

Assessment of optimal transport based FWI: 3D OBC Valhall case study

A. Pladys^{*1}, R. Brossier¹, M. Irnaka¹, N. Kamath¹, L. Métivier^{1,2},

¹Univ. Grenoble Alpes, ISTERre, F-38058 Grenoble, France

²Univ. Grenoble Alpes, CNRS, LJK, F-38058 Grenoble, France

SUMMARY

Optimal transport (OT) distances have been recently proposed to mitigate the non-convexity of the L^2 misfit function in FWI. However, OT is only applicable to positive and normalized data. To overcome this difficulty, we have proposed two strategies, one based on the Kantorovich-Rubinstein (KR) norm, which extends a specific OT distance to the comparison of signed data, the other based on the interpretation of the discrete graph space of the data through OT. In this study, we compare these two approaches for the inversion of 3D OBC data from the Valhall field, using a visco-acoustic time-domain FWI algorithm. Starting from a crude initial velocity model, both KR and graph space approaches provide more reliable results than L^2 , the best results being obtained with the graph space approach. Thanks to a recently developed numerical approach, the computational cost increase is limited in this case to approximately 15 % compared to standard L^2 FWI.

INTRODUCTION

Full waveform inversion (FWI) is a high-resolution seismic imaging data fitting based procedure, now routinely used for exploration and regional scale targets (Fichtner et al., 2010; Plessix and Perkins, 2010; Stopin et al., 2014; Bozdağ et al., 2016). Despite successful applications, and its ability to account for an increasingly complex physics of wave propagation (from simple acoustic to visco-elastic approximation), one major issue still relies on the non-convexity of the L^2 misfit function which is minimized. This makes FWI strongly dependent on the accuracy of the initial model, as well as the lowest frequency available in the data (phase ambiguity/cycle skipping issue) (Virieux and Operto, 2009).

Among recent attempts to mitigate this issue, optimal transport (OT) based misfit functions have been proposed (Engquist and Froese, 2014; Métivier et al., 2016b; Yang et al., 2018b; Métivier et al., 2018). The convexity of the OT distance with respect to shifted patterns makes it a good candidate to improve the convexity of FWI misfit functions. However, OT theory is founded for the comparison of probability distributions, hence not directly applicable for the comparison of seismic data which are oscillatory, non-positive functions.

We have proposed two strategies to overcome this issue. The first consists in considering a specific OT distance, which naturally extends to the comparison of signed functions, known as the Kantorovich Rubinstein (KR) norm (Métivier et al., 2016a,b). The second consists in comparing the discrete graph of the data through OT rather than the data itself, referred to as the GSOT approach in the following. While the KR approach makes it possible to compare directly 2-dimensional shot gathers and widen the valley of attraction of the misfit function, it loses the convexity property of OT. The GSOT preserves this convexity property; however, it is applied trace by trace and requires to solve a 2D OT problem for each trace, yielding possible large computational cost (Métivier et al., 2018).

We have recently designed a numerical strategy which makes possible the application of the GSOT strategy to realistic size FWI problems (Métivier et al., 2019). In this study, we want to compare the KR and GSOT approach on the 3D OBC Valhall data, to assess their interest in designing more robust FWI schemes on field data. We apply these techniques in the frame of time-domain visco-acoustic FWI. Starting from a crude initial velocity model, the KR and GSOT approach provide more reliable results than L^2 , the best results being obtained with the GSOT approach. The computational cost increase is limited in this case to approximately 15% compared to L^2 FWI.

METHODS

We first review the KR and GSOT misfit function formulations, as well as their corresponding adjoint sources for the gradient computation following the adjoint state strategy (Plessix, 2006). We denote the observed and calculated data set by $d_{obs}(x_r, x_s, t)$ and $d_{cal}[m](x_r, x_s, t)$ respectively, where (x_r, x_s) denotes the spatial position of the receivers and sources, t is the time index, and $[m]$ indicates the dependency of the calculated data with respect to the model parameters m .

