Wavefield-Reconstruction Inversion - WRI

Felix J. Herrmann

SLIM

University of British Columbia

Strategy

Derive an alternative *extended* formulation:

- *fits* data for *poor* starting models
- less prone to local minima
- computationally feasible
- relaxes the physics while staying solidly grounded

[Richter, ’81]

SLIM

Equation error approach

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

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

via

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

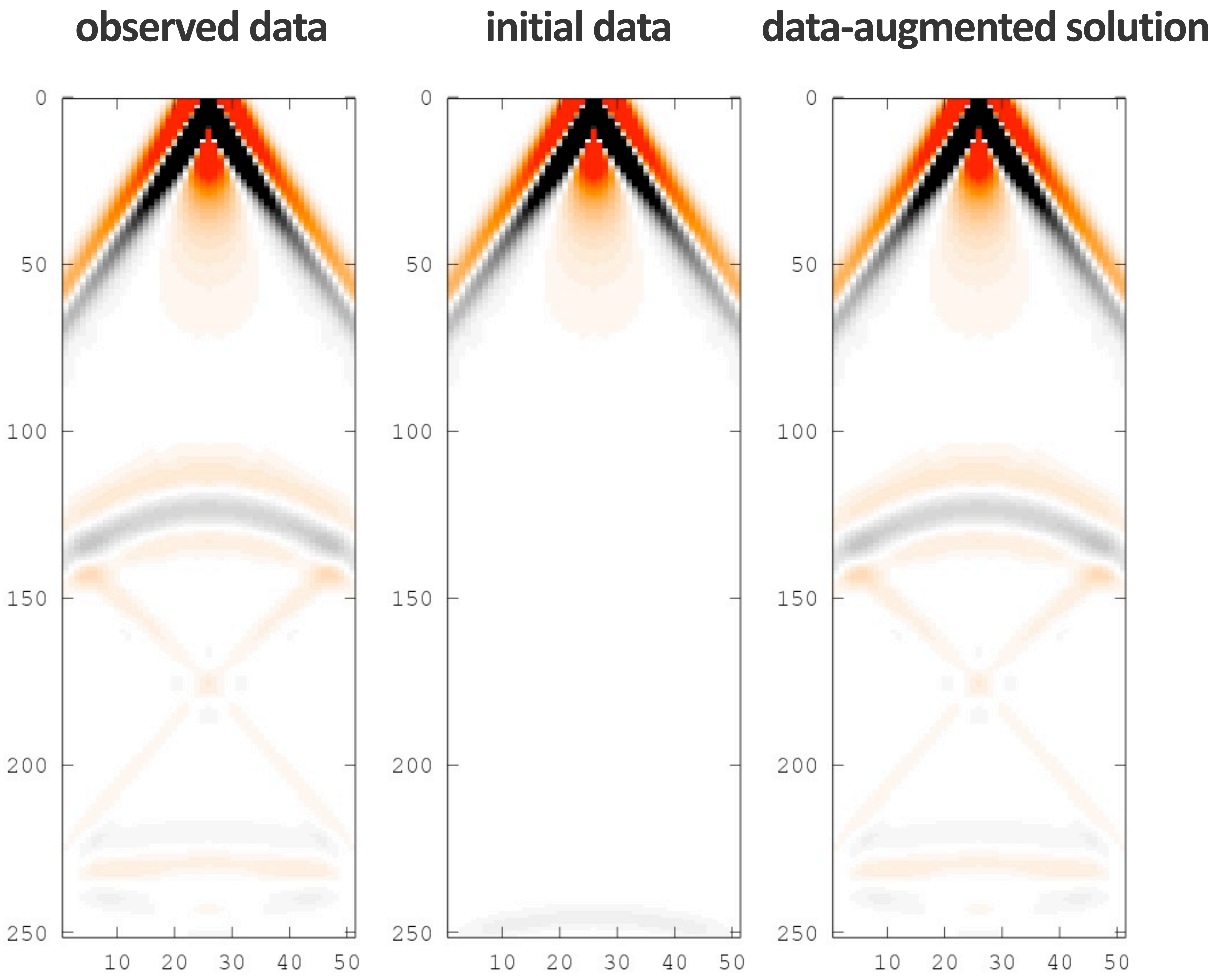
which is *linear* in $\mathbf{m}$.

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

wave-equation X wavefield = source

versus

$$\begin{pmatrix} \text{wave-equation} \\ \text{- - - - - - - - - - - -} \\ \text{sampling operator} \end{pmatrix} \times \text{wavefield} = \begin{pmatrix} \text{source} \\ \text{- - - - - - - - - - - -} \\ \text{data} \end{pmatrix}$$



observed data
initial data
data-augmented solution
0
50
100
150
200
250
10
20
30
40
50

[van Leeuwen & FJH, 2013]

SLIM

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

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

[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

source

Helmholtz equation

- 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

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

- 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 dependence on earth model **(m)**
- dense (GN) Hessian, involves additional PDE solves
- **reliance on accurate starting models to avoid cycle skipping**

[Bertsekas, '96; Wright, '00; van Leeuwen & FJH, '13]

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 $\lambda \uparrow \infty$

- **no** need to store all the fields (**u**)
- **no** adjoint solves
- sparse approximation of Gauss-Newton Hessian for small λ
- less non-linear in **m**
- need to solve data-augmented wave equation

[Aria Abubakar et. al. ’09]

Related work

Contrast-source formulation

- combined objective is similar
- but does not eliminate wavefields via variable projection [Golub ’03, van Leeuwen & Aravkin ’12]
- **requires storage of wavefields for all sources**

Tomographic extensions

- **sensitivities to “noise” & relative strengths of events**
- WRI uses wave equation itself to “focus”

[Aravkin & van Leeuwen, ’12; van Leeuwen & FJH, ’13]

Variable projection

Solve data-augmented wave equation for each source

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

Define reduced objective with proxy wavefields

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

[van Leeuwen & FJH, ’13]

SLIM

WRI method

for each source i

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

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

end

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

correlation proxy wavefield & PDE residual

Conventional method

for each source i

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

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

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

end

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

correlation wavefield & data residual

Tristan van Leeuwen and Felix J. Herrmann, "Mitigating local minima in full-waveform inversion by expanding the search space", *Geophysical Journal International*, vol. 195, p. 661-667, 2013
Tristan van Leeuwen, Felix J. Herrmann, and Bas Peters, "A new take on FWI: wavefield reconstruction inversion", in *EAGE*, 2014.
Bas Peters, Felix J. Herrmann, and Tristan van Leeuwen, "Wave-equation based inversion with the penalty method: adjoint-state versus wavefield-reconstruction inversion"

