Robust full-waveform inversion with graph-space optimal transport: Application to 3D ocean-bottom cable Valhall data

Arnaud Pladys, Romain Brossier, Nishant Kamath, Ludovic Métivier

Geophysics

Robust FWI with graph space optimal transport: application to 3D OBC Valhall data

Right running head: GSOT FWI: application to 3D Valhall

Arnaud Pladys*, Romain Brossier*, Nishant Kamath† and Ludovic Métivier‡*

January 20, 2022

*Univ. Grenoble Alpes, ISTerre, F-38000 Grenoble, France
†Formerly Univ. Grenoble Alpes, now CGG Crawley Research Center, Crawley, United Kingdom
‡Univ. Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France

Robust FWI with graph space optimal transport: application to 3D OBC Valhall data

(January 20, 2022)

Running head: **GSOT FWI: application to 3D Valhall**

ABSTRACT

Improving full-waveform inversion to make it more robust to cycle-skipping has been the subject of a large number of studies. From the several families of approaches developed, one of the most documented consists in modifying the least-squares distance defining the discrepancy between observed and calculated data. From all the propositions made to improve and replace the least-squares distance, only a few of them have been applied to field data. One of the methods proposed recently, the graph space optimal transport distance, presents appealing properties for field data applications. We compare it with the least-squares distance in an analysis performed on the three-dimensional ocean bottom cable data from the Valhall field. This data has already been at the heart of several full-waveform inversion studies, making it an excellent candidate to evaluate the properties of this new misfit function. We first perform this comparison starting the inversion from the reflection traveltime tomography model used in previous studies. We then perform a second comparison from a crude, linearly-varying-in-depth one-dimensional velocity model. Starting from this model, least-squares-based full-waveform inversion fails to provide a meaningful estimate of the pressure-wave velocity model due to cycle skipping. We illustrate how the graph-space optimal transport-based full-waveform inversion mitigates this issue. A meaningful estimate of the pressure-wave velocity model is obtained in the zone sampled by both diving and reflected waves, down to almost two kilometers depth. To our knowledge, this is the first application of a graph space optimal transport-based full-waveform inversion to three-dimensional field data.

INTRODUCTION

Full waveform inversion (FWI) is a seismic imaging method that aims to reconstruct high-resolution models (up to half the shortest wavelength) of the mechanical properties of the subsurface (Devaney, 1984; Pratt and Shipp, 1999; Plessix and Perkins, 2010; Raknes et al., 2015; Górszczyk et al., 2017). The method is an iterative process based on minimizing a misfit function between observed and calculated data over a space of model parameters describing the subsurface. The improvement of resolution that FWI provides over standard tomography methods makes it possible to significantly improve depth-migration images or directly produce interpretable images of the subsurface physical properties. This method is used at multiple scales; from global and regional scales (Fichtner et al., 2010; Tape et al., 2010; Bozdağ et al., 2016) to seismic exploration targets for oil & gas industry (Plessix and Perkins, 2010; Stopin et al., 2014; Operto et al., 2015) and even near-surface scale (Bretaudeau et al., 2013; Groos et al., 2014; Schäfer et al., 2013; Irnaka et al., 2019). A thorough review of FWI and its applications is given in Virieux et al. (2017).

As powerful as this method is, it suffers from a significant shortcoming in its classical formulation: the non-convexity with respect to time-shifts of the least-squares (L^2) misfit function used to calculate the distance between observed and synthetic data. This non-convexity of the misfit function is an issue as the iterative process used in FWI is based on local-optimization algorithms. This leads to the so-called cycle-skipping issue. This limitation of FWI in its classical formulation has been documented since FWI has been introduced (Gauthier et al., 1986), and it has been of great interest to overcome it.

Numerous studies proposing different approaches have been published. One of the historical approaches to overcome this limitation in practical cases is to rely on a data hierarchy workflow. This approach consists in interpreting first the lowest frequency available, generally $2-4$ Hz for seismic

exploration targets, then progressively introducing higher frequency data following a multi-scale approach (Pratt, 1999; Bunks et al., 1995; Sirgue and Pratt, 2004). A second level of data-hierarchy can be defined by modifying the temporal and/or offset selection of the data used during inversion (Shipp and Singh, 2002; Wang and Rao, 2009; Brossier et al., 2009). The idea is to reduce the number of propagated wavelengths that are interpreted simultaneously. Current industrial applications generally rely on these two levels of data hierarchy, combined with a robust starting model, obtained, for instance, through reflection traveltime tomography or stereotomography (Lambaré, 2008).

Nonetheless, the conditions to apply this workflow are not always satisfied. For instance, low-frequency data around $2-4$ Hz are not always available or of sufficient quality. Moreover, obtaining low frequency increases the cost of the acquisition and can also compromise the quality of the high frequency needed to obtain high-resolution model reconstructions. Accurate initial model building can also be a time-consuming and challenging task requiring strong human expertise as it generally requires accurate traveltime and/or reflected event picking. Besides, prior information coming from geology and sonic logs are often needed. These constraints make FWI less robust and reduce its potential range in terms of applications.

Several methods have been introduced to improve robustness to cycle-skipping. The first group can be named as "extension strategies" and relies on introducing supplementary degrees of freedom to the FWI problem (Symes, 2008, 2015; Huang et al., 2018; van Leeuwen and Herrmann, 2013; Wang et al., 2016; Aghamiry et al., 2020), which can be used to artificially match the data at early iterations of the FWI process, avoiding cycle-skipping.

The second group consists in reformulating the FWI problem using an alternative measure of the distance between the observed and calculated data, namely a different misfit function. Numerous ap-

proaches have been proposed, such as cross-correlation (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010) and deconvolution based misfit function (Luo and Sava, 2011; Warner and Guasch, 2016), or by modifying the signal itself, making the L^2 norm between this new observable more convex with, for instance, instantaneous envelope (Fichtner et al., 2008; Bozdağ et al., 2011). We want to keep in mind that replacing the L^2 norm is not an easy task, as, despite its simplicity, the L^2 misfit presents excellent and interesting properties. First, it is robust to Gaussian noise. Second, it presents an excellent resolution power, translating into high-resolution reconstruction that FWI is well known for. Third, it is straightforward to implement and the computational cost of the misfit function evaluation is negligible compared to most of the proposed alternative misfit functions. These advantages have made the L^2 misfit the "state of the art" for FWI at exploration scales and could explain why L^2 is still widely used even if many alternative misfit functions have been proposed to mitigate the cycle-skipping issue.

Indeed, there is a discrepancy between the many propositions for alternative misfit functions compared to the number of actual field data applications. We think that this discrepancy could be explained by the - often not deeply discussed - intrinsic limitations of these alternative formulations. For instance, cross-correlation-based misfit functions (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010) might have difficulties handling complex data, when observed and synthetic traces contain multiple arrivals (not necessarily the same number) with different time shifts, some being in phase, other being out of phase. Deconvolution-based strategies (Luo and Sava, 2011; Warner and Guasch, 2016) require a penalization/weighting function, which can be difficult to set. Such settings are often case-dependent, making FWI less of an automated process. Instantaneous envelope intrinsically modifies the signal shape and discards information coming from the phase (which is essential to interpret the polarity of reflected events in the data correctly).

A new class of misfit functions based on optimal transport (OT) has been introduced recently.

The motivation is to benefit from the convexity of the optimal transport distance with respect to translation and dilation, which provides a misfit function convex with respect to time and amplitude shifts, a good proxy towards convexity with respect to velocities perturbations (Engquist and Froese, 2014; Métivier et al., 2018). Another important motivation to use OT as a misfit function is the ability to take into account the coherency of the seismic signal in an adequate space, be it a common shot or receivers gather. However, OT can only be applied to positive quantities and cannot be directly applied to seismic traces. To circumvent this difficulty, three main strategies have been developed.

The first one proposes to bring back the problem to the comparison of positive quantities by modifying the signal before solving the OT problem (Engquist and Froese, 2014; Qiu et al., 2017; Yang et al., 2018b; Yang and Engquist, 2018). A nonlinear transform is applied to the data in a trace-by-trace framework to transform each of them as probability measures. However, modifying the signal and altering the polarity information might be detrimental to a stable and satisfactory reconstruction of the subsurface mechanical properties.

The second one relies on the dual formulation of a specific instance of optimal transport distance, namely the 1-Wasserstein distance (Métivier et al., 2016a,b,c). This formulation can be naturally extended to the comparison of signed data. It benefits from its ability to be applied directly to 2D and 3D data, taking into account the coherency of the seismogram in the receiver and/or sources direction. However, even if the attraction valley to the global minimum is enlarged compared with the least-squares approach, the application of the 1-Wasserstein distance to signed data loses the convexity with respect to time-shift, which was the original motivation to use OT in the framework of FWI (see Métivier et al., 2018 for a review on different OT formulations). This strategy has been successfully applied to several field datasets (Poncet et al., 2018; Messud and Sedova, 2019; Sedova et al., 2019; Carotti et al., 2020; Hermant et al., 2020).

Finally, the third one considers each discrete seismic trace as point clouds and computes the optimal transport distance between point clouds associated with synthetic and observed traces. This method is called the graph space optimal transport (GSOT) and presents the main characteristic of preserving the convexity with respect to time shifts (Métivier et al., 2018, 2019). GSOT has already been successfully applied to 3D synthetic and field data (He et al., 2019; Pladys et al., 2019; Li et al., 2019; Górszczyk et al., 2019).

In the following table 1 we summarize the pros and cons of the three discussed strategies, following three criteria: computational efficiency, data distortion (nonlinear transforms to make each trace a probability distribution, which severely affect the data), and ability to be applied to multi-dimensional data.

125 [Table 1 about here.]

As GSOT is a promising candidate to tackle cycle-skipping on field datasets, this study focuses on applying the GSOT strategy against the classical L^2 misfit on a 3D OBC dataset from the North Sea, the Valhall field data. This dataset has been one of the first used to make proof of concept of the resolution power that FWI can bring on field data as shown in Sirgue et al. (2010). Since then, this dataset has been used several times for FWI application (Prieux et al., 2011; Gholami et al., 2013; Prieux et al., 2013; Operto and Miniussi, 2018; Kamath et al., 2021). This dataset can been seen as a "calibrated reference" for testing FWI formulations, such as frequency-domain FWI in Operto et al. (2015) or in time-domain with attenuation in Kamath et al. (2021). Here, we are using this dataset to compare the GSOT misfit function to the conventional L^2 norm through a time-domain 3D visco-acoustic VTI FWI.

To make this comparison, we first consider a canonical case where the initial model is the same as the one used in the aforementioned studies, which is derived from reflection tomography. This initial model ensures FWI converges toward a plausible estimation of the subsurface using conventional L^2 misfit function.

Then, we introduce a very "crude" initial model in which calculated data are shifted by more than one cycle compared to observed data, which is the typical case scenario for cycle-skipping. We show how L^2-based FWI fails from this "crude" starting model, whereas GSOT-based FWI manages to correctly interpret data to provide a plausible reconstruction of the subsurface (down to almost 2 km depth, which corresponds to the zone of the data sampled by both diving and reflected waves). This constitutes one of the first applications of the graph-space optimal transport misfit function to 3D field data at the exploration scale.

In the next section, we present the modeling and inversion algorithm used for our FWI application. Then we detail the global methodology of GSOT for FWI. This is followed by a presentation

of the Valhall field application, from the geological situation to the initial model and dataset presentation. We detail our FWI workflow and analyze the results in two cases: from the reflection traveltime tomography initial model and then from a 1D initial model. Results are then discussed, followed by conclusion and perspectives, which are given in a final Section.

METHODOLOGY

Modeling

This study is performed in the frame of 3D time-domain FWI. We rely on the anisotropic visco-acoustic time-domain modeling and inversion algorithm developed by Yang et al. (2018a), based on the following partial differential equations:

$$\begin{cases} \rho\partial_t v_x &= \partial_x g \\ \rho\partial_t v_y &= \partial_y g \\ \rho\partial_t v_z &= \partial_z q \\ \partial_t g &= c_{11}(\partial_x v_x + \partial_y v_y) + c_{13}\partial_z v_z - \sum_{\ell=1}^{L} Y_\ell[c_{11}\xi_\ell^g + c_{13}\xi_\ell^q] \\ \partial_t q &= c_{13}(\partial_x v_x + \partial_y v_y) + c_{33}\partial_z v_z - \sum_{\ell=1}^{L} Y_\ell[c_{13}\xi_\ell^g + c_{33}\xi_\ell^q] \\ \partial_t \xi_\ell^g &= -\omega_\ell \xi_\ell^g + \omega_\ell(\partial_x v_x + \partial_y v_y), \;\; \ell = 1, 2, \ldots, L \\ \partial_t \xi_\ell^q &= -\omega_\ell \xi_\ell^q + \omega_\ell \partial_z v_z, \;\; \ell = 1, 2, \ldots, L\,. \end{cases} \tag{1}$$

In system 1, c_{11}, c_{13}, and c_{33} are the stiffness tensor coefficients, ρ is the density, v_x, v_y, v_z are the horizontal and vertical displacement velocities respectively, while g and q are related to the normal stress components σ_{xx}, σ_{yy} and σ_{zz} through

$$\begin{aligned} g &= \sigma_{xx} = \sigma_{yy} \\ q &= \sigma_{zz}\,. \end{aligned} \tag{2}$$

This simplification is due to the VTI approximation. Similarly, the memory variables ξ_ℓ^q and ξ_ℓ^g are related to the memory variables ξ^{xx}, ξ^{yy} and ξ^{zz} associated with the normal stress components through

$$\begin{aligned} \xi_\ell^g &= \xi^{xx} + \xi^{yy} \\ \xi_\ell^q &= \xi^{zz} \,. \end{aligned} \tag{3}$$

These memory variables are used to model the viscosity of the medium following the generalized Maxwell body theory. Each represents one relaxation mechanism. We use three relaxation mechanisms to approximate a constant attenuation within the considered frequency band (L=3). The variables Y_ℓ are therefore calibrated depending on the target quality factor representing the attenuation in the considered media. This calibration is done through the solution of a least-squares problem. The details of this calibration can be found in Yang et al. (2016a) for instance.

In the VTI approximation, the stiffness tensor coefficients are related to the vertical P-wave velocity, the density, and the Thomsen anisotropy parameters ϵ and δ through

$$\begin{aligned} c_{11} &= \rho V_P^2 (1 + 2\epsilon) \\ c_{33} &= \rho V_P^2 \\ c_{13} &= \rho V_P^2 (1 + 2\delta) \,. \end{aligned} \tag{4}$$

The discretization of this system of partial differential equations is performed using a fourth-order in space and second-order in time staggered grid finite-difference method (Virieux, 1986; Levander, 1988).

A flat free surface condition is applied on top of the model to represent the water/air interface. Sponge layers (Cerjan et al., 1985) are applied on the other faces of the model to mimic a medium of infinite extensions in these directions. This numerical method is used instead of perfectly matched layers (PML, Bérenger, 1994) mainly to ensure stability, as anisotropy generates

instabilities (wavefield amplification) when PML are used. The combination of PML and attenuation through relaxation mechanisms is also not trivial in terms of implementation.

Finally, windowed sync interpolation is used to simulate source and receivers off-grid points accurately (Hicks, 2002).

Inversion

General formulation

FWI is an iterative process which relies on the minimization of a misfit function. Classically, it is the L^2 misfit function defined as follows

$$f_{L^2}[m] = \sum_s \sum_r h\big(d_{cal}[m](x_r,t;x_s), d_{obs}(x_r,t;x_s)\big)\,, \tag{5}$$

where

$$h(d_1(t), d_2(t)) = \frac{1}{2}\int_{t=0}^{T} |d_1(t) - d_2(t)|^2 \mathrm{d}t\,. \tag{6}$$

The observed and synthetic trace calculated in model m, associated with source x_s and receiver x_r, are denoted by $d_{obs}(x_r,t;x_s)$ and $d_{cal}[m](x_r,t;x_s)$ respectively.

To solve this minimization problem, we rely on a local optimization scheme. We use a preconditioned quasi-Newton l-BFGS algorithm (Nocedal, 1980), implemented in the SEISCOPE optimization toolbox (Métivier and Brossier, 2016).

We compute the gradient following the adjoint state method (Plessix, 2006). This method makes it possible to easily replace the L^2 norm as a change of the misfit function only translates to a modification of its associated adjoint-source to obtain the gradient. See Yang et al. (2018a) for a review. Since the method relies on reverse time propagation of the wavefield, which is numerically unstable with attenuation, our code relies on the checkpoint-assisted reverse forward simulation (CARFS)

strategy proposed by Yang et al. (2016b) to provide a stable and yet efficient implementation for large scale problems with attenuation.

The preconditioner we use is the wavefield preconditioner presented in Kamath et al. (2021) An anisotropic non-stationary Gaussian smoothing is also applied to the gradient. Inversions are performed for P-wave velocities (V_P) only.

Optimal transport for FWI

Optimal transport (OT) distances are derived from the OT theory introduced by the French mathematician Gaspard Monge more than two centuries ago (Monge, 1781). More precisely, they rely on the OT relaxation proposed by Kantorovich (1942). The distances, also called Wasserstein distances, have an intrinsic property of particular interest for the definition of inverse problems: they are convex with respect to translation and dilation of the compared quantities. This convexity with respect to translation has been, in particular, the motivation to introduce it in the framework of FWI to obtain a distance measurement convex with respect to time-shifts (Engquist and Froese, 2014). However, OT distances are defined for comparing probability distributions, which are by definition positive and normalized. So this new distance cannot be directly applied to seismic data, which is oscillatory (a generalization of OT to signed distribution is still an open question from a mathematical point of view, see Ambrosio et al. (2011) and Mainini (2012) for instance.

This study focuses on a recent proposition made to apply OT to seismic data: the graph space optimal transport (GSOT) strategy, proposed in Métivier et al. (2019).

This formulation of OT distance should preserve the signal unmodified while also preserving the convexity to shifted patterns. It relies on the idea of comparing the discrete graph of the data rather than the data itself: each 1D trace in time becomes a point cloud of Dirac delta functions

(of amplitude 1) in a 2D space made of the time dimension, and a new amplitude dimension. This transformation does not affect the signal shape but makes it possible to deal with positive mass (the Dirac delta functions). Thus the OT distance can be applied directly.