KR strategy

The KR misfit function is based on the dual of the 1-Wasserstein distance, using an ℓ_1 ground distance. It is expressed as

$$f_{KR}[m] = \sum_s g_s(d_{cal}[m], d_{obs}),$$

$$g_s(d_{cal}, d_{obs}) = \max_{\varphi_s \in \text{Lip}_1} \int_{x_r} \int_t \varphi_s(x_r, t) (d_{cal}(x_s, x_r, t) - d_{obs}(x_s, x_r, t)) dt dx_r, \quad (1)$$

where Lip_1 is the set of 1-Lipschitz functions for the ℓ_1 distance

$$\text{Lip}_1 = \{ \varphi(x_r, t), |\varphi(x_r, t) - \varphi(x'_r, t')| < |x_r - x'_r| + |t - t'| \}. \quad (2)$$

The adjoint source is given by

$$\frac{\partial g_s}{\partial d_{cal}} = \bar{\varphi}_s(x_r, t), \quad (3)$$

where

$$\bar{\varphi}_s(x_r, t) = \arg \max_{\varphi_s \in \text{Lip}_1} \int_{x_r} \int_t \varphi_s(x_r, t) (d_{cal}(x_s, x_r, t) - d_{obs}(x_s, x_r, t)) dt dx_r. \quad (4)$$

The computation of the KR misfit function and its corresponding adjoint source thus requires to solve a single constrained maximization problem per shot gather. This is performed through the proximal splitting algorithm ADMM (Combettes and Pesquet, 2011). At each iteration, a Poisson's problem is solved, for which we use an FFT-based solver. This yields an $O(N \log N)$ algorithm per iteration to be applied independently to each shot gather, where $N = N_r \times N_t$ with N_r the number of receivers and N_t the number of time samples.

GSOT strategy

The GSOT strategy compares the discrete graph of the observed and calculated data for each trace independently. We introduce

$$s_{cal,s,r}[m](t) = d_{cal}[m](x_r, x_s, t), \quad s_{obs,s,r}(t) = d_{obs}(x_r, x_s, t). \quad (5)$$

The corresponding misfit function is formulated as

$$f_{GSOT}[m] = \sum_s \sum_r h(s_{cal,s,r}[m], s_{obs,s,r}),$$

$$h(s_{cal}, s_{obs}) = \min_{\sigma \in S(N_t)} \sum_{i=1}^{N_t} c_{i\sigma(i)}(s_{cal}, s_{obs}), \quad (6)$$

with c_{ij} the ℓ^2 distance between the discrete points of the graph $(t_i, s_{cal}(t_i))$ and $(t_j, s_{obs}(t_j))$

$$c_{ij}(s_{cal}, s_{obs}) = |t_i - t_j|^2 + \eta^2 |s_{cal}(t_i) - s_{obs}(t_j)|^2, \quad (7)$$

and $S(N_t)$ the ensemble of permutations of $(1 \dots N_t)$. The function h corresponds to the 2-Wasserstein distance between the discrete graph of the calculated trace $s_{cal}(t)$ and the observed trace $s_{obs}(t)$.

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

$$\eta = \frac{\tau}{A}, \quad (8)$$

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 amplitude discrepancy between observed and calculated data.

The adjoint source of the misfit function $f_{GSOT}[m]$ is computed from $\frac{\partial h}{\partial d_{cal}}$. We prove in Métivier et al. (2019) the following equality: Denoting σ^* the minimizer in (6), we have

$$\frac{\partial h}{\partial d_{cal}} = 2 \left(d_{cal} - d_{obs}^{\sigma^*} \right), \quad (9)$$

where

$$d_{obs}^{\sigma^*}(t_i) = d_{obs}(t_{\sigma^*(i)}) \quad (10)$$

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^* . Similarly as the KR approach, the solution of the problem (6) provides the information to compute both the misfit function and the adjoint source.

To solve (6) efficiently, we use the auction algorithm (Bertsekas and Castanon, 1989), dedicated to the solution of linear sum assignment problems (LSAP) such as (6). Despite a relatively high computational complexity in $O(N_i^3)$, it is quite efficient for small dense instances of LSAP. Resampling the data close to the Nyquist frequency yields such small scale problems making the GSOT feasible for realistic scale FWI applications, as we will see in the application presented here.

