Noise robust and time-domain formulations of Wavefield Reconstruction Inversion

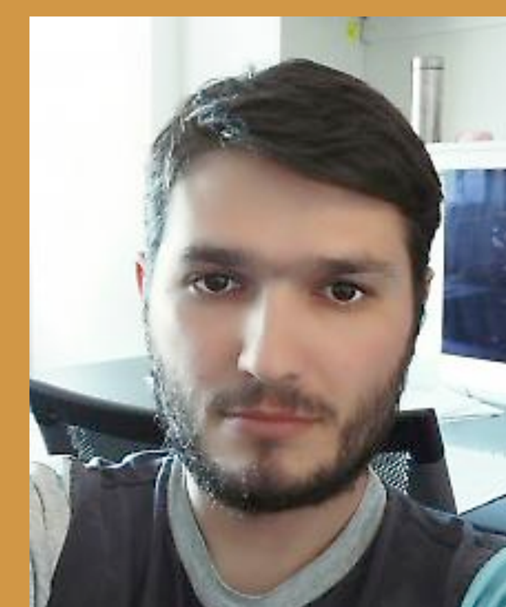
Felix J. Herrmann

SLIM
University of British Columbia

Noise robust and time-domain formulations of Wavefield Reconstruction Inversion

Mathias Louboutin, Peter Bas, Rongrong Wang, Emmanouil Daskalakis



SLIM
University of British Columbia

Motivation

Full-waveform inversion (FWI):

- hampered by poor data & parasitic local minima
- ill-posed ⟷ missing frequencies & finite aperture
- high-contrast high-velocity inclusions
- noise in data & errors in modeling

Notoriously difficult inverse problem:

- non-convex
- extremely large scale

Heuristic strategy

Extend & Project

- avoid local minima via extensions
- variable project to fit data

Squeeze

- impose physics & constrain the model

Cycle & Relax

- do multiple warm restarts while relaxing constraints

Heuristic strategy

Extend & Project

- avoid local minima via extensions
- variable project to fit data

Squeeze

- impose physics & constrain the model

Cycle & Relax

- do multiple warm restarts while relaxing constraints

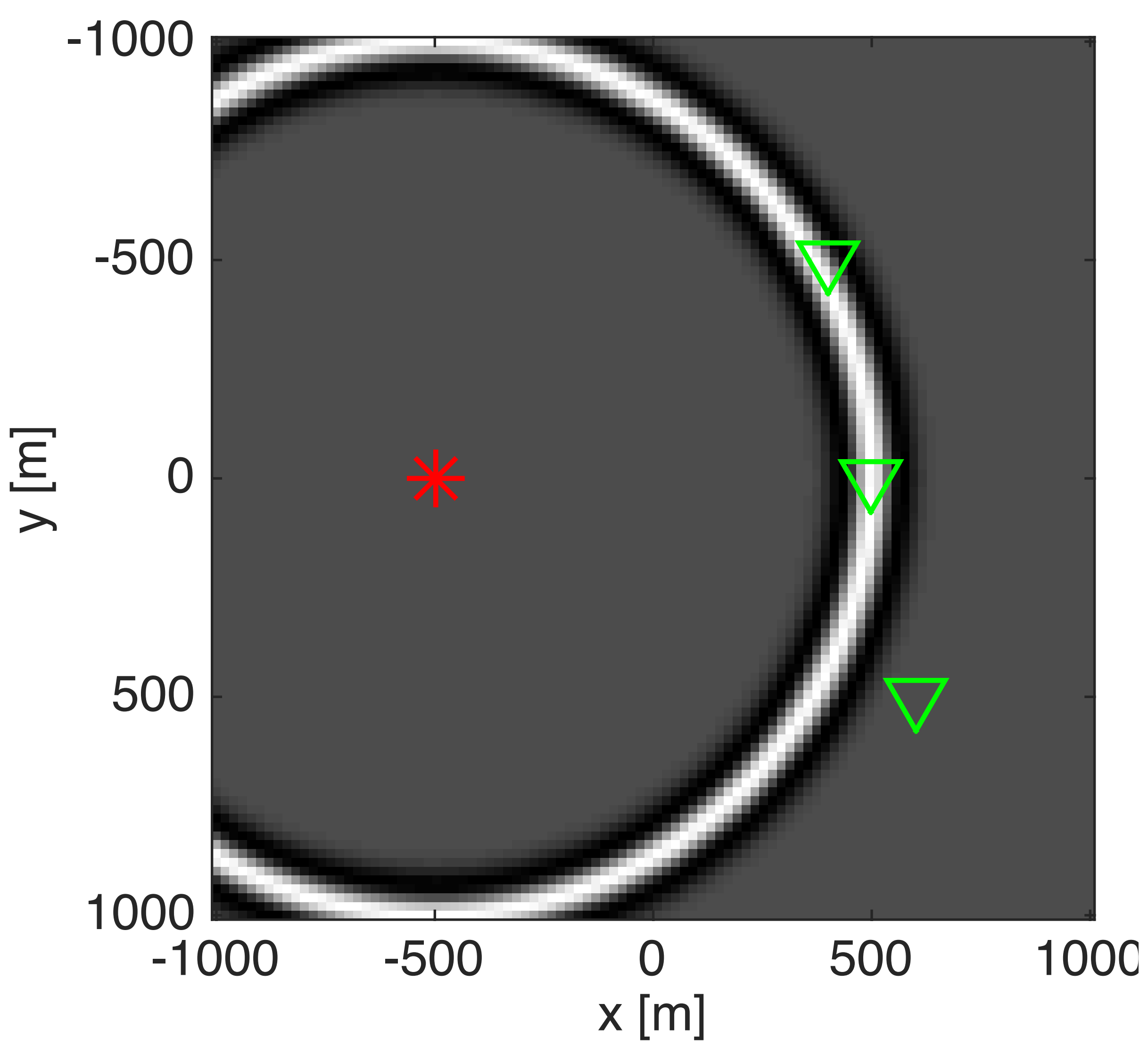
No, this is not a yoga class...

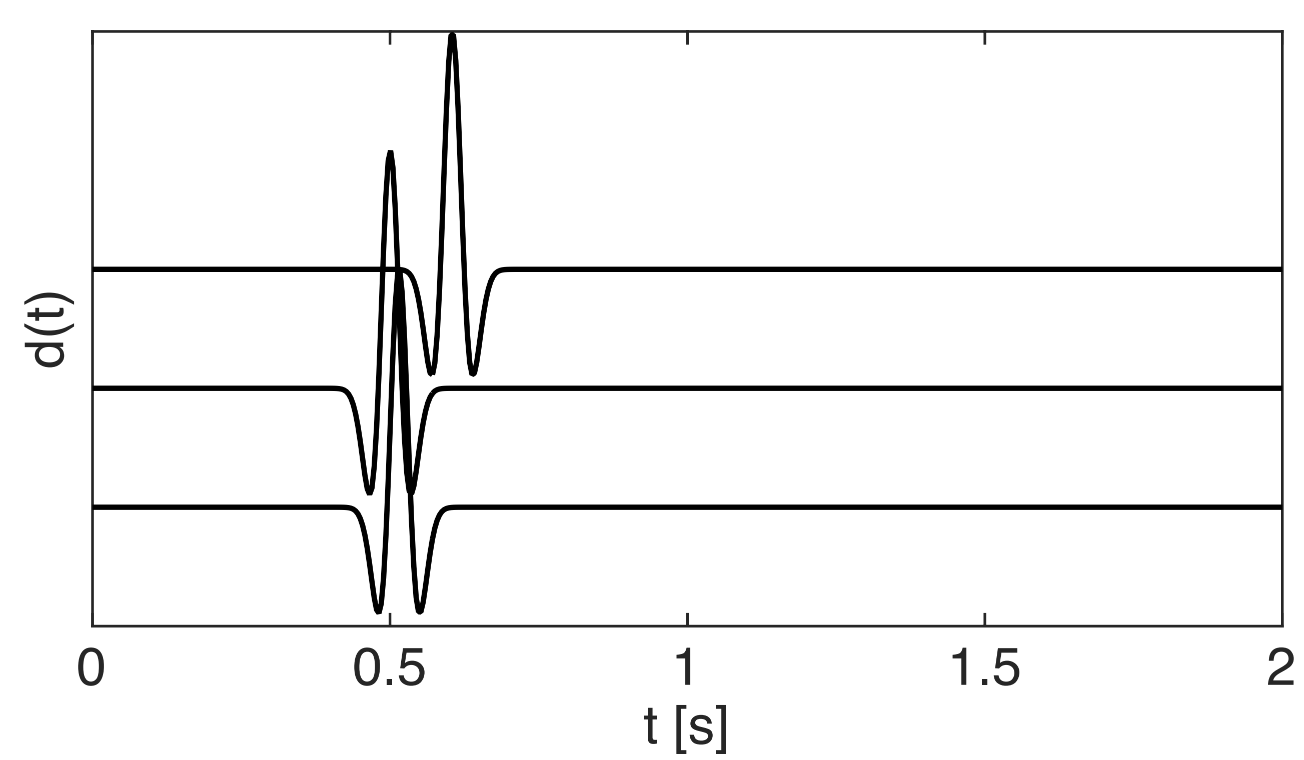
SLIM

Stylized example

Consider
$$F(c)q = \begin{pmatrix} \mathcal{F}^{-1} e^{\imath\omega\|x_1 - x_s\|_2/c} \mathcal{F} q \\ \mathcal{F}^{-1} e^{\imath\omega\|x_2 - x_s\|_2/c} \mathcal{F} q \\ \mathcal{F}^{-1} e^{\imath\omega\|x_3 - x_s\|_2/c} \mathcal{F} q \end{pmatrix}$$

- seems harmless
- **not so** – oscillatory because of missing low frequencies





Stylized example – extension

Replace $$\min_{m} \frac{1}{2} \sum_{j=1}^{N_s} \|F(m)q_j - d_j\|_2^2$$

by $$\min_{m,\Delta q} \frac{1}{2} \sum_{j=1}^{N_s} \|F(m)(q_j + \Delta q_j) - d_j\|_2^2 + \lambda^2 \|\Delta q_j\|_2^2,$$

with $\Delta q = [\Delta q_1; \Delta q_2, \ldots, \Delta q_{N_s}]$ slack variables

- coincides with original solution since

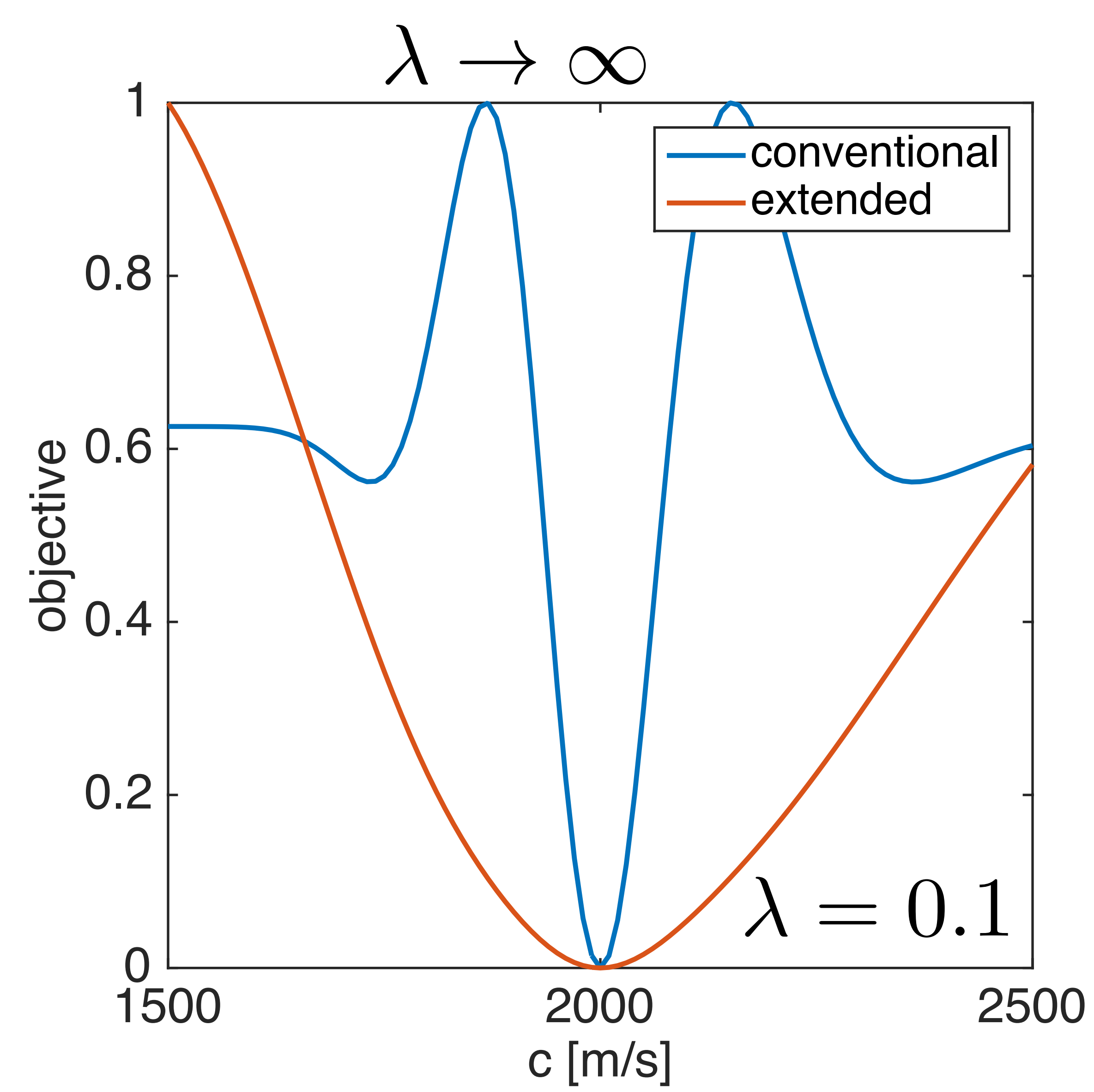
$$\lambda \uparrow \infty, \ \|\Delta q\|_2 \downarrow 0$$

Proxy for extensions of AWI, WRI, and their variants...

Stylized example – extension

Solve by **projecting out** the slack variables

- modify objective for model parameters
- avoids cycle skips



Tristan van Leeuwen and Felix J. Herrmann, "A penalty method for PDE-constrained optimization in inverse problems", *Inverse Problems*, vol. 32, p. 015007, 2015.

Guanghui Huang, William Symes, and Rami Nammour, "Matched source waveform inversion: Space-time extension" SEG Technical Program Expanded Abstracts 2016. September 2016, 1426-1431

Extensions

Wavefield Reconstruction Inversion:

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|A(\mathbf{m})\mathbf{u}-\mathbf{q}\|_2^2$$

- weak constraints
- "analysis" form

Matched Source Waveform Inversion:

$$\min_{\mathbf{m},\mathbf{q}} \frac{1}{2}\|PA^{-1}(\mathbf{m})\mathbf{q}-\mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|W\mathbf{q}\|_2^2$$

- weighted "synthesis" form & weights W focusses

Tristan van Leeuwen and Felix J. Herrmann, "**A penalty method for PDE-constrained optimization in inverse problems**", *Inverse Problems*, vol. 32, p. 015007, 2015.

Guanghui Huang, William Symes, and Rami Nammour, "**Matched source waveform inversion: Space-time extension**" SEG Technical Program Expanded Abstracts 2016. September 2016, 1426-1431

SLIM

Extensions

Wavefield Reconstruction Inversion:

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|A(\mathbf{m})\mathbf{u}-\mathbf{q}\|_2^2$$

- weak constraints
- "analysis" form

Matched Source Waveform Inversion:

$$\min_{\mathbf{m},\mathbf{q}} \frac{1}{2}\|PA^{-1}(\mathbf{m})\mathbf{q}-\mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|W\mathbf{q}\|_2^2$$

- weighted "synthesis" form & weights W focusses

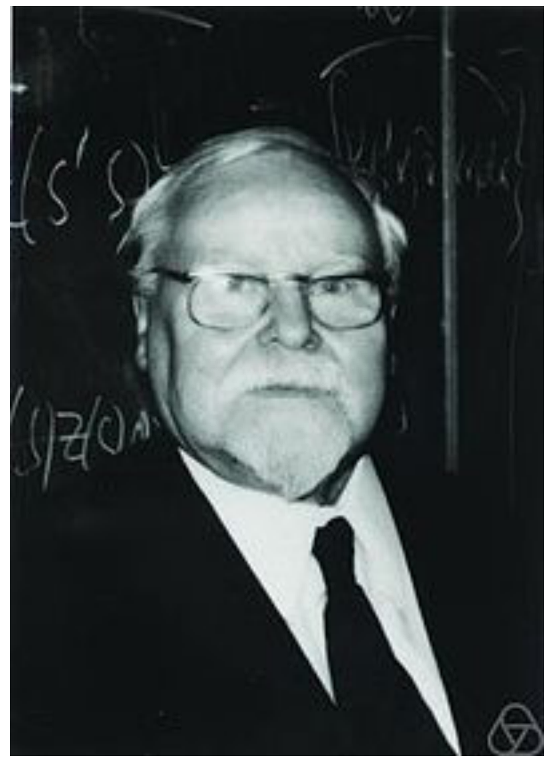
Challenge is to set the Lagrange multiplier!

