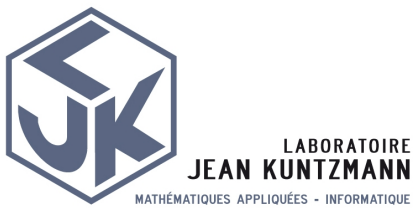
Optimal Transport for FWI : dealing with oscillatory signals

Aude Allain[1], R. Brossier[2], Q. Mérigot[3], L. Métivier[1,2], E. Oudet[1], J.Virieux[2]

Wednesday 3rd May, 2017



[1] LJK, CNRS, Univ. Grenoble Alpes, France
[2] ISTerre, Univ. Grenoble Alpes, France
[3] CEREMADE, CNRS, Univ. Paris Dauphine, France

05-03-2017, OILWS2 workshop, IPAM, UCLA, USA

Outline

Motivation for using optimal transport for FWI

A first implementation: a distance based on the dual form of W_1

Towards a generalization of the approach

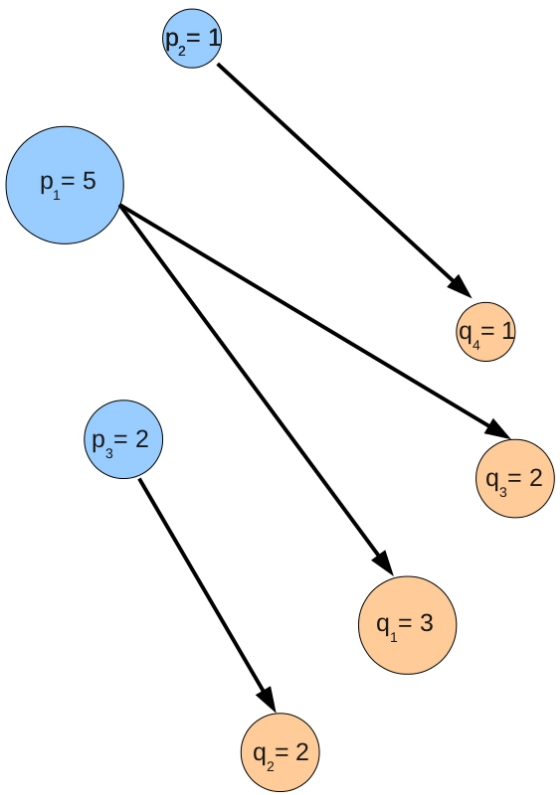
Conclusion & perspectives

Outline

Optimal transport: an optimal assignment problem

- The formalism of optimal transport goes back to (1781) and is due to Gaspard Monge, French Mathematician (one funder of the Ecole Normale and Ecole Polytechnique)
- Initial purpose: find the more efficient way to move sand piles p_i to fill holes q_j during the building of a bridge

- The formalism of optimal transport goes back to (1781) and is due to Gaspard Monge, French Mathematician (one funder of the Ecole Normale and Ecole Polytechnique)
- Initial purpose: find the more efficient way to move sand piles p_i to fill holes q_j during the building of a bridge



The matrix representing this assignment is denoted by

$$\gamma = \begin{pmatrix} 3 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 2 & 0 & 0 \end{pmatrix} \qquad (1)$$

- Kantorovich (1942) relaxation: Optimal transport consists in finding the matrix γ (also known as transport plan) satisfying the **linear programming** problem

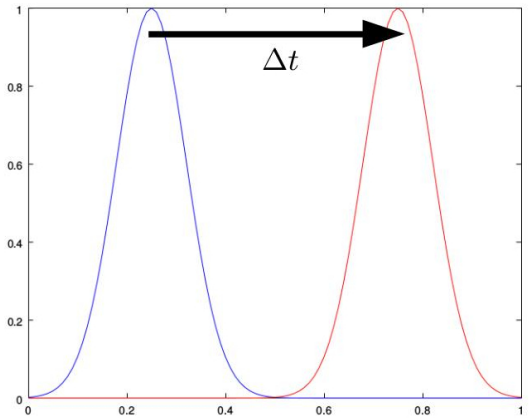
$$\min_{\gamma_{ij} \in \Pi(p,q} \sum_{ij} \gamma_{ij} \|x_i - y_j\|,$$
$$\text{where } \Pi(p,q) = \left\{ \gamma_{ij} \geq 0, \sum_j \gamma_{ij} = p_i, \sum_i \gamma_{ij} = q_j \right\} \tag{2}$$

  and $\|x_i - y_j\|$ denotes a distance between x_i and y_j which is called the ground distance (often the Euclidean distance)

- Assume p_i and q_j are discretization of data d_{cal} and d_{obs}
- The minimal cost in the sense of the optimal transport defines a distance between the distribution d_{cal} and d_{obs}
- This distance is also referred to as the Wasserstein distance $W_p, p \geq 1$

$$W_p(d_{cal}, d_{obs}) = \left(\min_{\gamma_{ij} \in \Pi(d_{cal}, d_{obs})} \sum_{ij} \gamma_{ij} \|x_i - y_j\|^p \right)^{1/p} \tag{3}$$

- Consider d_{cal} and d_{obs} are two 1D Gaussian distribution shifted by Δt



- Then we have

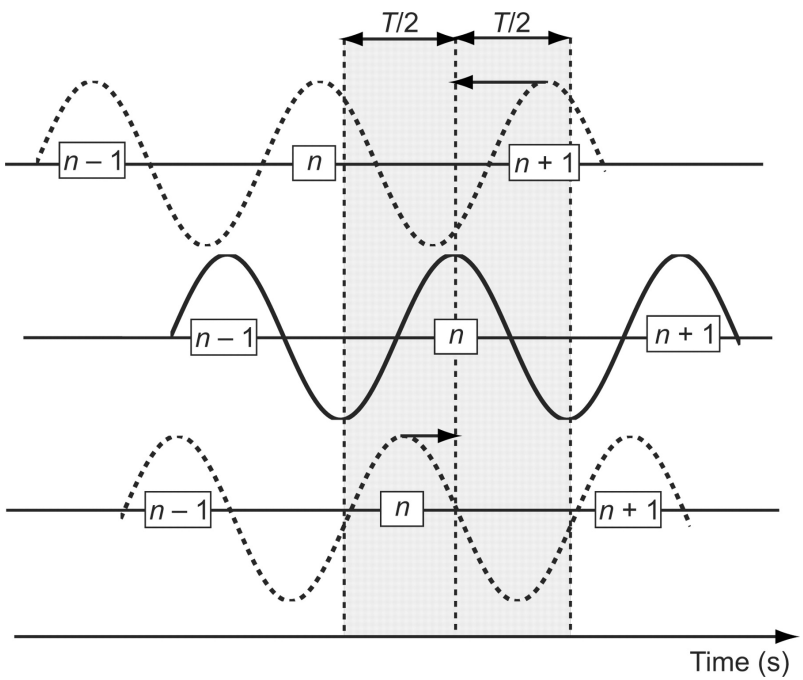
$$W_p^p(d_{cal}, d_{obs}) = |\Delta t|^p, \ p \geq 1 \tag{4}$$

Convexity of the Wasserstein distance with respect to Δt

- Large-scale to medium-scale velocity perturbations (1000 m to 100 m at the exploration scale) are mainly responsible for **time-shifts**

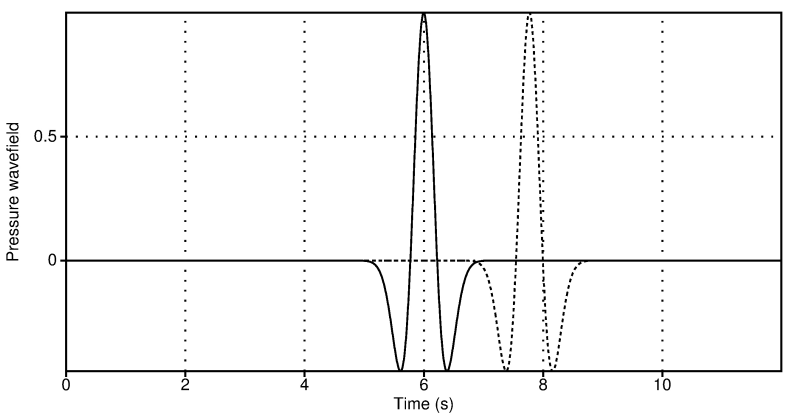
  $\Rightarrow$ from an **inverse problem** point of view, recovering these velocity perturbations require to correctly interpret these **time-shifts**

- This is difficult when the data is compared with a L^2 **distance** because of **cycle skipping/phase ambiguity**

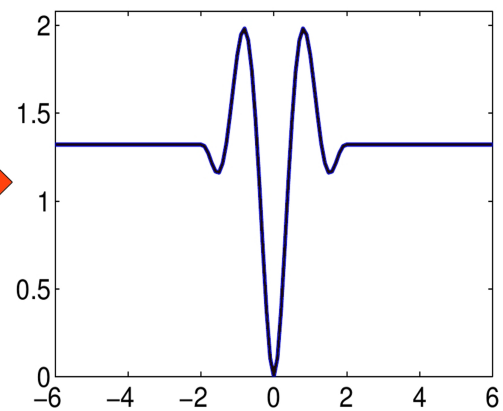


Capturing the time shift: how to measure the misfit ?

- Simple test: computing the L^2 misfit between two Ricker with respect to **time-shifts** yields a **multi-modal** misfit function

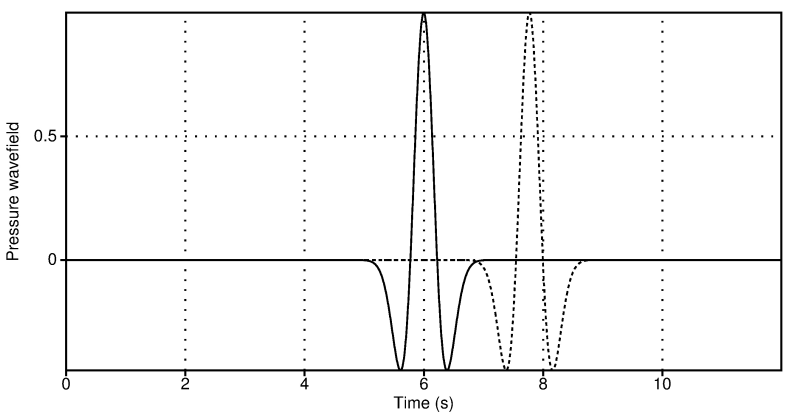


Shifted Ricker signal

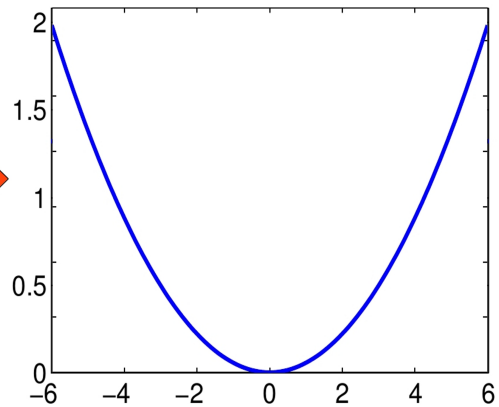


L2 misfit function

- Computing the optimal transport misfit between two Ricker with respect to **time-shifts** yields a **convex** misfit function (Engquist and Froese, 2014)



Shifted Ricker signal



OT misfit function

In the general case, computing the Wasserstein distance between d_{cal} and d_{obs} requires two assumptions to be satisfied

- positivity
  $d_{cal} \geq 0, \quad d_{obs} \geq 0$
- mass conservation
  $\sum_i (d_{cal})_i = \sum_j (d_{obs})_j$

What about seismic data

Consider d_{cal} and d_{obs} represent two discretized seismic traces

- d_{cal} and d_{obs} are oscillatory: **positivity assumption breaks down**
- the zero frequency of the signals is 0: **the mass conservation holds**

$$\int d_{cal}(t)dt = \int d_{obs}(t)dt = 0 \tag{5}$$

Another issue: computational complexity

Linear programming algorithms (simplex) for the computation of the Wasserstein distance require at least $O(N^3)$ operations for size N discrete distributions

For acoustic time-domain FWI

- **In 2D**, the data is discretized with O(10^3) time discretization points and O(10^2) to receivers $\Longrightarrow N = O(10^5)$
- **In 3D**, the data is discretized with O(10^3) time discretization points and O(10^4) to receivers $\Longrightarrow N = O(10^7)$

Besides the positivity problem, we need an efficient numerical method to solve the optimal transport problem

Outline

To overcome the positivity problem, we focus on the dual formulation of the W_1 distance (Santambrogio, 2015)

$$\begin{aligned}
&\min_{\gamma_{ij}\in\Pi(d_{cal},d_{obs})} \sum_{ij} \gamma_{ij}\|x_i - y_j\| \\
&\Longleftrightarrow \\
&\max_{\varphi,\psi} \sum_i \varphi_i(d_{cal})_i + \sum_j \psi_j(d_{obs})_j,\ \varphi_i + \psi_j \leq \|x_i - y_j\|, \quad \text{(standard dual)} \\
&\Longleftrightarrow \\
&\max_{\varphi\in \mathrm{Lip}_1} \sum_i \varphi_i \left((d_{cal})_i - (d_{obs})_i\right) \quad \text{(only for } W_1\text{)}
\end{aligned} \tag{6}$$

where Lip_1 is the space of 1-Lipschitz functions for the ground cost $\|x-y\|$

$$\mathrm{Lip}_1 \left\{\varphi : x \in X \longrightarrow \mathbb{R},\ \forall (x_i, x_j)\ \in X \times X,\ |\varphi_i - \varphi_j| \leq \|x_i - x_j\|\right\} \tag{7}$$