SLIM

WRI method

for each source i

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

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

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

end

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

diagonal Hessian
=
pseudo Hessian

Conventional method

for each source i

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

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

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

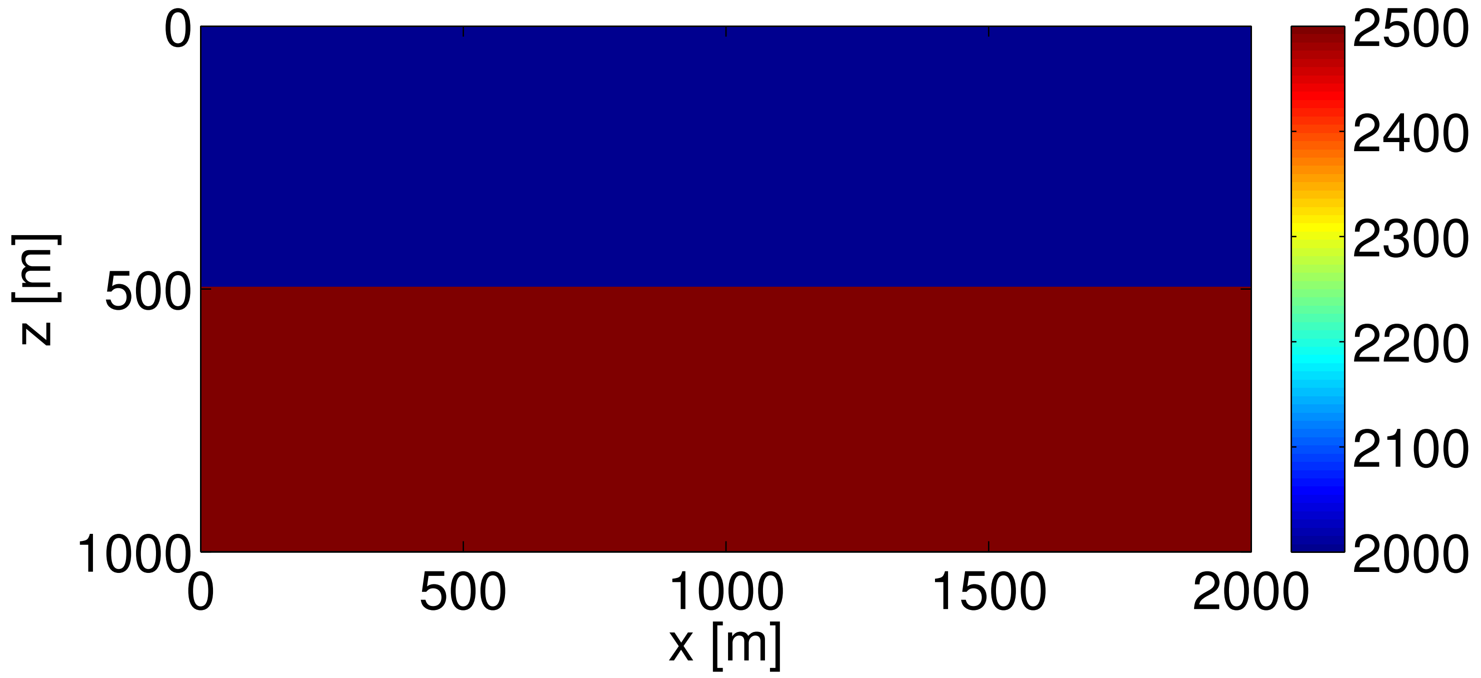
end

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

dense Hessian
&
too expensive

One reflector example

true model

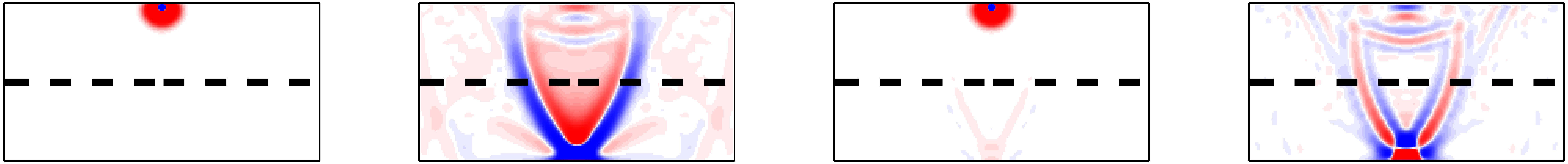


Wavefields in *homogeneous* background

FWI

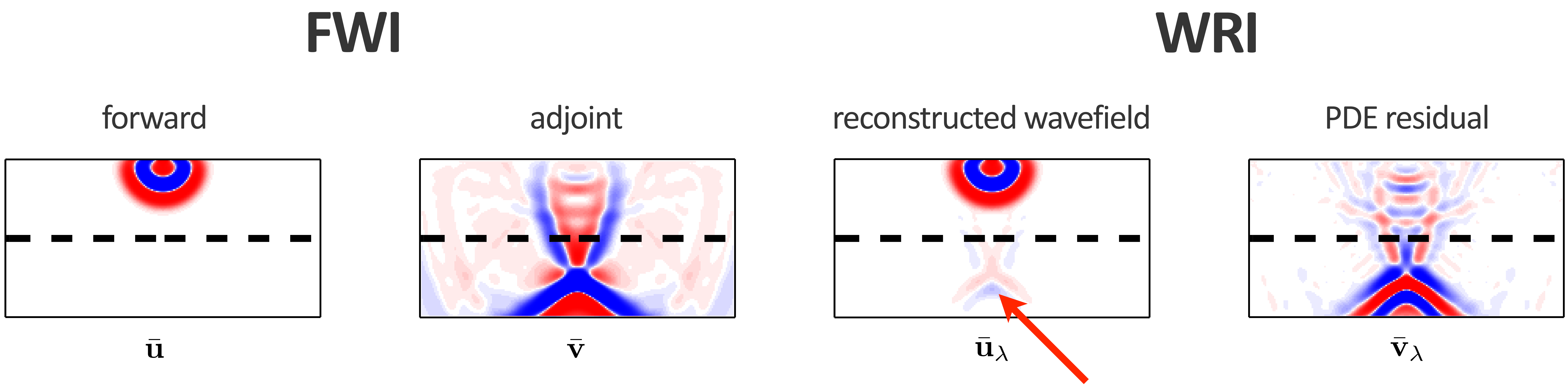
WRI

forward | adjoint | reconstructed wavefield | PDE residual



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

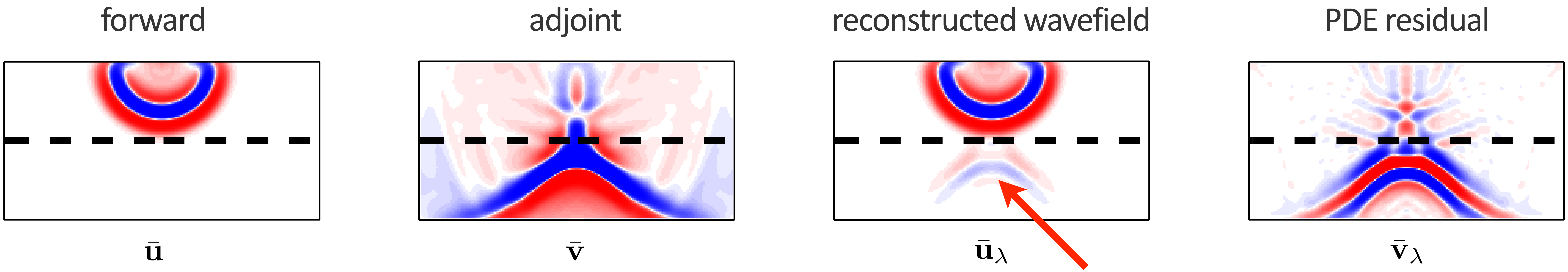
Wavefields in *homogeneous* background



Wavefields in *homogeneous* background

FWI

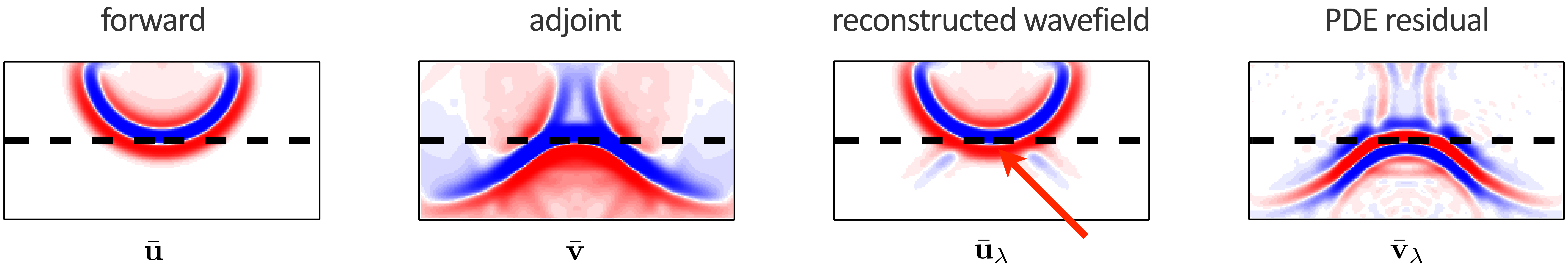
WRI



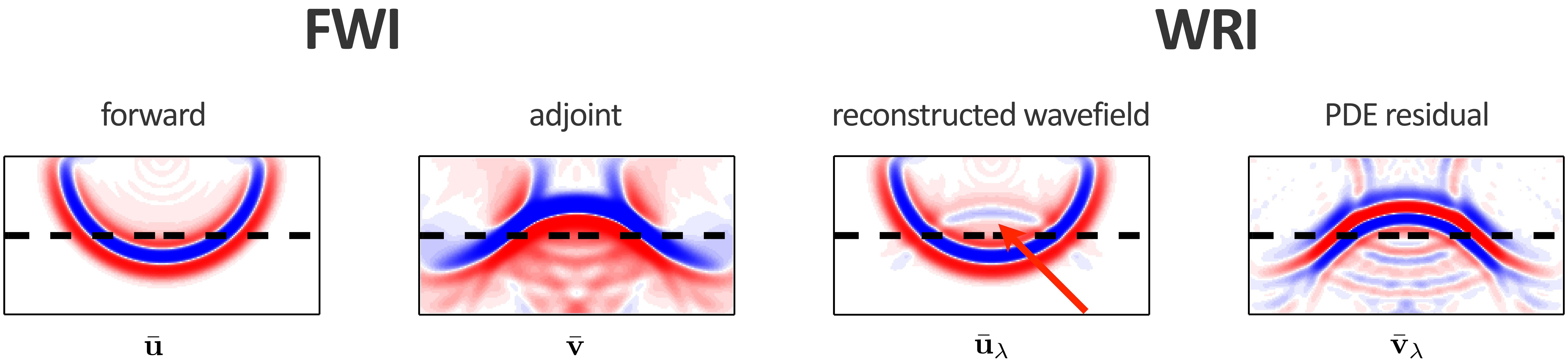
Wavefields in *homogeneous* background

FWI

WRI



Wavefields in *homogeneous* background



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

SLIM

Observations

Objective of fitting *both* data & wave equation

- introduces (reflection) events in our wavefield reconstructions
- we use these events to update the velocity model with the wave equation

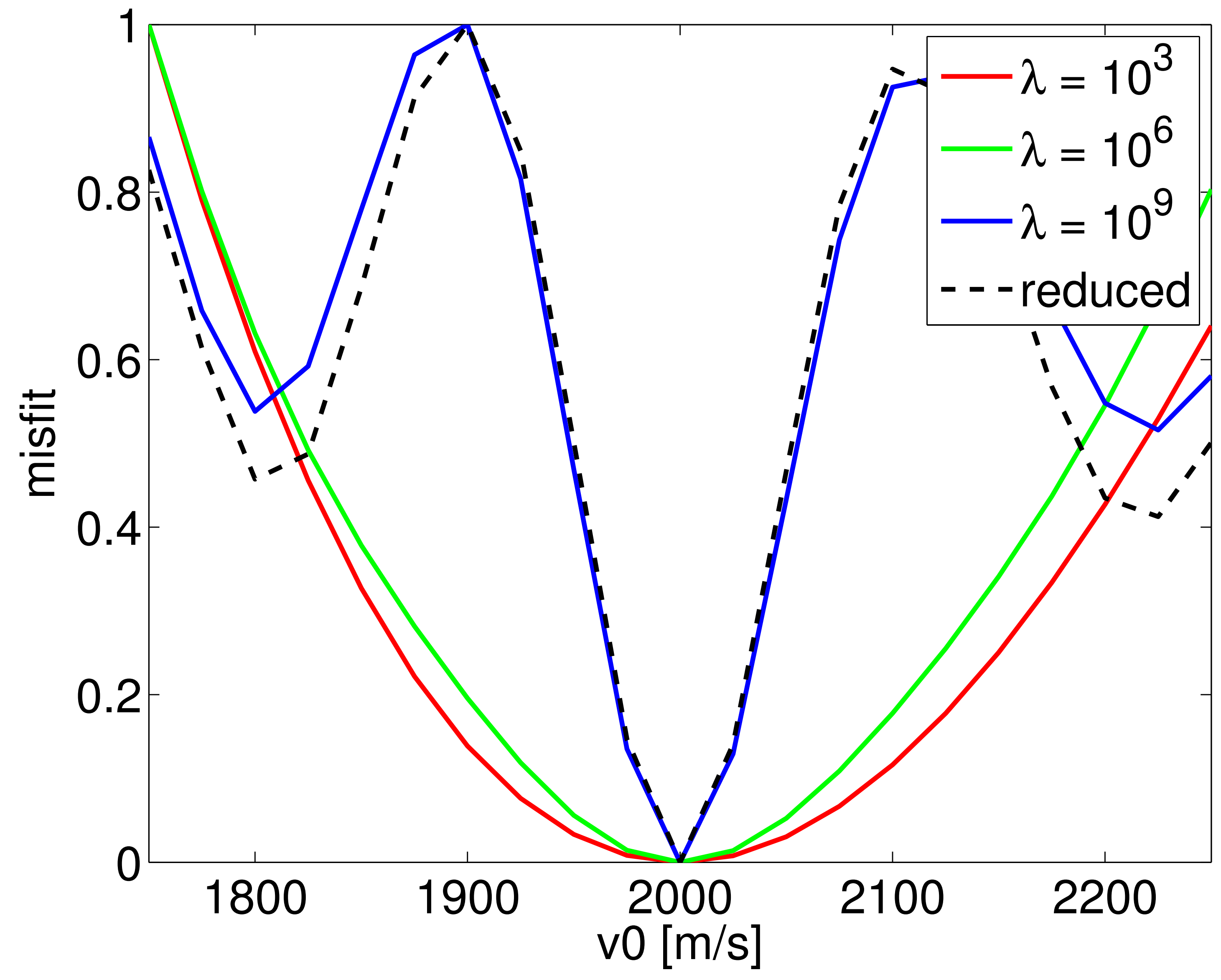
Corresponds to a variable-projection approach solving for **u** & **m**, respectively.

Differs from reduced/adjoint formulation where

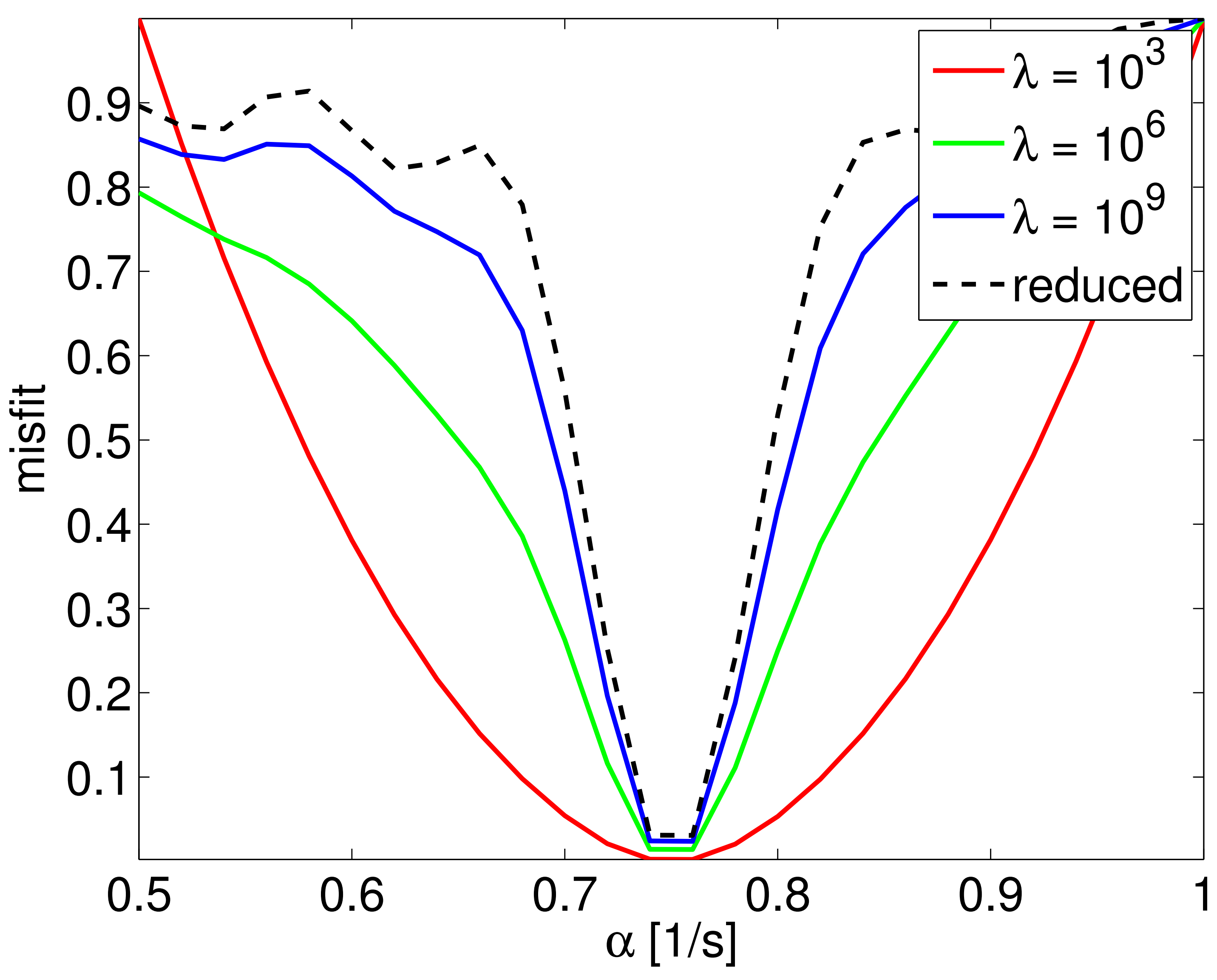
- these events are absent
- velocity is highly nonlinear in the data

