Relaxing the physics: A penalty method for full-waveform inversion

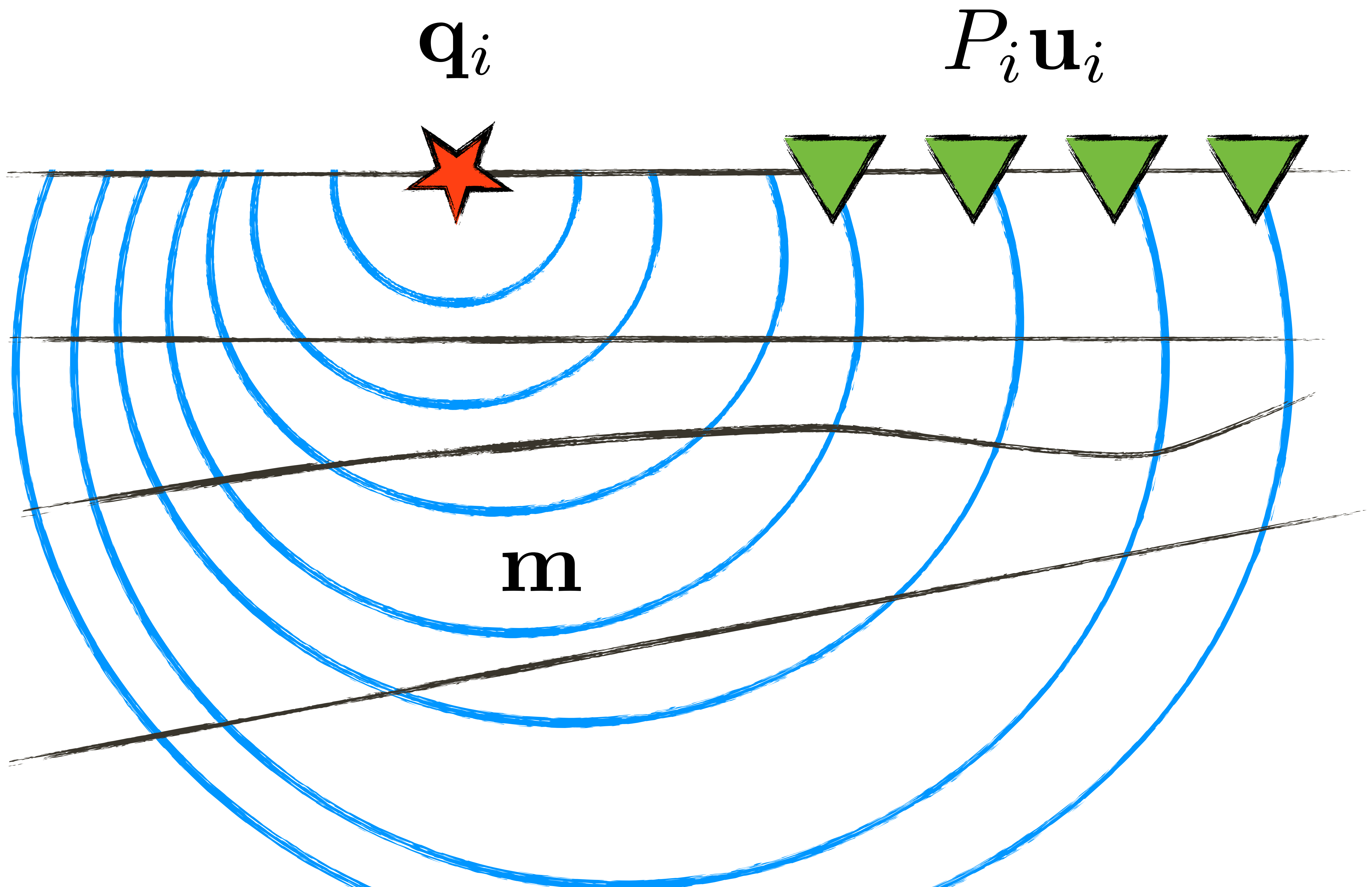
Tristan van Leeuwen
Felix Herrmann, Bas Peters

SLIM
University of British Columbia

Mitigating local minima in full-waveform inversion by expanding the search space, 2013, Tristan van Leeuwen, Felix J. Herrmann, Geophysical Journal International, Volume, 195, 661-667

Waveform inversion

Retrieve the medium parameters from *partial* measurements of the solution of the wave-equation: $A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$



Waveform inversion

- If we know the wavefields *everywhere*, we solve for $\mathbf{m}$ from

$$A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

- The challenge is reconstructing the wavefields from partial measurements

- Repeat in alternating fashion

[equation error approach]

Wavefield reconstruction

The wavefields...

- fit the data

$$P\mathbf{u}_i \approx \mathbf{d}_i$$

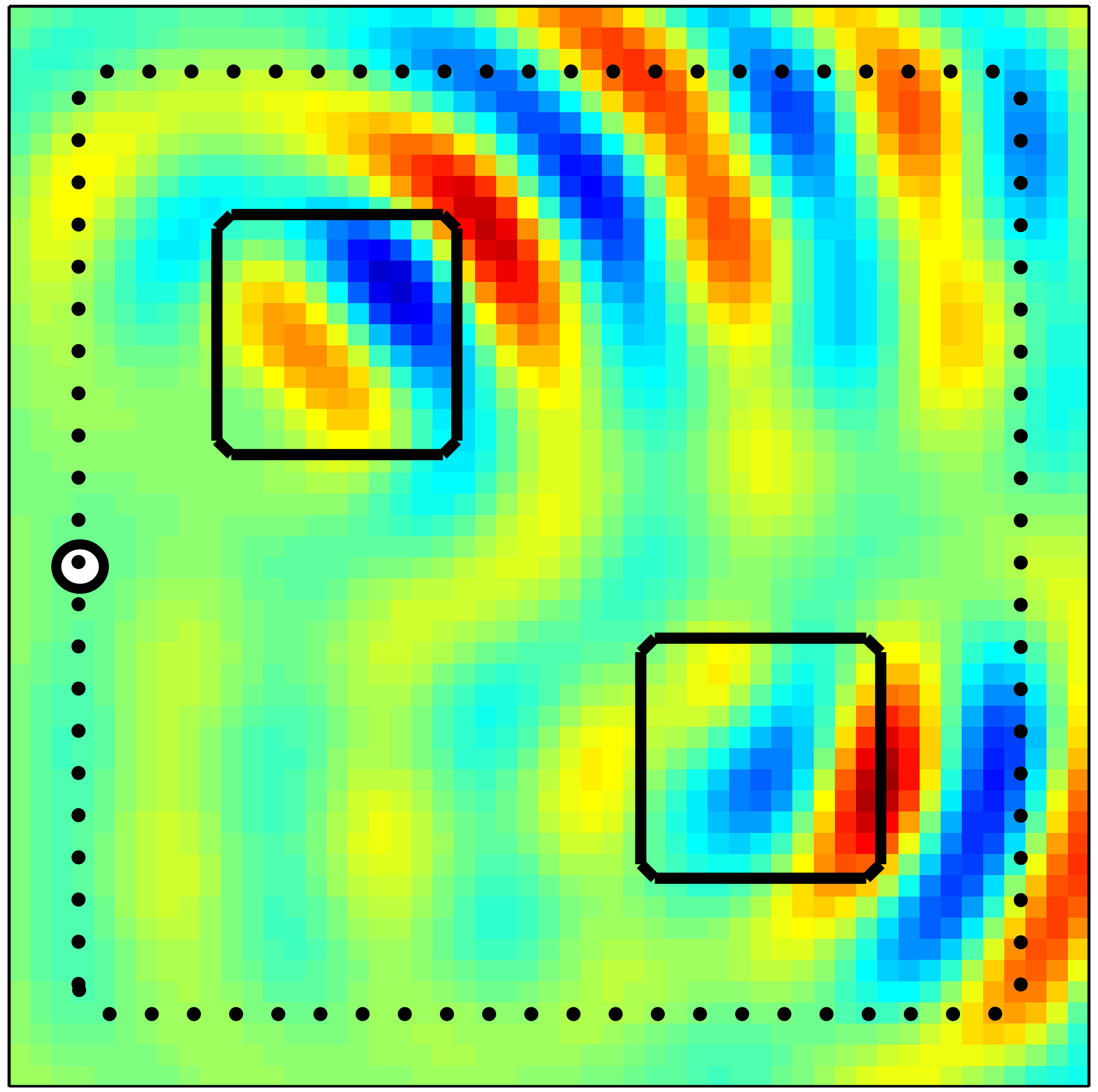
- solve a wave-equation

$$A(\mathbf{m})\mathbf{u}_i \approx \mathbf{q}_i$$

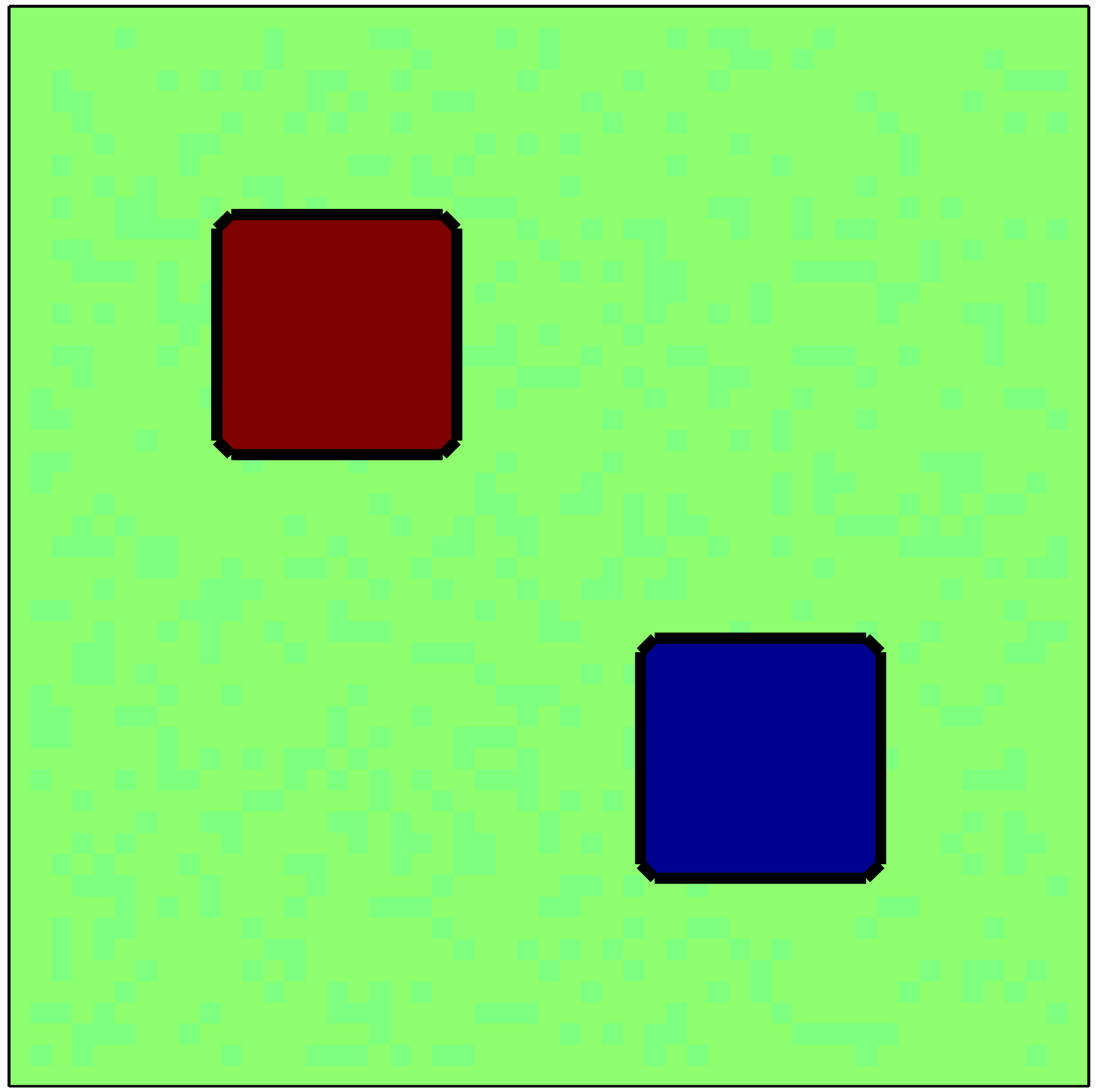
Try a least-squares solution!

$$\begin{pmatrix} P_i \\ A(\mathbf{m}) \end{pmatrix} \mathbf{u}_i \approx \begin{pmatrix} \mathbf{d}_i \\ \mathbf{q}_i \end{pmatrix}$$

