Overview research at the SINBAD Consortium

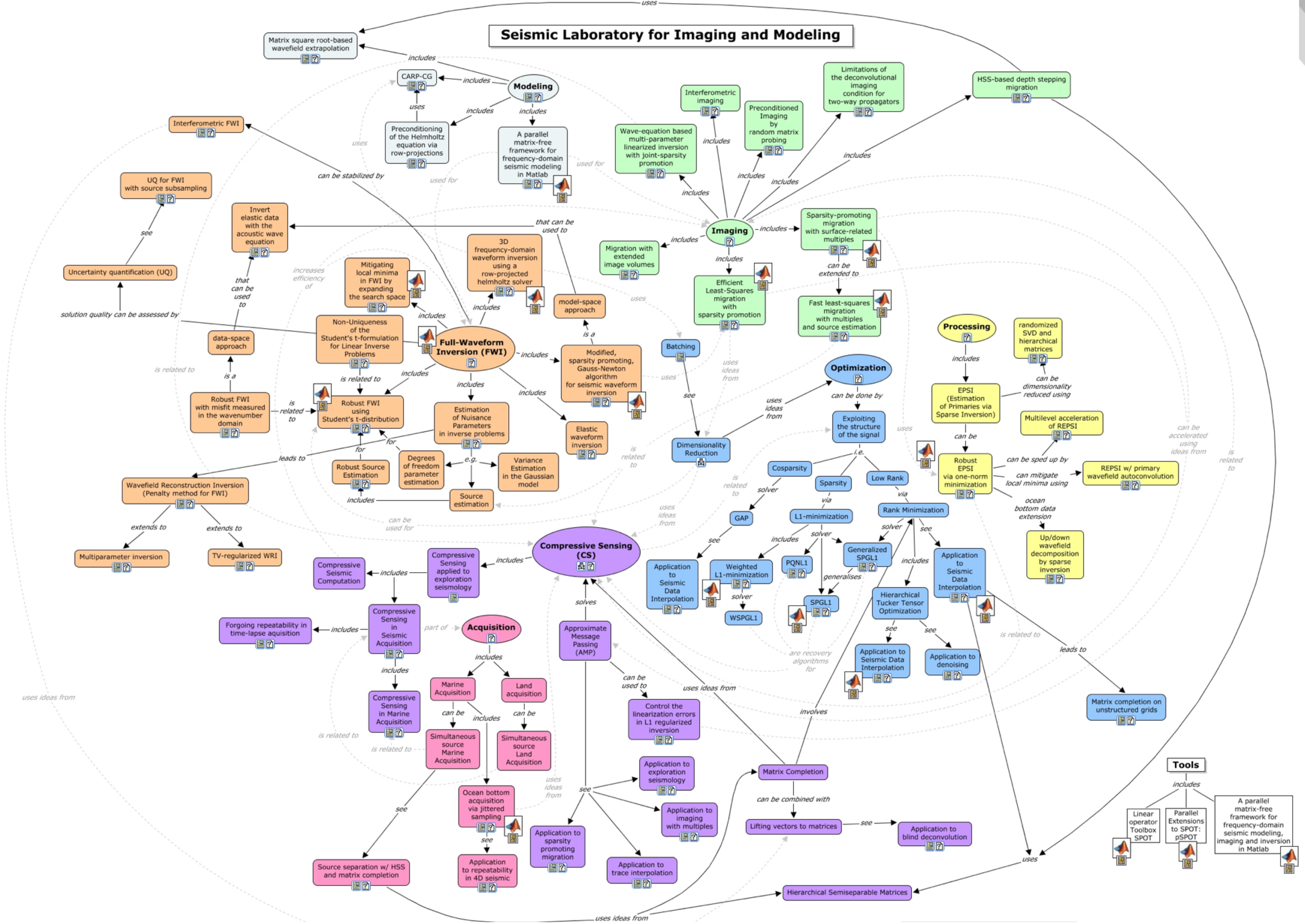
Felix J. Herrmann

Schlumberger Gould, Cambridge, March 17, 2016

SLIM
University of British Columbia

The team...

Total of 25 (under)graduate students, PDFs, visitors, faculty, and staff...


Seismic Laboratory for Imaging and Modeling
Modeling
Matrix square root-based wavefield extrapolation
CARP-CG
Preconditioning of the Helmholtz equation via row-projections
A parallel matrix-free framework for frequency-domain seismic modeling in Matlab
Interferometric FWI
UQ for FWI with source subsampling
Uncertainty quantification (UQ)
Invert elastic data with the acoustic wave equation
Mitigating local minima in FWI by expanding the search space
3D frequency-domain waveform inversion using a row-projected helmholtz solver
model-space approach
data-space approach
Non-Uniqueness of the Student's t-formulation for Linear Inverse Problems
Full-Waveform Inversion (FWI)
Modified, sparsity promoting, Gauss-Newton algorithm for seismic waveform inversion
Robust FWI with misfit measured in the wavenumber domain
Robust FWI using Student's t-distribution
Estimation of Nuisance Parameters in inverse problems
Elastic waveform inversion
Robust Source Estimation
Degrees of freedom parameter estimation
Variance Estimation in the Gaussian model
Source estimation
Wavefield Reconstruction Inversion (Penalty method for FWI)
Multiparameter inversion
TV-regularized WRI
Imaging
Interferometric imaging
Preconditioned Imaging by random matrix probing
Limitations of the deconvolutional imaging condition for two-way propagators
HSS-based depth stepping migration
Wave-equation based multi-parameter linearized inversion with joint-sparsity promotion
Sparsity-promoting migration with surface-related multiples
Migration with extended image volumes
Efficient Least-Squares migration with sparsity promotion
Fast least-squares migration with multiples and source estimation
Batching
Dimensionality Reduction
Processing
randomized SVD and hierarchical matrices
EPSI (Estimation of Primaries via Sparse Inversion)
Multilevel acceleration of REPSI
Robust EPSI via one-norm minimization
REPSI w/ primary wavefield autoconvolution
Up/down wavefield decomposition by sparse inversion
Optimization
Exploiting the structure of the signal
Cosparsity
Sparsity
Low Rank
GAP
L1-minimization
Rank Minimization
Application to Seismic Data Interpolation
Weighted L1-minimization
PQNL1
Generalized SPGL1
SPGL1
WSPGL1
Hierarchical Tucker Tensor Optimization
Application to Seismic Data Interpolation
Application to denoising
Application to Seismic Data Interpolation
Matrix completion on unstructured grids
Compressive Sensing (CS)
Compressive Seismic Computation
Compressive Sensing applied to exploration seismology
Compressive Sensing in Seismic Acquisition
Forgoing repeatability in time-lapse aquisition
Acquisition
Compressive Sensing in Marine Acquisition
Marine Acquisition
Land acquisition
Simultaneous source Marine Acquisition
Simultaneous source Land Acquisition
Ocean bottom acquisition via jittered sampling
Application to repeatability in 4D seismic
Source separation w/ HSS and matrix completion
Approximate Message Passing (AMP)
Control the linearization errors in L1 regularized inversion
Application to exploration seismology
Application to imaging with multiples
Application to sparsity promoting migration
Application to trace interpolation
Matrix Completion
Lifting vectors to matrices
Application to blind deconvolution
Hierarchical Semiseparable Matrices
Tools
Linear operator Toolbox SPOT
Parallel Extensions to SPOT: pSPOT
A parallel matrix-free framework for frequency-domain seismic modeling, imaging and inversion in Matlab
uses
includes
can be stabilized by
used for
that can be used to
see
solution quality can be assessed by
that can be used to
is a
is related to
leads to
for
e.g.
extends to
can be extended to
can be done by
uses ideas from
i.e.
solver
via
generalises
are recovery algorithms for
can be
can be dimensionality reduced using
can be sped up by
can mitigate local minima using
ocean bottom data extension
can be accelerated using ideas from
solves
part of
can be used to
involves
can be combined with
increases efficiency of
can be used for

Seismic Laboratory for Imaging and Modeling |

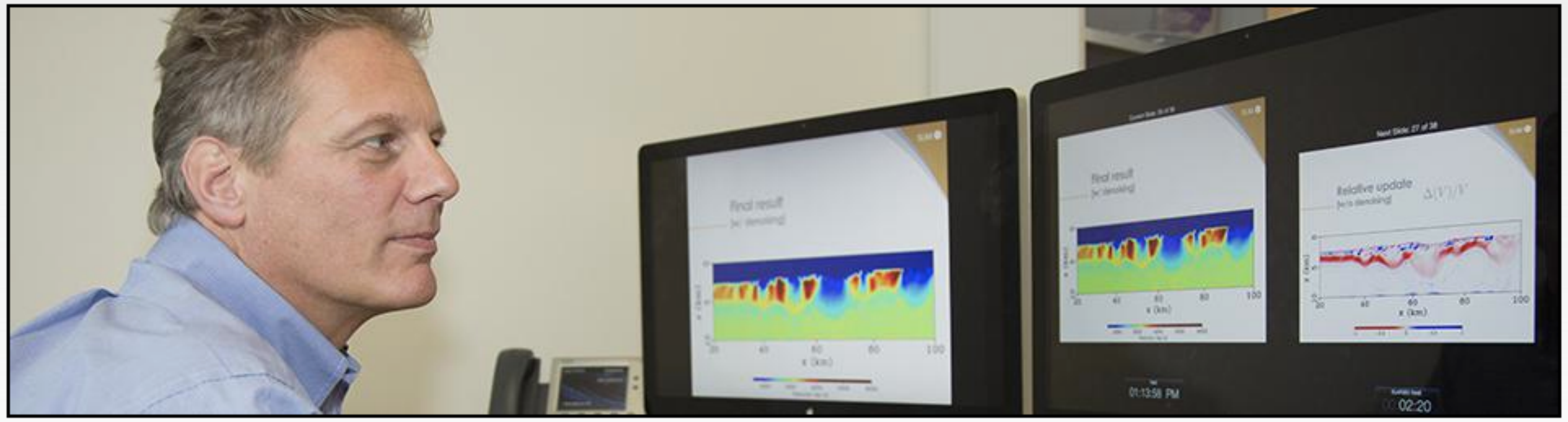
https www.slim.eos.ubc.ca

News ▾ Apple ▾ Reviewing ▾ Funding ▾ Professional ▾ SLIM ▾ EOS ▾ Fun ▾ Software ▾ Research ▾ Hobbies ▾ UBC ▾ Teaching ▾ Travel ▾ Research ▾ Agenda ▾ Your Telecom

Seismic Laboratory for Imaging and Modeling

UBC

ABOUT US | PROJECTS | EVENTS | PUBLICATIONS | RESEARCH | SOFTWARE | CONSORTIUM | INTERNAL



SEARCH

Upcoming events

Mon, Aug 31st, 2015
Inaugural Full-Waveform Inversion Workshop, Brazil

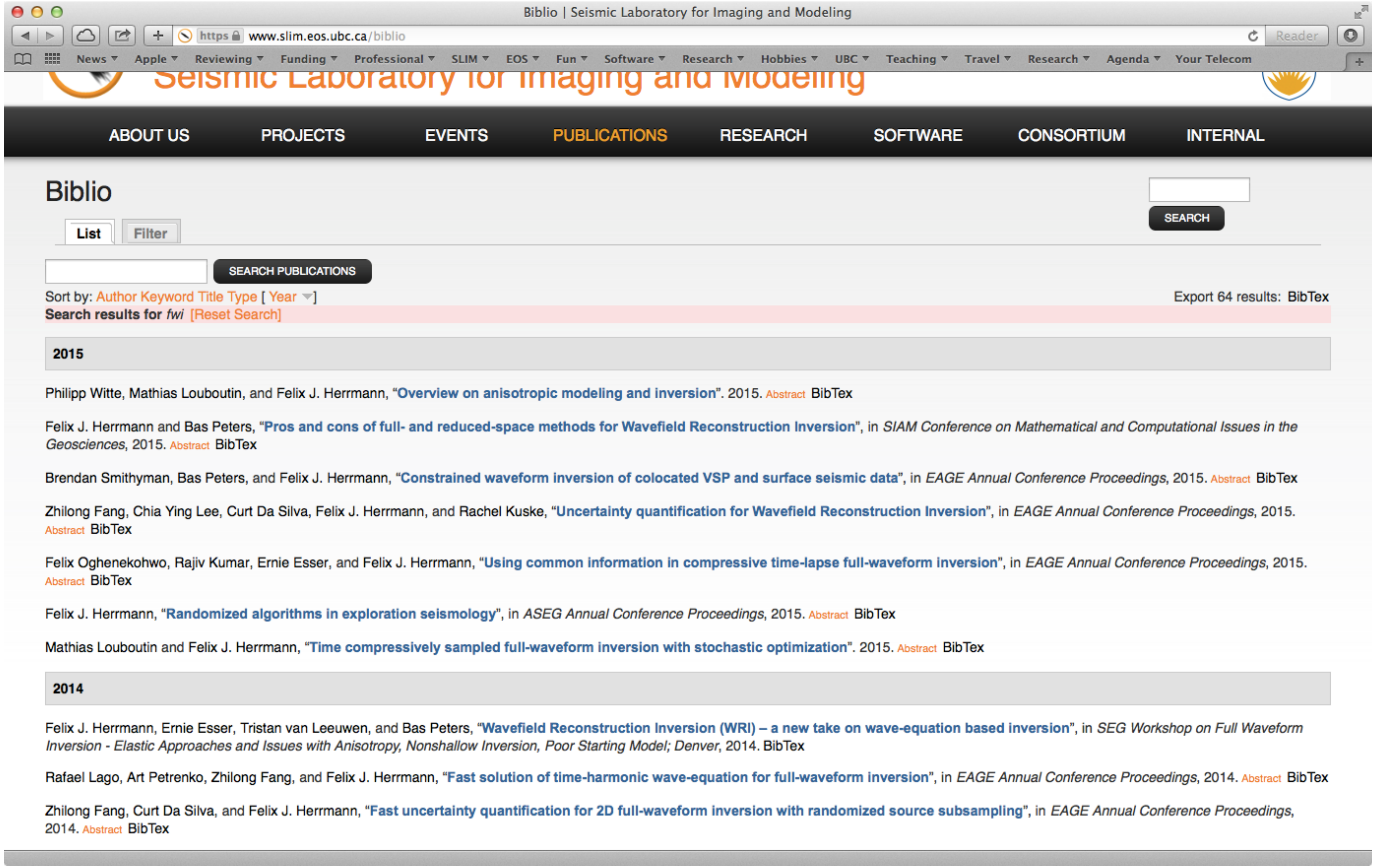
Wed, Sep 9th, 2015
Hansruedi Maurer, ETH Zurich "The curse of dimensionality in exploring the subsurface" 4:00 PM, ESB 5104 - 2207 Main Mall, UBC Campus

more

SINBAD Consortium Meeting Fall 2015

New Publications

- **Affordable full subsurface image volume—an application to WEMVA** *Conference (EAGE Workshop on Wave Equation based Migration Velocity Analysis, Madrid)*
- **Irregular grid tensor completion** *Conference (Workshop on Low-rank Optimization and Applications, University of Bonn, Germany)*
- **Wavefield-denoising and source encoding** *Conference (SIAM Conference on Mathematical and Computational Issues in the Geosciences, Stanford University, California)*
- **Sparsity promoting seismic imaging and full-waveform inversion** *Thesis (PhD)*
- **Total variation regularization strategies in full waveform inversion for improving robustness to noise, limited data and poor initializations** *Tech Report*
- **Sparse least-squares seismic imaging with source estimation utilizing multiples** *Conference (PIMS Workshop on Advances in Seismic Imaging and Inversion, University of Alberta, Edmonton)*
- **A new take on compressive time-lapse seismic acquisition, imaging and inversion** *Conference (PIMS Workshop on Advances in Seismic Imaging and Inversion, University of Alberta, Edmonton)*
- **Compressive time-lapse seismic data processing using shared information** *Conference (CSEG,*



Biblio | Seismic Laboratory for Imaging and Modeling

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

News Apple Reviewing Funding Professional SLIM EOS Fun Software Research Hobbies UBC Teaching Travel Research Agenda Your Telecom

Seismic Laboratory for Imaging and Modeling

ABOUT US PROJECTS EVENTS PUBLICATIONS RESEARCH SOFTWARE CONSORTIUM INTERNAL

Biblio

SEARCH

List Filter

SEARCH PUBLICATIONS

Sort by: Author Keyword Title Type [Year]

Export 64 results: BibTex

Search results for *fwi* [Reset Search]

2015

Philipp Witte, Mathias Louboutin, and Felix J. Herrmann, "**Overview on anisotropic modeling and inversion**". 2015. Abstract BibTex

Felix J. Herrmann and Bas Peters, "**Pros and cons of full- and reduced-space methods for Wavefield Reconstruction Inversion**", in *SIAM Conference on Mathematical and Computational Issues in the Geosciences*, 2015. Abstract BibTex

Brendan Smithyman, Bas Peters, and Felix J. Herrmann, "**Constrained waveform inversion of colocated VSP and surface seismic data**", in *EAGE Annual Conference Proceedings*, 2015. Abstract BibTex

Zhilong Fang, Chia Ying Lee, Curt Da Silva, Felix J. Herrmann, and Rachel Kuske, "**Uncertainty quantification for Wavefield Reconstruction Inversion**", in *EAGE Annual Conference Proceedings*, 2015. Abstract BibTex

Felix Oghenekohwo, Rajiv Kumar, Ernie Esser, and Felix J. Herrmann, "**Using common information in compressive time-lapse full-waveform inversion**", in *EAGE Annual Conference Proceedings*, 2015. Abstract BibTex

Felix J. Herrmann, "**Randomized algorithms in exploration seismology**", in *ASEG Annual Conference Proceedings*, 2015. Abstract BibTex

Mathias Louboutin and Felix J. Herrmann, "**Time compressively sampled full-waveform inversion with stochastic optimization**". 2015. Abstract BibTex

2014

Felix J. Herrmann, Ernie Esser, Tristan van Leeuwen, and Bas Peters, "**Wavefield Reconstruction Inversion (WRI) – a new take on wave-equation based inversion**", in *SEG Workshop on Full Waveform Inversion - Elastic Approaches and Issues with Anisotropy, Nonshallow Inversion, Poor Starting Model; Denver*, 2014. BibTex

Rafael Lago, Art Petrenko, Zhilong Fang, and Felix J. Herrmann, "**Fast solution of time-harmonic wave-equation for full-waveform inversion**", in *EAGE Annual Conference Proceedings*, 2014. Abstract BibTex

Zhilong Fang, Curt Da Silva, and Felix J. Herrmann, "**Fast uncertainty quantification for 2D full-waveform inversion with randomized source subsampling**", in *EAGE Annual Conference Proceedings*, 2014. Abstract BibTex

Software releases

https://sinbad.eos.ubc.ca/SoftwareReleases/highlights

Available at

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

SLIM

Seismic Laboratory for Imaging and Modeling

ABOUT US | PROJECTS | EVENTS | PUBLICATIONS | RESEARCH | SOFTWARE | CONSORTIUM | INTERNAL

SEARCH

Downloads

Bug Report

Applications in SINBAD software release

Thu Mar 3 16:21:33 PST 2016

1. **Acquisition**
   a. 2D ocean-bottom marine acquisition via jittered sampling [Read More] [GitHub]
   b. Rank minimization based source-separation in time-jittered marine acquisition [Read More] [GitHub]
   c. Source separation via SVD-free rank minimization in the hierarchical semi-separable representation [Read More] [GitHub]
   d. Time-jittered blended marine acquisition on non-uniform grids [Read More] [GitHub]
   e. Joint recovery method for time-lapse seismic data [Read More] [GitHub]
2. **Imaging**
   a. Efficient least-squares imaging with sparsity promotion and compressive sensing [Read More] [GitHub]
   b. Fast imaging with surface-related multiples by sparse inversion (update in master branch) [Read More] [GitHub]
   c. Fast least-squares imaging with source estimation using multiples (update in master branch) [Read More] [GitHub]
   d. Time domain LSRTM with sparsity promotion (new in master branch) [Read More] [Video] [GitHub]
   e. Wavefield reconstruction imaging [Read More] [GitHub]
3. **Modeling**
   a. Tutorial for 2D Frequency-domain acoustic modeling and imaging [Read More] [GitHub]
   b. 3D Frequency-Domain Modeling Kernel [Read More] [GitHub]
   c. Tutorial for time-domain 2D/3D acoustic modeling [Read More] [GitHub]

SLIM

SINBAD Consortium (SLIM)

github.com/SINBADconsortium/

News Apple Reviewing Funding Professional SLIM EOS Fun Software Research Hobbies UBC Teaching Travel Research Agenda Your Telecom

Santa Teresa Hotel – Rio de Janeiro – Deluxe Tropical Design Hotel... | Restaurante Térèze – R. Felício dos Santos, 15 – Santa Teresa, Rio... | SINBAD Consortium (SLIM)

Search GitHub   Pull requests   Issues   Gist

SINBAD Consortium (SLIM)

SLIM's repositories for SINBAD consortium members

UBC EOS, Vancouver, BC, Canada   https://www.slim.eos.ubc.ca

Repositories   People 40   Teams 5   Settings

Filters   Find a repository...   + New repository

SLIM-release-apps PRIVATE

Matlab ★ 4 0

Main SLIM software release to SINBAD sponsors - containing all applications, algorithms, tools, and utilities

Updated 4 days ago

SLIM-release-developers PRIVATE

HTML ★ 2 0

SLIM developer notes and templates

Updated 27 days ago

SLIM-release-comp PRIVATE

Shell ★ 2 1

3rd-party software for multi-user installation of SLIM software release to SINBAD sponsors - not required by some of appliactions from SLIM-release-apps

Updated on Aug 6

People 40 >

HM

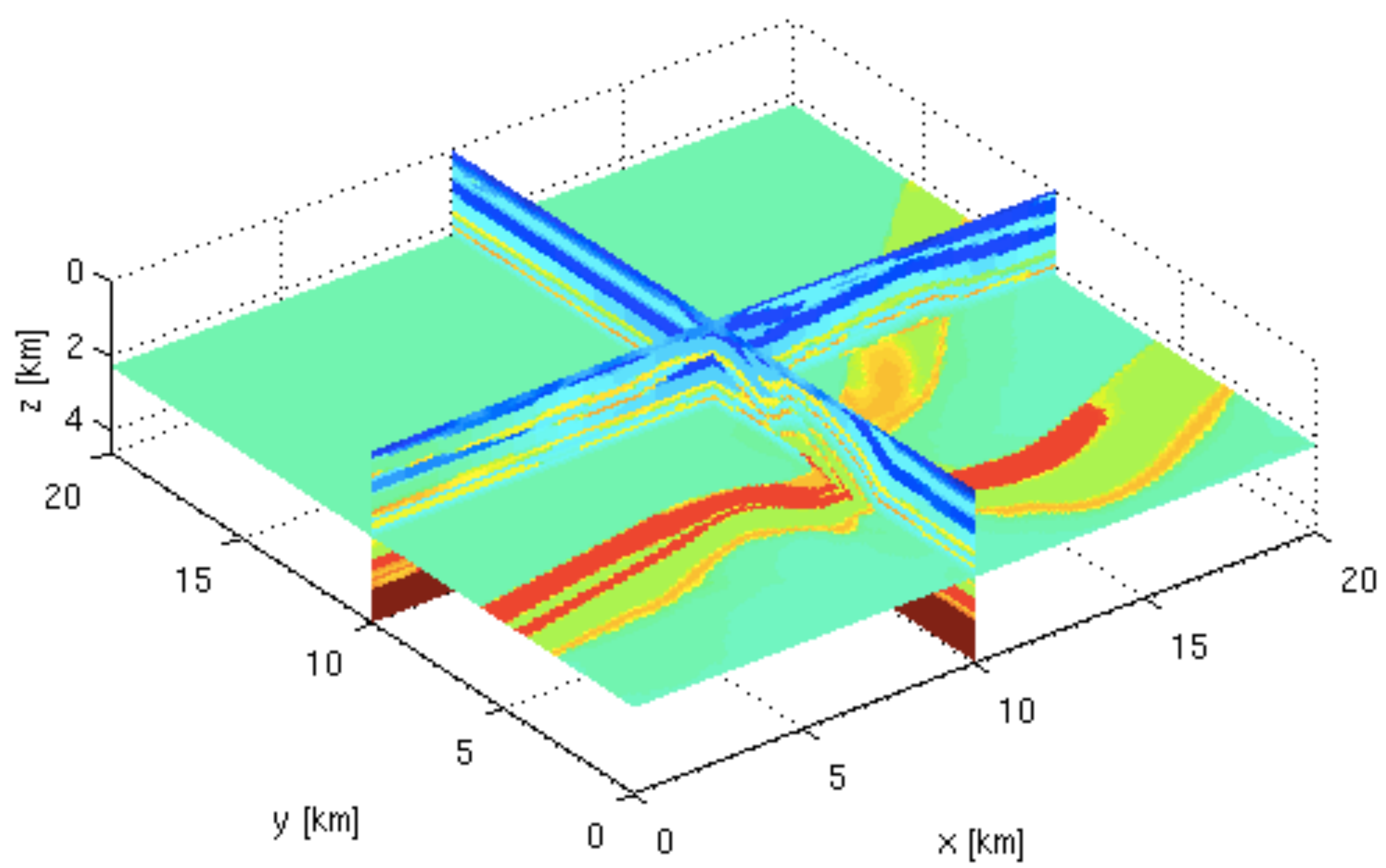
Invite someone

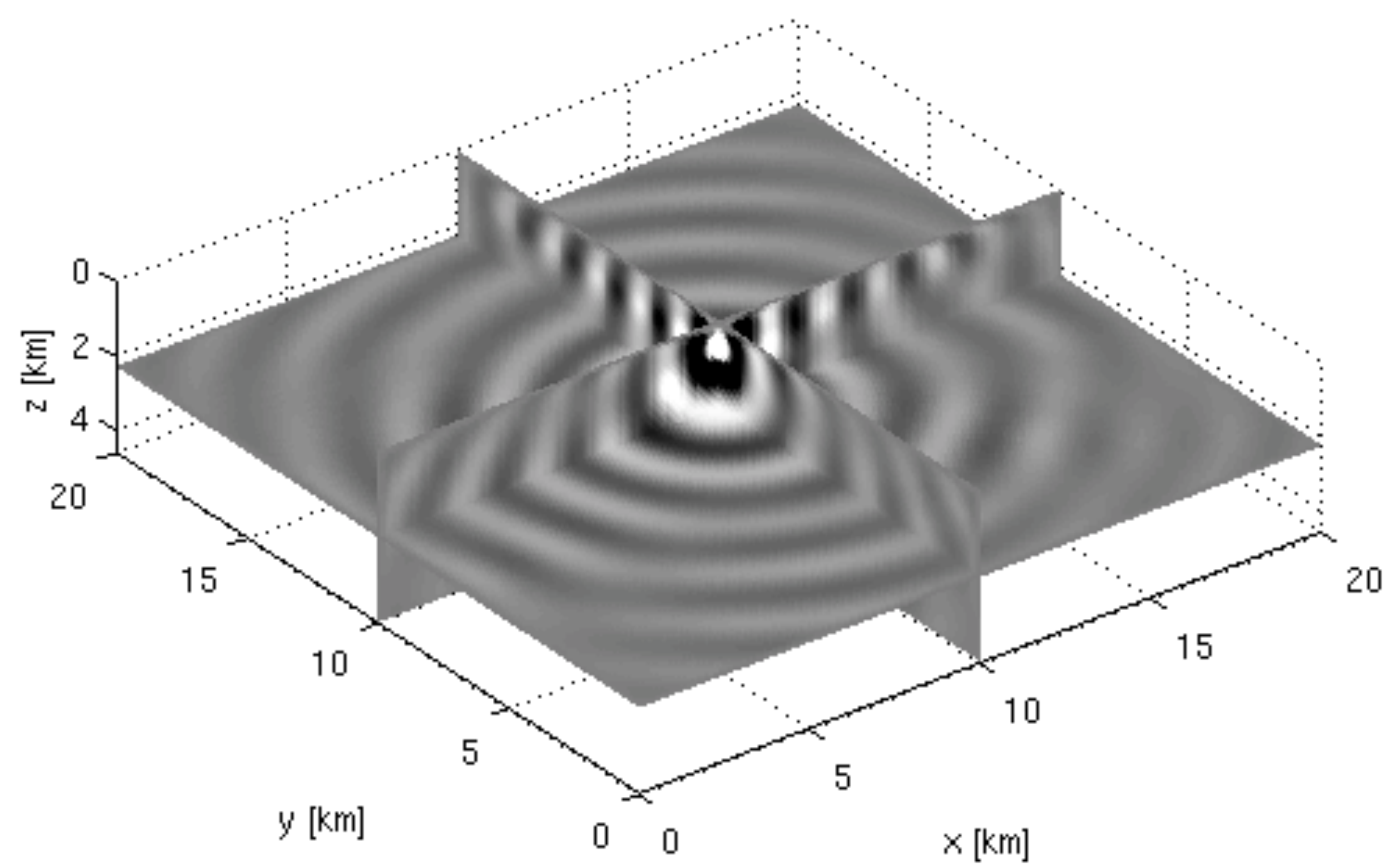
3D Frequency-domain FWI with batching: results

Contents

CARP-CG

Here we present some results of the Helmholtz solver on the overthrust model. The model and a wavefield for 2 Hz are shown below.

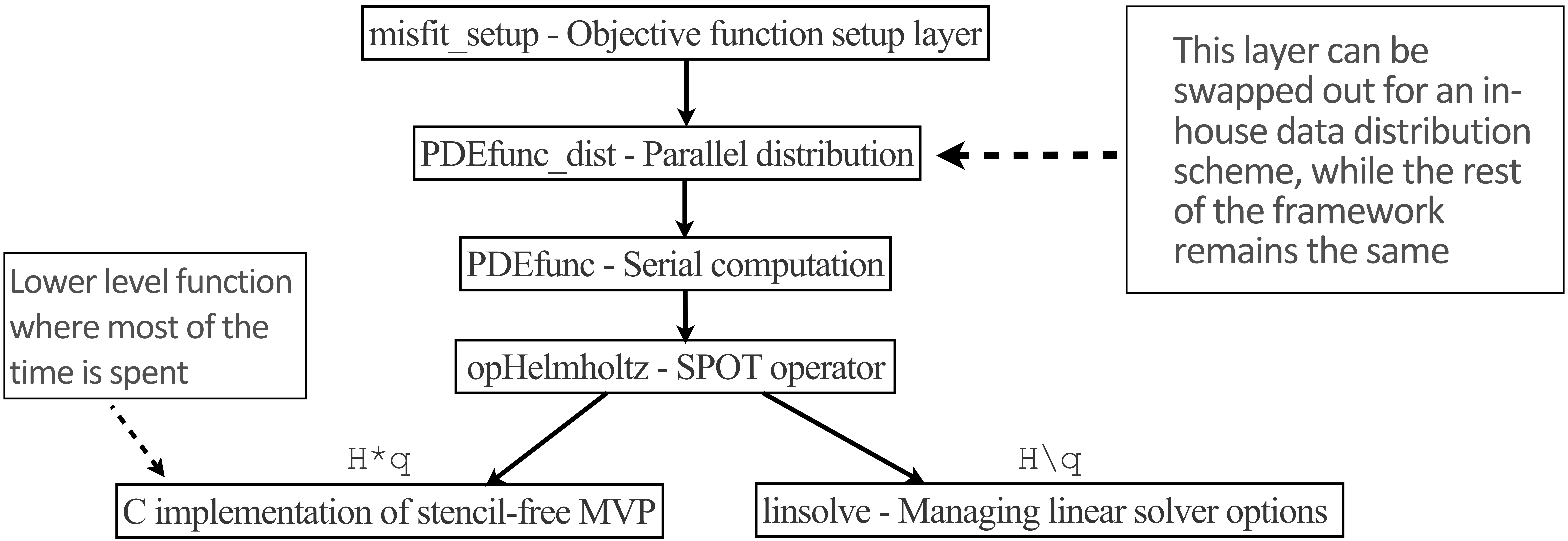




We compute the wavefield for various frequencies with a fixed number of gridpoints per wavelength. The convergence histories are shown below

A flexible and scalable software framework

By designing our software environment in a modular fashion, we have developed a system that is *flexible*, *efficient*, *scalable*, and provable *correct*. We have the freedom to easily mix and match Helmholtz discretizations, preconditioners, linear solvers, optimization algorithms and more.



Today's agenda

Time-lapse randomized marine acquisition
(15 minutes)

(Time-lapse) reverse-time migration w/ multiples, source estimation & gaps
(25 minutes)

Constrained full-waveform inversion
(20 minutes)

Randomized acquisition

Drivers:

- Wave-equation inversions call for dense, wide-azimuth & long-offset surveys
- control on environmental impact
- economics

Solution:

- rethink sampling technologies for land & marine using insights from Compressive Sensing
- remove sub-sampling-related artifacts by carrying out structure-promoting inversions
- Compressive Sensing = increased acquisition productivity

Time-lapse randomized marine acquisition

Haneet Wason & Felix Oghenekohwo

SLIM
University of British Columbia

Felix Oghenekohwo, Haneet Wason, Ernie Esser, and Felix J. Herrmann, "**Cheap time lapse with distributed Compressive Sensing–-exploiting common information among the vintages**". 2016.

Motivation

Seemingly *innocent* remark by Craig J. Beasley at SBGf meeting:

"Should we repeat or not repeat in randomized marine acquisition?"

Motivation

Seemingly *innocent* remark by Craig J. Beasley at SBGf meeting:

"Should we repeat or not repeat in randomized marine acquisition?"

"How sensitive is the recovery to minor errors in exact repeatability?"

Findings – a preview

Increased **exact** repetition amongst surveys leads to

- deteriorated recovery of prestack vintages themselves
- improved recovery of time-lapse prestack differences

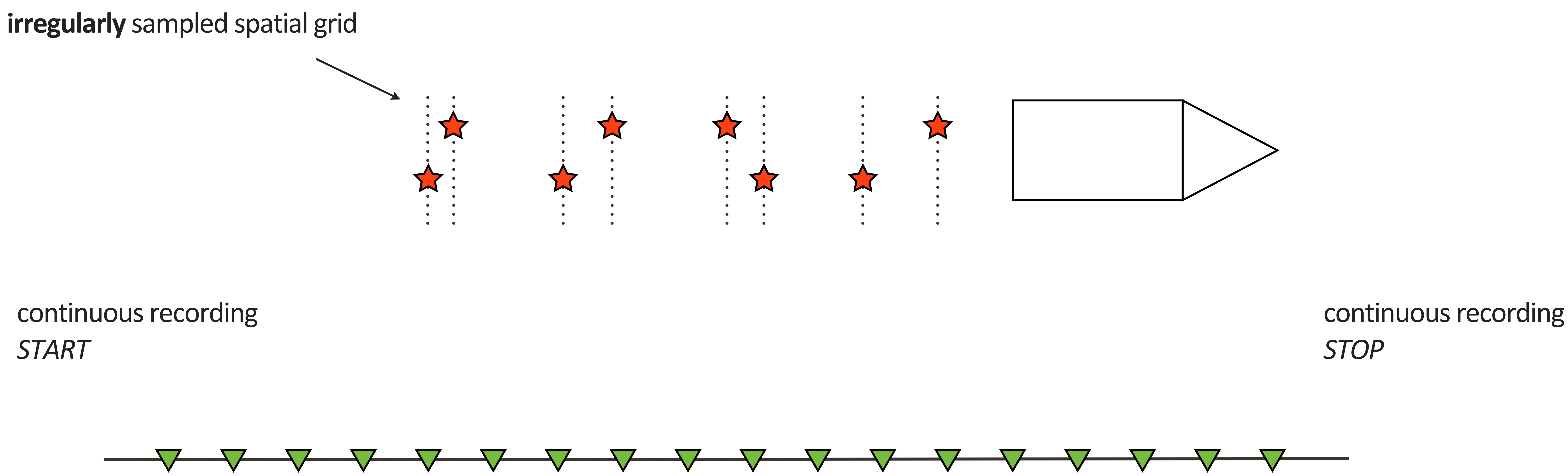
Small, but known, source location perturbations lead to

- improved recovery of the prestack vintages
- deteriorated recovery of prestack time-lapse differences

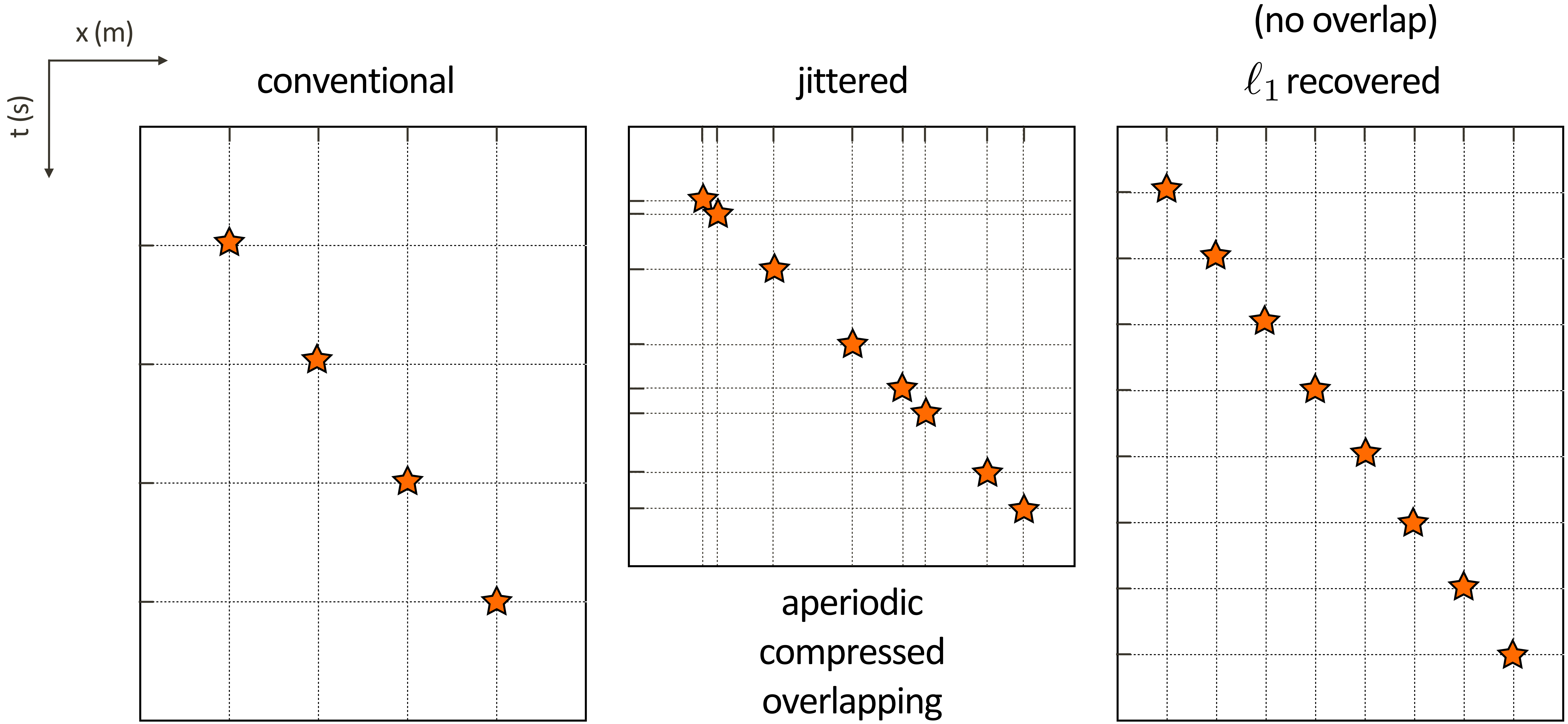
Tentative conclusions

- do not bother to repeat as long as you know where you were precisely
- instead aim to increase variability albeit natural variability already helps...

Time-jittered marine acquisition



Randomized jitter sampling in marine



periodic - sparse - no overlap

aperiodic
compressed
overlapping
irregular

periodic & dense

Time-lapse seismic

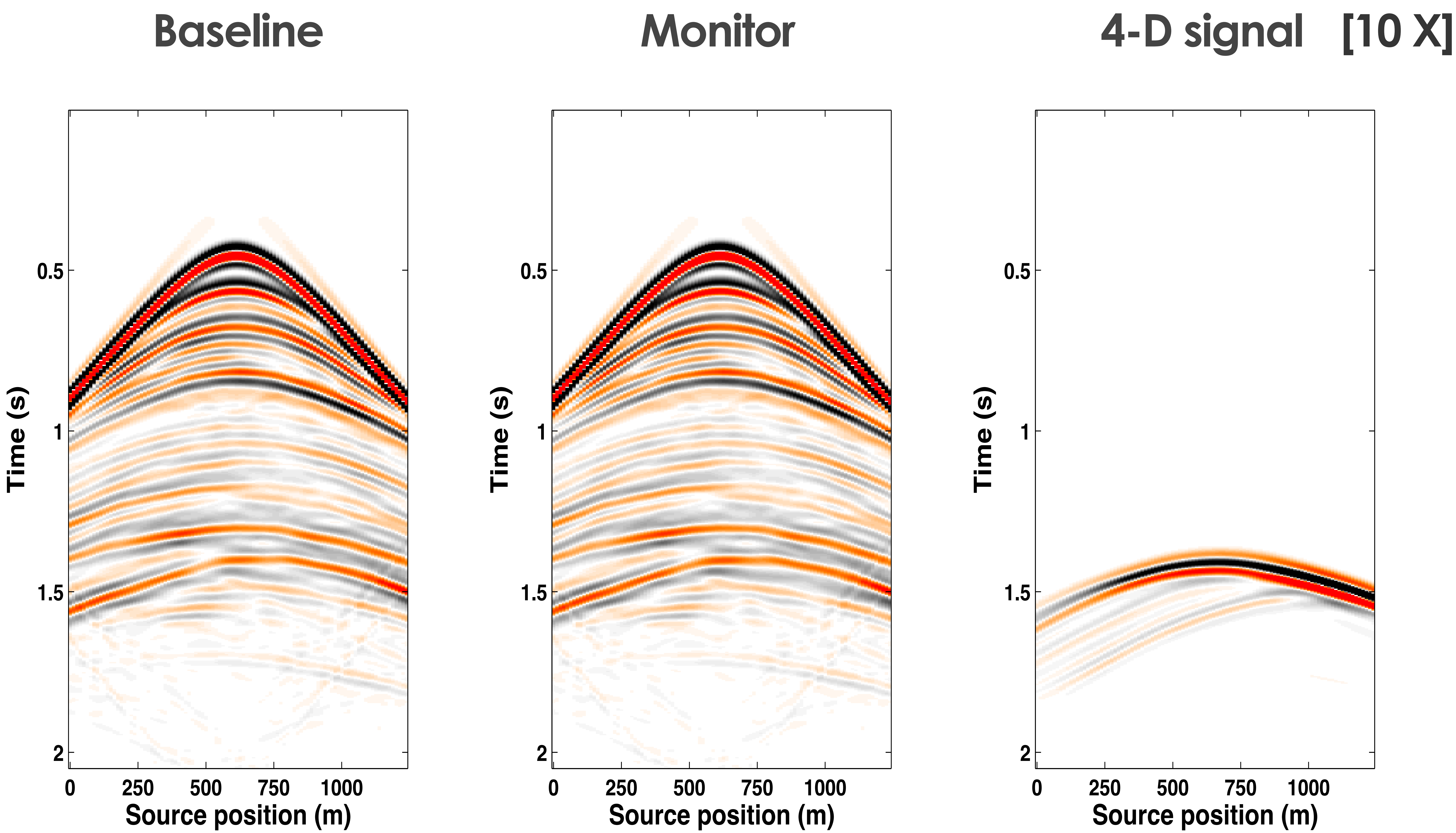
Current acquisition paradigm:

- repeat **expensive** dense acquisitions & “independent” processing
- compute differences between baseline & monitor survey(s)
- hampered by practical challenges to ensure repetition

New compressive sampling paradigm:

- **cheap** subsampled acquisition, e.g., via time-jittered marine subsampling
- may offer possibility to relax insistence on repeatability
- exploits insights from distributed compressed sensing

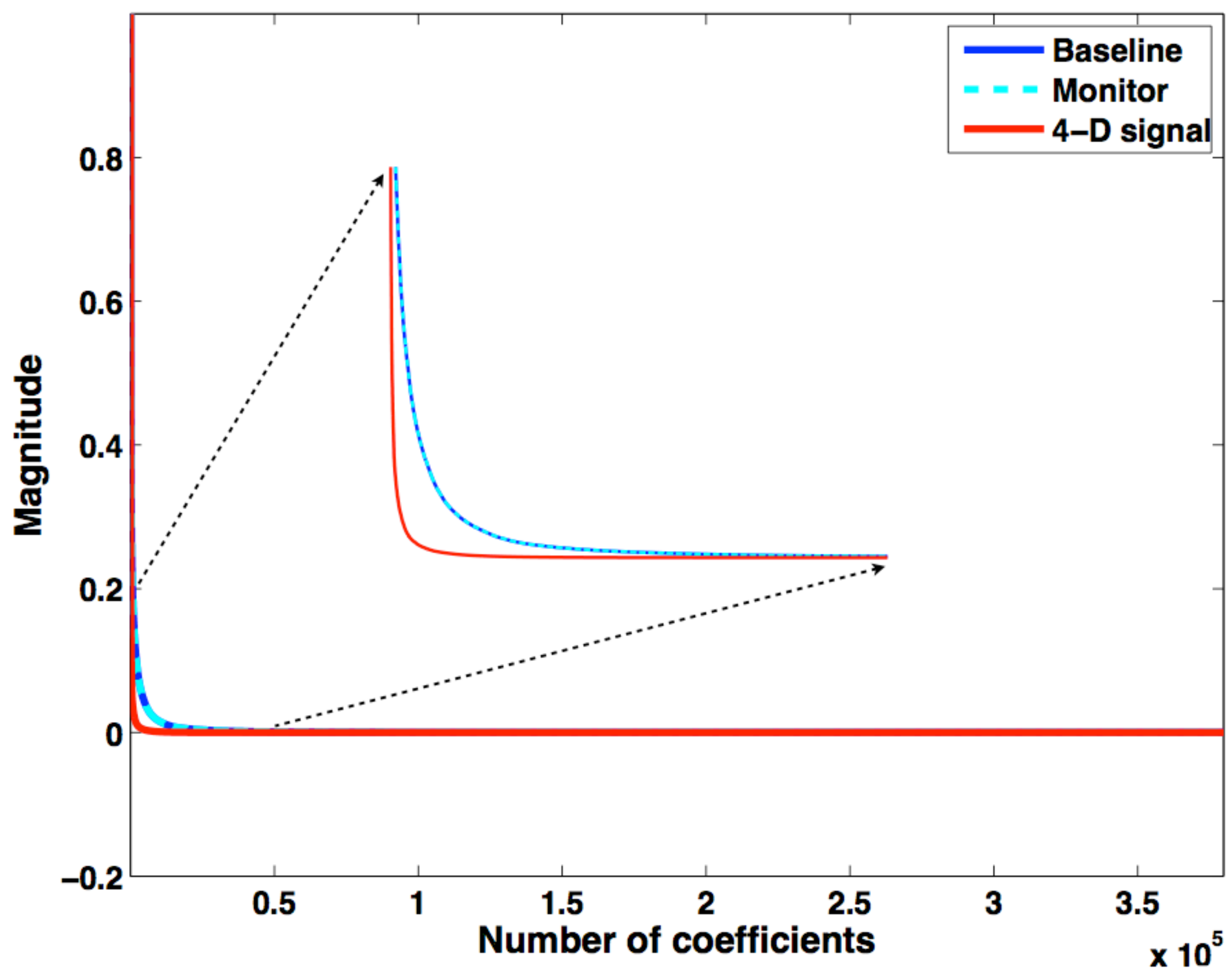
Time-lapse data



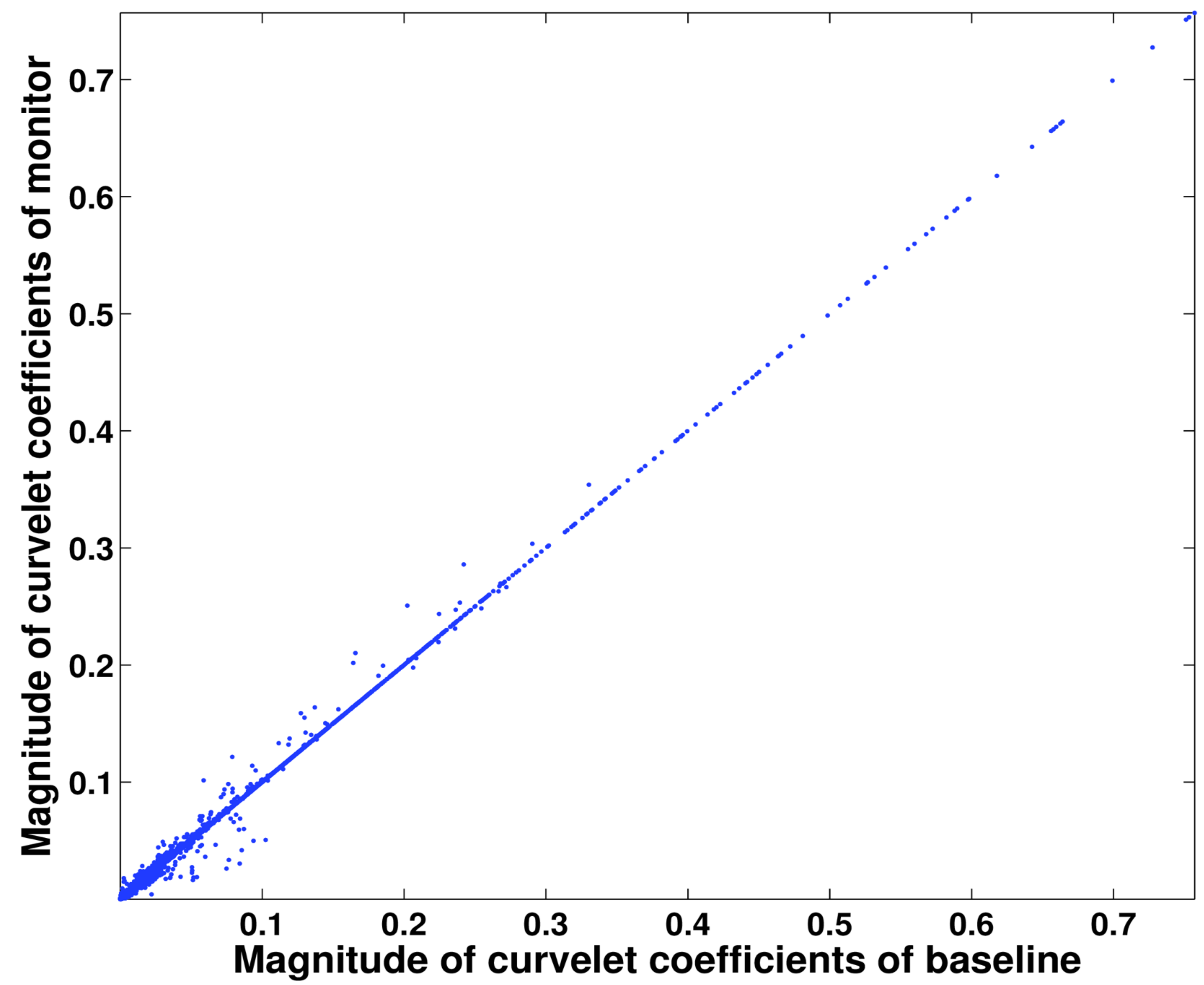
time samples: **512**
receivers: **100**
sources: **100**

sampling
time: **4.0 ms**
receiver: **12.5 m**
source: **12.5 m**

Sparse structure via curvelets



significant correlation between the vintages



Dror Baron, Marco F. Duarte, Shriram Sarvotham, Michael B. Wakin, Richard G. Baraniuk, "**An Information-Theoretic Approach to Distributed Compressed Sensing**" (2005)

Distributed compressed sensing

– joint recovery model (JRM)

$$\overbrace{\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_1 & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{A}_2 \end{bmatrix}}^{\mathbf{A}} \overbrace{\begin{bmatrix} \mathbf{z}_0 \\ \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix}}^{\mathbf{z}} = \overbrace{\begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{bmatrix}}^{\mathbf{b}}$$

baseline

monitor

vintages

$$\mathbf{x}_1 = \mathbf{z}_0 + \mathbf{z}_1$$
$$\mathbf{x}_2 = \mathbf{z}_0 + \mathbf{z}_2$$

differences

common component

different vintages share common information

Time-lapse seismic
– w/ & w/o repetition

In an *ideal* world $(\mathbf{A}_1 = \mathbf{A}_2)$

- JRM simplifies to $(\mathbf{b}_2 - \mathbf{b}_1) = \mathbf{A}_1(\mathbf{x}_2 - \mathbf{x}_1)$
- expect good recovery when difference is sparse
- but relies on “exact” repeatability…

In the *real* world $(\mathbf{A}_1 \neq \mathbf{A}_2)$

- no absolute control on surveys
- errors in the shot/receiver positions
- noise…

SLIM

Context

Acquire randomized subsamplings for the baseline and monitor surveys

Aim: recovery of both vintages & time-lapse signal from incomplete data

Questions:

- Process/recover independently or jointly to exploit common features of surveys?
- Should we repeat the surveys when doing randomized subsampling?

