Randomized Algorithms in Exploration Seismology

Felix J. Herrmann

SLIM

University of British Columbia

Carry home messages

Randomization of field-data acquisition = Compressive Sensing leads to

- new insights how to acquire data economically w/ reduced environmental impact
- improved wavefield reconstructions via structure-promoting inversions
- has been employed & validated on field dataset trials and is in production

Randomization of wave-equation based inversion leads to

- new insights how to invert exceedingly large data volumes
- fast & therefore economically viable RTM & FWI
- has been employed, validated, and is in production

Relaxing the physics, randomization & structure promotion lead to

- new insights how to avoid local minima
- reduced reliance on accuracy starting models...

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

Compressive sensing in a nutshell

SLIM

University of British Columbia

Felix J. Herrmann, Michael P. Friedlander, and Ozgur Yilmaz, “Fighting the Curse of Dimensionality: Compressive Sensing in Exploration Seismology”, *Signal Processing Magazine, IEEE*, vol. 29, p. 88-100, 2012
Felix J. Herrmann, “Randomized sampling and sparsity: Getting more information from fewer samples”, *Geophysics*, vol. 75, p. WB173-WB187, 2010

SLIM

Compressive sensing paradigm

Find representations that reveal structure

- transform-domain sparsity (e.g., Fourier, curvelets, etc.)
- rank revealing transforms (e.g. midpoint-offset domain)

Sample to break this structure

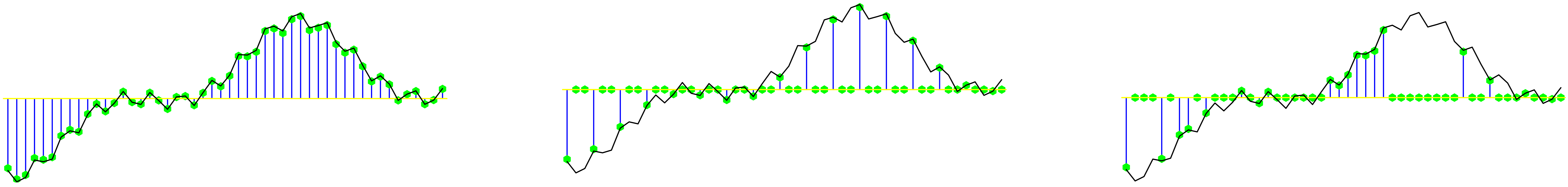
- randomized acquisition (e.g., jittered sampling, time dithering, source encoding, etc.)
- destroys sparsity or low-rank structure