Waveform inversion

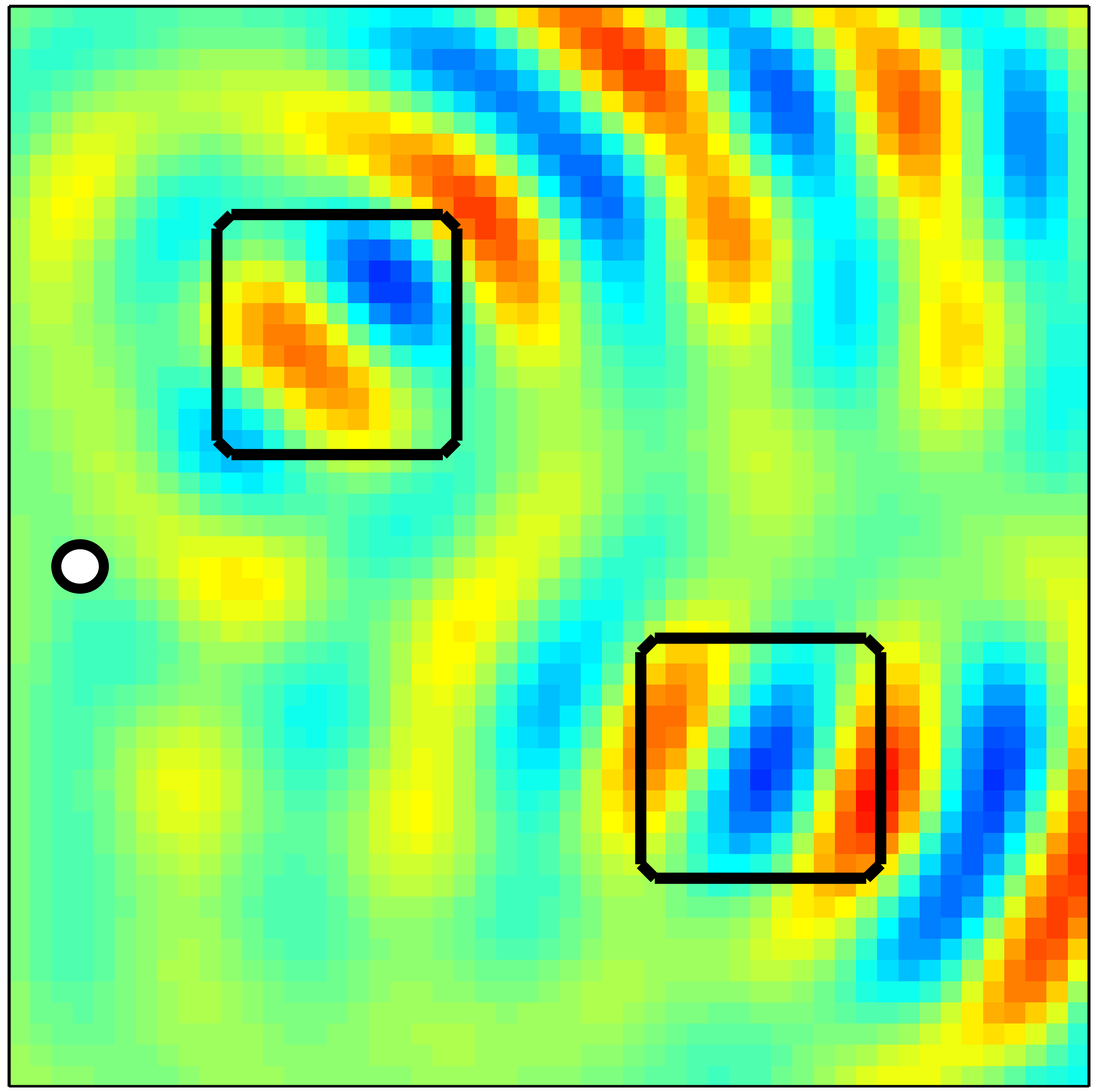


true wavefield

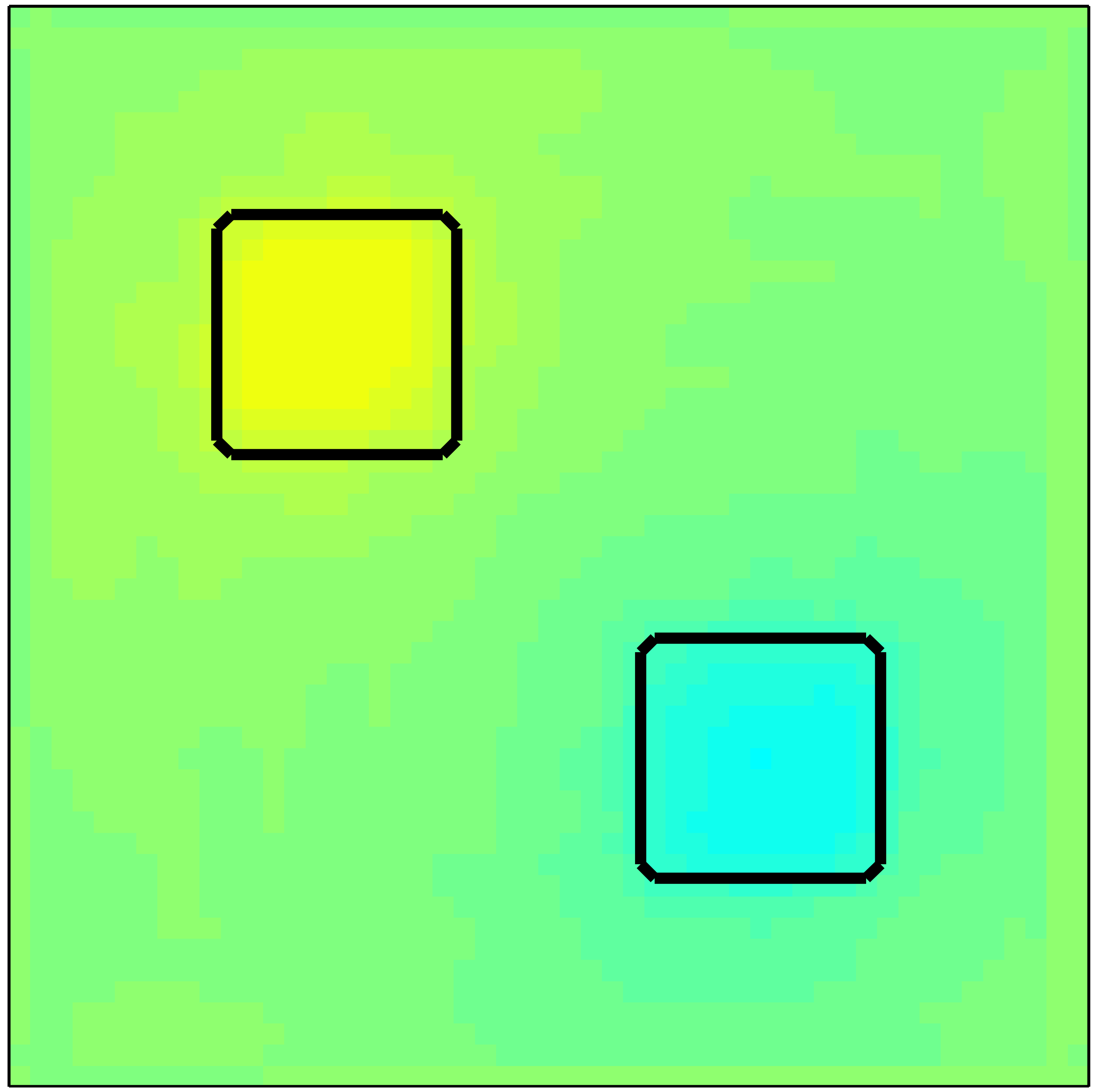


true model

First iteration

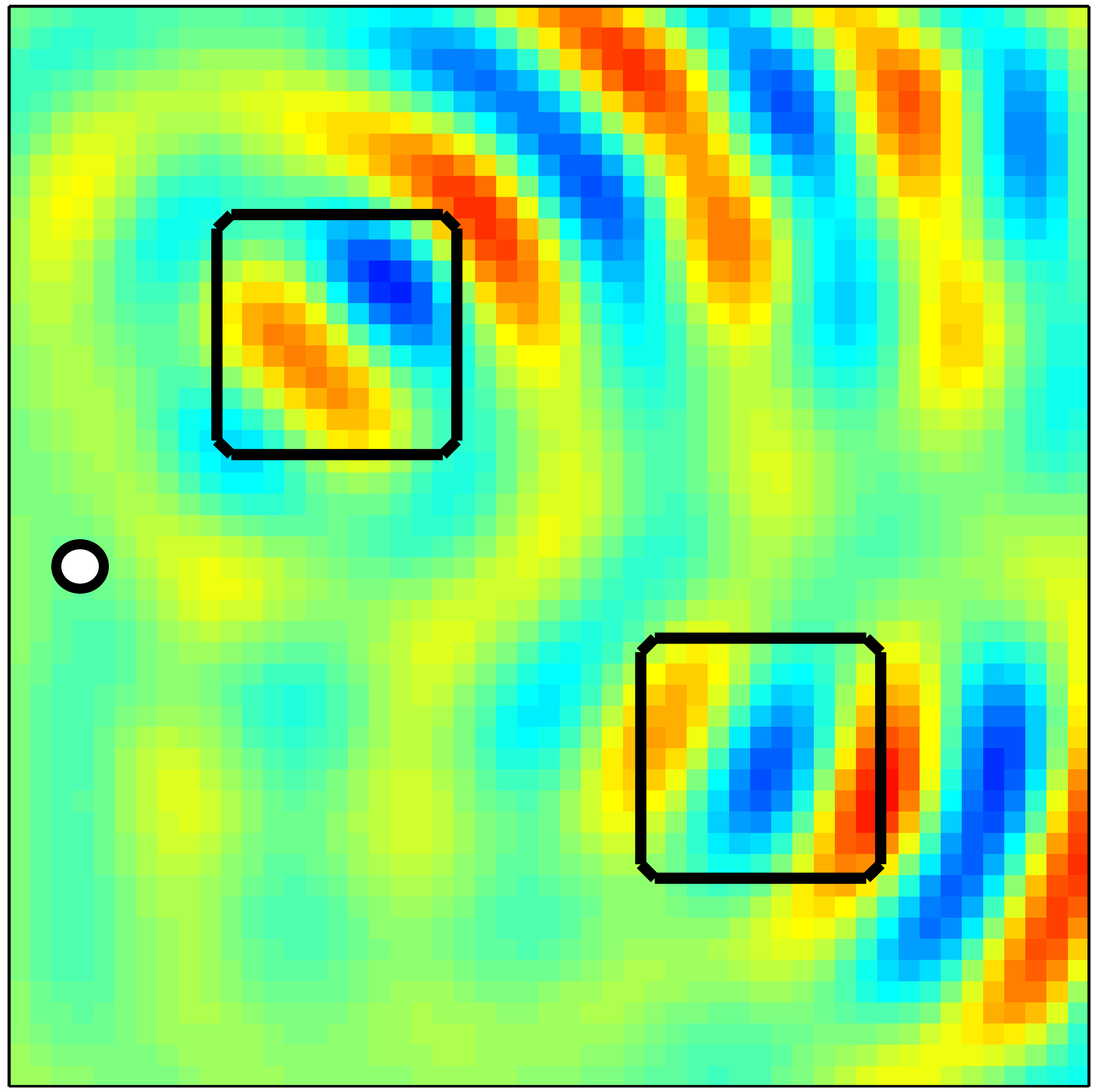


wavefield

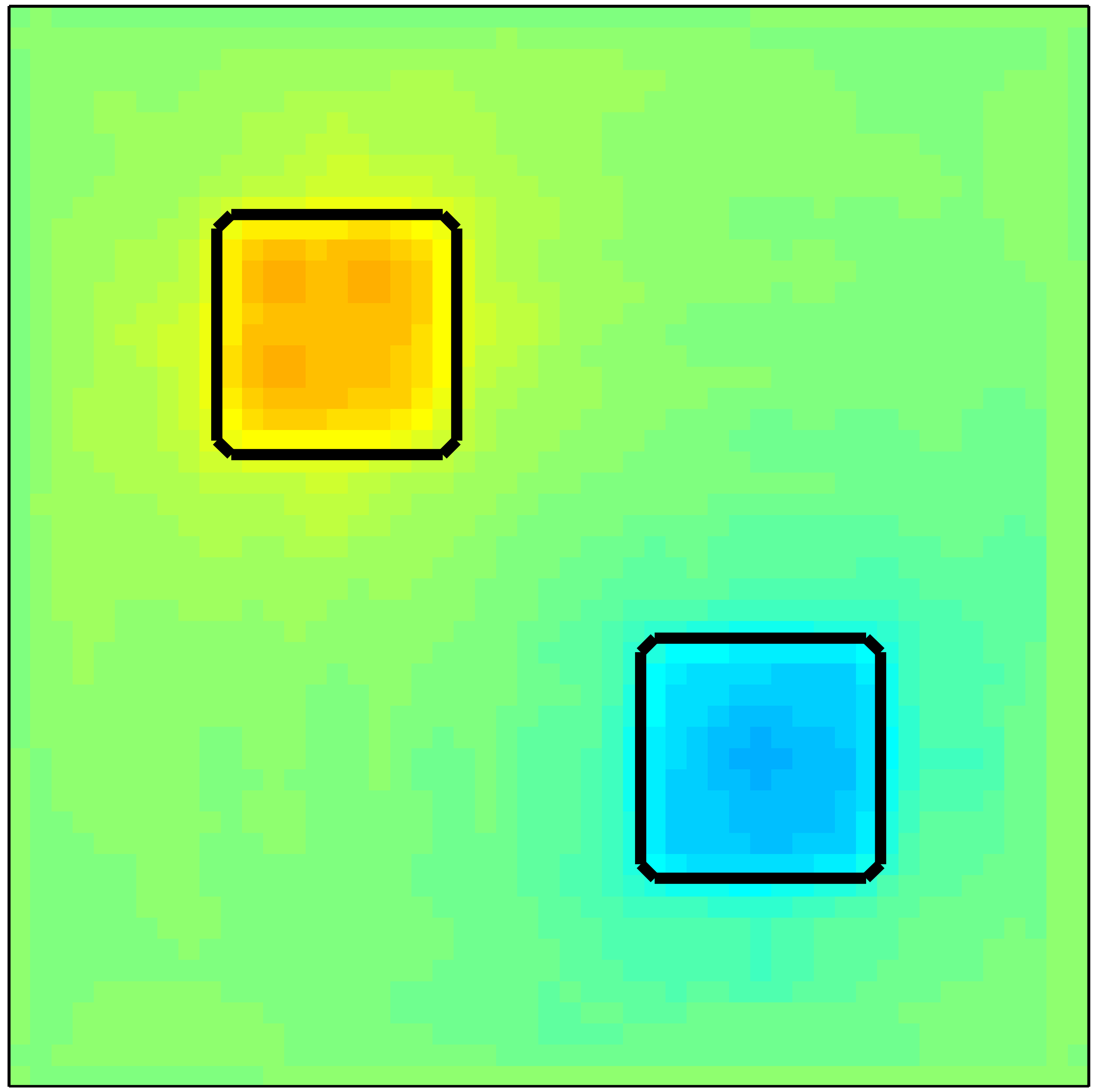


model

Second iteration

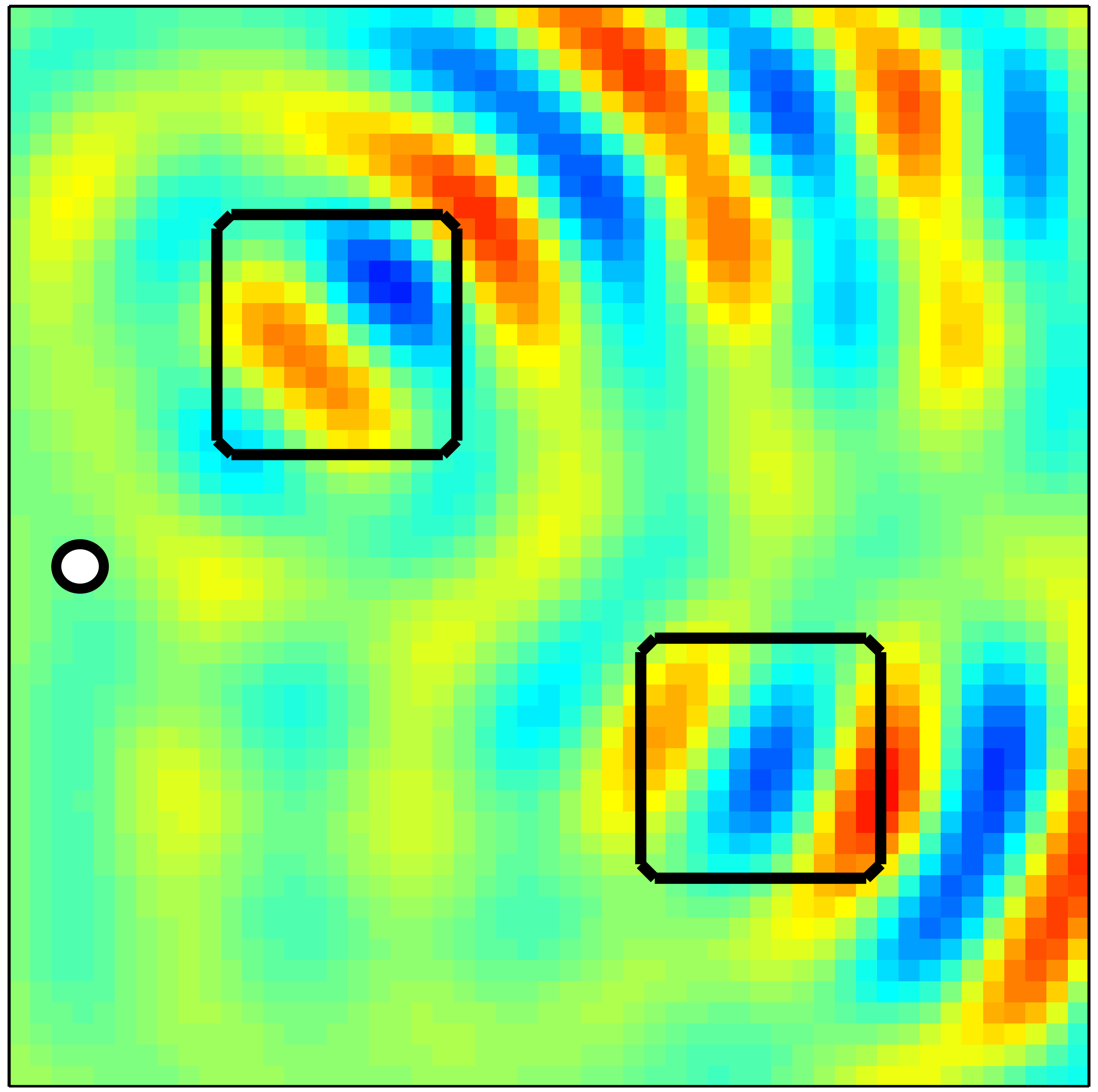


wavefield

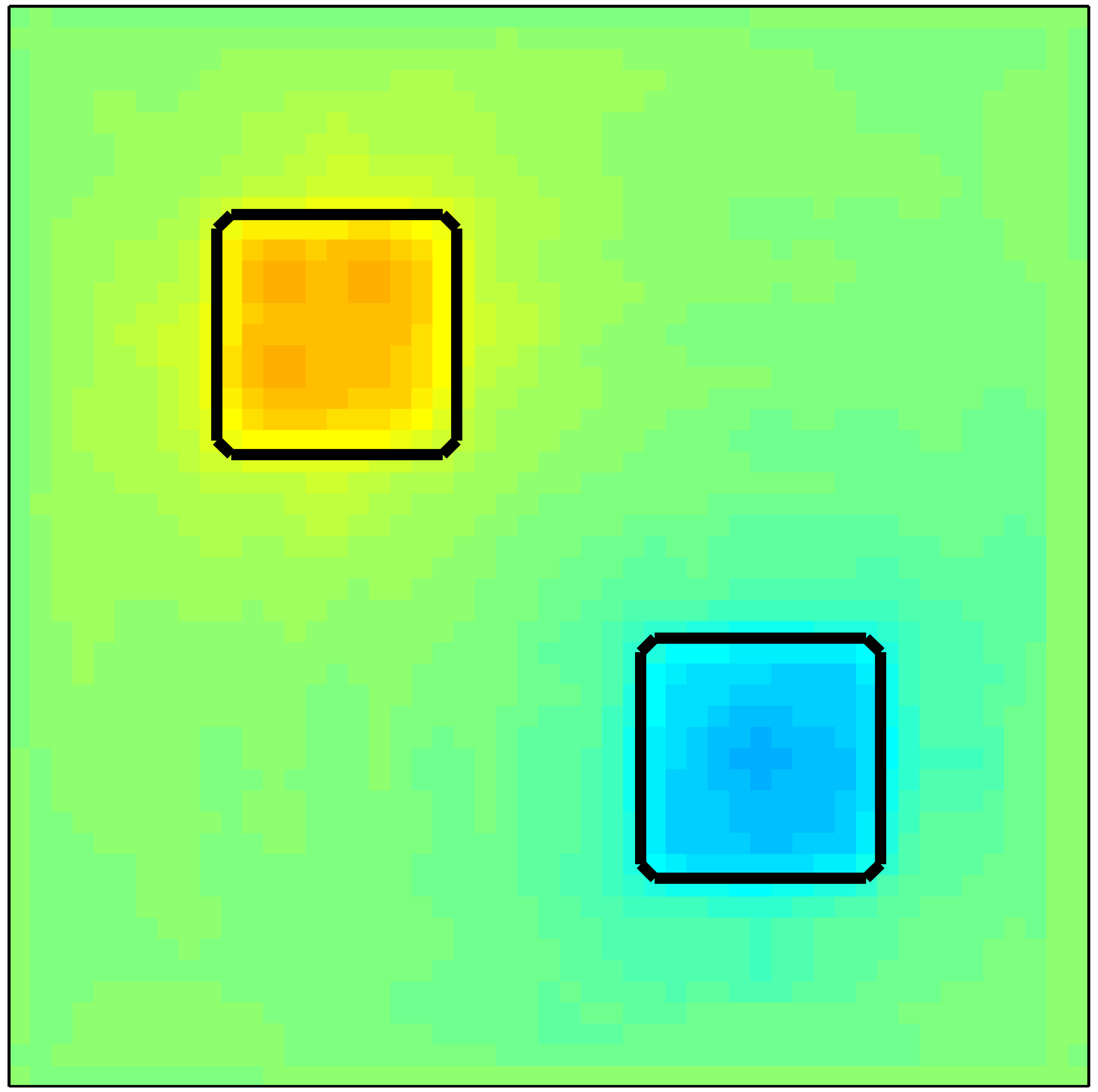


model

Third iteration

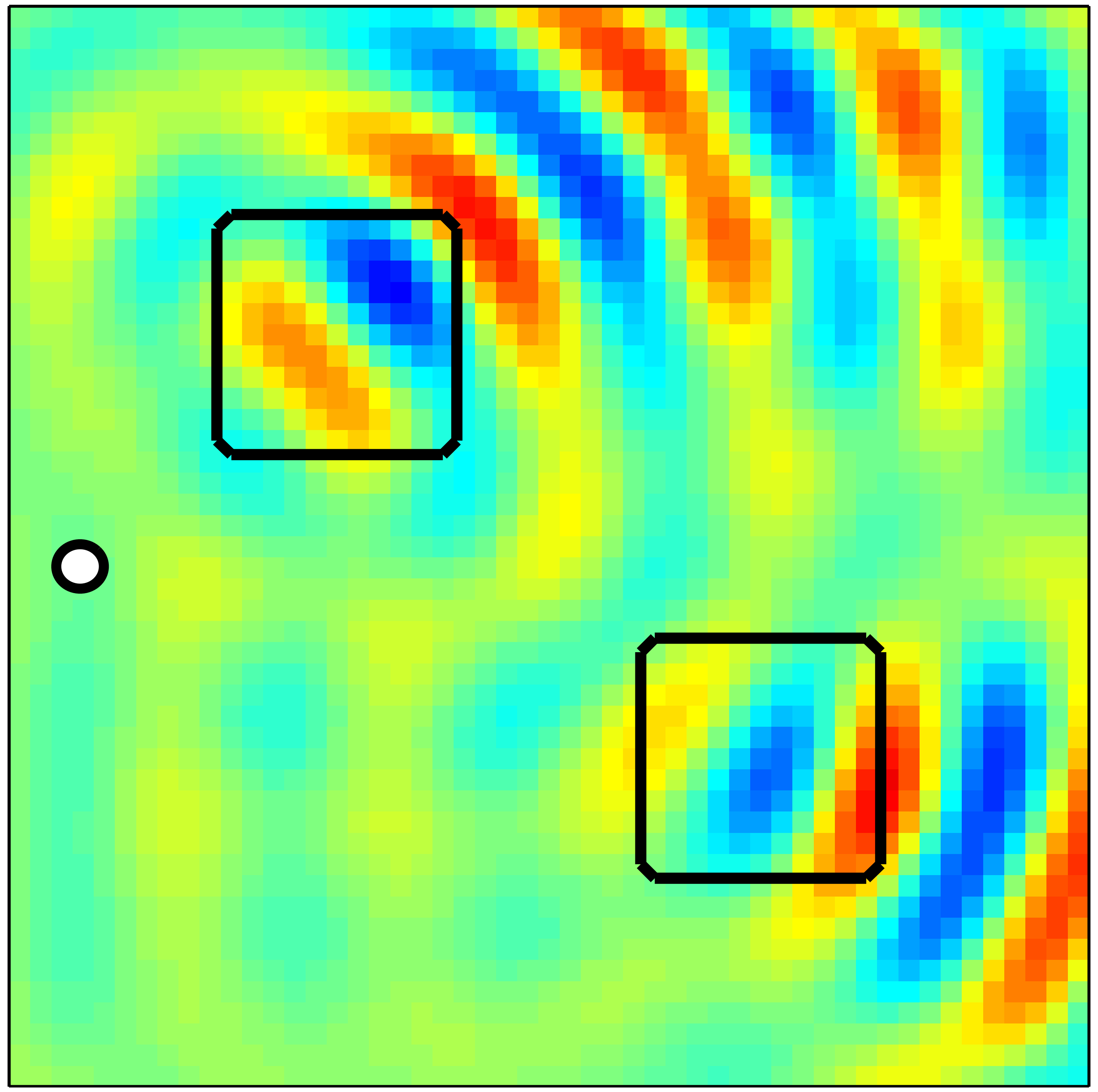


wavefield

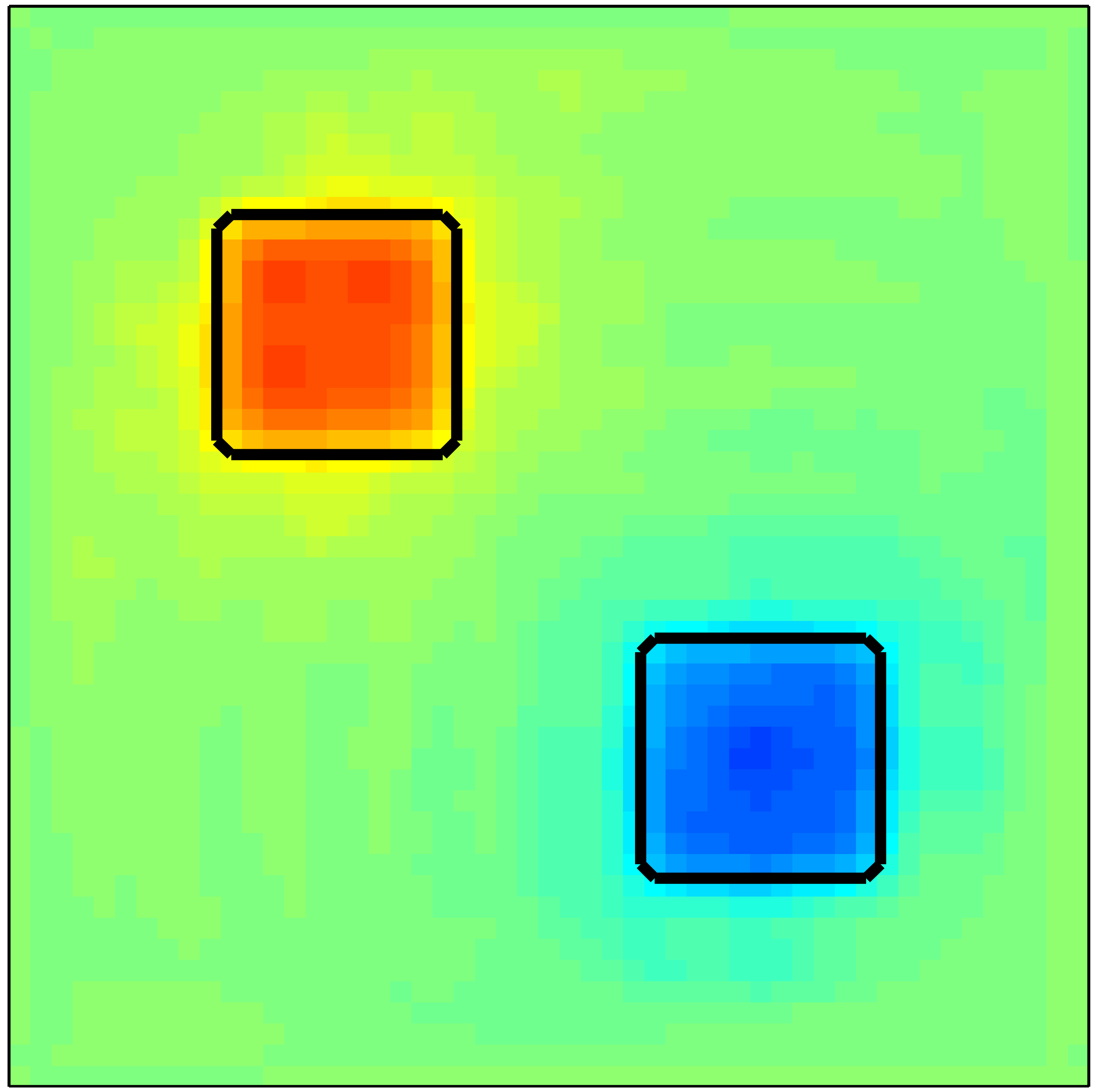


model

Fourth iteration

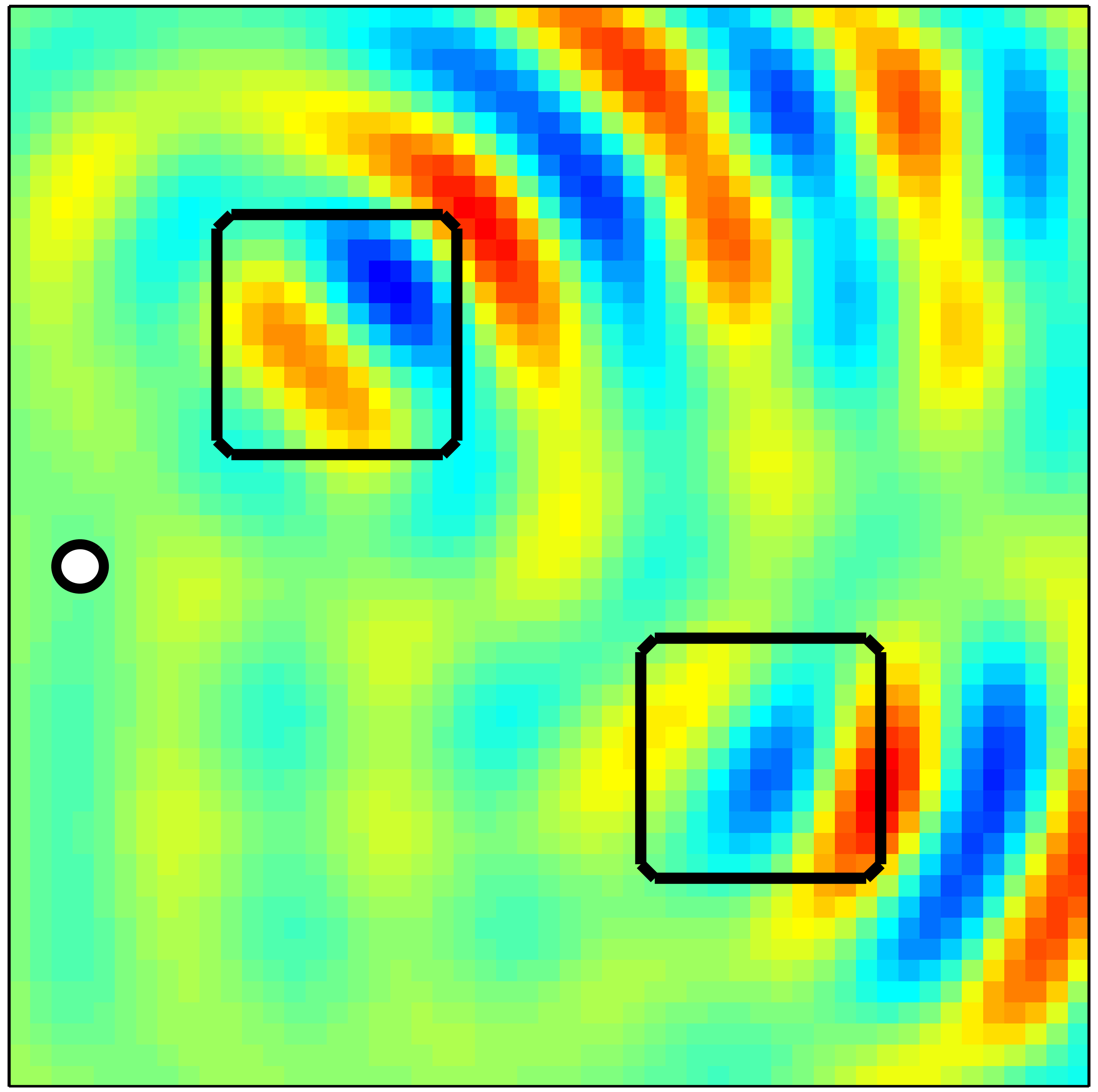


wavefield

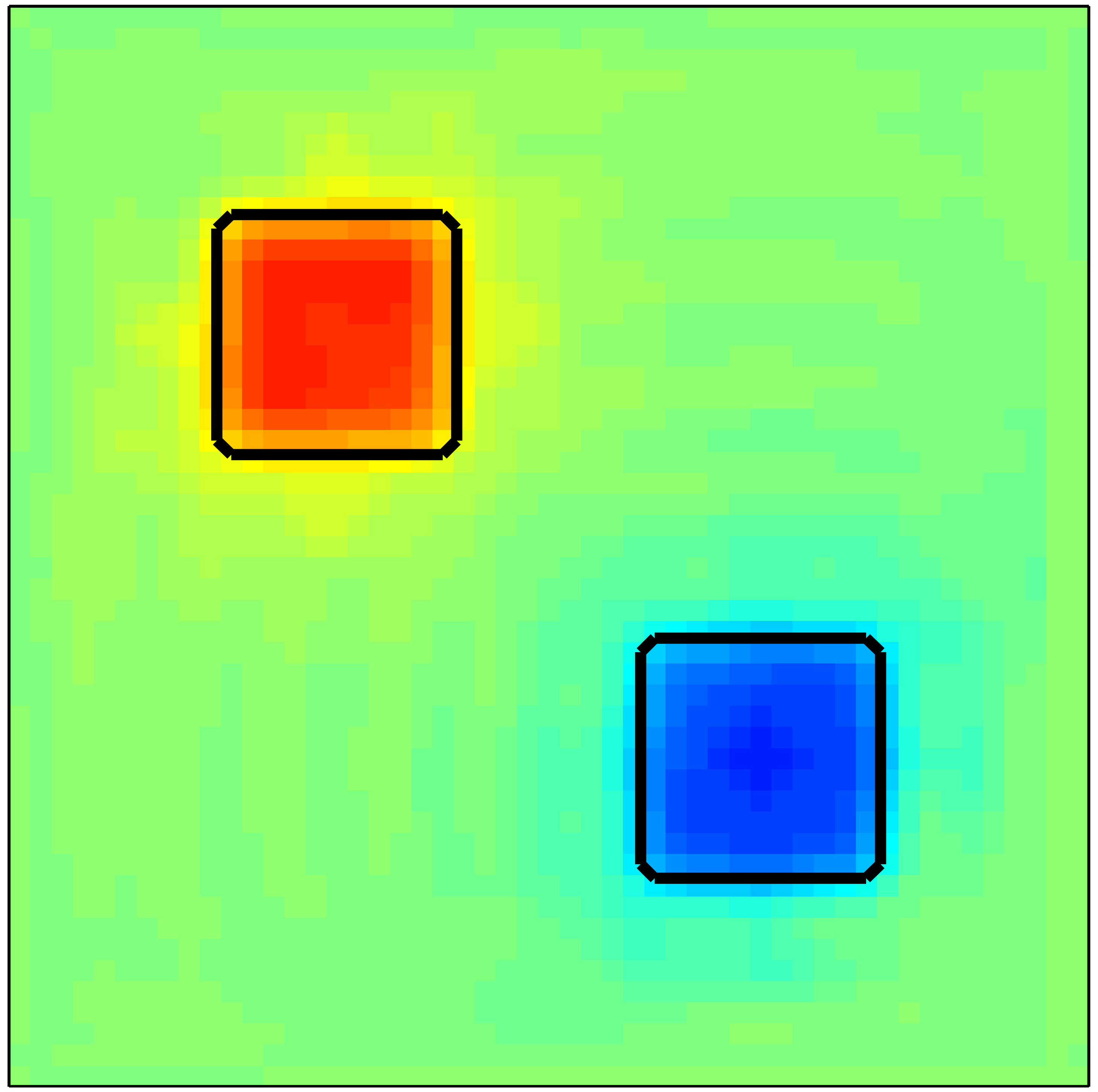


model

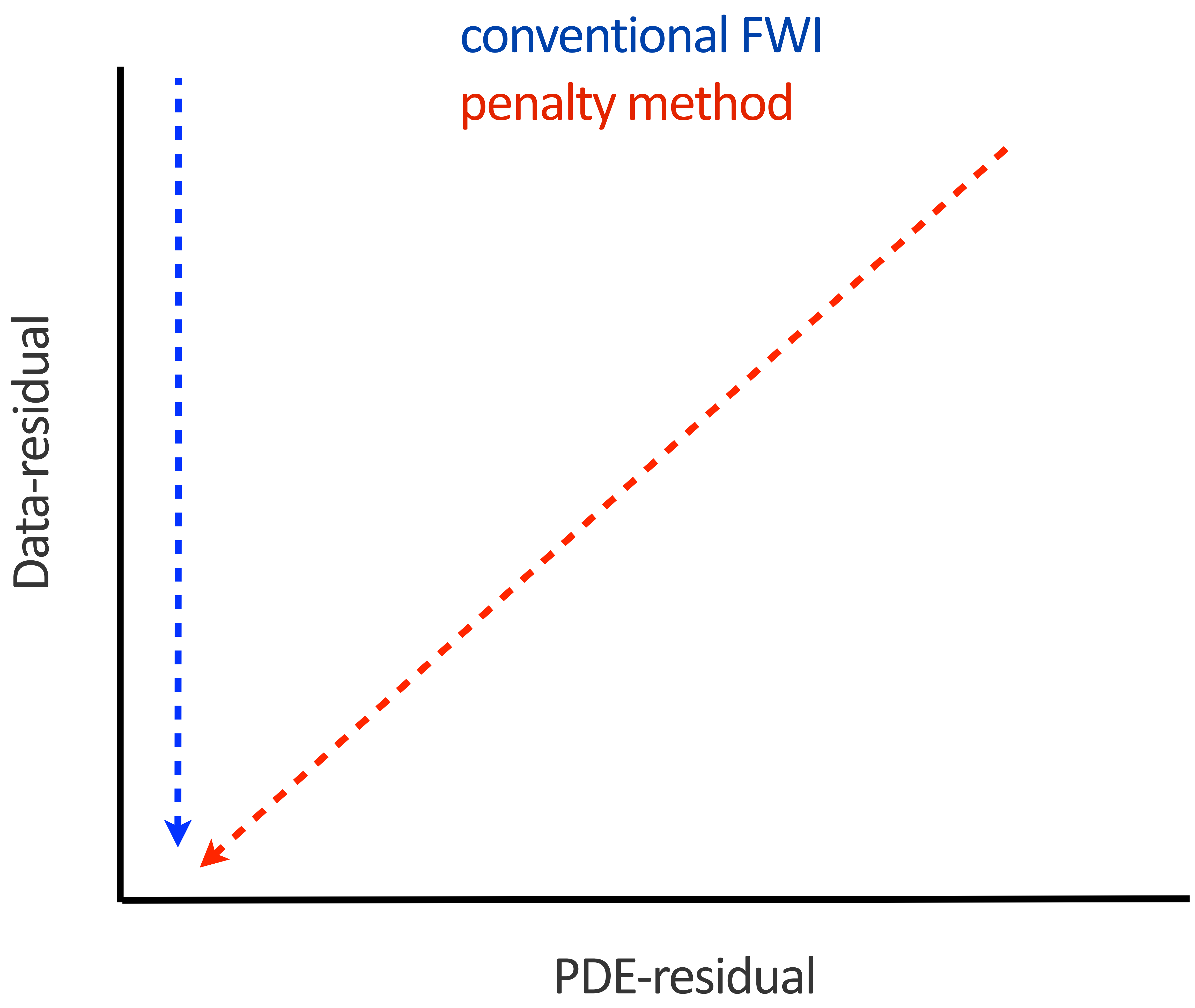
Fivth iteration



wavefield



model


conventional FWI
penalty method
Data-residual
PDE-residual

Overview

- PDE-constrained optimization
- Penalty formulation
- Physical intuition
- Numerical examples
- Conclusions & Future work

PDE-constrained optimization

$$\min_{\mathbf{m},\mathbf{u}} \sum_{i}^{M} ||P_i \mathbf{u}_i - \mathbf{d}_i||_2^2 \quad \text{s.t.} \quad A_i(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

SLIM

PDE-constrained optimization

Reformulate as unconstrained problem via the Lagrangian

$$\mathcal{L}(\mathbf{m}, \mathbf{u}, \mathbf{v}) = ||P\mathbf{u} - \mathbf{d}||_2^2 + \mathbf{v}^*(A(\mathbf{m})\mathbf{u} - \mathbf{q})$$

Leads to large-scale root-finding problem

- avoids having to solve the PDE explicitly
- sparse (GN) Hessian
- *all-at-once* approach based on KKT system requires storing *all* variables
- does not scale to industry-scale seismic problems

[Heinkenschloss, ’98 , Haber, ’00]

PDE-constrained optimization

Elimination of the constraint leads to

$$\min_{\mathbf{m}} \phi_{\mathrm{red}}(\mathbf{m}) = ||PA(\mathbf{m})^{-1}\mathbf{q} - \mathbf{d}||_2^2$$

- no need to store wavefields (block-elimination)
- suitable for black-box optimization (e.g., LBFGS)
- need to solve PDEs
- very non-linear in
- dense (GN) Hessian, involves PDE solves

[Tarantola ’84]

Penalty formulation

The traditional formulation:

- accounts for errors in the data
- considers physics as infallible

The penalty formulation:

- accounts for errors in both data *and* physics

Penalty formulation

Add constraint as penalty

$$\min_{\mathbf{m},\mathbf{u}} \phi_\lambda(\mathbf{m},\mathbf{u}) = ||P\mathbf{u} - \mathbf{d}||_2^2 + \lambda^2||A(\mathbf{m})\mathbf{u} - \mathbf{q}||_2^2$$

coincides with original problem when $\lambda \uparrow \infty$

[Wright ’00]

SLIM

Penalty formulation

Eliminate the wavefield by *variable projection*

$$\nabla_{\mathbf{u}}\phi_\lambda(\mathbf{m},\mathbf{u}) = 0$$

Define a *reduced* penalty objective:

$$\phi_{\text{pen}}(\mathbf{m}) = \phi_\lambda(\mathbf{m},\mathbf{u}(\mathbf{m}))$$

with

$$\nabla\phi_{\text{pen}} = \nabla_{\mathbf{m}}\phi_\lambda$$
$$\nabla^2\phi_{\text{pen}} = \nabla^2_{\mathbf{m}}\phi_\lambda - \nabla^2_{\mathbf{m},\mathbf{u}}\phi_\lambda\left(\nabla^2_{\mathbf{u}}\phi_\lambda\right)^{-1}\nabla^2_{\mathbf{u},\mathbf{m}}\phi_\lambda$$

[Golub ’73, Aravkin ’12]

Penalty approach

- no need to store all the fields
- no adjoint solves
- sparse approximation of Hessian for small λ
- less non-linear in $\mathbf{m}$
- need to solve overdetermined PDE
- not clear how to pick λ
- ...