Computation of the dual W_1 with ℓ_1 ground cost

To design a fast solver, we focus on the ℓ_1 ground cost

- The discretization of the problem is, for a 3D data cube indexed by

  $1 \leq n \leq N_t$ **for time** t_n

  $1 \leq i \leq N_x$, $1 \leq j \leq N_y$ **for the receiver position** x_i, y_j

$$\begin{aligned} &\max_{\varphi_{nij}} \sum_{nij} \varphi_{nij} \left((d_{cal})_{nij} - (d_{obs})_{nij} \right), \\ &|\varphi_{nij} - \varphi_{n'i'j'}| < |t_n - t_{n'}| + |x_i - x_{i'}| + |y_j - y_{j'}|, \end{aligned} \tag{8}$$

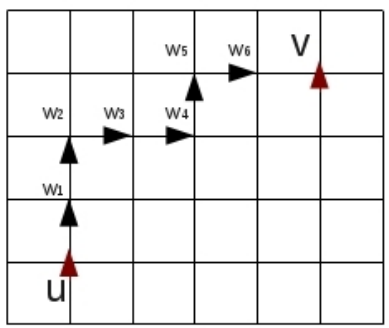
- The number of unknowns is $N = N_t \times N_x \times N_y$

O(N^2) global linear constraints: too high complexity for efficient numerical algorithms

- Thanks to the sum of absolute values in problem (8), the global system is equivalent to

$$\max_{\varphi_{nij}} \sum_{nij} \varphi_{nij} \left((d_{cal})_{nij} - (d_{obs})_{nij} \right), \tag{9}$$

$$\begin{cases} |\varphi_{n+1ij} - \varphi_{nij}| < |t_{n+1} - t_n|, \\ |\varphi_{ni+1j} - \varphi_{nij}| < |y_{i+1} - y_i|, \\ |\varphi_{nji+1} - \varphi_{nij}| < |z_{j+1} - z_j|, \end{cases}$$



O(N) local linear constraints: far better complexity (Métivier et al., 2016c)

Computation of the dual W_1 with ℓ_1 ground cost

- The problem (9) is reformulated as the convex non-smooth problem

$$\begin{aligned} &\max_{\varphi} f_1(\varphi) + f_2(A\varphi), \\ &f_1(\varphi) = \sum_{nij} \varphi_{nij} \left((d_{cal})_{nij} - (d_{obs})_{nij} \right), \quad f_2(\varphi) = -i_K(\varphi), \end{aligned} \tag{10}$$

- The matrix A accounts for the constraints

$$A = \left[D_x \;\; D_y \;\; D_z \;\; \frac{1}{\lambda} I_N \right]^T \in \mathbb{M}_{P,N}(\mathbb{R}) \tag{11}$$

where D_x, D_y, D_z are the forward finite-differences operators

$$(D_x\varphi)_{ijk} = \frac{\varphi_{i+1,j,k} - \varphi_{ijk}}{h_x}, \quad (D_y\varphi)_{ijk} = \frac{\varphi_{i,j+1,k} - \varphi_{ijk}}{h_y}, \quad (D_z\varphi)_{ijk} = \frac{\varphi_{i,j,k+1} - \varphi_{ijk}}{h_z} \tag{12}$$

while K is the unit hypercube and i_K the indicator function of K

$$i_K(v) = \begin{cases} 0 & \text{if} \quad v \in K \\ +\infty & \text{if} \quad v \notin K, \end{cases} \tag{13}$$

Computation of the dual W_1 with ℓ_1 ground cost

- A proximal splitting algorithm can be use to solve this problem: we choose the Simultaneous Direction Method of Multipliers (SDMM) for its good convergence properties (Combettes and Pesquet, 2011)
- The most computational demanding task of the algorithm is the resolution of a linear system involving a matrix which is equivalent to a second-order finite difference discrete of the Laplacian operator with homogeneous Neumann boundary condition
- Fast-solvers can be used to invert this matrix (Poisson's problem), either based on

  Fast Fourier Transform (Swarztrauber, 1974): complexity $\mathrm{O}(N \log N)$

  Multigrid algorithms (Brandt, 1977; Adams, 1989): complexity in $\mathrm{O}(N)$

A complete description of this numerical strategy is given in (Métivier et al., 2016c) with 2D and 3D examples

- We assume the acoustic approximation
- For the L^2 **distance**, the gradient of the misfit function is

$$\nabla f(x) = \frac{-2}{v_P(x)^3} \int_0^T \partial_{tt} p(x,t) \lambda(x,t) dt \tag{14}$$

where

- $v_P(x)$: P-wave velocity
- $p(x,t)$: pressure wavefield
- $\lambda(x,t)$: adjoint wavefield *backpropagation of the* L^2 **residuals** $d_{obs} - d_{cal}$

$$\partial_{tt} \lambda - c^2 \Delta \lambda = d_{obs} - d_{cal} \tag{15}$$

- We assume the acoustic approximation
- For the **KR distance**, the gradient of the misfit function is

$$\nabla f(x) = \frac{-2}{v_P(x)^3} \int_0^T \partial_{tt} p(x,t) \lambda(x,t) dt \tag{16}$$

where

- $v_P(x)$: P-wave velocity
- $p(x,t)$: pressure wavefield
- $\lambda(x,t)$: adjoint wavefield *backpropagation of the KR residuals* $\overline{\varphi}$

$$\partial_{tt}\lambda - c^2 \Delta \lambda = \overline{\varphi} \tag{17}$$

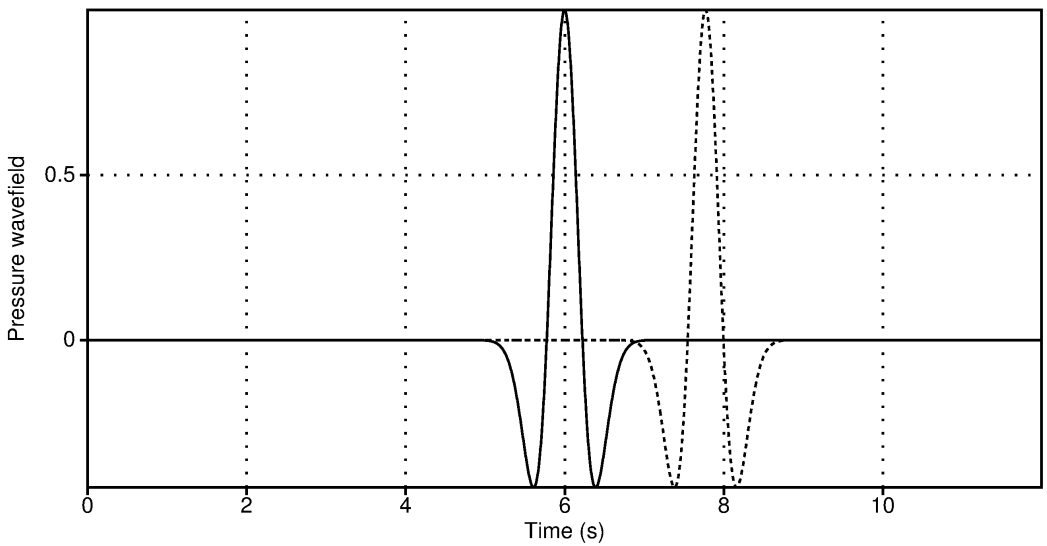
$$\overline{\varphi_{nij}} = \arg\max_{\varphi_{nij}} \sum_{nij} \varphi_{nij} \left((d_{cal})_{nij} - (d_{obs})_{nij} \right), \tag{18}$$

$$\begin{cases} |\varphi_{n+1ij} - \varphi_{nij}| < |t_{n+1} - t_n|, \\ |\varphi_{ni+1j} - \varphi_{nij}| < |y_{i+1} - y_i|, \\ |\varphi_{nji+1} - \varphi_{nij}| < |z_{j+1} - z_j|, \end{cases}$$

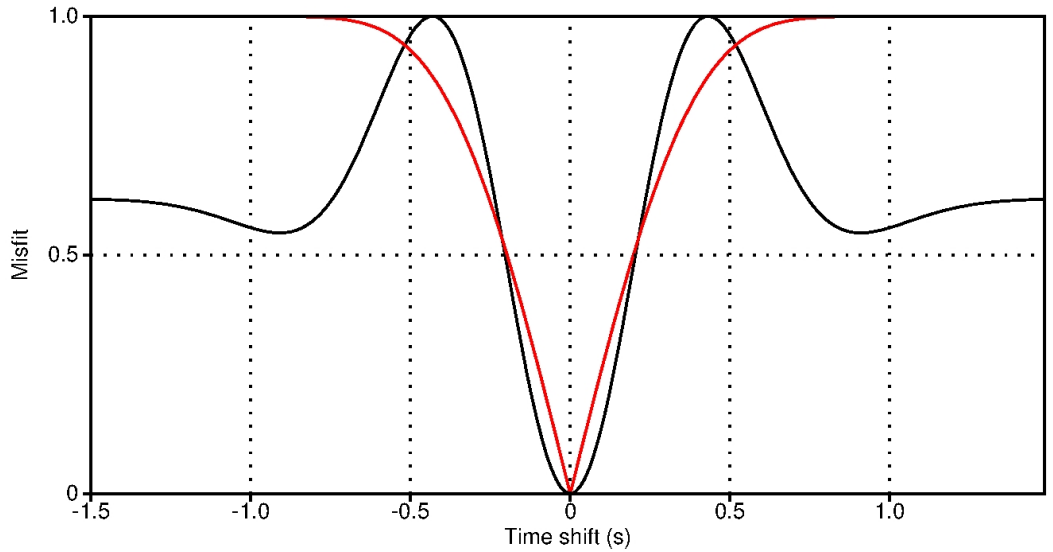
For the optimal transport distance, the corresponding adjoint source is equal to the function $\overline{\varphi}$ which is the solution of the constrained maximization problem

Numerical illustration: 1D Time shifted Ricker signals

- We come back to 1D time-shifted Ricker functions



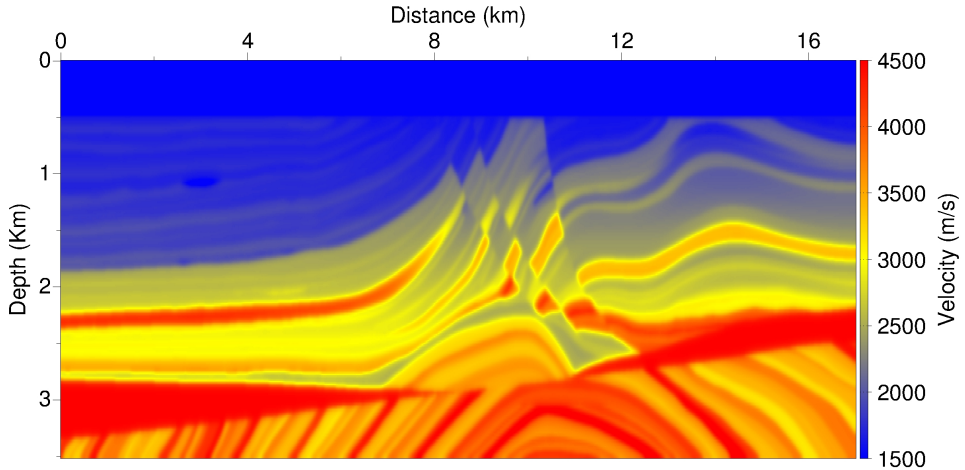
- Misfit function shape



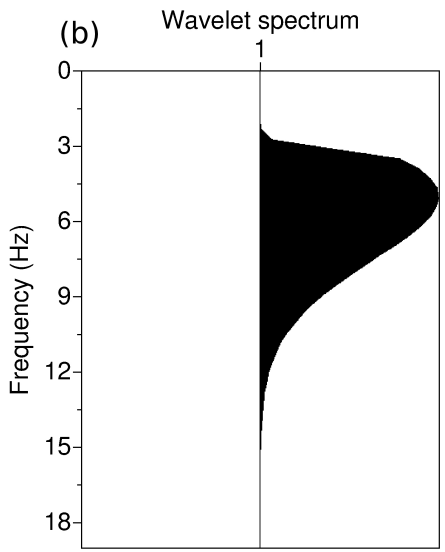
L^2 distance function with respect to the time-shift.

