Relax the *physics* & *expand* the *search* space - *FWI* via *W*avefield *R*econstruction *I*nversion

Felix J. Herrmann

SLIM
University of British Columbia

van Leeuwen, T and Herrmann, F J (2013). Mitigating local minima in full-waveform inversion by expanding the search space. Geophysical Journal International.
van Leeuwen, T and Herrmann, F J (2013). A penalty method for PDE-constrained optimization. Submitted for publication
van Leeuwen, T and Herrmann, F J (2013). US Provisional Patent Application No. 61/815,533. A Penalty Method for PDE-Constrained Optimization with Applications to Wave-Equation Based Seismic Inversion

Relax the *physics* & *expand* the *search* space - *FWI* via *W*avefield *R*econstruction *I*nversion

Ernier Esser, Tristan van Leeuwen*, and Bas Peters



* now @ CWI Amsterdam

SLIM
University of British Columbia

van Leeuwen, T and Herrmann, F J (2013). Mitigating local minima in full-waveform inversion by expanding the search space. Geophysical Journal International.
van Leeuwen, T and Herrmann, F J (2013). A penalty method for PDE-constrained optimization. Submitted for publication
van Leeuwen, T and Herrmann, F J (2013). US Provisional Patent Application No. 61/815,533. A Penalty Method for PDE-Constrained Optimization with Applications to Wave-Equation Based Seismic Inversion

A *new* take on FWI: *Wavefield Reconstruction Inversion*

Felix J. Herrmann

SLIM
University of British Columbia

A *new* take on FWI: *Wavefield Reconstruction Inversion*

Tristan van Leeuwen* and Bas Peters


* now @ CWI Amsterdam

SLIM
University of British Columbia

van Leeuwen, T and Herrmann, F J (2013). Mitigating local minima in full-waveform inversion by expanding the search space. Geophysical Journal International.
van Leeuwen, T and Herrmann, F J (2013). A penalty method for PDE-constrained optimization. Submitted for publication
van Leeuwen, T and Herrmann, F J (2013). US Provisional Patent Application No. 61/815,533. A Penalty Method for PDE-Constrained Optimization with Applications to Wave-Equation Based Seismic Inversion

Motivation

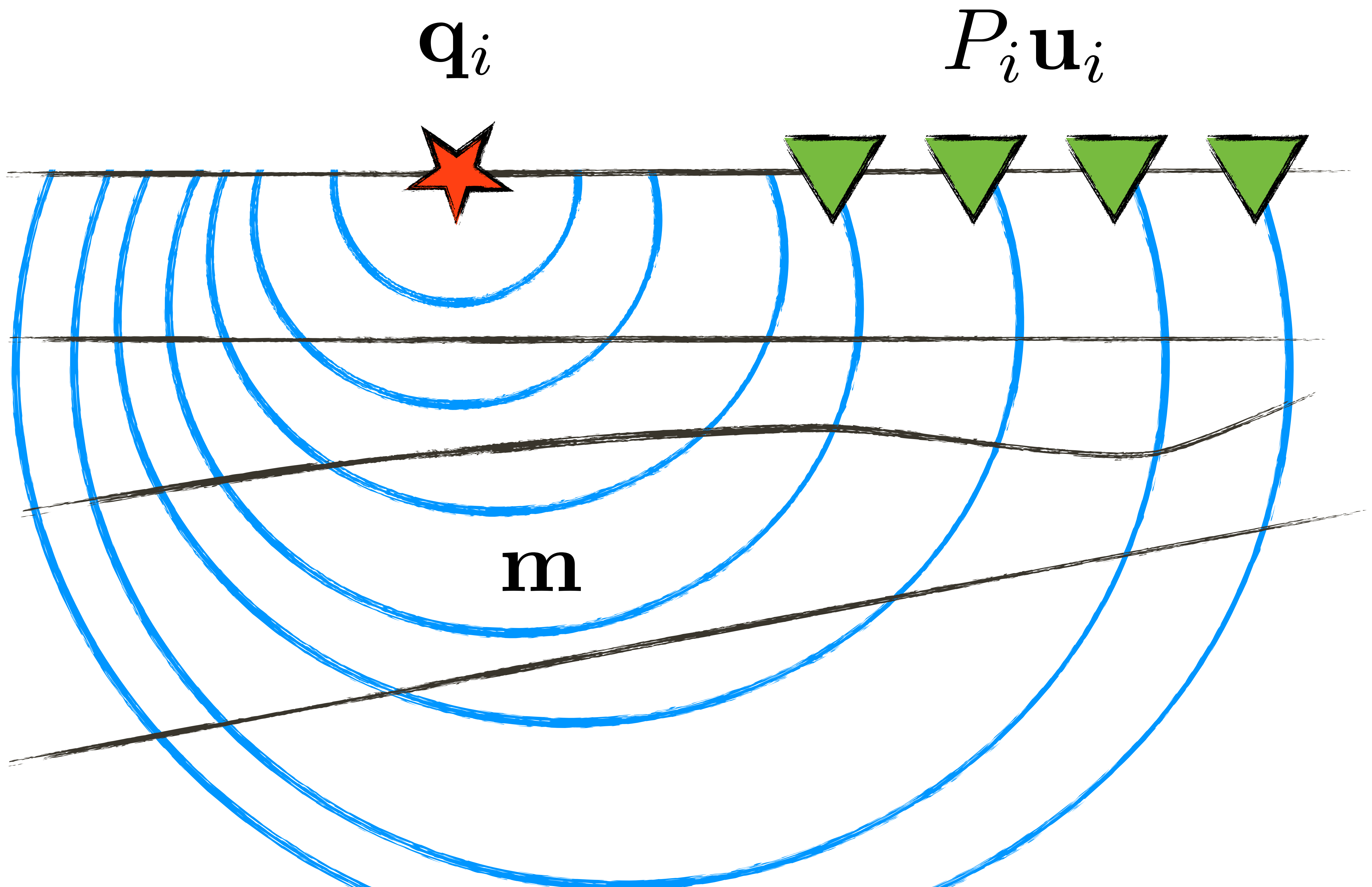
Full-waveform inversion is plagued with local minima

Derive an alternative formulation

- less prone to local minima
- computationally feasible
- relaxes the physics while staying solidly grounded

Waveform inversion

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



[Tarantola, ’84; Pratt, ’98; Haber, ’00; Plessix, ’06]

Waveform inversion

Adjoint-state/reduced-space methods:

- Optimize over earth models to minimize the misfit between observed and simulated data while solving the wave equation exactly for each earth model.

Full-space or all-at-once methods:

- Optimize over earth models & wavefields jointly to minimize the misfit between observed and simulated data subject to wavefields that are consistent with the wave equation.

SLIM

Waveform inversion

Both approaches assume *flawless* wave physics–i.e.,

“known” physics   “known” source

$$\underbrace{A(\mathbf{m})}_{\text{“known” physics}}\ \mathbf{u}_i = \underbrace{\mathbf{q}_i}_{\text{“known” source}}$$

“unknown” wavefield

- holds *exactly* for each source *i*
- *differ* on *insisting* wave equations to *hold* for *each* iteration
- *different* unknowns: **m** $\longleftrightarrow$ **m** & **u**

[Richter, '81]

Equation error approach

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

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

via

$$\min_{\mathbf{m}} \|A(\mathbf{m})P_i^{-1}\mathbf{d}_i - \mathbf{q}_i\|_2^2 \qquad \left(\text{cf. } \min_{\mathbf{m}} \|P_i A(\mathbf{m})^{-1}\mathbf{q}_i - \mathbf{d}_i\|_2^2\right)$$

The challenge is to reconstruct wavefields from partial measurements...

[van Leeuwen & FJH, 2013]

WRI – Wavefield Reconstruction Inversion

For $\mathbf{m}$ fixed, reconstruct wavefields by jointly fitting observed shots

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

and wave-equations

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

via least-squares solutions of the data-augmented wave-equation

$$\min_{\mathbf{u}_i} \left\| \begin{pmatrix} P_i \\ A(\mathbf{m}) \end{pmatrix} \mathbf{u}_i - \begin{pmatrix} \mathbf{d}_i \\ \mathbf{q}_i \end{pmatrix} \right\|_2^2$$

followed by fixing $\mathbf{u}_i$ and solving

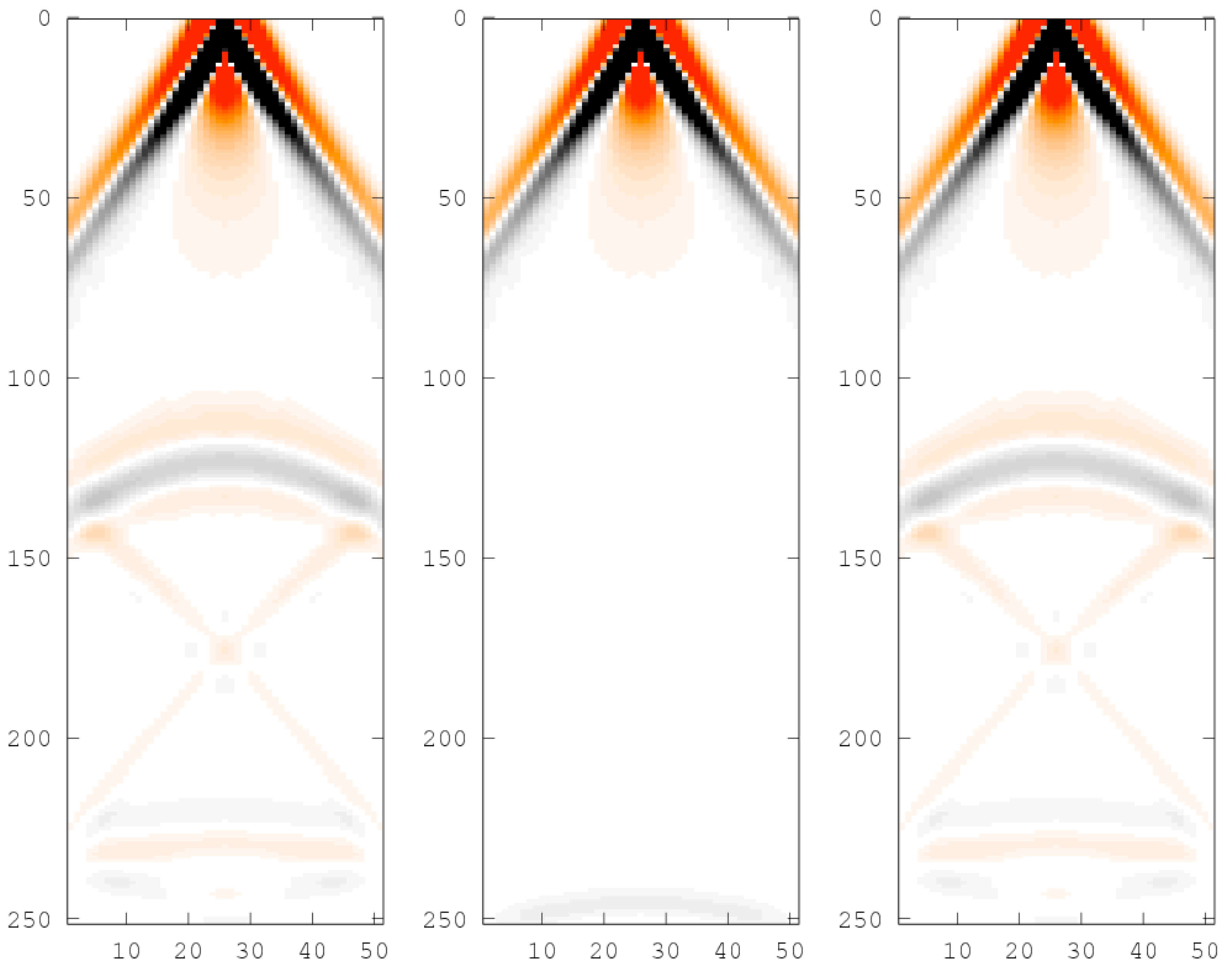
$$\min_{\mathbf{m}} \|A(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\|_2^2$$

