Robust and efficient waveform-based velocity-model building by optimal transport in the pseudotime domain: Methodology

Giuseppe Provenzano, Romain Brossier, Ludovic Métivier

Robust and efficient waveform-based velocity-model-building by optimal-transport in the pseudotime domain: methodology

Giuseppe Provenzano, Romain Brossier *, Ludovic Métivier †*

ABSTRACT

Full waveform inversion (FWI) aims at a broadband reconstruction of the subsurface physical properties by fitting the entire recorded wavefield. In realistic exploration seismic surveys, however, conventional FWI often fails to retrieve the deep velocity model due to the limited penetration depth of diving waves. Joint FWI (JFWI) unifies reflection-waveform inversion (RWI) and early-arrival waveform inversion (EWI) to reconstruct simutaneously the shallow and deep subsurface kinematics. However, a number of factors limit the appeal of JFWI velocity-model-building: 1) conflict between fixed reflectivity and evolving kinematics, creating phase ambiguity at short offsets; 2) susceptibility to cycle skipping at mid-to-long offsets, thus reliance on the quality of the starting model; 3) cost of building and updating the reflective model. We present a fully operational JFWI-based methodology that systematically addresses the aforementioned issues. JFWI is re-formulated in the pseudotime domain, in order to enforce consistency between velocity and reflectivity in a cost-effective fashion, without repeated least-square migrations. A JFWI graph space optimal transport (GSOT) objective function is designed to avert cycle skipping, while non-uniqueness is mitigated at no extra cost by smoothing the velocity gradient along the structures extracted from the reflective model. A dedicated asymptotic-based preconditioner is developed for impedance waveform inversion, making it possible to obtain sharp and balanced reflective images in a fraction of the time. We demonstrate that Pseudotime GSOT-JFWI retrieves complex velocity macromodels from limited-offset datasets with minimal pre-processing, starting from non-informative initial solutions. Compared to depth-domain JFWI, the computing cost is reduced significantly, along with a simpler and less subjective design of data weighting and inversion strategy. Pseudotime GSOT-JFWI provides FWI with the necessary low-wavenumbers to converge to the broadband model, reducing the need for accurate starting models, on the road to a fully waveform-based imaging workflow.

INTRODUCTION

Full waveform inversion (FWI, Lailly, 1983; Tarantola, 1984, 1986) aims at a broadband reconstruction of the subsurface physical properties by iteratively fitting the full recorded wavefield. It has higher resolution compared to ray-based methods and, unlike the latter, does not require phase identification and selection (Virieux and Operto, 2009). In its original intent, FWI combines of tomography and migration in a purely data-driven procedure (Claerbout, 1985; Mora, 1989) whereby both kinematic and reflective components of the

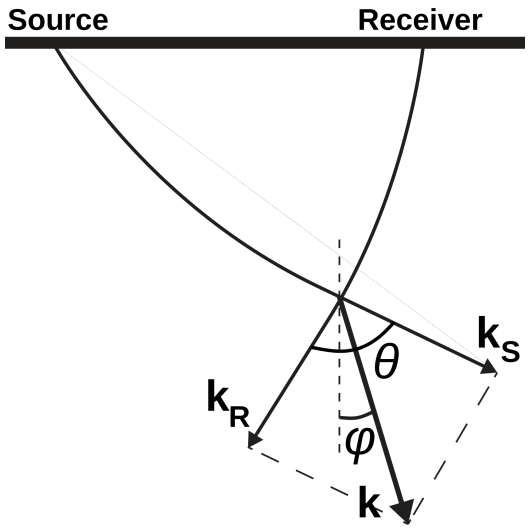


Figure 1: Wavenumbers resolved in FWI. The wavenumber vectors associated with the propagation direction of monochromatic source and receiver wavefields at the image point are denoted by $\mathbf{k}_S$ and $\mathbf{k}_R$, respectively. The local (geological) dip angle is denoted by ϕ and the aperture (or illumination) angle by θ. The vector $\mathbf{k} = \mathbf{k}_S + \mathbf{k}_R$ is the spectral component mapped into the subsurface model by FWI. The wider θ, the longer the wavelength imaged. [From Zhou et al. (2018)].

subsurface can be obtained. So far, FWI has been applied successfully on a range of scales and targets, from regional seismology (e.g., Fichtner and Villaseñor, 2015; Davy et al., 2017; Górszczyk et al., 2021) to exploration for subsurface fluids reservoirs (e.g., Hicks et al., 2016; Operto and Miniussi, 2018), down to engineering-scale site-characterization, on marine (e.g., Provenzano et al., 2018) as well as land settings (e.g., Irnaka et al., 2022).

However, the resolving power of FWI depends on specific data characteristics, namely source-receiver aperture and source-bandwidth (Jannane et al., 1989) (Figure 1). Low wavenumbers (long spatial wavelengths) are key for FWI to converge to a broadband P-wave velocity (V_p) model (Jannane et al., 1989), and its ability to retrieve them relies on the availability of low frequencies and transmission-regime, wide illumination angle wavepaths (Sirgue and Pratt, 2004). Such ideal illumination conditions, hence full wavenumber coverage, are achieved only in transmission regime settings, e.g. in medical imaging, where the target is surrounded by sources and receivers (e.g. Guasch et al., 2020; Marty et al., 2021). In seismic datasets, acquired with sources and receivers deployed at the surface, limited source-receiver offsets and limited-bandwidth, FWI retrieves low-wavenumber information only at the shallow depths sampled by diving waves (Virieux and Operto, 2009). At deep targets reached only by reflections, FWI has instead little sensitivity to model kinematics, and behaves similarly to a least-squares migration, reconstructing only the reflective component part of the subsurface mechanical properties (Woodward, 1992; Jannane et al., 1989). Under these conditions, FWI succeeds only if the starting model accurately represents the subsurface velocity trends, covering the intermediate-wavenumber resolution gap (Mora, 1989).

Acoustic sources containing ultra-low frequencies (< 2 Hz) can be used to circumvent

this limitation (e.g., Brenders et al., 2018), along with ocean-bottom-cable wide-angle and ultra-long offset streamer acquisitions to increase the penetration depth of diving waves (e.g., Vigh et al., 2013). However, these are expensive and generally non-affordable in limited-budget surveys and site-monitoring, for which repeated acquisitions are required over time (e.g., Eiken, 2019). Furthermore, though advances in the acquisition technology are of invaluable importance, they are limited to new surveys ideally designed for FWI, while data-driven methodologies remain of interest for the application of waveform-imaging to legacy data (e.g., Raknes et al., 2015).

Reflection waveform inversion (RWI) uses an initial reflective model to update the tomographic component of V_p beyond the depths sampled by diving waves (e.g. Xu et al., 2012; Brossier et al., 2015; Vigh et al., 2019; Yao et al., 2020). In a reflective subsurface, interaction between second-order scattering and first order wavefields takes place at wide angles, producing low-wavenumber contributions to the sensitivity kernel. RWI isolates those from the migration isochrones to form a tomographic V_p gradient (Zhou et al., 2015; Yao et al., 2020) ideally exploiting the reflections as virtual deep sources generating a transmission-regime wavefield (Mora, 1989). Joint FWI (JFWI, Zhou et al., 2015) enriches the RWI sensitivity kernel with the shallow contribution from diving waves, through an objective function enforcing the fit to both early-wave arrivals and reflections. This improves the robustness and stability of the reflection-based updates and, by improving the shallow kinematic characterization, significantly enhances the imaging quality at all depths (Zhou et al., 2018).

In reflection seismology, ensuring consistency between scatterers position and wavefield kinematics is key for efficient V_p-model building (e.g., Sambolian et al., 2019). Reflection-based waveform inversion makes no exception, and conflict between updating V_p and fixed reflectors position in depth leads the inversion to non-physical solutions (Yao et al., 2020). Cycles of short-offset reflectivity updates by least-squares migration and mid-to-long offset V_p inversion are thus required by the conventional V_p-building JFWI workflow (Zhou et al., 2015). This, however, comes at the price of a significant computational burden and high subjectivity of the offset windowing strategy.

Several methodologies have thus been proposed to enforce V_p-reflectivity consistency in reflection waveform tomographic methods. Data-domain approaches, such as migration-based traveltime-tomography (MBTT, Doherty and Claerbout, 1976; Snieder et al., 1989; Chavent et al., 1994; Alkhalifah et al., 2001) have been applied to RWI (Chavent, 2017; Kryvohuz et al., 2019) to honor the zero-offset traveltime seismic invariant. In order to do so, they require building and updating an intermediate reflectivity, between the data and the model space, with the same dimensions as the input data. An alternative pseudo-time (or vertical traveltime, τ) domain has been proposed by Plessix (2013) for FWI. The approach is based on chain-rule derivation of the FWI gradient from the spatial (z, x, y) to the (τ, x, y) domain, thus enforcing short-spread phase consistency with minimal cost and limited memory-requirements. More recently, Yang et al. (2021) decouple the high- and low- wavenumbers of FWI by reformulating the wave-equation to update simultaneously V_p, and a vector-reflectivity related to the spatial derivative of P-wave impedance (I_p). Alternatively, Valensi and Baina (2021) account for the velocity-reflectivity coupling by introducing an additional migration/demigration term to the RWI gradient that enforces consistency in time.

While the V_p-reflectivity conflict produces phase-ambiguity at short-offsets, inaccurate starting model kinematics may lead to cycle-skipping in JFWI when using a conventional least-squares (L^2) objective function. In this case, if the predicted traveltime is larger than half a dominant period, velocity updates are driven away from the global optimum (Virieux and Operto, 2009). Objective functions robust to cycle skipping are therefore highly recommended in RWI/JFWI, as both methods aim at the reconstruction of V_p long-wavelengths by fitting limited-bandwidth waveforms. In the literature so far, objective functions based on cross-correlation time-shifts (e.g., Brossier et al., 2015; Wang et al., 2019; Yao et al., 2020) have been shown to mitigate cycle skipping in RWI, though limited by cross-talk in multi-arrival data, where they may require appropriate windowing (Hale, 2009; Pladys et al., 2021).

This paper presents a JFWI-based methodology that overcomes the above-mentioned issues within a robust and efficent workflow, applicable with minimal pre-processing and starting from highly inaccurate initial models.

A graph-space optimal transport (GSOT, Métivier et al., 2019) JFWI objective function is designed in order to tackle the cycle skipping issue at intermediate-to-long offsets. The latter transforms the predicted and observed seismic reflection and refraction waveforms in 2D positive distributions, to which optimal transport (OT, Kantorovich, 1942) is applied. This approach has been shown to efficiently mitigate the risk of cycle skipping, reducing the sensitivity of FWI to the presence of low frequency components in the data and accurate initial model design (Métivier et al., 2019; Górszczyk et al., 2021; Pladys et al., 2021). As opposed to traveltime-based objective functions (Chi et al., 2015), GSOT can retain the amplitude information contained in the waveforms, effectively behaving as a generalization of the L^2-norm distance in which amplitude and time-shift information both contribute to defining the distance between predicted and observed data. This makes it capable of capturing time-shifts, while being more robust than cross-correlation and deconvolution-based objective functions to the presence of multiple/mixed seismic phases and missing events in the predicted data (Pladys et al., 2021). GSOT is thus an ideal candidate for complex datasets inversion starting from inaccurate initial models (e.g., Górszczyk et al., 2021). Preliminary results presented in Li et al. (2019); Provenzano et al. (2020) have shown how GSOT may work well in combination with JFWI in realistic synthetic examples.

The JFWI-problem is reformulated in the pseudotime domain (Plessix, 2013; Brossier et al., 2015) to enforce velocity-reflectivity consistency, thus attenuating phase ambiguities at short offsets, in a computationally and memory cheap fashion. As hinted in Chavent (2017), despite the assumption of near-vertical propagation, the combination of pseudotime with a robust objective function (in our case, GSOT), has the potential to alleviate cycle skipping at all offsets. The choice of pseudotime over alternative methods is driven by three main reasons. Firstly, it has minimal computing cost compared to both augmented objective-function (e.g., Valensi and Baina, 2021) and MBTT-based approaches (Chavent et al., 1994), which require additional migration/demigration steps at each iteration. In the case of MBTT, the optimization for the large-dimensionality data-space reflectivity results in an additional computational burden, in the order of 3 times FWI, and a worse-posed problem (Chavent, 2017). Futhermore, the lower memory requirement compared to

MBTT-based FWI makes pseudotime more convenient to handle large 3D dataset in time-domain. Finally, as opposed to the approach of Yang et al. (2021), it makes it possible to use physically-interpretable parameters to represent reflectivity (e.g. density (ρ) or P-wave impedance (I_p)), and allows for a straigthforward and flexible implementation in conventional full waveform modeling engines.

In order to address the non-uniqueness of the problem using data-driven information, the proposed approach takes advantage of a structure-oriented regularization strategy, whereby the V_p-updates are constrained by the geometries extracted from the reflective impedance (I_p) at a negligible extra-cost. This is performed via a Bessel smoothing (Trinh et al., 2017) of the JFWI gradient, in a rotated coordinate-system defined by the I_p structural tensor at each iteration. Compared with convolution-based smoothing (e.g. Gaussian window smoothing), the Bessel smoothing approach is faster and scales better with increasing coherent lengths, thus being more suitable for 3D applications.

The reconstruction of a reflective model, necessary to feed JFWI with deep-reaching reflections, is a significantly time-consuming task, whose cost is equivalent to a least-squares migration. This work addresses it as non-linear waveform inversion problem for P-wave impedance (I_p), in a $[V_p, I_p]$ parametrization. The latter is an ideal candidate for reflection-based inversion, since it minimizes the cross-talk between kinematic and reflective parameters (Zhou et al., 2015; Operto et al., 2013) by virtue of weakly overlapping radiation patterns. In order to reduce the cost of the I_p image building, we implement and adapt to time-domain FWI an efficient asymptotic-based preconditioner (Qin et al., 2015) with true deconvolution imaging condition in the discrete-Fourier domain. Thereby, a rapid and balanced image of the subsurface I_p can be obtained by waveform inversion of short-spread reflections (I_pWI) prior to JFWI, without requiring data de-multiple and de-ghosting. Despite having its roots in an asymptotic approximation of the wave propagation, such a preconditioner can be computed using full wavefield quantities, without the need of using an eikonal solver to compute ray quantities (Li and Chauris, 2018).

In the following sections, we illustrate how the advances introduced in this work jointly contribute to signficantly improving the robustness and reliability of JFWI (Zhou et al., 2015), and to increasing the efficiency of I_pWI, towards a mature and widely-applicable waveform-based V_p model building strategy.

After presenting our methodology in the theoretical section, we apply it to two realistic synthetic case studies, starting from very simple and inaccurate starting models. We demonstrate that: 1) GSOT-based objective function and pseudotime transformation jointly contribute to annihilating cycle skipping at both long and short offset; 2) velocity inversion in pseudotime does not require dedicated reflection offset selection unlike in depth-domain, and dramatically reduces the need to iteratively re-migrate the reflectivity, hence the computational cost of V_p inversion; 3) the asymptotic preconditioner ensures a significant speed-up in the reflectivity inversion step; 4) structure-oriented smoothing introduces useful constraints in velocity reconstruction, mitigating the problem non-uniqueness at a negligible extra-cost.

In a follow-up paper, we will present in details the application of this methodology to a 3D industrial field data from the North Sea, starting the inversion from a crude 1D model. This will illustrate how a high resolution 3D P-wave velocity model can be inferred from

3D field data using only full wavefield quantities, without the help of traveltime picking or computation through asymptotic approximations of the wave equation, relaxing the requirements of conventional imaging and moving towards a fully stand-alone FWI workflow.

THEORY

In this section, we first recall the JFWI formalism, followed by a description of the core of the velocity-model-building methodology, namely the GSOT-JFWI objective function and the pseudotime transformation. Then, the proposed structure-oriented regularization strategy for JFWI is introduced. Finally, we present the asymptotic preconditioner for efficient Impedance reconstruction (Impedance Waveform Inversion, I_pWI). The workflow is summarized in algorithm 1.

Joint FWI of diving and reflected waves

JFWI theory

JFWI (Zhou et al., 2015), jointly inverts early arrivals and reflected waves in a subsurface containing discontinuities, to obtain tomographic V_p updates both in the shallow subsurface sampled by diving waves and the deeper parts covered by reflection wavepaths. This is combined with a velocity-impedance (V_p-I_p) parameterization enhancing the scale separation between low and high wavenumbers of the model, by virtue of non-overlapping virtual source radiation patterns (Operto et al., 2013).