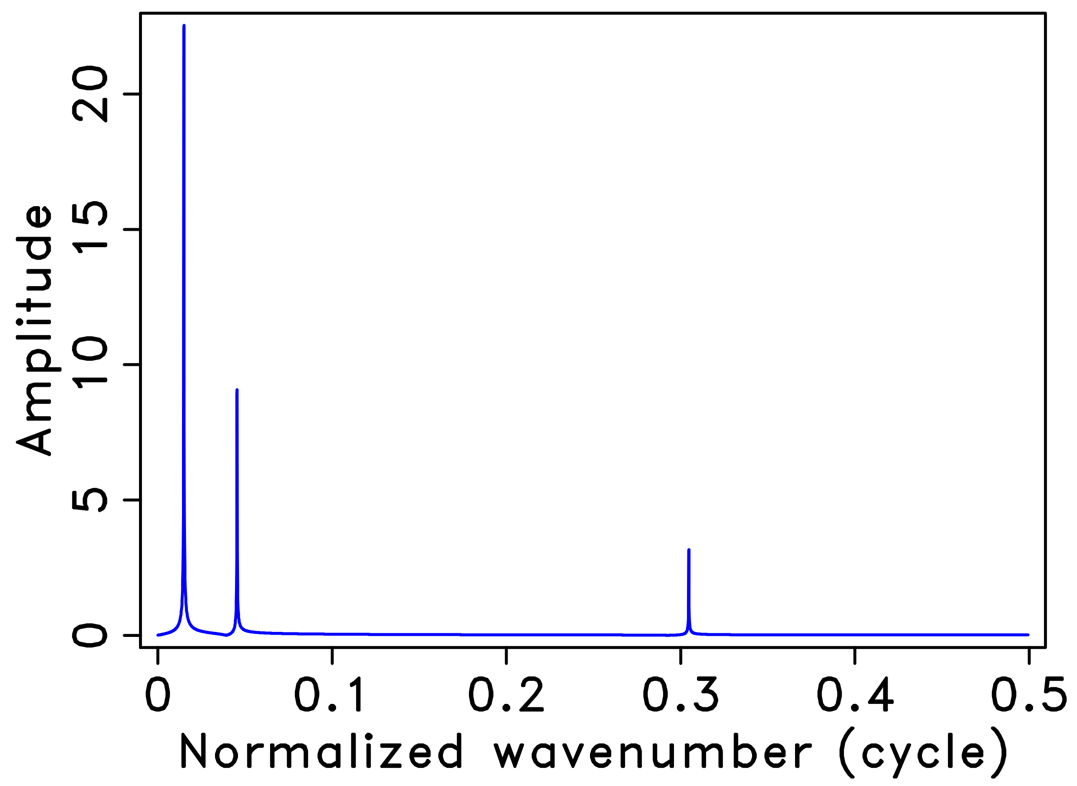
Recover this structure by promoting

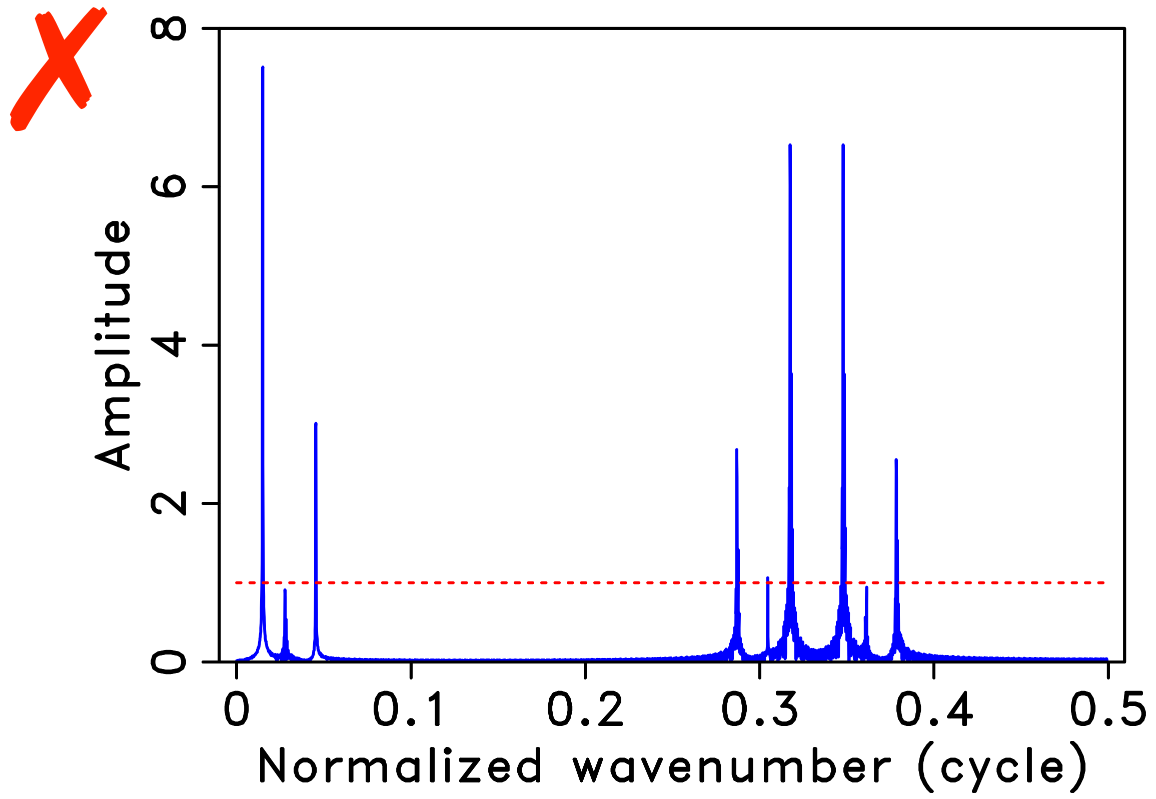
- sparsity via one-norm minimization
- rank revealing nuclear-norm minimization (one-norm singular values)