The corresponding misfit function is formulated as

$$f_{GSOT}[m] = \sum_s \sum_r h\big(d_{cal}[m](x_r, t; x_s), d_{obs}(x_r, t; x_s)\big)\,, \tag{7}$$

where this time

$$h(d_1, d_2) = \min_{\sigma \in S(N_t)} \sum_{i=1}^{N_t} c_{i\sigma(i)}(d_1, d_2)\,. \tag{8}$$

$S(N_t)$ denotes the ensemble of permutations of $\{1, \ldots, N_t\}$, and c_{ij} the L^2 distance between the discrete points of the graph $(t_i, d_1(t_i))$ and $(t_j, d_2(t_j))$:

$$c_{ij}(d_1, d_2) = |t_i - t_j|^2 + \psi^2 |d_1(t_i) - d_2(t_j)|^2\,. \tag{9}$$

The function h corresponds to the 2-Wasserstein distance between the discrete graph of the "calculated" trace $d_1(t)$ and the "observed" trace $d_2(t)$.

The scaling parameter ψ controls the convexity of the misfit function f_{GSOT} with respect to time shifts. In practice, we define it as

$$\psi = \frac{\tau}{A}\,, \tag{10}$$

where τ is a user-defined parameter corresponding to the maximum expected time shift between observed and calculated data in the initial model, and A is the maximum peak amplitude difference between observed and calculated data in the initial model. It is automatically computed prior to the inversion. This ensures the convexity of the GSOT distance for time up to approximately τ.

A visual illustration of the GSOT concept is given in Figure 1 which we reproduce from Métivier et al. (2019). In this Figure, the optimal permutation (assignment) σ between two Ricker functions, interpreted as point clouds, is plotted for different τ parameters. A larger τ value induces an optimal

assignment coupling points along the time axis, rendering the misfit function convex to time shifts. A small τ value induces an optimal assignment coupling points along the amplitude axis, rendering the misfit function equivalent to a least-squares misfit function.

[Figure 1 about here.]

The final cost function we use for the purpose of FWI application with N_s shots containing N_r receivers is defined as:

$$\min_m f_{GSOT}[m] = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} w^{s,r} h\left(d_{cal}^{s,r}[m], d_{obs}^{s,r}\right), \tag{11}$$

where $w^{s,r}$ is a trace-by-trace weighting factor, typically used to restore the AVO trend in the data. This trend is removed from the trace-by-trace GSOT approach, as the amplitude of each trace is treated separately through the normalization factor ψ. In practice, we compute $w^{s,r}$ as the L^2 energy of the corresponding observed trace

$$w^{s,r} = \left(\frac{1}{T}\int_0^T |d_{obs}^{s,r}(t)|^2 dt\right)^{1/2}. \tag{12}$$

The adjoint source of the misfit function $f_{GSOT}[m]$ is computed from $\frac{\partial h}{\partial_{cal}}$ using the adjoint-state strategy (Plessix, 2006). The following result is proved in Métivier et al. (2019). Denoting σ^* the minimizer in eq. 7, we have

$$\frac{\partial h}{\partial_{cal}} = 2\left(d_{cal} - d_{obs}^{\sigma^*}\right), \tag{13}$$

where

$$d_{obs}^{\sigma^*}(t_i) = d_{obs}(t_{\sigma^*(i)}). \tag{14}$$

The GSOT approach can thus be viewed as a generalization of the L^2 distance: The adjoint source is equal to the difference between calculated and observed data at time samples connected by the

optimal assignment σ^*. The solution of the problem eq. 7 provides the information to compute both the misfit function and the adjoint source.

To solve eq. 7 efficiently, we use the auction algorithm (Bertsekas and Castanon, 1989), dedicated to the solution of linear sum assignment problems such as eq. 7. Despite a relatively high computational complexity in $O(N_t^3)$, it is quite efficient for small instances of such problems. Resampling the data close to the Nyquist frequency (at the exploration scale, under acoustic approximation, traces are often around few hundreds of time steps after resampling) yields such small-scale problems making the GSOT feasible for realistic scale FWI applications, as seen in the application presented here.

FIELD DATA PRESENTATION AND FULL-WAVEFORM INVERSION WORKFLOW

Geological situation, dataset and initial models

Geological situation

The Valhall field is located in the southern part of the Norwegian sector in the North Sea, approximately 300 km southwest of Stavanger (Fig. 2). It is a shallow environment with a nearly constant water depth of 70 m. Valhall reservoir lies along the Lindesnes Ridge, which trends NNW (Munns, 1985; Leonard and Munns, 1987). The field has been discovered in 1975 and is used since then for oil production. It is characterized as an anticlinal in chalk in the Upper Cretaceous Hod and Tor formations, which form the reservoir at a depth of approximately 2400 m. Trapped gas in Tertiary shale is present above the reservoir (Sirgue et al., 2010; Prieux et al., 2011, 2013; Operto et al., 2015). The Tertiary overburden is relatively simple and free of com-

plex structure (Hall et al., 2002). Chalk compaction resulting from pressure depletion and water weakening has led to seabed subsidence (Field: VALHALL - Norwegianpetroleum.no `https://www.norskpetroleum.no/en/facts/field/valhall`).

[Figure 2 about here.]

Dataset presentation

The seismic data are 4-components acquired by ocean-bottom cables (OBC), with wide aperture/azimuth acquisition. The covered zone is a surface of 145 km^2. Twelve receiver cables are deployed on the seabed, containing 2048 receivers with an inline spacing of 50 m and a cable spacing of 300 m. A total of 50824 shots are available, located 5 m below the sea surface. The layout of this 3D acquisition is presented in Figure 3. In this study, we use only the pressure component of one acquisition performed in 2011 as part of the Valhall Life of Field Seismic (LoFS) project (Barkved et al., 2003).

[Figure 3 about here.]

From the raw data provided in SEGY format, only a simple pre-processing is applied. As our FWI code relies on source parallelization, source-receiver reciprocity is applied to process the hydrophone as explosive sources and the shots as hydrophones, hence sensibly reducing the impact on computer resources. The data is then de-spiked before a quality control over the complete dataset: the energy (RMS) of each gather is calculated to manually remove faulty gathers (the one with a large variation of RMS amplitude). The last step is to create frequency bands for the inversion using a minimum-phase band-pass filter in two distinct bands: $2.5 - 5$ Hz and $2.5 - 7$ Hz (referred to as band 1 and band 2 in the following). This goes in hand with time-decimation; from $\Delta t = 4$ ms to $\Delta t = 8$ ms on the first band, and from $\Delta t = 4$ ms to $\Delta t = 5$ ms on the second band (Fig. 4).

[Figure 4 about here.]

Initial models

An initial V_P model was made accessible to us thanks to AkerBP. It has been obtained through reflection traveltime tomography. It is referred to as TOMO initial and is presented in Figure 5. This initial model has been used in several publications using this Valhall dataset and has proven its capacity to give satisfactory results with L^2-based FWI as it predicts the arrival within half a period of the considered starting frequencies ($2.5 - 5$ Hz), avoiding cycle-skipping issue (Prieux et al., 2011; Operto et al., 2015; Operto and Miniussi, 2018). The associated density model is derived from V_P TOMO using Gardner's law (Gardner et al., 1974), defined as $\rho = 309.6 * V_P^{0.25}$. This relation is a fair average for brine-saturated rock (excluding evaporites), which is coherent with the expected geology of the Valhall field. AkerBP also provided us the anisotropy model ϵ and δ , and while their structure is not complicated, they are of significant influence in the modeling. The anisotropic parameter η define by Alkhalifah (1998) as

$$\eta = \frac{\epsilon - \delta}{1 - 2\delta} \tag{15}$$

is shown in Figure 7. We can see that maximum anisotropy reach values around 15% near the reservoir. Finally, to introduce attenuation, a simple two-layer Q_P model, with 1000 in the water column and 200 in the sediments, is used (Operto et al., 2015). While it is a simple model, it has been proven to be of great importance to explain the data. Using a more complex Q_P model (for example derived from V_P) only results in marginal improvements. Moreover, it introduces complexity and uncertainties that we prefer not to deal with (Kamath et al., 2021).

As the TOMO initial model is good enough to match the data within half a period of the $2.5 - 5$ Hz frequency-band used to start FWI (Fig. 8), it does not represent any challenge regarding the

cycle-skipping issue. This is why we introduce a new "crude" V_P starting model called 1D initial presented in Figure 6. It is a purely 1D vertical starting model, based on a linearly increasing profile with one main interface around 2400 m depth. This 1D starting model generates strong cycle-skipping, as made visible in Figure 9. As traditional L^2-based FWI cannot tackle this cycle-skipping in the data, it should be a good candidate to benchmark the capability of GSOT FWI. Associated with this V_P 1D model, the density model is derived using the preceding Gardner's law. The anisotropy model ϵ and δ , as the attenuation model Q_P are kept similar to those used in the TOMO setup.

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

Full-waveform inversion workflow

To obtain the final FWI reconstructed V_P model, we rely on several elements that compose our complete workflow. Each part of it plays a critical role in obtaining the best results possible, from wavelet estimation, data selection, to post-processing the FWI model. The generic workflow that we use in this article is presented in Figure 10. We detail the different parts composing the workflow in the following sections.

[Figure 10 about here.]

Source wavelet estimation

The first step before running FWI is to obtain a proper source wavelet that is used during the inversion to generate synthetic data. This is a crucial step as an incorrect wavelet estimation could induce artifacts into the reconstructed model.

The wavelet is estimated by solving a linear deconvolution problem in the frequency domain, following the methodology described in Pratt (1999). The wavelet inversion is performed on a single random subset of 240 shot-gathers. The sample of 240 wavelets (one per shot-gather) is then averaged to produce a single wavelet, assuming that all hydrophones have the same coupling response.

To minimize the cross-talk between the V_P model and the wavelet, we rely on a carefully designed data weighting strategy, focusing on short-offset only as presented in Figure 11. The data weighting consists in using 400 m of offset at full amplitude, then ramping down to zero at 1200 m offset. A tail mute is applied to remove the Schölte waves. Before being used for inversion, the wavelet is then manually checked and tapered to ensure its causality. With this methodology, the wavelet is only estimated at the beginning of each frequency band and kept fixed during FWI steps. The wavelet for the first frequency band and its associated spectrum is presented in Figure 11. We can see that no oscillations are present after 2.7 s, and wavelets generated from TOMO or 1D initial models are similar, which validates that this careful data selection mitigates the potential leakage in the wavelet estimation, which could come from the P-wave velocity model. Finally, the wavelet spectrum is coherent with the data (2.5 to 5 Hz).

[Figure 11 about here.]

Random shot subsampling and frequency continuation

The FWI workflow relies on a frequency continuation approach. Inversion is performed first on the frequencies ranging between $2.5 - 5$ Hz, then on a second band between $2.5 - 7$ Hz. Respectively, the model grid spacing is set to 70 m and 50 m in the three dimensions, ensuring at least five grid points for the smallest wavelength. These parameters are the same as in Operto et al. (2015) and Kamath et al. (2021).

As previously mentioned, the dataset contains 2048 shot-gathers. Since our FWI code relies on source-parallelization, the complete dataset could not be reasonably fit in a single inversion on the HPC facility we have access to. Therefore, we rely on a source-subsampling strategy, similar to the one described in Warner et al. (2013). This makes it possible to divide the dataset into batches of pseudo-randomly selected shot-gathers (120 in our case). The pseudo-random selection implies that the previously used shot-gather could not be selected in the next subsample until all of the available ones are used once. The subsample of shot-gathers is changed every time the memory limit for l-BFGS is reached (3 iterations is the maximum memory of l-BFGS in this study). The source-subsampling strategy makes it possible to fit the FWI problem on relatively small HPC facilities while mitigating the acquisition foot-print in the reconstructed model.

Hierarchical data weighting strategy

The first data weighting strategy tried on the dataset for FWI is simple: only remove the Schölte waves and muted the trace near-zero offset in a radius of 350 m (see the approach in Kamath et al. 2021). While this direct and straightforward approach (using all the data directly) could be justified with TOMO initial models as it is not supposed to generate cycle-skipping, this approach could not be applied to tackle a crude initial model as the 1D one, even with the GSOT misfit function.

Because of the strong cycle-skipping generated with the 1D initial model, we need to rely on a more careful data weighting strategy to maximize the capability of the GSOT. This leads to introduce a six-step data weighting strategy, presented in Figure 12. The three first steps only focus on diving waves with a strict time windowing while restricting to offset to the first 4 km (later referred to as **DW SO** for **D**iving **W**ave - **S**hort **O**ffsets), 8 km offset (later referred to as **DW MO** for **D**iving **W**ave - **M**edium **O**ffsets), and full offset (later referred to as **DW FO** for **D**iving **W**ave - **F**ull **O**ffsets). Then, the three next steps release the time windowing progressively, starting with 8 km offset (later referred to as **RT1 MO** for **R**elease **T**ime 1 - **M**edium **O**ffsets), then on full offset (later referred to as **RT1 FO** for **R**elease **T**ime 1 - **F**ull **O**ffsets); to finally finish with full offset and a complete release of time windowing (later referred to as **RT2 FO** for **R**elease **T**ime 2 - **F**ull **O**ffsets). Please also note that for each data weighting, the Schölte waves are additionally muted, as we cannot model them in the acoustic approximation.

Model post-processing

The last step of the FWI workflow is a post-processing applied to the updated V_P model at the end of each FWI step. It consists in cutting the model using a stencil (based on the shape of the acquisition) and extrapolating the value outside the stencil with a nearest-neighbors algorithm. This is performed for each depth slice. The stencil shaped is adapted for each depth slice to consider the maximum illumination the acquisition can provide at depth. This aims to remove the area on the edges of the model that are never updated during FWI due to the lack of illumination and remove artificial low-velocity zones created on the border of the well-illuminated zone.

To be consistent and perform a fair comparison between our two starting models TOMO and 1D , the complete workflow detailed before is applied to both starting models. By doing so, we validate

399 that our workflow can tackle the dataset properly and provide satisfactory results. Then, changing
400 only the misfit function from L^2 to GSOT, we reconduct the complete inversion process to check
401 how GSOT compares to L^2 in this controled environment.

402 [Figure 12 about here.]

FULL-WAVEFORM INVERSION RESULTS

403 Starting from the reflection tomography model

404 The first FWI results presented in this article are based on the TOMO initial model.

405 The final reconstructed V_P at 5 Hz is shown in Figure 13 using the L^2 misfit function, and in
406 Figure 14 using the GSOT misfit function. Respectively, the 7 Hz results are shown in Figure 15 and
407 Figure 16. Starting from the TOMO initial model, and as we do not expect cycle skipping, we keep
408 the τ parameter of GSOT to 0.2 s. This value is low enough to always allow for fast convergence
409 (similar to L^2) while preserving the potential improvement that GSOT can provide. Using a smaller
410 value of τ would make results similar to L^2 while making τ larger would degrade the capacity of
411 convergence of FWI due to a flatter attraction basin which is not needed in this case. The results are
412 almost identical between both formulations on this setup. Some minor differences can, however, be
413 observed. On the shallow depth slice (a) at 200 m, a small reduction of the acquisition imprint is
414 observable using the GSOT misfit function. On the depth slice (c) at 1 km, a slight improvement
415 of contrast between the low velocity anomaly and the sediment background is also visible. Finally,
416 on the vertical slices (d-g), more lateral coherency in the geological structures is visible using the
417 GSOT misfit function. The difference globally remains marginal, but this similarity between the
418 L^2 and GSOT misfit functions results is satisfying in itself. Indeed, alternative misfit functions

generally bring some drawbacks, such as loss of resolution power, which is not the case here. When comparing the V_P profile extracted from the reconstructed model to the sonic log filtered in the $0-7$ Hz frequency band (Fig. 17), we can see that the GSOT result is almost perfectly following L^2 result, which is consistent with the observation made directly on the V_P model. We can observe how FWI improves the fit to the sonic log over the initial model. This validates that our FWI workflow provides robust and reliable results.

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

[Figure 16 about here.]

[Figure 17 about here.]

The data fit presented in Figure 18 shows that GSOT can improve over the L^2 FWI data fit (which is already very satisfying). On the *rec A cable A* 2D CRG (through the low velocity anomaly), we can observe some data fit improvement with GSOT, for example, at -6.5 km offset and 7 s. On the *rec B cable B* 2D CRG, in an area with relatively mild variations in velocities and well away from the "gas cloud", we can observe that, globally, the data fit is almost the same for L^2 and GSOT, with all phases correctly explained. This is not surprising as this CRG focuses on the part of the model with sedimentary geology only, which the initial model better explains. However, we can still see the advantages of GSOT this time with more arrivals present for late time, as visible in the -7 to -4 km offset, from 5 to 7 s.

[Figure 18 about here.]

This "reference" result validates two essential aspects of this study. This first one is that our FWI workflow is adapted to the dataset and can provide satisfactory results when using L^2-based FWI when a good enough initial model is used as the TOMO model. This point is of first importance as it validates that our FWI workflow is consistent with the literature on this dataset. The second conclusion is that the GSOT misfit function can tackle field data problems without significant issues. Modifying the misfit function and keeping all other parameters similar, GSOT can provide comparable results to those provided by the L^2 FWI. We can even see slight improvement both in terms of the P-wave velocity model estimation and data fit with the GSOT misfit function. These encouraging results make it possible to push the analysis forward: can GSOT help tackle the cycle-skipping issue and improve FWI robustness compared to L^2-based FWI?

Starting from the 1D initial model

To validate the capability of GSOT to tackle large cycle-skipping, we use the 1D initial model that we introduced previously. This 1D initial model generates cycle-skipping, even on the mid-offset diving waves (mainly at -4 and -8 km offset), as clearly illustrated in Figure 9.

First, we compare the data fit obtained with L^2 and GSOT at the 2nd step of our FWI workflow as presented in Figure 19. For this first two steps of the FWI workflow, τ is set to 0.35 s, which is enough to handle the cycle-skipping generated in the data. First-order observation may indicate a similar data fit with GSOT and L^2 , but a more in-depth analysis shows several differences. At offset higher than 5 km, we observe a degradation with out-of-phase arrivals for L^2 on the *rec A cable A* CRG, while GSOT results present a significant improvement on this part. Then, late arrivals events are better explained with GSOT. On the second *rec B cable B* CRG, we observe a more continuous

reconstruction of the first events while also reducing out of phase one. Late arrivals are also better reconstructed with GSOT. On the second CRG, the data fit obtained with GSOT is already quite good for such an early stage of the inversion. In Figure 20 is presented the same CRG, but without applying the data weighting used at this early stage (DW MO data weighting), but instead the final relaxed data weighting (RT2 FO). Here, we can see that we are starting to predict data at larger offset when using GSOT compared to L^2 .

[Figure 19 about here.]

[Figure 20 about here.]