In this parametrization, the scattered field is generated at discontinuities entirely accounted for in the impedance model, while velocity is smooth. We define I_p^0 as the smooth impedance model, obtained as the product of starting velocity and density ($V_p\rho$), and I_p^r as the reflective impedance model ($I_p^0 + dI_p$), obtained by impedance waveform inversion (I_pWI) in the starting V_p model. Accordingly, we compute a reflection dataset d^r_{cal} in a reflective subsurface (V_p, I_p^r) and a diving dataset d^e_{cal} in a smooth model (V_p, I_p^0), therefore containing only early wave arrivals:

$$\begin{aligned} &A(V_p, I_p^r)u^r = s; \quad u^r = u_0 + \delta u \\ &d^r_{cal}(V_p, I_p^r) = Ru^r(V_p, I_p^r) \\ \\ &A(V_p, I_p^0)u_0 = s \\ &d^e_{cal}(V_p, I_p^0) = Ru_0(V_p, I_p^0) \end{aligned} \tag{1}$$

where A is the generic partial differential equation (PDE) modeling operator, s is the source wavelet, and R extracts the computed diving and reflection wavefields at the receivers location. While the latter contains both an incident (u_0) and a scattered (δu) contribution, the former contains only refractions, appearing as early arrival waveforms in the seismograms. In this full waveform formulation, unlike in Born modeling, the scattered field δu in d^r_{cal} contains scattering of every order, making it amenable to predict data containing ghosts and multiples.

The JFWI objective function for V_p is a weighted sum of two terms measuring the misfit

between respectively early arrivals and reflections of observed d_{obs} and calculated data d_{cal}. It writes as

$$\chi[V_p] = \quad G(W^e d^e_{obs}, W^e d^e_{cal}) + G(W^r d^r_{obs}, W^r d^r_{cal}) \tag{2}$$

where: G indicates a general positive function measuring the distance between two datasets; W^e and W^r are weighting matrices applied respectively to early arrivals and reflections, ideally isolating and scaling the background field in d^e and the scattered field in d^r.

The sensitivity kernel of the objective function $\chi[V_p]$ reads (Zhou et al., 2015):

$$\nabla\chi[V_p] = u_0 \star \lambda^e + u_0 \star \delta\lambda^r + \delta u \star \lambda^r_0 + \delta u \star \delta\lambda^r \tag{3}$$

where the symbol $\star$ denotes the time-convolution operator; for simplicity, it also contains the partial derivative of the forward modeling operators with respect to the unknown, in this case V_p ($\partial_{V_p} A$) (Zhou et al., 2015; Plessix, 2006). While the forward wavefields (u) have been defined in equation1, the adjoint wavefields (λ) used in equation3 are solutions of the following equations:

$$\begin{aligned} A^\dagger(V_p, I^r_p)(\lambda^r) &= \frac{\partial G(W^r d^r_{obs}, W^r d^r_{cal})}{\partial d^r_{cal}} \;; \quad \lambda^r = \lambda^r_0 + \delta\lambda^r \\ A^\dagger(V_p, I^0_p)\lambda^e &= \frac{\partial G(W^e d^e_{obs}, W^e d^e_{cal})}{\partial d^e_{cal}} \end{aligned} \tag{4}$$

where $\dagger$ indicates the adjoint operator and the superscripts e, r respectively identify the adjoint wavefields computed using the diving and reflection adjoint sources; while λ^r contains an incident (λ^r_0) scattered ($\delta\lambda^r$) field, since it is computed in the reflective model (V_p, I^r_p), λ^e only contains a background component, because it is a function of the smooth model (V_p, I^0_p). Thus, the computation of the JFWI gradient requires the solution of two adjoint equations, whose adjoint source terms (right-hand-sides) are the derivative of the objective function with respect to the state variables, d^e_{cal} and d^r_{cal} (Plessix, 2006). In the L^2 case, those amount to the reflection and diving data residuals at the receivers location, weighted and reversed in time.

Using the quantities defined in equation4, the four-term gradient in equation3 can therefore be interpreted as follows (Zhou et al., 2015):

1. the first term $u_0 \star \lambda^e$ is the diving wave first Fresnel zone obtained by cross-correlation of the incident field u_0 and the background adjoint λ^e generated by the diving wave adjoint source (EWI component), both computed in the smooth model.

2. the second term $u_0 \star \delta\lambda^r_0$ and third term $\delta u \star \lambda^r_0$, are the reflection rabbit ears resulting from the cross-correlation of the incident wavefields (u_0, λ^r_0) and second order scattered wavefields ($\delta\lambda^r$,δu) at the model discontinuities (RWI component).

3. the last term represents higher-order scattering effects, among them undesired migration isochrones, which in the V_p, I_p parametrisation are attenuated.

JFWI gradient implementation

The FWI gradient in a reflective model, however, along with the desired tomographic contributions in equation3, contains migration isochrones that contribute to building the high-wavenumbers of the V_p-model. Those high-frequency contributions result from the direct interaction between the incident source wavefield (u_0) and the downgoing reflection adjoint wavefield (λ_0^r), and are one order of magnitude larger than the reflection rabbit ears (Virieux and Operto, 2009). Following Zhou et al. (2015), in this work the JFWI gradient in equation3 is obtained via a full-waveform based approach which eliminates the first-order migration isochrones in two steps:

1. The first step computes a gradient in the reflective subsurface using the reflection adjoint source, and therefore contains both the rabbit ears and the migration isochrone. $\eta_1 = u_0 \star \delta\lambda^r + \delta u \star \lambda_0^r + u_0 \star \lambda_0^r$
2. The second step computes a gradient in a non-reflective subsurface (smooth I_p^0) using the diving wave adjoint source *minus* the reflection adjoint source, resulting in a banana kernel and an opposite polarity migration isochrone. $\eta_2 = u_0 \star \lambda^e - u_0 \star \lambda_0^r$
3. The sum $\eta_1 + \eta_2$ eliminates the migration isochrone $u_0 \star \lambda_0^r$ and combines the diving and reflection tomographic kernels, obtaining the JFWI gradient in equation3.

This procedure exactly obtains the tomographic JFWI gradient in equation 3 at the cost of two FWI gradient computations, similarly to Born modeling approaches. Higher-order scattering effects, $\delta u \star \delta\lambda^r$ (last term of equation3) may contain higher-order migration isochrones, whose contribution is made smaller by employing the chosen V_p, I_p parametrisation (Operto et al., 2013; Zhou et al., 2015).

Graph-space optimal transport

Cycle skipping is likely to occur at intermediate-to-long offsets in least-squares V_p waveform inversion when the traveltime prediction error is larger than half a dominant period. Among different misfit function alternatives to L^2, the graph-space optimal transport (GSOT) distance has shown interesting properties to mitigate cycle skipping for conventional FWI applications (Górszczyk et al., 2021; Pladys et al., 2022). The conceptual core of GSOT is the re-mapping of each seismic trace as a two-dimensional discrete distribution of K unit-weight points (point cloud) in a time-amplitude space (graph space). Using this representation, an optimal transport distance is applied to measure the misfit between predicted and observed traces. Such distances have the specificity to be convex with respect to dilation and translation when applied to positive measures. Applying them to the graph of predicted and observed traces therefore guarantees convexity with respect to time-shifts between traces, reducing the risk of cycle skipping (Pladys et al., 2021).

Each individual predicted $d_{cal}(t)$ and observed trace $d_{obs}(t)$ is associated to its discrete graph $(t_i, d_{cal}(t_i))$, $(t_i, d_{obs}(t_i))$, $i = 1, \dots K$ assuming the same time discretization is employed. As described in Métivier et al. (2019) and Métivier and Brossier (2021), the

2−Wasserstein (optimal transport) distance between $d_{cal}(t)$ and $d_{obs}(t)$ thus amounts to the solution of the optimal assignment problem

$$G_{gsot}(d_{cal}, d_{obs}) = \min_{\sigma \in S(K)} \sum_{i=1}^{K} c_{i,\sigma(i)}(d_{cal}, d_{obs}) \tag{5}$$

where $S(K)$ is the ensemble of permutations of $\{1, \ldots, K\}$, and

$$c_{ij}(d_{cal}, d_{obs}) = \frac{A^2}{\Delta t^2}|t_i - t_j|^2 + |d_{cal}(t_i) - d_{obs}(t_j)|^2. \tag{6}$$

In 6, Δt represents the maximum expected time shift and A is the maximum amplitude (peak-to-peak) discrepancy between the observed and predicted trace. Among all possible permutations, the optimal permutation σ^* solution of the problem 5 minimizes the work needed to transport the initial distribution to the target one, in the analogy with optimal transport for engineering problems (Kantorovich, 1942). Considering predicted and observed traces containing shifted phases, the transport work will depend monotonically on the time shifts between phases thanks to the terms $|t_i - t_j|^2$ in the transport cost c_{ij}, hence the convexity of the distance with respect to time shifts.

One important property of the GSOT strategy is the relative simplicity of the gradient building step once the misfit function is defined. For a misfit function

$$G_{gsot}(d_{cal}, d_{obs}) = \sum_{i=1}^{K} \frac{A^2}{\Delta t^2}|t_i - t_{\sigma*}|^2 + |d_{cal}(t_i) - d_{obs}(t_{\sigma*})|^2. \tag{7}$$

the gradient can be obtained, thanks to the adjoint state strategy, as the time-convolution between an incident field and an adjoint field (as in equation3), whose source term is defined by

$$\frac{\partial G_{gsot}(d_{cal}, d_{obs})}{\partial d_{cal}} = d_{cal}(t_i) - d_{obs}(t_{\sigma^*(i)}). \tag{8}$$

The latter equation illustrates how the GSOT misfit function can be seen as a generalization of the least-squares misfit function: as soon as the optimal assignment σ^* is the identity, the GSOT misfit function becomes equivalent to the least-squares distance and the adjoint source becomes the conventional residual between observe and predicted data.

The GSOT-JFWI objective function

This is adapted to JFWI by replacing each term of the objective function (equation2) by the respecive 2−Wasserstein (optimal transport) distance:

$$\chi_{gsot}[V_p] = \quad G_{gsot}(A_e, \Delta_t^e, W^e, d_{obs}^e, d_{cal}^e) + G_{gsot}(A_r, \Delta_t^r, W^r, d_{obs}^r, d_{cal}^r) \tag{9}$$

The two GSOT misfit terms, for reflections and diving waves, are scaled independently through the respective maximum amplitude discrepancies, A_r and A_e. This is necessary in order not to bias the estimation of either 2-Wasserstein distance when the amplitude

range of diving waves significantly differs from the one of the reflections. Potentially, two different maximum expected time shifts $(\Delta t^r, \Delta t^e)$ can also be used. For each of the two propagation regimes, a different optimal assigment map σ^* is therefore computed. Thus, the JFWI adjoint wavefields in equation4 with the GSOT adjoint sources read:

$$\begin{aligned} A^\dagger(V_p, I_p^r)\lambda^r &= W^{rT}W^r(d^r_{cal}(t_i) - d^r_{obs}(t_{\sigma^{r,*}(i)})) \\ A^\dagger(V_p, I_p^0)\lambda^e &= W^{eT}W^e(d^e_{cal}(t_i) - d^e_{obs}(t_{\sigma^{e,*}(i)})) \end{aligned} \tag{10}$$

where a specific optimal assignment plan σ^* is used for the reflection and refraction terms of the objective function. The expressions for the JFWI gradient are otherwise equivalent to the general case presented in the JFWI theoretical section.

The intrinsic flexibility of GSOT can be exploited during inversion, starting with large Δt to ensure robustness against cycle skipping, and reducing it as model kinematics improves and it becomes desirable to approximate the resolution of L^2-norm.

Pseudotime domain

While GSOT minimizes the risk of cycle-skipping at intermediate-to-long offsets, phase ambiguity at near-zero offsets may occur in reflection-based velocity-model-building as a consequence of kinematics-reflectivity conflict. The impedance model I_p^r in equation1 depends on V_p, namely the position and geometry of the reconstructed reflective discontinuities are a function of the current wavefield kinematics. However, such a dependency is not accounted for in depth-domain reflection-based FWI, nor is in JFWI, which uses I_p^r as a passive parameter. Hereafter, in the text we will refer to the reflective impedance model simply as I_p.

In a depth-domain implementation, since V_p updates are driven by moveout residuals, but the depth of the reflectors is fixed, convergence can be attained only by weighting out the short-spread reflection, which would be out of phase and lead the inversion to a local minimum (Plessix, 2013; Yao et al., 2020). Repeated I_pWI reflectivity reconstructions are therefore interleaved within successive steps of depth-domain JFWI. This significantly increases the computational burden and is dependent on the offset weighting strategy.

To overcome this difficulty and address the velocity-depth ambiguity in reflection-based velocity model building, we propose an alternative approach that enforces $V_p - I_p$ consistency by reformulating JFWI in the pseudotime domain (Plessix, 2013; Brossier et al., 2015). The relationships between pseudotime (τ) and depth (z) at each horizontal position are functions of V_p:

$$\tau(z) = \int_0^z dz'/V_p^z(z'); \qquad z(\tau) = \int_0^\tau V_p^t(\tau')d\tau' \tag{11}$$

where the superscripts ^{t} and ^{z} denote respectively the pseudotime and depth quantities. The depth domain V_p^z, I_p^z model is mapped to the pseudotime domain τ, discretized using

$\Delta\tau = \Delta z/Vp_{max}$ and reaching $\tau_{max} = z_{max}/Vp_{min}$.

The gradient (equation3) is computed in the depth domain, and reformulated in pseudotime using the derivation chain rule (Plessix, 2013, their equation B6):

$$\nabla_{\tau(i)}\chi = \nabla_{z(i)}\chi - \int_{z_i}^{z_{max}} \frac{dV_p^z}{dz}\frac{1}{V_p^z(z)}\nabla_z\chi dz \tag{12}$$

The latter step is analogous to a re-parametrization of the inverse problem, from a depth-domain variable to its pseudotime equivalent. After each V_p^t update in the τ domain, the V_p^t, I_p^t model is re-mapped to the modeling (z) domain consistently with the new V_p^t, using equation11 and interpolating the value at $z(\tau)$ into the modeling grid. Reflectors are thus repositioned, and I_p remains fixed in pseudotime while passively being updated in depth at no extra modellig cost. Thereby, the near-zero offset waveforms are kept in phase, honoring the zero-offset seismic invariant and reducing the need of cycling through successive V_p-JFWI and I_pWI inversion. In defining the pseudotime grid for marine datasets in which seabed position and sea-column velocities are known, it is beneficial to tune $\Delta\tau$ so that the seabed reflection one-way time corresponds to a grid point in the $\tau(z)$ axis.

Note that the depth to pseudotime transformation implies 1D wave propagation, and is therefore suitable for models showing smooth lateral variations, which is usually the case for JFWI velocity-macromodels, and moderate structural dips. On the other hand, merely repositioning the reflectors in depth does not account for changes in amplitude due to changes in model kinematics and thus illumination. Therefore, the need for re-runs of I_pWI after significant velocity updates is not removed altogether, especially when starting from particularly inaccurate initial velocity models.

Structure-oriented regularization

Besides its non-linearity, waveform-based V_p inversion is inherently an ill-conditioned inverse problem, i.e. potentially an infinite number of models in the chosen parameterization may fit the data within an acceptable data-fit threshold. It is therefore a common practice to introduce prior information to reduce the size of the solution space, most commonly by constraining the regularity of the solution *i.e.* the smoothness of the V_p variation in space. This is implemented through the application of specific smoothing filters to the model update at each inversion iteration. Among this category, the most widespread is the Gaussian filter, which can be made non-stationary to adapt to it the expected local resolution, as done for instance in Operto et al. (2006). The local resolution can be indeed estimated from diffraction tomography analysis (Devaney, 1984; Wu and Toksöz, 1987; Sirgue, 2003). A potential difficulty with this technique is the absence of information regarding the underlying geological structure of the model, which can lead to remove from the V_p model spatial variations associated with features one would like to preserve.