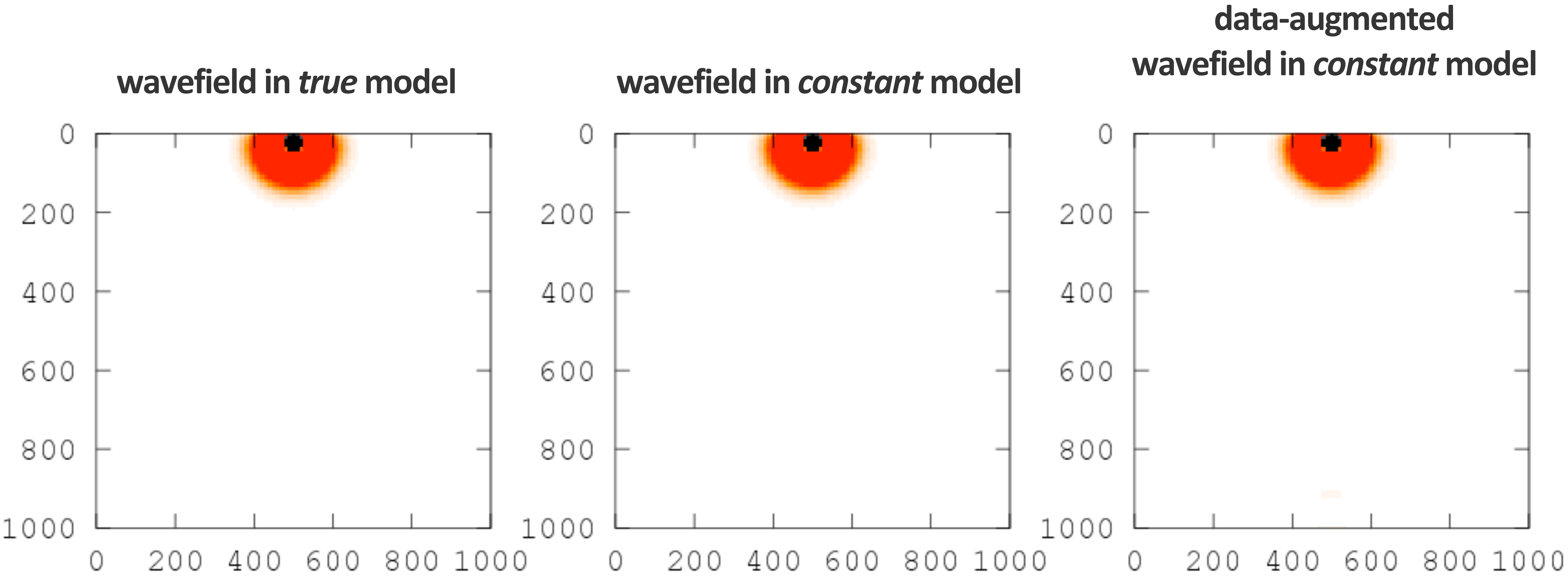
wave-equation X wavefield = source

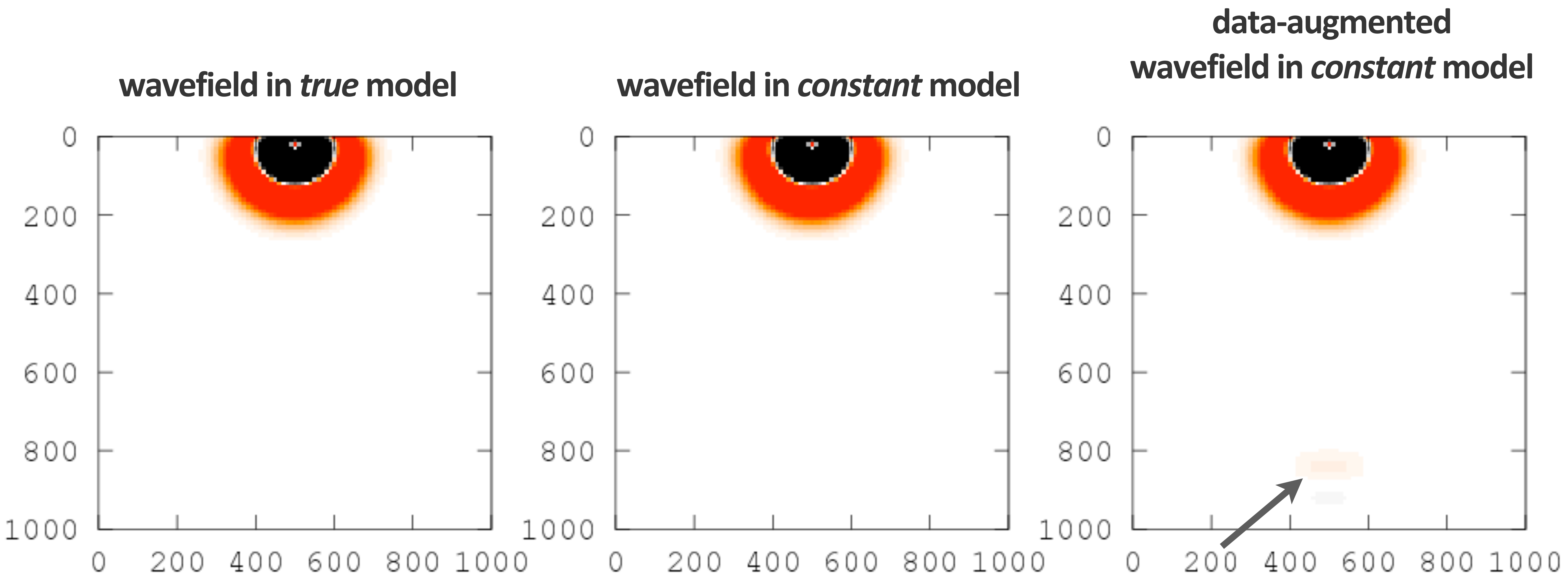
versus

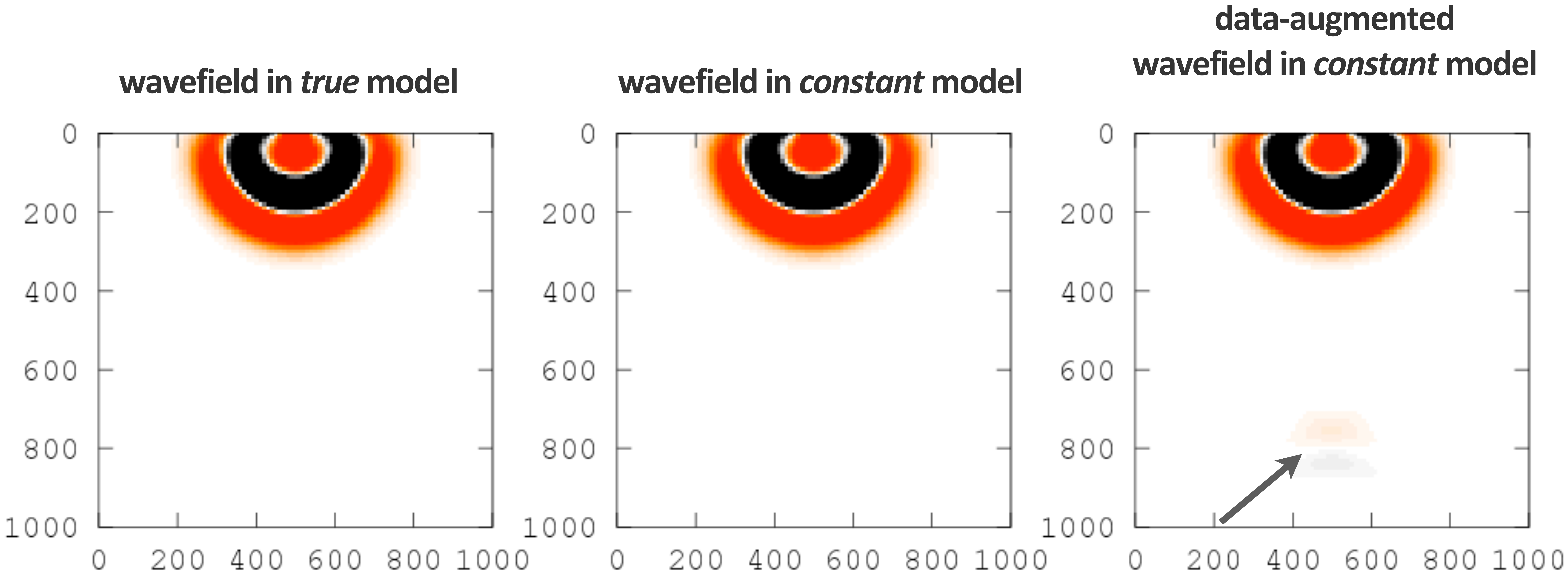
(wave-equation / sampling operator) X wavefield = (source / data)

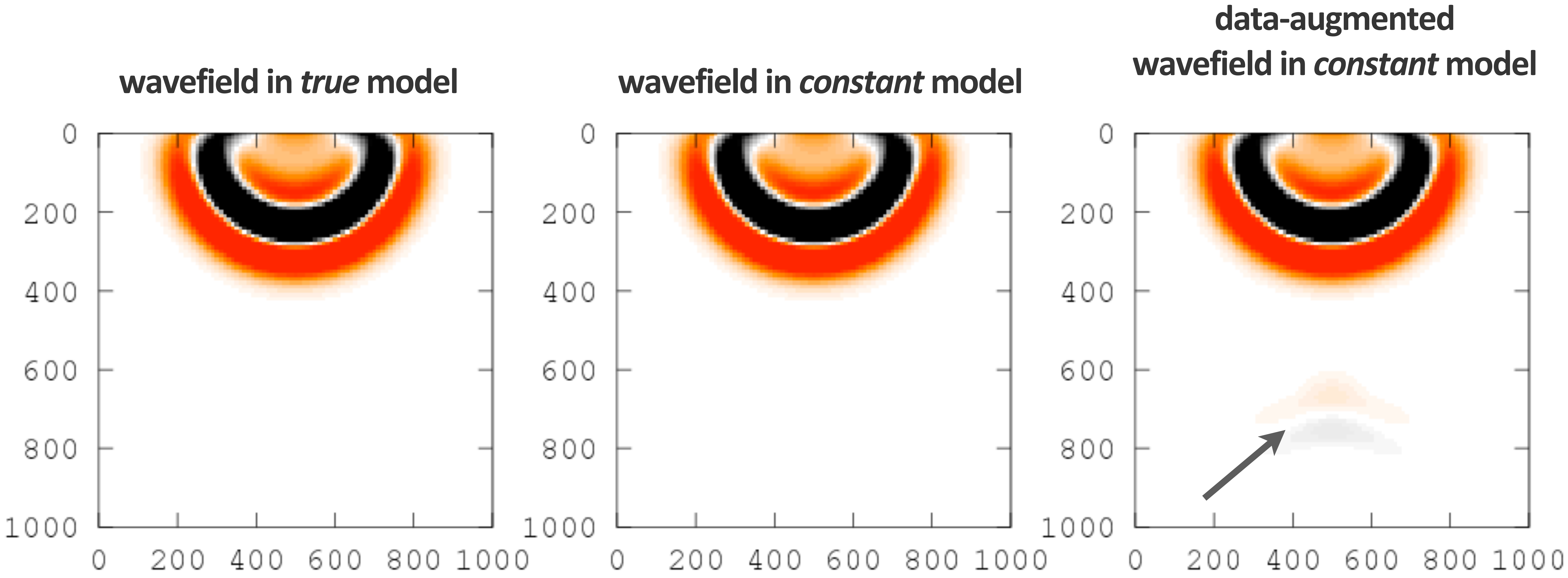
observed data | **initial data** | **data-augmented solution**

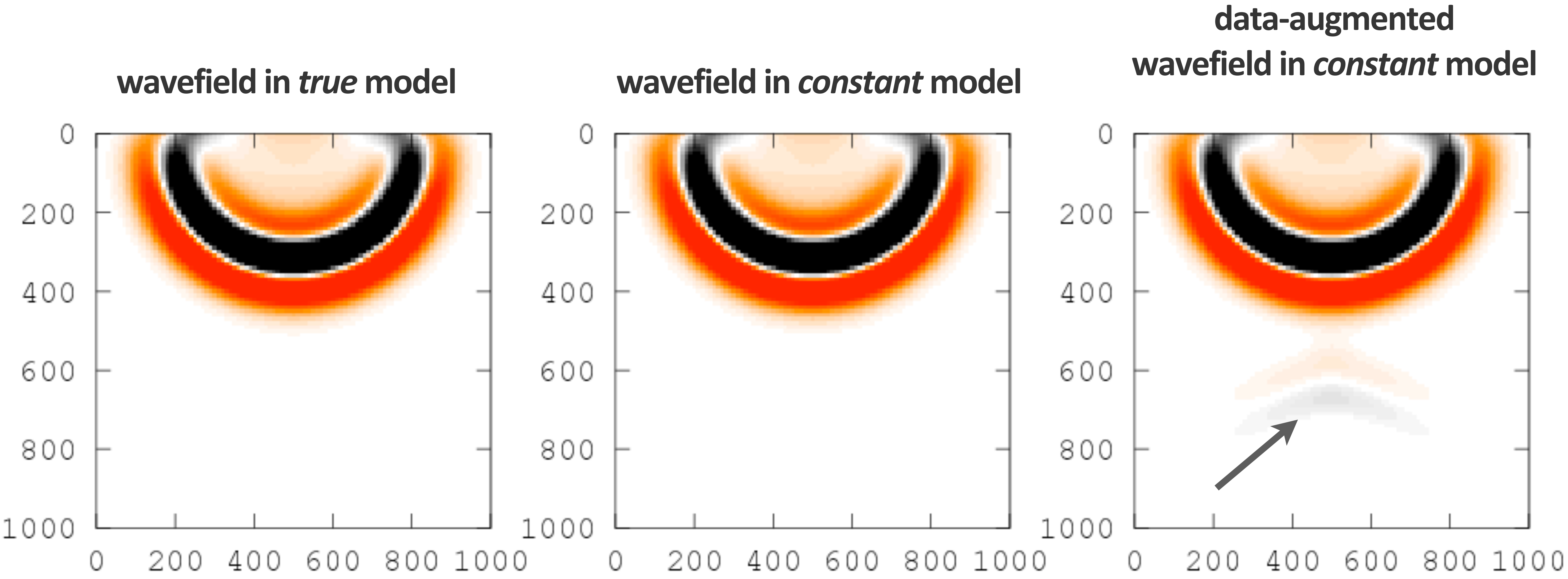


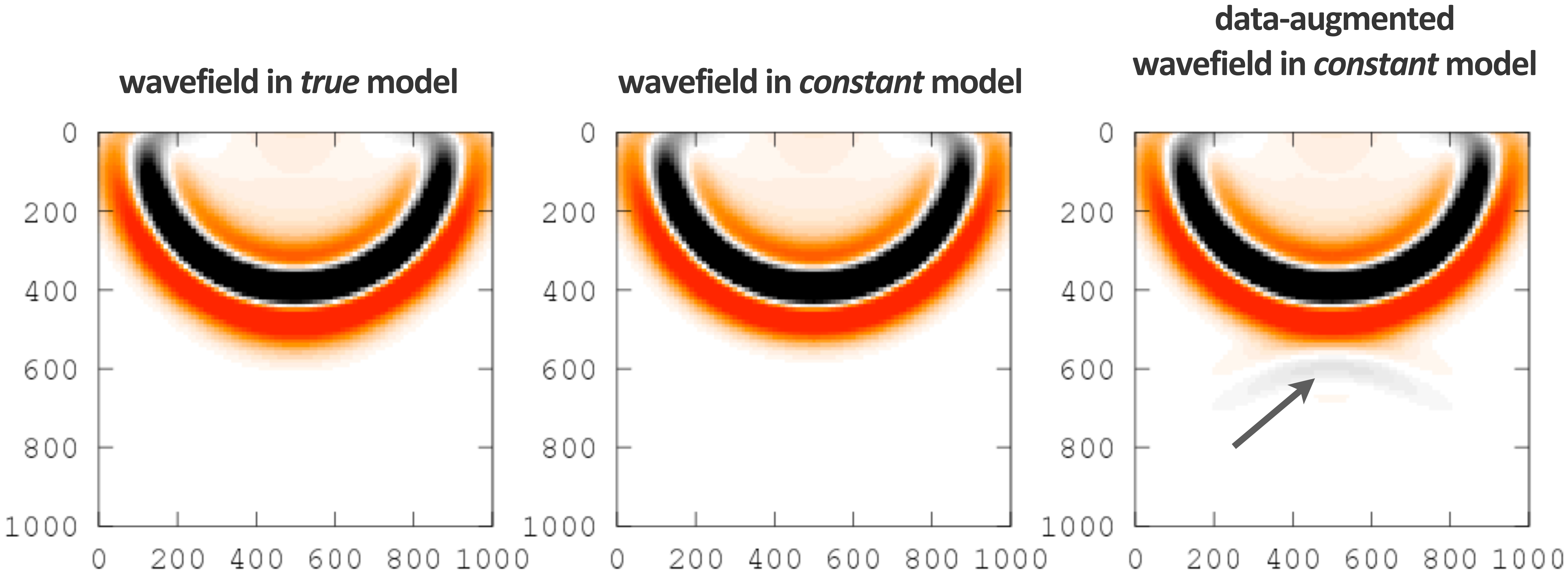

wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model


wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model
0
200
400
600
800
1000
0 200 400 600 800 1000

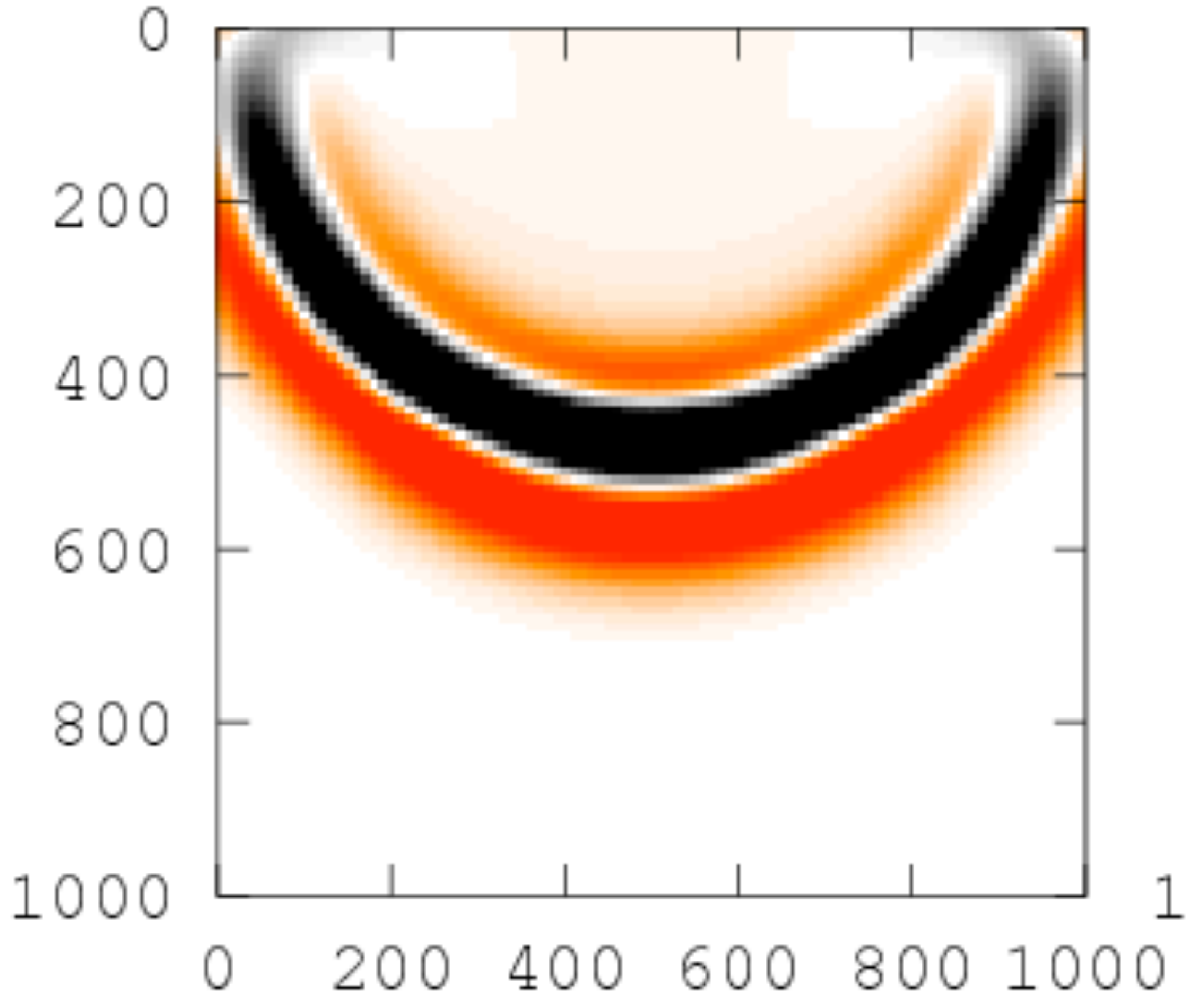

wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model
0
200
400
600
800
1000
0 200 400 600 800 1000


wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model
0
200
400
600
800
1000
0 200 400 600 800 1000

