

Optimal transport full-waveform inversion: from theory to industrial applications with examples from the Sultanate of Oman

J eremie Messud^{1*}, Diego Carotti¹, Olivier Hermant¹, Anna Sedova¹ and Gilles Lambar e¹.

Summary

The optimal transport problem was formulated more than 200 years ago to calculate the optimal way of transporting piles of sand. Due to interesting properties of its solutions with respect to shifts between the compared distributions, optimal transport has recently been adapted to full-waveform inversion to mitigate the cycle-skipping issue. Various formalisms have been proposed. Here we present an overview of these approaches, emphasizing more specifically the approach based on the bi-dimensional Kantorovich-Rubinstein norm, which has led to numerous successful full-waveform inversion applications. We illustrate these successes with two onshore case studies from the Sultanate of Oman.

Introduction

For the last decade, full-waveform inversion (FWI) has emerged as a central component of the subsurface velocity model building (VMB) toolbox (Virieux and Operto, 2009). Thanks to extensions of the original formulation (Virieux and Operto, 2009), it has become the most promising research direction for delivering multi-parameter high-resolution visco-elastic models of the subsurface. To date, it has demonstrated unequalled capabilities for producing subsurface models in areas probed by diving and reflected waves, enhancing the resolution when reflected waves are inverted (Huang et al., 2021).

However, numerous challenges remain. These are linked to the limitations of conventional FWI when the dataset lacks low frequencies (potential cycle-skipping issue) or long offsets (poor penetration of diving waves), or when inappropriate physics is used (acoustic instead of elastic modelling, see e.g. Perez-Solano and Plessix (2019)), or when different parameters need to be decoupled (Allemand et al., 2020), or when reflection data is used to recover the long-wavelength components of the model (Gomes

and Chazalnoel, 2017). In these contexts, FWI still needs to be used within a VMB toolbox, in combination with other tools such as ray-based tomography which can provide an initial subsurface model (velocity and anisotropy) and a final flattening of common image gathers (CIGs) (Allemand et al., 2019, 2020).

Conventional FWI is based on the least-squares (LS) misfit, which can easily trap the solution into a local minimum during the local optimization process, especially with the oscillatory nature of seismic data (cycle-skipping) (Virieux and Operto, 2009). This motivates the search for alternative misfits exhibiting more convexity with respect to kinematic information, i.e. more sensitivity to time shifts between the events in the observed and computed traces, together with a relaxed sensitivity to the amplitude information. Hence, many proposed alternative misfits are explicit functionals of the time shifts estimated between the events of the observed and computed traces (Jiao et al., 2015; Warner and Guash, 2016; Wang et al., 2016; Zhang et al., 2018).

In this context, misfits based on the optimal transport (OT) framework gained attention in the field of geophysics (Engquist

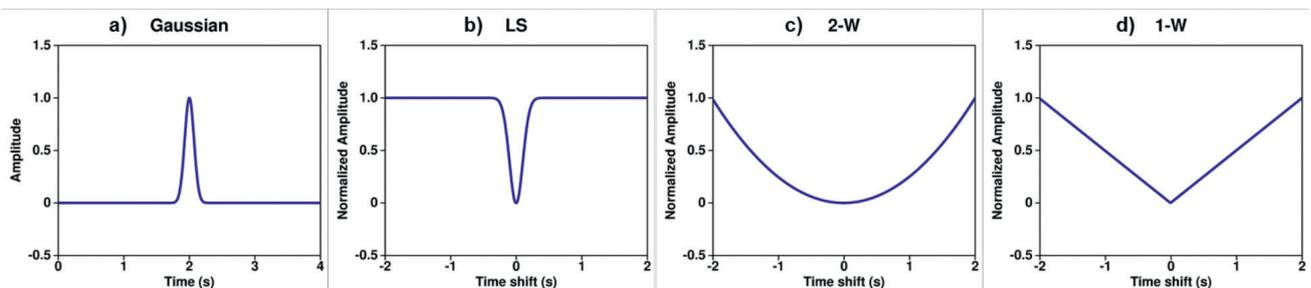


Figure 1 OT-based misfits for two shifted PDFs. (a) A PDF profile $f(t)$ (a gaussian 'wavelet'). Value of the misfit depending on the time shift, s , between $f(t)$ and $g(t) = f(t-s)$ for: (b) Least Square, (c) the 2-Wasserstein distance and (d) the 1-Wasserstein distance.

¹ CGG

* Corresponding author, E-mail: jeremie.messud@cgg.com

DOI: 10.3997/1365-2397.fb2021090

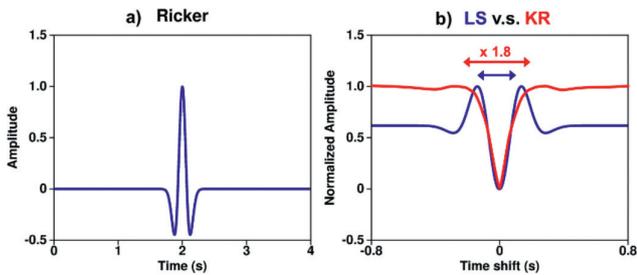
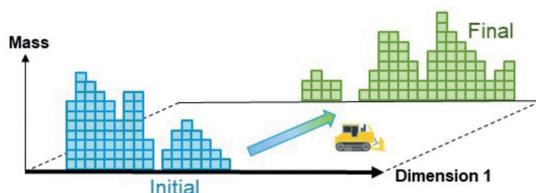


Figure 2 Comparison of the KR norm and LS values for 2 traces containing shifted Ricker wavelets. (a) A wavelet profile $f(t)$ (a Ricker); (b) Value of the misfit depending on the time shift, s , between $f(t)$ and $g(t) = f(t-s)$ for LS and the KR norm, showing that the width of the global minimum valley almost doubles for KR.

et al., 2016; Métivier et al., 2016a). OT originally aims to find the optimal way to transport piles of sand by minimizing the ‘expended energy’ (Monge, 1781). It can be based on various formulations usually related to Wasserstein distances, which defines precisely this optimal cost based on the choice of a geometrical distance (for example Euclidian distance for 2-Wasserstein distance in the ‘data coordinate space’ or ‘taxi cab distance’ for the 1-Wasserstein distance) (see equations (4) and (5)). While OT has many appealing features for the comparison of probability density functions (PDFs) containing shifted (or translated) ‘events’ (Figure 1), it also implies conditions that are difficult to satisfy when considering seismic traces. Indeed, unlike PDFs, seismic traces are signed and do not necessarily have the same ‘mass’. Therefore, adaptations had to be introduced for the application of OT to seismic acquisition.