Local minima

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



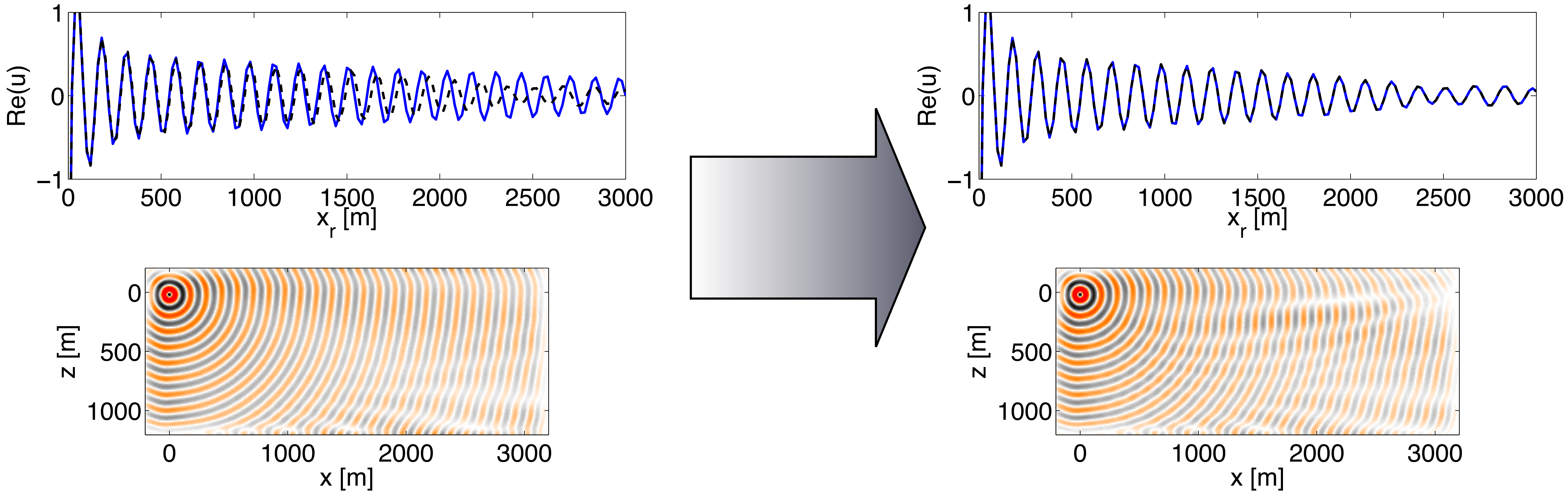
constant



slope

FWI vs WRI

Solutions of data-augmented system force data fits… **no longer cycle skipped!**



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]

Examples
– FWI is known to fail

Velocity models with

- low-velocity “kick backs”
- high-contrast high-velocity unconformities

Solve WRI w/ poor starting models using

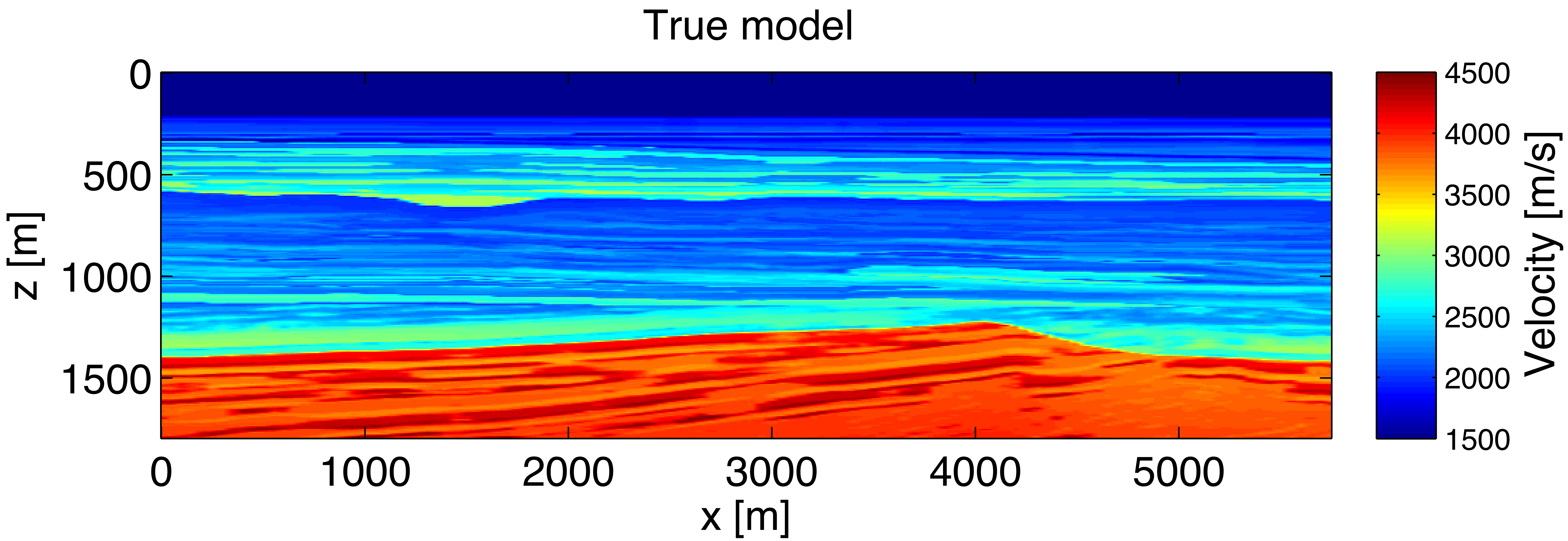
- multiple frequency sweeps w/ warm starts
- additional convex constraints

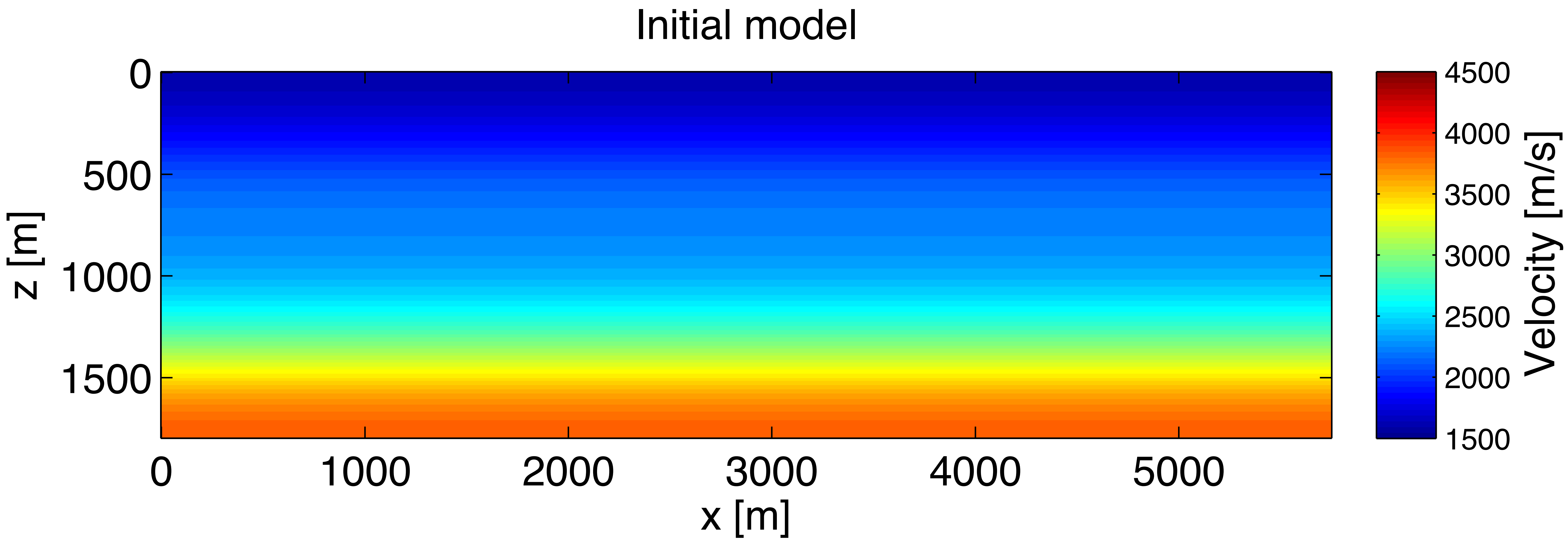
Use WRI to leverage reflected energy during the inversions…

BG Compass model

- **Challenges:** velocity kick backs & detailed geology
- Isotropic acoustic data & poor starting model
- Invert for slownesses w/ acoustic kernel
- 24 frequency batches {5 6}, {6 7}, ... , {28 29} Hz w/ 5 frequencies each
- 103 sources/receivers w/ 55m sample interval
- l-BFGS with 15 iterations per frequency band
- Two frequency sweeps