Penalty vs. Reduced

Penalty method

for each source i

$$\textit{solve} \begin{pmatrix} P \\ \lambda A(\mathbf{m}) \end{pmatrix} \mathbf{u} \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

$$\mathbf{g} = \mathbf{g} + \lambda^2\omega^2 \mathsf{diag}(\mathbf{u}_i)^*(A(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i)$$

end

Conventional method

for each source i

$$\textit{solve}\ A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

$$\textit{solve}\ A(\mathbf{m})^*\mathbf{v}_i = P^*(P\mathbf{u}_i - \mathbf{d}_i)$$

$$\mathbf{g} = \mathbf{g} + \omega^2 \mathsf{diag}(\mathbf{u}_i)^*\mathbf{v}_i$$

end

Penalty vs. reduced

	# PDE's	Storage	Gauss-Newton update
penalty	M	N	solve sparse SPSD system in N unknowns
reduced	$2M$	$2N$	solve matrix-free linear system in N unknowns, requires $3M$ PDE-solves per mat-vec
all-at-once	0	$(2M+1)N$	solve sparse symmetric, possibly indefinite system in $(2M+1)N$ unknowns

Table 1: Leading order computation and storage costs per iteration of different methods; M denotes the number of experiments and N denotes the number of gridpoints. For large-scale 3D seismic inverse problems we typically have $M = \mathcal{O}(10^4) - \mathcal{O}(10^7)$ and $N = \mathcal{O}(10^6) - \mathcal{O}(10^9)$

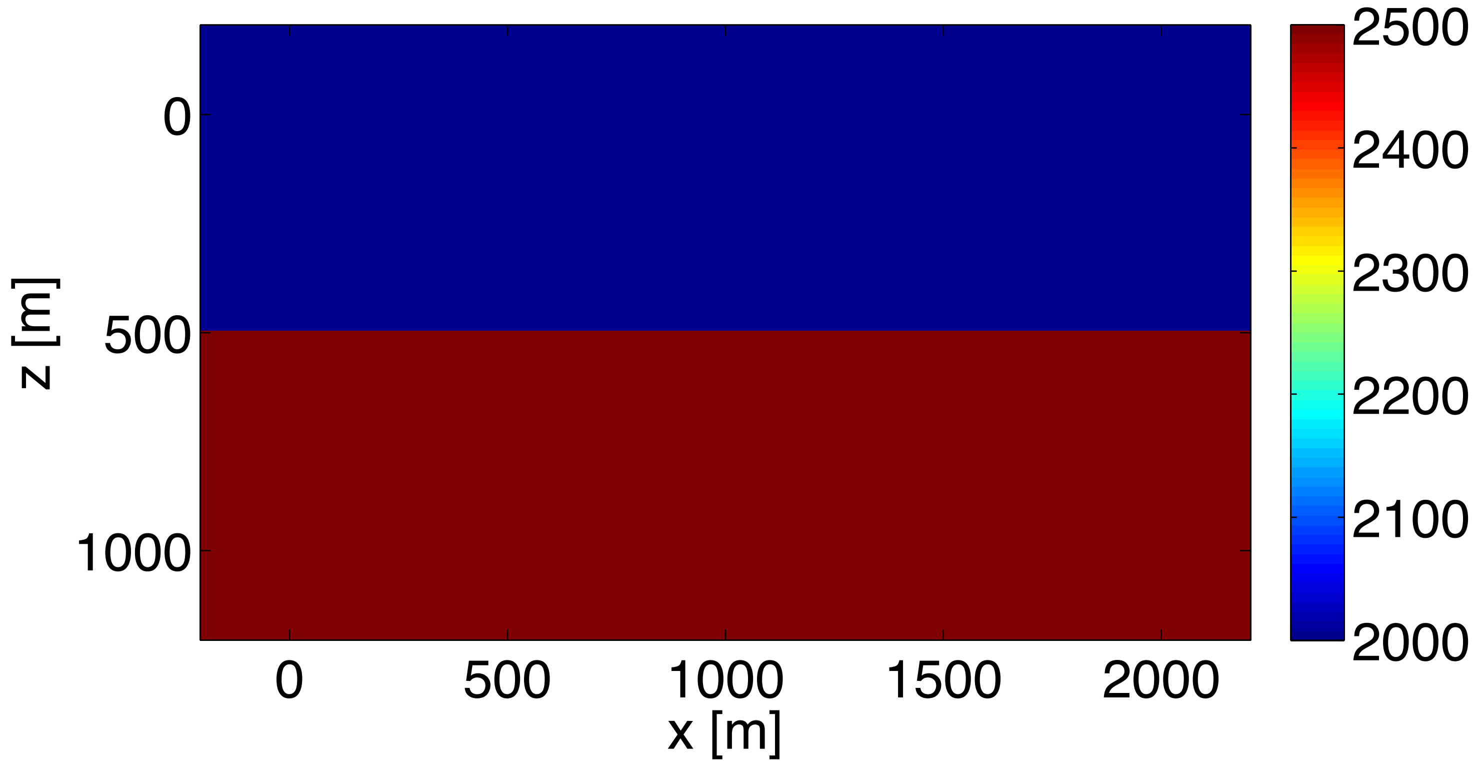
Physical intuition

Low-rank updates

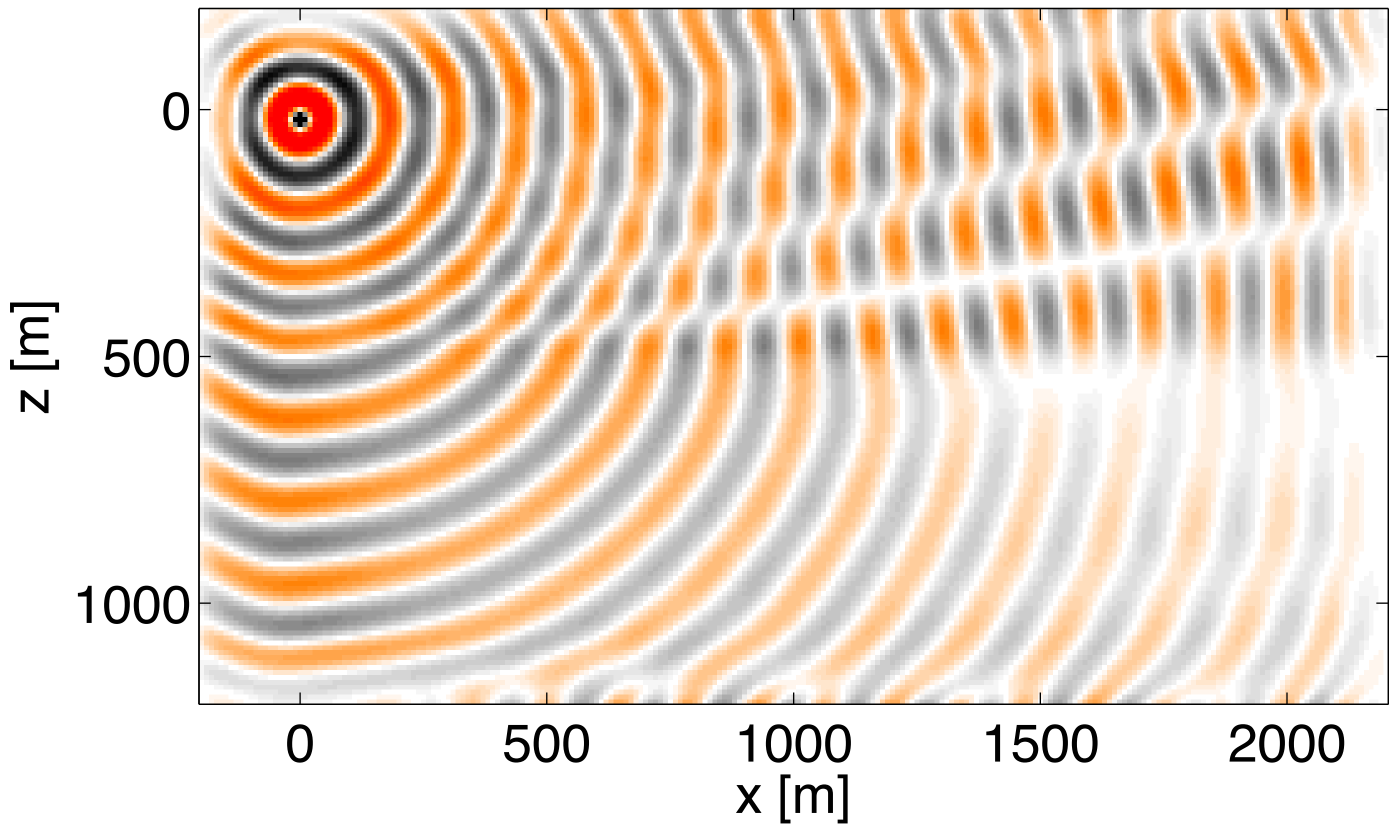
Expand the inverse as

$$\left(\mathbf{A}^*\mathbf{A} + \sum_i \mathbf{p}_i\mathbf{p}_i^*\right)^{-1} = \left(\mathbf{A}^*\mathbf{A}\right)^{-1} - \sum_i \mathbf{p}_i\mathbf{p}_i^* A^{-1} - A^{-1}\mathbf{p}_i\mathbf{p}_i^* + \left(\mathbf{p}_i^* A^{-1}\mathbf{p}_i\right)\mathbf{p}_i\mathbf{p}_i^*$$

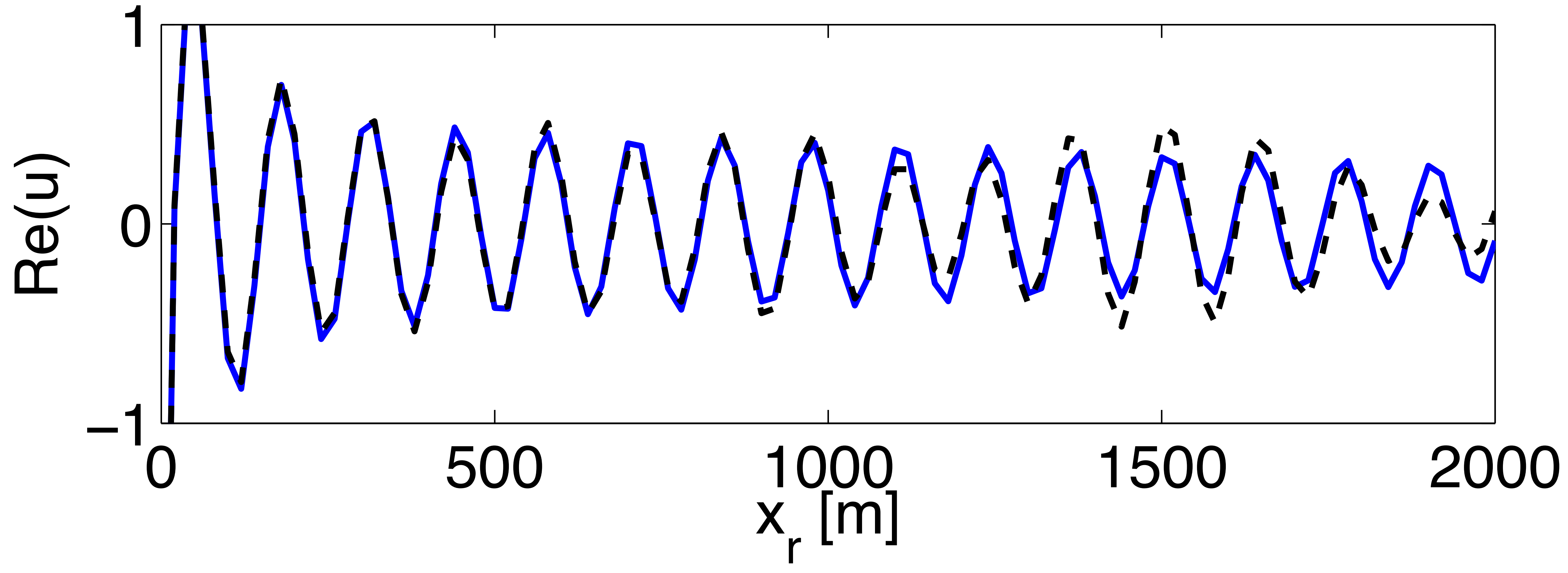
So, adding *n* data constraints yields a rank-*n* update of the wavefield

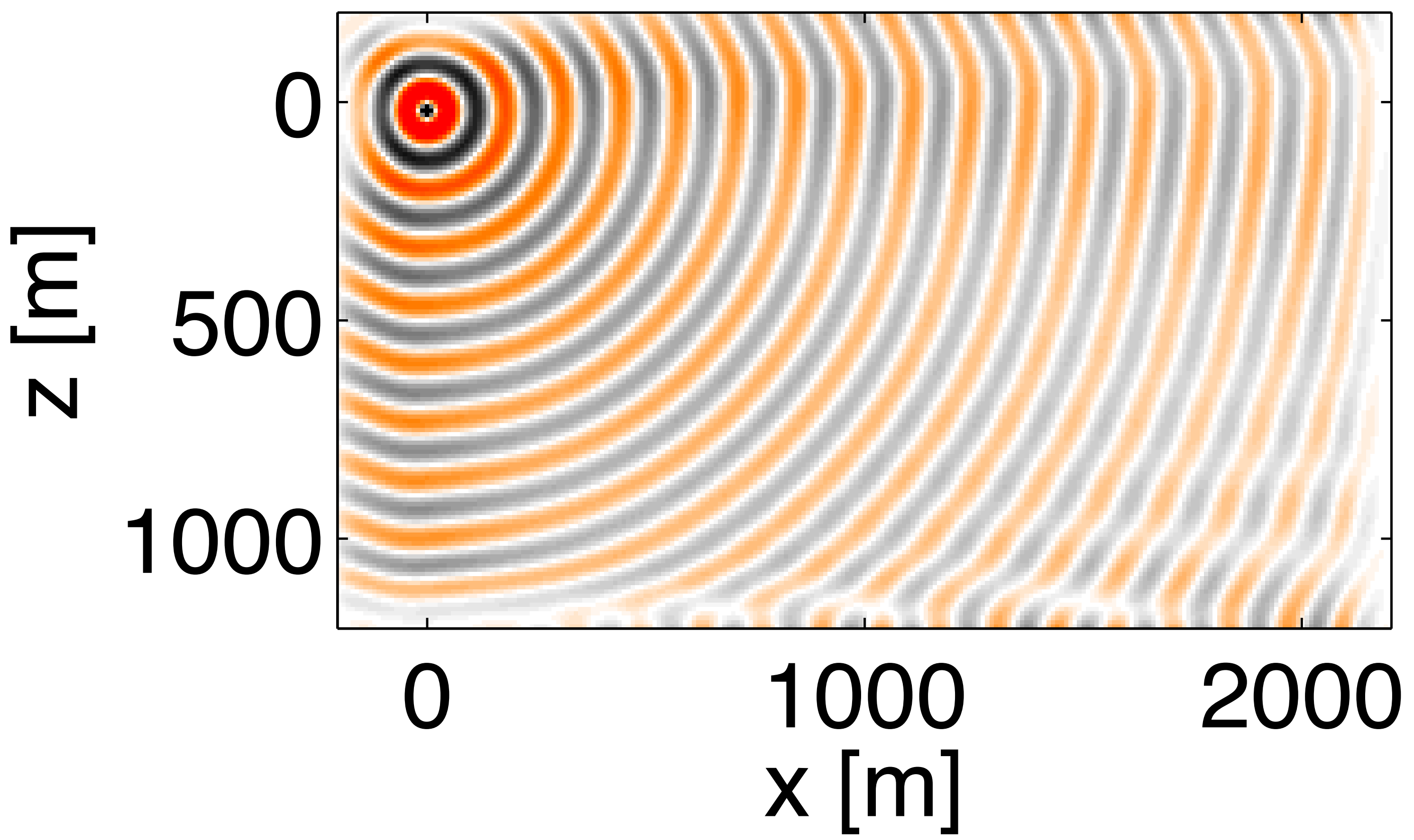


velocity model



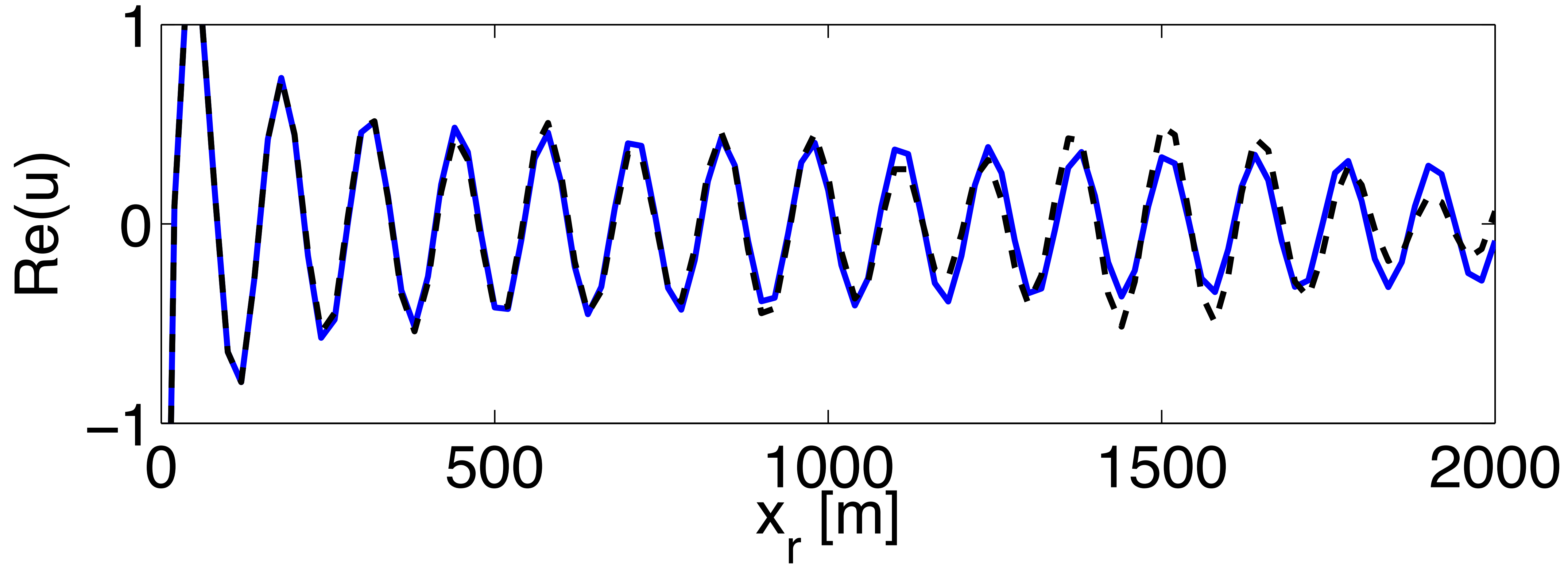
corresponding wavefield

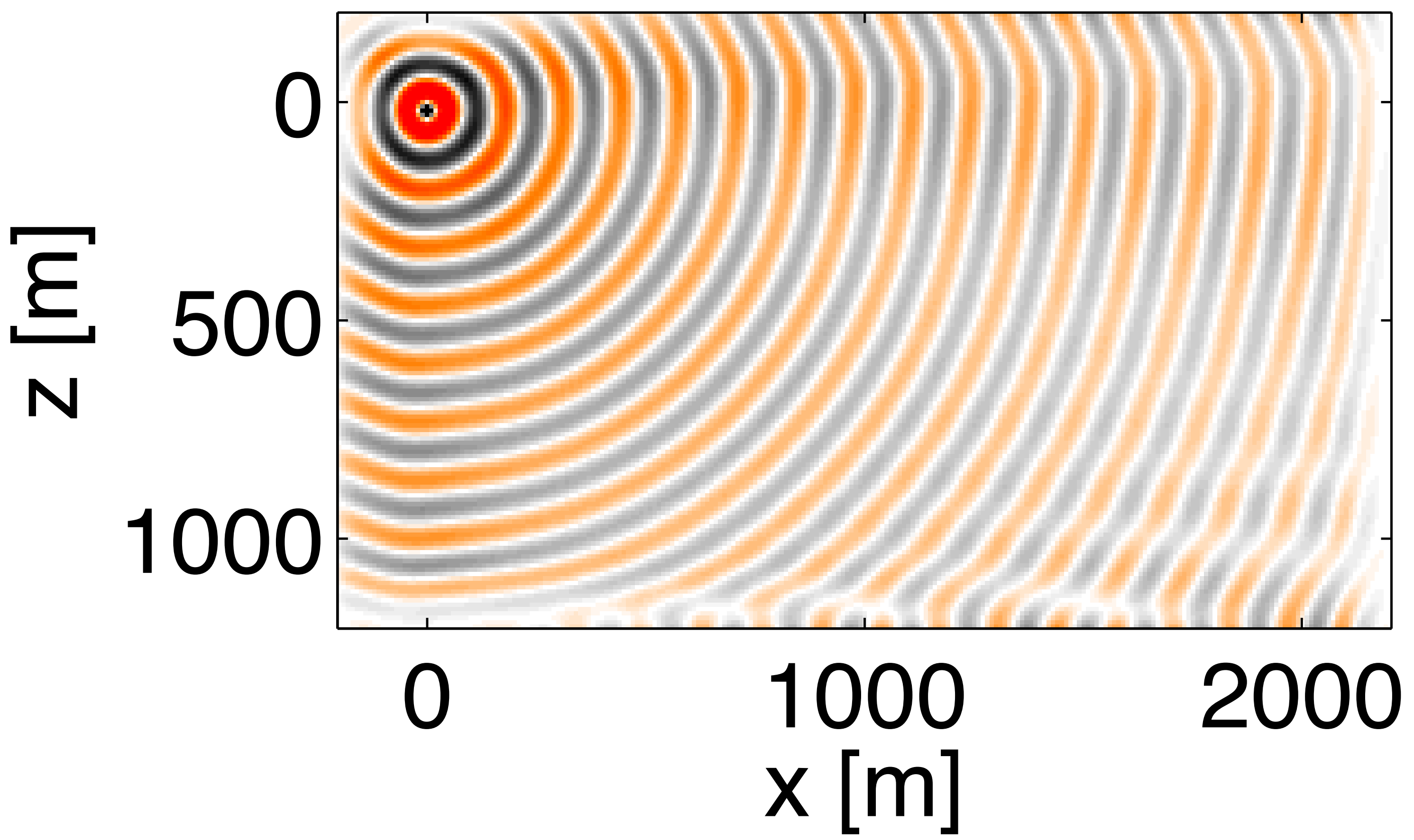




no constraints

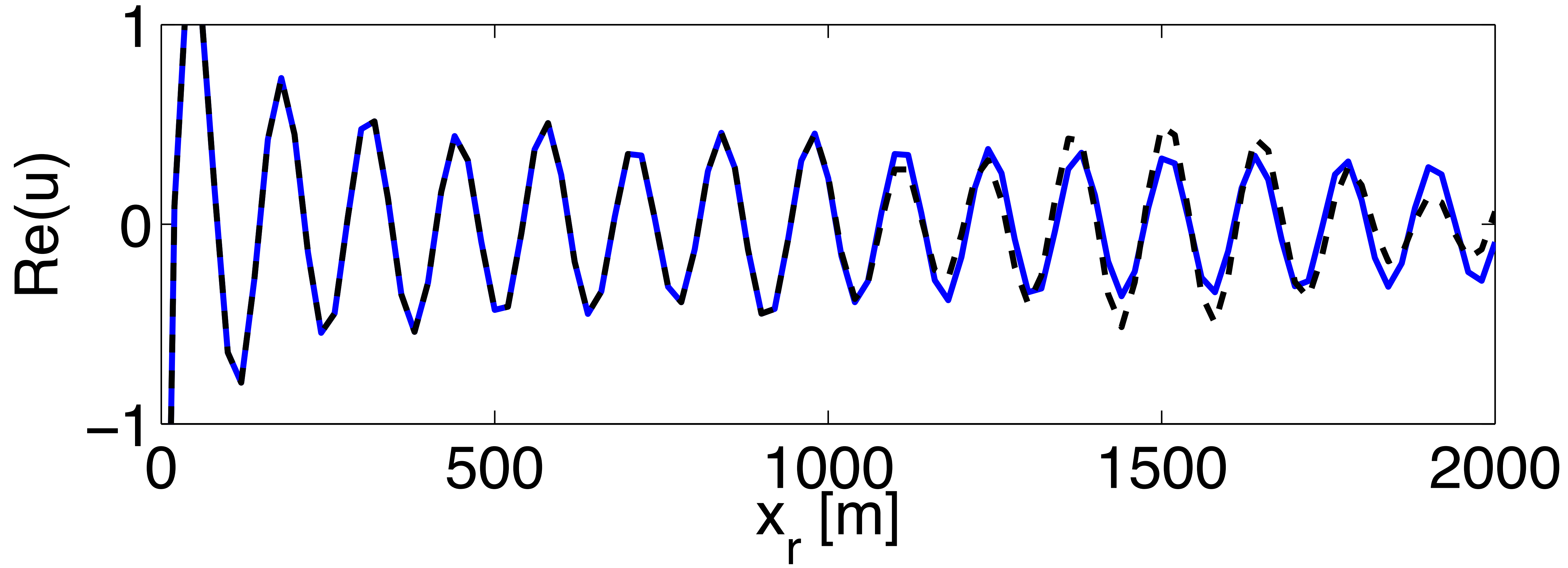
wavefield for constant background velocity