Tikhonov regularization

Add quadratic penalty terms:

$$\underset{\mathbf{m}}{\text{minimize}}\, f(\mathbf{m}) + \frac{\alpha}{2}\|R_1\mathbf{m}\|^2 + \frac{\beta}{2}\|R_2\mathbf{m}\|^2$$

- well-known & successful technique
- is differentiable
- not an exact penalty
- gradient & Hessian may become ill-conditioned
- requires non-trivial choices for hyper parameters
- not easily extended to edge-preserving ℓ_1 - norms & bound constraints
- **no guarantees that all model iterates are regularized**



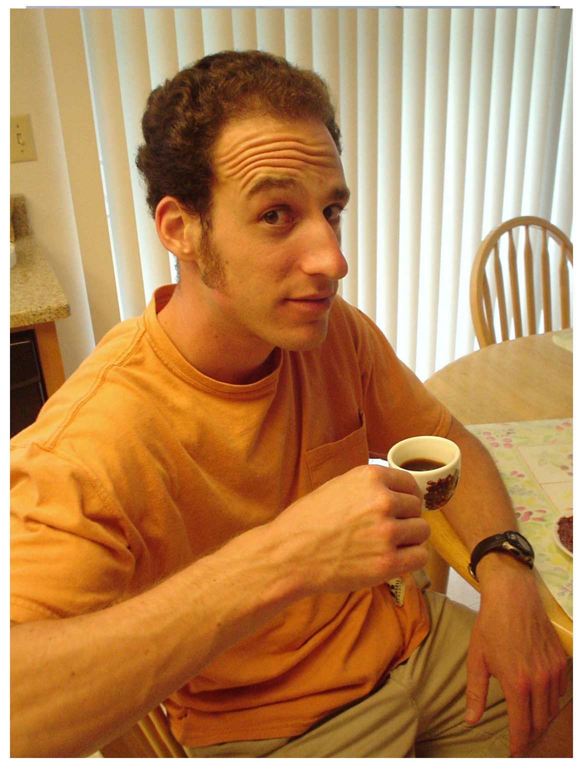
Andrey Tikhonov
1906–1993

Ernie Esser, Xiaoqun Zhang, and Tony F. Chan. A General Frame- work for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science. SIAM Journal on Imaging Sciences, 3(4):1015–1046, 2010.

SLIM

Ernie Esser

$$\underset{\mathbf{m}}{\text{minimize}}\, f(\mathbf{m}) \quad \text{subject to} \quad \mathbf{m} \in \mathcal{C}$$



Ernie Esser
1980 – 2015

Ernie Esser, Xiaoqun Zhang, and Tony F. Chan. A General Frame- work for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science. SIAM Journal on Imaging Sciences, 3(4):1015–1046, 2010.

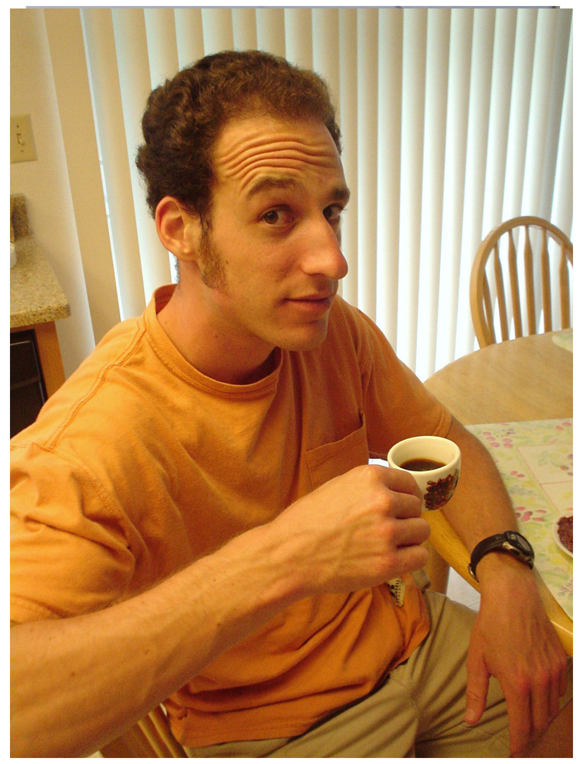
SLIM

Ernie Esser

$$\underset{\mathbf{m}}{\text{minimize}}\, f(\mathbf{m}) \quad \text{subject to} \quad \mathbf{m} \in \mathcal{C}$$

Algorithm 1 A Scaled Gradient Projection Algorithm :

$n = 0$; $m^0 \in C$; $\rho > 0$; $\epsilon > 0$; $\sigma \in (0,1]$;
H symmetric with eigenvalues between $\lambda_H^{\min}$ and $\lambda_H^{\max}$;
$\xi_1 > 1$; $\xi_2 > 1$; $c_0 > \max(0, \rho - \lambda_H^{\min})$;
while $n = 0$ or $\frac{\|m^n - m^{n-1}\|}{\|m^n\|} > \epsilon$
  $\Delta m = \arg\min_{\Delta m \in C - m^n} \Delta m^T \nabla F(m^n) + \frac{1}{2}\Delta m^T (H^n + c_n \mathrm{I})\Delta m$
  if $F(m^n + \Delta m) - F(m^n) > \sigma(\Delta m^T \nabla F(m^n) + \frac{1}{2}\Delta m^T (H^n + c_n \mathrm{I})\Delta m)$
    $c_n = \xi_2 c_n$
  else
    $m^{n+1} = m^n + \Delta m$
    $c_{n+1} = \begin{cases} \frac{c_n}{\xi_1} & \text{if } \frac{c_n}{\xi_1} > \max(0, \rho - \lambda_H^{\min}) \\ c_n & \text{otherwise} \end{cases}$
    Define H^{n+1} to be symmetric Hessian approximation
      with eigenvalues between $\lambda_H^{\min}$ and $\lambda_H^{\max}$
    $n = n + 1$
  end if
end while



Ernie Esser
1980 – 2015

Friday, October 6, 2017

Bello, L., and Raydan, M., 2007, Convex constrained optimization for the seismic reflection tomography problem: Journal of Applied Geophysics, **62**, 158–166

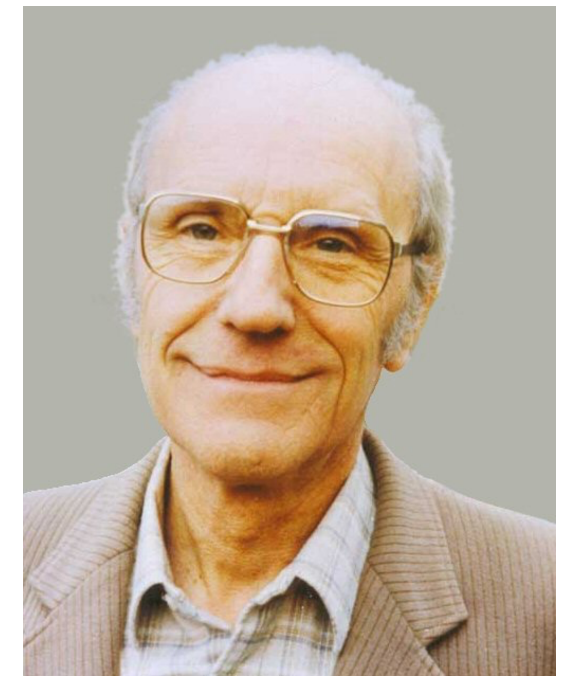
L. Métivier and R. Brossier. The seiscope optimization toolbox: A large-scale nonlinear optimization library based on reverse communication. Geophysics, 81:F11-F25, 2016

Regularization w/ constraints

Add multiple constraints:

$$\underset{\mathbf{m}}{\text{minimize}}\, f(\mathbf{m}) \quad \text{subject to} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

- not well-known in our community
- requires understanding of latest optimization techniques
- does not affect gradient & Hessian
- easier parameterization
- able to uniquely project onto intersection of multiple constraint sets
- constraints do not need to be differentiable
- **constraints are satisfied at *every* model iterate**



Jean Jacques Moreau
1923–2014

Billette, F., and Brandsberg-Dahl, S., 2005

SLIM

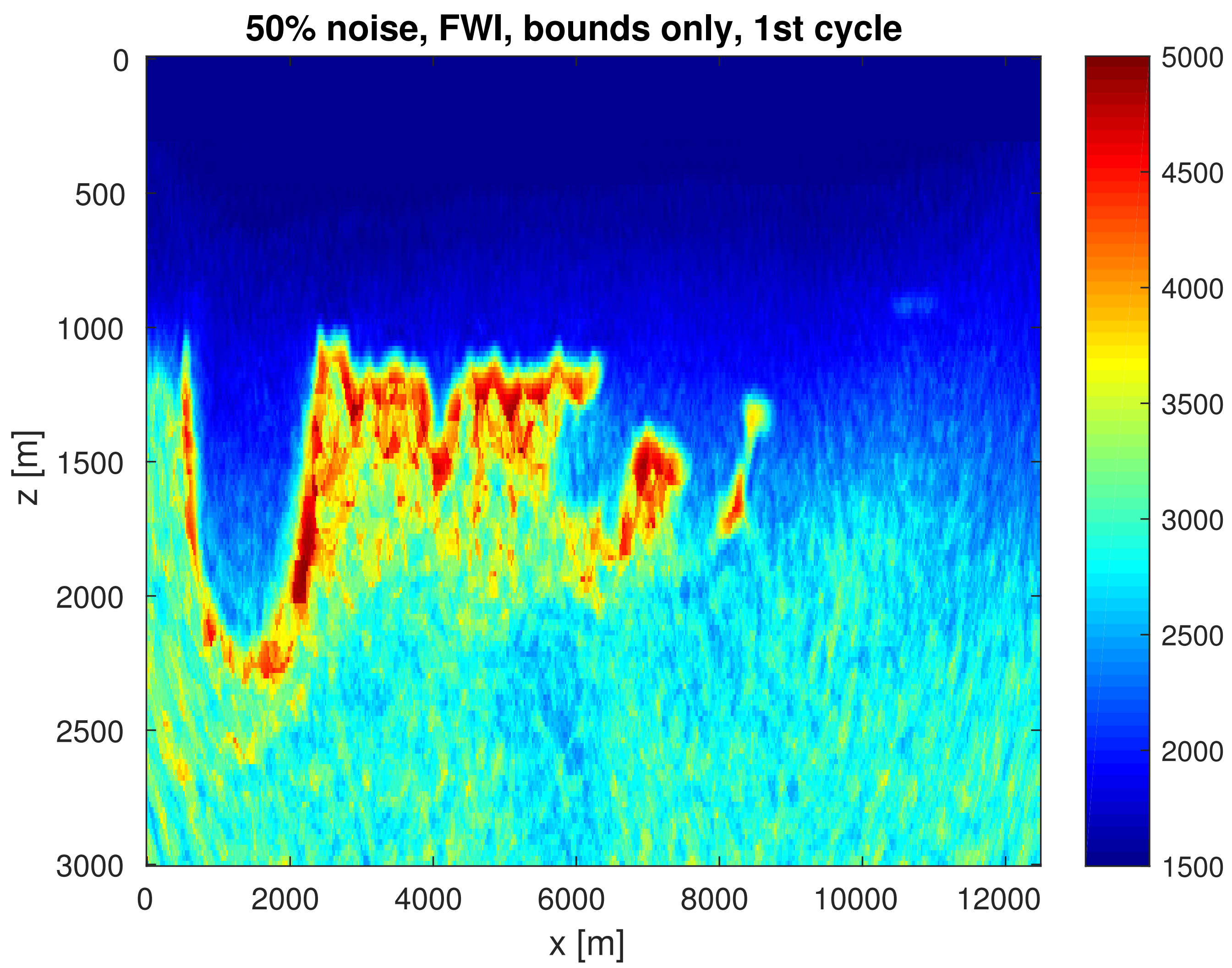
Reduced (2.5 X) BP model – modelling parameters

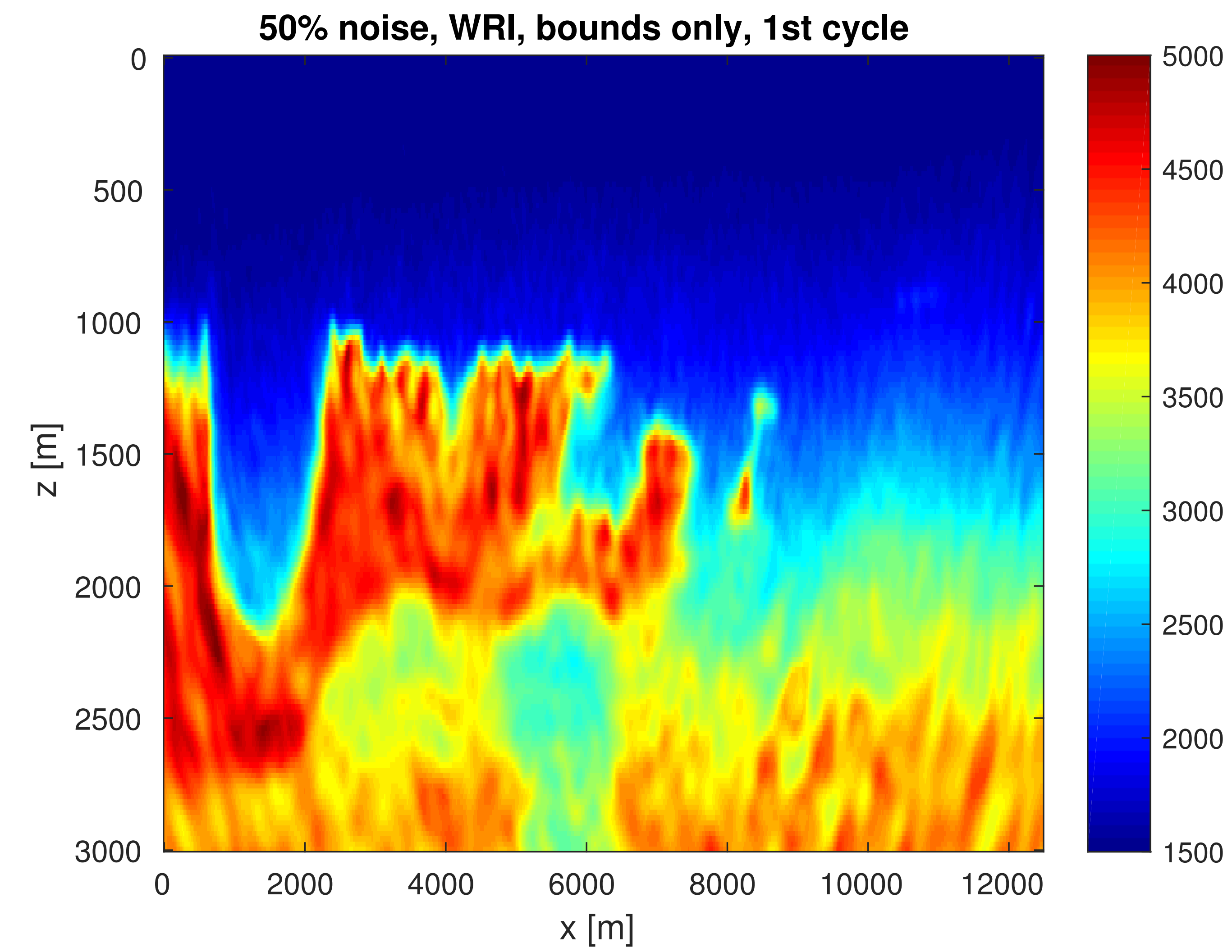
- number of sources: 132; number of receivers: 311
- receiver spacing: 40m, source spacing: 80m, max offset 11.5 km
- grid size: 20 m
- known Ricker wavelet sources with 15Hz peak frequency
- data available starting at 3 Hz
- 8 simultaneous shots w/ Gaussian weights w/ redraws
- starting model = smoothed true model
- inversion crime but poor data $\|\mathrm{noise}\|_2/\|\mathrm{signal}\|_2 = 0.5$