APPLICATION TO 3D OBC DATA FROM THE VALHALL FIELD

Geological context, data, and preprocessing

The Valhall field is located in the North Sea and has been in production since 1982 (Barkved et al., 2010).

The Valhall permanent 3-D wide aperture/azimuth seismic array (LoFS) has been installed in 2003 (Barkved et al., 2004). It covers a surface of 145 km². The acquisition is composed of 12 cables laying on the seafloor at 70 m equipped with approximately 2500 four-component receivers. Data are acquired by firing a seismic source every 50 m by 50 m at a depth of 5 m. The total recording time is 8 s.

The data have been preprocessed internally (Kamath et al., 2018). The processing consists of applying source-receiver reciprocity to decrease the number of sources to 2046, then removing noisy traces, de-spiking, and filtering the data to a 2.5-5 Hz frequency band.

While the KR approach can be formally applied to the entire 3D shot-gathers, it is easier, from implementation and computational cost point of view, to apply it to multiple 2D lines (Poncet et al., 2018). The Poisson's solver which is used within the KR approach also requires to work on a regular discretization of the shot gather. In this multiple 2D strategy, this implies a regular spacing between receivers, which is not guaranteed by the original acquisition. For this reason, we have interpolated the dataset on multiple 2D lines. For each time sample of the data, the interpolation is constructed by triangulating the input data, and on each triangle performing linear barycentric interpolation. We use this interpolated dataset with the three methods (L^2 , KR, and GSOT) for fair comparison even if it is only required for the KR approach.

FWI workflow

We consider FWI based on a time-domain VTI visco-acoustic solver (Yang et al., 2018a). The code is parallelized over shots with MPI and for finite-difference modeling and OT computations with OpenMP. The initial model for ε and δ are provided by AkerBP and are derived from reflection tomography (TOMO). The quality factor Q_p model used is simple: 1000 in the shallow water layer (70 m depth), 200 below. We consider two initial V_p models (Fig.1). The first is derived from TOMO, the second is a smoothed version of this TOMO model, using a Gaussian smoothing with 500 m correlation lengths in the three spatial directions. The initial density models ρ are derived from these models using Gardner's law.

To decrease the computational cost, we rely on source subsampling, with 120 sources used per cycle, each source being used only once in the whole process (Warner et al., 2013). Each cycle corresponds to 3 iterations of FWI, 51 iterations are required to use all the 2046 shot-gather. For a fair comparison between misfit functions, we use the same selection of sources, keeping a trace of the seed generating the pseudo-random source selection.

A source estimation is performed at the beginning of each cycle. The minimization of the misfit function is performed using the l -BFGS algorithm of the SEISCOPE toolbox (Métivier and Brossier, 2016). The memory l is set to 3 according to the previous settings.

FWI results

Starting from the TOMO model, we obtain satisfying and comparable results using the three different misfit functions (Fig.3a-l). This constitutes our reference to compare the improvement gained using optimal transport approaches. In the first slice at $z = 210$ m (Fig.3a-c), we observe the reconstruction of what appears to be channel-like deposits. We can notice that the imprint of the acquisition is smoothed using both KR and GSOT approaches, especially with GSOT. The second slice at $z = 980$ m (Fig.3d-f) is taken within the low-velocity anomaly. The KR approach exhibits little more artifacts; this could be related by the application of the method line by line. The GSOT results seem less noisy, the low-velocity anomaly being however less resolved than with the L^2 misfit function. In Figure 3g-l, constant x slices for $x = 3010$ m and $x = 3990$ m provide fairly similar results with the three misfit function, indicating that they have converged towards a similar model.