After the first two steps of FWI (DW SO and DW MO), we obtain the reconstructed model presented in Figures 21 and 22. It is clear here that L^2-based FWI result displays heavy artifacts on the reconstructed V_P model for depth larger than 300 m. Even if the very shallow part of the model is correctly reconstructed, the deeper part of the model is not (see shallow slice (a) compared with deeper slices (b) and (c)). Starting from a crude 1D initial model, L^2 reaches its limit and is likely affected by strong cycle-skipping. This is why we stop the L^2 inversion at this stage of our workflow: pushing forward the inversion by introducing more data does not help. Conversely, GSOT-based FWI can provide promising and meaningful V_P updates, with the recovery of correct background velocities at depth and even key features such as the definition of strong low velocity anomalies (slices (c) (d) (f)).

[Figure 21 about here.]

[Figure 22 about here.]

As GSOT results are encouraging after only two passes on the first frequency band, we apply our complete workflow and perform the complete inversion similar to our reference inversion, with six passes on 5 Hz data and 7 Hz data. Regarding the τ parameter for GSOT, as said earlier, the first two steps on the first frequency band (2.5 to 5 Hz) were performed using $\tau = 0.35$ s. The next four steps on the first band used a reduced $\tau = 0.2$ s as it is enough to tackle the shift present in the data (which already illustrates the improvement that GSOT achieves in the first step of the inversion). For the second frequency band (2.5 to 7 Hz), only the first step of the workflow uses a slightly relaxed τ of 0.25 s, while the remaining five steps use $\tau = 0.2$ s. Again, τ is relaxed at the begining to mitigate cycle-skipping, while a smaller τ is used afterward to preserve resolution power and speed up convergence.

We present the final data fit at 7 Hz for GSOT in Figure 23. We can see that a relatively good data fit is obtained, with most of the arrivals correctly explained. Still, some late arrivals are out of phase, mainly for large offsets (larger than 6 km)

[Figure 23 about here.]

We obtain the final reconstructed V_P presented in Figure 24. The results are promising and show a clear improvement in resolution compared to the early results at 5 Hz, with the main target structures retrieved above 2 km depth. The shape of the low velocity anomaly is correctly retrieved (slice (c) (d) (f)). Lateral resolution is very significantly improved, allowing the definition of narrow low-velocity (150 m wide) anomalies not resolved in the L^2 inversion (slice (e)).

To assess the quality of the final reconstructed V_P model, we can represent the value of the zero-lag cross-correlation between the observed data and synthetic data for different common-receiver gathers. When the data fit is converging toward a perfect match the (normalized) cross-correlation value converges to 1. We have performed this analysis for two common-receiver gathers (rec A and

rec B) in Figure 25. The improvement in data-fit is clear after FWI (Figure 25 (b) and (d)) compared to the data-fit obtained with the initial 1D model (Figure 25 (a) and (c)). This validates that final FWI results are not likely affected by remaining cycle-skipping.

Comparing the sonic log filtered in the $0-7$ Hz frequency band to the early L^2 and final GSOT results (Fig. 26) clearly illustrates that only GSOT-based FWI performs meaningful updates of the model that follows the trends of the sonic log 2 and 3. The case of sonic log 1 is interesting: as it is close to the target, it exhibits the down-shift of approximately 150 m observed earlier in the V_P model.

A comparison between the sonic log and GSOT-based FWI results for the two starting models is shown in Figure 27. For the two logs outside the center target area (Log 2 & 3), a good agreement of the FWI results is observed, with reconstructed V_P models following the same trend until 1.6 to 2 km. We want to remind that the illumination only constrains the model above 1.4 km depth when starting from the 1D initial model. In the Log 1, near the target, even when starting from the TOMO initial model, we can observe that the reconstruction is degrading under 1.6 km depth.

[Figure 24 about here.]

[Figure 25 about here.]

[Figure 26 about here.]

[Figure 27 about here.]

The results are still not perfect, and the main issue is the presence of a low-velocity update around 500 m depth, as made visible on slice (b). This low-velocity update (not present in the reference FWI results starting from TOMO initial model) introduces a down vertical shift of layer

under this perturbation. This is why slice (c) is extracted 100 m under (1.1 km instead of 1 km depth). This vertical shift does not affect the shape of the low velocity anomaly but only its depth. One possible way to avoid this artifact would be to modify the early stage of the inversion, for example, with different data-selections or modifications of the initial model for a slightly better one (for example, based on sonic log information). While this would have probably improved the results, we decided to keep a crude 1D initial model to stay as generic as possible and assess the capacity of GSOT in a setup without prior information.

Computational costs

Computational cost analysis is performed for one gradient estimation for both L^2 and GSOT misfit functions. Computation is performed on Haswell E5-2690V3@2.6 GHz Intel nodes containing 24 cores and 64 Gb of memory per node. We use 4 cores per source as our finite difference scheme uses an OpenMP parallelization.

The computational costs on the first frequency band ($2.5 - 5$ Hz) and second frequency band ($2.5 - 7$ Hz) are summerized in the table 2.

539 [Table 2 about here.]

The gradient column corresponds to the time spent to calculate the gradient (containing all required wave modelings), while the misfit column isolates the time spent computing the misfit function. The total time is the sum of gradient and misfit, and the ratio compares L^2 (put at 100%) with GSOT. This analysis shows that while the computational complexity of the solution of the gradient estimation scales to $O(\omega^4)$), the computation complexity of the GSOT computation is in $O(\omega^3)$, as noted in Métivier et al. (2019).

The overhead cost induced by GSOT is therefore reduced on the higher frequencies, which are the most expensive ones computationally speaking (for example, with a maximum frequency of 5 Hz, the first frequency band is relatively fast to compute, making a 20% overhead acceptable). This is one key feature that makes the GSOT misfit function able to tackle field data applications as higher frequency drastically increases the computational cost, and GSOT overhead will become smaller.

DISCUSSION

Facing challenging field data applications with alternative misfit functions is not widely documented in the literature. One of the only other alternative misfit functions that has been applied successfully to field data is adaptive waveform inversion (AWI) (Warner and Guasch, 2015; Ravaut et al., 2017; Debens et al., 2017; Roth et al., 2018; Guasch et al., 2019; Warner et al., 2019) or Kantorovich-Rubinstein optimal transport (KROT) (Poncet et al., 2018; Messud and Sedova, 2019; Sedova et al., 2019).

Regarding the final results obtained with the GSOT misfit function, we first validate that it can improve over L^2-based FWI in a controled environment (starting from a good enough V_P TOMO) with improved datafit and improved lateral coherency and reduced acquisition imprint on the model

side. The computational overhead induced by GSOT stays limited thanks to the computational complexity of GSOT being one order of magnitudes smaller than the complexity of the gradient estimation for a given maximum frequency. This behavior translates into a good scaling property when facing high-frequency data, which is now one trend for field data FWI applications. When tackling a difficult 1D initial model, which generates cycle-skipping, we show that GSOT-based FWI provides good model updates and good datafit where classical L^2-based FWI fails. The control of the convexity provided by the GSOT misfit function is performed through the choice of the τ parameter. This parameter is directly linked to the observed time-shift between calculated and observed data, making it easy to tune, depending on the initial fit of the data and the expected maximum time shift. We adapt it from 0.2 s to 0.35 s in the initial stage of the workflow when we switch from the initial tomography model to the initial 1D model. We emphasize the fact that the τ parameter is simple to define and small modification (± 0.2 s) does not translate into drasticaly different results.

We also emphasize that results under 2 km suffer from a substantial lack of illumination and are therefore limited to the resolution of our initial model, explaining why no meaningful updates are present for depth superior to 2 km. This limitation in terms of depth reconstruction (under 2 km) is not surprising as the FWI alone is not expected to present enough illumination. In this case, it would require the use of reflected wave inversion (RWI) or joint full-waveform inversion (JFWI). Some preliminary but encouraging results were obtained combining GSOT and JFWI (Provenzano et al., 2020): GSOT adds the convexity necessary to predict the reflected data, enabling robust velocity updates of the model at depth.

CONCLUSION

In this work, we focus on the application of a new misfit function: graph space optimal transport. This formulation shows a clear improvement over L^2 in a controled environment, unleashing the potential of FWI to perform meaningful updates when starting from a crude, 1D initial model. This clear improvement in cycle-skipping robustness, combined with a simple setup (only one *physical* parameter to define) and a reasonable computational overhead, illustrates that GSOT is a good candidate to improve FWI robustness and therefore make FWI more accessible and easily applicable.

ACKNOWLEDGMENTS

This study was partially funded by the SEISCOPE consortium (`http://seiscope2.osug.fr`), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, SHELL, SINOPEC, SISPROBE and TOTAL. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (`https://ciment.ujf-grenoble.fr`), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche, and the HPC resources of CINES/IDRIS/TGCC under the allocation 046091 made by GENCI.

We thank AKERBP ASA and their partner Pandion Energy for providing the dataset and permission to present this work and the help of Ross Milne from AKERBP.

REFERENCES

Aghamiry, H., A. Gholami, and S. Operto, 2020, Accurate and efficient wavefield reconstruction in the time domain: Geophysics, **85(2)**, A7–A12.

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

Ambrosio, L., E. Mainini, and S. Serfaty, 2011, Gradient flow of the Chapman Rubinstein Schatzman model for signed vortices: Annales de l'Institut Henri Poincaré (C) Non Linear Analysis, **28**, 217–246.

Barkved, O., A. Bærheim, D. Howe, J. Kommedal, and G. Nicol, 2003, Life of Field Seismic Implementation - Another "first at valhal": Presented at the 65^{th} EAGE Workshop, Stavanger.

Bérenger, J.-P., 1994, A perfectly matched layer for absorption of electromagnetic waves: Journal of Computational Physics, **114**, 185–200.

Bertsekas, D. P., and D. Castanon, 1989, The auction algorithm for the transportation problem: Annals of Operations Research, **20**, 67–96.

Bozdağ, E., D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire, 2016, Global adjoint tomography: first-generation model: Geophysical Journal International, **207**, 1739–1766.

Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

Bretaudeau, F., R. Brossier, D. Leparoux, O. Abraham, and J. Virieux, 2013, 2D elastic full waveform imaging of the near surface: Application to synthetic and a physical modelling data sets: Near Surface Geophysics, **11**, 307–316.

Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore structures by 2D

elastic frequency-domain full-waveform inversion: Geophysics, **74**, WCC105–WCC118.

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

Carotti, D., O. Hermant, S. Masclet, M. Reinier, J. Messud, A. Sedova, and G. Lambaré, 2020, Optimal transport full waveform inversion - applications: Presented at the 82^{th} Annual EAGE Meeting (Amsterdam), European Association of Geoscientists & Engineers.

Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef, 1985, A nonreflecting boundary condition for discrete acoustic and elastic wave equations: Geophysics, **50**, 2117–2131.

Debens, H. A., F. Mancini, M. Warner, and L. Guasch, 2017, Full-bandwidth adaptive waveform inversion at the reservoir: SEG Technical Program Expanded Abstracts 2017, 1378–1382.

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

Engquist, B., and B. D. Froese, 2014, Application of the Wasserstein metric to seismic signals: Communications in Mathematical Science, **12**, 979–988.

Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2008, Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain: Geophysical Journal International, **175**, 665–685.

———, 2010, Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle: Earth and Planetary Science Lettters, **290**, 270–280.

Gardner, G. F., L. Gardner, and A. Gregory, 1974, Formation velocity and density—the diagnostic basics for stratigraphic traps: Geophysics, **39**, 770–780.

Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: numerical results: Geophysics, **51**, 1387–1403.

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., L. Métivier, and R. Brossier, 2019, Mitigating the nonlinearity of the crustal scale full waveform inversion through the graph space optimal transport misfit function: AGU Fall Meeting Abstracts, S41A–03.

Górszczyk, A., S. Operto, and M. Malinowski, 2017, Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern nankai trough revisited: Journal of Geophysical Research: Solid Earth, **122**, 4601–4630.

Groos, L., M. Schäfer, T. Forbriger, and T. Bohlen, 2014, The role of attenuation in 2D full-waveform inversion of shallow-seismic body and Rayleigh waves: Geophysics, **79**, R247–R261.

Guasch, L., M. Warner, and C. Ravaut, 2019, Adaptive waveform inversion: Practice: Geophysics, **84(3)**, R447–R461.

Hall, S., J. Kendall, and O. Barkved, 2002, Fractured reservoir characterization using P-wave AVOA analysis of 3D OBC data: The Leading Edge, 777–781.

He, W., R. Brossier, and L. Métivier, 2019, 3D elastic FWI for land seismic data: A graph space OT approach: SEG Technical Program Expanded Abstracts 2019, 1320–1324.

Hermant, O., A. Aziz, S. Warzocha, and M. Al Jahdhami, 2020, Imaging complex fault structures on-shore oman using optimal transport full waveform inversion: **2020**, 1–5.

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

Huang, G., R. Nammour, and W. W. Symes, 2018, Source-independent extended waveform inversion based on space-time source extension: Frequency-domain implementation: Geophysics, **83**, R449–R461.

669 Irnaka, T., R. Brossier, L. Métivier, T. Bohlen, and Y. Pan, 2019, Towards 3d 9c elastic full waveform
670 inversion of shallow seismic wavefields - case study ettlingen line: 81^{th} Annual EAGE Confer-
671 ence & Exhibition, London, Expanded Abstracts, 81^{th} Annual EAGE Conference & Exhibition,
672 London, EAGE, We P01 04.
673 Kamath, N., R. Brossier, L. Métivier, A. Pladys, and P. Yang, 2021, Multiparameter full-waveform
674 inversion of 3D ocean-bottom cable data from the Valhall field: Geophysics, **86**, B15–B35.
675 Kantorovich, L., 1942, On the transfer of masses: Dokl. Acad. Nauk. USSR, **37**, 7–8.
676 Lambaré, G., 2008, Stereotomography: Geophysics, **73(5)**, VE25–VE34.
677 Leonard, R., and J. Munns, 1987, Valhall field in geology of norwegian oil and gas fields: Graham
678 and Trotman.
679 Levander, A. R., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–
680 1436.
681 Li, Y., R. Brossier, and L. Métivier, 2019, Joint FWI for imaging deep structures: A graph-space
682 OT approach: SEG Technical Program Expanded Abstracts 2019, 1290–1294.
683 Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion:
684 SEG Technical Program Expanded Abstracts, **30**, 2788–2792.
685 Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.
686 Mainini, E., 2012, A description of transport cost for signed measures: Journal of Mathematical
687 Sciences, **181**, 837–855.
688 Messud, J., and A. Sedova, 2019, Multidimensional optimal transport for 3d FWI: Demonstration
689 on field data: Presented at the Expanded Abstracts, 81^{th} Annual EAGE Meeting (London).
690 Métivier, L., A. Allain, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2018, Optimal transport
691 for mitigating cycle skipping in full waveform inversion: a graph space transform approach:
692 Geophysics, **83**, R515–R540.

693 Métivier, L., and R. Brossier, 2016, The seiscope optimization toolbox: A large-scale nonlinear
694 optimization library based on reverse communication: Geophysics, **81**, F11–F25.
695 Métivier, L., R. Brossier, Q. Mérigot, and E. Oudet, 2019, A graph space optimal transport distance
696 as a generalization of L^p distances: application to a seismic imaging inverse problem: Inverse
697 Problems, **35**, 085001.
698 Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016a, Increasing the robustness
699 and applicability of full waveform inversion: an optimal transport distance strategy: The Leading
700 Edge, **35**, 1060–1067.
701 ———, 2016b, Measuring the misfit between seismograms using an optimal transport distance: Ap-
702 plication to full waveform inversion: Geophysical Journal International, **205**, 345–377.
703 ———, 2016c, An optimal transport approach for seismic tomography: Application to 3D full wave-
704 form inversion: Inverse Problems, **32**, 115008.
705 Monge, G., 1781, Mémoire sur la théorie des déblais et des remblais: Histoire de l'Académie Royale
706 des Sciences de Paris.
707 Munns, J. W., 1985, The Valhall field: a geological overview: Marine and Petroleum Geology, **2**,
708 23–43.
709 Nocedal, J., 1980, Updating Quasi-Newton Matrices With Limited Storage: Mathematics of Com-
710 putation, **35**, 773–782.
711 Operto, S., and A. Miniussi, 2018, On the role of density and attenuation in 3D multi-parameter
712 visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea: Geophysical
713 Journal International, **213**, 2037–2059.
714 Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and
715 J. Virieux, 2015, Efficient 3-D frequency-domain mono-parameter full-waveform inversion of
716 ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic

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

718 Pladys, A., R. Brossier, M. Irnaka, N. Kamath, and L. Métivier, 2019, Assessment of optimal trans-
719 port based FWI: 3d OBC valhall case study: SEG Technical Program Expanded Abstracts 2019,
720 1295–1299.

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

723 Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom seis-
724 mometer dataset: First Break, **28**, 71–78.

725 Poncet, R., J. Messud, M. Bader, G. Lambaré, G. Viguier, and C. Hidalgo, 2018, Fwi with optimal
726 transport: a 3D implementation and an application on a field dataset: Presented at the Expanded
727 Abstracts, 80^{th} Annual EAGE Meeting (Copenhagen).

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

730 Pratt, R. G., and R. M. Shipp, 1999, Seismic waveform inversion in the frequency domain, part II:
731 Fault delineation in sediments using crosshole data: Geophysics, **64**, 902–914.

732 Prieux, V., R. Brossier, Y. Gholami, S. Operto, J. Virieux, O. Barkved, and J. Kommedal, 2011,
733 On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study: Geo-
734 physical Journal International, **187**, 1495–1515.

735 Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013, Multiparameter full waveform inversion of
736 multicomponent OBC data from Valhall. Part 1: imaging compressional wavespeed, density and
737 attenuation: Geophysical Journal International, **194**, 1640–1664.

738 Provenzano, G., R. Brossier, L. M´etivier, and Y. Li, 2020, *in* Joint FWI of diving and reflected
739 waves using a graph space optimal transport distance: Synthetic tests on limited-offset surface
740 seismic data: 780–784.