1st cycle cycle

$$\|\text{noise}\|_2 / \|\text{signal}\|_2 = 0.5$$

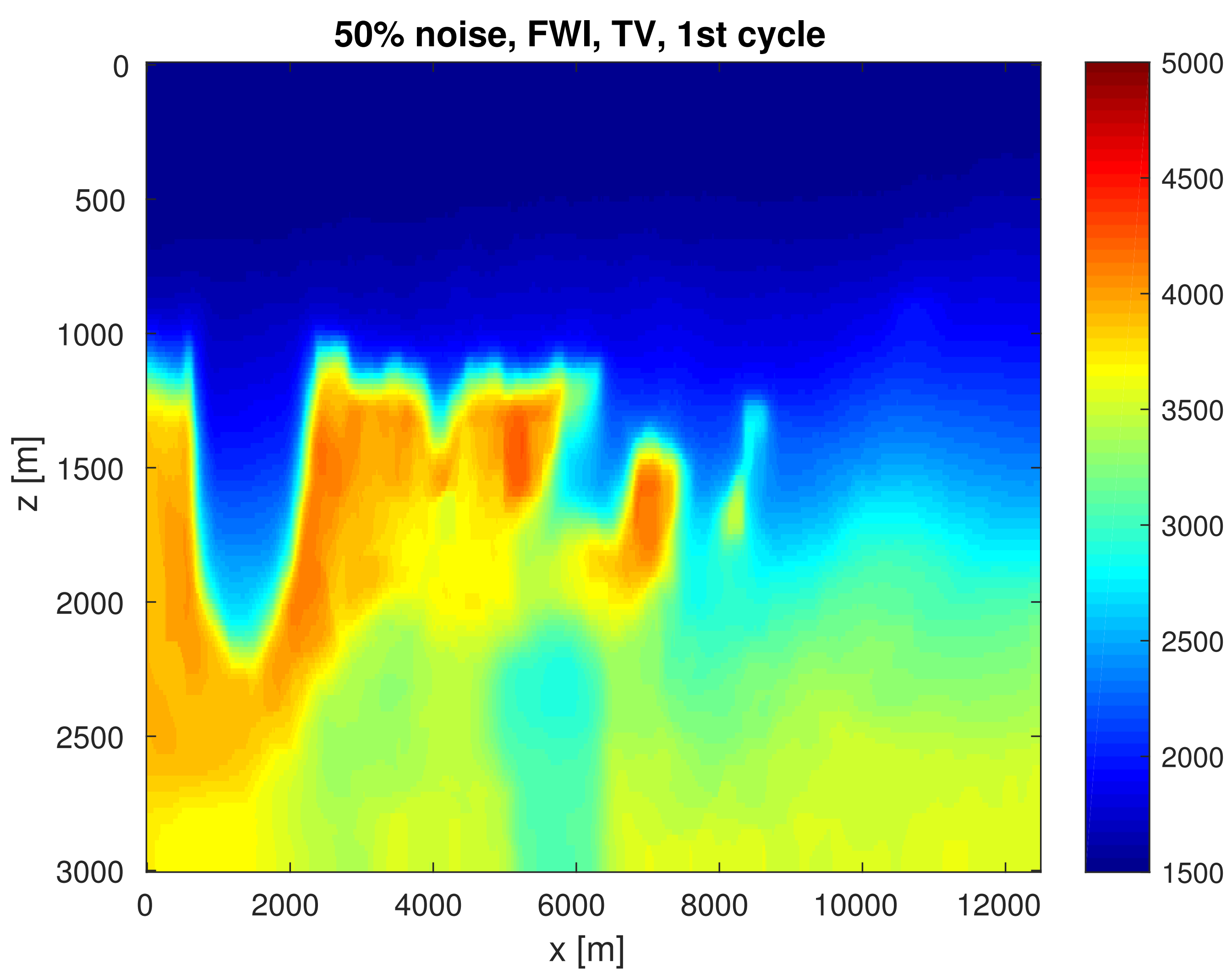
bounds only

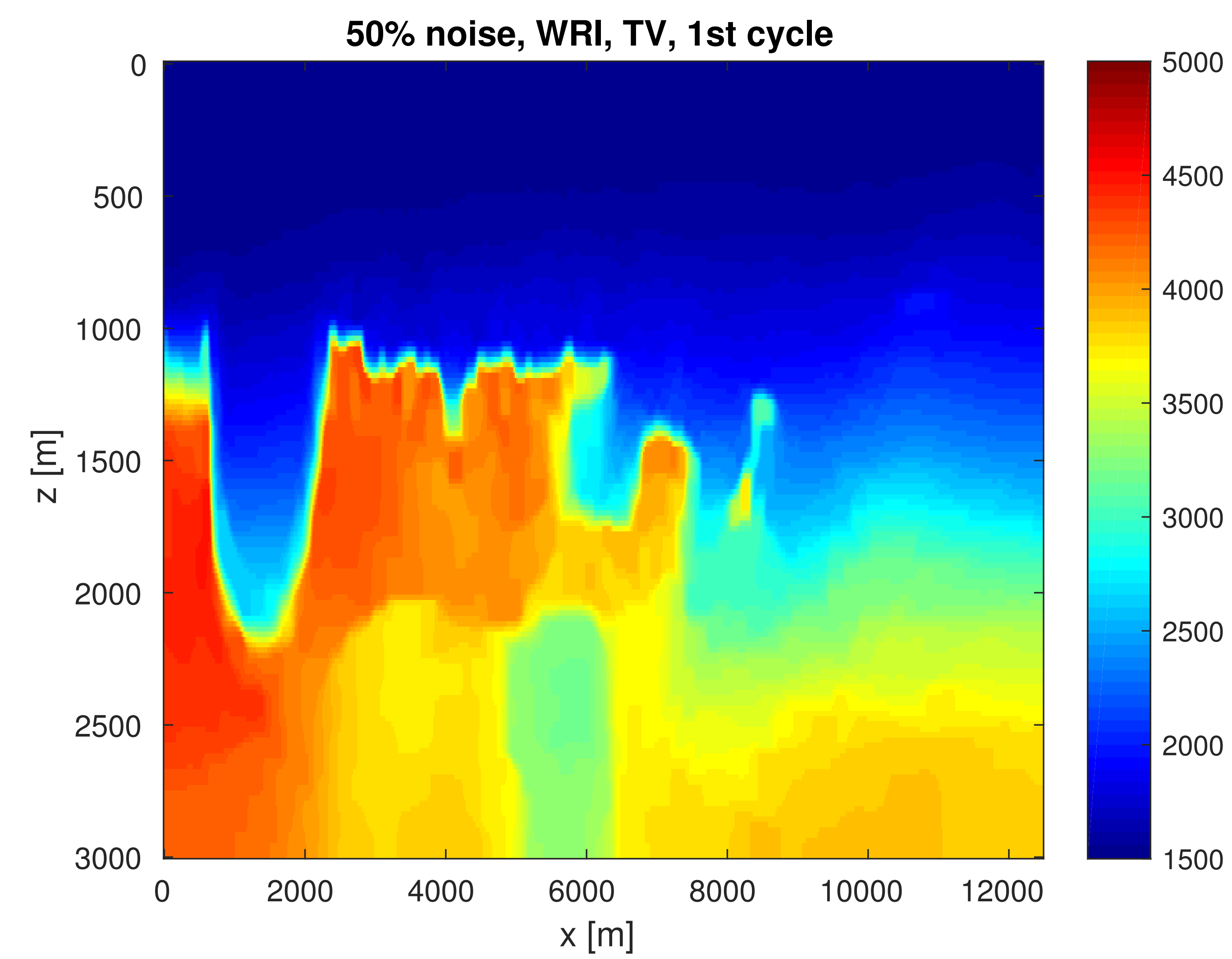




FWI **WRI**

bounds & TV

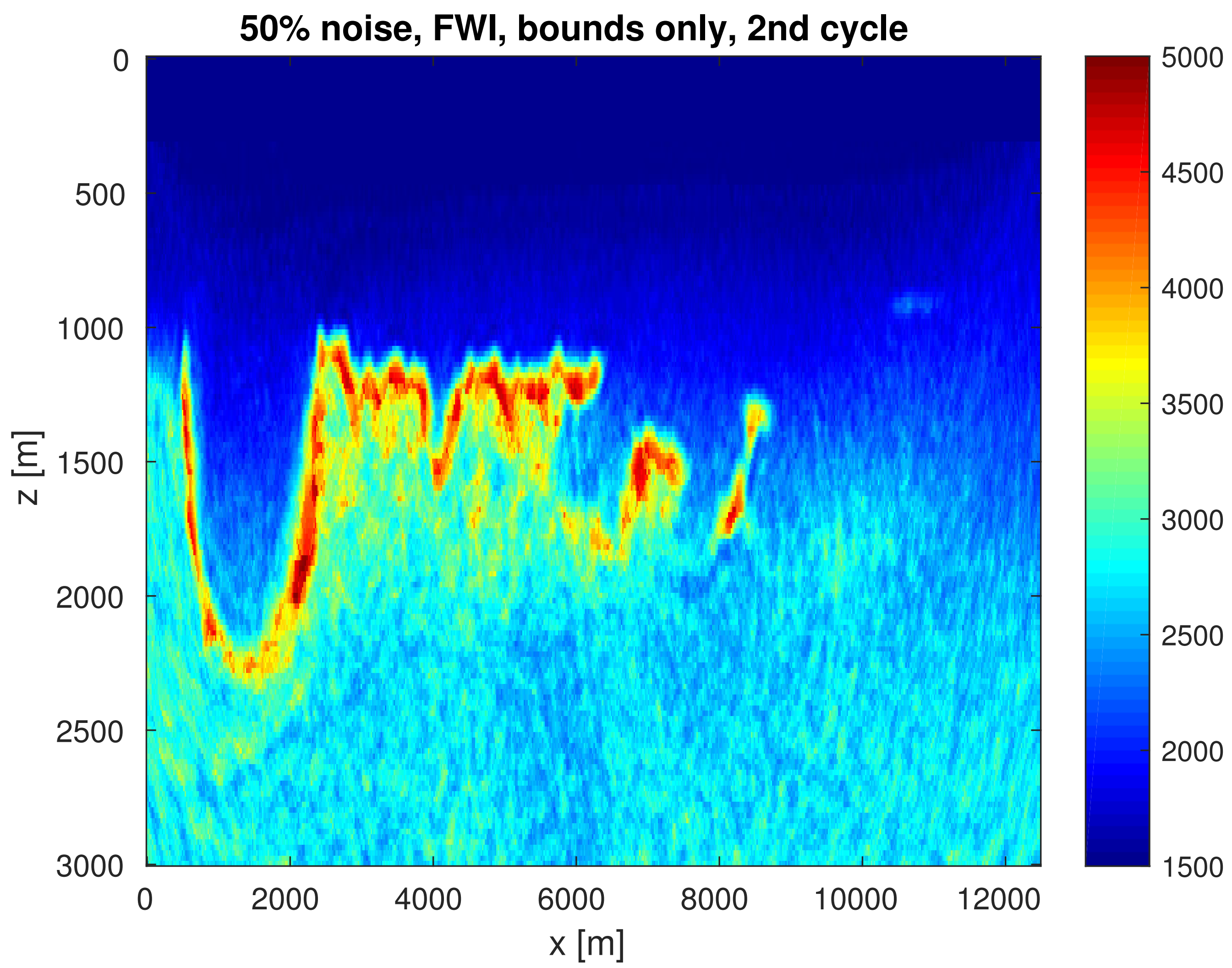


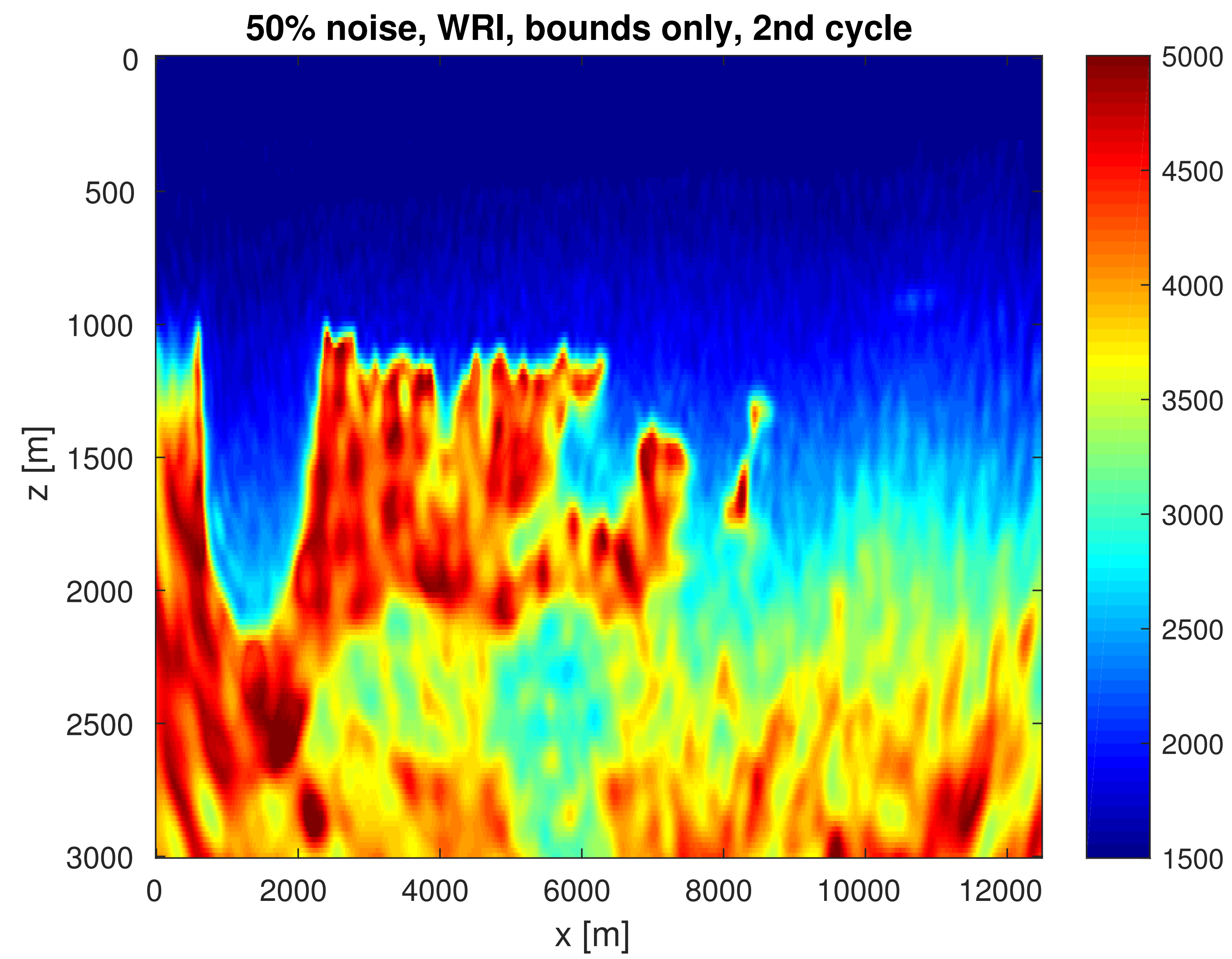


2nd cycle

$$\|\text{noise}\|_2 / \|\text{signal}\|_2 = 0.5$$

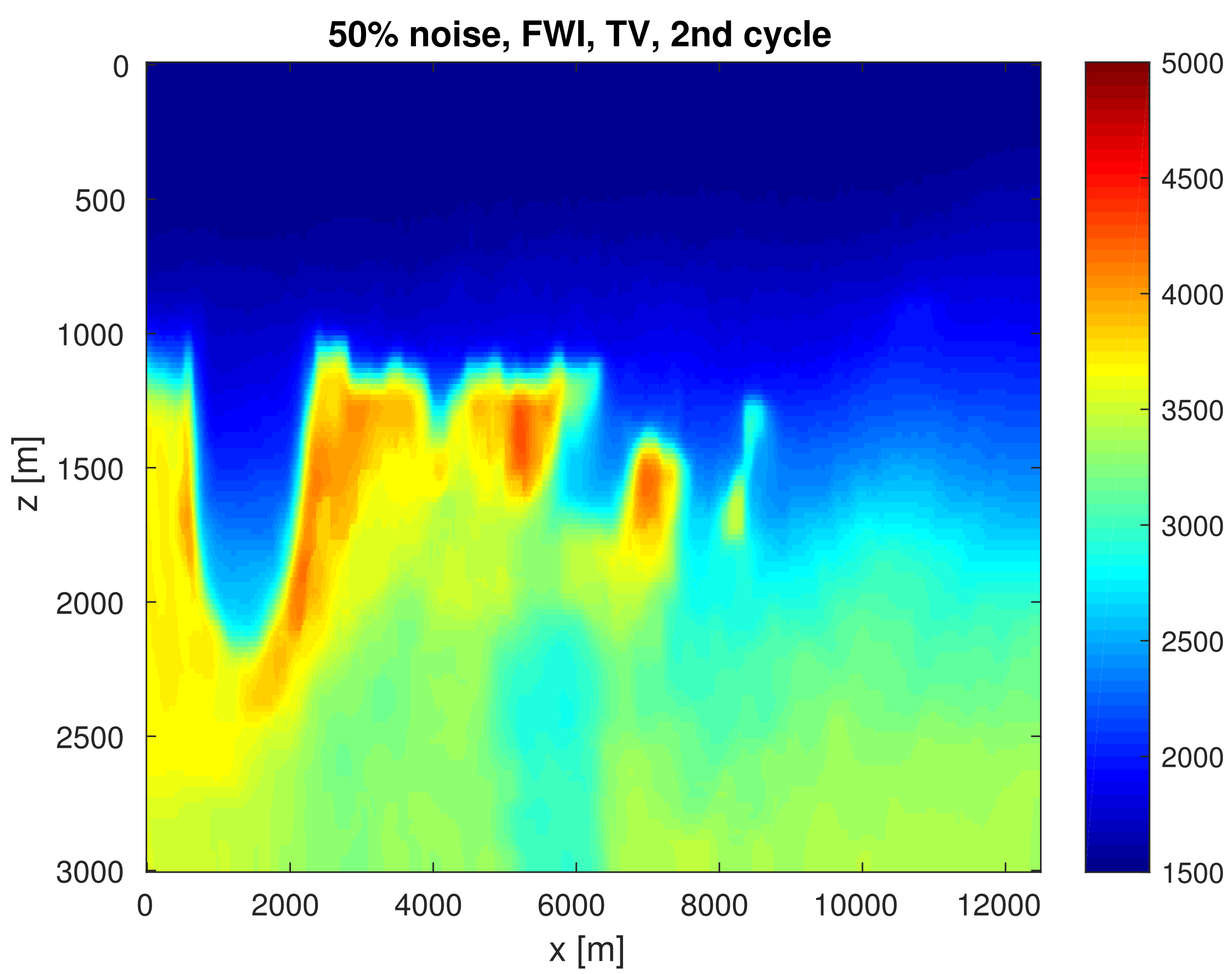
bounds only

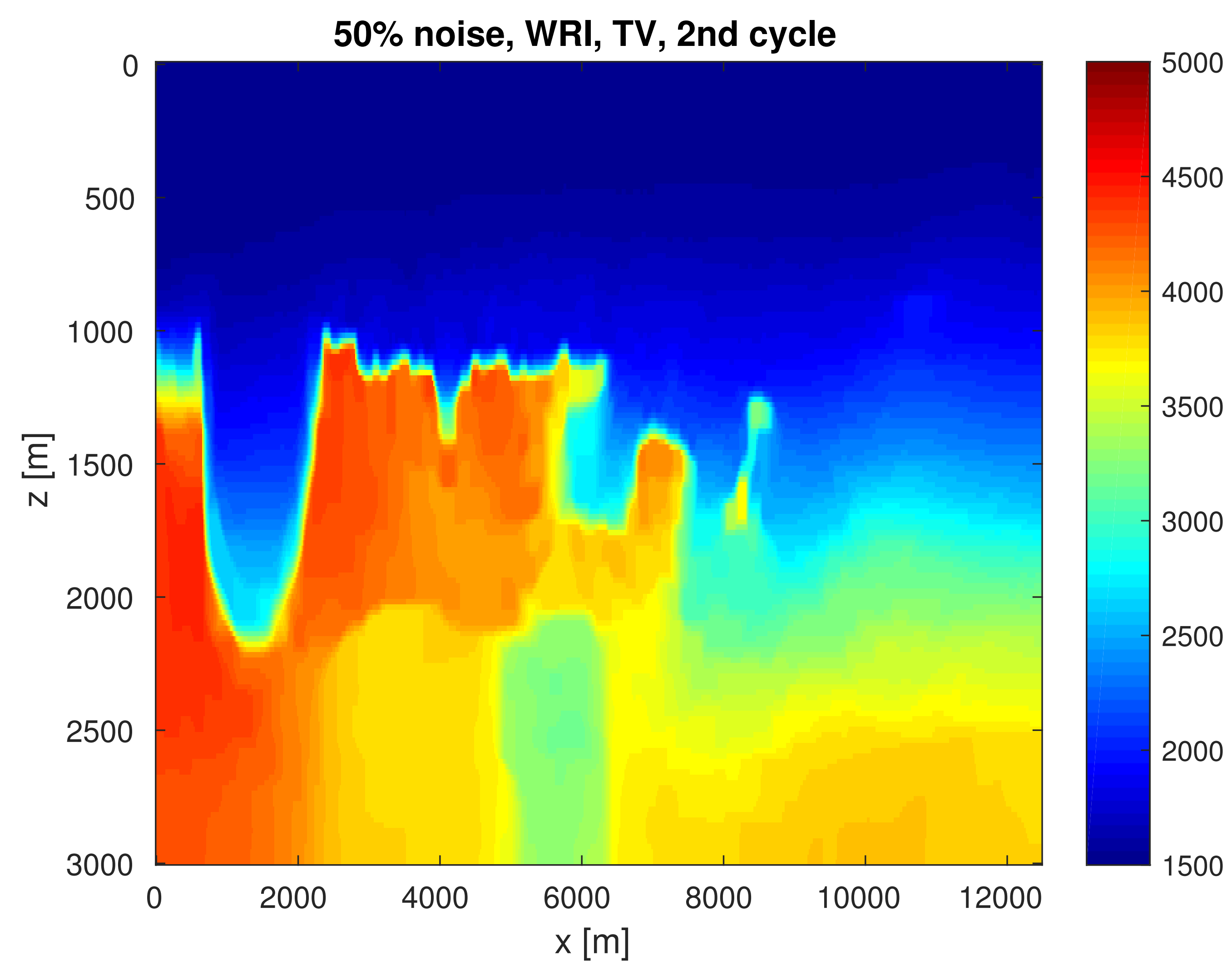