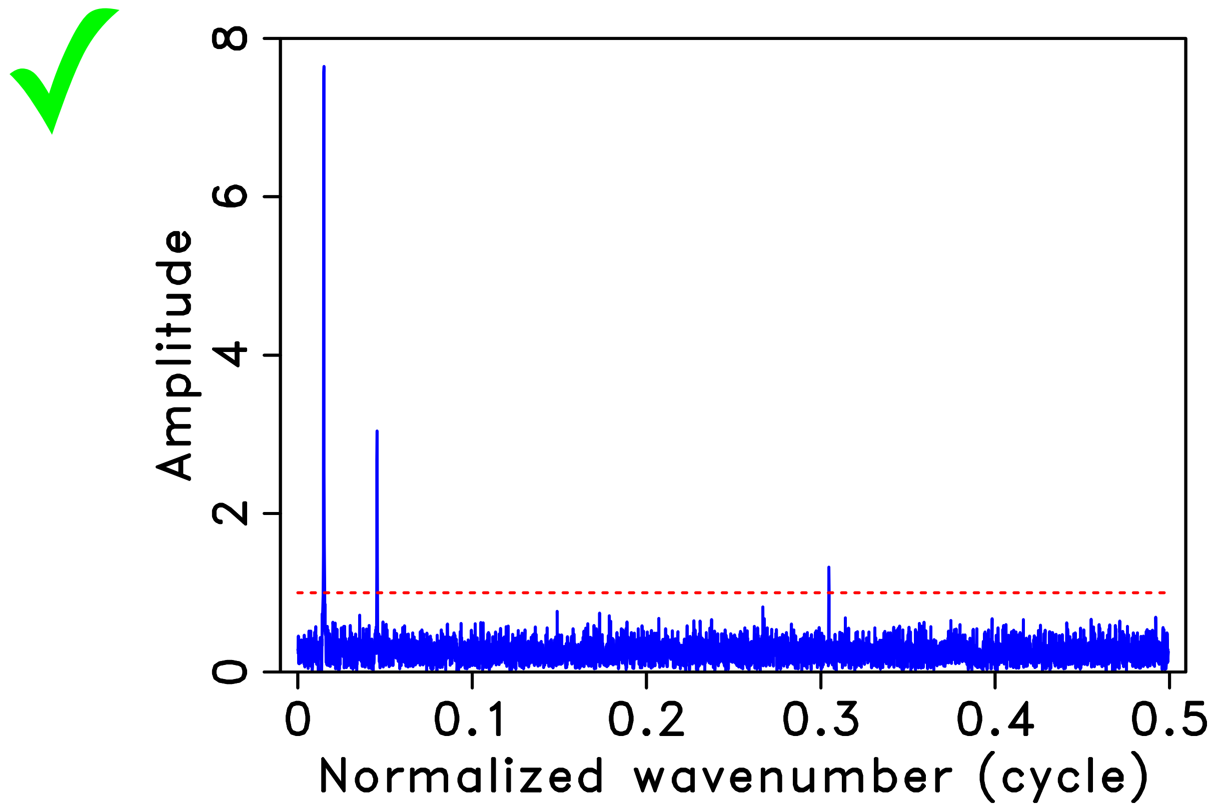
Felix J. Herrmann and Gilles Hennenfent, "Non-parametric seismic data recovery with curvelet frames", *GJI*, vol. 173, p. 233-248, 2008.
Gilles Hennenfent and Felix J. Herrmann, "Simply denoise: wavefield reconstruction via jittered undersampling", *Geophysics*, vol. 73, p. V19-V28, 2008.
Felix J. Herrmann, "Randomized sampling and sparsity: Getting more information from fewer samples", *Geophysics*, vol. 75, p. WB173-WB187, 2010.

