion. *SEG Technical Program Expanded Abstracts*, *30*(1), 2788–2792. Retrieved
739 from <http://link.aip.org/link/?SGA/30/2788/1> doi: 10.1190/1.3627773
- 740 Luo, Y., & Schuster, G. T. (1991). Wave-equation travelttime inversion. *Geophysics*, *56*(5),
741 645–653.
- 742 Malinowski, M., Operto, S., & Ribodetti, A. (2011). High-resolution seismic attenua-
743 tion imaging from wide-aperture onshore data by visco-acoustic frequency-domain full
744 waveform inversion. *Geophysical Journal International*, *186*(3), 1179–1204.
- 745 Messud, J., & Sedova, A. (2019). Multidimensional optimal transport for 3d FWI: Demon-
746 stration on field data. In *Expanded abstracts, 81th annual eage meeting (london)*.
- 747 Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J. (2018). Optimal
748 transport for mitigating cycle skipping in full waveform inversion: a graph space trans-
749 form approach. *Geophysics*, *83*(5), R515–R540. Retrieved from [https://doi.org/](https://doi.org/10.1190/geo2017-0807.1)
750 [10.1190/geo2017-0807.1](https://doi.org/10.1190/geo2017-0807.1) doi: 10.1190/geo2017-0807.1

- 751 Métivier, L., & Brossier, R. (2016). The SEISCOPE optimization toolbox: A large-scale
752 nonlinear optimization library based on reverse communication. *Geophysics*, *81*(2),
753 F11-F25.
- 754 Métivier, L., Brossier, R., Mérigot, Q., & Oudet, E. (2019a). A graph space optimal
755 transport distance as a generalization of L^p distances: application to a seismic imaging
756 inverse problem. *Inverse Problems*, *35*(8), 085001.
- 757 Métivier, L., Brossier, R., Mérigot, Q., & Oudet, E. (2019b). Graph space optimal transport
758 for FWI: Auction algorithm, application to the 2d valhall case study. In *Expanded*
759 *abstracts, 81th annual EAGE conference & exhibition, london* (p. Tu R08 03). EAGE.
760 doi: 10.3997//2214-4609.201900870
- 761 Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J. (2016a). Measuring the
762 misfit between seismograms using an optimal transport distance: Application to full
763 waveform inversion. *Geophysical Journal International*, *205*, 345–377.
- 764 Métivier, L., Brossier, R., Mérigot, Q., Oudet, E., & Virieux, J. (2016b). An optimal
765 transport approach for seismic tomography: Application to 3D full waveform inversion.
766 *Inverse Problems*, *32*(11), 115008.
- 767 Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie*
768 *Royale des Sciences de Paris*.
- 769 Morgan, J., Warner, M., Bell, R., Ashley, J., Barnes, D., Little, R., . . . Jones, C. (2013).
770 Next-generation seismic experiments: wide-angle, multi-azimuth, three-dimensional,
771 full-waveform inversion. *Geophysical Journal International*, *195*, 1657–1678.
- 772 Nakanishi, A., Shiobara, H., Hino, R., Kodaira, S., Kanazawa, T., & Shinamura, H. (1998).
773 Detailed subduction structure across the eastern Nankai Trough obtained from ocean
774 bottom seismograph profiles. *Journal of Geophysical Research*, *103*(B11), 27151–
775 27168.
- 776 Nocedal, J. (1980, September). Updating quasi-Newton matrices with limited storage. *Math-*
777 *ematics of Computation*, *35*(151), 773–773. doi: 10.1090/s0025-5718-1980-0572855-7
- 778 Nocedal, J., & Wright, S. J. (2006). *Numerical optimization* (2nd ed.). Springer.
- 779 Operto, S., Brossier, R., Gholami, Y., Métivier, L., Prioux, V., Ribodetti, A., & Virieux, J.
780 (2013). A guided tour of multiparameter full waveform inversion for multicomponent
781 data: from theory to practice. *The Leading Edge, Special section Full Waveform*
782 *Inversion*(September), 1040-1054.
- 783 Operto, S., & Miniussi, A. (2018). On the role of density and attenuation in 3D multi-
784 parameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the
785 North Sea. *Geophysical Journal International*, *213*, 2037–2059. doi: 10.1093/gji/
786 ggy103
- 787 Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., . . . Virieux,
788 J. (2015). Efficient 3-D frequency-domain mono-parameter full-waveform inversion of
789 ocean-bottom cable data: application to Valhall in the visco-acoustic vertical trans-
790 verse isotropic approximation. *Geophysical Journal International*, *202*(2), 1362-1391.
- 791 Operto, S., Virieux, J., Dessa, J. X., & Pascal, G. (2006). Crustal imaging from multifold
792 ocean bottom seismometers data by frequency-domain full-waveform tomography: ap-
793 plication to the eastern Nankai trough. *Journal of Geophysical Research*, *111*(B09306),
794 doi:10.1029/2005JB003835.