Overcoming this difficulty requires not only to adapt the smoothing strategy but also to obtain relevant information on the geological structure of the investigated zone. In the proposed methodology, we implement a technique to extract structural information from the reconstructed I_p model, already available at all times during JFWI-based V_p inversion. In order to extract the local dip of the reflectors in the I_p model, we first compute its perturbation image $dI_p = I_p^r - I_p^0$. The local derivatives of dI_p are computed and a covariance

matrix of the latter is built at each point within the model, to obtain a measure of the local coherence of the structures, or structural tensor. In two-dimensions, it reads:

$$C_{x,z} = \begin{pmatrix} g_{xx} & g_{xz} \\ g_{zx} & g_{zz} \end{pmatrix} \tag{13}$$

where

$$g_{ij}|_{x,z} = \frac{\partial dI_p}{\partial i} \cdot \frac{\partial dI_p}{\partial j}\bigg|_{x,z} \tag{14}$$

Thus, we have obtained a local structure matrix from which we wish to extract the direction of strongest coherent structural variation. In order to do so, the principal components of C are extracted by eigenvalue analysis:

$$\begin{aligned} C_{x,y} &= VSV^T \\ V_{x,z} &= \begin{pmatrix} v_{1x} & v_{2x} \\ v_{1z} & v_{2z} \end{pmatrix} \\ S_{x,z} &= \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \end{aligned} \tag{15}$$

where v_1 in the eigenvector matrix V is the direction corresponding to the largest eigenvalue (s_1) in S, thus normal to the local structural dip, and v_2 is instead parallel to the layer orientation, corresponding to the lowest eigenvalue (s_2). For each point of $dIp(x, z)$, it is therefore possible to extract the dip angle matrix with respect to the horizontal (x) direction as:

$$\phi_{x,z} = \tan^{-1} \frac{v_{2z}}{v_{2x}} \tag{16}$$

Smoothing is applied to structure tensor before eigenvalue decomposition, and the resulting dip angle matrix is again smoothed with user-defined lengths, in the order of the propagated wavelength. Note that in three dimensions, this formulation would require both dip and azimuth matrices to be extracted from the 3x3 structure tensor.

Once this information is extracted from the impedance, it can be injected in a specific directional Bessel smoothing filter, introduced in (Trinh et al., 2017). For each point in a two-dimensional subsurface, a directional Bessel filter is obtained as the solution of an elliptical partial differential equation (Trinh et al., 2017) in the ϕ-rotated coordinate system $[v_1, v_2]$. In two-dimensions, we have:

$$B_f(v_1, v_2) - (L_2^2 \frac{\partial^2}{\partial v_2^2} + L_1^2 \frac{\partial^2}{\partial v_1^2}) B_f(v_1, v_2) = \delta(v_1, v_2) \tag{17}$$

where L_1 and L_2 are the user-defined local filter lengths in the v_1, v_2 directions, respectively perpendicular and parallel to the reflector dip, and $\delta(v_1, v_2)$ is a delta Dirac function with the same dimensions as the gradient. The lengths vary as a function of the propagated wavelength, thus adapting to the local resolution determined by the modelled frequency and the V_p values at each point; though, ideally, equation17 requires them to be homogeneous,

the smoothly varying V_p resulting from JFWI only weakly undermine the validity of the approach.

By virtue of the linearity of the latter equation, the solution of $F[B_f * \gamma] = \delta * \gamma$, where γ is the un-smoothed gradient, is the smoothed gradient $\gamma_s = B_f * \gamma$. This can be interpreted as the steady-state diffusion equation using the un-smoothed gradient as source term. Therefore, γ_s can be obtained as the solution of a linear system, not only incorporating structural information, but also with higher efficiency than convolution-based smoothings. The smoothed gradient can be obtained by discretising equation17 and solving the linear system in a matrix-free fashion, with a conjugate-gradient solver, thanks to its simmetry (Trinh et al., 2017). The regularization is applied to the depth-domain gradient, before re-parametrizatization to the pseudotime domain (equation12).

Compared to convolution-based approaches, whose computational complexity scales almost cubically with the filter lengths, the PDE-based smoothing scales sublinearly (Trinh et al., 2017), which makes it amenable to 3D applications. However, the inaccurate V_p model at early iterations produces deformed reflectors' geometries, thus inaccurate structure tensors. In depth-domain inversion, it is therefore necessary to apply structure-oriented smoothing only after some loops of JFWI and I_pWI, in order to avoid error propagation from the I_pWI image to the JFWI-V_p update (Yao et al., 2019). In pseudotime, instead, the impact of inaccurate initial reflective images in depth is limited to the very first iterations, since the reflectors in I_p are simultaneously repositioned as V_p is updated.

Asymptotic preconditioning for I_p reconstruction

While the pseudotime approach alleviates the need of iteratively recomputing I_p within the velocity model building workflow, the reconstruction of the initial I_p (I_pWI) remains a significantly time-consuming task, whose cost is equivalent to a least-squares migration. To speed up this process, we propose to use a specific preconditioned gradient based on asymptotic direct inverse operators.

Direct inverse operators retrieve the true-amplitude subsurface image using inverse scattering operators based on the asymptotic high-frequency approximation of the linearized wavefield (Beylkin, 1985). In this work, instead, the asymptotic formulation is used as a preconditioner to scale and weight the gradient of the I_pWI objective function, defined as the least-squares distance between short-spread reflected wavefields:

$$\chi[I_p] = \frac{1}{2} \left|\left| \left(W^r_{Ip} \left(d^r_{cal} \left[V_p, I_p \right] - d^r_{obs} \right) \right) \right|\right|^2_2 \tag{18}$$

where W^r_{Ip} is a weighting operator applied to select the near-offset reflections , while the projection operator is included in d^r_{cal}. The use of L^2 misfit function is here justified by the limited risk of cycle skipping in short-spread waveform inversion; in this L^2 case, the adjoint source is simply the data residuals at the receivers, and the adjoint wavefield (λ) is the solution of the following equation:

$$A^\dagger(V_p, I_p)\lambda = {W^r_{Ip}}^T W^r_{Ip}(d^r_{cal} - d^r_{obs}) \tag{19}$$

where T indicates the transpose operator. Since both forward and adjoint operators are function also of the reflective part of the model, the predicted data will contain scattering of any order, unlike Born modeling approaches. Accordingly, multiple and ghost reflections are not eliminated from the field data in pre-processing. Under these conditions, the choice of a non-linear minimization process over direct methods is a necessity (Virieux et al., 2016; Li and Chauris, 2018).

In Qin et al. (2015) and Li and Chauris (2018), the direct inverse of δu to obtain the model perturbation δm is defined in frequency-domain using only pressure wavefields, thus not requiring the computation of rays and associated incidence and take-off angles:

$$\begin{aligned}\delta m(x) &= \sum_{shot}\sum_{k} 4\delta\omega_k \frac{\nabla_x u^*(x,\omega_k)\nabla_x \Lambda(x,\omega_) + (i\omega_k/V_p)^2 u^*(x,\omega_k)\Lambda(x,\omega_k)}{-i\omega_k^3 u^*(x,\omega_k)u(x,\omega_k)} \\ \Lambda(x,\omega_k) &= \sum_{rec} \frac{\partial \zeta_x(rec,\omega_k)}{\partial_z}\lambda(rec,\omega_k)\end{aligned} \tag{20}$$

where a summation is performed over shots and positive angular frequencies ω_k and $*$ indicates complex conjugation. The source wavefield and the modified adjoint wavefield are indicated respectively by u and Λ. The latter is the convolution of the vertical derivative of the adjoint Green function (ζ) at the receivers positions and the perturbation wavefield δu. This enhances narrow take-off angle at the receiver positions, while the horizontal derivatives at the scattering points (x) enhance the narrow illumination angles (Qin et al., 2015), mapping into high-wavenumbers of the retrieved perturbation model. Finally, a deconvolution imaging condition is applied (denominator), which compensates for the wavefield illumination and attenuates imprint of the limited-bandwidth of the source wavelet (Li and Chauris, 2018).

We reformulate equation20 to adapt and optimize it for our full-waveform engine based on a first-order formulation of the wave equation. Since in the latter we have access to particle velocity and pressure wavefields, it is convenient expressing the horizontal derivative of the pressure wavefields in equation 20 as a function of local particle velocity, using

$$\rho\partial_t v_{\mathbf{x}} = \partial_{\mathbf{x}} P. \tag{21}$$

This makes it possible to use time-differentiation in the frequency domain (multiplication by $i\omega$), instead of discretizing the spatial derivatives at the scattering point within the modeling grid. Finally, using a small take-off angle approximation to simplify the expression for Λ, and reformulating for I_p (Li et al., 2019), we can write:

$$\Gamma_{\chi[I_p]} = \sum_{shot}\sum_{k} \mathbb{R}\frac{\frac{i\omega_k\rho^2}{V_p}{u_v}^*(\omega_k,x)\lambda_v(\omega_k,x) - \frac{i\omega_k}{V_p^3}u_p^*(\omega_k,x)\lambda_p(\omega_k,x)}{u_p^*(\omega_k,x)u_p(\omega_k,x) + \epsilon(shot)} \tag{22}$$

where the subscripts v and p indicate respectively the particle velocity and pressure wavefields or the primary u and adjoint λ fields. To stabilize the spectral division, a damping (water-level) parameter ϵ is defined for each shot as $1e^{-4}$ of the maximum amplitude of the absolute-valued incident wavefield (Schleicher et al., 2008).

The latter expression is used as preconditioned gradient in our non-linear optimization problem for the objective function defined in equation 18 (Virieux et al., 2016; Farshad and Chauris, 2021). Both the conventional adjoint-based gradient and the preconditioned gradient Γ are therefore required at each iteration for the optimization (Métivier and Brossier, 2016,, their Eqs. 6 and 7). As observed by Métivier et al. (2015), who alternatively approached the problem by modifying the objective function through migration/demigration of residuals, this amounts to having an inverse operator with a near diagonal Hessian.

In order to adapt it to our time-domain FWI approach, we implement the proposed preconditioned gradient using discrete-fourier-transform (DFT). Source and receiver DFT wavefields are built on-the-fly during the incident field computation and stored at a set of discrete frequencies to compute Γ, as done by Yong et al. (2022) for the computation of Hessian-vector products in the Truncated-Newton algorithm. Unlike in previous time-domain implementations (Li et al., 2019), this effectively implements the deconvolution imaging condition in equation 20, without relying on the assumption of frequency-independence of the incident wavefield (Schleicher et al., 2008). A significant frequency decimation is applied on the positive frequencies, to make the approach memory-affordable also in 3D. As shown in Yong et al. (2022), a number of frequencies lower than Nyquist might be used, as long as it is sufficient to avoid wrap-around effects.

Summary velocity-model-building workflow

In summary (Alg.1), once the reflection and diving propagation regimes have been identified and the weighting matrices built, the workflow starts with the reconstruction of I_p in the initial model, employing the asymptotic preconditioning strategy. At this stage, either a steepest-descent or conjugate gradient algorithm can be used to update I_p using jointly preconditioned and conventional gradient.

Starting from the reconstructed reflective model, GSOT-JFWI is performed in the pseudotime domain, updating V_p and, simultaneously and at no-extra cost, I_p. JFWI requires the solution of six wave equations, if the incident source wavefield are not stored, three for each step (η_1 and η_2) of the JFWI gradient computation, and thus twice as many as for conventional FWI. The structure-oriented smoothing can be adopted to constrain the V_p solution. The GSOT-JFWI stage requires the definition of the maximum expected time-shifts, while the structure oriented smoothing requires the user to choose the coherence lengths as a function of the propagated wavelengths. Steepest-descent, conjugate gradient, or quasi-newton methods can be used for the optimization (Métivier and Brossier, 2016).

Algorithm 1 Velocity model building workflow. Note that, during JFWI, I_p^r is updated consistently with V_p in depth, while the inversion is performed for V_p^t at fixed I_p^t in pseudotime. In the following synthetic case studies, the optimization follows either a non-linear conjugate gradient or steepest-descent algorithm (Métivier and Brossier, 2016)

- $k = 0$, *BEGIN* $\mathbf{I_p WI}$
- Model data and compute $\chi[I_p](0)$ ▷ Eq. 18
while $[\chi[I_p(k)] \leq threshold]$ .OR. $[k = Iter_{max}]$ **do**
    - Compute $\nabla\chi[I_p(k)]$ and $\Gamma[I_p(k)]$ ▷ Eq. 22
    - Optimization for $dI_p(k) \rightarrow I_p(k+1) = I_p(k) + dI_p(k)$
    - Model data and compute $\chi[I_p(k+1)]$ ▷ Eq. 18
    - $k = k + 1$
end while ▷ $I_p^0 \rightarrow I_p^r$

==========

- $k = 0$, *BEGIN* **JFWI**
- Compute structure tensor of dI_p^r ▷ Eq. 15
- Convert to pseudotime $V_p \rightarrow V_p^t$ and $I_p^r \rightarrow I_p^t$ ▷ Eq. 11
- Model data and compute $\chi[V_p(0), I_p^r]$ ▷ Eq. 2
while $[\chi[V_p(k), I_p^r(k)] \leq threshold]$ .OR. $[k = Iter_{max}]$ **do**
    - Compute $\nabla\chi[V_p(k)]$ ▷ Eq. 3
    - Convert to pseudotime $\nabla\chi[V_p(k)] \rightarrow \nabla\chi[V_p^t(k)]$ ▷ Eq. 12
    - Optimization for $dV_p^t(k) \rightarrow V_p^t(k+1) = V_p^t(k) + dV_p^t(k)$
    - Convert to depth $V_p^t(k+1) \rightarrow V_p(k+1)$ and $I_p^t \rightarrow I_p^r(k+1)$ ▷ Eq. 11
    - Model data and compute $\chi[V_p(k+1), I_p^r(k+1)]$ ▷ Eq. 2
    - $k = k + 1$
end while ▷ $V_p^0 \rightarrow V_p$

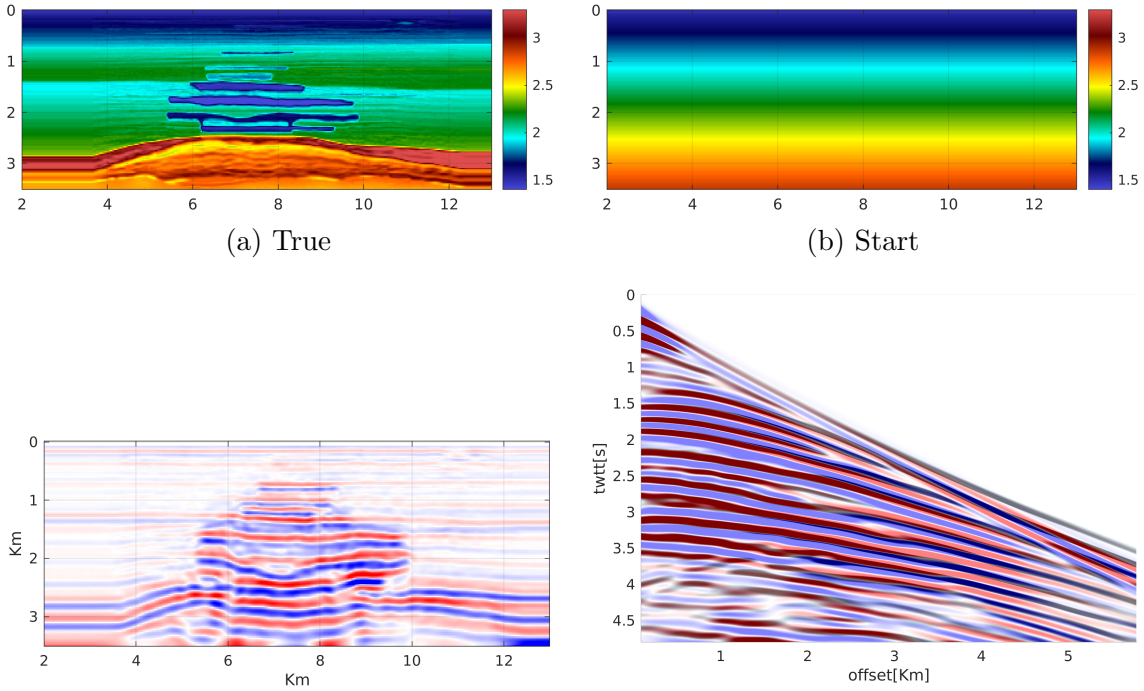
SYNTHETIC CASE STUDIES

Valhall 2D synthetic reflection dataset

Data and initial model

The first synthetic application presented is on the Valhall 2D acoustic model (V_p and ρ), comprising a multi-layered low-velocity zone and a deeper high velocity anticline, inspired by the Valhall oil field in the North Sea (e.g., Gholami et al., 2013; Zhou et al., 2015). However, instead of an ocean-botton cable survey, we simulate a surface streamer acquisition, making velocity reconstruction at deep targets more challenging. 128 shots are fired at 110 m spacing and the wavefield is recorded by a 229 channels surface streamer with group interval equal to 25 m, and absolute offset ranging between 80 and 5700 m. Since maximum target depth is about 3 Km, such an acquisition is expected to be unsuitable for conventional FWI to yield a broadband velocity reconstruction (Woodward, 1992).

