

# Joint FWI for imaging deep structures: a graph-space OT approach

Yubing Li<sup>\*1</sup>, Romain Brossier<sup>1</sup>, Ludovic Métivier<sup>1,2</sup>

<sup>1</sup>Univ. Grenoble Alpes, ISTerre, F-38058 Grenoble, France

<sup>2</sup>Univ. Grenoble Alpes, CNRS, LJK, F-38058 Grenoble, France

## SUMMARY

Joint full waveform inversion (JFWI) aims at building a velocity macromodel of the subsurface by combining early arrivals and reflection waveform inversions. JFWI requires an explicit separation between early-arrivals and reflections, which is accomplished by applying time-windows to the data in practice. The JFWI approach is formulated as a workflow in which one repeatedly alternates two steps: the velocity macromodel is reconstructed assuming a known perturbation model, then the perturbation model is updated using the previously retrieved velocity as the background model. The perturbation model is used as an input to build the low-wavenumber sensitivity kernel along the two-way reflection paths. JFWI followed by FWI can further enrich the high-wavenumber contents of the subsurface model. However, JFWI, as FWI, suffers from cycle-skipping issues because one attempts to fit the data using the  $\ell^2$  misfit function. Optimal transport (OT) distances have been recently proposed to mitigate the non-convexity of the  $\ell^2$  misfit function in seismic imaging. Nevertheless, OT is initially designed to compare probability distributions, which is not the case for the original seismic data due to the oscillatory and signed natures. To overcome this difficulty, one possibility is to compare the discrete graph of data through OT. In this study, we assess the graph-space OT based JFWI approach for the inversion of 2D streamer synthetic data with a limited offset range (6 km) for the Marmousi model. Starting from a 1D linear model, GSOT-JFWI is less prone to cycle skipping than in the  $\ell^2$ -JFWI case, and consequently provides a sufficient initial velocity macromodel for subsequent FWI.

## INTRODUCTION

Full waveform inversion (FWI) is a powerful tool to build high-resolution velocity model of the subsurface (Virieux and Operto, 2009), especially for the areas that are properly sampled by early arrivals such as diving waves and super-critical reflections. However, the limited penetration depth of early arrivals prevents an accurate imaging of the structural target at deep part. Alternatively, to reach such targets, reflection waveform inversion (RWI) is a promising technique to build the velocity macromodel beyond the penetration depths of early arrivals (Xu et al., 2012; Brossier et al., 2015). It is inspired by the migration-based traveltimes tomography approach (Chavent et al., 1994). RWI relies on the scale separation which either assumes an explicit split between low and high frequency content of the velocity model (Xu et al., 2012) or introduces the velocity-impedance parameterization to naturally facilitate the separation (Zhou et al., 2015). In the latter case, RWI is alternated with the impedance waveform inversion (IpWI) of short-offset data, a least-squares migration-like procedure, that produces an impedance model used as an input of reflectivity information to build the sensitivity kernel along the reflection path. For the sake of efficiency, Li et al. (2019) em-

phasize the importance of preconditioning IpWI to reduce the computational cost of such workflow. Zhou et al. (2015) propose to extend RWI to joint full waveform inversion (JFWI) of early arrivals and reflections, which involves two different wavepaths simultaneously. Consequently, JFWI tends to preferentially sample both the vertical and horizontal components of the wavenumbers of the subsurface model.

JFWI, as in the FWI case, is prone to the convergence towards local minima, because one attempts to fit the data by minimizing the  $\ell^2$  misfit function which is non-convex. To mitigate this issue, one can exploit an alternative misfit function. Among others, optimal transport (OT) based functions have been investigated (Engquist and Froese, 2014; Métivier et al., 2016; Yang et al., 2018; Métivier et al., 2018) in the context of classical FWI. Interestingly, OT distance exhibits convexity with respect to shifted patterns. Nevertheless, OT is initially designed to compare probability distributions such that it is not directly applicable to seismic data due to its oscillatory and signed natures.