FWI **WRI**

bounds & TV

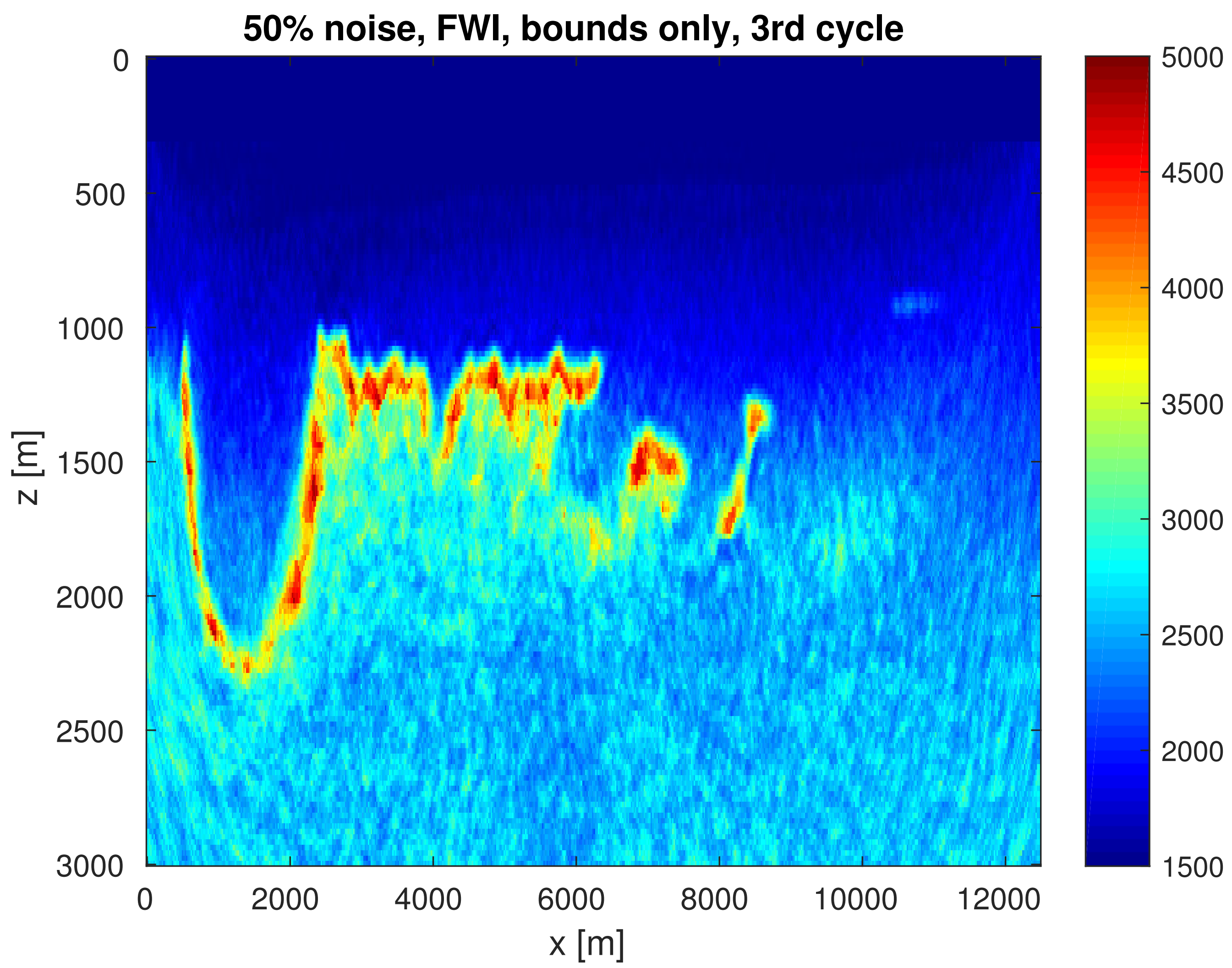


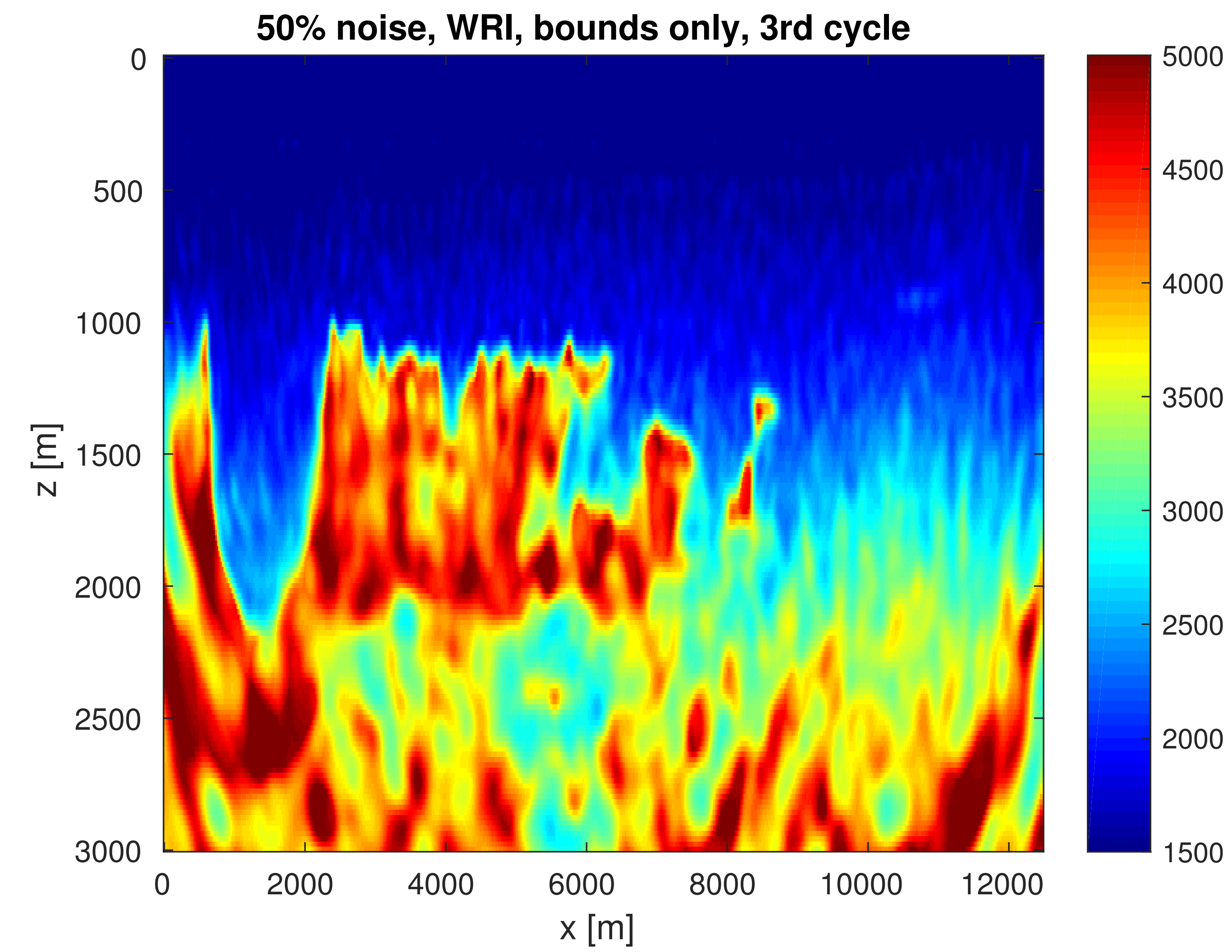


3rd cycle

$$\|\text{noise}\|_2 / \|\text{signal}\|_2 = 0.5$$

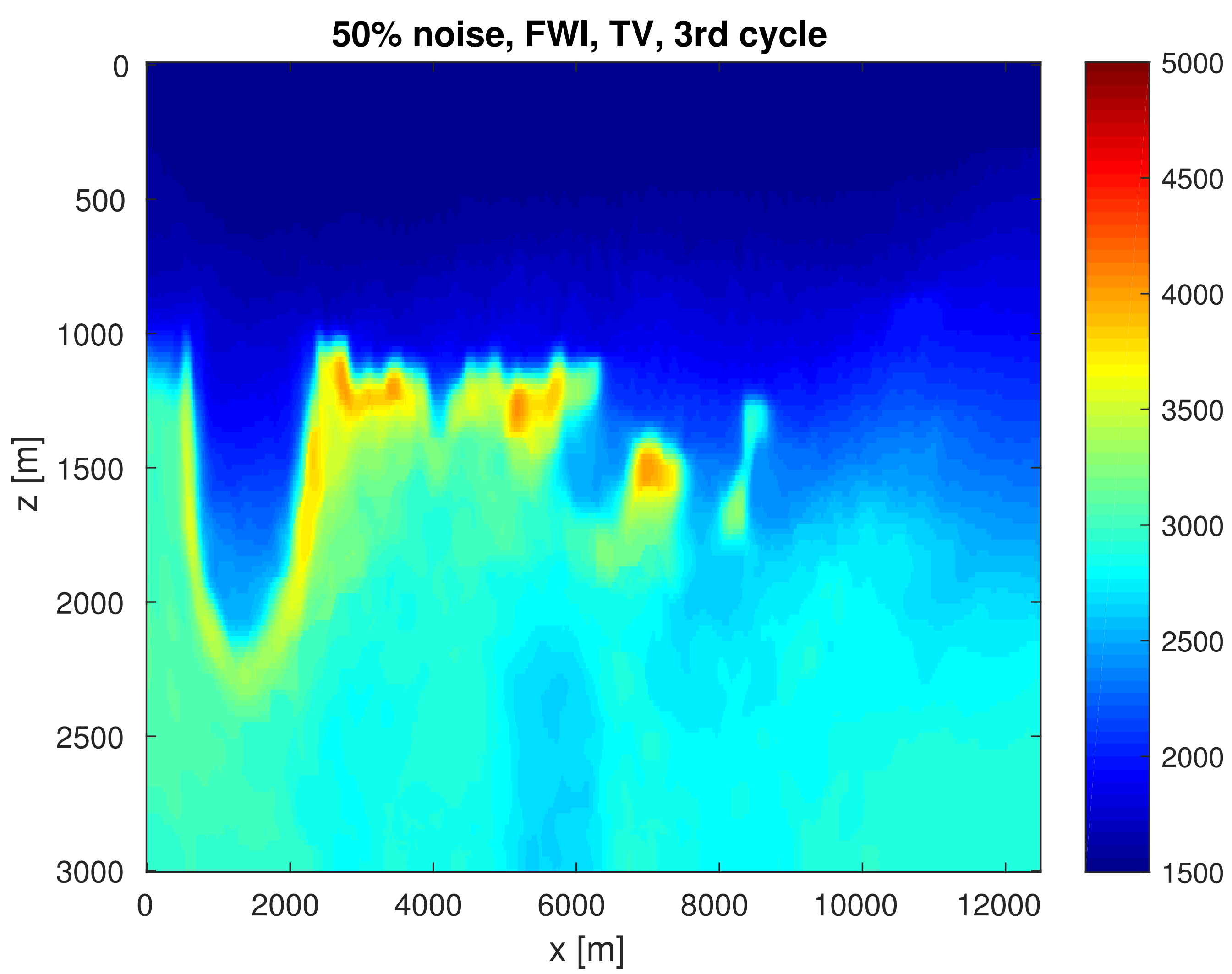
bounds only

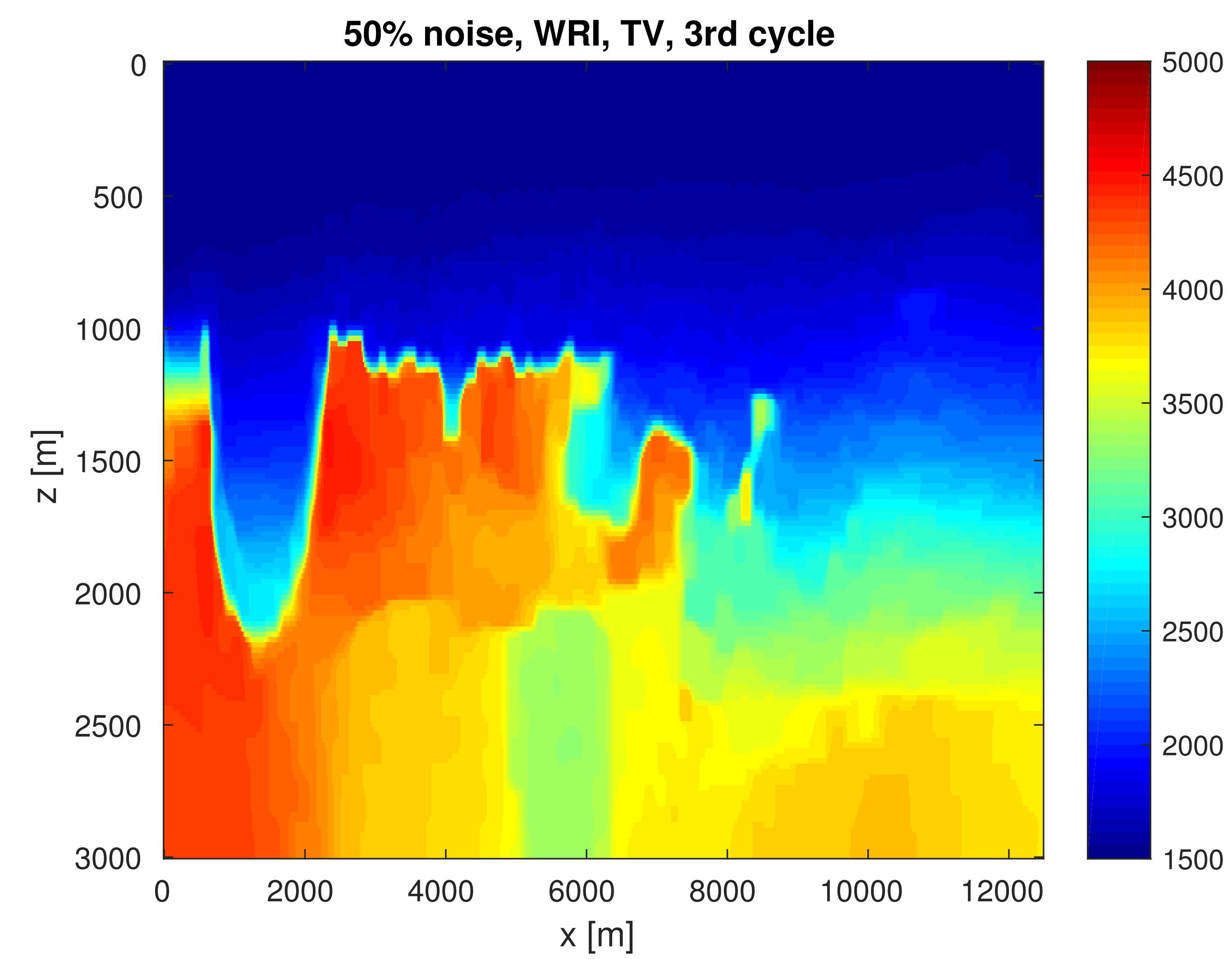




FWI **WRI**

bounds & TV





Today's agenda

Deal w/ "noise" by

- by handling source-side noise & modeling errors
- automatically select penalty parameter by exploiting duality