The source signature is a zero-phase Ricker with central frequency equal to 6.25 Hz. Modeling is performed in the time domain with a 4th order finite-difference scheme ($dx = 25\ m$, $dt = 4\ ms$), perfectly-matching-layer (PML) absorbing boundaries and free-surface (Yang et al., 2018); therefore, the observed data contain surface-related multiples and



(a) True (b) Start

c) Initial reflectivity $I_p - I_{p(smooth)}$ d) Predicted data

Figure 2: Models (a-b), reconstructed reflectivity (c) and data (d). In the example common-shot gather (d), real data in black overlaid to synthetic in red and blue; data perfectly in phase would appear as blue and black, as red is covered by black

ghosts, which are kept in pre-processing. The starting model is 1D (Figure 2), and it is made up by a shallow water layer (70 m depth, $V_p = 1500\ m/s$, $\rho = 1000\ kg/m^3$) and a subsurface in which V_p increases linearly with depth. Note in Figure 2 how the data predicted by this model, with the reconstructed reflectivity, is prone to cycle skipping both in the diving waves and long-offset (> 3 km) reflections. While true ρ is derived from true V_p via a Gardner's relationship, starting ρ below seafloor is constant and equal to 2000 kg/m^3, therefore the initial I_p is simply a scaled version of V_p.

I_p-reconstruction

The V_p-model building workflow begins with the inversion of the short-spread (offset < 500 m) reflections (I_pWI) in the starting V_p. Figure 3 shows the convergence history of non-linear conjugate gradient (Nocedal and Wright, 2006) I_pWI with the proposed asymptotic preconditioner against linear depth preconditioning and a pseudo-Hessian preconditioner compensating for wavefield illumination (Kamath et al., 2021). The convergence speed-up is remarkable, despite the presence of shallow water multiples and ghosts. The number of frequencies stored in the DFT scheme is 38, with a sampling rate $\Delta f = 0.35$ suitable to avoiding wrap-around effects, despite a compression of a factor 1.8 compared to a Nyquist sampling for a maximum frequency of 15 Hz. The reconstructed I_p (Figure 2c) is then used as the reflective contributor of the starting JFWI model (data predicted for this initial

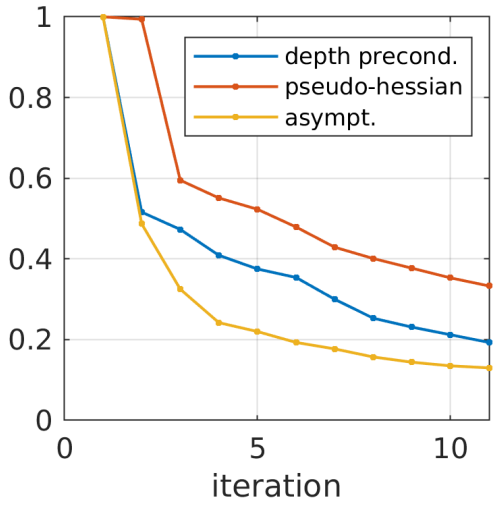


Figure 3: Convergence of I_pWI in initial V_p model using different preconditioners.

reflective model are shown in Figure 2d).

V_p-macromodel building

This section compares the performance of: 1) GSOT-JFWI; 2) Pseudotime GSOT-JFWI; 3) Pseudotime L^2-JFWI; 4) Pseudotime GSOT-RWI, where the diving component of JFWI is set to zero and only reflections are used. Finally, the impact of structure-oriented smoothing is also assessed.

Data windows are designed to separate early arrivals and reflected waves, in order to build the two components of the JFWI objective function in equation2; the limit between the two propagation regimes is simply given by the direct-wave arrival time, plus the wavelet duration (Figure 2d). In JFWIs, a scaling factor equal to 0.66 is applied to the diving wave component of the adjoint source, in order to balance the diving and reflected contributions to the two-term objective function (equation 3). 25 iterations of steepest descent with linear depth preconditioning are performed. The choice of a steepest-descent optimization over quasi-newton methods (Nocedal and Wright, 2006) is justified by the need of a robust V_p update in the presence of multiples and multi-scattering, in order to avoid overfitting, at the expense of convergence speed (e.g., Zhou et al., 2018). In the GSOT case, Δt (equation 6) is set to 0.4 s to privilege convexity with respect to traveltime differences; this is larger than the maximum expected time-shift between predicted and observed data, but is helpful to attribute higher weights to small amplitude events in the graph space (Métivier and Brossier, 2021).

In Figure 4 the advantages of pseudotime over a conventional depth formulation are apparent, GSOT JFWI failing to retrieve the correct velocity trends. Pseudotime L^2-JFWI identifies the low velocity zone, but is cycle skipped in the shallow part, where velocity is erroneously decreased instead of increased, especially in portions of the model sampled preferentially by refracted arrivals. Pseudotime GSOT-JFWI, on the other hand, combines

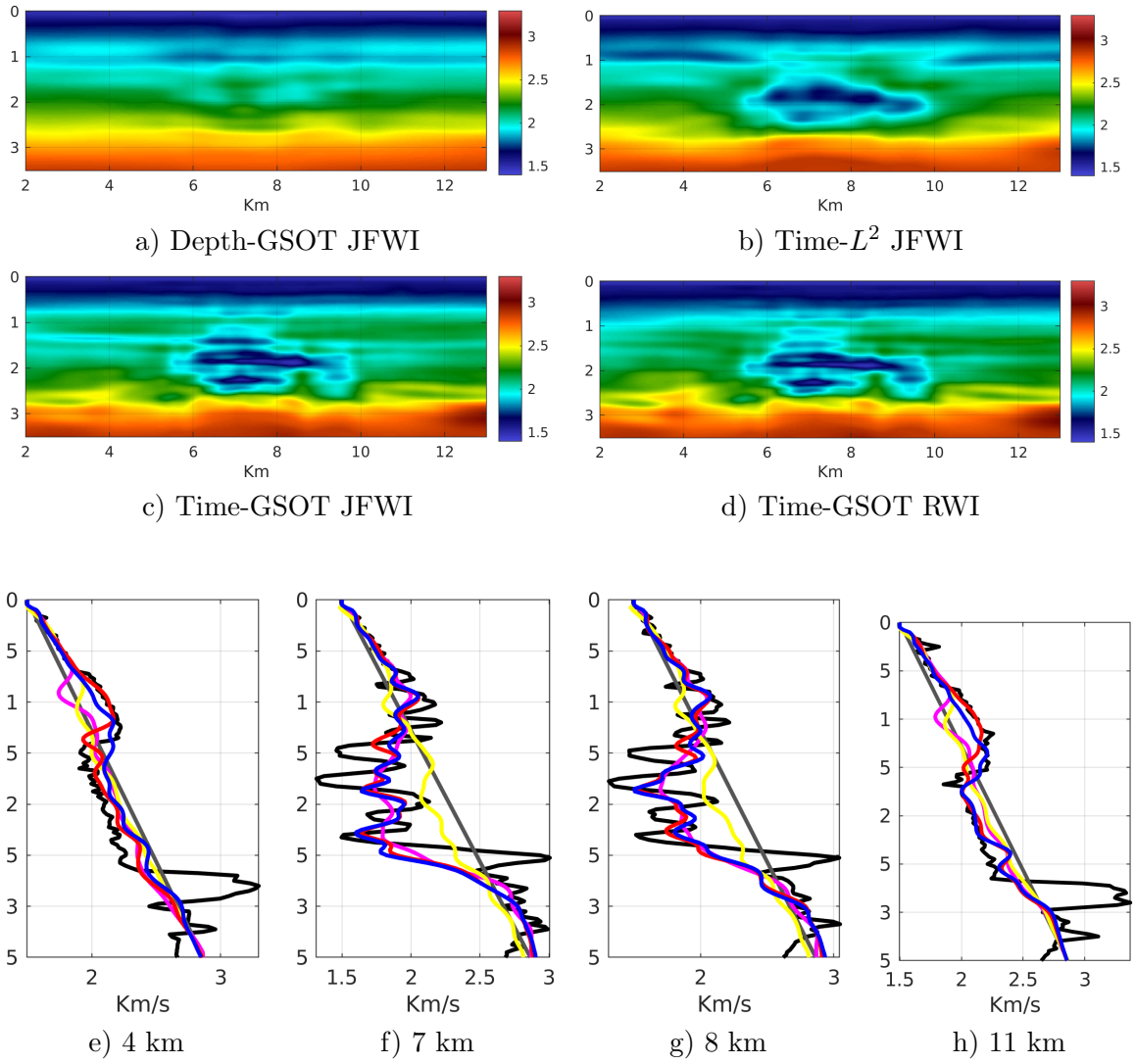
a convex misfit function with an enforced kinematic-reflectivity consistency, and retrieves an accurate V_p model starting from the 1D initial guess. The benefit of jointly inverting reflections and diving waves in JFWI is a better reconstruction of the shallow part of the model with respect to RWI.

Structure-oriented smoothing is used in conjunction with pseudotime GSOT-JFWI. In Figure 5 the local dip angle matrix extracted from the I_p structure tensor is shown, after appropriate attenuation of the high-wavenumber oscillations. Smoothing the JFWI gradient along the corresponding v_1, v_2 directions injects structural information in the V_p update, cleaning it up from anomalies inconsistent with the true subsurface. This is apparent in the smoother transitions between high and low velocity anomalies (Figure 5). Though more sensible, this model does not correspond to a lower final misfit function, in a clear example of ill-posedness in waveform-inversion: among an ensemble of models with acceptable data-fit, here we constrain the inversion to move towards a geologically meaningful one, where velocity varies slower within strata than across them.

To assess the benefit of structure-oriented smoothing at this stage, we plot the final I_p reflectivity, as it results from the passive pseudotime updates (Figure 6). JFWI constrained by structures almost perfectly reconstructs the reflective interfaces geometries, outperforming unconstrained JFWI, which yields a I_p reflectivity with unphysical oscillation symptomatic of data overfitting in V_p model building. It is also worth pointing out that JFWI produces less-deformed structures than RWI: the diving-wave constraint in the shallow part allows for better imaging at all depths, as shown in Zhou et al. (2015). Note that, unlike in the latter paper and Provenzano et al. (2020), the pseudotime formulation allows to get rid of the inner loop of I_pWI, since I_p is passively repositioned consistently with each V_p update. This produces a significant saving in computing time, while reducing the need of subjective and time consuming offset weightings (e.g., Yao et al., 2020; Valensi and Baina, 2021).

Data comparison: why does depth-domain fail?

In order to understand why depth-domain JFWI fails in this case, it is instructive to look at the predicted data in the reflective initial and final $V_p - I_p$ models. A common-shot gather is extracted around the central part of the acquisition, where the performances of the pseudotime methodologies are comparable, while depth-domain JFWI fails to update the velocity model. In Figure 7, we compare initial (a), GSOT-JFWI (b), Pseudotime GSOT-RWI (c), Pseudotime GSOT-JFWI (d), Pseudotime L^2-JFWI (e), Pseudotime with structure-oriented smoothing (f). In a) the near-offsets are in phase, while the mid-to-long ones are not, as a consequence of the wrong model kinematics. In b) V_p inversion is driven by long-offsets mismatches and reduces them, but, as soon as the V_p update is no longer compatible with the fixed-depth reflectivity, the misfit at short offset increases (more red wiggles than initial model) and the inversion is stuck in a local minimum. In c-to-f, the pseudotime formulation mitigates this issue and data-fit is improved in a broad offset range (blue-dominated plots). Data prediction is almost invariant with respect to the structure-constraints (d and f): as pointed out in the previous subsection, this is an example of non-uniqueness of the JFWI problem; RWI (c) has worse data prediction in the refracted waves, since they do not contribute to the objective function. L^2-JFWI (e), though not cycle-skipped in this instance, has higher misfit at offsets larger than 5 km in refractions and at offset larger than 4 km in reflections, as a results of an objective function non convex

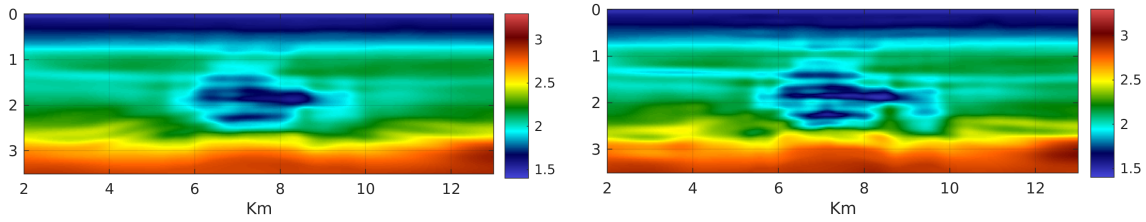


a) Depth-GSOT JFWI b) Time-L^2 JFWI

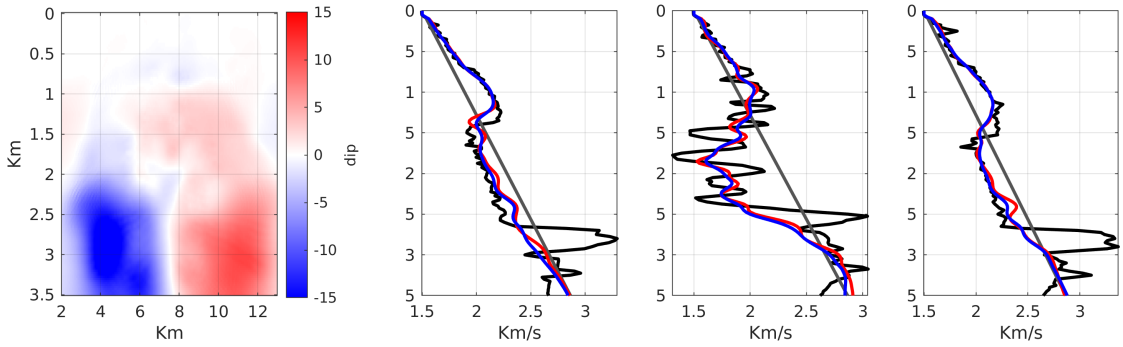
c) Time-GSOT JFWI d) Time-GSOT RWI

e) 4 km f) 7 km g) 8 km h) 11 km

Figure 4: Comparison of V_p-model building methodologies. In the vertical profiles: pseudotime L^2 in magenta, pseudotime GSOT in red, pseudotime GSOT RWI in blue, depth-GSOT in yellow, starting model in grey and true one in black.

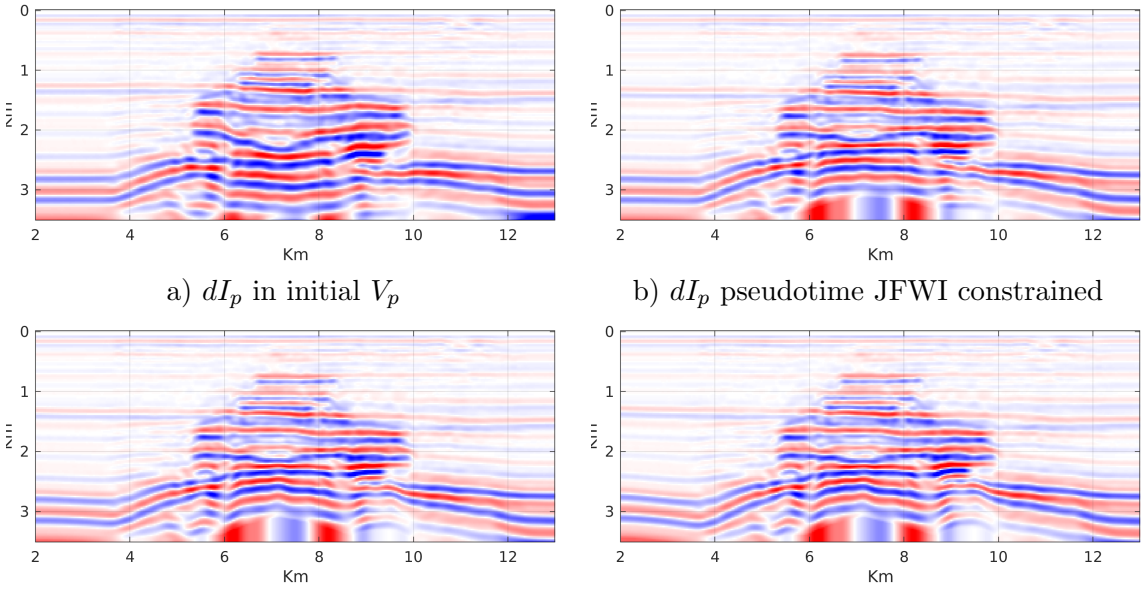


a) Time GSOT-JFWI constrained b) Time GSOT-JFWI unconstrained



e) Dip angle from reflectivity f) 4 km g) 8 km h) 11 km

Figure 5: Dip angle extracted from the I_p image and constrained V_p compared with unconstrained JFWI results. In the vertical profiles: constrained in blue, unconstrained in red, true in black and starting in grey.



a) dI_p in initial V_p b) dI_p pseudotime JFWI constrained

c) dI_p pseudotime JFWI d) dI_p pseudotime RWI

Figure 6: I_p reflectivity updated consistently with V_p-model building in the pseudotime domain

against time-shifts.

