A dual formulation for time-domain wavefield reconstruction inversion

Gabrio Rizzuti[*,1], Mathias Louboutin[1], Rongrong Wang[2], Emmanouil Daskalakis[3], and Felix J. Herrmann[1]

[1] Georgia Institute of Technology
[2] Michigan State University
[3] University of British Columbia

SLIM

Houston, SIAM GS2019
03/11/2019

PDE-constrained optimization

$$\inf_{m,u} \frac{1}{2}||d - R\,u||^2 \quad \text{s.t.} \quad A(m)u = S^* f$$

d	:	data	$A(m)$	:	wave equation
f	:	source	R	:	receiver interp.
m	:	model	S	:	source interp.
u	:	wavefield			

PDE-constrained optimization (all-at-once full-space)

$$\mathcal{L}(m,u,v) = \frac{1}{2}||d - R\,u||^2 + \langle v, S^* f - A(m)\,u \rangle$$

[Haber, E., and Ascher, U. M., 2001; Biros, G., and Ghattas, O., 2005; Grote et al, 2011]

Pros:
- ✓ **simultaneous** update of model, wavefield, multiplier
- ✓ **cheap** evaluation and gradient computation
- ✓ sparse Hessian

Cons:
- ✗ **storage of fields and multipliers**
  (for each source and time/frequency sample)
- ✗ **updates** computationally demanding

PDE-constrained optimization (reduced space)

$$J(m) = \frac{1}{2}||d - F(m)\, S^* f||^2 \qquad F(m) = RA(m)^{-1}$$

forward map

[Tarantola, A., ‘84; Haber, E., et al, 2000; Epanomeritakis, I., et al, 2008]

Pros:
- ✓ low(er) **storage** requirements

Cons:
- ✗ highly **non-convex** (needs good starting model)
- ✗ requires exact **PDE solutions**
- ✗ dense Hessian

PDE-constrained optimization (penalty method)

$$J_\lambda(m,u) = \frac{1}{2}||d - R\,u||^2 + \frac{\lambda^2}{2}\underbrace{||S^* f - A(m)\,u||^2}_{\text{PDE penalty}}$$

[van den Berg, P. M., and Kleinman, R. E., 1997; van Leeuwen, T. and Herrmann, F. J., 2013]

PDE-constrained optimization (wavefield reconstruction inversion)

$$\bar{J}_\lambda(m) = \frac{1}{2}||d - R\,\bar{u}||^2 + \frac{\lambda^2}{2}||S^* f - A(m)\,\bar{u}||^2$$

$$\begin{bmatrix} R \\ \lambda A(m) \end{bmatrix} \bar{u} = \begin{bmatrix} d \\ \lambda S^* f \end{bmatrix}$$

augmented wave equation

[van Leeuwen, T. and Herrmann, F. J., 2013]

Pros:
- ✓ no simultaneous field **storage** for all sources
- ✓ sparse Gauss-Newton Hessian

Cons:
- ✗ continuation strategy on λ
- ✗ need **augmented PDE solution:**
  - **frequency** domain: proper preconditioning?
  - **time** domain: no explicit time-marching scheme

Attainable model extension

Question: feasible model extension?

- convenient handling of slack variables
- feasible computation of augmented PDE (e.g., can be solved by time-marching)

Wavefield Reconstruction Inversion (time domain approximation)

$$J_\lambda(m) = \frac{1}{2}||d - F(m)\,\bar{q}||^2 + \frac{\lambda^2}{2}||S^* f - \bar{q}||^2$$

$$\begin{bmatrix} F(m) \\ \lambda I \end{bmatrix} \bar{q} = \begin{bmatrix} d \\ \lambda S^* f \end{bmatrix}$$

augmented wave equation

[Wang et al, 2016]

$\bar{q}$ ~ generalized source
(see also [Huang et al, 2018])

First order Taylor expansion wrt $1/\lambda^2$:

$$\bar{q} \approx S^* f + \frac{1}{\lambda^2} F(m)^* (d - F(m) S^* f), \quad \lambda \to +\infty$$

Denoising reformulation of WRI

$$\inf_{m,u} \underbrace{\frac{1}{2}||S^* f - A(m)u||^2}_{\text{PDE misfit}} \quad \text{s.t.} \quad \underbrace{||d - R\,u|| \leq \varepsilon}_{\text{data constraint}}$$

[Wang, R., and Herrmann, F. J., 2017]

Dual formulation of WRI - Lagrangian

$$\inf_{m,u} \frac{1}{2} \underbrace{||S^* f - A(m)u||^2}_{\text{PDE misfit}} \quad \text{s.t.} \quad \underbrace{||d - R\,u|| \leq \varepsilon}_{\text{data constraint}}$$

Saddle-point problem

$$\inf_{m,u} \sup_{y} \mathcal{L}(m,u,y)$$

$$\mathcal{L}(m,u,y) = \frac{1}{2}||S^* f - A(m)u||^2 + \langle y, d - R\,u \rangle - \varepsilon ||y||$$

Dual formulation of WRI – Augmented wave equation

SLIM

Solving for u ...

$$A(m)\bar{u} = S^* f + F(m)^* y$$

augmented wave equation

Data interfere with the system!