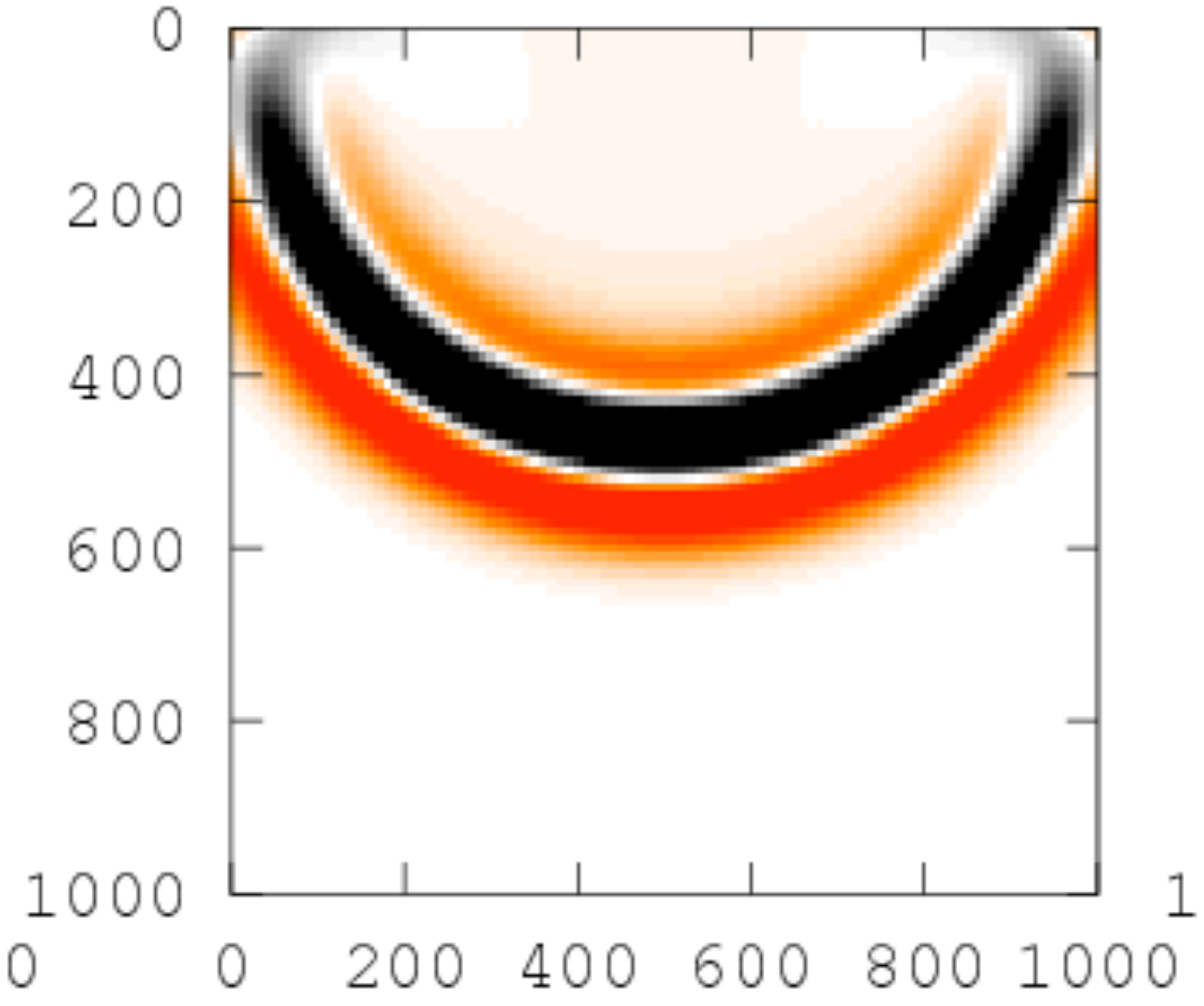

wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model
0
200
400
600
800
1000
0 200 400 600 800 1000


wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model
0
200
400
600
800
1000
0 200 400 600 800 1000

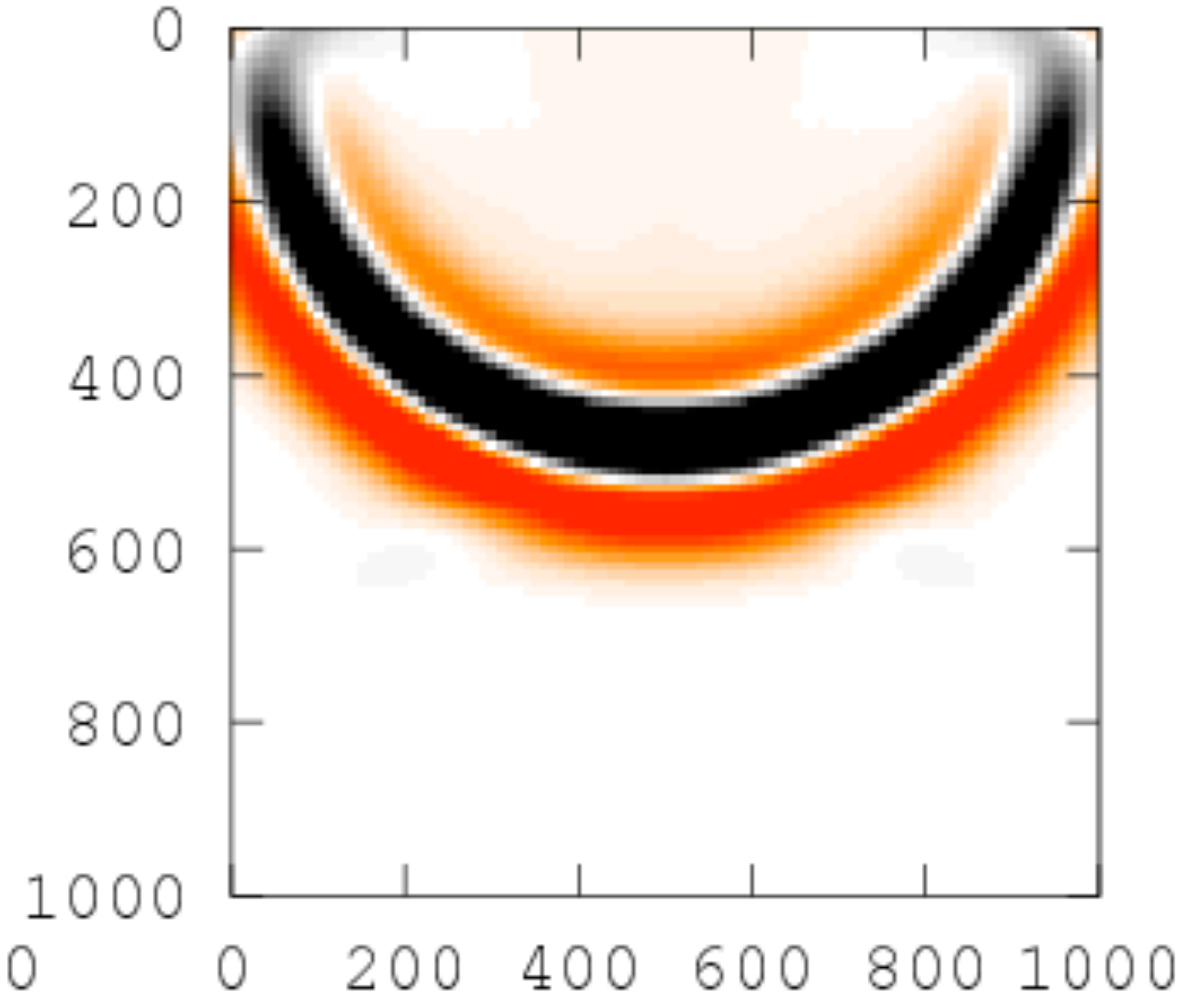
wavefield in *true* model

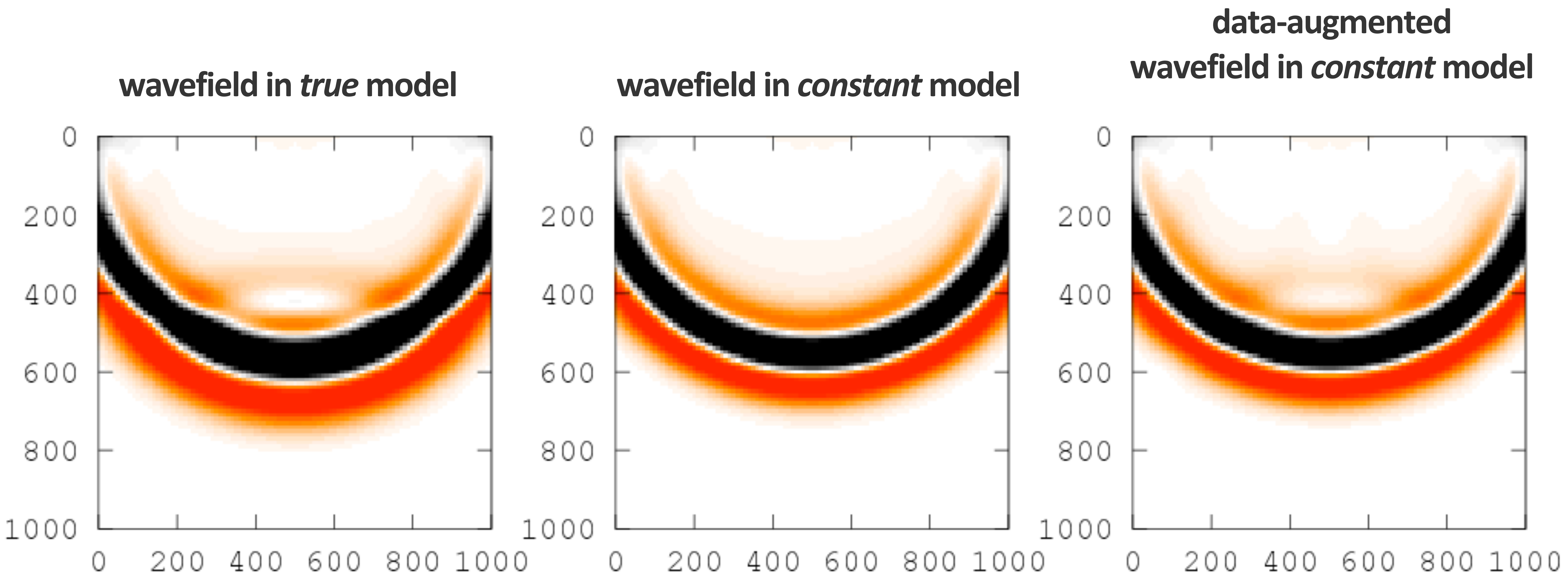


wavefield in *constant* model



data-augmented
wavefield in *constant* model

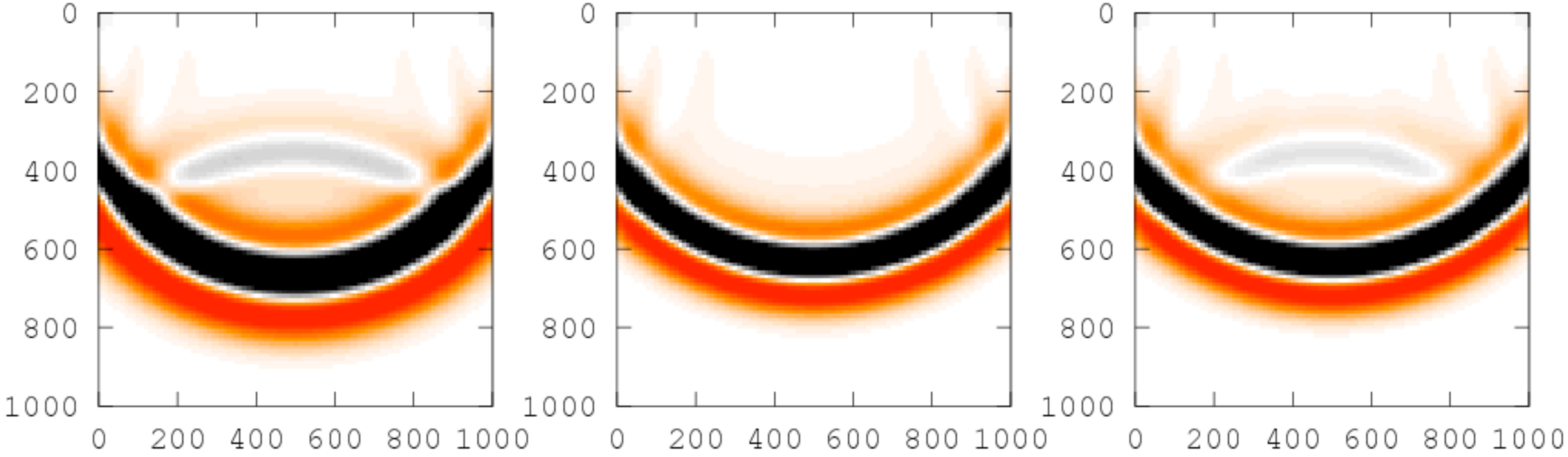



wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model
0
200
400
600
800
1000
0 200 400 600 800 1000

wavefield in *true* model

wavefield in *constant* model

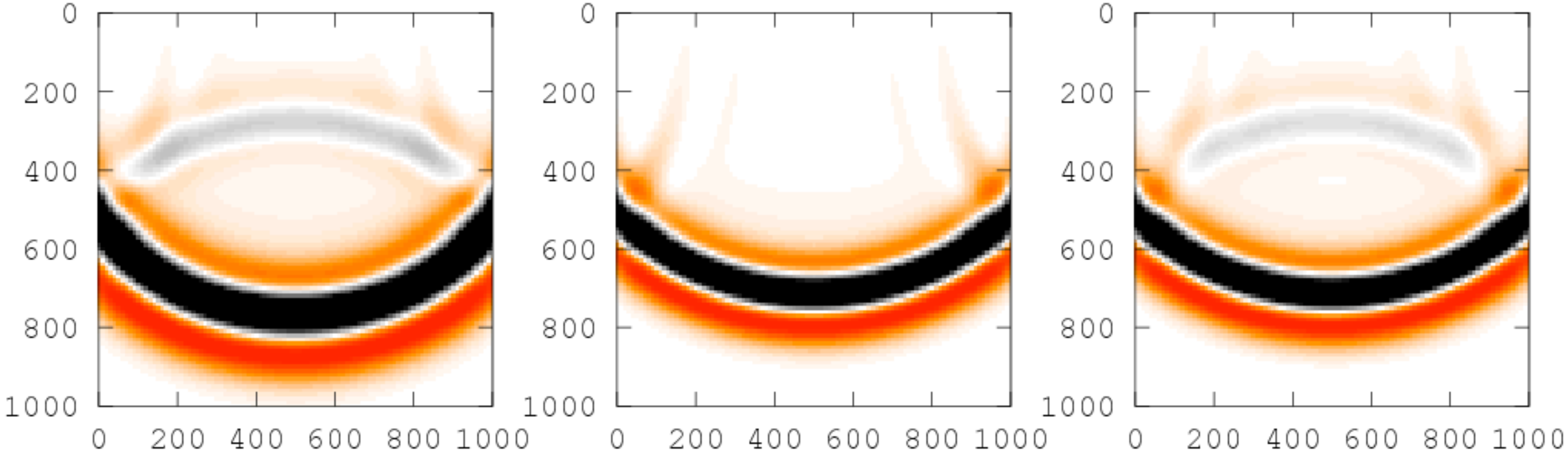
data-augmented
wavefield in *constant* model