741 Qiu, L., J. Ramos-Martìnez, A. Valenciano, Y. Yang, and B. Engquist, 2017, Full-waveform inver-
742 sion with an exponentially encoded optimal-transport norm: SEG Technical Program Expanded
743 Abstracts 2017, 1286–1290.
744 Raknes, E. B., B. Arntsen, and W. Weibull, 2015, Three-dimensional elastic full waveform inversion
745 using seismic data from the sleipner area: Geophysical Jounal International, **202**, 1877–1894.
746 Ravaut, C., F. Maao, J. Mispel, A. Osen, M. Warner, L. Guasch, and T. Nangoo, 2017, Imaging
747 beneath a gas cloud in the north sea without conventional tomography: EAGE, 79th Conference
748 and Exhibition, Expanded abstracts, We A3 04.
749 Roth, T., T. Nangoo, N. Shah, M. Riede, C. Henke, and M. Warner, 2018, Improving seismic image
750 with high resolution velocity model from awi starting with 1d initial model - case study barents
751 sea.
752 Schäfer, M., L. Groos, T. Forbriger, and T. Bohlen, 2013, 2D full waveform inversion of recorded
753 shallow seismic Rayleigh waves on a significantly 2D structure: Presented at the Proceedings
754 of 19th European Meeting of Environmental and Engineering Geophysics, Expanded Abstracts,
755 Bochum, Germany.
756 Sedova, A., J. Messud, H. Prigent, G. Royle, and G. Lambaré, 2019, Acoustic land full waveform
757 inversion on a broadband land dataset: the impact of optimal transport: Presented at the Expanded
758 Abstracts, 81^{th} Annual EAGE Meeting (London).
759 Shipp, R. M., and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture
760 marine seismic streamer data: Geophysical Journal International, **151**, 325–344.
761 Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full
762 waveform inversion: the next leap forward in imaging at Valhall: First Break, **28**, 65–70.
763 Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging : a strategy for selecting
764 temporal frequencies: Geophysics, **69**, 231–248.

765 Stopin, A., R.-E. Plessix, and S. Al Abri, 2014, Multiparameter waveform inversion of a large
766 wide-azimuth low-frequency land data set in Oman: Geophysics, **79**, WA69–WA77.
767 Symes, W., 2015, Algorithmic aspects of extended waveform inversion: Presented at the 77th EAGE
768 Conference and Exhibition 2017-Workshops.
769 Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospect-
770 ing, **56**, 765–790.
771 Tape, C., Q. Liu, A. Maggi, and J. Tromp, 2010, Seismic tomography of the southern California
772 crust based on spectral-element and adjoint methods: Geophysical Journal International, **180**,
773 433–462.
774 Thurin, J., 2020, Uncertainties estimation in full waveform inversion using ensemble methods: The-
775 ses, Université Grenoble Alpes.
776 van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by
777 expanding the search space: Geophysical Journal International, **195(1)**, 661–667.
778 van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation
779 traveltime tomography: Geophysical Journal International, **182**, 1383–1394.
780 Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite difference
781 method: Geophysics, **51**, 889–901.
782 Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An intro-
783 duction to Full Waveform Inversion, *in* Encyclopedia of Exploration Geophysics: Society of
784 Exploration Geophysics, R1–1–R1–40.
785 Wang, C., D. Yingst, P. Farmer, and J. Leveille, 2016, Full-waveform inversion with the recon-
786 structed wavefield method, *in* SEG Technical Program Expanded Abstracts 2016: 1237–1241.
787 Wang, Y., and Y. Rao, 2009, Reflection seismic waveform tomography: Journal of Geophysical
788 Research, **114**, 1978–2012.

789 Warner, M., and L. Guasch, 2015, Robust adaptive waveform inversion: SEG Technical Program
790 Expanded Abstracts 2015, 1059–1063.
791 ———, 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, R429–R445.
792 Warner, M., T. Nangoo, A. Pavlov, and C. Hidalgo, 2019, Extending the velocity resolution of
793 waveform inversion below the diving waves using awi: **2019**, 1–5.
794 Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch,
795 C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: Geophysics,
796 **78**, R59–R80.
797 Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016a, A review on the systematic formulation of
798 3D multiparameter full waveform inversion in viscoelastic medium: Geophysical Journal Inter-
799 national, **207**, 129–149.
800 ———, 2016b, Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-
801 forward simulation method: Geophysics, **81**, R349–R362.
802 Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018a, A Time-Domain Preconditioned
803 Truncated Newton Approach to Multiparameter Visco-acoustic Full Waveform Inversion: SIAM
804 Journal on Scientific Computing, **40**, B1101–B1130.
805 Yang, Y., and B. Engquist, 2018, Analysis of optimal transport and related misfit functions in full-
806 waveform inversion: GEOPHYSICS, **83**, A7–A12.
807 Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018b, Application of optimal transport and the
808 quadratic Wasserstein metric to full-waveform inversion: Geophysics, **83**, R43–R62.

LIST OF FIGURES

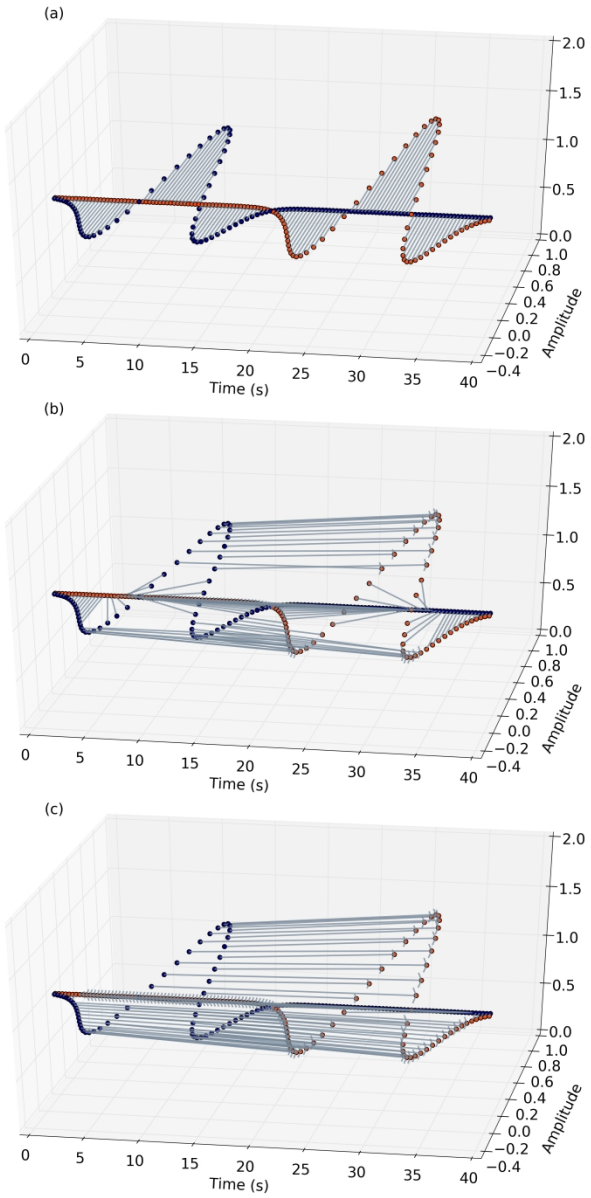

Figure 1. 3D representation of the discrete graph of a reference Ricker function (red points) and a shifted in time Ricker function (blue points) scaled in amplitude by a factor 0.8. The gray arrows represent the assignment solution of the LSAP problem, which depends on the value of the parameter $ au$. Top $ au=0.4$ s, middle $ au=4$ s, bottom $ au=20$ s.

1259x2404mm (72 x 72 DPI)

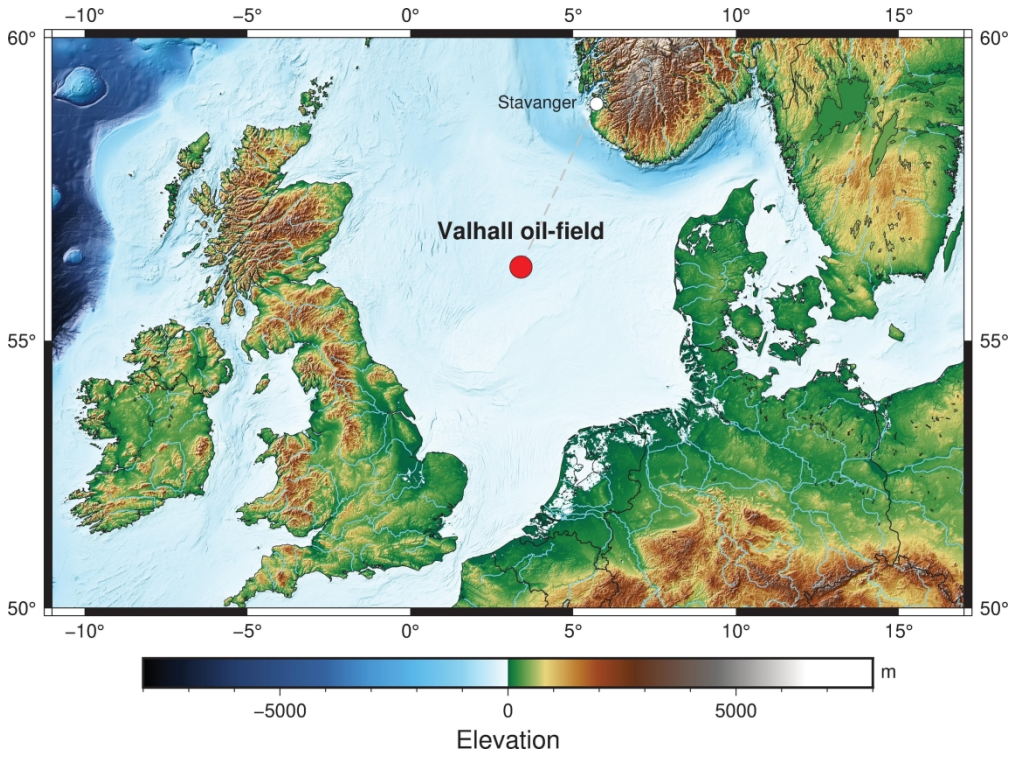

Figure 2: Location of the Valhall field on the North Sea (from \citealp{Thurin_2020_PhD})

222x165mm (300 x 300 DPI)

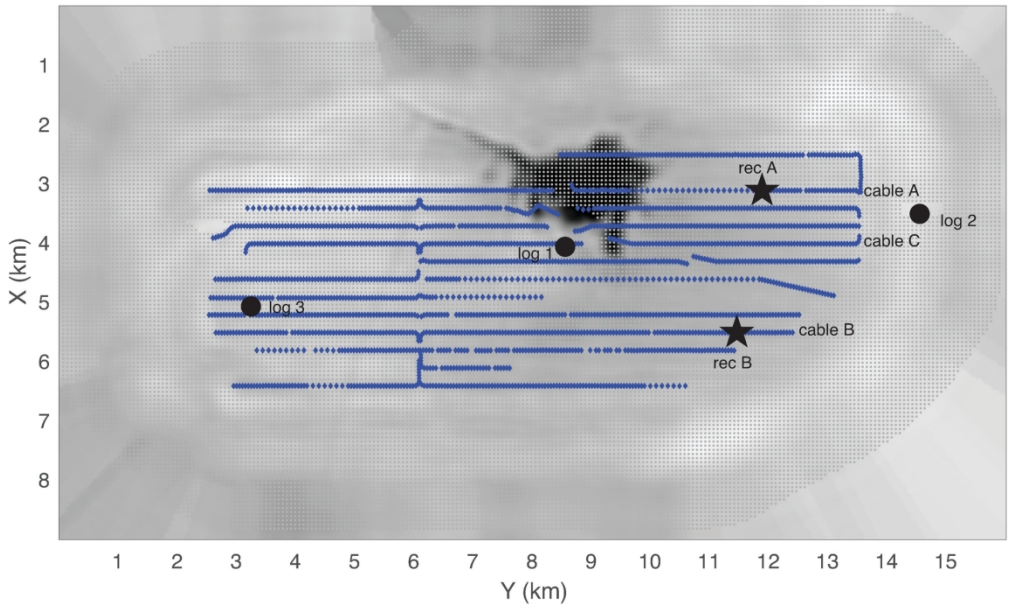

Figure 3: Layout of the Valhall acquisition overlapped on an horizontal \vp slice at 1~km obtained by GSOT-based FWI (from this study). Location of sources (gray dots) and receivers (blue diamonds). Two receivers positions (A and B) are located with black stars. Cables A ($x=2950$~m), B ($x=5530$~m) and C ($x=3080$~m) are identifed. Black dots denote the position of the three \vp sonic-logs.

168x101mm (300 x 300 DPI)

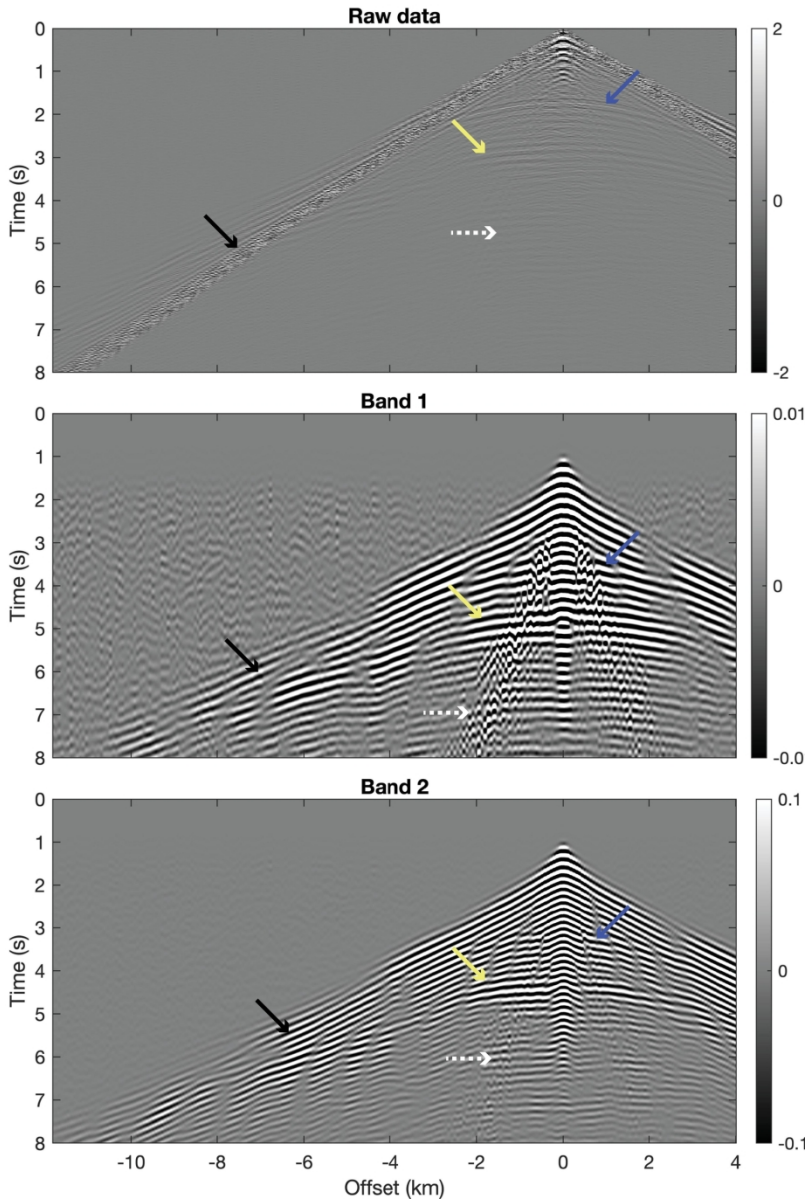

Figure 4: 2D common-receiver gathers extracted for receiver A along cable A for: raw data (top), band 1 data $2.5-5$~Hz (middle) and band 2 data $2.5-7$~Hz (bottom). White dashed arrows point on the Sch\"{o}lte waves which are muted for the inversion. Blue and yellow arrows respectively point on the reflexion from the low velocity zone and the reflexion on the top of the reservoir. Black arrows point on the diving waves.

149x224mm (300 x 300 DPI)

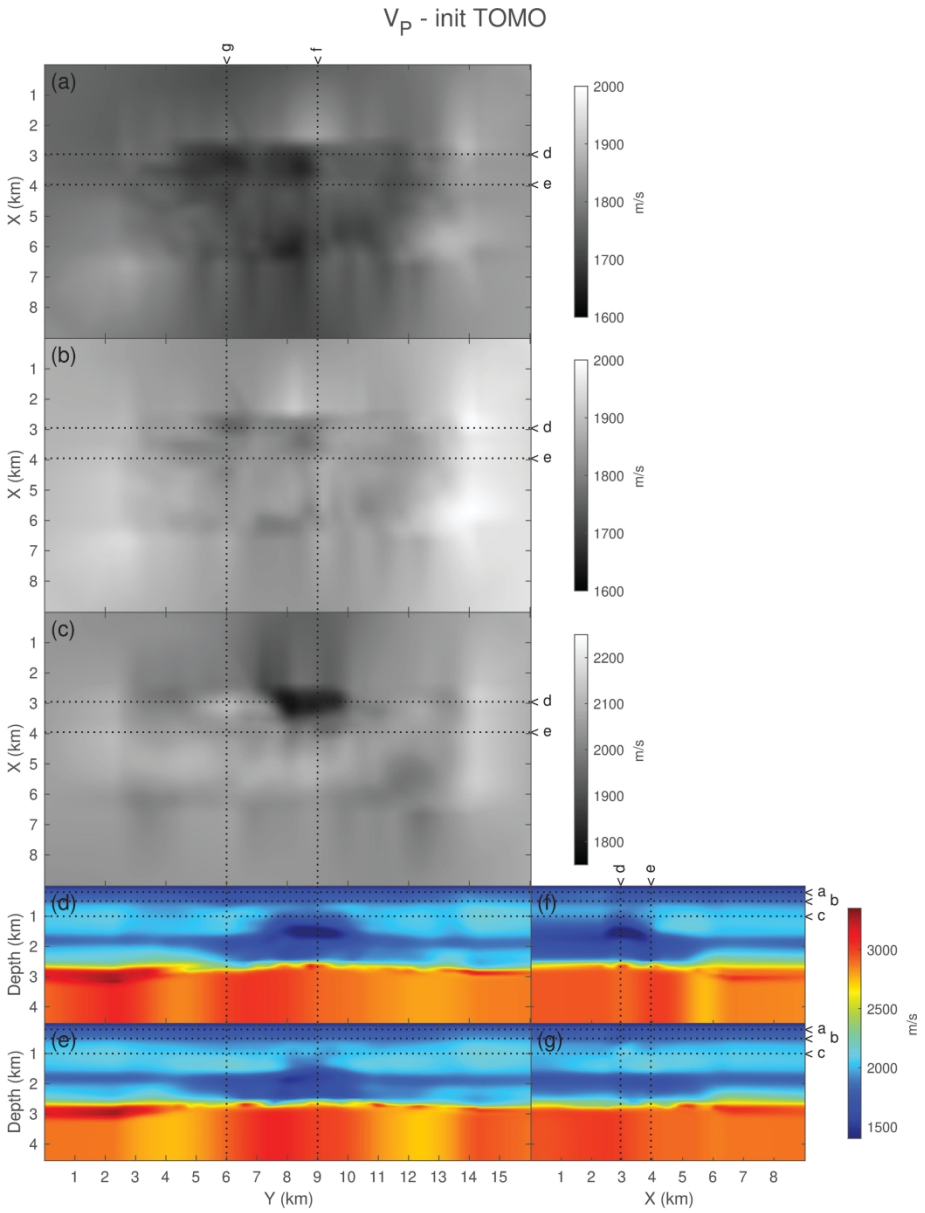

Figure 5: Slices of the initial model \vpa. (a-c) Horizontal slices at (a) 200~m depth, (b) 500~m depth and (c) 1~km depth. (d-e) Inline vertical slices for (d) $x=2.95$~km and (e) $x=3.95$~km. (f-g) Cross-line vertical slices at (f) $y=9$~km and (g) $y=6$~km. Note that the depth slices used a grayscale colormap with two different velocity ranges, while the vertical slices used a ``jet'' type colormap with a fixed velocity range for all slices.