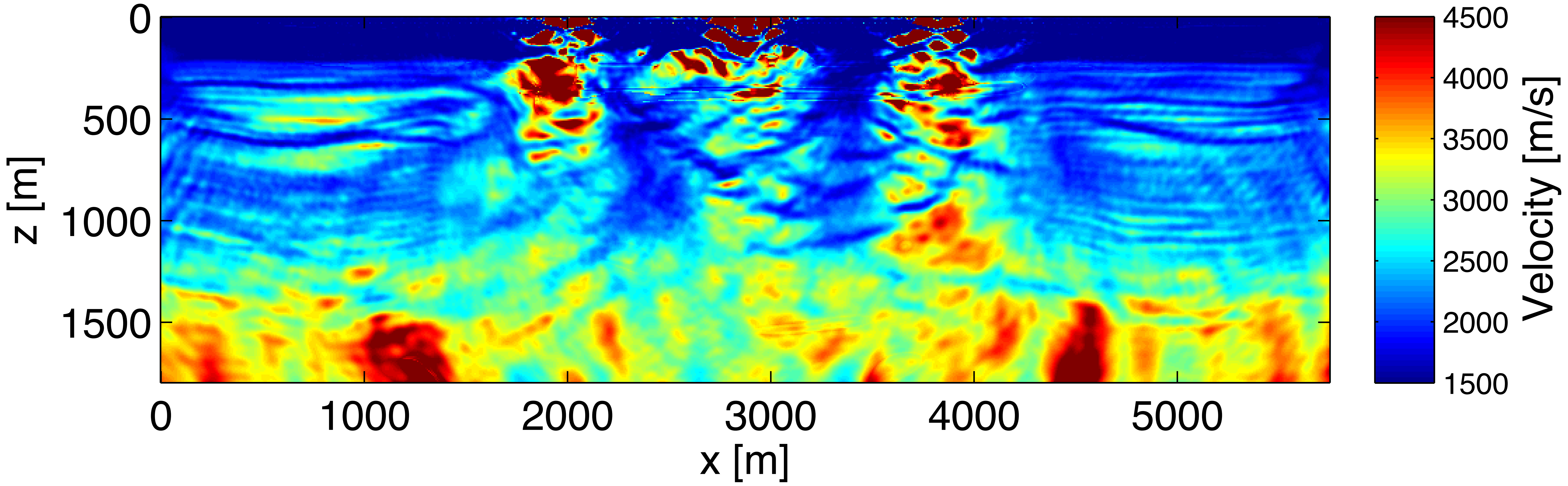
True & initial model



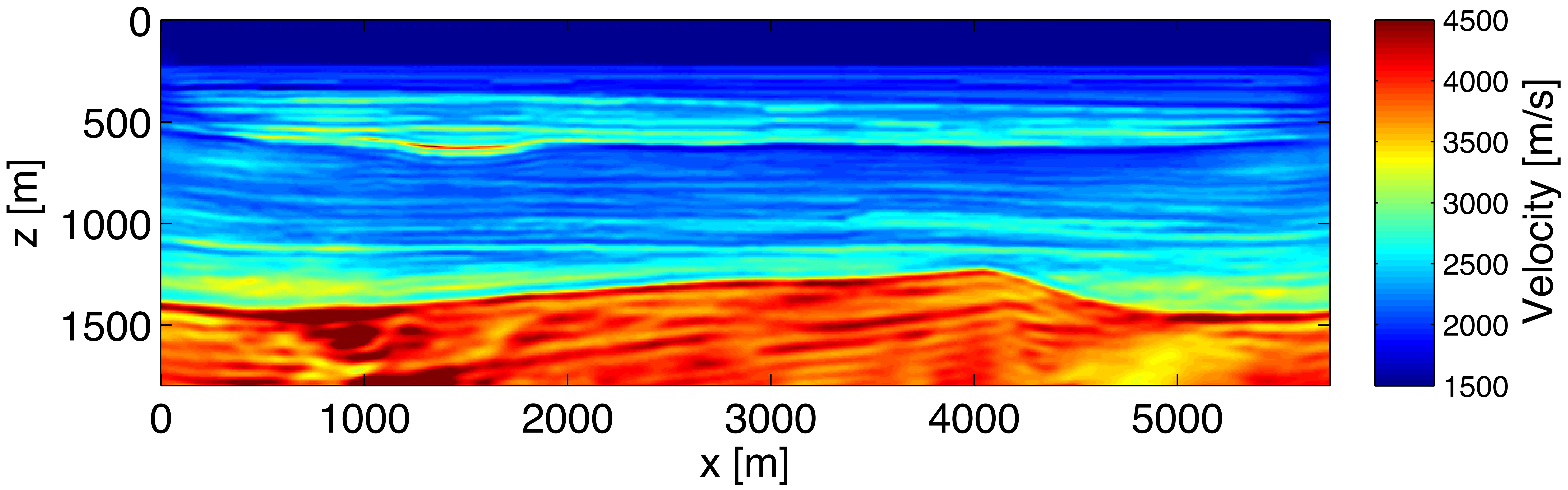


FWI vs WRI

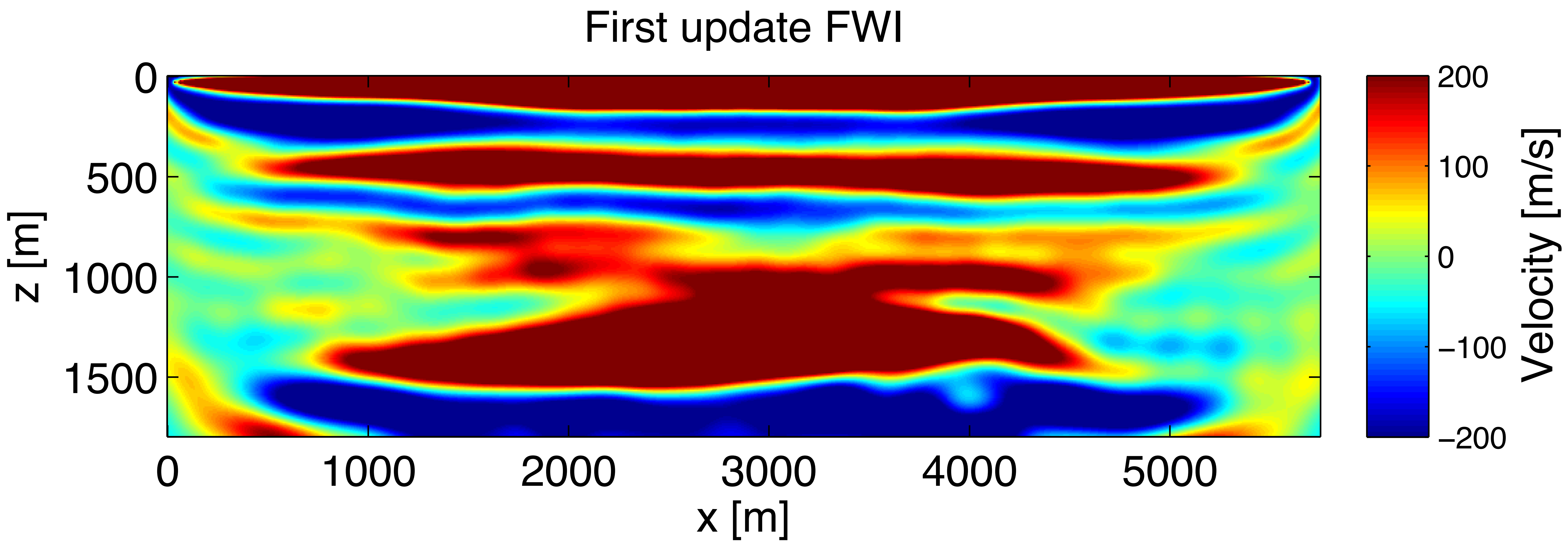
Result FWI

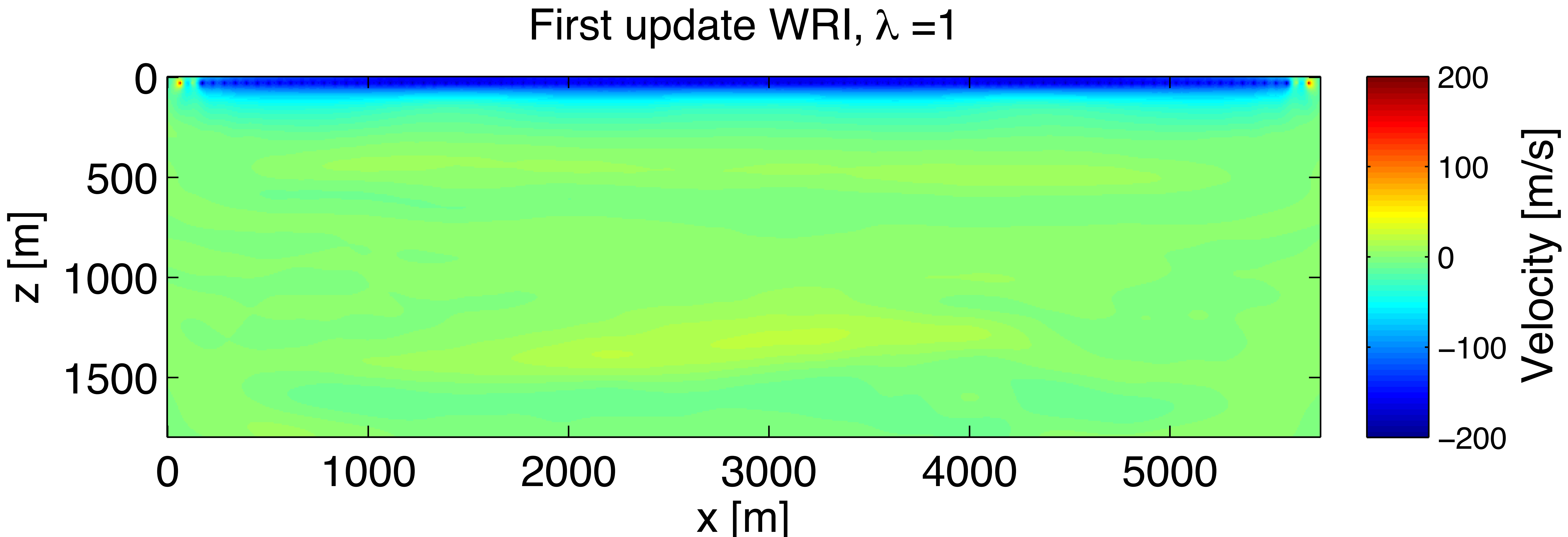


Result WRI, λ =1

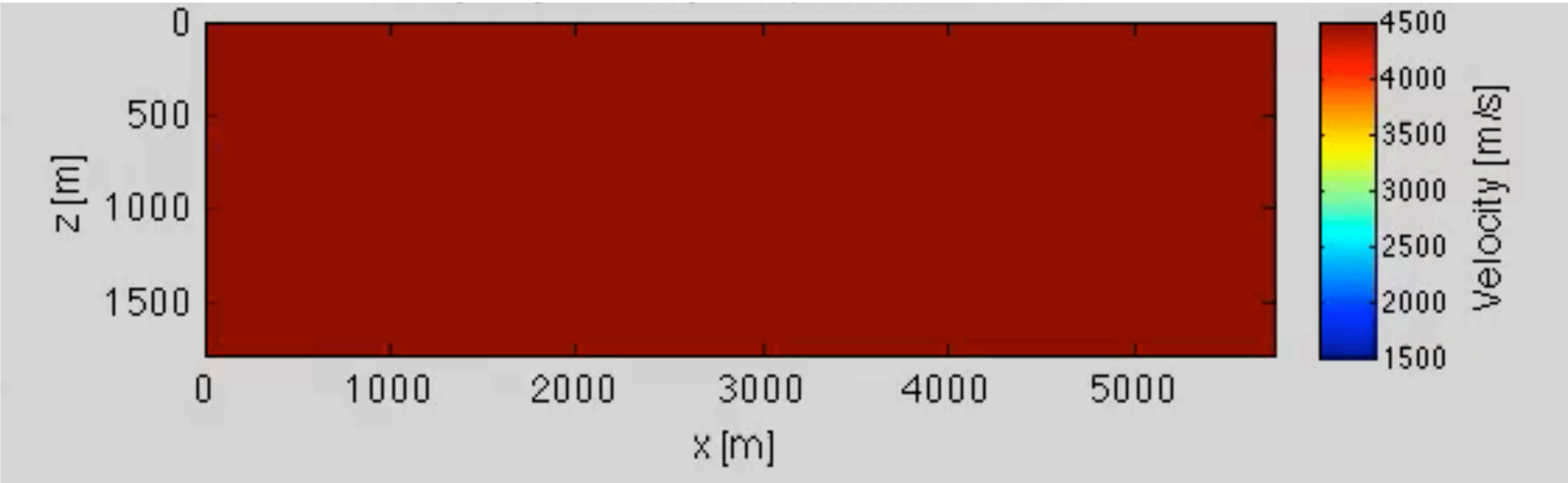


Gradients


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

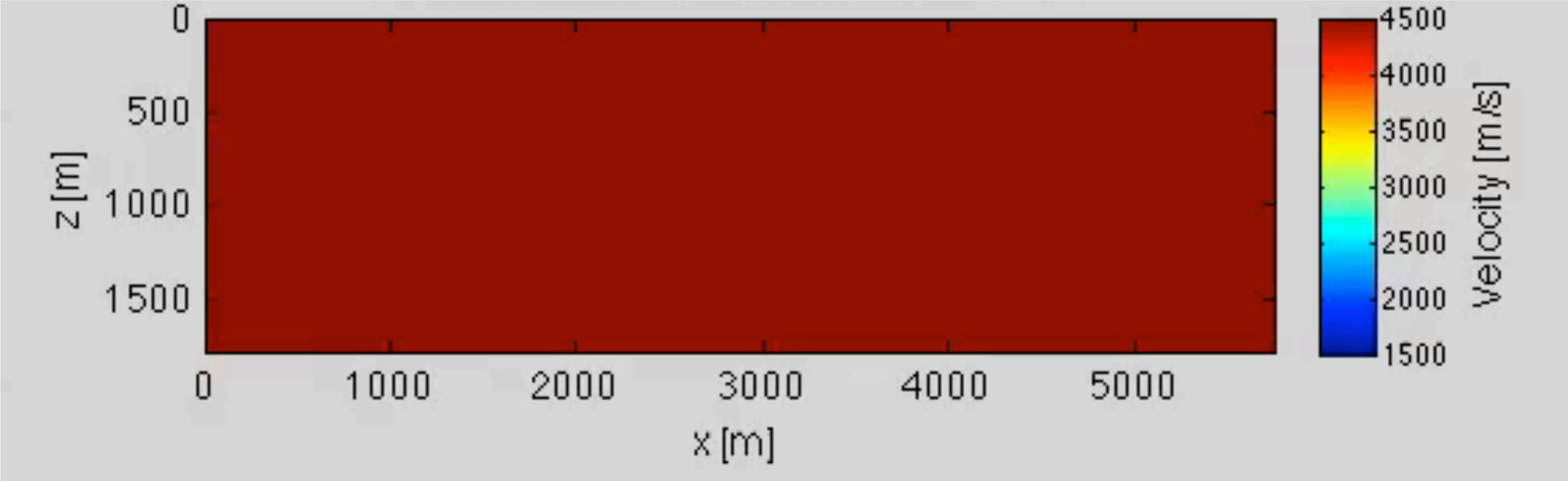

First update WRI, λ =1
z [m]
0
500
1000
1500
0
1000
2000
3000
4000
5000
x [m]
200
100
0
−100
−200
Velocity [m/s]

First sweep

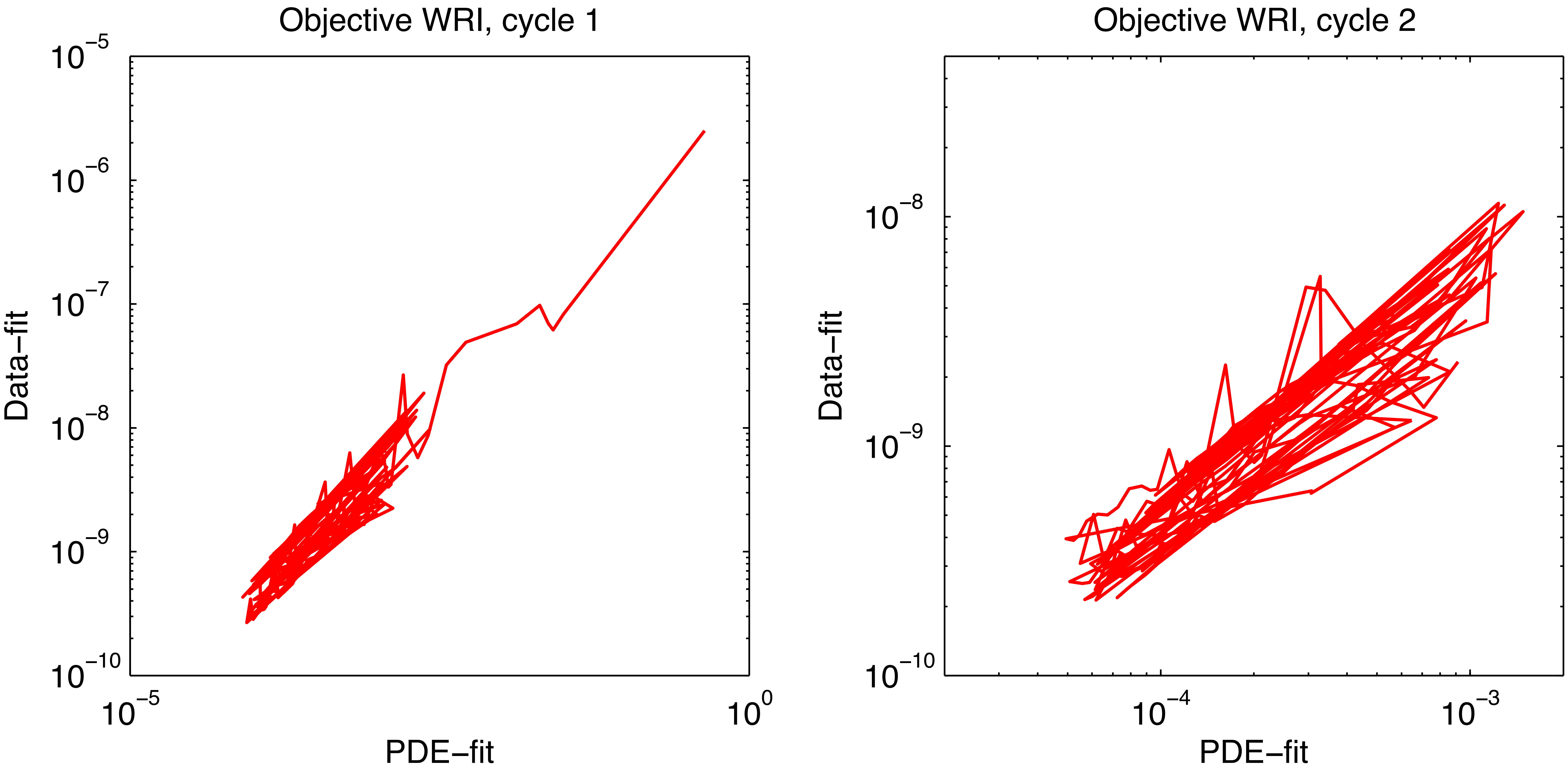


0
500
1000
1500
z [m]
0
1000
2000
3000
4000
5000
x [m]
4500
4000
3500
3000
2500
2000
1500
Velocity [m/s]

