A denoising formulation of Full-Waveform Inversion

Rongrong Wang and Felix J. Herrmann

SLIM
University of British Columbia

Outline

1. Motivation

2. Formulation of the proposed method

3. Algorithm development

4. Numerical tests

SLIM

Motivation

Noises in observed data consist of:

modeling error

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

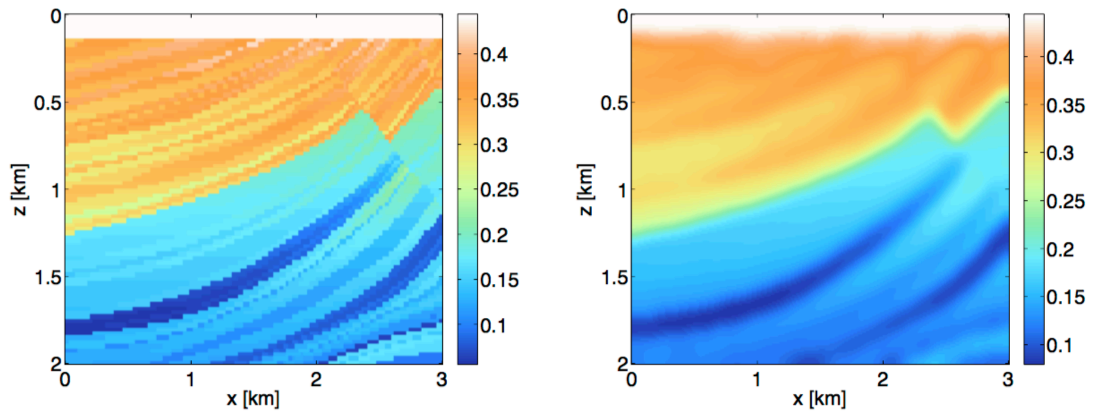