Among other solutions, Métivier et al. (2018) propose to compare the data represented in the discrete graph-space through OT (GSOT). Such approach preserves the appealing convex property of OT, and interestingly, it can be implemented trace-by-trace for solving realistic size waveform inversion problems at affordable computational costs (Métivier et al., 2019), saying it is feasible to be extended to 3D. In this study, we compare  $\ell^2$  and GSOT approaches in the context of JFWI for the Marmousi model. We utilize a streamer acquisition with a limited offset range (0-6 km) to evaluate the robustness of the method. Starting from a 1D linearly increasing velocity model, the GSOT based JFWI approach fits the data better than in the  $\ell^2$  case, and consequently provides better initial model for subsequent FWI. Moreover, the result supports the theory that GSOT does not decrease the resolution of the reconstructed model, as GSOT is a generalization of the  $\ell^p$  distance.

## METHODOLOGY

We first review the preconditioned IpWI method that we have introduced in Li et al. (2019) for the efficient impedance reconstruction, and the JFWI approach for the velocity macromodel update. Then, we present the GSOT formulation and its associated adjoint source used in the adjoint-state technique for computing the gradient (Plessix, 2006).

### Preconditioned IpWI approach

Consider  $u$ ,  $V_P$  and  $I_P$  as wavefield, velocity and impedance, respectively. Density  $\rho$  is linked to  $V_P$  and  $I_P$  through  $I_P = \rho V_P$ . The perturbations can be expressed as  $(\delta I, \delta V, \delta u) = (I_P - I_0, V_P - V_0, u - u_0)$ . In the acoustic isotropic medium, Li et al. (2019) define a shot-dependent parameter  $\xi(x, s) = 2\delta I / (V_0 I_0^2)$ , then derive an approximate inverse in the time-

## Joint FWI for imaging deep structures: a graph-space OT approach

domain by considering the small scattering angle case, reading

$$\xi(x, s) \approx \frac{\nabla S(s, x, t) \star \nabla R(s, x, t) + \frac{1}{V_0^2} \frac{\partial^2}{\partial t^2} S(s, x, t) \star R(s, x, t)}{G_0(s, x, t) \star G_0(s, x, t)}, \quad (1)$$

where  $s$ ,  $r$  and  $x$  are shot, receiver and imaging positions, respectively,  $\star$  the zero lag cross correlation, and  $G_0$  the Green's function. Incident and adjoint wavefields are defined in the frequency-domain only for simplicity

$$S(s, x, \omega) = \frac{4}{(i\omega)^3} \frac{1}{\Omega^*(\omega)} G_0(s, x, \omega), \quad (2)$$

$$R(s, x, \omega) = \int dr \frac{\partial G_0^*(r, x, \omega)}{\partial r_z} \delta u(s, r, \omega), \quad (3)$$

with  $\Omega$  the source wavelet depending on angular frequency  $\omega$ . Equations (1) to (3) are presented only in 3D for simplicity. In practice, we fix  $V_P$  and invert for  $I_P$  with only short-offset reflections by replacing the gradient of conventional IpWI by equation (1) with a summation over all shots.

### JFWI approach

With  $I_P$  reconstructed by preconditioned IpWI, we minimize the following misfit function to update the  $V_P$  macromodel

$$C[V_P] = f(w^e d, w^e p[V_P, I_P]) + f(w^r d, w^r p[V_P, I_P]) \quad (4)$$

where  $d$ ,  $p$  denote the observed and predicted data, respectively.  $w^e$  and  $w^r$  are weighting functions to roughly select early arrivals and reflections, respectively.  $f$  evaluates the difference between two input data, and refers to the  $\ell^2$  distance in the conventional case. The gradient of the misfit function with respect to the  $V_P$  macromodel reads

$$\nabla_{V_P} C = \sum_{s,r} (u_0 \star \delta \lambda^r + \delta u \star \lambda_0^r + \delta u \star \delta \lambda^r + u_0 \star \lambda_0^e), \quad (5)$$

where  $\lambda_0^e$  and  $\lambda_0^r$  denote the background adjoint wavefields generated by the difference between observed and predicted data in the early arrival and reflection cases, respectively.  $\delta \lambda^e$  and  $\delta \lambda^r$  are the scattered adjoint wavefields associated with the early arrived and reflected wave differences, respectively. By difference, we simply refer to the adjoint source, which is the data residual in the  $\ell^2$  case. See Zhou et al. (2015) for more details about the JFWI theory.