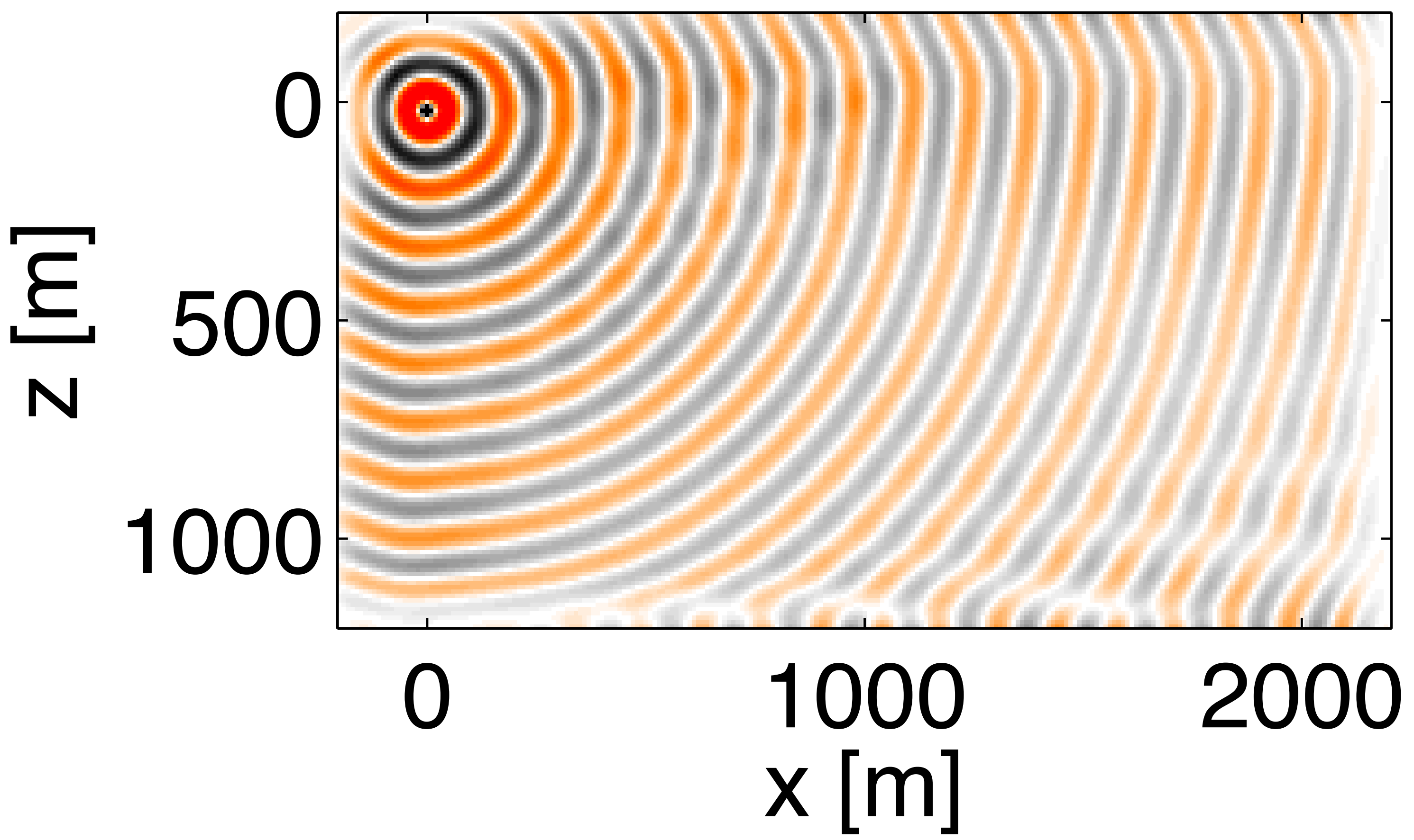




1 receiver

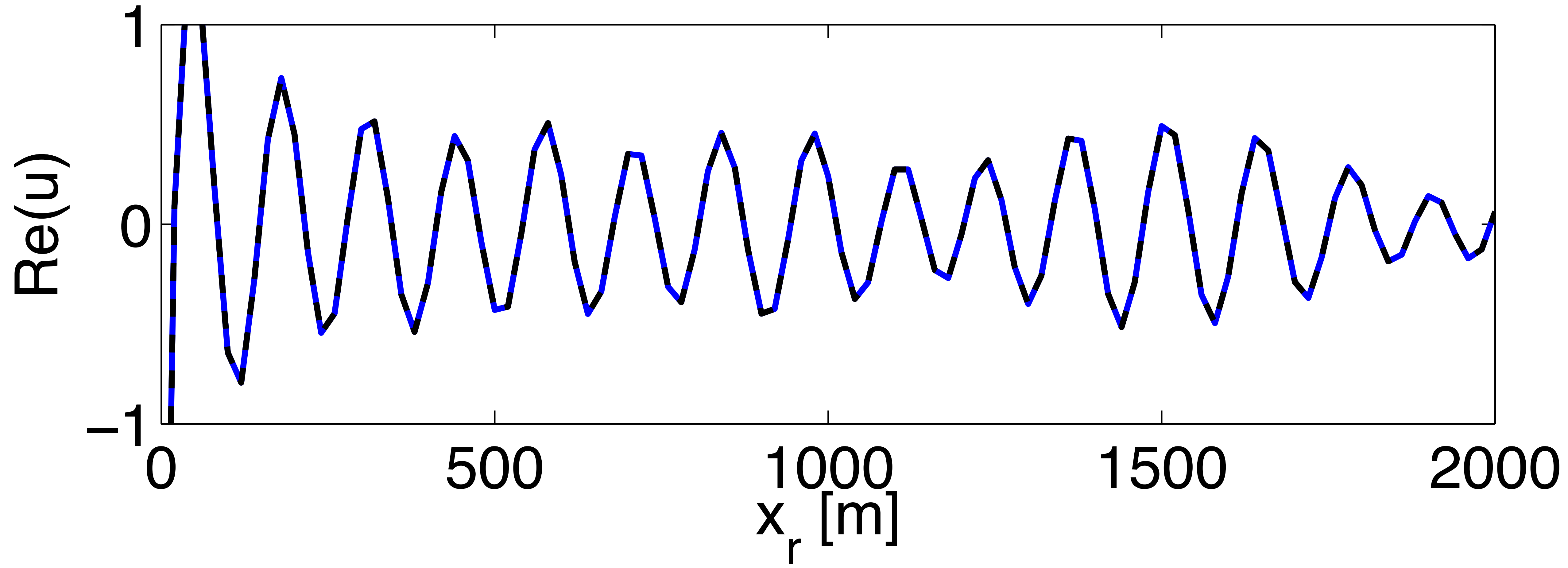
wavefield for constant background velocity

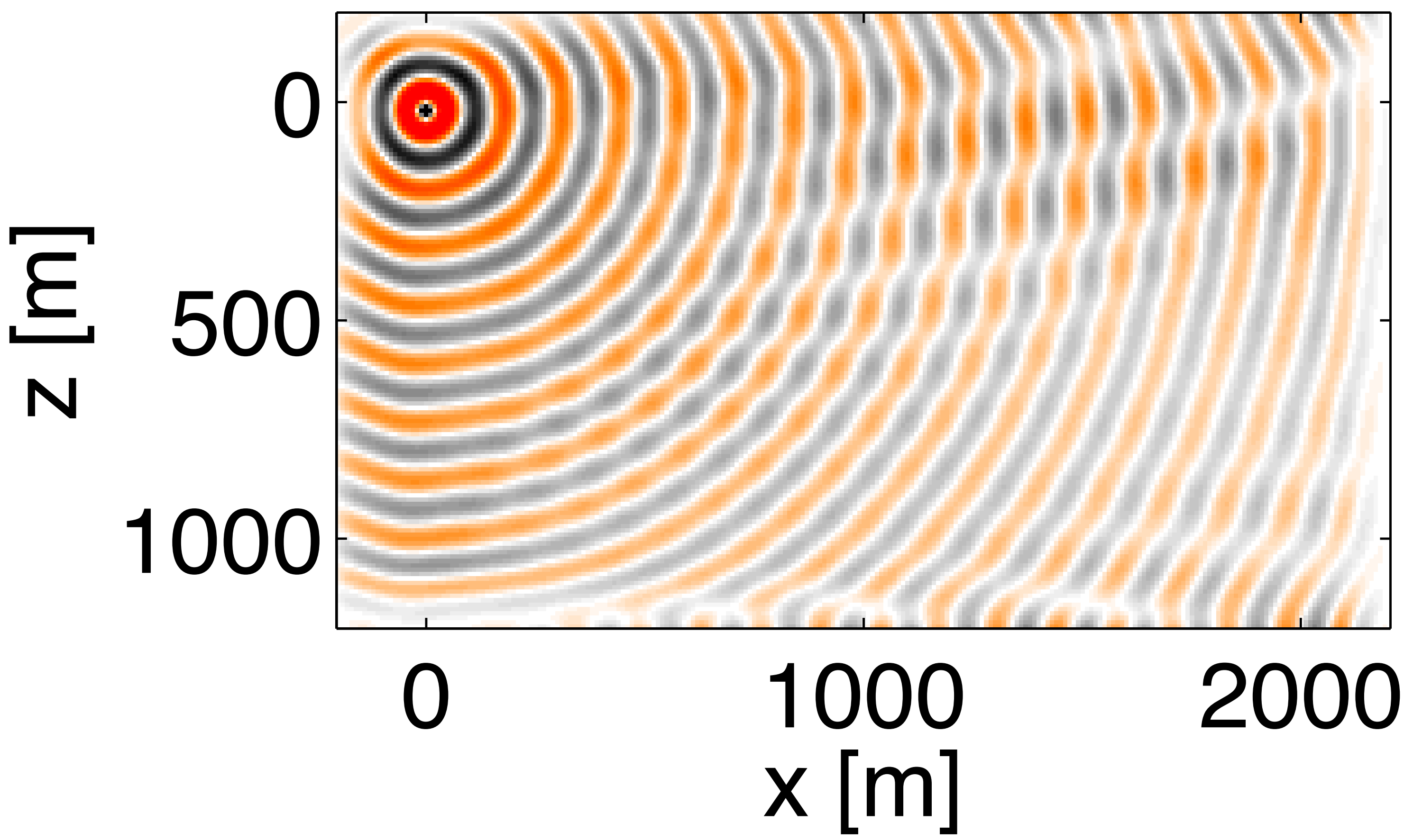




50 receivers

wavefield for constant background velocity

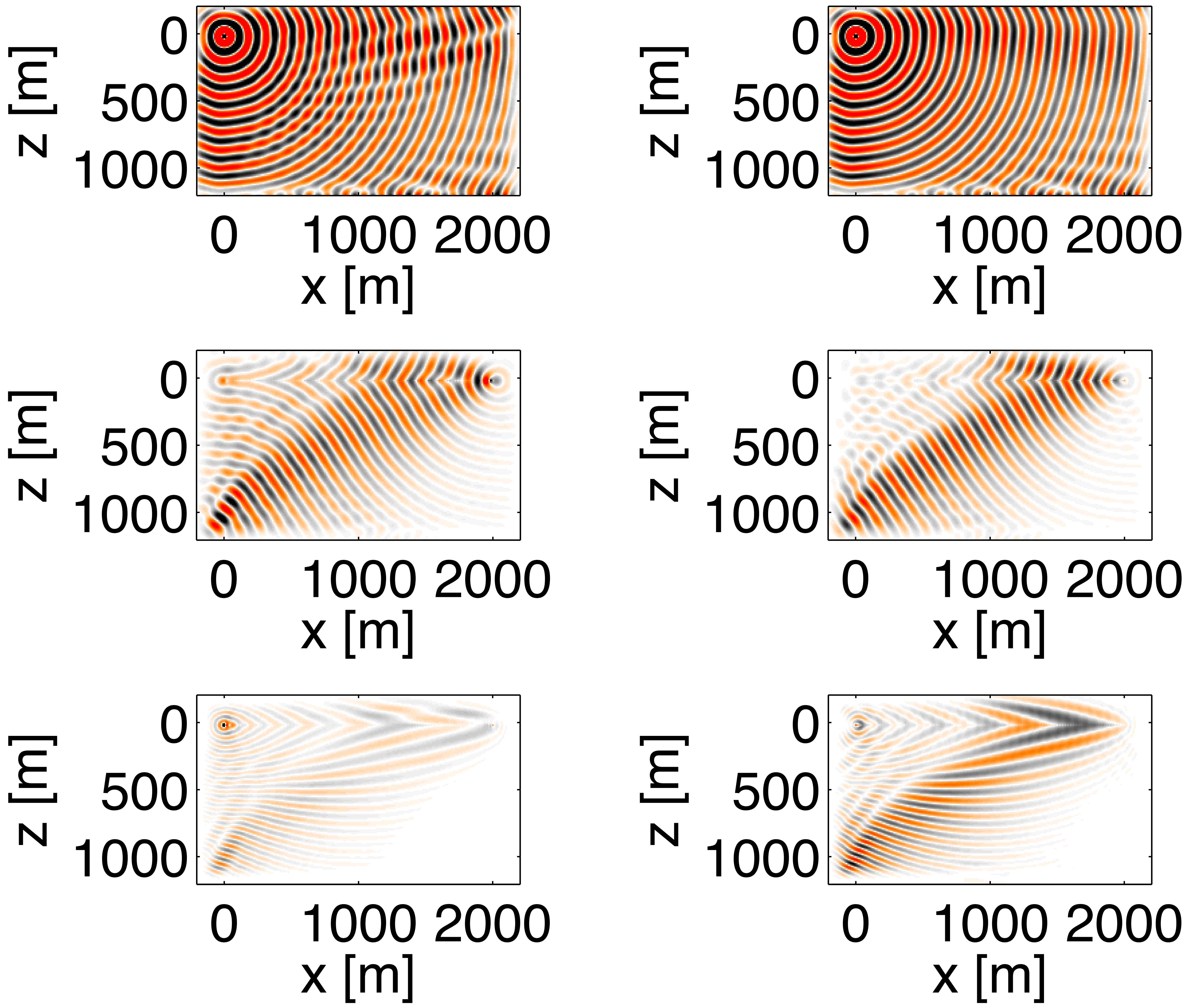




all receivers

wavefield for constant background velocity

SLIM



“source wavefield”

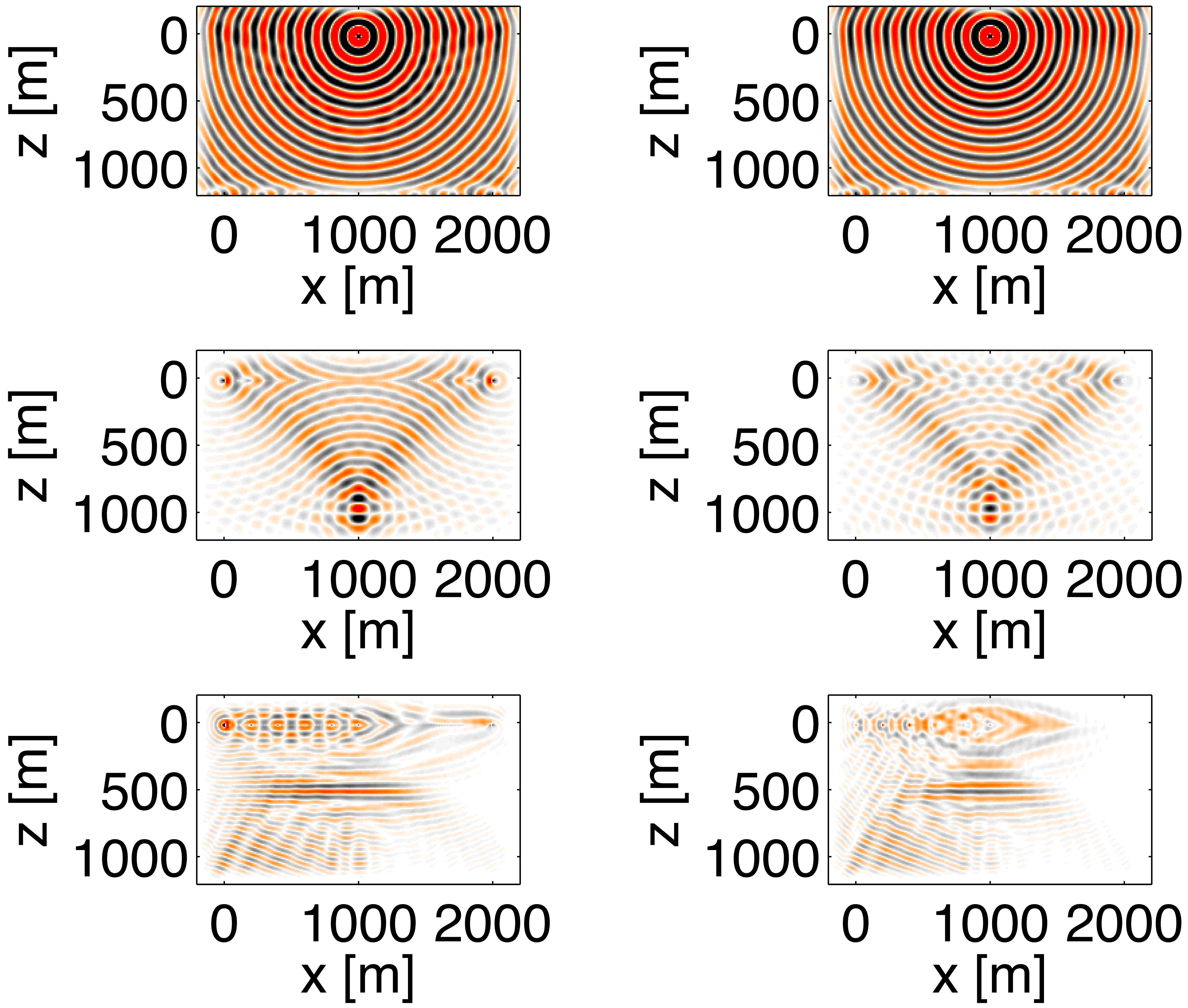
“receiver wavefield”

“correlation”

penalty

conventional

SLIM



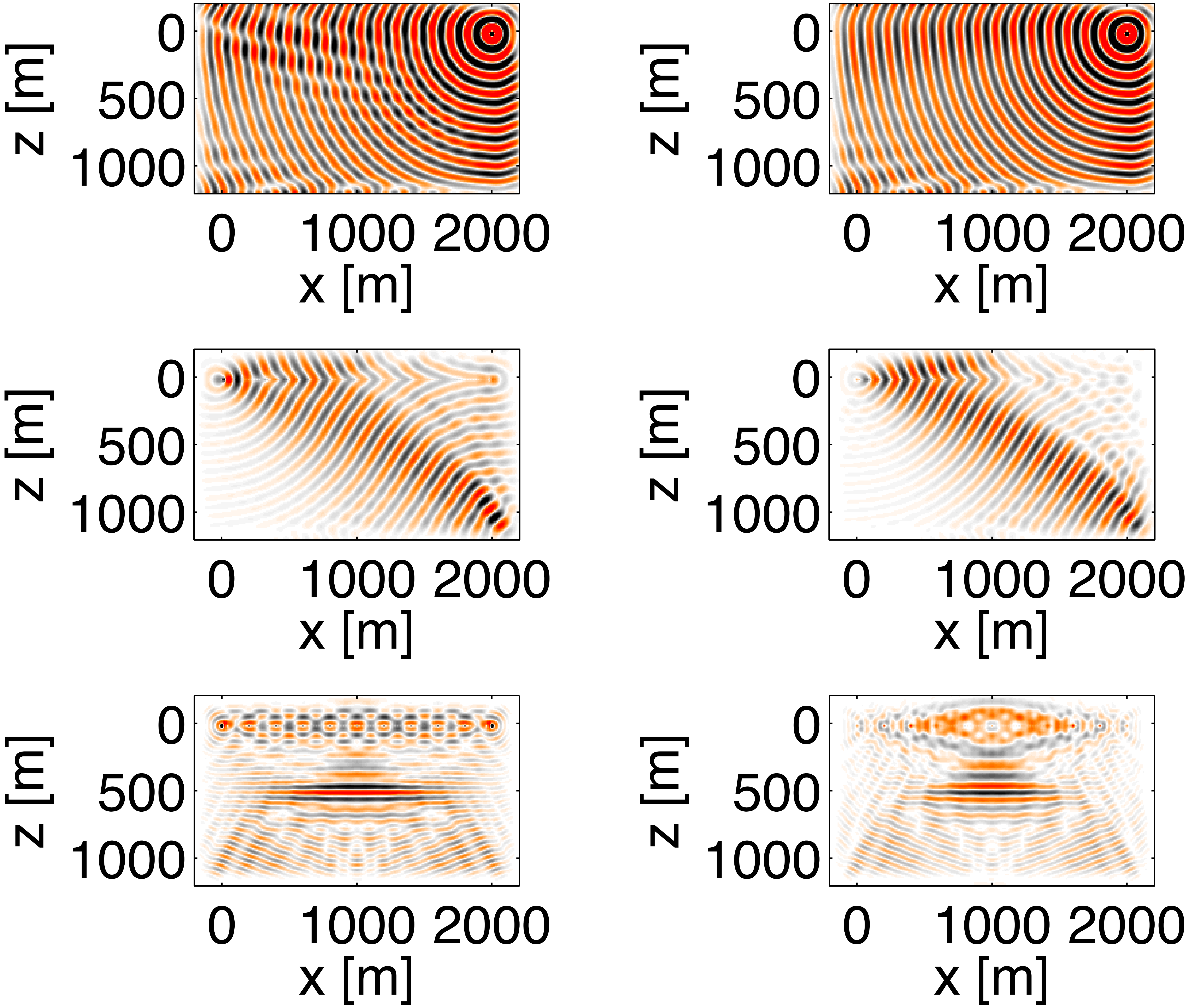
penalty

conventional

“source wavefield”

“receiver wavefield”

“correlation”



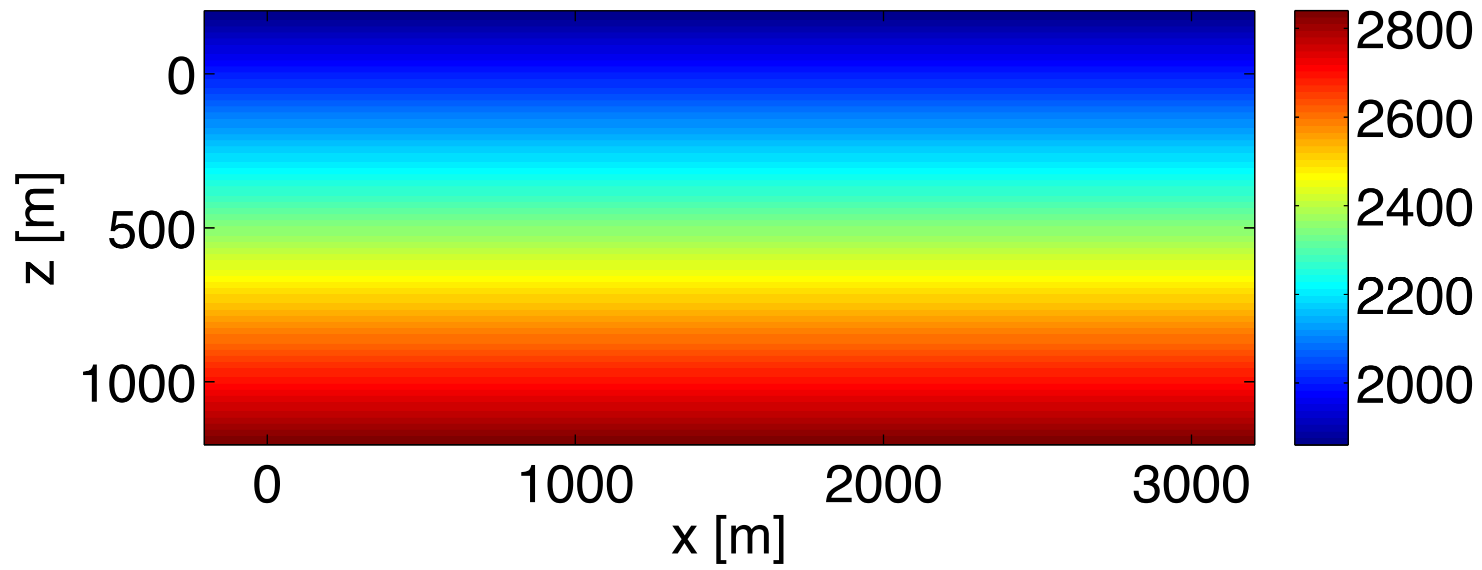
“source wavefield”

“receiver wavefield”

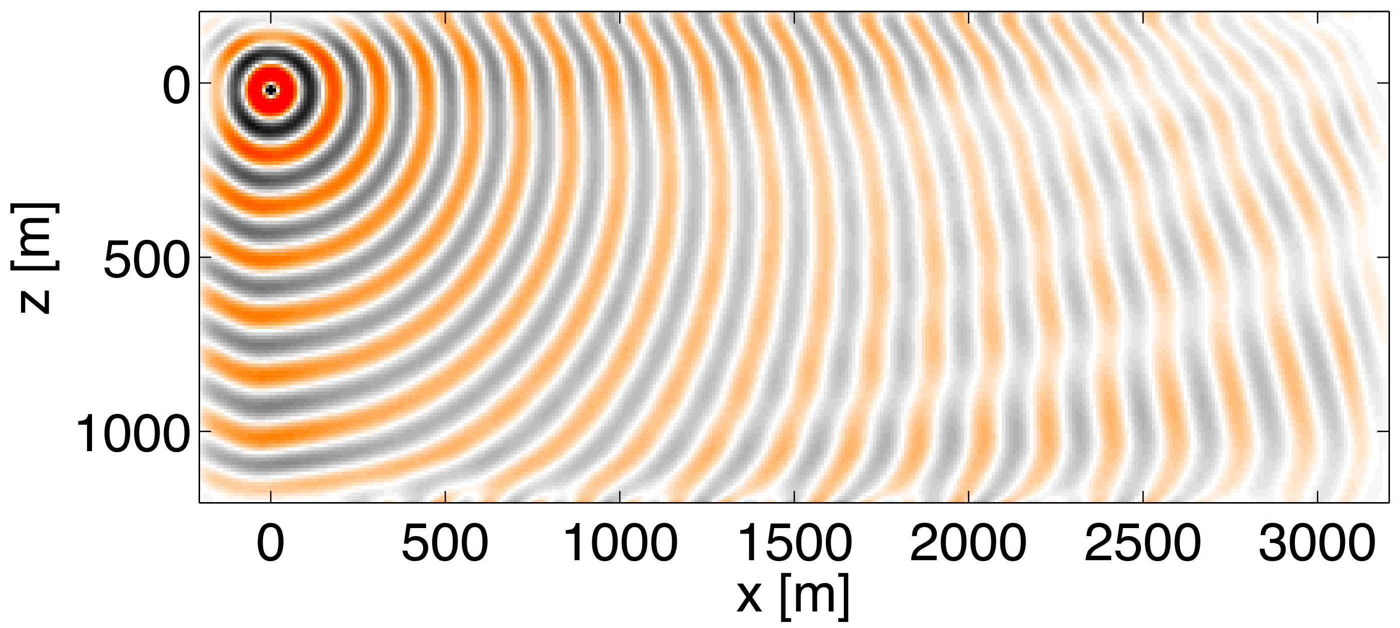
“correlation”