Synthetic seismic case study

Time-jittered marine acquisition on the grid

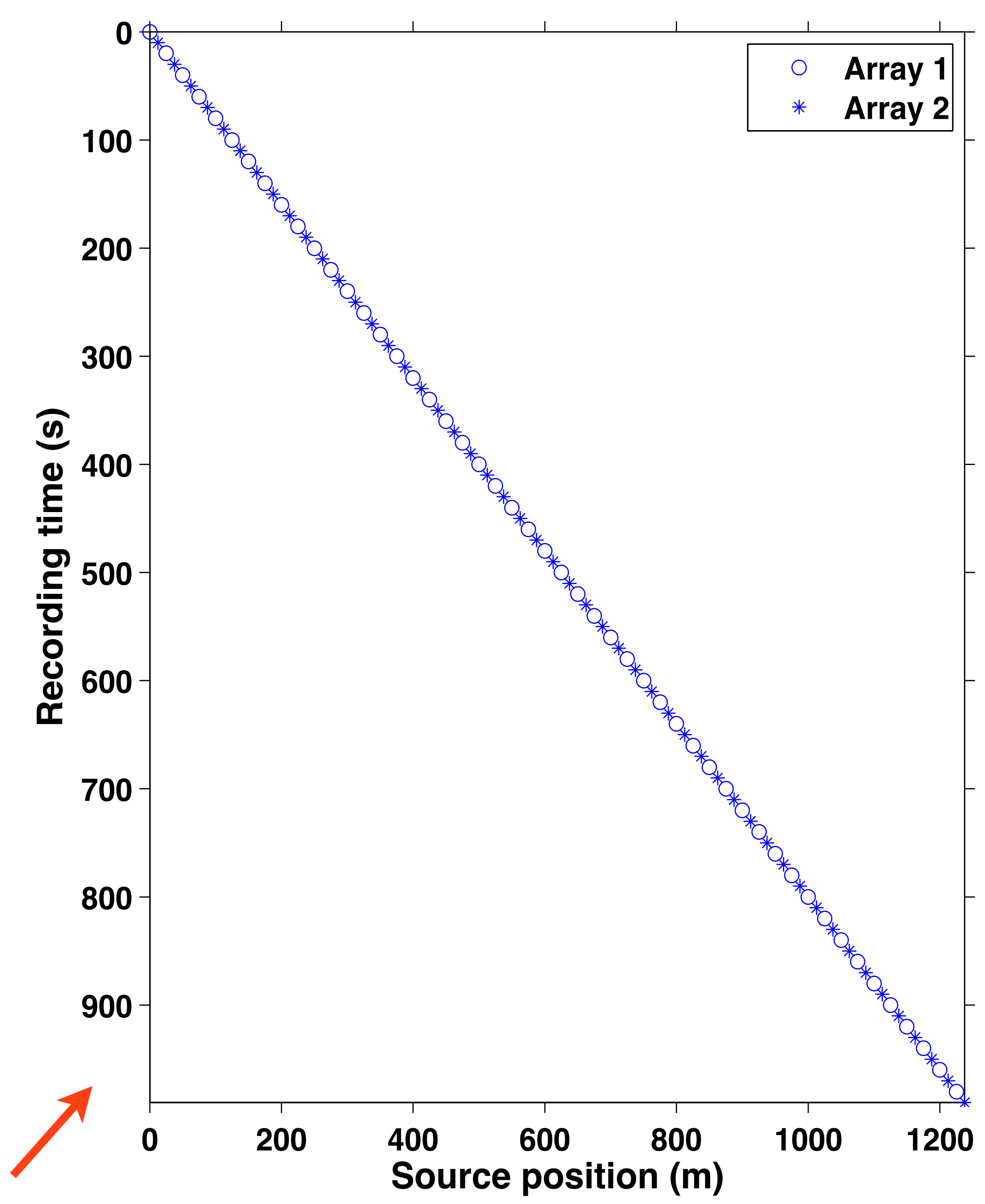
% repetition => “exact” repetition

No errors in the shot/receiver locations

Conventional vs. time-jittered sources
– subsampling ratio = 2 (2 source arrays)

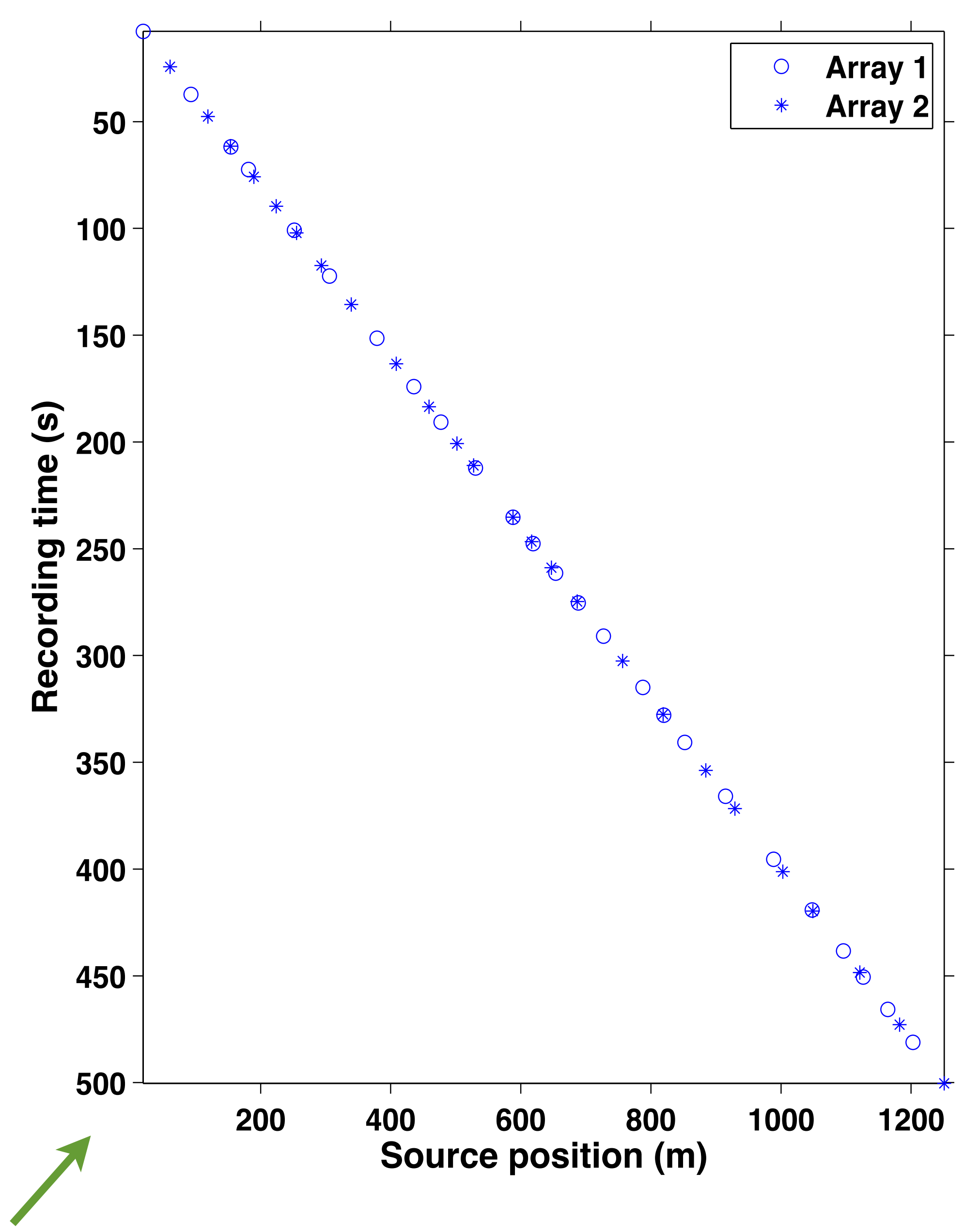
SLIM

conventional



***“unblended”* shot gathers**
number of shots = **100** (per array)
shot record length: 10.0 s
spatial sampling: **12.5 m**
vessel speed: **1.25 m/s**
recording time = 100 x 10.0 = **1000.0 s**

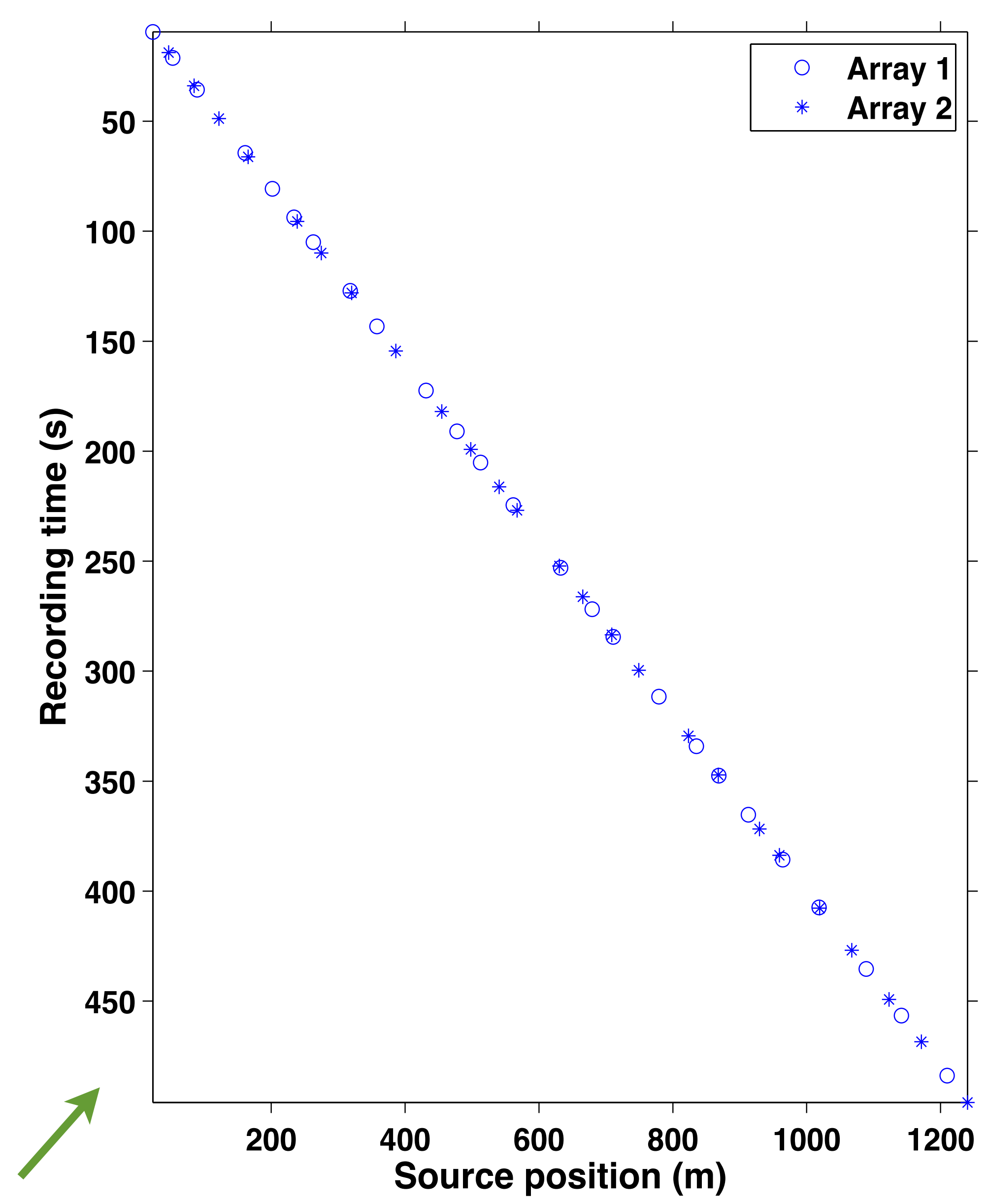
jittered acquisition 1 (baseline)



[BLENDING & SUBSAMPLING]
spatial subsampling factor = 2

spatial sampling **increase** factor = 2
[DEBLENDING & INTERPOLATION]

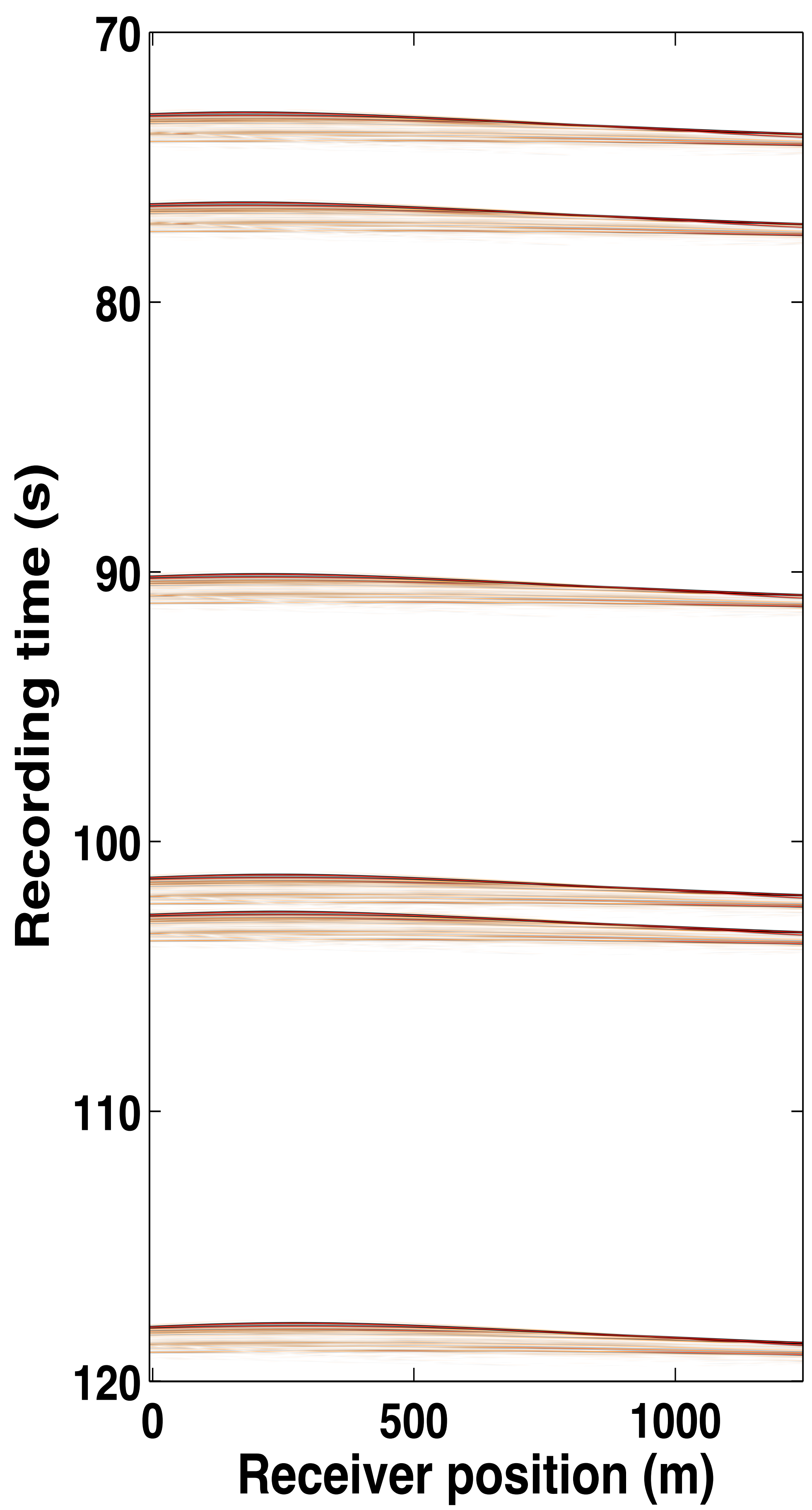
jittered acquisition 2 (monitor)



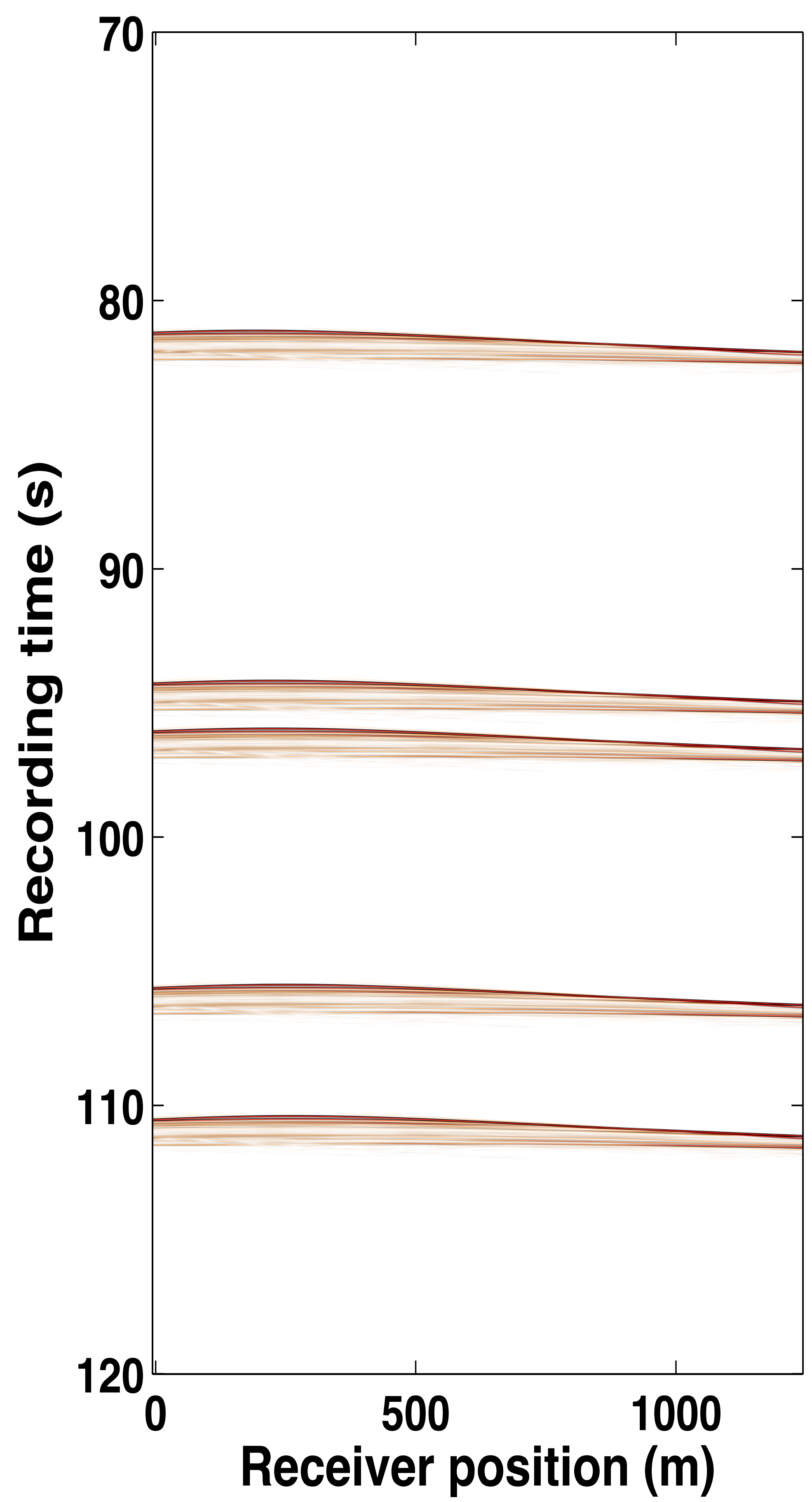
***“blended”* shot gathers**
number of shots = 100/2 = **50** (25 per array)
spatial sampling: **50.0 m (*jittered*)**
vessel speed: **2.50 m/s**
recording time ≈ 1000.0 s/2 = **500.0 s**

Measurements
– subsampled and blended

Baseline



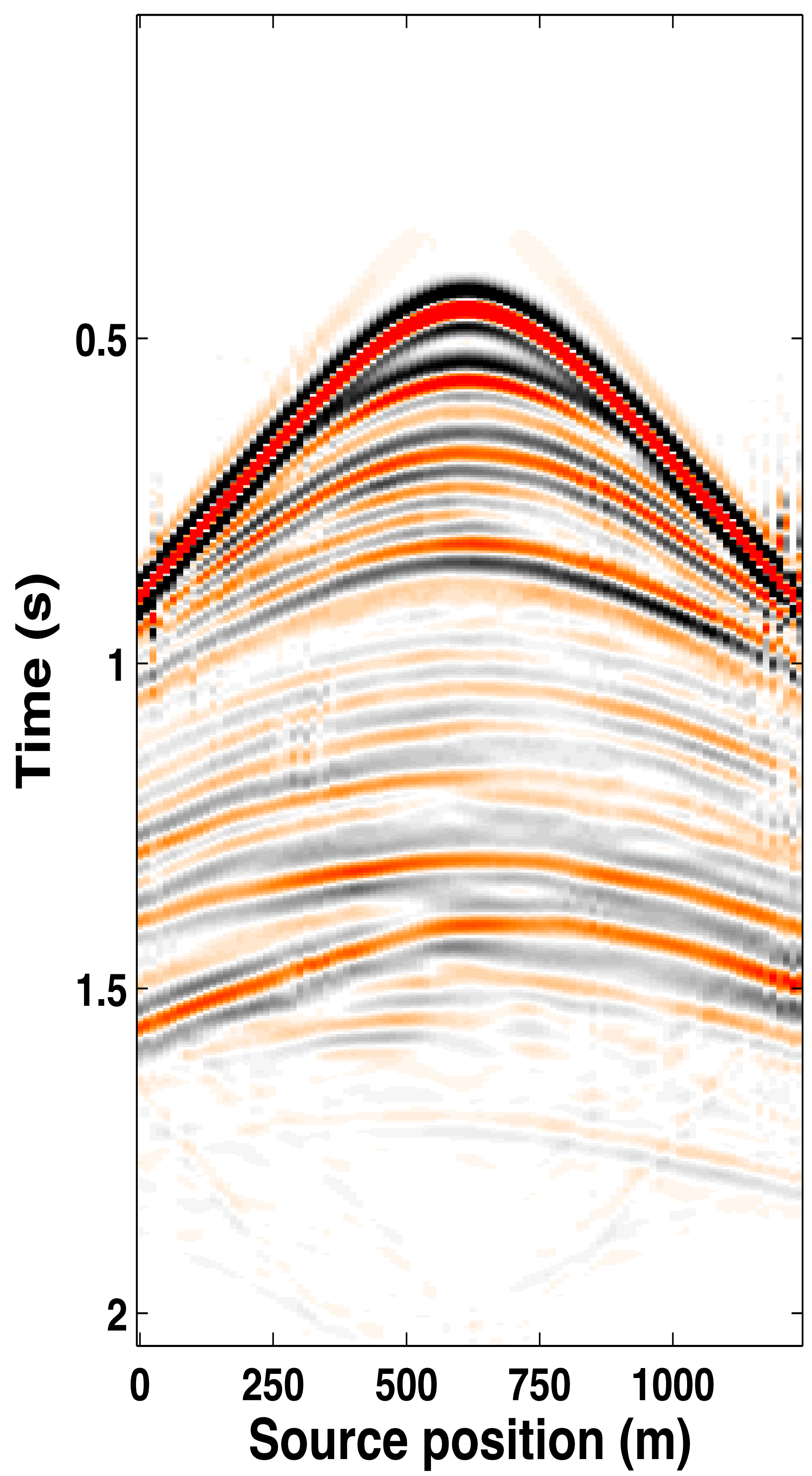
Monitor



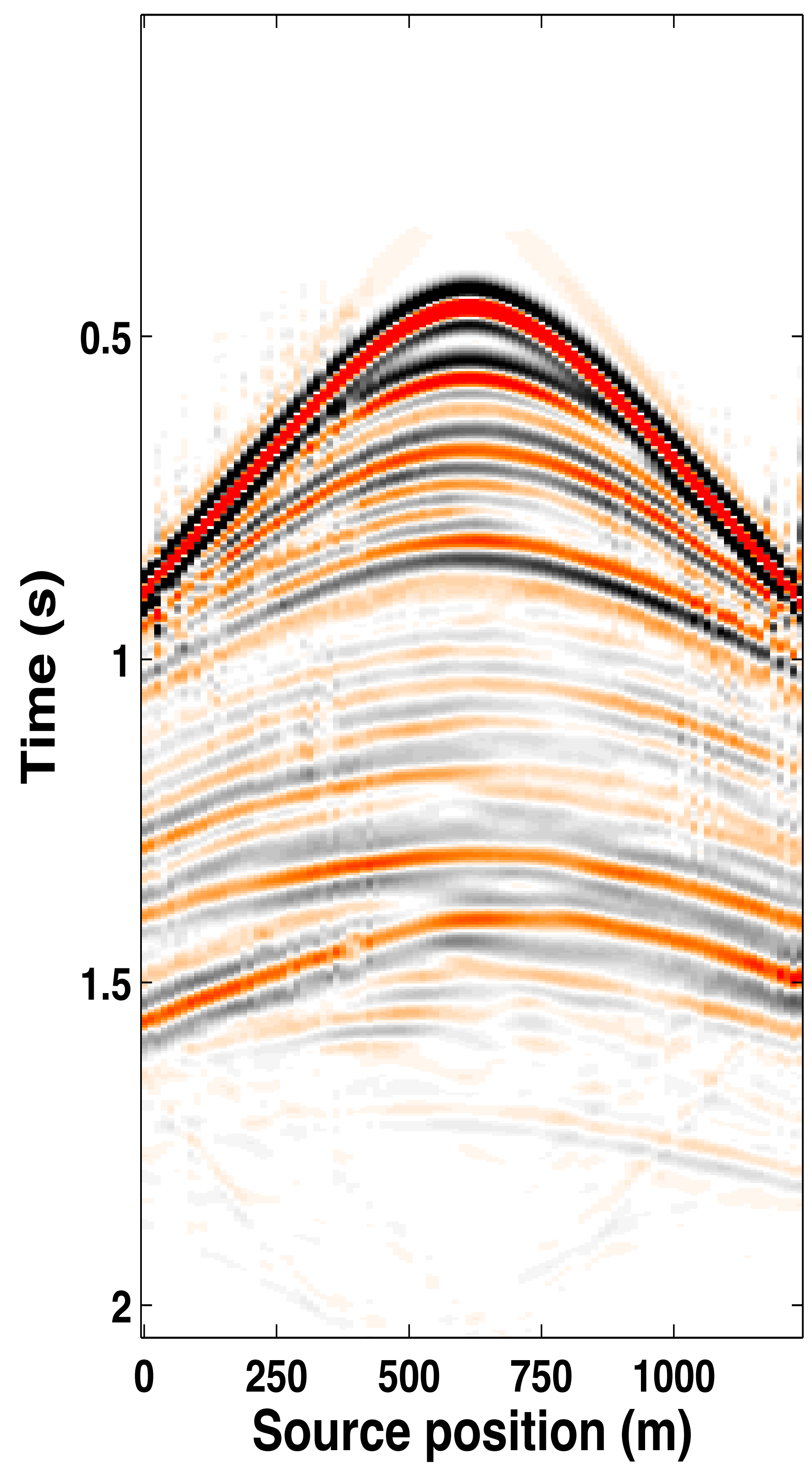
Monitor recovery

– Independent recovery

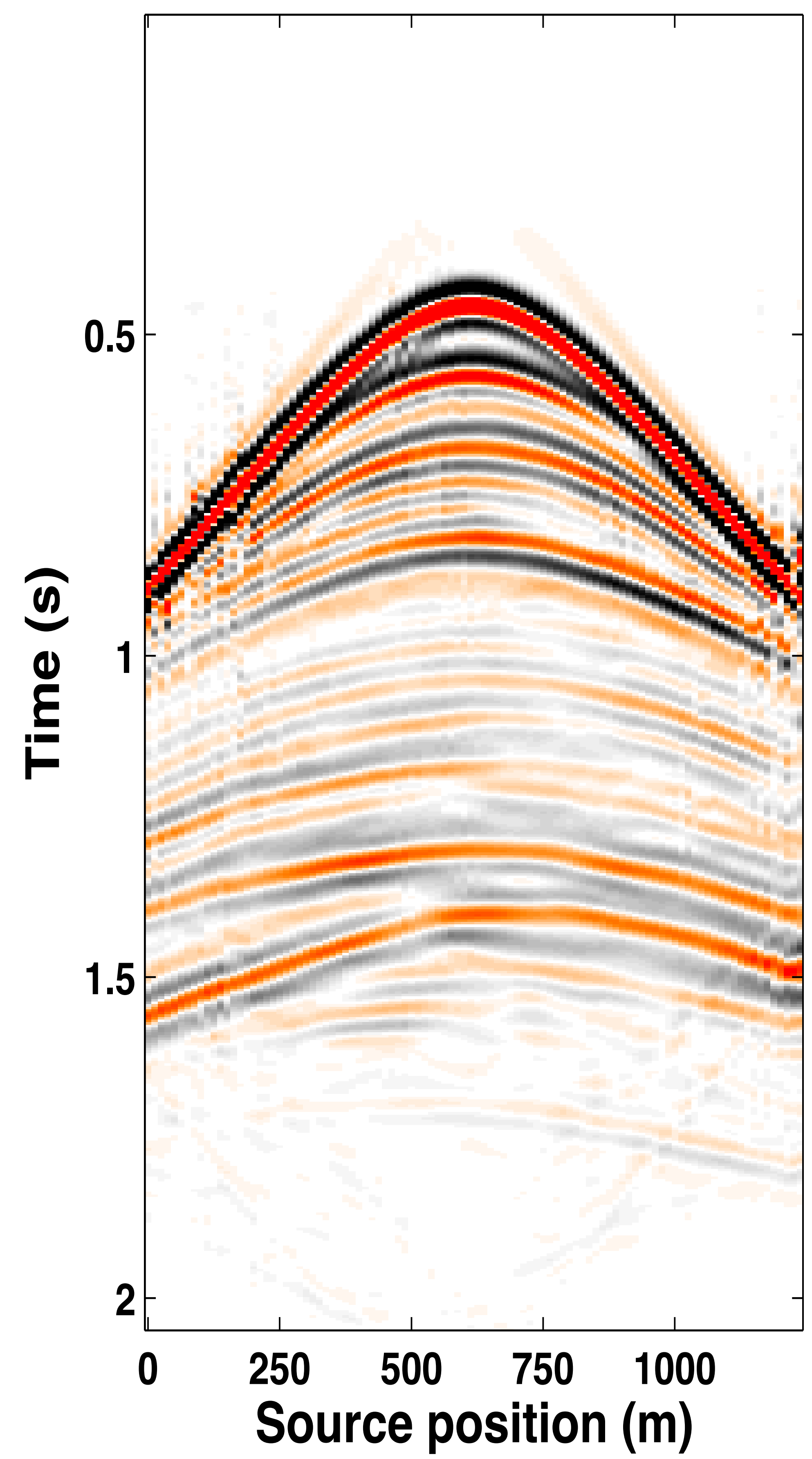
100% overlap
[11.6 dB]



50% overlap
[11.0 dB]



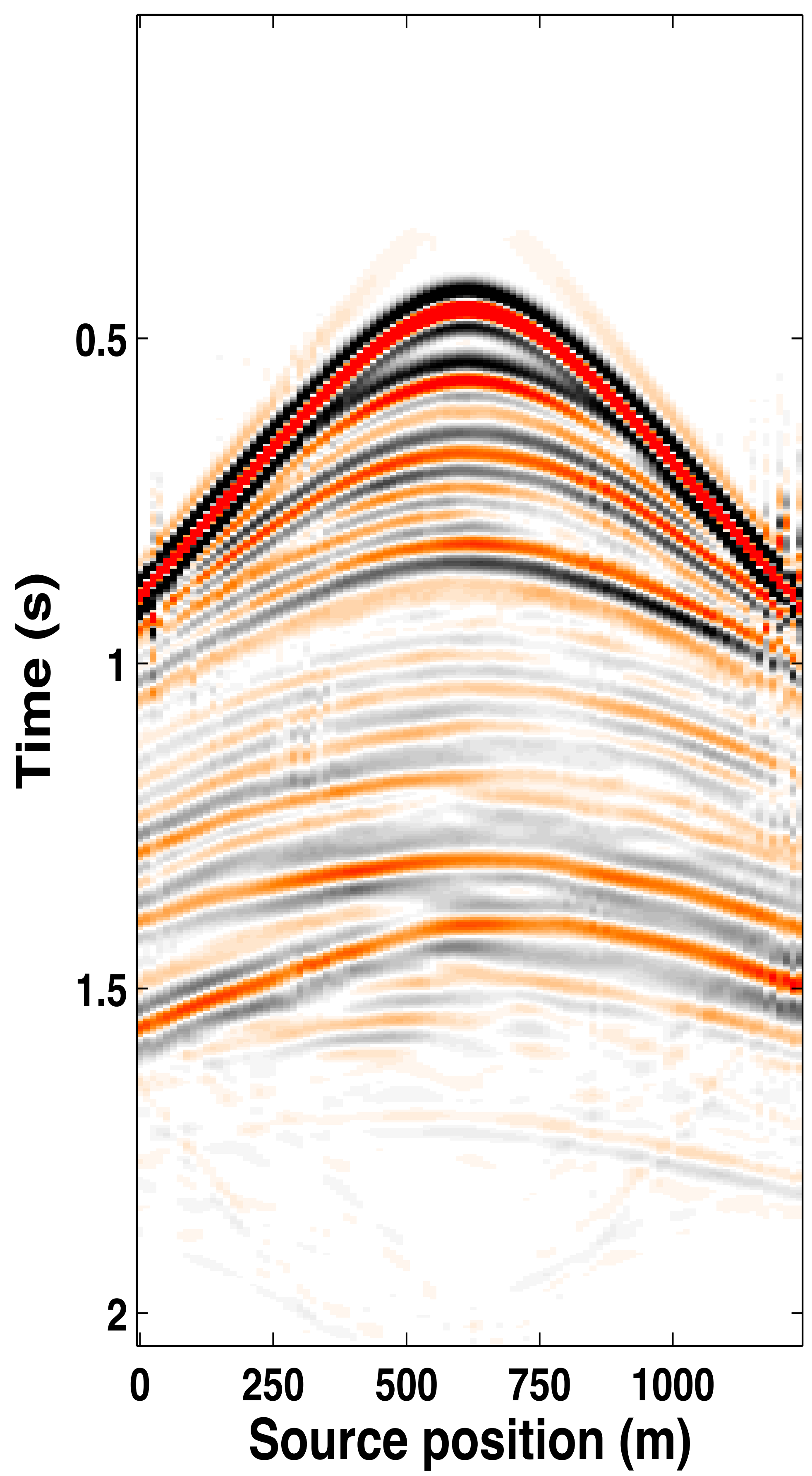
25% overlap
[10.3 dB]



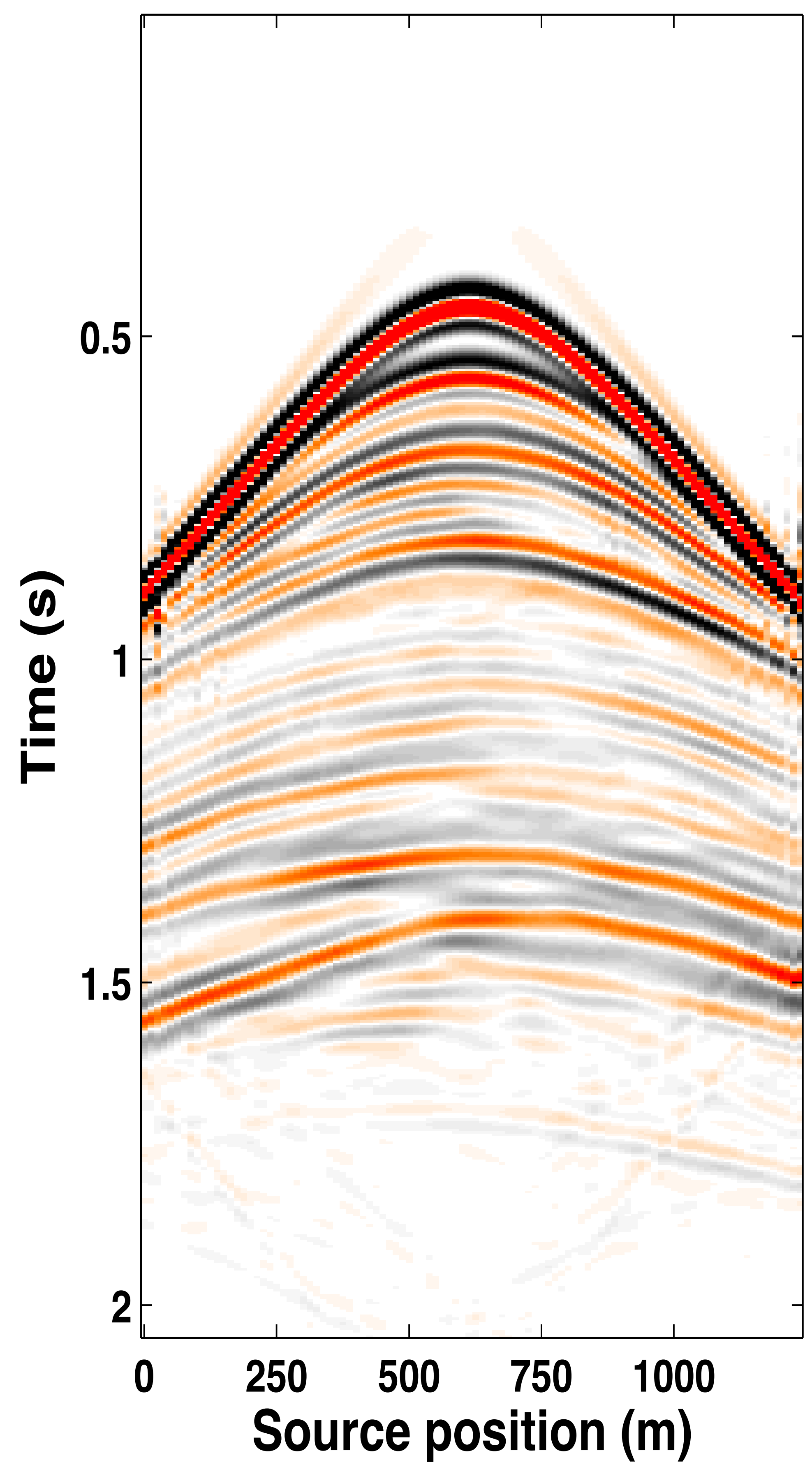
Monitor recovery

– Joint recovery

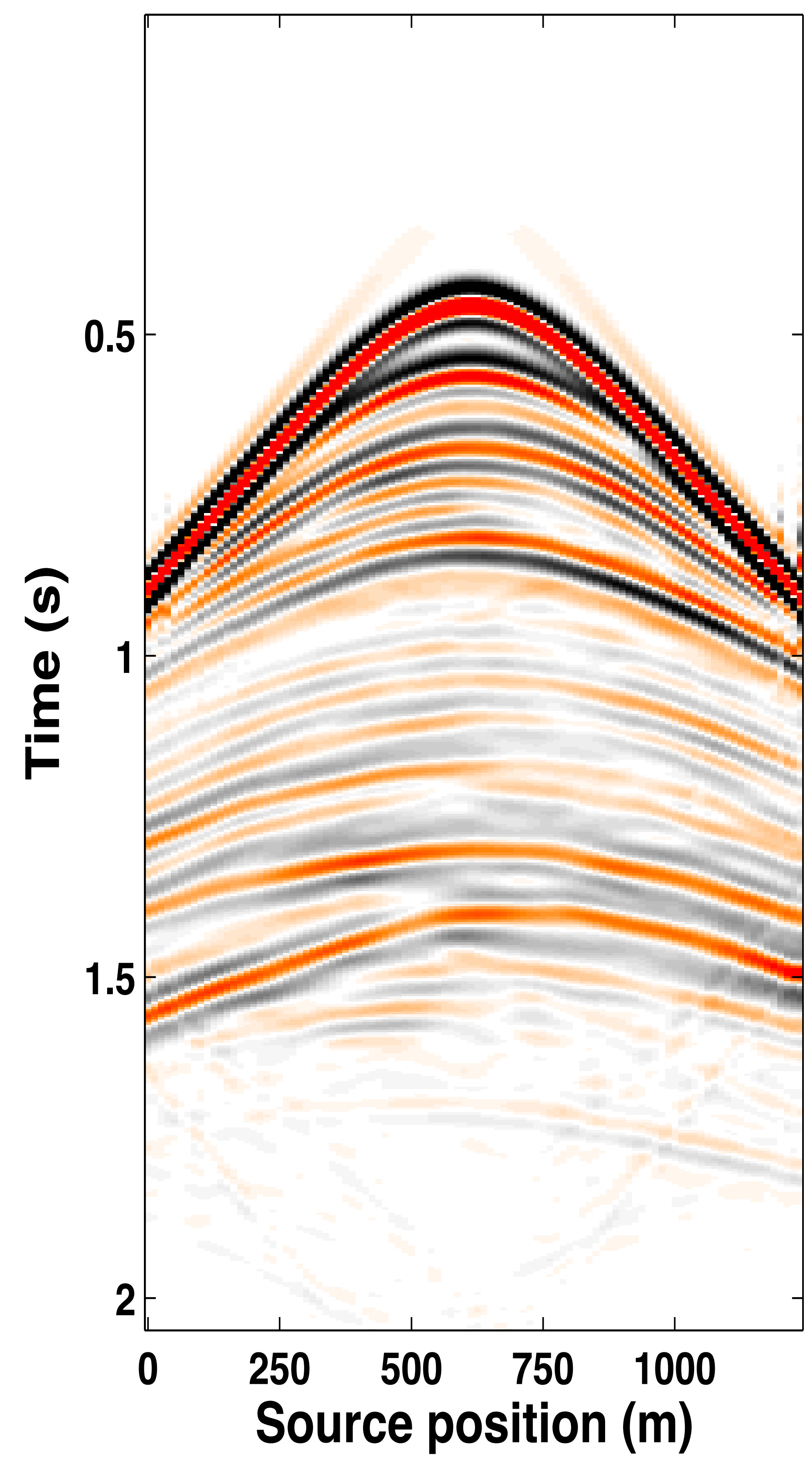
100% overlap
[11.6 dB]



50% overlap
[15.7 dB]



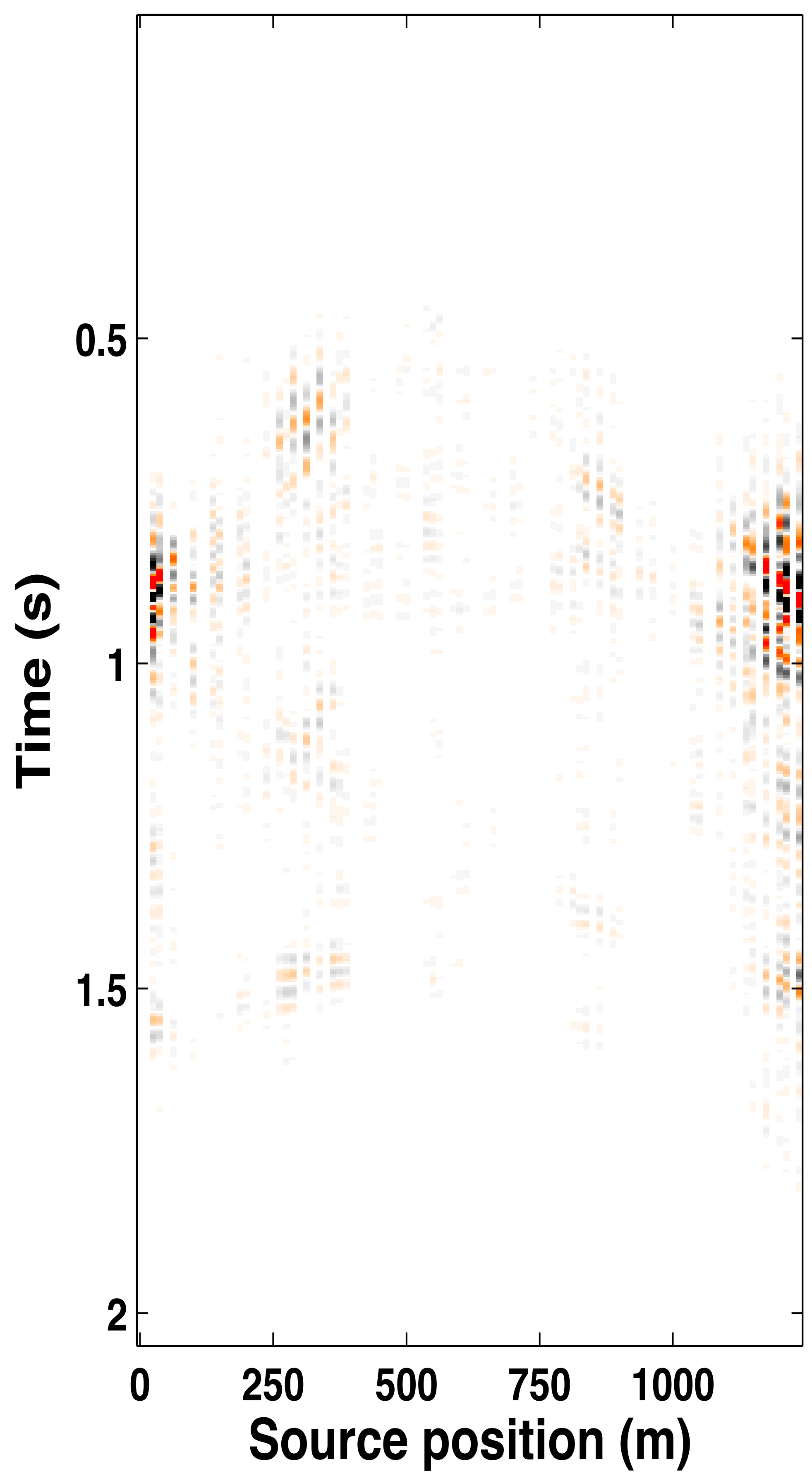
25% overlap
[18.6 dB]