W_1 distance function with respect to the time shift

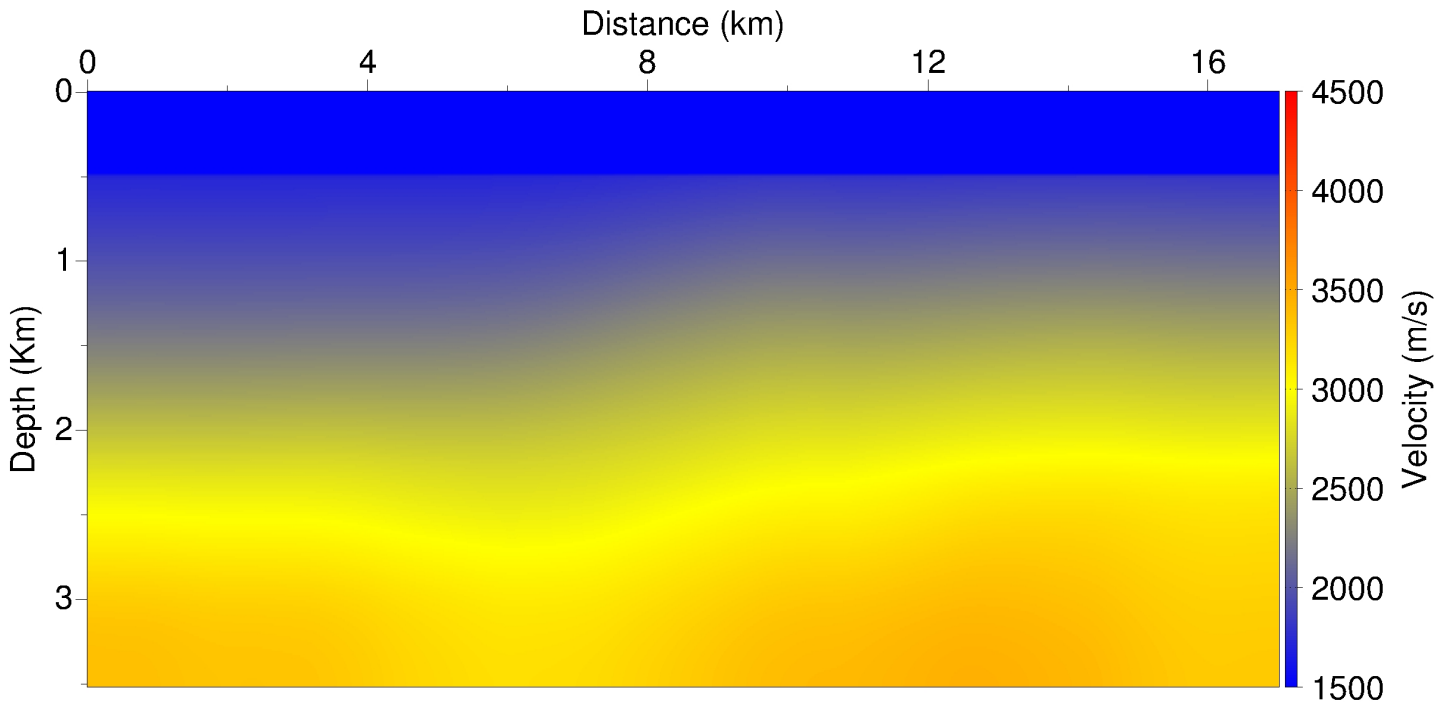
We recover a single minimum but lose the convexity



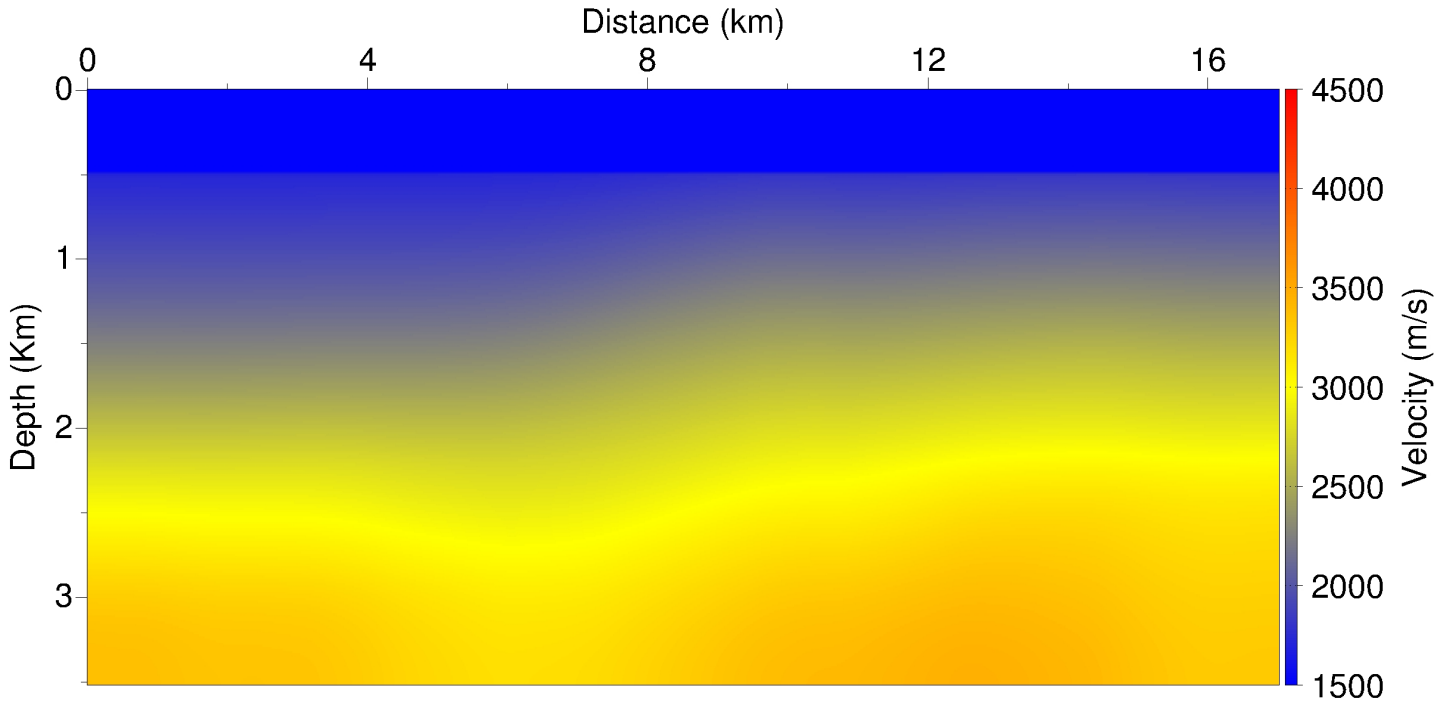
- **surface** acquisition with 128 sources each 125 m and 168 receivers each 100 m
- **acoustic** modeling engine to generate synthetic data
- **high pass filter of data**: no energy below 3 Hz



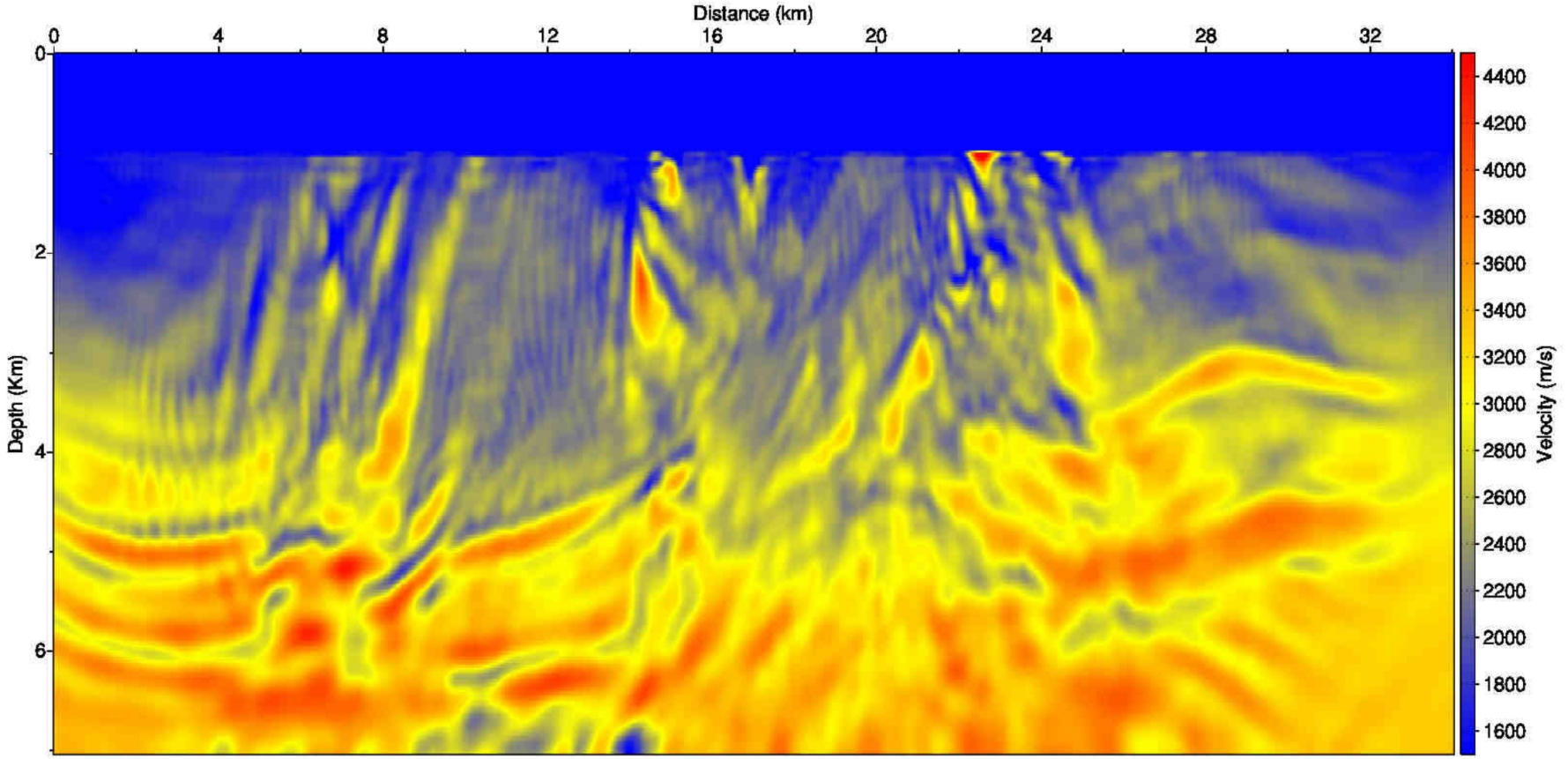
Inversion using the standard L^2 distance starting from the crude velocity model



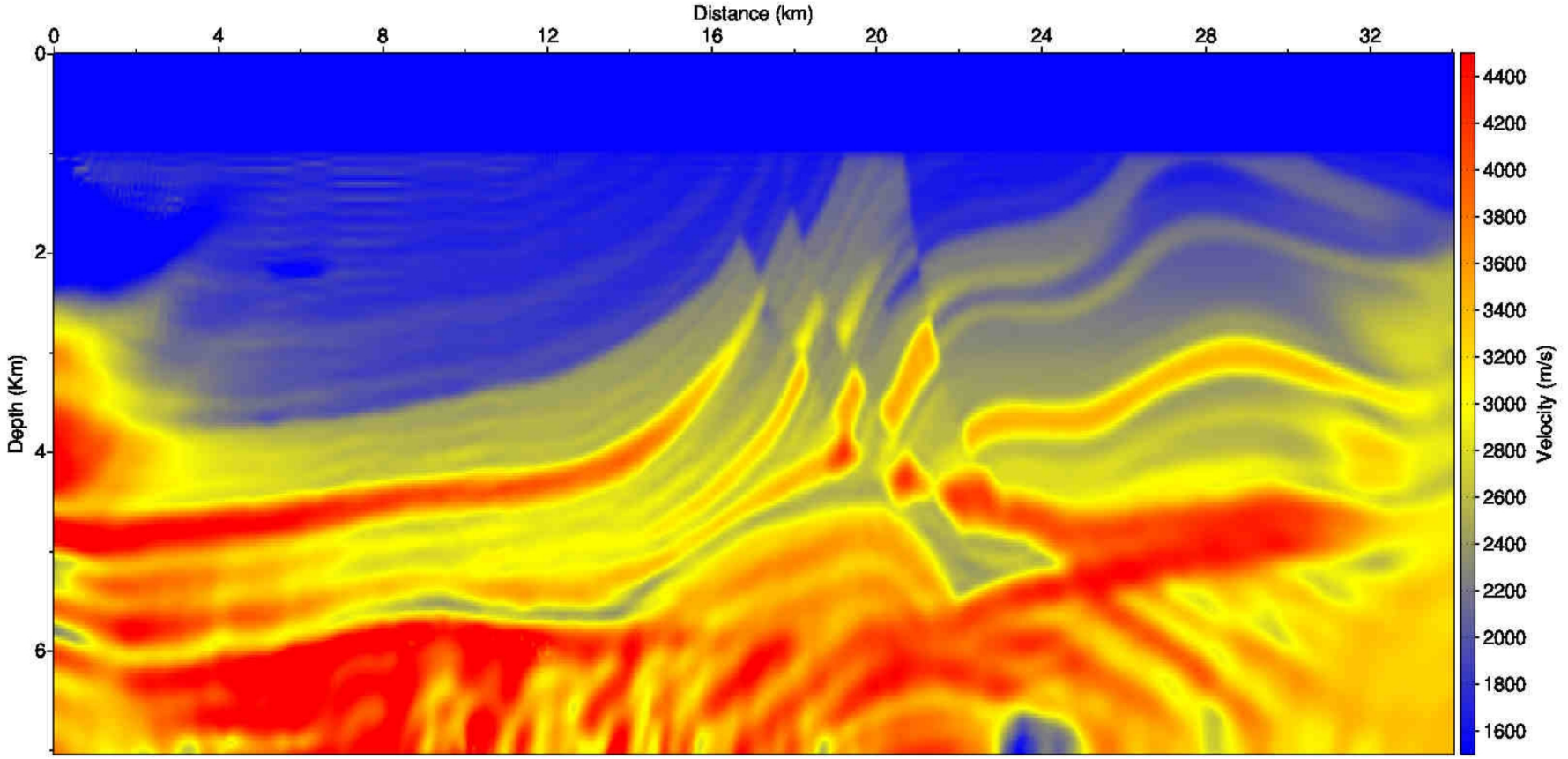
Inversion using the W_1 distance starting from the crude velocity model