Periodic vs random subsampling
– sparse time-harmonic signals









Sparsity-promoting recovery

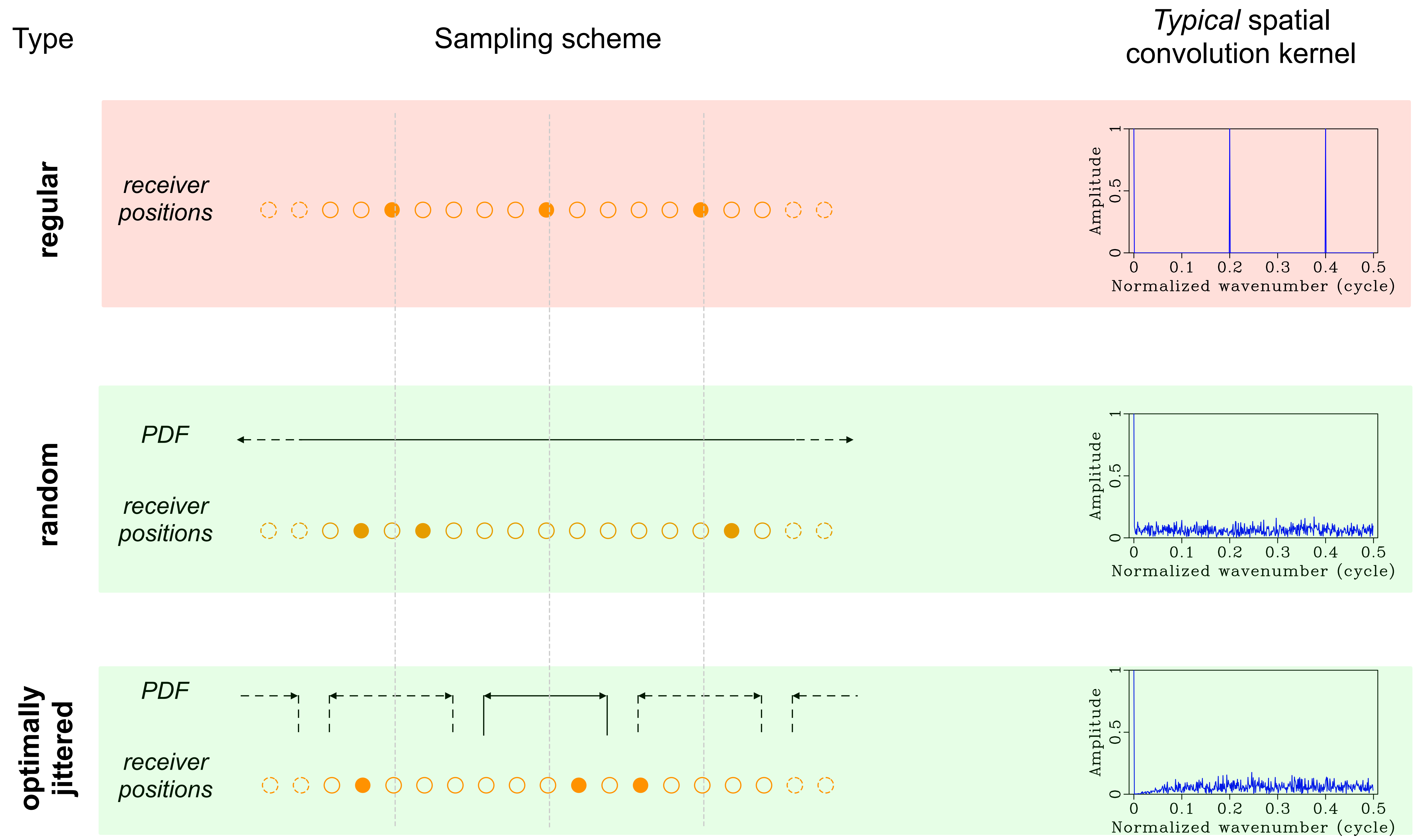
$$\tilde{\mathbf{x}} = \underbrace{\arg\min_{\mathbf{x}} \|\mathbf{x}\|_1}_{\text{support detection}} \quad \text{subject to} \quad \underbrace{\mathbf{A}\mathbf{x} = \mathbf{b}}_{\text{data-consistent amplitude recovery}}$$