Move extensions to 3D

- time-domain WRI
- by exploiting duality

A denoising formulation of Full-Waveform Inversion

Rongrong Wang and Felix J. Herrmann

SLIM

University of British Columbia

Motivation

Noises in observed data consist of:

- spatial and spectral discretization errors
- inaccurate PDE modeling, boundary reflections, multiples
- source estimation error, unknown & interfering sources
- trace truncation error
- timing error
- receiver location error
- measurement noise

modeling error (spatial and spectral discretization errors through trace truncation error)

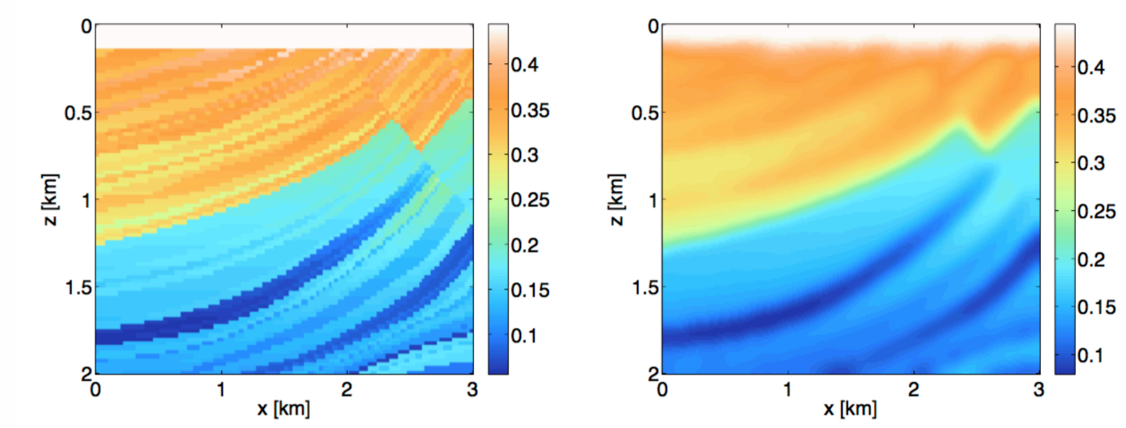
acquisition error (timing error through measurement noise)

interior of the domain (spatial and spectral discretization errors through source estimation error)

boundary (trace truncation error through measurement noise)

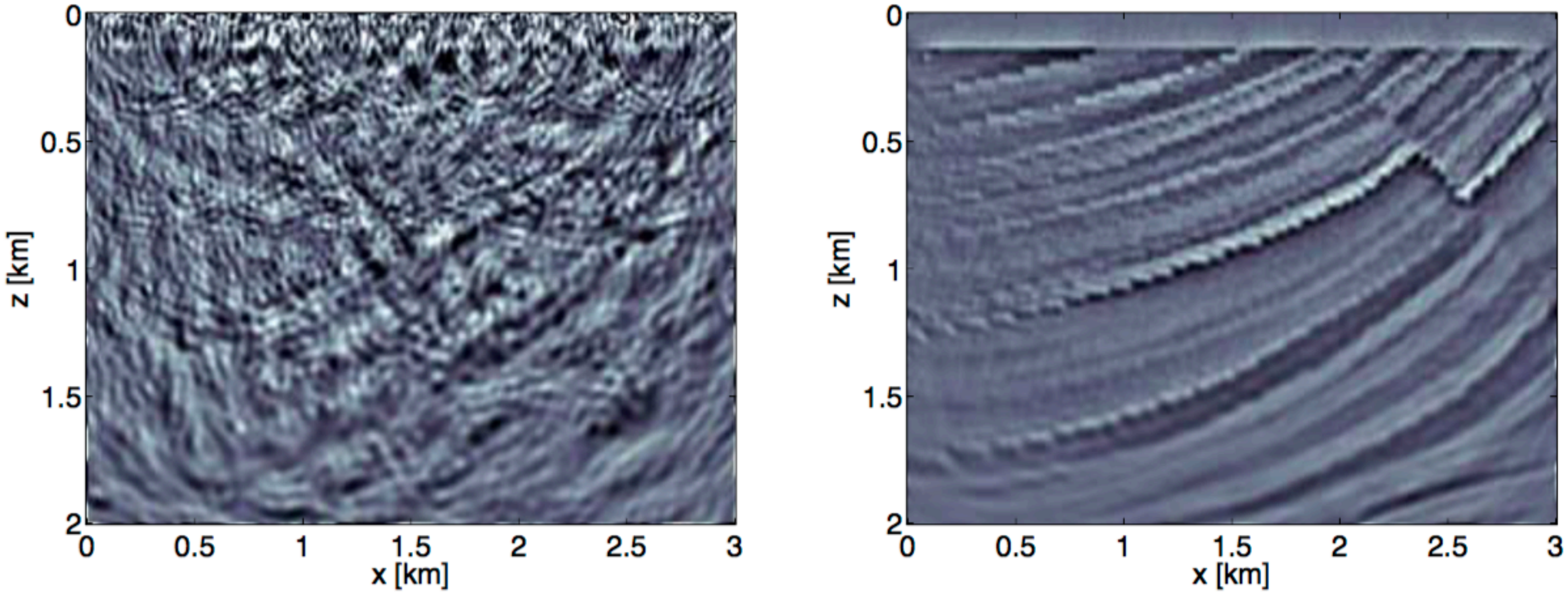
[Aravkin, A., van Leeuwen, T., & Herrmann, F J, 2012]

True Initial



Motivation—the Failures of FWI

When measurement noise is ``spiky"



(a) (b)

Model misfit for FWI

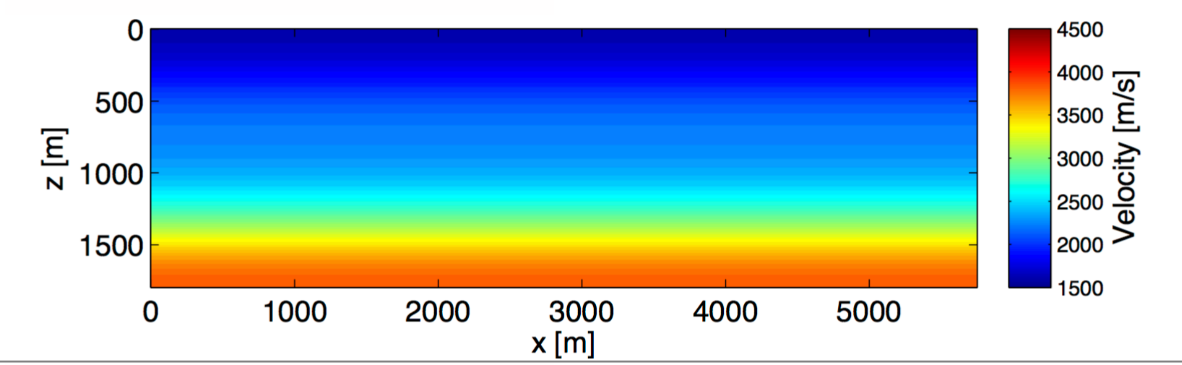
Model misfit for inversion with Student's t penalty

[Peters, A., & Herrmann, F J, 2014]

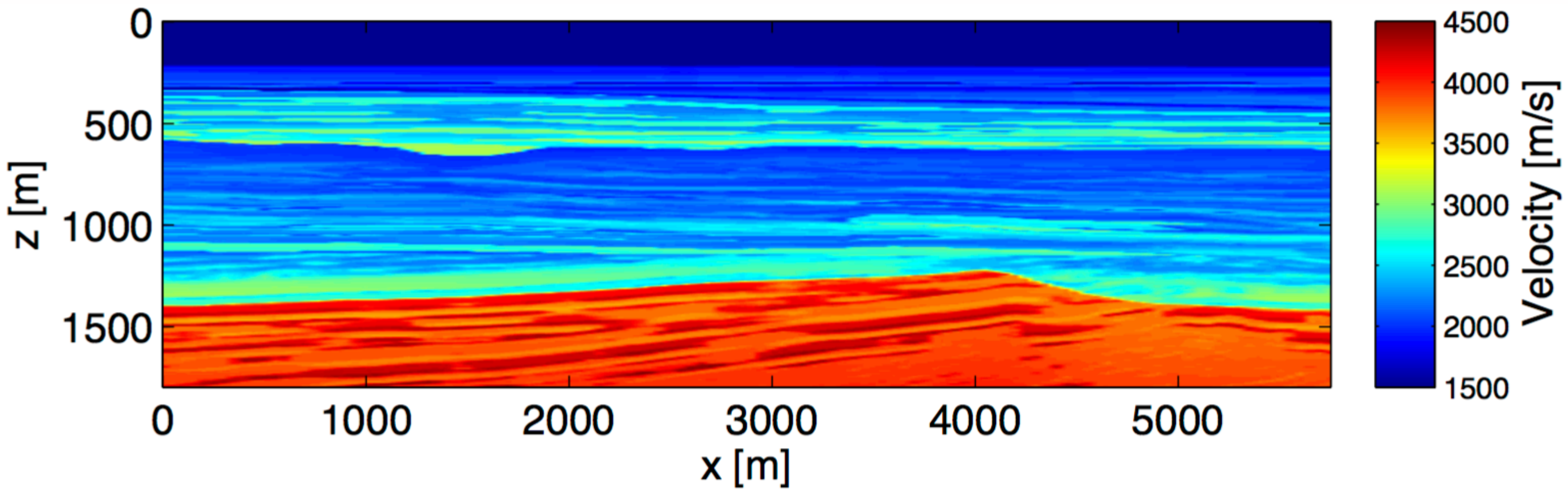
Motivation—the Failures of FWI

When water velocity is wrong

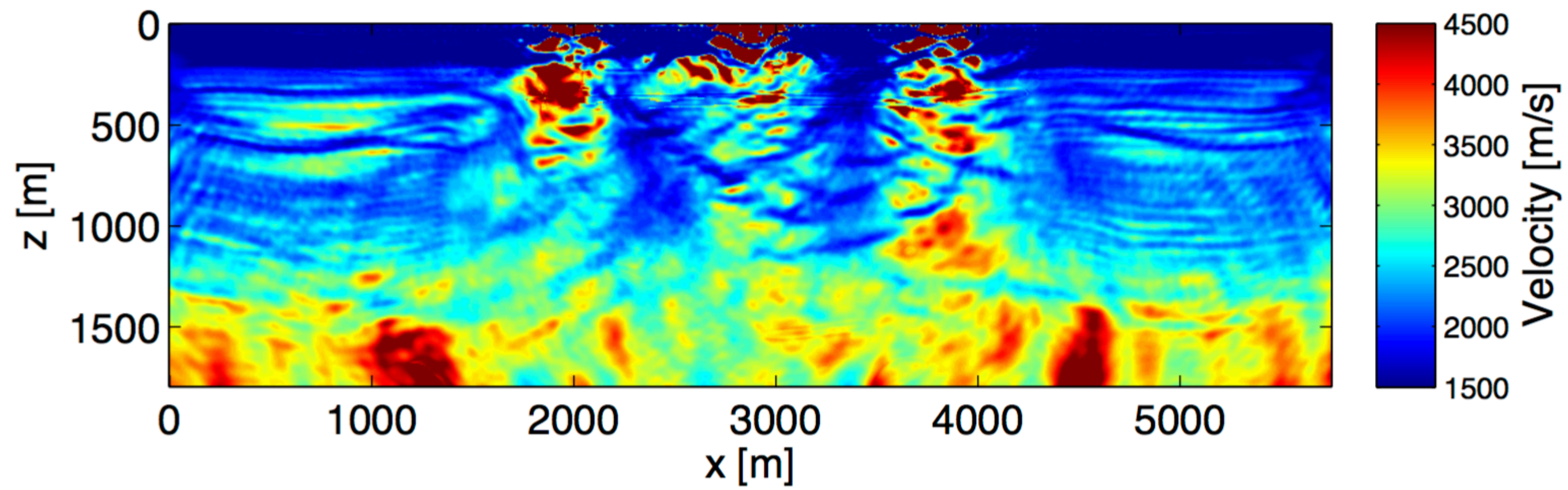
Initial



True model



FWI inversion



FWI & its relaxation

FWI requires strict satisfaction of the PDE:

$$\min_{m, u_i, i=1,\ldots,n_s} \sum_{i}^{n_s} \|P_{\Omega_i} u_i - d_i\|_2^2$$

$$\text{subject to } A(m)u_i = q_i, i = 1, ..., n_s$$

P_{Ω_i} : restriction operator

q_i : ith source

A : Time stepping or Helmholtz operator

d_i : Observed data for the ith source

u_i : wavefield associated to the ith source

- Implicitly assumes that noise is Gaussian distributed along sources & receivers
- Neglects modeling errors
- Cannot accommodate prior information on noise level
- Becomes problematic when water velocity is wrong

FWI & its relaxation

Direct relaxation of PDE constraint:

$$\min_{m,u_i,i=1,\ldots,n_s} \sum_i \|P_{\Omega_i} u_i - d_i\|_2$$

$$\text{subject to } \|A(m)u_i - q_i\|_2 \leq \epsilon, i = 1, ..., n_s$$

FWI & its relaxation

Direct relaxation of PDE constraint:

$$\min_{m, u_i, i=1,\dots,n_s} \sum_i \|P_{\Omega_i} u_i - d_i\|_2$$

Hard to choose!

$$\text{subject to } \|A(m)u_i - q_i\|_2 \leq \epsilon, i = 1, \dots, n_s$$

FWI & its relaxation

Direct relaxation of PDE constraint

$$\min_{m,u_i,i=1,\dots,n_s} \sum_i \|P_{\Omega_i} u_i - d_i\|_2$$

Hard to choose!

$$\text{subject to } \|A(m)u_i - q_i\|_2 \leq \epsilon, i = 1, \dots, n_s$$

Flip the objective and the constraint

$$\min_{m,u_i,i=1,\dots,n_s} \|A(m)u_i - q_i\|_2^2$$

$$\text{subject to } \|P_{\Omega_i} u_i - d_i\|_2 \leq \epsilon_i, i = 1, \dots, n_s$$

FWI & its relaxation

Direct relaxation of PDE constraint

$$\min_{m, u_i, i=1,\ldots,n_s} \sum_i \|P_{\Omega_i} u_i - d_i\|_2$$

$$\text{subject to } \|A(m)u_i - q_i\|_2 \leq \epsilon, i = 1, \ldots, n_s$$

Hard to choose!

Flip the objective and the constraint

$$\min_{m, u_i, i=1,\ldots,n_s} \|A(m)u_i - q_i\|_2^2$$

$$\text{subject to } \|P_{\Omega_i} u_i - d_i\|_2 \leq \epsilon_i, i = 1, \ldots, n_s$$

Noise level

FWI & its relaxation

Direct relaxation of PDE constraint

$$\min_{m,u_i,i=1,\ldots,n_s} \sum_i \|P_{\Omega_i} u_i - d_i\|_2$$

Hard to choose!

$$\text{subject to } \|A(m)u_i - q_i\|_2 \le \epsilon, i = 1, \ldots, n_s$$

Flip the objective and the constraint

$$\min_{m,u_i,i=1,\ldots,n_s} \|A(m)u_i - q_i\|_2^2$$

Noise level

$$\text{subject to } \|P_{\Omega_i} u_i - d_i\|_2 \le \epsilon_i, i = 1, \ldots, n_s$$

Decompose wavefield variables

$$u_i = P_{\Omega_i^c}^T P_{\Omega_i^c} u_i + P_{\Omega_i}^T P_{\Omega_i} u_i$$

Boundary part

Interior part

FWI & its relaxation

Direct relaxation of PDE constraint

$$\min_{m,u_i,i=1,\ldots,n_s} \sum_i \|P_{\Omega_i} u_i - d_i\|_2$$

$$\text{subject to } \|A(m)u_i - q_i\|_2 \leq \epsilon, i = 1, \ldots, n_s$$

Hard to choose!