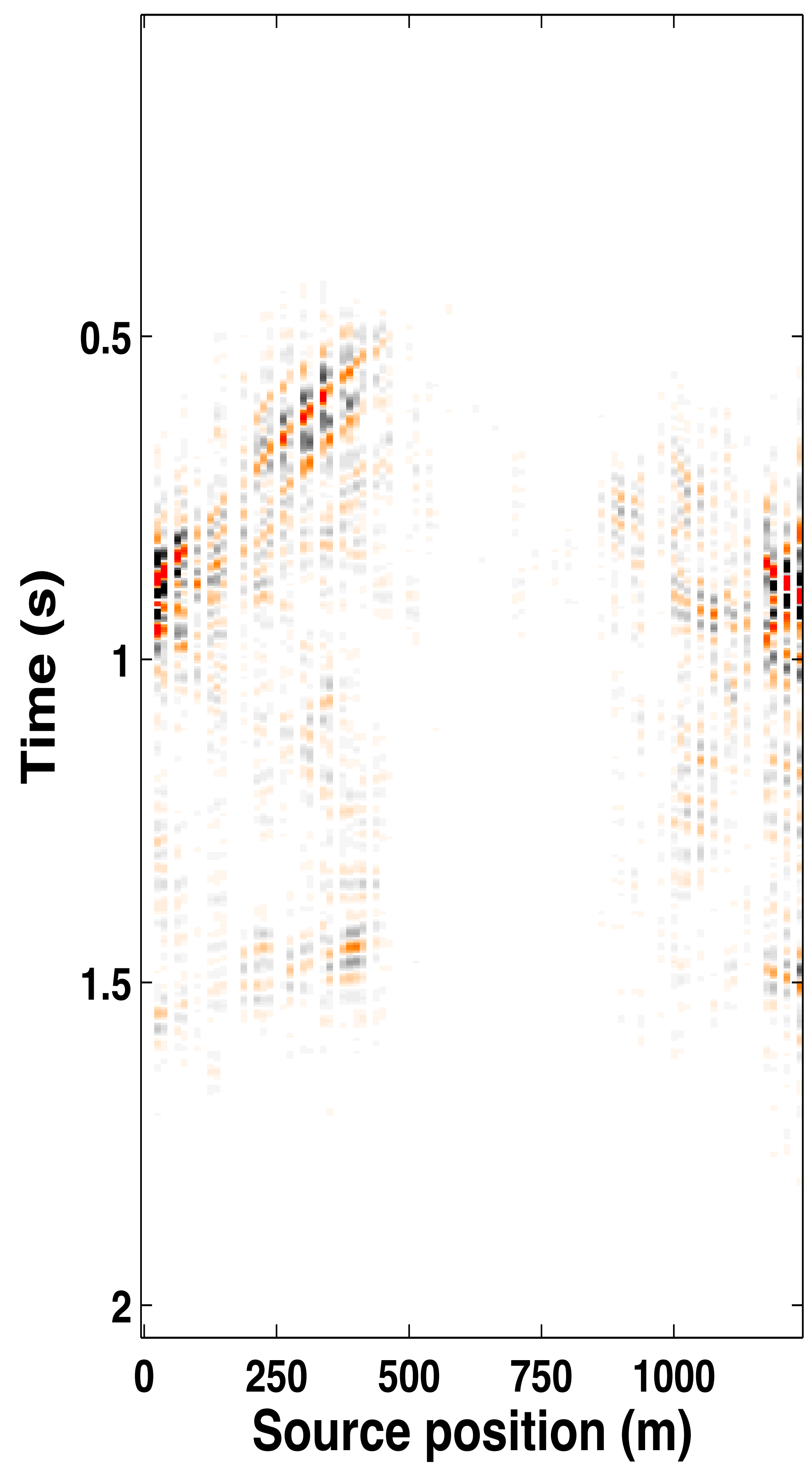
Monitor residual

– Independent residual

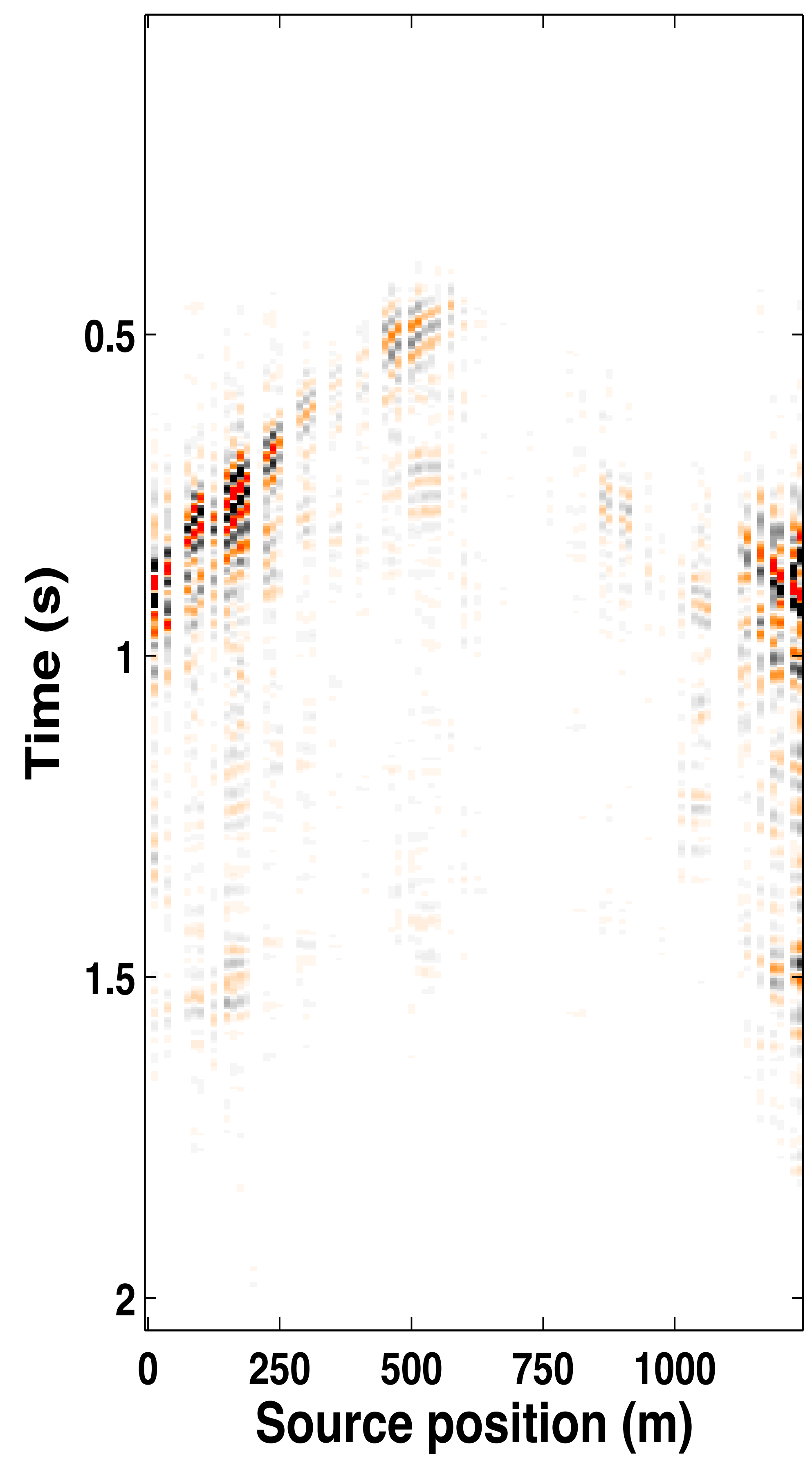
100% overlap
[11.6 dB]



50% overlap
[11.0 dB]

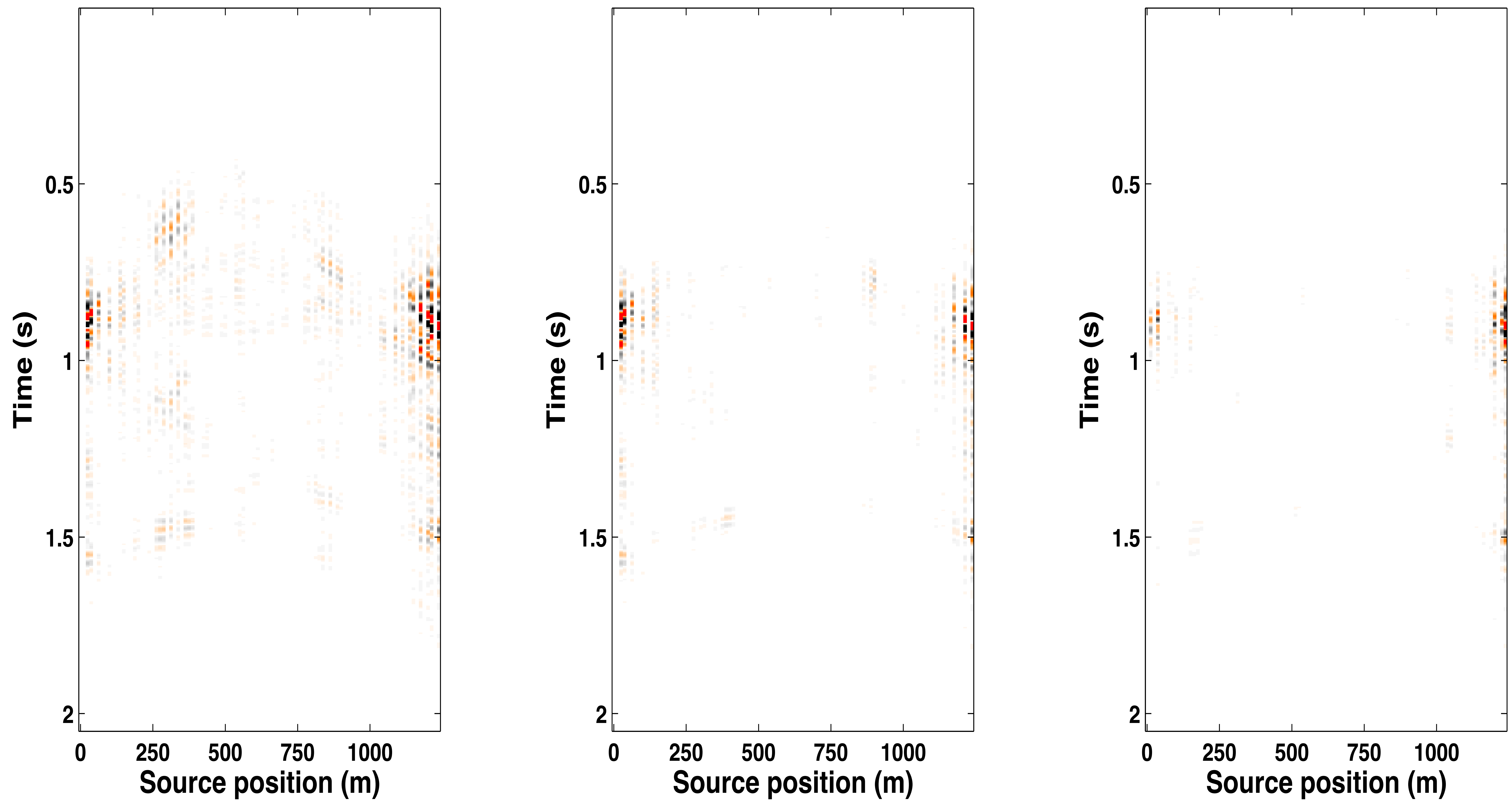


25% overlap
[10.3 dB]



Monitor residual
– Joint residual

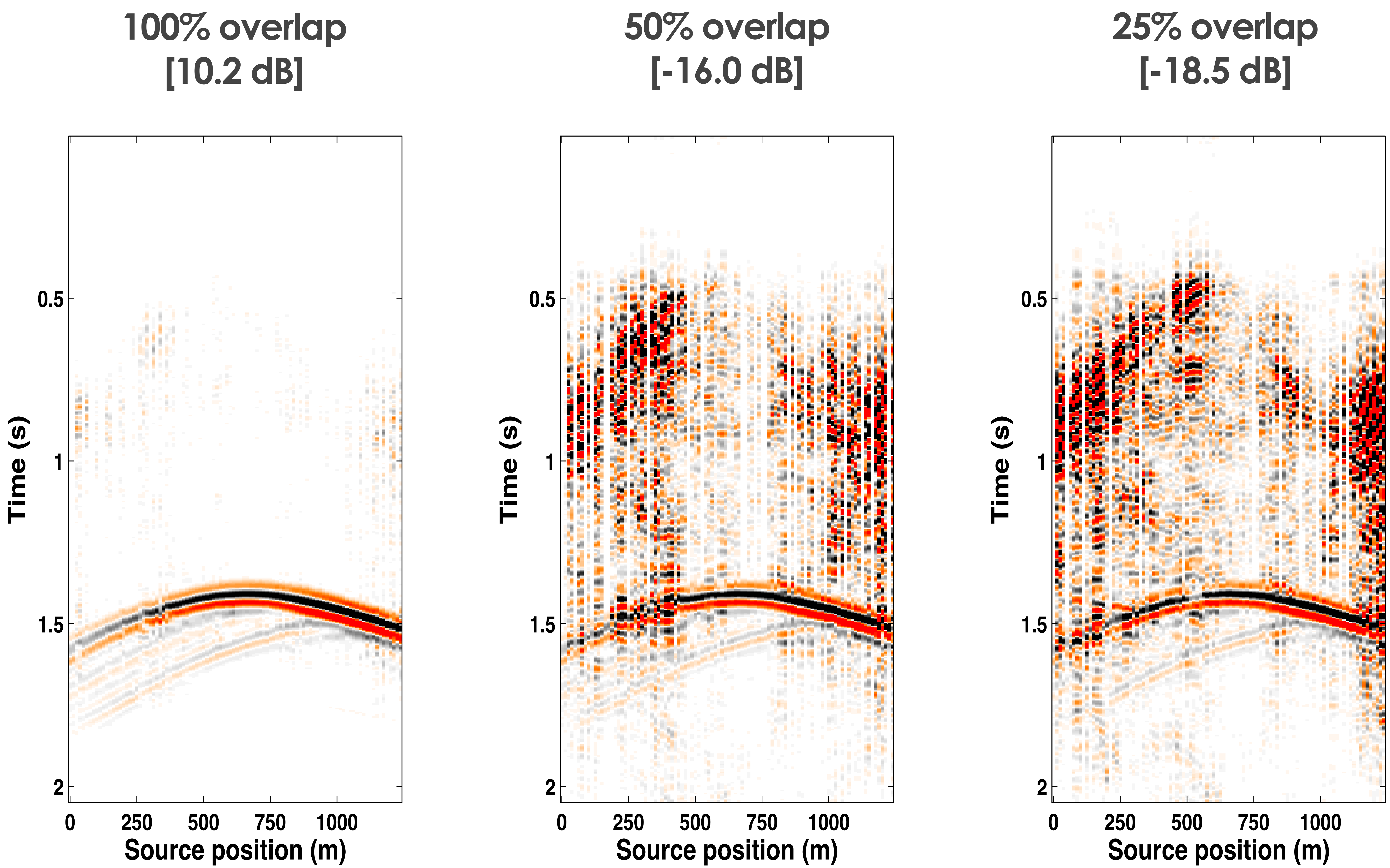
100% overlap [11.6 dB] | 50% overlap [15.7 dB] | 25% overlap [18.6 dB]



4-D recovery

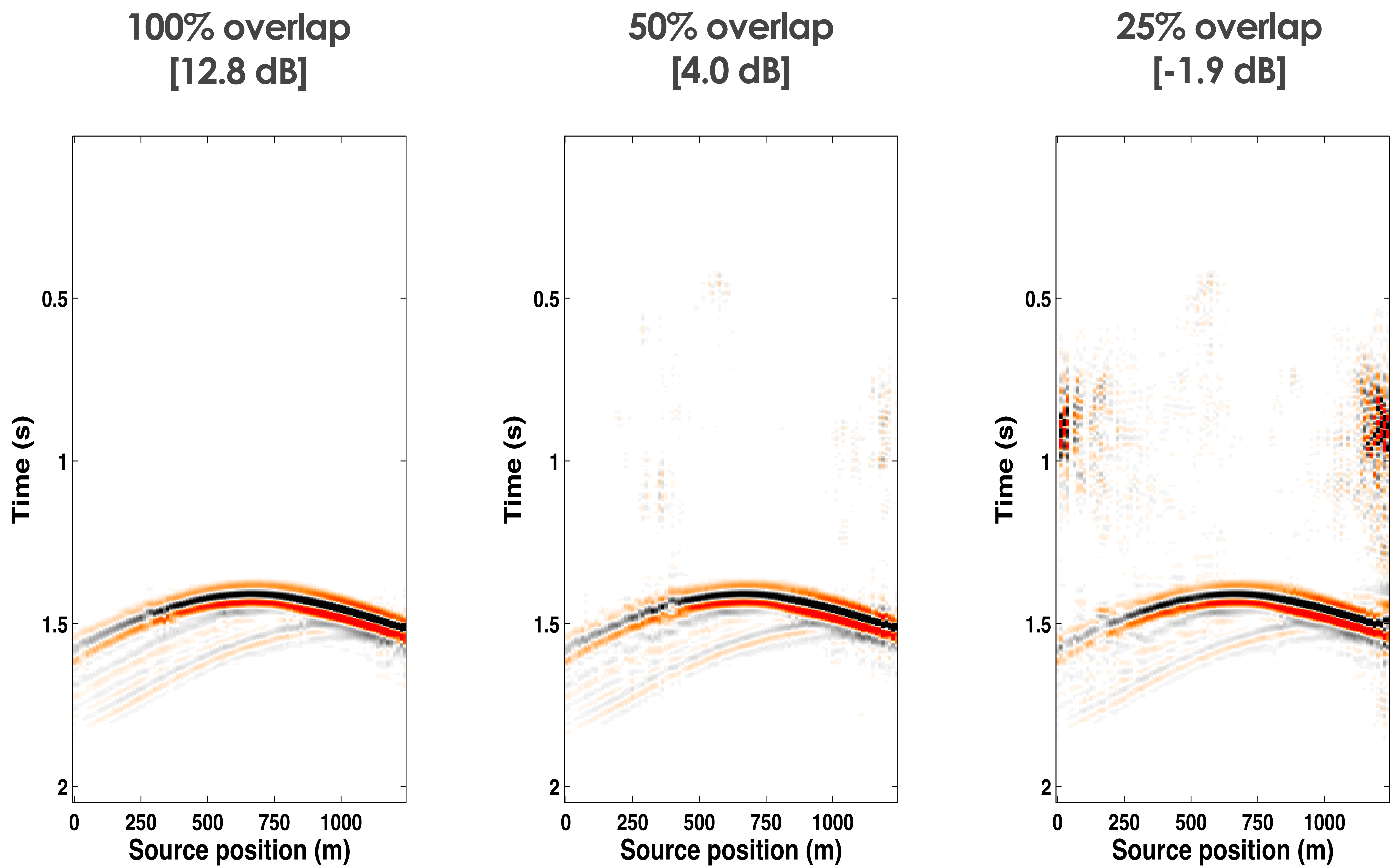
– Independent recovery

[colormap scale: 10 X]



4-D recovery
– Joint recovery

[colormap scale: 10 X]

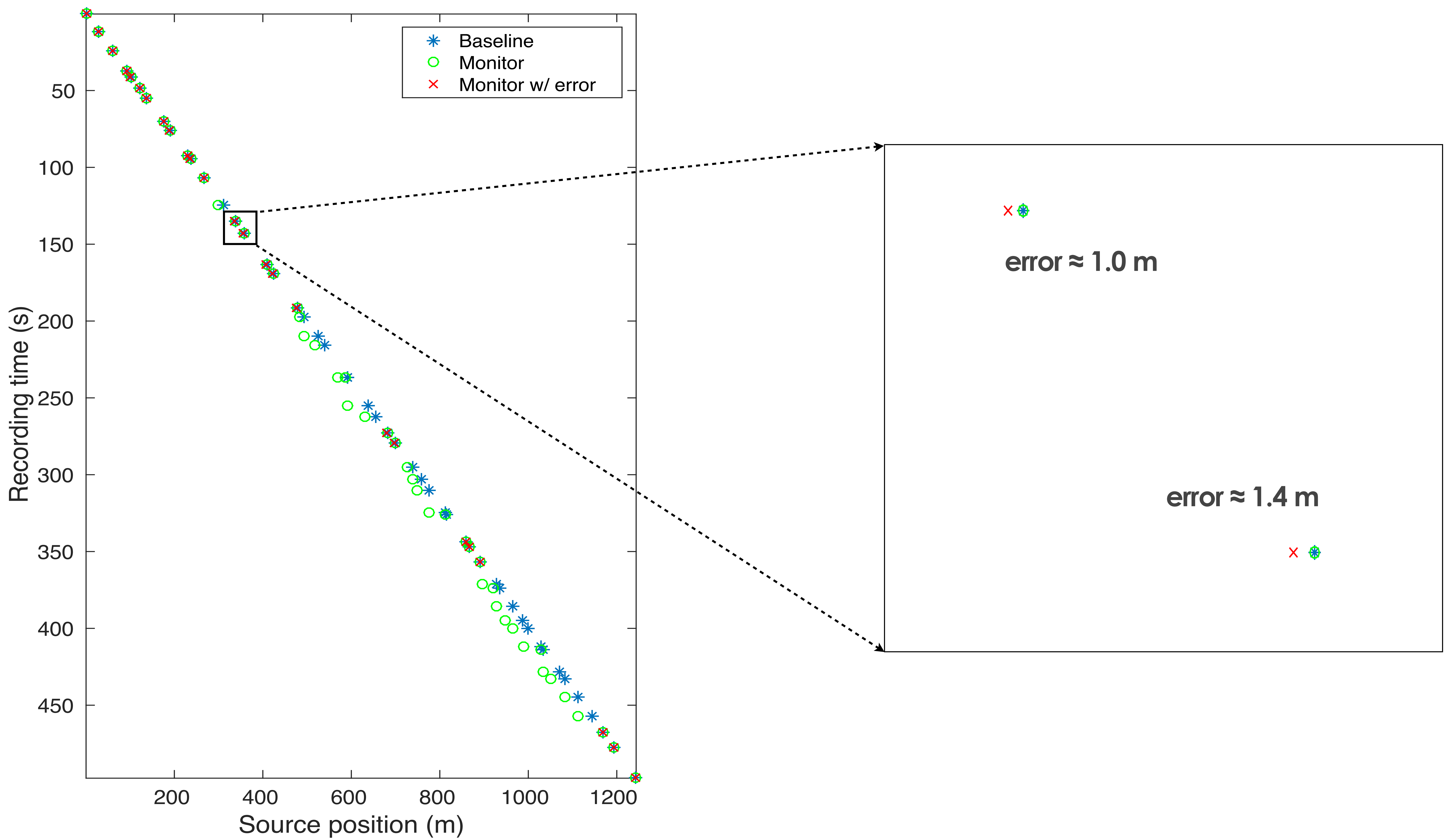


Notion of repetition

Time-jittered marine acquisition off the grid

With & without errors in shot locations

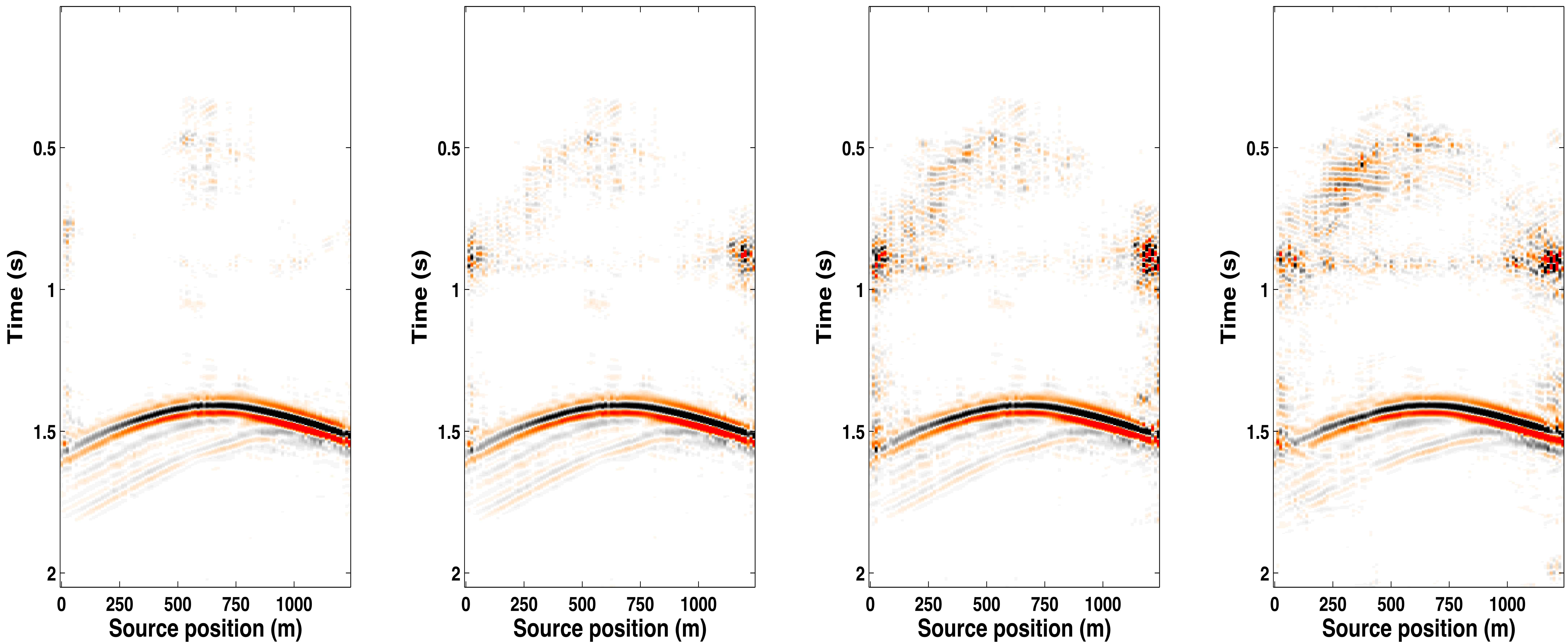
4-D time-jittered marine acquisition



4-D recovery - JRM
– **50% overlap** in acquisition matrices

0% overlap

no error [12.2 dB] | error ≈ 1.0 m [8.5 dB] | error ≈ 2.8 m [3.8 dB] | [2.0 dB]



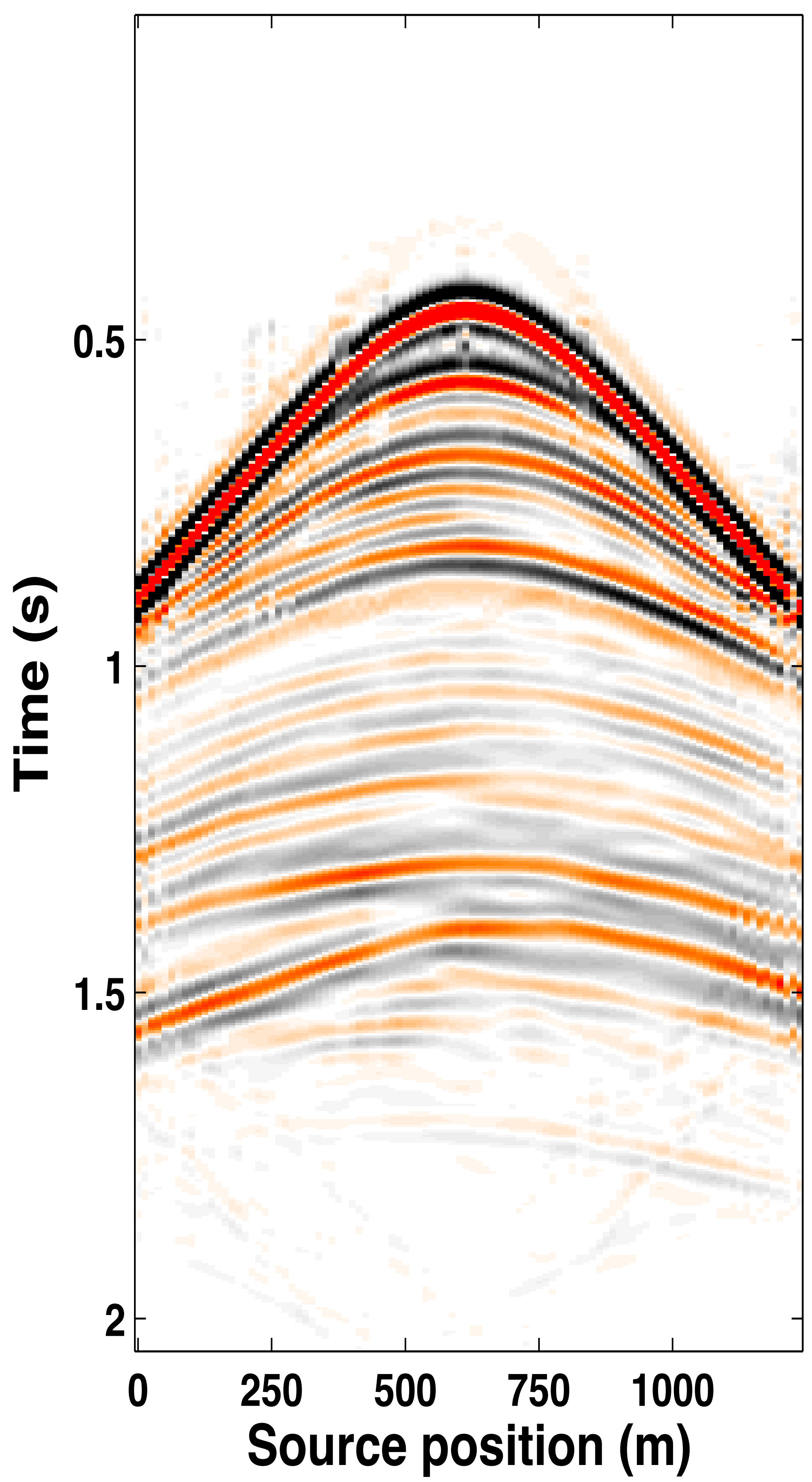
On the contrary,

location errors improve recovery of the vintages!

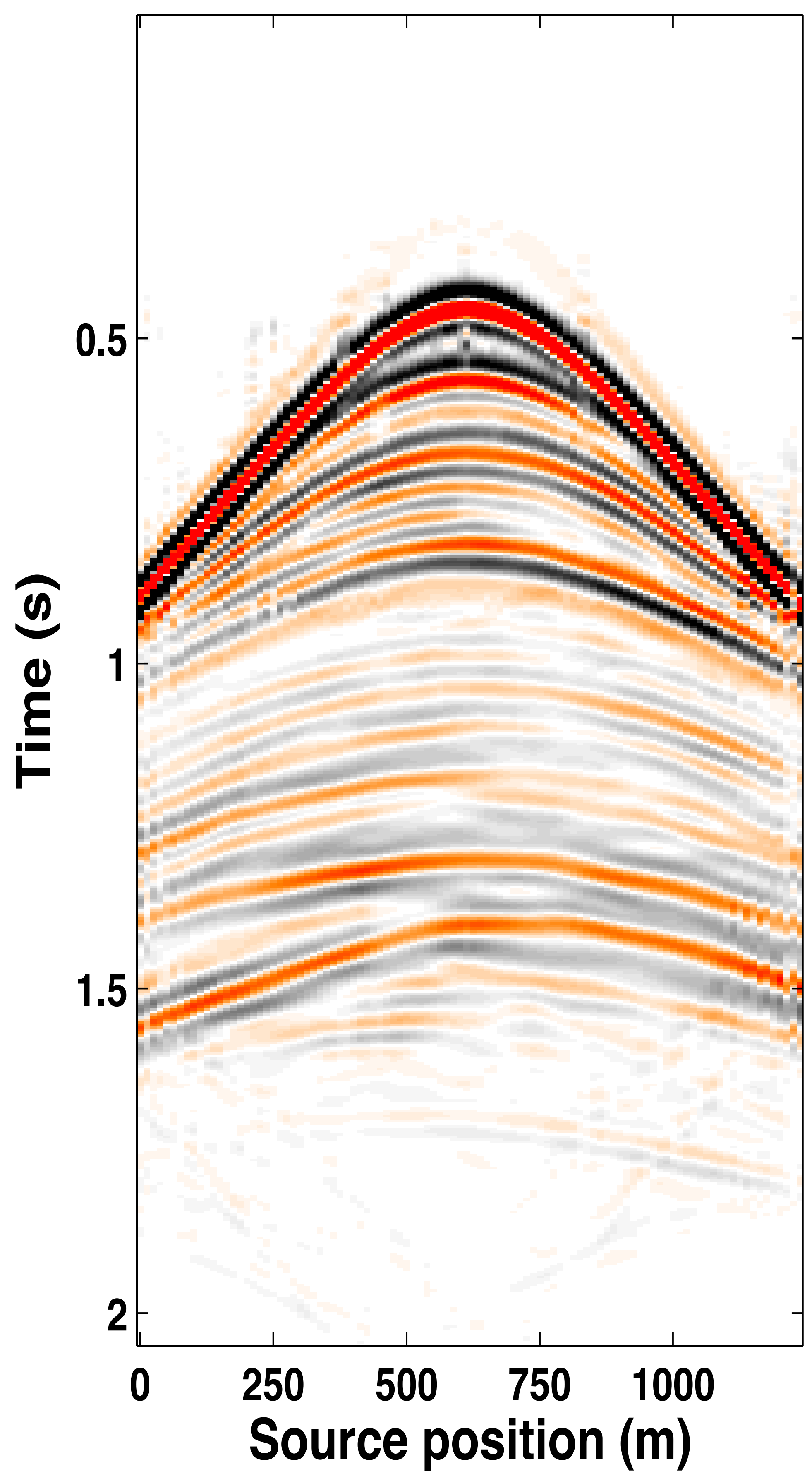
Monitor recovery - JRM

– 50% overlap in acquisition matrices

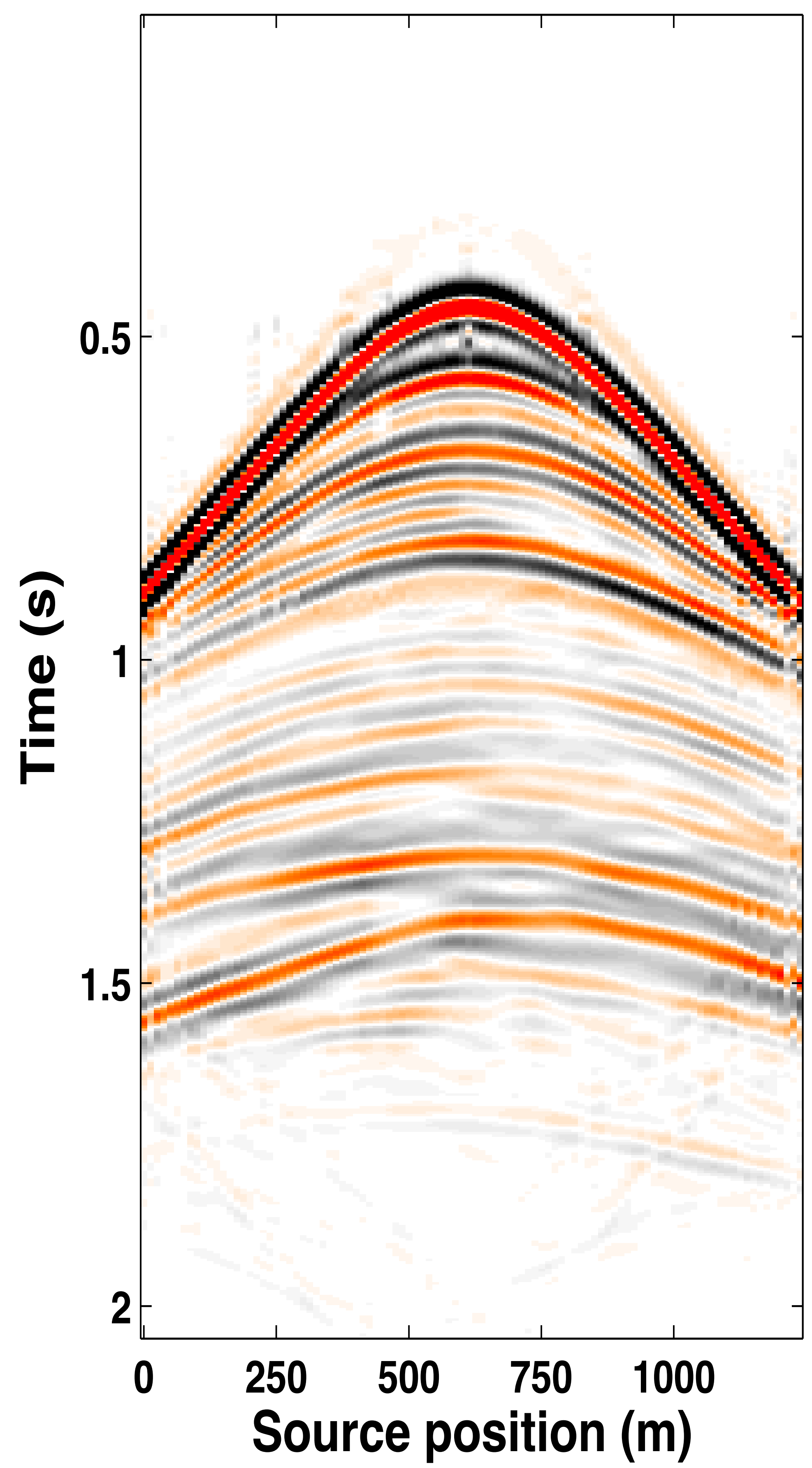
no error
[13.9 dB]



error ≈ 1.0 m
[14.5 dB]

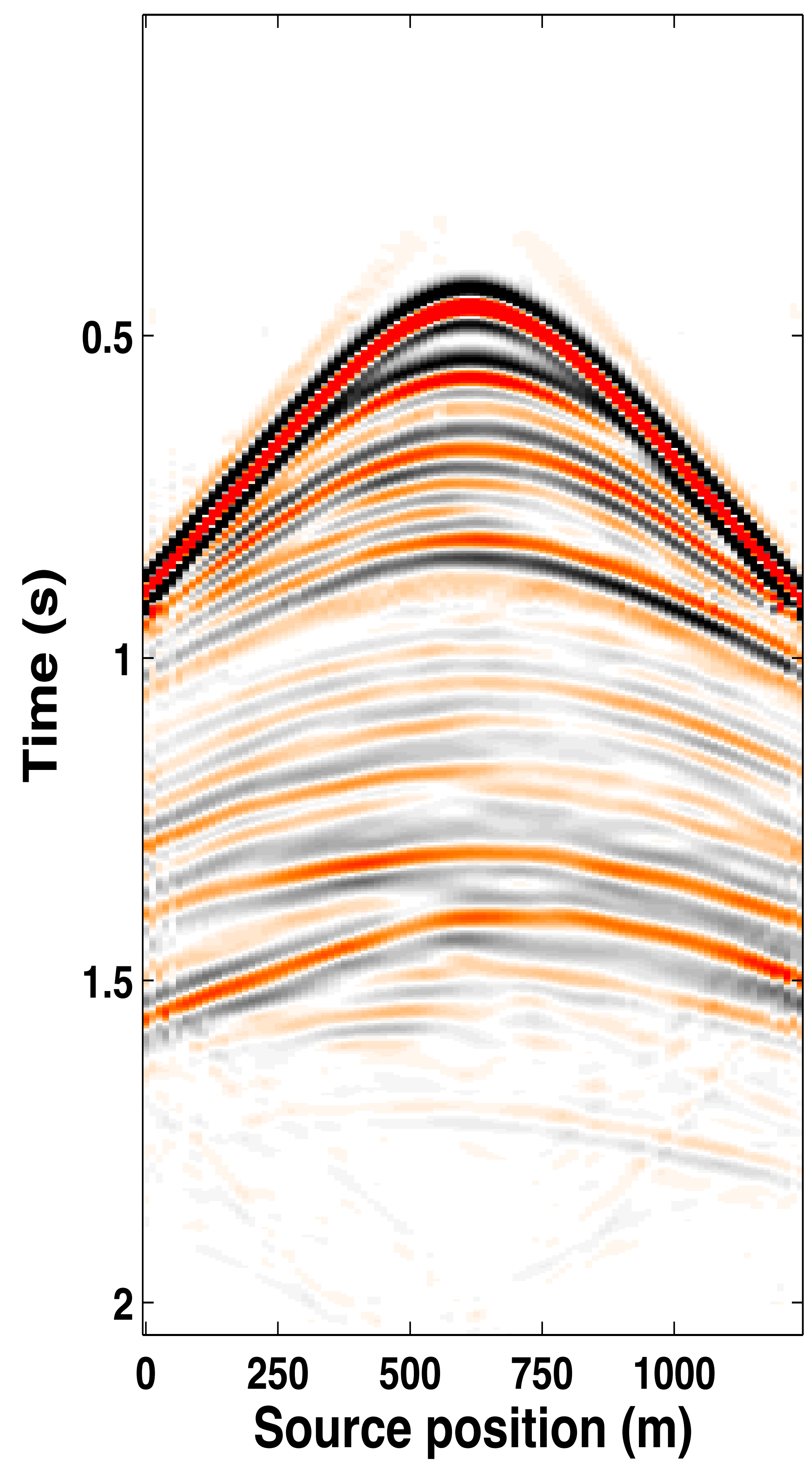


error ≈ 2.8 m
[15.5 dB]



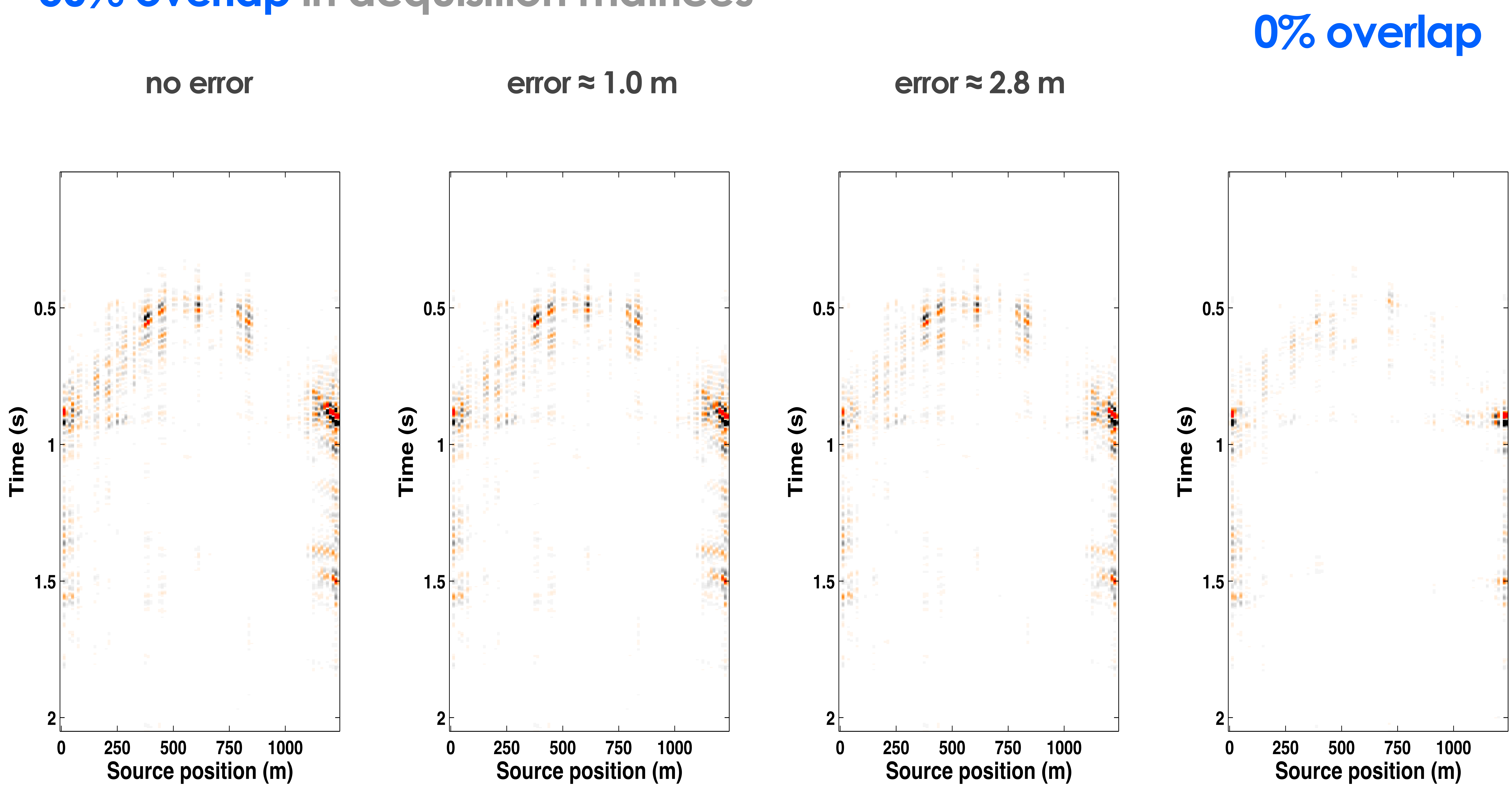
0% overlap

[18.3 dB]



Monitor residual - JRM

– **50% overlap** in acquisition matrices



Observations

In the given context of randomized subsampling, errors in the shot locations

- deteriorate recovery of the time-lapse signal
- improve recovery of the vintages

“Exact” repeatability of the surveys seems essential for good recovery of prestack time-lapse signals

However, most time-lapse studies involve poststack attibutes suggesting not to repeat in the field...

Randomized computations

Drivers:

- wave-equation inversions are computationally prohibitively expensive
- withstands their widespread adaptation
- challenges development of resilient workflows, inclusion of more complex wave physics, and assessment of risk

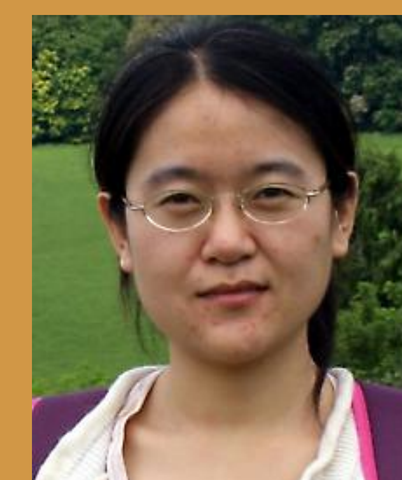
Solution:

- remove insistence of "touching all data" for each iteration while still leveraging the fold
- work on small randomized subsets of data
  (random batches of shots / randomized composite shots)
- control sub-sampling related artifacts via averaging or structure promotion
- randomized computations = increased imaging productivity

(Time-lapse) reverse-time migration w/ multiples, source estimation & gaps

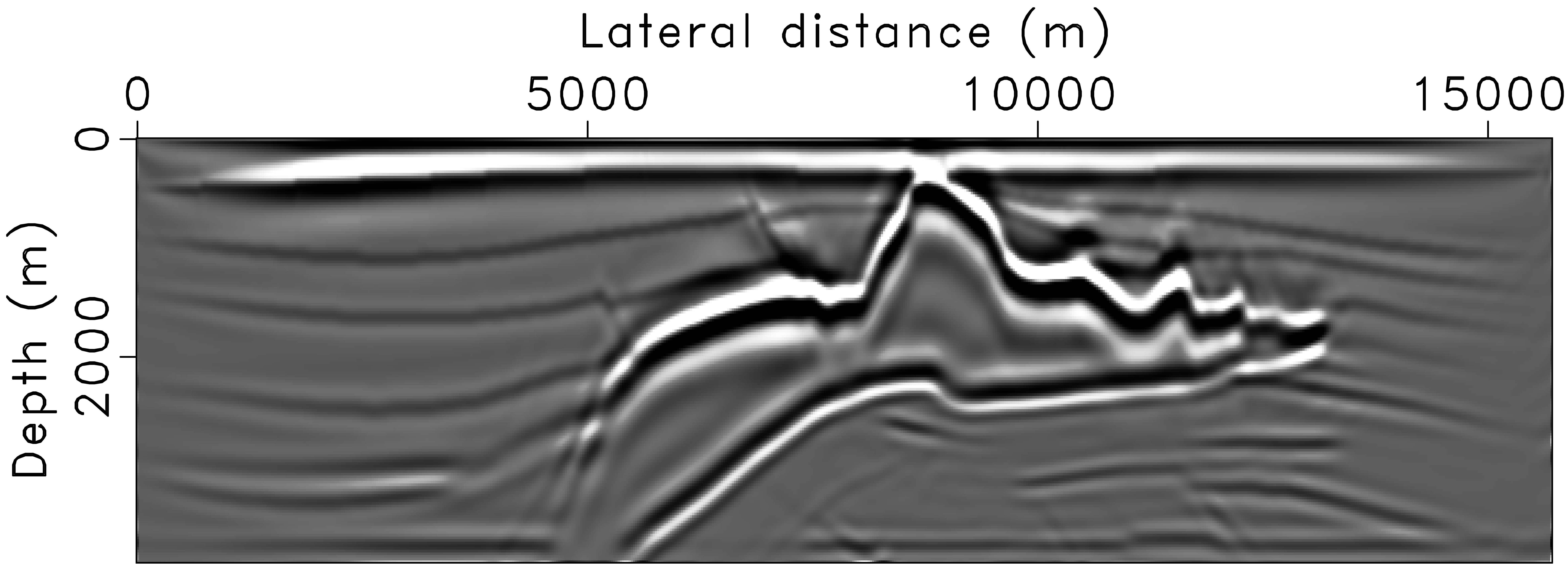
Xiang Li, Felix Oghenekohwo, Ning Tu, Phillip Witte, and Mengmeng Yang



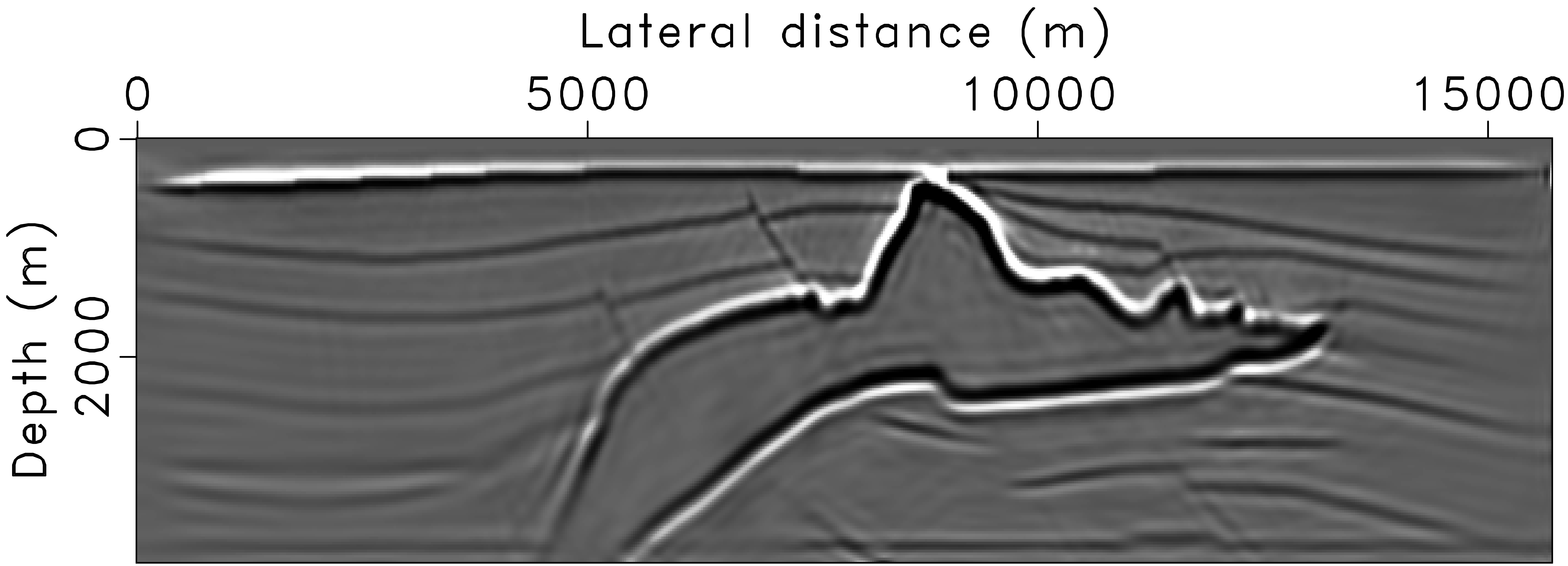
SLIM
University of British Columbia

From processing to inversion



RTM imaging via **adjoint**, high-pass filtered to remove low-wavenumber RTM artifacts

From processing to inversion



SPLSM image via **inversion**, # of wave-equation solves roughly equals 1 RTM w/ all data

Motivation

Wave-equation based imaging (migration) is expensive

- insists on touching all data (= all RHS's)
- low resolution

Linearized wave-equation based inversion is prohibitively expensive

- touches all data for each iteration
- restores amplitudes (corrects for GN-Hessian)
- high-resolution when exploiting structure (e.g. sparsity)

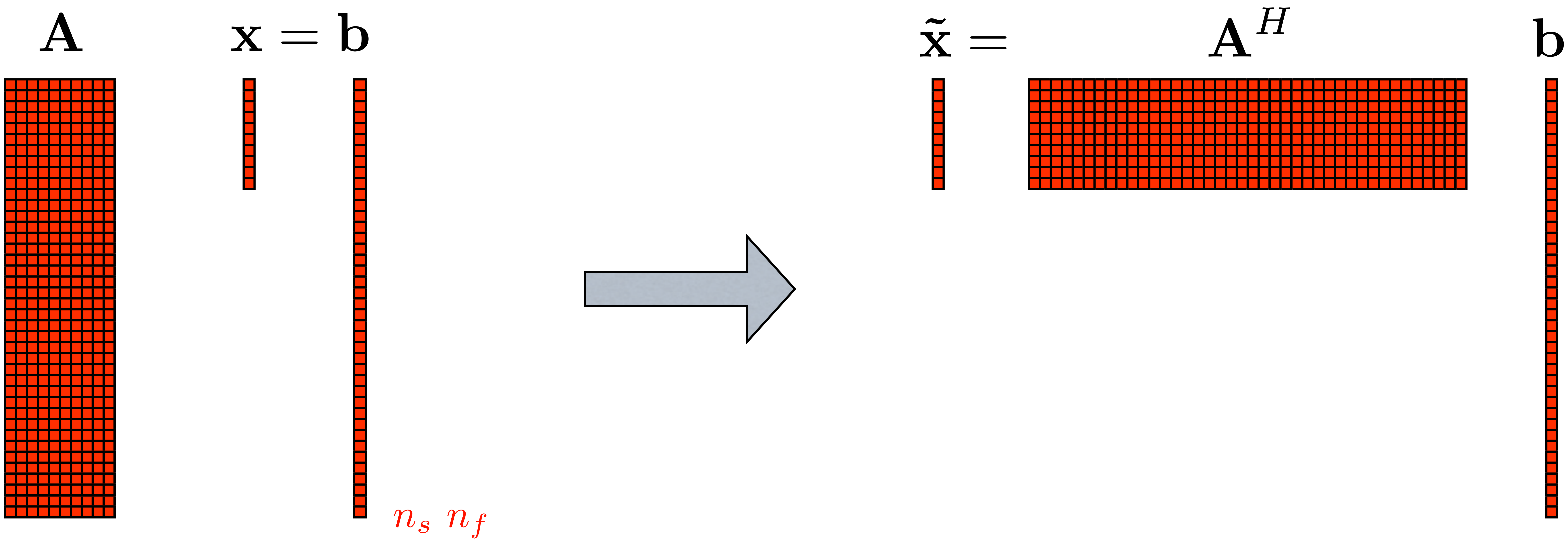
Leverage randomized sampling techniques...