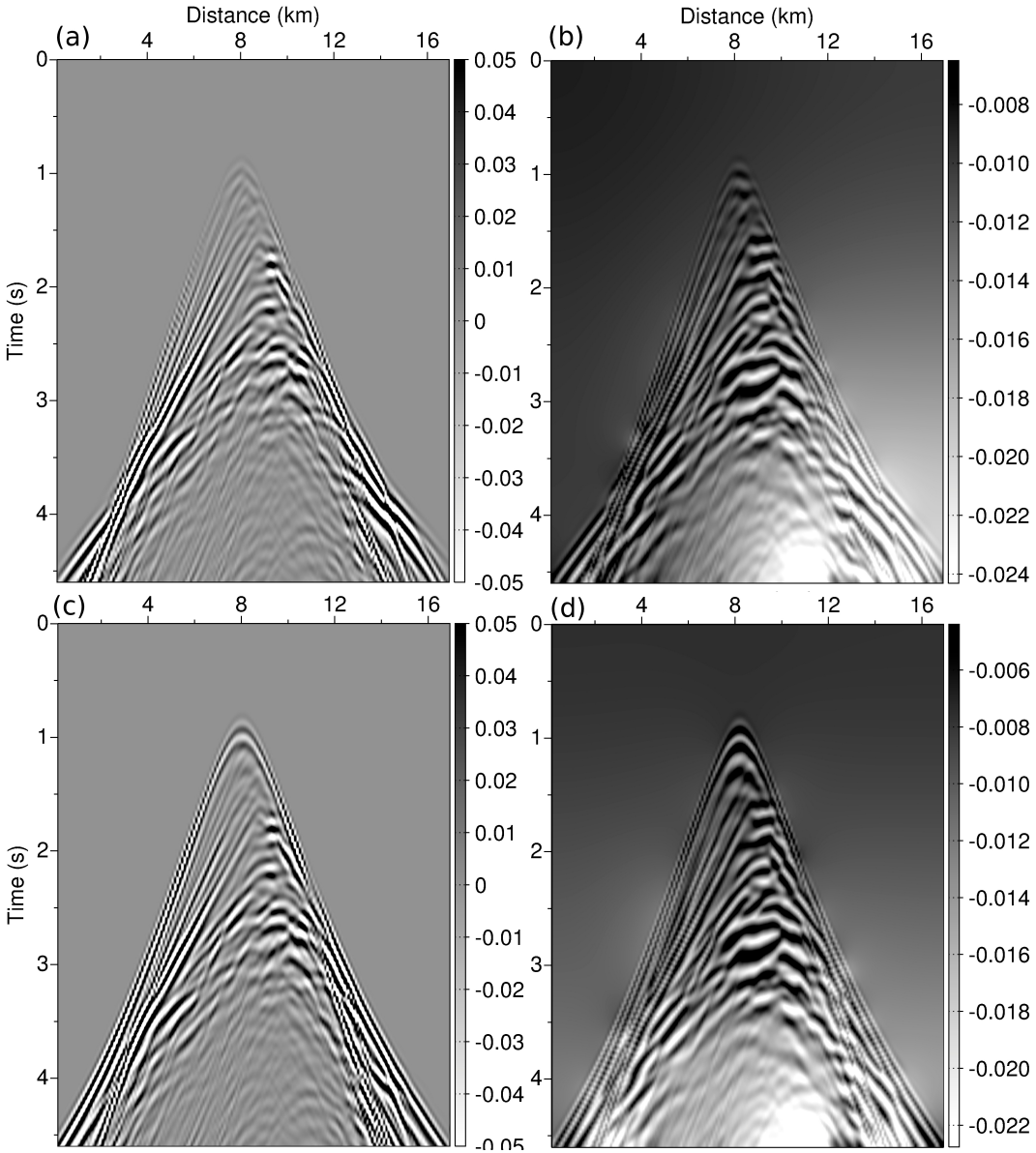
Inversion using the L^2 distance starting from a crude velocity model



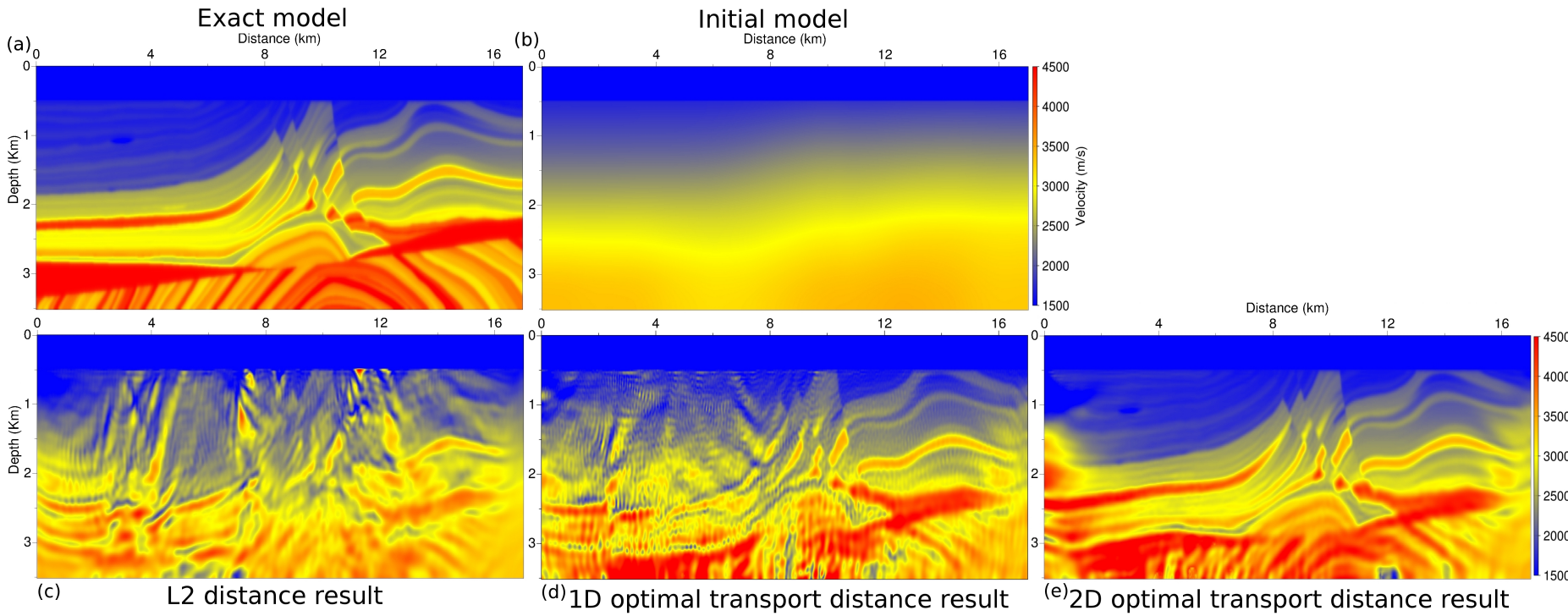
Inversion using the W_1 distance starting from a crude velocity model



The Marmousi case study : example of adjoint-source



The Marmousi case study : 2D OT or multi-1D OT ?



Computation overhead per gradient (FISHPACK and 50 SDMM iterations): **3.8 s** (19%)

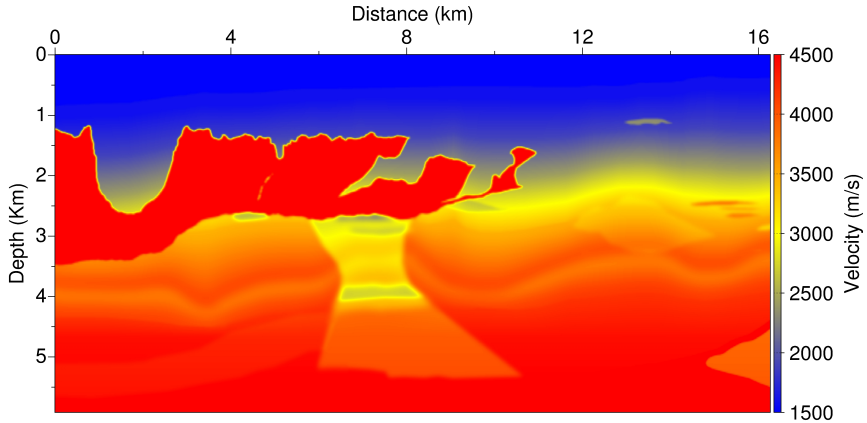
- L^2 gradient computation time **20,6 s**
- KR gradient computation time **24,4 s**

Number of iterations

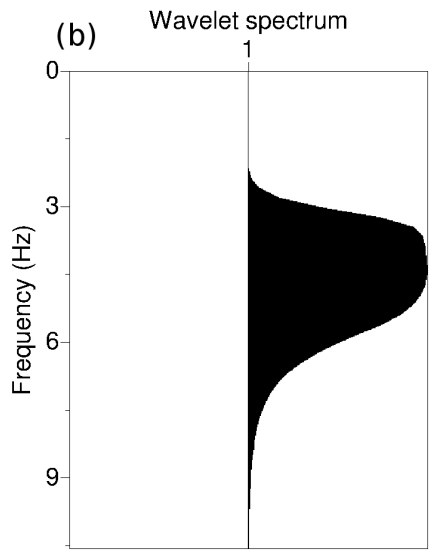
- L^2 inversion number of iterations: 83
- KR inversion number of iterations: 439

Reconstructing salt structures through a layer stripping approach

Exact model

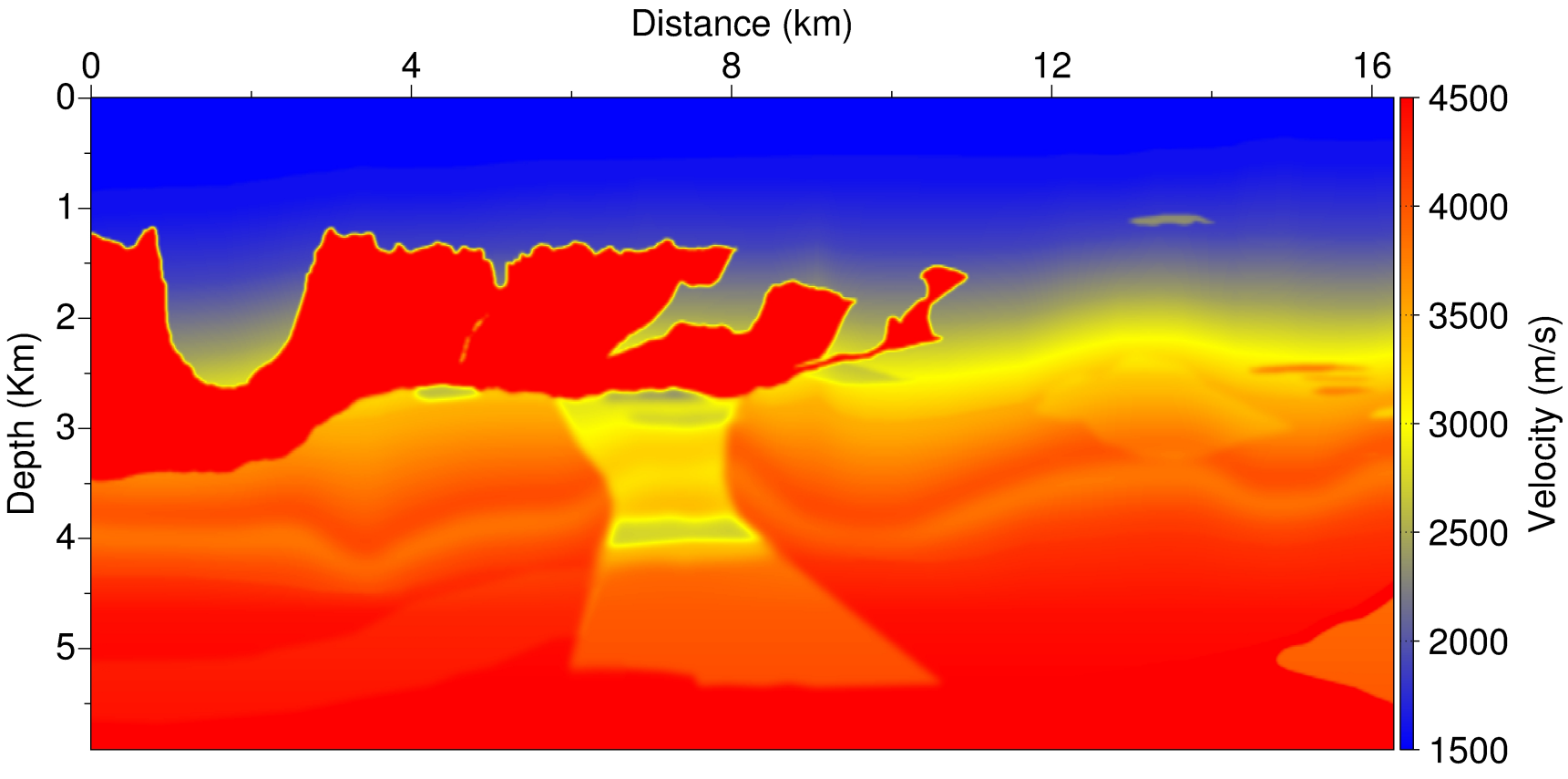


- **surface** acquisition with 128 sources each 125 m and 161 receivers each 100 m
- **acoustic** modeling engine to generate synthetic data
- **high pass filter of data**: no energy below 3 Hz