recovered data: $\tilde{\mathbf{d}} = \mathbf{S}^H \tilde{\mathbf{x}}$

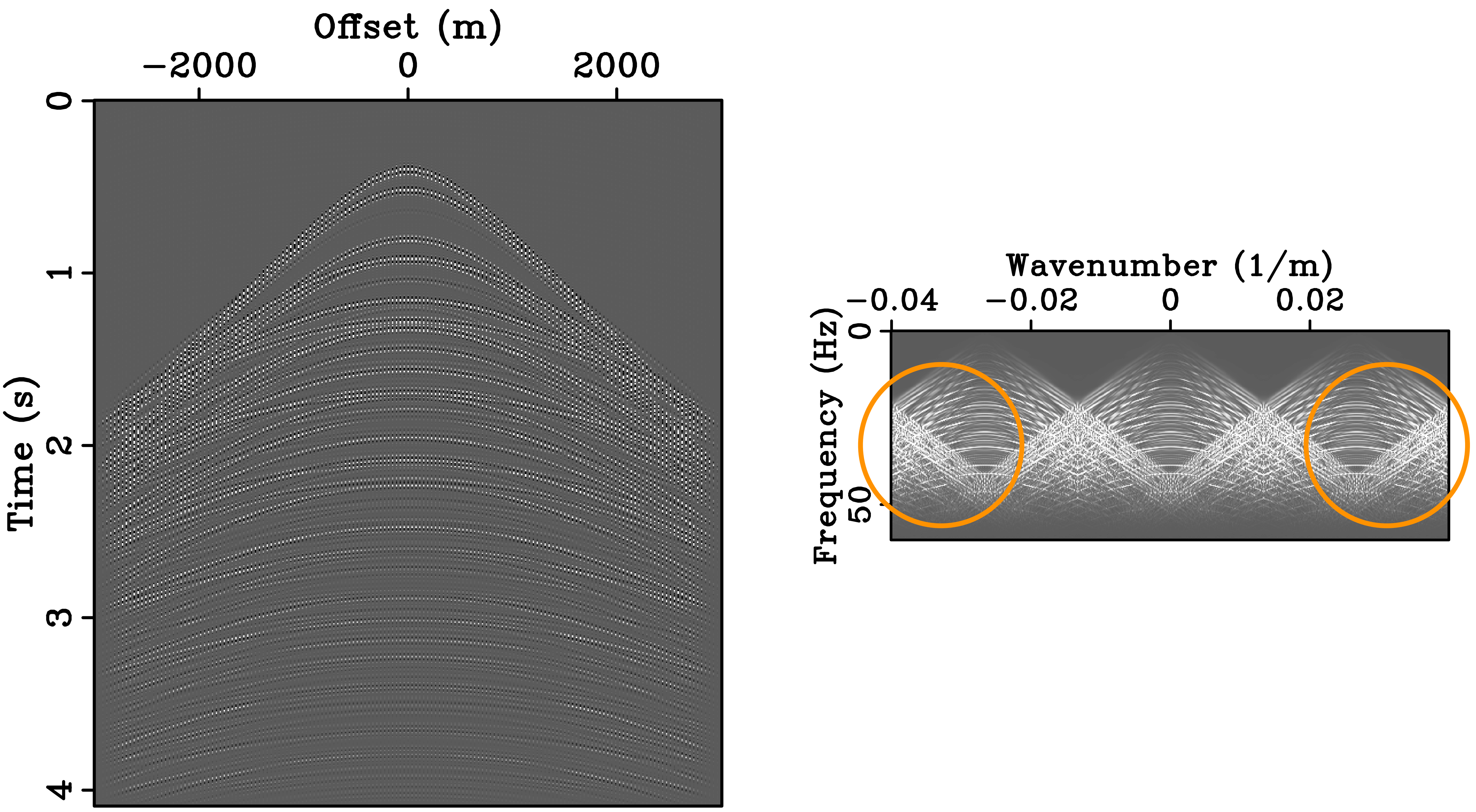
$\mathbf{S^H}$	transform domain synthesis matrix
$\mathbf{A}$	measurement matrix : $\mathbf{MS^H}$, $\mathbf{M}$ is a measurement matrix
$\mathbf{b}$	randomly sampled data
$\tilde{\mathbf{x}}$	estimated (curvelet) coefficients for recovered wavefields