Flip the objective and the constraint

$$\min_{m,u_i,i=1,\ldots,n_s} \|A(m)u_i - q_i\|_2^2$$

$$\text{subject to } \|P_{\Omega_i} u_i - d_i\|_2 \leq \epsilon_i, i = 1, \ldots, n_s$$

Noise level

Decompose wavefield variables

$$u_i = P_{\Omega_i^c}^T \overbrace{P_{\Omega_i^c} u_i}^{v_i} + \underbrace{P_{\Omega_i}^T P_{\Omega_i} u_i}_{b_i}$$

Interior part

Boundary part

FWI & its relaxation

Direct relaxation of PDE constraint

$$\min_{m, u_i, i=1,\ldots,n_s} \sum_i \|P_{\Omega_i} u_i - d_i\|_2$$

Hard to choose!

$$\text{subject to } \|A(m)u_i - q_i\|_2 \leq \epsilon, i = 1, \ldots, n_s$$

Flip the objective and the constraint

$$\min_{m, u_i, i=1,\ldots,n_s} \|A(m)u_i - q_i\|_2^2$$

Noise level

$$\text{subject to } \|P_{\Omega_i} u_i - d_i\|_2 \leq \epsilon_i, i = 1, \ldots, n_s$$

Decompose wavefield variables

$$u_i = \underbrace{P_{\Omega_i^c}^T \overbrace{P_{\Omega_i^c} u_i}^{v_i}}_{\text{Interior part}} + \underbrace{P_{\Omega_i}^T \overbrace{P_{\Omega_i} u_i}^{b_i}}_{\text{Boundary part}}$$

Interior part

Boundary part

$$\min_{m, b_i, v_i, i=1,\ldots,n_s} \|A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i) - q_i\|_2^2$$

$$\text{subject to } \|b_i - d_i\|_2 \leq \epsilon_i, i = 1, \ldots, n_s$$

The denoising formulation (FWIDN)

***Denoising* formulation of FWI:**

$$\min_{m, b_i, v_i, i=1,\ldots,n_s} \|A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i) - q_i\|_2^2$$
$$\text{subject to } \|b_i - d_i\|_2 \leq \epsilon_i, i = 1, \ldots, n_s$$

Pros:

- allows noise levels ϵ_i to vary with sources, and allows $\epsilon_i = 0$
- ensures reasonable PDE fidelity while preventing overfit
- all pros of WRI

Cons: algorithmically & computationally more demanding

FWI-DN – a more general form

Weighted/preconditioned least-squares objective:

$$\min_{m,b_i,v_i,i=1,\dots,n_s} \|\mathcal{D}_z(A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i) - q_i)\|_2^2$$

$$\text{subject to } \|b_i - d_i\|_2 \leq \epsilon_i, i = 1, \dots, n_s$$

- $\mathcal{D}_z$ reshapes PDE misfit distribution
- Imposes looser PDE constraint at shallow part where the model is “noisier”
- Examples of $\mathcal{D}_z$: linear depth weighting, two-level depth weighting

$$\mathcal{D}_z f(x,z) = z f(x,z) \qquad \mathcal{D}_z f(x,z) = \chi_{z<z_0} f(x,z) + 2\chi_{z\geq z_0} f(x,z)$$

Solving FWI-DN

Strategy: alternatively update m and $b_i, i = 1, ..., n_s$

At iteration k,

1. fix m^k, solve for $b_i^{k+1}, i = 1, ..., n_s$ from

$$(b_i^{k+1}, v_i^{k+1}) = \arg \min_{b_i, v_i, i=1,\dots,n_s} \|\mathcal{D}_z(A(m^k)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i) - q_i)\|_2^2 \qquad (P_d)$$
$$\text{subject to } \|b_i - d_i\|_2 \leq \epsilon_i, i = 1, ..., n_s$$

2. for fixed $b_i^{k+1}, i = 1, ..., n_s$, update m^k by solving T steps of

$$\min_{m, v_i, i=1,\dots,n_s} \sum_{i=1}^{n_s} \|\mathcal{D}_z(A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i^{k+1}) - q_i)\|_2^2 \qquad (P_m)$$

Solving for (P_d) — a denoising step

(P_d) Fix m^k, solve for b_i^{k+1} from

$$(b_i^{k+1}, v_i^{k+1}) = \arg\min_{b_i, v_i} \|\mathcal{D}_z(A(m^k)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i) - q_i)\|_2^2$$

$$\text{subject to } \|b_i - d_i\|_2 \leq \epsilon_i,$$

Solving for (P_d) — a denoising step

(P_d) Fix m^k, solve for b_i^{k+1} from

$$(b_i^{k+1}, v_i^{k+1}) = \arg\min_{b_i, v_i} \|\mathcal{D}_z(A(m^k)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i) - q_i)\|_2^2$$
$$\text{subject to } \|b_i - d_i\|_2 \leq \epsilon_i,$$

The Lagrangian dual of (P_d) is

$$\max_{\lambda \geq 0} \phi(\lambda)$$

where

$$\phi(\lambda) = \min_{u_i} \|\mathcal{D}_z(A(m)u_i - q_i)\|_2^2 + \lambda \|P_{\Omega_i} u_i - d_i\|_2^2 - \lambda \epsilon_i$$

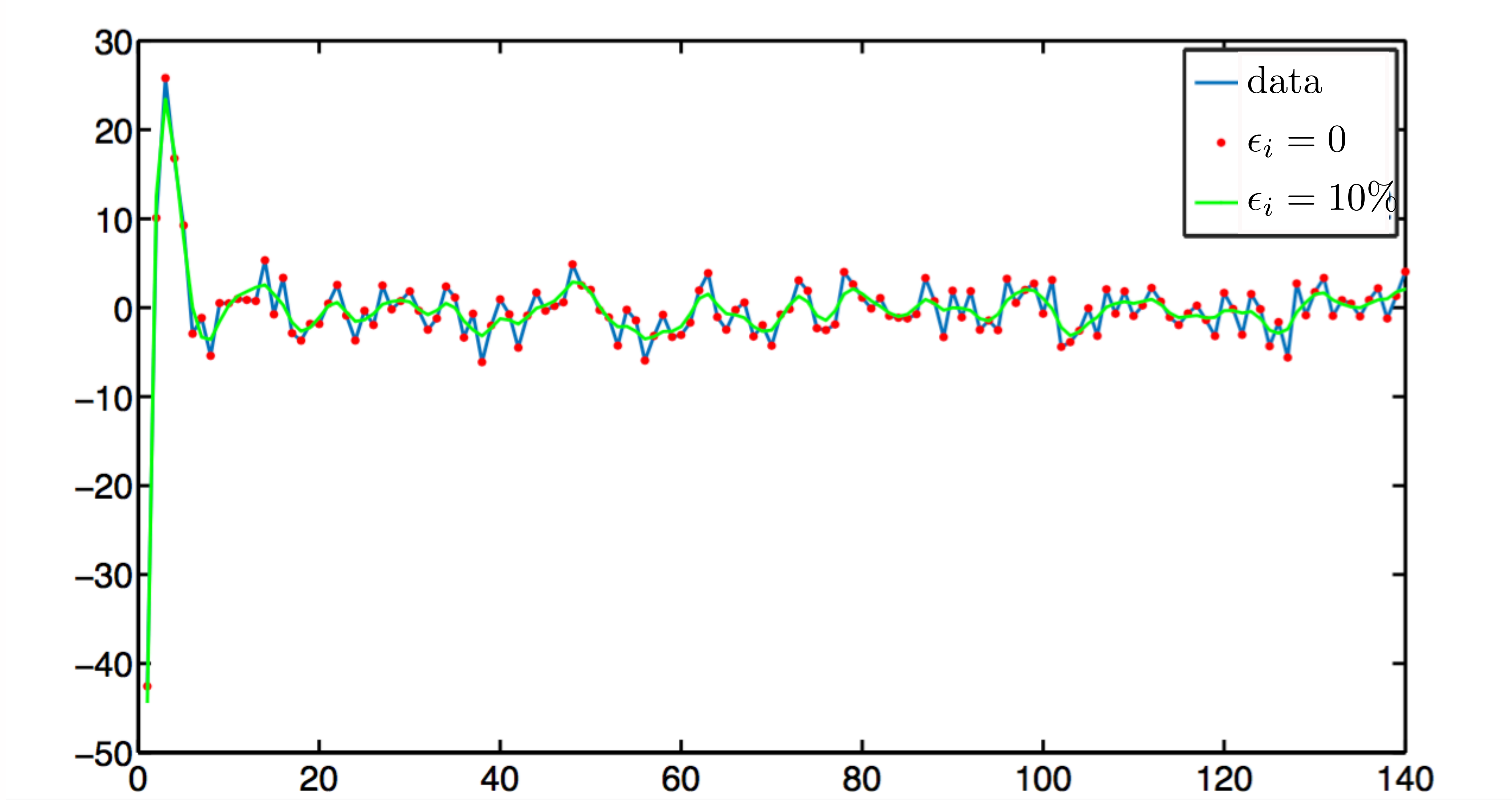
Strong duality principle [More, 1993] guarantees primal & dual optimality agree

Solving for (P_d) — a denoising step

(P_d) Fix m^k, solve for b_i^{k+1} from

$$(b_i^{k+1}, v_i^{k+1}) = \arg\min_{b_i, v_i} \|\mathcal{D}_z(A(m^k)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i) - q_i)\|_2^2$$
$$\text{subject to } \|b_i - d_i\|_2 \leq \epsilon_i,$$

Text



Denoising effect of (P_d)

The Lagrangian dual of (P_d) is

$$\max_{\lambda \geq 0} \phi(\lambda)$$

where

$$\phi(\lambda) = \min_{u_i} \|\mathcal{D}_z(A(m)u_i - q_i)\|_2^2 + \lambda \|P_{\Omega_i} u_i - d_i\|_2^2 - \lambda \epsilon_i$$

Strong duality principle [More, 1993] guarantees primal & dual optimality agree.

Solving for (P_d) — a denoising step

$\phi(\lambda)$ has closed-form gradient & Hessian

$$\phi'(\lambda) = \|P_{\Omega_i}\bar{u}_i(\lambda) - d_i\|_2^2 - \epsilon_i$$

$$\phi''(\lambda) = -2(P_{\Omega_i}\bar{u}_i(\lambda) - d_i)^T P_{\Omega_i} C^{-1} P_{\Omega_i}^T (P_{\Omega_i}\bar{u}_i - d_i)$$

where

$$C = A(m)^T \mathcal{D}_z^T \mathcal{D}_z A(m) + \lambda P_{\Omega_i}^T P_{\Omega_i} \quad \bar{u}_i(\lambda) = \begin{bmatrix} \mathcal{D}_z(A(m^k)) \\ \sqrt{\lambda} P_{\Omega_i} \end{bmatrix}^\dagger \begin{bmatrix} \mathcal{D}_z(q_i) \\ \sqrt{\lambda} d_i \end{bmatrix}$$

Newton steps for λ

$$\lambda^{k+1} = \lambda^k - \phi'(\lambda)/\phi''(\lambda)$$

After finding the minimizer λ^*, the primal optimizers are

$$v_i^{k+1} = P_{\Omega_i^c}\bar{u}_i(\lambda^*), \quad v_i^{k+1} = P_{\Omega_i}\bar{u}_i(\lambda^*)$$

Solving for (P_m)

For fixed $b_i^{k+1}, i = 1, ..., n_s$, update m^k by solving T steps of

$$(P_m) \quad \min_{m, v_i, i=1,\dots,n_s} \sum_{i=1}^{n_s} \|\mathcal{D}_z(A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i^{k+1}) - q_i)\|_2^2$$

Solving for (P_m)

For fixed $b_i^{k+1}, i = 1, ..., n_s$, update m^k by solving T steps of

$$(P_m) \quad \min_{m, v_i, i=1,\ldots,n_s} \sum_{i=1}^{n_s} \|\mathcal{D}_z(A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i^{k+1}) - q_i)\|_2^2$$

Solve by variable projection

[Aravkin and van Leeuwen, 2012]

Solving for (P_m)

For fixed $b_i^{k+1}, i = 1, ..., n_s$, update m^k by solving T steps of

$$(P_m) \quad \min_{m, v_i, i=1,\dots,n_s} \sum_{i=1}^{n_s} \|\mathcal{D}_z(A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i^{k+1}) - q_i)\|_2^2$$

Solve by variable projection

[Aravkin and van Leeuwen, 2012]

$$\min_{m, v_i, i=1,\dots,n_s} \sum_{i=1}^{n_s} \|\mathcal{D}_z(A(m)(P_{\Omega_i^c}^T v_i + P_{\Omega_i}^T b_i^{k+1}) - q_i)\|_2^2 \equiv f(m, v_1, ... v_{n_s})$$

$$\Longleftrightarrow \min_{m, v_i, i=1,\dots,n_s} f(m, v_1, ..., v_{n_s}) = \min_m f(m, \bar{v}_1, ..., \bar{v}_{n_s}) \quad (\bar{v}_1, ..., \bar{v}_{n_s}) = \arg \min_{v_1,\dots,v_{n_s}} f(m, v_1, ..., v_{n_s})$$

Algorithm and Complexity

Inputs: $m_0, d_i, q_i, \quad i = 1, ..., n_s$, T, K

For $\omega = \omega_1, ..., \omega_n$ do

solve (P_d) using T iterations of Newton updates on λ

perform K gradient or L-BFGS updates on m towards the minimizer of (P_m)

Endfor

On average, 1 update of m requires: 2 PDE solves for FWI

2 PDE solves for WRI

3-5 PDE solves for FWI-DN

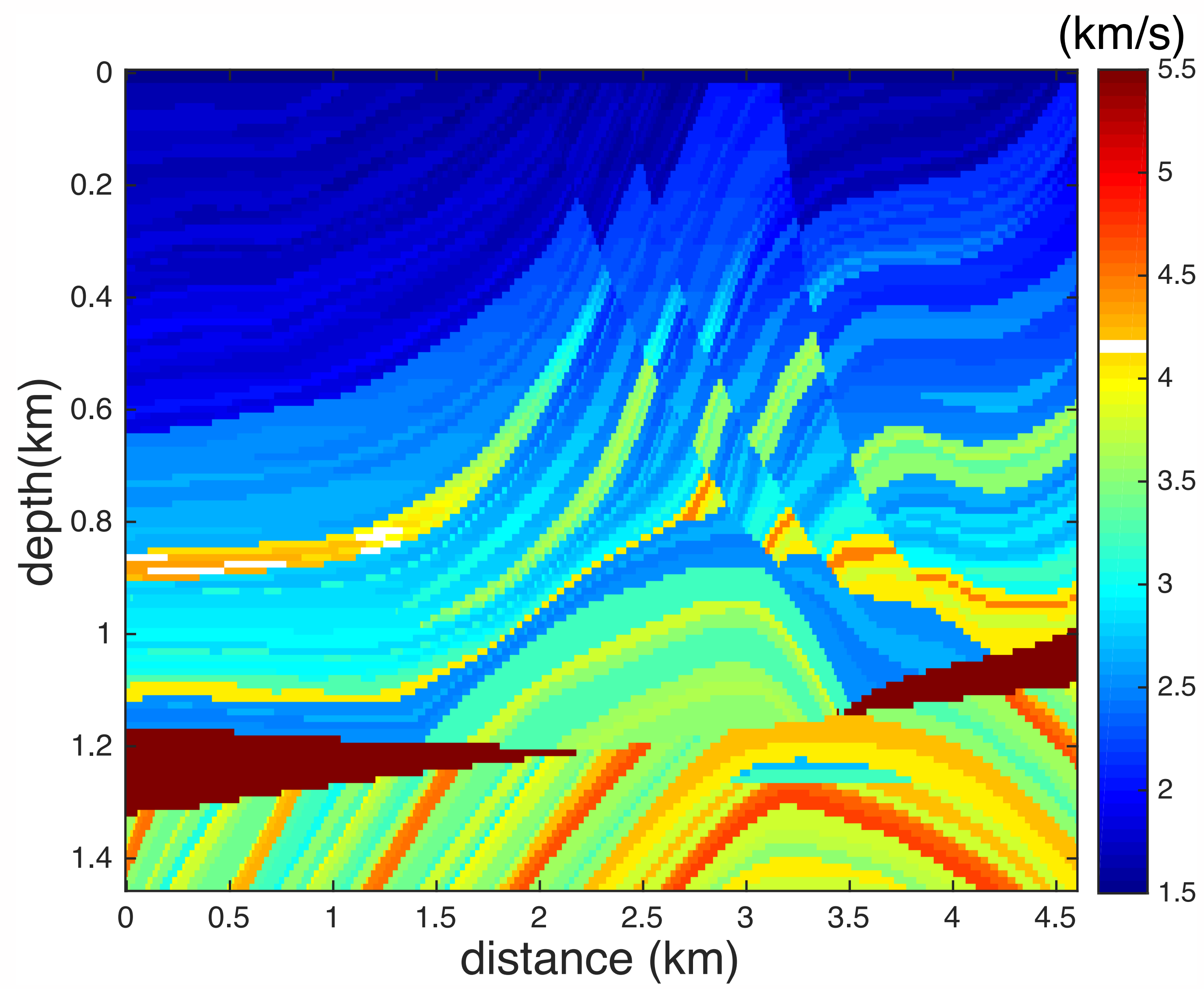
Case Study

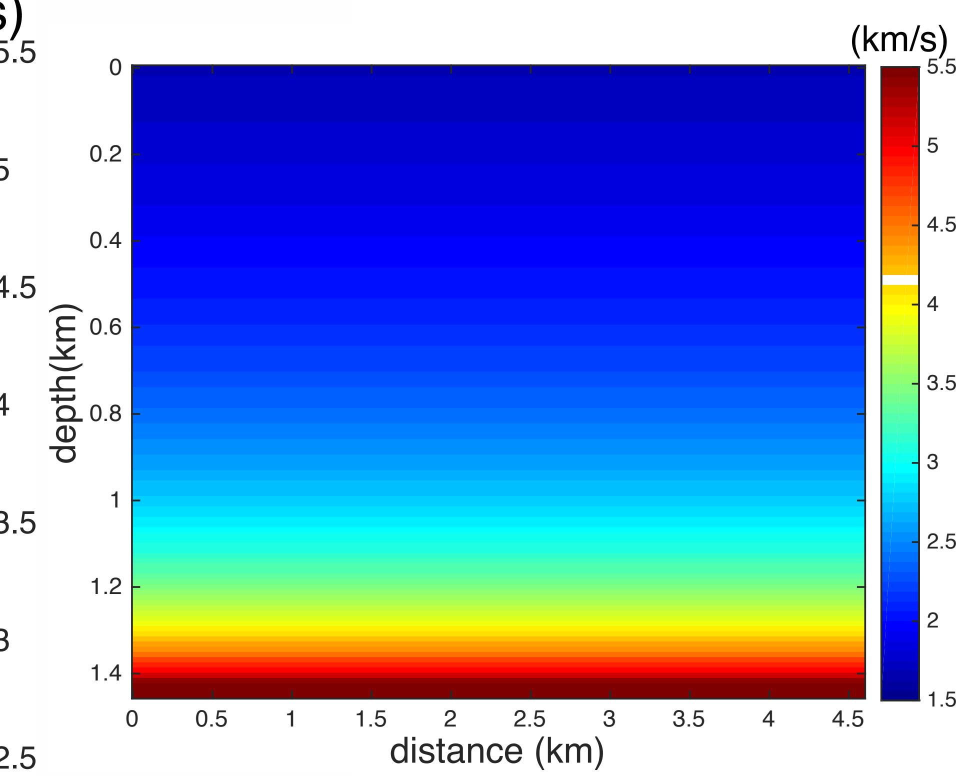
Test 1: robustness under non-uniform noise along sources