A first strategy proposed by Engquist and Froese (2014), Engquist et al. (2016), Yang and Engquist (2018), and Yang et al. (2018) is based on a 2-Wasserstein distance, and ad-hoc positive transformations and normalizations of the traces (adding a constant, taking the envelop, comparing separately the positive and negative parts of the traces, etc.). As none of these transformations proved to be fully satisfactory, a second family of approaches was proposed by Métivier et al. (2016a, 2016b, 2016c). Inspired by the 1-Wasserstein distance, they used the so-called Kantorovich-Rubinstein (KR) norm. A transformation of seismic traces is no longer required, while the resulting misfit convexity with respect to time-shifts remains significantly improved compared to the convexity of the LS misfit (Figure 2). They also proposed

Mono-dimensional OT



Bi-dimensional OT

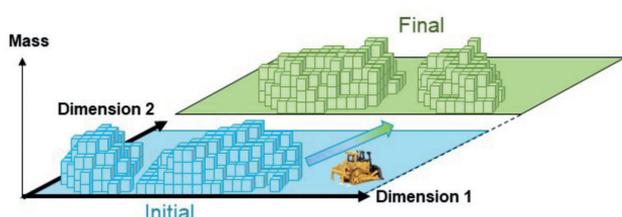


Figure 3 Mapping between some initial and final piles of sand. Schematic mono-dimensional (top) and bi-dimensional (bottom) OT cases.

an efficient approach for the numerical implementation of the KR norm, allowing for industrial use of the bi-dimensional version that can compare a set of observed and computed traces (and not only single traces) (Figure 3).

The use of the bi-dimensional KR norm within FWI has led to numerous successful industrial applications, e.g. Poncet et al. (2018), Messud and Sedova (2019), Sedova et al. (2019), Hermant et al. (2019), Carotti et al. (2020) and Hermant et al. (2020). For these case studies, an interesting reduction in sensitivity to cycle-skipping compared to LS FWI results is observed, together with an improved structural consistency of the inverted models. The latter can be related to the use of bi-dimensional implementation of the KR norm. As we will illustrate, these features can be interpreted in the light of the corresponding ‘adjoint-source’ or data-space gradient of the misfit, which is back-propagated before being converted into the model-space gradient (Virieux and Operto, 2009). Compared to the LS FWI adjoint-source, the KR FWI adjoint-source exhibits enhanced low frequencies and better amplitude balancing, hence its robustness to cycle-skipping. Also, when the bi-dimensional KR formulation is used, this produces an additional increase in coherency along the events (Messud and Sedova, 2019; Messud et al., 2021).

In this paper we briefly review the basis of bi-dimensional KR FWI, emphasizing the adjoint-source features. We then present two applications of land datasets from the Sultanate of Oman where these features are key for solving the VMB challenges.

Optimal transport full-waveform inversion

FWI aims to match the observed data $d_{obs}(s,r,t)$ with the data $d[m](s,r,t)$ computed using a given subsurface model $m(x)$. s,r,t represent the data space parameters, i.e. the shot and receiver positions and the time, which index the data space D . x represents the position in the subsurface.

Various misfits $J(d_{obs}, d[m])$ have been proposed for assessing the match between observed and computed data. Solving the FWI problem aims to minimize the chosen misfit which, due to the cost of calculating $d[m]$ (i.e. of the full wave modelling), is solved most of the time by an iterative local optimization scheme.

In this context, the model-space gradient of the misfit, $\frac{\partial J}{\partial m}(x)$, is

a key component of the scheme. Irrespective of the misfit used, the model-space gradient can be broken down into

$$\left(\frac{\partial J}{\partial m}\right)^t = \left(\frac{\partial d}{\partial m}\right)^t \left(\frac{\partial J}{\partial d}\right)^t \tag{1}$$

where the first term $\frac{\partial d}{\partial m}(s,r,t;x)$ is the kernel of the first-order

Born operator, and the second term $\frac{\partial J}{\partial d}(s,r,t)$ is the adjoint-

source which ‘lives’ in D . Equation (1) shows that changing the misfit affects the model-space gradient only through the adjoint-source. Looking at adjoint-sources is therefore a useful way of analysing the impact of the misfit choice.

In practical terms, computing the model-space gradient is performed by (Virieux and Operto, 2009):

- injecting the adjoint-source at the receiver positions,

- back-propagating the resulting wavefield in the subsurface model,
- cross-correlating (at time $t = 0$) with the wavefield forward-propagated from the source position.

The simplest misfit, used in conventional FWI, is the LS misfit,

$$J_{L_2}(d_{obs}, d[\mathbf{m}]) = \frac{1}{2} \|d_{obs} - d[\mathbf{m}]\|_2^2, \quad (2)$$

where $\|\cdot\|_2$ denotes the L_2 norm that induces a distance between vectors in the data space D . In the case of the LS misfit, the adjoint-source equals the data ‘residual’:

$$\frac{\partial J_{L_2}}{\partial d[\mathbf{m}]} = \Delta d[\mathbf{m}] = d_{obs}(\mathbf{x}) - d[\mathbf{m}]. \quad (3)$$

OT-based misfits can be interesting for FWI as they exhibit a wider convexity with respect to time shifts. Various formulations of OT applied to FWI exist, and all are related to Wasserstein distances. The p-Wasserstein distance for two PDFs $\rho_1(s, r, t)$ and $\rho_2(s, r, t)$ is

$$J_{w_p}^p(\rho_1, \rho_2) = \min_{\mathcal{T}} \int d\mathbf{e} \, c_p^p(\mathbf{e}, \mathbf{E}_{\rho_1, \rho_2}(\mathbf{e})) \rho_2(\mathbf{e}) \quad (4)$$

subject to the constraint

$$\mathbf{E}_{d_1, d_2} \in \mathcal{T} \text{ the set of maps on } \mathbf{e} \text{ that rearrange } \rho_2 \text{ into } \rho_1. \quad (5)$$

$\mathbf{e} = (s, r, t)$ denotes a data coordinate space vector and c_p denotes the L_p distance between vectors in the data coordinate space. Equations (4) and (5) seek at the minimum cost to transport mass from ρ_2 to ρ_1 from the c_p^p point of view. Note that with seismic data being recorded in space and time, these dimensions need to be normalized within the Lp distance with a ratio involving an apparent velocity (Messud and Sedova, 2019).

An important aspect of the OT formalism is that ρ_1 and ρ_2 are required to be PDFs, i.e. positive with equal masses, implying that OT cannot be readily applied to seismic data. Yang et al., (2018) and Qiu et al., (2017) proposed using ad-hoc transformations of the observed and computed data to make them positive with the same mass. They chose the $p=2$ case in eq. (4), i.e. the squared 2-Wasserstein distance. Since solving equations (4) and (5) in the multi-dimensional data coordinate space case is computationally demanding, most of their applications consider a mono-dimensional coordinate space, where \mathbf{e} is parameterized by time only (source and receiver positions being fixed). Then, equations (4) and (5) are solved for each trace independently.