- 795 Peters, B., & Herrmann, F. J. (2017, 01). Constraints versus penalties for
796 edge-preserving full-waveform inversion. *The Leading Edge*, *36*(1), 94-100.
797 Retrieved from [https://www.slim.eos.ubc.ca/Publications/Public/Journals/
798 TheLeadingEdge/2016/peters2016cvp/peters2016cvp.html](https://www.slim.eos.ubc.ca/Publications/Public/Journals/TheLeadingEdge/2016/peters2016cvp/peters2016cvp.html) doi: 10.1190/
799 tle36010094.1
- 800 Pladys, A., Brossier, R., & Métivier, L. (2020). Graph space optimal transport based FWI:
801 3D OBC valhall case study. In *Seg technical program expanded abstracts 2020*.
- 802 Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a
803 functional with geophysical applications. *Geophysical Journal International*, *167*(2),
804 495-503.
- 805 Plessix, R. E., & Perkins, C. (2010). Full waveform inversion of a deep water ocean bottom

- 806 seismometer dataset. *First Break*, 28, 71-78.
- 807 Poncet, R., Messud, J., Bader, M., Lambaré, G., Viguier, G., & Hidalgo, C. (2018). Fwi
808 with optimal transport: a 3D implementation and an application on a field dataset.
809 In *Expanded abstracts, 80th annual eage meeting (copenhagen)*.
- 810 Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part I: theory
811 and verification in a physical scale model. *Geophysics*, 64, 888–901.
- 812 Pratt, R. G. (2008). Waveform tomography - successes, cautionary tales, and future direc-
813 tions. In *Presented at the 70th annual EAGE conference & exhibition, roma* (p. WO11
814 - Full-Waveform Inversion: current status and perspectives).
- 815 Provenzano, G., Brossier, R., & Métivier, L. (2020). Joint FWI of diving and reflected
816 waves using a graph space optimal transport distance: synthetic tests on limited-offset
817 surface seismic data. In *Seg technical program expanded abstracts 2020*.
- 818 Qin, Y., & Singh, S. C. (2017, apr). Detailed seismic velocity of the incoming subducting sed-
819 iments in the 2004 great sumatra earthquake rupture zone from full waveform inversion
820 of long offset seismic data. *Geophysical Research Letters*, 44(7), 3090–3099. Retrieved
821 from <https://doi.org/10.1002/2016gl072175> doi: 10.1002/2016gl072175
- 822 Qiu, L., Ramos-Martínez, J., Valenciano, A., Yang, Y., & Engquist, B. (2017). Full-
823 waveform inversion with an exponentially encoded optimal-transport norm. In *Seg*
824 *technical program expanded abstracts 2017* (p. 1286-1290).
- 825 Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., & dell’Aversana, P. (2004).
826 Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data
827 by frequency-domain full-wavefield inversions: application to a thrust belt. *Geophysical*
828 *Journal International*, 159, 1032–1056.
- 829 Sambolian, S., Gorszcyk, A., Operto, S., Ribodetti, A., F., B. T., , & Virieux, J. (2019).
830 Building initial model for FWI from ultra long-offset obn data by first-arrival travel-
831 time + slope tomography. In *81th annual eage meeting (london)*.
- 832 Sambolian, S., Operto, S., Ribodetti, A., Tavakoli, B., & Virieux, J. (2019). Parsimonious
833 slope tomography based on eikonal solvers and the adjoint-state method. *Geophysical*
834 *Journal International*, 218(1), 456–478. doi: 10.1093/gji/ggz150
- 835 Sedova, A., Messud, J., Prigent, H., Royle, G., & Lambaré, G. (2019). Acoustic land full
836 waveform inversion on a broadband land dataset: the impact of optimal transport. In
837 *Expanded abstracts, 81th annual eage meeting (london)*.
- 838 Shin, C., Yoon, K., Marfurt, K. J., Park, K., Yang, D., Lim, H. Y., . . . Shin, S. (2001). Effi-
839 cient calculation of a partial derivative wavefield using reciprocity for seismic imaging
840 and inversion. *Geophysics*, 66(6), 1856–1863.
- 841 Shutova, S., Carotti, D., & Hermant, O. (2020). Extending the potential of acoustic optimal
842 transport fwi in the south of oman. In *82th Annual EAGE Meeting (Amsterdam)*.
843 European Association of Geoscientists & Engineers. Retrieved from [https://www](https://www.earthdoc.org/content/papers/10.3997/2214-4609.202012113)
844 [.earthdoc.org/content/papers/10.3997/2214-4609.202012113](https://www.earthdoc.org/content/papers/10.3997/2214-4609.202012113) doi: [https://doi](https://doi.org/10.3997/2214-4609.202012113)
845 [.org/10.3997/2214-4609.202012113](https://doi.org/10.3997/2214-4609.202012113)
- 846 Sirgue, L., & Pratt, R. G. (2004). Efficient waveform inversion and imaging : a strategy for
847 selecting temporal frequencies. *Geophysics*, 69(1), 231–248.