In Figure 3m-x, we present the V_p model obtained starting from the second initial V_p model. We still observe in Figure 3m-o that the imprint of the acquisition is less visible using OT. In the second slice, at $z = 980$ m, the L^2 result presents some high-velocity values in the center of the low-velocity anomaly (Fig. 3p). In Figure 3q and Figure 3r, the results from the KR and GSOT approaches do not exhibit this high-velocity anomaly. Still, both KR and GSOT reconstruct a clearer view of this low-velocity anomaly at this depth. The constant x slice in Figure 3s-u exhibits a clear example of local minima by cycle skipping with the L^2 misfit function. Instead of the low-velocity anomaly, the L^2 FWI has wrongly increased the velocity to fit the data (Fig.3s). This is not the case both with KR and GSOT. Besides, we observe that the sharp velocity contrast below the low-velocity anomaly is better reconstructed with GSOT than with KR (Fig. 3t). Figure 3v provides another view of the cycle skipped L^2 reconstruction with the presence of a high-velocity artifact. Again, both KR and GSOT reconstructions do not exhibit these artifacts. The anticline interface below the low-velocity anomaly is better continuously reconstructed with GSOT.

The data fit for final reconstructed V_p using L^2 , KR and GSOT starting from the second initial V_p model are presented in Figure 2. We can observe an improvement using KR and GSOT compared to L^2 in the area highlighted by the black arrow and the dashed black arrow. Both KR and GSOT can retrieve the shape of the gap present in the data with more accuracy than the L^2 .

From a computational cost point of view, the increase, compared to L^2 is between 15-20 % for KR and 10-15% for GSOT for this low-frequency application. Note that the OT overhead should decrease with frequency as the computational complexity for modeling is in ω^4 while being in ω^3 for GSOT.

CONCLUSION

The application of both KR and GSOT approaches to the Valhall 3D field data show how the two strategies help to mitigate the cycle skipping. The best results are obtained using the GSOT approach, which is expected, as it preserves the convexity property of the OT distance with respect to time shifts, which is not the case with KR. Besides, the computational cost increase compared to L^2 is limited. This is very encouraging and prompts us to investigate the interest of these techniques for other applications and in other contexts in the future (visco-elastic FWI, RWI).

Assessment of optimal transport based FWI: 3D OBC Valhall case study

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>) and 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.