Bearing in mind the properties of the 1-Wasserstein distance, Métivier et al. (2016a, 2016b, 2016c) chose $p=1$ in equation (4). As proposed by Métivier et al., (2016a), by reformulating the problem using the Kantorovich–Rubinstein (KR) dual formulation and adding a bounding constraint it is possible to use the seismic data directly without any transformation. Rather than minimizing equation (4) for $\mathbf{E}_{d_1, d_2} \in \mathcal{T}$ we look for a λ -bounded 1-Lipschitz function, $\varphi(\mathbf{e})$, maximizing the so-called KR norm

$$J_{KR}(d_{obs}, d[\mathbf{m}]) = \max_{\varphi} \int d\mathbf{e} \, \varphi(\mathbf{e}) \Delta d[\mathbf{m}](\mathbf{e}) \quad (6)$$

subject to

$$|\varphi(\mathbf{e}_1) - \varphi(\mathbf{e}_2)| \leq c_1(\mathbf{e}_1, \mathbf{e}_2) \quad \text{and} \quad |\varphi(\mathbf{e})| \leq \lambda. \quad (7)$$

φ is the solution of the maximization problem and can be demonstrated to represent the adjoint-source (Métivier et al., 2016b; Messud et al., 2021). The first constraint on φ in equation (7) is called 1-Lipschitz for the metric c_1 . It imposes that changes in φ are sufficiently slow, which emphasizes low frequencies in φ . The second constraint, controlled by the parameter λ , makes it possible to stabilize the problem when seismic traces are used directly. This approach does not require any transformation of traces and has the advantage of an efficient numerical implementation. Métivier et al. (2016b, 2016c) proposed using the Simultaneous Descent Method of Multipliers to iteratively resolve equations (6) and (7), which is efficient even for a bi-dimensional data coordinate space case, where \mathbf{e} is parameterized by time and a chosen receiver ‘line direction’ (the best sampled one; for instance, the inline direction). Note that, to our knowledge, no bi-dimensional numerical implementation had been developed for industrial applications with approaches based on the 2-Wasserstein distance for the computational cost reason mentioned above.

While the LS adjoint-source is the residual $\Delta d(s, r, t)$, equation (3), the KR adjoint-source is the function $\varphi(s, r, t)$ which maximizes the integral in equation (6). It is obviously a function of the residual and tends to be a piecewise linear function especially in areas of non-null signal (see Figure 4 for the mono-dimensional case and Messud et al. (2021) for extensive details about the impact of the choice of parameters c_1 and λ).

Figure 5 compares the LS adjoint-source to the corresponding bi-dimensional KR adjoint-source for a marine data set (with a mute applied as commonly performed in FWI applications). As for mono-dimensional KR, the bi-dimensional KR adjoint-source exhibits more low frequencies and more balanced amplitudes between the events. Figure 6 compares, in another marine case, the LS adjoint-source to the mono and bi-dimensional KR adjoint-sources. In the corresponding data, random noise leads to deterioration in the continuity of the LS adjoint-source. Even if mono-dimensional KR shows more low frequencies and enhanced amplitude balancing, it does not exhibit an improved lateral coherency along the events.

In contrast, bi-dimensional KR further improves the lateral coherency and even the amplitude balancing as we have shown and detailed in Messud and Sedova (2019) and Messud et al. (2021). This is an important point, especially for applications to land data where the signal-to-noise ratio (SNR) is often low. All these features enhance the kinematic information in the adjoint-source, which is important to overcome the non-convexity issue and especially cycle-skipping, and to explain the success of bi-dimensional KR FWI in industrial applications.

From now on, bi-dimensional KR FWI will be denoted by OT FWI to make it clear that we are dealing with a version of OT.

Land data example from the north of the Sultanate of Oman with challenging structural complexity

OT FWI is now applied to a land dataset from the north of the Sultanate of Oman. The geological context is a complex strike-slip fault system, causing sharp lateral velocity variations in the

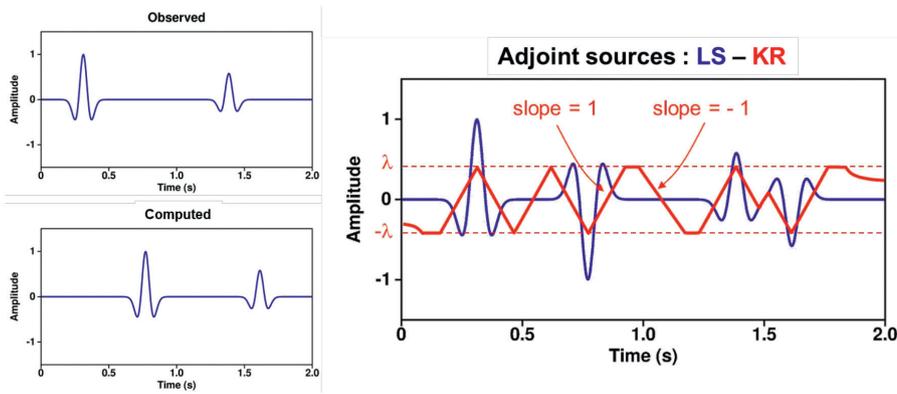


Figure 4 Comparison of LS and KR adjoint-sources for two traces (observed and computed) with shifted Rickers.

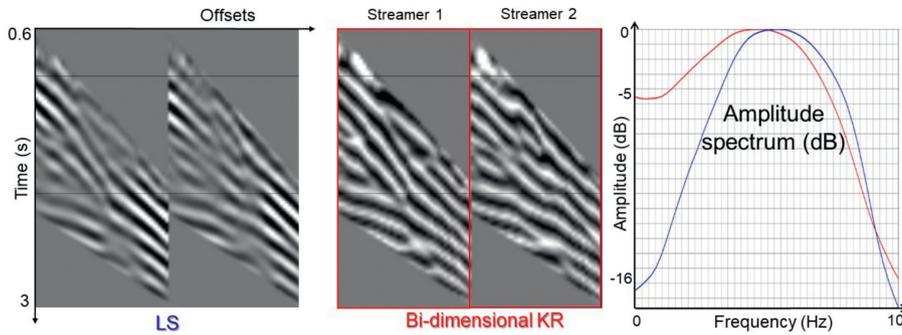


Figure 5 Marine data adjoint-sources at 6 Hz (with a mute). LS (left), bi-dimensional KR (middle) and corresponding amplitude spectra (right).