In order to assess the efficacy of the pseudotime-GSOT scheme in the presence of structural dips that might challenge the vertical traveltime assumption, the same comparison is presented for a shot gather acquired above the left flank of the deep anticline (Figure 8). In this instance, with a local structural dip reaching 15 degrees, pseudotime GSOT-JFWI shows no deterioration of performance compared to the nearly-flat case (Figure 7)

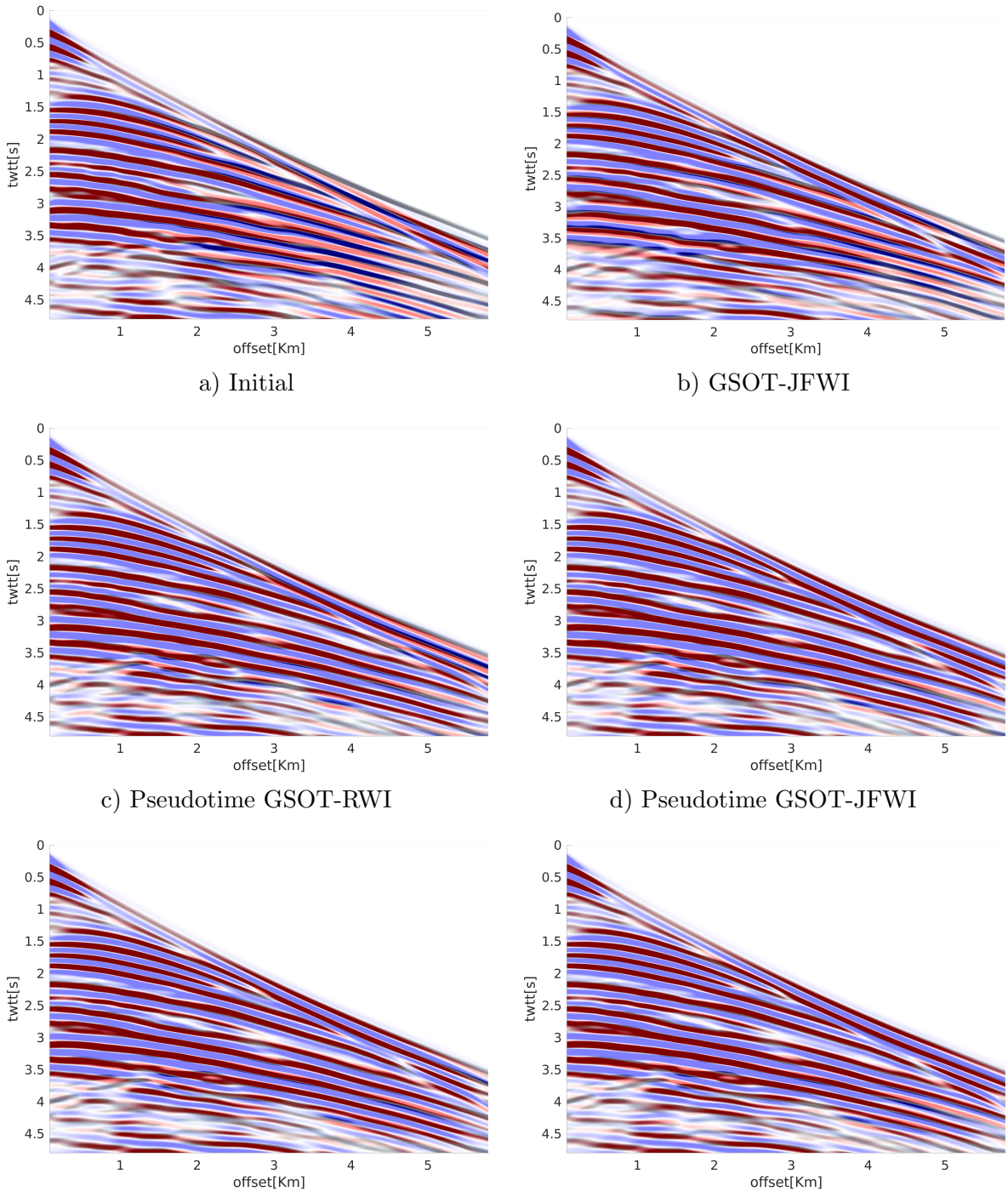
V_p-FWI

After pseudotime JFWI, parametrisation is switched to $V_p - \rho$ and one-parameter V_p-FWI is run in depth domain, in order to invert for the broadband V_p model (Operto et al., 2013). Density is set back to the initial uniform model, because in the $V_p - I_p$ parametrisation ρ moves to non-meaningful values when inverting for V_p only. Under these conditions, FWI yields a broadband reconstruction of the true V_p model (Figure 9), thanks to the accurate low-wavenumbers contained in the starting model, which avoid cycle skipping and ensure convergence. This appears to be true even using a L^2-norm at this stage, while the benefits of using GSOT (Δt=0.25 s) at this stage are marginal, thanks to a starting model that predicts traveltimes within less than half a period (see data fit in Figure 6). On the contrary, starting from the 1D model, and using the same GSOT parameters as in JFWI, V_p-FWI fails to provide meaningful velocity updates at depths larger than 1 km, yielding a migration-like image in the inaccurate kinematic model. This shows that not only does broadband FWI require convex misfit function, but it also relies on dedicated strategies such as JFWI to obtain tomographic updates in areas of the subsurface sampled by reflections only. Finally, the rate of data misfit reduction with iterations is much faster when starting from the JFWI-based models, and particularly from the structure-constrained pseudotime GSOT-JFWI (Figure 10), showing that structures in the I_pWI image at the V_p-building stage are precious information at no extra cost that ought to be exploited (Figure 5).

The Chevron 2014 blind reflection dataset

Data and initial model

The Chevron-2014 benchmark dataset simulates a realistic two-dimensional reflection survey in a shallow water environment and an isotropic elastic subsurface. A surface streamer records the wavefield at source-receiver offsets ranging between 0 and 8 km and group interval of 25 m. The wide source band is contaminated by frequency-dependent noise, resulting in a lack of usable frequencies < 3 Hz. Although the true model has not yet been disclosed, the available 1D starting Vp-model (Figure 11) suggests the presence of a low velocity layer (LVL) at about 2 km depth, with a maximum target depth of 6 km. A virtual V_p log is available for quality-check in the right part of the model at depths ranging between 1 and 2 km. Low offset/depth ratio, lack of low frequencies and the presence of a LVL jointly contribute to limiting the broadband Vp-reconstruction to the shallow (< 2 km depth) portion of the subsurface (Vigh et al., 2016). In addition, the 1D starting model exposes the L^2-based inversion to cycle-skipping (e.g., Sun and Alkhalifah, 2018). In this test, the inversion is run in the isotropic acoustic approximation discretized with a 25 m spatial step up to 15 Hz to avoid numerical dispersion, the temporal sampling interval chosen



a) Initial

b) GSOT-JFWI

c) Pseudotime GSOT-RWI

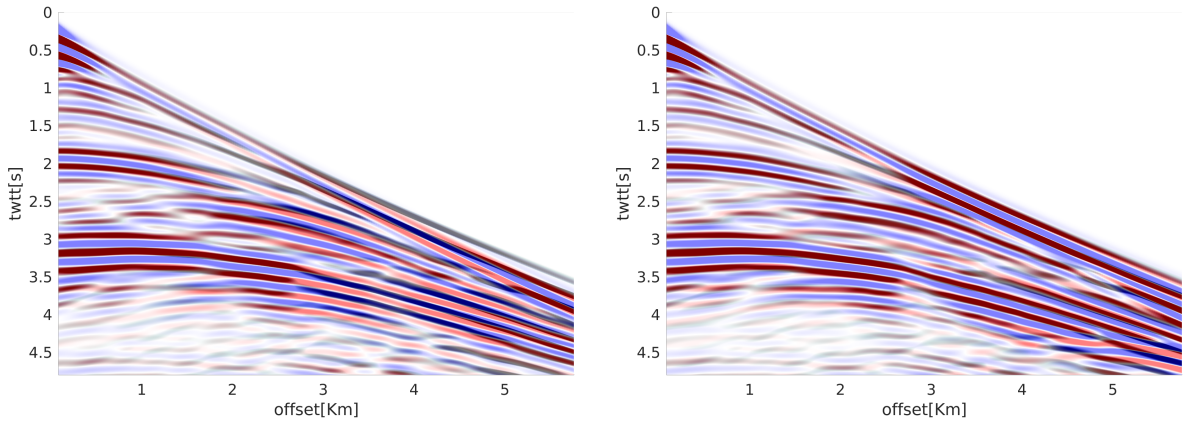
d) Pseudotime GSOT-JFWI

e) Pseudotime L^2-JFWI

f) Pseudotime GSOT-JFWI constrained

Figure 7: Example common-shot gather fit. Real data in black overlaid to synthetic in red and blue; data perfectly in phase would appear as blue and black, as red is covered by black

to honor CFL (Courant-Friedrichs-Lewy) stability conditions. PML boundary conditions are applied, except at the top of the model, in order to simulate free-surface condition and



a) Initial b) Pseudotime GSOT-JFWI

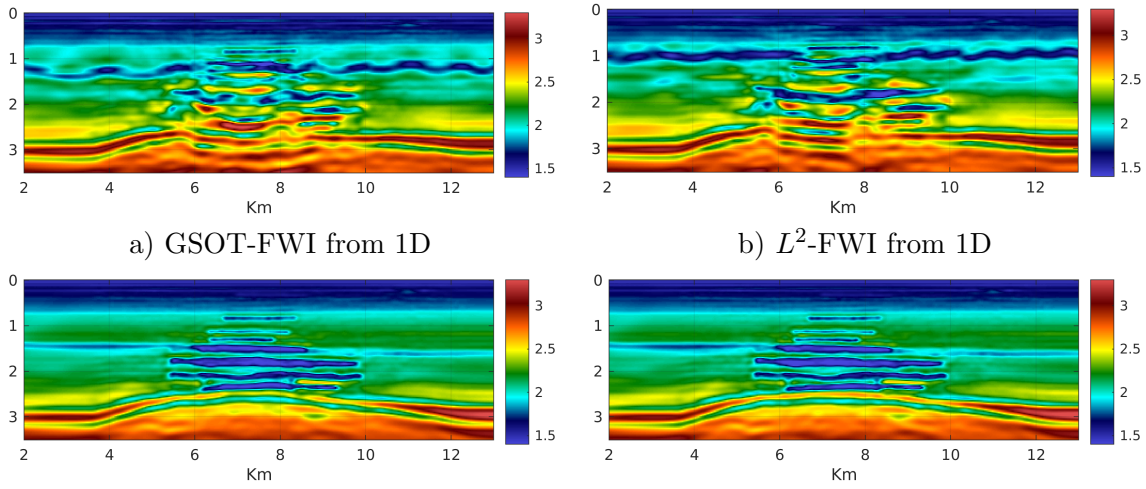
Figure 8: Example common-shot gather fit. As in Figure 7, but for a shot acquired above the left flank of the deep anticline.

reproduce both the up- and downgoing wavefield components of the field dataset.

V_p-macromodel building

In order to minimize the risk of cycle skipping in this more realistic test, we employ a multi-scale approach (Bunks et al., 1995) in three frequency bands (3-6,3-8,3-10 Hz). For each band, the V_p-model building workflow is similar to the one designed for the Valhall synthetic case, with an initial asymptotic I_pWI performed in 6 non-linear conjugate gradient iterations, followed by 30 steepest-descent JFWI iterations. A frequency-domain wavelet estimation (Pratt, 1999) is performed for each band using a subset of the short-offset direct arrivals to minimize the impact of the model inaccuracy on the source signature estimation. Diving/reflection separation is designed for each frequency band, to account for the bandwidth-dependent wavelet duration after the direct arrival onset. I_pWI is run on short spread reflection data, with offset shorter than 1 km, while offsets lower than 500 m are weighted out for JFWI, to avoid the inversion to be biased by the near-field high amplitudes. In the GSOT case, the maximum expected time-shift Δt (equation 6) is initially set to 0.35 s to enhance convexity with respect to traveltime differences, and later reduced to 0.15 s in the latest band, to approximate the resolution of the L^2 norm, as model kinematics is expected to have been improved in the previous bands.

In Figure 12 the results of V_p-model building are presented. Depth-domain JFWI (Figure 12a), provides sensible updates at depths lower than 2 km, where three localized low-V_p pockets are identified, but yields a non-realistic reconstruction of the LVL extent and geometry. In the pseudotime-JFWI cases (Figure 12b-d) LVL velocity and slope is updated as well as the shallow part of the subsurface. Starting from the one-dimensional model, a deep anticline-like higher-V_p structure is retrieved. While L^2 norm pseudotime inversion (Figure 12b) reconstructs a similar LVL as GSOT-JFWI (Figure 12c-d), the left part of the model, dominated by long offsets diving waves, shows signs of cycle skipping by leakage of the LVL upwards and downwards. Finally, the structure oriented smoothing applied to pseu-



a) GSOT-FWI from 1D

b) L^2-FWI from 1D

c) GSOT-FWI from GSOT-JFWI

d) L^2-FWI from GSOT-JFWI

Figure 9: V_p FWI

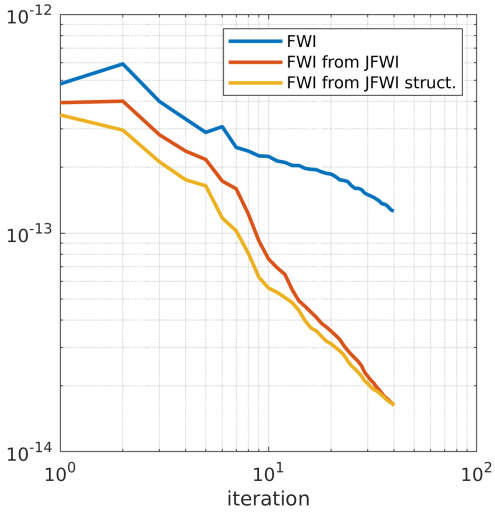
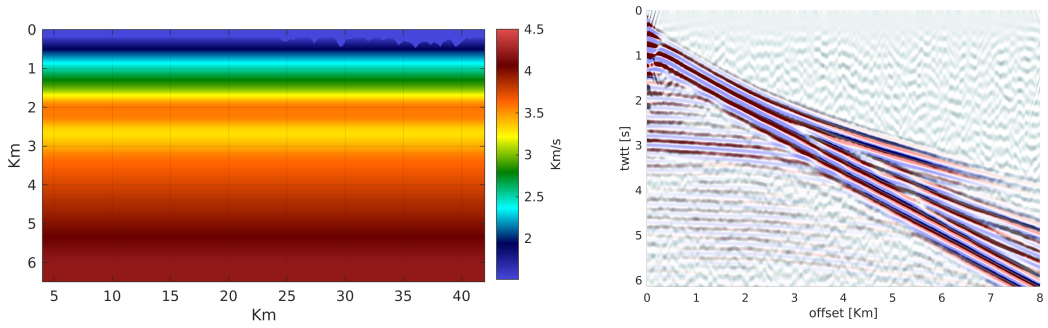


Figure 10: Convergence of V_p-FWI starting from 1D, JFWI, and structure-constrained JFWI models.