penalty

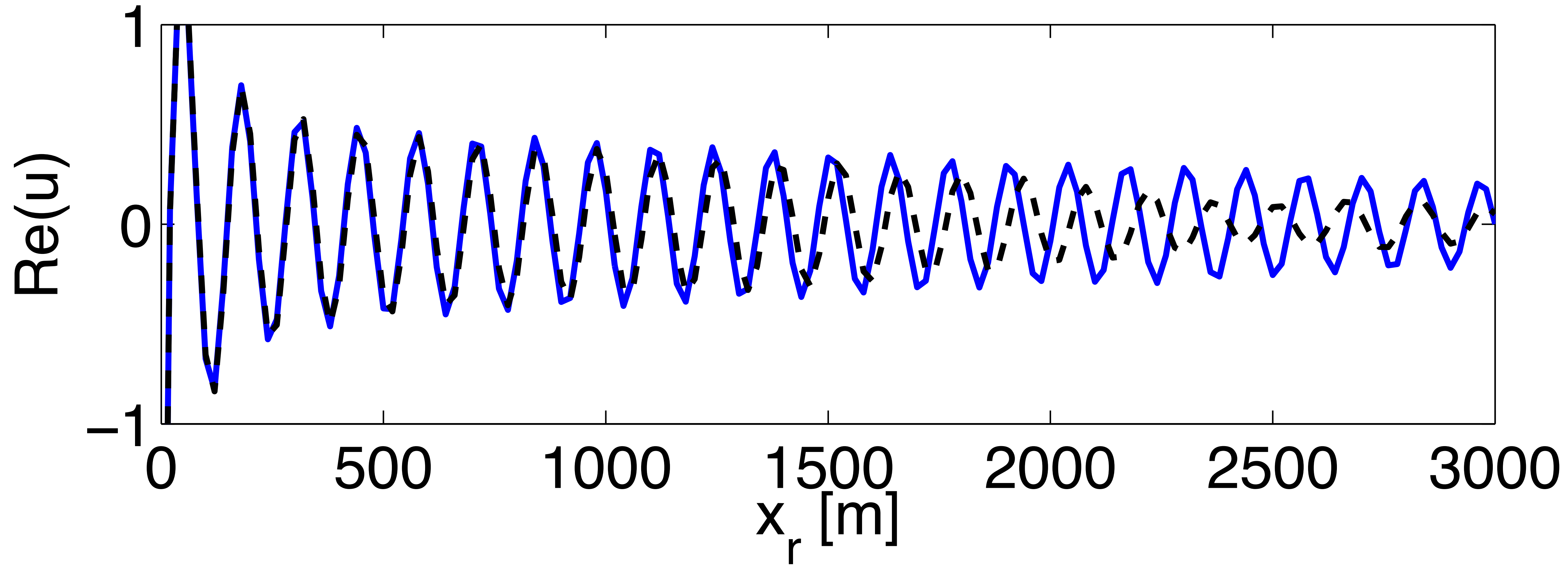
conventional

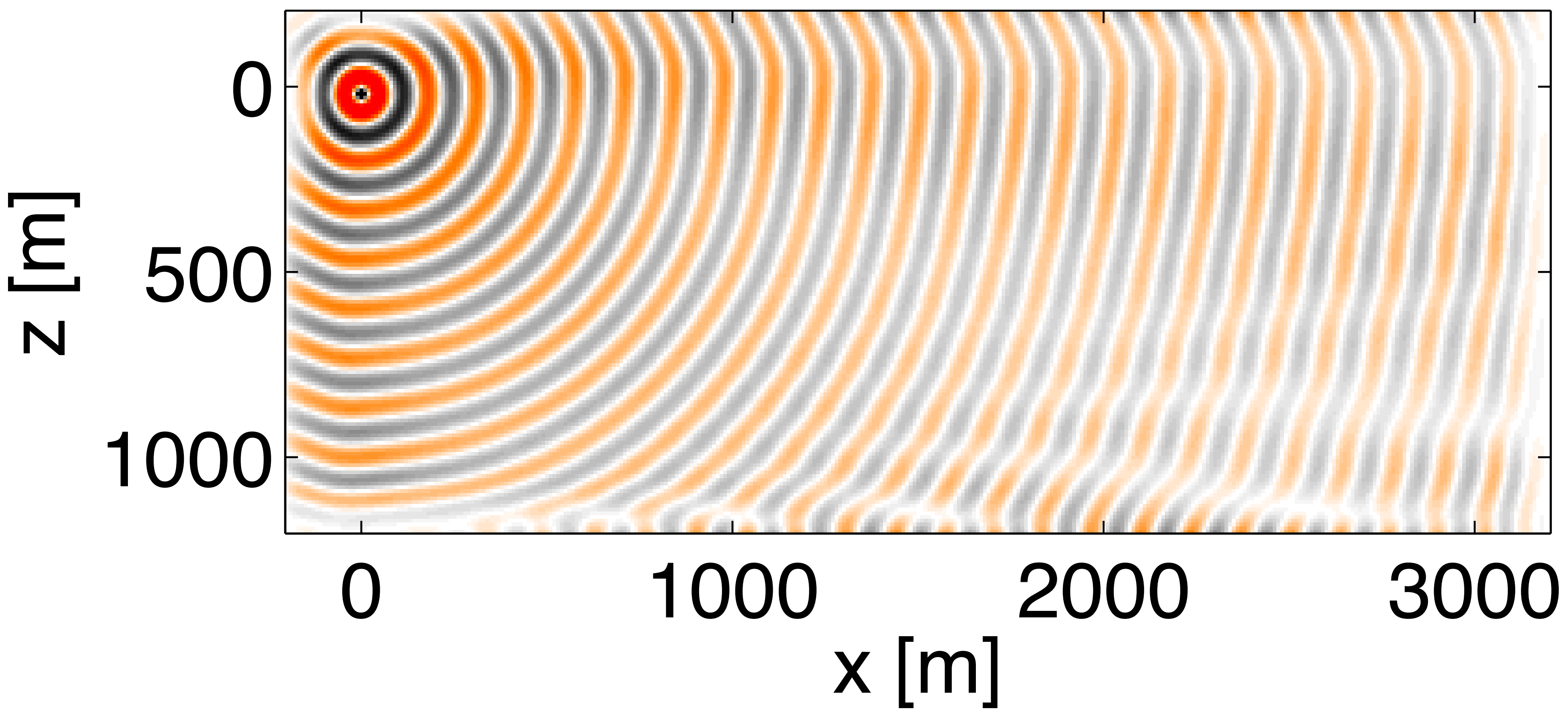


velocity model



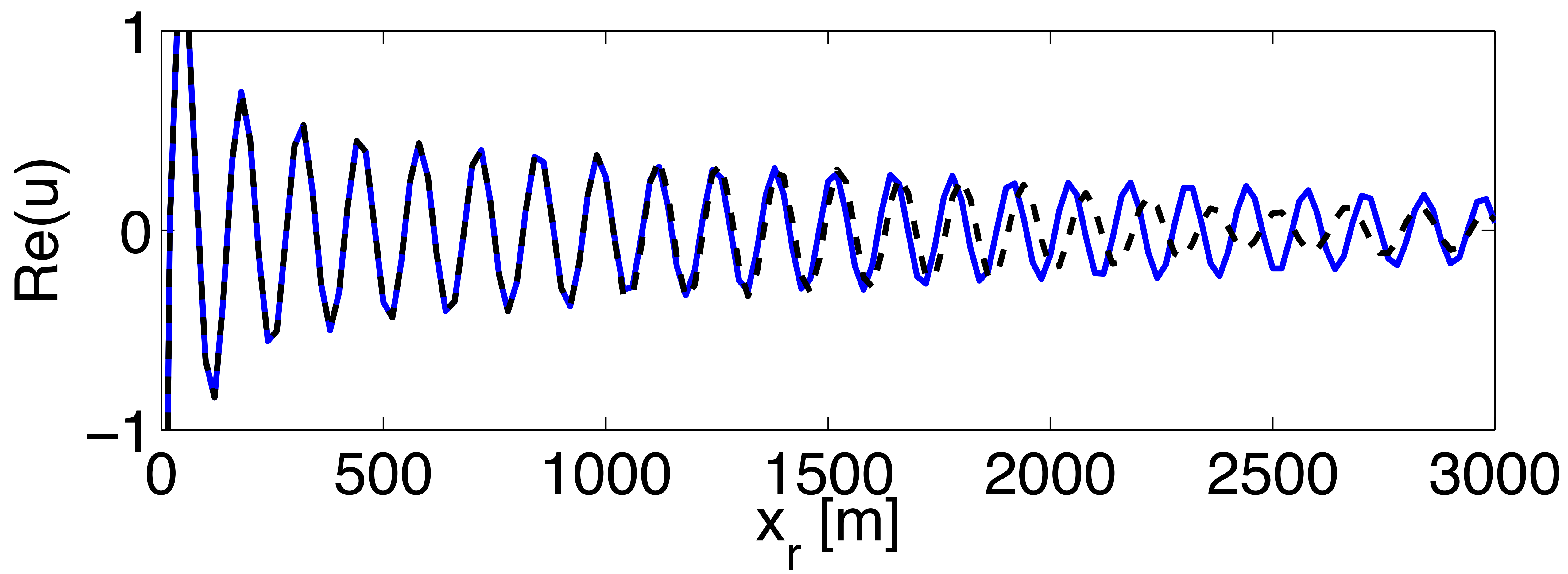
corresponding wavefield

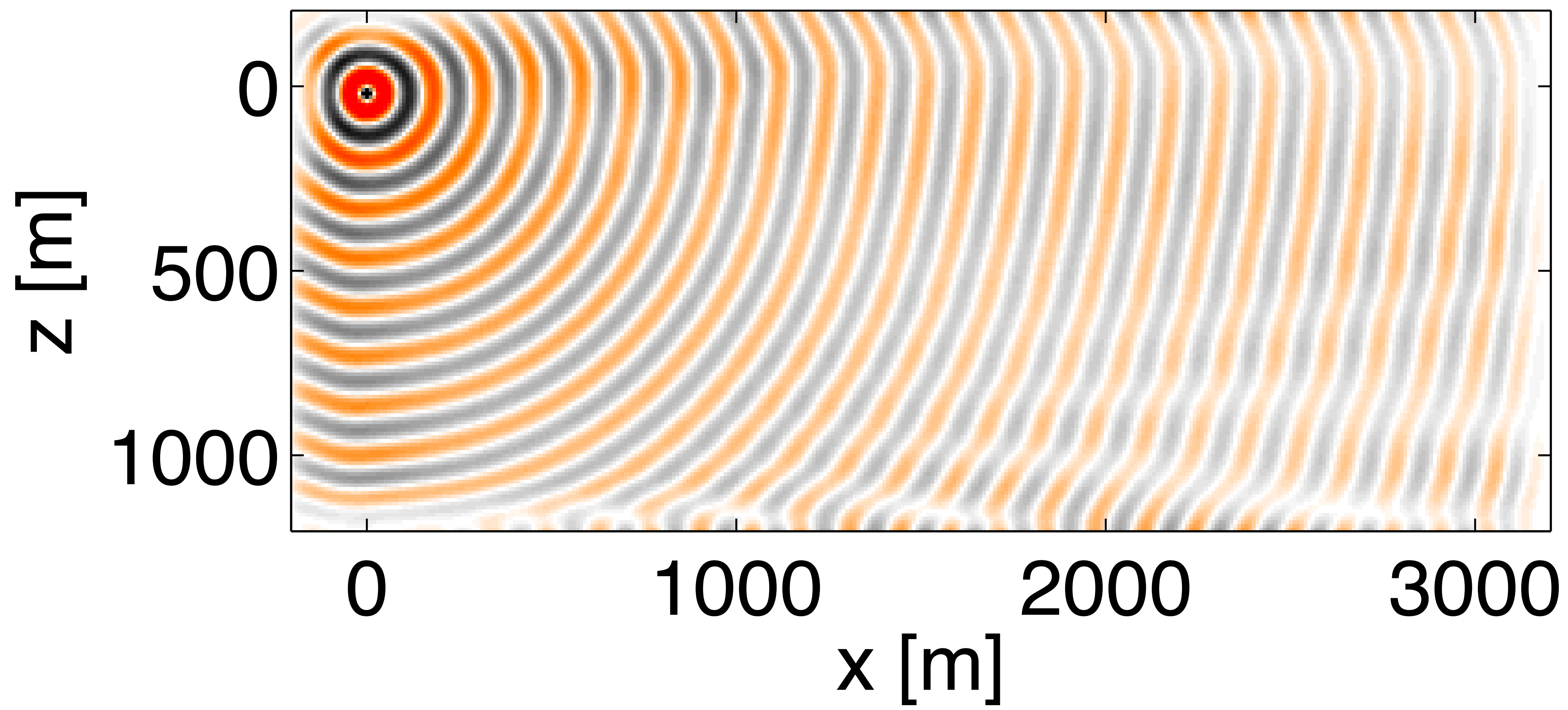




no constraints

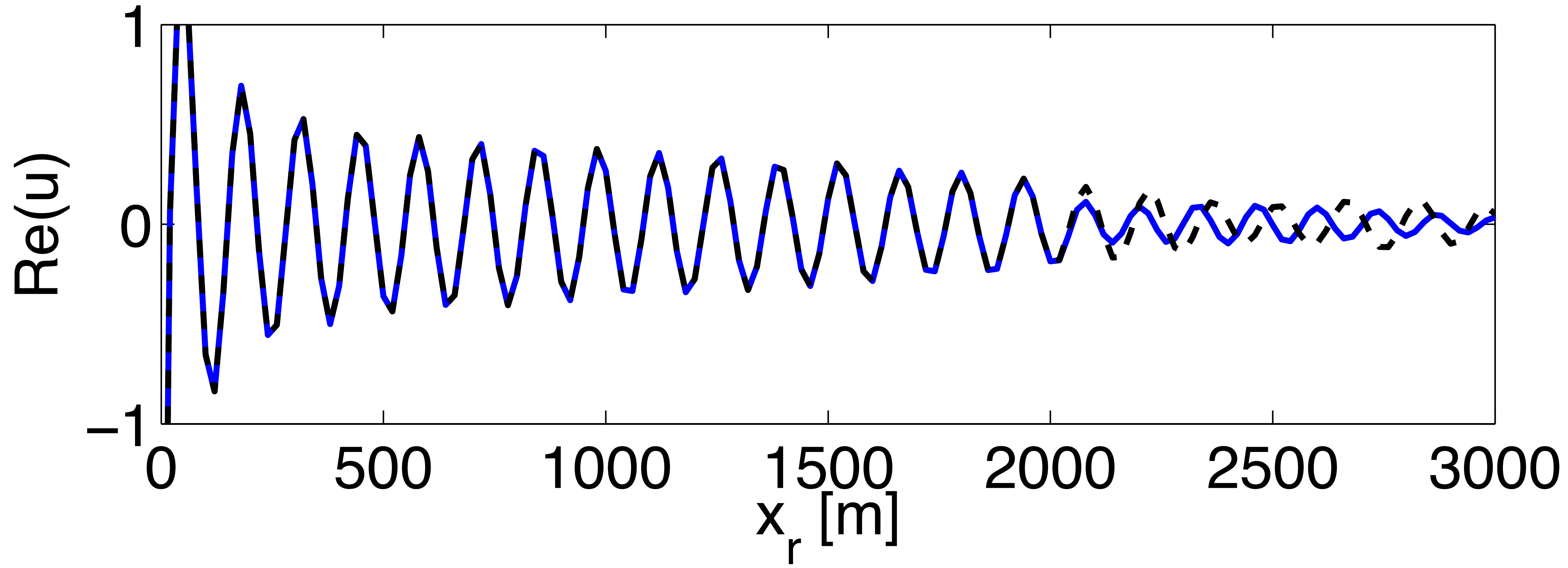
wavefield for constant background velocity

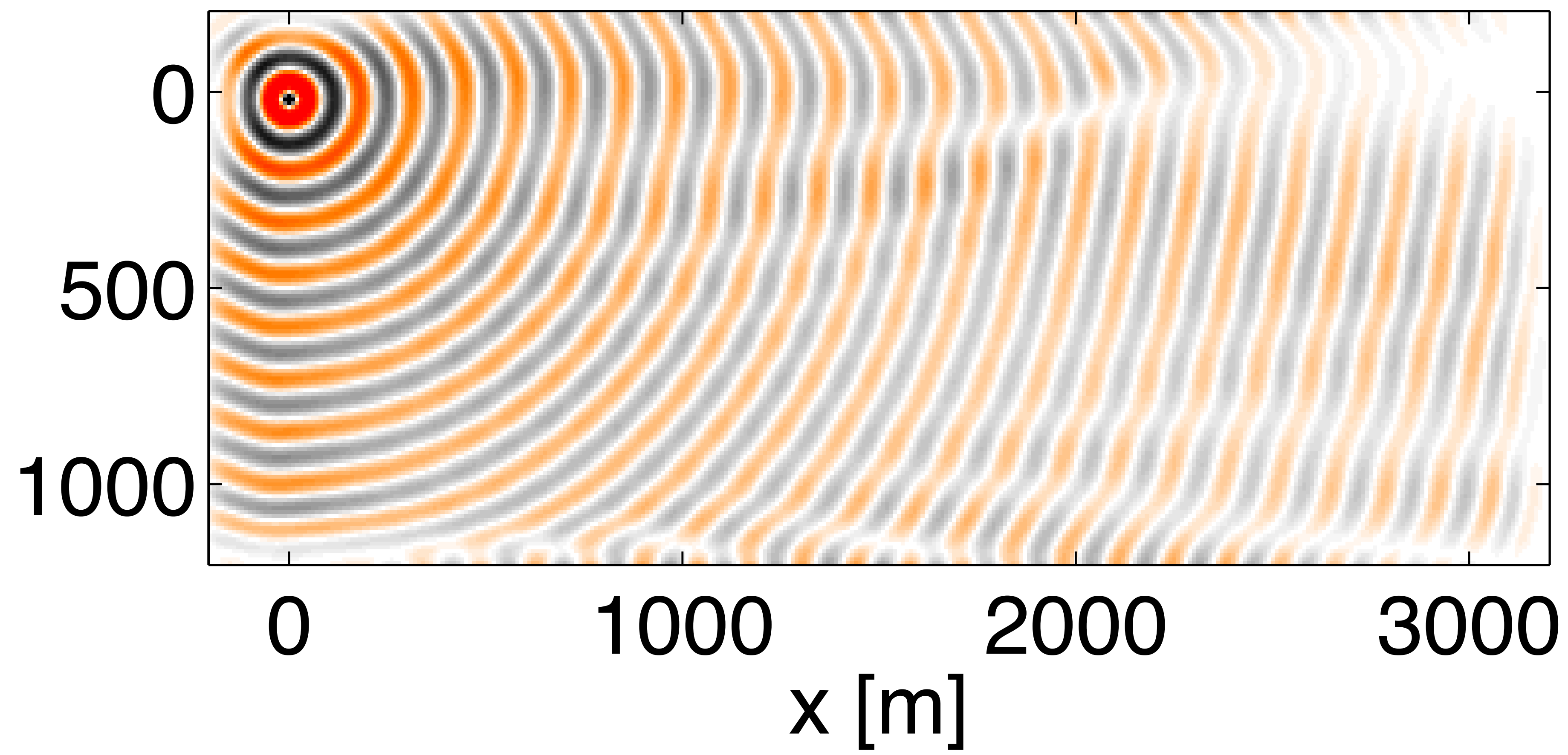




50 receivers

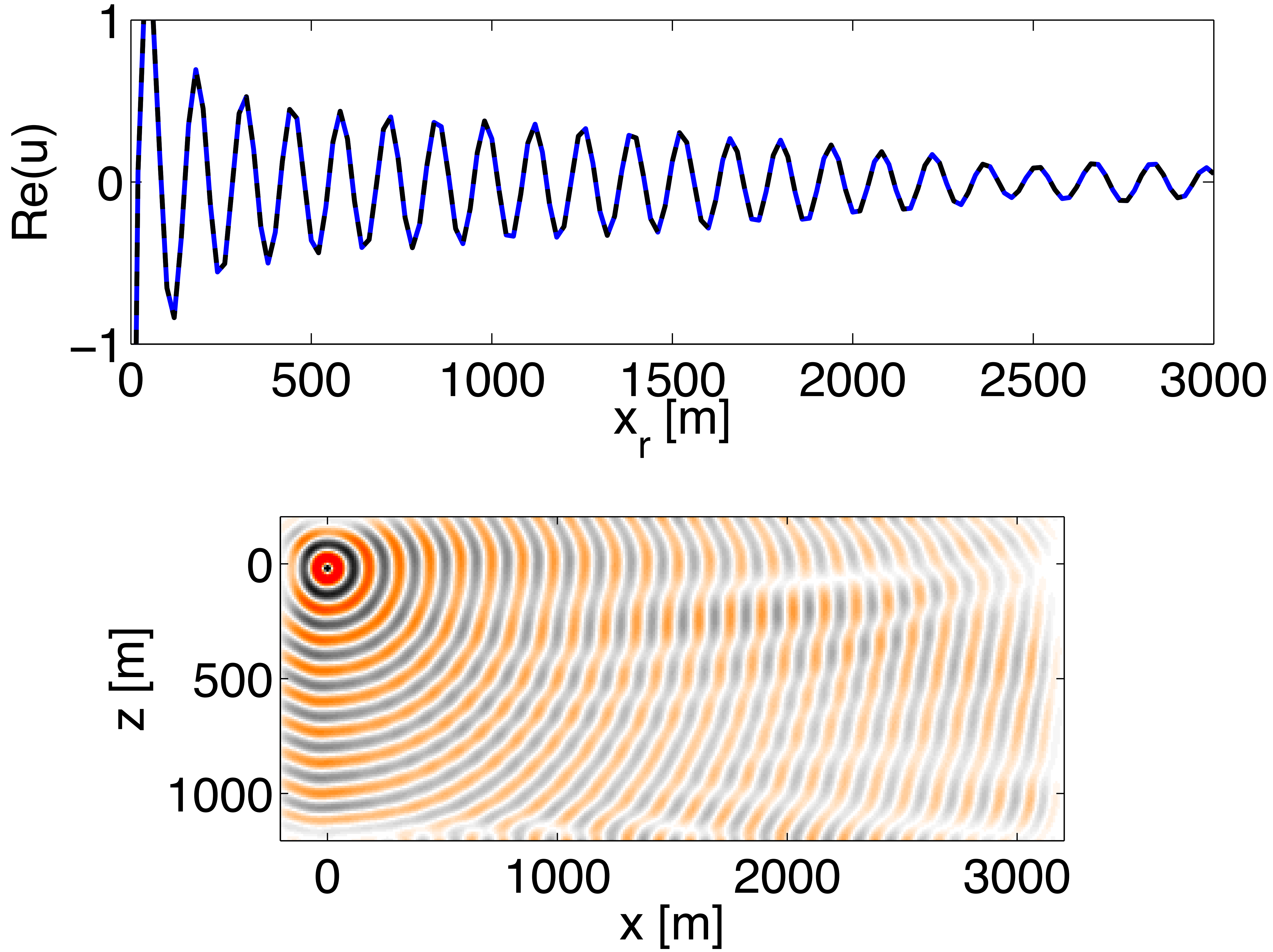
wavefield for constant background velocity





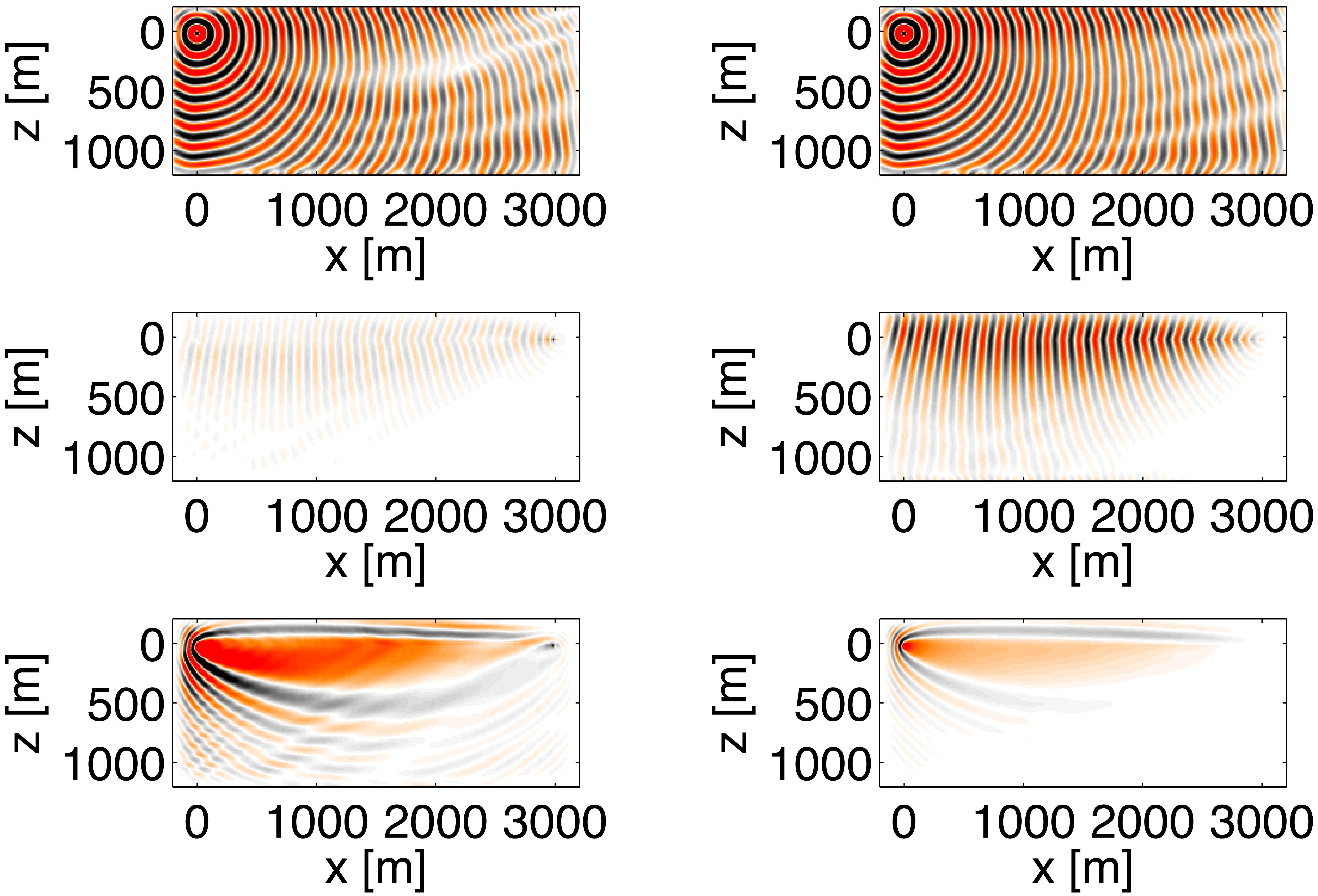
150 receivers

wavefield for constant background velocity



all receivers

wavefield for constant background velocity



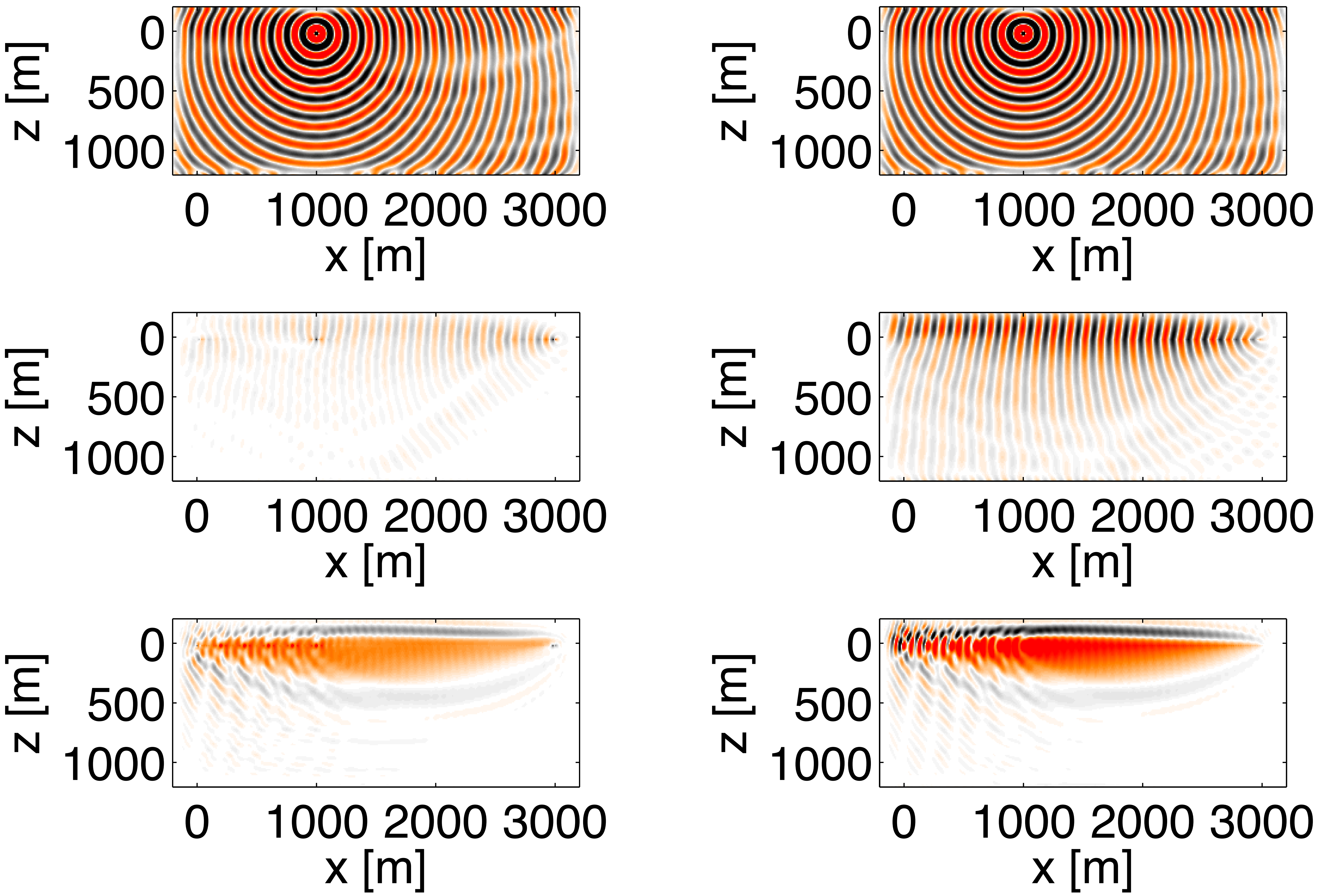
“source wavefield”

“receiver wavefield”

“correlation”

penalty

conventional



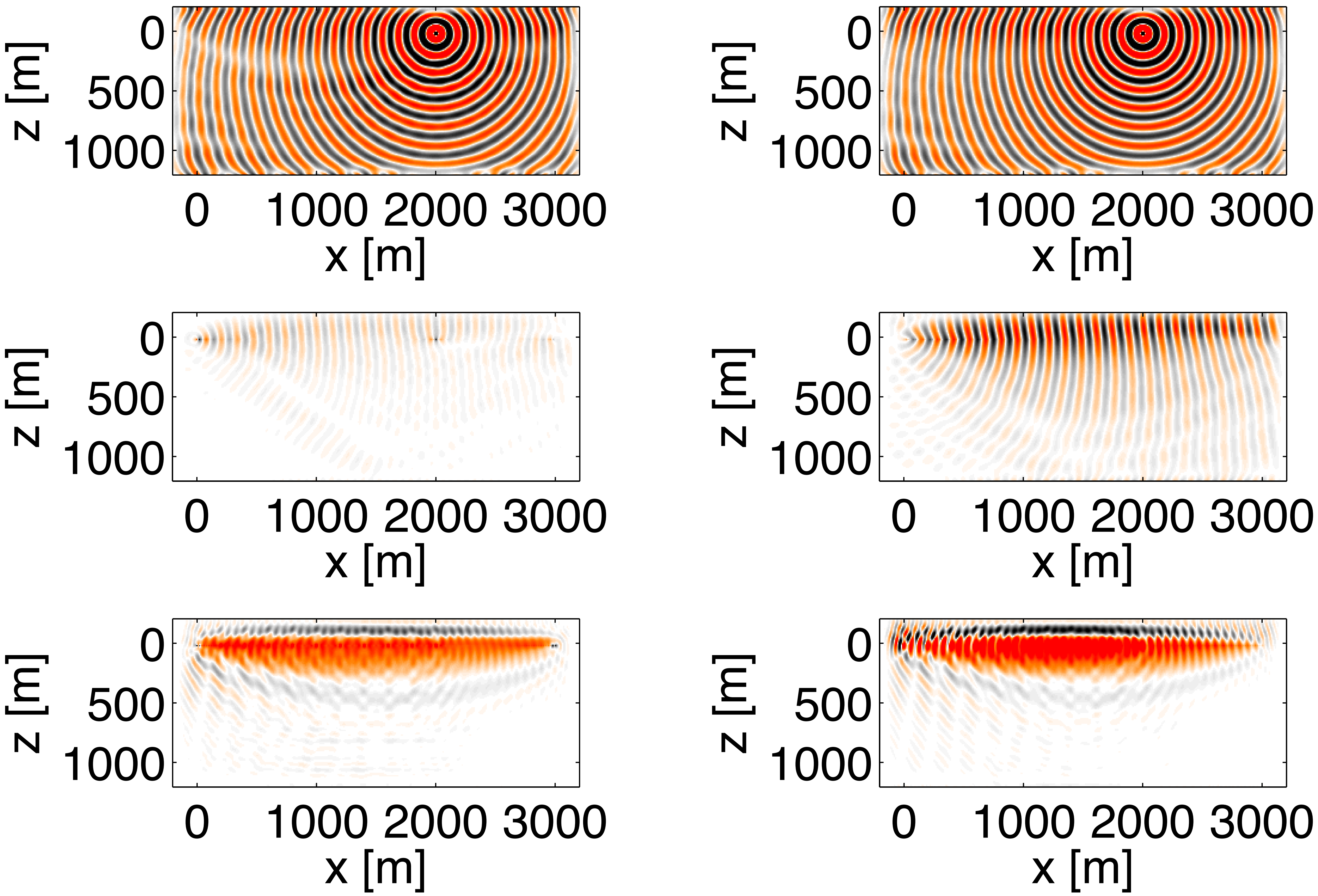
"source wavefield"