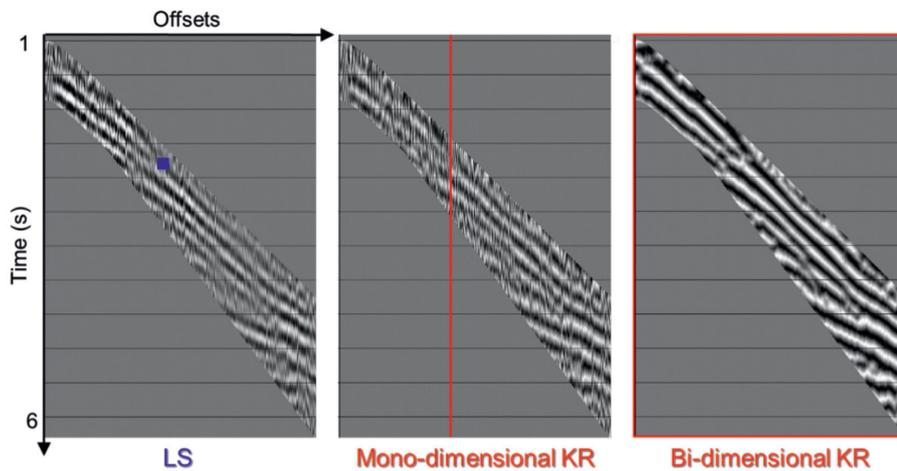


Figure 6 Marine data adjoint-sources at 4 Hz (with a mute). LS (left), mono-dimensional KR (middle) and bi-dimensional KR (right). The blue point indicates that LS performs a sample-wise comparison (left), the red line indicates that mono-dimensional KR performs a trace-by-trace comparison (middle), and the red box indicates that bi-dimensional KR performs a panel-by-panel (here common streamer) comparison (right).

faulted area. Until recently, imaging beneath the faulted area was challenging due to the wavefield complexity induced by the lateral velocity variations. Previous VMB attempts using ray-based tomography failed due to difficult horizon interpretation and residual move-out picking on migrated gathers.

The seismic survey of interest was acquired in 2010 using an acquisition design developed by PDO and Shell (Mahrooqi et al., 2012), with distance-separated simultaneous sweeping (DS³, Bouska, 2010). The broadband 3D vibroseis survey used a nine-second sweep starting from 1.5 Hz, with full-azimuth and large offsets that were processed up to 13 km. The shot sampling is dense and regular, with a 50 m increment in both x and y directions. The receiver lines are spaced 250 m apart, with 25 m spacing between receiver stations (Figure 7a).

The workflow implemented for FWI is adapted from Sedova et al. (2017), with more details in Hermant et al. (2020). It uses two different sets of input data, as shown in Figure 7b. The long-offset dataset, optimizing the penetration

of diving and post-critical waves, is used to update the velocity model from a starting frequency of 2 Hz and then up to 9 Hz. The second dataset, optimized for reflections, is then used to update the velocity model from 6 Hz up to 16 Hz. Figure 8 gives additional details.

The results obtained by OT FWI led to an accurate update of the velocity model, up to 16Hz, with an unprecedented resolution and imaging uplift in this area of the Sultanate of Oman. Both lateral and vertical velocity variations are resolved in the graben area (Figure 9) and the imaging is clearly improved, especially when using RTM imaging (Figure 10).

To better evaluate the uplift brought by OT FWI, a comparison with LS FWI using the same data set and workflow settings was made. As expected, OT FWI results in a more structurally consistent velocity model, providing improved imaging of the deep reflectors and better focusing of the major fault (Figure 11). Cross-correlation QCs between observed data and data computed with the updated velocity, for both the

LS FWI and OT FWI models, show that the data similarity is greatly enhanced by OT FWI (Figure 12).

Land data example from the south of the Sultanate of Oman with challenging near-surface geology

The next example is from the south of the Sultanate of Oman, known for its complex near-surface geology. The presence of strong lateral and vertical velocity contrasts creates elastic effects in the recorded surface seismic data. Although these effects may prevent acoustic FWI from converging (Perez-Solano and Plessix, 2019), we show that under a certain set of conditions, a geologically plausible acoustic velocity model can be inverted up to 7 Hz. The fundamental ingredients for our results are a

dedicated pre-processing sequence prior to FWI, and the use of OT FWI (Sedova et al., 2019; Hermant et al., 2019; Messud and Sedova, 2019), which help to partially mitigate the elastic effects and exploit the full potential of acoustic FWI.

The data were acquired with a vibrator sweep starting at 1.5 Hz and with standard 10 Hz vertical single-component geophones. The dataset covers an area of approximately 700 km² with a maximum offset of 12 km in the east-west direction and 2 km in the north-south direction. The area is known for poor SNR, as illustrated in Figure 13 with a receiver gather example on which barely any signal can be identified apart from the ground roll. After application of the dedicated signal processing workflow for land FWI presented earlier (including surface-wave attenuation, joint low-rank sparse inversion and

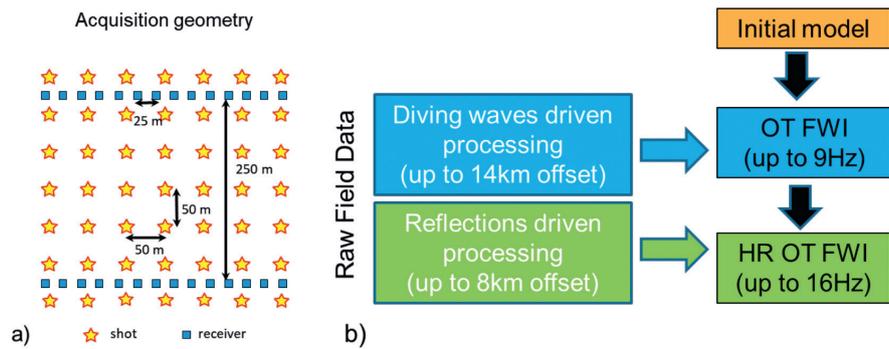


Figure 7 (a) Schematic description of the acquisition layout (b); Workflow implemented for high-resolution (HR) OT FWI using two input datasets, one for diving and post-critical waves, and one for reflected waves.

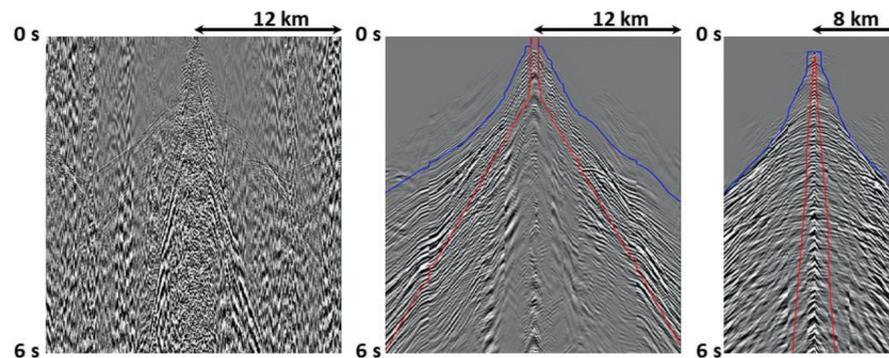


Figure 8 Receiver gather, raw data (left); Receiver gather after dedicated processing to enhance diving and post-critical waves (centre); Receiver gather after a conventional processing sequence for reflected waves and additional de-noise applied before FWI (right). The blue and red lines indicate the outer and inner mutes used in FWI, respectively.