[Herrmann & Li, ’12; Ning & Herrmann, ’15]

Migration

Seismic imaging is linear, separable but extreme large scale

- overdetermined, ill-conditioned & inconsistent system
- so far: solved by applying a (scaled) adjoint w/ many PDE solves



Migration

Solving $\mathbf{A}\mathbf{x} = \mathbf{b}$ with $\tilde{\mathbf{x}} = \mathbf{A}^H\mathbf{b}$ → image

Alternatively: least-squares solution via matrix-free iterations to

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$$

with solution $\mathbf{x} = (\underbrace{\mathbf{A}^H\mathbf{A}}_{\text{GN Hessian}})^{-1}\mathbf{A}^H\mathbf{b}$

Inverting GN Hessian expensive → solve with linear optimization

[Shen et. al. '01]

SLIM

Migration with sparsity promotion

Normal least-squares solution:

- does not exploit structure in $\mathbf{x}$
- requires many iterations (= data passes/# of PDE solves)

Sparsity-promoting inversion

- "classic" sparse recovery:

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{x}\|_1 \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}$$

- Basis Pursuit (BP)
- designed for underdetermined systems
  (but we will later see it works for randomly sampled systems too!)

SLIM

ISTA
Iterative Shrinkage Thresholding Algorithm

$$
\begin{array}{ll}
1. & \textbf{for } k = 0, 1, \cdots \\
2. & \quad \mathbf{z}_{k+1} = \mathbf{x}_k - t_k \mathbf{A}^*(\mathbf{A}\mathbf{x}_k - \mathbf{b}_k) \\
3. & \quad \mathbf{x}_{k+1} = S_\lambda(\mathbf{z}_{k+1}) \\
4. & \textbf{end for}
\end{array}
$$

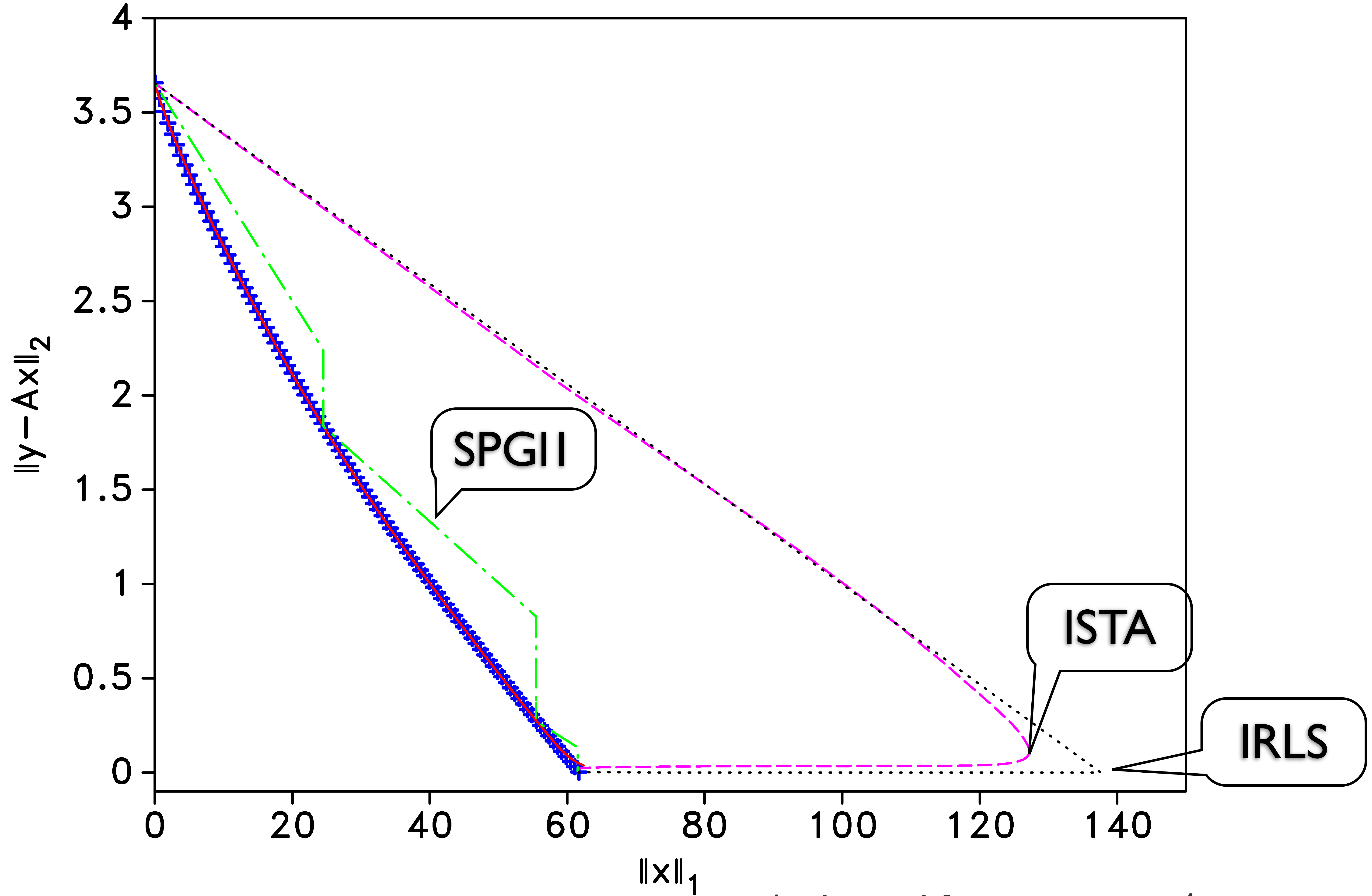
*where $S_\lambda(x) = \text{sign}(x) \cdot \max(|x| - \lambda, 0)$ is soft thresholding and t_k are step lengths

- simple but converges slowly, especially for λ small
- BP corresponds to non-trivial limit $\lambda \to 0^+$
- requires (complicated) continuation strategies for λ

Gilles Hennenfent, Ewout van den Berg, Michael P. Friedlander, and Felix J. Herrmann, "**New insights into one-norm solvers from the Pareto curve**", *Geophysics*, vol. 73, p. A23-A26, 2008.

SLIM

Solution paths



*adapted from 10.1190/1.2944169

Ewout van den Berg and Michael P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions", *SIAM Journal on Scientific Computing*, vol. 31, p. 890-912, 2008

SLIM

Observations

Contributions from "optimizers" yielded robust solvers such as SPGl1

- relatively fast because of continuation methods that relax the constraint
- black boxes with clever state-of-the-art "tricks"

But, their

- convergence is too slow for realistic seismic problems w/ expensive matvecs & IO
- implementation is rather complicated & somewhat inflexible
- design is not optimized for overdetermined problems

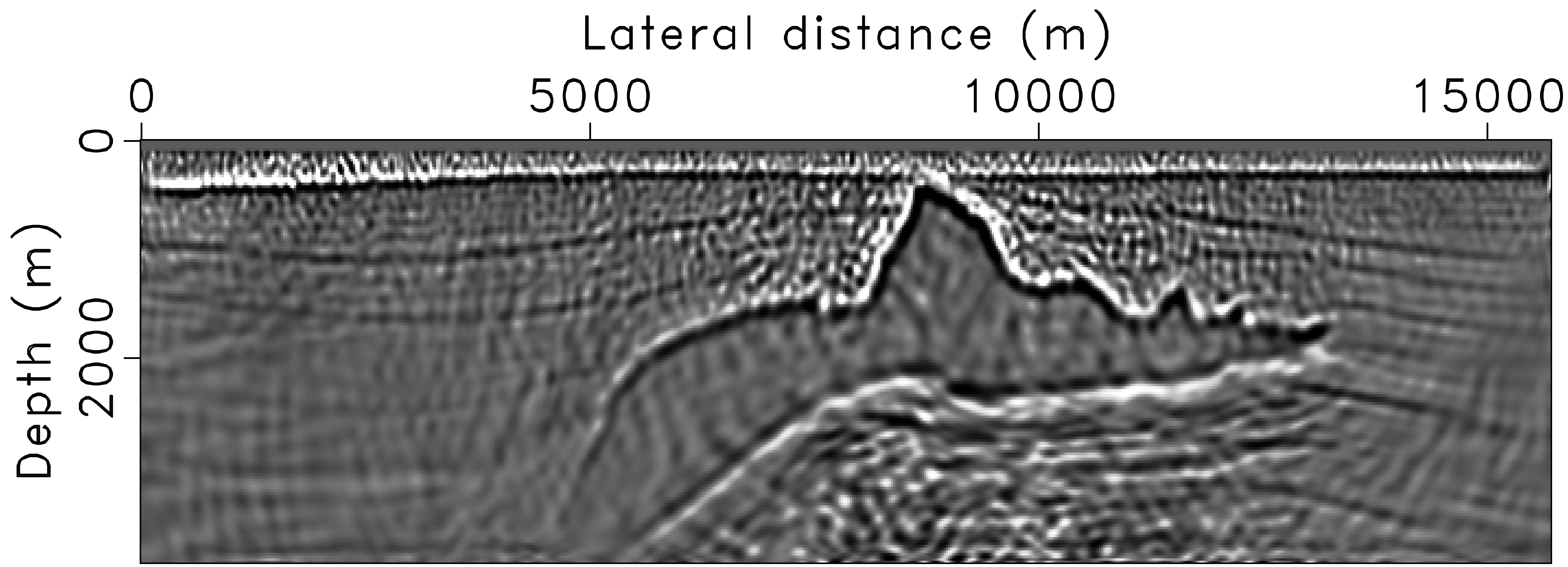
Suggests use of Stochastic Average Approximation (SAA) to reduce costs...

Eldad Haber, Matthias Chung, and Felix J. Herrmann, "**An effective method for parameter estimation with PDE constraints with multiple right hand sides**", *SIAM Journal on Optimization*, vol. 22, 2012.

SLIM

SPLSM w/ CS

slow convergence of SAA

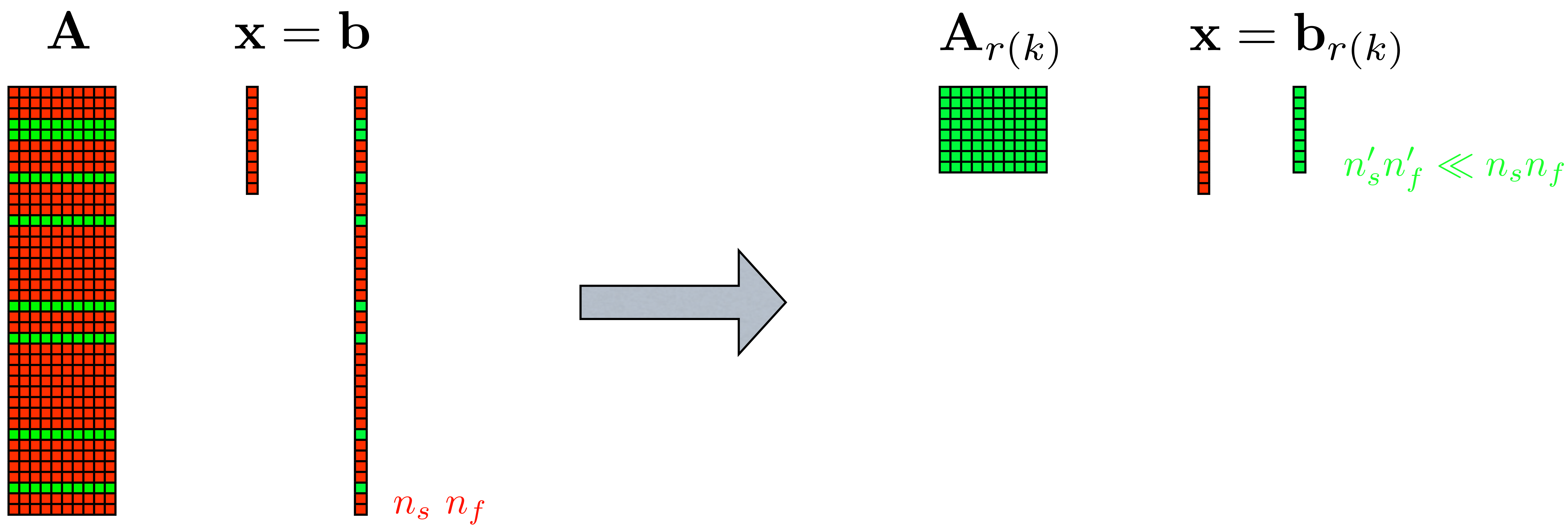


SPLSM image via **inversion** w/ **fixed** randomization

Randomized L1 solvers

For large scale seismic problems we are interested in:

- reducing time consuming I/O & PDE solves
- working on (random) subsets of data = row blocks of $\mathbf{A}$



Felix J. Herrmann and Xiang Li, "Efficient least-squares imaging with sparsity promotion and compressive sensing", *Geophysical Prospecting*, vol. 60, p. 696-712, 2012
Felix J. Herrmann, "Accelerated large-scale inversion with message passing", in *SEG Technical Program Expanded Abstracts*, 2012, vol. 31, p. 1-6.

SLIM

Randomized L1 solvers

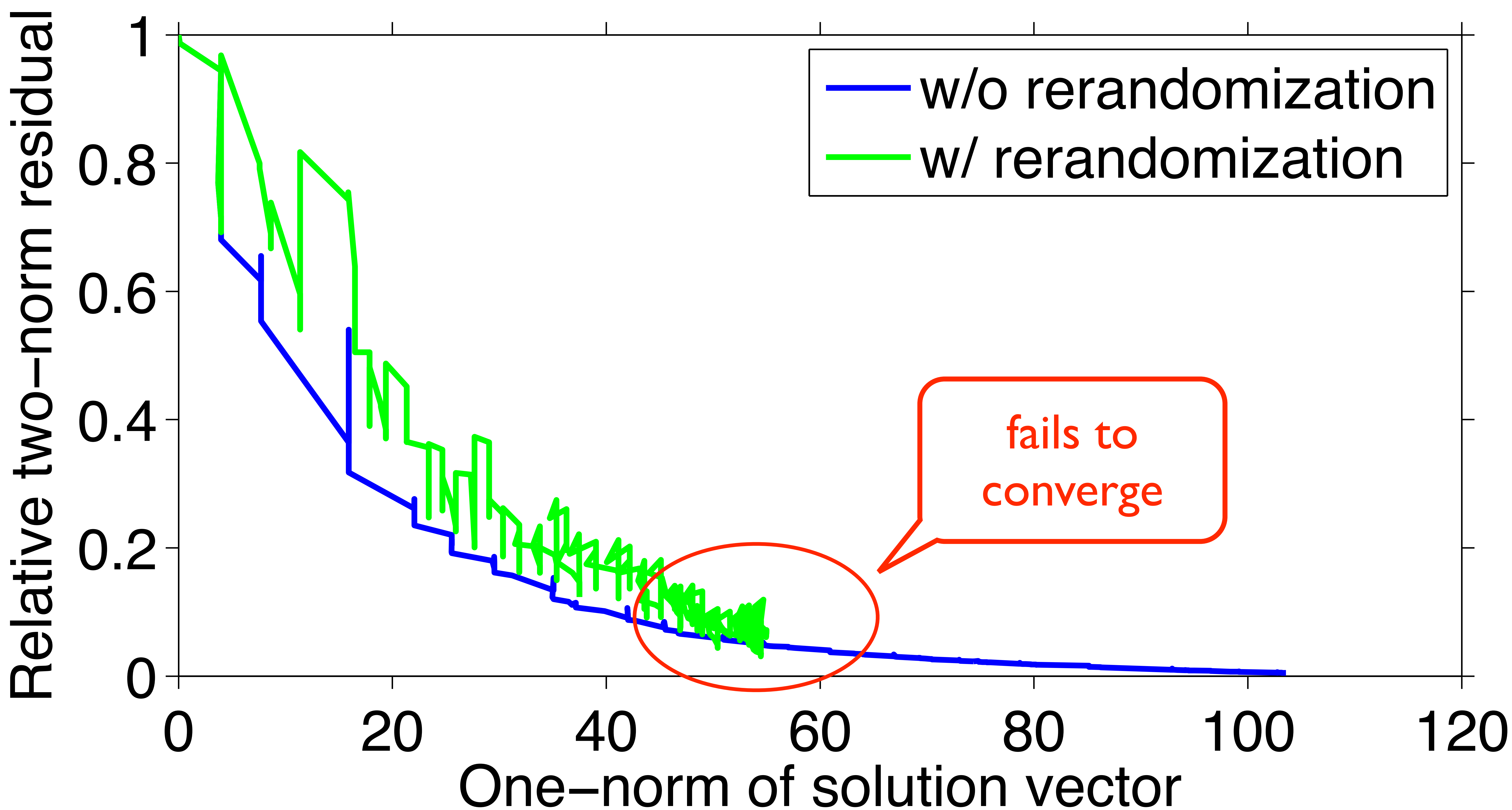
Randomized iterative soft thresholding algorithm (RISTA):

1. $\textbf{for } k = 0, 1, \cdots$
2. $\quad \mathbf{z}_{k+1} = \mathbf{x}_k - t_k \mathbf{A}^*_{r(k)} (\mathbf{A}_{r(k)} \mathbf{x}_k - \mathbf{b}_k)$
3. $\quad \mathbf{x}_{k+1} = S_{\lambda_k}(\mathbf{z}_{k+1})$
4. $\textbf{end for}$

*where $S_\lambda(x) = \mathrm{sign}(x) \cdot \max(|x| - \lambda, 0)$ is soft thresholding and t_k are step lengths

- relates to "approximate" message passing theory (Montanari, '09)
- reduces I/O, works on small subset of data
- only converges for special $\mathbf{A}^*, \mathbf{A}$ and tuned λ s

RISTA solution path



SLIM

Relaxed sparsity objective

Instead, consider

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}\|^2 \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

- strictly convex objective known as “elastic” net in machine learning
- equivalent to Basis Pursuit for “large enough” λ
- corresponds to [Lorentz et. al.,‘14]
  - sparse Kaczmarz for single-row $\mathbf{A}_k$’s
  - linearized Bregman for full $\mathbf{A}$’s

RISKA

Randomized IS Kaczmarz Algorithm w/ linearized Bregman

1. **for** $k = 0, 1, \cdots$
2. $\quad \mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^*(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$
3. $\quad \mathbf{x}_{k+1} = S_\lambda(\mathbf{z}_{k+1})$
4. **end for**

*where $t_k = \frac{\|\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|^2}{\|\mathbf{A}_k^*(\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k\|^2}$ are the step lengths

- exceedingly simple flexible “three line” algorithm
- gradient descend on the dual problem, which provably converges
- total different role for λ

SLIM

Linearized Bregman

Extension to handle noisy data

$$\begin{aligned}\underset{\mathbf{x}}{\text{minimize}} \quad & \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}\|^2 \\ \text{subject to} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{b}\| \leq \sigma\end{aligned}$$

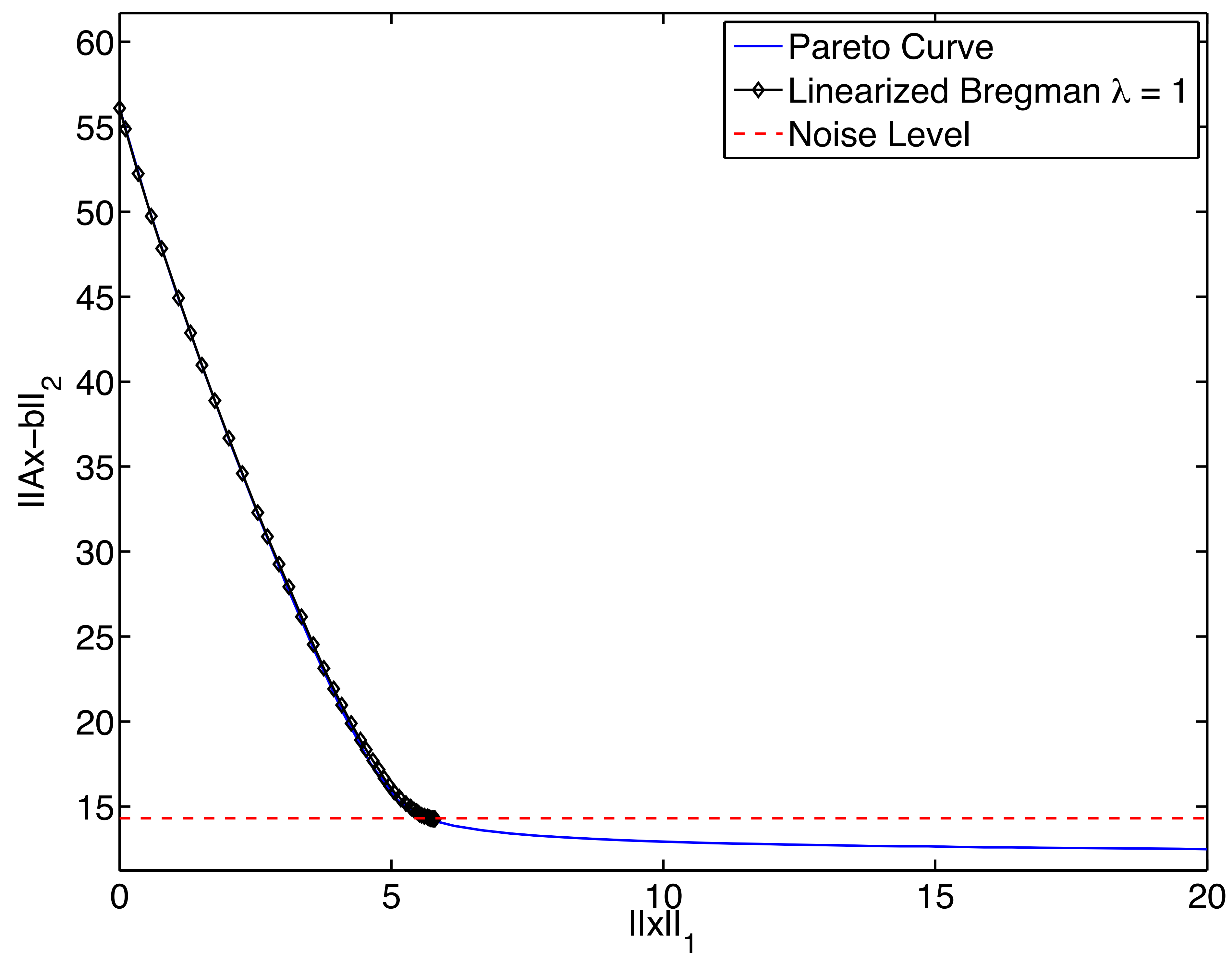
via projections onto norm balls

1. $\textbf{for } k = 0, 1, \cdots$
2. $\quad \mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}^*_{r(k)} \mathcal{P}_\sigma(\mathbf{A}_{r(k)}\mathbf{x}_k - \mathbf{b}_{r(k)})$
3. $\quad \mathbf{x}_{k+1} = S_\lambda(\mathbf{z}_{k+1})$
4. $\textbf{end for}$

*where $\mathcal{P}_\sigma(\mathbf{A}_{r(k)}\mathbf{x}_k - \mathbf{b}_{r(k)}) = \max\{0, 1 - \frac{\sigma}{\|\mathbf{A}_{r(k)}\mathbf{x}_k - \mathbf{b}_{r(k)}\|}\} \cdot (\mathbf{A}_{r(k)}\mathbf{x}_k - \mathbf{b}_{r(k)})$

Linearized Bregman solution path

Converges (up to noise level), even when working on randomized subsets of data + without difficult strategies for λ



Least-squares migration

Switch to seismic notation: $\delta \mathbf{d}_{ij} = \nabla \mathbf{F}_{ij}(\mathbf{m}_0, \mathbf{q}_{ij})\delta \mathbf{m}$

$\delta \mathbf{m}$: Model perturbation
$\delta \mathbf{d}$: data residual
$\nabla \mathbf{F}$: Born modeling operator
i, j : frequency, source index

Instead of applying $\nabla \mathbf{F}^H$ (migration)

$$\delta \mathbf{m} = \sum_{ij} \nabla \mathbf{F}_{ij}^H(\mathbf{m}_0, \mathbf{q}_{ij})\delta \mathbf{d}_{ij}$$

➡ Solve least-squares problem w/ sparsity constraint (SPLSM) using linearized Bregman

Fast SPLSM w/ CS
w/ randomized source subsets

$$\underset{\mathbf{x}}{\text{minimize}} \quad \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}\|^2$$

$$\text{subject to} \quad \sum_{ij} \|\nabla\mathbf{F}_{ij}(\mathbf{m}_0, \mathbf{q}_{ij})\mathbf{C}^*\mathbf{x} - \delta\mathbf{d}_{ij}\| \leq \sigma$$

By iterating

$\mathbf{C}$: Curvelet transform

1. **for** $k = 0, 1, \cdots$
2. $\Omega \in [1 \cdots n_f]$, $\Sigma \in [1 \cdots n_s]$ for $\#\Omega \ll n_f$, $\#\Sigma \ll n_s$
3. $\mathbf{A}_k = \{\nabla\mathbf{F}_{ij}(\mathbf{m}_0, \bar{\mathbf{q}}_{ij})\mathbf{C}^*\}_{i\in\Omega, j\in\Sigma}$ with $\bar{\mathbf{q}}_{ij} = \sum_{l=1}^{n_s} w_l \mathbf{q}_{i,l}$
4. $\mathbf{b}_k = \{\delta\bar{\mathbf{d}}_{ij}\}_{i\in\Omega, j\in\Sigma}$ with $\delta\bar{\mathbf{d}}_{ij} = \sum_{l=1}^{n_s} w_l \delta\mathbf{d}_{i,l}$
5. $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^* \mathcal{P}_\sigma(\mathbf{A}_k\mathbf{x}_k - \mathbf{b}_k)$
5. $\mathbf{x}_{k+1} = S_\lambda(\mathbf{z}_{k+1})$
6. **end for**

Fast SPLSM w/ CS
experimental setup

Data:

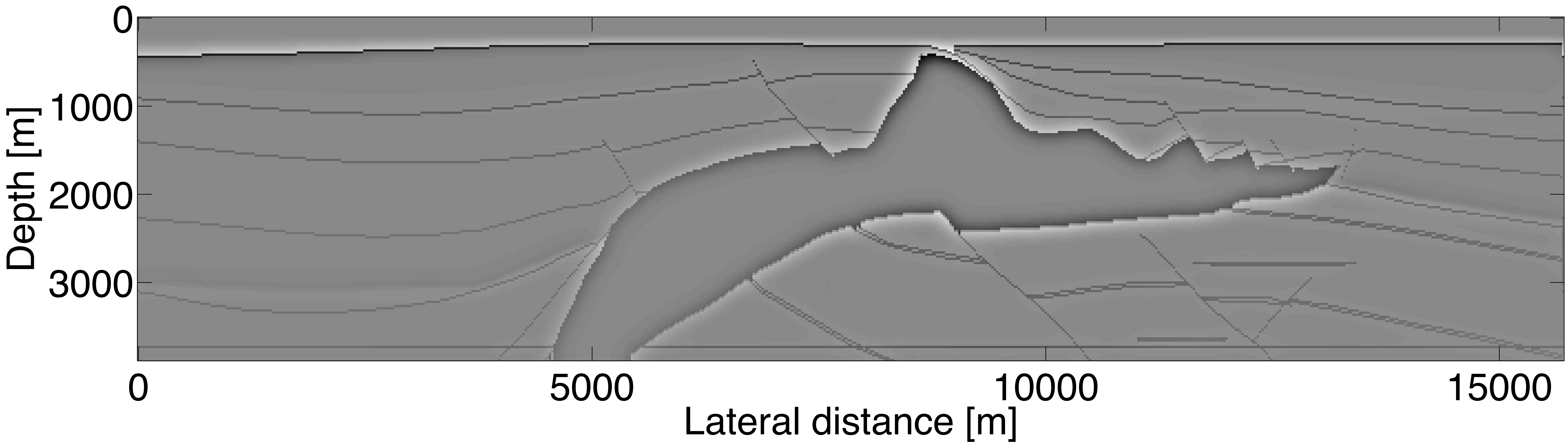
- 320 sources and receivers
- 72 frequency slices ranging from 3 – 12 Hz
- $\delta \mathbf{d} = \mathbf{F}(\mathbf{m}) - \mathbf{F}(\mathbf{m}_0)$, generated with separate modeling engine

Experiments:

- one pass through the data with different batch/block sizes
- choose λ according to $\max\left(t_1 \cdot \mathbf{A}_1^* \mathbf{b}_1\right)$ and number of iterations
- no source estimation – use correct source for linearized inversions

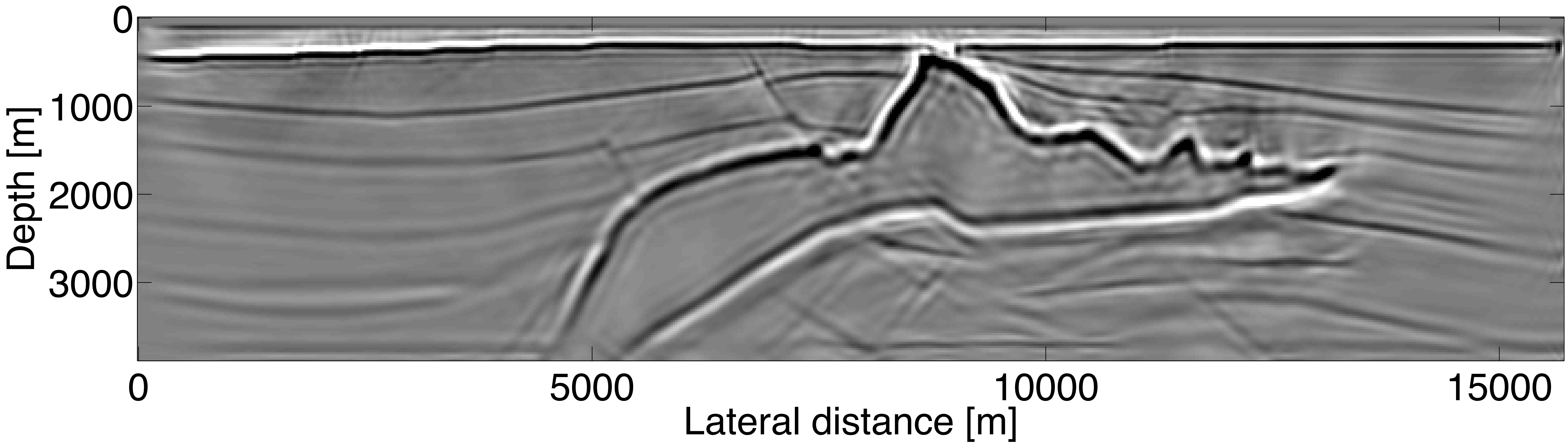
Fast SPLSM w/ CS

True model perturbation



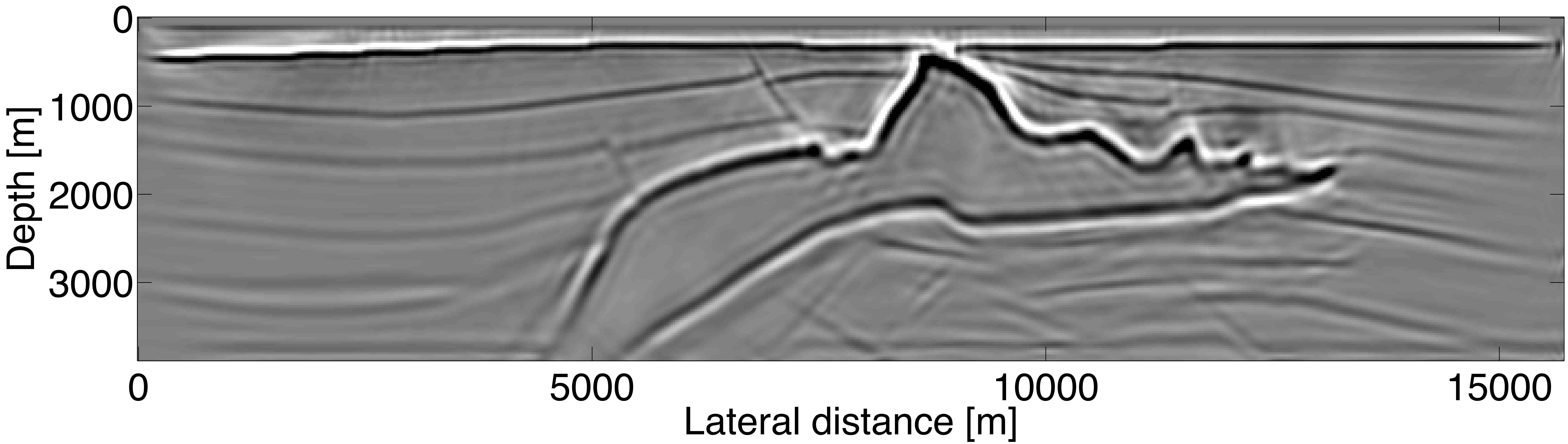
Fast SPLSM w/ CS

360 iterations, each w/ 8 frequencies/source experiments



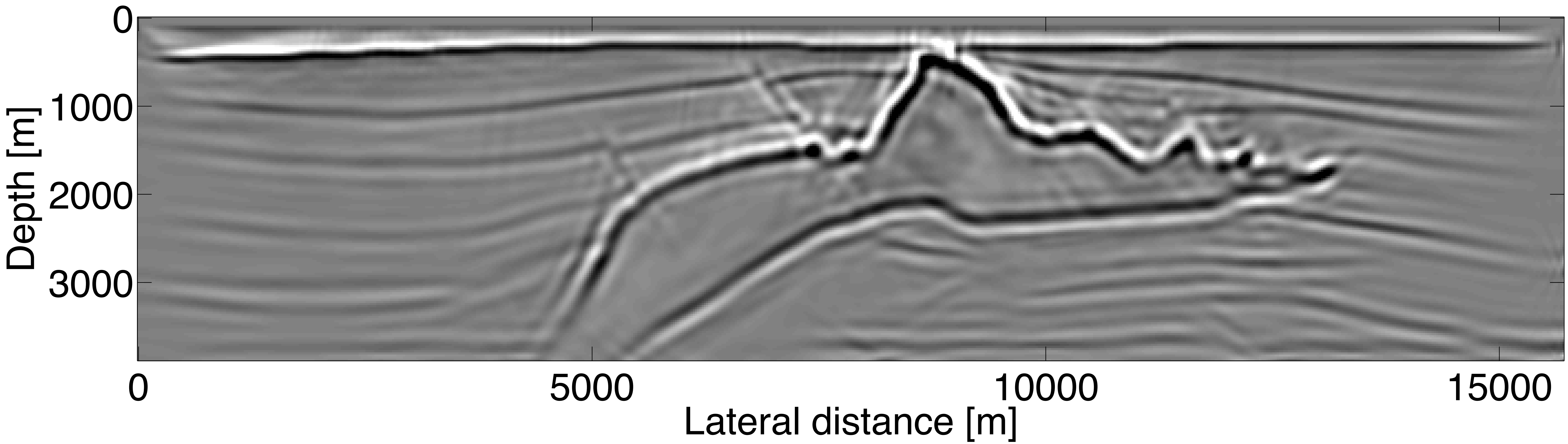
Fast SPLSM w/ CS

90 iterations, each w/ 16 frequencies/source experiments



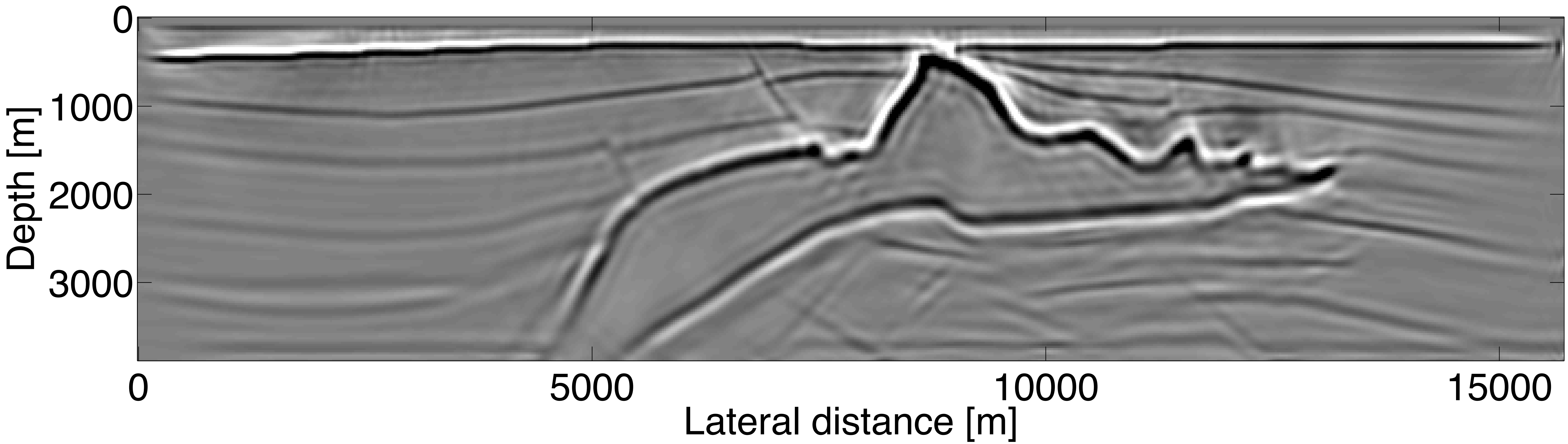
Fast SPLSM w/ CS

23 iterations, each w/ 32 frequencies/source experiments



Fast SPLSM w/ CS

90 iterations, each w/ 16 frequencies/source experiments



Aleksandr Y. Aravkin and Tristan van Leeuwen, "Estimating nuisance parameters in inverse problems", *Inverse Problems*, vol. 28, 2012
Ning Tu, Aleksandr Y. Aravkin, Tristan van Leeuwen, Tim T.Y. Lin, and Felix J. Herrmann, "Source estimation with surface-related multiples—-fast ambiguity-resolved seismic imaging". 2015

SLIM

Fast SPLSM w/ CS
w/ source estimation w/ variable projection

$$\underset{\mathbf{x}}{\text{minimize}} \quad \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}\|^2$$

$$\text{subject to} \quad \sum_{ij}\|\nabla\mathbf{F}_{ij}(\mathbf{m}_0,\mathbf{q}_{ij})\mathbf{C}^*\mathbf{x} - \delta\mathbf{d}_{ij}\| \leq \sigma$$

By iterating

1. $\textbf{for } k = 0, 1, \cdots$
2. $\Omega \in [1\cdots n_f],\ \Sigma \in [1\cdots n_s]$ for $\#\Omega \ll n_f,\ \#\Sigma \ll n_s$
3. $\mathbf{A}_k = \{\nabla\mathbf{F}_{ij}(\mathbf{m}_0,\bar{\mathbf{q}}_{ij})\mathbf{C}^*\}_{i\in\Omega,j\in\Sigma}$ with $\bar{\mathbf{q}}_{ij} = \sum_{l=1}^{n_s} w_l\mathbf{q}_{i,l}$
4. $\mathbf{b}_k = \{\delta\bar{\mathbf{d}}_{ij}\}_{i\in\Omega,j\in\Sigma}$ with $\delta\bar{\mathbf{d}}_{ij} = \sum_{l=1}^{n_s} w_l\delta\mathbf{d}_{i,l}$
5. $\bar{\mathbf{q}}_{ij} = \frac{<A_k x_k, b_k>}{<A_k x_k, A_k x_k>},\ \mathbf{A}_k = \{\nabla\mathbf{F}_{ij}(\mathbf{m}_0,\bar{\mathbf{q}}_{ij})\mathbf{C}^*\}_{i\in\Omega,j\in\Sigma}$
6. $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k\mathbf{A}_k^*\mathcal{P}_\sigma(\mathbf{A}_k\mathbf{x}_k - \mathbf{b}_k)$
7. $\mathbf{x}_{k+1} = S_\lambda(\mathbf{z}_{k+1})$
8. $\textbf{end for}$

Fast SPLSM w/ source estimation

experimental setup

Data:

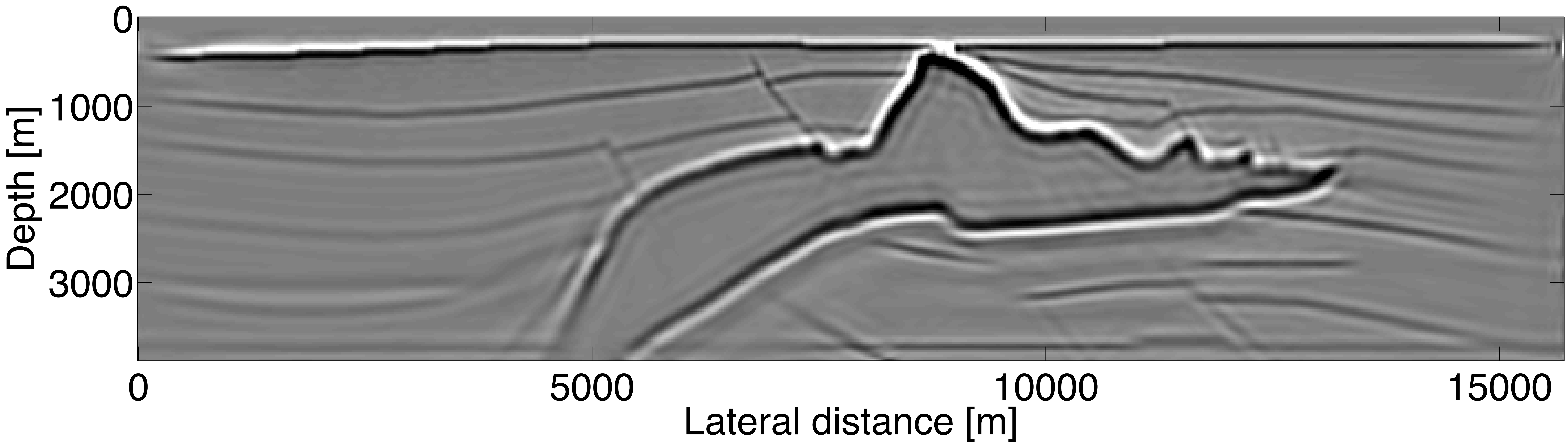
- 320 sources and receivers
- 72 frequency slices ranging from 3 - 12 Hz
- $\delta \mathbf{d} = \mathbf{F}(\mathbf{m} - \mathbf{m}_0)$ inverse crime data

Experiments:

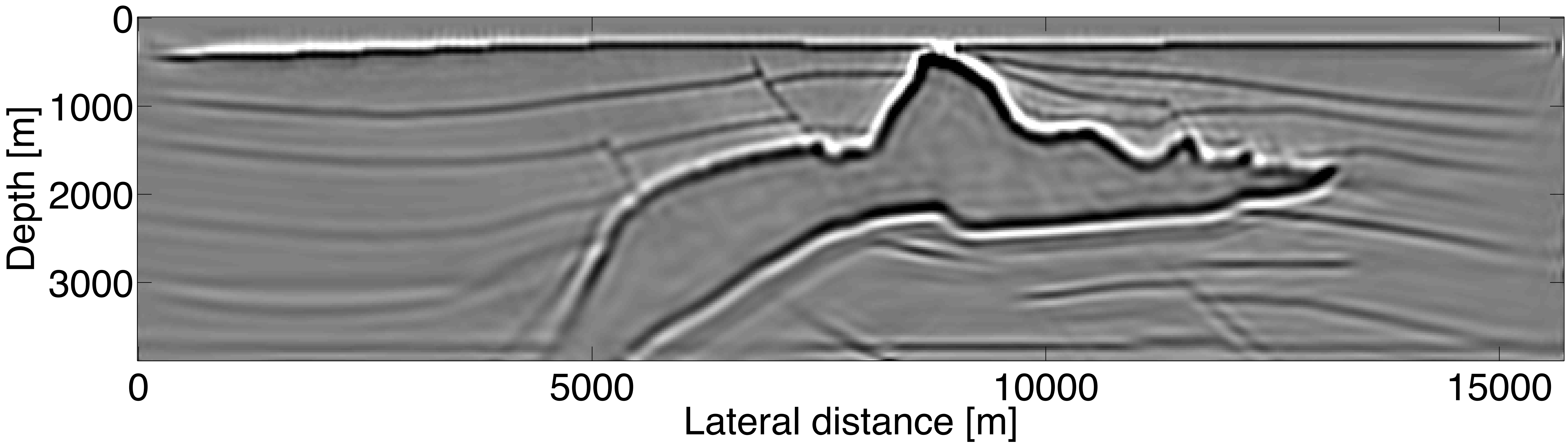
- one pass through the data with the same block size
- simultaneous sources
- choose λ according to $\max\left(t_1 \cdot \mathbf{A}_1^* \mathbf{b}_1\right)$
- source estimation with delta Dirac as initial guess
- estimated source scaled w.r.t. true source

Fast SPLSM w/ source estimation

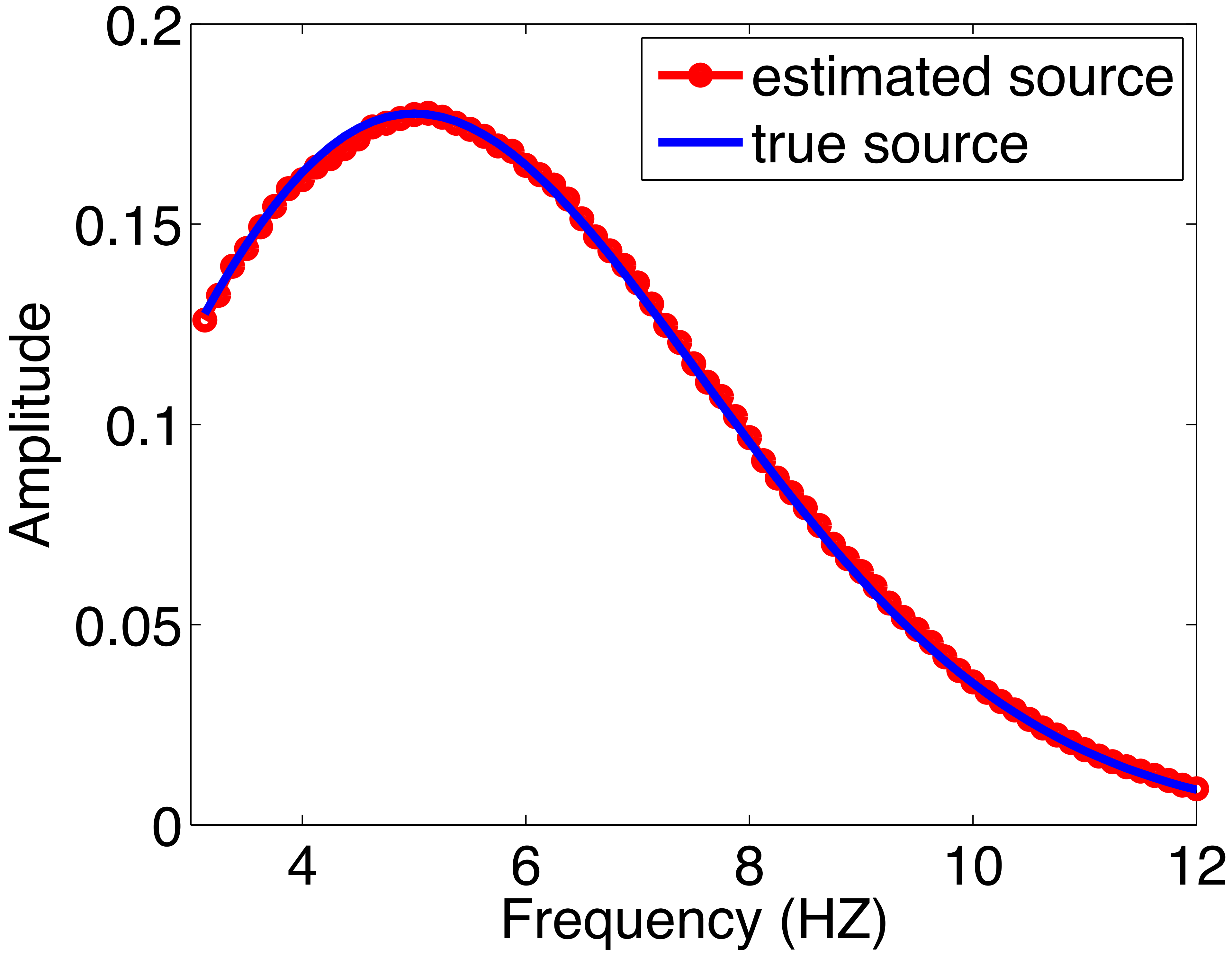
80 iterations, each w/ 72 frequencies/4sim. shots & true source