"receiver wavefield"

"correlation"

penalty

conventional



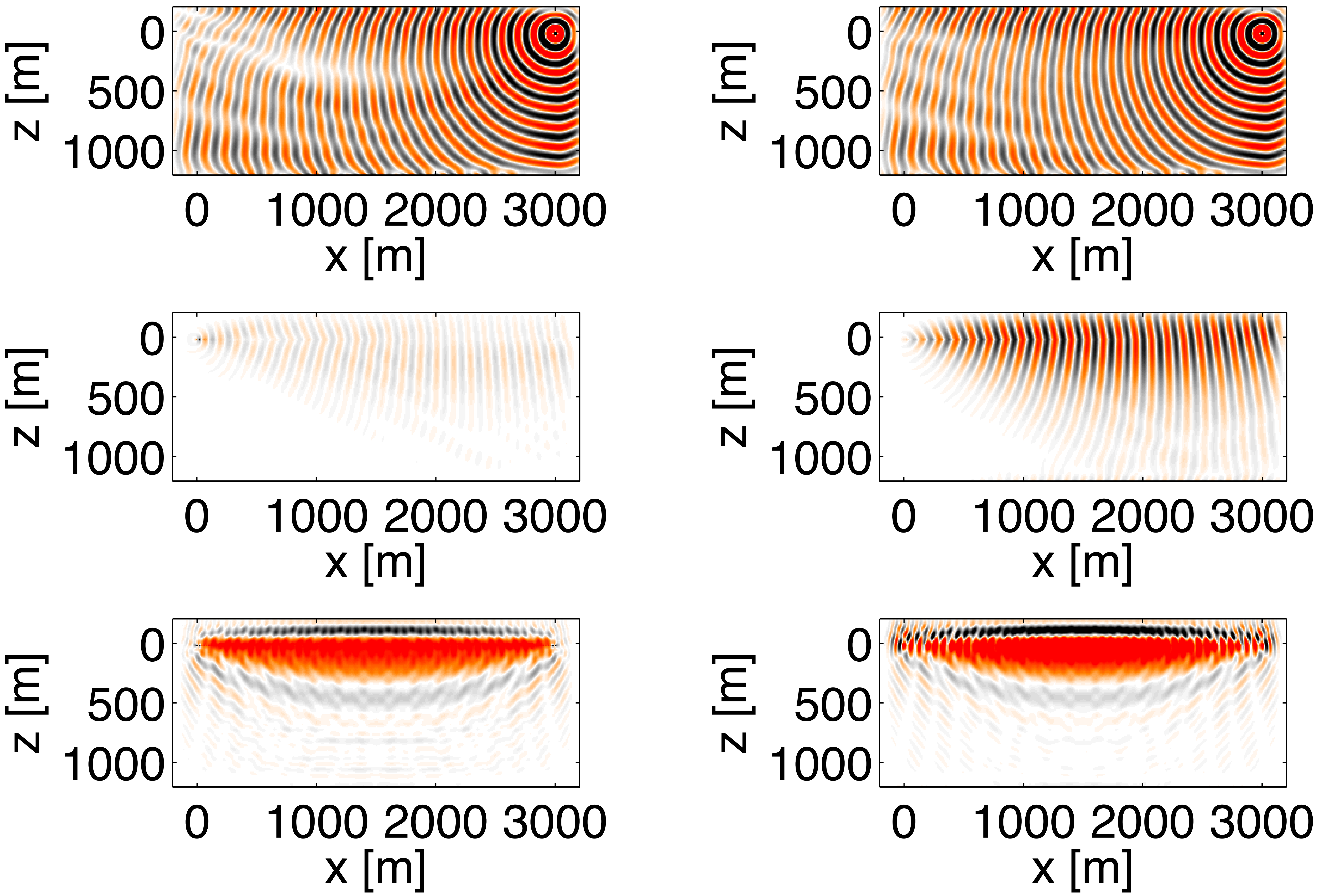
"source wavefield"

"receiver wavefield"

"correlation"

penalty

conventional



“source wavefield”

“receiver wavefield”

“correlation”

penalty

conventional

Connections

Extended modelling

The penalty formulation

$$\min_{\mathbf{m},\mathbf{u}} ||P\mathbf{u} - \mathbf{d}||_2^2 + \lambda^2 ||A(\mathbf{m})\mathbf{u} - \mathbf{q}||_2^2$$

can be interpreted as

$$\min_{\tilde{\mathbf{m}}} \text{misfit}(\tilde{\mathbf{m}}) + \text{annihilator}(\tilde{\mathbf{m}})$$

with $\tilde{\mathbf{m}} = (\mathbf{m}, \mathbf{u})$

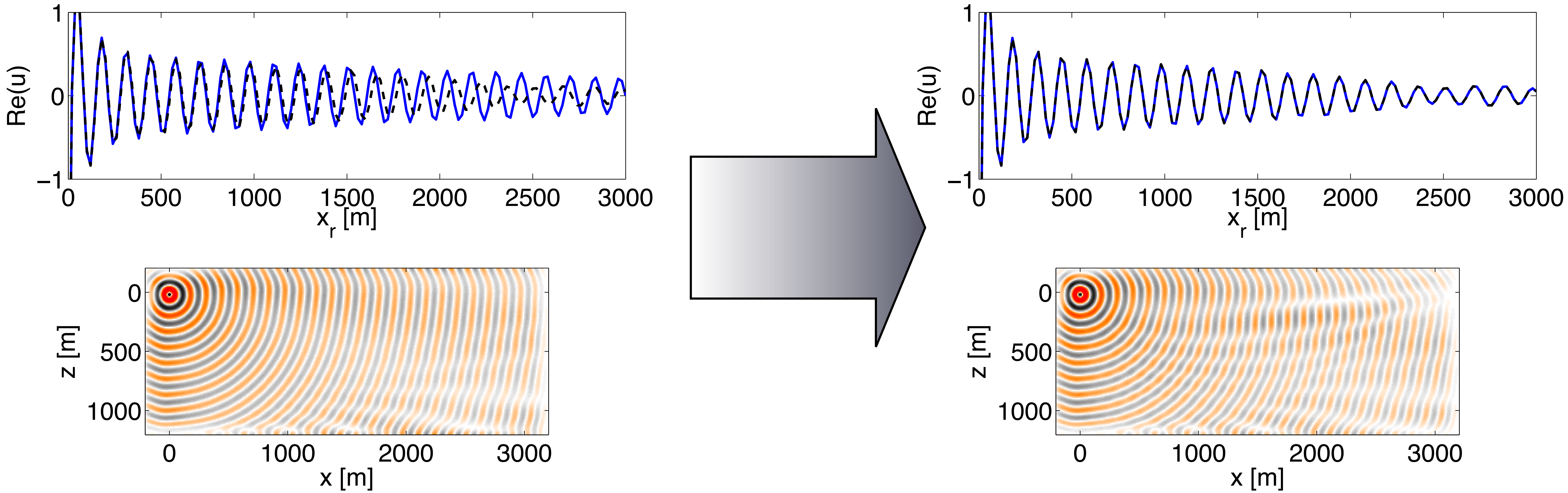
For a physically plausible model we have

$$\text{annihilator}(\tilde{\mathbf{m}}) = 0$$

[Symes, personal communication]

Warping

The overdetermined WE is a way of warping



[Baek ’13, Ma ’13]

ADMM

$$\min_{\mathbf{m},\mathbf{u}} \max_{\mathbf{v}} ||P\mathbf{u} - \mathbf{d}||_2^2 + \mathbf{v}^*(A(\mathbf{m})\mathbf{u} - \mathbf{q}) + \frac{\lambda}{2}||A(\mathbf{m}_k)\mathbf{u} - \mathbf{q}||_2^2$$

Alternating updates

$$\mathbf{u}_{k+1} = \mathrm{argmin}_{\mathbf{u}} ||P\mathbf{u} - \mathbf{d}||_2^2 + \mathbf{v}_k^*(A(\mathbf{m}_k)\mathbf{u} - \mathbf{q}) + \frac{\lambda}{2}\|A(\mathbf{m}_k)\mathbf{u} - \mathbf{q}\|_2^2$$

$$\mathbf{m}_{k+1} = \mathrm{argmin}_{\mathbf{m}} ||P\mathbf{u}_{k+1} - \mathbf{d}||_2^2 + \mathbf{v}_k^*(A(\mathbf{m}_k)\mathbf{u}_{k+1} - \mathbf{q}) + \frac{\lambda}{2}\|A(\mathbf{m})\mathbf{u}_{k+1} - \mathbf{q}\|_2^2$$

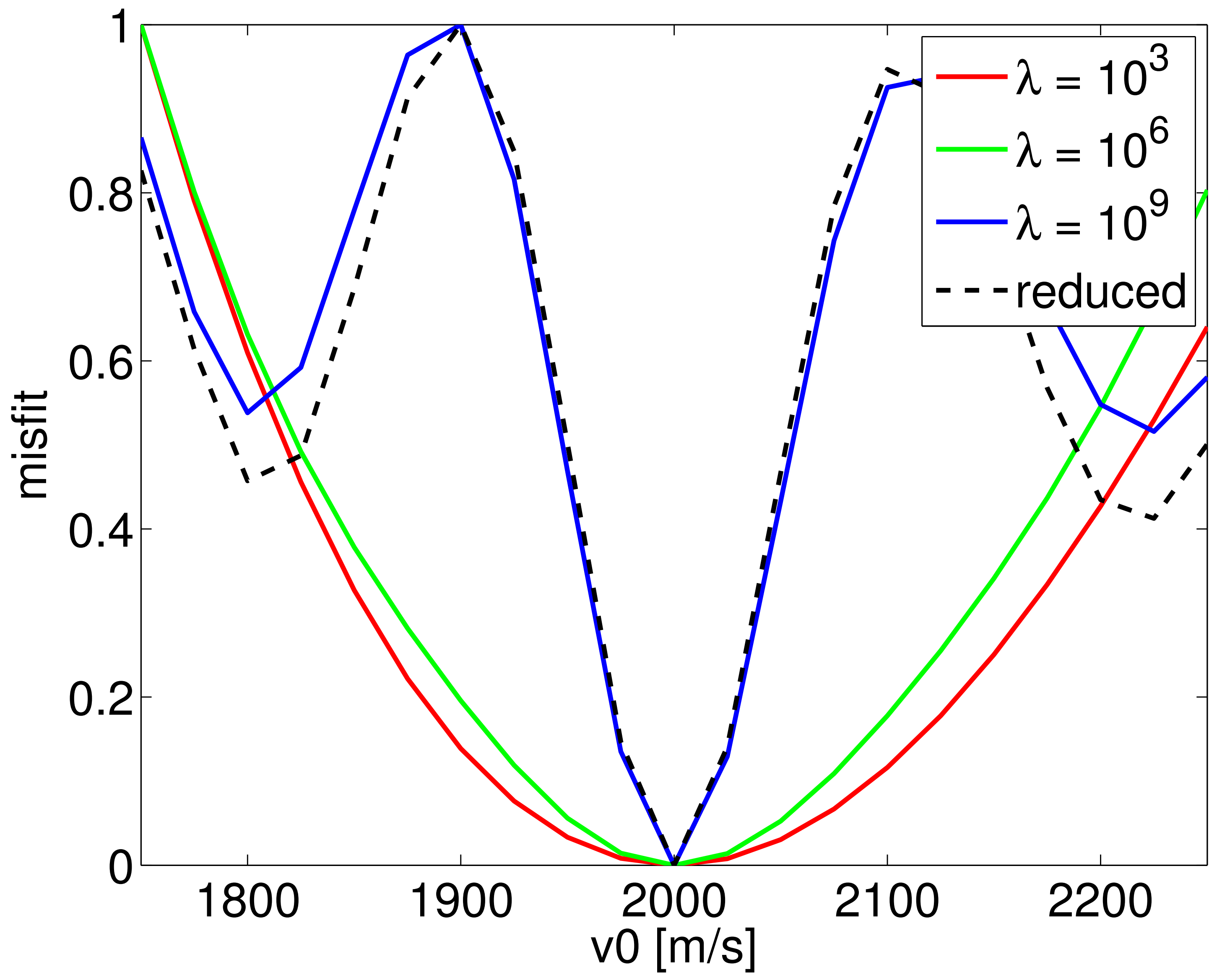
$$\mathbf{v}_{k+1} = \mathbf{v}_k + \lambda\left(A(\mathbf{m}_{k+1})\mathbf{u}_{k+1} - \mathbf{q}\right)$$

So first iteration starting from $\mathbf{v}_k = 0$ is similar to penalty method.

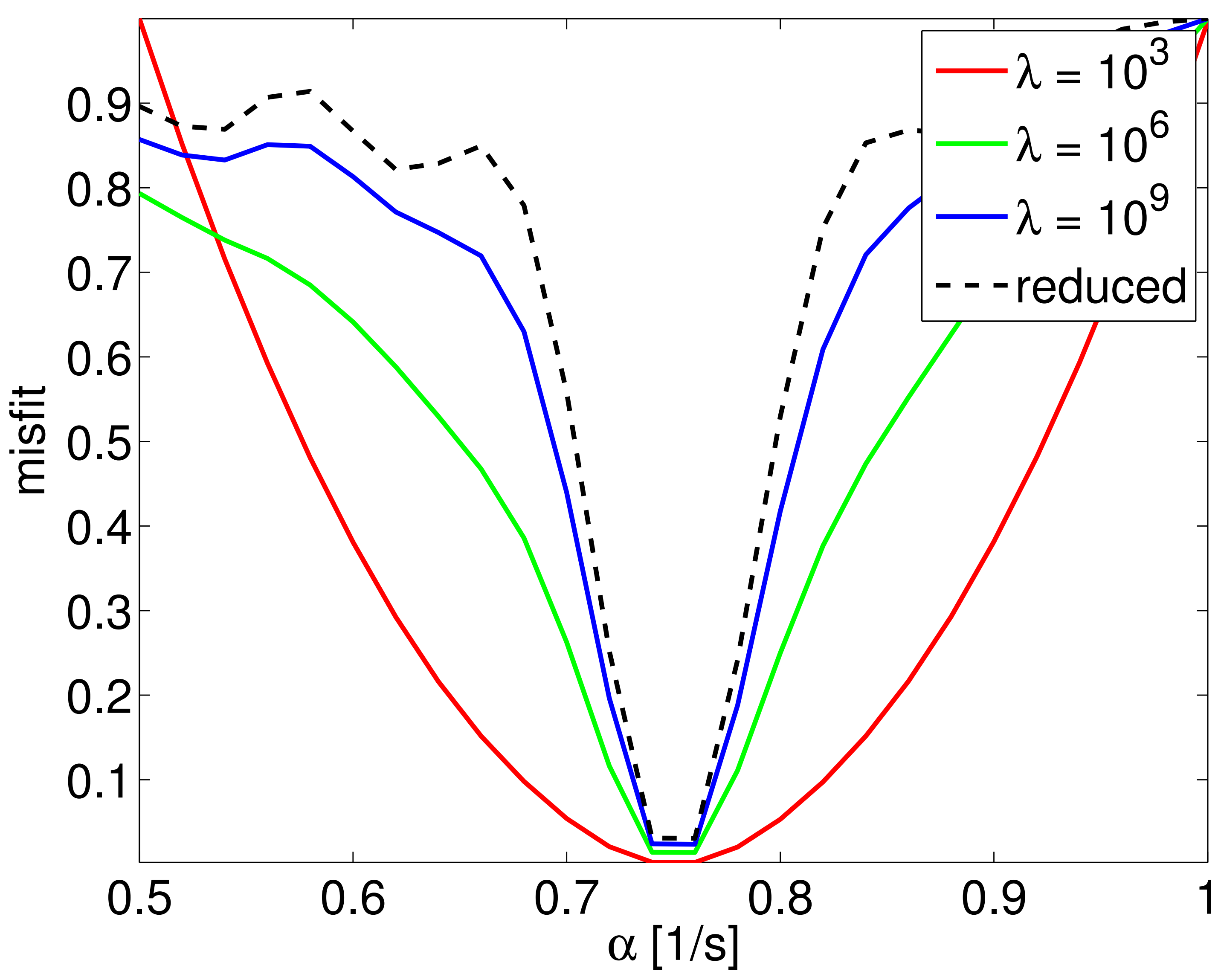
Numerical examples

Local minima

single shot, single frequency data for *linear* velocity profile $v(z) = v_0 + \alpha z$, misfit as function of (v_0, α)



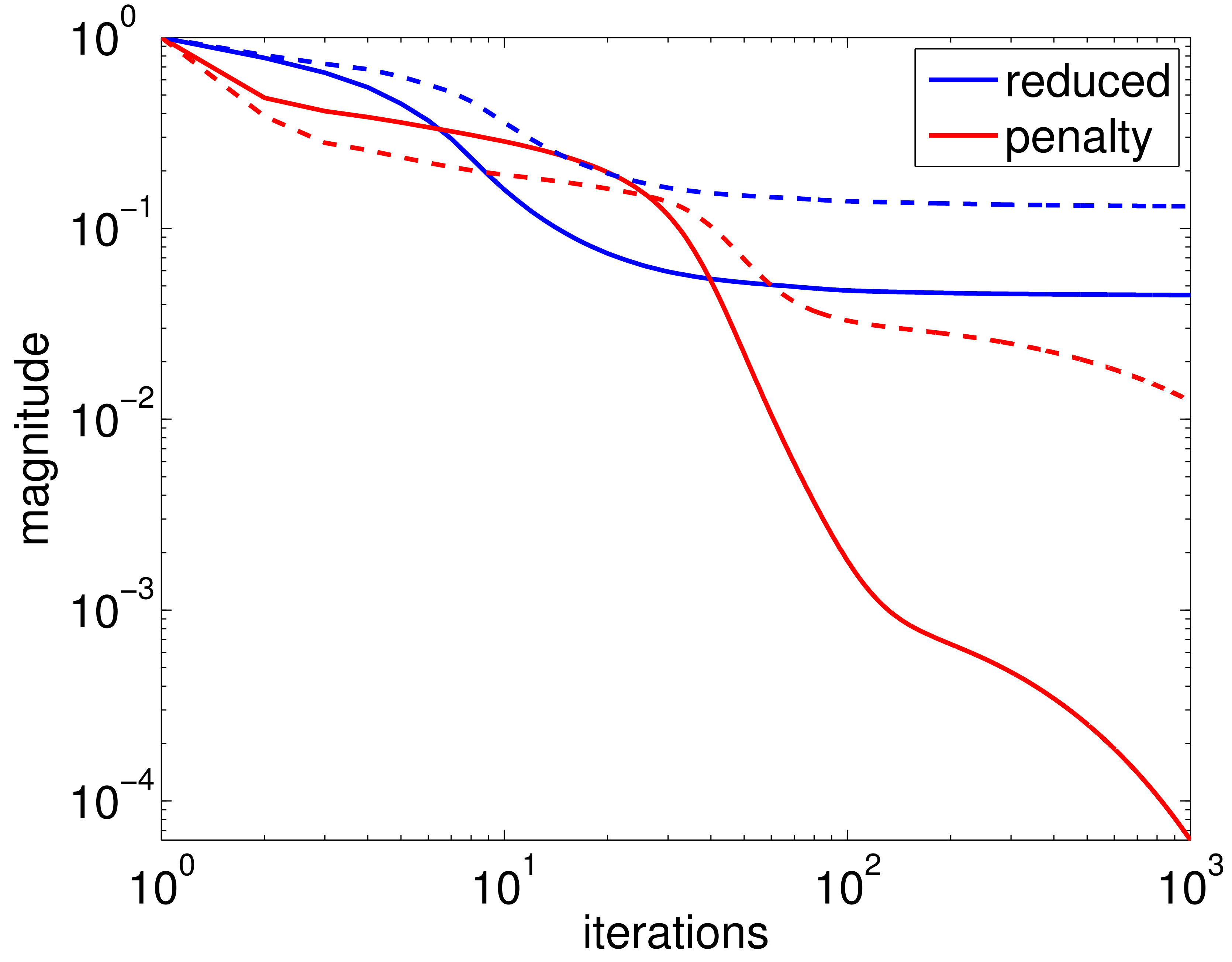
constant



slope

Local minima

Invert *single* shot, *single* frequency data for 1D velocity profile using *steepest* descent.

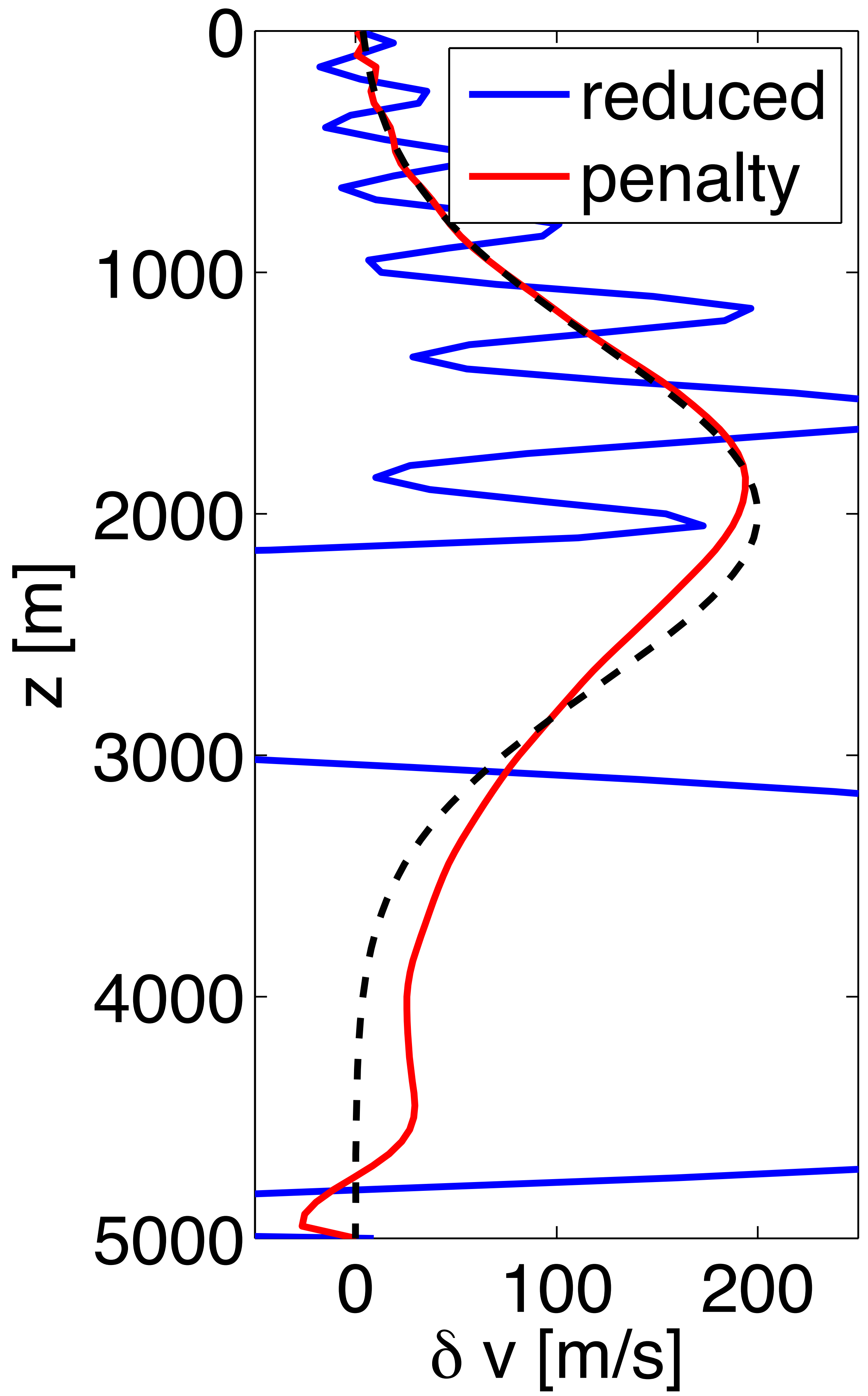


- misfit
-- gradient

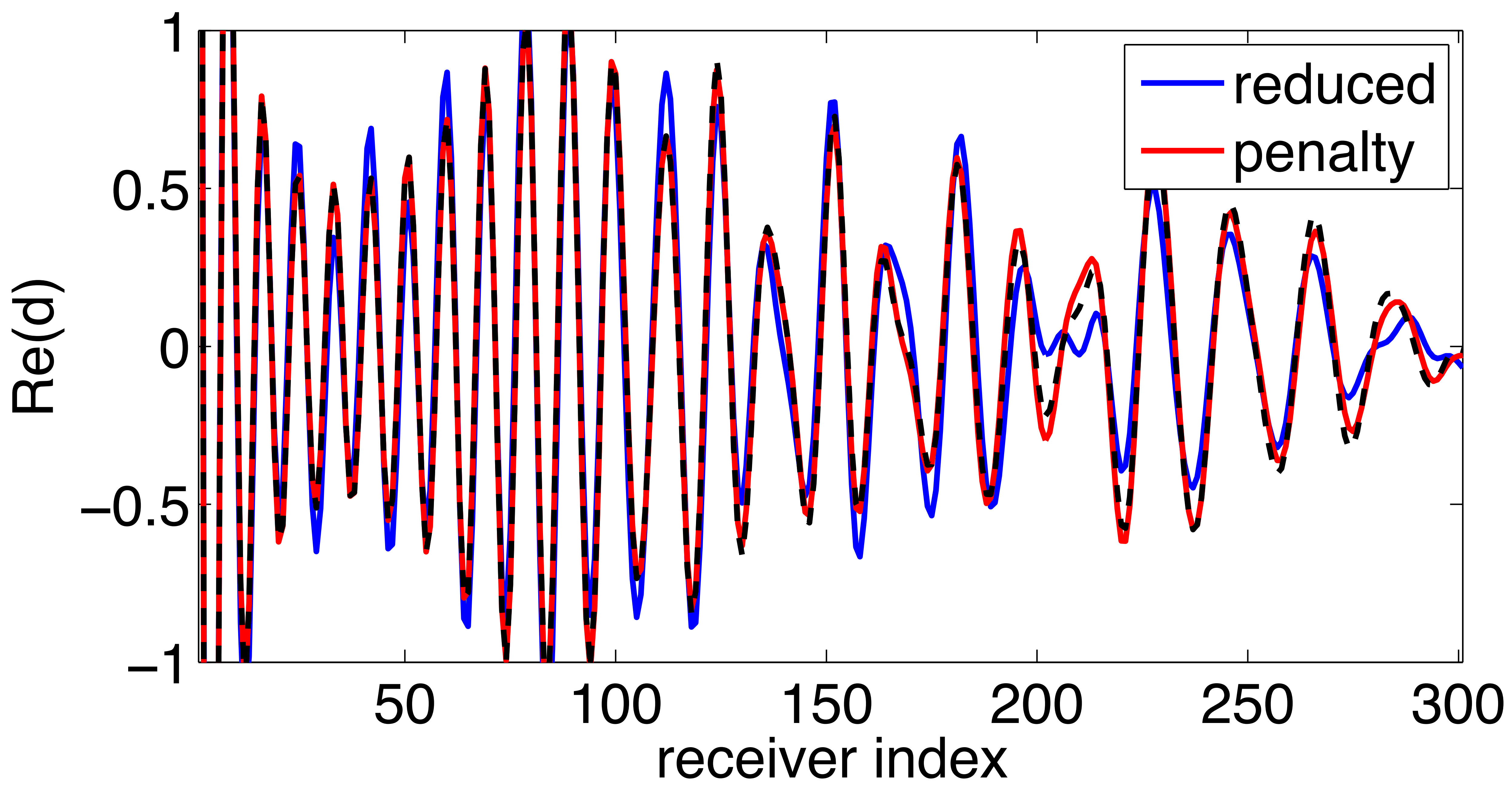
misfit & gradient as function # iterations