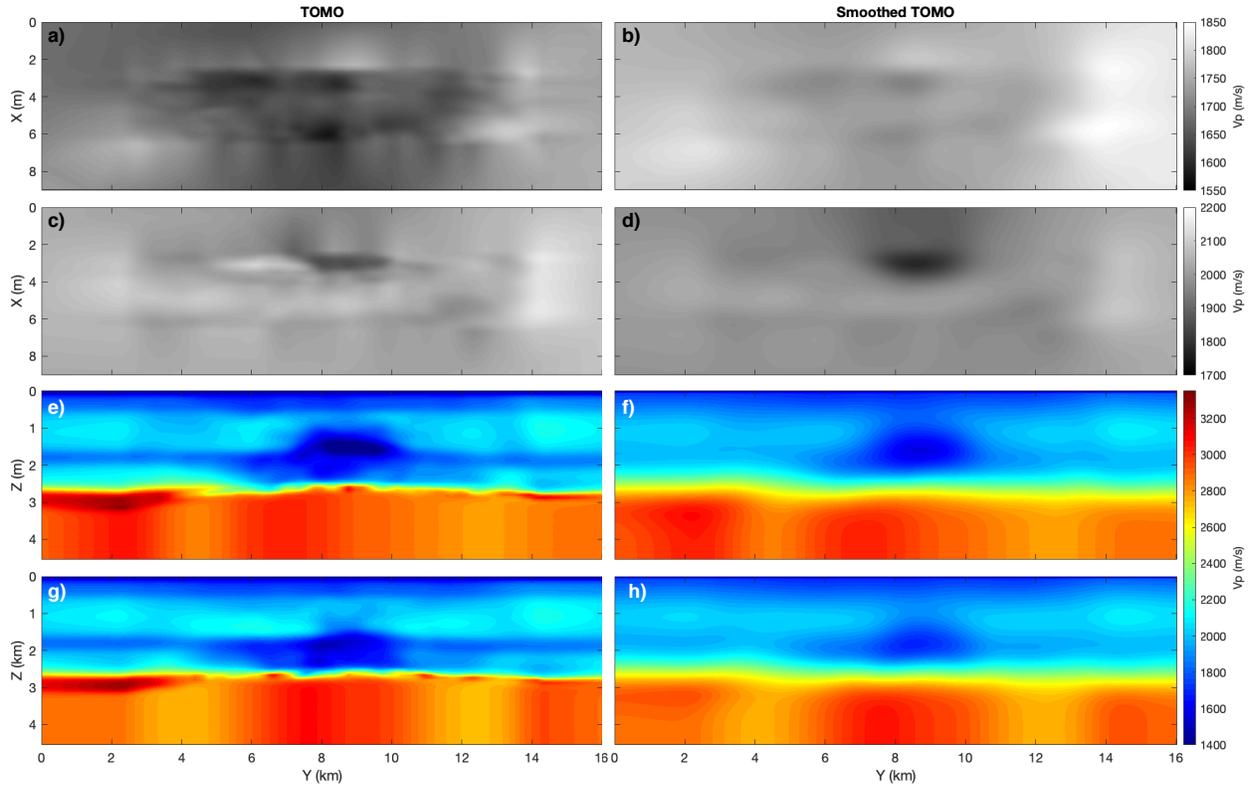


Figure 1: Initial V_p models. The first column (a,c,e,g) corresponds to TOMO V_p . The second column (b,d,f,h) corresponds to the Gaussian smoothed of TOMO V_p . The first line (a,b) in this V_p panel corresponds to a horizontal slice at $z = 210$ m, the second line (c,d) to a horizontal slice at $z = 980$ m. The third line (e,f) corresponds to a vertical slice at $x = 3010$ m and the last line (g,h) to a vertical slice at $x = 3990$ m.

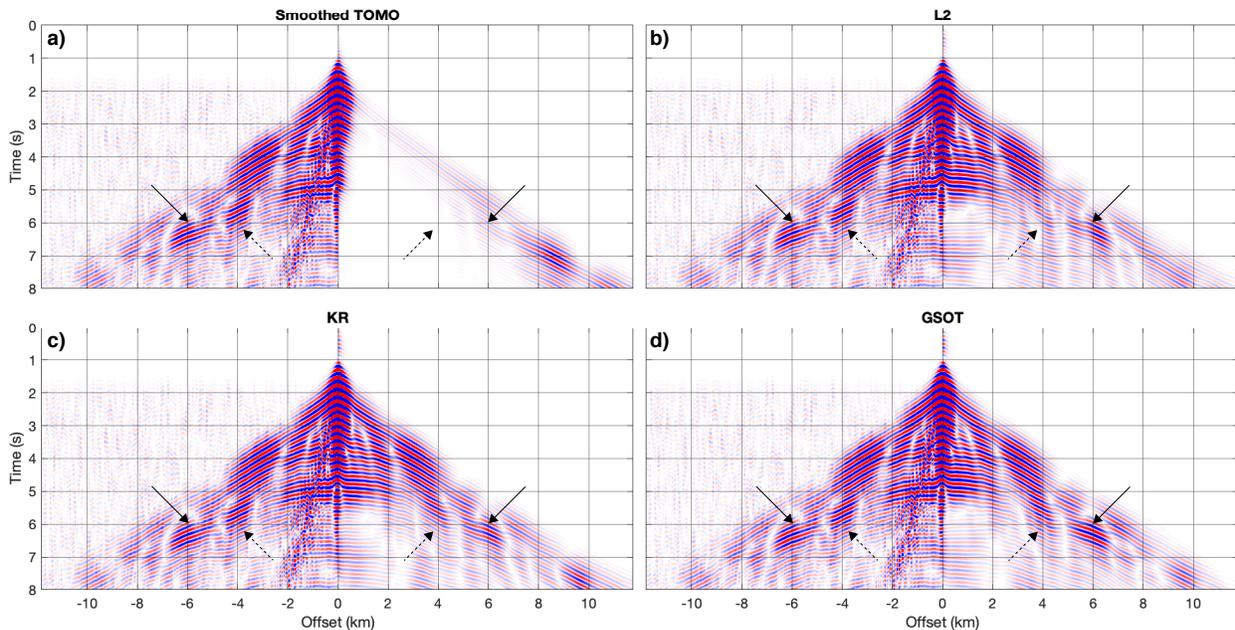


Figure 2: Common shot gather panel with recorded traces on the left and mirrored-modeled traces on the right. a) correspond to a CSG for smoothed TOMO initial model. a-c) correspond to a CSG inside the final reconstructed V_p after 51 FWI iterations for respectively L^2 , KR and GSOT.