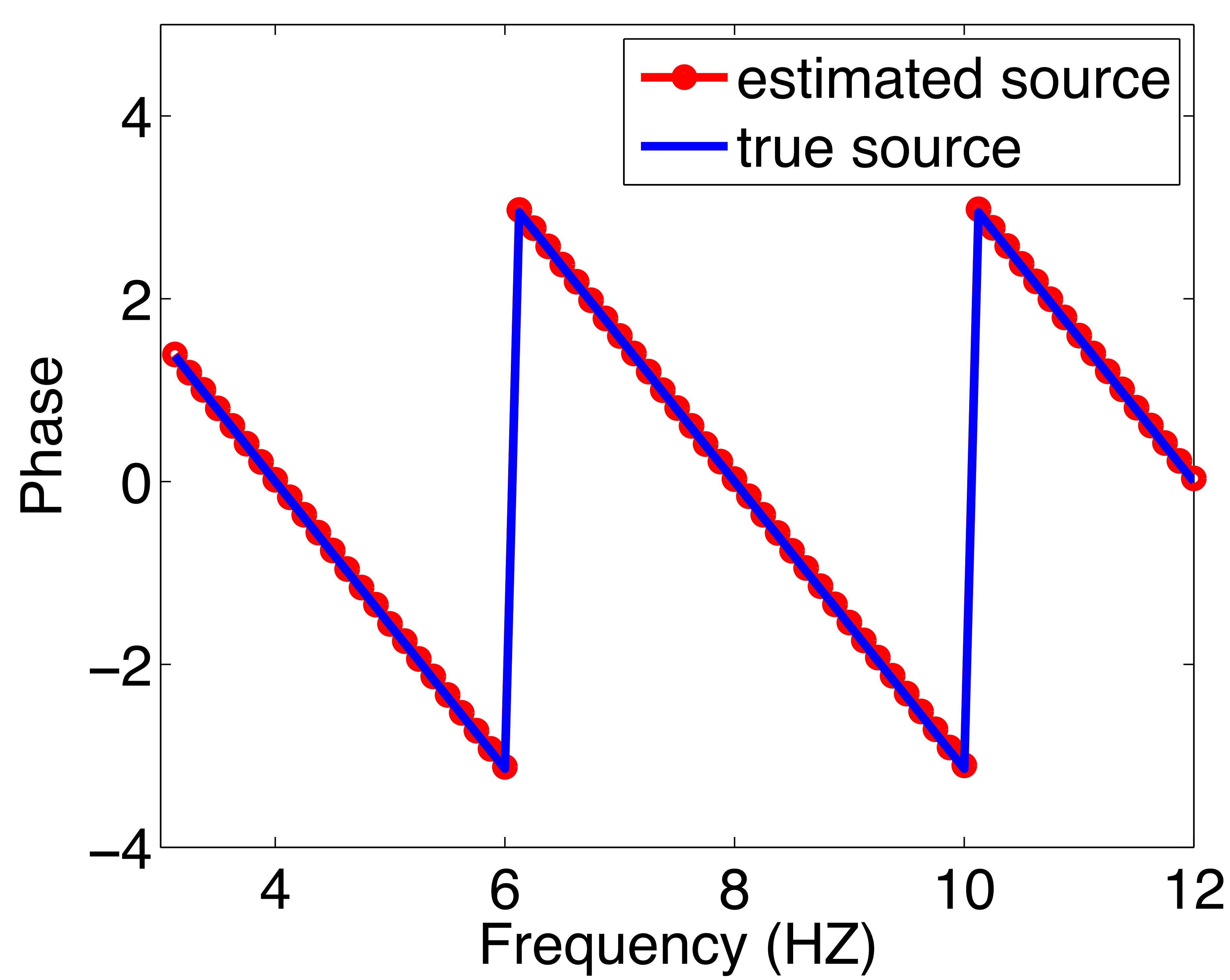


Fast SPLSM w/ source estimation
estimated source



Fast SPLSM w/ source estimation
– estimated source





Ning Tu and Felix J. Herrmann, "Fast imaging with surface-related multiples by sparse inversion", *Geophysical Journal International*, vol. 201, p. 304-317, 2015

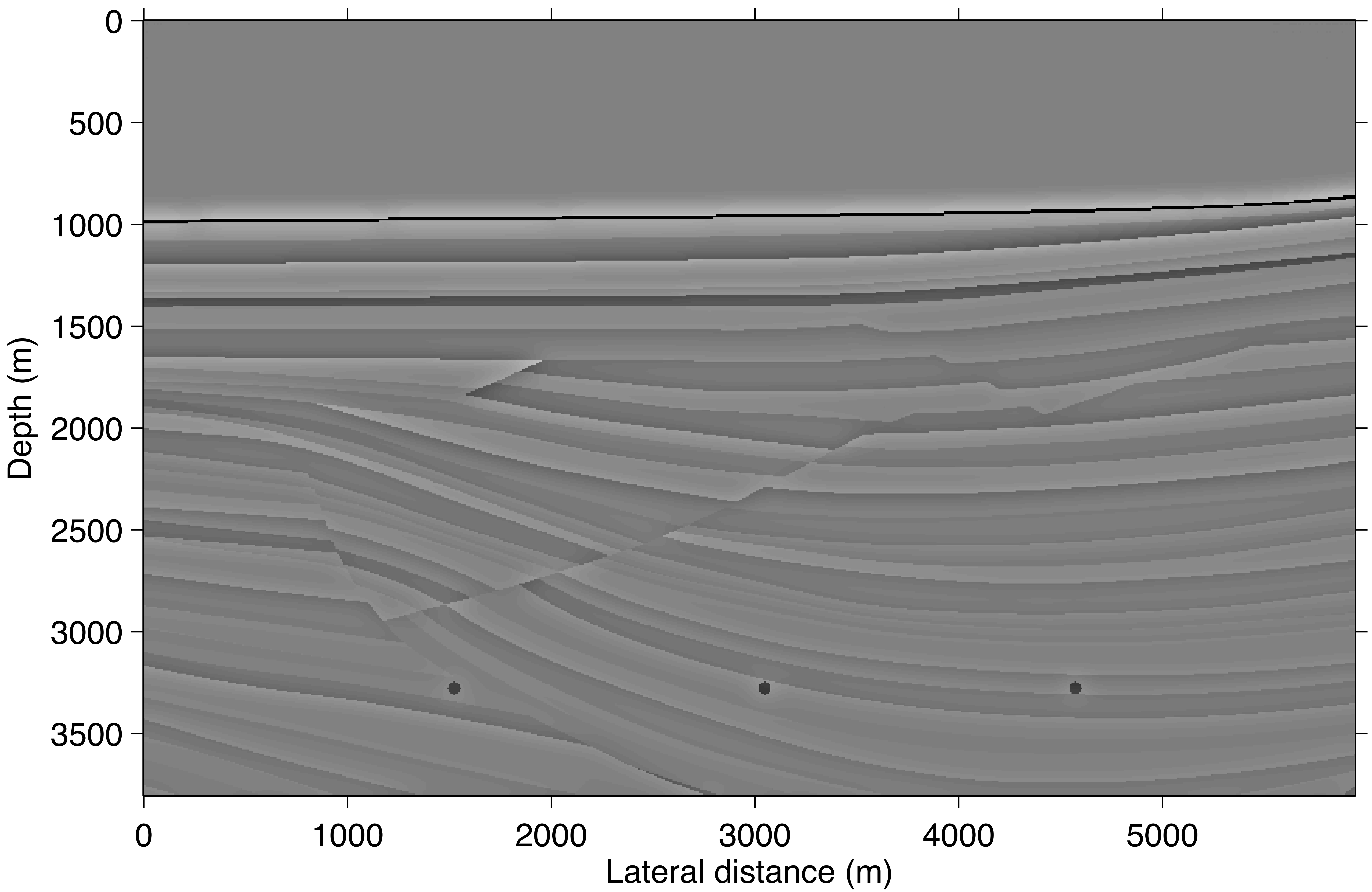
SLIM

Extension
imaging w/ surface-related multiples

Incorporate predictor of surface-related multiples via areal sources

$$f(\mathbf{x}, \boldsymbol{w}) \doteq \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\delta\bar{\mathbf{d}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, s_i\bar{\mathbf{q}}_j - \delta\bar{\mathbf{d}}_{i,j}]\mathbf{C}^*\mathbf{x}\|_2^2$$

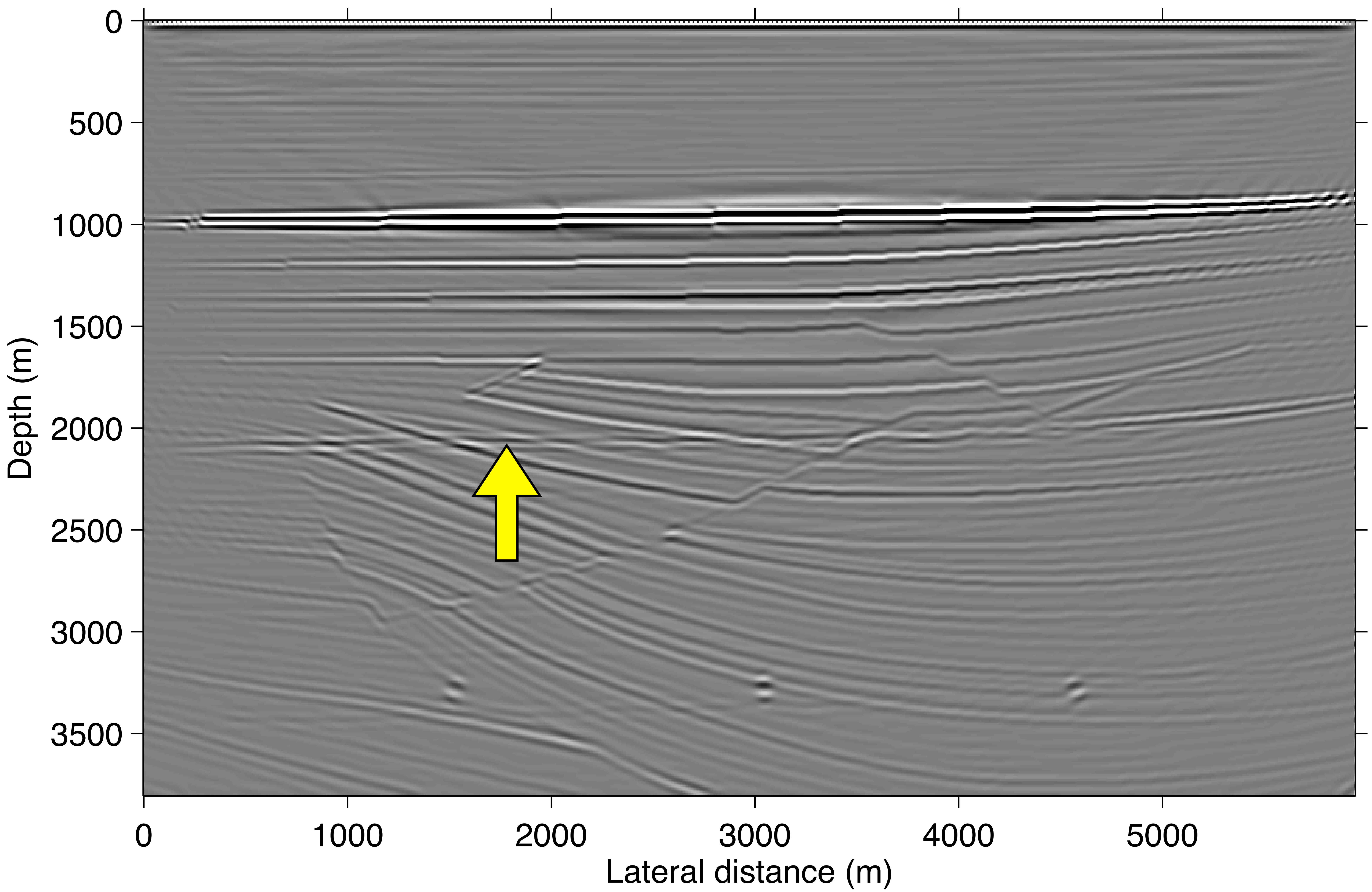
True image



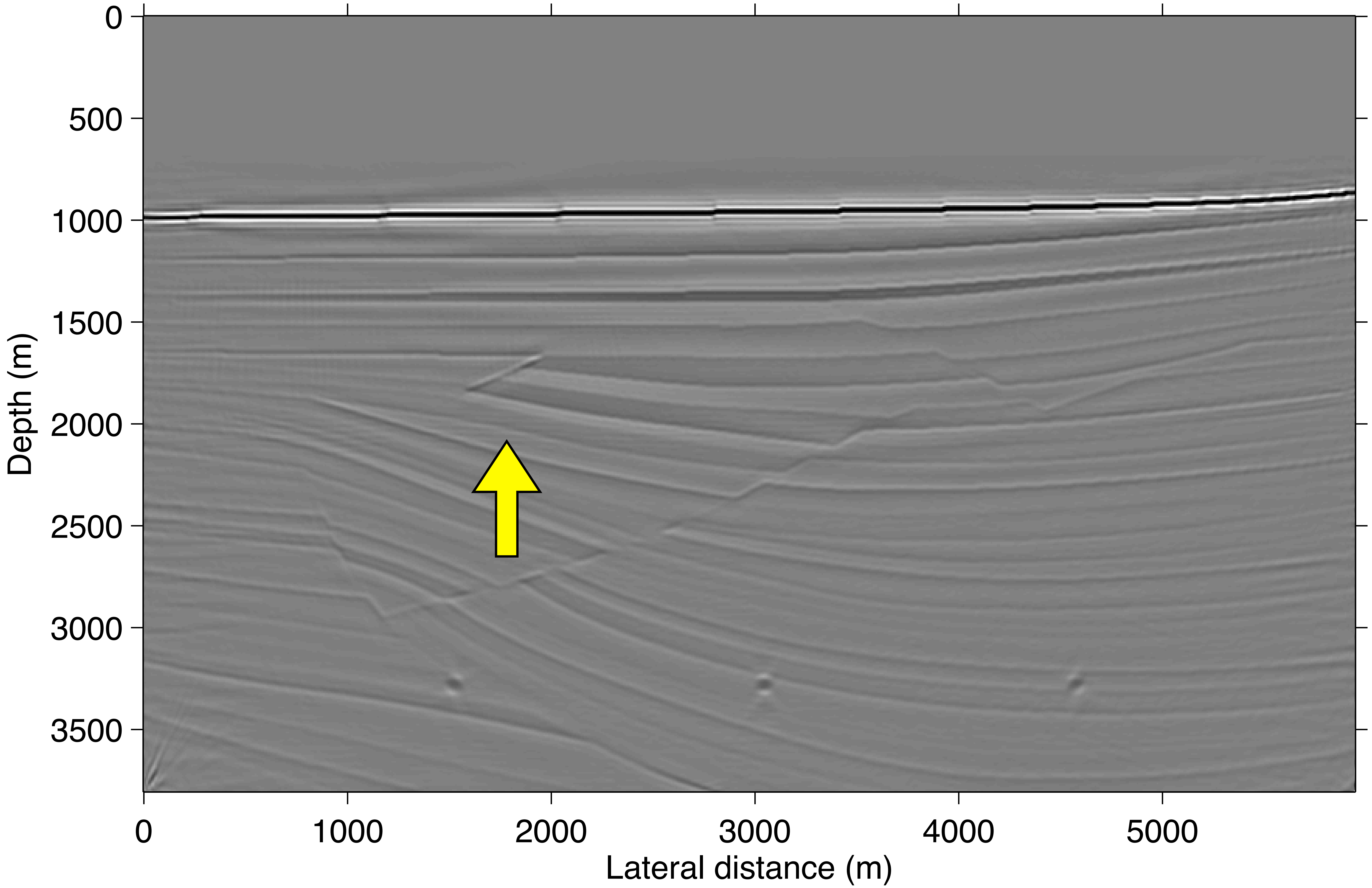
0
500
1000
1500
2000
2500
3000
3500
Depth (m)
0
1000
2000
3000
4000
5000
Lateral distance (m)

RTM w/ multiples

SLIM

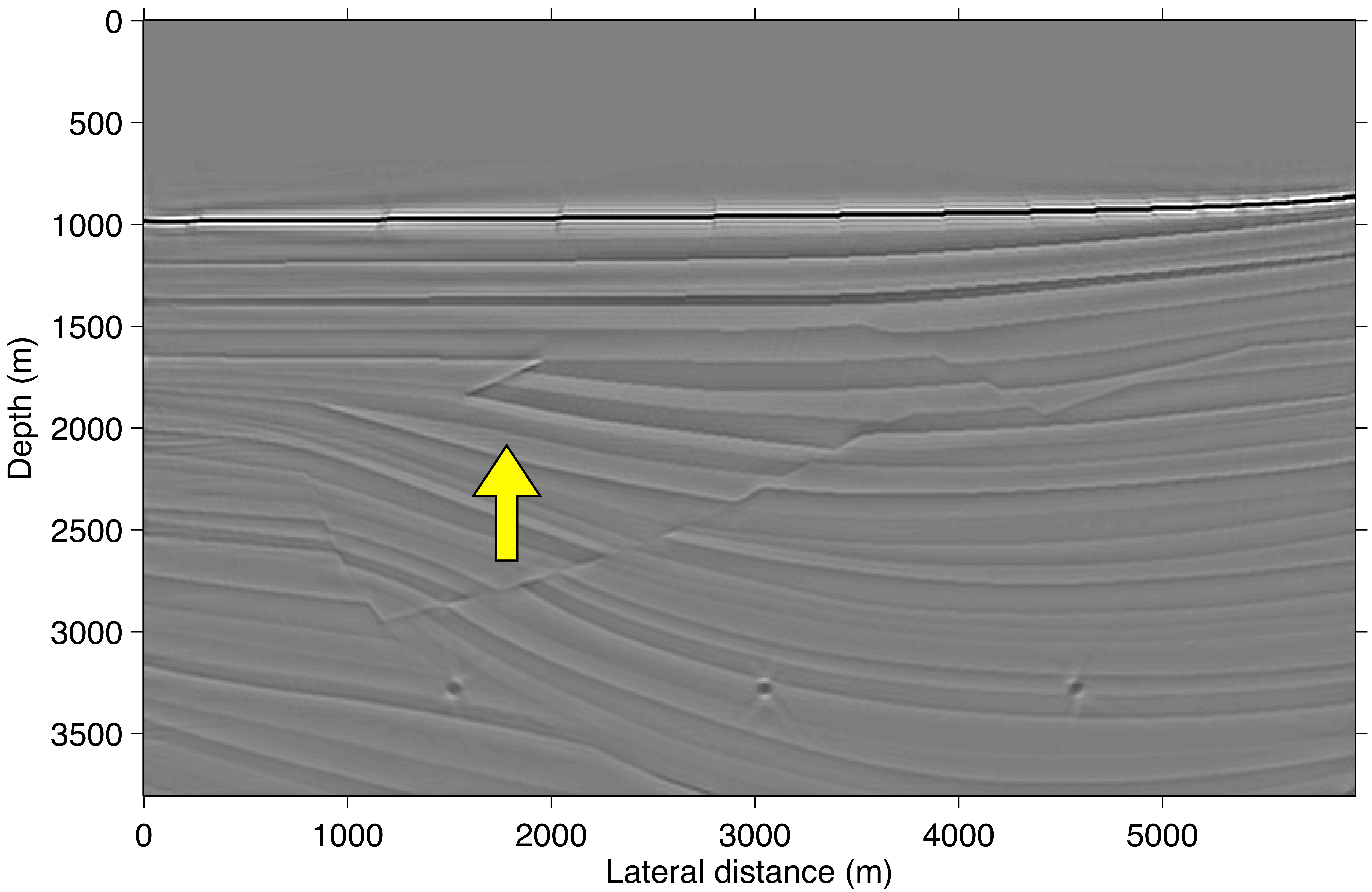


Fast SPLSM w/ multiples by SPGl1



Simulation cost **~1 RTM** using all the data

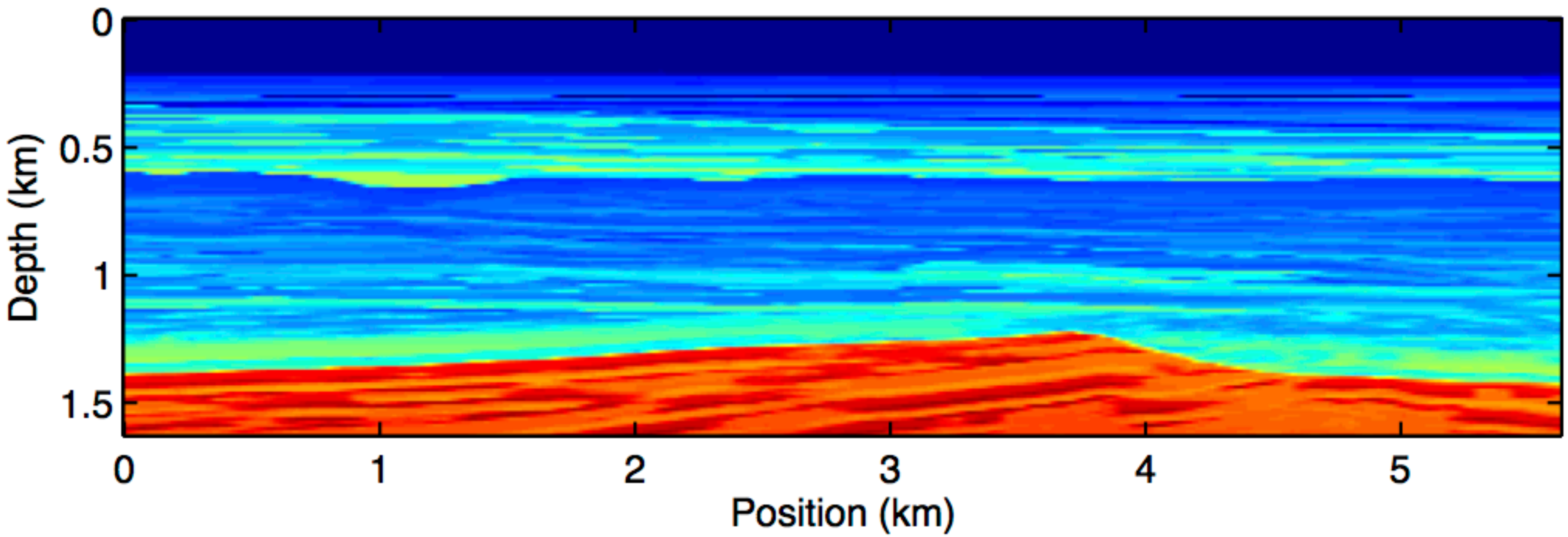
Fast SPLSM w/ multiples by **RISKA**



Simulation cost **~1 RTM** using all the data

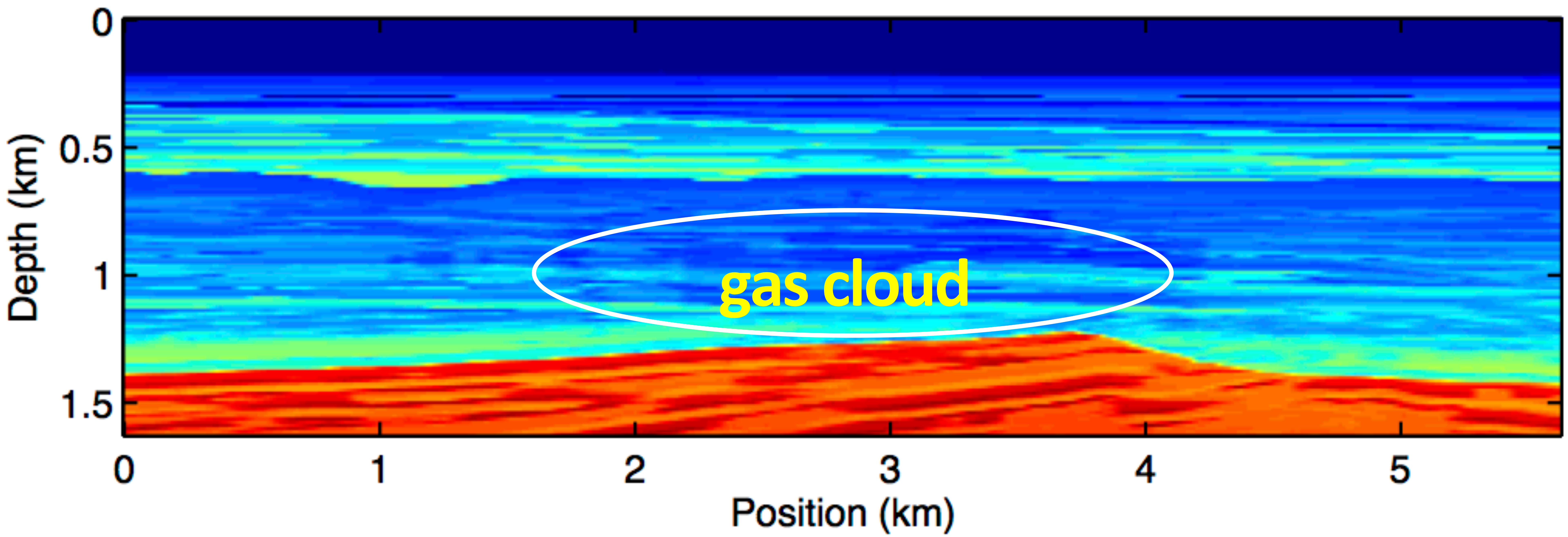
Time-lapse seismic

Baseline

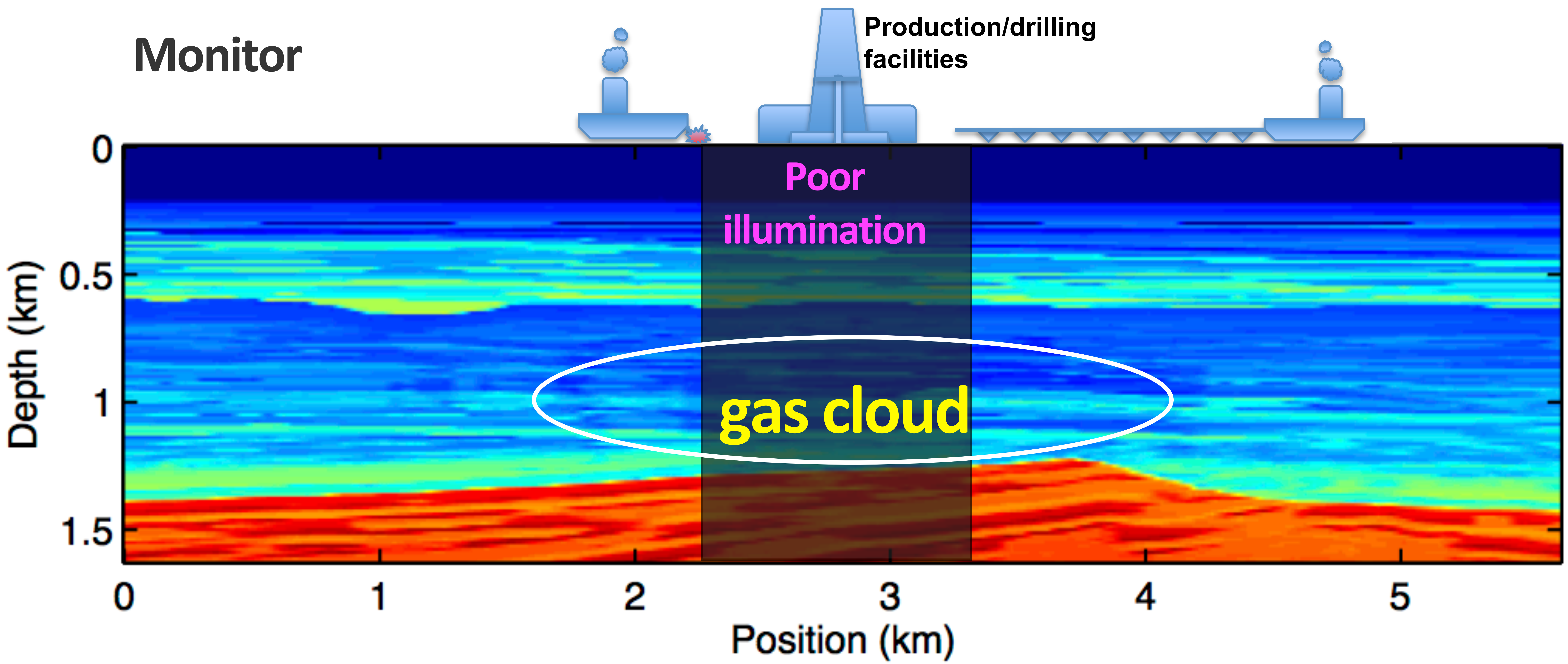


Time-lapse seismic

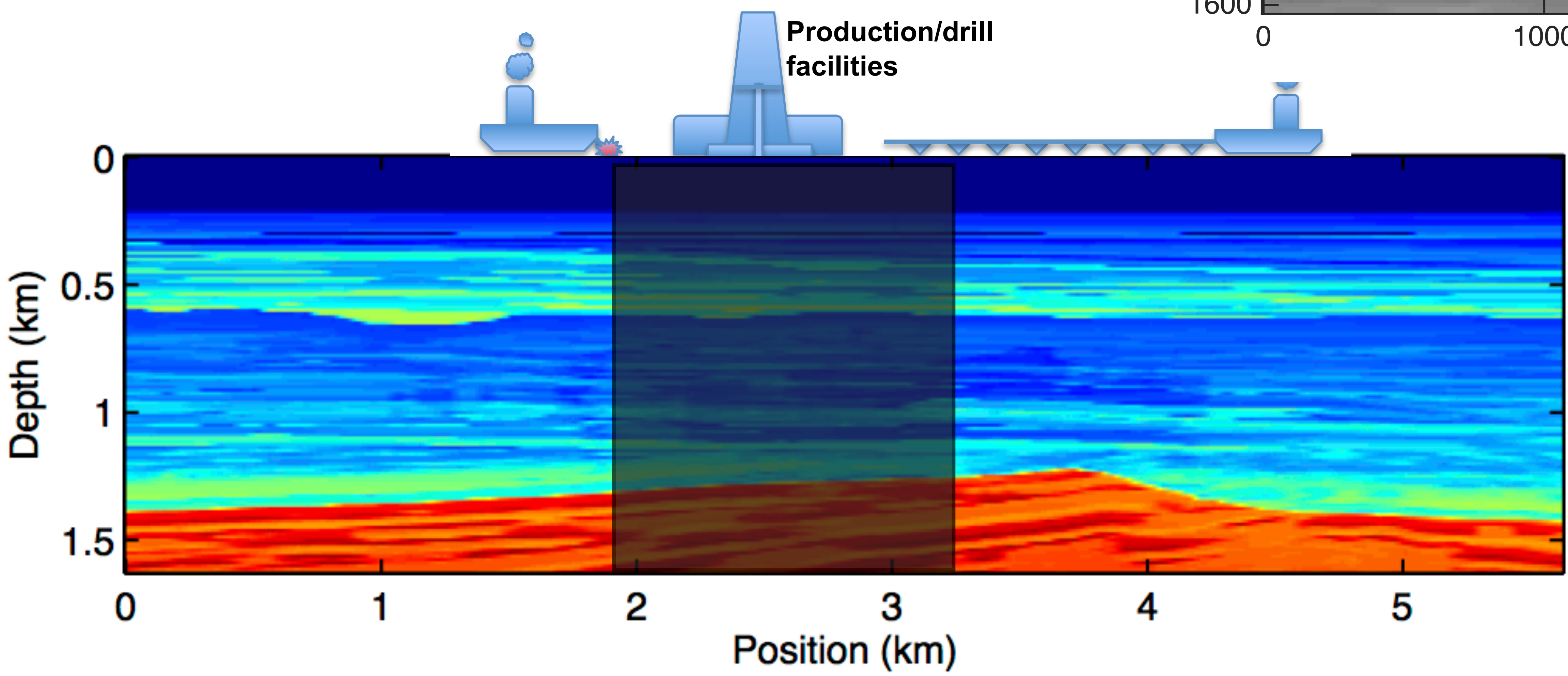
Monitor



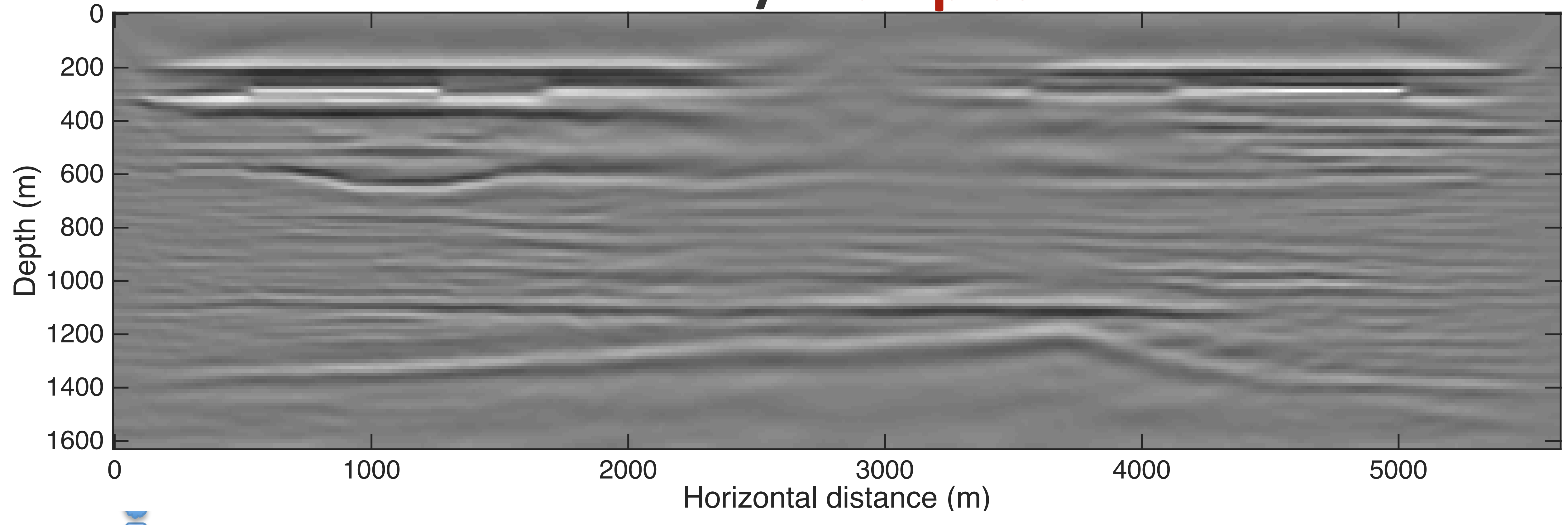
Time-lapse seismic



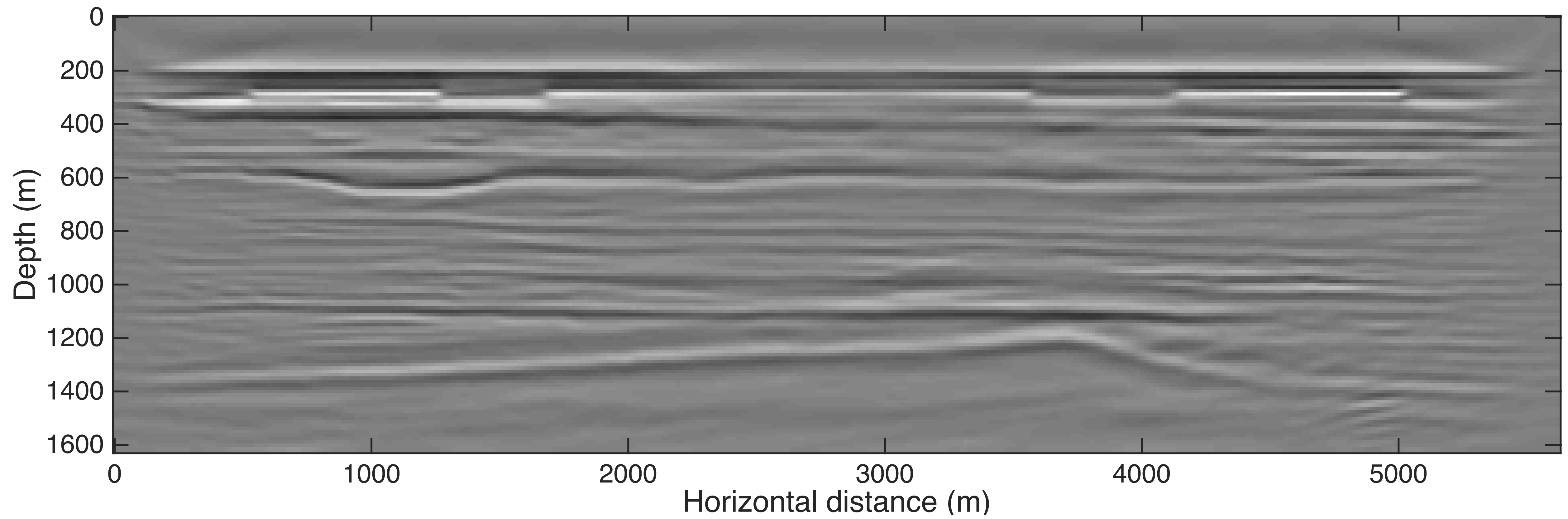
Large monitor acquisition gap



Sparsity-promoting migration w/ multiples



Joint sparsity-promoting migration w/ multiples



Time-stepping LSRTM with Sparsity Promotion

SLIM

University of British Columbia

Example w/o source estimation

Data:

- 320 sources (25 m spacing), 800 receivers (10 m spacing), OBN
- 4 s recording time
- 30 Hz peak frequency

LSRTM:

- 40 iterations /w 8 random shots per iteration
  (1 data pass, 640 PDE solves)

Sparsity promoting LSRTM

- Problem formulation

$$\text{minimize} \quad \lambda||\mathbf{C}\delta\mathbf{m}||_1 + \frac{1}{2}||\mathbf{C}\delta\mathbf{m}||_2^2$$
$$\text{subject to } ||\,\mathbf{J}\delta\mathbf{m} - \delta\mathbf{d}||_2 \leq \sigma$$

$\delta\mathbf{m}$: model perturbation/image
$\delta\mathbf{d}$: linearized data (single scattered data)
$\mathbf{J}$: linearized forward modeling operator (Jacobian)
$\mathbf{C}$: curvelet transform

Sparsity promoting LSRTM

- Problem formulation

$$\text{minimize} \quad \lambda||\mathbf{C}\delta\mathbf{m}||_1 + \frac{1}{2}||\mathbf{C}\delta\mathbf{m}||_2^2$$
$$\text{subject to } ||\,\mathbf{J}\delta\mathbf{m} - \delta\mathbf{d}||_2 \leq \sigma$$

- Preconditioning

$$\delta\mathbf{m} = \mathbf{M}_R^{-1}\mathbf{x}$$

$$\mathbf{M}_L^{-1}\mathbf{J}\mathbf{M}_R^{-1}\mathbf{x} = \mathbf{M}_L^{-1}\delta\mathbf{d}$$

SLIM

Preconditioning

- Left-hand preconditioning (data space)

$$\mathbf{M}_L^{-1} = \mathbf{T}_d \mathbf{F}$$

$\mathbf{T}_d$: Topmute

$\mathbf{F}$: Fractional integration $\partial_{|t|}^{-1/2}$

- Right-hand preconditioning (model space)

$$\mathbf{M}_R^{-1} = \mathbf{T}_m \mathbf{A}$$

$\mathbf{T}_m$: Topmute

$\mathbf{A}$: Depth scaling

Linearized Bregman

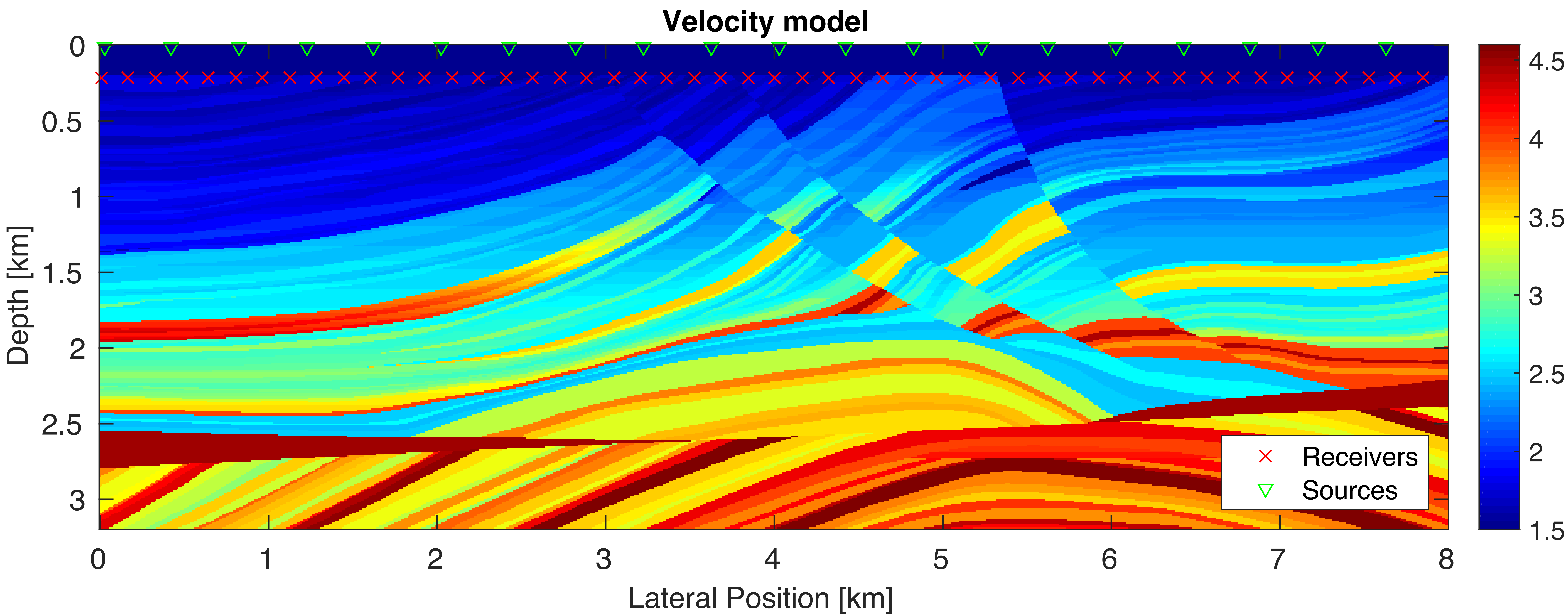
$$\begin{aligned}&\text{minimize} \quad \lambda||\mathbf{Cx}||_1 + \frac{1}{2}||\mathbf{Cx}||_2^2 \\ &\text{subject to } ||\underbrace{\mathbf{M}_L^{-1}\mathbf{J}\mathbf{M}_R^{-1}}_{\hat{\mathbf{J}}}\mathbf{x} - \underbrace{\mathbf{M}_L^{-1}\delta\mathbf{d}}_{\mathbf{b}}||_2 \leq \sigma\end{aligned}$$

Algorithm:

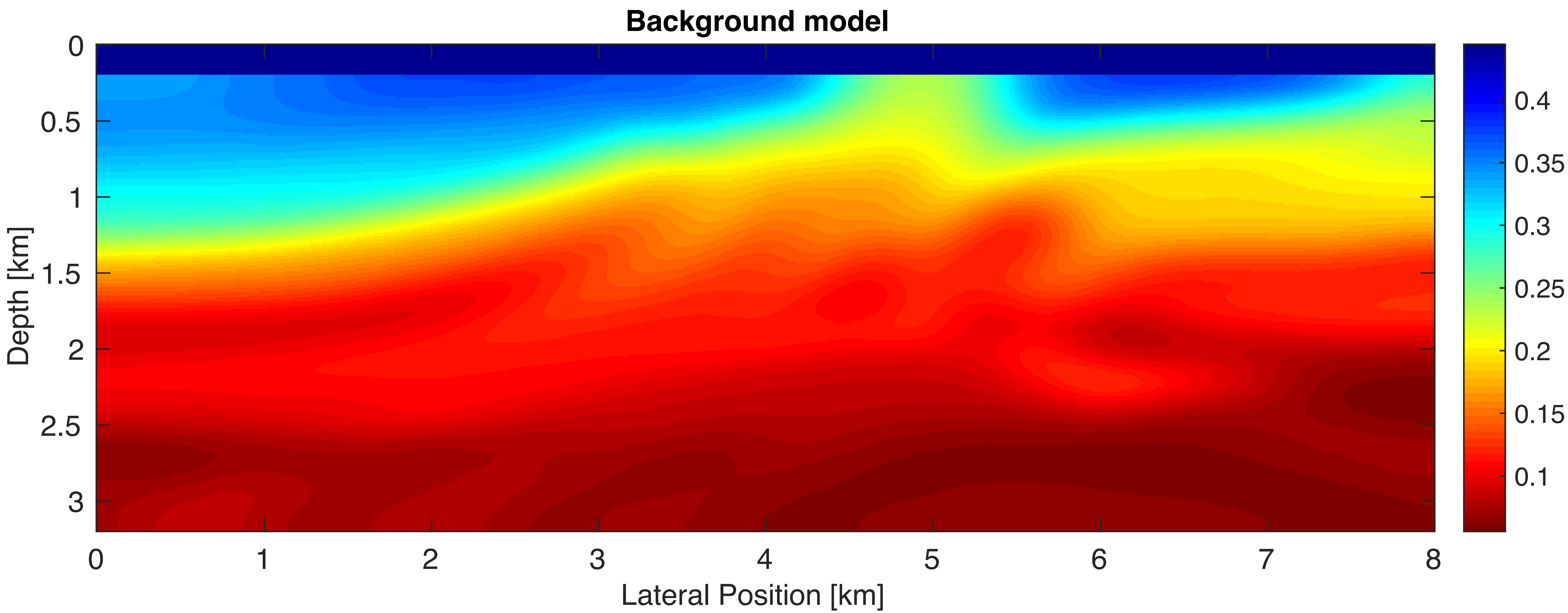
1. **for** $k = 0, 1, \cdots$
2. $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \hat{\mathbf{J}}^*_{r(k)}(\hat{\mathbf{J}}_{r(k)}\mathbf{x}_k - \mathbf{b}_{r(k)}) \cdot \max(0, 1 - \frac{\sigma}{||\hat{\mathbf{J}}_{r(k)}\mathbf{x}_k - \mathbf{b}_{r(k)}||_2})$
3. $\mathbf{x}_{k+1} = \mathbf{C}^* S_\lambda(\mathbf{C}\mathbf{z}_{k+1})$
4. **end for**

$$S_\lambda(\mathbf{x}) = \max(0, |\mathbf{x}| - \lambda) \cdot \text{sign}(\mathbf{x}) \qquad t_k = \frac{||\hat{\mathbf{J}}_{r(k)}\mathbf{x}_k - \mathbf{b}_{r(k)}||_2^2}{||\hat{\mathbf{J}}^*_{r(k)}(\hat{\mathbf{J}}_{r(k)}\mathbf{x}_k - \mathbf{b}_{r(k)})||_2^2}$$