Local minima

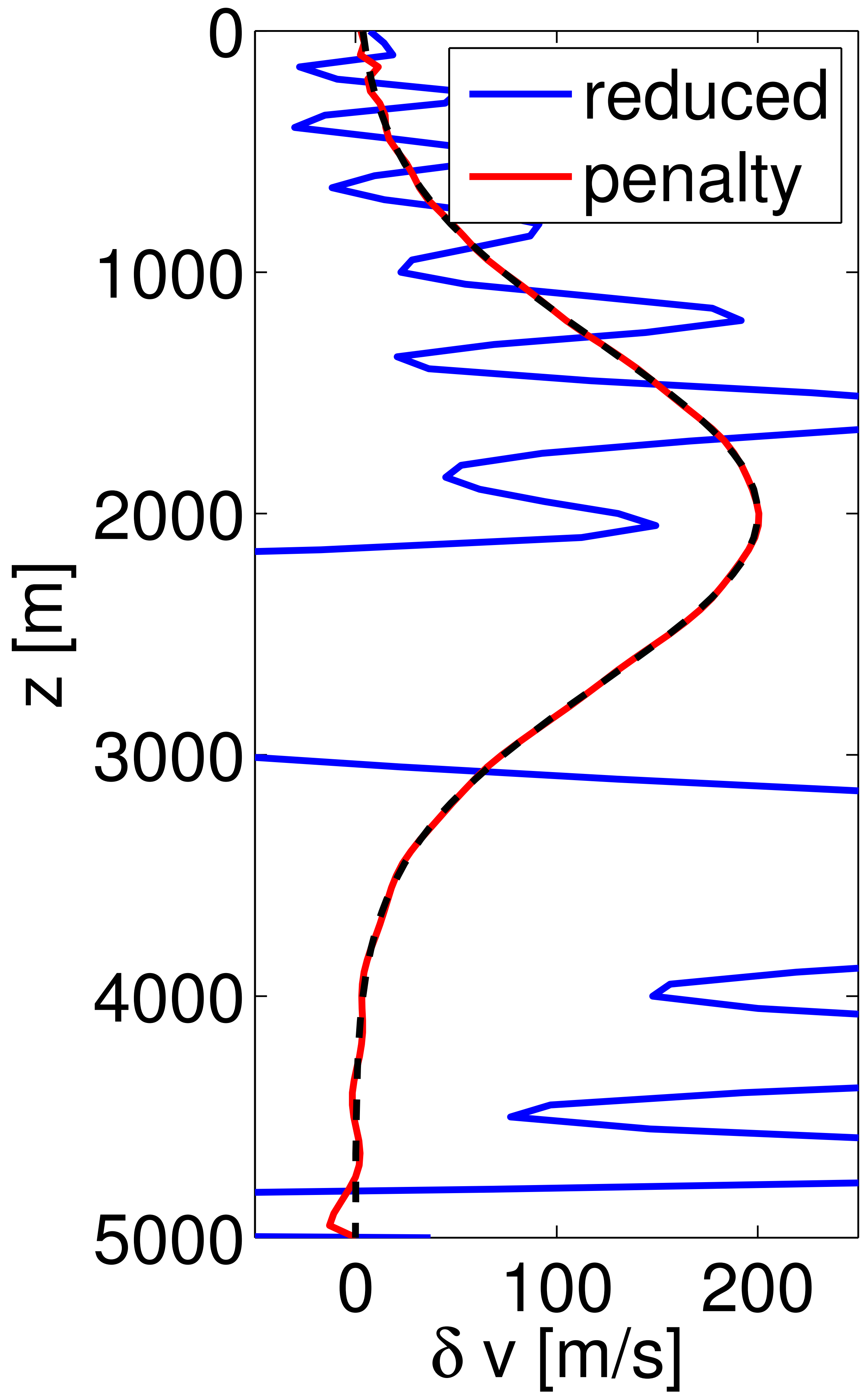
SLIM



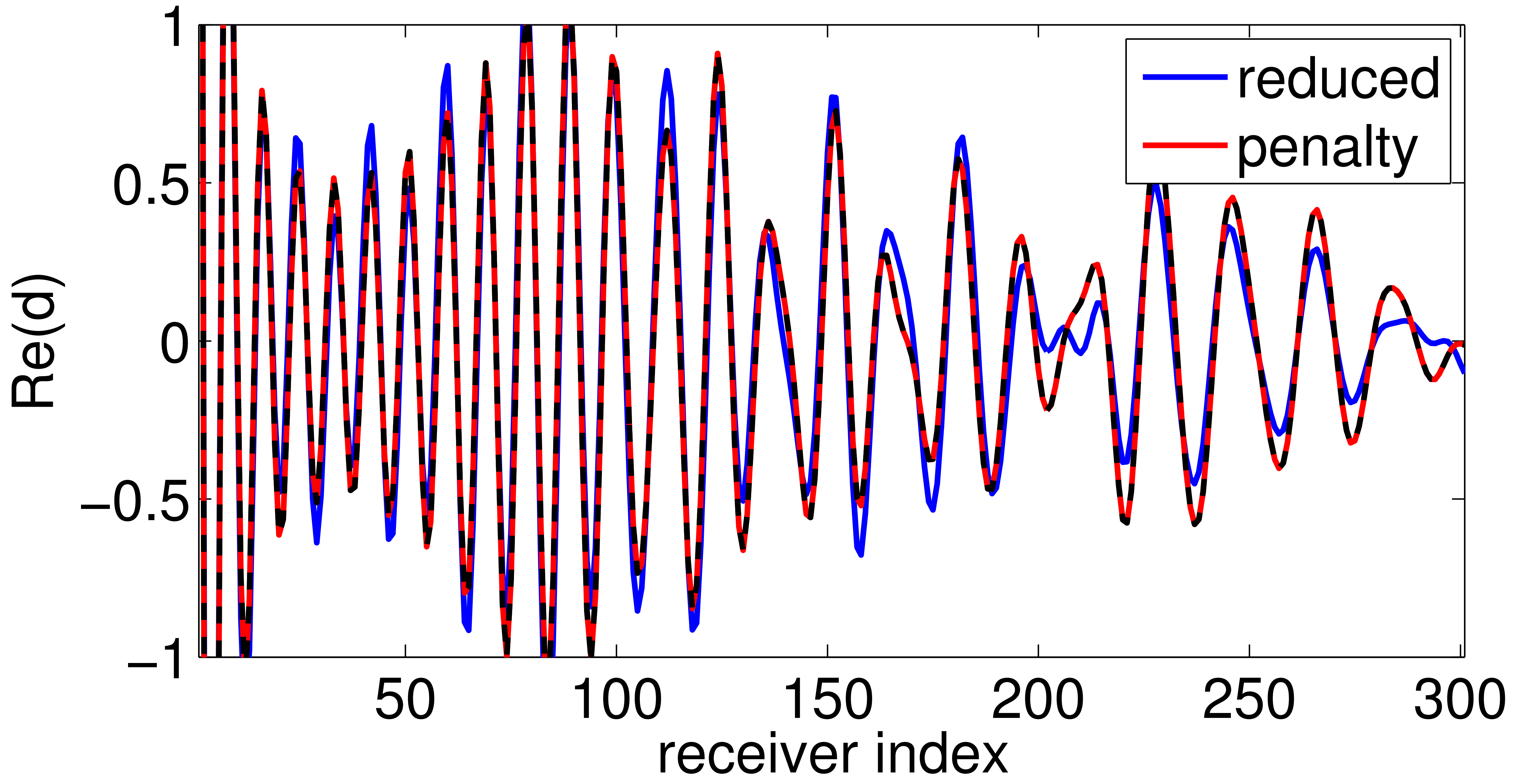
After 100 iterations



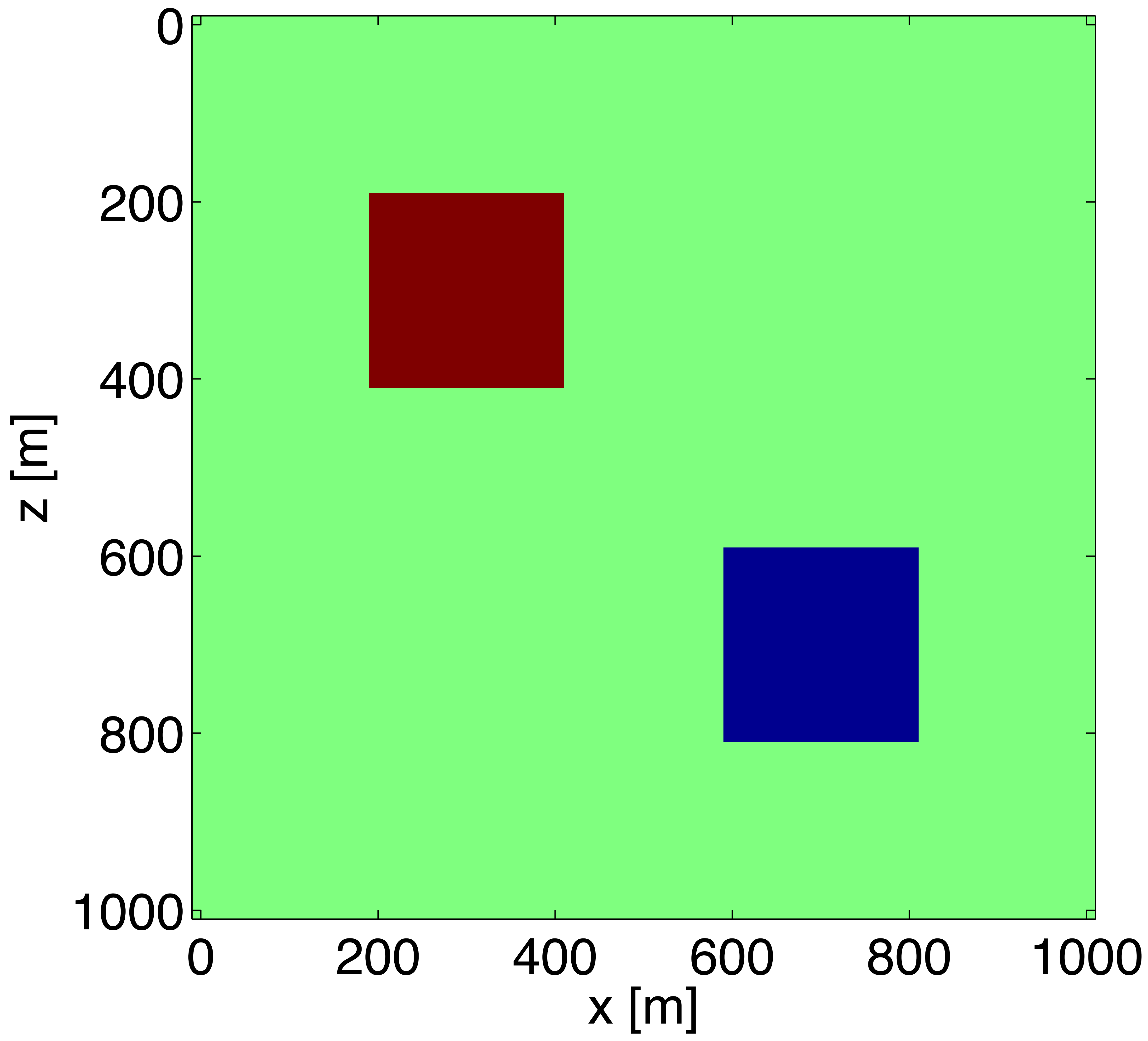
Local minima



After 1000 iterations



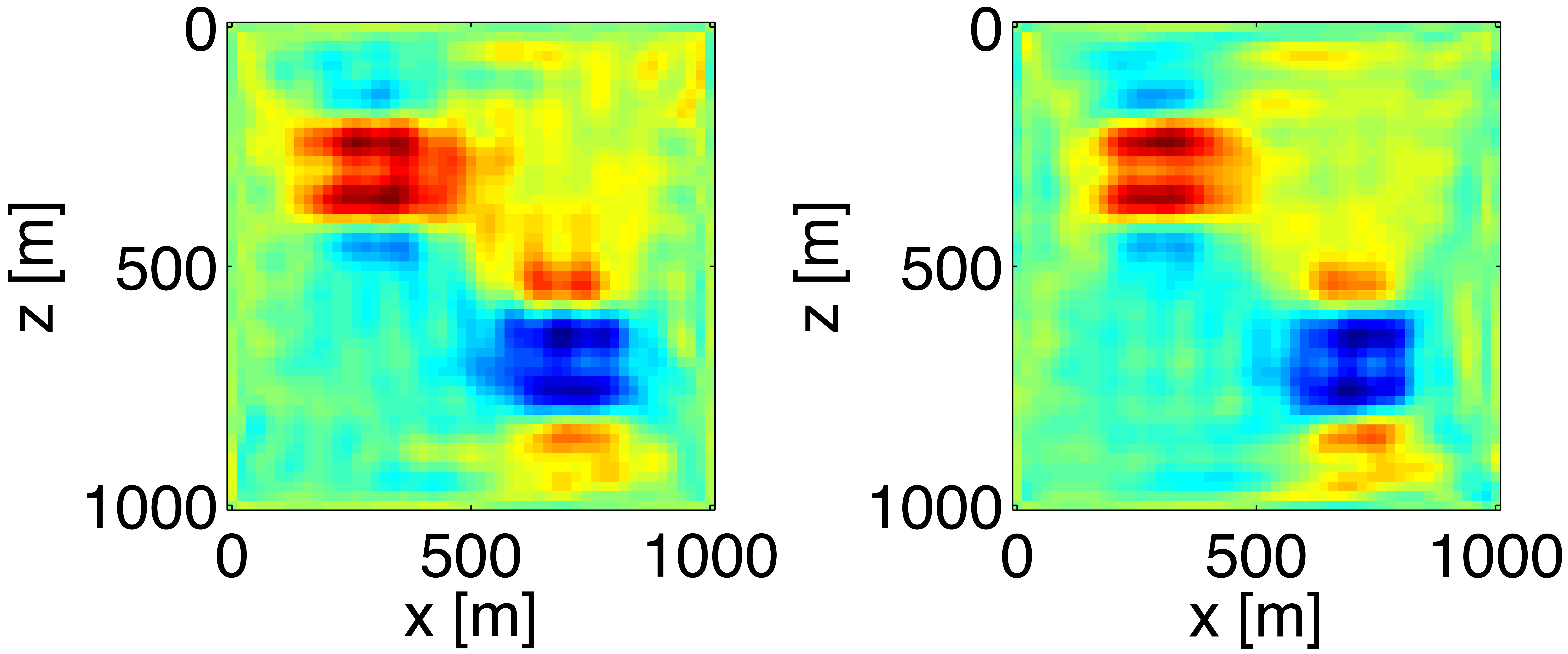
Cross well



original medium

- 2D cross-well configuration
- 5 point Helmholtz FD operator with abc
- Direct solver for PDEs (A\b)
- invert a single frequency (10 Hz)
- L-BFGS

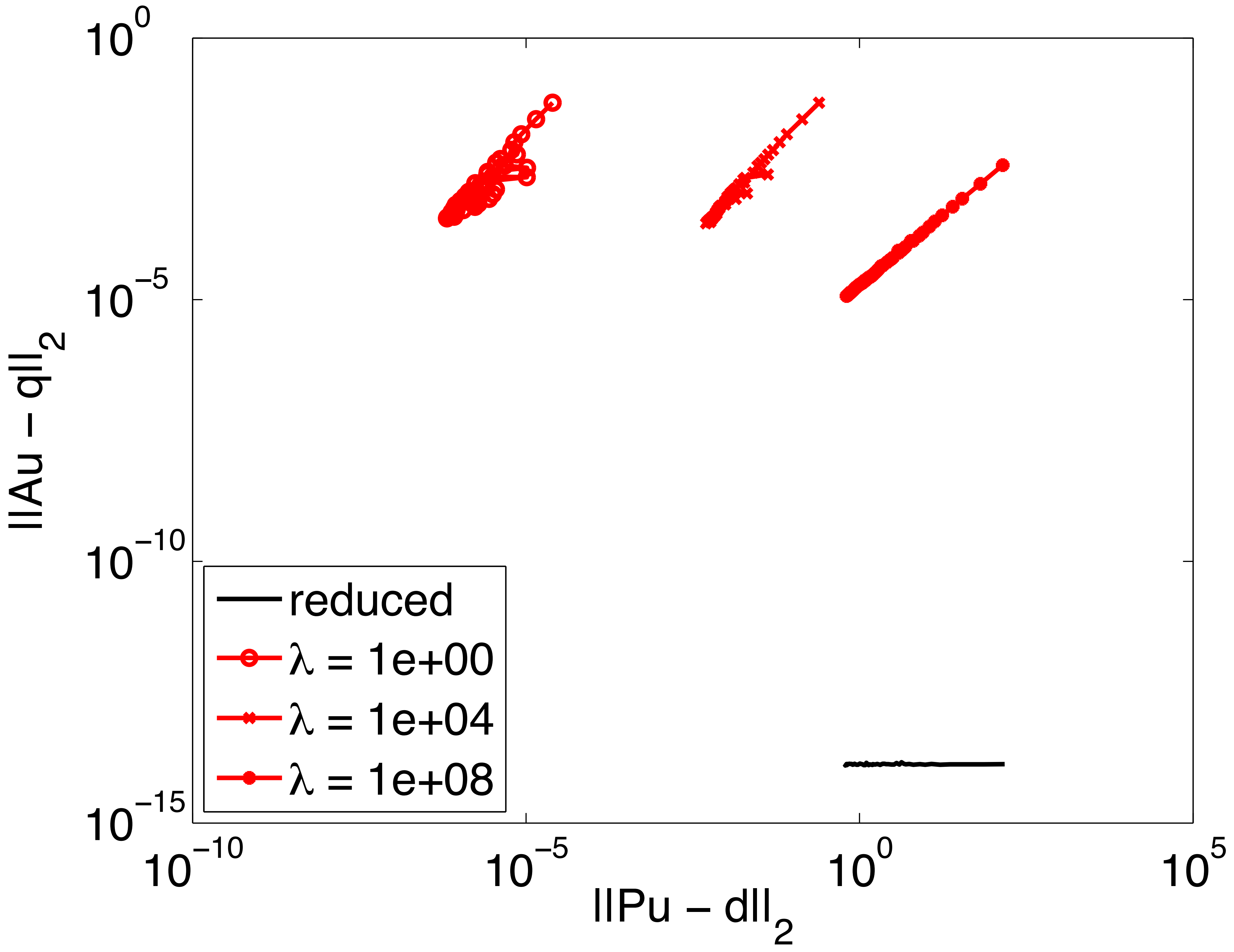
Cross well FWI

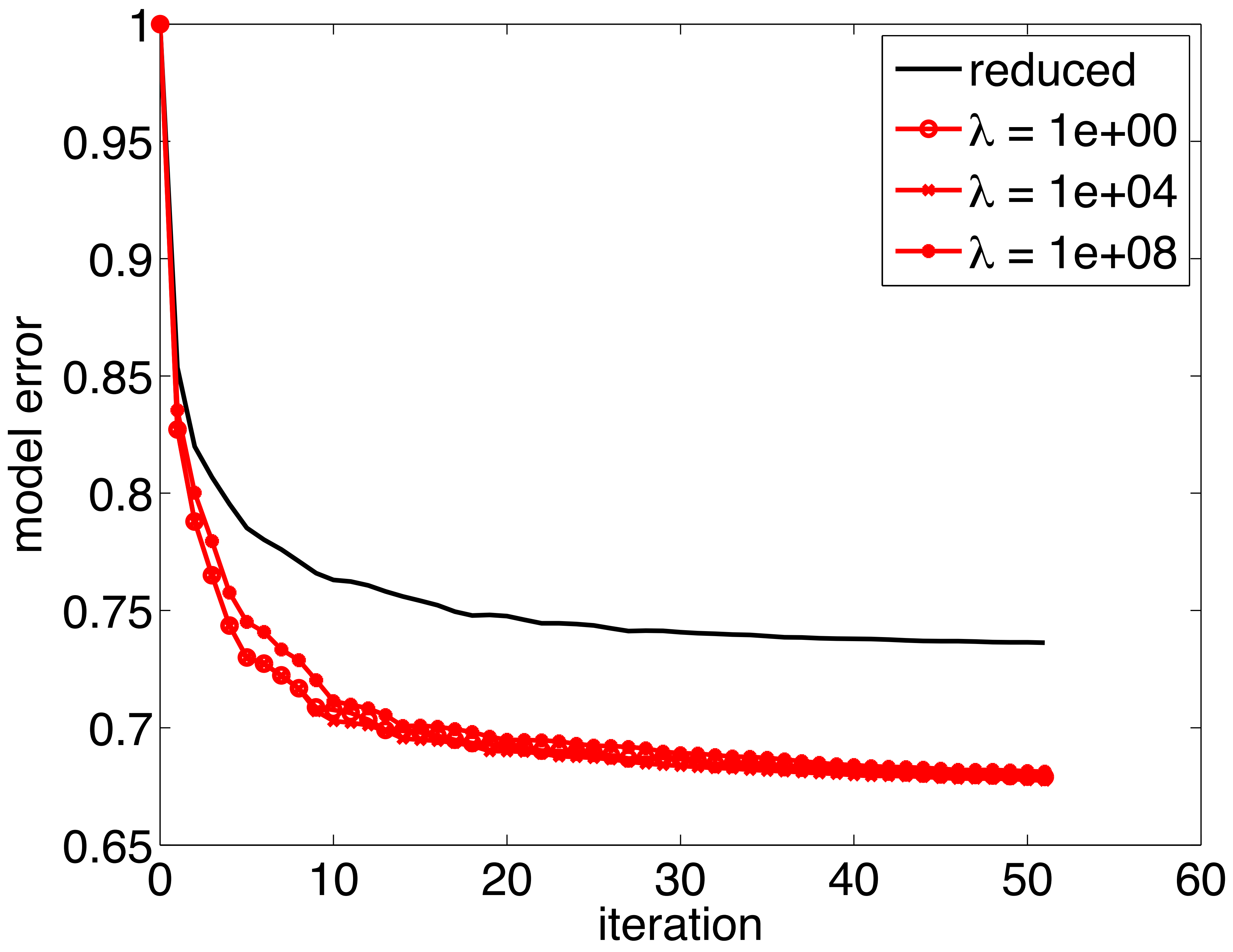


conventional

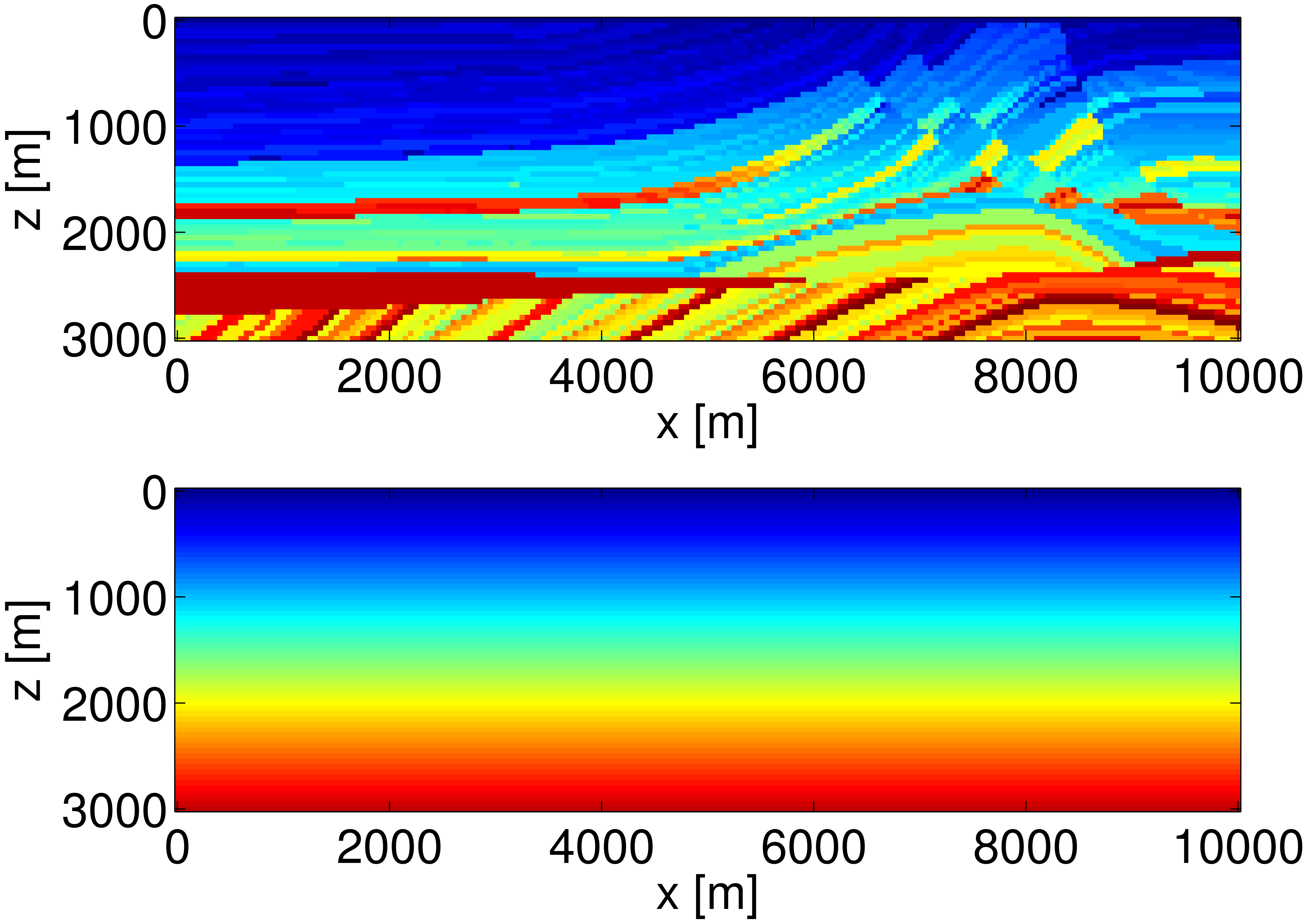
penalty

Cross well FWI





Waveform inversion

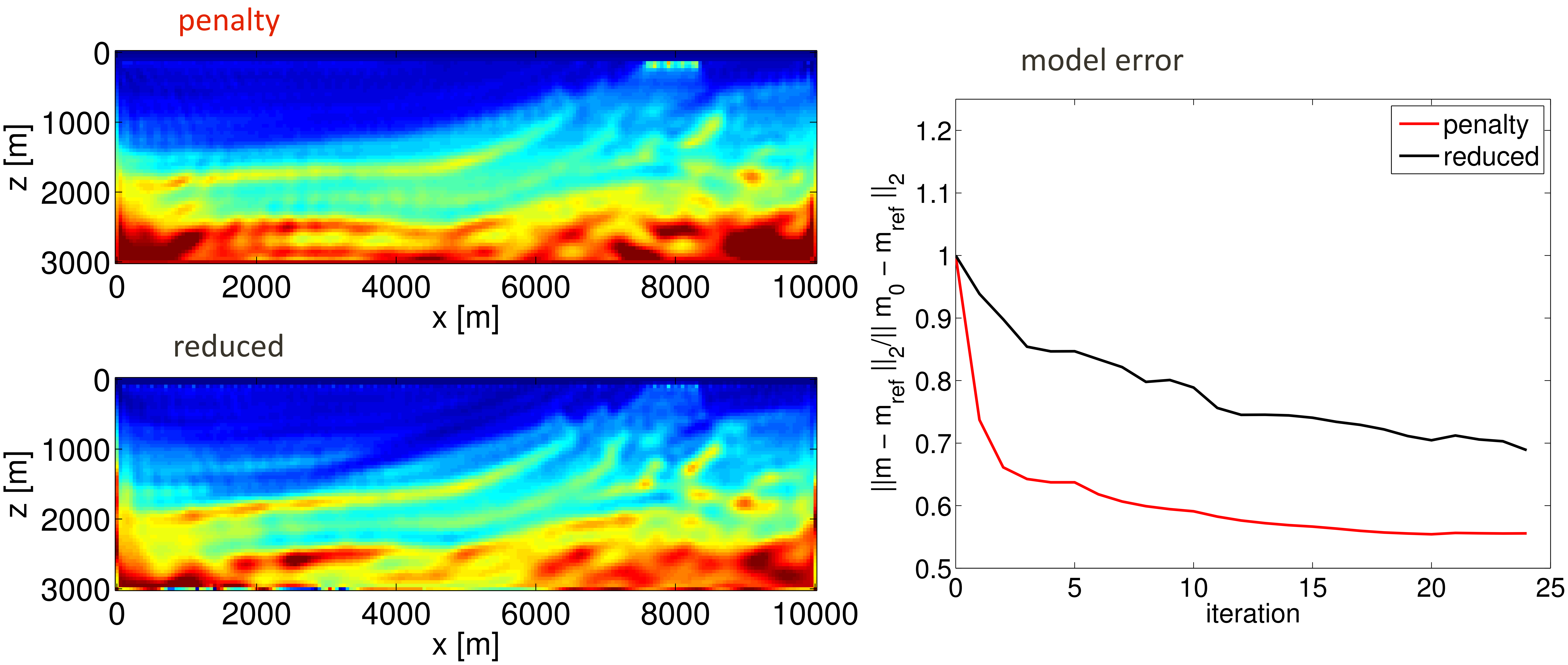


original + initial model

- 2D reflection configuration
- 5 point Helmholtz FD operator with abc
- Direct solver for PDEs (A\b)
- frequency continuation
- L-BFGS