Gilles Hennenfent and Felix J. Herrmann, "**Simply denoise: wavefield reconstruction via jittered undersampling**", *Geophysics*, vol. 73, p. V19-V28, 2008.

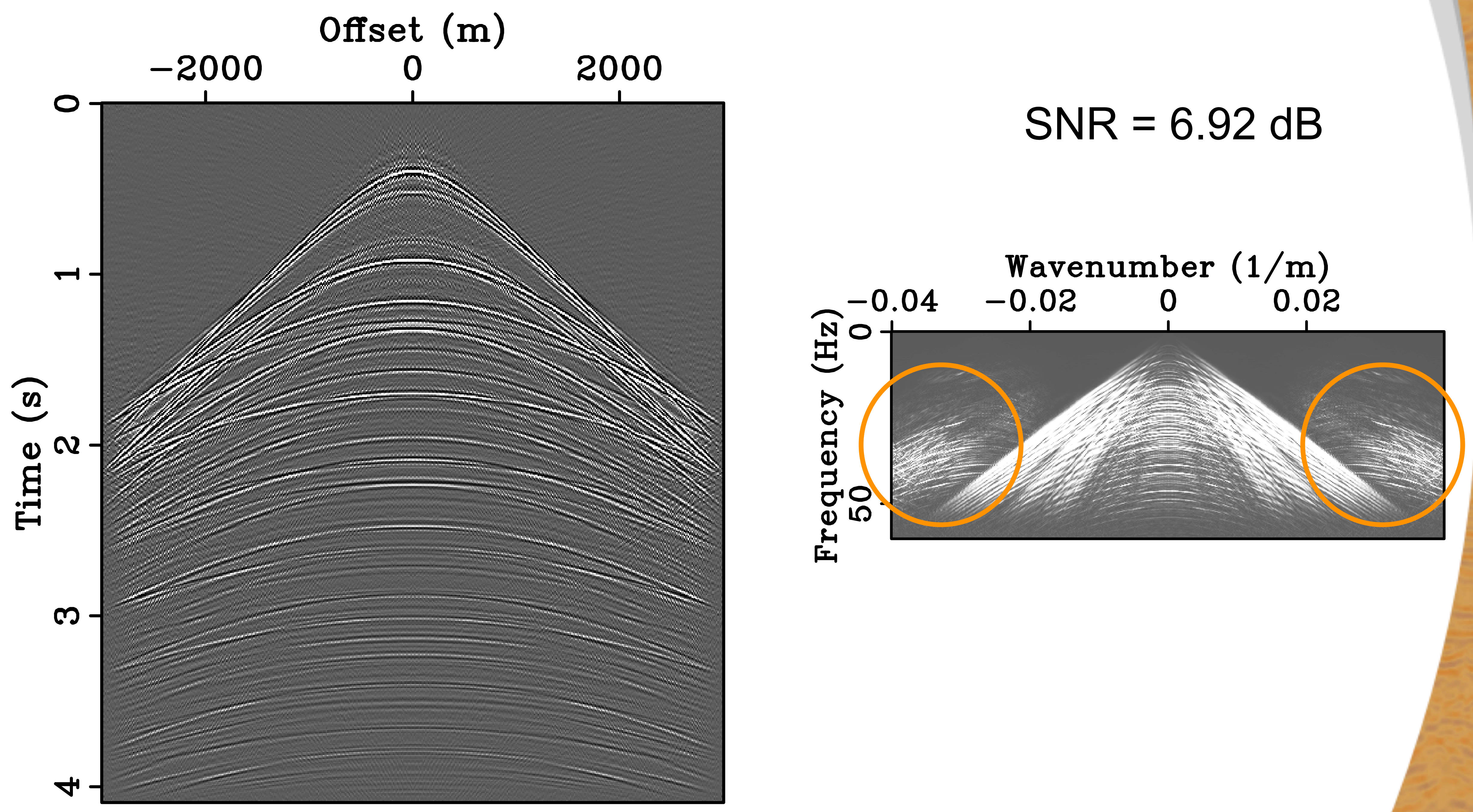
Jittered sampling



Periodic sampling