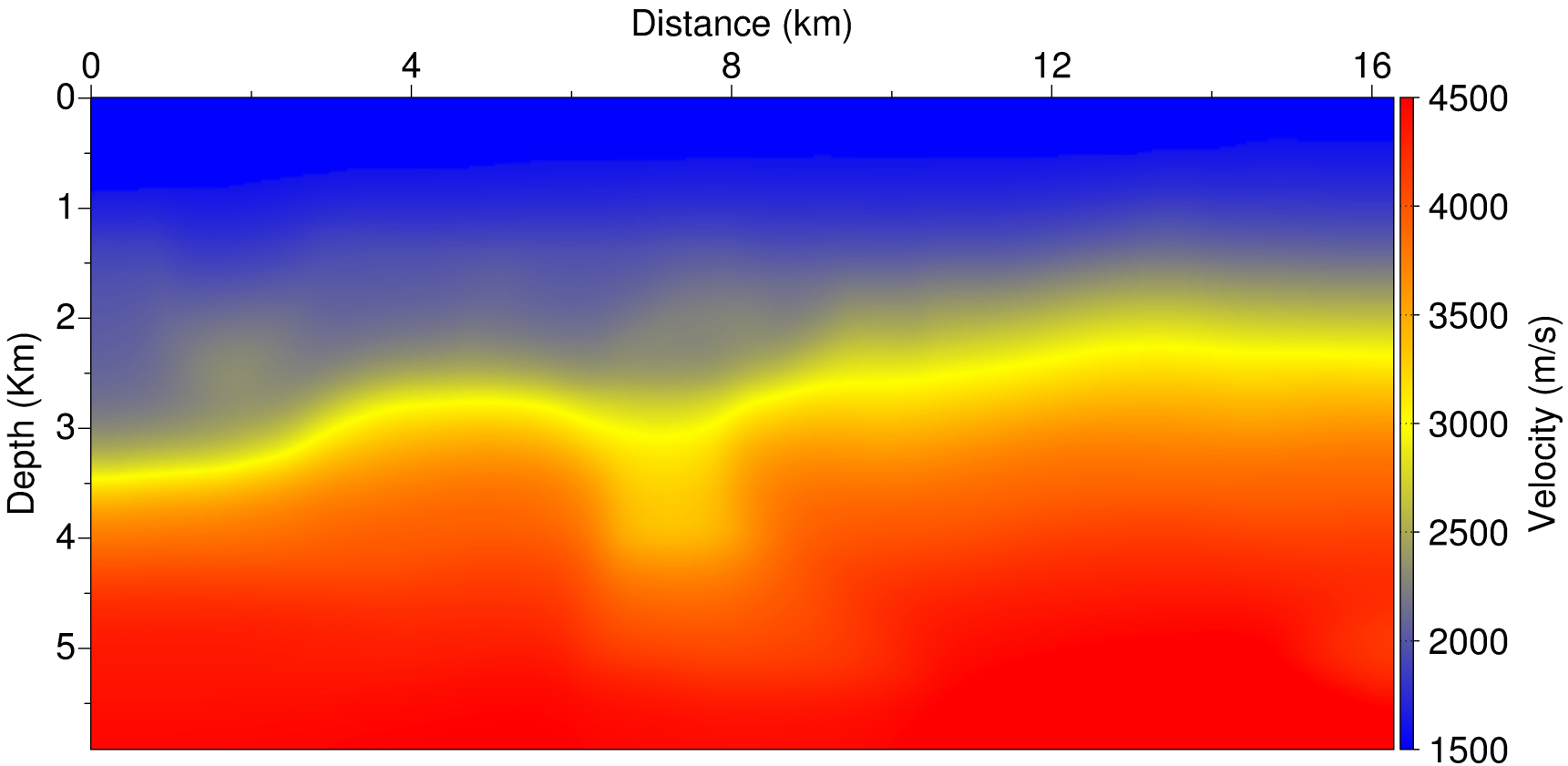
Reconstructing salt structures through a layer stripping approach

Exact model



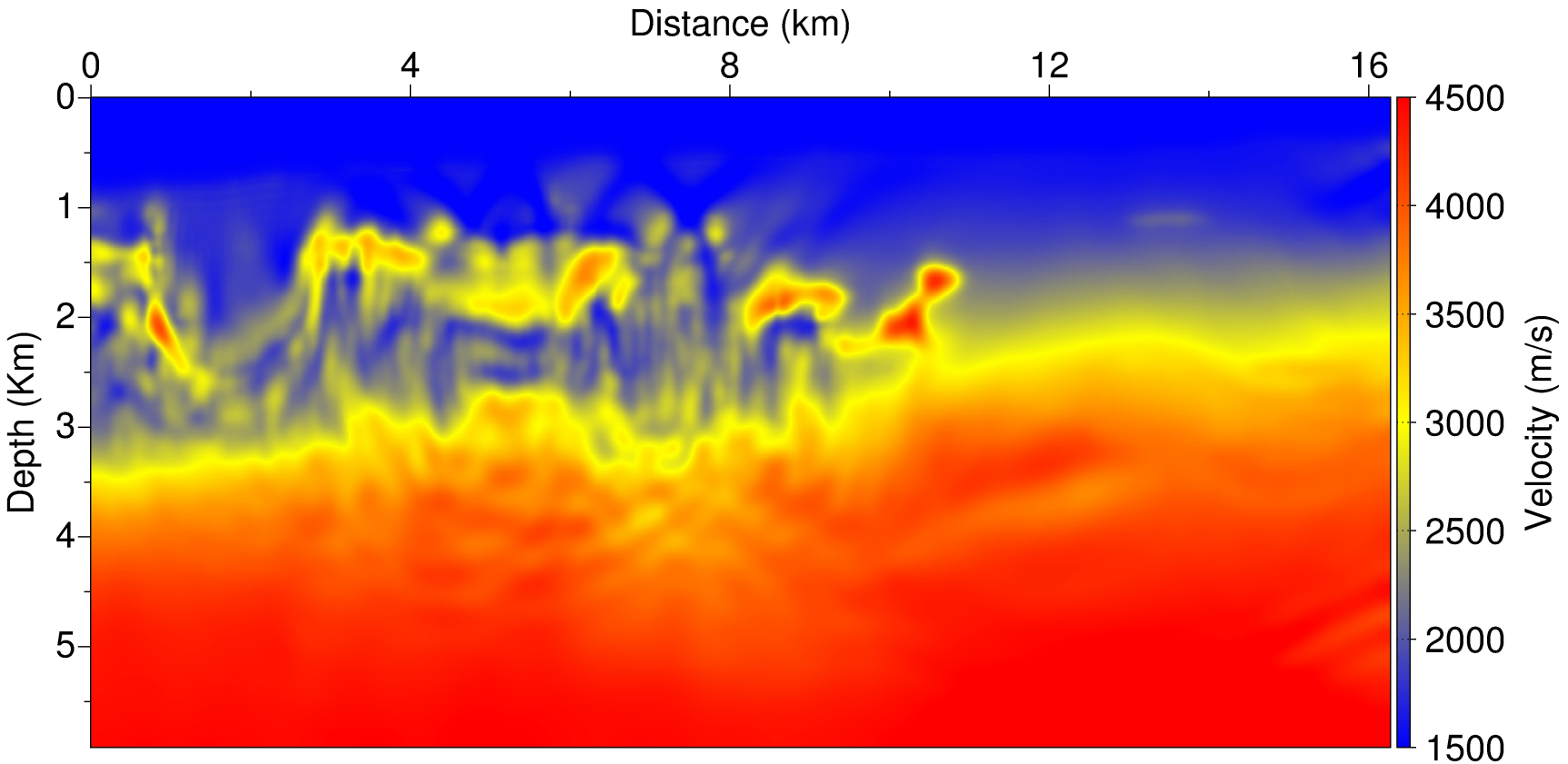
Reconstructing salt structures through a layer stripping approach

Initial model



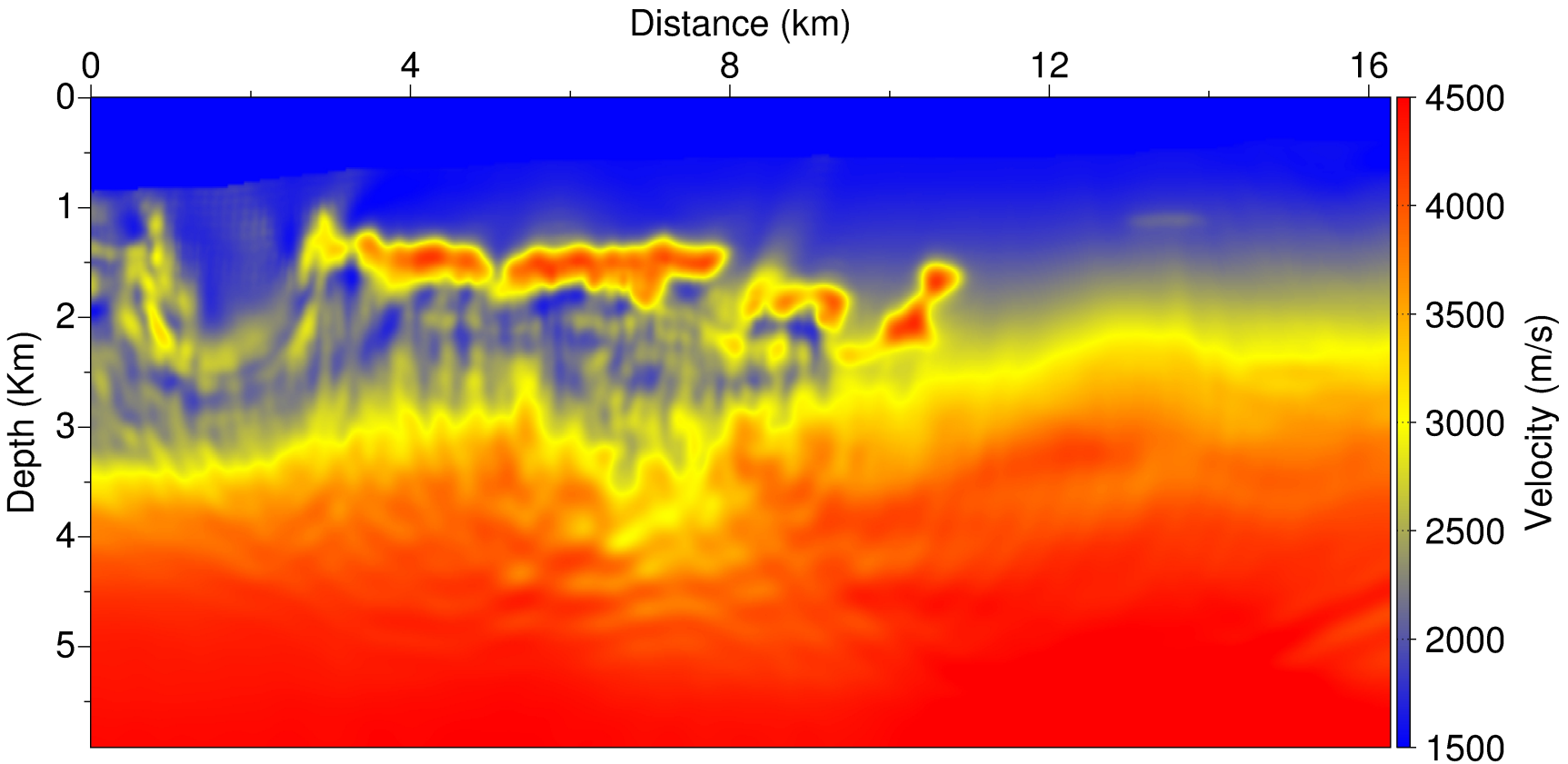
Reconstructing salt structures through a layer stripping approach

Step 1 with L^2 distance



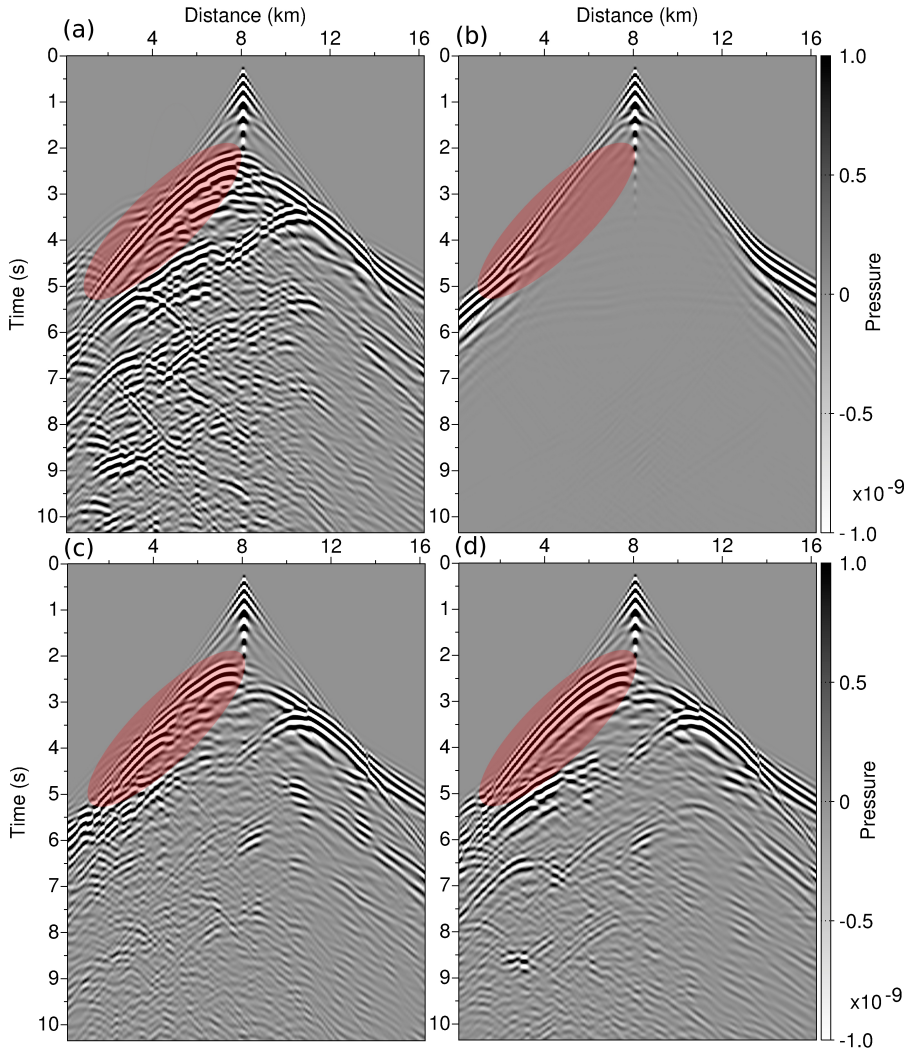
Reconstructing salt structures through a layer stripping approach