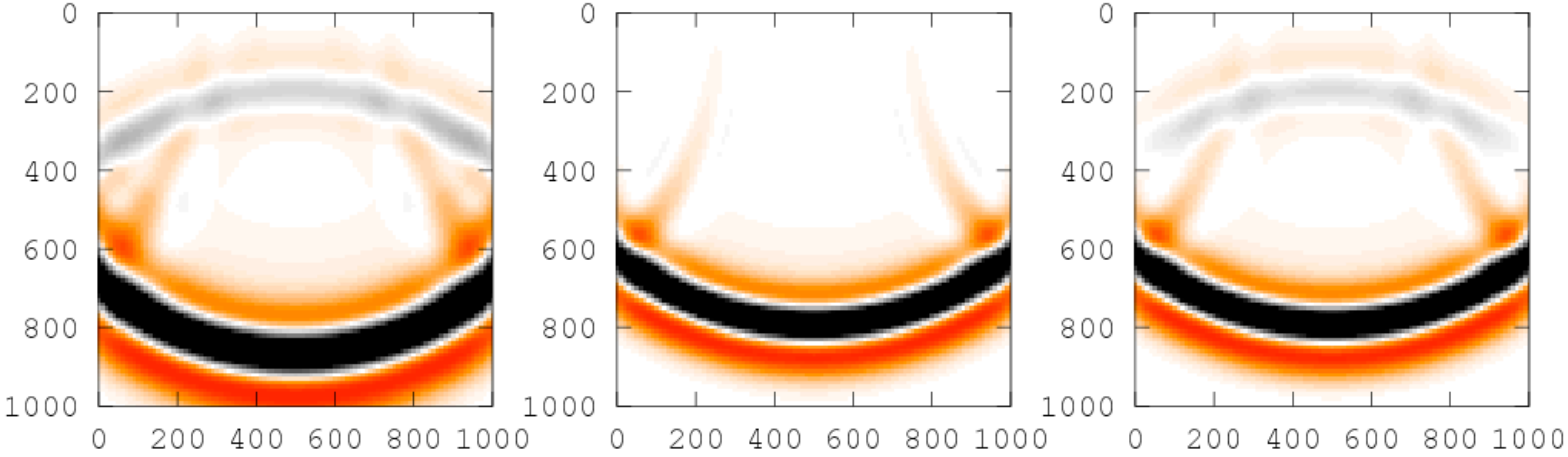
wavefield in *true* model

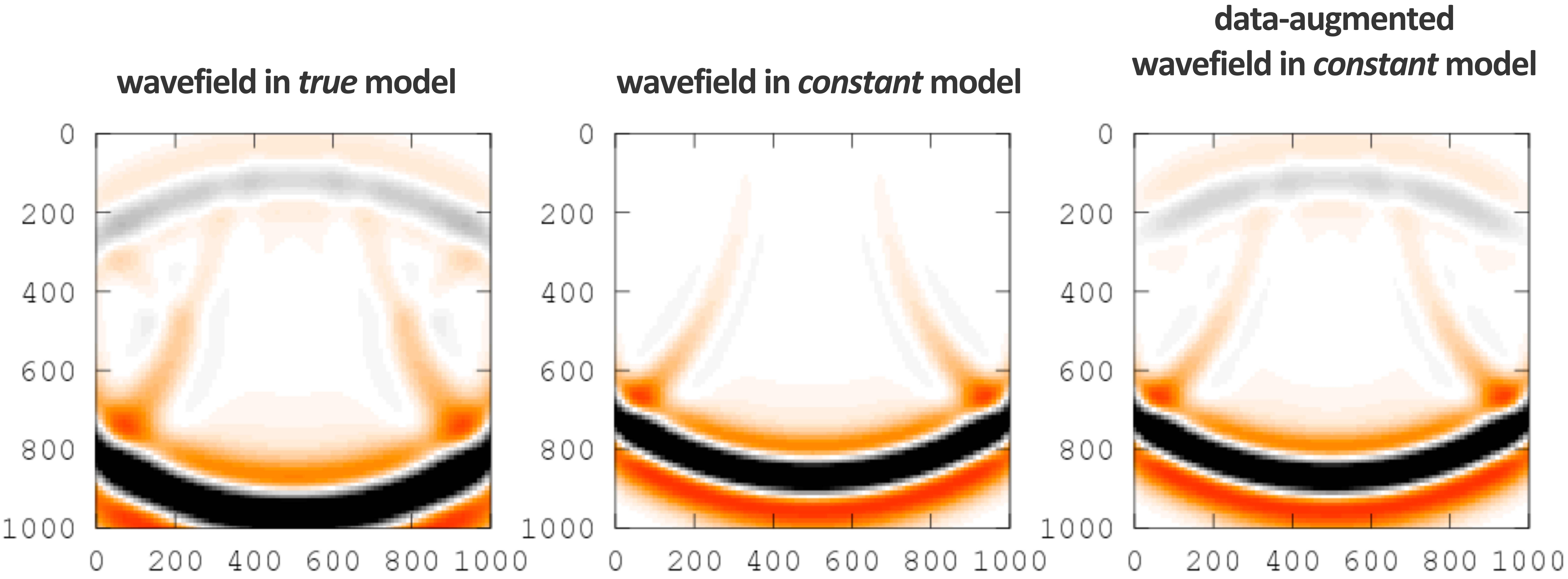
wavefield in *constant* model

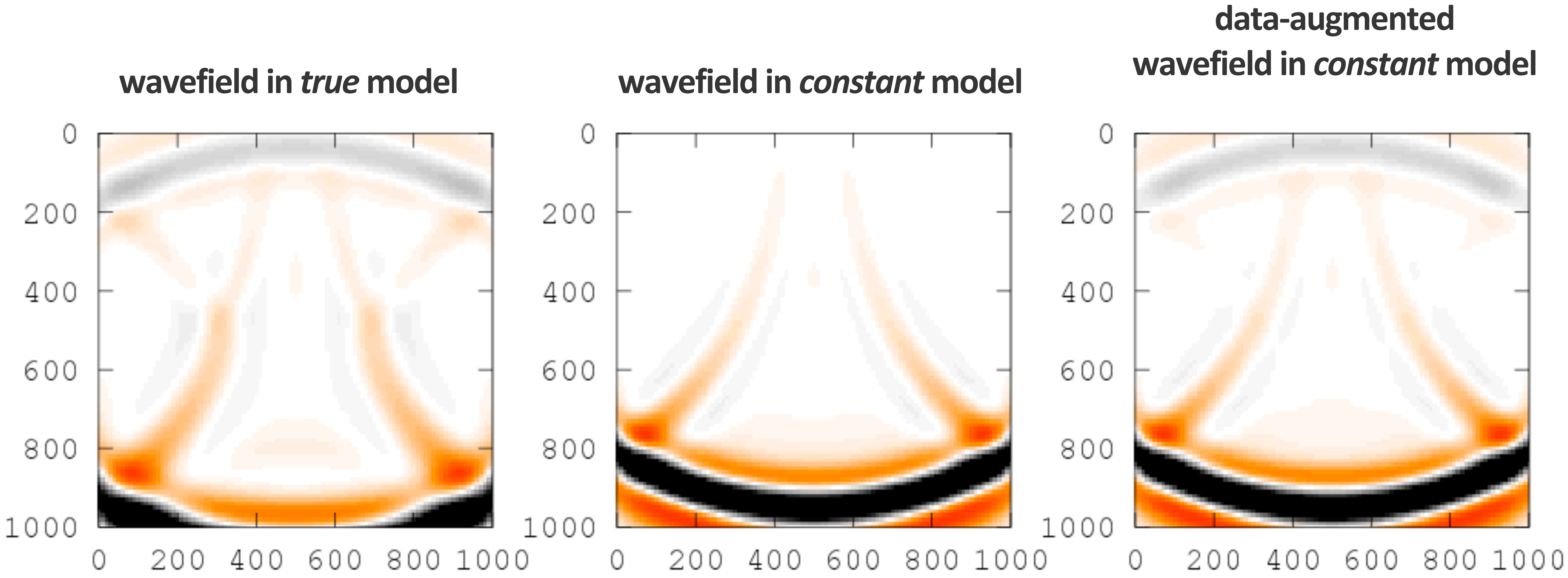
data-augmented
wavefield in *constant* model



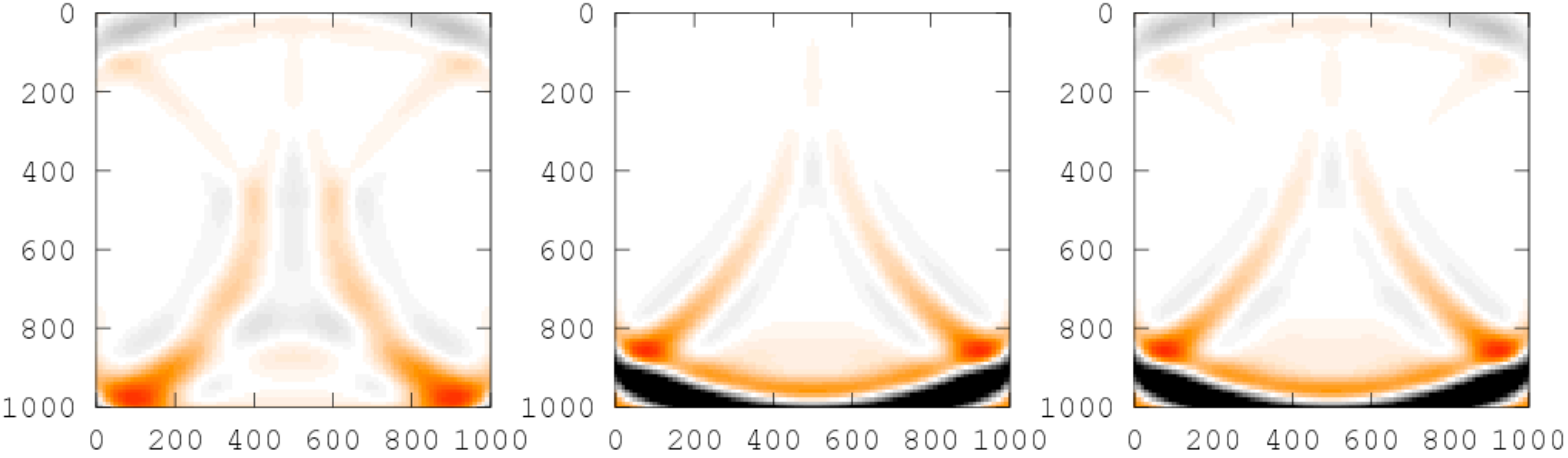
wavefield in *true* model | wavefield in *constant* model | data-augmented wavefield in *constant* model



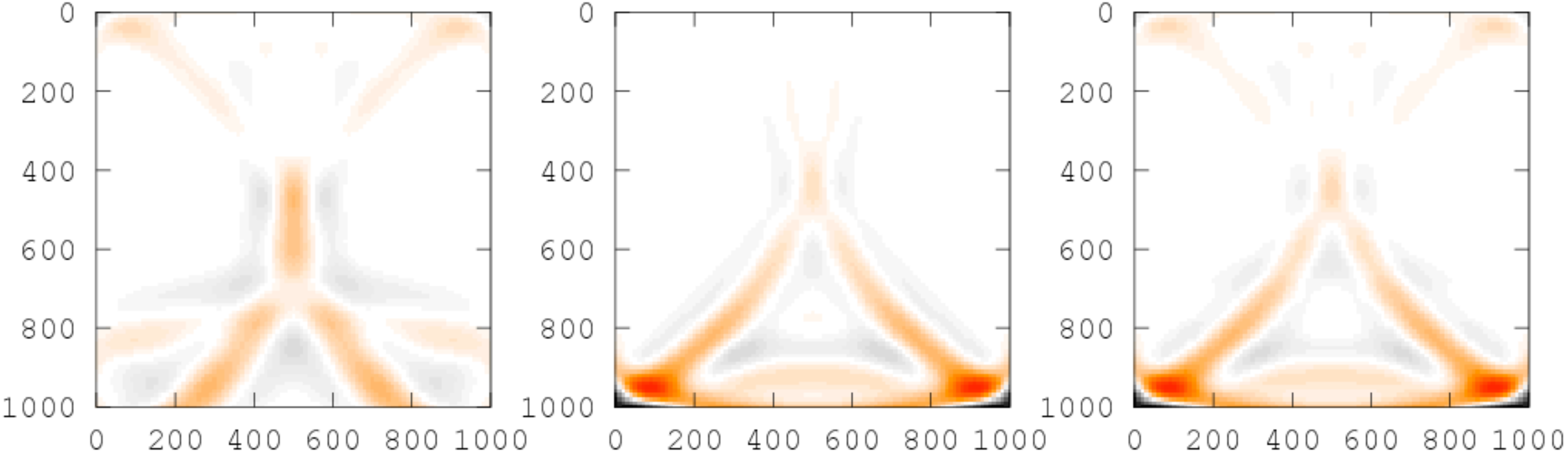

wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model
0
200
400
600
800
1000
0 200 400 600 800 1000


wavefield in *true* model
wavefield in *constant* model
data-augmented
wavefield in *constant* model
0
200
400
600
800
1000
0 200 400 600 800 1000

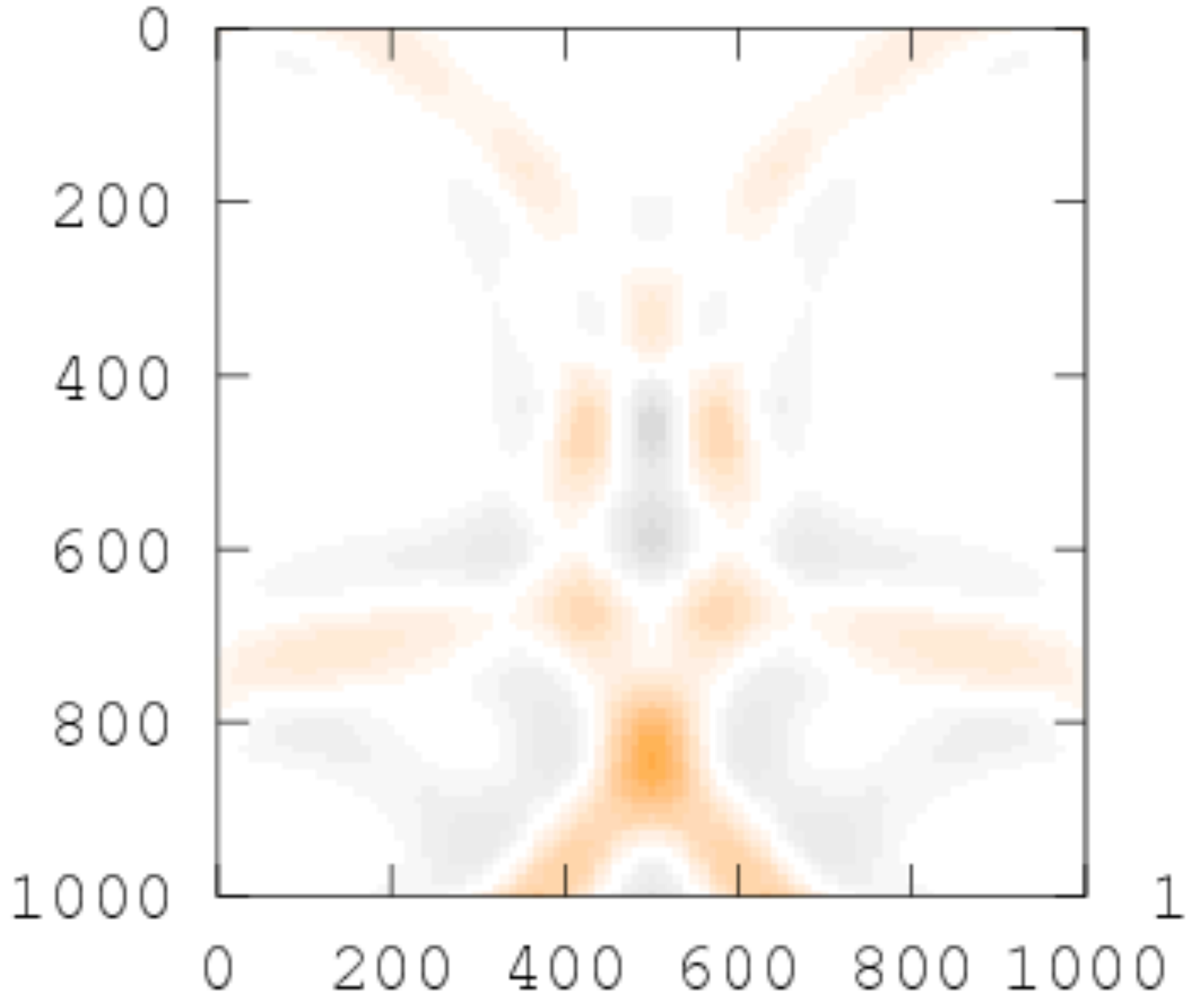
wavefield in *true* model | **wavefield in *constant* model** | **data-augmented wavefield in *constant* model**