Second sweep



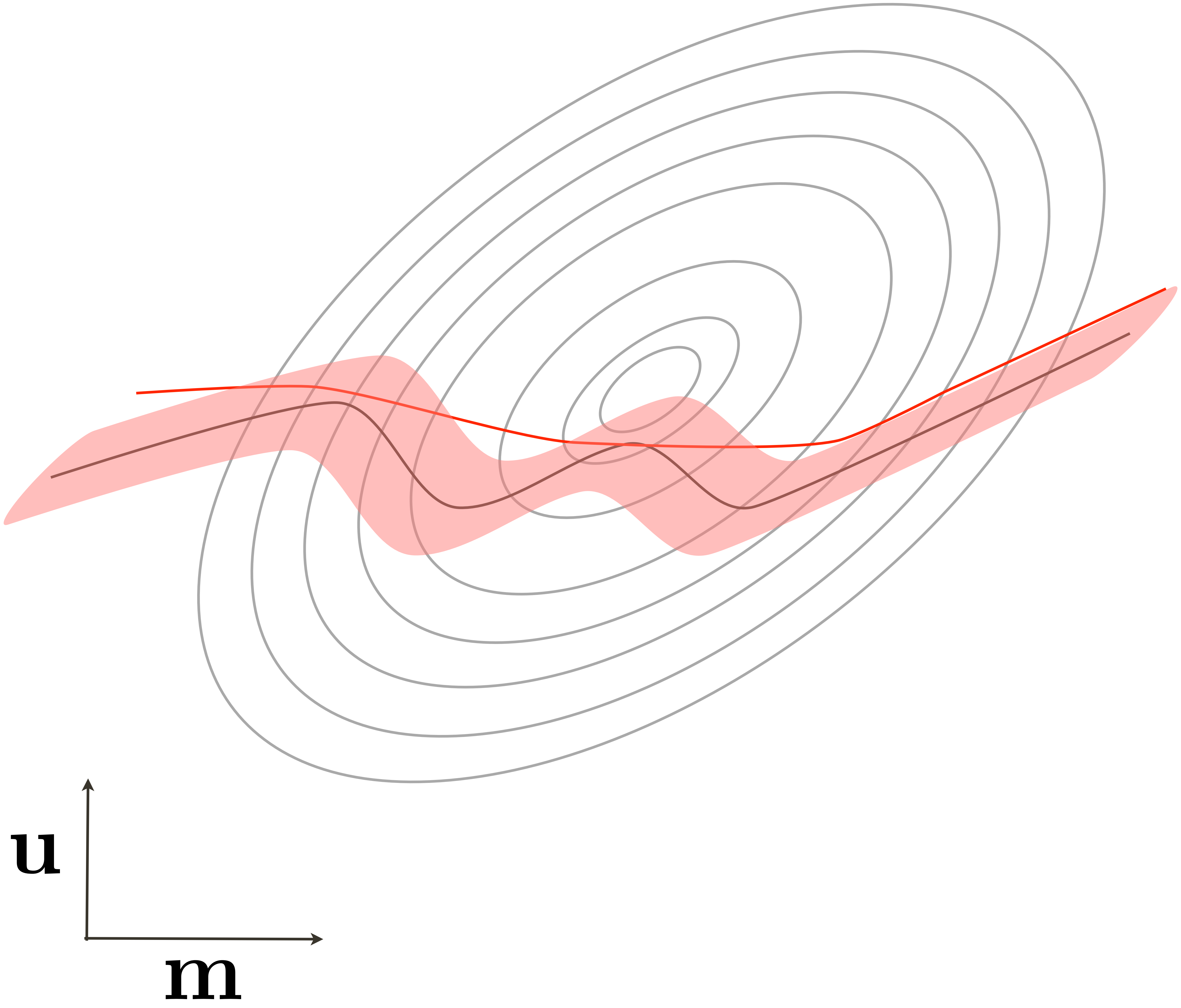
Objective function value



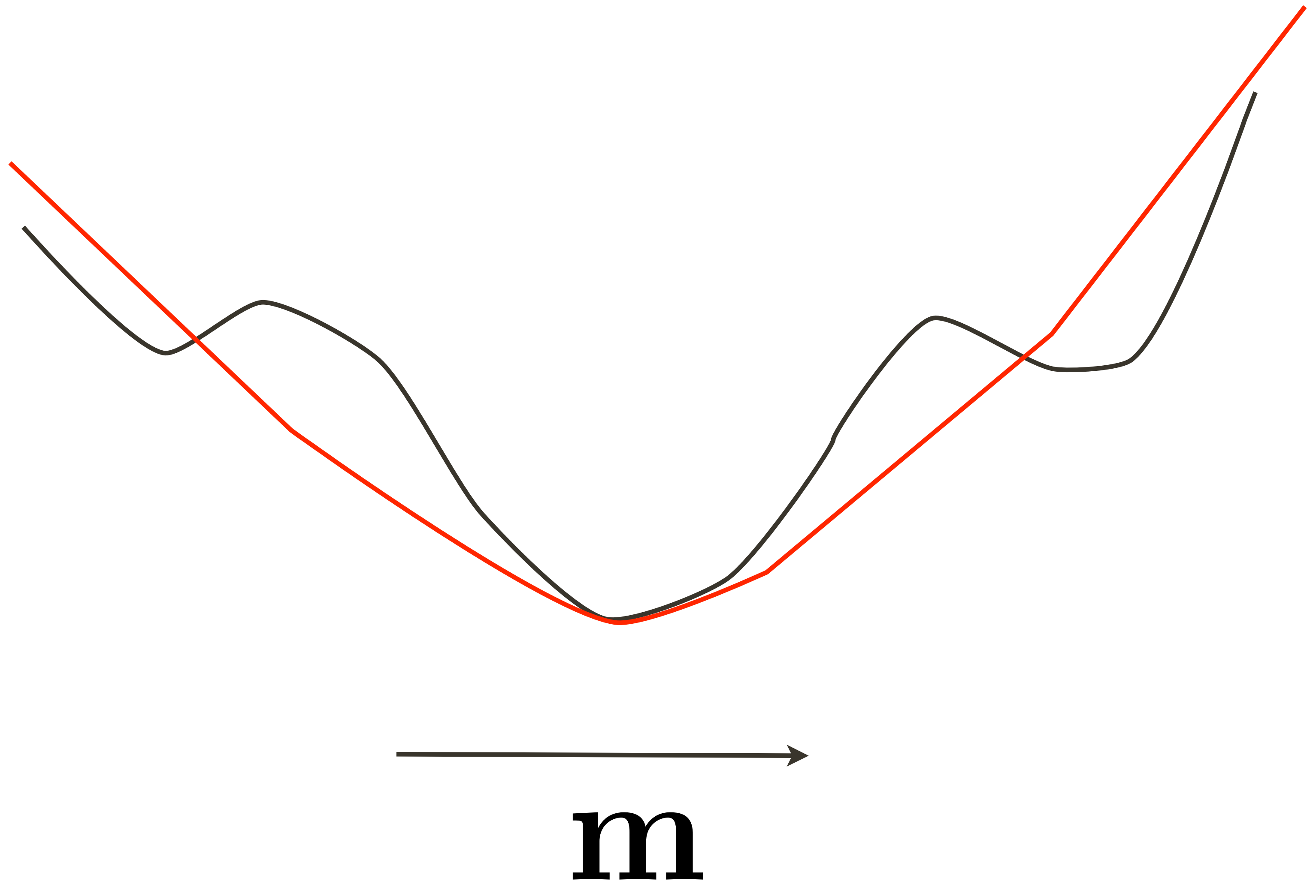
Data fit increases at some iterates

WRI vs. FWI

Larger # of degrees of freedom



"more convex"



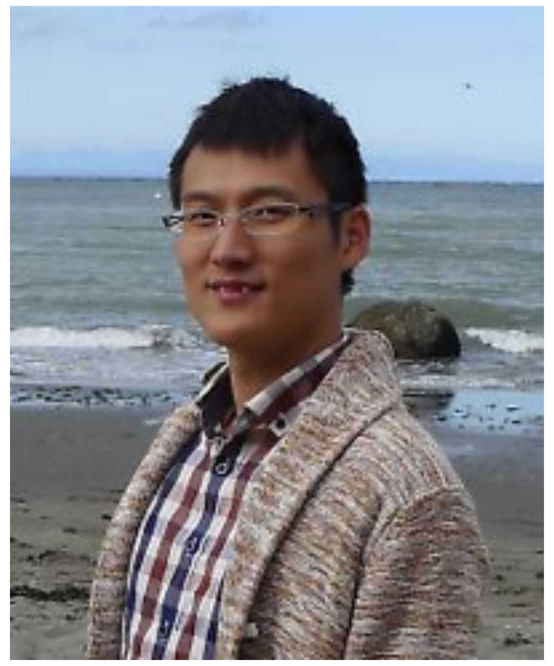
[van Leeuwen & FJH, ’13; ’15, Peters et. al. ’14]

Chevron blind test data

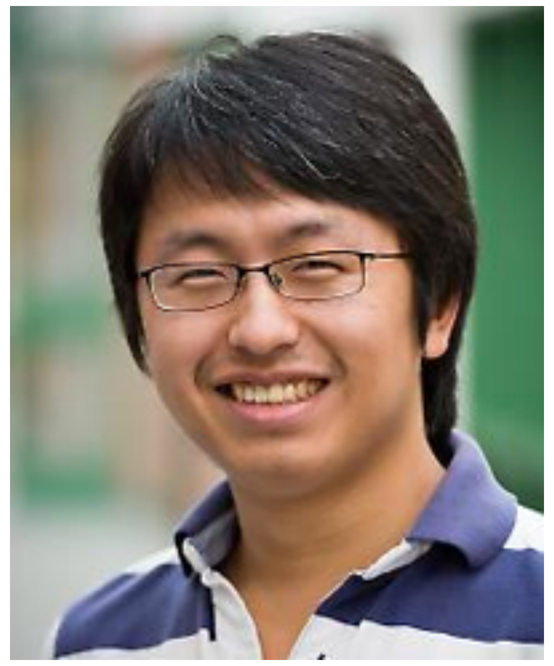
Inversion strategy:

1. Frequency domain WRI with Source estimation;
2. Frequency bands: [3:0.2:5]Hz, [3:0.2:7]Hz, [3:0.2:9]Hz, [3:0.2:11]HzHz;
3. Batch sizes of random frequency subsets: 3, 6, 10, 10;
4. Batch size of random source subsets: 300;
5. Optimization solver: l-BFGS with 30 iterations per frequency band;
6. 2passes of WRI from frequency 3-11 Hz;
7. Grid size: 20m;
8. Minimum offset used: 1000m;
9. **No pre-processing !!!**