209x276mm (300 x 300 DPI)

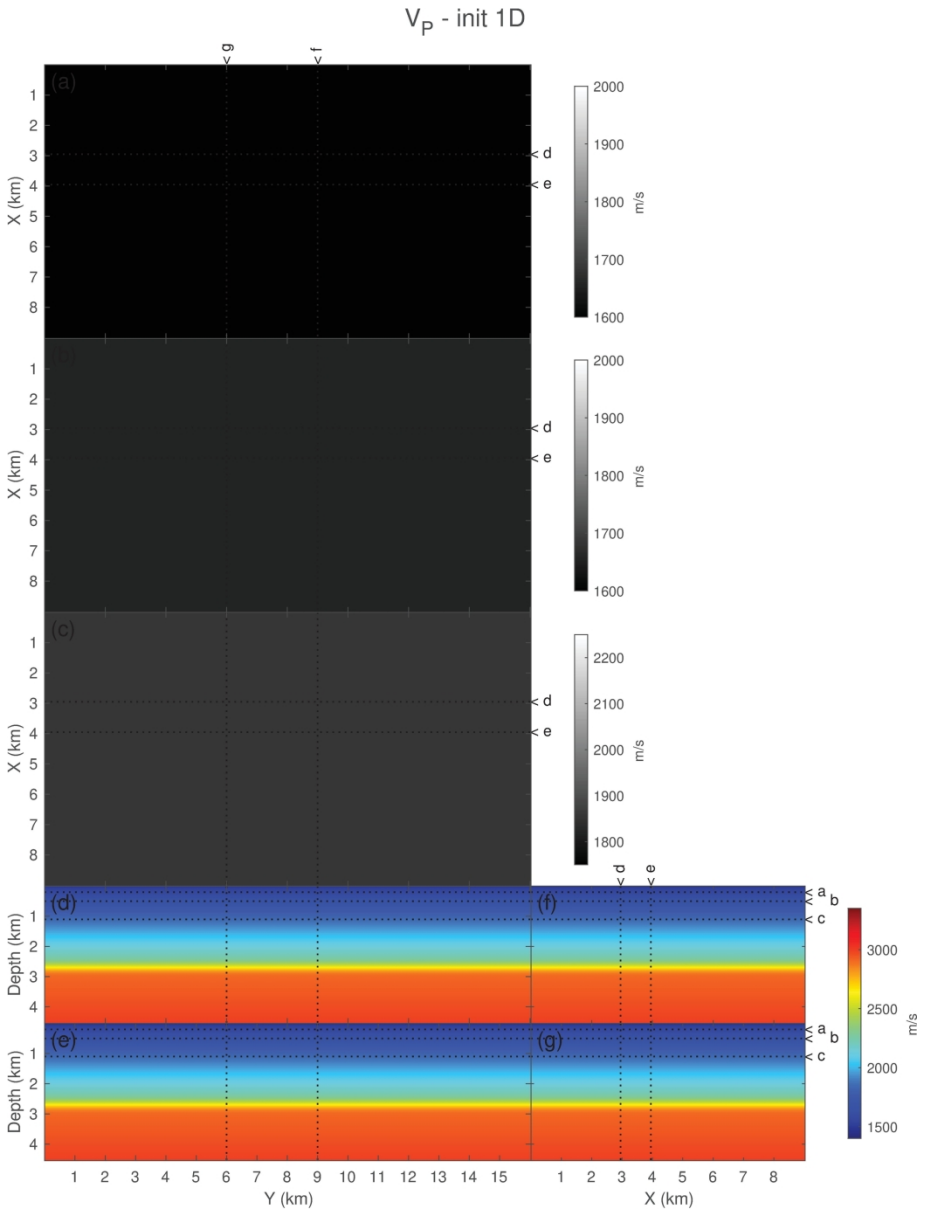

Figure 6: Same as \Cref{init:VP_initTOMO} for initial model \vpb.

209x276mm (300 x 300 DPI)

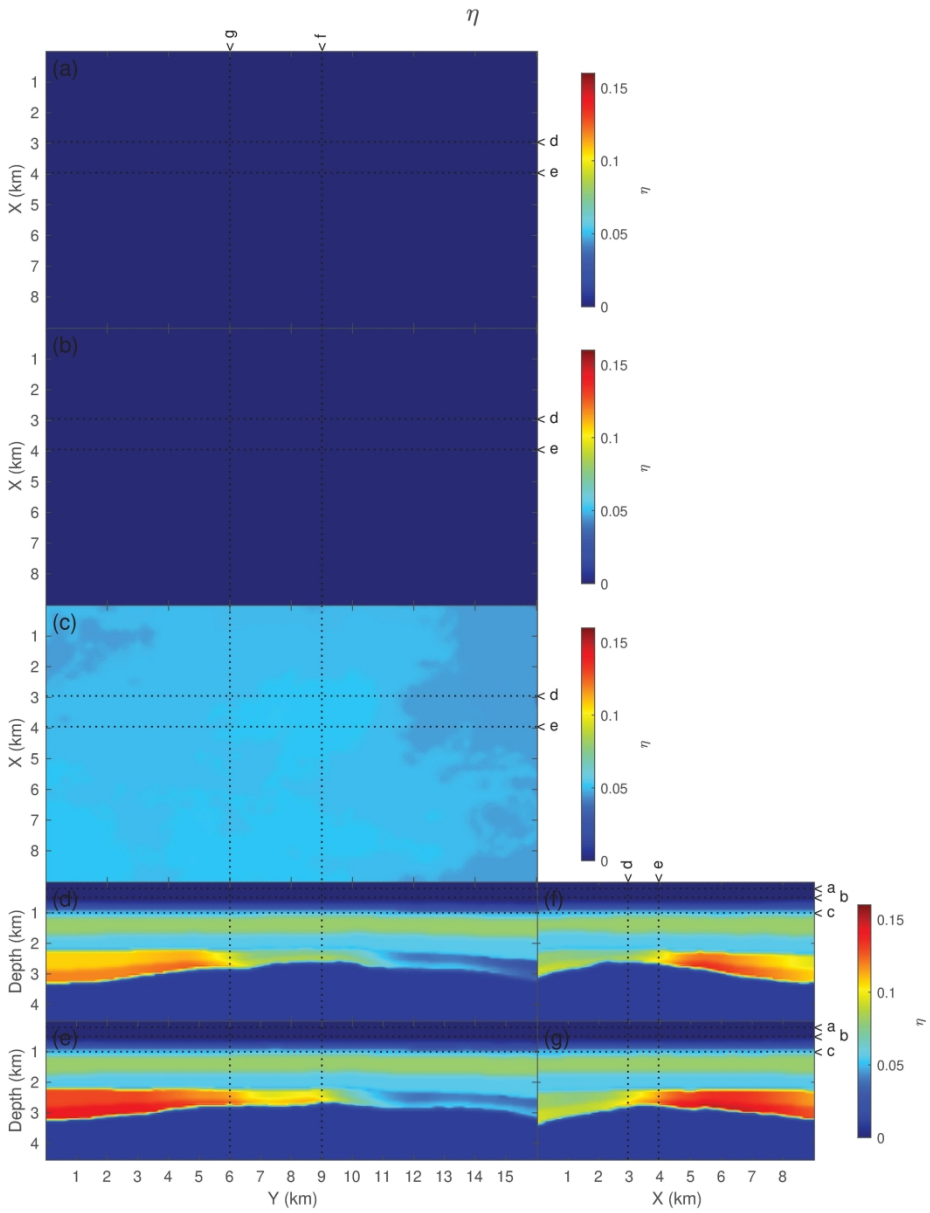

Figure 7: Same as \Cref{init:VP_initTOMO} for anisotropic parameter η.

209x273mm (300 x 300 DPI)

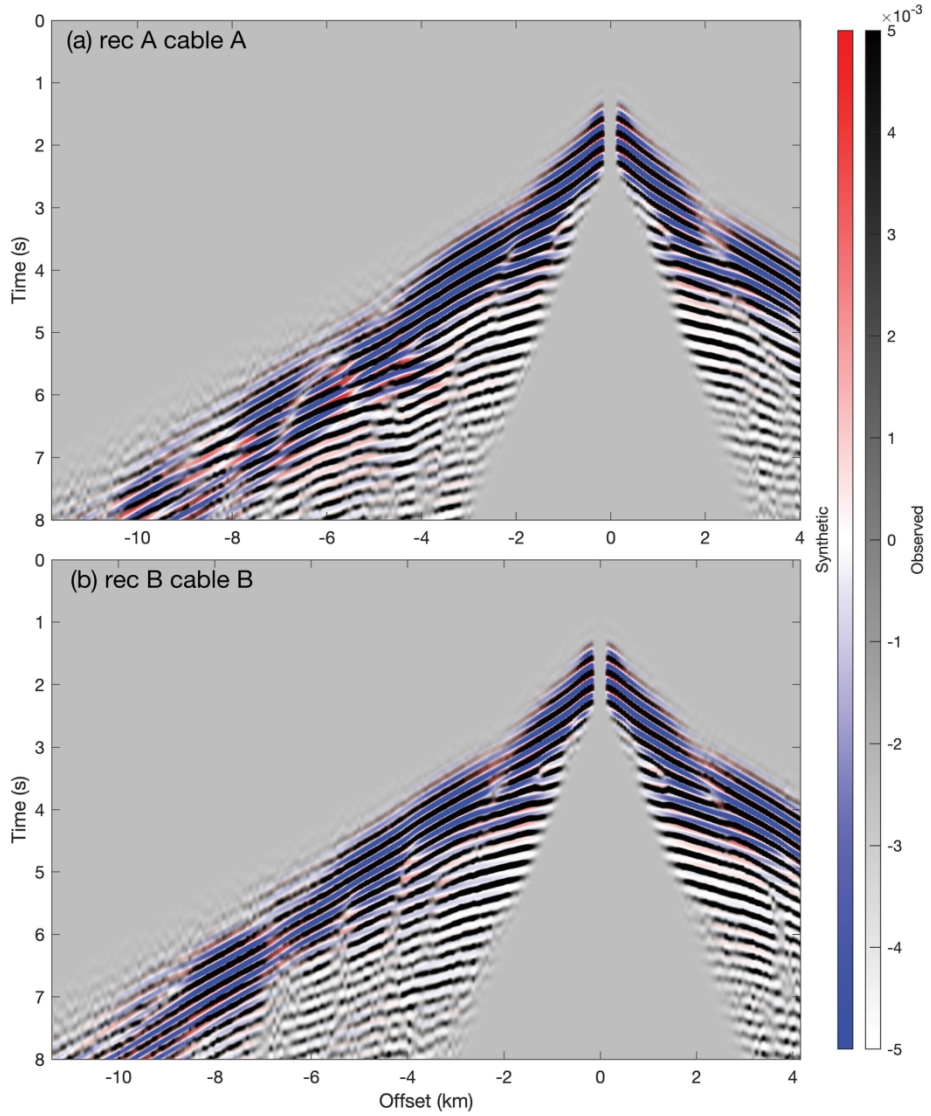

Figure 8: 2D common-receiver gathers at 5~Hz. Synthetic data generated into the initial model \moda are displayed in a blue/white/red color scale, field data are overlapped in grayscale with transparency. The best result is achieved when black and blue are the only colors visible. Red and white are shown when data are not in phase. (a) receiver A along cable A (through the low velocity anomaly). (b) receiver B along cable B.

200x229mm (300 x 300 DPI)

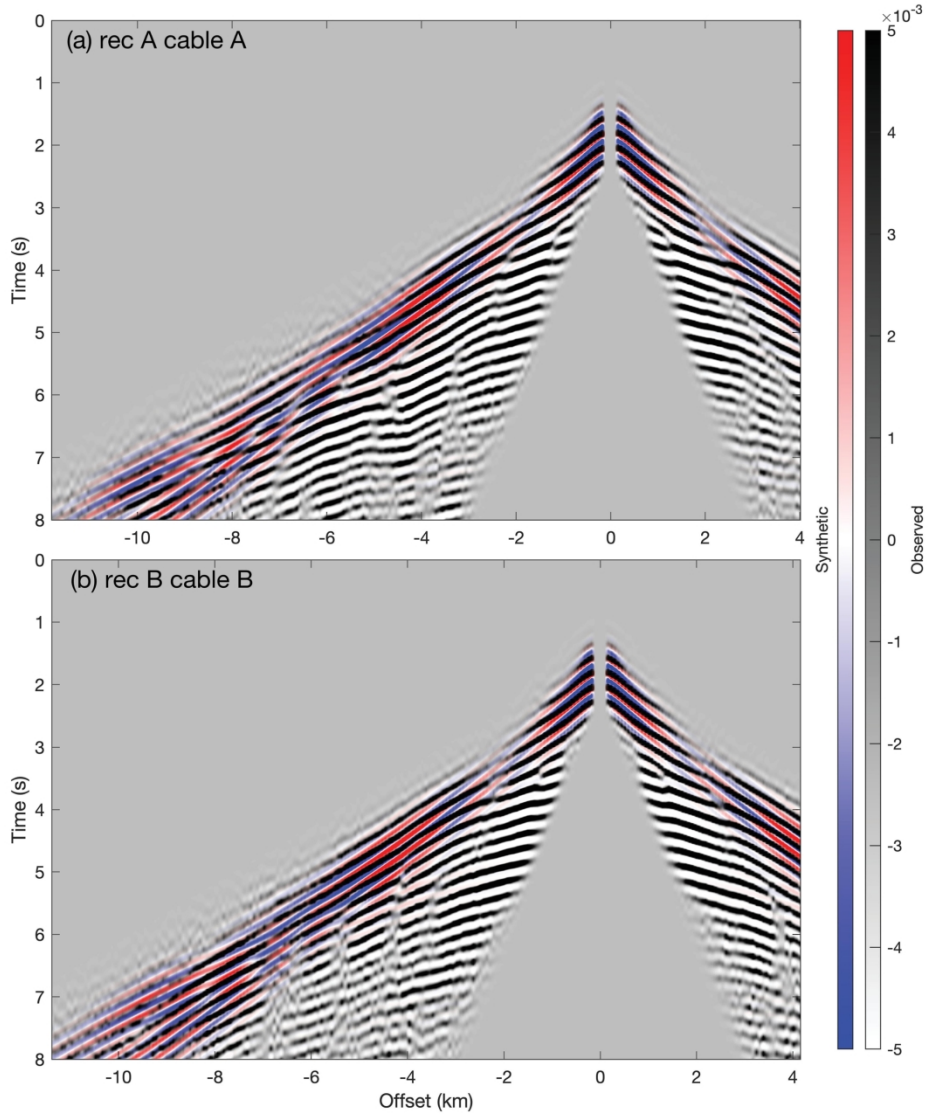

Figure 9: Same as \Cref{crg:CRG_5Hz_initTOMO} into the \modb initial model.

200x229mm (300 x 300 DPI)

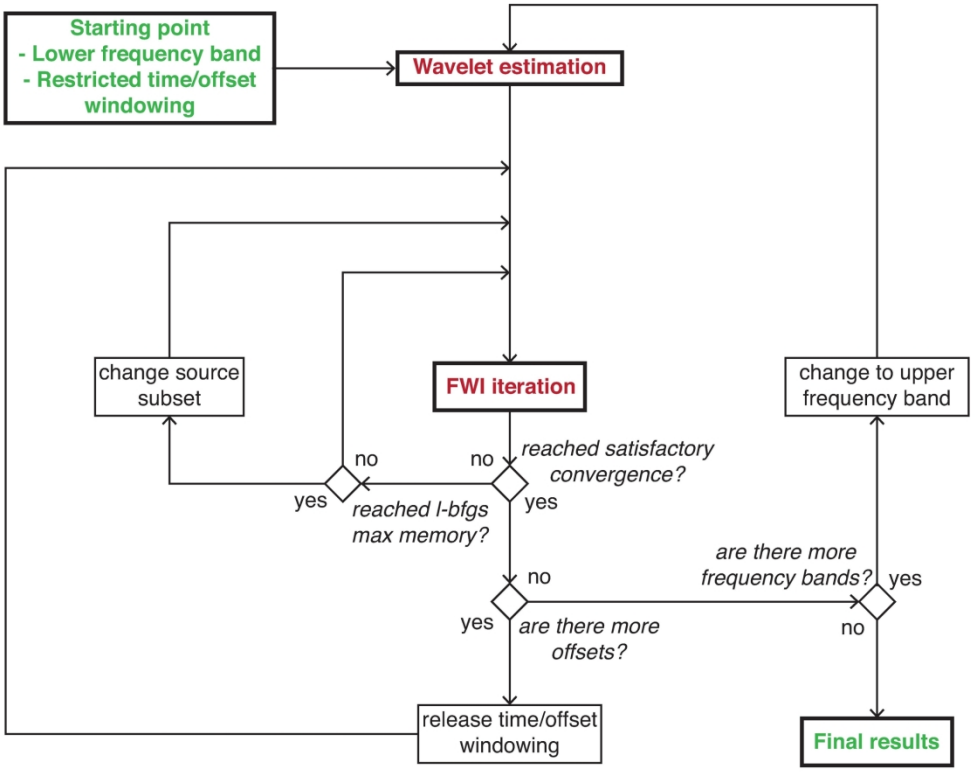

Figure 10: Complete FWI workflow used on the OBC Valhall dataset. At the core of the process lies the FWI iterations. Then several loops are nested one into another, from internal FWI iterations, source-subsampling, data selection, to finally the outer one of frequency continuation.