### GSOT strategy

Métivier et al. (2019) propose the GSOT strategy to compare the discrete graph of the observed and predicted data trace-by-trace. By graph-space, it means that a seismic trace  $d(t)$  is represented as the ensemble of  $K$  discrete points defined by  $(t_1, d_1), \dots, (t_K, d_K)$ . Assuming predicted data  $p$  in model  $m$ , the GSOT misfit function reads

$$f_{gsot}[m] = \sum_s \sum_r w_{obs}^{s,r} h(p[m], d), \quad (6)$$

$$h(p, d) = W^2((t, p), (t, d)) = \min_{\sigma \in S(K)} \sum_{i=1}^{N_i} c_{i\sigma(i)}(p, d), \quad (7)$$

where  $w_{obs}^{s,r}$  scales the contribution of each trace by the energy of the corresponding trace in the observed data to preserve the amplitudes-versus-offset effect.  $h$  corresponds to the

2-Wasserstein ( $W^2$ ) distance between discrete graphs of predicted and observed data.  $S(K)$  is the ensemble of permutations of  $(1 \dots K)$ , and  $c_{ij}$  the  $\ell^2$  distance between the discrete points  $(t_i, p(t_i))$  and  $(t_j, d(t_j))$

$$c_{ij}(p, d) = |t_i - t_j|^2 + \eta^2 |p(t_i) - d(t_j)|^2. \quad (8)$$

The scaling parameter  $\eta$  controls the extent to which we emphasize time shifts. In practice, it is defined as

$$\eta = \frac{\tau}{A}, \quad (9)$$

where  $\tau$  is user-defined depending on the maximum expected time shift between observed and predicted data in the initial model, and  $A$  the maximum amplitude discrepancy between observed and predicted data. To derive the minimizer  $\sigma^*$ , we solve the linear sum assignment problem (LSAP) of equation (6) using the auction algorithm (Bertsekas and Castanon, 1989). The approach is highly efficient for small dense instances of LSAP, which can be satisfied by resampling the seismic data according to the Nyquist frequency. Once the minimizer  $\sigma^*$  for (6) is derived, Métivier et al. (2019) prove that the adjoint source of  $f_{gsot}(m)$  reads

$$\frac{\partial h}{\partial p(t_i)} = 2 \left( p(t_i) - d(t_{\sigma^*(i)}) \right). \quad (10)$$

The  $W^2$  based GSOT approach can thus be considered as a generalization of the  $\ell^2$  distance (extendable to  $\ell^p$ ): the adjoint source is equivalent to the difference between calculated and observed data at time samples connected by the optimal assignment  $\sigma^*$ . Consequently, besides the convexity, we expect GSOT to have almost the same resolution as in the  $\ell^2$  case by appropriately choosing  $\tau$ . In practice, we implement the GSOT based JFWI by simply integrating equations (6) and (10) with (4) and (5), respectively.

### APPLICATION TO 2D SYNTHETIC STREAMER DATA

To generate the observed data, we trigger 64 shots located at the surface from 0.1 to 10.18 km in the true Marmousi model (Figures 1a and 1b). The shot interval is 0.16 km. For each shot, receivers are deployed on the right side of the shot with very limited offsets ranging from 0 to 6 km. The source function is a Ricker wavelet, of which the peak frequency is 4 Hz and the contents below 2 Hz are filtered out. For initial model, we generate a linearly increasing  $V_P$  model with a water layer on top: the velocity for areas shallower than 0.2 km is 1.5 km/s and the velocity at bottom is 4.1 km/s (Figure 1c). The initial impedance model is generated through  $I_P = \rho V_P$  where the density is a two-layer model: 1 g/cm<sup>3</sup> above 0.2 km and 2 g/cm<sup>3</sup> for the rest part (Figure 1d). With such configurations, we carry out 4 experiments to compare:  $\ell^2$ -FWI,  $\ell^2$ -JFWI followed by  $\ell^2$ -FWI, GSOT-JFWI followed by  $\ell^2$ -FWI, and GSOT-JFWI followed by GSOT-FWI. Note that such acquisition with limited offsets introduces more difficulties than in the conventional fixed spread case, as the diving waves only sample the model until around 1 km depth.