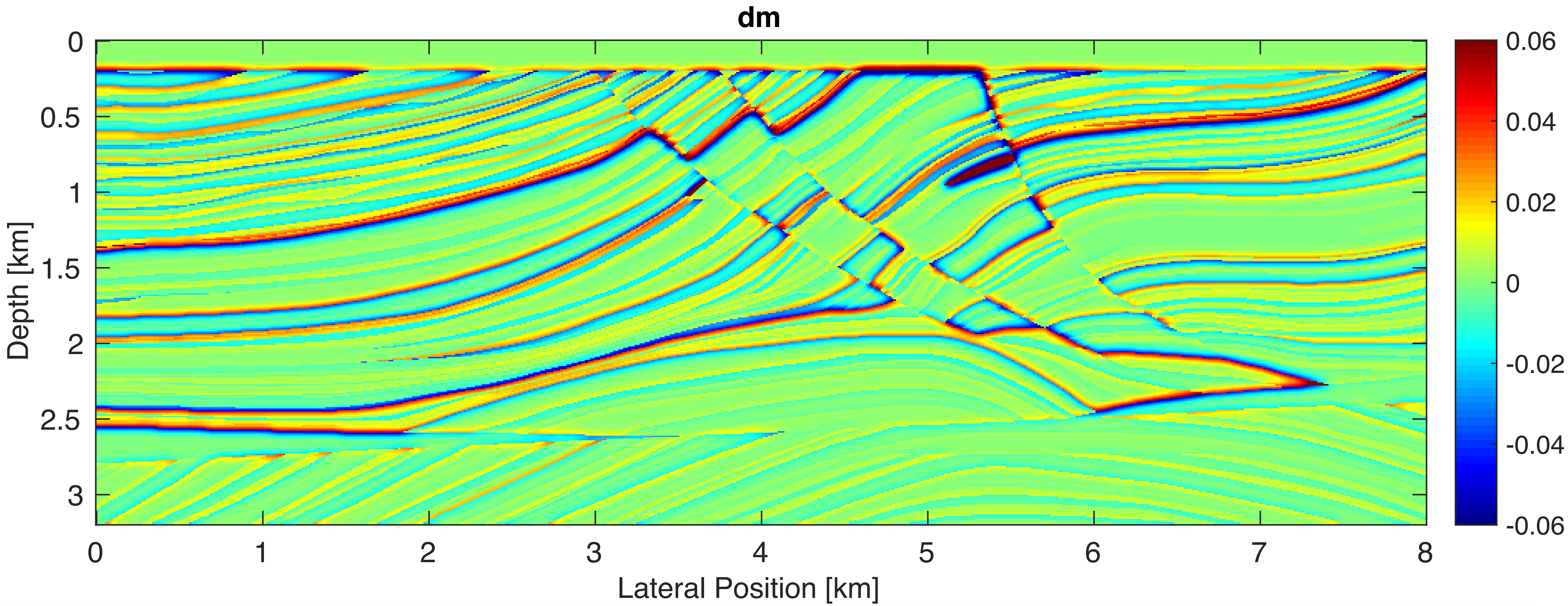
Velocity model & perturbation



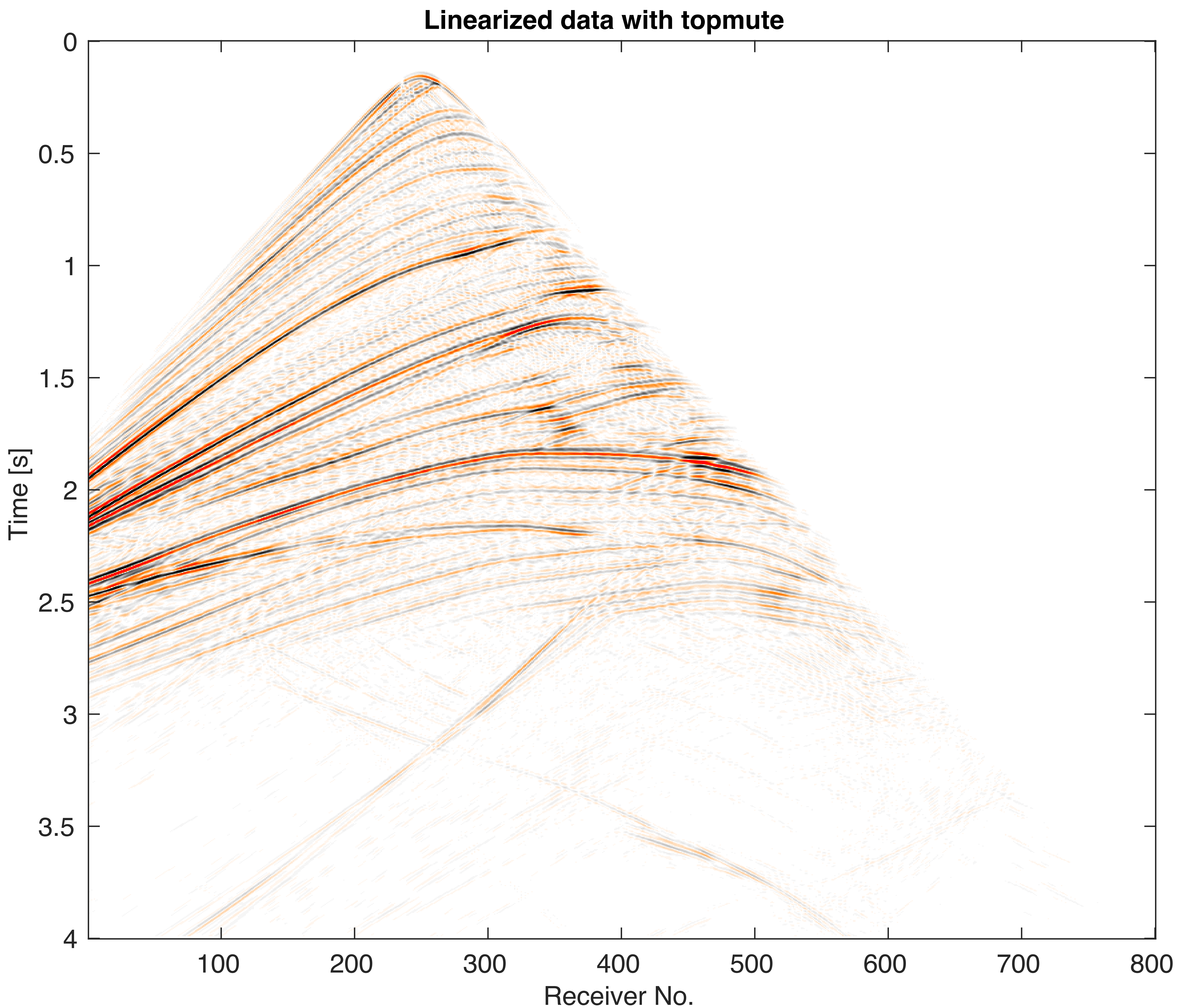
Velocity model & perturbation



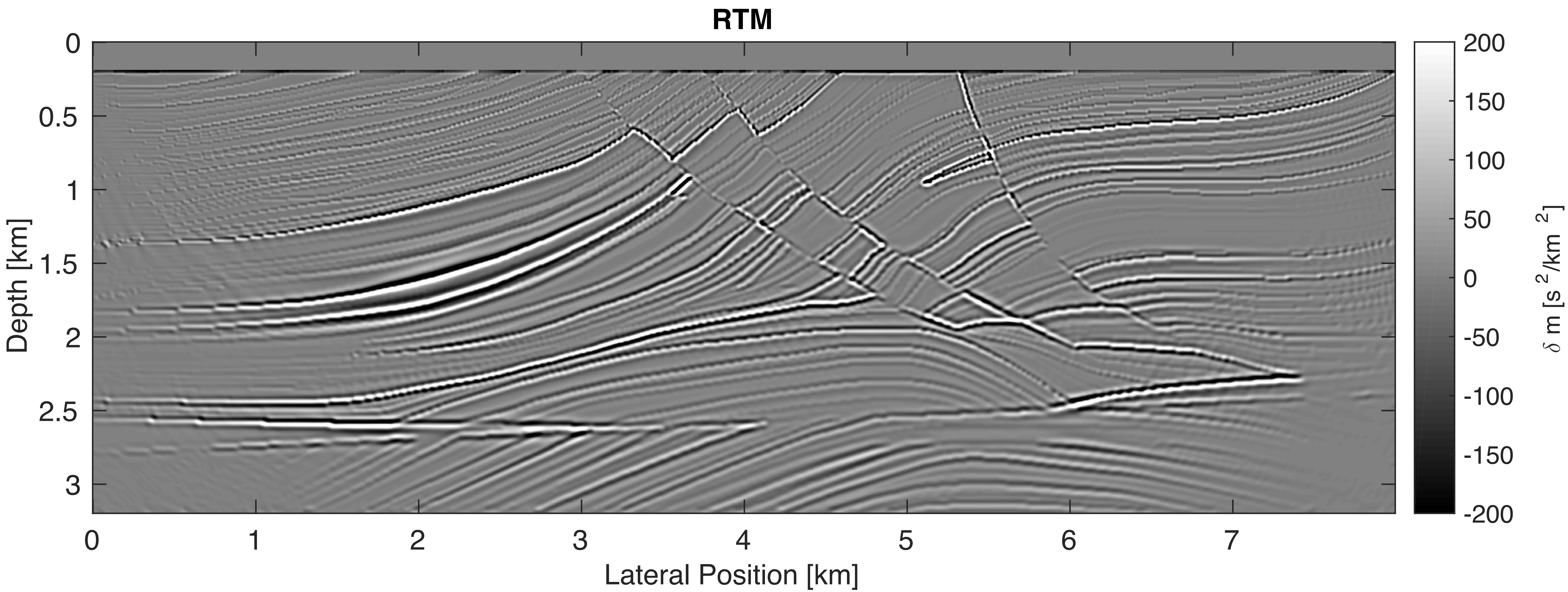
Velocity model & perturbation



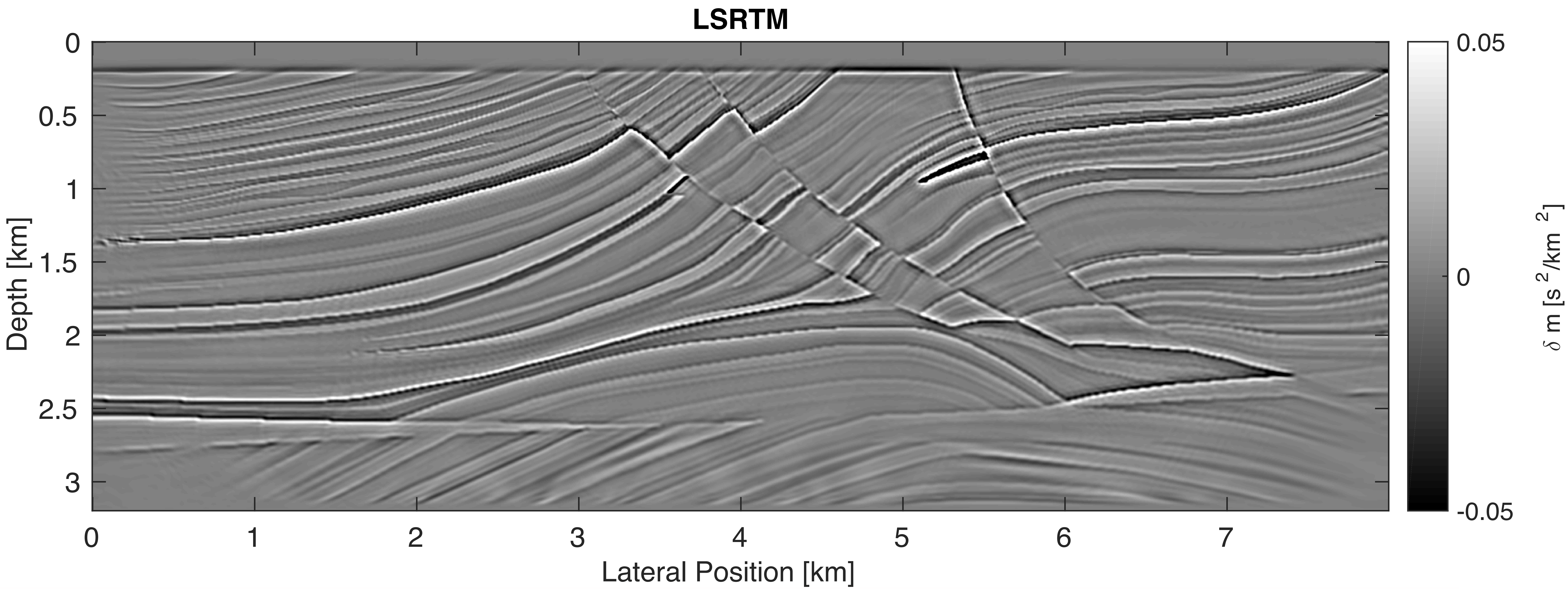
Data preprocessing



Results

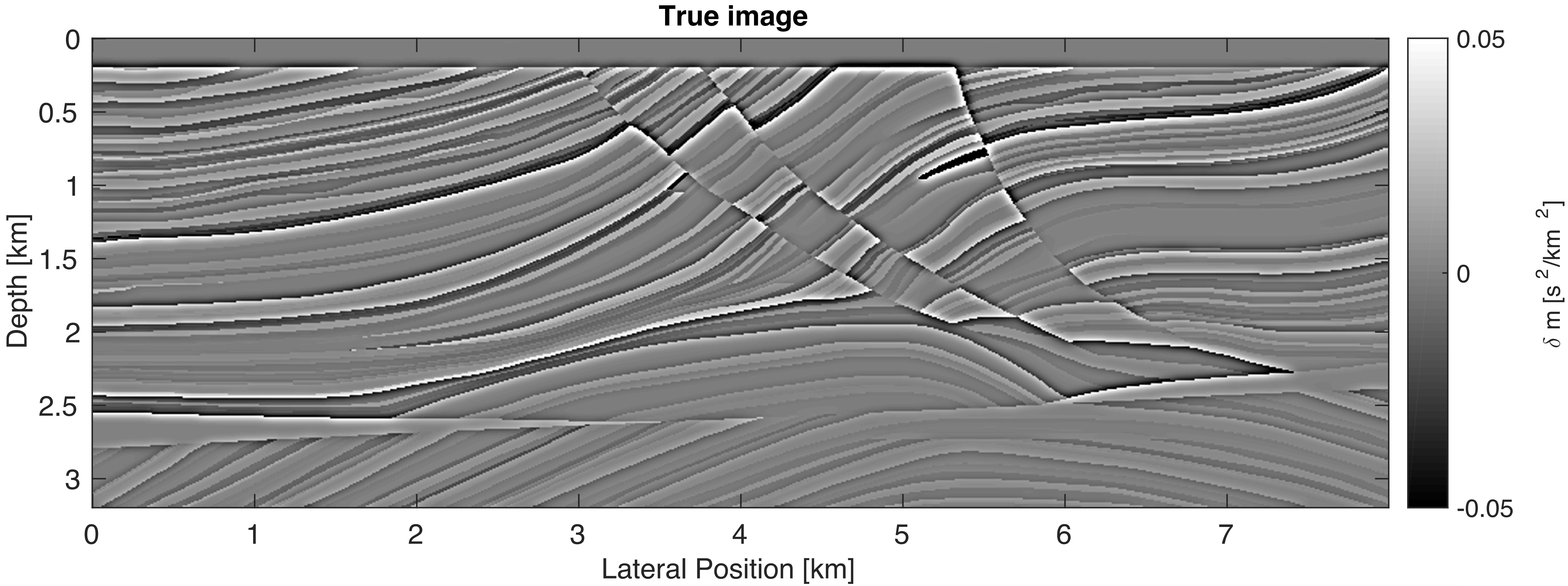

RTM
Depth [km]
0
0.5
1
1.5
2
2.5
3
0
1
2
3
4
5
6
7
Lateral Position [km]
200
150
100
50
0
-50
-100
-150
-200
δ m [s ²/km ²]

Results



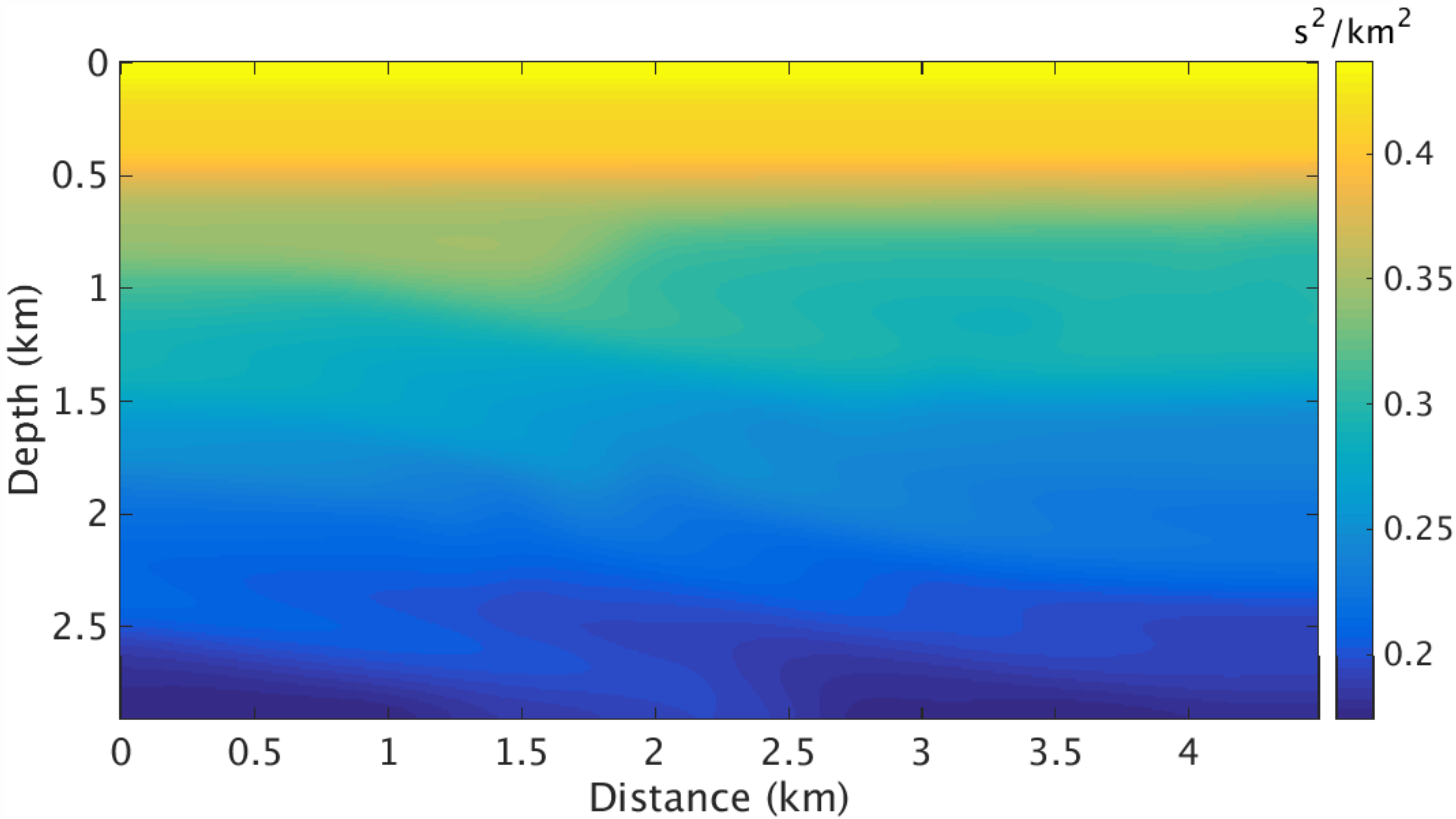
LSRTM
0
0.5
1
1.5
2
2.5
3
Depth [km]
0
1
2
3
4
5
6
7
Lateral Position [km]
0.05
0
-0.05
δ m [s ²/km ²]

Results

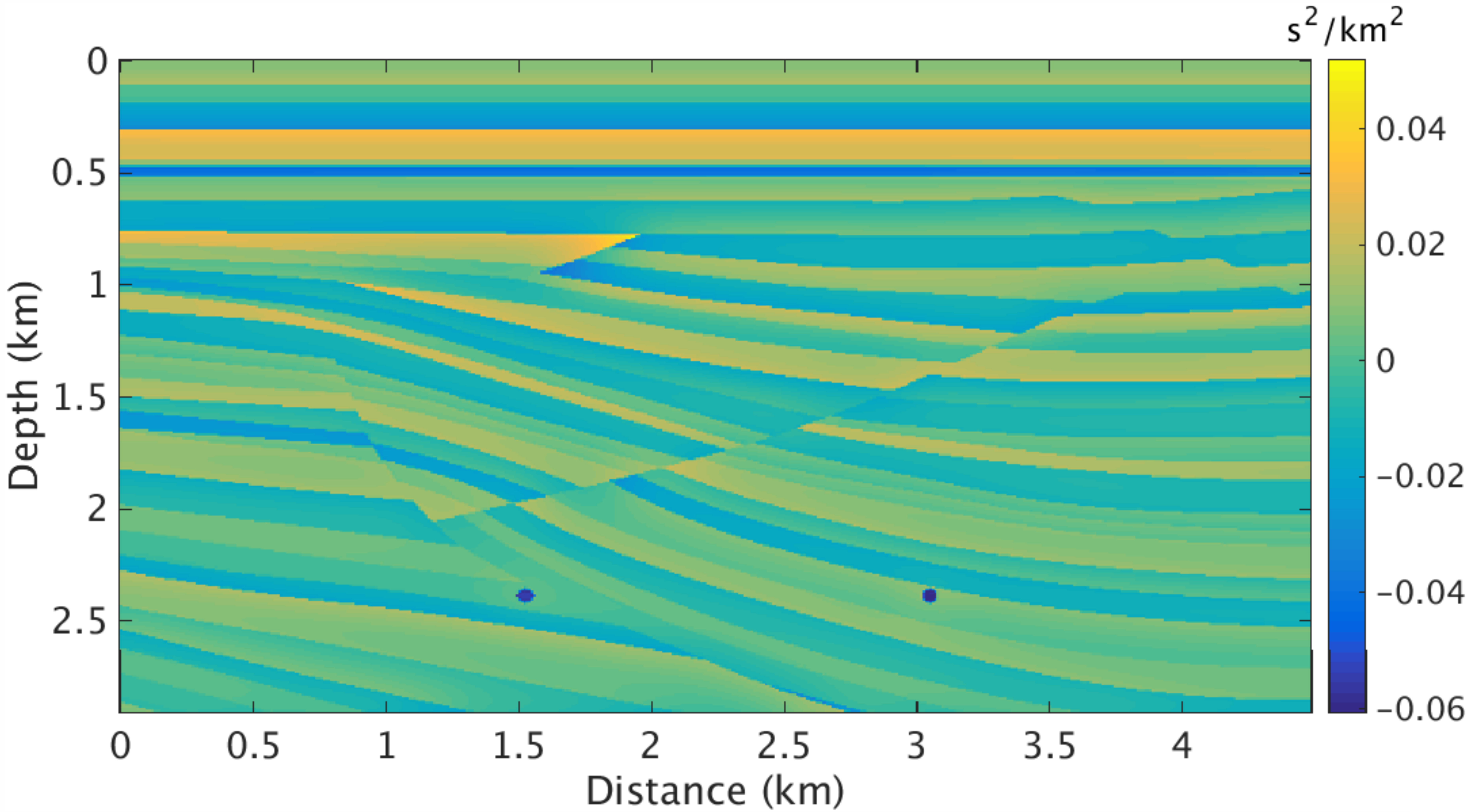


True image
0
0.5
1
1.5
2
2.5
3
Depth [km]
0
1
2
3
4
5
6
7
Lateral Position [km]
0.05
0
-0.05
δ m [s ²/km ²]

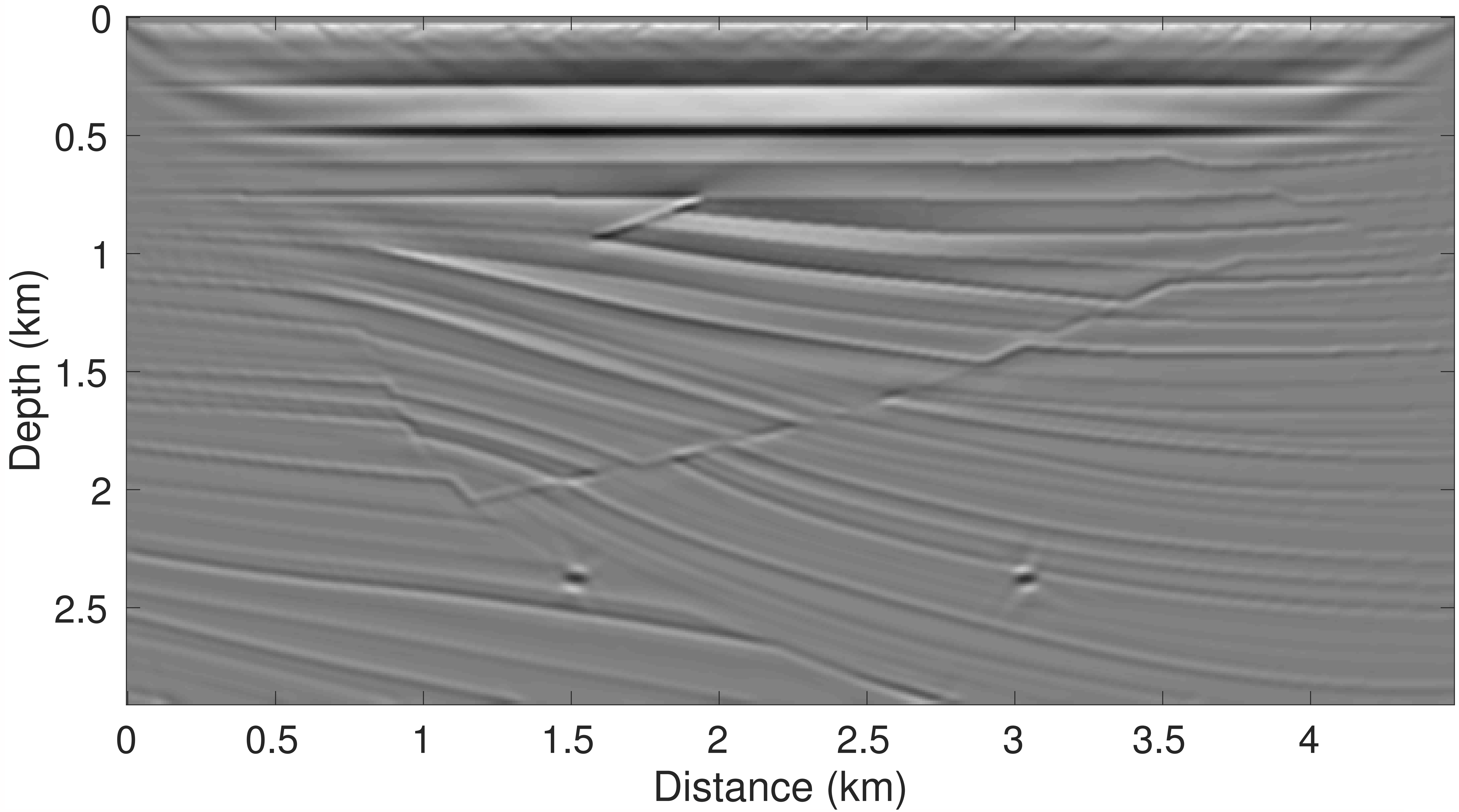
Source estimation – background model



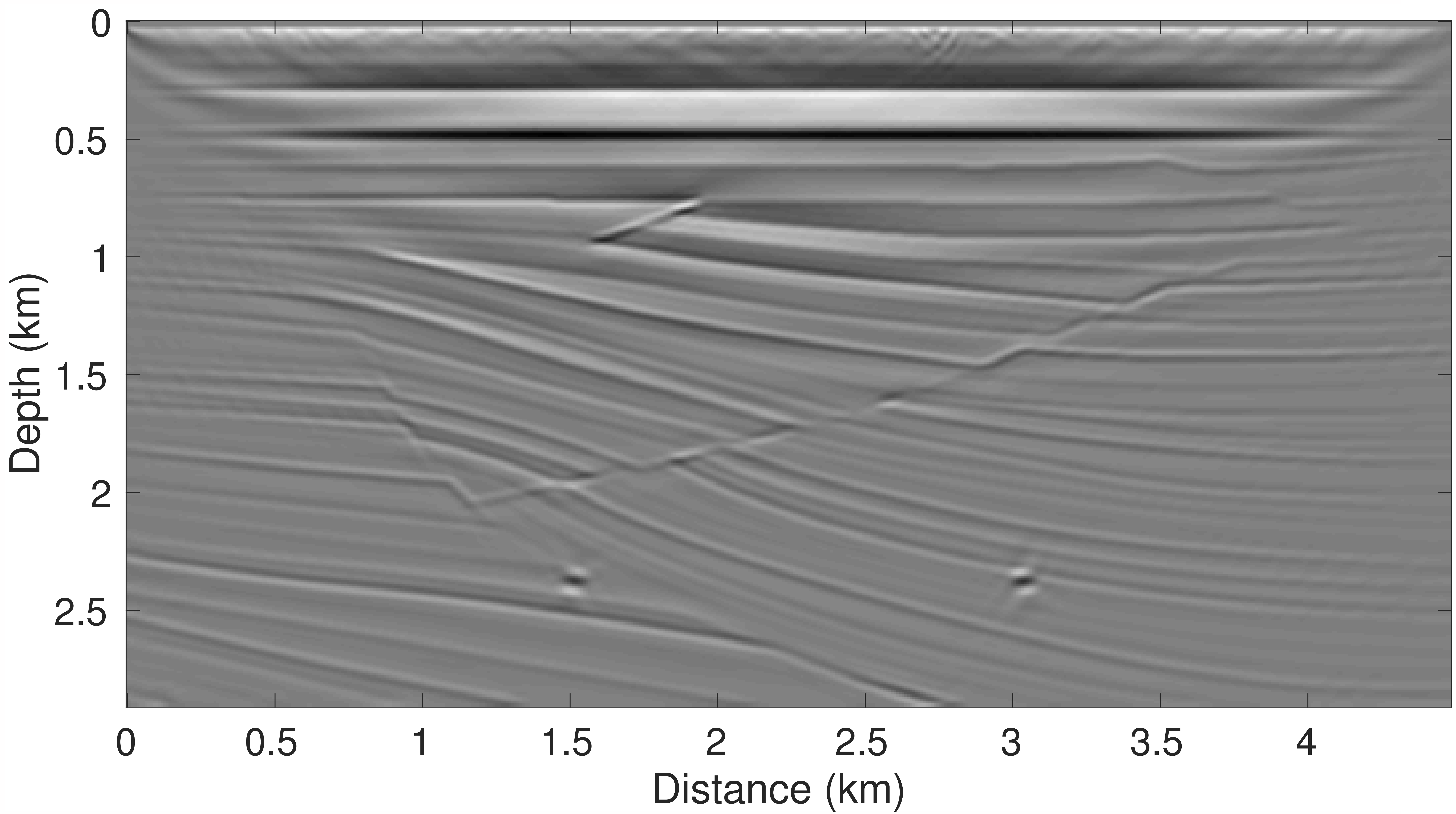
Source estimation – perturbation



Images – w/ known source



Images – w/ estimated source



Observations

Compressive imaging leads to

- a simple parallel algorithm w/ flexible degree of parallelism
- hifi artifact-free images from data w/ multiples

Randomizations lead to fast & computationally affordable RTM

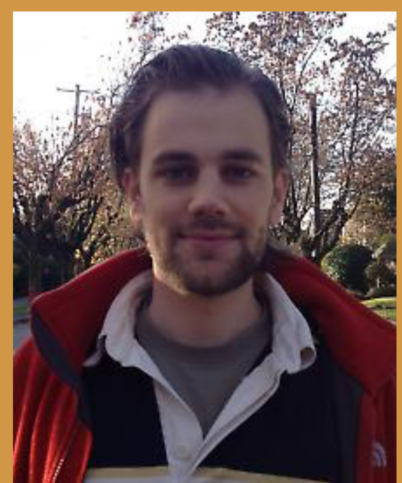
- touches data only once or twice / reduces # of PDEs solves & IO
- iterative system not yet fully analyzed
- issue w/ scaling ambiguity

But, requires

- densely sampled data
- good velocity models...

Constrained full-waveform inversion

Joint work w/ Ernie Esser, Bas Peters, Zhilong Fang, Tristan van Leeuwen, Mathias Louboutin


SLIM
University of British Columbia

John "Ernie" Esser (May 19, 1980 – March 8, 2015)

Stylized example

Simplistic forward model

$$\mathbf{d} = F(\mathbf{c})\mathbf{q} \equiv \mathbf{c} * \mathbf{q}$$

w/ vanilla inversion

$$\underset{\mathbf{c}\in\mathbb{R}^m}{\text{minimize}} \; \frac{1}{2}\|F(\mathbf{c})\mathbf{q} - \mathbf{d}\|_2^2$$

leads to nowhere if the source **q** misses low frequencies...

Stylized example w/ constraints

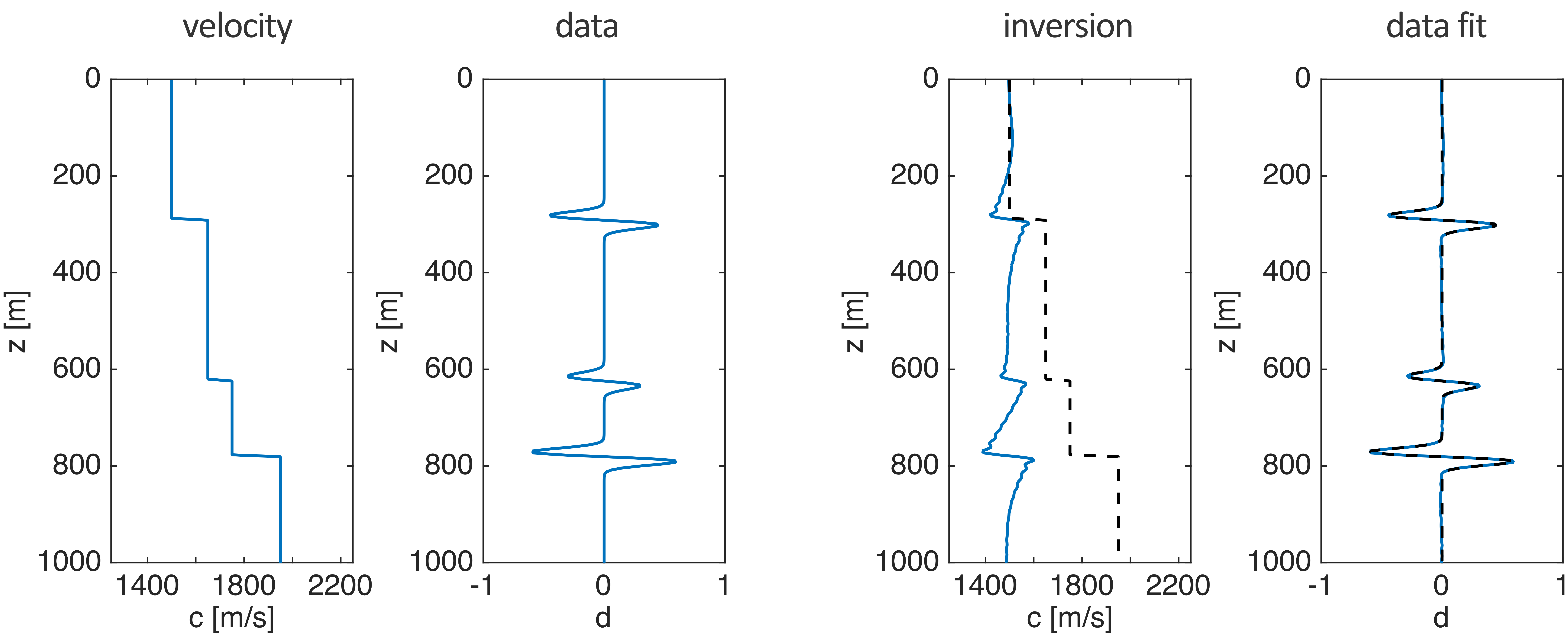
However, imposing constraints

$$\underset{\mathbf{c}\geq\mathbf{c}_0}{\text{minimize}}\ \frac{1}{2}\|F(\mathbf{c})\mathbf{q}-\mathbf{d}\|_2^2 \quad \text{subject to} \quad \mathbf{D}\mathbf{c}\geq\mathbf{0}$$

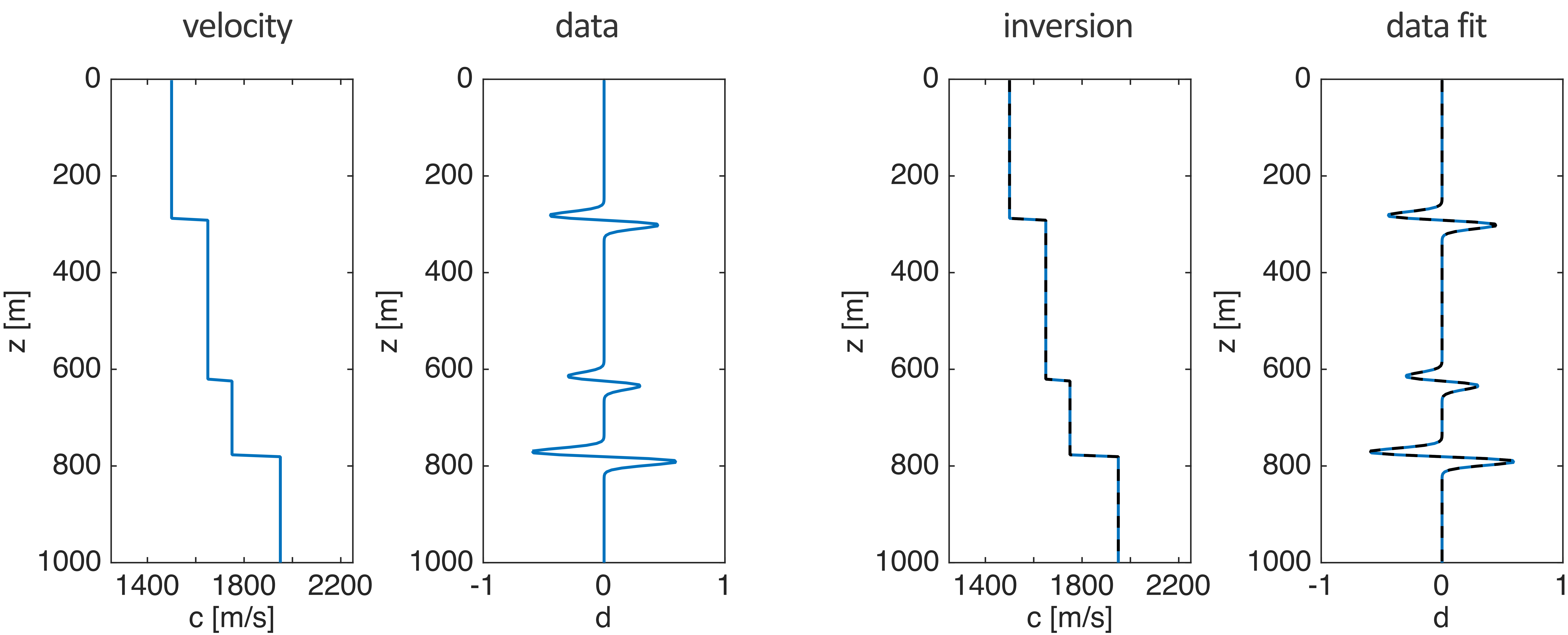
- minimal velocity
- monotonic increasing gradient of the velocity

on the model fully recovers the model...

Inversion w/o constraints



Inversion w/ constraints



Ernie Esser, Lluís Guasch, Tristan van Leeuwen, Aleksandr Y. Aravkin, and Felix J. Herrmann, "**Total variation regularization strategies in full waveform inversion for improving robustness to noise, limited data and poor initializations**". 2015

SLIM

Strategy

Extend the search space

- "less" nonlinear
- ensures data fit & avoids cycle skips

"Squeeze" the extension by

- enforcing the wave equation to compute model updates
- imposing *asymmetric* constraints that encode "rudimentary" properties of the geology
- relaxing the constraints to allow data fits & details to enter the solution

Leverage frequency continuation & warm starts where

- *sparsity-promoting asymmetric* constraints limit adverse affects of local minima
- there is hope as long as progress is made towards the solution during each cycle

Outcome: an automatic multi-cycle optimization-driven workflow

WRI – outer iterations

WRI method

for each source i

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

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

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

end

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

diagonal Hessian = pseudo Hessian

replace by inner loop that imposes convex constraints

Conventional method

for each source i

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

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

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

end

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

dense Hessian & too expensive

Including convex constraints into WRI

expensive but fixed

cheap

damped

$$\Delta \mathbf{m} = \arg\min_{\Delta \mathbf{m} \in R^N} \Delta \mathbf{m}^T \mathbf{g}^n + \frac{1}{2} \Delta \mathbf{m}^T H^n_{GN} \Delta \mathbf{m} + c_n \Delta \mathbf{m}^T \Delta \mathbf{m}$$

$$\text{such that} \quad \mathbf{m}^n + \Delta \mathbf{m} \in C$$

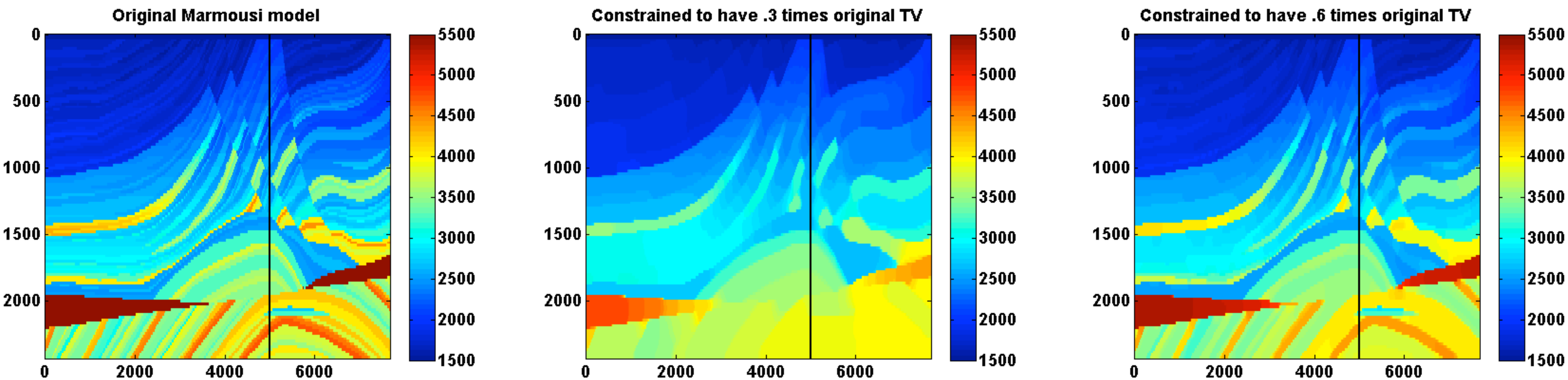
- guarantees $\mathbf{m}^{n+1} \in C$
- more difficult to compute
- feasible if it is easy to project onto
- naive projections $\mathbf{m}^{m+1} = \Pi_C \left(\mathbf{m}^n - (H^n)^{-1} \mathbf{g}^n \right)$ are not guaranteed to converge [Bertsekas ’99]

[Oldenburg ’83; Akçelik ’08; Anagaw ’11; Maharramov ’14; Esser & FJH ’14]

Projections onto convex sets

$v_{\min} = 1500,\ v_{\max} = 5500,\ \text{and}\ \tau = \{0.3\tau_0,\ 0.6\tau_0\}$

$$\Pi_C(\mathbf{m}_0) = \arg\min_{\mathbf{m}} \frac{1}{2}\|\mathbf{m} - \mathbf{m}_0\|^2 \quad \text{subject to} \quad \mathbf{m}_i \in [B_i^l, B_i^u] \text{ and } \|\mathbf{m}\|_{TV} \leq \tau$$



Proposed algorithm

Solve

$$\underset{\mathbf{m}}{\text{minimize}}\, \Phi(\mathbf{m}) \quad \text{subject to} \quad \mathbf{m}^{n+1} \in C_{\text{box}} \cap C_{\text{TV}}$$

by iterating

$$\Delta\mathbf{m} = \arg\min_{\Delta\mathbf{m}} \Delta\mathbf{m}^T \mathbf{g}^n + \frac{1}{2}\Delta\mathbf{m}^T (H^n + c_n \mathrm{I})\Delta\mathbf{m}$$

$$\text{subject to} \quad \mathbf{m}_i^n + \Delta\mathbf{m}_i \in [B_i^l, B_i^u] \text{ and } \|\mathbf{m}^n \Delta\mathbf{m}\|_{TV} \leq \tau$$

$$\mathbf{m}^{n+1} = \mathbf{m}^n + \Delta\mathbf{m}$$

Solving the convex subproblems

Find saddle point of

$$\mathcal{L}(\Delta\mathbf{m},\mathbf{p}) = \Delta\mathbf{m}^T\mathbf{g}^n + \frac{1}{2}\Delta\mathbf{m}^T(H^n + c_n\mathrm{I})\Delta\mathbf{m} + g_B(\mathbf{m}^n + \Delta\mathbf{m}) + \mathbf{p}^T D(\mathbf{m}^n + \Delta\mathbf{m}) - \tau\|\mathbf{p}\|_{\infty,2}$$

with indicator functions for

Bound constraint

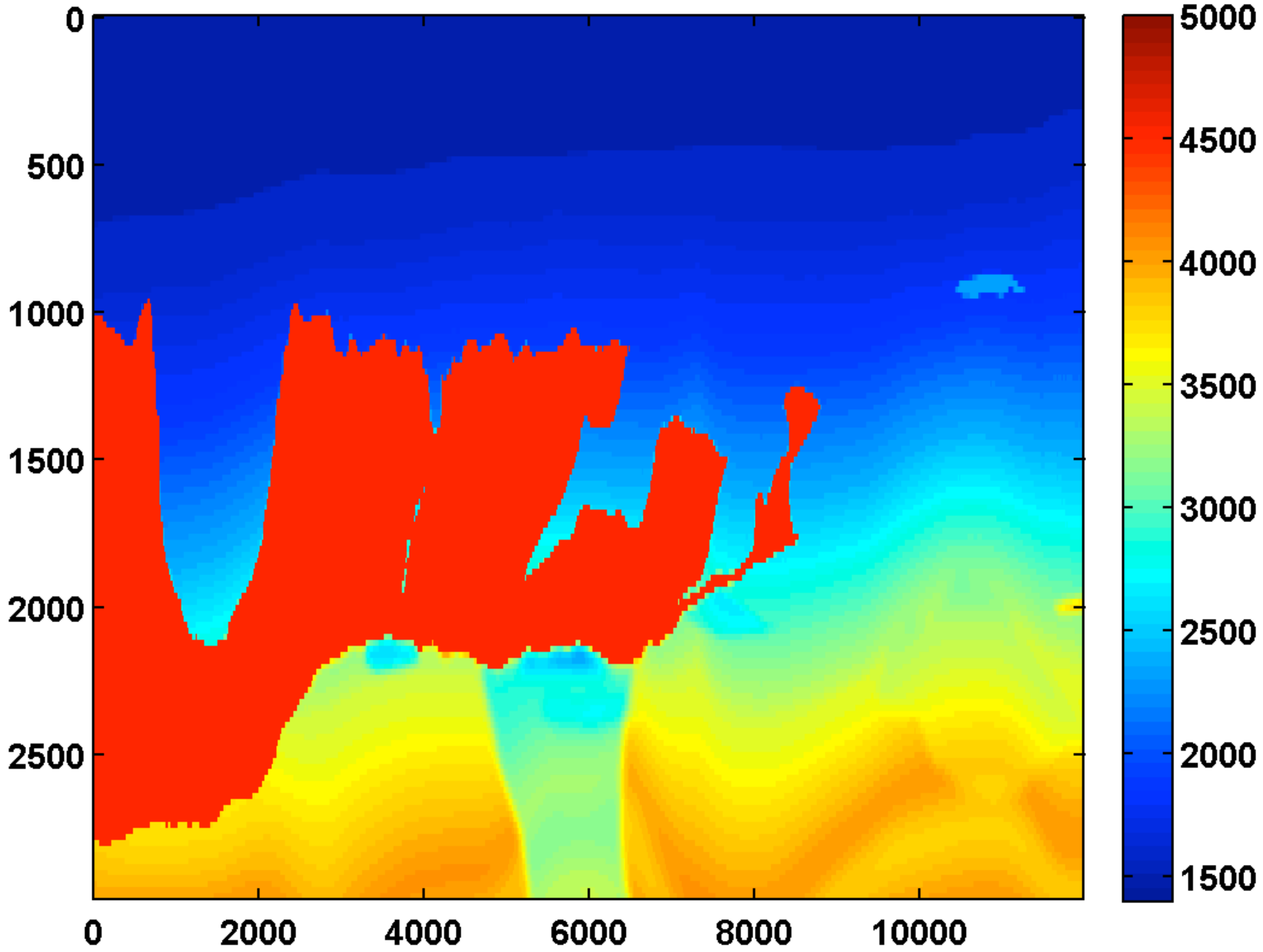
$$g_B(\mathbf{m}) = \begin{cases} 0 & \text{if } m_i \in [B_i^l, B_i^u] \\ \infty & \text{otherwise} \end{cases}$$

TV-norm constraint

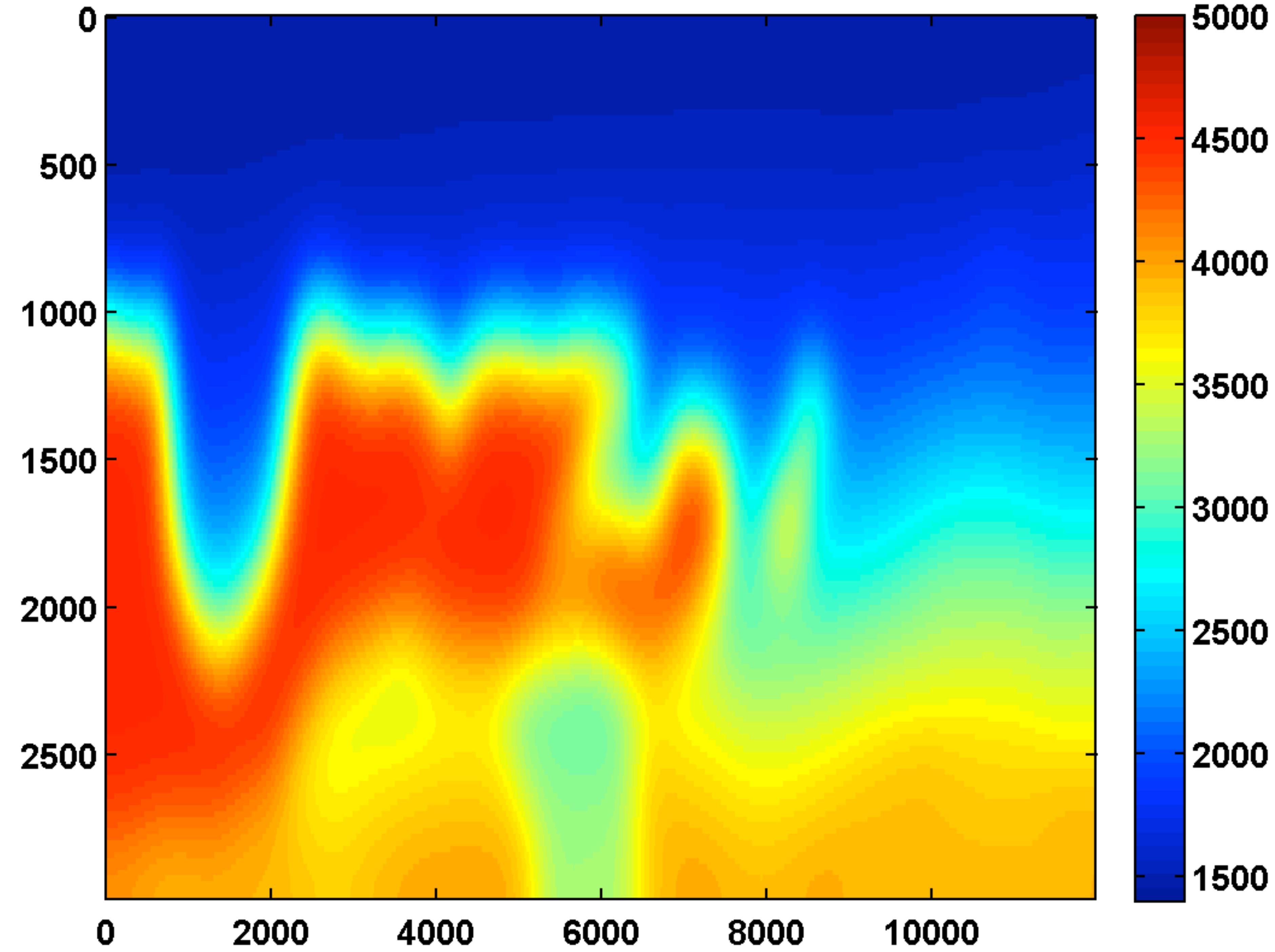
$$\sup_{\mathbf{p}} +\mathbf{p}^T D(\mathbf{m}^n + \Delta\mathbf{m}) - \tau\|\mathbf{p}\|_{\infty,2} = \begin{cases} 0 & \text{if } \|D(\mathbf{m}^n + \Delta\mathbf{m})\|_{1,2} \le \tau \\ \infty & \text{otherwise} \end{cases}$$