196x155mm (300 x 300 DPI)

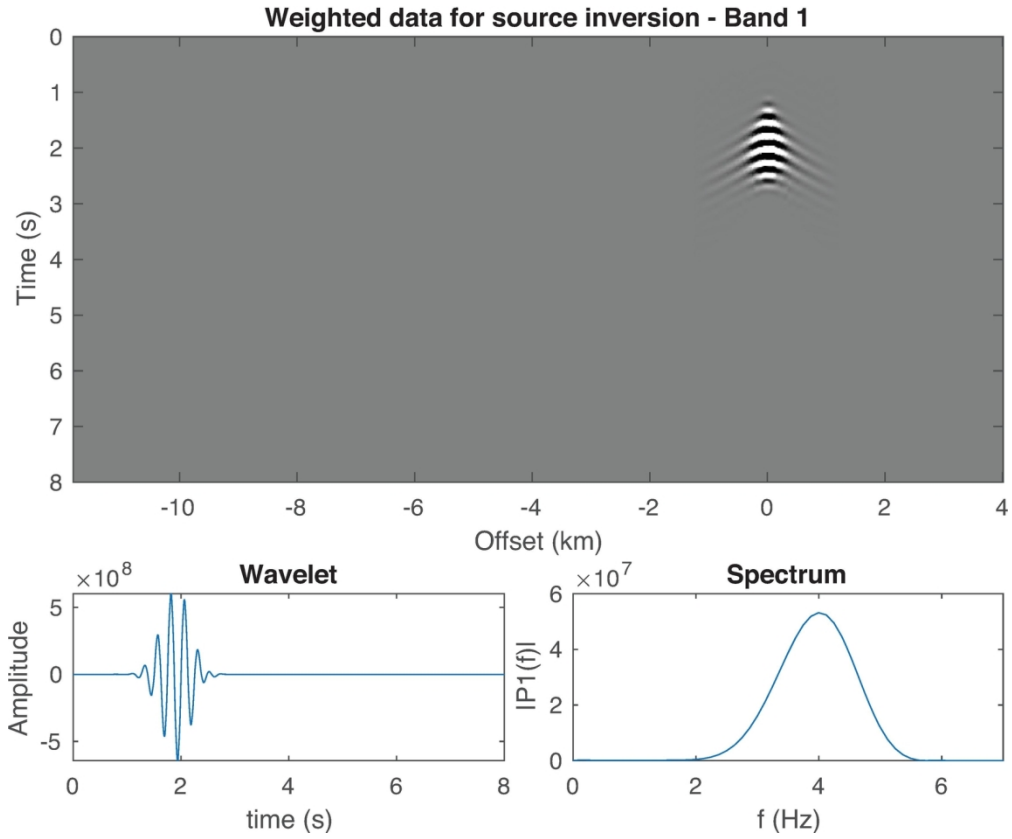

Figure 11: On top, weighted data for source inversion displayed on a 5~Hz 2D common-receiver gather (receiver A cable A). On the bottom, the estimated wavelet (left) and associated spectrum (right).

148x123mm (300 x 300 DPI)

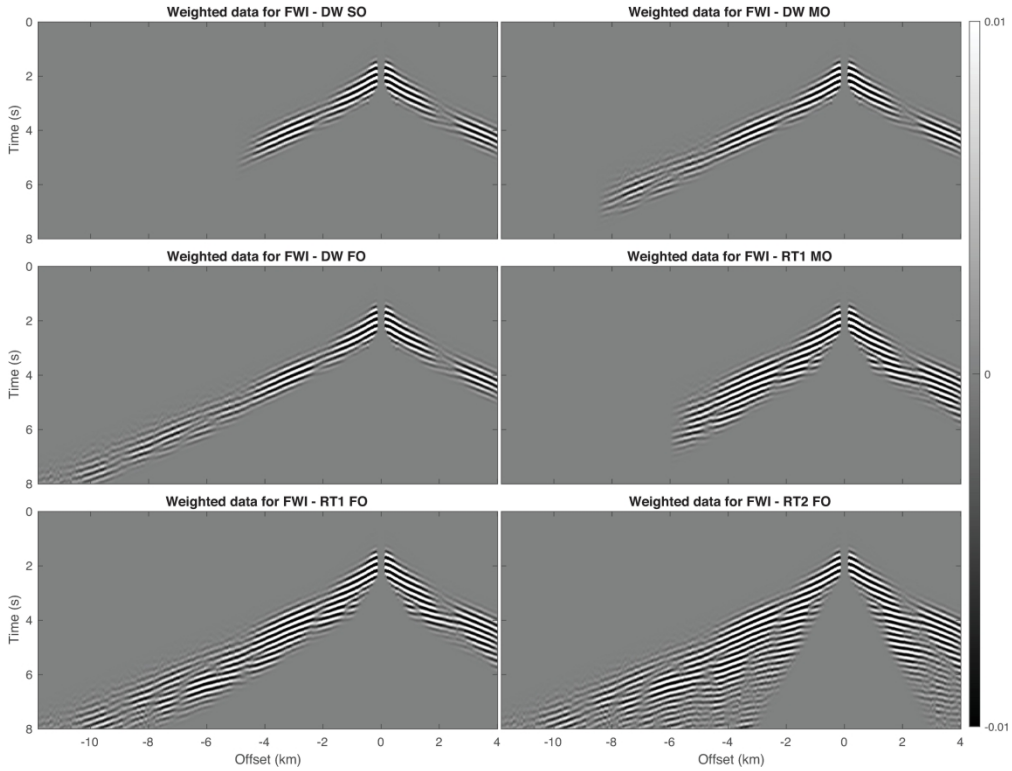

Figure 12: 2D common-receiver gathers extracted for receiver A along cable A with different data weighting applied on them. From top to bottom: first break \& short offset (DW SO), first break \& medium offset (DW MO), first break \& full offset (DW FO), first time release \& medium offset (RT1 MO), first time release \& full offset (RT1 FO), second time release \& full offset (RT2 FO).

288x219mm (300 x 300 DPI)

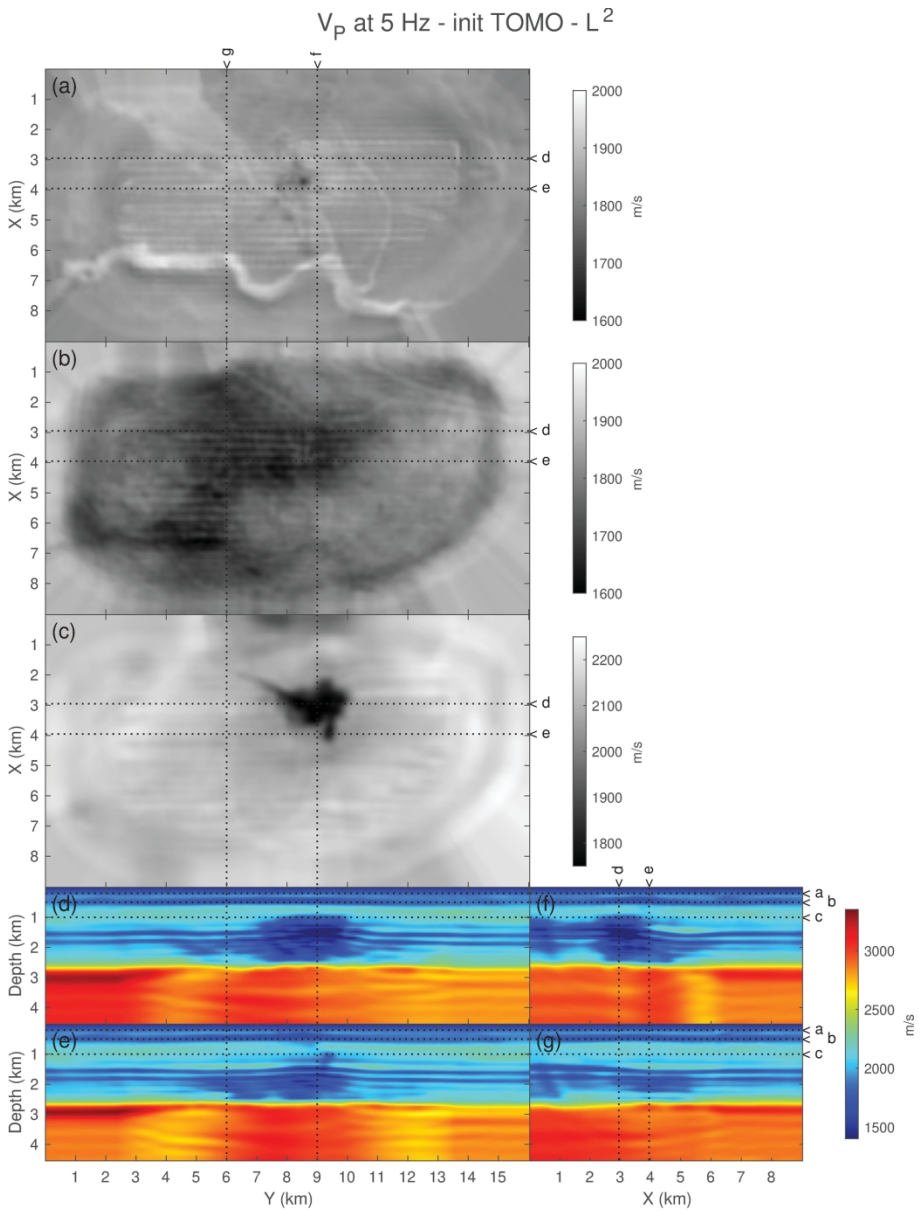

Figure 13: Slices of the 5~Hz FWI reconstructed \vp using \ls misfit function starting from \moda initial model. (a-c) Horizontal slices at (a) 200~m depth, (b) 500~m depth and (c) 1~km depth. (d-e) Inline vertical slices for (d) $x=2.95$~km and (e) $x=3.95$~km. (f-g) Cross-line vertical slices at (f) $y=9$~km and (g) $y=6$~km.

209x277mm (300 x 300 DPI)

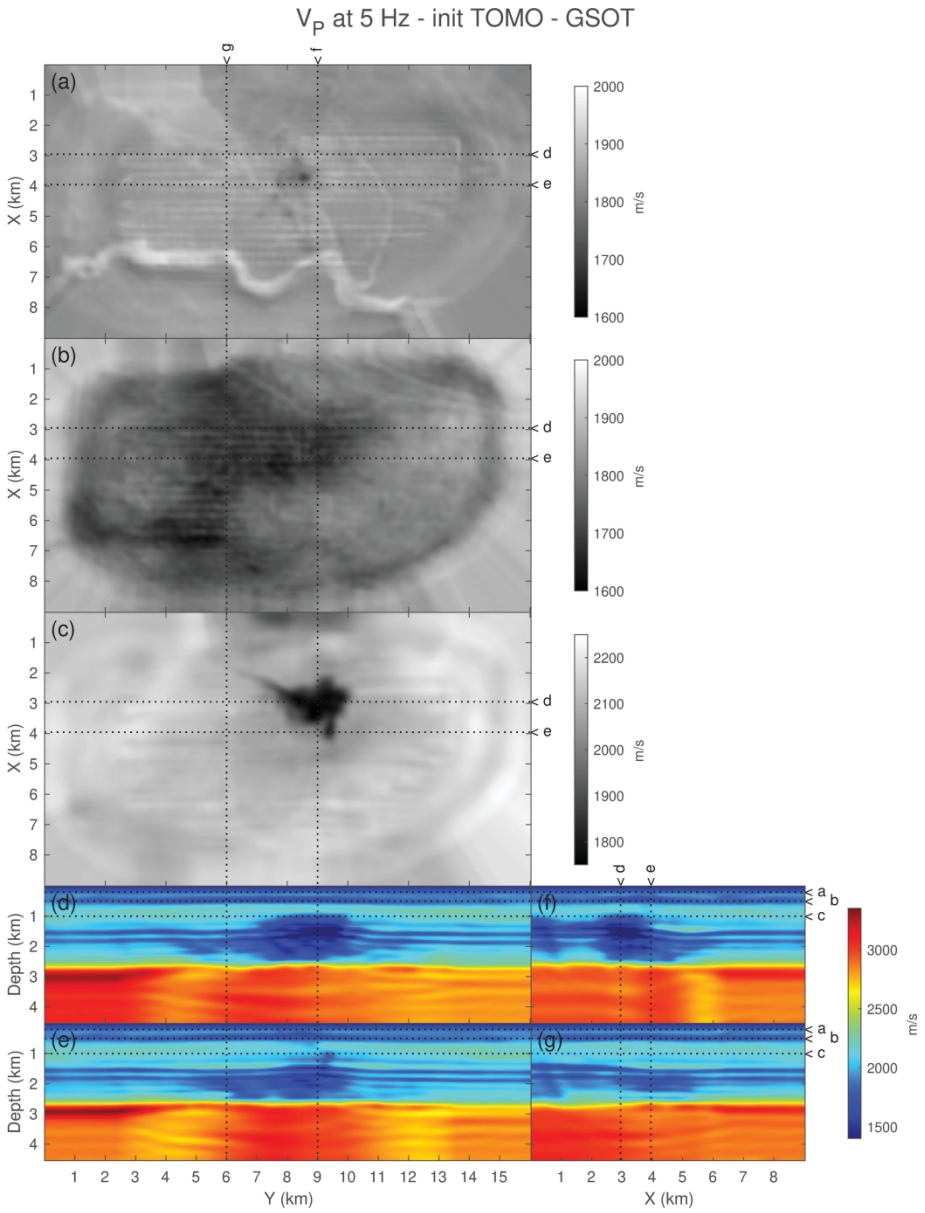

Figure 14: Same as \Cref{vp:VP_initTOMO_5Hz_L2} using GSOT misfit function.

209x276mm (300 x 300 DPI)

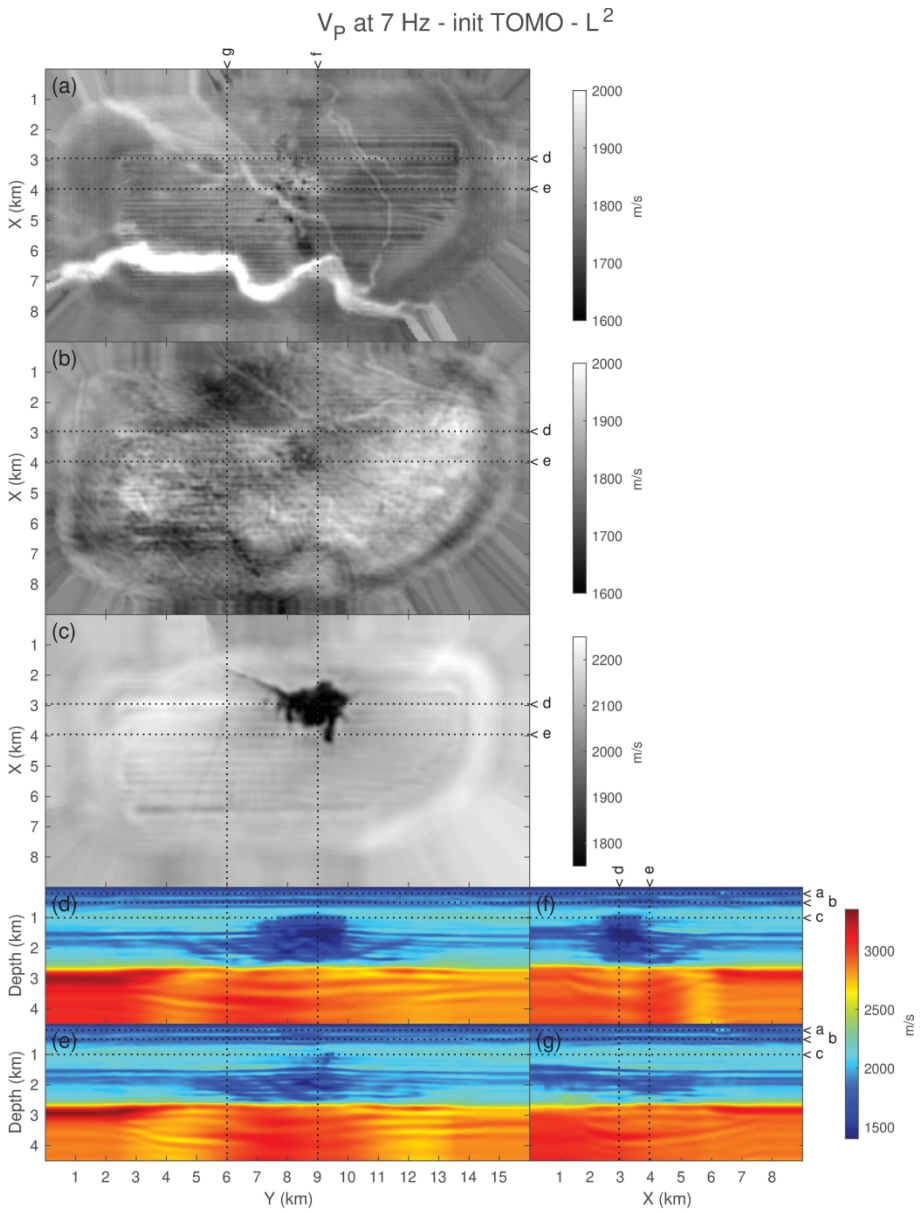

Figure 15: Same as \Cref{vp:VP_initTOMO_5Hz_L2} for 7~Hz results.