acquisition error

- timing error
- receiver location error
- measurement noise

interior of the domain

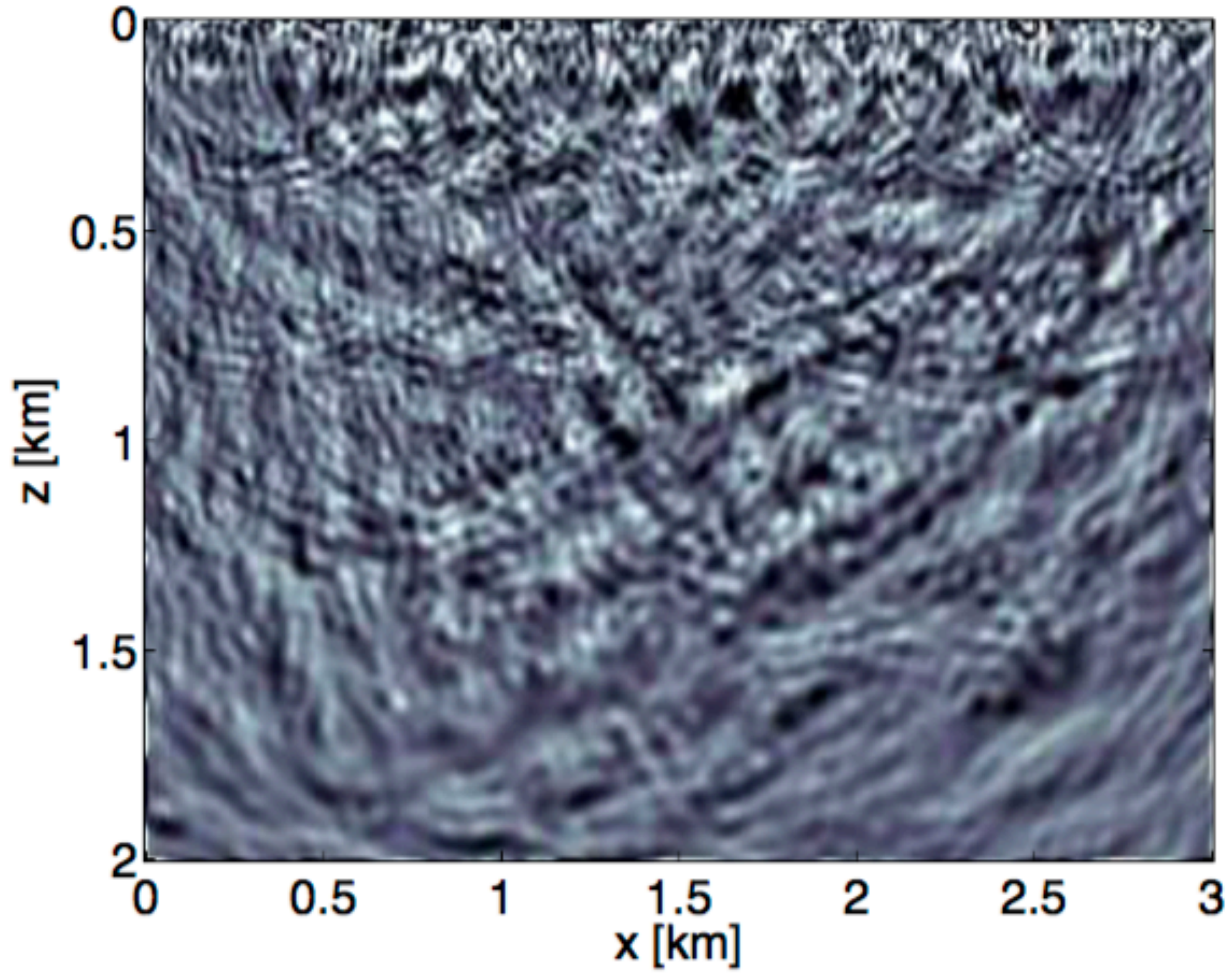
boundary

SLIM



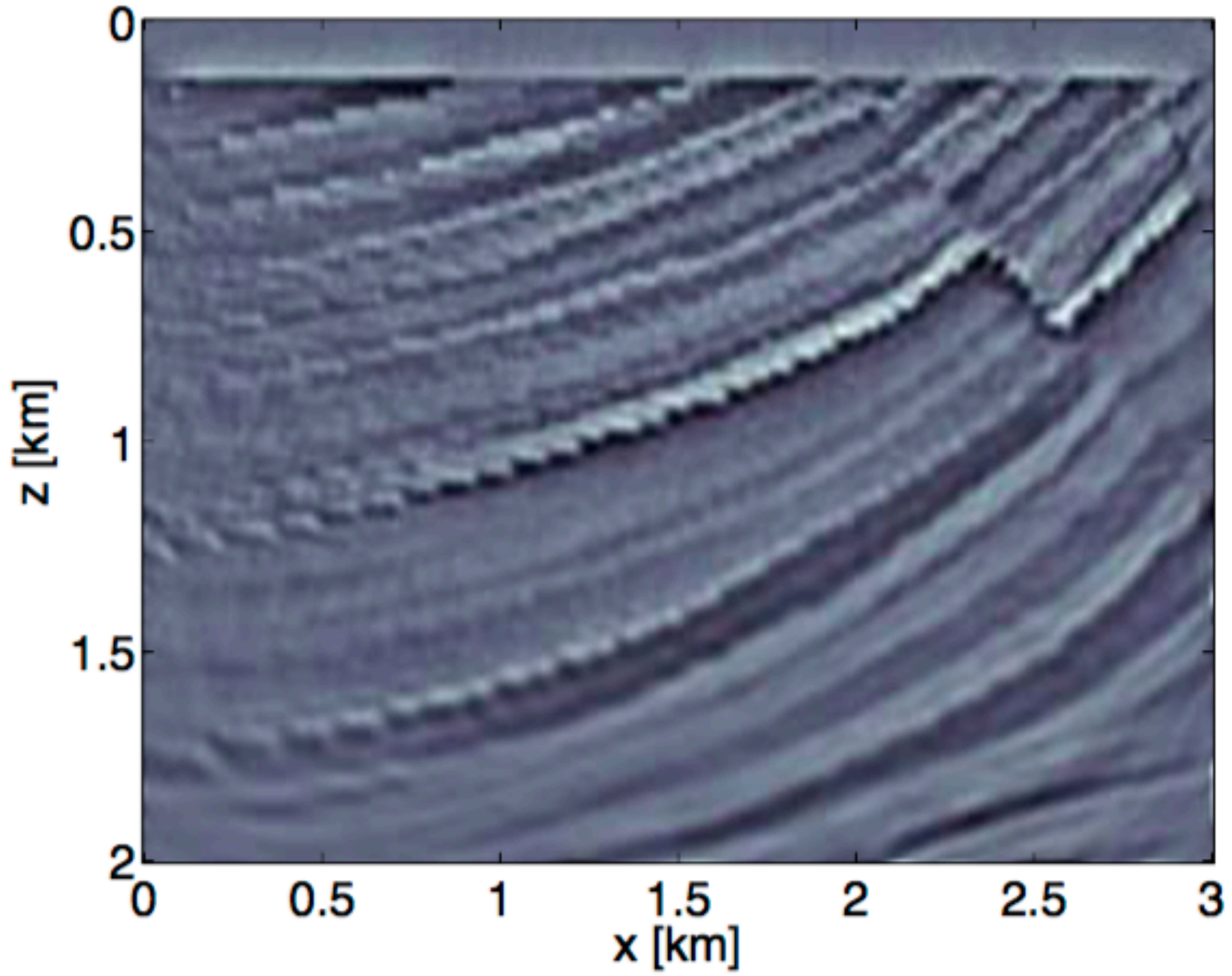
Motivation—the Failures of FWI

When measurement noise is ``spiky"



(a)

Model misfit for FWI



(b)

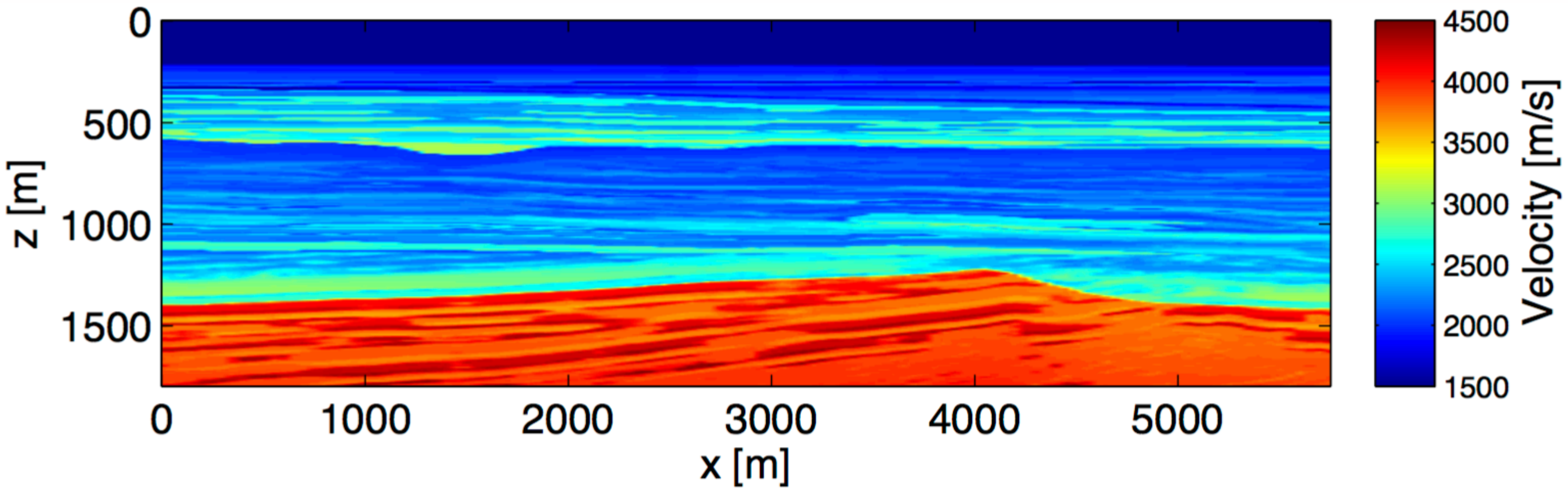
Model misfit for inversion with Student's t penalty