Saddle-point problem

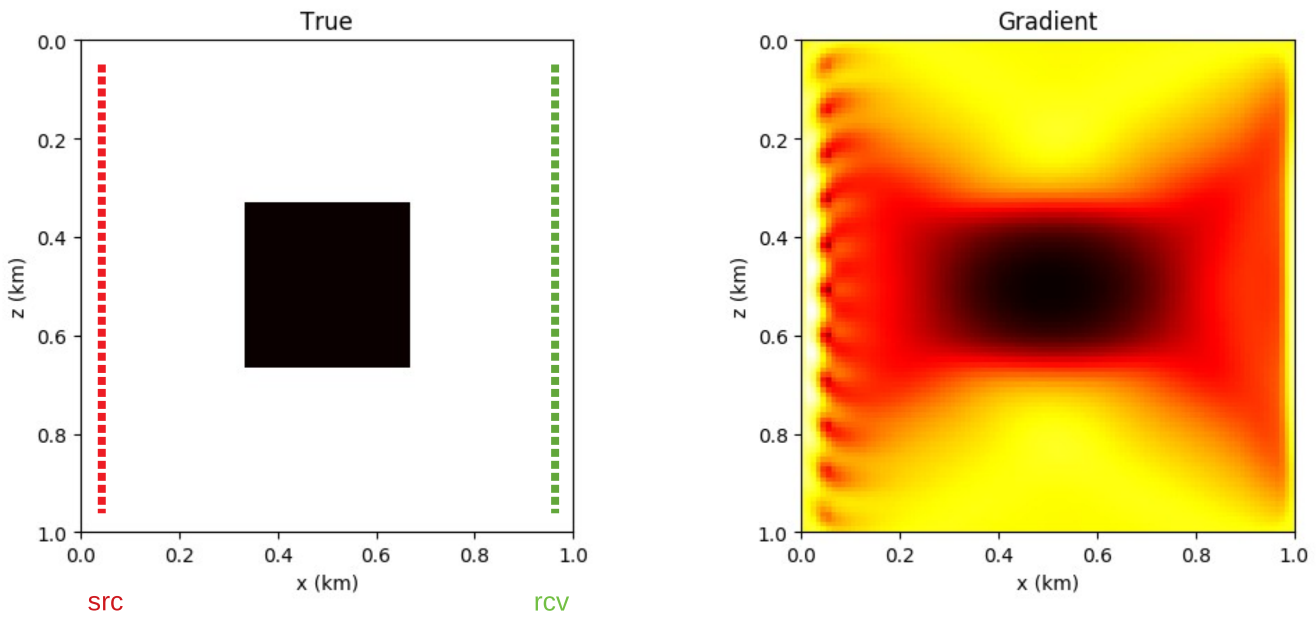
$$\inf_{m,u} \sup_{y} \mathcal{L}(m,u,y) \geq \inf_{m} \sup_{y} \inf_{u} \mathcal{L}(m,u,y)$$

$$\mathcal{L}(m,u,y) = \frac{1}{2}||S^* f - A(m)u||^2 + \langle y, d - Ru \rangle - \varepsilon ||y||$$

Dual formulation of WRI – Augmented wave equation

Transmission data problem:

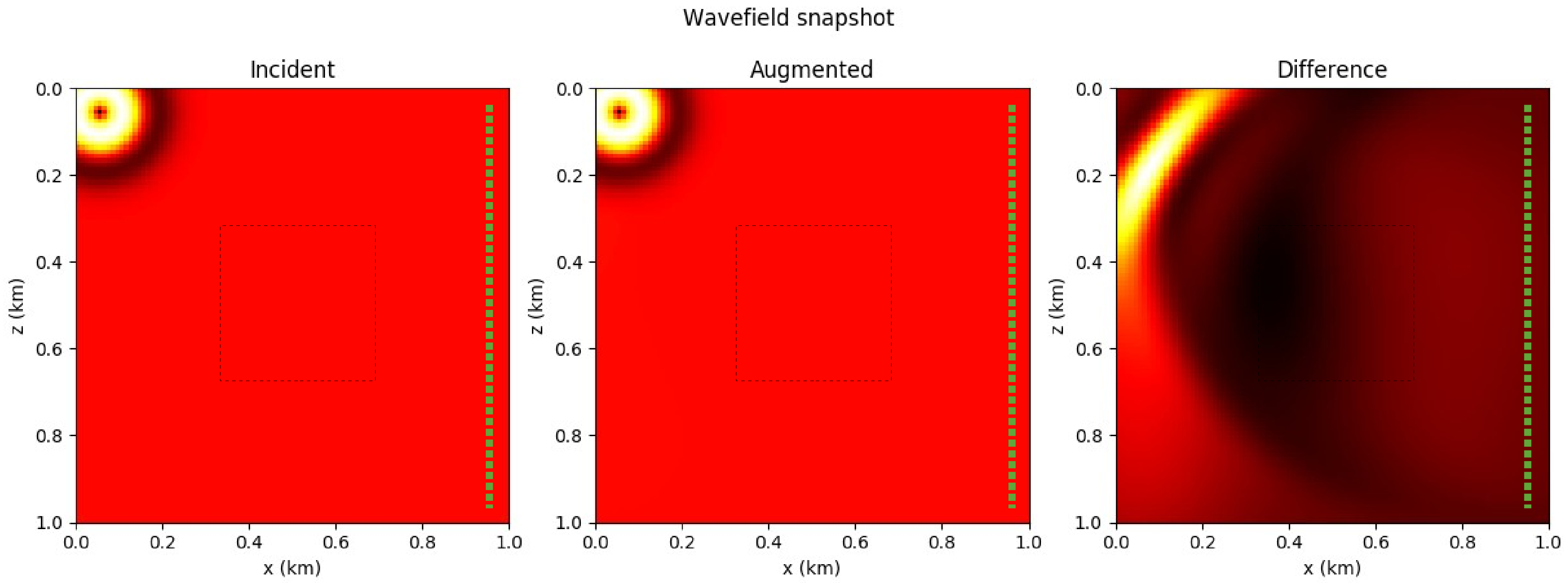
- 1 km x 1 km model
- background velocity: 2000 m/s, inclusion 2100 m/s
- 11 sources (10 Hz peak), 51 receivers
- y = background data residual



Dual formulation of WRI – Augmented wave equation

Transmission data problem:

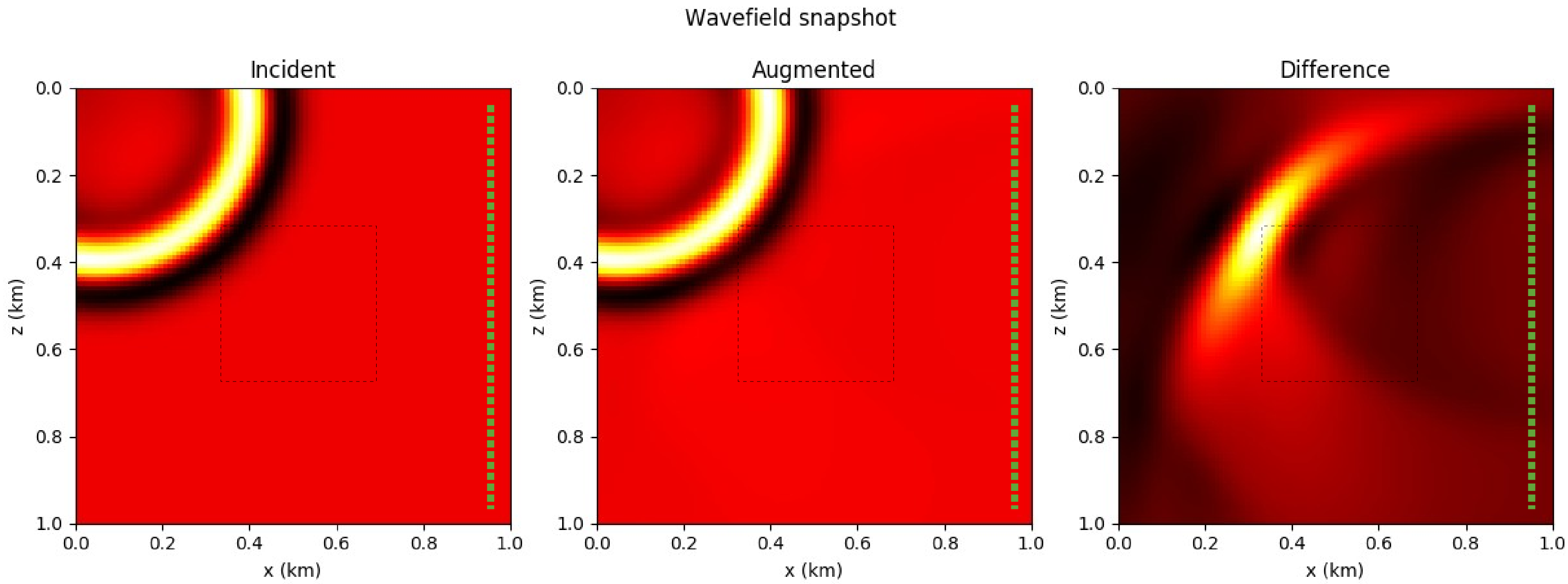
- 1 km x 1 km model
- background velocity: 2000 m/s, inclusion 2100 m/s
- 11 sources (10 Hz peak), 51 receivers
- y = background data residual



Dual formulation of WRI – Augmented wave equation

Transmission data problem:

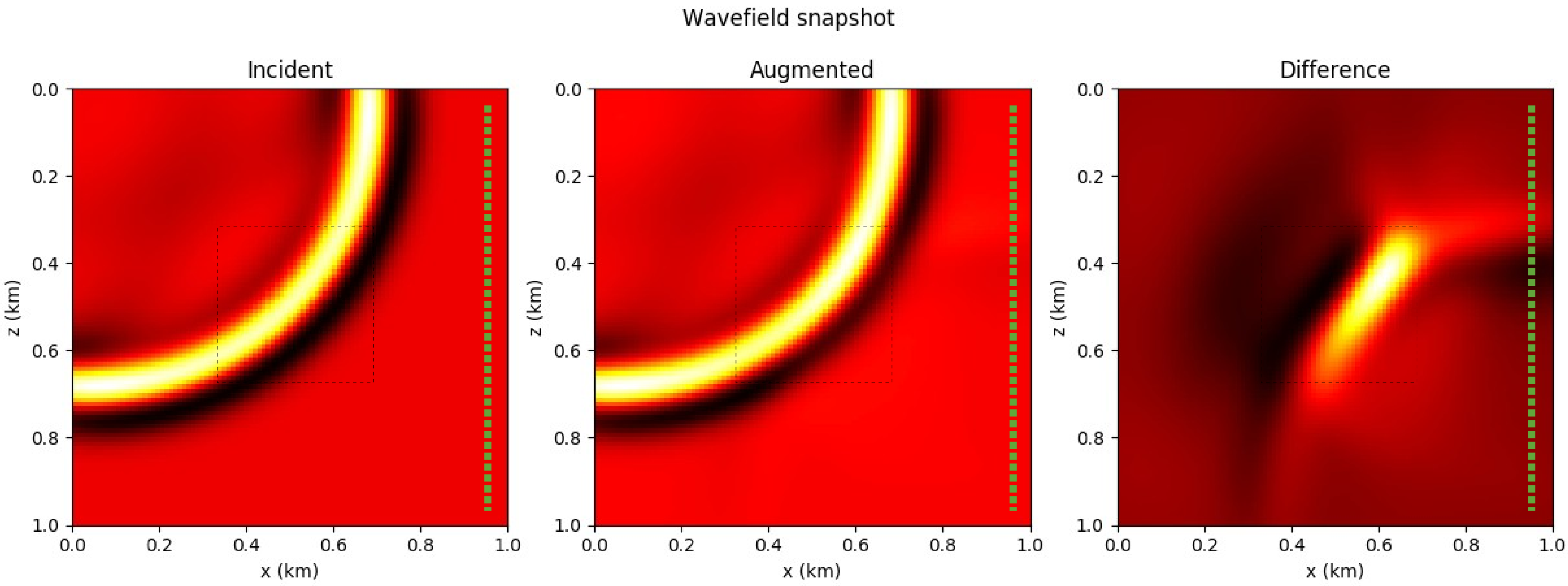
- 1 km x 1 km model
- background velocity: 2000 m/s, inclusion 2100 m/s
- 11 sources (10 Hz peak), 51 receivers
- y = background data residual



Dual formulation of WRI – Augmented wave equation

Transmission data problem:

- 1 km x 1 km model
- background velocity: 2000 m/s, inclusion 2100 m/s
- 11 sources (10 Hz peak), 51 receivers
- y = background data residual



Dual formulation of WRI – Objective and gradients

Dual saddle-point formulation

$$\inf_m \sup_y \bar{\mathcal{L}}(m,y)$$

$$\bar{\mathcal{L}}(m,y) = -\frac{1}{2}||F(m)^* y||^2 + \langle y, d - F(m)S^* f\rangle - \varepsilon||y||$$

Gradients

$$\nabla_m \bar{\mathcal{L}} = -\mathrm{Jac}(m, S^* f + F(m)^* y)^* y$$

~ conventional FWI gradient

$$\nabla_y \bar{\mathcal{L}} = d - F(m)(S^* f + F(m)^* y)$$

generalized-source data residual

$$\mathrm{Jac}(m,q) = \frac{\mathrm{d}}{\mathrm{d}m}(\cdot \mapsto F(\cdot)\, q)\Big|_m$$

Dual formulation of WRI (recap)

Dual saddle-point formulation

$$\inf_m \sup_y \bar{\mathcal{L}}(m,y)$$

$$\bar{\mathcal{L}}(m,y) = -\frac{1}{2}||F(m)^* y||^2 + \langle y, d - F(m)S^* f\rangle - \varepsilon||y||$$

Obtained **model extension** along **data space**:

$$\bar{\mathcal{L}} : M \times D \to \mathbb{R}$$

- ✓ amenable to **time-domain** methods
- ✓ extra variable **storage is affordable**
- ✓ **no continuation strategy** needed for relaxation parameter

- ✗ extra **time complexity** (2x PDE solutions, wrt FWI)
- ✗ dense Hessian
- ✗ need to figure out best **optimization strategy** (work in progress ...)