209x277mm (300 x 300 DPI)

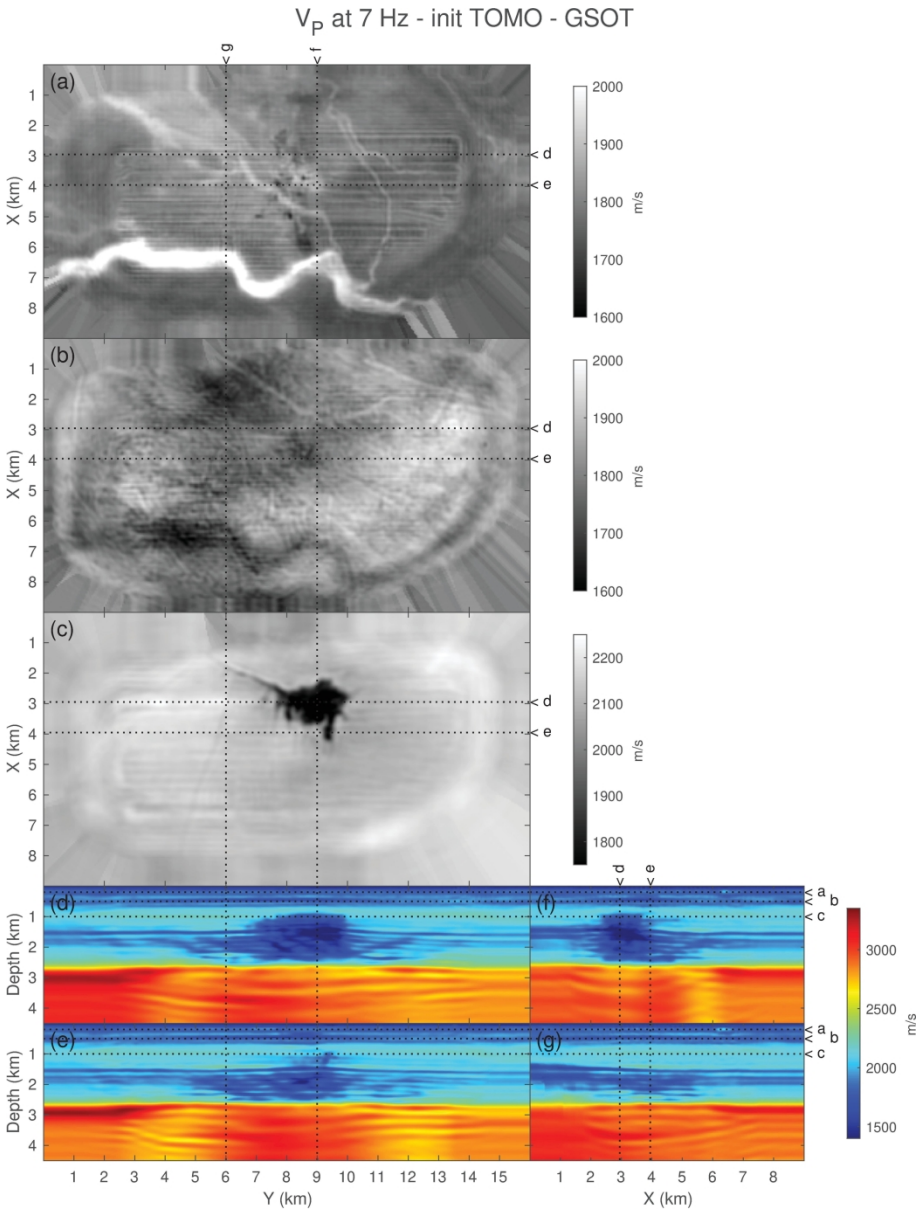

Figure 16: Same as \Cref{vp:VP_initTOMO_5Hz_GSOT} for 7~Hz results.

209x276mm (300 x 300 DPI)

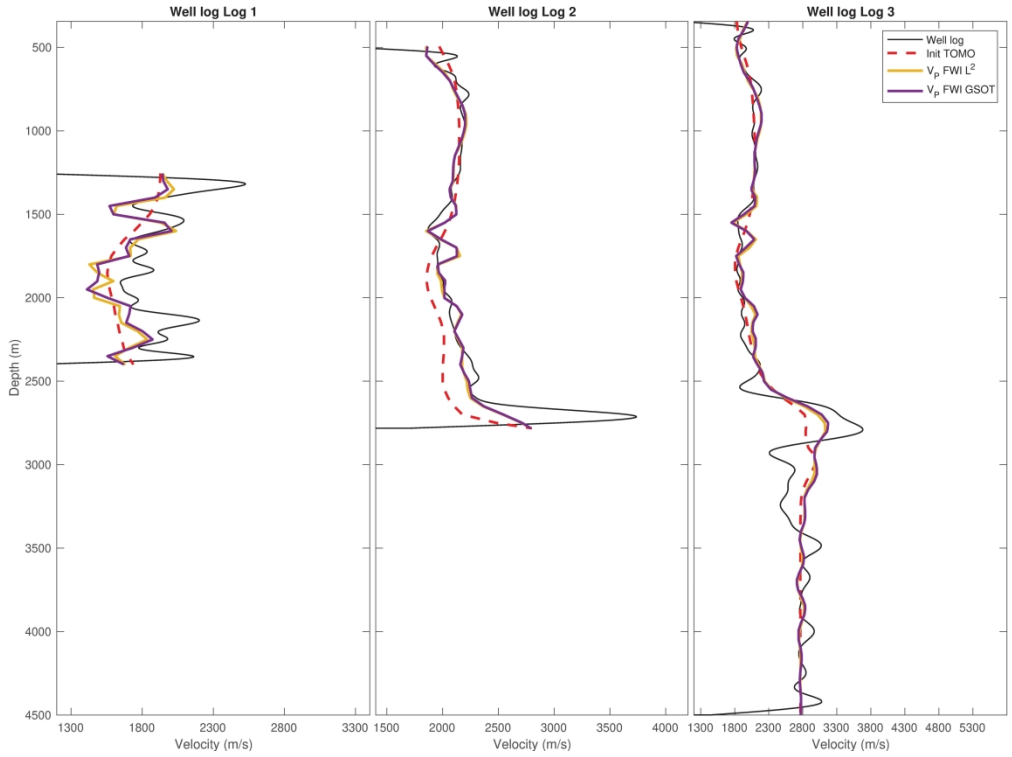

Figure 17: Comparison of \vp profiles extracted from the \moda initial model (dashed red), 7~Hz FWI models using \ls (solid yellow) and 7~Hz FWI model using GSOT (solid purple) with sonic log filtered in the $0-7$~Hz frequency band (solid black). Left subfigure corresponds to Log 1 at the center of the target. Middle subfigure to Log 2, and right subfigure to Log 3 (far away from the target).

299x223mm (300 x 300 DPI)

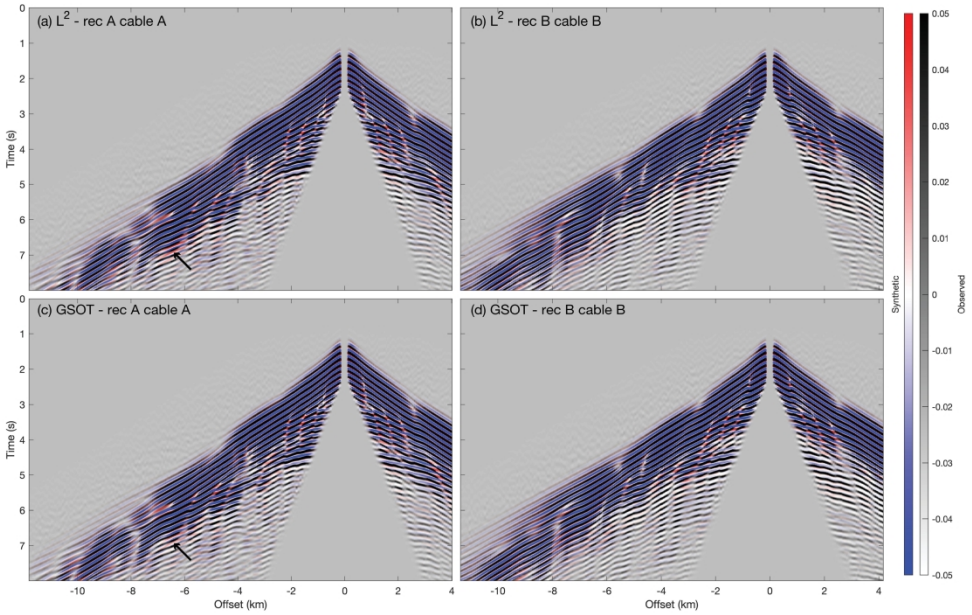

Figure 18: 2D common-receiver gathers at 7~Hz starting from the \moda initial model. Synthetic data (blue/white/red color scale) generated into the final reconstructed \vp using (a,b) \ls misfit function, (c,d) GSOT misfit function. (a,c) receiver A along cable A (through the low velocity anomaly). (b,d) receiver B along cable B. Field data are overlapped in grayscale with transparency. Black arrows point to improvement obtained with GSOT.

354x229mm (300 x 300 DPI)

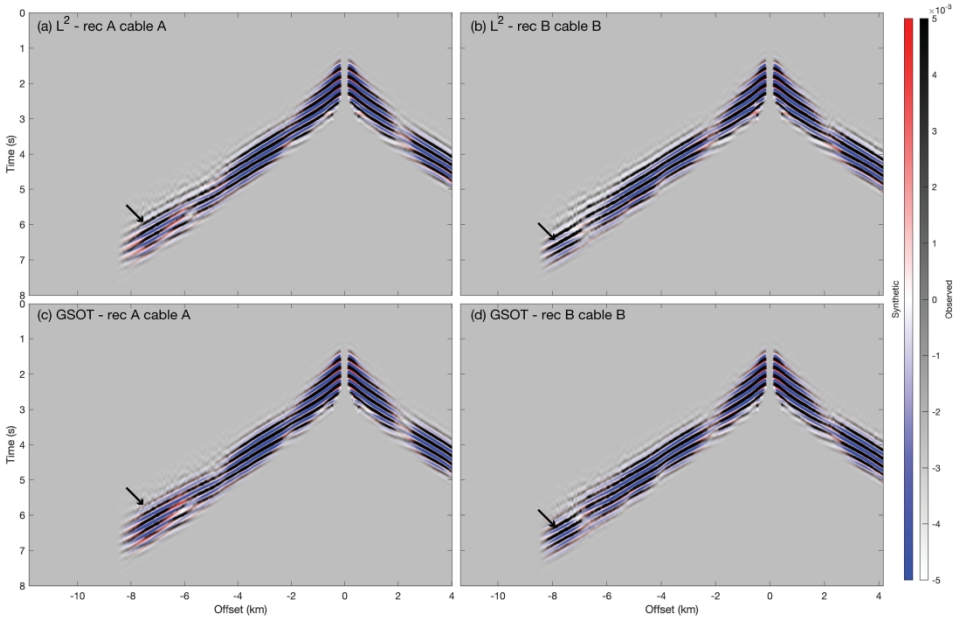

Figure 19: 2D common-receiver gathers at 5~Hz starting from the \modb initial model with data weighting apply (DW MO). Synthetic data (blue/white/red color scale) generated into the final reconstructed \vp at 2nd workflow step using: (a,b) \ls misfit function, (c,d) GSOT misfit function. (a,c) receiver A along cable A (through the low velocity anomaly). (b,d) receiver B along cable B. Field data are overlapped in grayscale with transparency. Black arrows point to area where GSOT improves the datafit.

354x229mm (300 x 300 DPI)

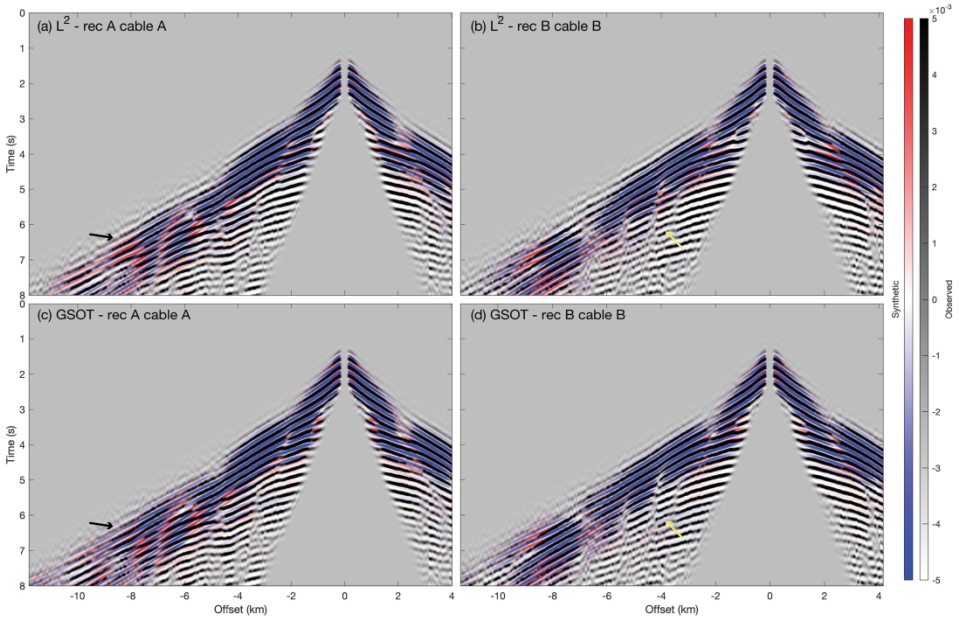

Figure 20: Same as \Cref{crg:CRG_5Hz_init1D_L2-GS_muted} but with final relaxed data weighting (RT2 FO) for display. This exhibits the improvement of datafit obtained in area which are not yet inverted. Here, black arrows point to improve fit and coherency of the diving wave with GSOT, whereas yellow arrows point to improved data-fit of reflected events with GSOT.

354x229mm (300 x 300 DPI)

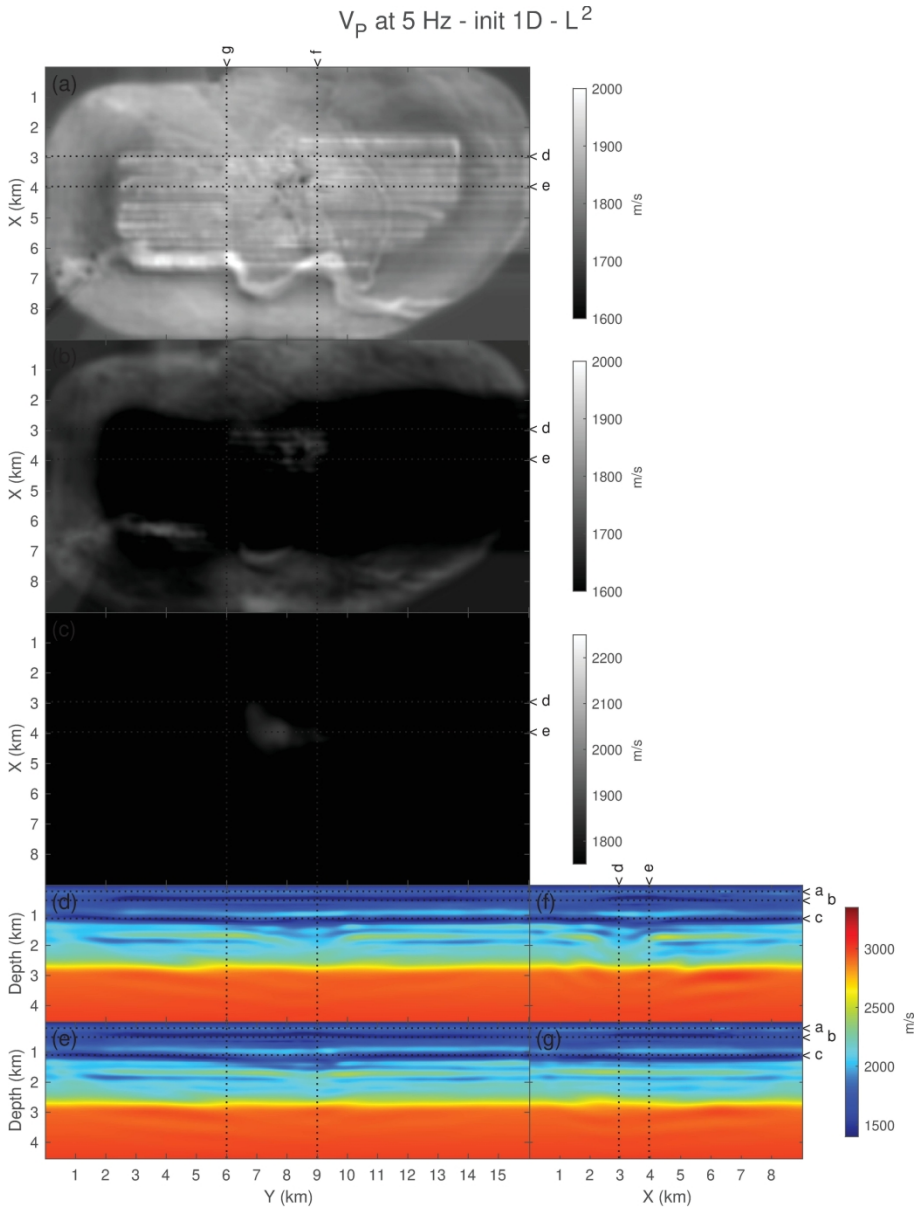

Figure 21: Slices of the 5~Hz FWI reconstructed \vp using \ls misfit starting from \modb initial model. (a-c) Horizontal slices at (a) 200~m depth, (b) 500~m depth and (c) 1.1~km depth. (d-e) Inline vertical slices for (d) $x=2.95$~km and (e) $x=3.95$~km. (f-g) Cross-line vertical slices at (f) $y=9$~km and (g) $y=6$~km.

209x277mm (300 x 300 DPI)

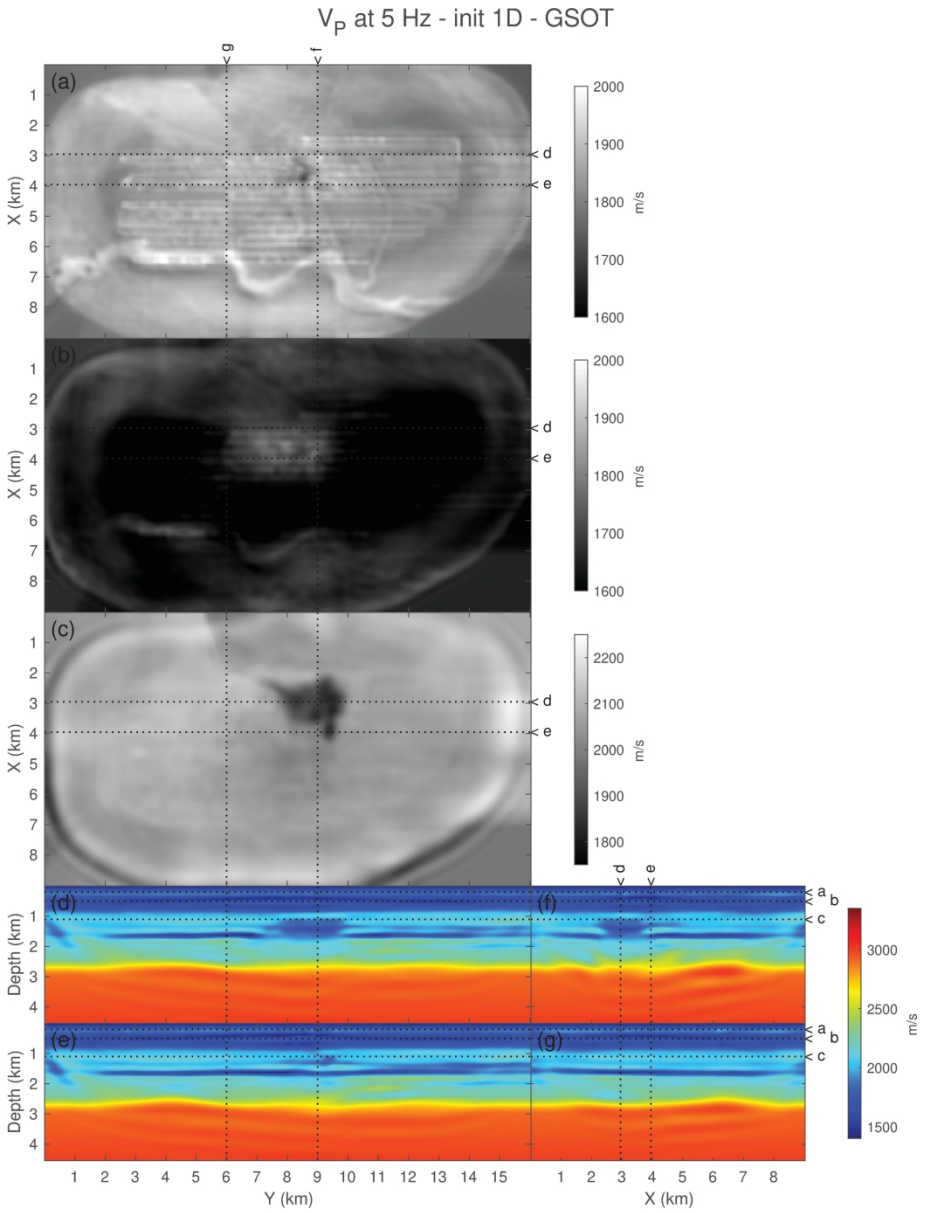

Figure 22: Same as \Cref{vp:VP_init1D_5Hz_L2} using GSOT. Here, results are consistant until 1.5 to 2~km depth compared to \lsb FWI. Characteristic structures of the Valhall field are recovered.