- 848 Stopin, A., Plessix, R.-E., & Al Abri, S. (2014). Multiparameter waveform inversion of a
849 large wide-azimuth low-frequency land data set in Oman. *Geophysics*, 79(3), WA69-
850 WA77.
- 851 Sun, B., & Alkhalifah, T. (2019). The application of an optimal transport to a pre-
852 conditioned data matching function for robust waveform inversion. *GEOPHYSICS*,
853 84(6), R935-R957. Retrieved from <https://doi.org/10.1190/geo2018-0413.1> doi:
854 10.1190/geo2018-0413.1
- 855 Takougang, E. M. T., & Calvert, A. J. (2011). Application of waveform tomography to
856 marine seismic reflection data from the Queen Charlotte Basin of western canada.
857 *Geophysics*, 76(2), B55-B70.
- 858 Thorpe, M., Park, S., Kolouri, S., Rohde, G., & Slepčev, D. (2016, sep). A transportation
859 l^p distance for signal analysis. *ArXiv e-prints*.
- 860 Treister, E., & Haber, E. (2017). Full waveform inversion guided by travelttime tomography.

- 861 *SIAM Journal on Scientific Computing*, 39(5), S587-S609.
- 862 Trinh, P. T., Brossier, R., Métivier, L., Virieux, J., & Wellington, P. (2017). Bessel smoothing filter for spectral element mesh. *Geophysical Journal International*, 209(3), 1489–1512. Retrieved from [+http://dx.doi.org/10.1093/gji/ggx103](http://dx.doi.org/10.1093/gji/ggx103) doi: 10.1093/gji/ggx103
- 863
- 864
- 865
- 866 Tromp, J. (2019). Seismic wavefield imaging of earth’s interior across scales. *Nature Reviews Earth & Environment*, 1–14.
- 867
- 868 van Leeuwen, T., & Herrmann, F. J. (2013). Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, 195(1), 661–667. doi: doi:10.1093/gji/ggt258
- 869
- 870
- 871 van Leeuwen, T., & Mulder, W. A. (2010). A correlation-based misfit criterion for wave-equation travelttime tomography. *Geophysical Journal International*, 182(3), 1383–1394.
- 872
- 873
- 874 Virieux, J., & Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6), WCC1-WCC26.
- 875
- 876 Warner, M., & Guasch, L. (2016). Adaptive waveform inversion: Theory. *Geophysics*, 81(6), R429-R445.
- 877
- 878 Yang, P., Brossier, R., Métivier, L., & Virieux, J. (2016a, October). A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium. *Geophysical Journal International*, 207(1), 129-149. doi: 10.1093/gji/ggw262
- 879
- 880
- 881 Yang, P., Brossier, R., Métivier, L., & Virieux, J. (2016b). Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method. *Geophysics*, 81(6), R349-R362. doi: 10.1190/geo2016-0082.1
- 882
- 883
- 884 Yang, P., Brossier, R., Métivier, L., Virieux, J., & Zhou, W. (2018). A Time-Domain Preconditioned Truncated Newton Approach to Multiparameter Visco-acoustic Full Waveform Inversion. *SIAM Journal on Scientific Computing*, 40(4), B1101-B1130. doi: 10.1137/17M1126126
- 885
- 886
- 887
- 888 Yang, Y., & Engquist, B. (2018). Analysis of optimal transport and related misfit functions in full-waveform inversion. *GEOPHYSICS*, 83(1), A7-A12. doi: 10.1190/geo2017-0264.1
- 889
- 890
- 891 Yang, Y., Engquist, B., Sun, J., & Hamfeldt, B. F. (2018). Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion. *Geophysics*, 83(1), R43-R62. Retrieved from <https://doi.org/10.1190/geo2016-0663.1> doi: 10.1190/geo2016-0663.1
- 892
- 893
- 894
- 895 Yilmaz, O. (2001). *Seismic data analysis*. Society of Exploration Geophysicists: processing, inversion and interpretation of seismic data.
- 896
- 897 Yu, Z., Li, J., Ding, W., Zhang, J., Ruan, A., & Niu, X. (2016, October). Crustal structure of the southwest subbasin, south china sea, from wide-angle seismic tomography and seismic reflection imaging. *Marine Geophysical Research*, 38(1-2), 85–104. doi: 10.1007/s11001-016-9284-1
- 898
- 899
- 900
- 901 Zelt, C., & Barton, P. J. (1998). Three-dimensional seismic refraction tomography: a comparison of two methods applied to data from the Faeroe basin. *Journal of Geophysical Research*, 103(B4), 7187–7210.
- 902
- 903
- 904 Zhu, H., & Fomel, S. (2016). Building good starting models for full-waveform inversion using adaptive matching filtering misfit. *Geophysics*, 81(5), U61-U72. doi: 10.1190/geo2015-0596.1
- 905
- 906