Dual formulation of WRI – Dual variable scaling

Dual saddle-point formulation

$$\inf_m \sup_y \bar{\mathcal{L}}(m,y)$$

$$\bar{\mathcal{L}}(m,y) = -\frac{1}{2}||F(m)^* y||^2 + \langle y, d - F(m)S^* f\rangle - \varepsilon||y||$$

Preconditioning/**Scaling issue** in Lagrangian/augmented wave equation:

$$A(m)\bar{u} = S^* f + \boxed{\alpha} F(m)^* y$$

Adaptive strategy?

Dual formulation of WRI – Dual variable scaling

Dual saddle-point formulation

$$\inf_m \sup_y \bar{\mathcal{L}}(m, y)$$

$$\bar{\mathcal{L}}(m, y) = -\frac{1}{2}||F(m)^* y||^2 + \langle y, \underbrace{d - F(m)S^* f}_{\text{r: data residual}} \rangle - \varepsilon ||y||$$

Solving for the scaling parameter

$$\bar{\bar{\mathcal{L}}}(m, y, \alpha) := \bar{\mathcal{L}}(m, \alpha y)$$

$$\bar{\bar{\bar{\mathcal{L}}}}(m, y) := \sup_\alpha \bar{\bar{\mathcal{L}}}(m, y, \alpha) = \bar{\bar{\mathcal{L}}}(m, y, \alpha(m, y))$$

$$\alpha(m, y) = \begin{cases} \dfrac{\langle y, r \rangle - \varepsilon \mathrm{sign}(\langle y, r \rangle)||y||}{||F(m)^* y||^2}, & |\langle y, r \rangle| \geq \varepsilon ||y|| \\ 0, & \text{otherwise} \end{cases}$$

Dual formulation of WRI (scaled)

Dual saddle-point formulation (scaled)

$$\inf_m \sup_y \bar{\bar{\mathcal{L}}}(m,y)$$

$$\bar{\bar{\mathcal{L}}}(m,y) = \begin{cases} \frac{1}{2}(|\langle \hat{y}, r\rangle| - \varepsilon||\hat{y}||)^2, & |\langle \hat{y}, r\rangle| \geq \varepsilon||\hat{y}|| \\ 0, & \text{otherwise} \end{cases} \quad \left(\hat{y} = \frac{y}{||F(m)^* y||}\right)$$

Spurious minima

Diving-wave example:

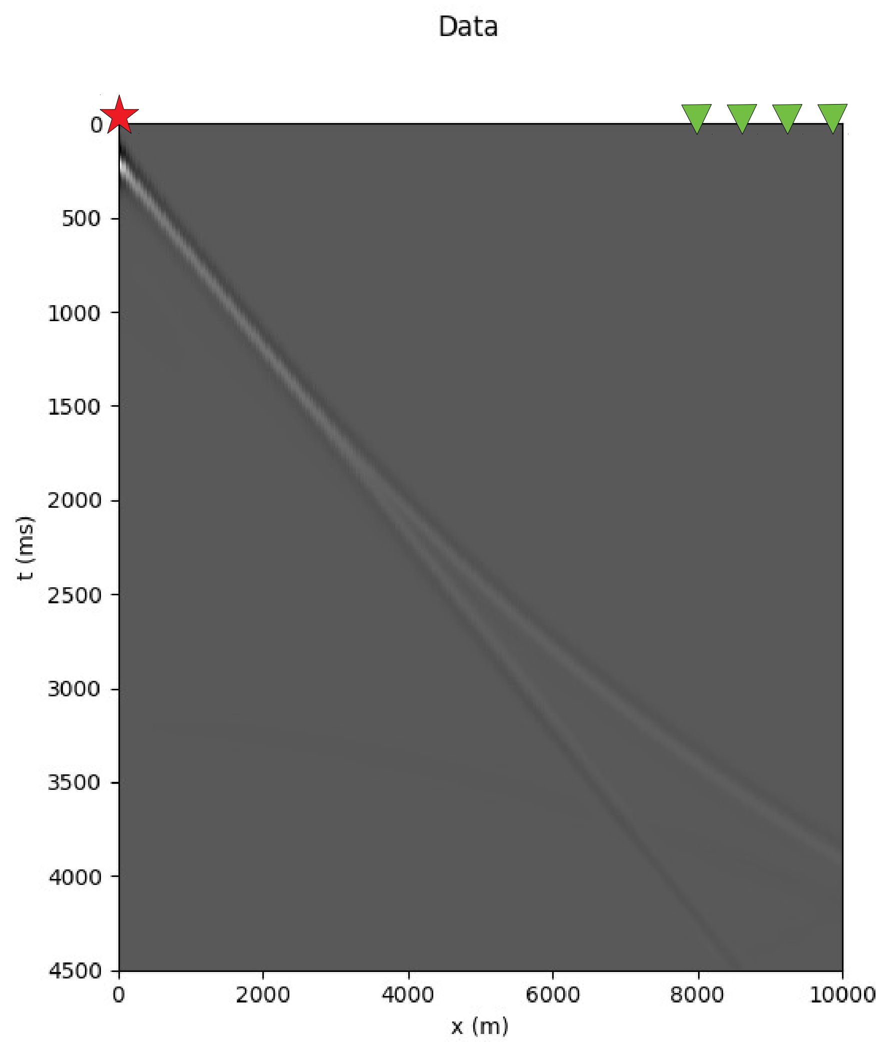
- 5 km x 10 km model
- linear velocity, varying slope:

$$v(x,z) = v_0 + \beta z$$

$$v_0 = 2\,\mathrm{km/s}, \quad 0.5 \leq \beta \leq 1$$

- single source (5 Hz peak), 21 receivers

➢ Plot misfit landscape as a function of slope



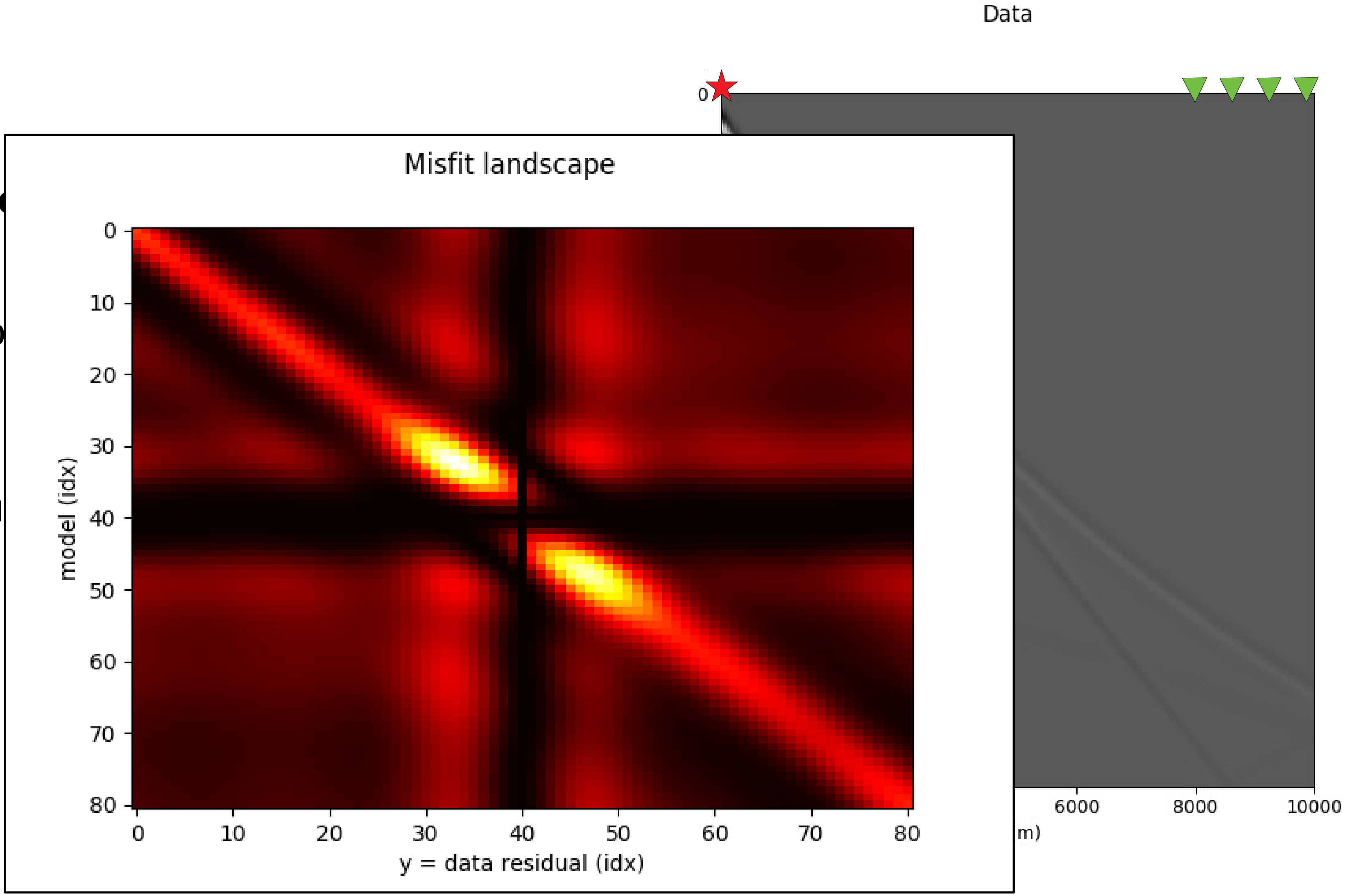
Spurious minima

Diving-wav

- 5 km x 10
- linear velo

- single sou

➢ Plot misfit



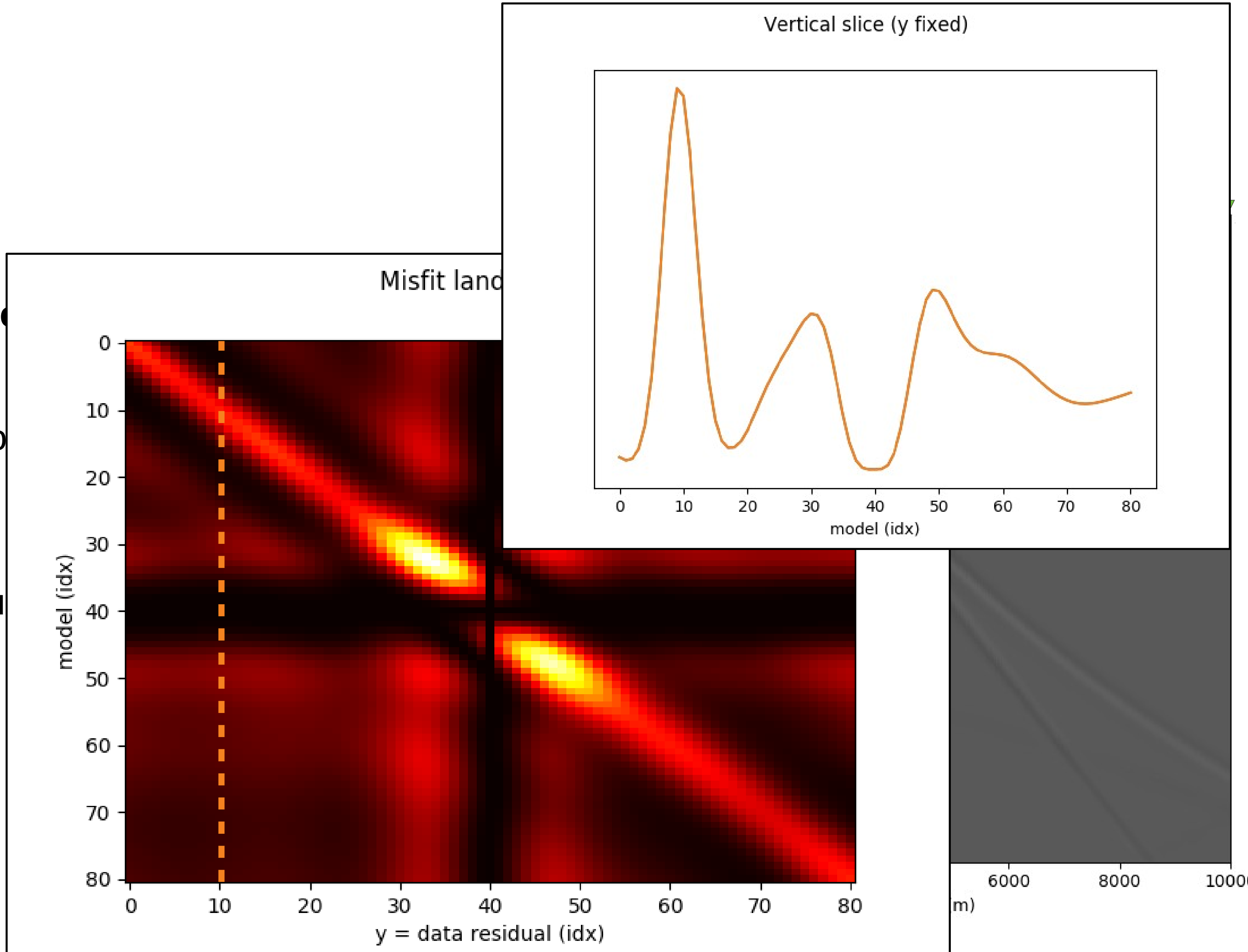
Spurious minima

Diving-wav

- 5 km x 10
- linear velo

- single sou

➢ Plot misfit



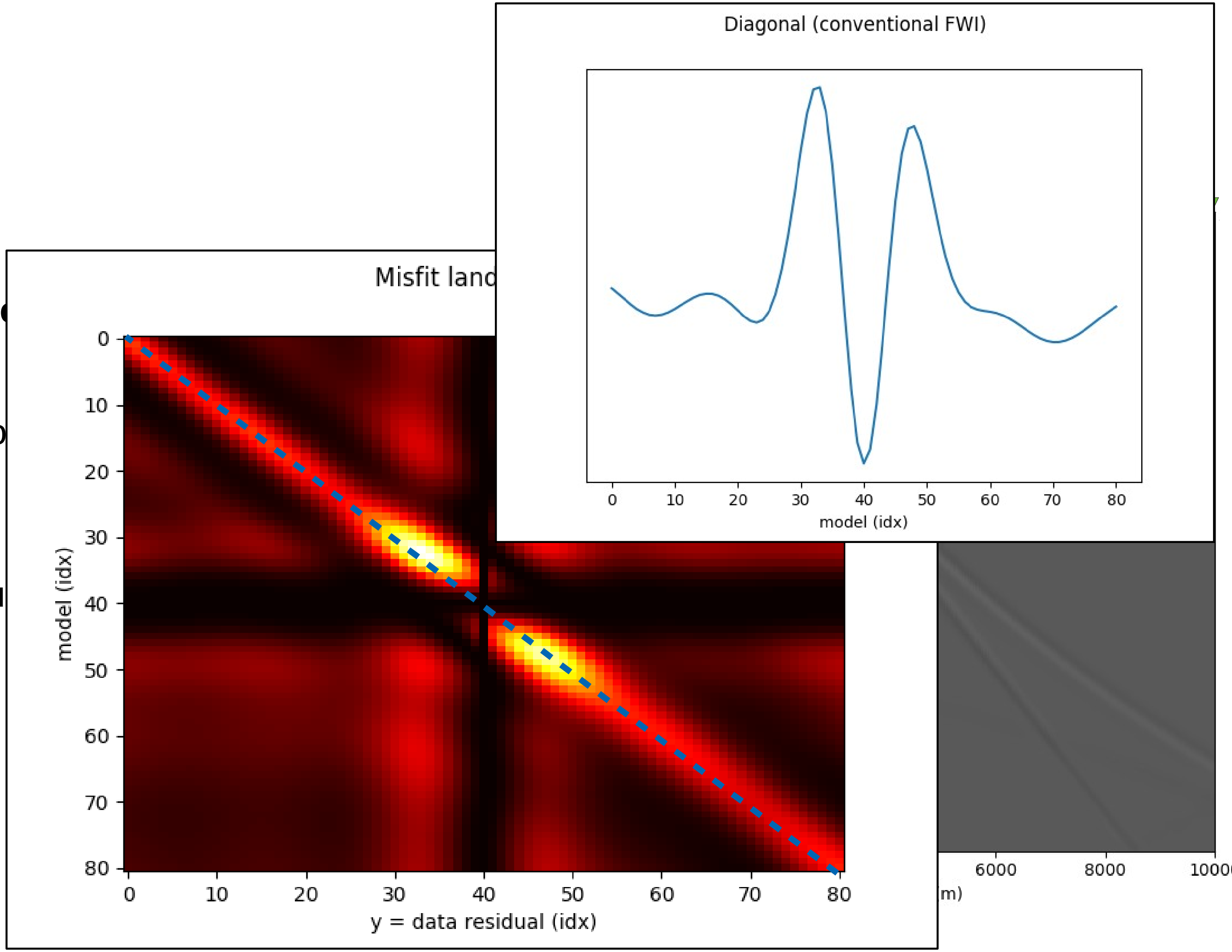
Spurious minima

Diving-wav

- 5 km x 10
- linear velo

- single sou

➢ Plot misfit



Reconstruction algorithm (sketch)

Alternating gradient descent

```
Given  m_0, y_0 = d - F(m_0)S^* f

for n = 1:Niter

    compute gradient wrt m:     ∇_{m_n} L̄̄

    update m (minimize):        m_n ← m_n - β∇_{m_n} L̄̄

    compute gradient wrt y:     ∇_{y_n} L̄̄

    update y (maximize):        y_n ← y_n + γ∇_{y_n} L̄̄
```

Given $m_0, y_0 = d - F(m_0)S^* f$

for n = 1:Niter

- compute gradient wrt m: $\nabla_{m_n} \bar{\bar{\mathcal{L}}}$
- update m (minimize): $m_n \leftarrow m_n - \beta \nabla_{m_n} \bar{\bar{\mathcal{L}}}$
- compute gradient wrt y: $\nabla_{y_n} \bar{\bar{\mathcal{L}}}$
- update y (maximize): $y_n \leftarrow y_n + \gamma \nabla_{y_n} \bar{\bar{\mathcal{L}}}$