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

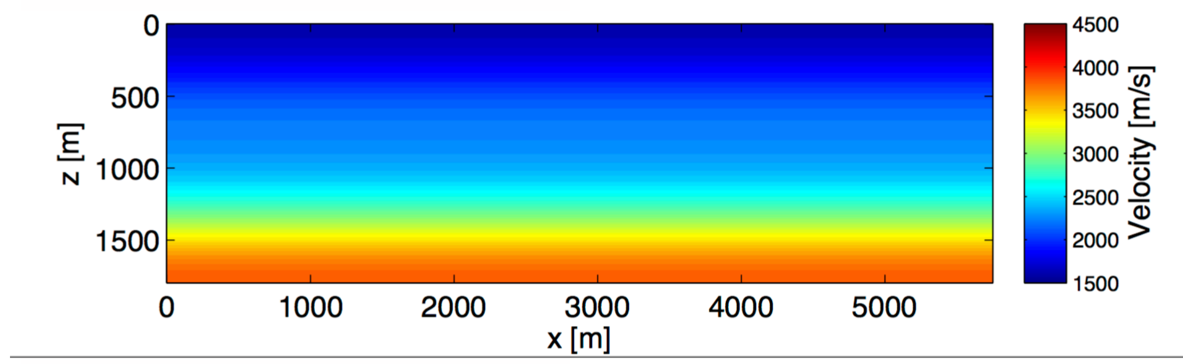
Motivation—the Failures of FWI

When water velocity is wrong

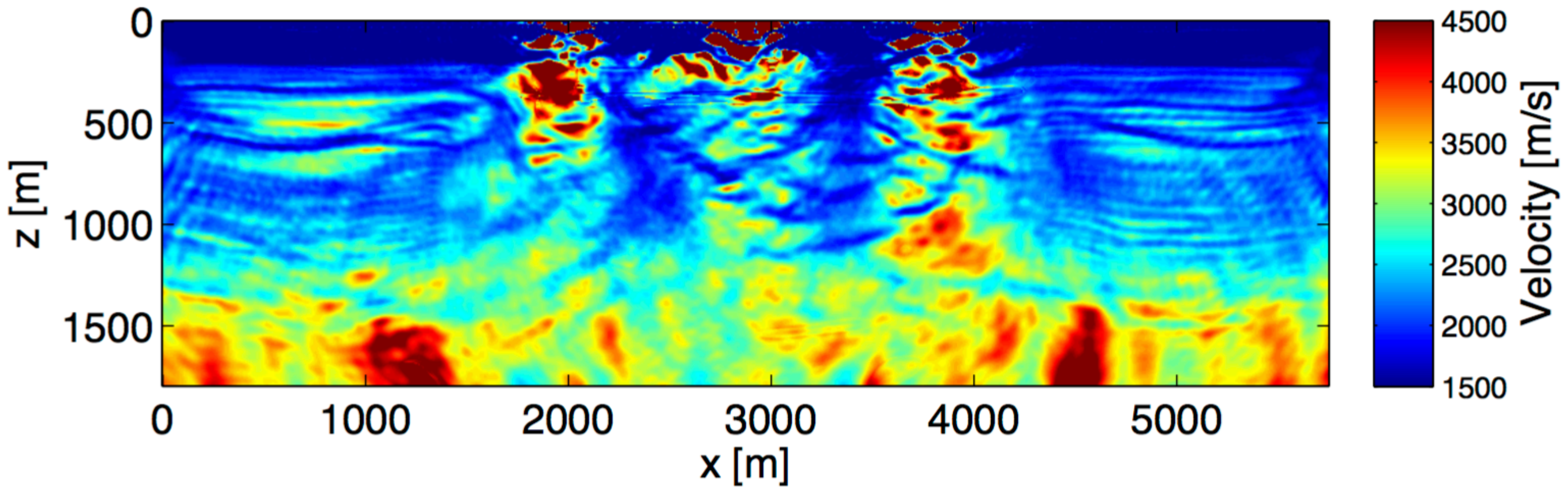
True model



Initial



FWI inversion



SLIM

FWI & its relaxation

FWI requires strict satisfaction of the PDE

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

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

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 assumed that noise is Gaussian distributed along sources and receivers
- Neglects modeling errors
- Cannot accommodate prior information of noise level
- Becomes problematic when water velocity is wrong

FWI & its relaxation

Direct relaxation of PDE constraint

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

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

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

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

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

Hard to choose!

Flip the objective and the constraint

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

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

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

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

Noise level

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

FWI & its relaxation

Direct relaxation of PDE constraint

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

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

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

Noise level

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

Decomposing the wavefields variable

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

Boundary part

Interior part

FWI & its relaxation

Direct relaxation of PDE constraint

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

Hard to choose!

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

Flip the objective and the constraint

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

Noise level

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

Decomposing the wavefields variable

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

Interior part

Boundary part

FWI & its relaxation

Direct relaxation of PDE constraint

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

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

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

Decomposing the wavefields variable

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

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)

We call it the 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
- Impose 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