### JFWI workflows

The general JFWI workflow consists of the following steps: (1) reconstruct the perturbation model  $I_P'$  by performing IpWI

## Joint FWI for imaging deep structures: a graph-space OT approach

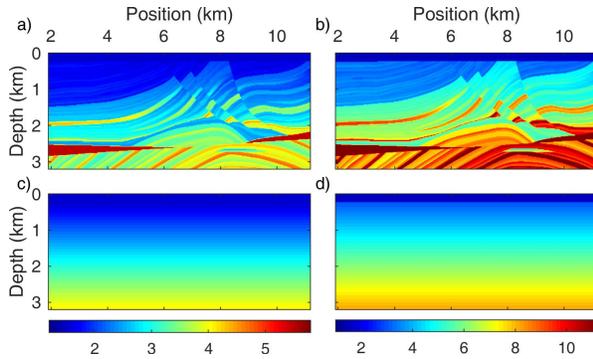


Figure 1: True  $V_P$  (a) and  $I_P$  (b) models, and linearly increasing initial  $V_P$  (c) and  $I_P$  (d) models.

for short-offset reflections in  $(I_P^0, V_P^{i-1})$ ; (2) update  $V_P^i$  macromodel by simultaneously considering the early arrivals and reflections with JFWI in  $(I_P^i, V_P^{i-1})$ ; (3) go back to step 1 until the convergence of step 2 is achieved. Note that  $I_P$  is reset to its initial value at the beginning of step 1. We use "cycle" to refer to one single loop of such workflow. We design the time windows for IpWI and JFWI as following: events arrived at receivers later than the time when direct arrivals stop are considered as reflections, and the others as early arrivals. The stopping time can be simply achieved by considering the offset and water velocity value for every trace.

In this study, we implement the JFWI workflow with specific strategies. For both  $\ell^2$ - and GSOT-JFWI cases, we perform 21 cycles, each of which consists of 3 preconditioned IpWI and 20 JFWI iterations. We always build the  $I_P$  model with reflections at offset less than 1 km. For first 5 cycles, we update the  $V_P$  macromodel with reflections at offsets between 1-6 km and all early arrivals. The reason is to honor the idea of Zhou et al. (2015) that  $V_P$  and  $I_P$  are responsible for short and large offset reflections, respectively. In the following cycles, we update the  $V_P$  macromodel with full-offset reflections and early arrivals for fitting the data overall. In the GSOT-JFWI case, we set  $\tau$  to 0.5 s for cycles 1-5, 0.3 s for cycles 6-16, and 0.2 s for remaining cycles. As for the subsequent GSOT-FWI,  $\tau$  is set to 0.2 s. Assume the source is known, we perform 400 iterations of  $\ell$ -BFGS for all subsequent FWI procedures. We always use all the frequency contents presented in the observed data.

### Inversion results

We first present the  $V_P$  and  $I_P$  models built by two types of JFWI approaches. The  $\ell^2$ -JFWI approach updates the  $V_P$  model in a less significant way (Figure 2a) than in the GSOT-JFWI case (Figure 2b), especially for the deep part of the subsurface. The reconstructed  $I_P$  perturbation models illustrate that GSOT-JFWI better retrieves the kinematic information. For example, the resulting  $I_P$  of  $\ell^2$ -JFWI cannot distinguish the discontinuous event (Figure 2c) as in the GSOT case (Figure 2d), as illustrated by the black dashed circles. In both cases, the major reflected events have been well interpreted. On the other hand, the  $\ell^2$  approach faces cycle skipping issues for diving waves, as marked by black arrows in Figure 3a. Indeed, the GSOT based approach have reduced the ambiguity for fitting these diving waves (Figure 3b), and thus introduced better ro-

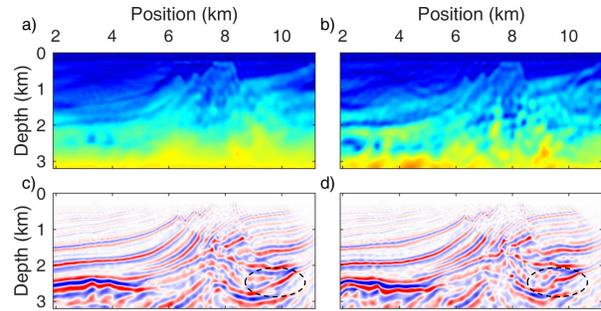


Figure 2:  $V_P$  models built by  $\ell^2$ -JFWI (a) and GSOT-JFWI (b) using the linearly increasing initial model, and  $\delta I$  (c,d) reconstructed in these  $V_P$  models, respectively.