Chevron blind test data



Zhilong Fang



Xiang Li

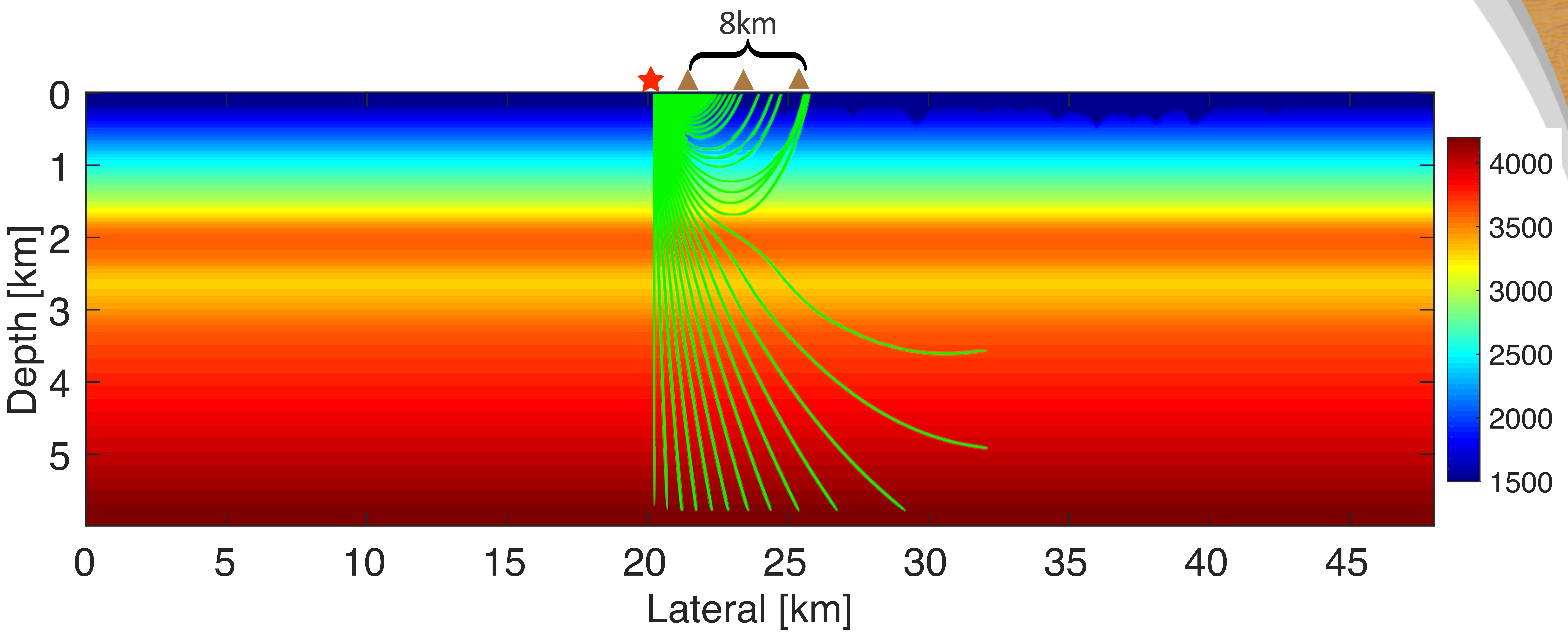
Bas Peters

Brendan Smithyman

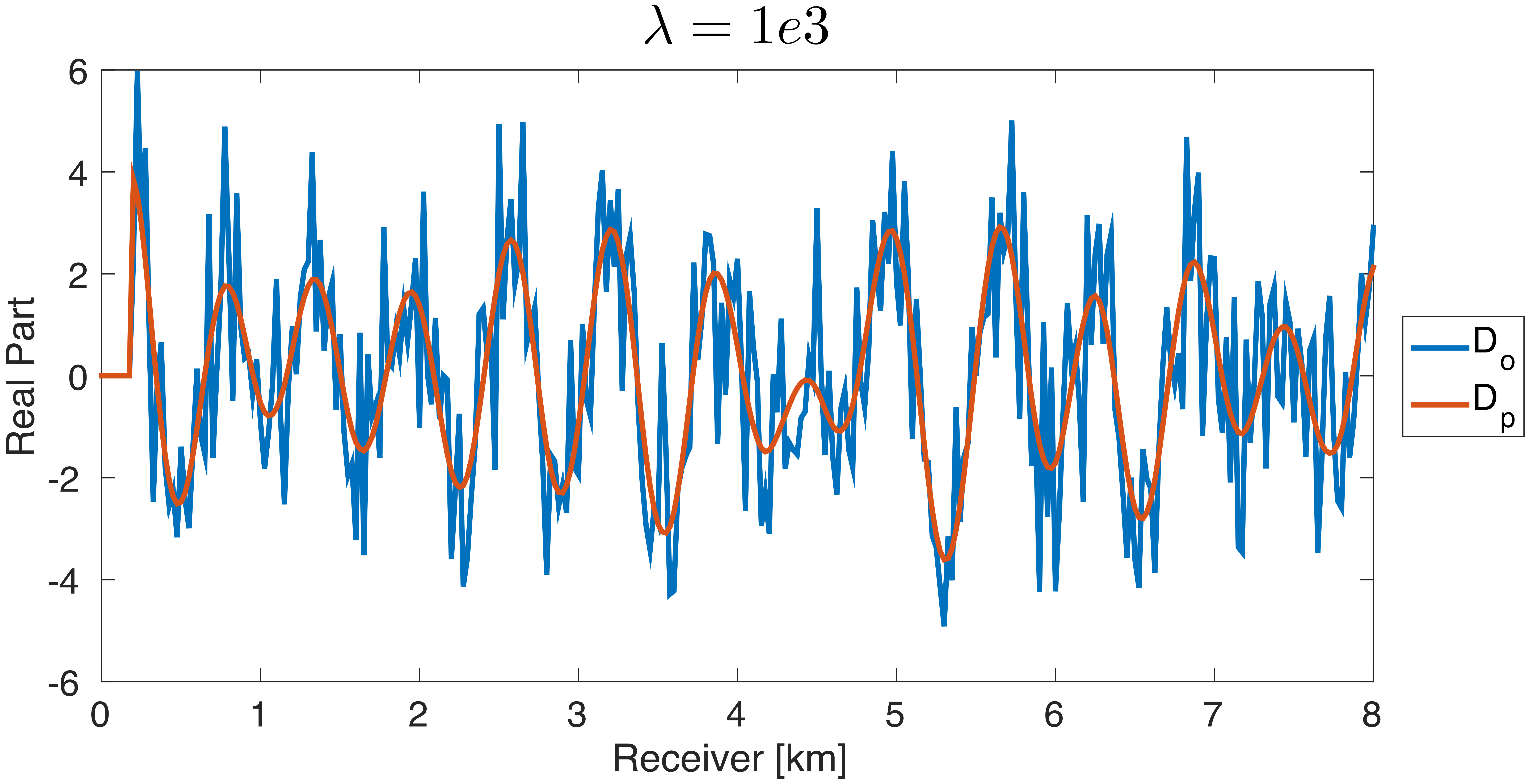
Mengmeng Yang

Felix J. Herrmann

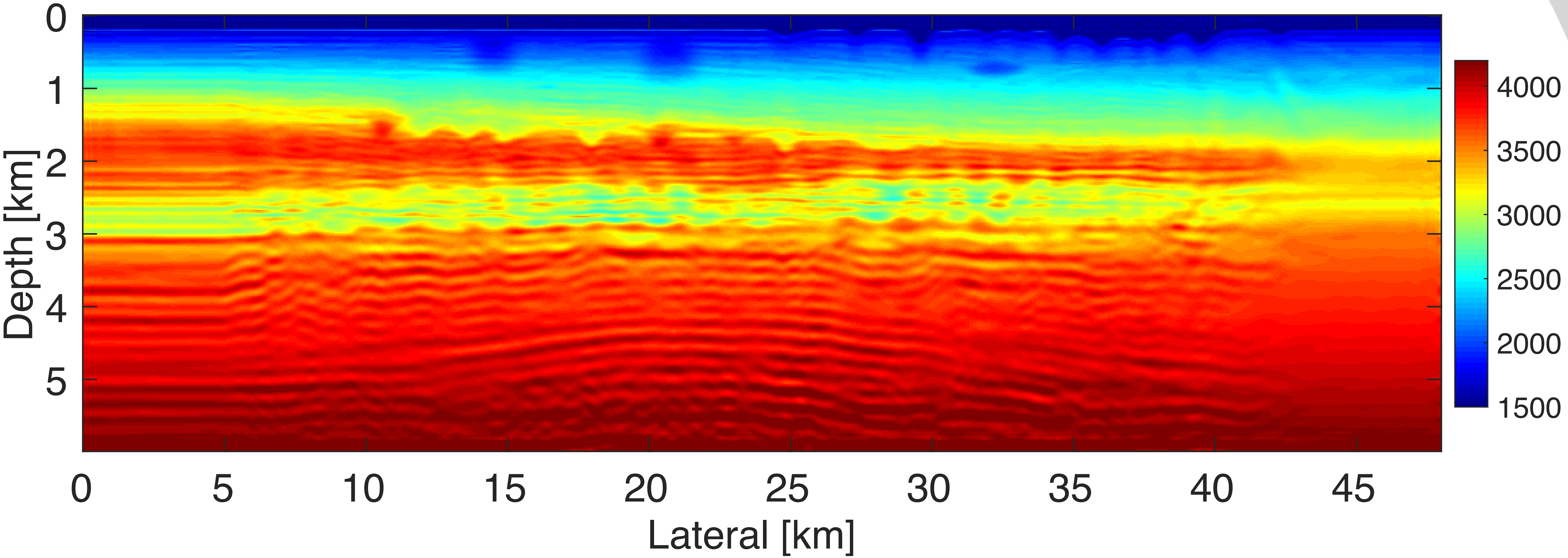
Initial model



Initial data fit
— @3 Hz & shot 800

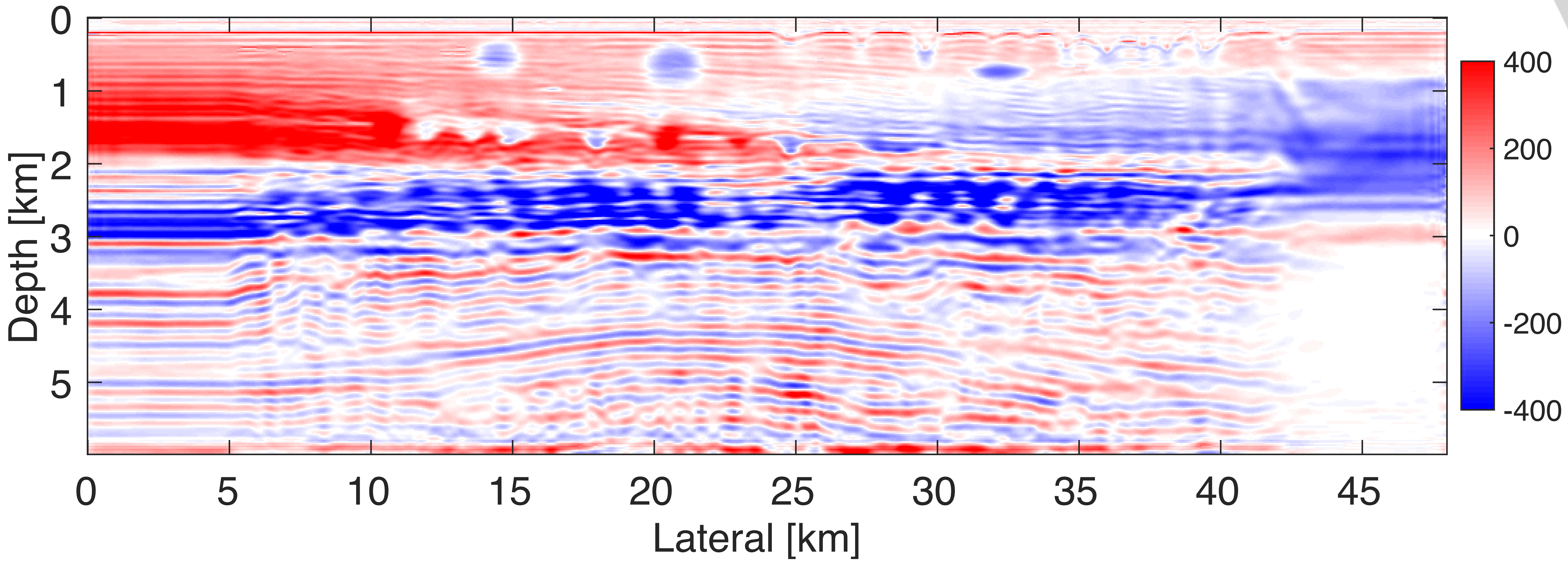


WRI

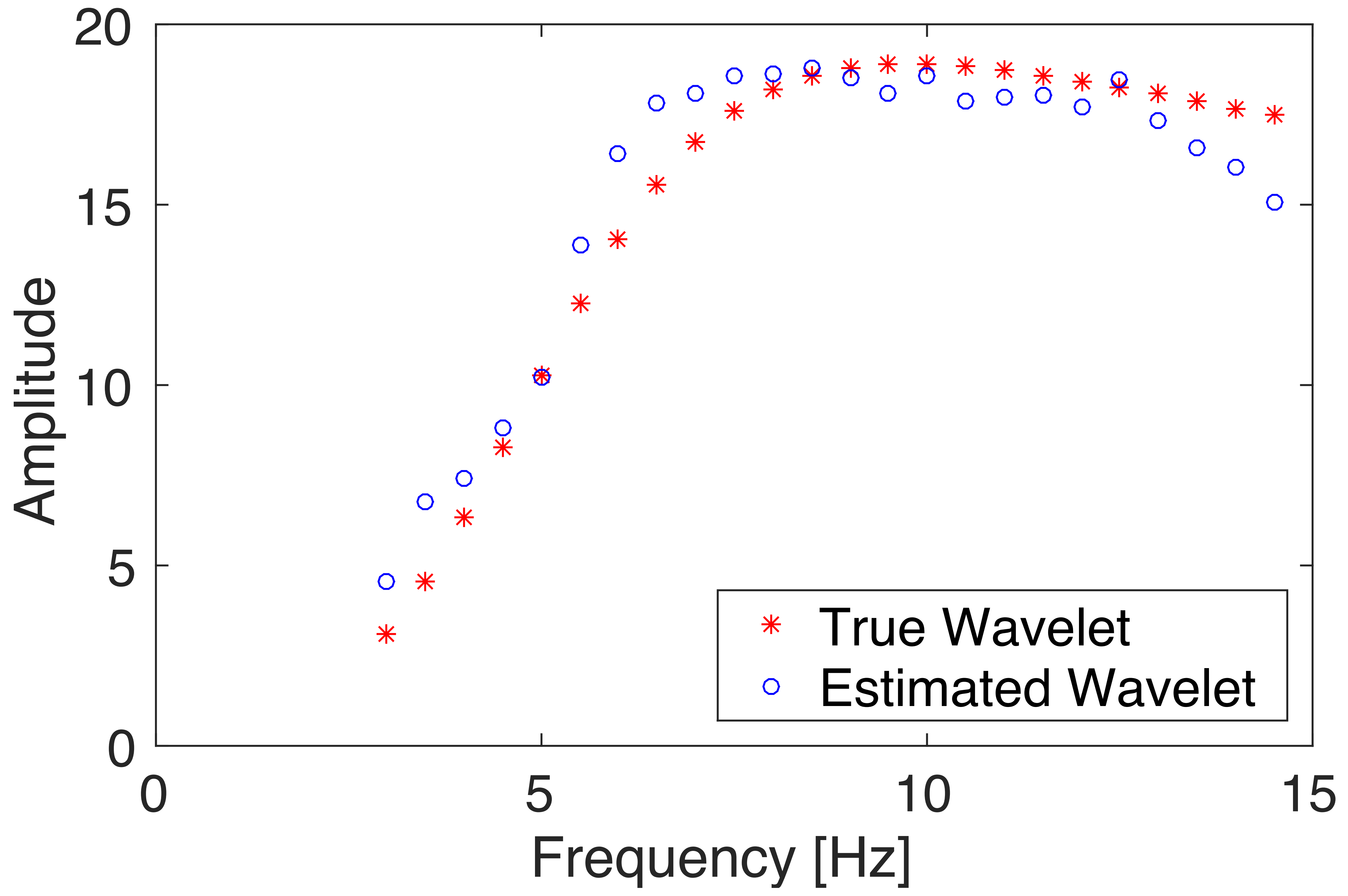


Depth [km]
0
1
2
3
4
5
0
5
10
15
20
25
30
35
40
45
Lateral [km]
4000
3500
3000
2500
2000
1500