Test 1

- Frequency continuation using batches [3,3.5][3.5,4]....[14.5,15]Hz
- source spacing : 240m
- receiver spacing : 48m
- SNR=0 for low frequency data 3-10Hz
- SNR=25dB for high frequency data 10-15Hz
- Linear depth weighting
- noise level $\epsilon_{s_j} = 3\epsilon_{s_i}$ where

  $i = 1, ..., \lfloor n_s/2 \rfloor, j = \lfloor n_s/2 \rfloor + 1, ..., n_s$





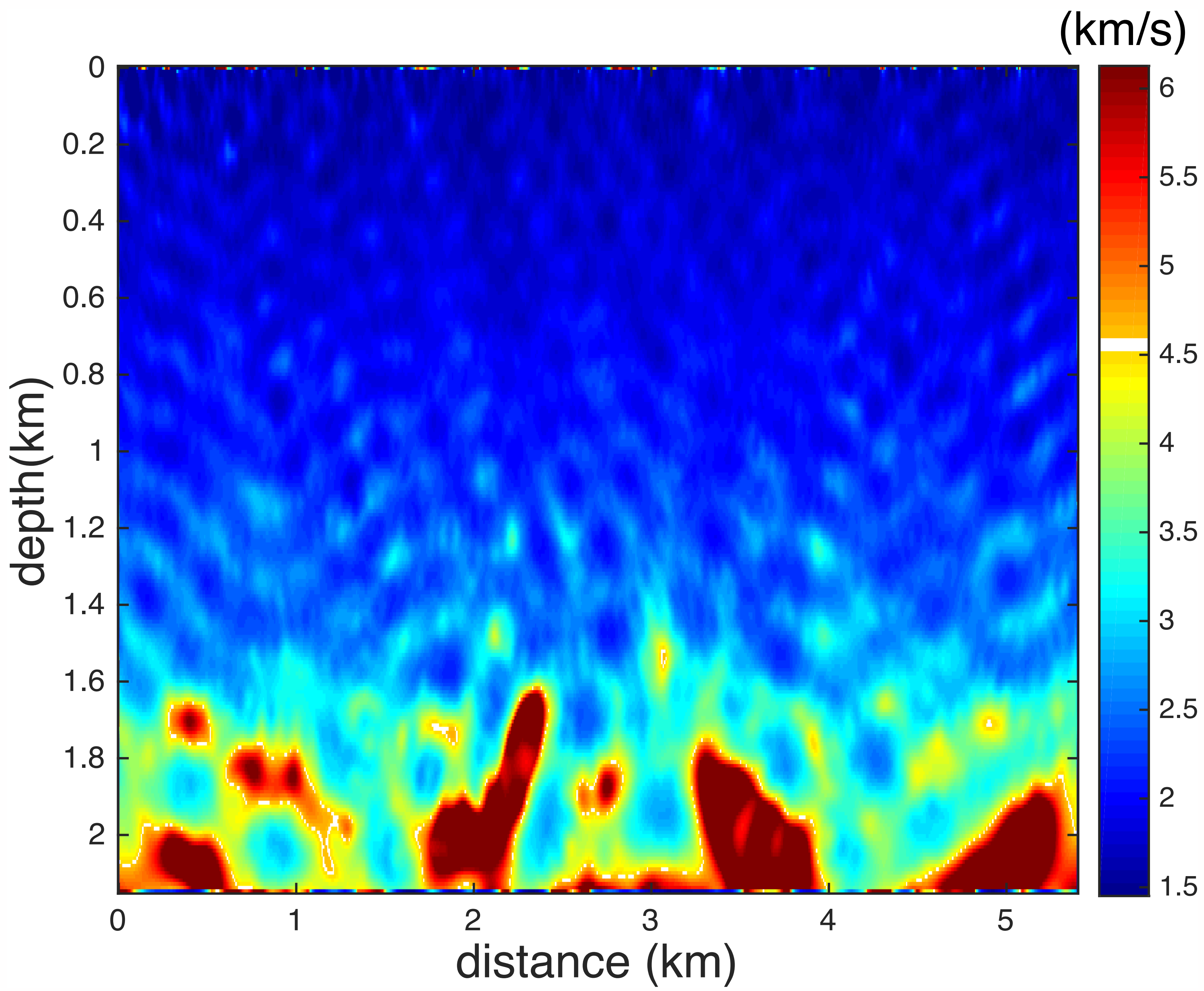
Test 1

Method for comparison: weighted FWI

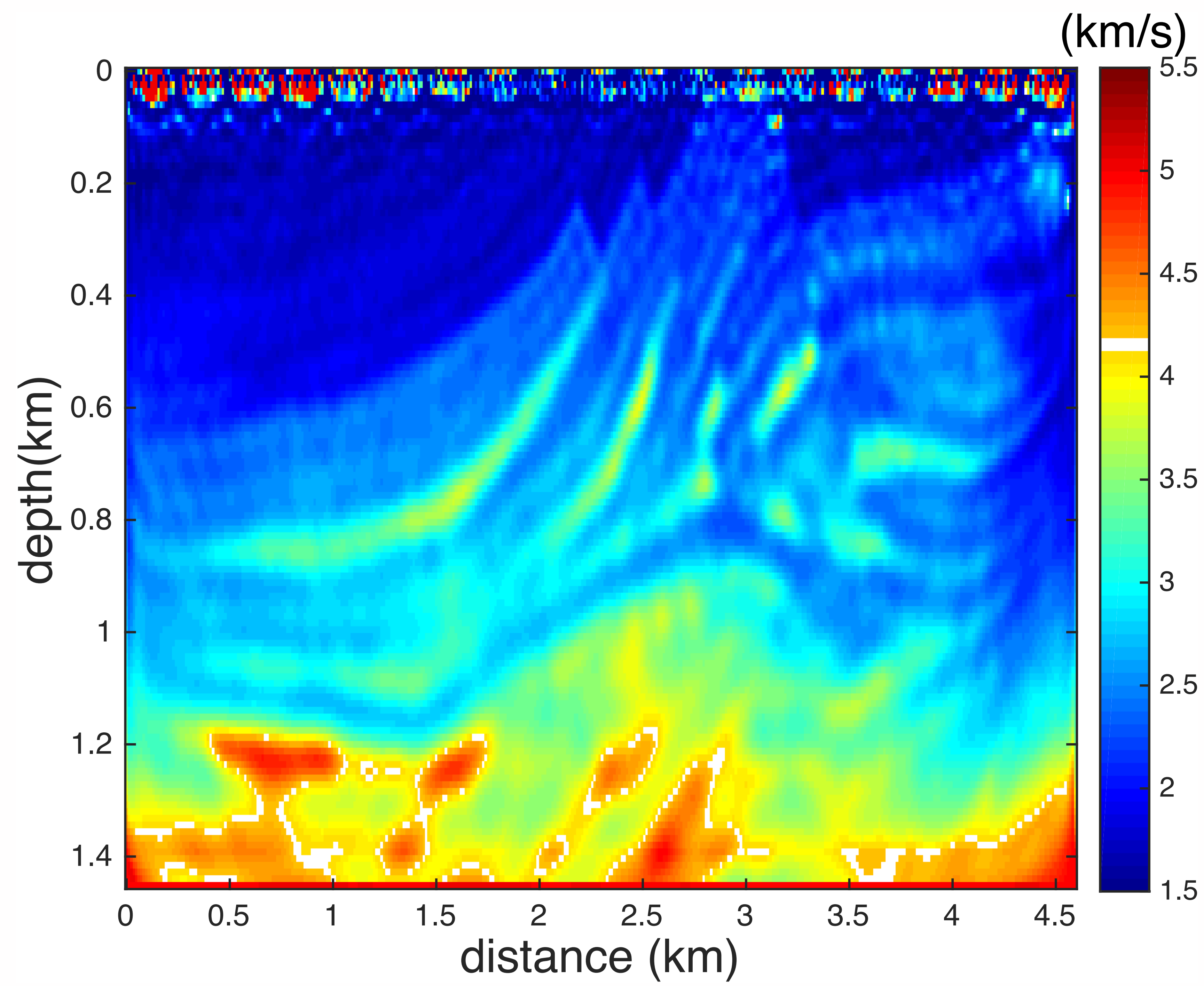
$$\min_m \sum_{i \in N_1} 9 \| P_{\Omega_i} A^{-1}(m) q_i - d_i \|_2^2 + \min_m \sum_{i \in N_2} \| P_{\Omega_i} A^{-1}(m) q_i - d_i \|_2^2$$

where $N_1 = \{1, ..., \lfloor \frac{n_s}{2} \rfloor\}, N_2 = \{\lfloor \frac{n_s}{2} \rfloor + 1, ..., n_s\}$