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

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

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

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

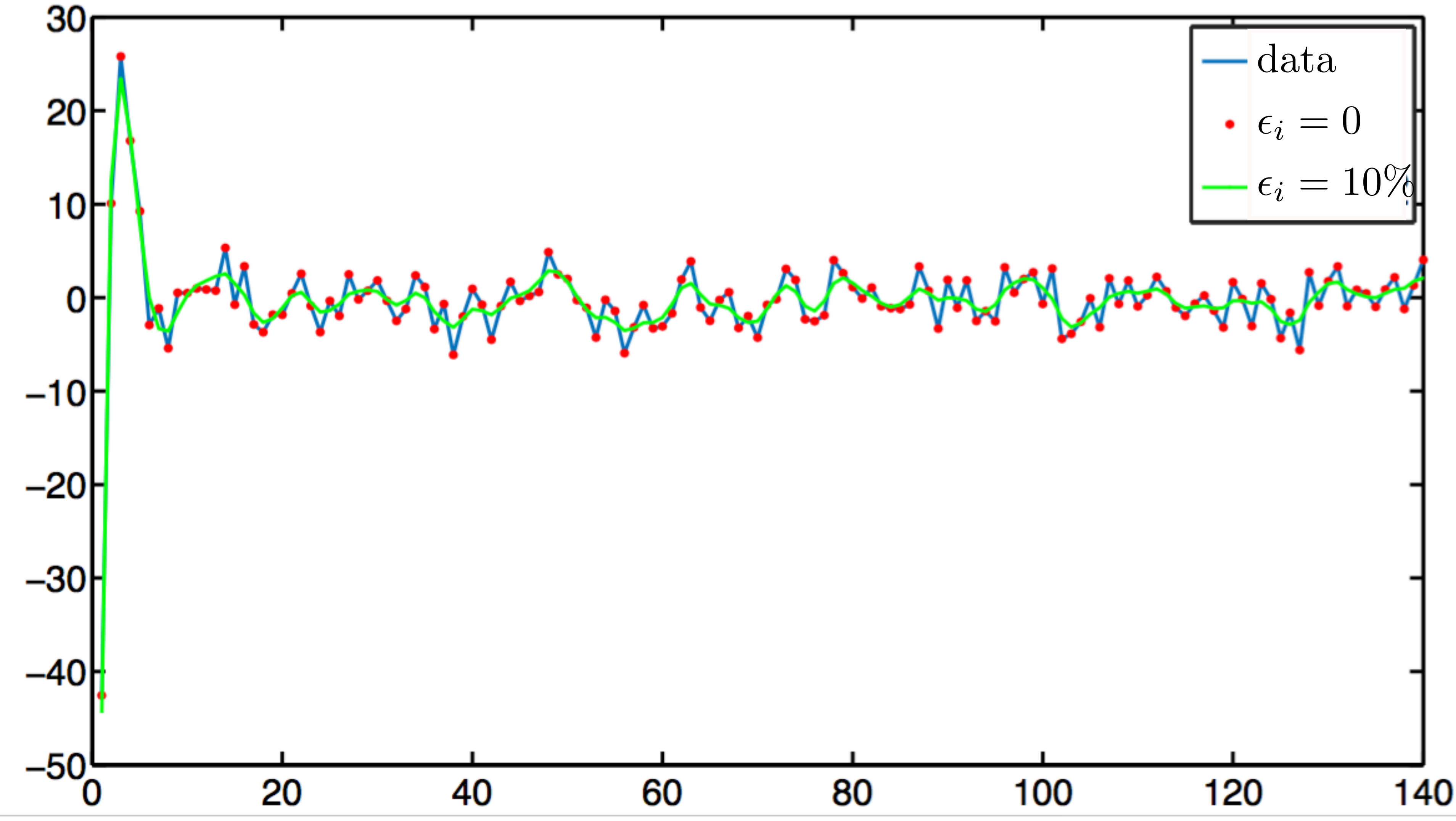
$$\phi(\lambda) = \min_{u_i} \|\mathcal{D}_z(A(m)u_i - q_i)\|_2^2 + \lambda\|P_{\Omega_i} u_i - d_i\|_2^2 - \lambda\epsilon_i$$

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

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



Denoising effect of (P_d)

The Lagrangian dual of (P_d) is

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

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

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

Solve by variable projection

[Aravkin and van Leeuwen, 2012]