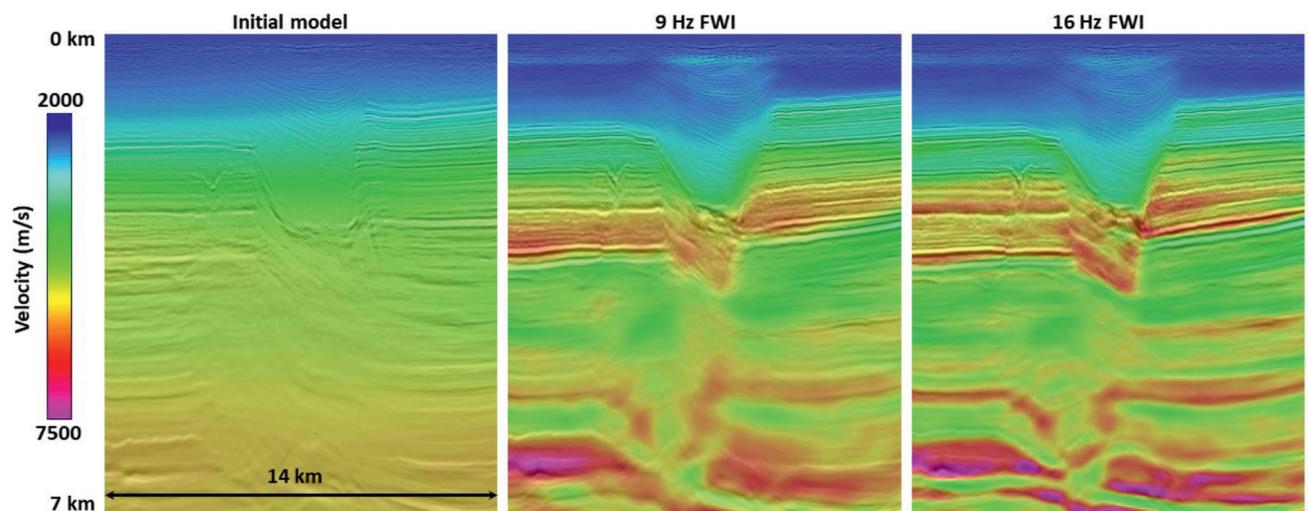


Figure 9 Velocity model and corresponding Kirchhoff PreSDM image overlaid. Initial model (left), model after 9 Hz OT FWI using diving and post-critical waves (centre), and 16 Hz OT FWI model using reflected waves (right).

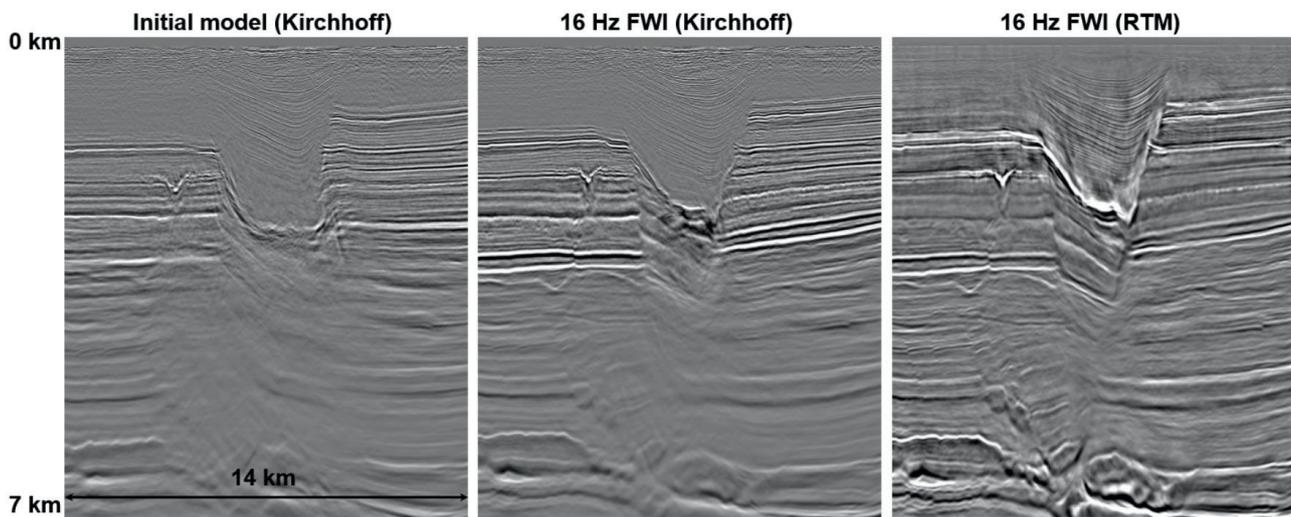


Figure 10 Imaging results. Kirchhoff PreSDM with initial model (left) and with 16 Hz OT FWI model (centre); RTM with 16 Hz OT FWI model (right). Note the imaging uplift with RTM, when using the high-resolution OT FWI model.

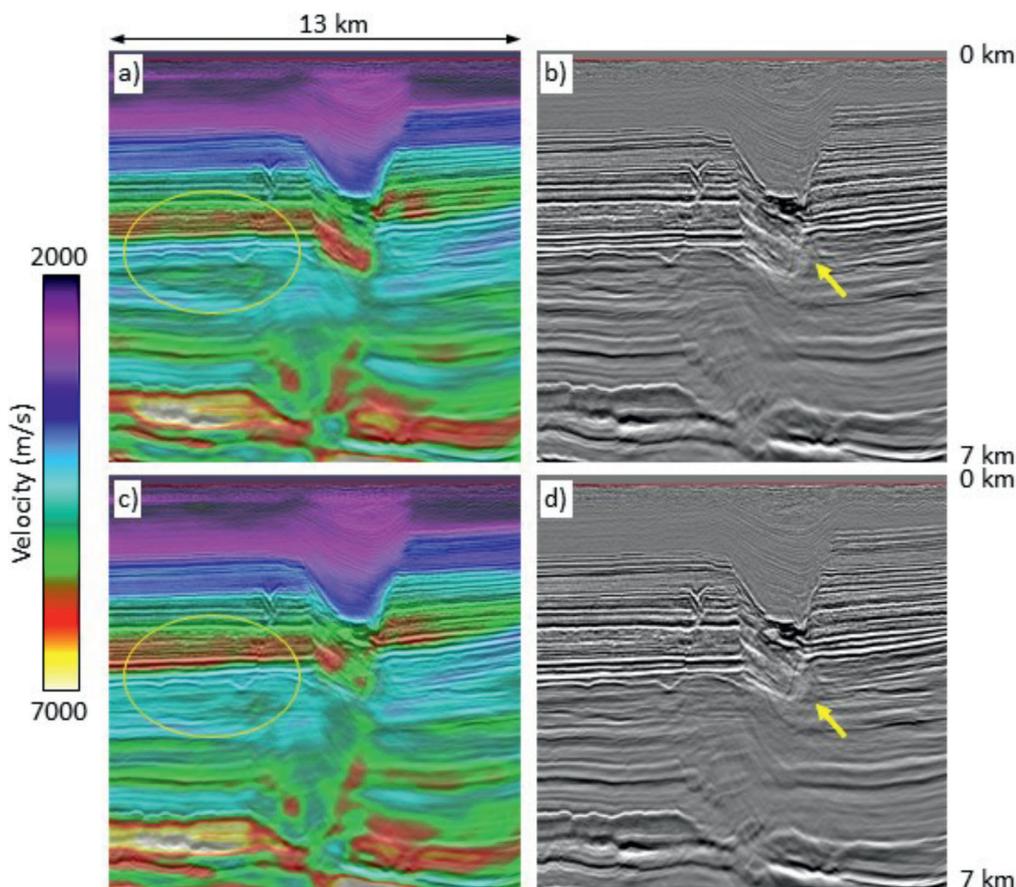


Figure 11 LS and OT FWI comparison. (a) LS FWI velocity model superimposed on Kirchhoff PreSDM (K-mig); (b) K-mig image with LS FWI model; (c) OT FWI velocity model superimposed on Kirchhoff PreSDM image; (d) K-mig image with OT FWI.