BP model

original model

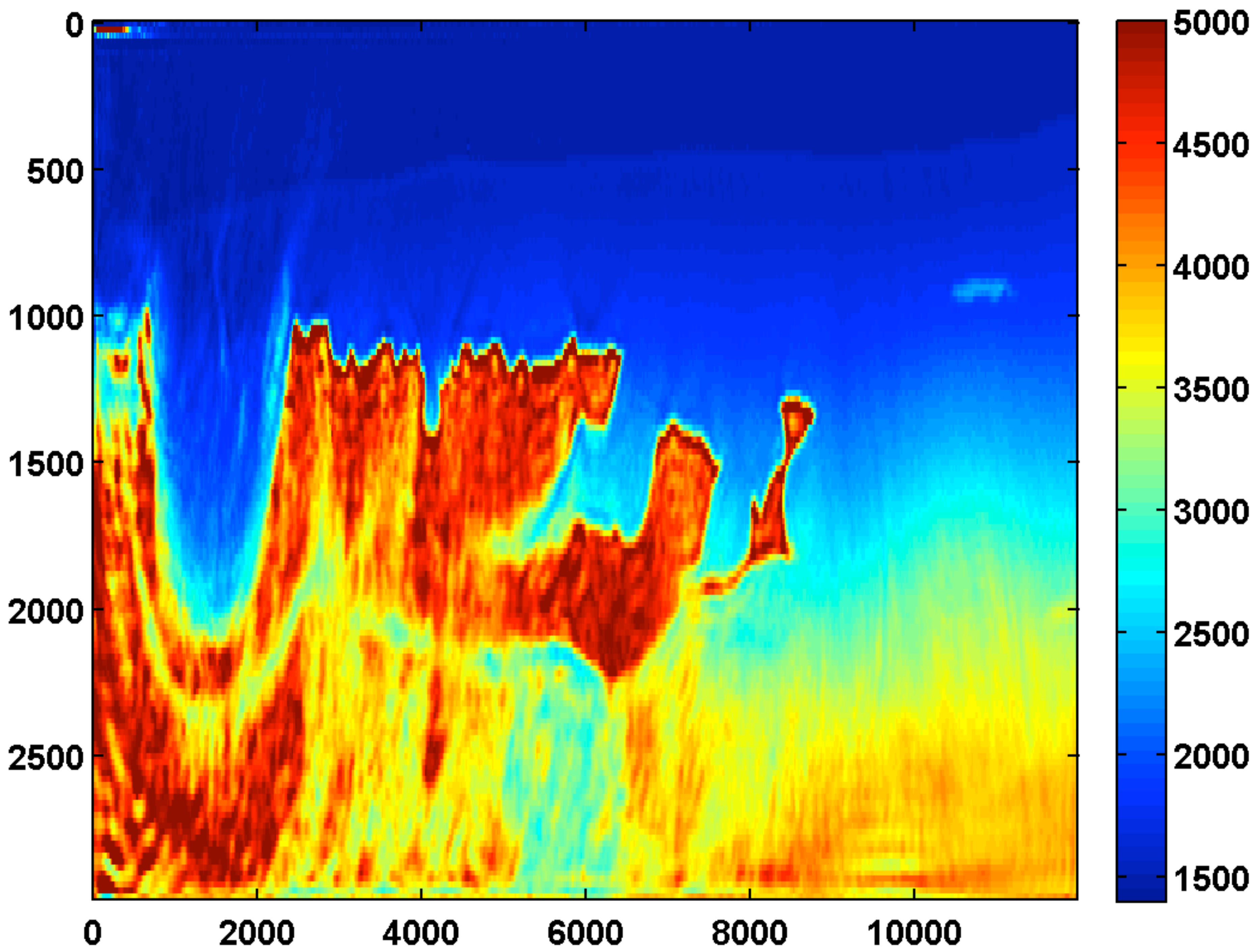


starting model

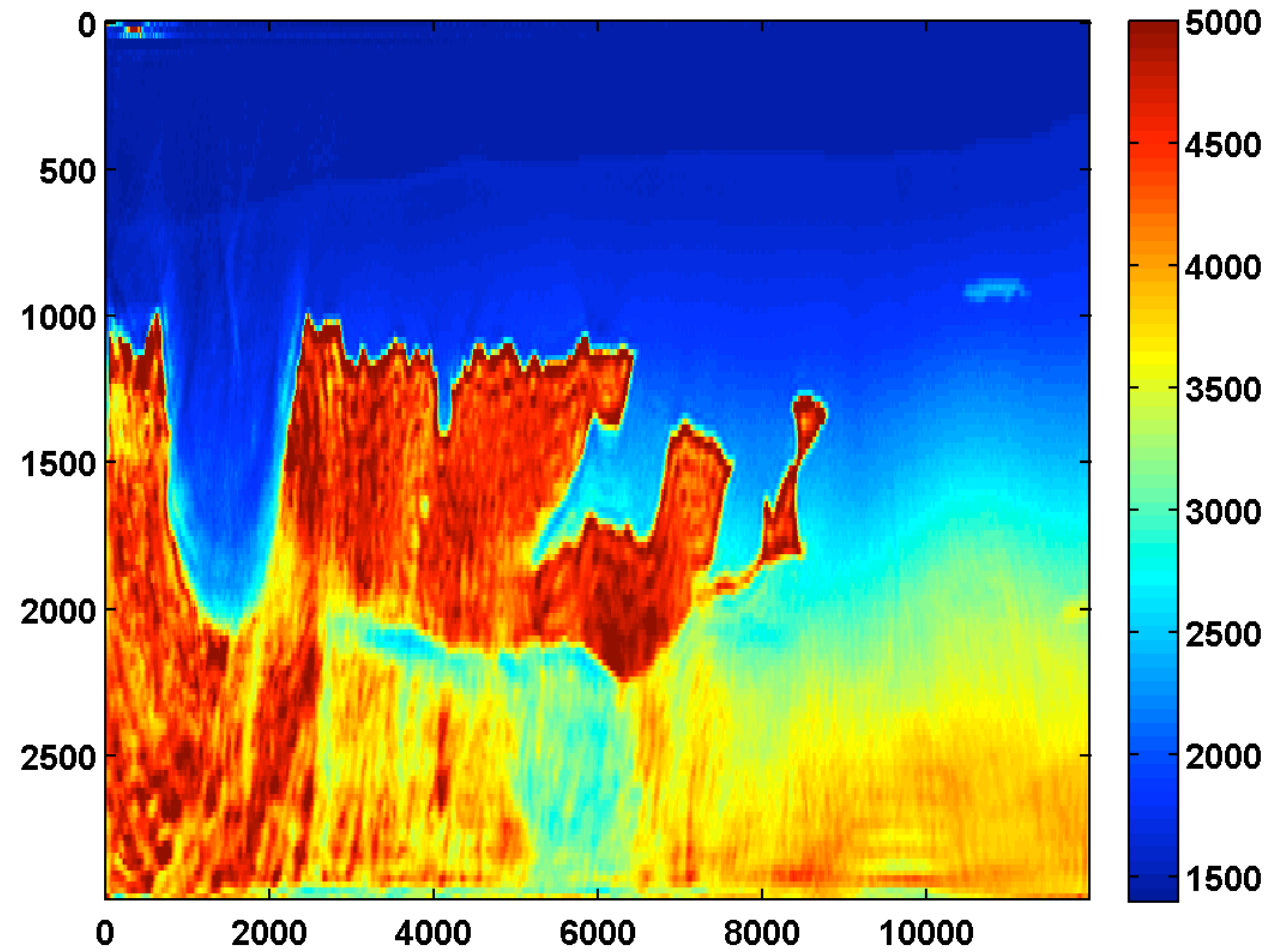


WRI w/o contraints

first sweep

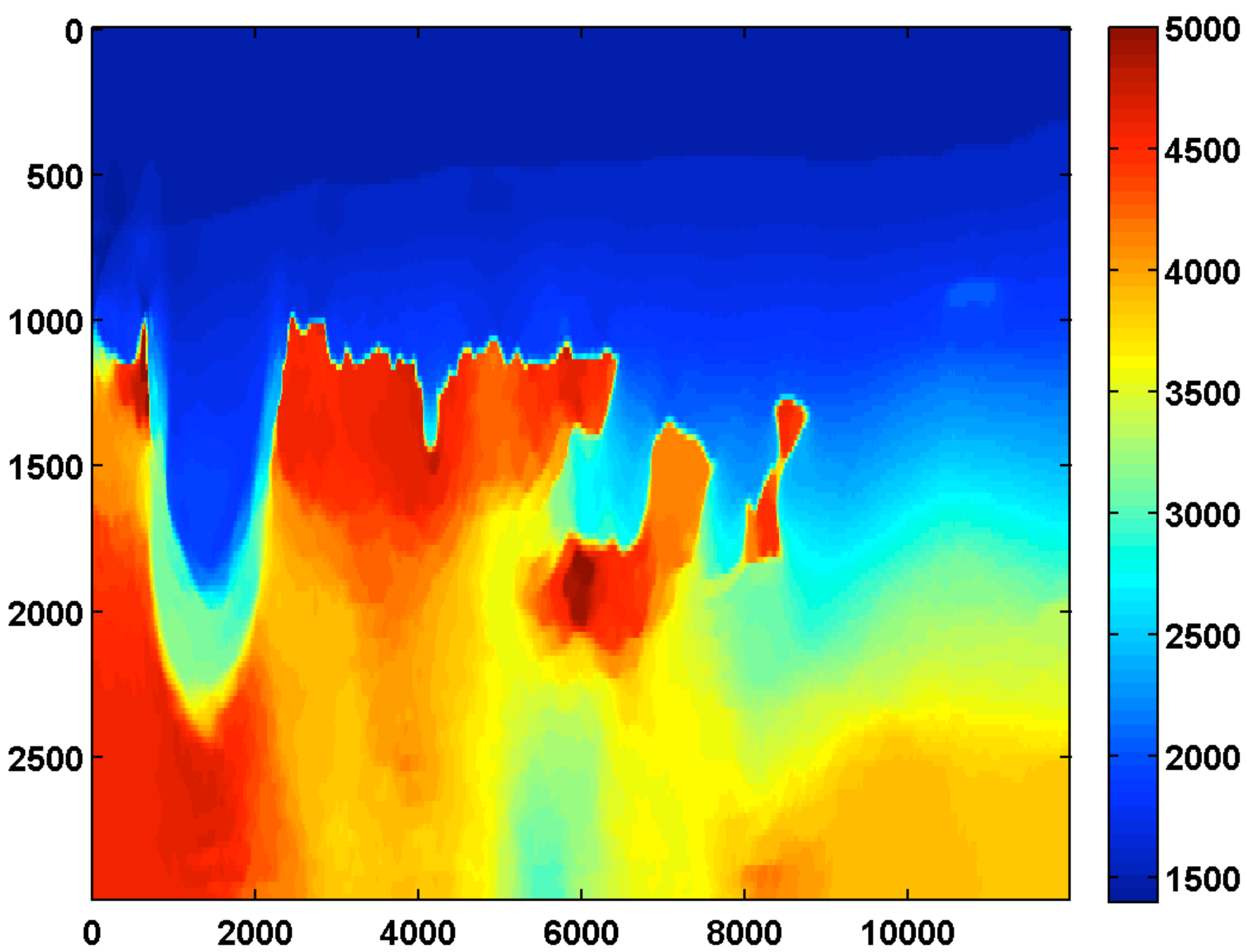


second sweep

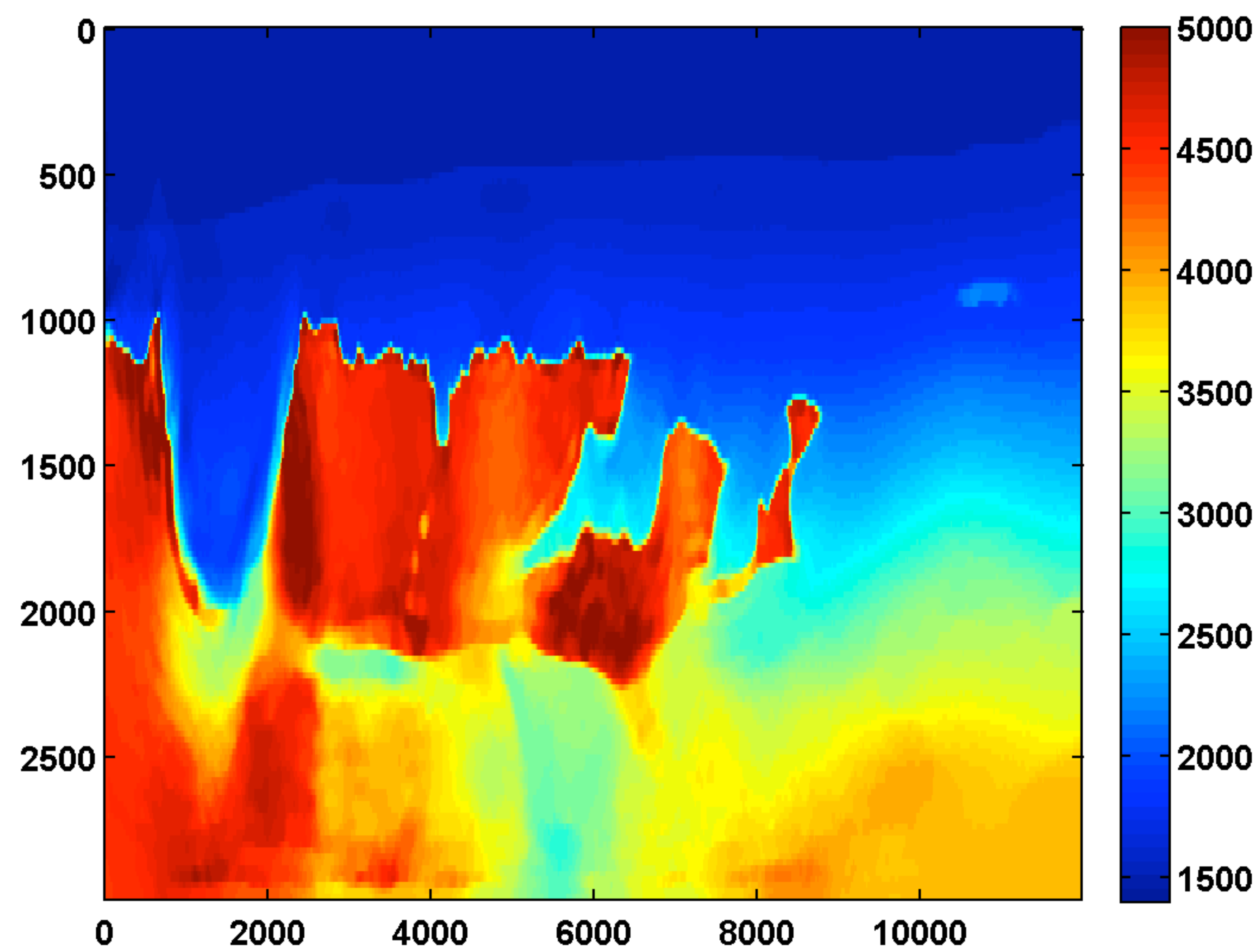


WRI w/ relaxed constraints

first sweep

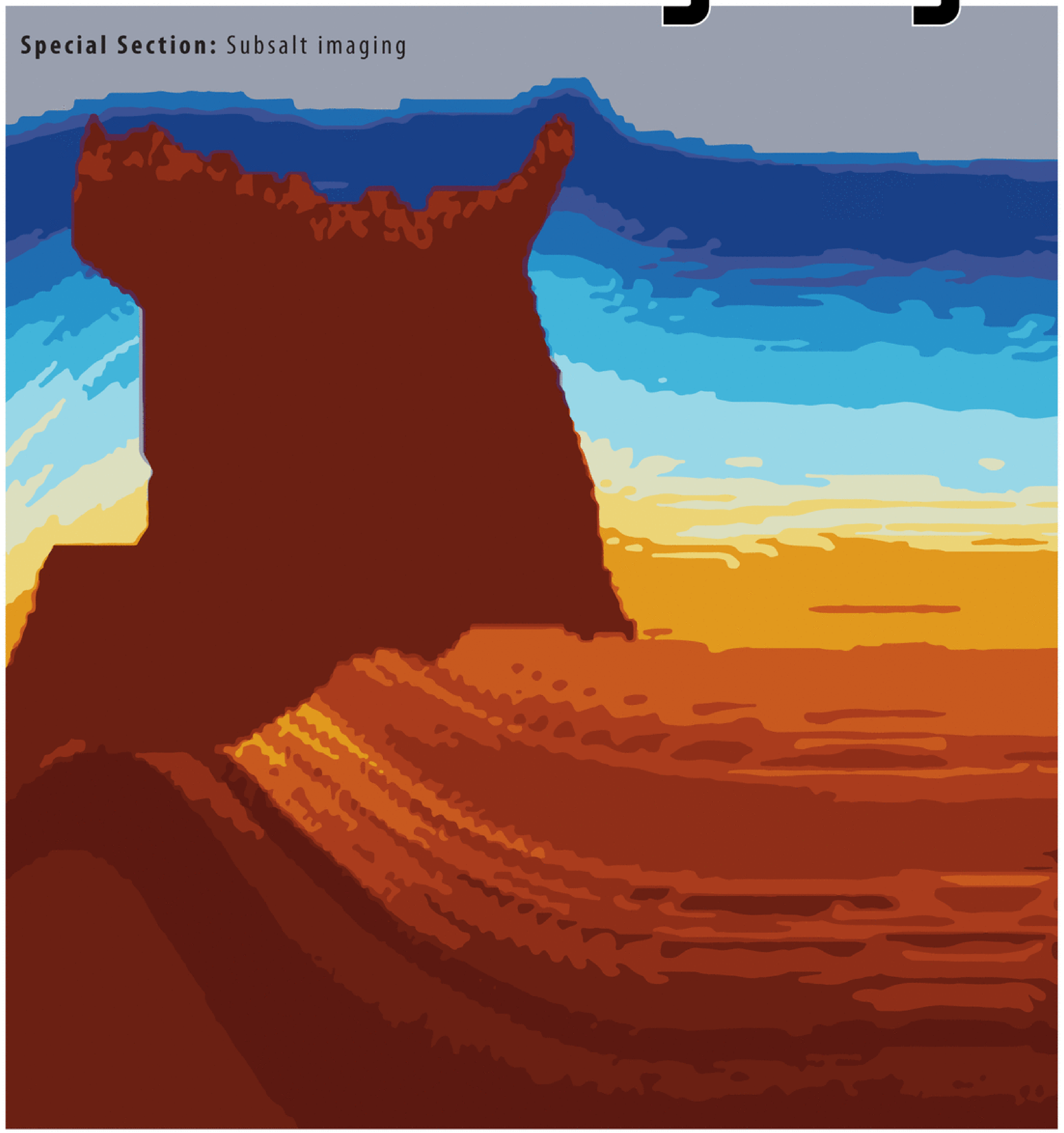


second sweep



Total Variation Regularized Wavefield Reconstruction Inversion [Read More] [GitHub]

The Leading Edge®



March 2016 Volume 35, No. 3

Society of Exploration Geophysicists
The international society of applied geophysics
ISSN 1070-485X

Design criteria – including constraints

Flexible optimization framework that

- incorporates rudimentary properties of the geology via constraints
- uniquely imposes multiple constraints on each FWI model iterate
- is fast & leaves the main loop w/ wave-equation solves alone
- works as a “black box” w/ existing FWI code bases

A few regularization strategies

Objective function: $f(\mathbf{m})$ (differentiable, time or frequency)

Tikhonov penalty: $\phi(\mathbf{m}) = f(\mathbf{m}) + \frac{\alpha}{2}\|R_1\mathbf{m}\|^2 + \frac{\beta}{2}\|R_2\mathbf{m}\|^2$

Gradient filtering: $\mathbf{m}_{k+1} = \mathbf{m}_k - \gamma F \nabla_{\mathbf{m}} f(\mathbf{m})$

Constrained formulation: $\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$

A few regularization strategies

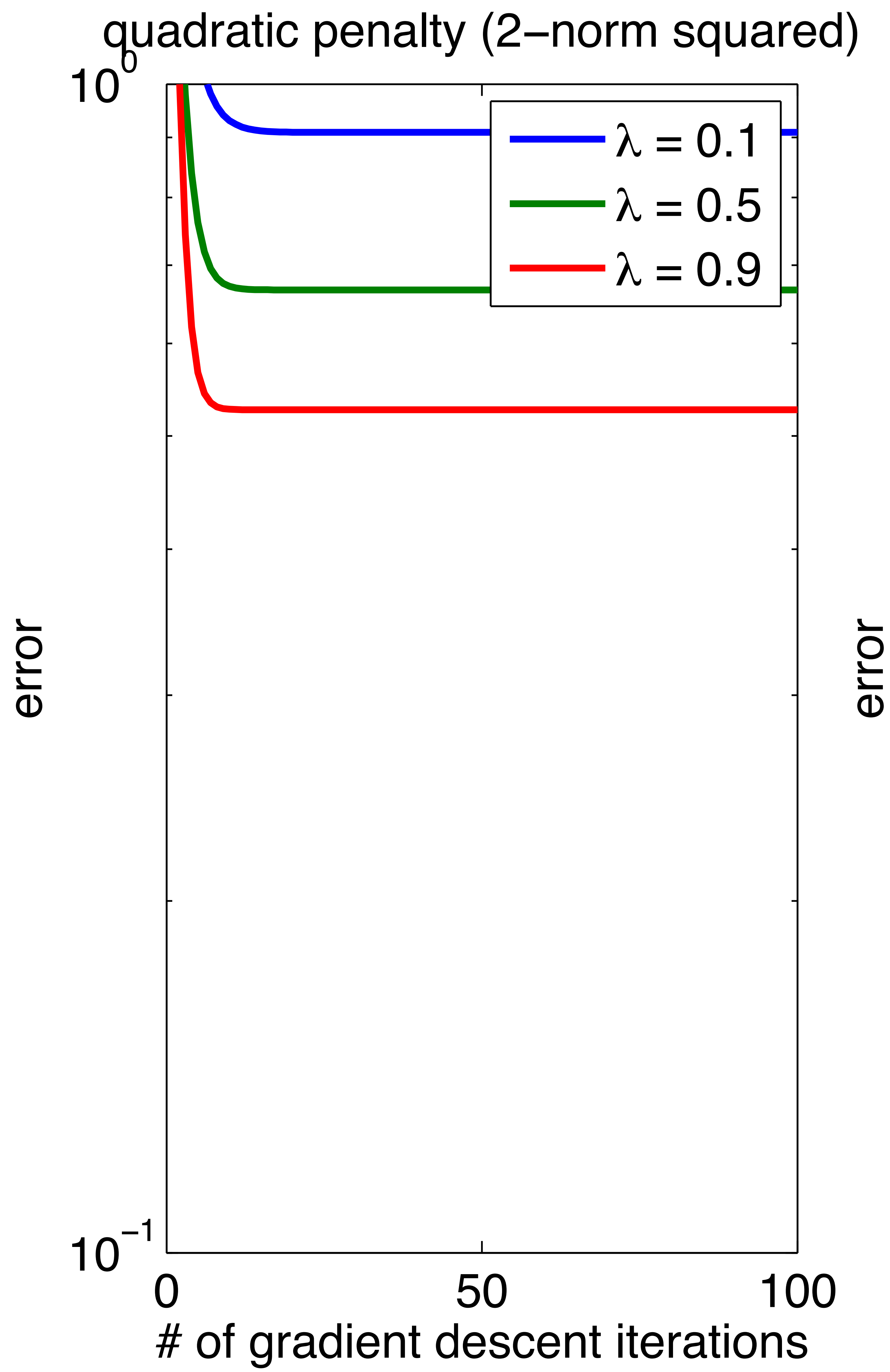
Tikhonov:

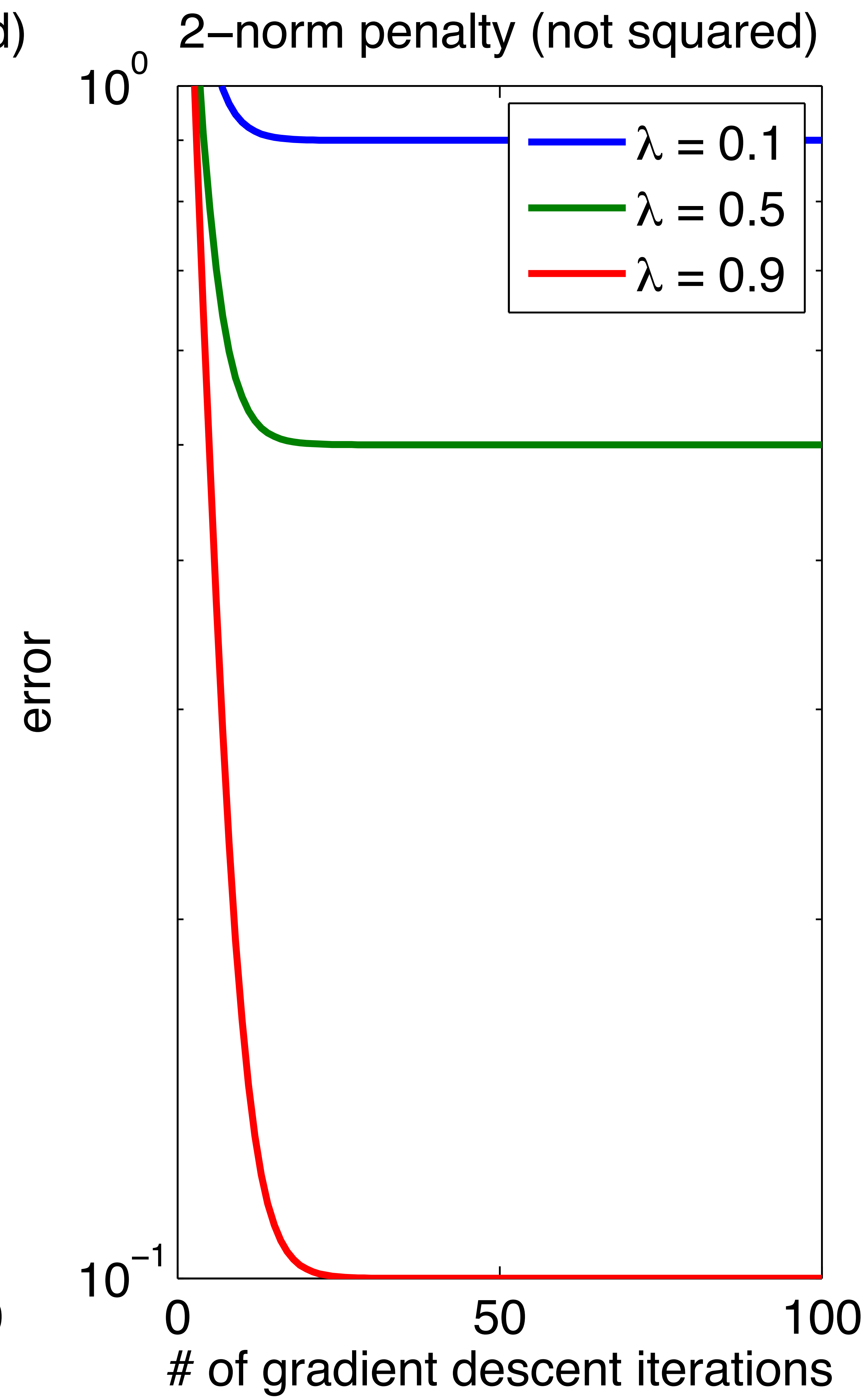
$$\phi(\mathbf{m}) = f(\mathbf{m}) + \frac{\alpha}{2}\|R_1\mathbf{m}\|^2 + \frac{\beta}{2}\|R_2\mathbf{m}\|^2$$

Potential problems:

- squared norm is not an exact penalty
- difficult/costly to determine penalty/tradeoff-parameters
- potentially ill-conditioned Hessian
- may not be obvious which constrained problem is solved for a given penalty parameter
- **no guarantees that all model iterates are regularized**

A few regularization strategies





exact versus non-exact penalty

Toy problem:

$$\min_x \frac{1}{2}\|x-1\|_2^2 \quad \text{s.t.} \quad x=2$$

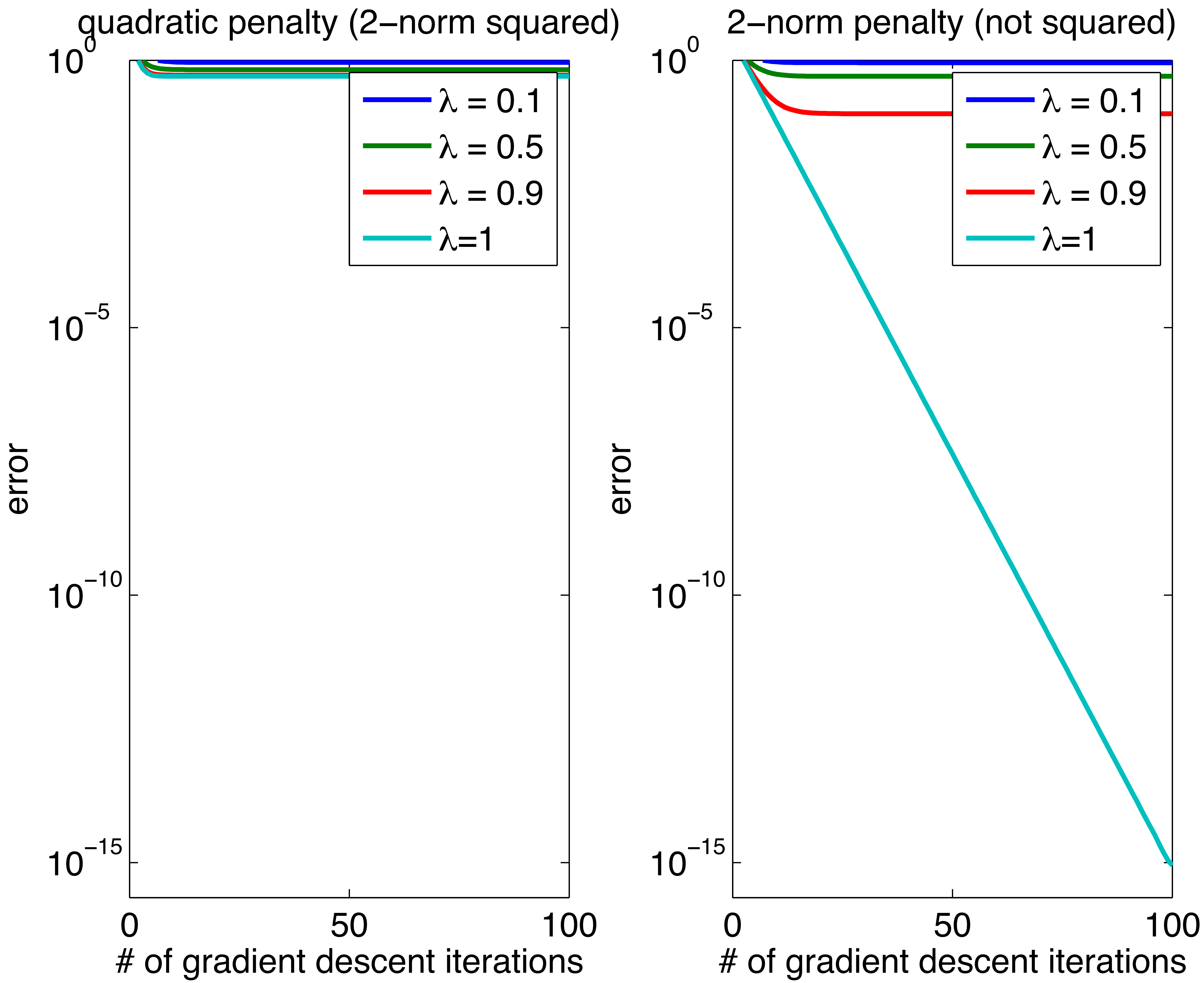
Quadratic-penalty:

$$\min_x \frac{1}{2}\|x-1\|_2^2 + \lambda\|x-2\|_2^2$$

2-norm penalty:

$$\min_x \frac{1}{2}\|x-1\|_2^2 + \lambda\|x-2\|_2$$

A few regularization strategies



exact versus non-exact penalty

Toy problem:

$$\min_x \frac{1}{2}\|x-1\|_2^2 \quad \text{s.t.} \quad x=2$$

Quadratic-penalty:

$$\min_x \frac{1}{2}\|x-1\|_2^2 + \lambda\|x-2\|_2^2$$

2-norm penalty:

$$\min_x \frac{1}{2}\|x-1\|_2^2 + \lambda\|x-2\|_2$$

SLIM

A few regularization strategies

Gradient filtering:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \gamma F \nabla_{\mathbf{m}} f(\mathbf{m})$$

If the gradient filter F is the inverse Hessian, this is just Newton's method

Can work if F is definite positive

Potential problems:

- filtered gradient may not be a gradient of the objective anymore
- **no obvious way to include multiple filters**

A few regularization strategies

Multiple constrained formulation:

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

"Find a model which satisfies all pieces of prior info simultaneously…"

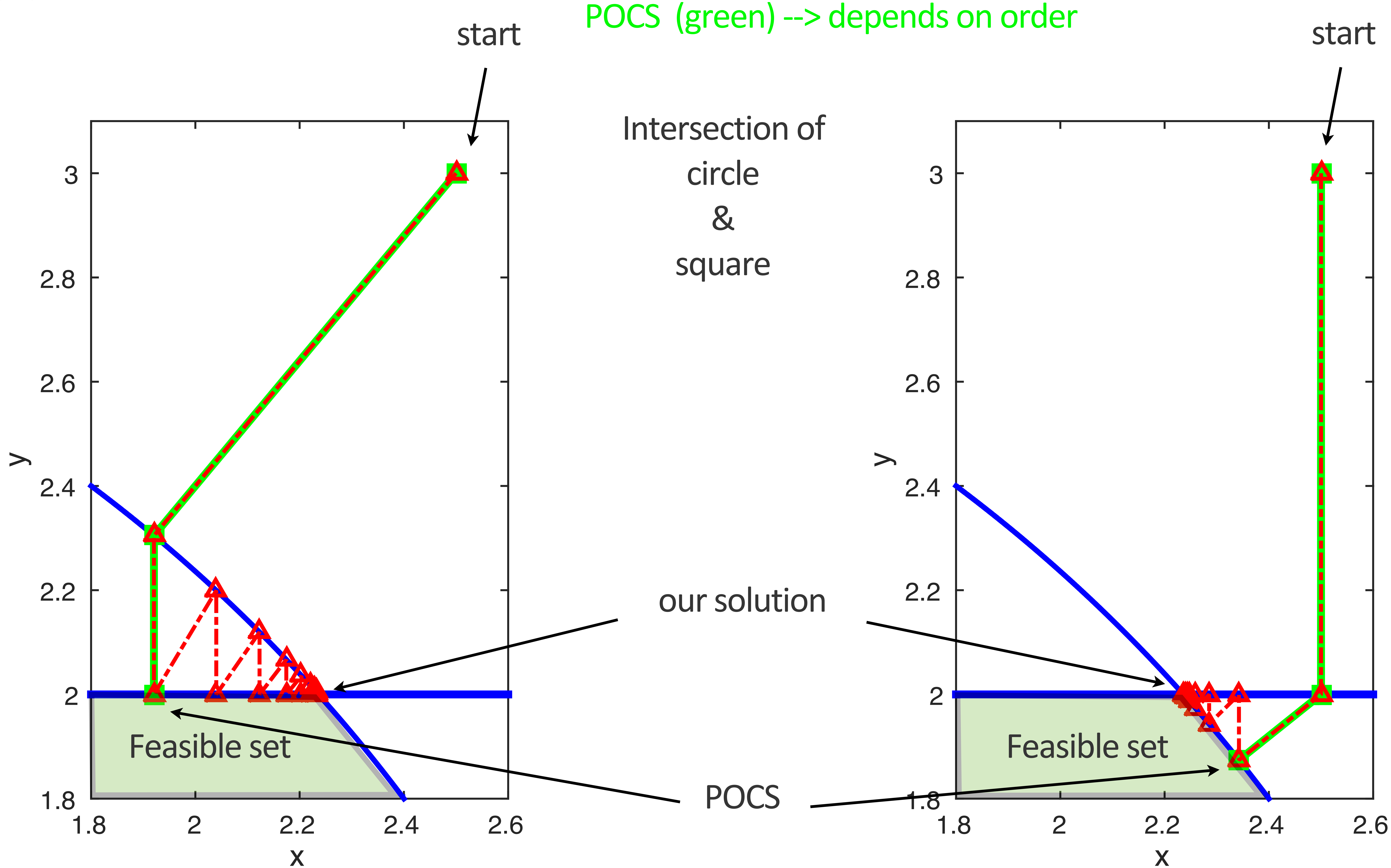
- constraints can be satisfied at every iteration
- works w/ gradient/quasi-Newton/Newton-type methods as a black box
- can define more than two constraint-sets
- no weights or other parameters required, just define the sets

Challenge: to impose multiple constraints uniquely…

POCS vs our method

SLIM (red) --> unique solution

POCS (green) --> depends on order



Prior information as intersection

Projections w/ proximal operators :

$$\mathcal{P}_{\mathcal{C}}(\mathbf{m}) = \arg\min_{\mathbf{x}} \|\mathbf{x} - \mathbf{m}\|_2^2 \quad \text{subject to} \quad \mathbf{x} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

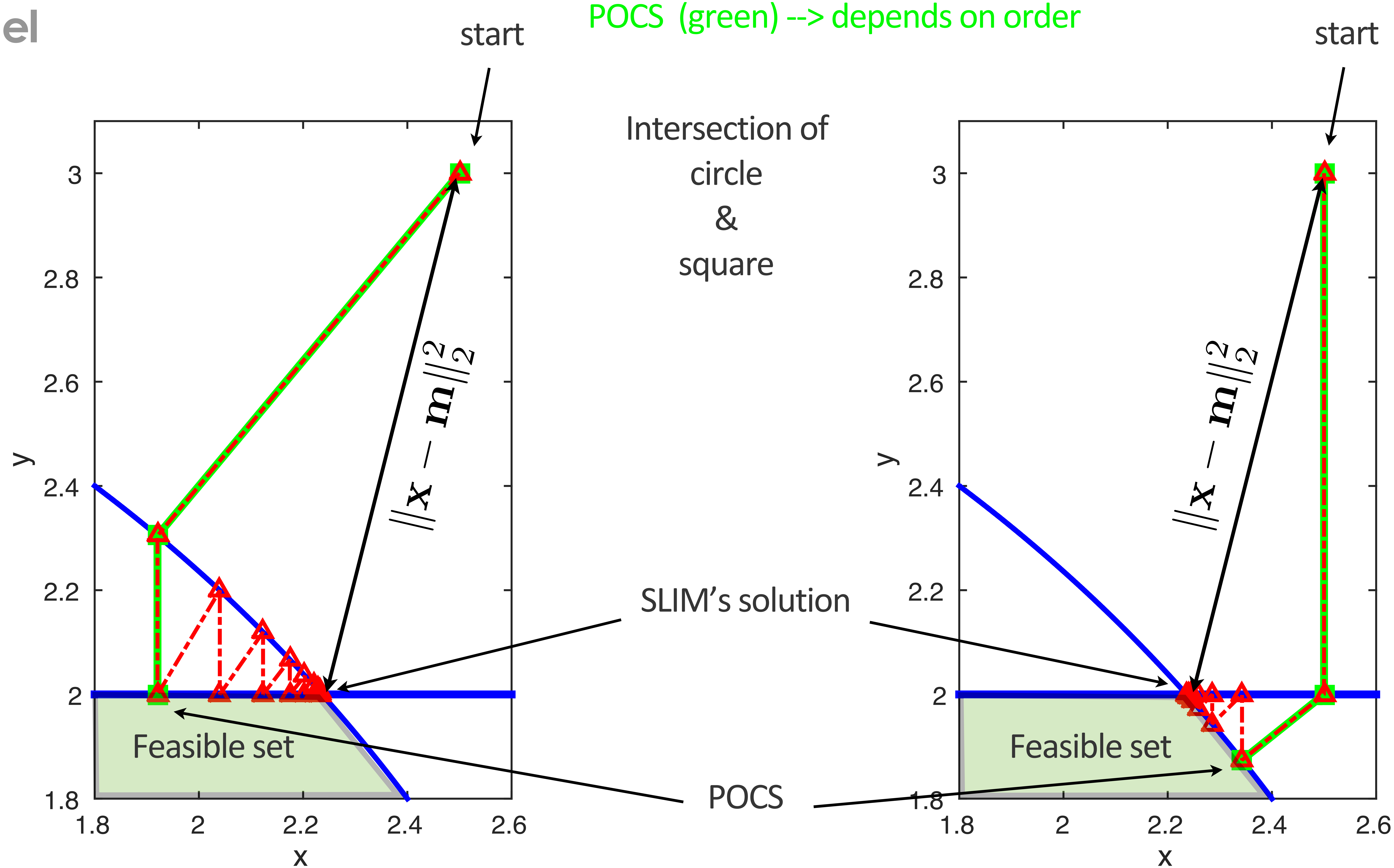
In words: *find closest (Euclidean minimum-distance projection) model subject to being in the intersection of the constraints.*

Important property: $\mathcal{P}_{\mathcal{C}}(\mathbf{m}) = \mathcal{P}_{\mathcal{C}}(\mathcal{P}_{\mathcal{C}}(\mathbf{m}))$

Our method– finds closest model

SLIM (red) --> unique solution

POCS (green) --> depends on order



SLIM

Multiple-constrained formulation

Alternating Projections Onto Convex Sets (POCS) lead to

- new models that can be far from the original model
- project on each set separately => ambiguous results (depends on the order)

Alternating projections w/ proximal operators

- finds the closest model subject to the constraints
- but needs extra work so we project onto the intersection
- **get unique projections on these intersections**

Algorithmic development

$$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

$\mathcal{C}_1 \bigcap \mathcal{C}_2$ is convex if $\mathcal{C}_1$ and $\mathcal{C}_2$ are convex

We would like the model to be in $\mathcal{C}_1 \bigcap \mathcal{C}_2$ at every iteration

One possibility:

$$\min_{\mathbf{m}} f(\mathbf{m}) + \iota_{\mathcal{C}_1}(\mathbf{m}) + \iota_{\mathcal{C}_2}(\mathbf{m})$$

$$\iota_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{if } x \notin \mathcal{C}. \end{cases}$$

Algorithmic development

$$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

$$\min_{\mathbf{m}} f(\mathbf{m}) + \iota_{\mathcal{C}_1}(\mathbf{m}) + \iota_{\mathcal{C}_2}(\mathbf{m})$$ $\longrightarrow$ not differentiable

Can use forward-backward splitting / proximal-gradient algorithms.

Algorithmic development

$$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

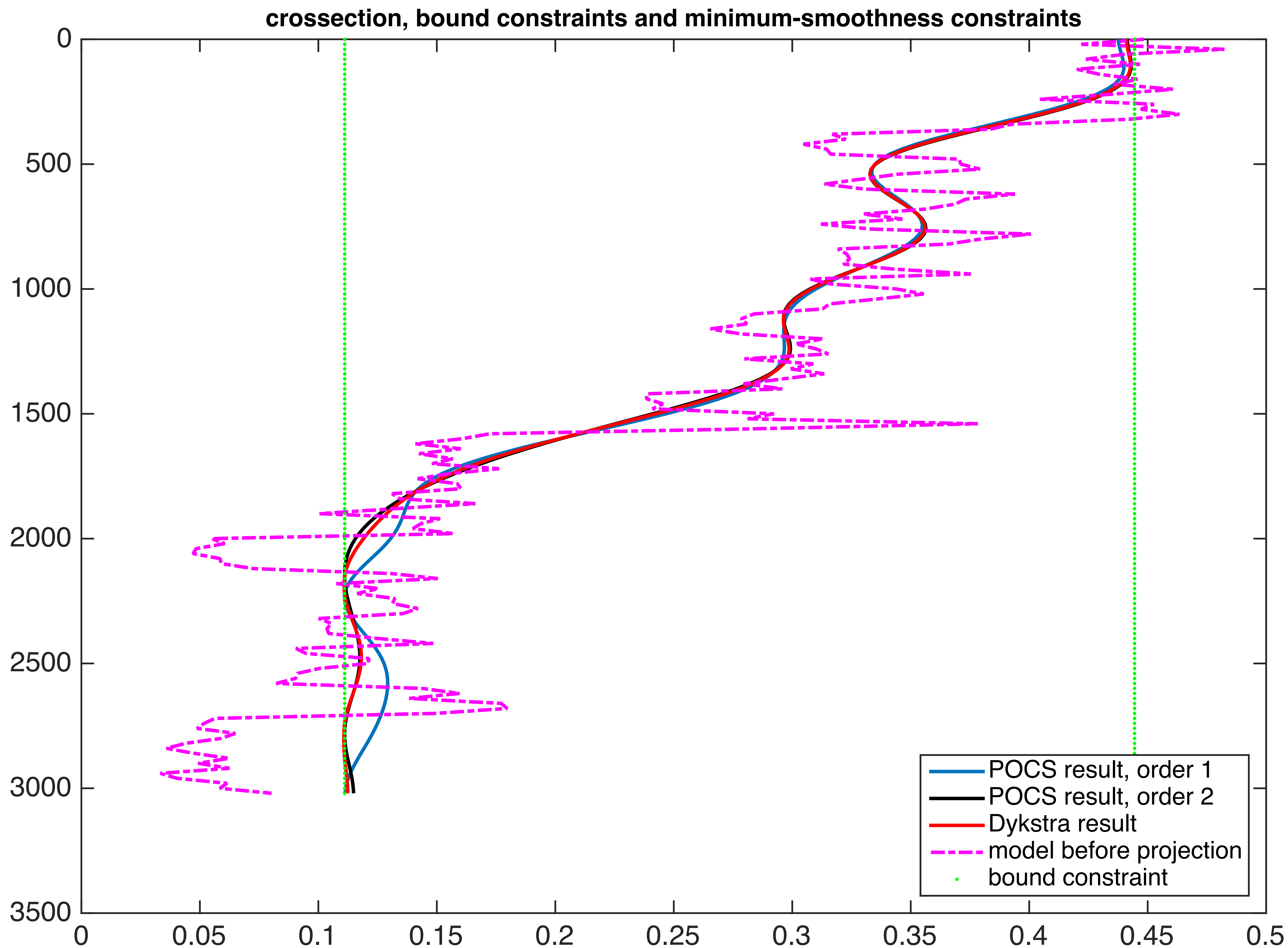
Project onto an intersection of convex sets:

- sometimes known analytically
- otherwise compute numerically; Dykstra's algorithm is used in this work

POCS vs our method

Projection of a noisy Marmousi model onto the intersection of bound constraints & minimum-smoothness constraints

also shows the POCS results which are very different.



Dykstra splitting

Toy example:

find projection onto intersection of a circle and a square

Algorithm 1 Dykstra.

$x_0 = \mathbf{m},\ p_0 = 0,\ q_0 = 0$

For $k = 0, 1, \ldots$

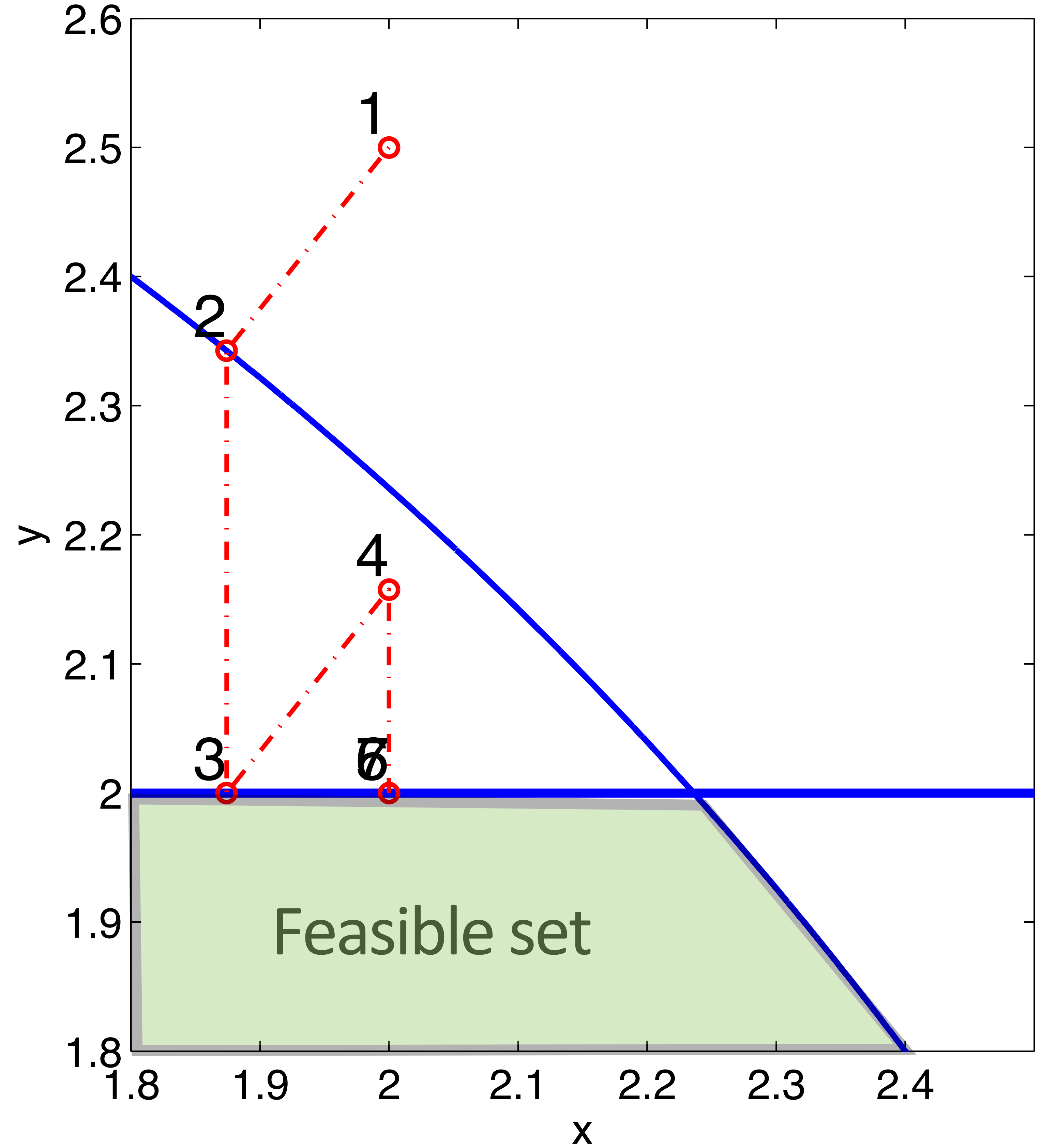
$$y_k = \mathscr{P}_{C_1}(x_k + p_k)$$

$$p_{k+1} = x_k + p_k - y_k$$

$$x_{k+1} = \mathscr{P}_{C_2}(y_k + q_k)$$

$$q_{k+1} = y_k + q_k - x_{k+1}$$

End



Dykstra splitting

Toy example:
find projection onto intersection of a circle and a square

Algorithm 1 Dykstra.