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

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

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

performs 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

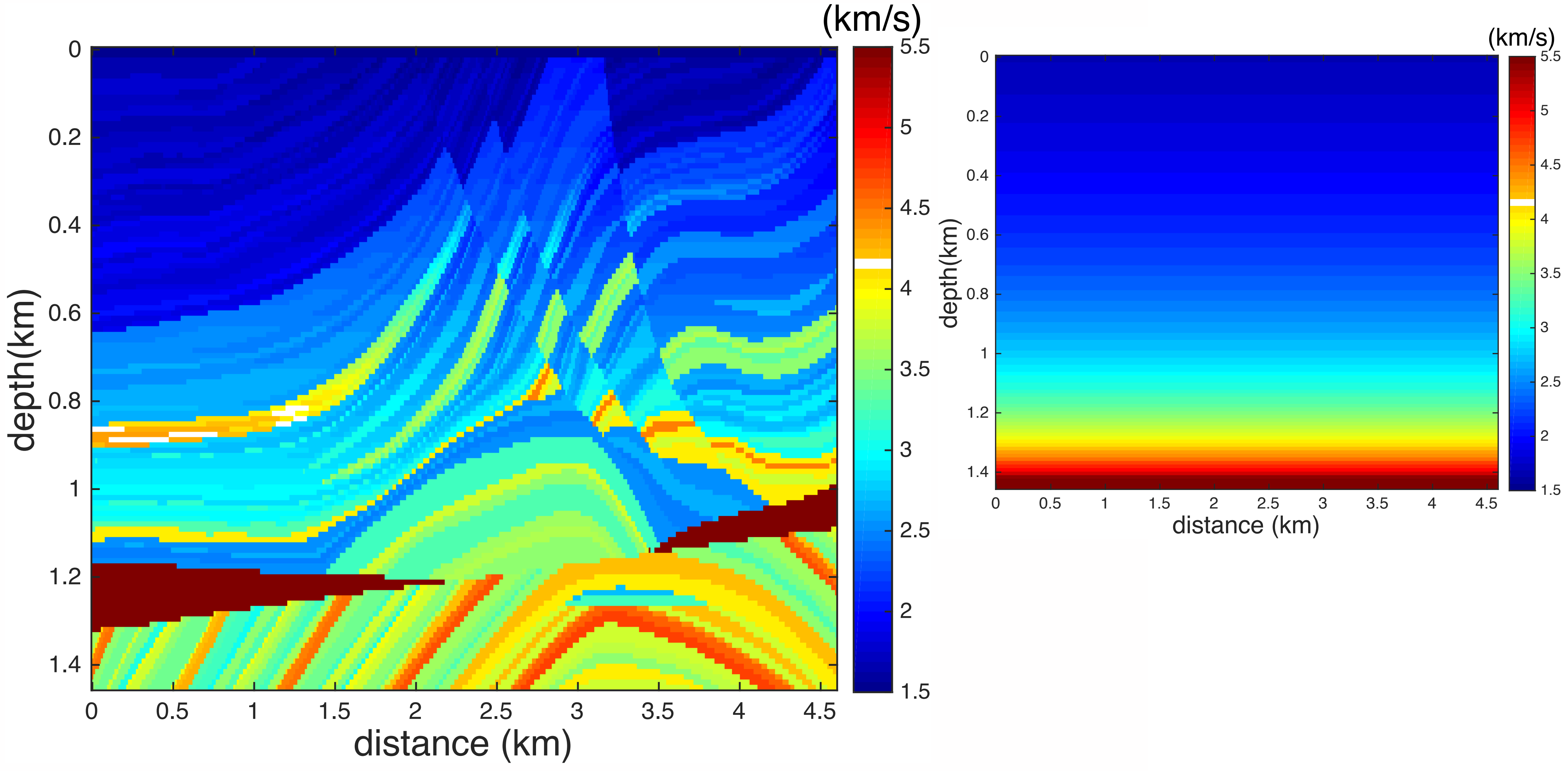
Case Study

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