3D linear noise filtering), the SNR improved, revealing the diving wave information. Additional analysis of the quality of the low frequencies using the phase stability QCs shown in Figure 14 helped to determine the optimal starting frequency for FWI.

It is well established that the starting velocity model is a key element for FWI. In these challenging SNR conditions, it was difficult to obtain an accurate model to initiate the FWI updates. It was decided to start with a smoothed version of a legacy tomographic model. The smooth initial model and the poor quality

of the low frequencies suggested the use of OT FWI, to avoid cycle-skipping issues and provide a better structural consistency of the model with the geology.

The results of the 7Hz OT FWI are shown in Figure 15 and compared to the result obtained with the legacy tomographic model. Without any a-priori information about the location of the shallowest fast layers, OT FWI was able to recover the Rus and Natih formations (yellow and green arrows, respectively, in Figure 15b). This 7 Hz FWI model resulted in imaging improvements, compared to the tomographic model, down

to 3 km (Figure 16). Events at the Natih level (yellow arrow in Figure 16d) are better flattened and more geologically plausible.

A comparison of observed and computed data (with the 7 Hz OT FWI updated model) also shows a good agreement, despite the poor match produced by the initial model (Figure 17). Perez-Solano and Plessix (2019) have shown that, except for the

first loop, there was a noticeable amplitude and phase mismatch between the elastic and acoustic modelling in this area. We show here that, despite the amplitude and phase mismatch between the observed data (elastic) and the computed acoustic data, our optimized acoustic OT FWI was able to match the kinematics of several diving waves, providing us with confidence in the quality of the inverted velocity model.

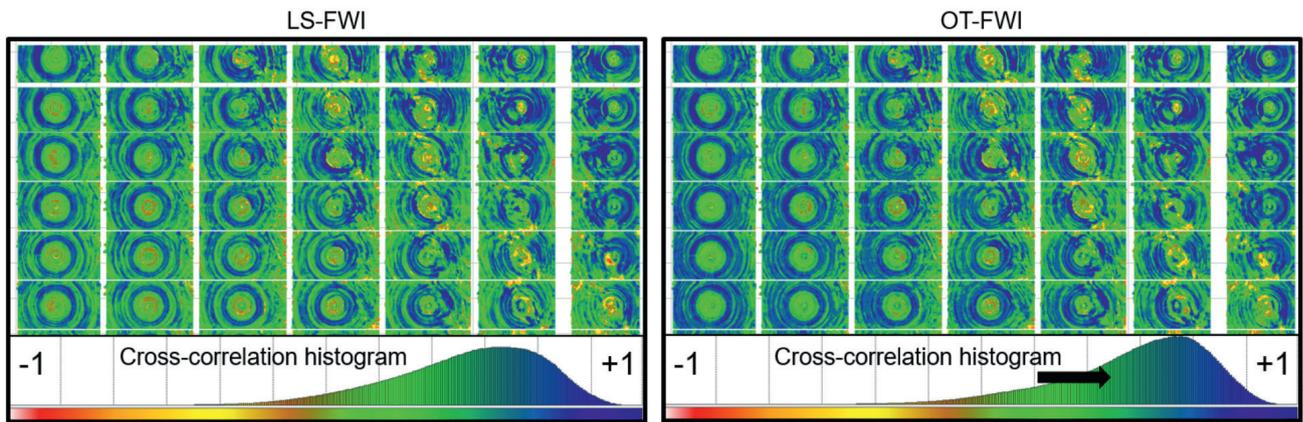


Figure 12 Cross-correlation QCs between observed data and data computed with the updated model. OT FWI cross-correlation (right) shows higher values than LS FWI cross-correlation (left), showing a better similarity measurement.

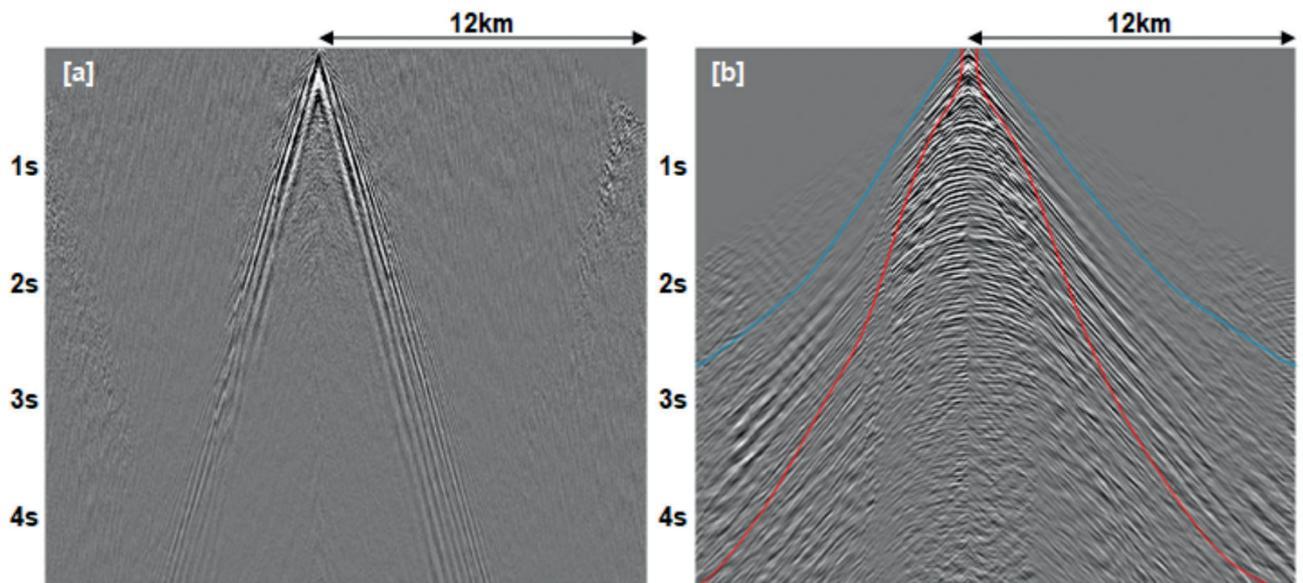


Figure 13 a) Receiver gather before denoise; b) Receiver gather after dedicated processing to enhance diving and post-critical waves. The blue and the red lines define the mute applied for FWI. Note that a gain of 20dB was applied in (b) with respect to (a).