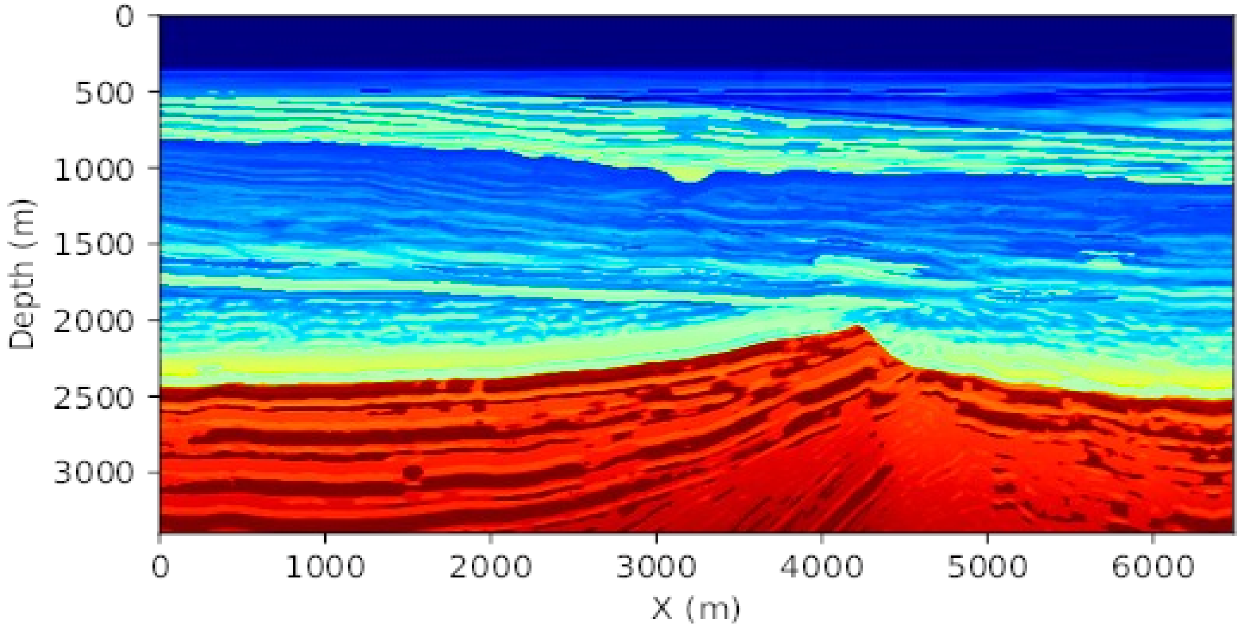
BG Compass (reflections/diving waves) **[preliminary results]**

BG compass:

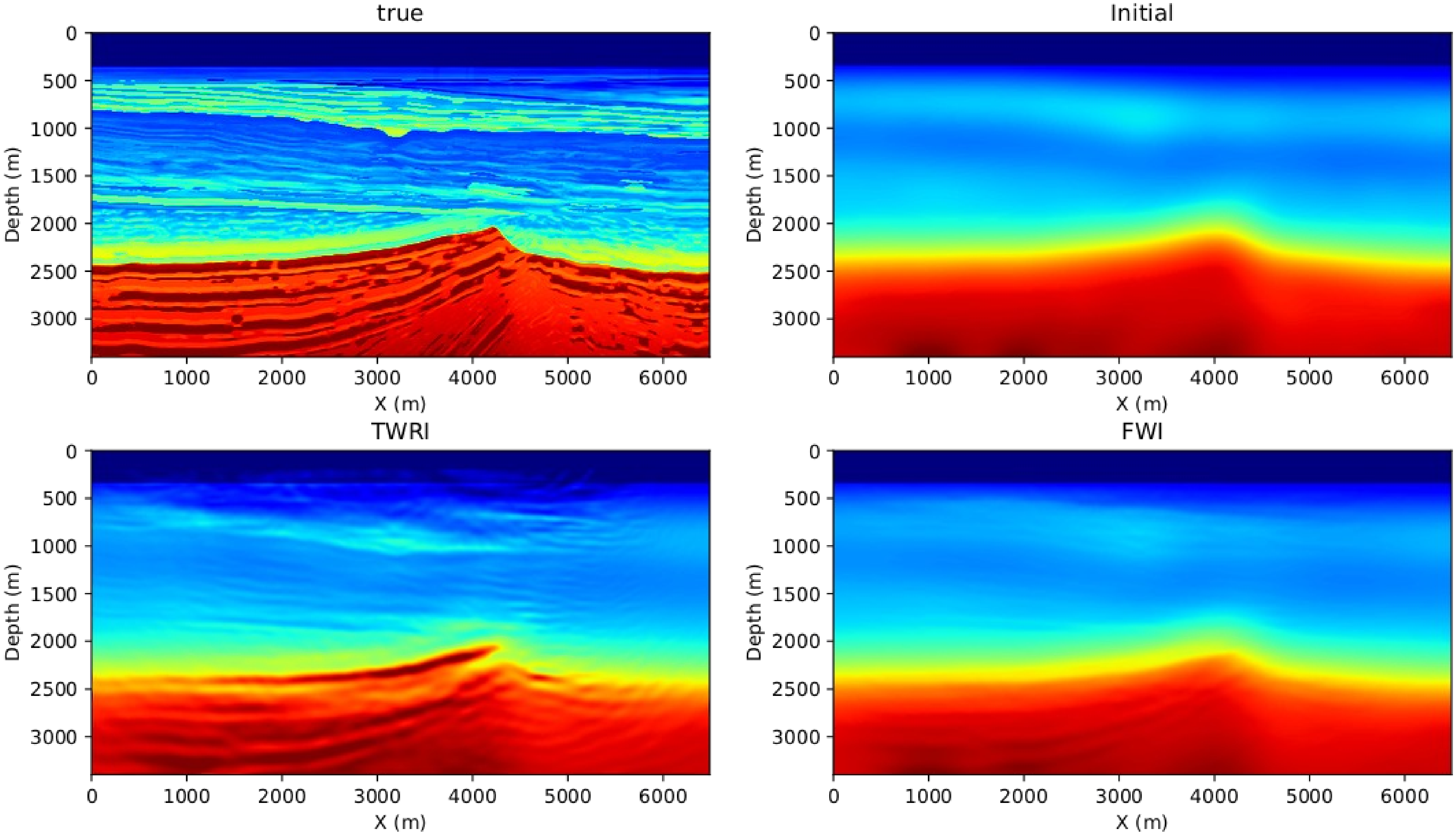
- ~ 3 km x 6 km model
- 25 sources (10 Hz peak), 251 receivers
- *challenging* for conventional FWI!