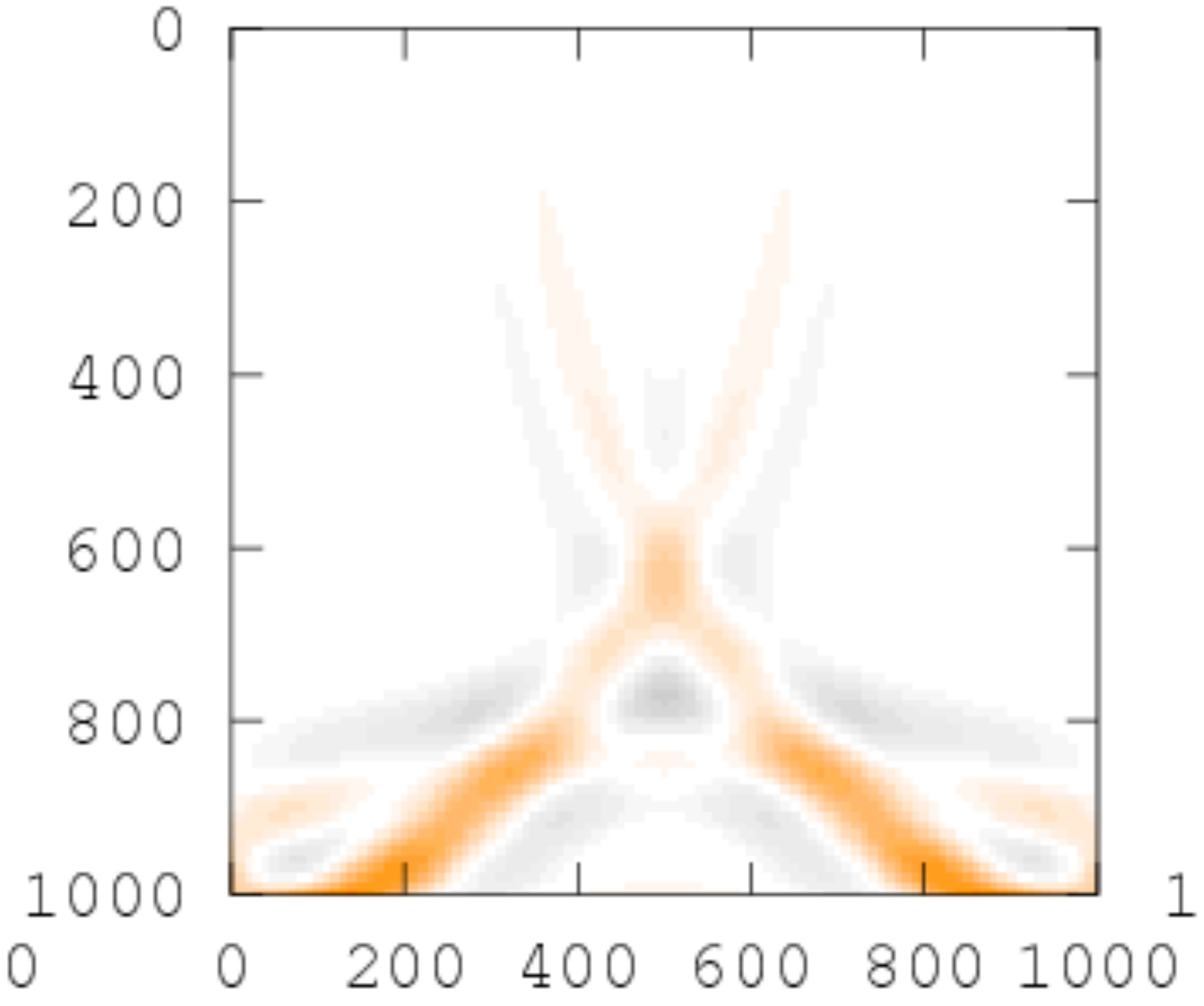
wavefield in *true* model | **wavefield in *constant* model** | **data-augmented wavefield in *constant* model**



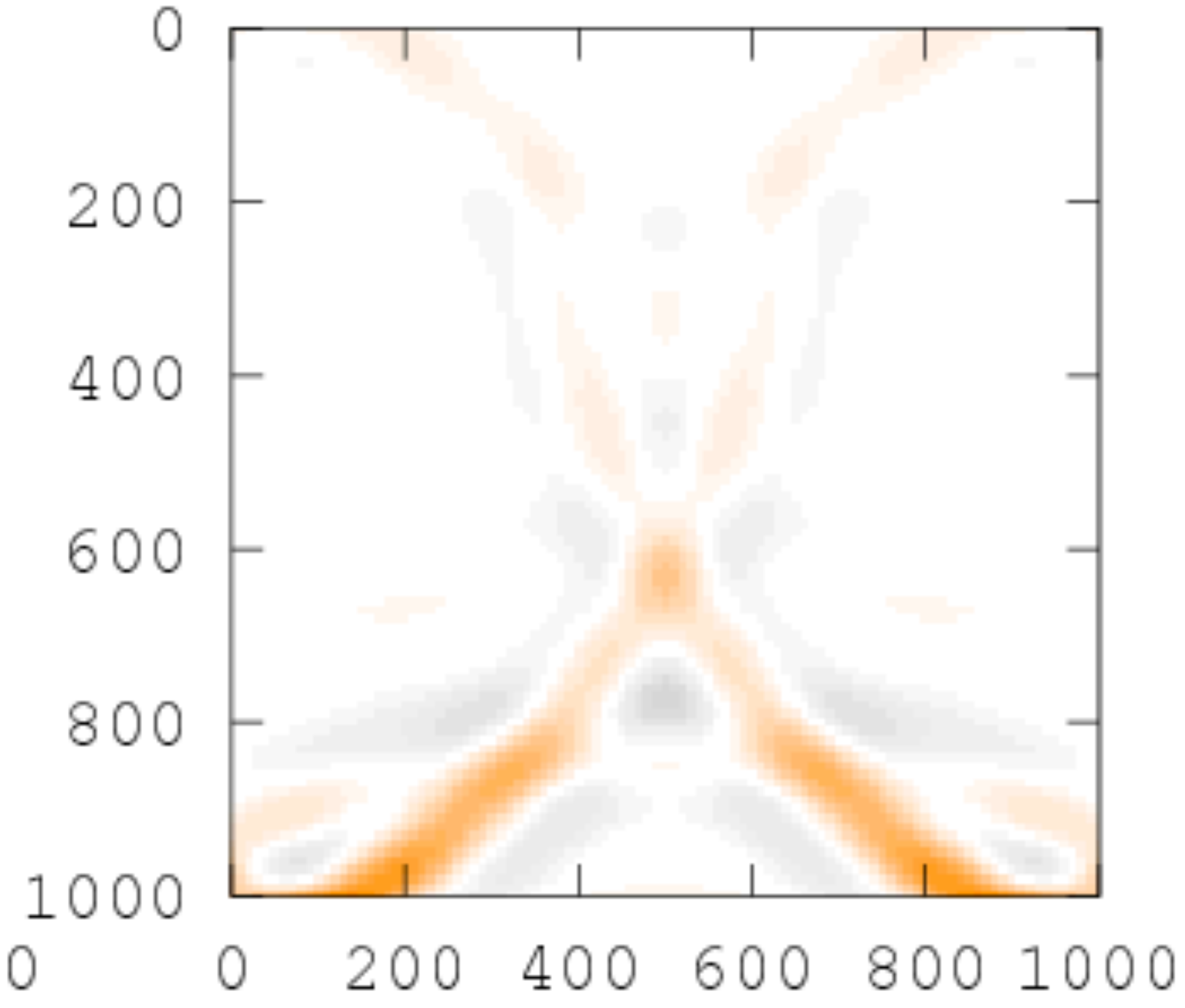
wavefield in *true* model



wavefield in *constant* model



data-augmented wavefield in *constant* model



[Heinkenschloss, ’98 , Haber, ’00]

PDE-constrained optimization
all-at-once full-space approach

simulated data — simulated wavefield

$$\min_{\mathbf{m},\mathbf{u}} \sum_{i=1}^{M} \|P_i \mathbf{u}_i - \mathbf{d}_i\|_2^2 \quad \text{s.t.} \quad A_i(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

observed data — Helmholtz equation — source

- avoids having to solve the PDE explicitly
- sparse (GN) Hessian
- requires storing all variables **(m,u)**
- does **not** scale to industry-scale seismic problems

[Tarantola ’84; Pratt, ’98; Plessix, ’06]

SLIM

Adjoint-state/reduced-space formulation

Elimination of the constraint leads for all sources to

$$\min_{\mathbf{m}} \phi_{\mathrm{red}}(\mathbf{m}) = \sum_{i=1}^{M} \|P_i A_i(\mathbf{m})^{-1}\mathbf{q}_i - \mathbf{d}_i\|_2^2$$

- no need to store all wavefields (block-elimination)
- suitable for black-box optimization (e.g., l-BFGS)
- need to solve forward & adjoint PDEs
- very non-linear in earth model (m)
- dense (GN) Hessian, involves additional PDE solves

SLIM

WRI – penalty formulation

Instead of eliminating, we add constraints as penalties–i.e.,

$$\min_{\mathbf{m},\mathbf{u}} \phi_\lambda(\mathbf{m},\mathbf{u}) = \sum_{i=1}^{M} \|P\mathbf{u}_i - \mathbf{d}_i\|_2^2 + \lambda^2 ||A_i(\mathbf{m})\mathbf{u}_i - \mathbf{q}_i\|_2^2$$

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

SLIM

Variable projection

Solve data-augmented wave equation for each source

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

Define reduced objective

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

[van Leeuwen & FJH, '13]

Gradient

Wavefield eliminated–i.e., $\nabla_{\bar{\mathbf{u}}}\phi_\lambda(\mathbf{m},\bar{\mathbf{u}}) = 0$ by solving

$$\left(\lambda^2 A^*(\mathbf{m})A(\mathbf{m}) + P^*P\right)\mathbf{u} = \lambda^2 A^*(\mathbf{m})\mathbf{q} + P^*\mathbf{d}$$

yielding the gradient

Jacobian of $A(\mathbf{m})\bar{\mathbf{u}}_\lambda$

$$\nabla\phi_\lambda(\mathbf{m}) = G(\mathbf{m},\bar{\mathbf{u}}_\lambda)^*\bar{\mathbf{v}}_\lambda$$

with

$$\bar{\mathbf{v}}_\lambda = \lambda^2(A(\mathbf{m})\bar{\mathbf{u}}_\lambda - \mathbf{q})$$ (PDE residual)

[van Leeuwen & FJH, ’13]

SLIM

Wavefield Reconstruction Inversion

WRI method

for each source i

$$\text{solve} \begin{pmatrix} P_i \\ \lambda A_i(\mathbf{m}) \end{pmatrix} \mathbf{u}_i \approx \begin{pmatrix} \mathbf{d}_i \\ \lambda \mathbf{q}_i \end{pmatrix}$$

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

end

Conventional method

for each source i

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

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

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

end

Wavefield Reconstruction Inversion

Penalty method

for each source i

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

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

end

correlation augmented wavefield & PDE residual

Conventional method

for each source i

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

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

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

end

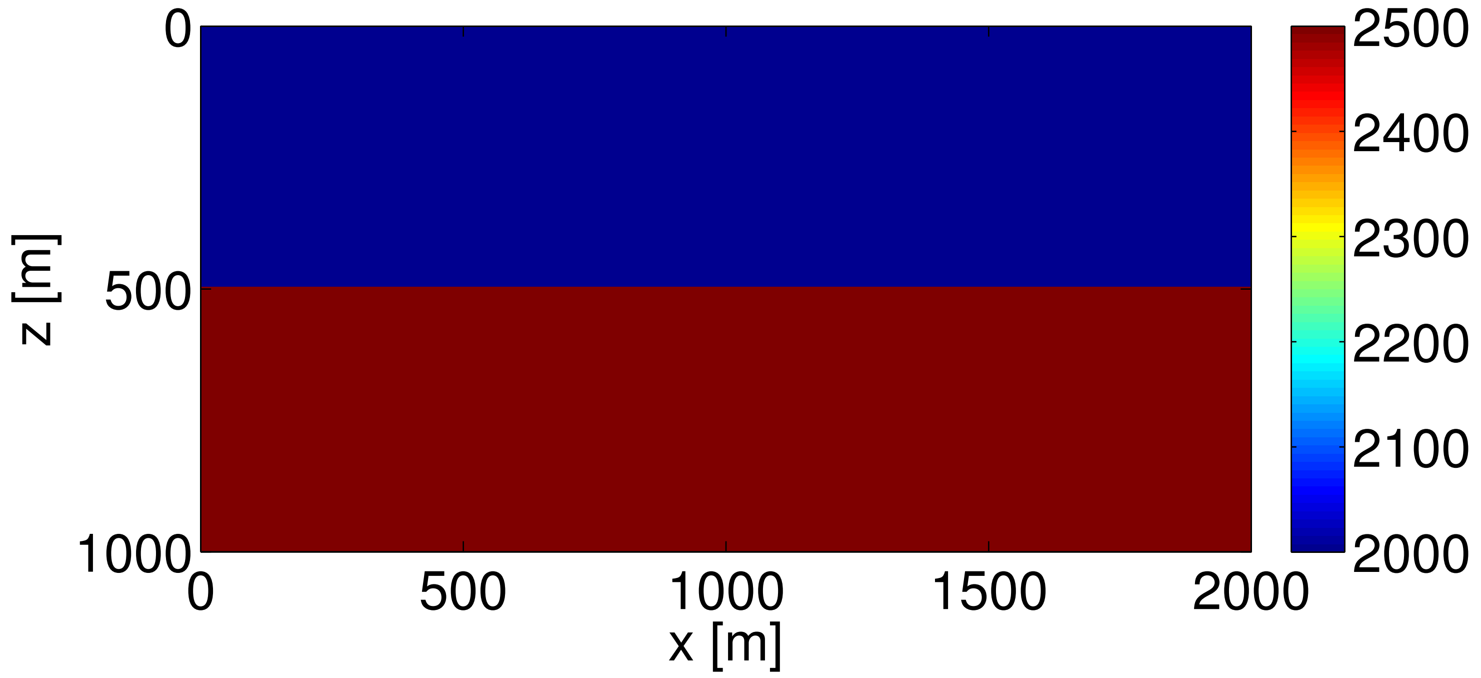
correlation wavefield & data residual

Wavefield Reconstruction Inversion

- **no** need to store all the fields (**u**)
- **no** adjoint solves
- sparse approximation of GN Hessian for small
- less non-linear in **m**
- need to solve overdetermined PDE
- not clear how to pick λ
- ...

One reflector example

true model

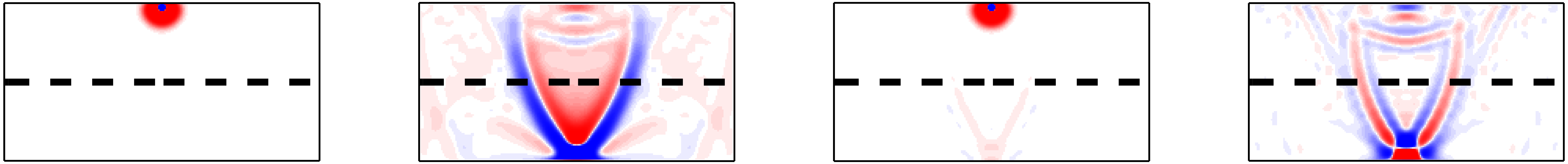


Wavefields in *homogeneous* background

FWI

WRI

forward | adjoint | reconstructed wavefield | PDE residual

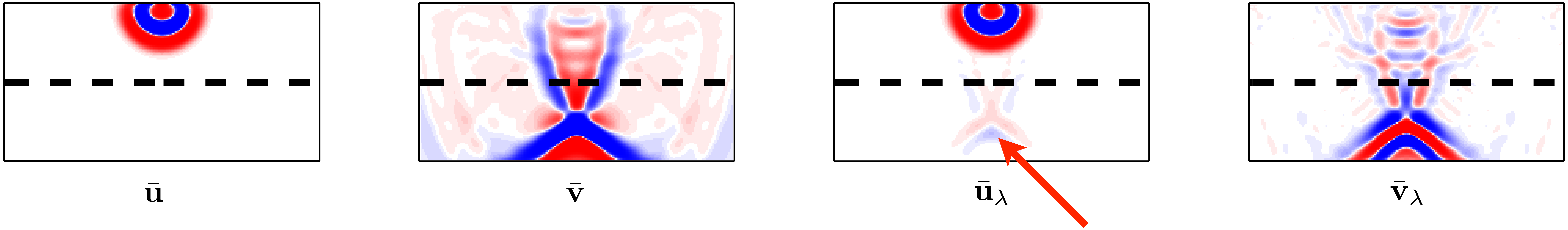


$\bar{\mathbf{u}}$ | $\bar{\mathbf{v}}$ | $\bar{\mathbf{u}}_{\lambda}$ | $\bar{\mathbf{v}}_{\lambda}$

Wavefields in *homogeneous* background

FWI **WRI**

forward | adjoint | reconstructed wavefield | PDE residual



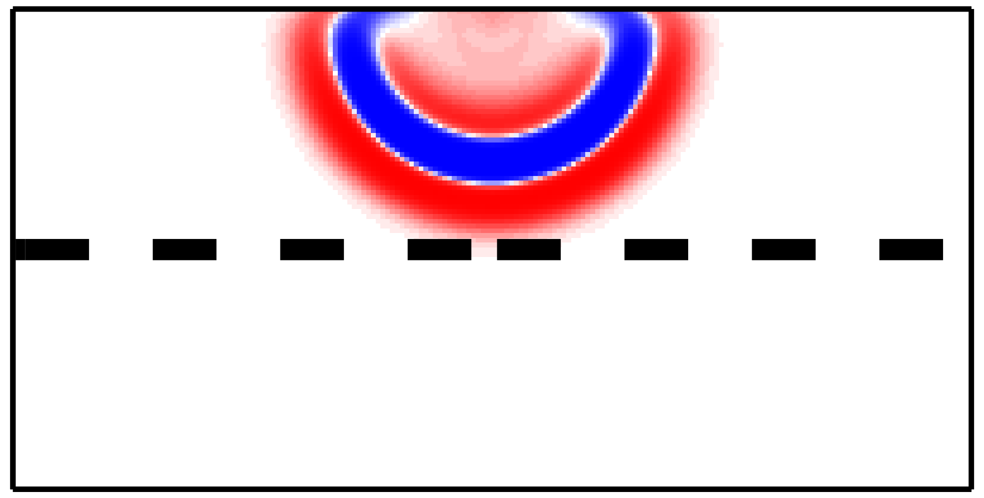
$\bar{\mathbf{u}}$ | $\bar{\mathbf{v}}$ | $\bar{\mathbf{u}}_\lambda$ | $\bar{\mathbf{v}}_\lambda$