a) Initial V_p model b) Example common shot gather fit

Figure 11: Initial V_p for the Chevron '14 blind test and example observed vs predicted common shot gather, showing phase mismatch (red wiggles appearing) at offsets larger than 3.5 km

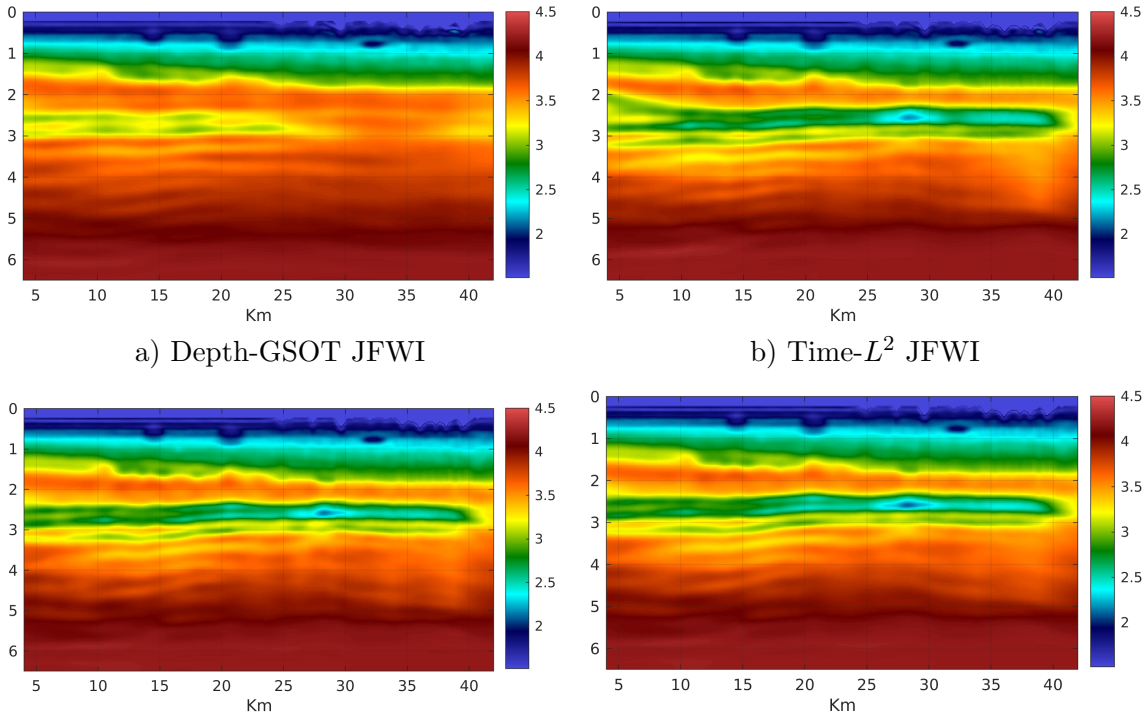
.

dotime GSOT-JFWI attenuates non-physical near-vertical anomalies at depths larger than 2.5 km, (Figure 12c-d), thereby introducing more realistic structures in the reconstructed macromodel (Figure 13).

In Figure 14 the stabilizing contribution of the diving wave component of JFWI is demonstrated by halving the diving wave weight and comparing the results with the low-pass filtered virtual well-log. Note how the updated V_p in this case (J-RWI, in Figure 14b as the reflection component dominates) shows obvious signs of velocity under- and over-estimation in pull-ups and -downs and, in particular, anomalously low velocities at the well-log location (Figure 14d). This might result from reflection overfitting, driven by the presence of converted waves and multiples in the elastic field dataset in the acoustic inversion. On the other hand, if only diving waves are used by making the contribution of reflections negligible (J-EWI in Figure 14a), the deep model update is, as a consequence, neglected and well-log poorly reconstructed at 2 km (Figure 14d); the velocity decrease at this depth is not captured by J-EWI, which remains grounded to the starting model due to lack of illumination. Jointly inverting for reflection and refraction in this dataset is thus key to obtain a stable macromodel update at all depths. In the following kinematic QCs, we demonstrate that the lower velocities detected by JFWI are key for correct imaging and that accounting for both diving and reflected waves is necessary to obtain a sensible macromodel update at all depths.

Data-fit and kinematics QC

In Figure 15 an example common-shot gather (CSG) extracted in the central part of the model is used to showcase the data-fit after JFWI at 6 Hz. Note how the pseudotime approaches (Figure15b-c) improve the waveform fit at all offsets compared to the prediction of the initial model (Figure15a), whereas the depth-domain inversion (Figure15d), while improving the average fit between predicted and observed gathers, deteriorates the phase alignment at near-offsets due to the fixed-depth reflectivity. In more detail, cycle-skipping in the L^2 gather is not apparent in this CSG, although waveform prediction is marginally



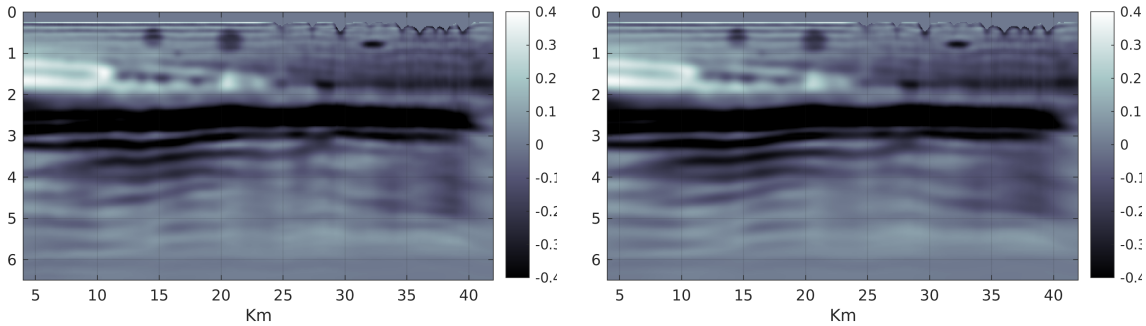
a) Depth-GSOT JFWI

b) Time-L^2 JFWI

c) Time-GSOT JFWI

d) Time-GSOT JFWI structure-smoothing

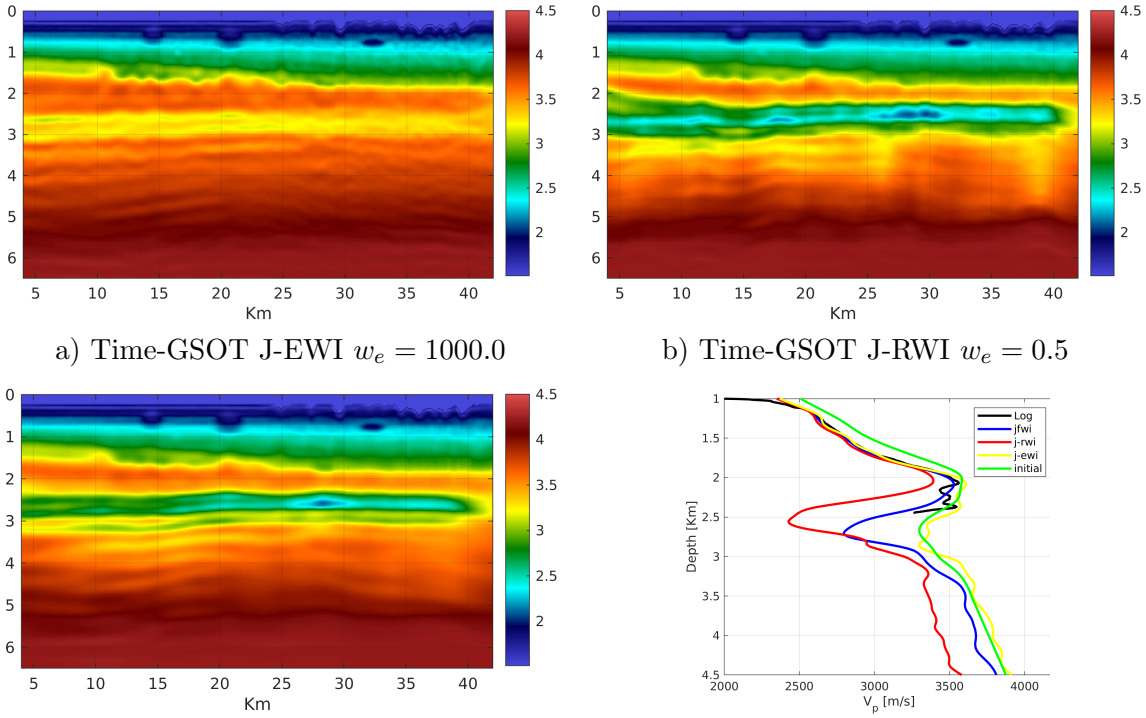
Figure 12: V_p-model building via JFWI.



a) dV_p JFWI

b) dV_p JFWI structure-smoothing

Figure 13: Pseudotime GSOT JFWI model update without (a) and with (b) structure-oriented smoothing



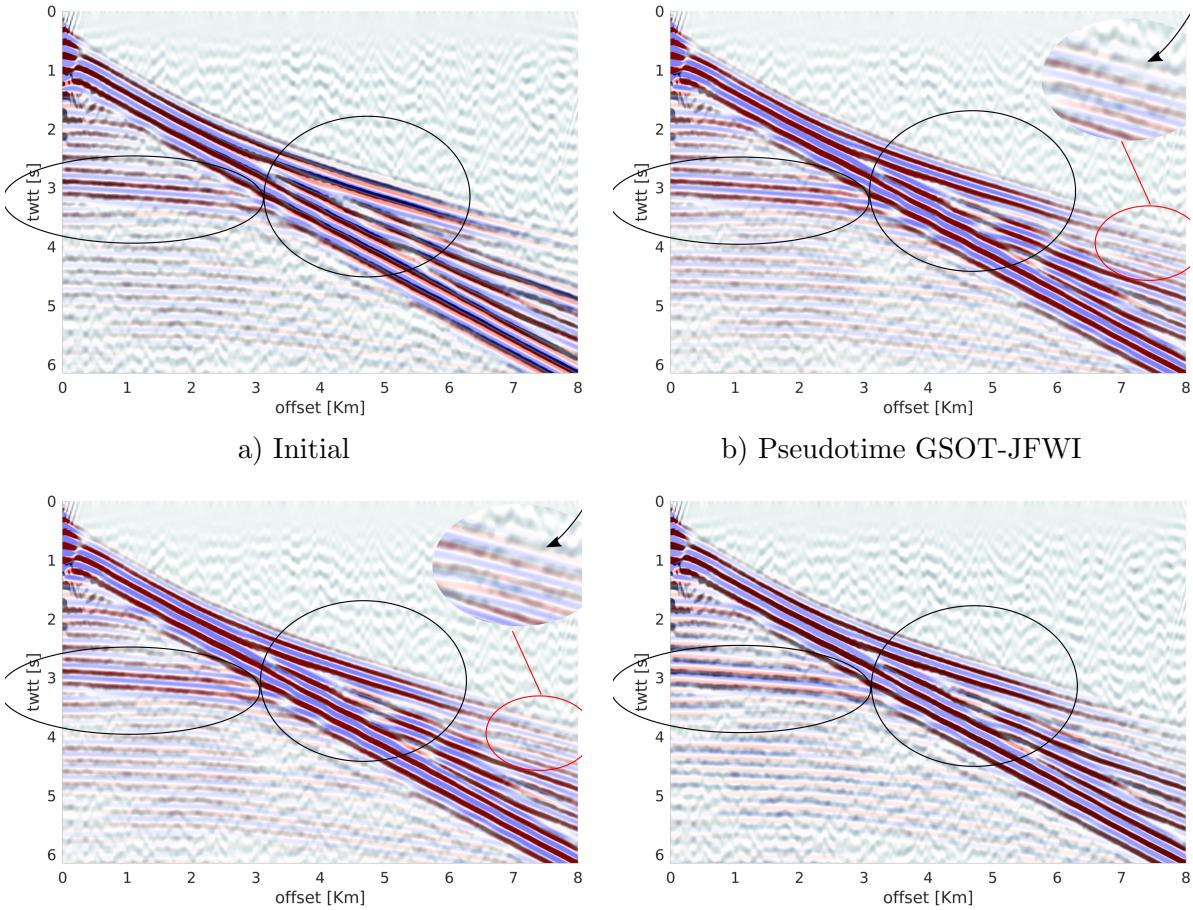
a) Time-GSOT J-EWI $w_e = 1000.0$

b) Time-GSOT J-RWI $w_e = 0.5$

c) Time-GSOT JFWI $w_e = 1.0$

d) well log comparison

Figure 14: Stabilizing effect of diving wave component in JFWI. The best well-log prediction is obtained when using the true weight of the diving wave component ($W_e = 1$ in the JFWI objective function). While reflections are necessary to reconstruct the true LVL values, their contribution alone might be unstable due to acoustic inversion vs elastic modeling.



a) Initial

b) Pseudotime GSOT-JFWI

c) Pseudotime L^2-JFWI

d) Depth GSOT-JFWI

Figure 15: Example 6 Hz common-shot gather band fit. Real data in black overlaid to synthetic in red and blue; data perfectly in phase would appear as blue and black, as red is covered by black. The long offset refractions in the pseudotime L^2 and GSOT cases are zoomed in, highlighting a better data fit in the GSOT case (red covered by black)

worse than the GSOT case.

The V_p-models obtained in the pseudotime domain are compared in the image domain to assess their kinematic accuracy (Figure16), in particular to assess wether or not the LVL reconstructed by the reflection component is necessary to appropriately image the susbsurface structures. Common image gather (CIGs) are computed by extended RTM on offset subsets from 0.2 to 5 km for: a) Initial model; b) pseudotime diving wave only GSOT WI; c) pseudotime GSOT JFWI. The solution of pseudotime JFWI best represents the model kinematics, both compared to the initial model and the one obtained by early-arrivals only JFWI (J-EWI), the latter showing clear signs of velocity overestimation (downward bending arrivals).

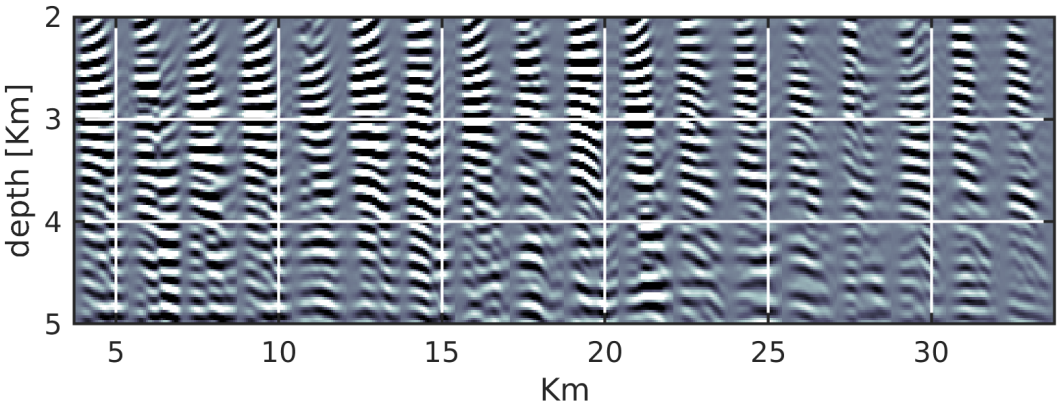
V_p-FWI

FWI starting from the JFWI models is performed with a multi-scale approach in 6 frequency bands from 6 to 15 Hz. In V_p-FWI, we re-set density to its initial value and keep it fixed. In Figure 17 the contribution of JFWI as starting point for V_p-FWI (Figure 17a-c) is apparent in its richer low wavenumber content and better imaging of the deep structures, compared to FWI starting from the initial model (Figure 17d). At this stage, GSOT-FWI (Figure 17a and c) is essentially undistinguishable from L^2 FWI (Figure 17b), because the risk of cycle skipping is already minimised by the accurate kinematics of the pseudotime GSOT JFWI V_p starting model. Similarly, the structure-oriented JFWI starting model does not significantly outperform the isotropic-smoothing one (Figure 17 a vs c), although a marginally higher convergence rate (Figure 18b) is obtained at the first FWI frequency band when starting from the constrained JFWI model. The final quality-check on the available V_p-log is comforting, and further confirms the importance of JFWI in fitting the log at 2.0 to 2.5 km depth (Figure 18a).