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



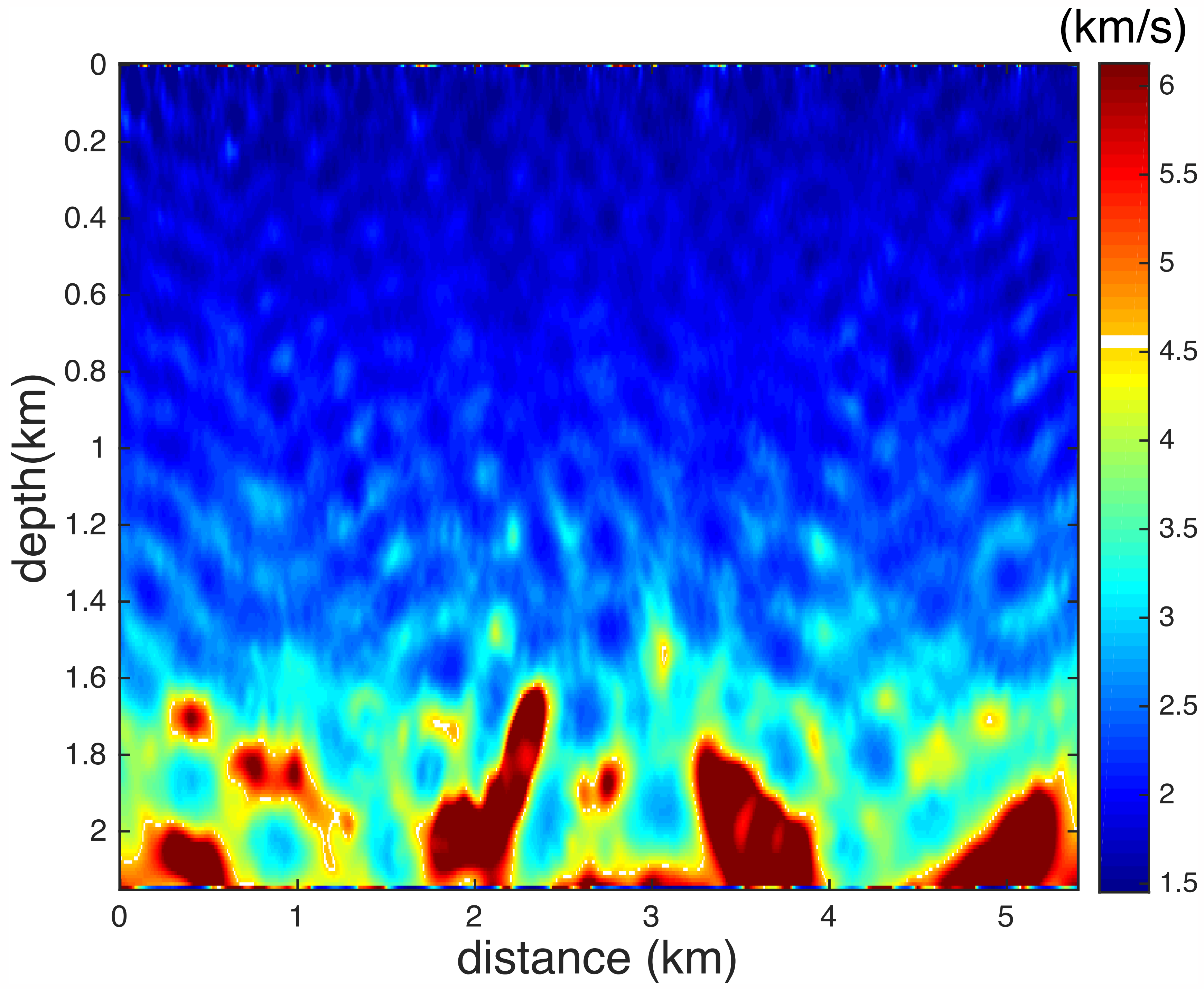
Test 1

Method for comparison: weighted FWI

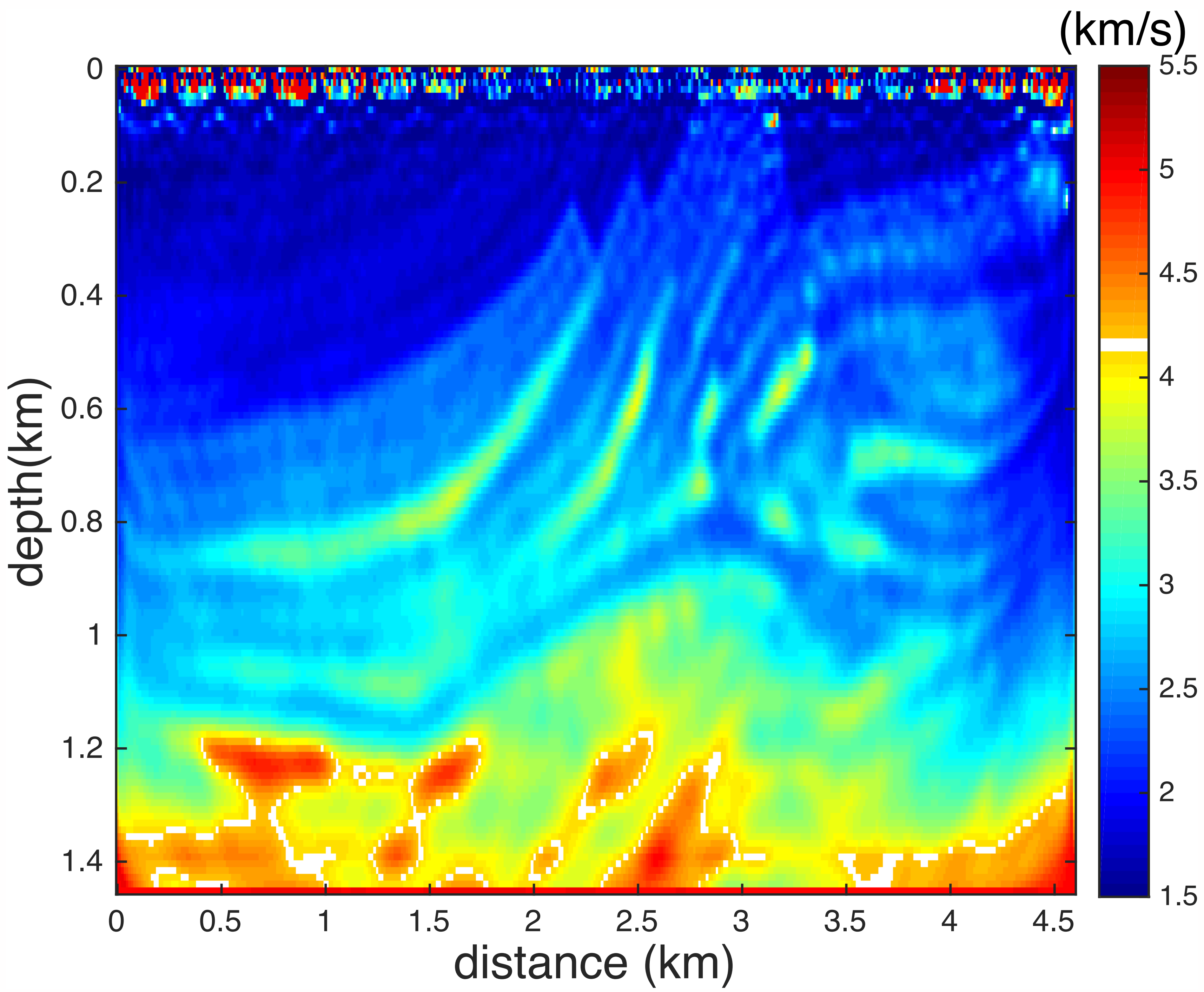
$$\min_{m} \sum_{i \in N_1} 9\|P_{\Omega_i} A^{-1}(m) q_i - d_i\|_2^2 + \min_{m} \sum_{i \in N_2} \|P_{\Omega_i} A^{-1}(m) q_i - d_i\|_2^2$$