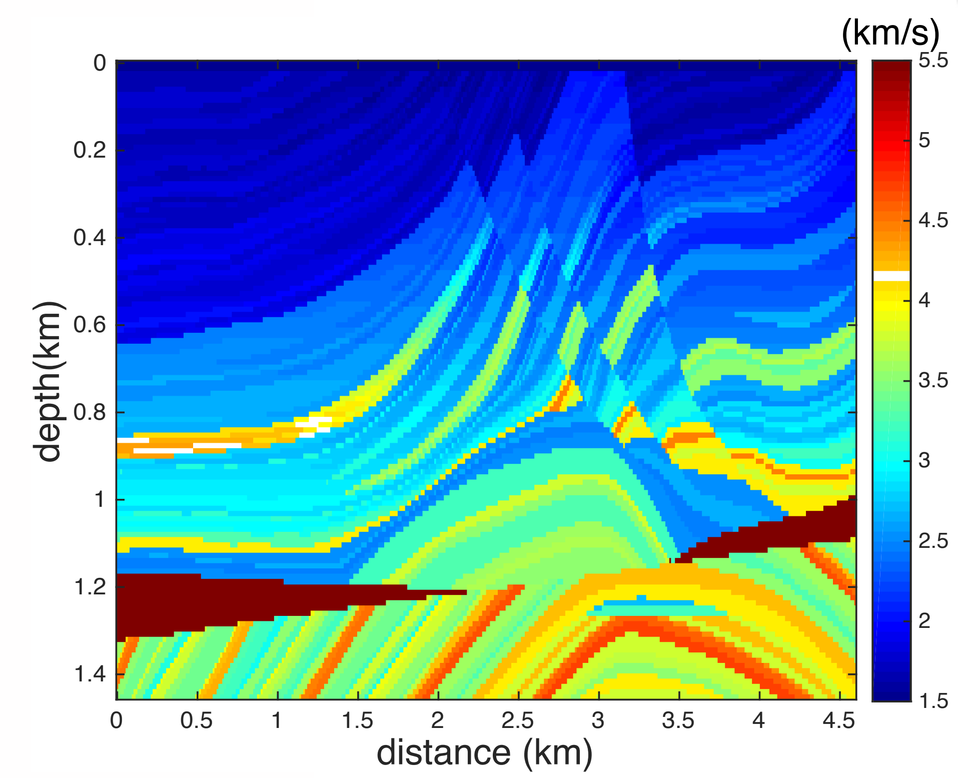
Test 1



Inverted model w/ weighted FWI



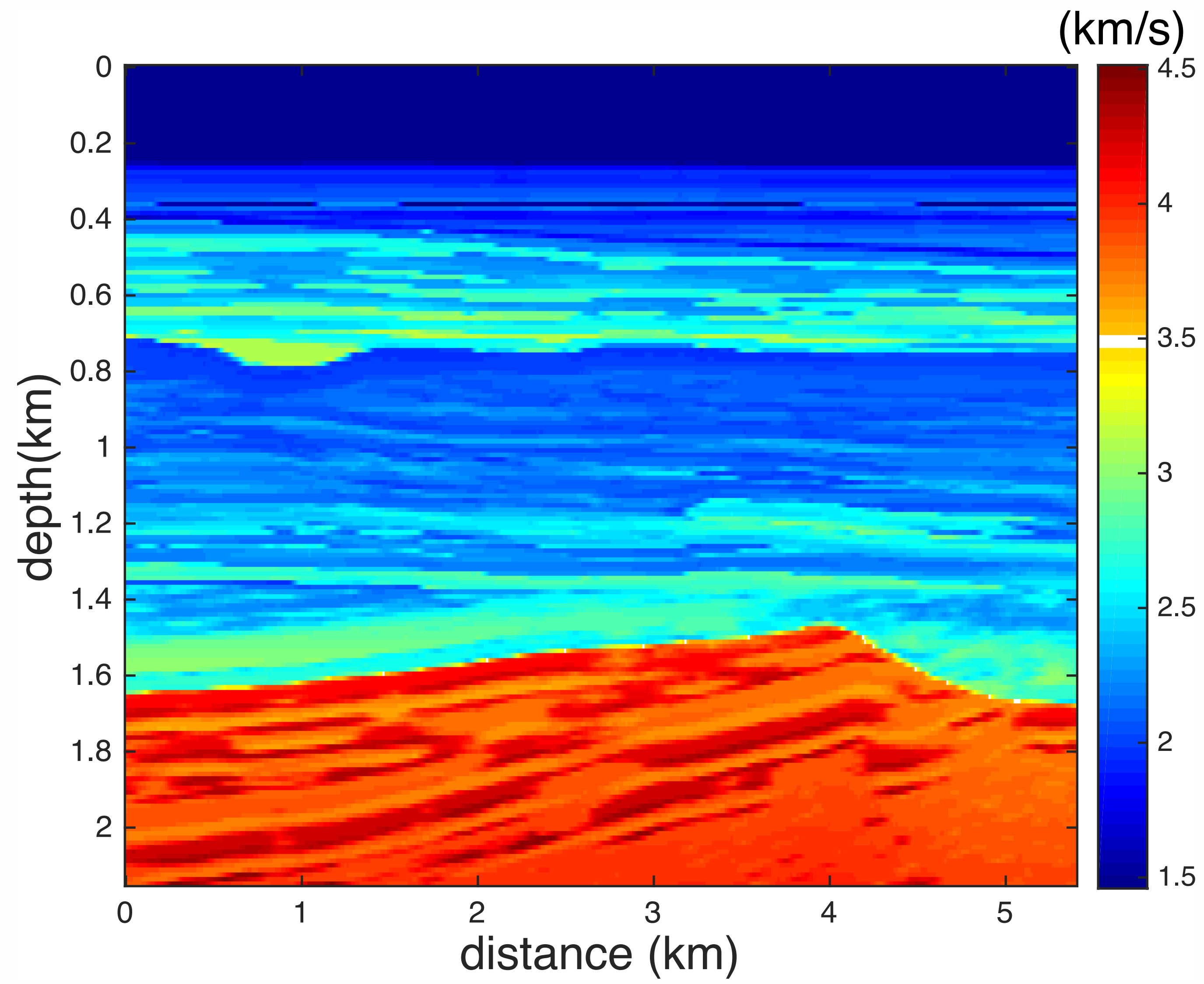
Inverted model w/ FWI-DN

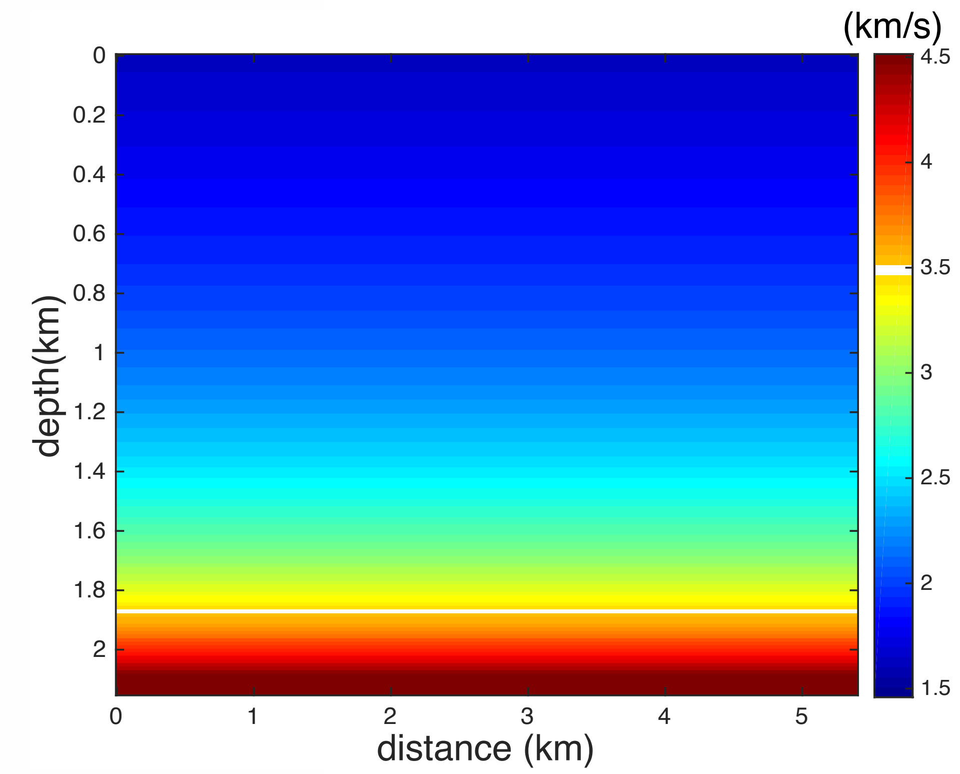


Test 2: robustness under modeling error

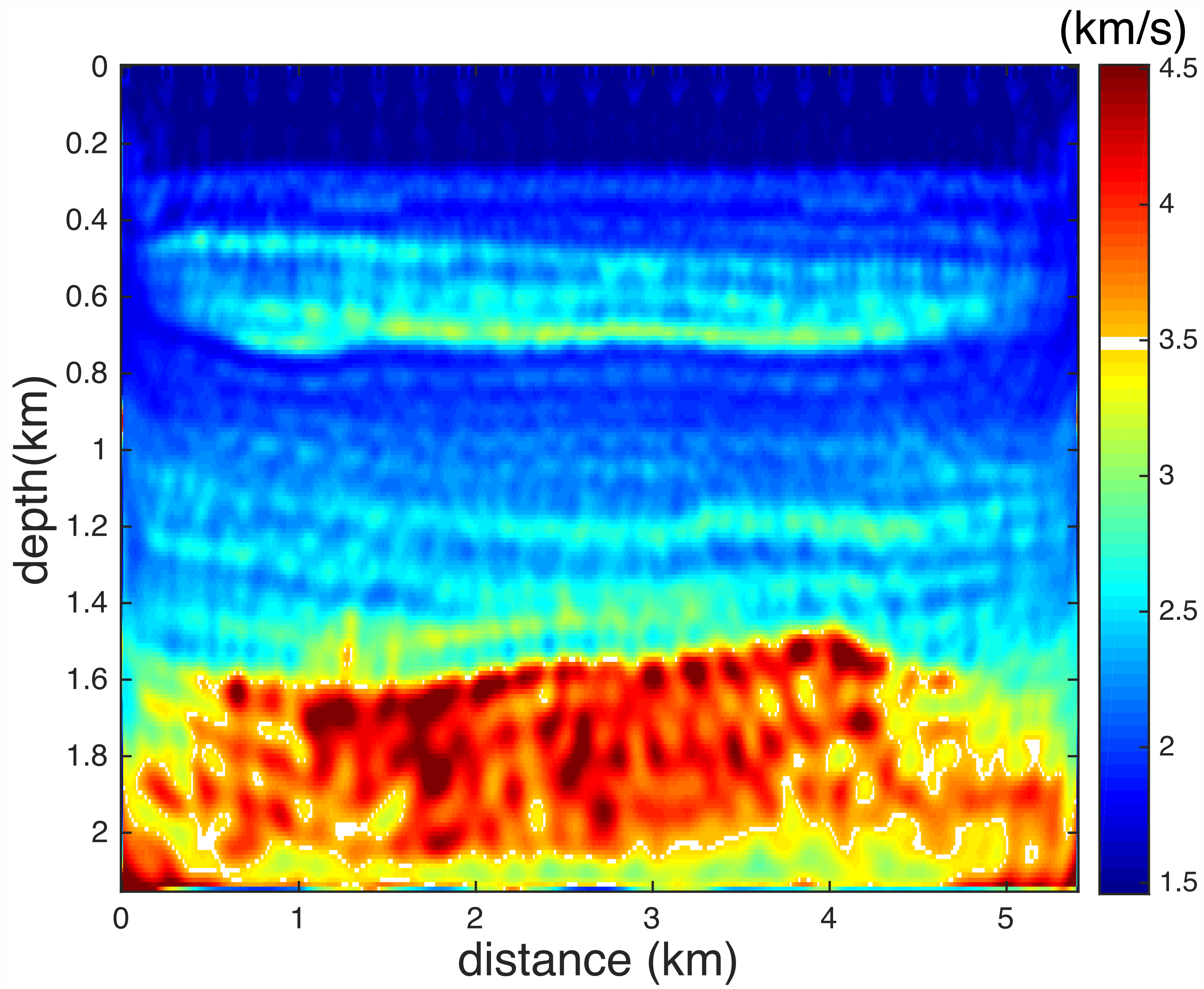
Test 2

- Frequency continuation using batches [3,3.5][3.5,4]....[14.5,15]Hz
- source spacing : 240m
- receiver spacing : 48m
- source depth: 12m
- True source q: ricker wavelet at 10Hz
- Source used for inversion: 0.8q
- Linear depth weighting

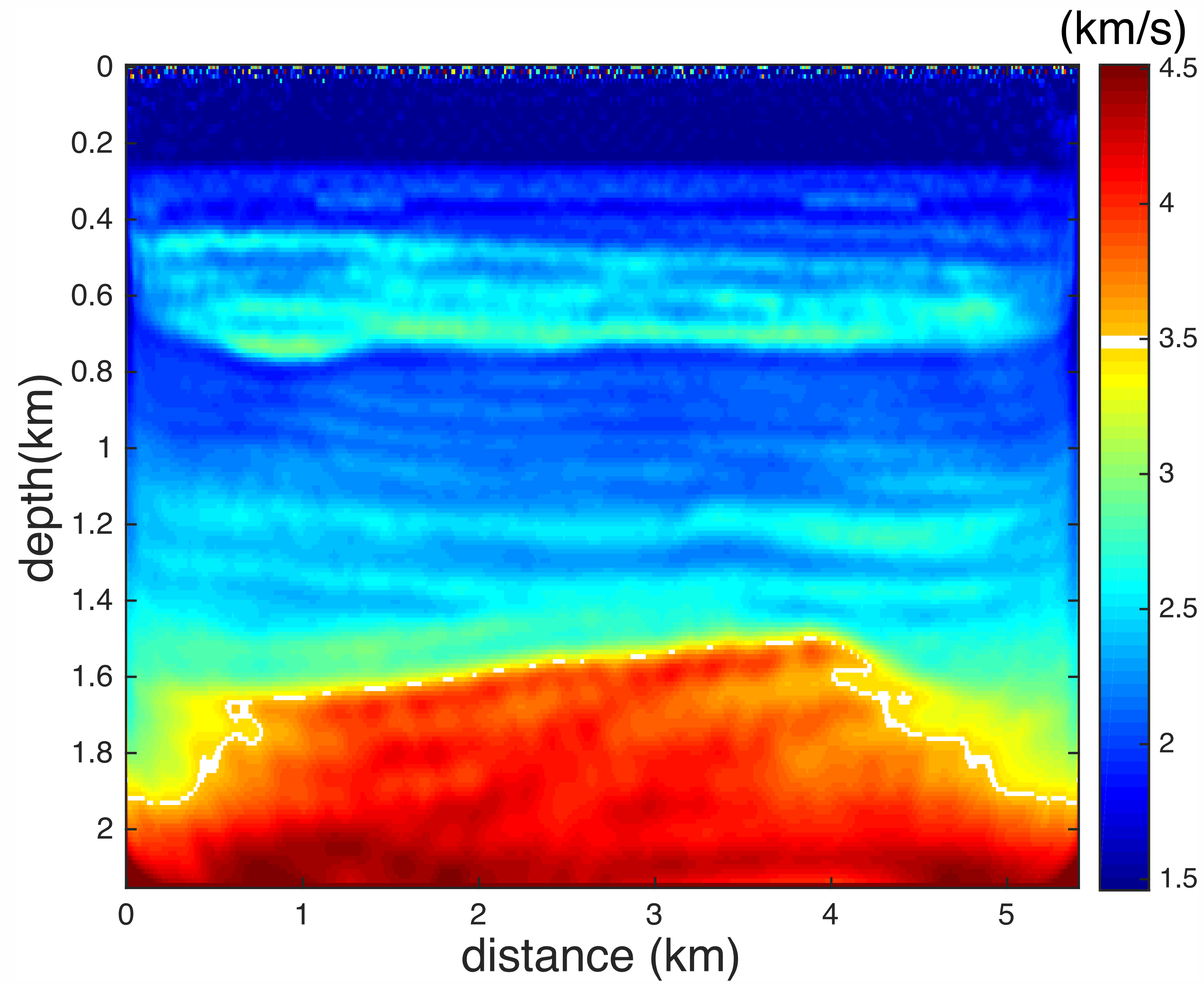




Test 2



Inverted model w/ FWI



Inverted model w/ FWI-DN with $\epsilon = 0$

Conclusion

- We proposed a denoising version of FWI
- We observed weighted/preconditioned PDE misfits dramatically increase robustness to modeling error
- The formulation makes incorporating prior knowledge of noise level convenient w/o increasing too much of complexity

Cost saving, time-domain dual formulation of WRI

Felix J. Herrmann, Mathias Louboutin, Peter Bas, Rongrong Wang, Emmanouil Daskalakis





SLIM
University of British Columbia

Time domain implementation through dual formulation

Primal formulation

$$\min_{\mathbf{U},\mathbf{m}} \|A(\mathbf{m})\mathbf{U} - \mathbf{Q}\|_2$$

$$\text{s.t. } \|P_\Omega \mathbf{U} - \mathbf{D}\|_2 \leq \epsilon$$

Dual formulation

$$\min_{\mathbf{m}} -\{\max_{\mathbf{y}} \frac{1}{2}\|F^T(\mathbf{y})\|_2^2 + \langle \mathbf{y}, \mathbf{D} - F\mathbf{Q}\rangle + \epsilon\|\mathbf{y}\|_2\}$$

$$\text{where } F = P_\Omega A^{-1}(\mathbf{m})$$

Variable Size: $O(N_x N_z N_T N_s)$ $O(N_r N_s N_T + N_x N_z)$

At optimal points: $A(\mathbf{m})\mathbf{U}^* = \mathbf{Q} - F^T\mathbf{y}^*$

Algorithm

$$G_{\mathbf{y}} = FF^T\mathbf{y} + \mathbf{D} - F\mathbf{Q} + \epsilon\frac{\mathbf{y}}{\|\mathbf{y}\|_2}$$

$$G_{\mathbf{m}} = \text{Jacobian}(\mathbf{m}, \mathbf{Q} - F^T\mathbf{y})$$

Step 1: L-BFGS on $\mathbf{y}$

Step 2: Gradient descent on $\mathbf{m}$

Algorithm

Using the following gradients:

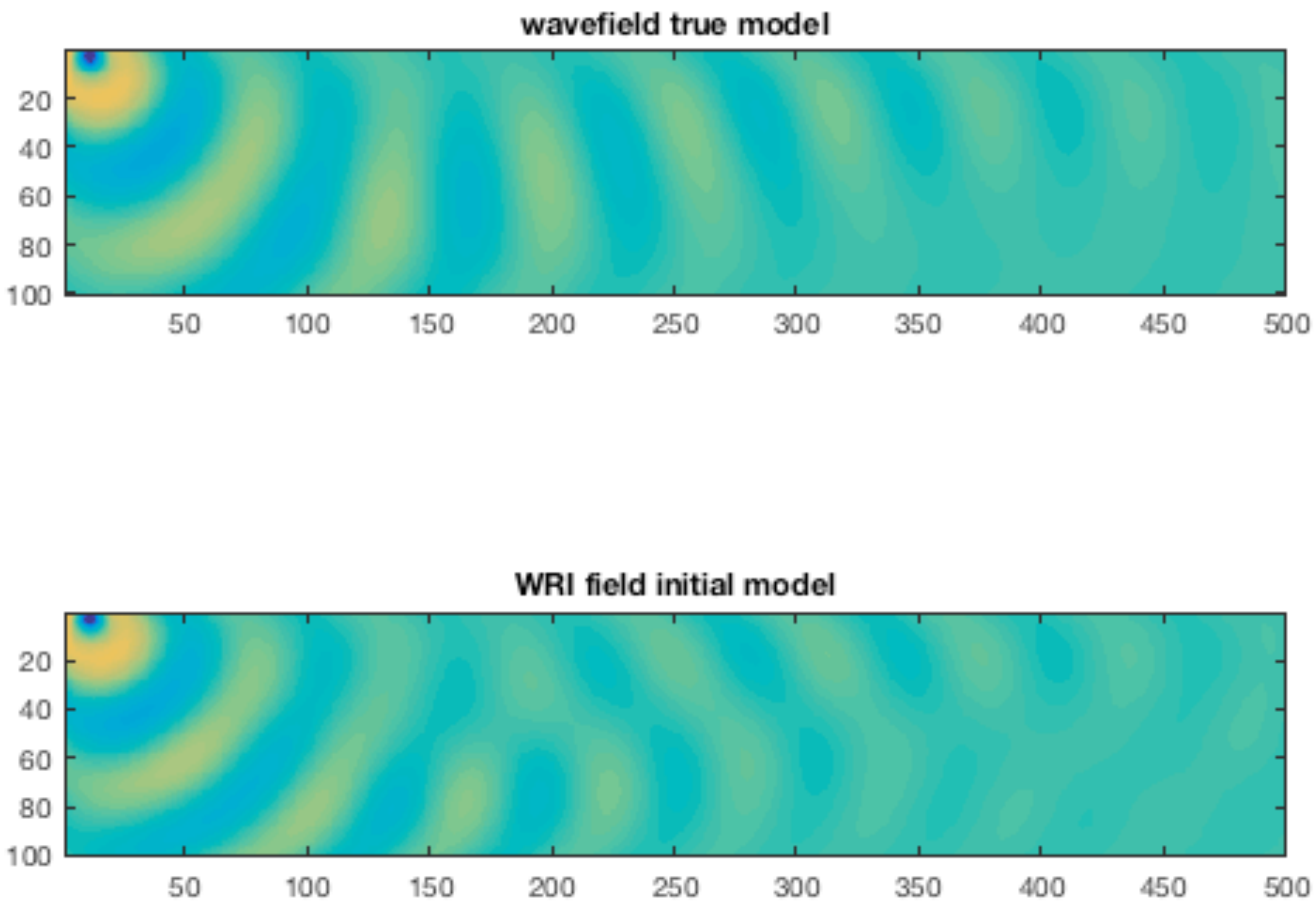
$$\frac{\partial p(m,y)}{\partial y} = FF^T y + D - FQ + \epsilon \frac{y}{||y||_2}$$

$$\frac{\partial p(m,y)}{\partial m} = J^T(m, \widetilde{Q})y$$

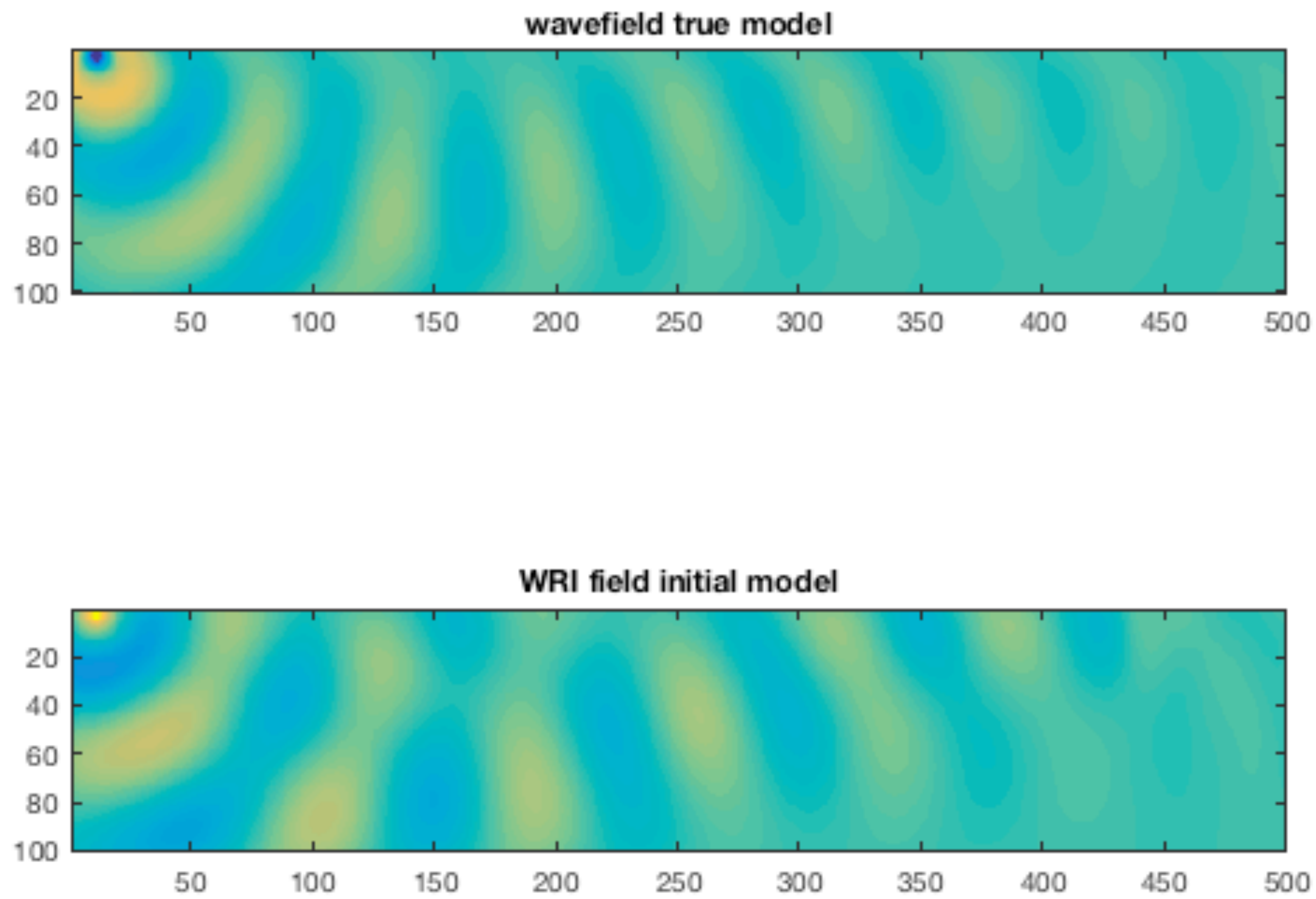
1. Solve with L-BFGS or gradient descent on the variable pair (m, y)

2. Alternating updates of the two variables

Preliminary results



Original time-harmonic WRI



New time-domain WRI