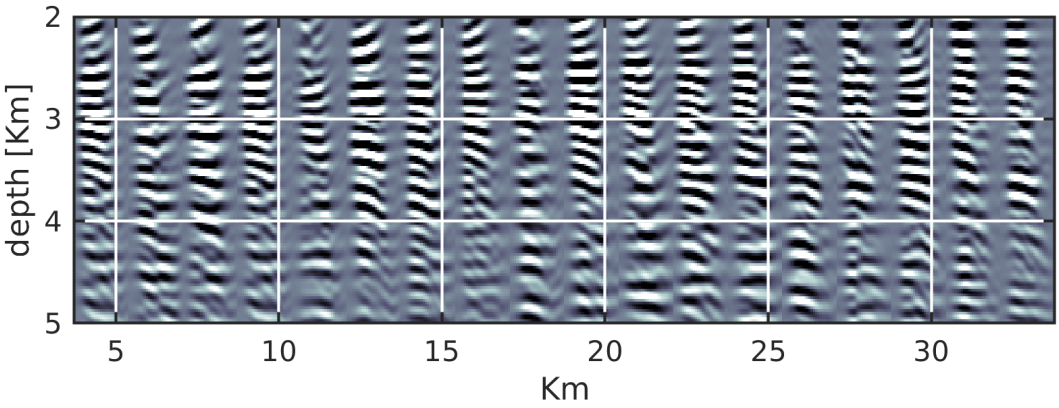
The stationarity of shot-by-shot wavelet estimation is then compared as final QC of the broadband FWI V_p models. In the correct model, the latter should be independent of the shot (Pratt, 1999; Operto et al., 2015), while in an incorrect one, source estimation would collect the space-dependent model inaccuracies in the results. For each of the FWI cases, we compute the L^2-norm misfit between the average estimated wavelet and the wavelet estimated for each shot. The results, shown in Figure 19, support the claim that FWI starting from the structure-constrained JFWI model ensures better performance than FWI-only, and reveal a better accuracy of the GSOT-FWI model over L^2-FWI, even when starting from the accurate JFWI model.

DISCUSSION

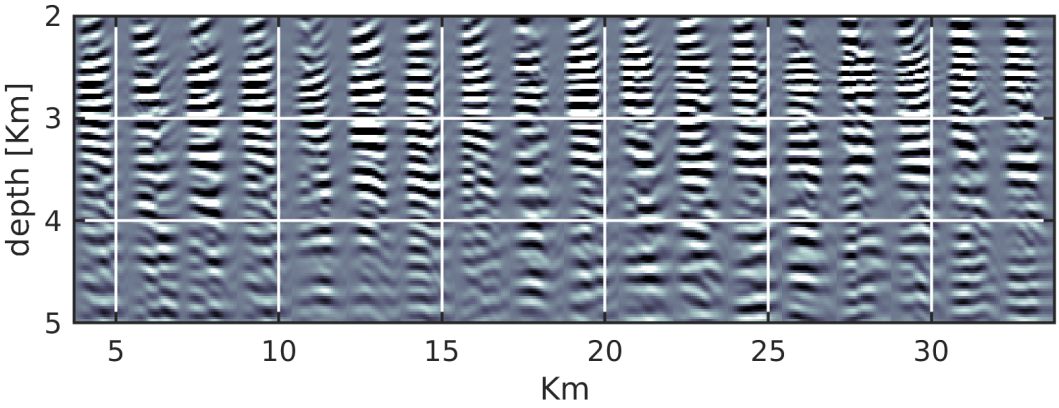
The results obtained in both synthetic tests point towards confirming the suitability of the proposed methodology with coherent results. It is however interesting to note how the diving wave component of the JFWI objective function, while it is only marginally important in the first example application, is paramount in stabilizing the inversion in the Chevron dataset. In this case, the acoustic approximation adopted to invert an elastic dataset introduces systematic errors in modeling the reflection amplitude-versus offset, as well as the impossibility to predict P-S-P converted waves. Our interpretation is that such an epistemic



a) RTM CIG gathers in initial model

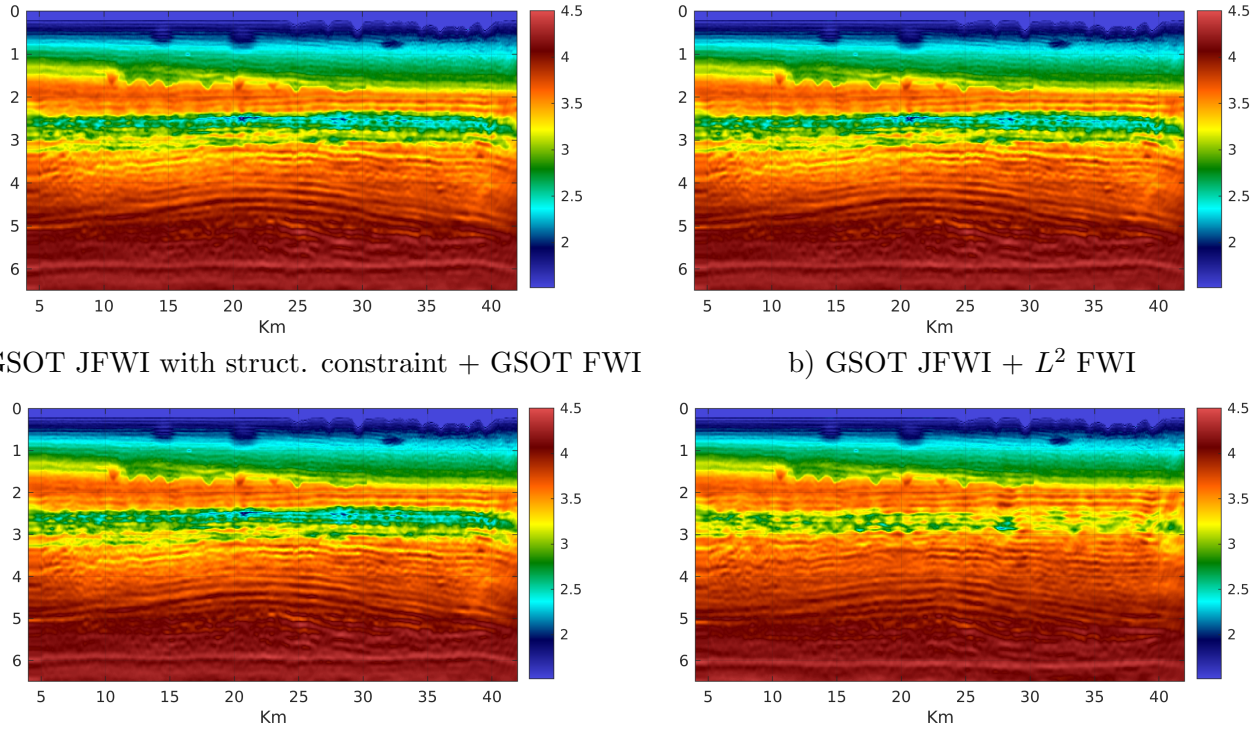


b) RTM CIG gathers in J-EWI model



c) RTM CIG gathers in JFWI model

Figure 16: Offset-domain RTM Common Image Gathers for the initial (a), Pseudotime GSOT JFWI with early wave only (J-EWI, b), and Pseudotme GSOT JFWI (c)



a) GSOT JFWI with struct. constraint + GSOT FWI

b) GSOT JFWI + L^2 FWI

c) GSOT JFWI + GSOT FWI

d) GSOT FWI from Init.

Figure 17: V_p-FWI results in the [2-3-12-14]Hz band

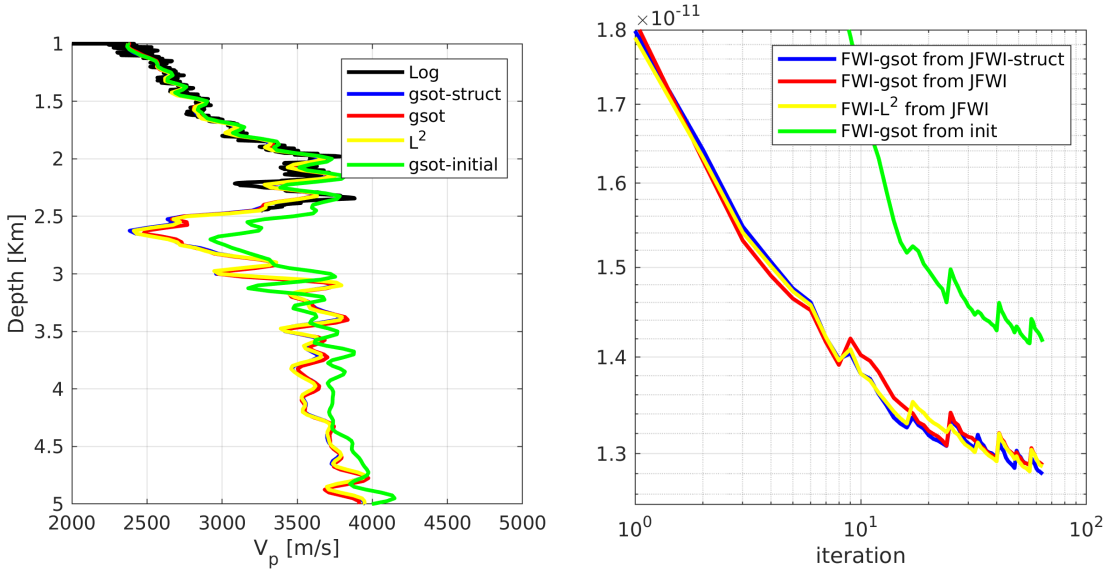
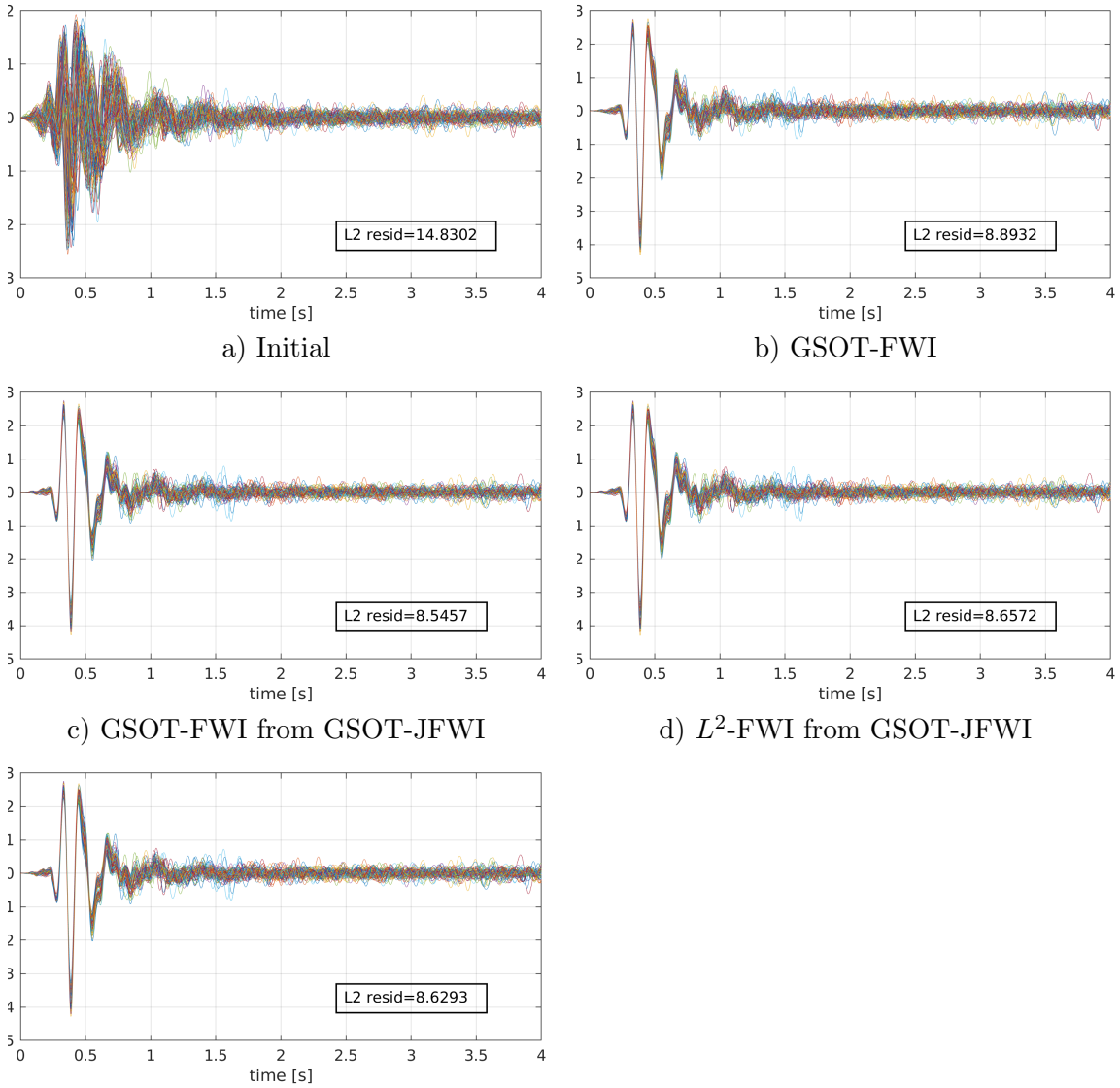


Figure 18: V_p-FWI results in the [2-3-12-14]Hz band. On the left, fit with V_p log at 39.5 km horizontal position; on the right convergence history for the first FWI frequency band.



a) Initial

b) GSOT-FWI

c) GSOT-FWI from GSOT-JFWI

d) L^2-FWI from GSOT-JFWI

e) GSOT-FWI from GSOT-JFWI no-struct

Figure 19: Shot-by-shot source wavelet estimation quality-check after FWIs. The mean L^2 value is calculated between the average wavelet and each shot's estimated source.

modeling error makes a reflection-only FWI unstable, regardless the objective function used. Introducing the early-wave component of JFWI, stabilizes the velocity estimation at 2 km depth and makes it possible to obtain sensible model reconstruction at all depths.

The acoustic approximation remains however a limitation of JFWI in an elastic world, despite the stabilizing effect of the diving-wave component in the objective function. A dedicated pre-processing might be necessary on datasets with important elastic imprints on the reflected waveform shape and polarity (Ostrander, 1984), while an amplitude-offset normalization may be easily applied to GSOT to compensate for amplitude-only variations (AVO). Other approaches may be envisaged that reproduce elastic effects by adding artificial source terms (Hobro et al., 2014) to the predicted acoustic data. Similarly, the impact of attenuation may be important on the amplitude and phase of reflections, and including a Quality-factor model will be key for JFWI in attenuating media.

As shown in Zhou et al. (2015), the computational cost of JFWI is twice as large as the one FWI, due to the two adjoint-state steps adopted to eliminate the migration isochrones from the FWI gradient and thus obtain the desired gradient in equation3. The two-step gradient computation uses full-waveform modeling to obtain the desired tomographic waveform-based V_p updates (Brossier et al., 2015; Zhou et al., 2015). In order to isolate the macromodel updates from the migration isochrones, reflection-based FWI methods in the literature adopt several alternative approaches, based on either scattering-angle filtering of the gradient or Born modeling of reflections (Yao et al., 2020). While the former would require in time-domain computing asymptotic quantities to extract and isolate specific illumination angles, the latter employs a linearized one-way wave propagation that, unlike the full-waveform approach of JFWI, does not account for possible multi-scattering, inter-bed multiples and free-surface effects.

Graph-space optimal transport has proven effective in averting cycle-skipping when starting from a simple initial guess of V_p. Though no offset windowing has been necessary in those datasets, a multi-scale approach was adopted to further minimize the risk of cycle-skipping. This is consistent with the real data application of GSOT-FWI shown in Górszczyk et al. (2021) and Pladys et al. (2022), who also show that GSOT does not altogether eliminate the need for data selection and weighting. The extra cost of GSOT with respect to L^2 norm in these 2D examples has been in the order of 20%. As shown in Métivier et al. (2019) and Pladys et al. (2022), such a computational burden is expected to become relatively less important with increasing frequency, thanks to a quadratically complexity, as opposed to the cubic one of the modeling (in two dimensions).