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

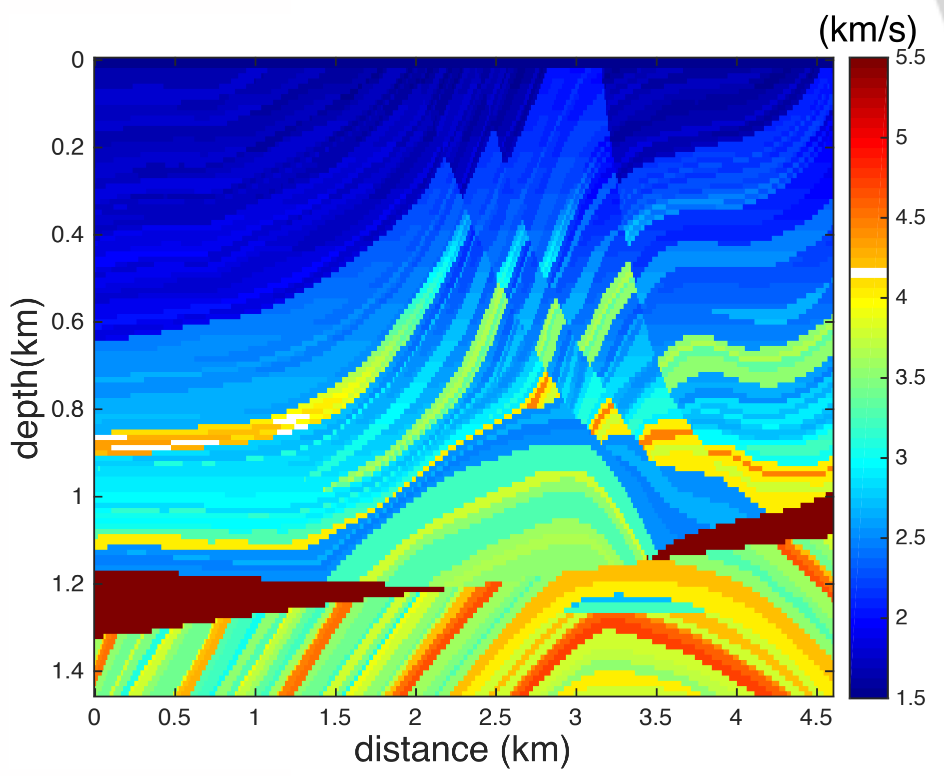
Test 1



Inverted model w/ weighted FWI



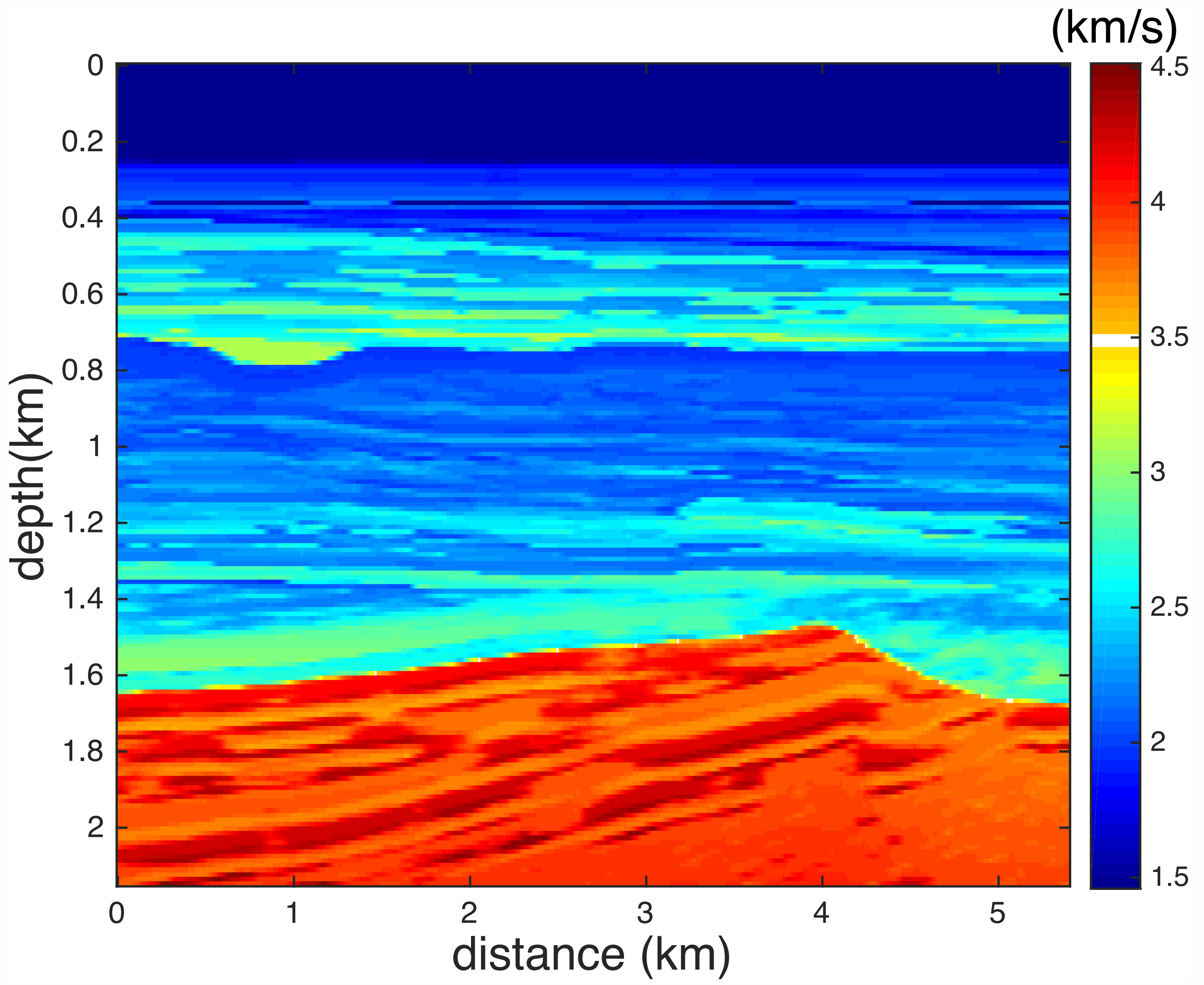
Inverted model w/ FWI-DN

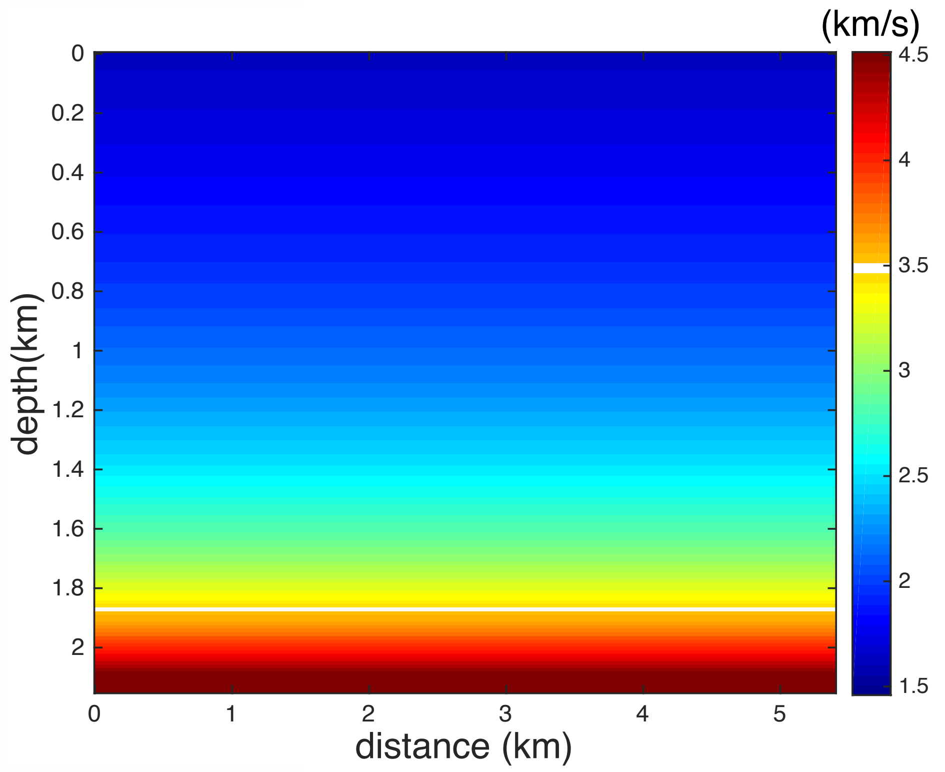


Test 2: robustness under modeling error