Step 1 with optimal transport distance



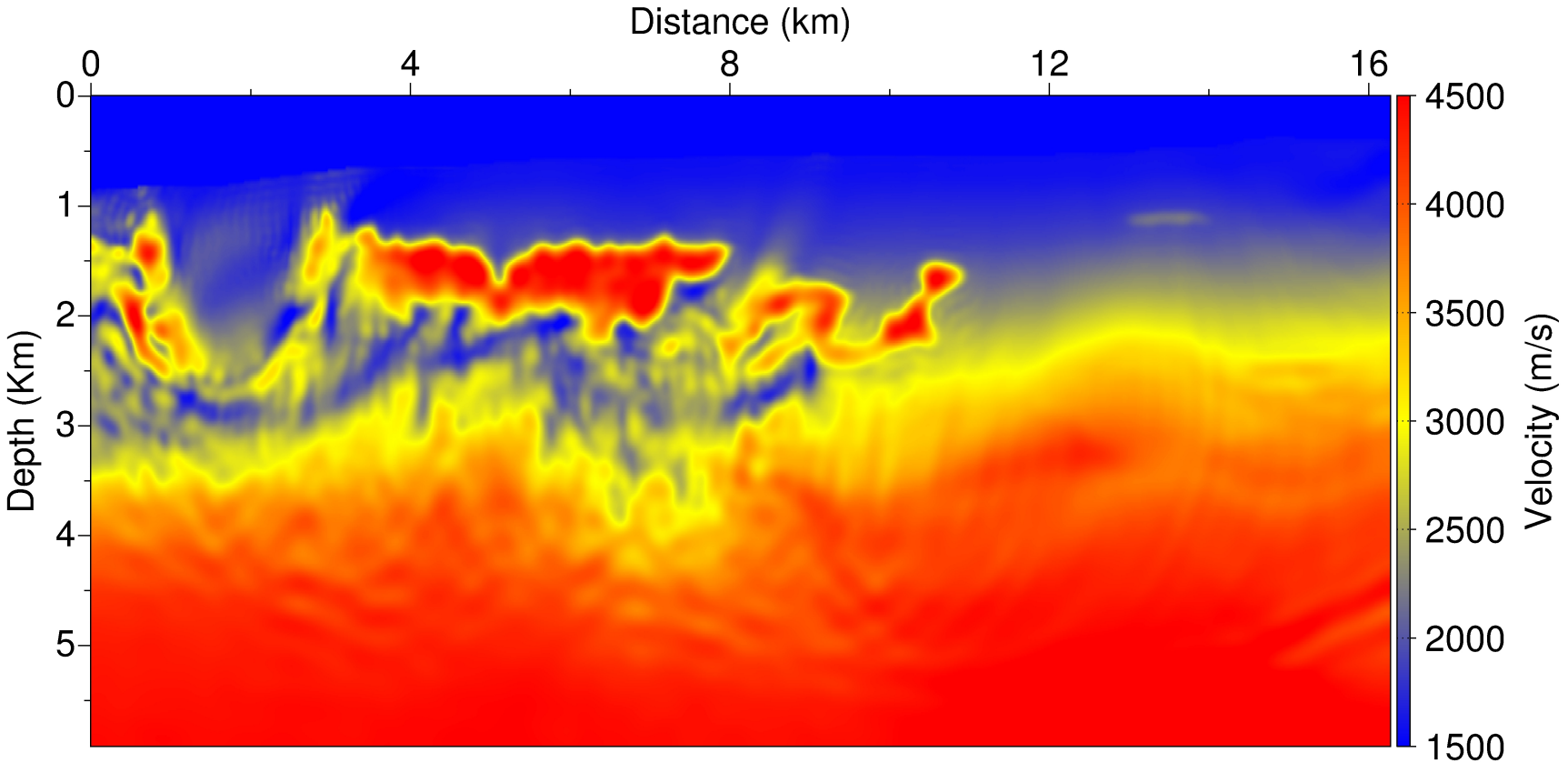
Reconstructing salt structures through a layer stripping approach

Data comparison



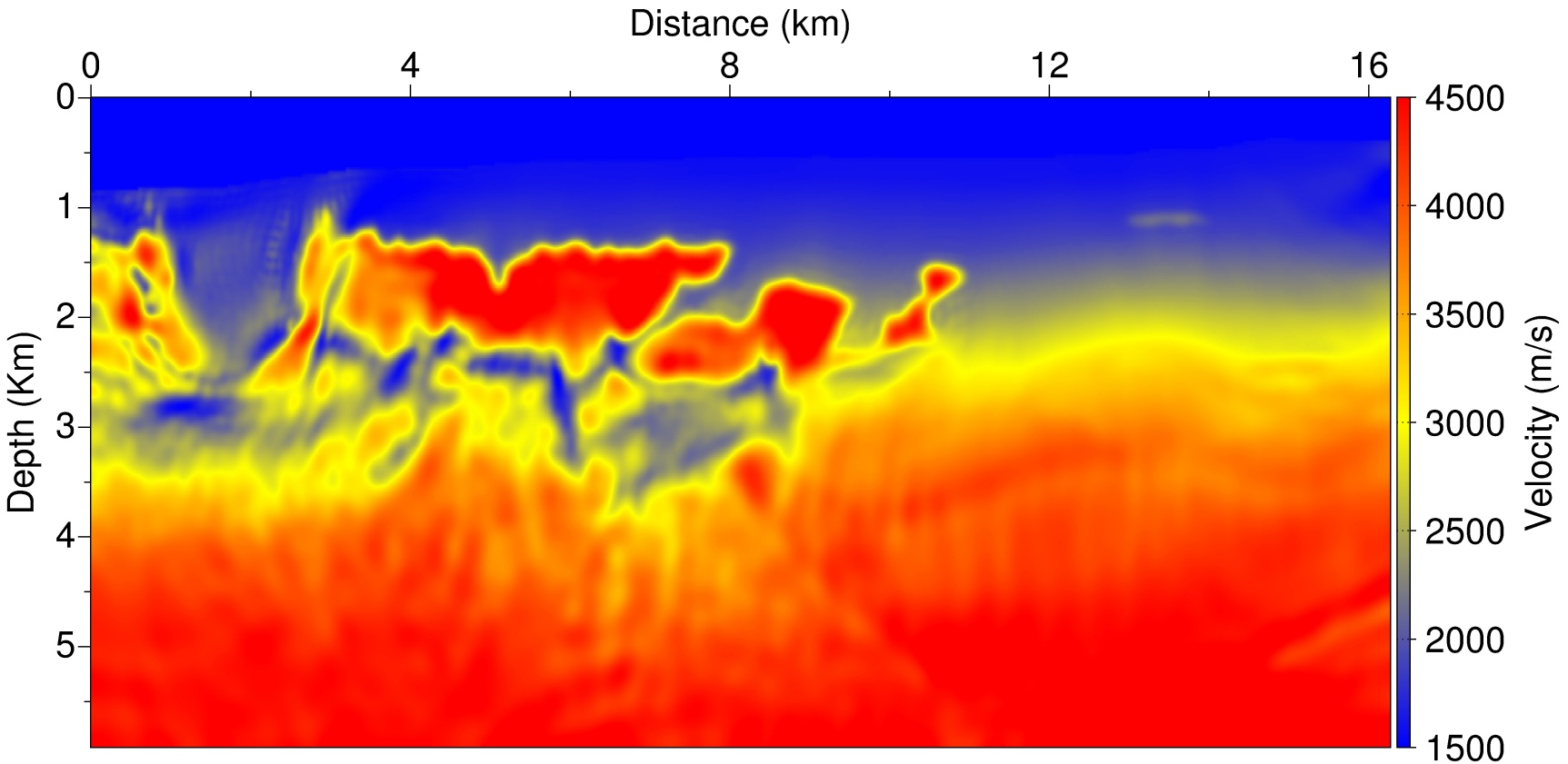
Reconstructing salt structures through a layer stripping approach

Step 2



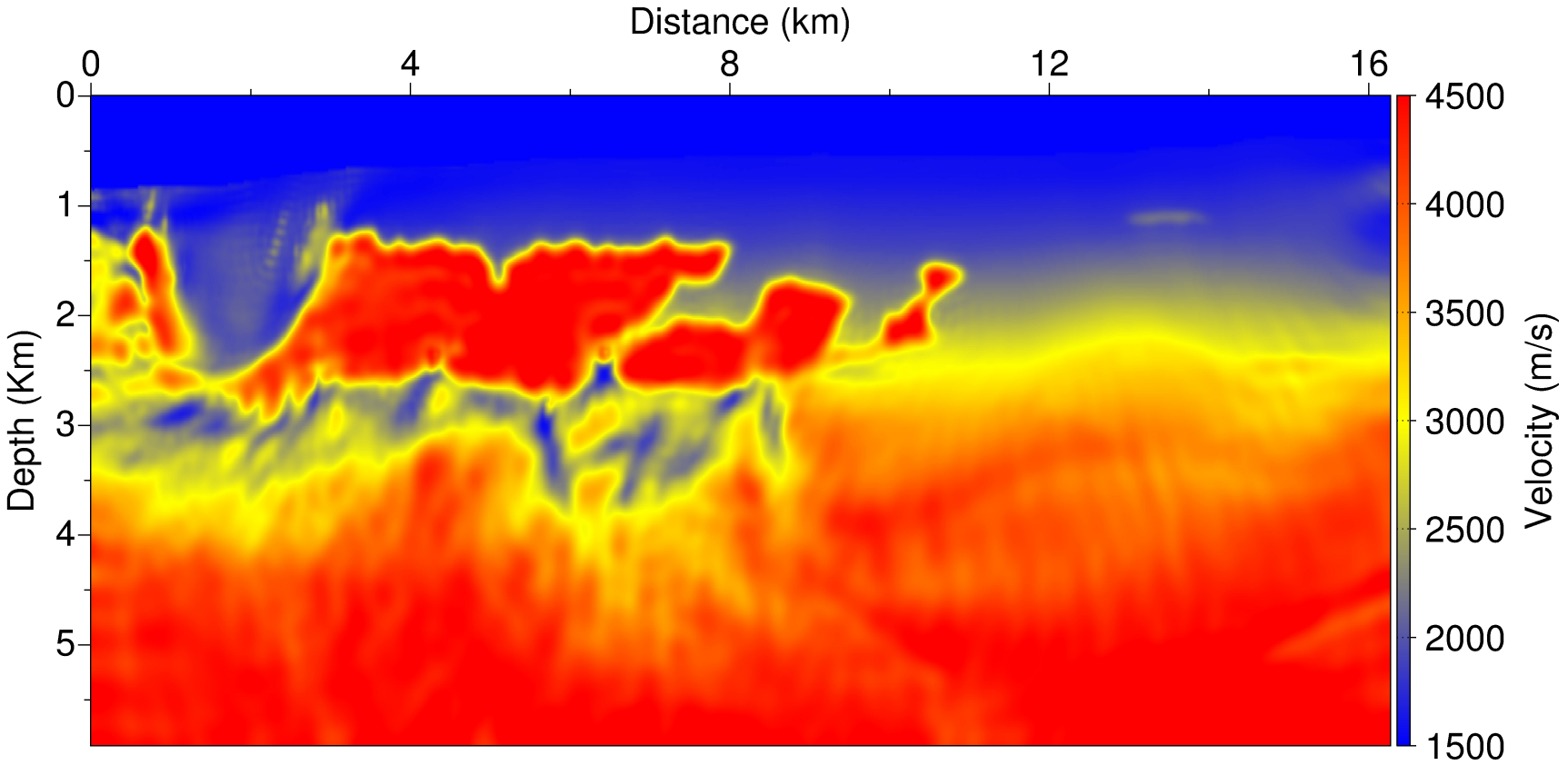
Reconstructing salt structures through a layer stripping approach