209x276mm (300 x 300 DPI)

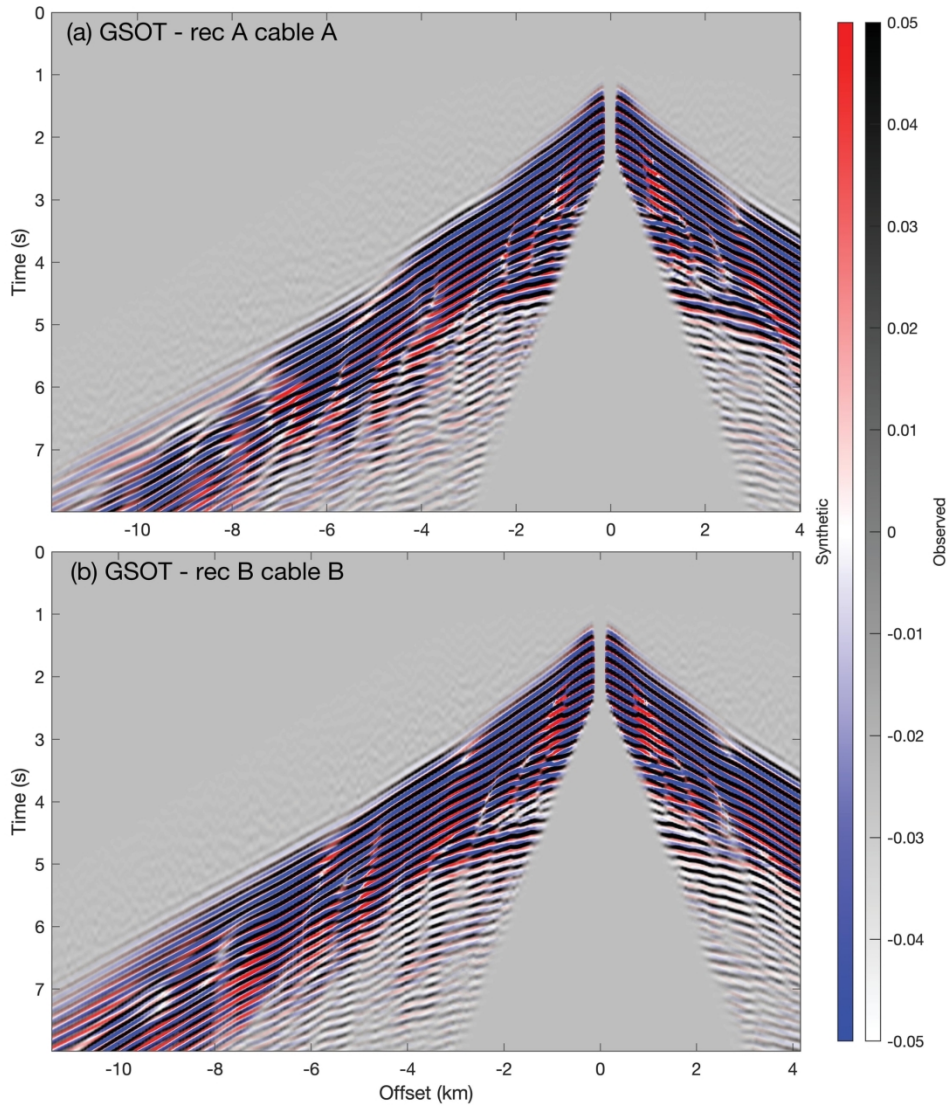

Figure 23: 2D common-receiver gathers at 7~Hz starting from the \modb initial model. Synthetic data (blue/white/red color scale) generated into the final reconstructed \vp using GSOT. (a) receiver A along cable A (through the low velocity anomaly). (b) receiver B along cable B. Field data are overlapped in grayscale with transparency.

200x229mm (300 x 300 DPI)

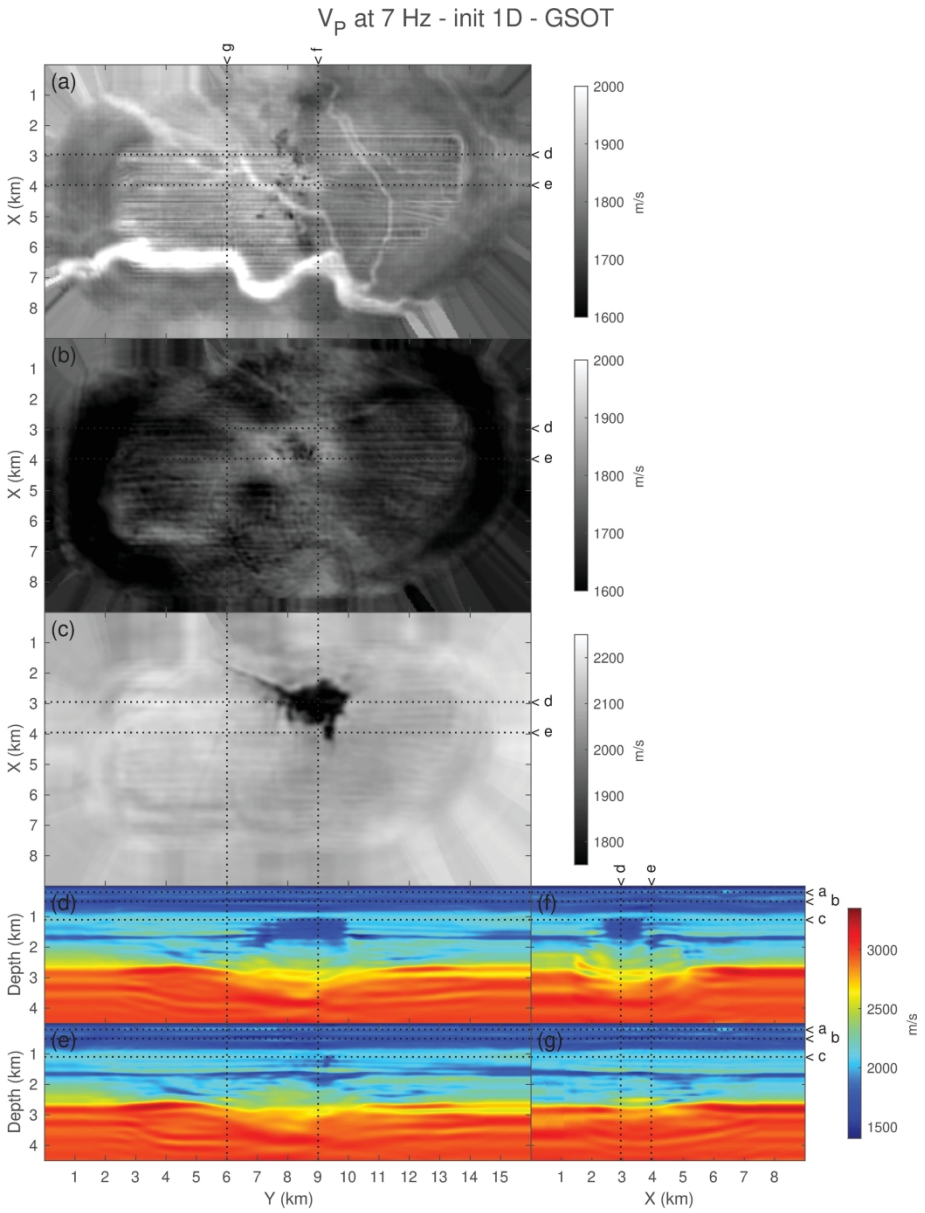

Figure 24: Slices of the 7~Hz FWI reconstructed \vp using GSOT misfit starting from \modb initial model. (a-c) Horizontal slices at (a) 200~m depth, (b) 500~m depth and (c) 1.1~km depth. (d-e) Inline vertical slices for (d) $x=2.95$~km and (e) $x=3.95$~km. (f-g) Cross-line vertical slices at (f) $y=9$~km and (g) $y=6$~km.

209x276mm (300 x 300 DPI)

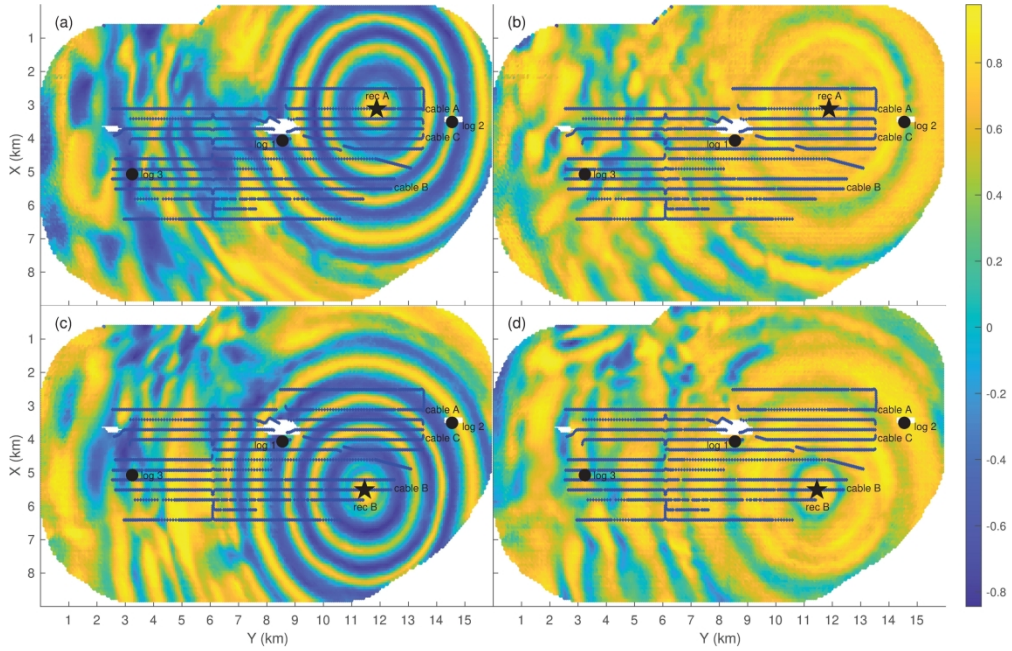

Figure 25: Cross-correlation analysis between field data and synthetic data. (a) common-receiver gather A and synthetic data in the \modb initial model. (b) common-receiver gather A and synthetic data in the final reconstructed GSOT-based FWI model. (c) common-receiver gather B and synthetic data in the \modb initial model. (d) common-receiver gather B and synthetic data in the final reconstructed GSOT-based FWI model.

274x176mm (300 x 300 DPI)

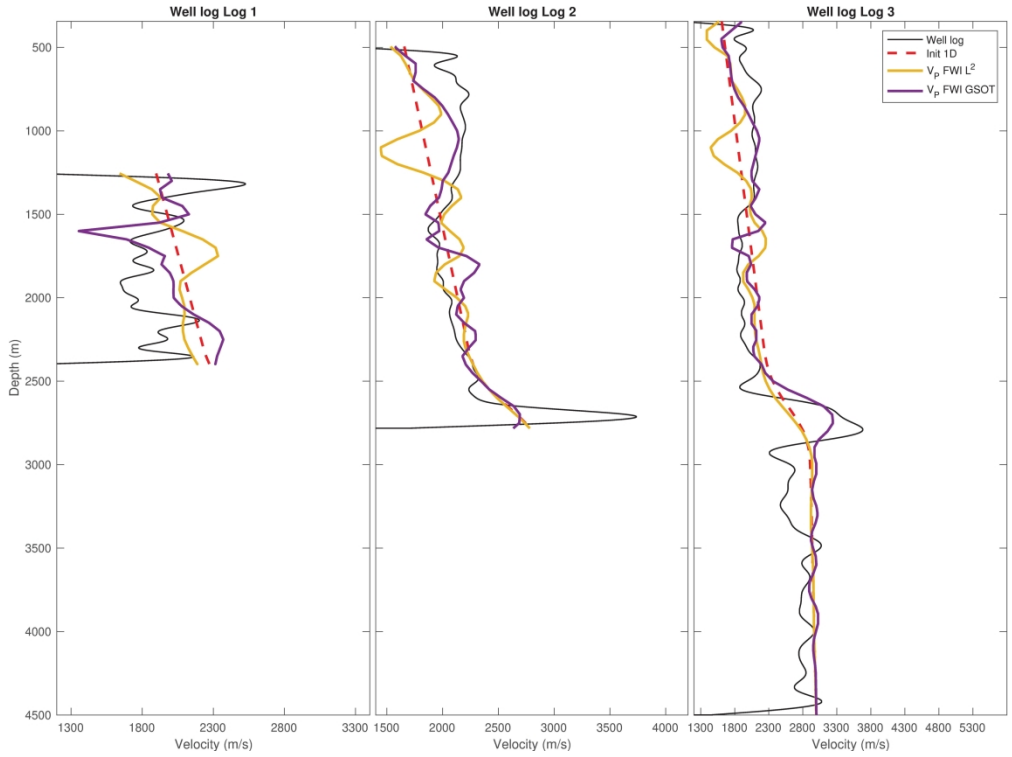

Figure 26: Comparison of \vp profiles extracted from the \modb initial model (dashed red), early FWI models using \ls (solid yellow) and 7~Hz FWI model using GSOT (solid purple) with sonic log filtered in the $0-7$~Hz frequency band (solid black). Left subfigure corresponds to Log 1 at the center of the target. Middle subfigure to Log 2, and right subfigure to Log 3 (far away from the target). Updates of velocity model obtained with GSOT are following the sonic logs trend until ≈ 2~km depth.

299x223mm (300 x 300 DPI)

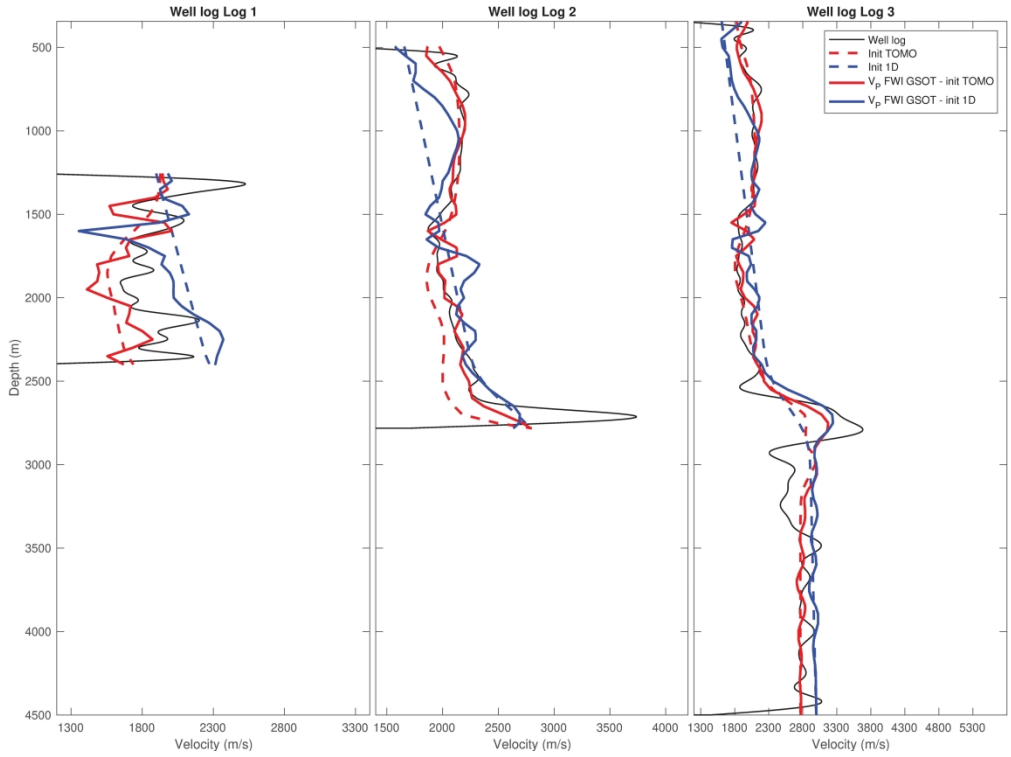

Figure 27: Comparison of \vp profiles extracted from the \moda and \modb initial models (respectively dashed red and dashed blue), GSOT-based FWI reconstructed models at 7~Hz starting from \moda and \modb initial models (respectively solid red and solid blue), with sonic log filtered in the $0-7$~Hz frequency band (solid black). Logs 2 and 3 show that results from the two different starting models are globaly following the same trend. Results from Log 1 passing through the target are following the same trend until 1.4~km depth.

299x223mm (300 x 300 DPI)

	computational efficiency	data distorsion	convexity	multi-D
Nonlinear transform OT	++	-	+	-
1-Wasserstein	-	+	-	+
GSOT	+	+	+	-

Table 1

	Gradient	Misfit	Total time	Ratio
L^2 - 5 Hz	243 s	**1 s**	254 s	100 %
GSOT - 5 Hz	243 s	**55 s**	308 s	121 %
L^2 - 7 Hz	898 s	**1 s**	912 s	100 %
GSOT - 7 Hz	898 s	**101 s**	1012 s	111 %

Table 2