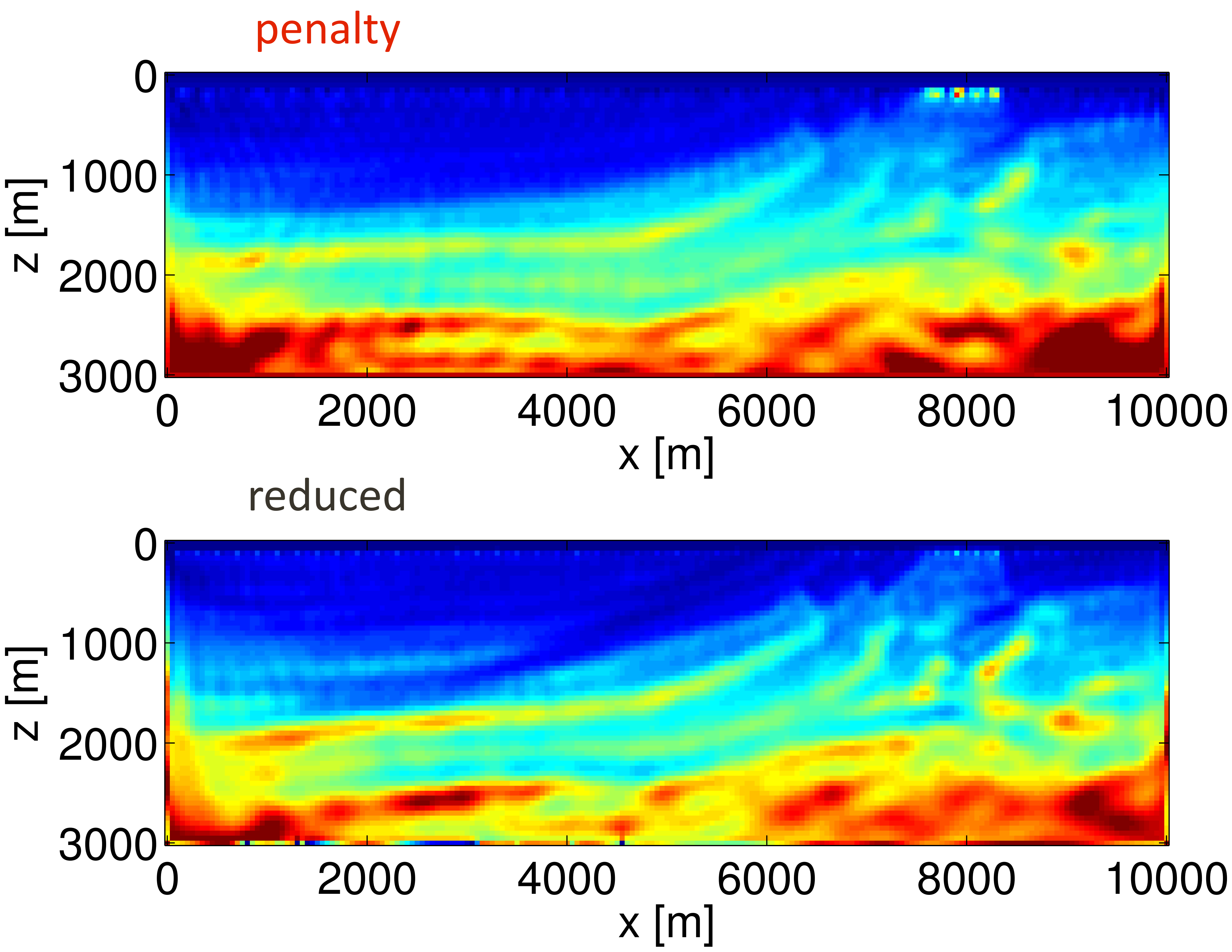
Waveform inversion

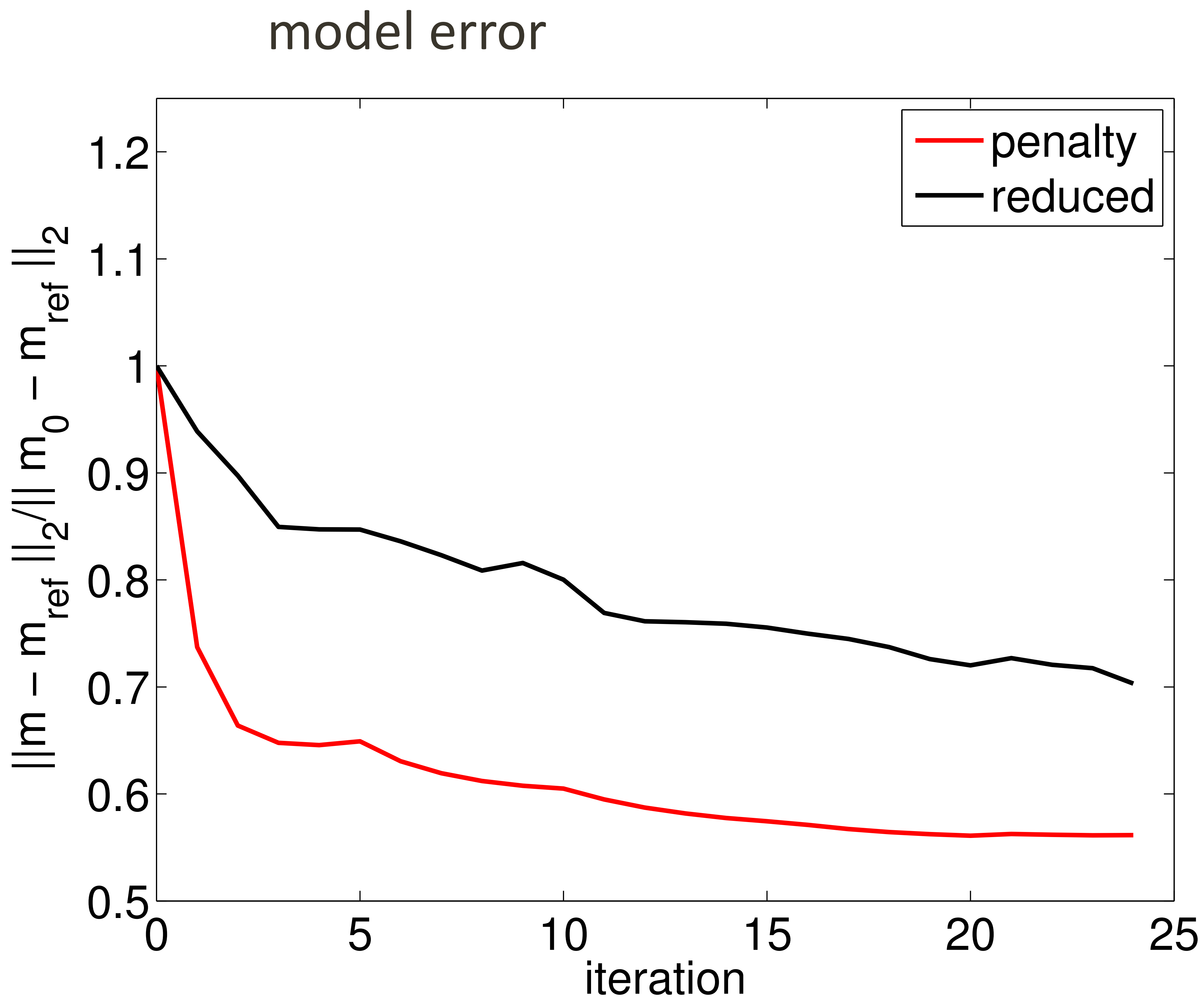
frequencies 1-5 Hz, no noise



Waveform inversion

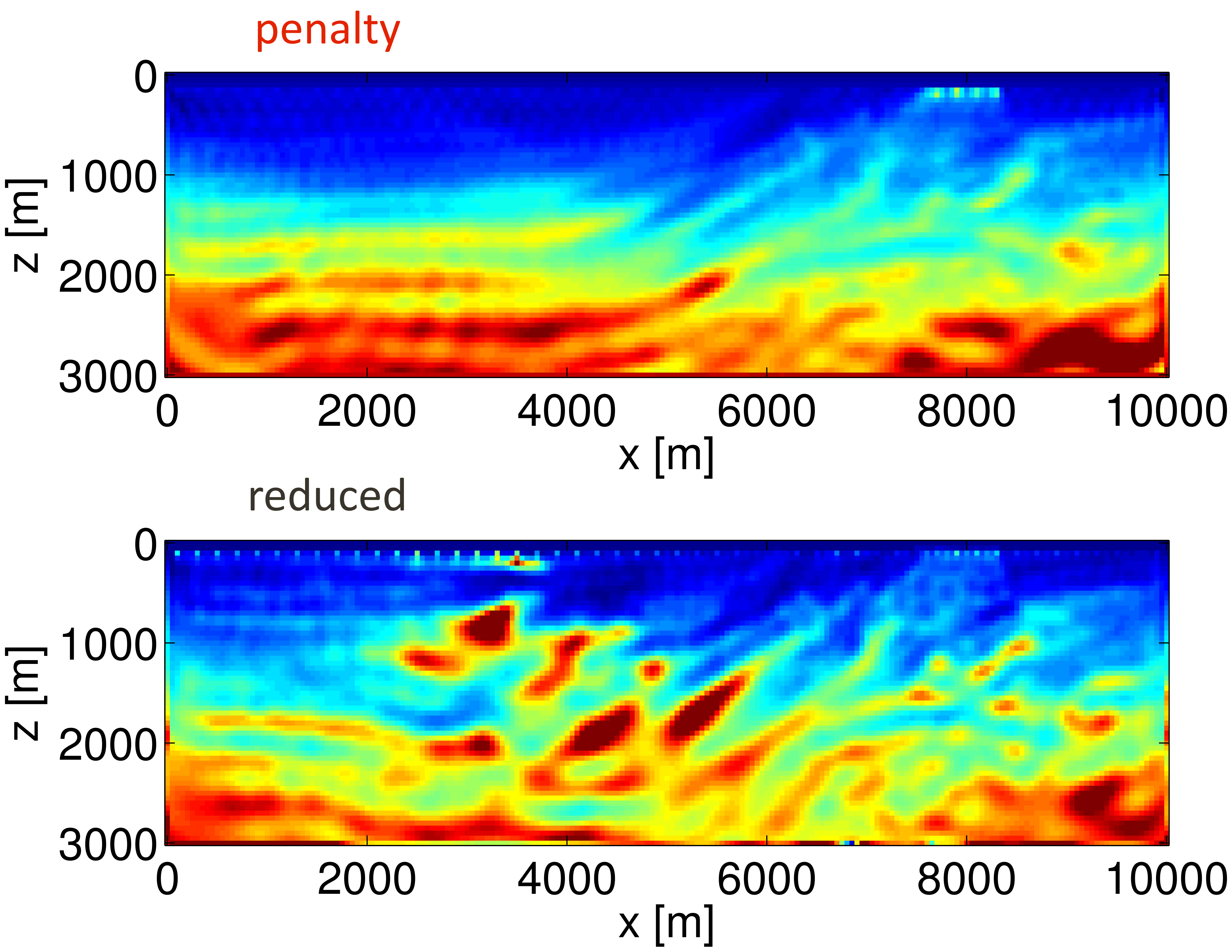
frequencies 1-5 Hz, 10 % Gaussian noise

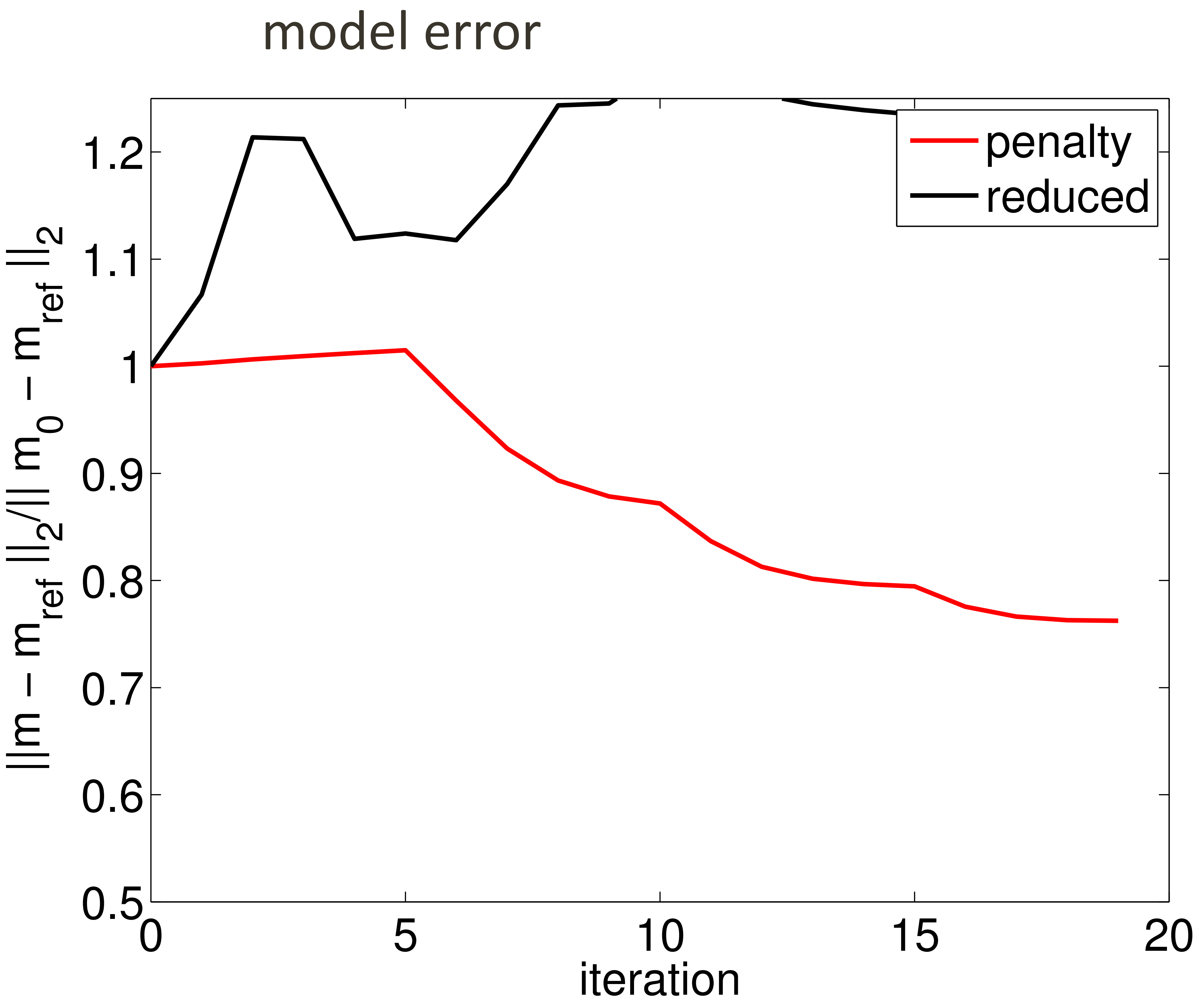


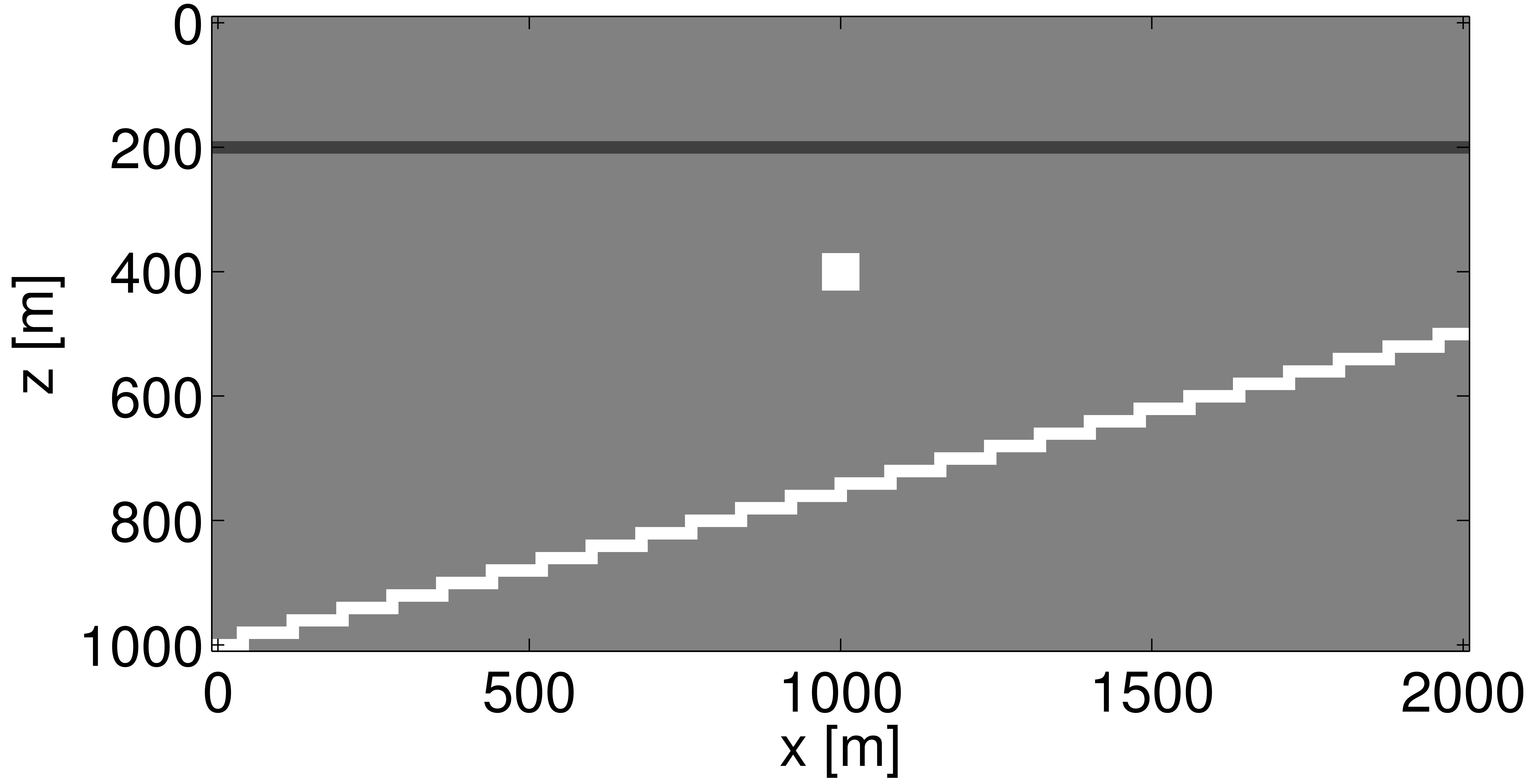


Waveform inversion

frequencies 2-5 Hz, 10 % Gaussian noise



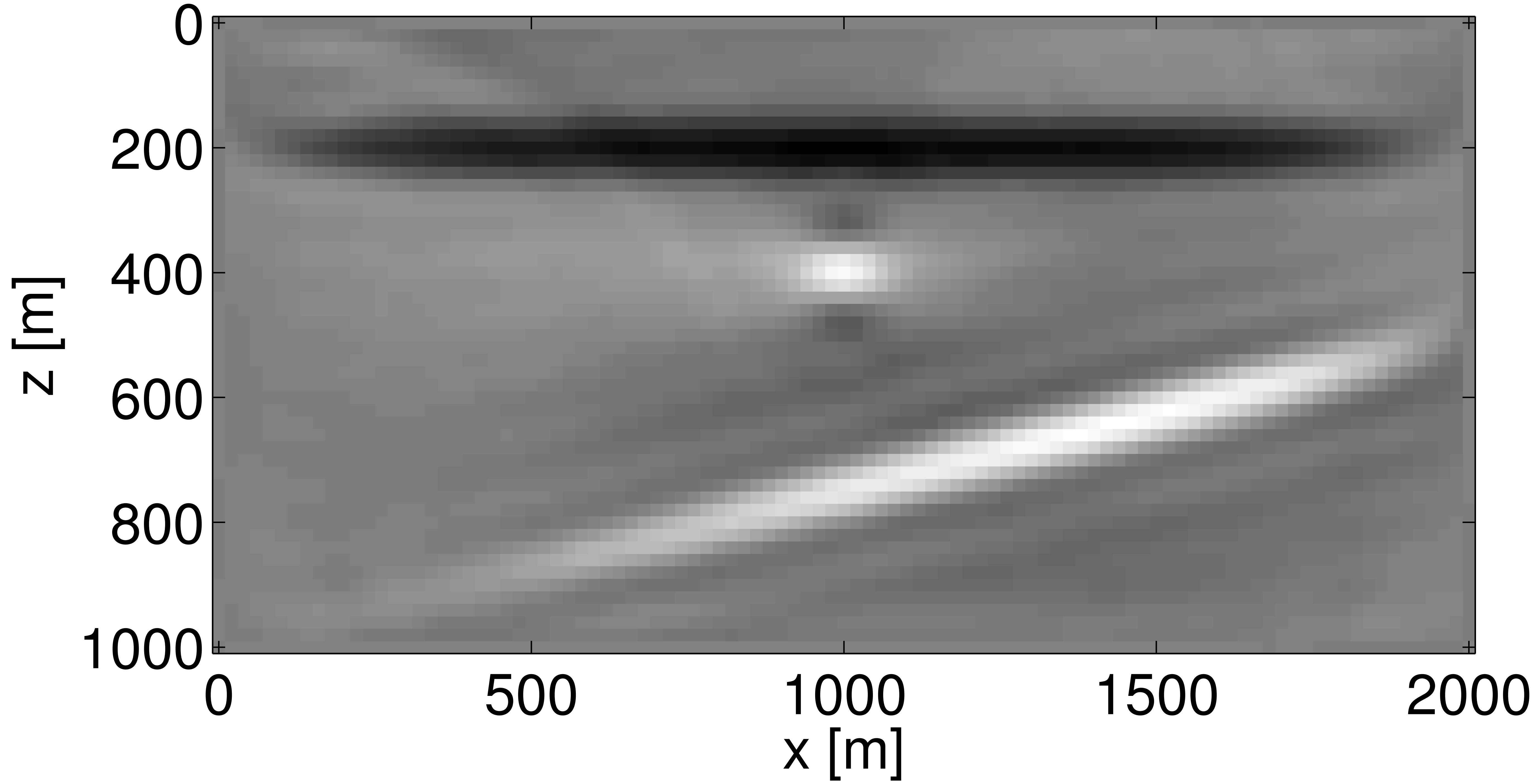




Original **image**

Imaging

The *gradient* of the reduced objective yields an *image* of the subsurface..



RTM-migration

Imaging

***Conventional* RTM:**

1. solve forward wave-equation
2. solve adjoint wave-equation
3. apply `imaging condition’

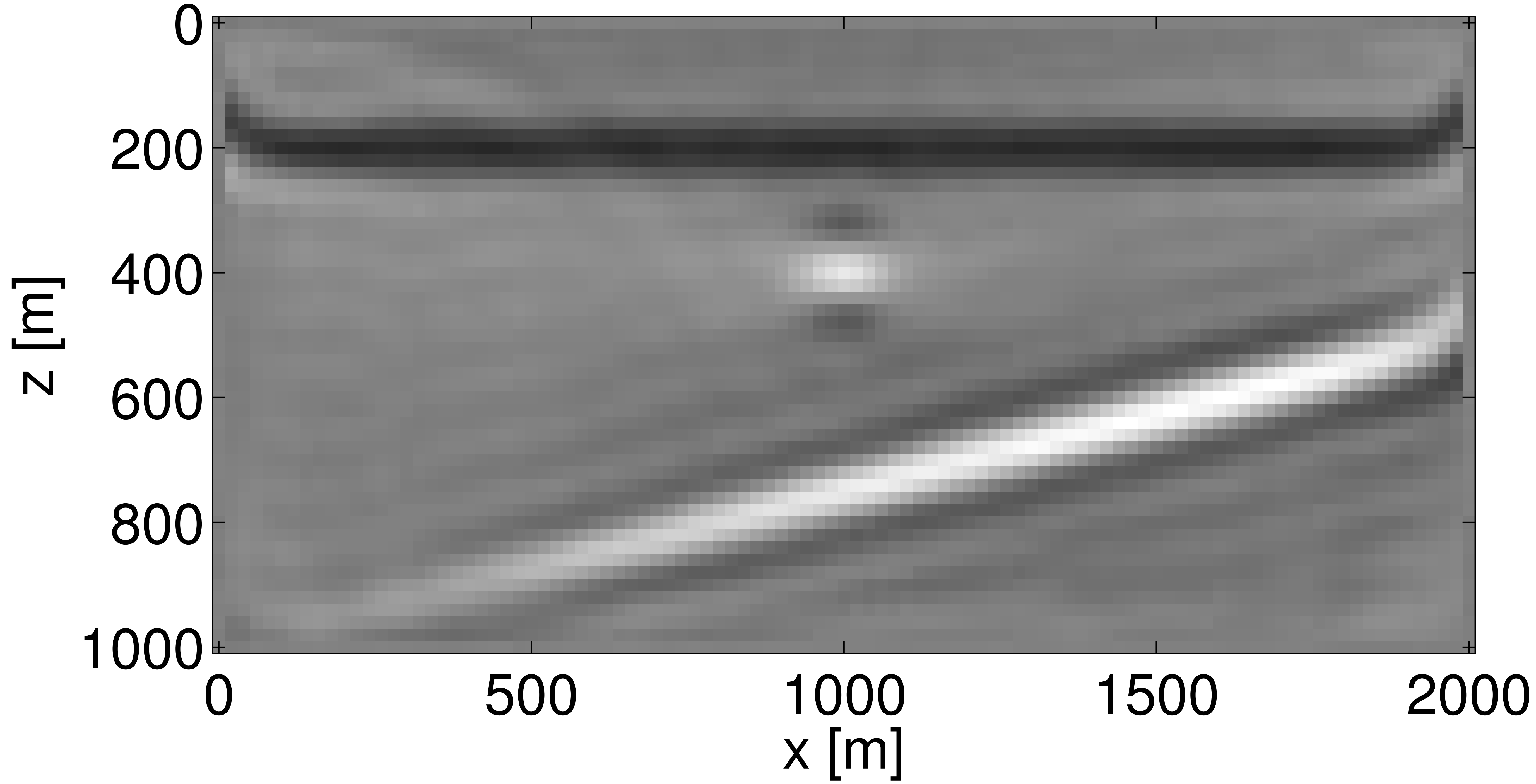


Image w/ penalty method

Imaging

***Penalty*-method reverse-time migration:**

1. solve overdetermined wave-equation
2. go for lunch
3. apply `imaging condition’

Conclusions

New inversion strategy for WE imaging & inversion;

- no explicit adjoint wavefields
- mitigates some of the non-linearity

Methods hinges on

- efficiently solving overdetermined WE
- choosing appropriate penalty parameter

Future work

- Solving the data-augmented WE
- Continuation in penalty parameter
- Extensive numerical experiments
- Theoretical analysis, connection to extended modelling
- application in time-domain
- multi-parameter inversion

Acknowledgements

Thank you for your attention !

https://www.slim.eos.ubc.ca/

SINBAD

NSERC CRSNG

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, WesternGeco, and Woodside.