Step 3



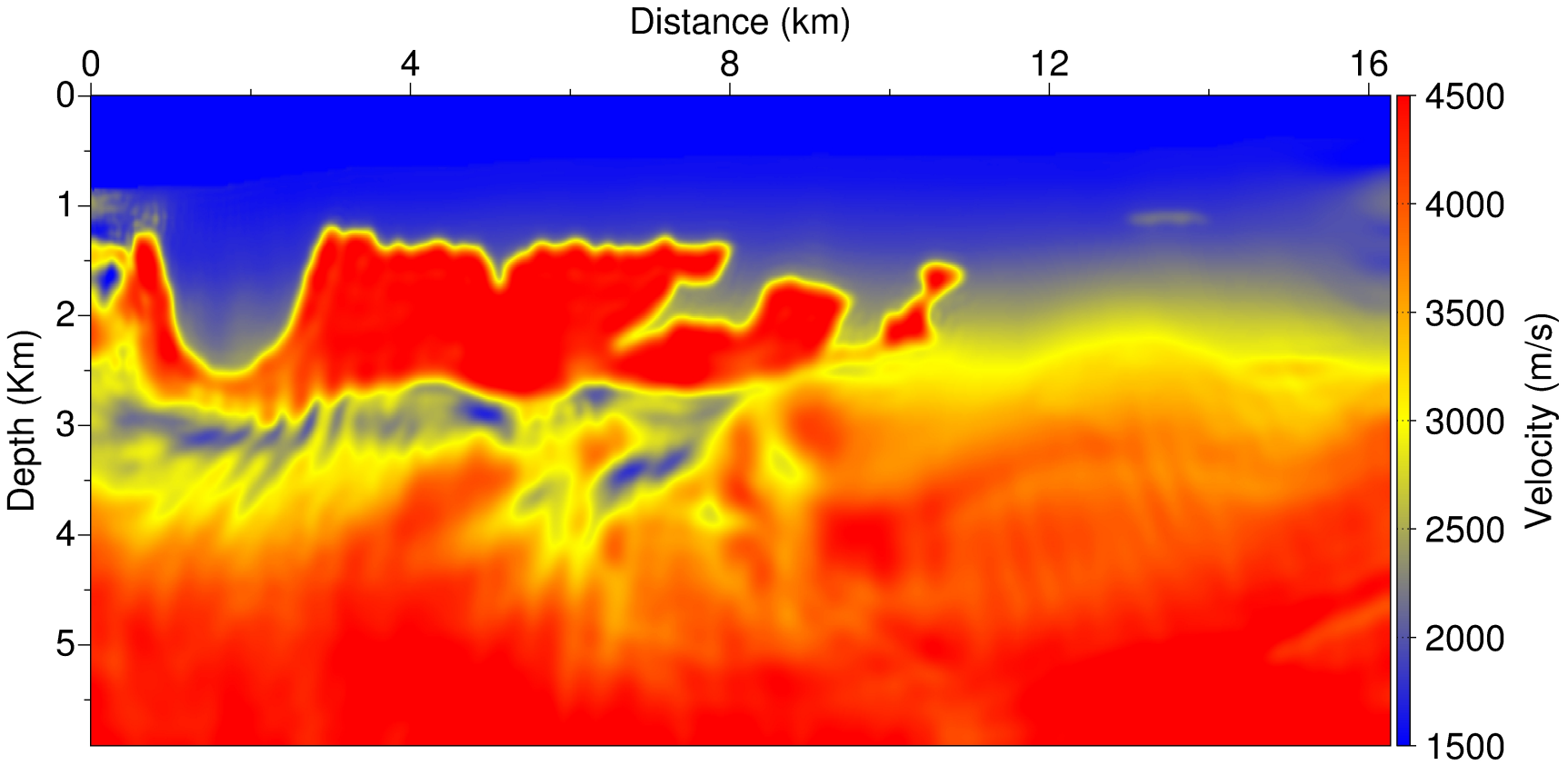
Reconstructing salt structures through a layer stripping approach

Step 4



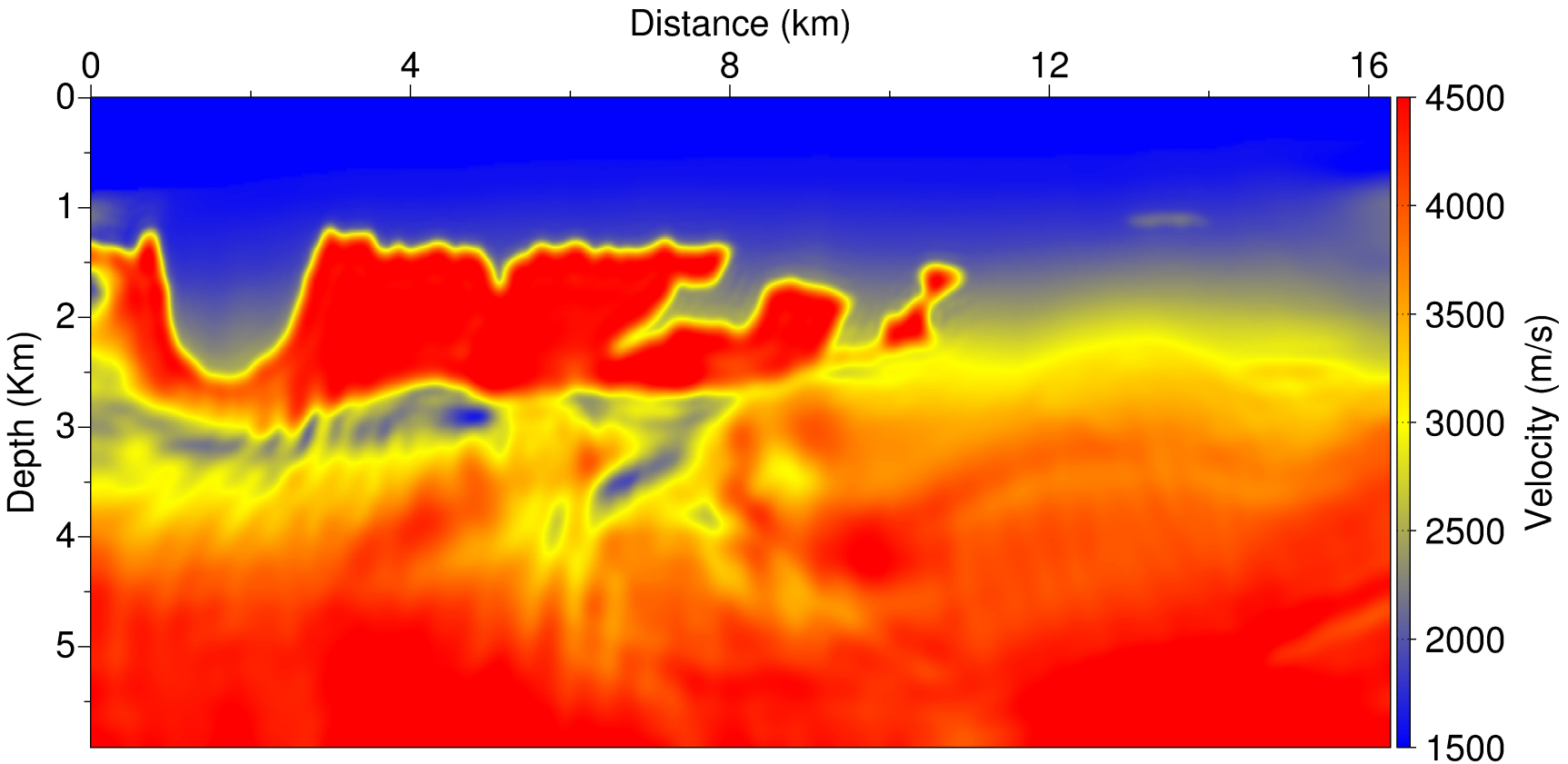
Reconstructing salt structures through a layer stripping approach

Step 5



Reconstructing salt structures through a layer stripping approach

Step 6

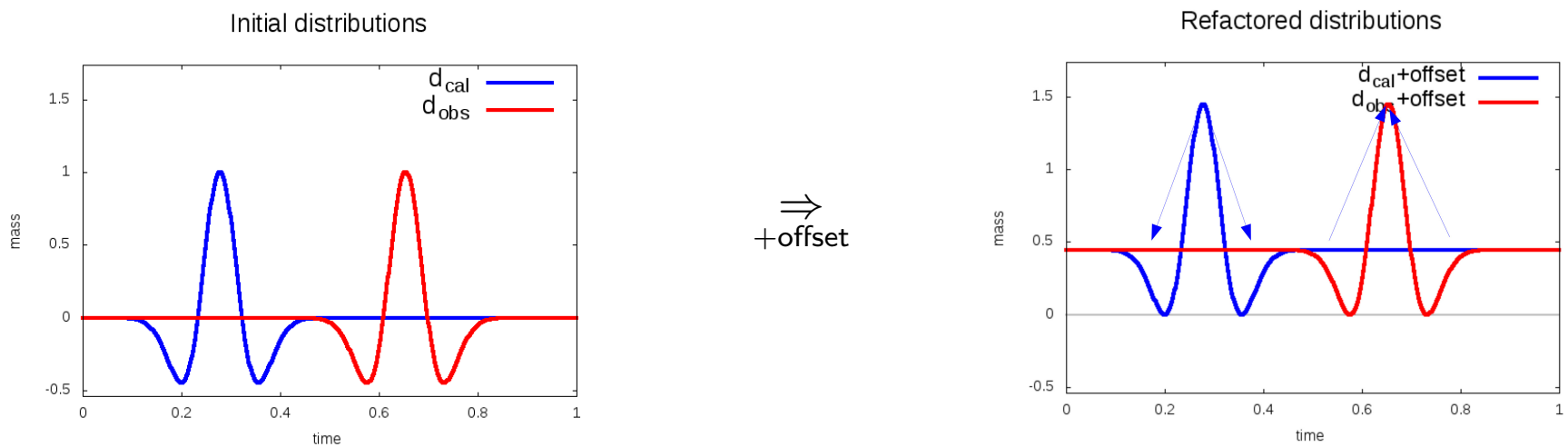


Outline

- Addition of a **positive constant mass** to make the signed measures strictly positive

$$\widetilde{W}(d_{cal}, d_{obs}) := W(d_{cal} + \alpha, d_{obs} + \alpha)$$

$$\text{with } \alpha > |\min(d_{cal}, d_{obs})|$$



Drawbacks: the transformation becomes local: no more transportation along the time axis: we loose the convexity with respect to time shifts

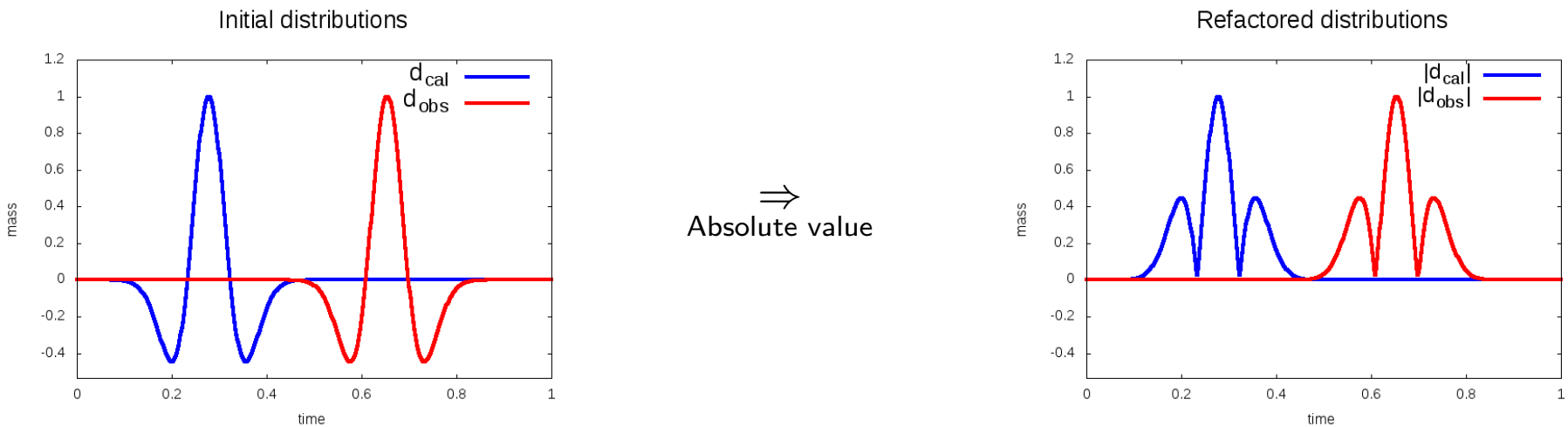
- In the literature, we have found 3 ideas based on the separation of positive and negative parts of the signals (Ambrosio et al., 2011; Mainini, 2012)

  1. taking the absolute value of the signal

  2. transport separately positive and negative part of the signal (Engquist and Froese, 2014)

  3. recombine the data using the decomposition in positive and negative part to compare positive measures with mass conservation (Mainini, 2012)

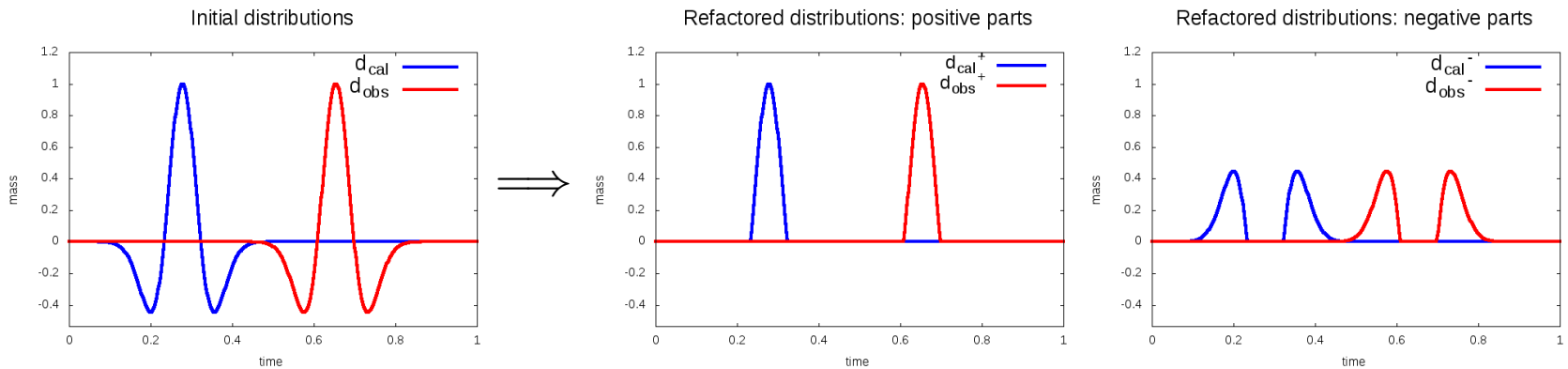
$$W_{absolute}(d_{cal}, d_{obs}) := W(|p|, |q|)$$



Drawbacks: loss of polarity information: no sensitivity to impedance contrast

$$W_{Engquist}(d_{cal}, d_{obs}) := W(d_{cal}^{+}, d_{obs}^{+}) + W(d_{cal}^{-}, d_{obs}^{-})$$

with $d_{cal} = d_{cal}^{+} - d_{cal}^{-}$ and $d_{obs} = d_{obs}^{+} - d_{obs}^{-}$



Drawbacks:

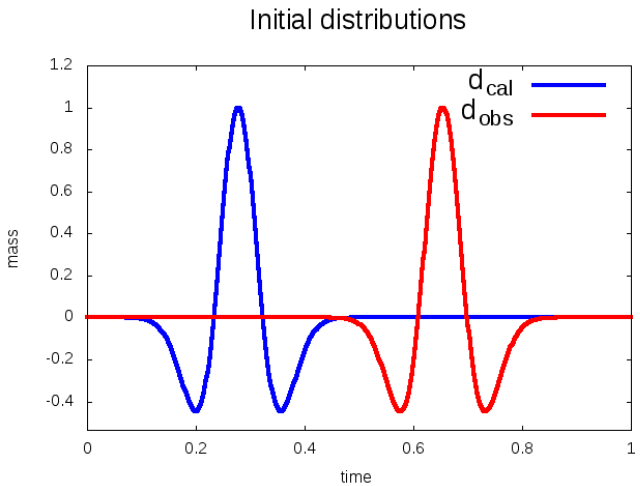
- we lose the mass conservation

$$\int p^{+} \neq \int q^{+}, \quad \int p^{-} \neq \int q^{-} \tag{19}$$

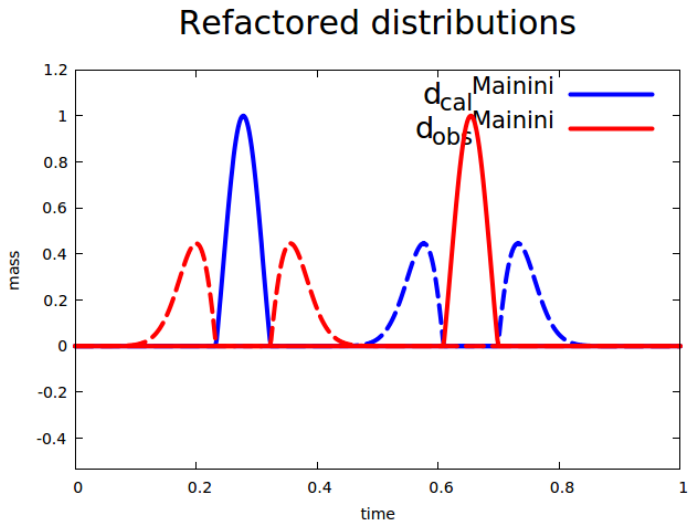
- artificial decorrelation between positive and negative part of the signal

$$W_{Mainini}(d_{cal}, d_{obs}) := W(d^+_{cal} + d^-_{obs}, d^+_{obs} + d^-_{cal})$$

with $d_{cal} = d^+_{cal} - d^-_{cal}$ and $d_{obs} = d^+_{obs} - d^-_{obs}$



$\Rightarrow$ Mainini



Advantages:

- positivity and mass conservation

$$\int d^+_{cal} - d^-_{cal} = \int d^+_{obs} - d^-_{obs} \Leftrightarrow \int d^+_{cal} + d^-_{obs} = \int d^+_{obs} + d^-_{cal} \quad (20)$$

- Important: the Mainini strategy for the W_1 distance is equivalent to the dual strategy we have employed so far

$$
\begin{aligned}
W^1_{Mainini}(d_{cal}, d_{obs}) &= W^1(d^+_{cal} + d^-_{obs}, d^-_{cal} + d^+_{obs}) \\
&= \max_{\varphi \in \mathrm{Lip}_1} \sum_{i=1}^{N} \varphi_i \left(d^+_{cal,i} + d^-_{obs,i} - (d^-_{cal,i} + d^+_{obs,i}) \right) \\
&= \max_{\varphi \in \mathrm{Lip}_1} \sum_{i=1}^{N} \varphi_i \left(\underbrace{d^+_{cal,i} - d^-_{cal,i}}_{d_{cal}} - (\underbrace{d^+_{obs_i} - d^-_{obs_i}}_{d_{obs}}) \right) \\
&= \max_{\varphi \in \mathrm{Lip}_1} \sum_{i=1}^{N} \varphi_i \left(d_{cal} - d_{obs} \right) \\
&= W^1(d_{cal}, d_{obs})
\end{aligned}
$$

- We have obtained encouraging results with W^1 and a ℓ_1 ground cost
- What about W^2 using the Mainini decomposition ? $\longrightarrow$ more convex / smoother w.r.t time shift?

To do so, we now need an efficient numerical strategy for general W^p problems

Conventional approaches for large scale transport problems:

- **Monge-Ampère** equation (Philippis and Figalli, 2014)
- **Benamou-Brenier** formulation (Benamou and Brenier, 2000)
- **Entropic regularization** (Cuturi, 2013; Benamou et al., 2015)

- We have obtained encouraging results with W^1 and a ℓ_1 ground cost
- What about W^2 using the Mainini decomposition ? $\rightarrow$ more convex / smoother w.r.t time shift?

To do so, we now need an efficient numerical strategy for general W^p problems

Conventional approaches for large scale transport problems:

- **Monge-Ampère** equation (Philippis and Figalli, 2014)
- **Benamou-Brenier** formulation (Benamou and Brenier, 2000)
- **Entropic regularization** (Cuturi, 2013; Benamou et al., 2015)

- **Entropic regularized problem** (Cuturi, 2013; Benamou et al., 2015)

$$W_\varepsilon^p(d_{cal}, d_{obs}) = \min_{\gamma \in \Pi(d_{cal}, d_{obs})} \sum_{i,j=1}^{N} C_{ij}\gamma_{ij} + \underbrace{\varepsilon}_{\text{regularization}} \underbrace{\sum_{i,j=1}^{N} \gamma_{ij}(\log(\gamma_{ij}) - 1)}_{\text{entropy}} \tag{21}$$

with:

- $\Pi(d_{cal}, d_{obs}) := \left\{ \gamma \in \mathbb{R}_{\geq 0}^{N \times N} \; ; \; \sum_{i=1}^{N} \gamma_{ij} = d_{cal,i} \; , \; \sum_{j=1}^{N} \gamma_{ij} = d_{obs,j} \right\}$
- $C_{ij} := \text{dist}(x_i, y_j)$
- $\varepsilon > 0$
- if $\gamma_{ij} = 0$, the convention is that $0 \log(0) = 0$

- An alternate projection algorithm based on the Sinkhorn iteration has been proposed to solve this problem (Bregman, 1967; Sinkhorn and Knopp, 1967; Cuturi, 2013; Benamou et al., 2015)

$$\gamma_\varepsilon^* = \lim_{n \to +\infty} \gamma^{(n)} = \lim_{n \to +\infty} \text{diag}\left(u^{(n)}\right) K \, \text{diag}\left(v^{(n)}\right) \tag{22}$$

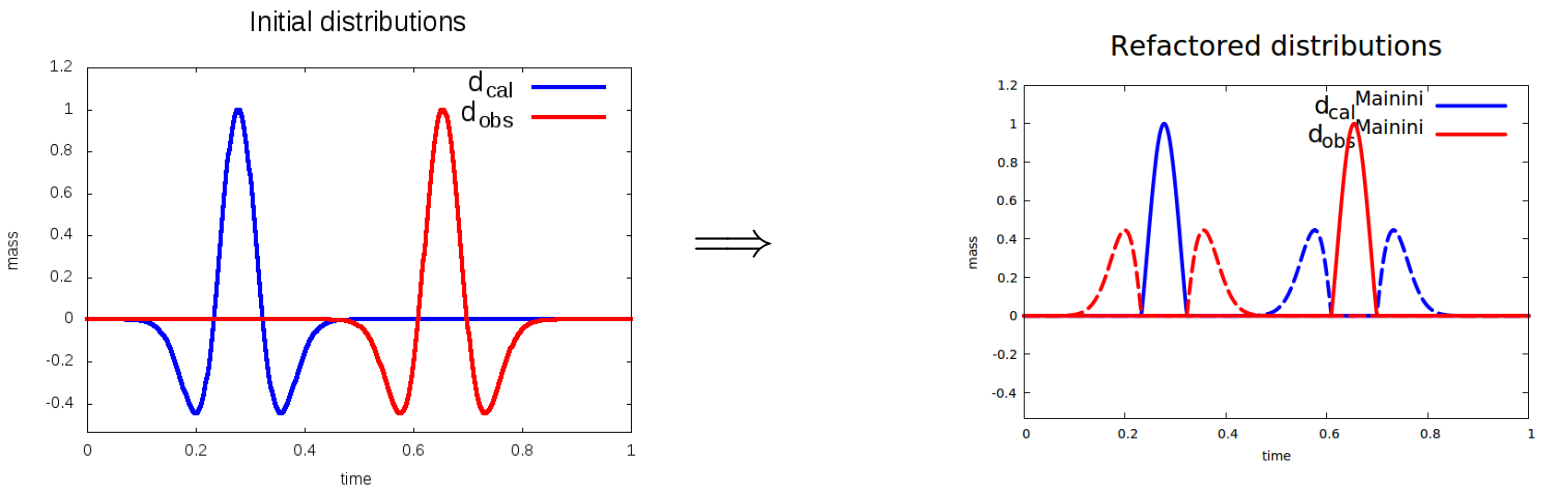
$$K_{ij} = e^{-\frac{C_{ij}}{\varepsilon}} \tag{23}$$

- Algorithm
  - $u_i^{(n+1)} = \dfrac{d_{cal,i}}{(K v^{(n)})_i}$
  - $v_j^{(n+1)} = \dfrac{d_{obs,j}}{(K^T u^{(n+1)})_j}$
  - $u^{(n)} \in \mathbb{R}^N$, $v^{(n)} \in \mathbb{R}^N$, $\mathsf{K} \in \mathbb{R}^{N \times N}$
  - $u^{(0)} = v^{(0)} = (1, ..., 1)^T \in \mathbb{R}^N$

The computational complexity depends on the product: Kv

- **naive implementation** $Kv \rightarrow \mathcal{O}(N^2)$
- exploiting the **symmetric Toeplitz** structure of $K \rightarrow \mathcal{O}(N \log(N))$ (FFT acceleration, no approximation)
- exploiting the **sparsity** of $K \rightarrow \mathcal{O}(N)$ (approximation)

- We come back to 1D time-shifted Ricker functions now with the Mainini (2012) approach



Comparison between Mainini cost with W_1 and W_2 computed through the entropic regularization approach (Cuturi, 2013; Benamou et al., 2015)

Outline

- Optimal transport appears as an interesting approach to mitigate cycle skipping in FWI
- The main difficulty is however to deal with non-positive meausres: the seismic data is oscillatory

- Optimal transport appears as an interesting approach to mitigate cycle skipping in FWI
- The main difficulty is however to deal with non-positive meausres: the seismic data is oscillatory

- A first implementation based on the dual form of the W_1 distance has been proposed in (Métivier et al., 2016a,b,c)
- In this presentation, we have shown that this is actually a special case of a more general approach proposed by Mainini (2012) to extend optimal transport to signed measures

- Optimal transport appears as an interesting approach to mitigate cycle skipping in FWI
- The main difficulty is however to deal with non-positive meausres: the seismic data is oscillatory

- A first implementation based on the dual form of the W_1 distance has been proposed in (Métivier et al., 2016a,b,c)
- In this presentation, we have shown that this is actually a special case of a more general approach proposed by Mainini (2012) to extend optimal transport to signed measures

- This opens the way to work with more general $W_p, p > 1$ distance, that might be more convex and/or smoother than W_1
- As a preliminary step, we have focused on the entropic regularization approach (Cuturi, 2013; Benamou et al., 2015), an efficient strategy to solve general large scale optimal transport problems

Thank you for your attention

- IDRIS and TGCC, French national computing centers
- CIMENT, Grenoble computing center
- SEISCOPE sponsors : `http://seiscope2.osug.fr`

Questions?

Adams, J. C. (1989). MUDPACK: Multigrid portable FORTRAN software for the efficient solution of linear elliptic partial differential equations. *Applied Mathematics and Computation*, 34(2):113–146.

Ambrosio, L., Mainini, E., and Serfaty, S. (2011). Gradient flow of the Chapman–Rubinstein–Schatzman model for signed vortices. *Annales de l'Institut Henri Poincaré (C) Non Linear Analysis*, 28(2):217–246.

Benamou, J.-D. and Brenier, Y. (2000). A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G. (2015). Iterative Bregman Projections for Regularized Transportation Problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138.

Brandt, A. (1977). Multi-level adaptive solutions to boundary-value problems. *Mathematics of Computation*, 31:333–390.

Bregman, L. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*.

Combettes, P. L. and Pesquet, J.-C. (2011). Proximal splitting methods in signal processing. In Bauschke, H. H., Burachik, R. S., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, volume 49 of *Springer Optimization and Its Applications*, pages 185–212. Springer New York.

Cuturi, M. (2013). Sinkhorn distances: lightspeed computation of optimal transportation distances. *Advances in Neural Information Processing Systems*.

Engquist, B. and Froese, B. D. (2014). Application of the wasserstein metric to seismic signals. *Communications in Mathematical Science*, 12(5):979–988.

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

Mainini, E. (2012). A description of transport cost for signed measures. *Journal of Mathematical Sciences*, 181(6):837–855.

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(12):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*, 205:345–377.

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.

Philippis, G. D. and Figalli, A. (2014). The monge-ampère equation and its link to optimal transportation. *BULLETIN (New Series) OF THE AMERICAN MATHEMATICAL SOCIETY*.

Santambrogio, F. (2015). *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Progress in Nonlinear Differential Equations and Their Applications. Springer International Publishing.

Sinkhorn, R. and Knopp, P. (1967). Concerning nonnegative matrices and doubly stochastic matrices. *Pacific Journal of Mathematics*.

Swarztrauber, P. N. (1974). A Direct Method for the Discrete Solution of Separable Elliptic Equations. *SIAM Journal on Numerical Analysis*, 11(6):1136–1150.