Wavefields in *homogeneous* background

FWI

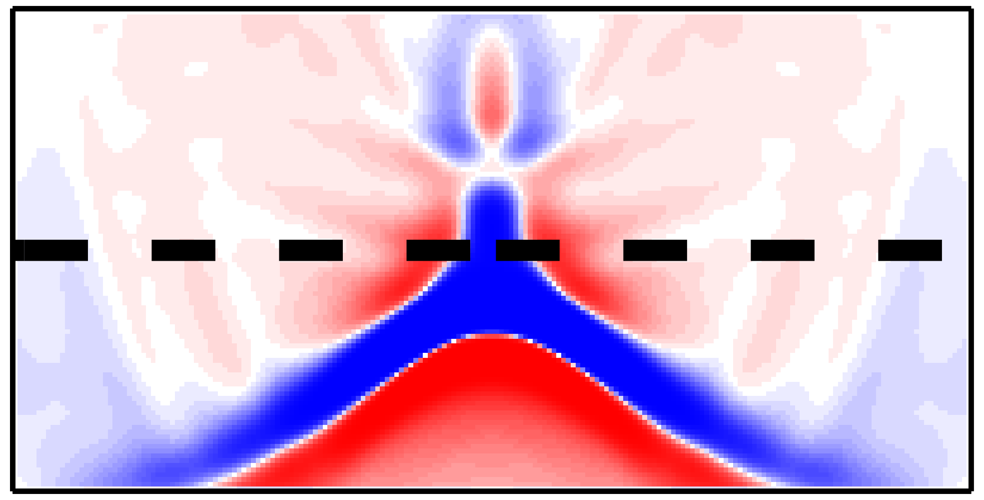
WRI

forward



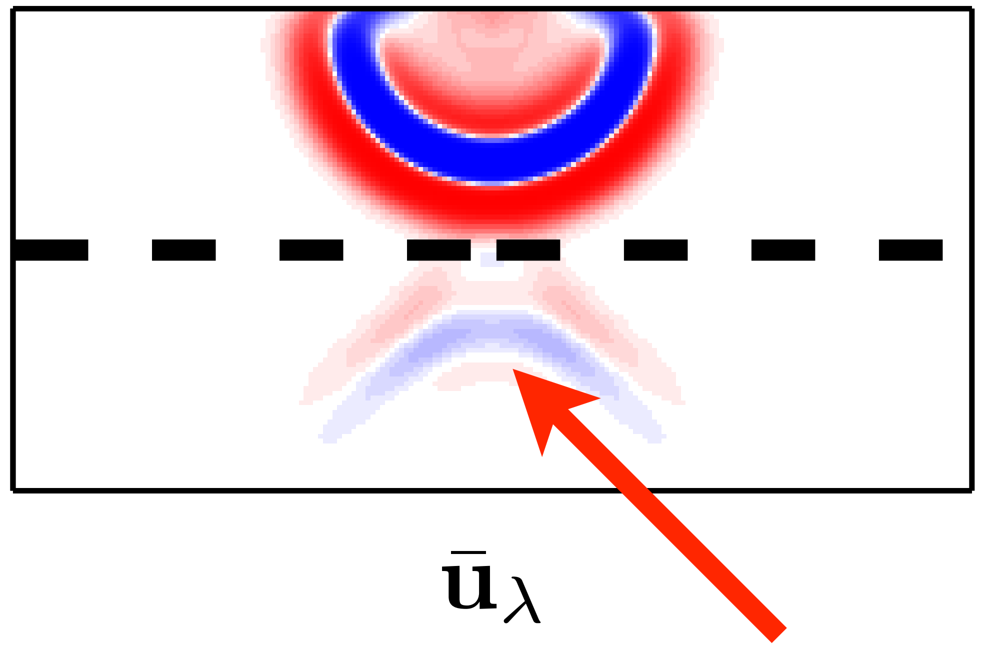
$\bar{\mathbf{u}}$

adjoint



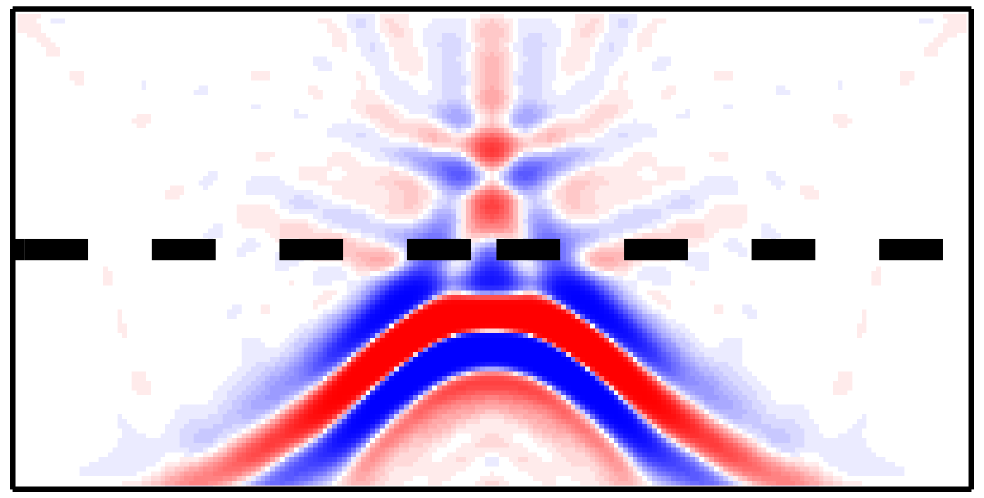
$\bar{\mathbf{v}}$

reconstructed wavefield



$\bar{\mathbf{u}}_\lambda$

PDE residual



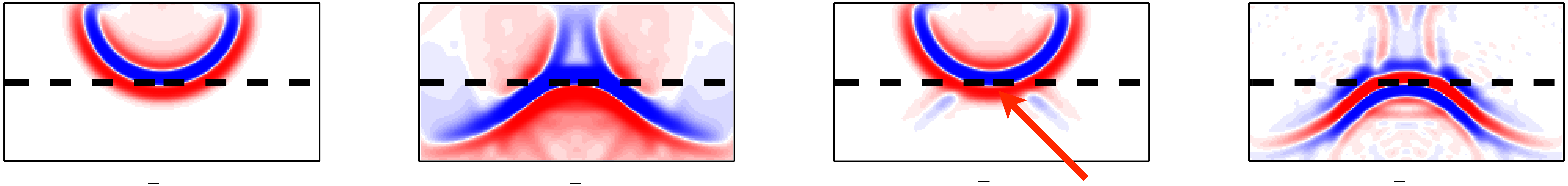
$\bar{\mathbf{v}}_\lambda$

Wavefields in *homogeneous* background

FWI

WRI

forward | adjoint | reconstructed wavefield | PDE residual

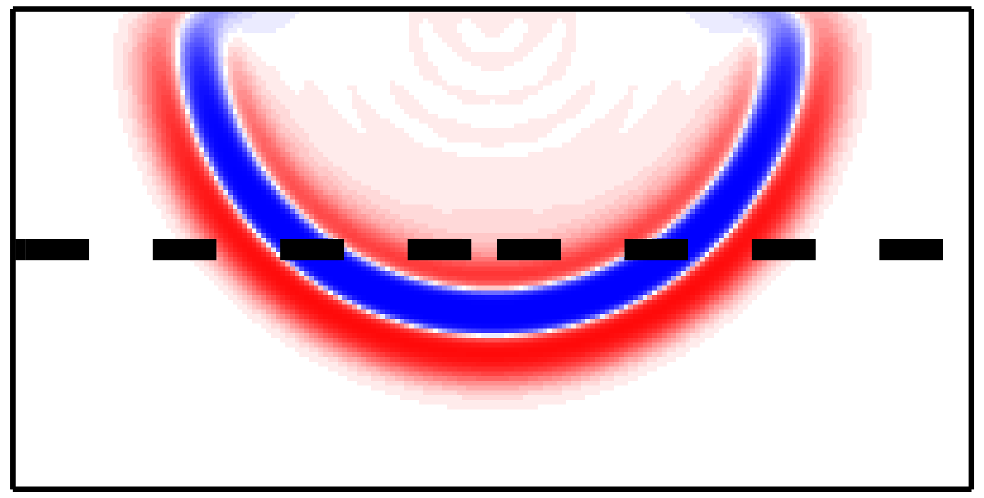


$\bar{\mathbf{u}}$ | $\bar{\mathbf{v}}$ | $\bar{\mathbf{u}}_\lambda$ | $\bar{\mathbf{v}}_\lambda$