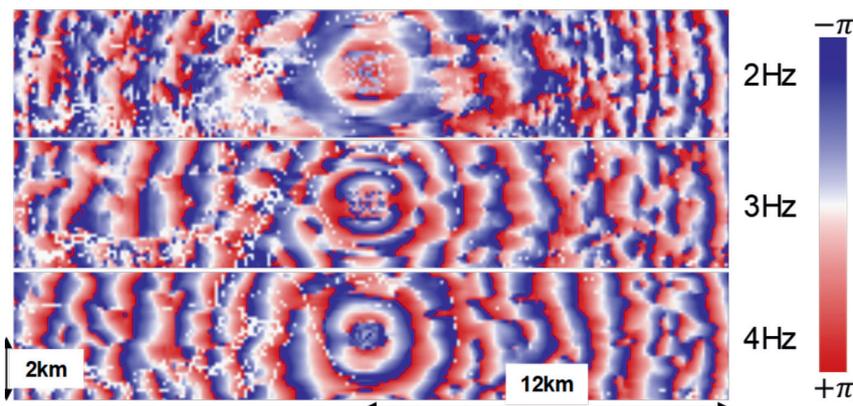


Figure 14 Phase stability QC for three different frequencies (2, 3 and 4 Hz) for the receiver gather shown in Figure 13b. This QC helped to choose the minimum reliable frequency for which we expect continuity of the signal (2.5Hz), to start FWI..

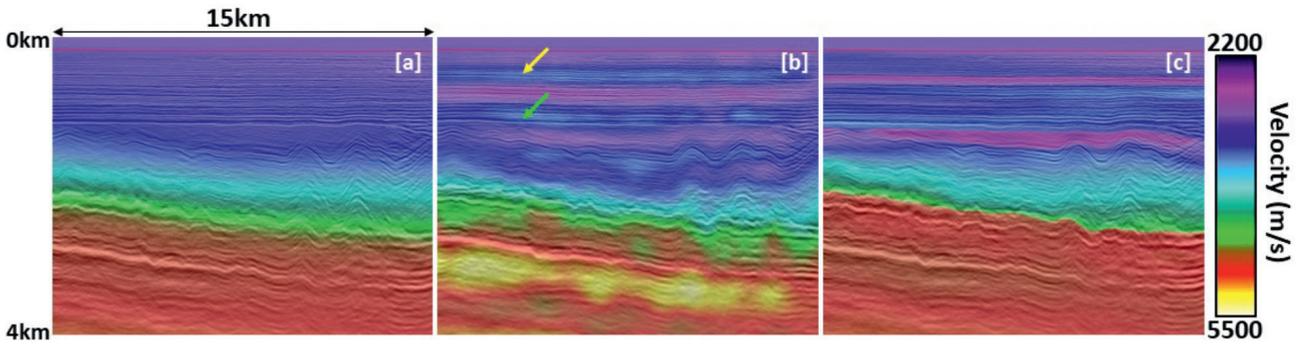


Figure 15 Velocity models superimposed with the corresponding K-mig stacks. a) Initial velocity model for OT FWI; b) OT FWI velocity model revealing Rus and Natih formations; c) Legacy tomography velocity model.

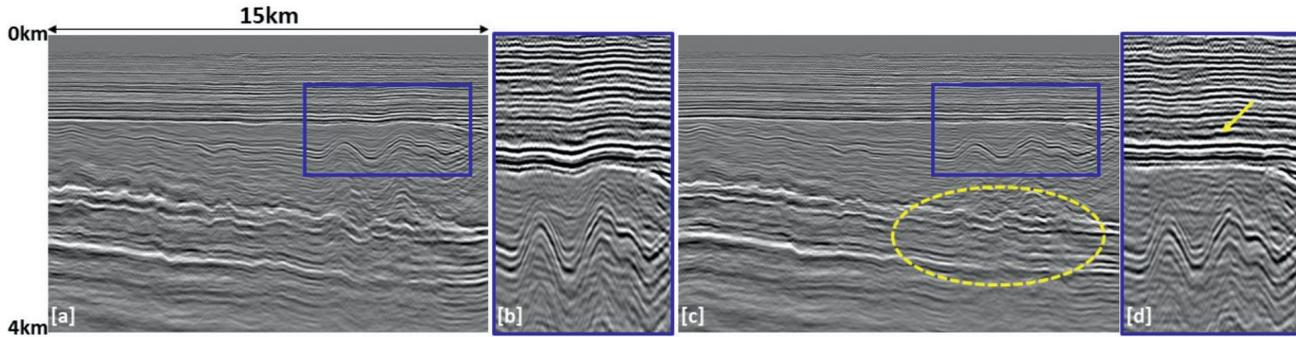


Figure 16 K-mig stacks using a) the legacy tomographic model and b) its zoom at Natih level; c) the OT FWI model and d) its zoom showing improved flattening of the Natih formation (yellow arrow) and better continuity for the deeper horizons (yellow ellipse).

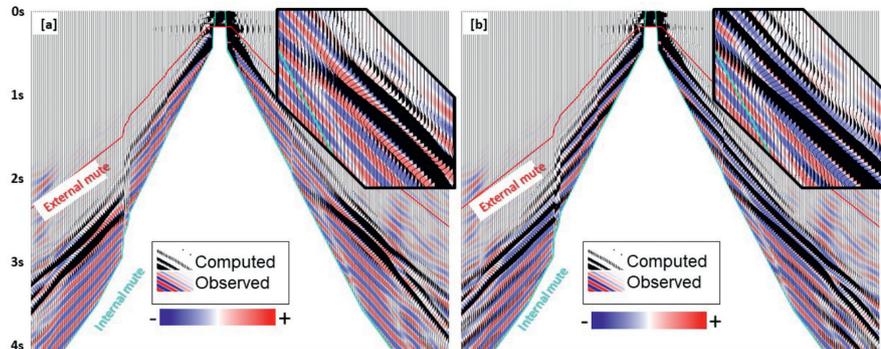


Figure 17 (a) Computed data (wiggles) superimposed on the observed data (blue-white-red) for the initial model and (b) for the OT FWI updated model at 7 Hz. The computed data matches the observed data when the black wiggles are over the red loops.

Conclusion

Optimal transport full-waveform inversion has fully demonstrated its advantages over least-squares full-waveform inversion in terms of non-convexity and cycle-skipping mitigation, especially when based on the Kantorovich-Rubinstein norm formulation. The associated adjoint-source is characterized by enhanced low frequencies and amplitude balancing. Its Kantorovich-Rubinstein norm bi-dimensional implementation (in data coordinate space) also improves the lateral coherency of the events in the adjoint-source. The land case studies from the Sultanate of Oman shown in this paper illustrate the advantages of these features which help to enhance the kinematic information present in the adjoint-source. This has made optimal transport full-waveform inversion a valuable approach for subsurface model building in land areas (Messud and Sedova, 2019; Sedova et al., 2019; Hermant et al., 2019; Carotti et al., 2020; Hermant et al., 2020).