$x_0 = \mathbf{m}$, $p_0 = 0$, $q_0 = 0$

For $k = 0, 1, \ldots$

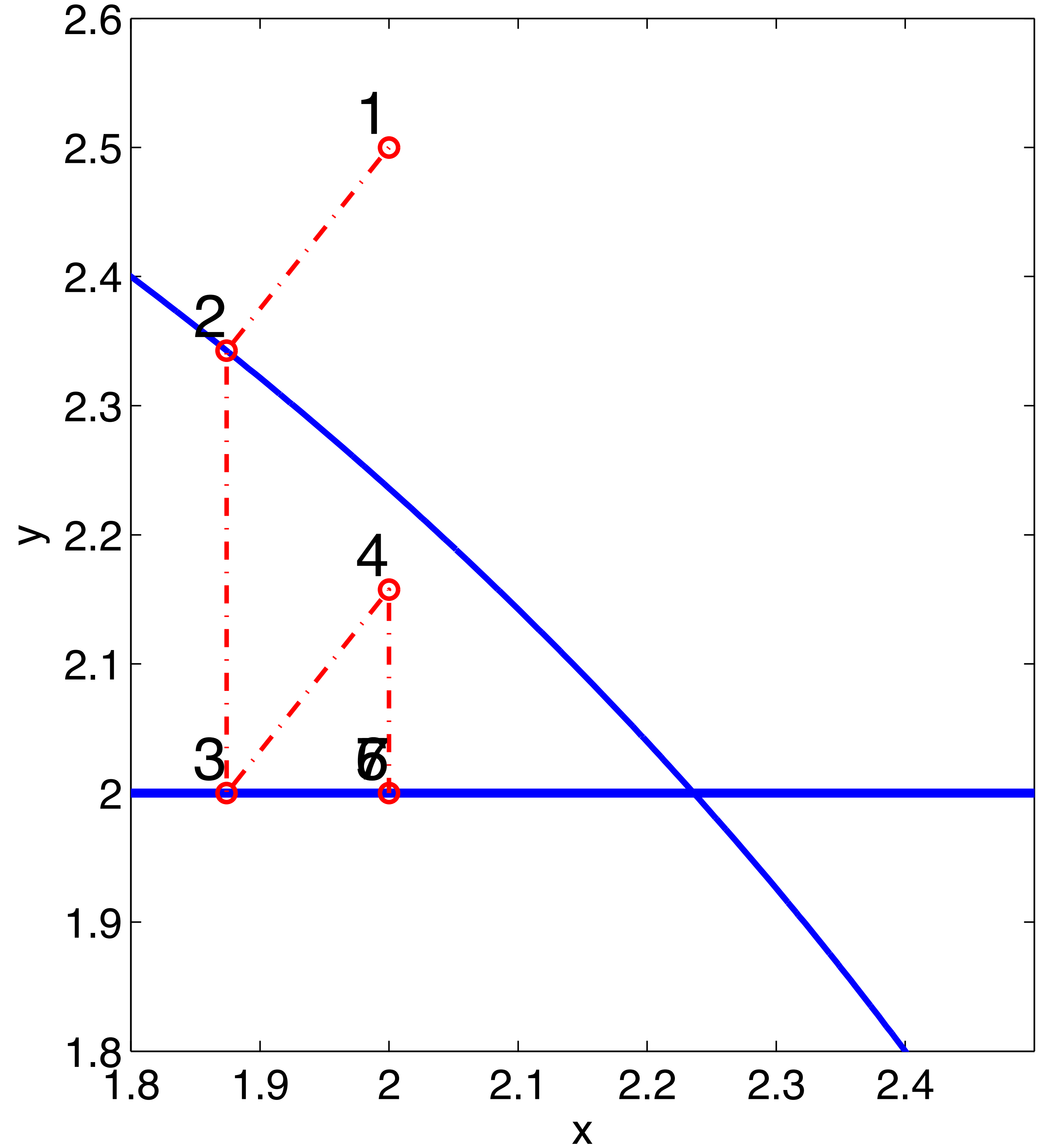
$\longrightarrow$ $y_k = \mathscr{P}_{C_1}(x_k + p_k)$

$p_{k+1} = x_k + p_k - y_k$

$\longrightarrow$ $x_{k+1} = \mathscr{P}_{C_2}(y_k + q_k)$

$q_{k+1} = y_k + q_k - x_{k+1}$

End



only need projection onto each set separately

Dykstra splitting

Toy example:

find projection onto intersection of a circle and a square

Algorithm 1 Dykstra.

$$x_0 = \mathbf{m},\ p_0 = 0,\ q_0 = 0$$

For $k = 0, 1, \ldots$

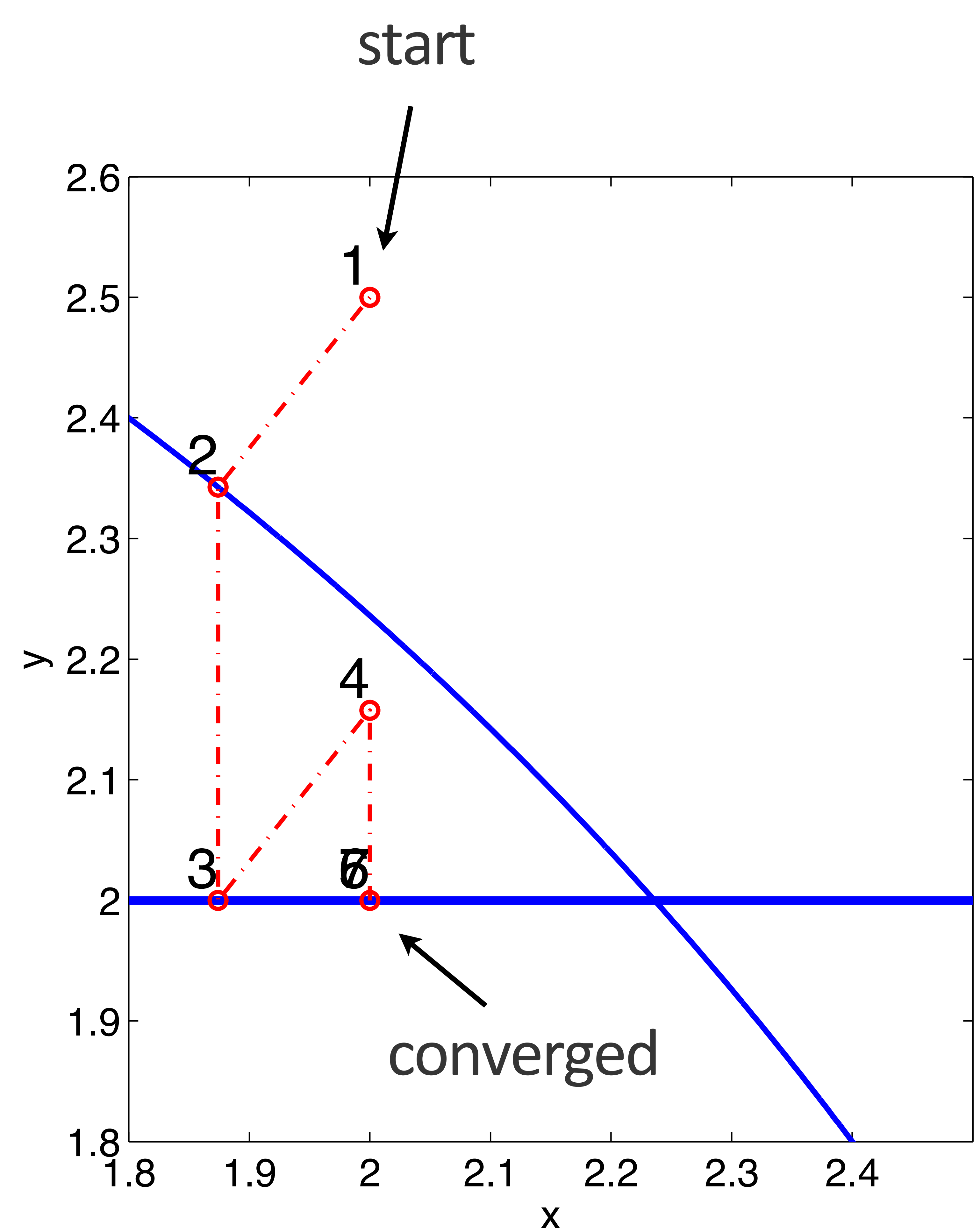
$$y_k = \mathscr{P}_{C_1}(x_k + p_k)$$

$$p_{k+1} = x_k + p_k - y_k$$

$$x_{k+1} = \mathscr{P}_{C_2}(y_k + q_k)$$

$$q_{k+1} = y_k + q_k - x_{k+1}$$

End

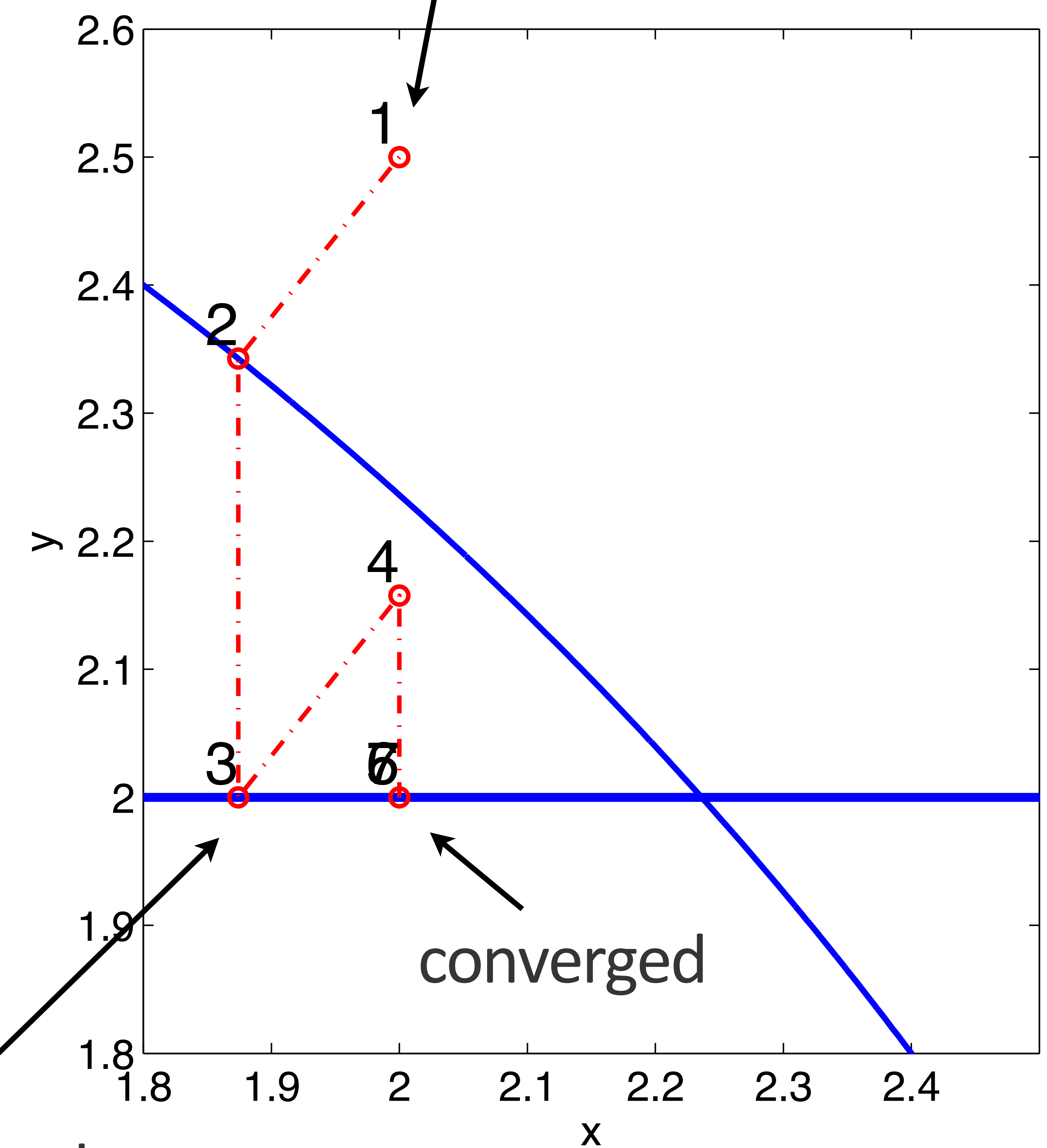


Dykstra splitting

Toy example:

find projection onto intersection of a circle and a square

start



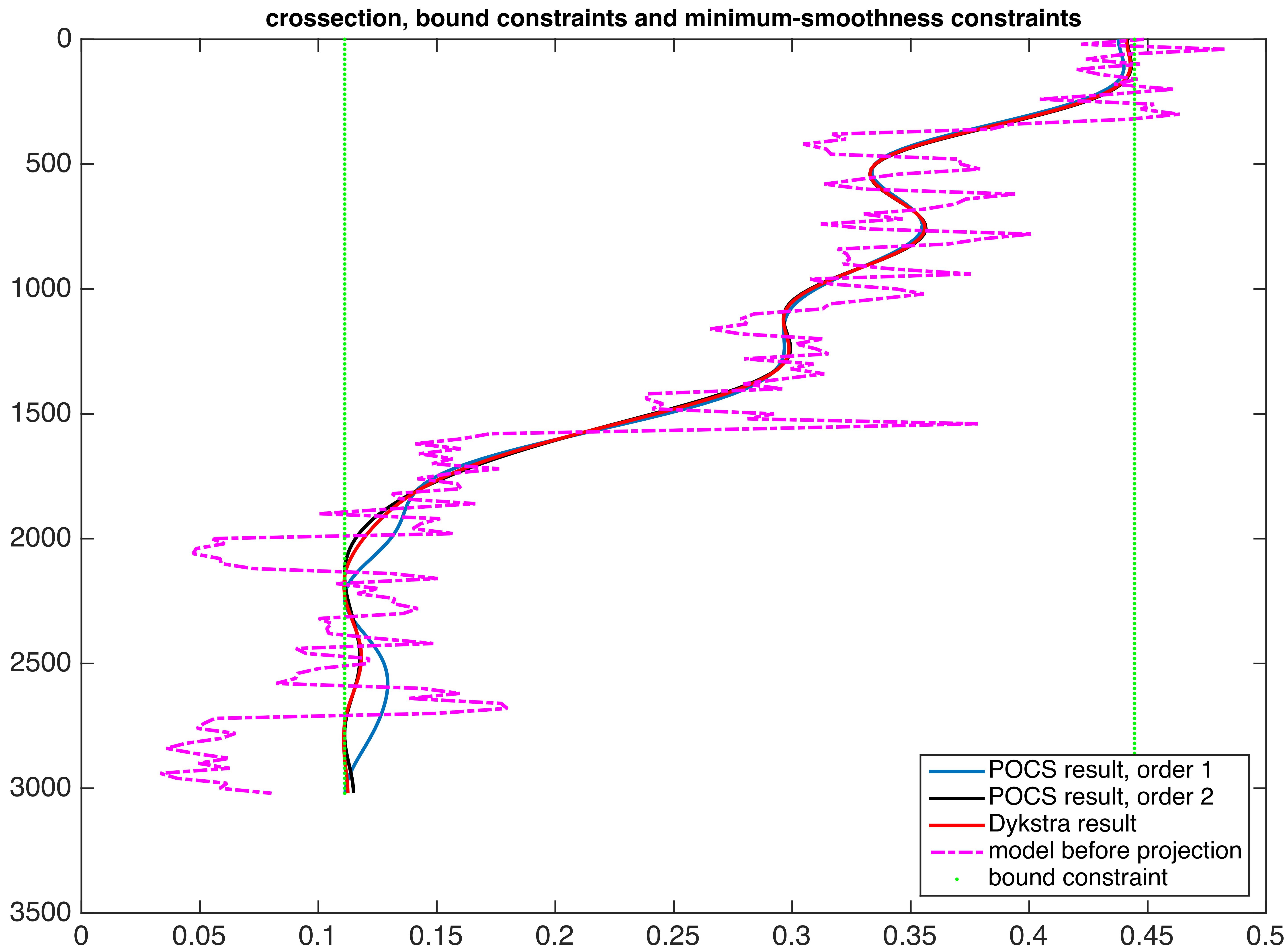
converged

POCS would converge here,
feasible point, not the projection onto

POCS vs Dykstra

Projection of a noisy Marmousi model onto the intersection of bound constraints & minimum-smoothness constraints

also shows the POCS results which are very different.



Dykstra splitting

Projection-onto-convex-sets (POCS) solves the convex feasibility problem:

$$\text{find} \quad x \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

Dykstra's algorithm solves:

$$\min_x \iota_{\mathcal{C}_1}(x) + \iota_{\mathcal{C}_2}(x) + \frac{1}{2}\|x-y\|^2$$

with indicator function:

$$\iota_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C}, \\ +\infty & \text{if } x \notin \mathcal{C}. \end{cases}$$

Dykstra splitting

Projection-onto-convex-sets (POCS) solves the convex feasibility problem:

$$\text{find} \quad x \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

Dykstra's algorithm solves:

$$\min_x \iota_{\mathcal{C}_1}(x) + \iota_{\mathcal{C}_2}(x) + \frac{1}{2}\|x - y\|^2$$

is equivalent to:

$$\min_x \frac{1}{2}\|x - y\|^2 \quad \text{s.t.} \quad x \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

Dykstra splitting

Projection-onto-convex-sets (POCS):

$$\text{find} \quad x \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

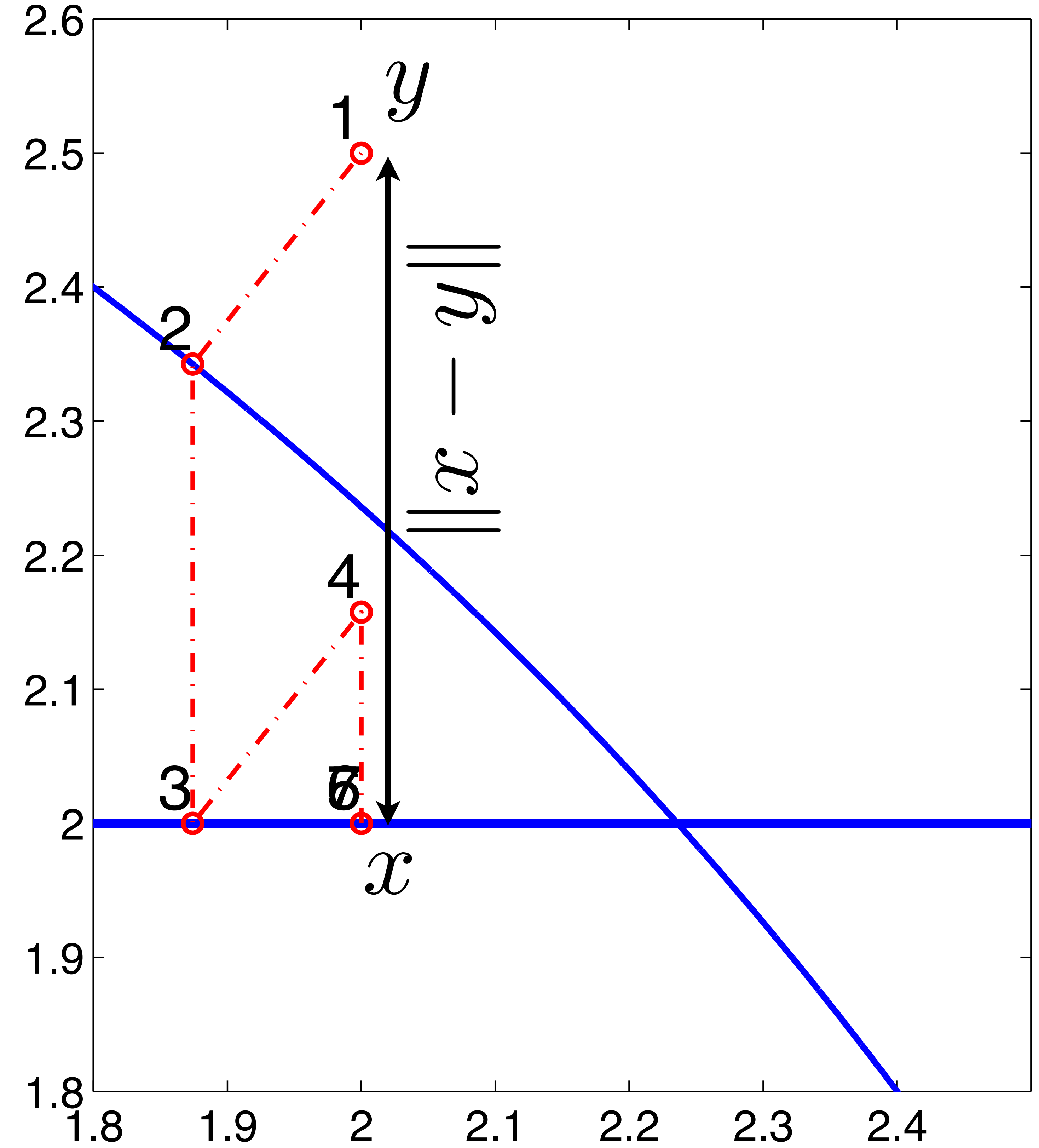
find any point in the intersection, may be the closest point

Dykstra's algorithm solves:

$$\min_x \iota_{\mathcal{C}_1}(x) + \iota_{\mathcal{C}_2}(x) + \frac{1}{2}\|x - y\|^2$$

is equivalent to:

$$\min_x \frac{1}{2}\|x - y\|^2 \quad \text{s.t.} \quad x \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$



Prior information as convex sets

example 1: (spatially varying) bound constraints:

$$\mathcal{C}_1 \equiv \{\mathbf{m} \mid \mathbf{b}_l \leq \mathbf{m} \leq \mathbf{b}_u\}$$

Projector: (element-wise)

$$\mathcal{P}_{\mathcal{C}_1}(\mathbf{m}) = \mathrm{median}\{\mathbf{b}_l, \mathbf{m}, \mathbf{b}_u\}$$

SLIM

Bound constraints on vertical derivative

$\mathcal{C} \equiv \{\mathbf{m}_i \mid \mathbf{b}_i^l \leq A\mathbf{m}_i \leq \mathbf{b}_i^u\}$ with $A = I_n \otimes D_z$

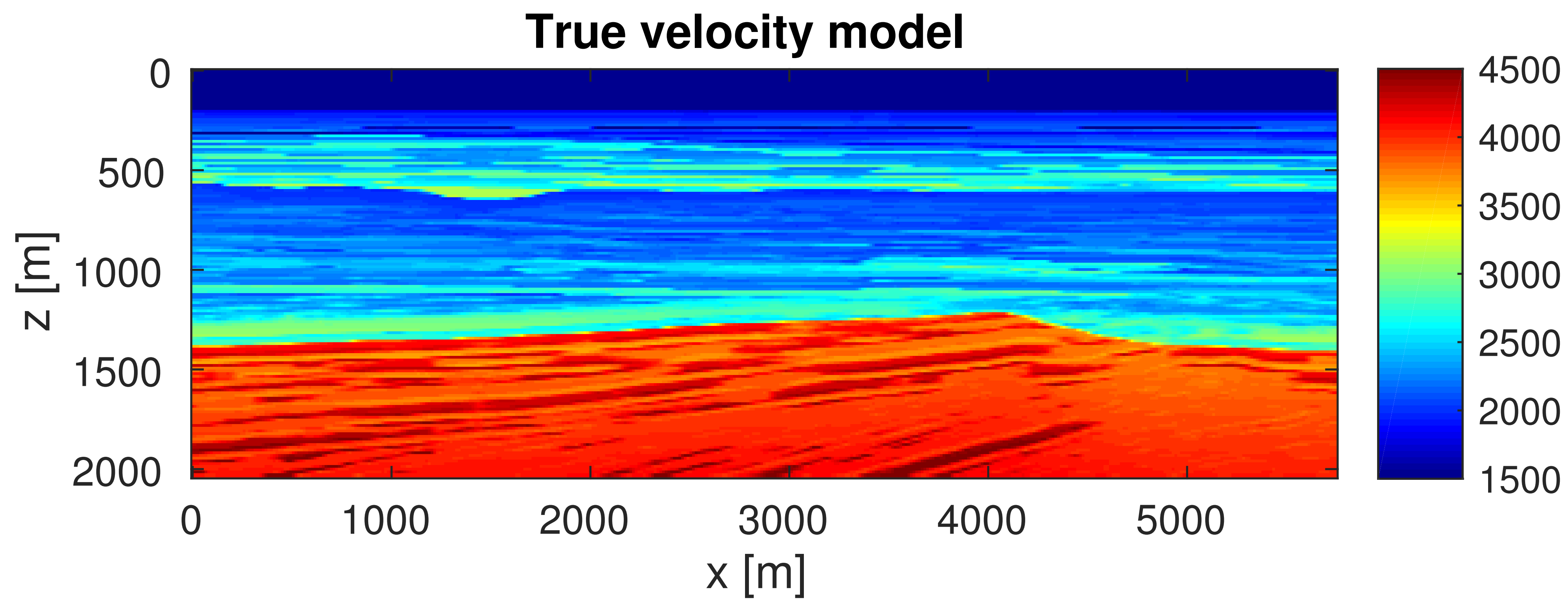
$$D_z = \frac{1}{h_z}\begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix}$$

Interpretation:

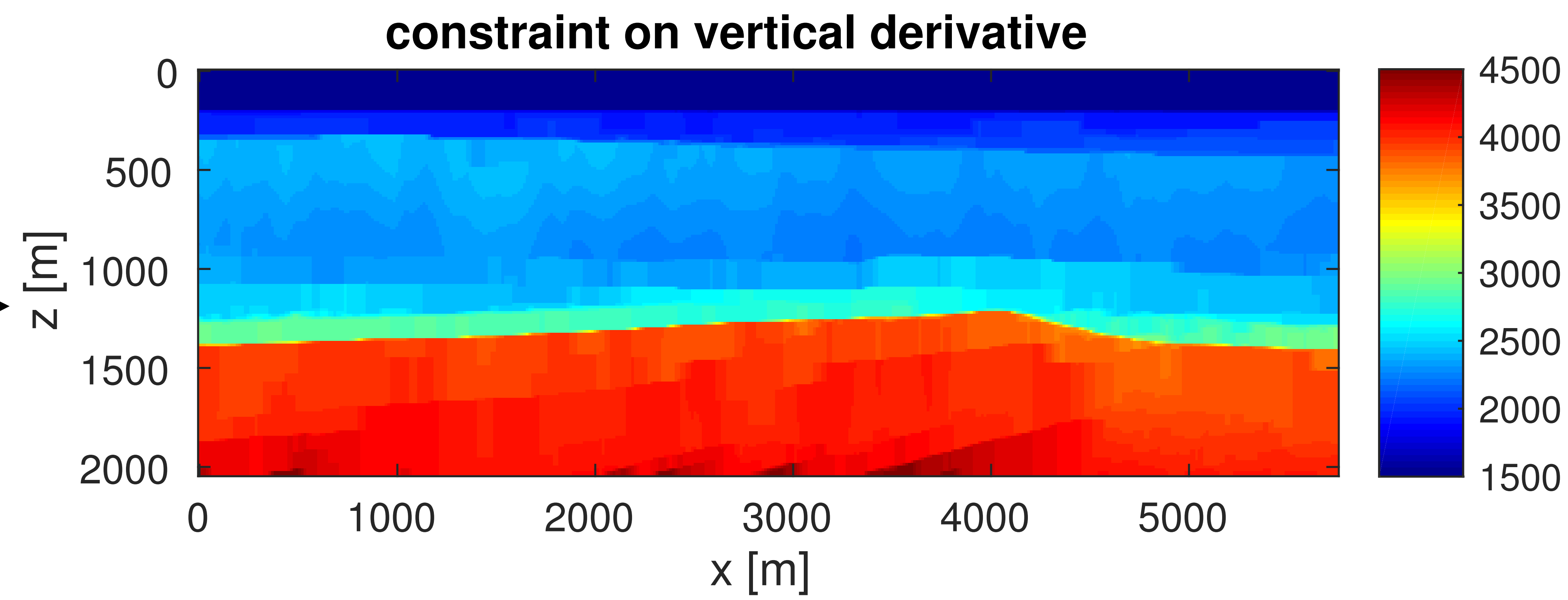
Limit the medium parameter variation per distance unit.

Can select different bounds for each gridpoint.

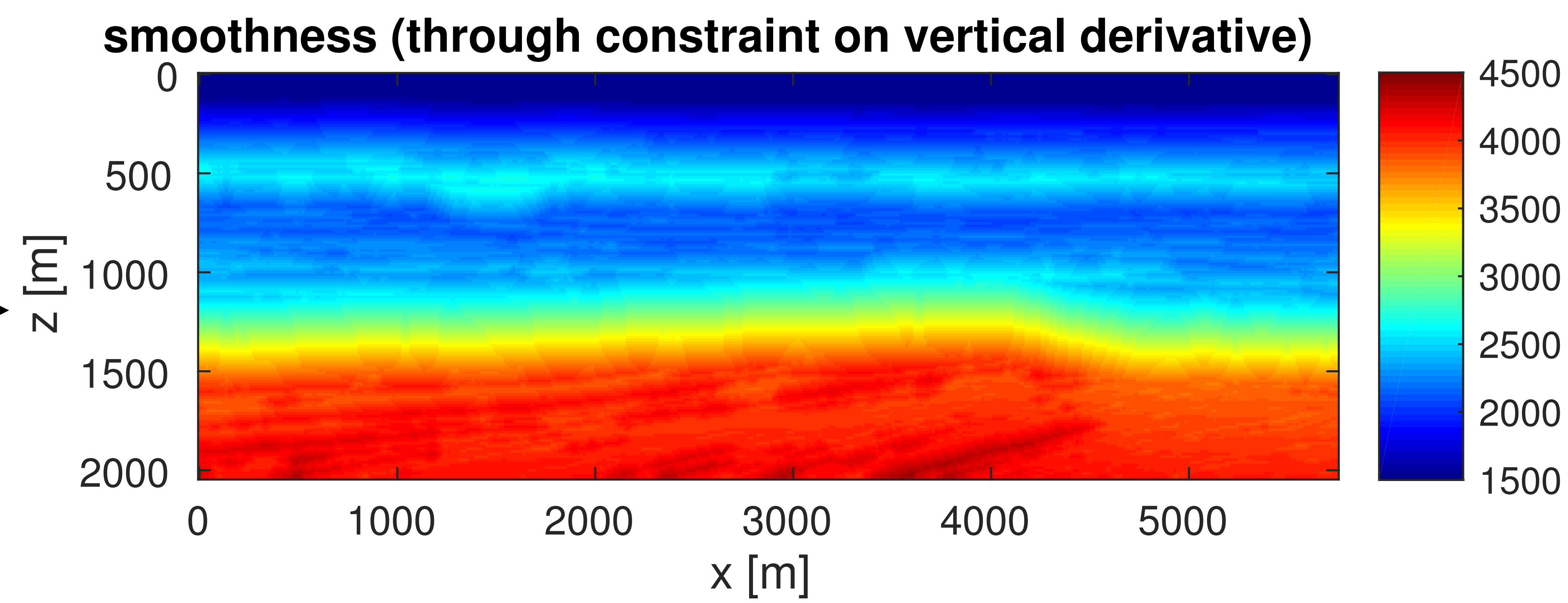
Smoothness & bound constraints
on vertical derivative



arbitrary medium parameter increase, limited medium parameter decrease with depth - induces monotonicity →



limited increase and limited decrease induces vertical smoothness →

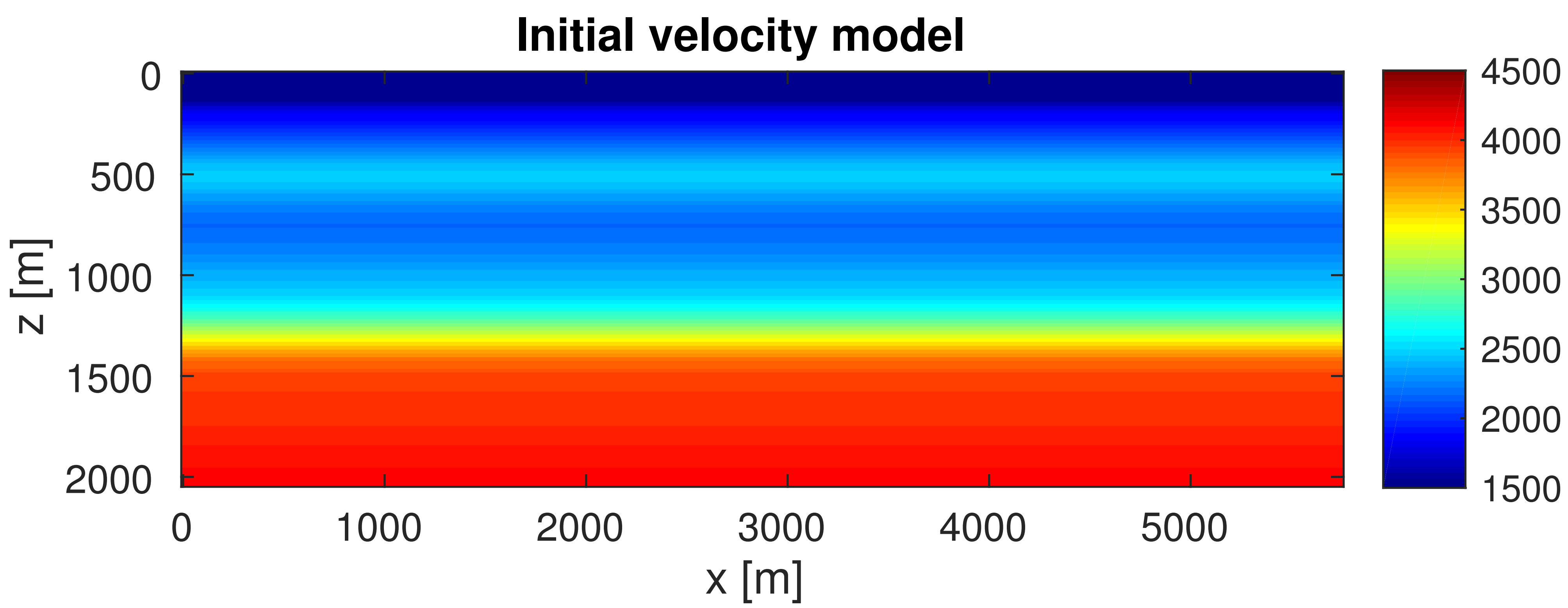


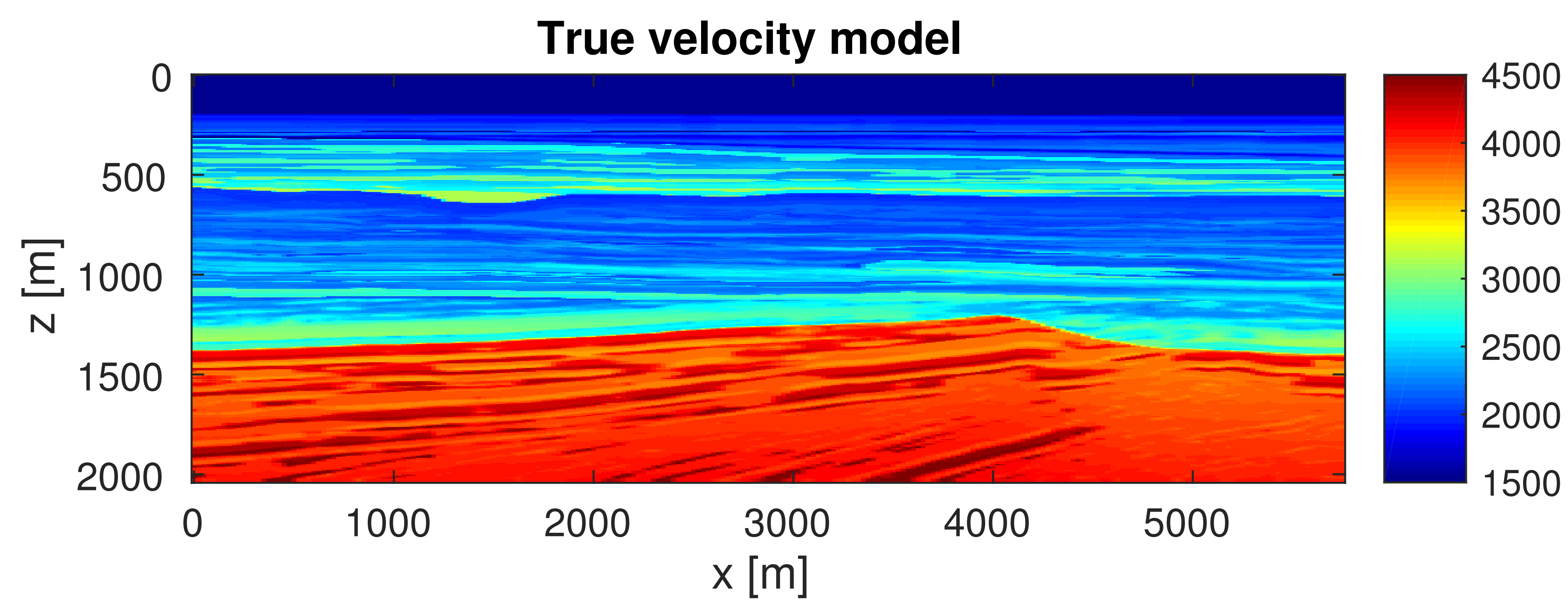
Frequency domain FWI example 1 - noisy data

Frequency batches:

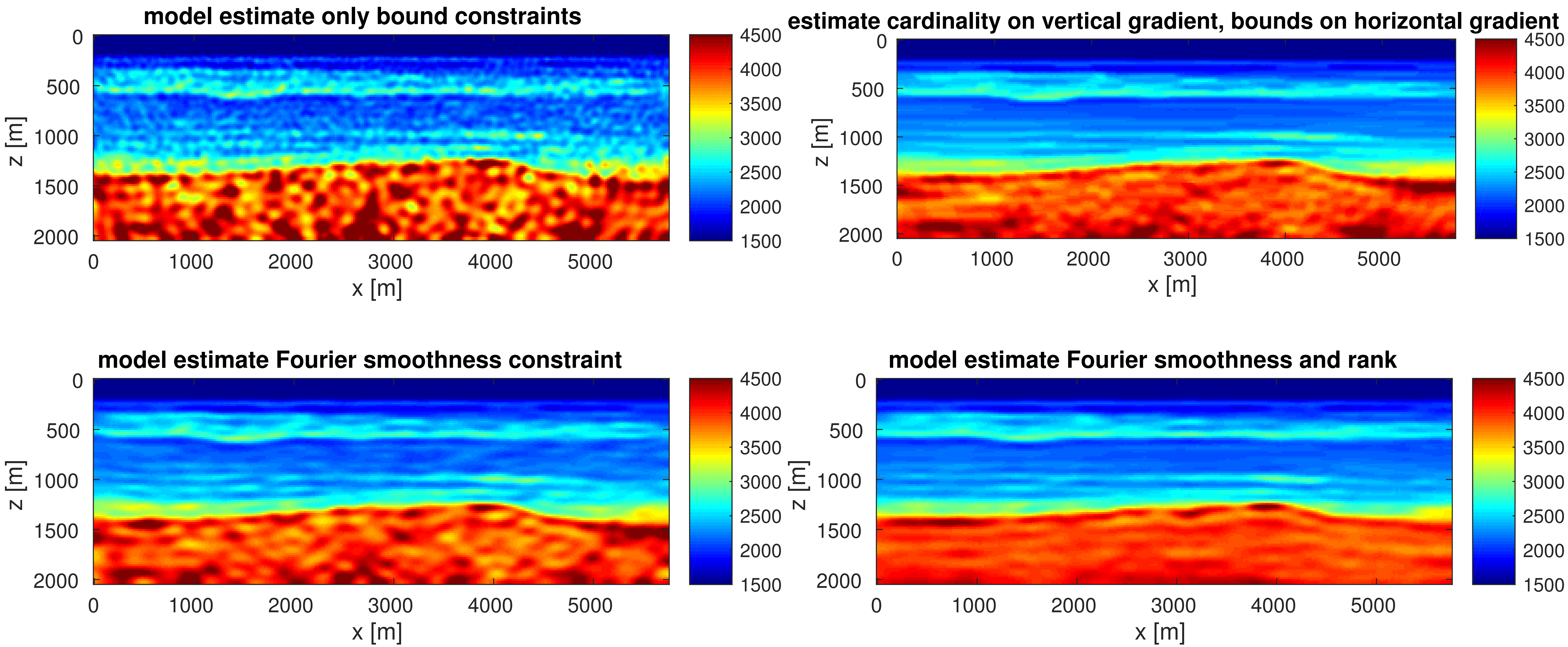
$$\{3, 3.33, 3.67, 4\}, \{4, 4.33, 4.67, 5\}, \{\ldots\}, \{12, 12.33, 12.67, 13\}$$

$$\|\text{noise}\|_2 / \|\text{signal}\|_2 = 1$$





Frequency domain FWI example 1 - noisy data

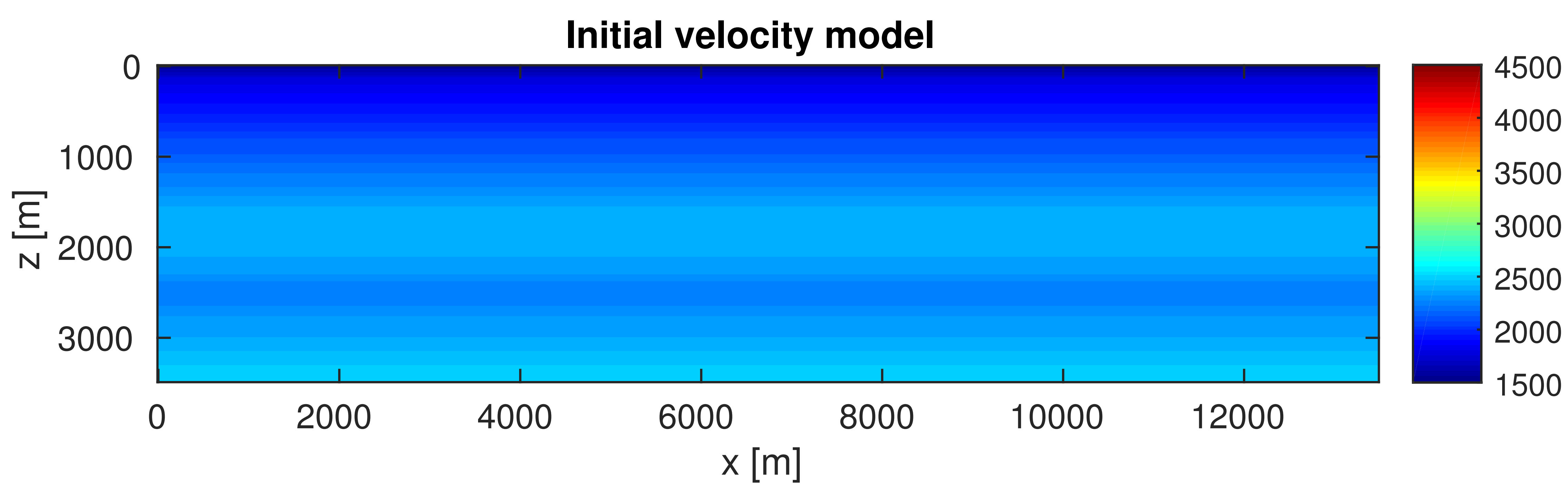


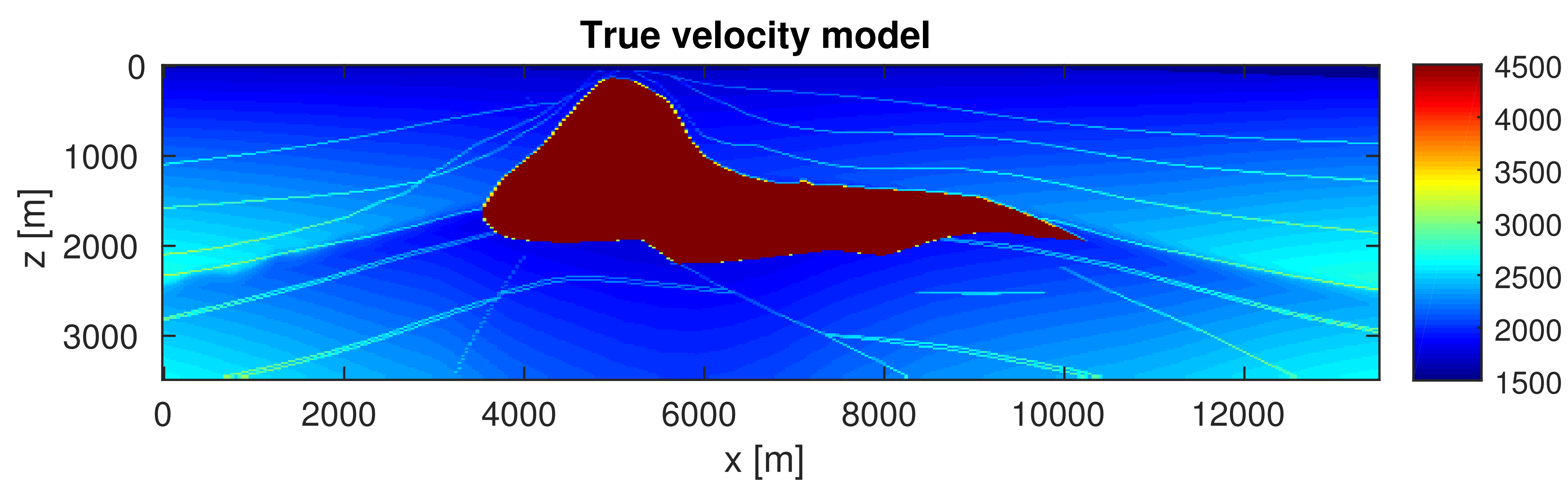
Frequency domain FWI example 2 - Salt structure

Frequency batches: $\{3, 3.33, 3.67, 4\}, \{4, 4.33, 4.67, 5\}, \{\ldots\}, \{12, 12.33, 12.67, 13\}$

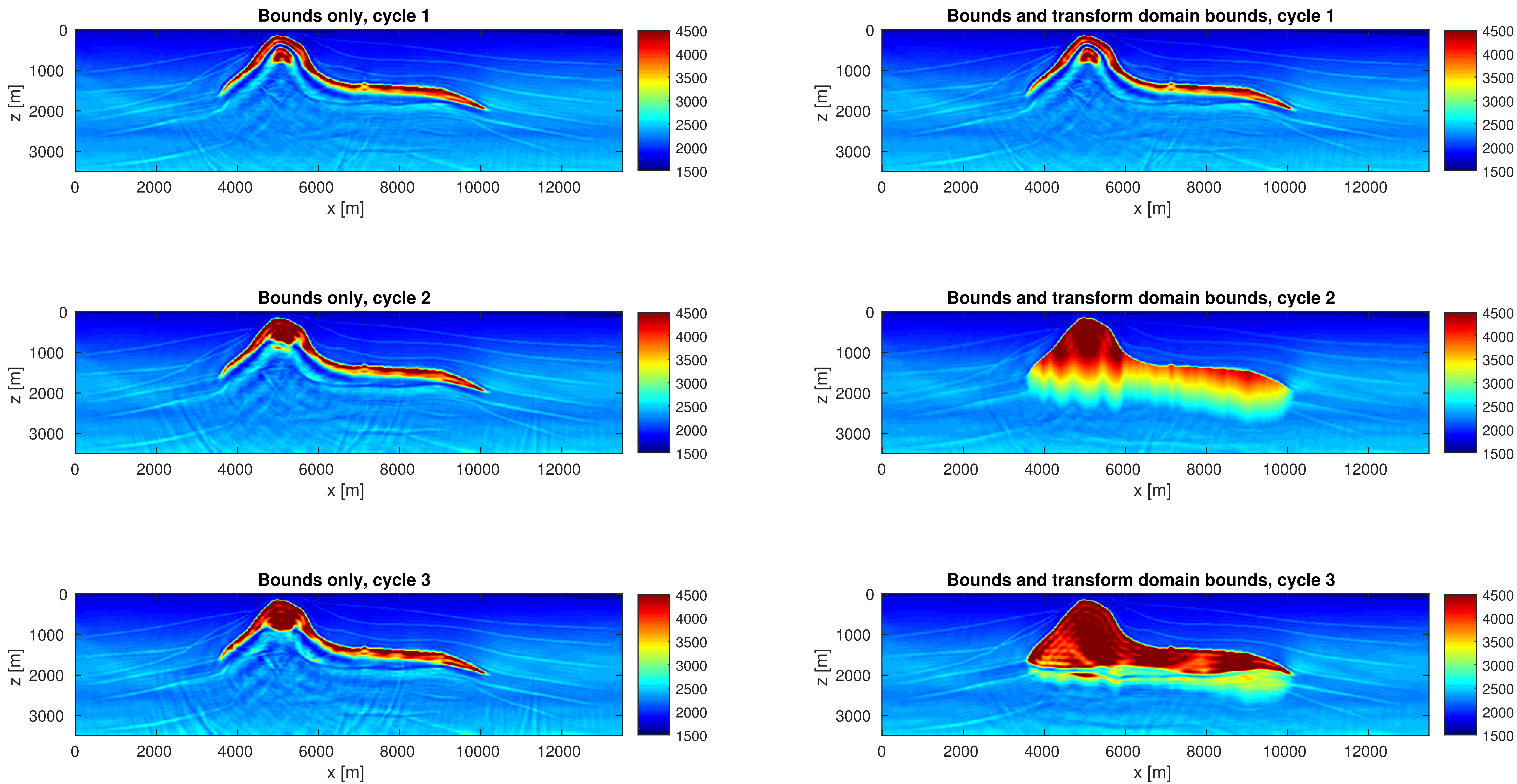
Strategy: 1st cycle bounds only -> 2nd cycle bounds & transform-domain bounds -> 3rd cycle bounds only

Bounds on vertical gradient set to allow arbitrary velocity jumps up & require smooth decrease of velocity with depth.





Frequency domain FWI example 2 - Salt structure



Algorithmic development

$$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

Projected-gradient: $\mathbf{m}_{k+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{m}_k - \gamma \nabla_{\mathbf{m}} f(\mathbf{m}_k))$

SLIM

Algorithmic development

$$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

Projected-gradient: $\mathbf{m}_{k+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{m}_k - \gamma \nabla_{\mathbf{m}} f(\mathbf{m}_k))$

Can this simply be accelerated using Hessian approximation $B(\mathbf{m}_k)$?

$$\mathbf{m}_{k+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{m}_k - \gamma B(\mathbf{m}_k)^{-1} \nabla_{\mathbf{m}} f(\mathbf{m}_k))$$

SLIM

Algorithmic development

$$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

Projected-gradient: $\mathbf{m}_{k+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{m}_k - \gamma \nabla_{\mathbf{m}} f(\mathbf{m}_k))$

Can this simply be accelerated using Hessian approximation $B(\mathbf{m}_k)$?

$$\mathbf{m}_{k+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{m}_k - \gamma B(\mathbf{m}_k)^{-1} \nabla_{\mathbf{m}} f(\mathbf{m}_k))$$

Generally not, when using the Euclidean projection and general $B(\mathbf{m}_k)$

SLIM

Algorithmic development

Projected-gradient: $\mathbf{m}_{k+1} = \mathcal{P}_{\mathcal{C}}(\mathbf{m}_k - \gamma \nabla_{\mathbf{m}} f(\mathbf{m}_k))$

Projected Quasi-Newton **[M. Schmidt et. al., 2009]**

- solves quadratic sub-problem with constraints using the spectral projected-gradient algorithm (inexactly)
- L-BFGS Hessian

Projected Newton-type:

- solves quadratic sub-problem with constraints
- efficient if approximate Hessian is ‘easy to invert’

Algorithmic development

Projected Newton-type:

- solves quadratic sub-problem with constraints:

$$Q(\mathbf{m}) = f(\mathbf{m}_k) + (\mathbf{m} - \mathbf{m}_k)^* \nabla_{\mathbf{m}} f(\mathbf{m}_k) + (\mathbf{m} - \mathbf{m}_k)^* B_k (\mathbf{m} - \mathbf{m}_k)$$

$$\mathbf{m}_{k+1} = \min_{\mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2} Q(\mathbf{m})$$

- efficient if approximate Hessian is ‘easy to invert’ (factored Hessian, sparse & well conditioned, diagonal)

Multiple algorithms can solve the constrained sub-problem

We use Alternating Direction Method of Multipliers (ADMM)

SLIM

Algorithmic development

Projected Newton-type:

- solves quadratic sub-problem with constraints:

$$\mathbf{m}_{k+1} = \min_{\mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2} Q(\mathbf{m})$$

- can be reformulated as: **[M. Schmidt et. al., 2011]**

$$\mathbf{y}_k = \mathbf{m}_k - B_k^{-1} \nabla_{\mathbf{m}} f(\mathbf{m}_k)$$

(unconstrained Newton-step)

$$\mathbf{m}_{k+1} = \min_{\mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2} \frac{1}{2} \|\mathbf{y}_k - \mathbf{m}\|_{B_k}^2$$

(projection w.r.t. metric induced by the approximate Hessian)

Algorithmic development

Projection methods do not modify the gradient or Hessian.

Instead, they find an updated model which still satisfies the constraints.

Workflow summary

1. Define convex feasible sets, possibly velocity & frequency dependent
2. Set up Dykstra's algorithm for projection onto intersections of sets
3. Set up an algorithm to solve the quadratic sub-problem with constraints (ADMM)
4. Solve waveform inversion problem using the a Projected Newton-type algorithm

Observations

Constraints are a powerful method to regularize FWI

- imposed on every iteration
- parameters intuitive to choose
- unambiguous results

Is leading to major breakthroughs as long we can design clever constraints...

Relation to other work

Using projection onto convex sets to constrain full-wavefield inversion

Patent number: 9140812 by Anatoly Baumstein from Exxonmobil

General Optimization Framework for Robust and Regularized 3D FWI

by S.R. Becker, L. Horesh, A.Y. Aravkin, E. van den Berg and S. Zhuk from IBM

The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication by Ludovic Métivier and Romain Brossier

Further reading

1. Bas Peters, Zhilong Fang, Brendan Smithyman, and Felix J. Herrmann, "Regularizing waveform inversion by projections onto convex sets –- application to the 2D Chevron 2014 synthetic blind-test dataset". 2015.
2. Ernie Esser, Lluís Guasch, Tristan van Leeuwen, Aleksandr Y. Aravkin, and Felix J. Herrmann, "Total variation regularization strategies in full waveform inversion for improving robustness to noise, limited data and poor initializations". 2015.
3. Bas Peters, Brendan Smithyman, and Felix J. Herrmann, "Regularizing waveform inversion by projection onto intersections of convex sets". 2015.
4. Regularization of nonlinear geophysical inverse problems using projection methods by Bas Peters, Brendan Smithyman, Felix J. Herrmann in preparation

Conclusions

Randomizations in acquisition make

- seismic surveys more economic
- reduces the environmental impact
- allows for recovery of fully-sampled data volumes

Randomization in computations make

- wave-equation based inversions more economic
- but still rely on underlying fold

Open problems

- combine randomized acquisition w/ wave-equation inversions to mitigate acquisition imprints
- build in adaptive sampling

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