Wavefields in *homogeneous* background

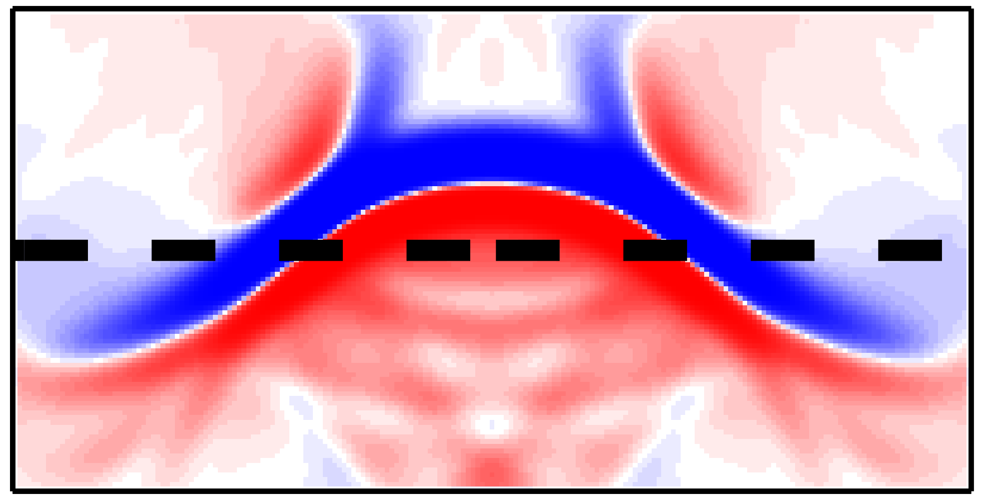
FWI

forward



$\bar{\mathbf{u}}$

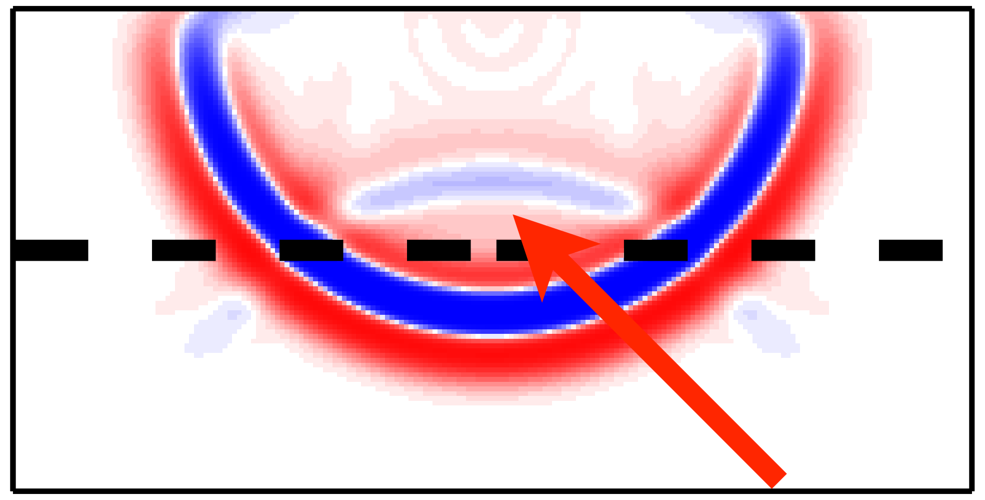
adjoint



$\bar{\mathbf{v}}$

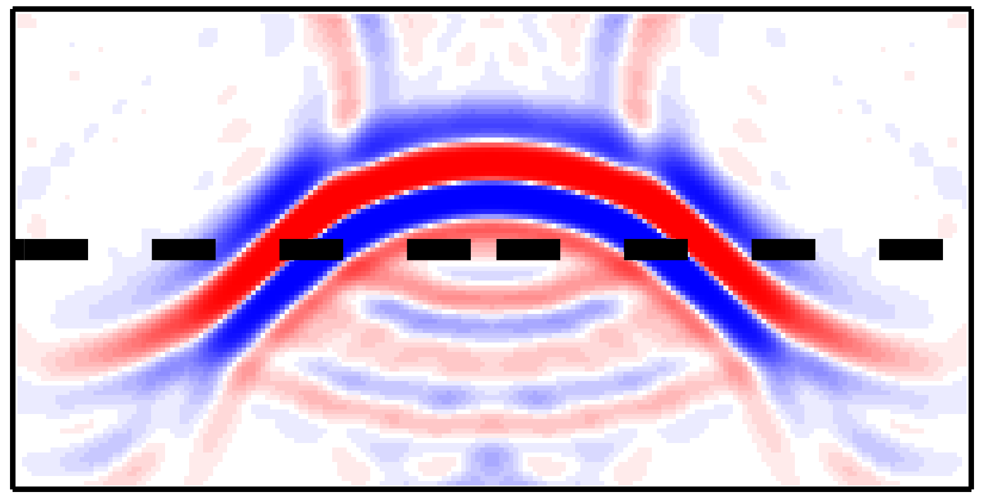
WRI

reconstructed wavefield



$\bar{\mathbf{u}}_{\lambda}$

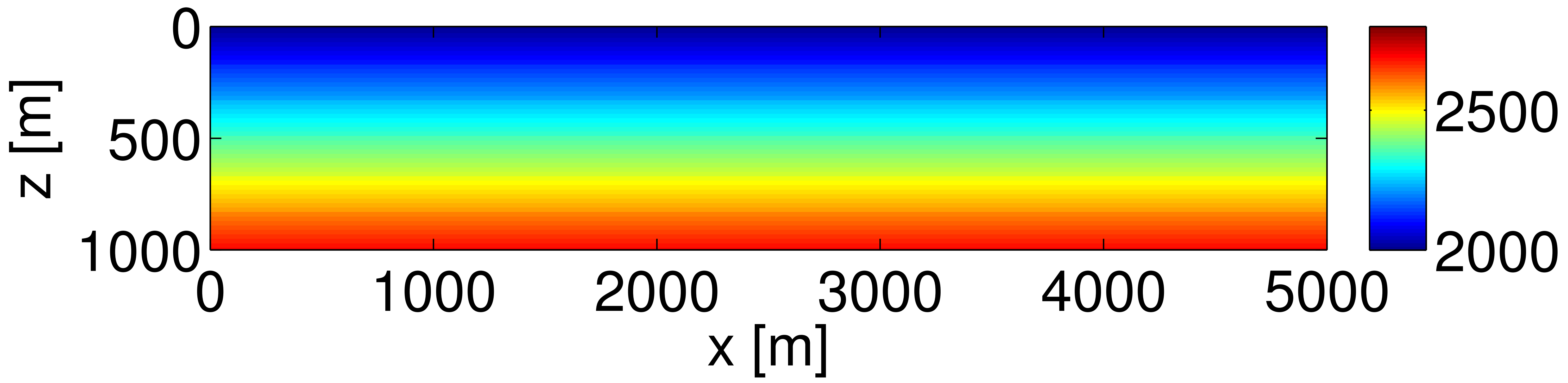
PDE residual

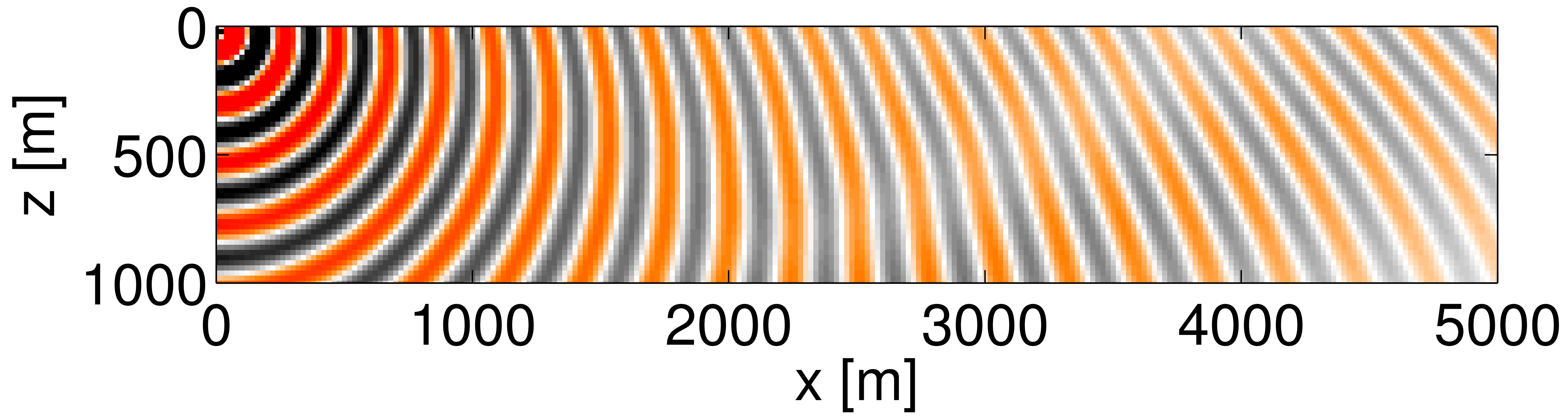


$\bar{\mathbf{v}}_{\lambda}$

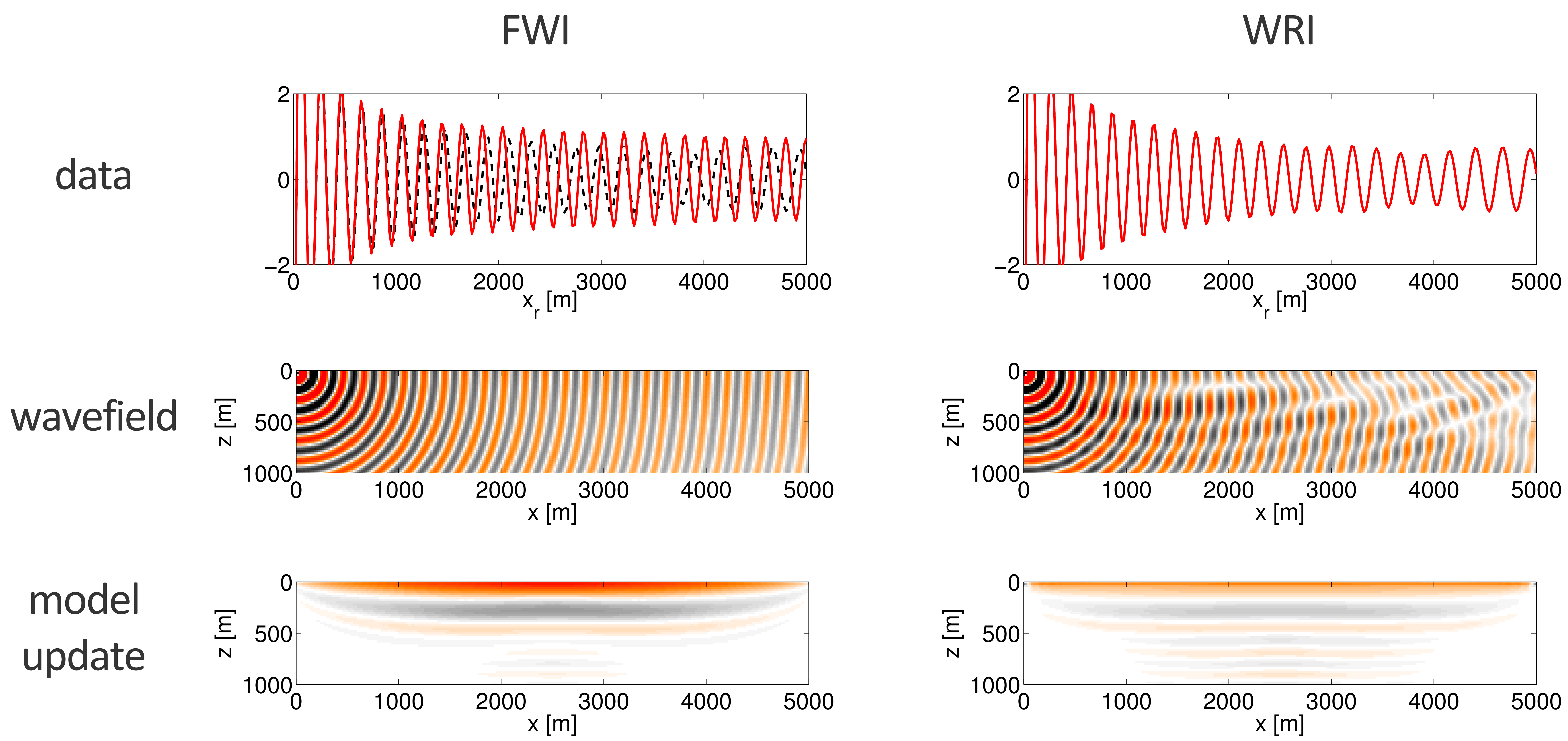
Diving wave example

true model and wavefield





Wavefields in *homogeneous* background



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

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

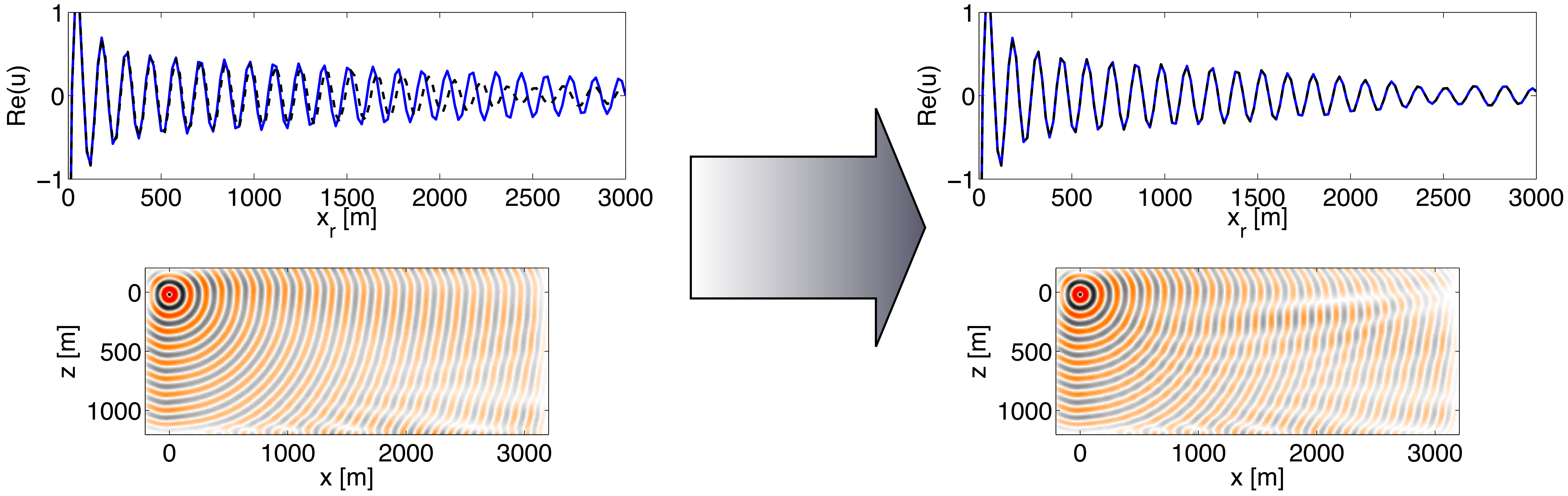
For a physically plausible model we have

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

[Symes, personal communication]

Warping

The overdetermined WE is a way of warping



[Baek ’13, Ma ’13]

[van Leeuwen & FJH, ’13]

WRI vs. FWI

Penalty method

for each source i

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

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

$$H_{GN} = H_{GN} + \lambda^2 \omega^4 \mathrm{diag}(\mathbf{u}_i)^* \mathrm{diag}(\mathbf{u}_i)$$

$$\mathbf{m} := \mathbf{m} - \alpha H_{GN}^{-1} \mathbf{g}$$

end

Conventional method

for each source i

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

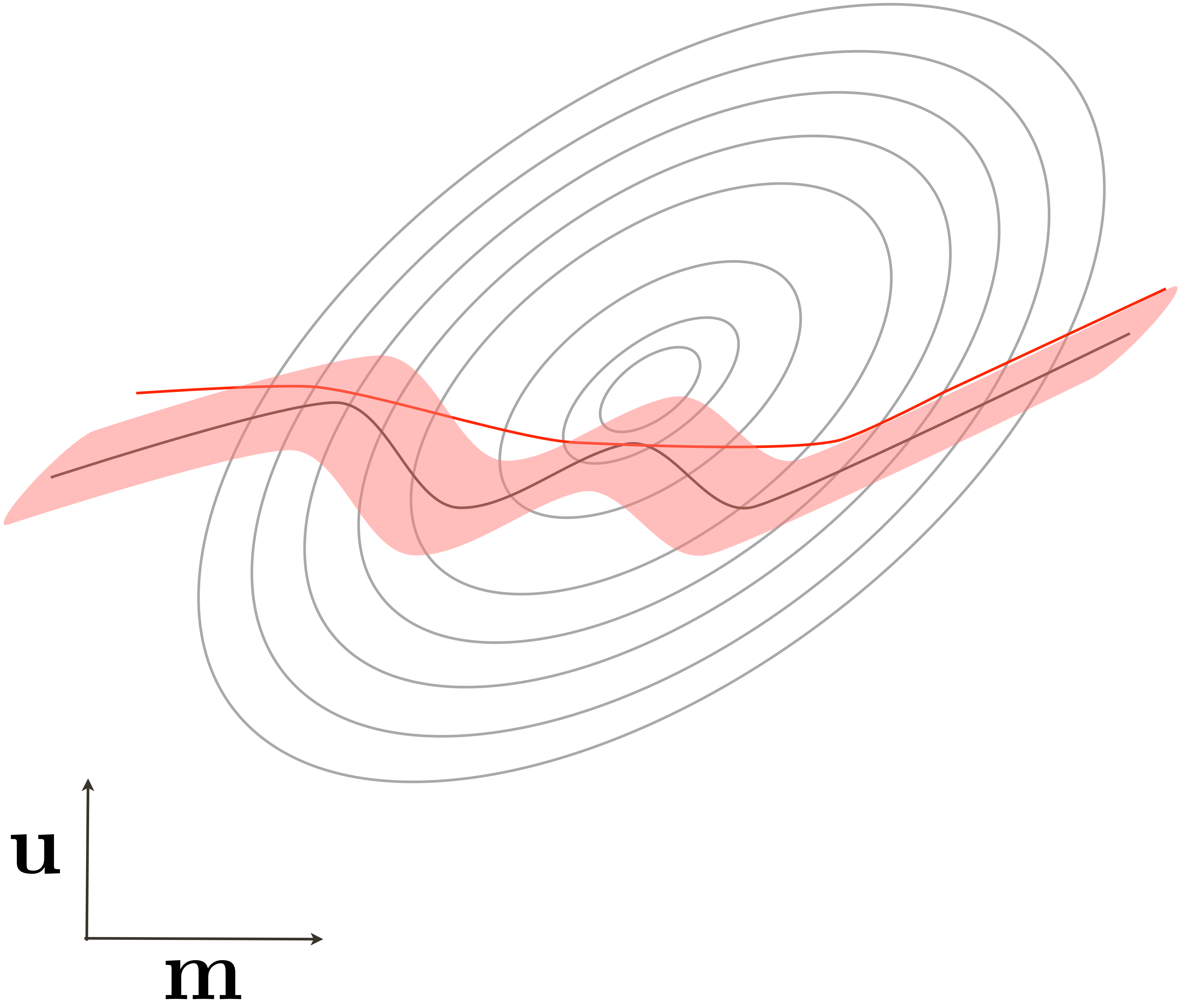
$$\textit{solve } A(\mathbf{m})^* \mathbf{v}_i = P^*(P\mathbf{u}_i - \mathbf{d}_i)$$

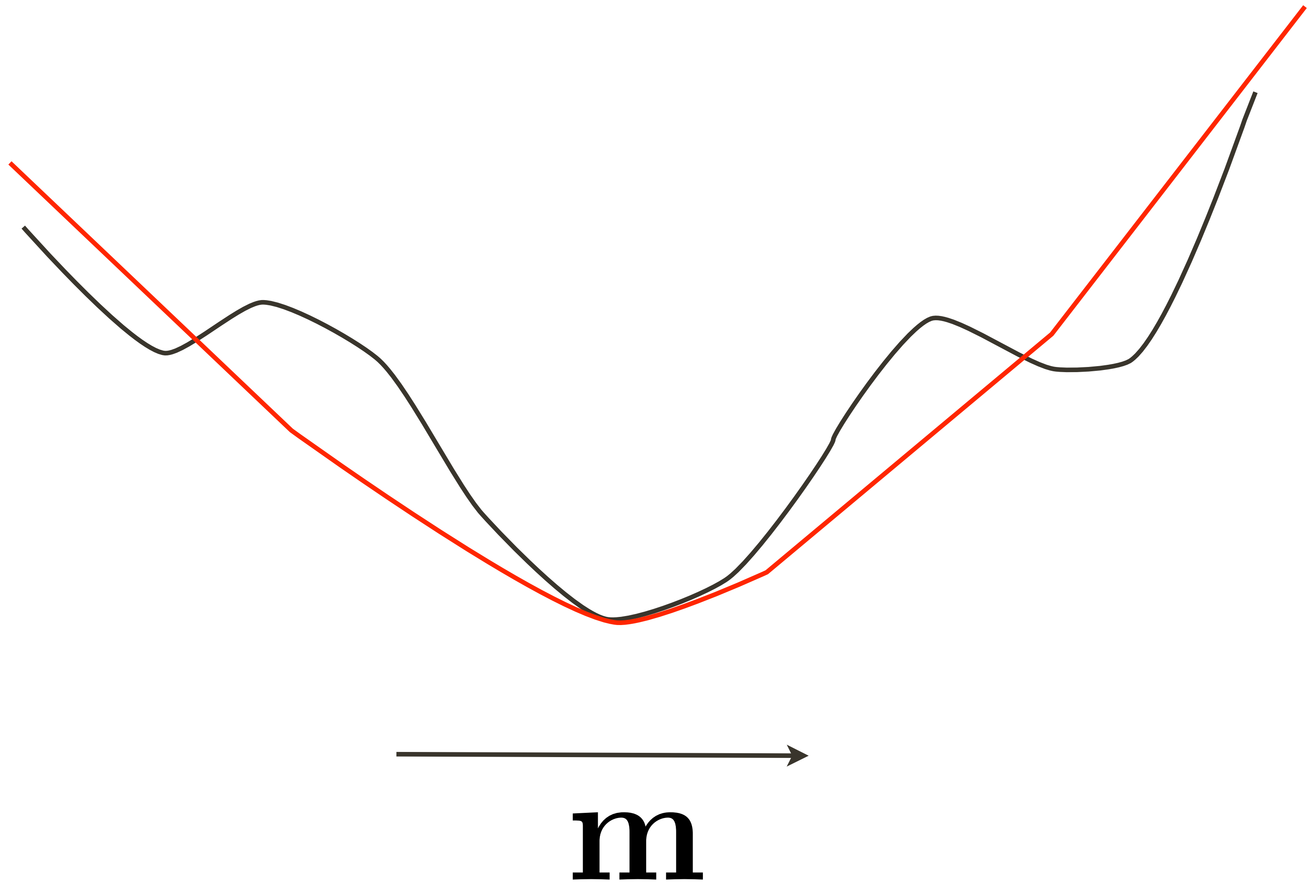
$$\mathbf{g} = \mathbf{g} + \omega^2 \mathrm{diag}(\mathbf{u}_i)^* \mathbf{v}_i$$