Acknowledgements

The authors would like to thank the Ministry of Energy and Minerals of the Sultanate of Oman and Petroleum Development

Oman for permission to publish the data examples in Figures 7 to 17. They also thank CGG for its support of the publication of this work.

References

Allemand, T., Sedova, A., Lambaré, G., Grenié, D. and Guillaume, P. [2020]. Full Waveform Inversion in an Anisotropic Earth: A Practical Workflow. 82nd EAGE Annual Conference & Exhibition, Extended abstract, Fr_Dome1_14.

Allemand, T., Lopes, L., Lambaré, G. and Leblanc, O. [2019]. High resolution Full Waveform Inversion: Two Pitfalls and a Remedy. 81st EAGE Annual Conference & Exhibition, Extended abstract, Tu_R08_07.

Bouska, J. [2010]. Distance separated simultaneous sweeping, for fast, clean, Vibroseis acquisition. *Geophysical Prospecting*, 58, 123-153.

Carotti, D., Hermant, O., Masclet, S., Reinier, M., Messud, J., Sedova, A., and Lambaré, G. [2020]. Optimal transport full waveform inversion - Applications. 82nd EAGE Conference and Exhibition, Expanded Abstracts, Th Dome1 17.

- Engquist, B. and Froese, B. [2014]. Application of the Wasserstein metric to seismic signals. *Communications in Mathematical Sciences*, **12**(5).
- Engquist, B., Froese, B.D. and Yang, Y. [2016]. Optimal transport for seismic full waveform inversion. *Communications in Mathematical Sciences*, **14**, 2309-2330.
- Gomes, A., and Chazalnoel, N. [2017]. Extending the reach of FWI with reflection data: Potential and challenges. 87th SEG Annual International Meeting, Expanded abstract, 1454-1459.
- Hermant, O., Aziz, A., Warzocha, S. and Al Jahdhami, M. [2020]. Imaging complex fault structures on-shore Oman using optimal transport full waveform inversion. 82nd EAGE Conference and Exhibition, Expanded Abstracts, We Dome1 19.
- Hermant, O., Sedova, A., Royle, G., Retailleau, M., Messud, J., Lambaré, G., Al Abri, S. and Al Jahdhami, M. [2019]. Broadband FAZ land data: an opportunity for FWI. 81st EAGE Conference and Exhibition, Extended abstract, WS08.
- Huang, R., Zhang, Z., Wu, Z., Wei, Z., Mei, J. and Wang, P. [2021]. Full-waveform inversion for full-wavefield imaging: Decades in the making. *The Leading Edge*, May 2021, p. 324-334.
- Jiao, K., Sun, D., Cheng, X. and Vigh, D. [2015]. Adjustive full waveform inversion. SEG Technical Program Expanded Abstracts, 1091-1095.
- Mahrooqi, S., Rawahi, S., Yarubi, S., Abri, S., Yahyai, A., Jahdhami, M., Hunt, K. and Shorter, J. [2012]. Land seismic low frequencies: Acquisition, processing and full wave inversion of 1.5-86 Hz. 82nd Annual International Meeting, SEG, Expanded Abstracts, 1-5.
- Messud, J. and Sedova, A. [2019]. Multidimensional Optimal Transport for 3D FWI: Demonstration on Field Data. 81st EAGE Conference and Exhibition, Extended abstract, Tu R08 02.
- Messud, J., Poncet, R. and Lambaré, G. [2021]. Optimal transport in full-waveform inversion: Analysis and practice of the multidimensional Kantorovich-Rubinstein norm, *Inverse Problems* **37**, 065012.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. and Virieux, J. [2016a]. Increasing the robustness and applicability of full waveform inversion: an optimal transport distance strategy. *The Leading Edge*, **35**:1060-1067.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. and Virieux, J. [2016b]. Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion. *Geophysical Journal International*, **205**(1), 345-377.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. and Virieux, J. [2016c]. An optimal transport approach for seismic tomography: application to 3D full waveform inversion. *Inverse Problems*, **32**(11), 115008.
- Monge, G. [1781]. Mémoire sur la théorie des déblais et des remblais. *Histoire de l'Académie Royale des Sciences*, pages 666-704.
- Perez-Solano, C. and Plessix, R-E. [2019]. Velocity-model building with enhanced shallow resolution using elastic waveform inversion - An example from onshore Oman. *Geophysics*, **84**(6) R977-R988.
- Poncet, R., Messud, J., Bader, M., Lambaré, G., Viguier, G. and Hidalgo, C. [2018]. FWI with optimal transport: a 3D implementation and an application on a field dataset. 80th EAGE Conference and Exhibition, Extended abstract, We A12 02.
- Qiu, L., Ramos-Martínez, J., Valenciano, A., Yang, Y. and Engquist, B. [2017]. Full waveform inversion with an exponentially-encoded optimal transport norm. 87th Annual International Meeting, SEG, Expanded Abstracts, 1286-1290.
- Sedova, A., Messud, J., Prigent, H., Masclet, S., Royle, G. and Lambaré, G. [2019]. Acoustic Land Full Waveform Inversion on a Broadband Land Dataset: The Impact of Optimal Transport. 81st EAGE Conference and Exhibition Extended abstract, Th R08 07.
- Sedova, A., Royle, G., Hermant, O., Retailleau, M. and Lambaré, G. [2017]. High-resolution land full-waveform inversion: a case study on a data set from the Sultanate of Oman. 79th Conference and Exhibition, EAGE, Expanded Abstracts, We A3 04.
- Virieux, J. and Operto, S. [2009]. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, **74**(6), WCC127-WCC152.
- Wang, M., Xie, Y., Xu, W. Q., Xin, K. F., Chuah, B. L., Loh, F. C., Manning, T. and Wolfarth, S. [2016]. Dynamic-warping full-waveform inversion to overcome cycle skipping. 86th SEG Annual International Meeting, Expanded abstract, 1273-1277.
- Warner, M. and Guash, L. [2016]. Adaptive waveform inversion: Theory. *Geophysics*, **81**, R429-R445.
- Yang, Y. and Engquist, B. [2018]. Analysis of optimal transport and related misfit functions in full-waveform inversion. *Geophysics*, **83**(1), A7-A12.
- Yang, Y., Engquist, B., Sun, J. and Froese, B. [2018]. Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion. *Geophysics*, **83**: R43-R62.
- Zhang, Z., Mei, J., Lin, F., Huang, R. and Wang, P. [2018]. Correcting for salt misinterpretation with full-waveform inversion. 88th SEG Annual International Meeting, Expanded abstract, 1143-1147.