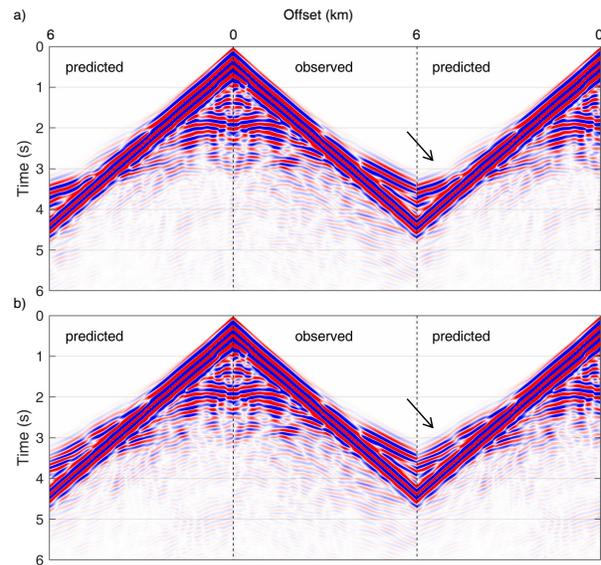


Figure 3: Observed and predicted data corresponding to  $\ell^2$ -JFWI (a) and GSOT-JFWI (b) results at  $x_s = 7.3$  km.

bustness against cycle skipping issues for JFWI.

Then, we present the FWI results by starting from the 1D linear  $V_P$  model and from the  $V_P$  models built by two JFWI approaches, respectively. Note that we reset the density model to its initial value during subsequent inversion to put the resolution back to velocity. The classical FWI approach is severely cycle skipped (Figure 4a). On the other hand,  $\ell^2$ -FWI procedures starting from JFWI results both provide sufficient  $V_P$  models of which the high-wavenumber contents are enriched (Figures 4b and 4c). Interestingly, GSOT-FWI starting from the GSOT-JFWI result (Figure 4d) produces almost the same result as in the GSOT-JFWI+ $\ell^2$ -FWI case. It supports the theory that GSOT is a generalization of  $\ell^2$  distance and consequently will not decrease the resolution of results with an appropriate choice of  $\tau$ . To carefully check the quality of reconstructed  $V_P$  models, we extract the vertical and horizontal profiles of different results.  $\ell^2$ -JFWI+ $\ell^2$ -FWI underestimate the velocity at areas deeper than 2 km compared to the GSOT-JFWI+ $\ell^2$ -FWI and GSOT-JFWI+GSOT-FWI cases, as presented in Figure 5a. This part is the most well-illuminated area for the acquisition considered and any inaccuracy can re-

## Joint FWI for imaging deep structures: a graph-space OT approach

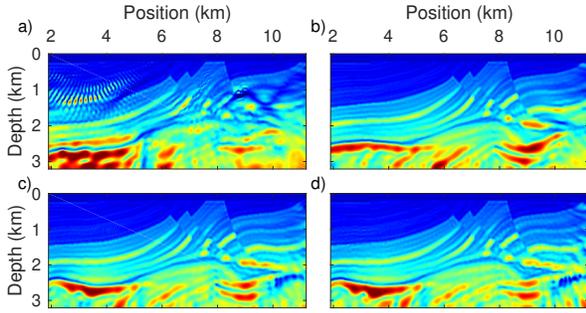


Figure 4:  $V_p$  models reconstructed by  $\ell^2$ -FWI (a),  $\ell^2$ -JFWI followed by  $\ell^2$ -FWI (b), GSOT-JFWI followed by  $\ell^2$ -FWI (c), and GSOT-JFWI followed by GSOT-FWI (d).