$$\mathbf{m} = \mathbf{m} - \alpha \mathbf{g}$$

end

WRI vs. FWI

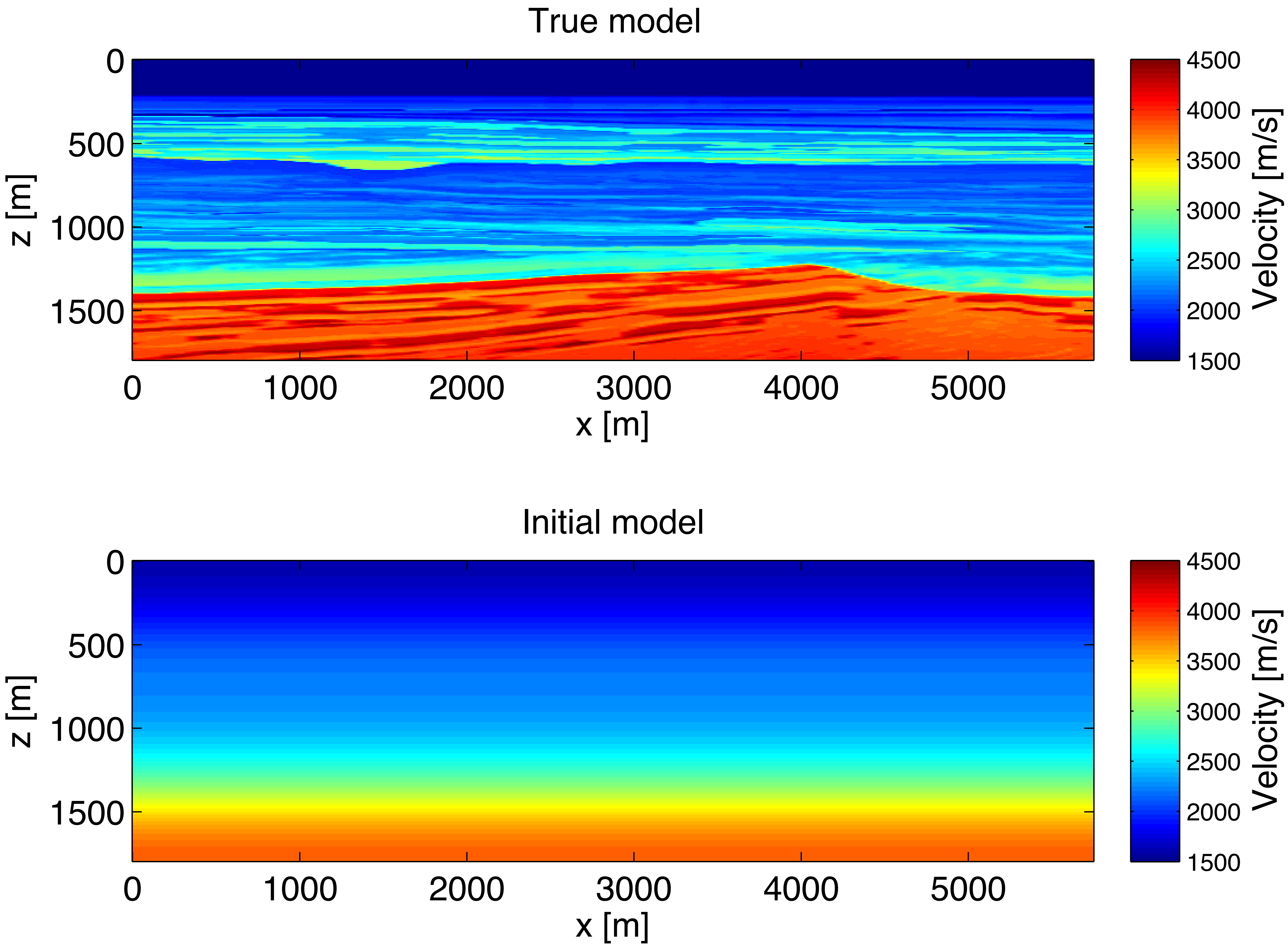




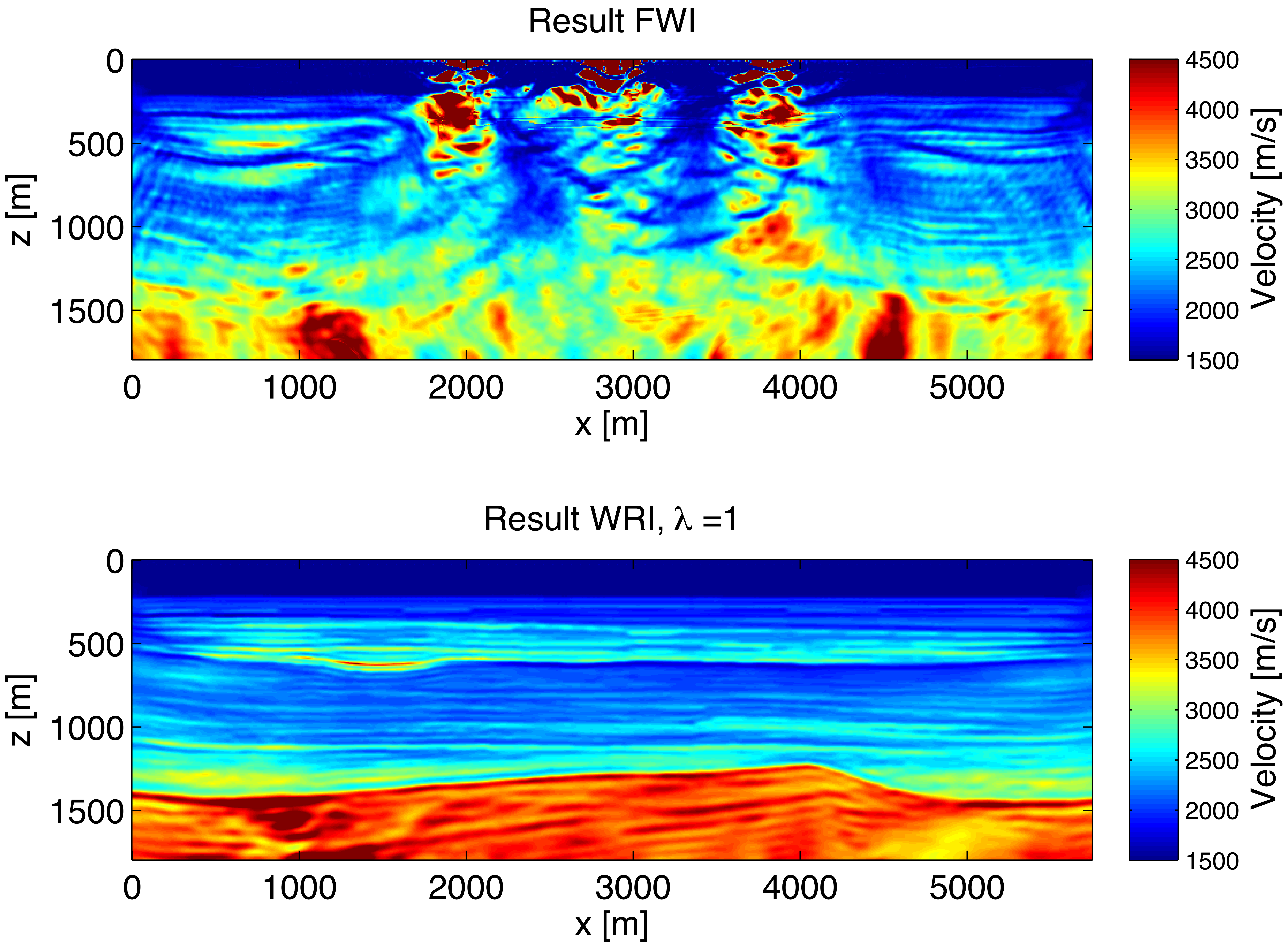
Example – BG Compass model

- *Low* frequencies missing, 24 frequency batches (15 iterations each) {5 6} ,{6 7},… ,{28 29} Hertz. Each interval contains 5 frequencies.
- 103 sources/receivers w/ 55m sample interval
- Inaccurate *initial* model

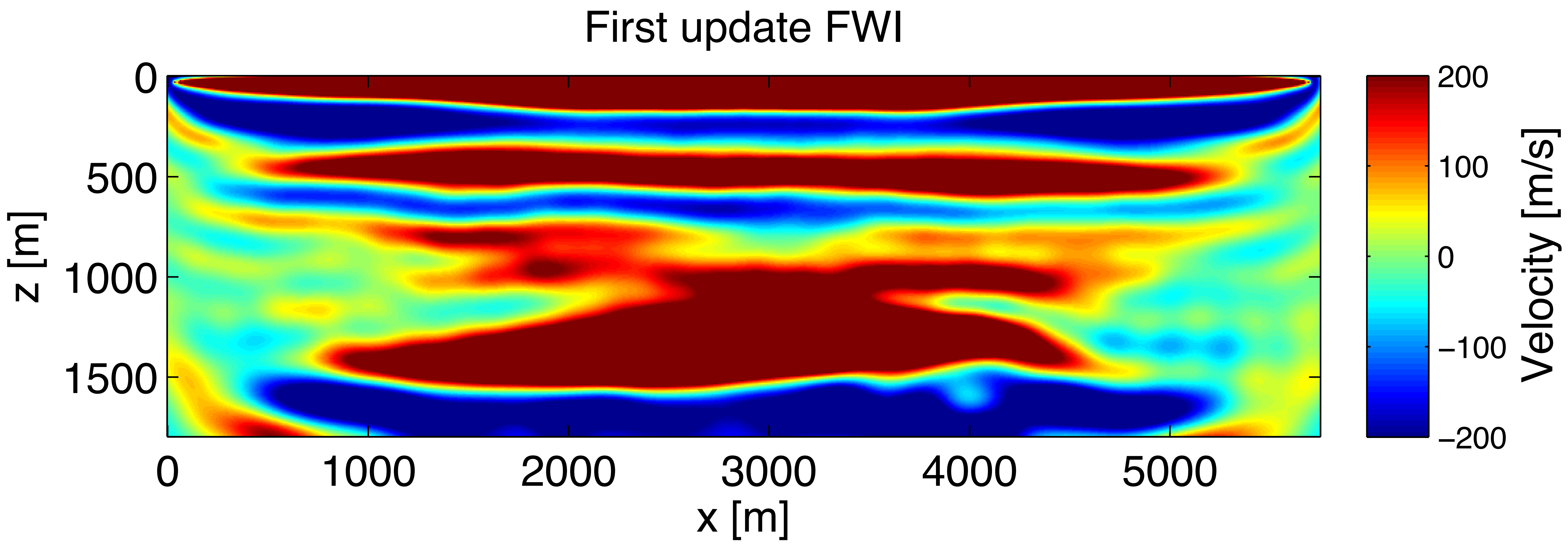
True & initial model

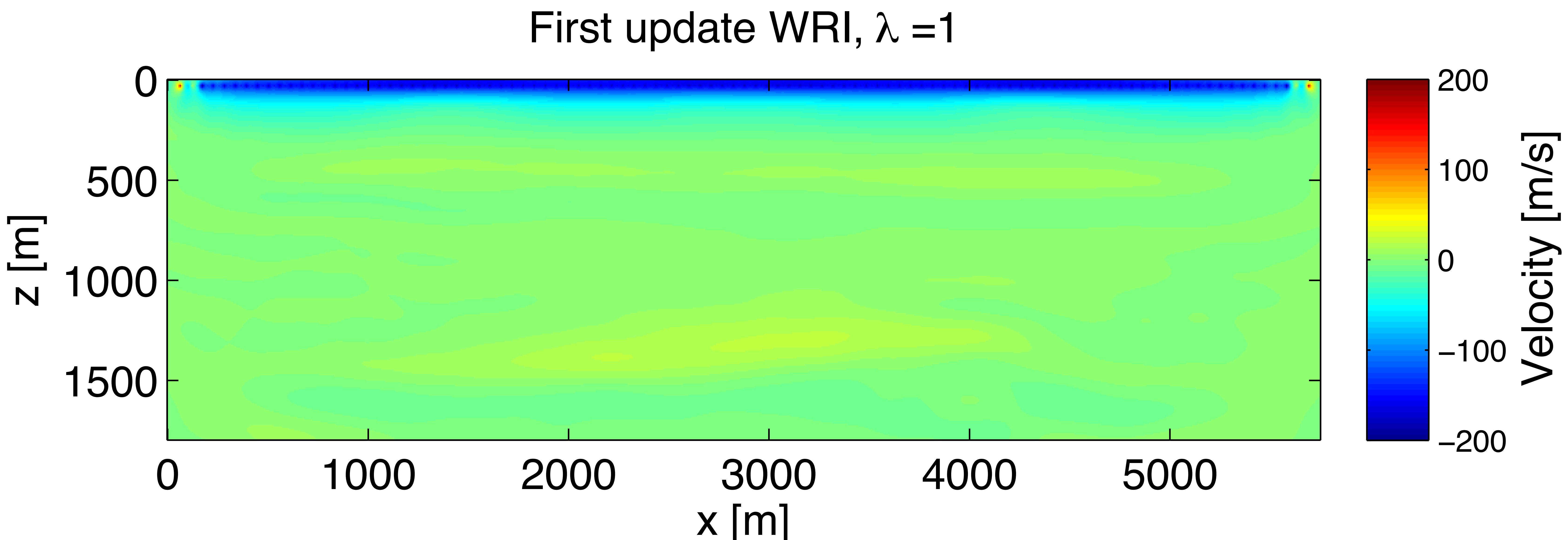


FWI vs WRI

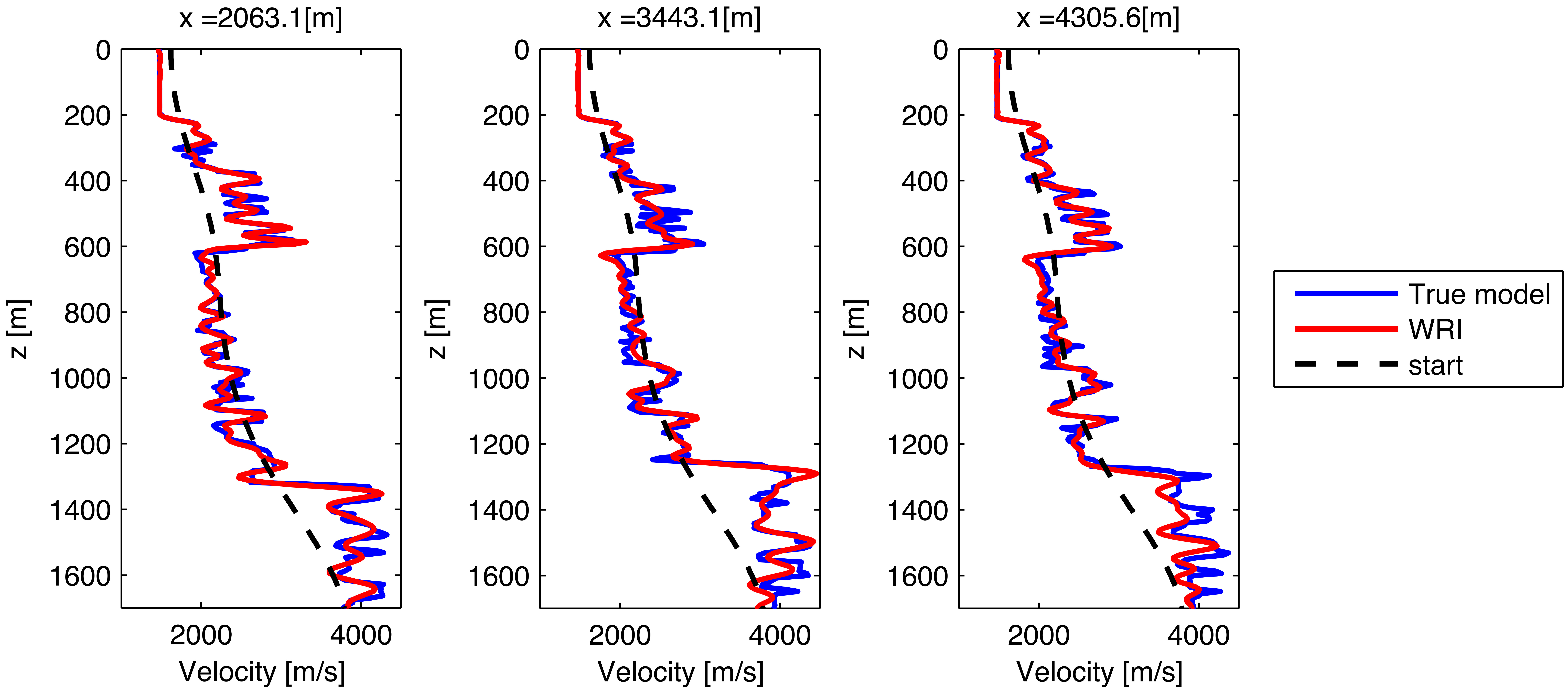

Result FWI
0
500
1000
1500
z [m]
0
1000
2000
3000
4000
5000
x [m]
4500
4000
3500
3000
2500
2000
1500
Velocity [m/s]
Result WRI, λ =1
0
500
1000
1500
z [m]
0
1000
2000
3000
4000
5000
x [m]
4500
4000
3500
3000
2500
2000
1500
Velocity [m/s]

Gradients


First update FWI
z [m]
x [m]
0
500
1000
1500
0
1000
2000
3000
4000
5000
200
100
0
−100
−200
Velocity [m/s]


First update WRI, λ =1
z [m]
x [m]
0
500
1000
1500
0
1000
2000
3000
4000
5000
200
100
0
−100
−200
Velocity [m/s]

Cross sections



Perspectives

Gauss-Newton Hessian is sparse

- fast evaluation of inverse
- possibility to include TV or one-norm minimization (Ernie's talk)
- possibility to exploit for multi-parameter (WRI)

Conclusions

New method for *wave*-equation based inversion:

- same *extended* search space as in *all-at-once* but with memory & CPU requirements as in *adjoint*-state approach
- *no* adjoints & *sparse* GN-Hessian approximation
- "*less* non-linear" and therefore *less* susceptible to *local* minima

Still somewhat *early* days in the development:

- *encouraging* 2-D results w/ ***less** sensitivity* to *initial* model
- ***cheap** scalings* of the *inverse* of the GN Hessian
- extension to 3D is challenging

Acknowledgements

Thank you for your attention !

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

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, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, WesternGeco, and Woodside.