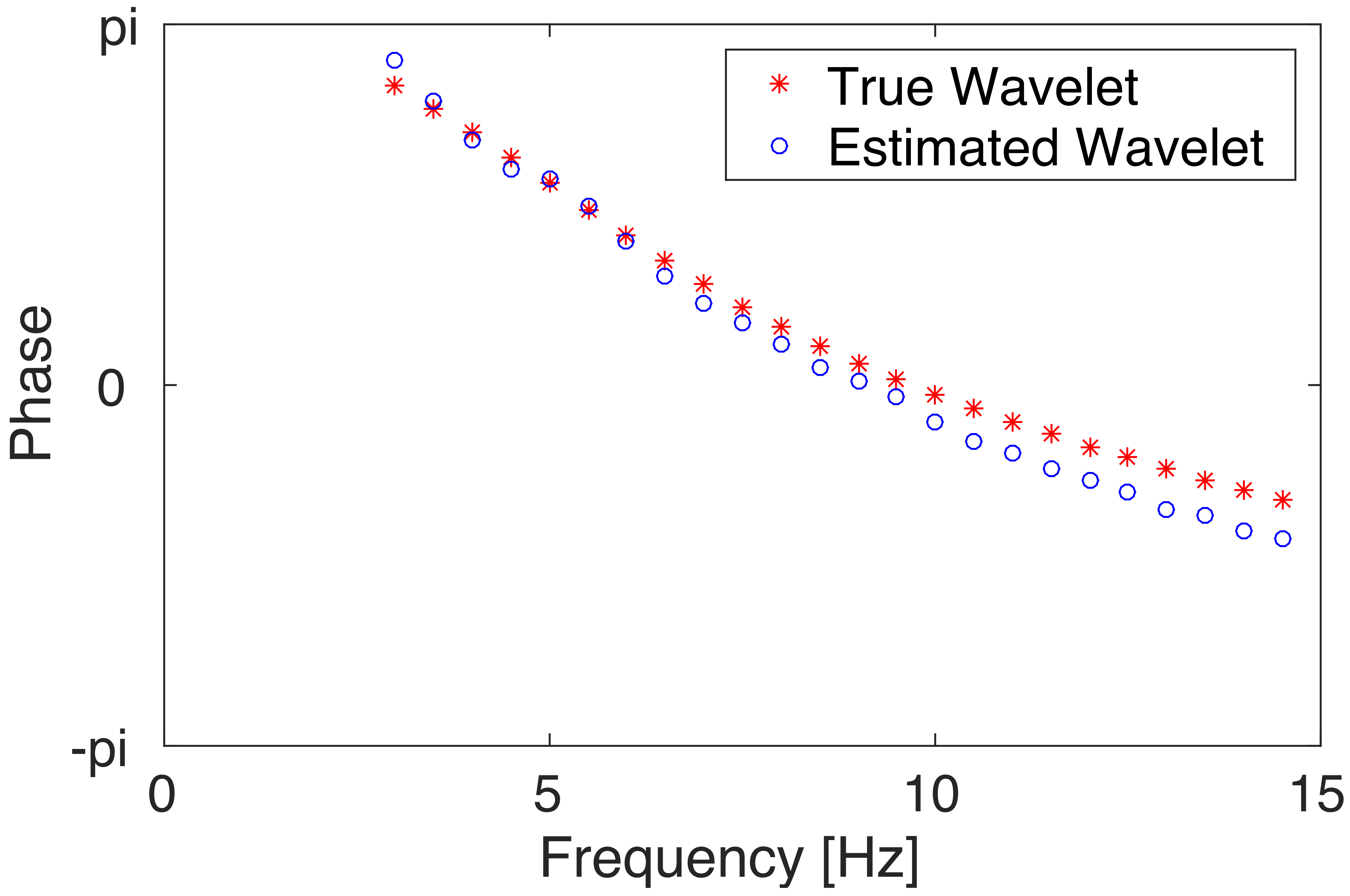
Model update



Source wavelet comparison

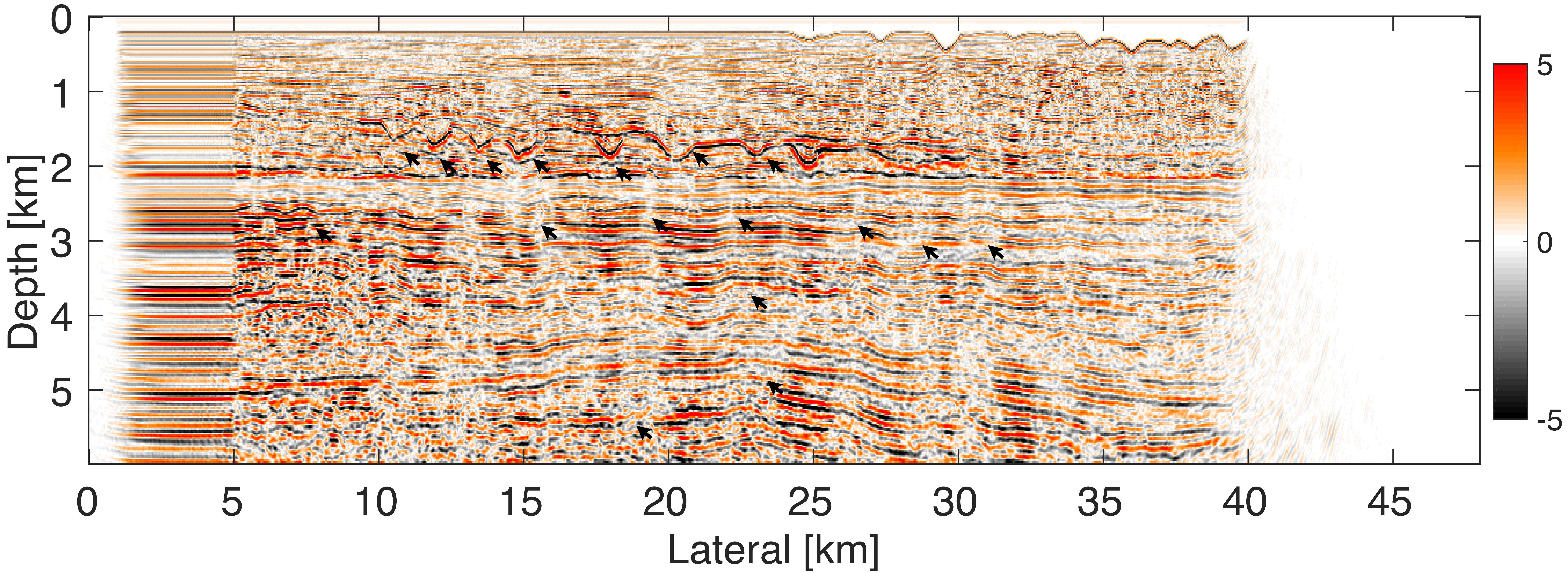


Amplitude

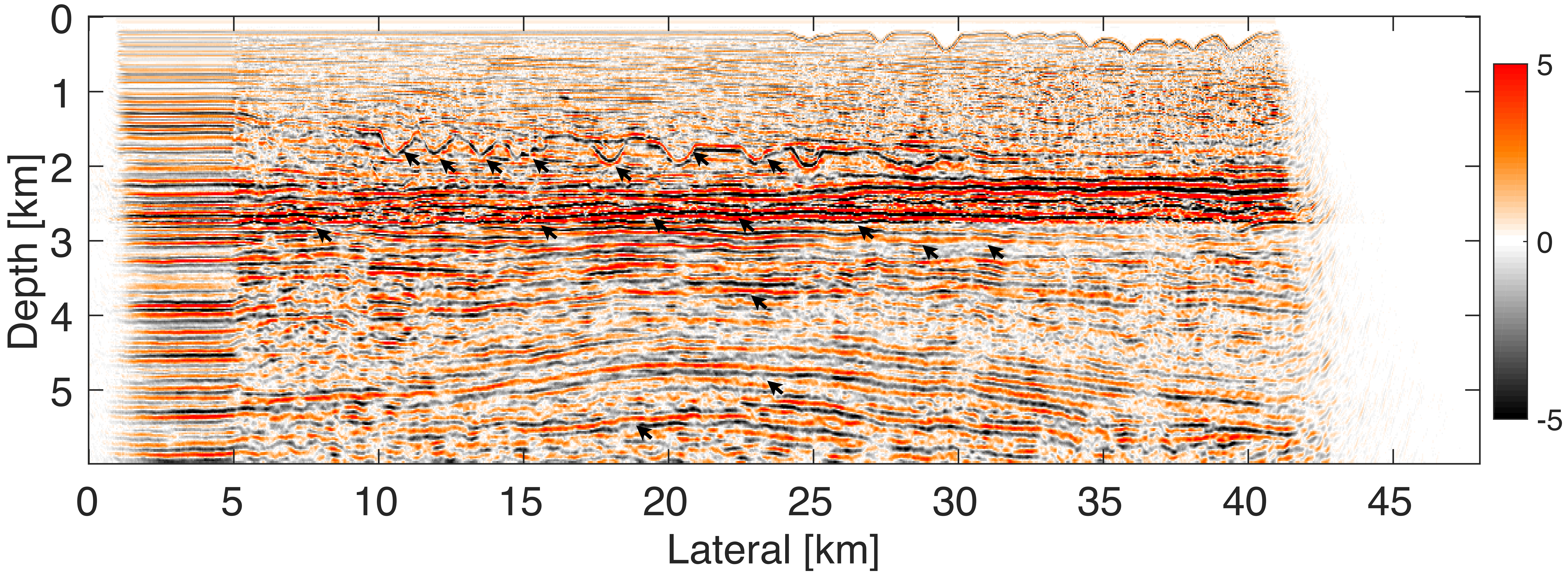


Phase

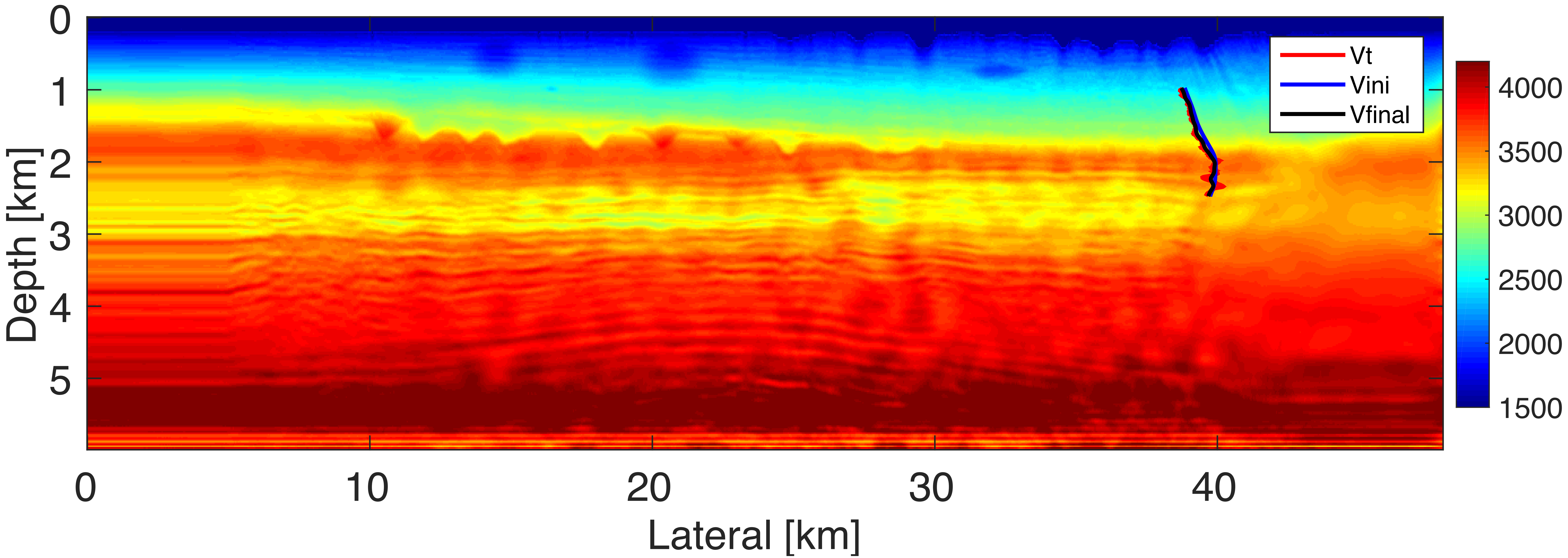
Kirchhoff migration
—Initial model



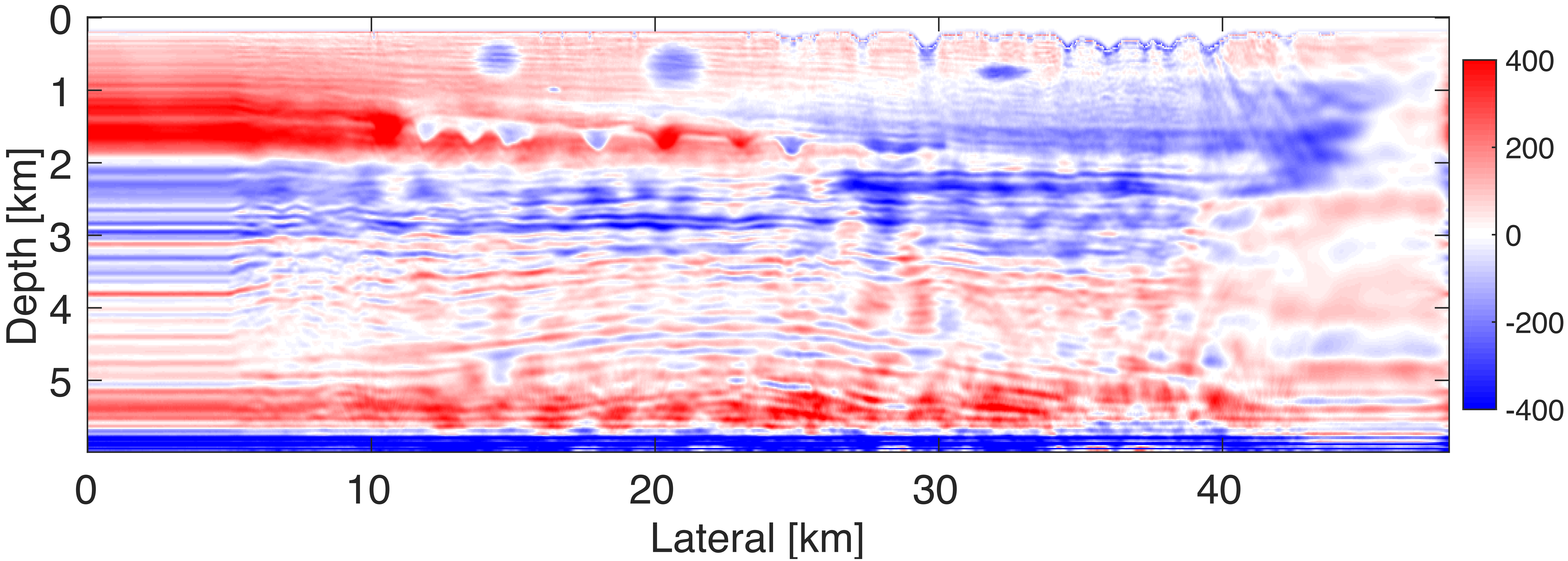
Kirchhoff migration
—Inversion result



WRI w/ density & curvelet sparsity



Model update



Observations

WRI obtains a reasonable inversion results for the velocity & source

No data preprocessing or extensive handholding needed

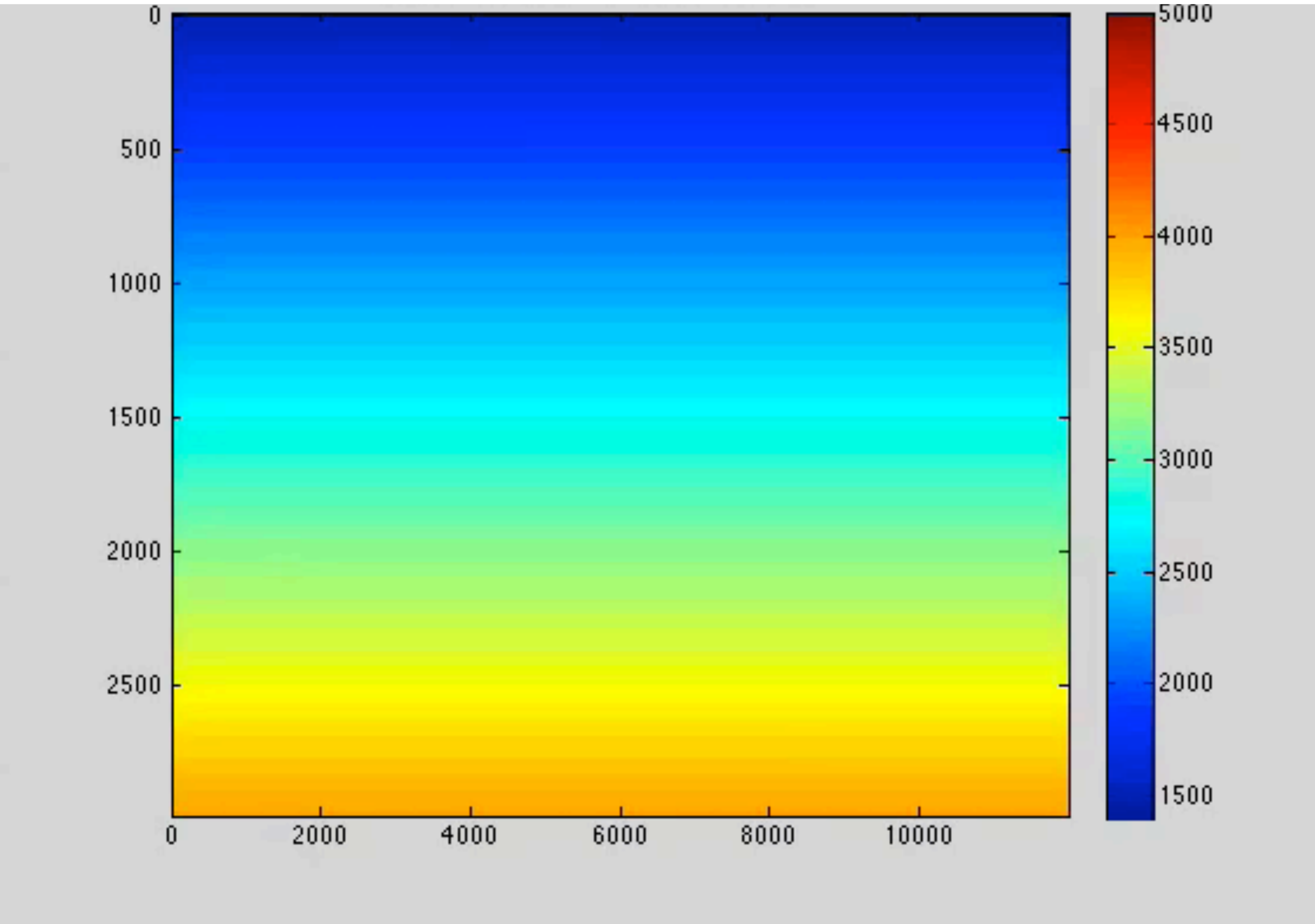
How does this method hold up in cases where we have high-contrast high-velocity unconformities such as salt?

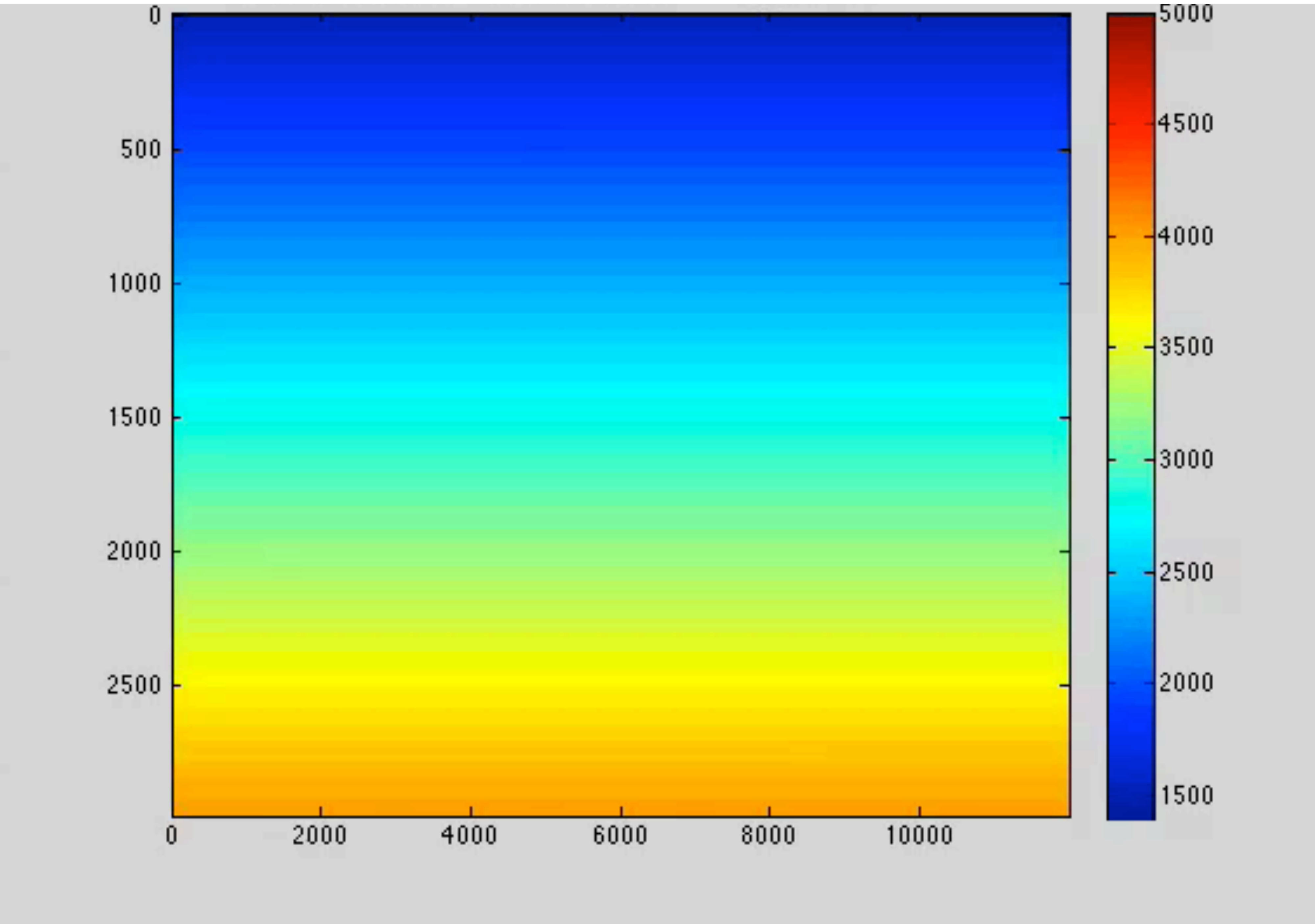
BP benchmark

- **Challenge:** high-contrast & high-velocity unconformity
- ***Added convex constraints***
- Isotropic acoustic “inverse-crime” data w/ known 15 Hz Ricker wavelet
- Invert for slownesses w/ acoustic kernel
- Frequency bands: 3–20Hz in overlapping batches of 2
- Number of sources: 126; number of receivers: 299
- Maximum number of outer iterations per frequency batch: 25
- Maximum number of inner iterations for convex subproblems: 2000
- 8 frequency sweeps that relax the constraints

WRI

w/ or w/o convex constraints





Conclusions

New method for wave-equation based inversion:

- benefits from same extended search space as in *all-at-once* but w/ memory & CPU requirements of adjoint-state approaches
- fits the observed data by design & is "less non-linear"
- therefore less susceptible to local minima

Experiments show that WRI succeeds where FWI fails because it uses

- reflected energy to invert for low-velocity kickbacks
- convex constraints by virtue of the (near) diagonal GN Hessian

Candidate for "automatic" salt flooding...

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, DownUnder GeoSolutions, Hess, Petrobras, PGS, Subsalt Ltd, WesternGeco, and Woodside.