sult from the approach adopted. The horizontal profiles (Figure 5b) are almost identical for both three cases, except that  $\ell^2$ -JFWI+ $\ell^2$ -FWI overestimates the velocity at around  $x = 10$  km. In the data calculated for the result of  $\ell^2$  FWI starting from 1D linear model (Figure 6a), it is visible that observed and predicted early arrivals are not in phase. For the same event marked by black arrows,  $\ell^2$ -JFWI+ $\ell^2$ -FWI significantly improves the consistency between two data, but the phase ambiguity still exists (Figure 6b). Differently, the two cases stratifying from GSOT-JFWI results are immune to cycle-skipping (Figures 6c and 6d). Consequently, we consider that GSOT-JFWI+ $\ell^2$ -FWI and GSOT-JFWI+GSOT-FWI have provided the best results in the sense that data are properly interpreted.

The computational cost for a preconditioned IpWI gradient is 1.5 times as in the conventional case. The JFWI gradient requires twice the computational time as in the FWI case. No additional memory requirements are introduced in both cases. Considering the same misfit function, the JFWI workflow requires around twice the computational time as in the FWI case until convergence for our experiment. Thanks to the efficient GSOT approach developed by Métivier et al. (2019), the computational cost increase for our configurations is limited to approximately 2.5 % compared to standard  $\ell^2$ -JFWI or  $\ell^2$ -FWI.

### CONCLUSIONS

We apply GSOT to JFWI in this study. With a very limited offset range of data, GSOT-based JFWI followed by FWI provides well-reconstructed  $V_p$  model of subsurface starting from the 1D linear model. It is remarkable that GSOT significantly reduces the cycle skipping issues for JFWI but does not decrease the resolution of results. Consequently, it is of great interest for JFWI to adopt GSOT to further improve the robustness. The proposed workflow can be seamlessly extended to 3D at affordable numerical costs. The future work consists of the application of GSOT based JFWI to 3D field data.

**Acknowledgements:** This study was partially funded by the SEISCOPE consortium (<https://seiscope2.osug.fr>), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC and TOTAL. This study was granted access to the HPC resources of CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>)

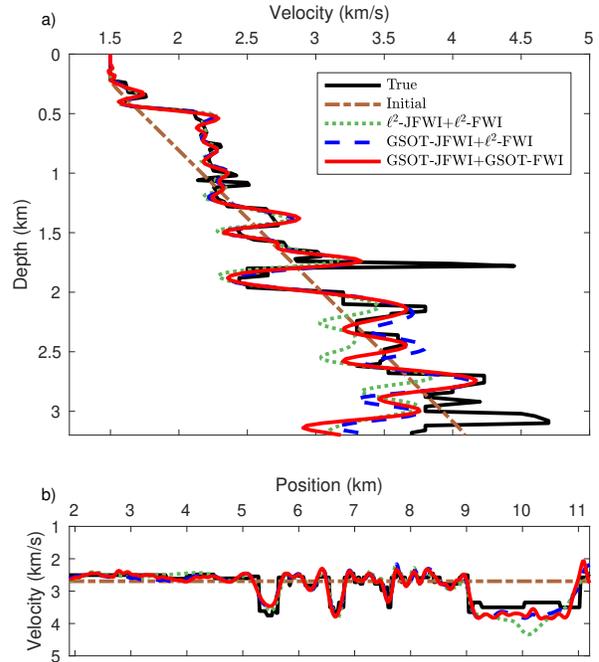


Figure 5: Vertical profiles at  $x = 7.4$  km (a) and horizontal profiles at  $z = 1.6$  km (b) for different final  $V_p$  models.