In the pseudotime formulation, reflectivity is passively repositioned in depth during V_p reconstruction, assuming near-vertical wave propagation (vertical traveltime). This approach has the merit of a negligible computing cost, while being limited to smoothly varying models, as is normally the case for macromodel reconstruction. Its applicability is reduced in complex subsurfaces with significantly dipping reflectors, while in strongly anisotropic media the depth adjustment based on vertical velocity may not capture the true model kinematics at non-normal incidence. Therefore, especially when starting from particularly non-informative initial models, interleaving stages of I_pWI between JFWI steps might still be necessary to ensure accuracy. Alternatively, Valensi and Baina (2021) propose to add a zero-offset invariance term to the RWI misfit function. This requires one Born modeling based migration of the zero-offset observed data and a demigration per V_p-gradient calculation, therefore providing a more general framework than pseudotime, though at an

important extra computing cost.

Thanks to our asymptotic approach, the cost of each I_pWI is significantly lower than gradient-based or quasi-newton inversions, at the expense of an extra memory requirement due to the storage of the decimated DFT wavefields. The asymptotic formulation of I_pWI is adopted as preconditioner of an indirect (iterative) inversion scheme, in other words acting as a set of weights for the adjoint-based FWI gradient that speeds up convergence to the true-amplitude reflectivity. Note that, unlike direct inversion approach, this is applicable to data containing multiples and ghosts and, generally, beyond the one-way Born approximation (Li and Chauris, 2018). It is worth pointing out that, however, the asymptotic approximation is valid on narrow reflection angle, narrow illumination angle data, and therefore appropriate short offset data weighting is necessary.

As for the structure-oriented smoothing, the availability of a reflective model makes it possible to extract structural information at almost no-extra cost, and constrain the intermediate wavenumbers of the model, partially compensating for the ill-posed nature of the velocity reconstruction inverse problem. Also in this case, however, it might be wise to start injecting structural information into V_p inversion only after some JFWI iterations, i.e. after the structural dips in the I_p image begin to approach the true ones.

In both the synthetic datasets presented, a depth-domain inner-loop approach had been recently adopted with reasonable results in previous publications, (e.g., Zhou et al., 2015; Provenzano et al., 2020). In this case, JFWI converges to models containing sensible features at the expense of carefully designing a mid-to-long offset window for V_p update and of recomputing I_pWI twice per frequency band to attempt to mantain reflectivity-kinematic consistency (Provenzano et al., 2021). The process is highly subjective and time consuming, and generally yields sub-optimal results, as pointed out in, e.g., Valensi and Baina (2021).

CONCLUSIONS

In this paper, a robust and efficient velocity-model-building methodology for limited-offset marine seismic reflection data is presented. It requires minimal data pre-processing and little initial knowledge of the subsurface. It is based on a joint graph-space optimal transport objective function comprising a refraction and a reflection term (GSOT-JFWI), and relies on a domain transformation from depth to pseudotime. While the joint full waveform inversion permits to extend velocity inversion beyond the depth sampled by diving waves, the combination of GSOT and pseudotime is uniquely suited to ensure phase-consistency (avoid cycle-skipping) at both short and long offsets in a cheap and robust fashion. Structure-derived constraints are injected in V_p reconstruction at no extra cost through anisotropic Bessel smoothing. The cost of building the initial reflective model is greatly reduced by a dedicated asymptotic preconditioner for impedance waveform inversion (I_pWI). Two complex synthetic examples demonstrate the efficacy of the methodology and the impact of the proposed advances compared to conventional depth-domain least-squares methods. While the pseudotime transformation might be sub-optimal for complex subsurfaces and strongly dipping interfaces it still has wide applicability in a wide range of geological context, and its benefits in terms of computing cost are large compared to migration/demigration approaches. On the other hand, important elastic effects affecting reflection phase and polarity might hamper the efficacy of JFWI due to its acoustic approximation. Notwithstanding these limitations, the proposed methodology paves the way towards a mature and widely ap-

plicable velocity-model-building strategy from reflection data and little a-priori knowledge. A follow-up paper will be aimed at demonstrating its effectiveness on a 3D ocean-bottom-cable dataset.

ACKNOWLEDGMENTS

REFERENCES

Alkhalifah, T., S. Fomel, and B. Biondi, 2001, The space-time domain: theory and modelling for anisotropic media: Geophysical Journal International, **144**, 105–113.

Beylkin, G., 1985, Imaging of discontinuities in the inverse scaterring problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, **26**, 99–108.

Brenders, A., J. Dellinger, C. Kanu, Q. Li, and S. Michell, 2018, *in* The Wolfspar-field trial: Results from a low-frequency seismic survey designed for FWI: SEG, 1083–1087.

Brossier, R., S. Operto, and J. Virieux, 2015, Velocity model building from seismic reflection data by full waveform inversion: Geophysical Prospecting, **63**, no. 2, 354–367.

Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Chavent, G., 2017, Data space reflectivity and the migration based travel time approach to fwi: Presented at the 79th EAGE Conference and Exhibition 2017-Workshops.

Chavent, G., F. Clément, and S. Gòmez, 1994, Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example: SEG Technical Program Expanded Abstracts 1994, 1179–1182.

Chi, B., L. Dong, and Y. Liu, 2015, Correlation-based reflection full-waveform inversion: Geophysics, **80**, R189–R202.

Claerbout, J., 1985, Imaging the Earth's interior: Blackwell Scientific Publication.

Davy, R. G., J. V. Morgan, T. A. Minshull, G. Bayrakci, J. M. Bull, D. Klaeschen, T. J. Reston, D. S. Sawyer, G. Lymer, and D. Cresswell, 2017, Resolving the fine-scale velocity structure of continental hyperextension at the Deep Galicia Margin using full-waveform inversion: Geophysical Journal International, **212**, 244–263.

Devaney, A., 1984, Geophysical diffraction tomography: Geoscience and Remote Sensing, IEEE Transactions on, **GE-22**, 3–13.

Doherty, S., and J. Claerbout, 1976, Structure independent velocity estimation: Geophysics, **41**, 850–881.

Eiken, O., 2019, Twenty years of monitoring CO2 injection at Sleipner, *in* Geophysics and Geosequestration: Cambridge University Press, 209–234.

Farshad, M., and H. Chauris, 2021, Accelerating the multi-parameter least-squares reverse time migration using an appropriate preconditioner: Computational Geosciences, **25**, 2071–2092.

Fichtner, A., and A. Villaseñor, 2015, Crust and upper mantle of the western mediterranean – constraints from full-waveform inversion: Earth and Planetary Science Letters, **428**, 52 – 62.

Gholami, Y., R. Brossier, S. Operto, V. Prieux, A. Ribodetti, and J. Virieux, 2013, Which parametrization is suitable for acoustic VTI full waveform inversion? - Part 2: application to Valhall: Geophysics, **78**, R107–R124.

Górszczyk, A., R. Brossier, and L. Métivier, 2021, 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, **126**, e2020JB021504.

Guasch, L., O. C. Agudo, M.-X. Tang, P. Nachev, and M. Warner, 2020, Full-waveform inversion imaging of the human brain: NPJ digital medicine, **3**, 1–12.

Hale, D., 2009, A method for estimating apparent displacement vectors from time-lapse seismic images: Geophysics, **74**, P99–P107.

Hicks, E., H. Hoeber, M. Houbiers, S. P. Lescoffit, A. Ratcliffe, and V. Vinje, 2016, Time-

lapse full-waveform inversion as a reservoir-monitoring tool − A North Sea case study: The Leading Edge, **35**, 850–858.

Hobro, J. W., C. H. Chapman, and J. O. Robertsson, 2014, A method for correcting acoustic finite-difference amplitudes for elastic effects: Geophysics, **79**, T243–T255.

Irnaka, T. M., R. Brossier, L. Metivier, T. Bohlen, and Y. Pan, 2022, 3D Multi-component Full Waveform Inversion for Shallow-Seismic Target: Ettlingen Line Case Study: Geophysical Journal International, **229**, 1017–1040.

Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, and D. Trezeguet, 1989, Wavelengths of Earth structures that can be resolved from seismic reflection data: Geophysics, **54**, 906–910.

Kamath, N., R. Brossier, L. Métivier, A. Pladys, and P. Yang, 2021, Multiparameter full-waveform inversion of 3D ocean-bottom cable data from the Valhall field: Geophysics, **86**, B15–B35.

Kantorovich, L., 1942, On the transfer of masses: Dokl. Acad. Nauk. USSR, **37**, 7–8.

Kryvohuz, M., H. Kuel, R. Plessix, and Y. Yang, 2019, Reflection full-waveform inversion with data-space lsrtm: Presented at the SEG International Exposition and 89th Annual Meeting.

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

Li, Y., R. Brossier, and L. Métivier, 2019, Joint FWI for imaging deep structures: A graph-space OT approach: SEG Technical Program Expanded Abstracts 2019, 1290–1294.

Li, Y., and H. Chauris, 2018, Coupling direct inversion to common-shot image-domain velocity analysis: Geophysics, **83**, R497–R514.

Marty, P., C. Boehm, and A. Fichtner, 2021, Acoustoelastic full-waveform inversion for transcranial ultrasound computed tomography: Medical Imaging 2021: Ultrasonic Imaging and Tomography, International Society for Optics and Photonics, 1160211.

Métivier, L., and R. Brossier, 2016, The seiscope optimization toolbox: A large-scale nonlinear optimization library based on reverse communication: Geophysics, **81**, F11–F25.

———, 2021, New insights on the graph space optimal transport distance for full waveform inversion: Presented at the SEG Technical Program Expanded Abstracts 2021.

Métivier, L., R. Brossier, Q. Mérigot, and E. Oudet, 2019, A graph space optimal transport distance as a generalization of L^p distances: application to a seismic imaging inverse problem: Inverse Problems, **35**, 085001.

Métivier, L., R. Brossier, and J. Virieux, 2015, Combining asymptotic linearized inversion and full waveform inversion: Geophysical Journal International, **201**, 1682–1703.

Mora, P. R., 1989, Inversion = migration + tomography: Geophysics, **54**, 1575–1586.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.

Operto, S., R. Brossier, Y. Gholami, L. Métivier, V. Prieux, A. Ribodetti, and J. Virieux, 2013, A guided tour of multiparameter full waveform inversion for multicomponent data: from theory to practice: The Leading Edge, **Special section Full Waveform Inversion**, 1040–1054.

Operto, S., and A. Miniussi, 2018, On the role of density and attenuation in 3D multi-parameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea: Geophysical Journal International, **213**, 2037–2059.

Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse

isotropic approximation: Geophysical Journal International, **202**, 1362–1391.

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

Ostrander, W. J., 1984, Plane-wave reflection coefficients for gas sands at non-normal angles of incidences: Geophysics, **49**, 1637–1648.

Pladys, A., R. Brossier, N. Kamath, and L. Métivier, 2022, Robust FWI with graph space optimal transport: application to 3D OBC Valhall data: Geophysics, **87**, 1–76.

Pladys, A., R. Brossier, Y. Li, and L. Métivier, 2021, On cycle-skipping and misfit function modification for full-wave inversion: Comparison of five recent approaches: Geophysics, **86**, R563–R587.

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

———, 2013, A pseudo-time formulation for acoustic full waveform inversion: Geophysical Journal International, **192**, 613–630.

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

Provenzano, G., R. Brossier, and L. Métivier, 2020, Joint FWI of diving and reflected waves using a graph space optimal transport distance: synthetic tests on limited-offset surface seismic data: Presented at the SEG Technical Program Expanded Abstracts 2020.

Provenzano, G., L. Metivier, and R. Brossier, 2021, Joint reflection and diving fwi using graph-space optimal transport and structure-guided smoothing on benchmark data: 82^{th} Annual EAGE Conference & Exhibition, Amsterdam, Expanded Abstracts, 82^{th} Annual EAGE Conference & Exhibition, Amsterdam, EAGE, Th P01 05.

Provenzano, G., M. E. Vardy, and T. J. Henstock, 2018, Decimetric-resolution stochastic inversion of shallow marine seismic reflection data: dedicated strategy and application to a geohazard case study: Geophysical Journal International, **214**, 1683–1700.

Qin, B., T. Allemand, and G. Lambaré, 2015, Full waveform inversion using preserved amplitude reverse time migration: SEG Technical Program Expanded Abstracts 2015, 1252–1257.

Raknes, E. B., B. Arntsen, and W. Weibull, 2015, Three-dimensional elastic full waveform inversion using seismic data from the sleipner area: Geophysical Journal International, **202**, 1877–1894.

Sambolian, S., S. Operto, A. Ribodetti, B. Tavakoli, and J. Virieux, 2019, Parsimonious slope tomography based on eikonal solvers and the adjoint-state method: Geophysical Journal International, **218**, 456–478.

Schleicher, J., C. J.C., and N. A., 2008, A comparison of imaging conditions for wave-equation shot-profile migration: Geophysics, **73**, S219–S227.

Sirgue, L., 2003, Inversion de la forme d'onde dans le domaine fréquentiel de données sismiques grand offset: PhD thesis, Université Paris 11, France - Queen's University, Canada.

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

Snieder, R., M. Y. Xie, A. Pica, and A. Tarantola, 1989, Retrieving both the impedance contrast and background velocity: a global strategy for the seismic reflection problem: Geophysics, **54**, 991–1000.

Sun, B., and T. Alkhalifah, 2018, Mitigate cycle skipping in fwi: a generalized instanatenous travel-time approach: Presented at the 80th EAGE Conference & Exhibition, 11-14 June 2018, Copenhagen, Denmark.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

———, 1986, A strategy for non linear inversion of seismic reflection data: Geophysics, **51**, 1893–1903.

Trinh, P. T., R. Brossier, L. Métivier, J. Virieux, and P. Wellington, 2017, Bessel smoothing filter for spectral element mesh: Geophysical Journal International, **209**, 1489–1512.

Valensi, R., and R. Baina, 2021, A time consistent waveform inversion (twin) method: Presented at the 82nd EAGE Annual Conference & Exhibition.

Vigh, D., X. Cheng, K. Jiao, and D. Sun, 2019, Keys to robust reflection-based full-waveform inversion: Presented at the 81st EAGE Conference & Exhibition 2019.

Vigh, D., K. Jiao, X. Cheng, D. Sun, and W. Lewis, 2016, Earth-model building from shallow to deep with full-waveform inversion: The Leading Edge, **35**, 1535–1554.

Vigh, D., N. Moldoveanu, K. Jiao, W. Huang, and J. Kapoor, 2013, Ultralong-offset data acquisition can complement full-waveform inversion and lead to improved subsalt imaging: The Leading Edge, **32**, 1116–1122.

Virieux, J., R. Brossier, L. Métivier, S. Operto, and A. Ribodetti, 2016, Direct and indirect inversions: Journal of Seismology, **20**, 1107–1121.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Wang, C., P. Farmer, D. Yingst, I. Jones, C. Calderon, and J. Brittan, 2019, Preconditioned reflection full waveform inversion for subsalt imaging: Presented at the 81st EAGE Conference & Exhibition 2019.

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

Wu, R. S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11–25.

Xu, S., D. Wang, F. Chen, G. Lambaré, and Y. Zhang, 2012, Inversion on reflected seismic wave: SEG Technical Program Expanded Abstracts 2012, 1–7.

Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018, A Time-Domain Preconditioned Truncated Newton Approach to Multiparameter Visco-acoustic Full Waveform Inversion: SIAM Journal on Scientific Computing, **40**, B1101–B1130.

Yang, Y., J. Ramos-Martinez, D. Whitmore, G. Huang, and N. Chemingui, 2021, Simultaneous velocity and reflectivity inversion: Fwi+ lsrtm: Presented at the 82nd EAGE Annual Conference & Exhibition.

Yao, G., N. V. da Silva, and D. Wu, 2019, Reflection-waveform inversion regularized with structure-oriented smoothing shaping: Pure and applied geophysics, **176**, 5315–5335.

Yao, G., D. Wu, and S. Wang, 2020, A review on reflection waveform inversion: Petroleum science, **17**, 334–351.

Yong, P., R. Brossier, and L. Métivier, 2022, Parsimonious truncated newton method for time-domain full-waveform inversion based on the fourier-domain full-scattered-field approximation: Geophysics, **87**, R123–R146.

Zhou, W., R. Brossier, S. Operto, and J. Virieux, 2015, Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation: Geophysical Journal International, **202**, 1535–1554.

Zhou, W., R. Brossier, S. Operto, J. Virieux, and P. Yang, 2018, Velocity model building by waveform inversion of early arrivals and reflections: a 2d ocean-bottom-cable study

with gas cloud effects: Geophysics, **83(2)**, R141–R157.