Assessment of optimal transport based FWI: 3D OBC Valhall case study

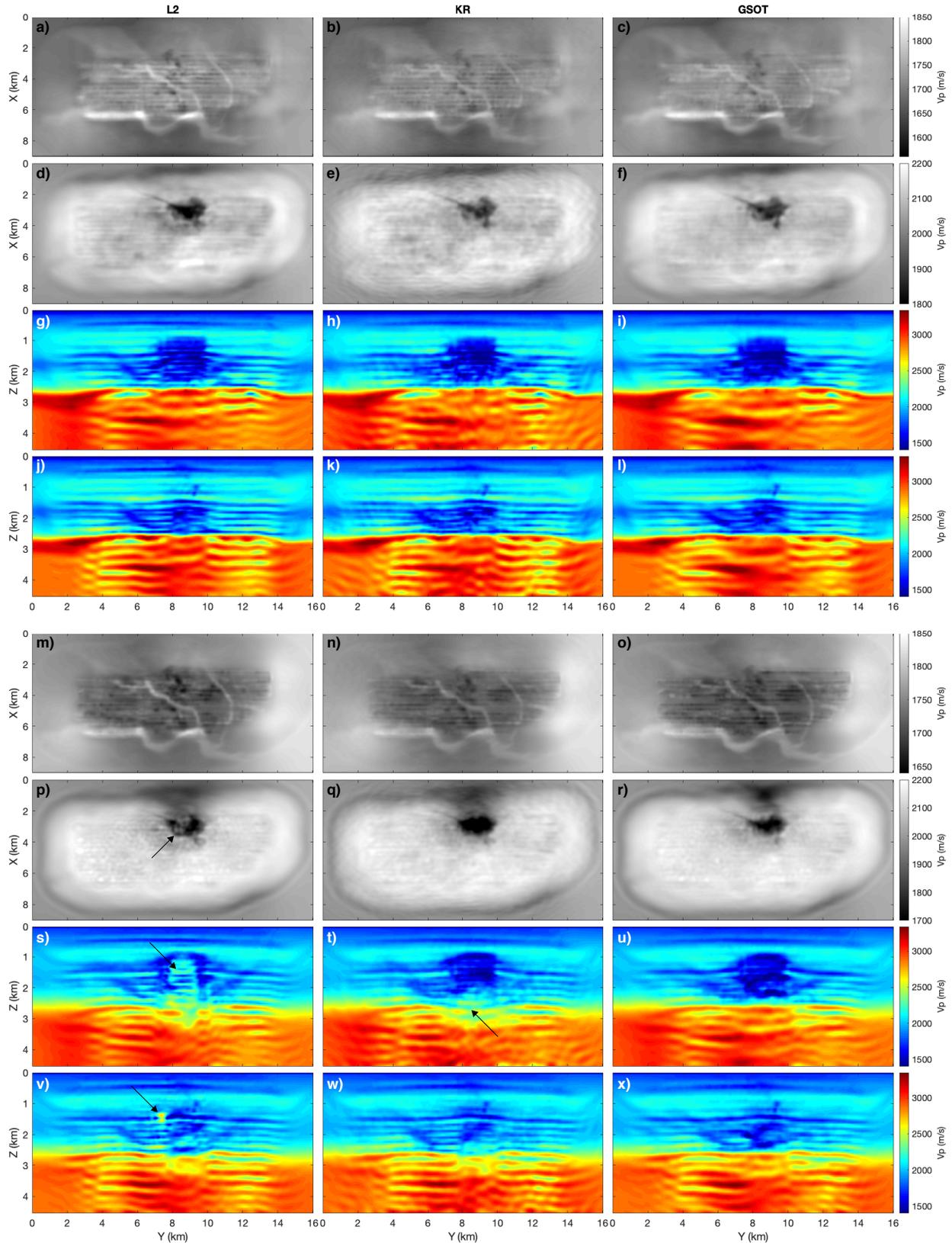


Figure 3: Slices of the 5 Hz FWI models for L^2 (first column), KR (second column) and GSOT (third column). The first four lines correspond to TOMO V_p initial model. The last four lines correspond to smoothed TOMO V_p initial model. a-c) and m-o) correspond to a horizontal slice at $z = 210$ m depth, d-f) and p-r) correspond to a horizontal slice at $z = 980$ m depth across the low-velocity zone. g-i) and s-u) correspond to inline vertical slices at $x = 3010$ m that pass through the low-velocity zone, and j-l) and v-y) at $x = 3990$ m pass near the periphery of the low-velocity zone.

REFERENCES

- Barkved, O., P. Heavey, J. H. Kommedal, J.-P. van Gestel, R. S. Ve, H. Pettersen, C. Kent, and U. Albertin, 2010, Business impact of full waveform inversion at Valhall: 80th Annual International Meeting, SEG, Expanded Abstracts, 925–929, doi: <https://doi.org/10.1190/1.3513929>.
- Barkved, O. I., J. H. Kommedal, and L. A. Thomsen, 2004, The role of multi-component seismic data in developing the Valhall field, Norway: 66th Annual International Conference and Exhibition, EAGE, Extended Abstracts.
- 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>.
- Bozdog, 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, doi: <https://doi.org/10.1093/gji/ggw356>.
- Combettes, P. L., and J.-C. Pesquet, 2011, Proximal splitting methods in signal processing, in *Fixed-point algorithms for inverse problems in science and engineering*: Springer, Springer Optimization and Its Applications 49, 185–212.
- 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>.
- Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2010, Full waveform tomography for radially anisotropic structure: New insights into present and past states of the Australasian upper mantle: *Earth and Planetary Science Letters*, **290**, 270–280, doi: <https://doi.org/10.1016/j.epsl.2009.12.003>.