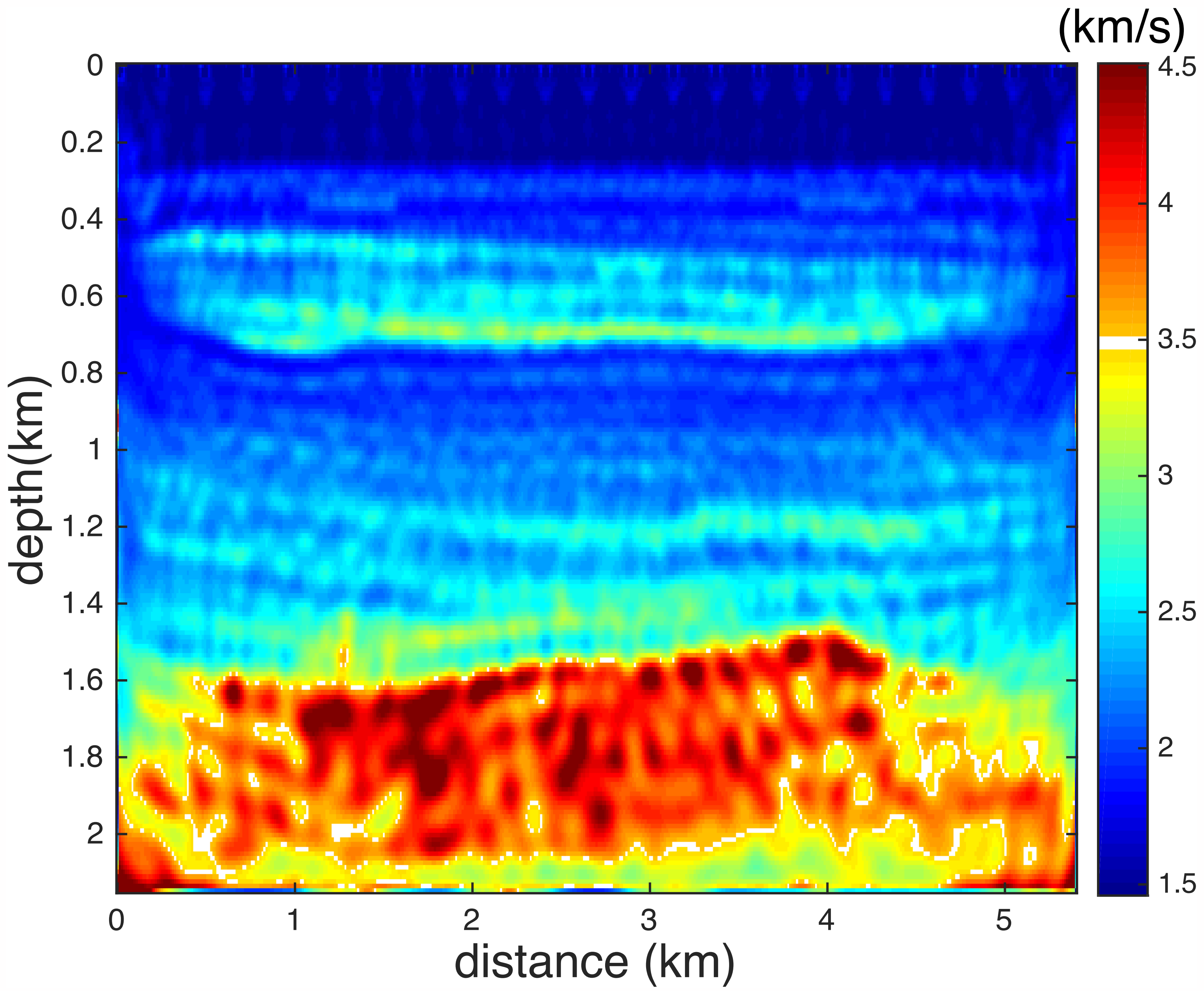
Test 2

- 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

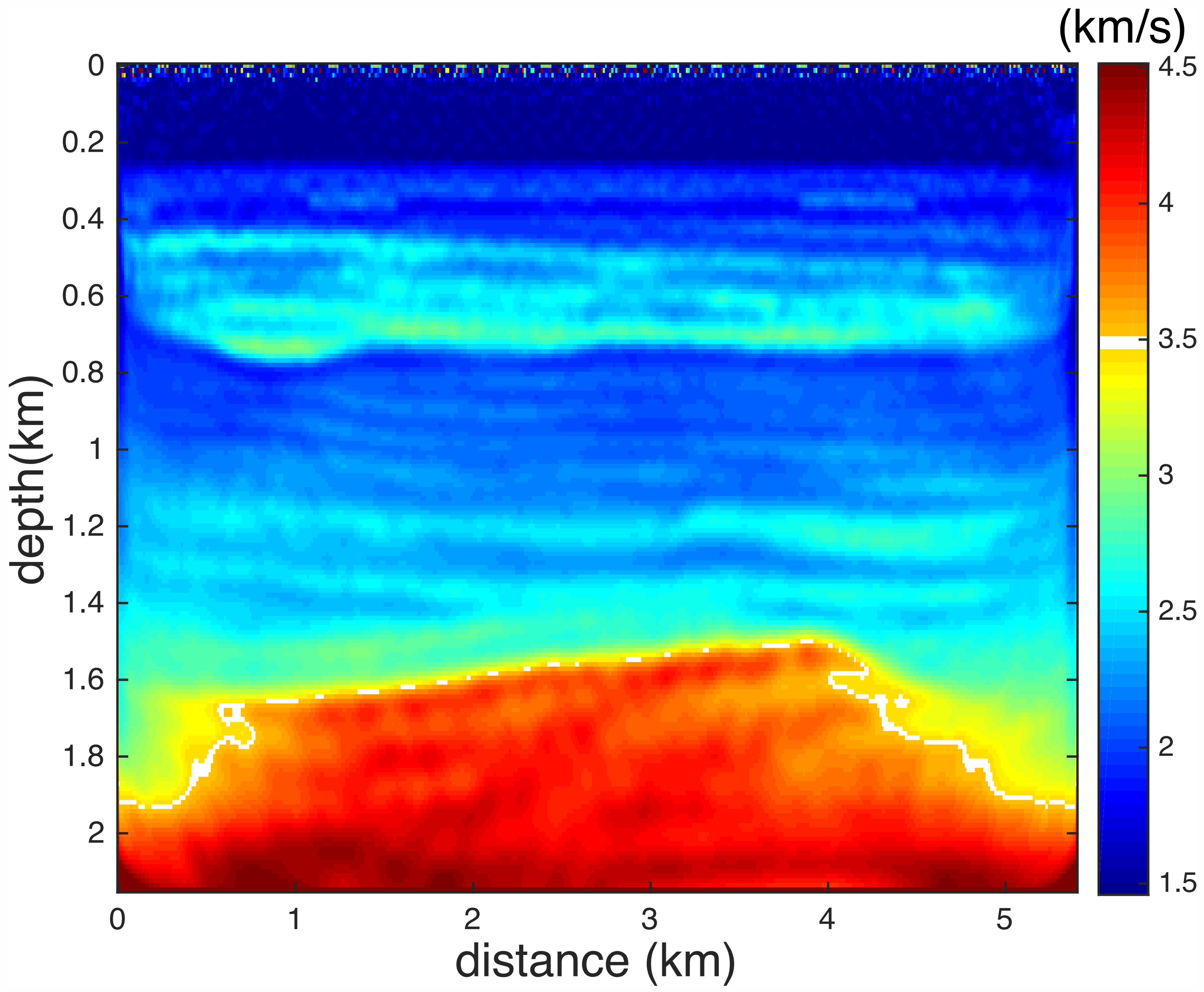




Test 2



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

Acknowledgements

This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.