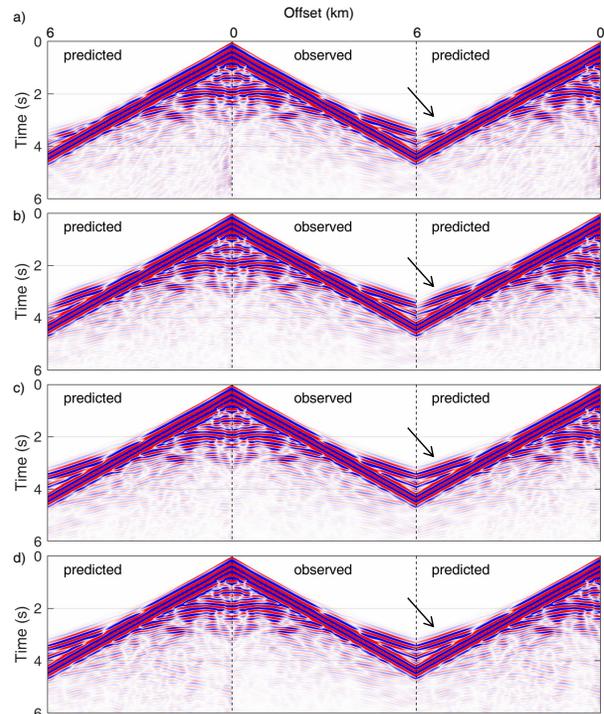


Figure 6: Observed and predicted data at  $x_s = 7.3$  km corresponding to the models reconstructed by  $\ell^2$ -FWI (a),  $\ell^2$ -JFWI followed by  $\ell^2$ -FWI (b), GSOT-JFWI followed by  $\ell^2$ -FWI (c), and GSOT-JFWI followed by GSOT-FWI (d).

and CINES/IDRIS/TGCC under the allocation 046091 made by GENCI.

## REFERENCES

- Bertsekas, D. P., and D. Castanon, 1989, The auction algorithm for the transportation problem: *Annals of Operations Research*, **20**, 67–96, doi: <https://doi.org/10.1007/bf02216923>.
- Brossier, R., S. Operto, and J. Virieux, 2015, Velocity model building from seismic reflection data by full waveform inversion: *Geophysical Prospecting*, **63**, 354–367, doi: <https://doi.org/10.1111/1365-2478.12190>.
- Chavent, G., F. Clément, and S. Gómez, 1994, Automatic determination of velocities via migration-based travelttime waveform inversion: A synthetic data example: 64th Annual International Meeting, SEG, Expanded Abstracts, 1179–1182, doi: <https://doi.org/10.1190/1.1822731>.
- Engquist, B., and B. D. Froese, 2014, Application of the Wasserstein metric to seismic signals: *Communications in Mathematical Science*, **12**, 979–988, doi: <https://doi.org/10.4310/cms.2014.v12.n5.a7>.
- Li, Y., R. Brossier, and L. Métivier, 2019, Combining approximate inverse and joint full waveform inversion for velocity model building: 81st Annual International Conference and Exhibition, EAGE, Extended Abstracts (London), submitted.
- Métivier, L., A. Allain, R. Brossier, Q. Méridot, E. Oudet, and J. Virieux, 2018, Optimal transport for mitigating cycle skipping in full waveform inversion: A graph space transform approach: *Geophysics*, **83**, no. 5, R515–R540, doi: <https://doi.org/10.1190/geo2017-0807.1>.
- Métivier, L., R. Brossier, Q. Méridot, and E. Oudet, 2019, A graph space optimal transport distance as a generalization of Lp distances: Application to a seismic imaging inverse problem: *Inverse Problems*, submitted.
- Métivier, L., R. Brossier, Q. Méridot, E. Oudet, and J. Virieux, 2016, An optimal transport approach for seismic tomography: Application to 3D full waveform inversion: *Inverse Problems*, **32**, 115008, doi: <https://doi.org/10.1088/0266-5611/32/11/115008>.
- 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, doi: <https://doi.org/10.1111/j.1365-246x.2006.02978.x>.
- Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: <https://doi.org/10.1190/1.3238367>.
- Xu, S., D. Wang, F. Chen, G. Lambaré, and Y. Zhang, 2012, Inversion on reflected seismic wave: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–7, doi: <https://doi.org/10.1190/segam2012-1473.1>.
- Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: *Geophysics*, **83**, no. 1, R43–R62, doi: <https://doi.org/10.1190/geo2016-0663.1>.
- 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, doi: <https://doi.org/10.1093/gji/ggv228>.