- Kamath, N., R. Brossier, L. Métivier, and P. Yang, 2018, 3D acoustic/viscoacoustic time-domain FWI of OBC data from the Valhall field: 88th Annual International Meeting, SEG, Expanded Abstracts, 1093–1097, doi: <https://doi.org/10.1190/segam2018-2997017.1>.
- 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., and R. Brossier, 2016, The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication: *Geophysics*, **81**, no. 2, F11–F25, doi: <https://doi.org/10.1190/geo2015-0031.1>.
- Métivier, L., R. Brossier, Q. Merigot, 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. Merigot, E. Oudet, and J. Virieux, 2016a, Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion: *Geophysical Journal International*, **205**, 345–377, doi: <https://doi.org/10.1093/gji/ggw014>.
- Métivier, L., R. Brossier, Q. Merigot, E. Oudet, and J. Virieux, 2016b, 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>.
- Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom seismometer dataset: *First Break*, **28**, 71–78.
- Poncet, R., J. Messud, M. Bader, G. Lambaré, G. Viguier, and C. Hidalgo, 2018, FWI with optimal transport: A 3D implementation and an application on a field dataset: 80th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: <https://doi.org/10.3997/2214-4609.201801029>.
- Stopin, A., R.-E. Plessix, and S. Al Abri, 2014, Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman: *Geophysics*, **79**, no. 3, WA69–WA77, doi: <https://doi.org/10.1190/geo2013-0323.1>.
- 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>.
- Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: *Geophysics*, **78**, no. 2, R59–R80, doi: <https://doi.org/10.1190/geo2012-0338.1>.
- Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018a, A time-domain preconditioned truncated Newton approach to multiparameter visco-acoustic full waveform inversion: *SIAM Journal on Scientific Computing*, **40**, B1101–B1130, doi: <https://doi.org/10.1137/17m1126126>.
- Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018b, 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>.