Optimization setup:

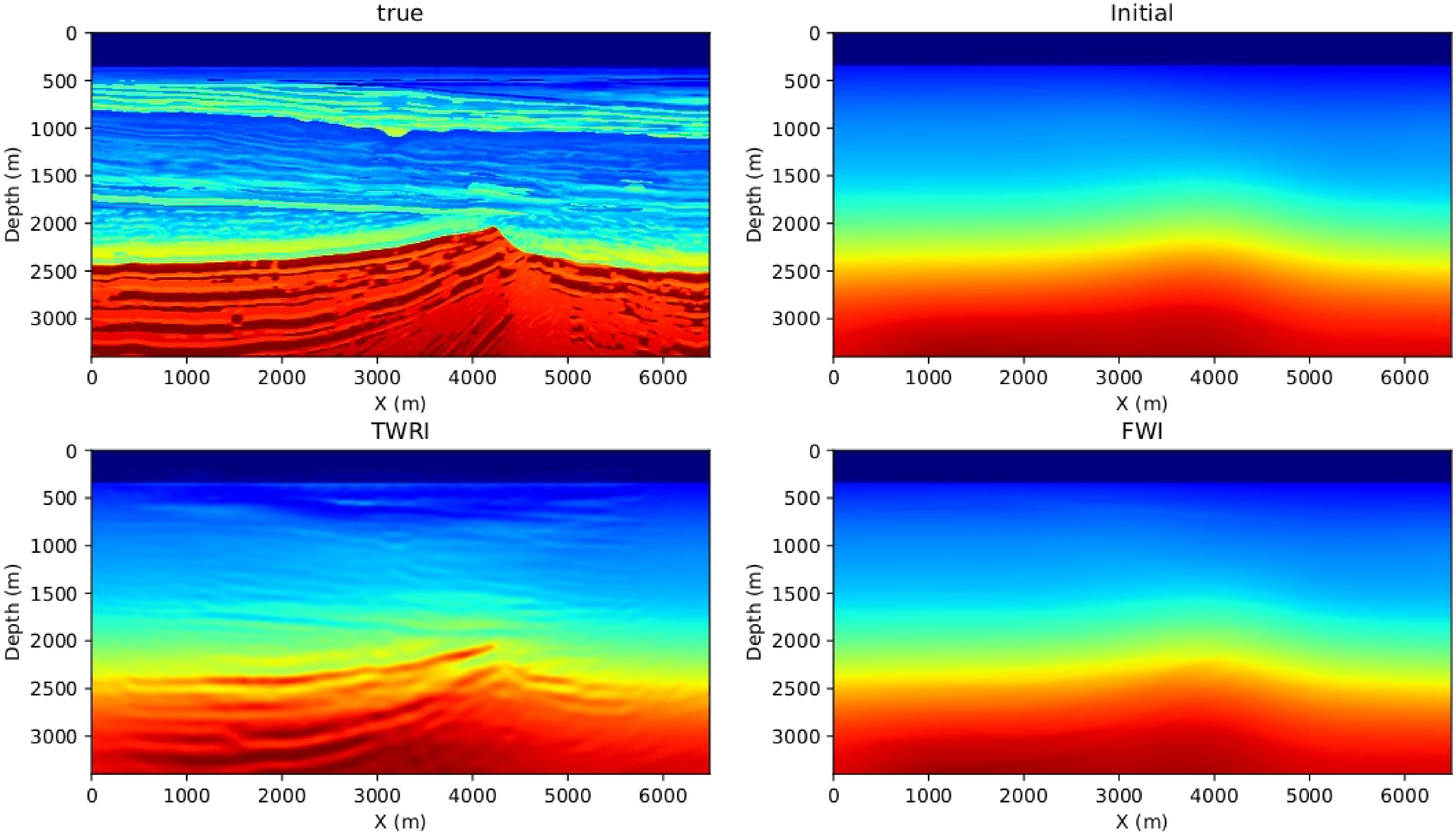
- *gradient descent* (fixed step length)
- just *10 iterations*



BG Compass **[preliminary results]**



BG Compass (worse background) **[preliminary results]**



Conclusions and road ahead

Presented a **reconstruction algorithm** potentially apt to large **3D** problems:

- based on **model extension** along data dimension: **robust to spurious minima**
- **storage** of slack variables is **affordable** (unlike full-space methods)
- computationally advantageous:
  PDEs can be solved in time domain with **explicit schemes** (unlike classical WRI)

To do:

- study of **optimization strategy**
- implement **constraints**

References

Haber, E., and Ascher, U. M., and Oldenburg, D., On optimization techniques for solving nonlinear inverse problems, Inverse Problems, 16 (2000)

Haber, E., and Ascher, U. M., Preconditioned all-at-once methods for large, sparse parameter estimation problems, Inverse Problems, 17 (2001)

Biros, G., and Ghattas, O., Parallel Lagrange-Newton-Krylov-Schur methods for PDE-constrained optimization. Part i: The Krylov-Schur solver, SIAM Journal on Scientific Computing, 27 (2005)

Grote, M. J., and Huber, J., and Schenk, O., Interior point methods for the inverse medium problem on massively parallel architectures, Procedia Computer Science, 4 (2011)

Tarantola, A., Inversion of seismic reflection data in the acoustic approximation, Geophysics 49(8) (1984)

Epanomeritakis, I., and Akcelik, V., and Ghattas, O., and Bielak, J., A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion, Inverse Problems, 24(3) (2008)

Kleinman, R. E., and van den Berg, P. M., A modified gradient method for two-dimensional problems in tomography, Journal of Computational and Applied Mathematics, 42 (1992)

van Leeuwen, T., and Herrmann, F. J., Mitigating local minima in full-waveform inversion by expanding the search space, Geophysical Journal International 195.1 (2013)

Peters, B., and Herrmann, F. J., and van Leeuwen, T., Wave-equation Based Inversion with the Penalty Method-Adjoint-state Versus Wavefield-Reconstruction Inversion, 76th EAGE Conference (2014)

Huang, G., and Nammour, R., and Symes, W. W., Volume source-based extended waveform inversion, Geophysics 83(5) (2018)

Wang, C., and Yingst, D., and Farmer, P., and Leveille, J., Full Waveform Inversion with the Reconstructed Wavefield Method, 86th EAGE Conference (2016)

Wang, R., and Herrmann, F. J., A denoising formulation of Full-Waveform Inversion, 87th SEG International Exposition (2017)