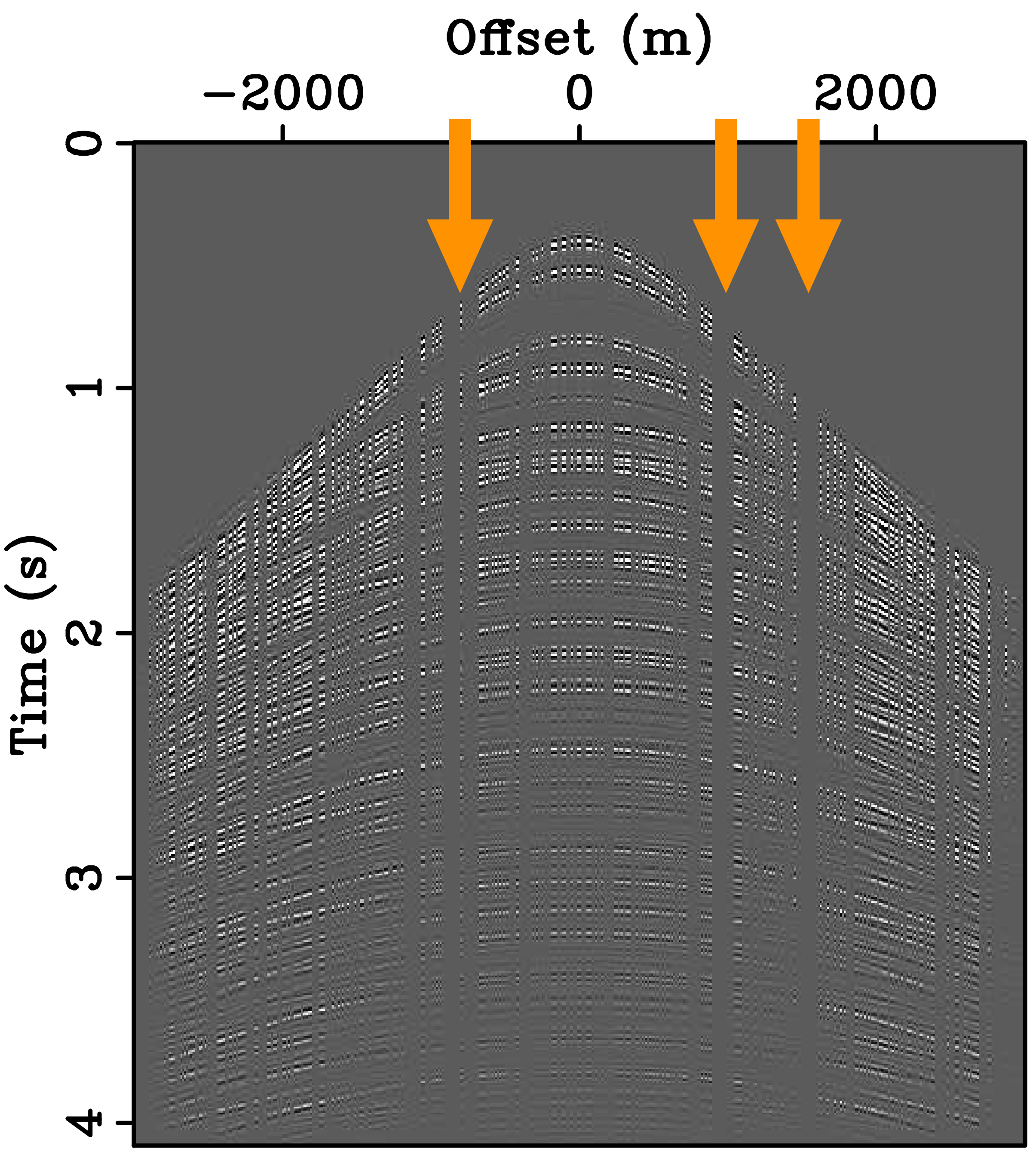


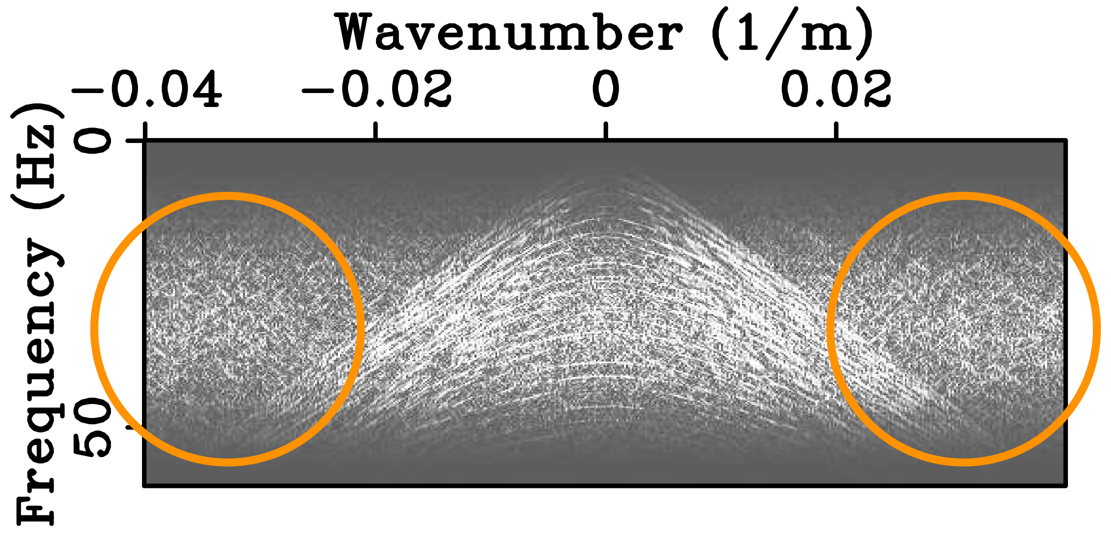
3-fold undersampled



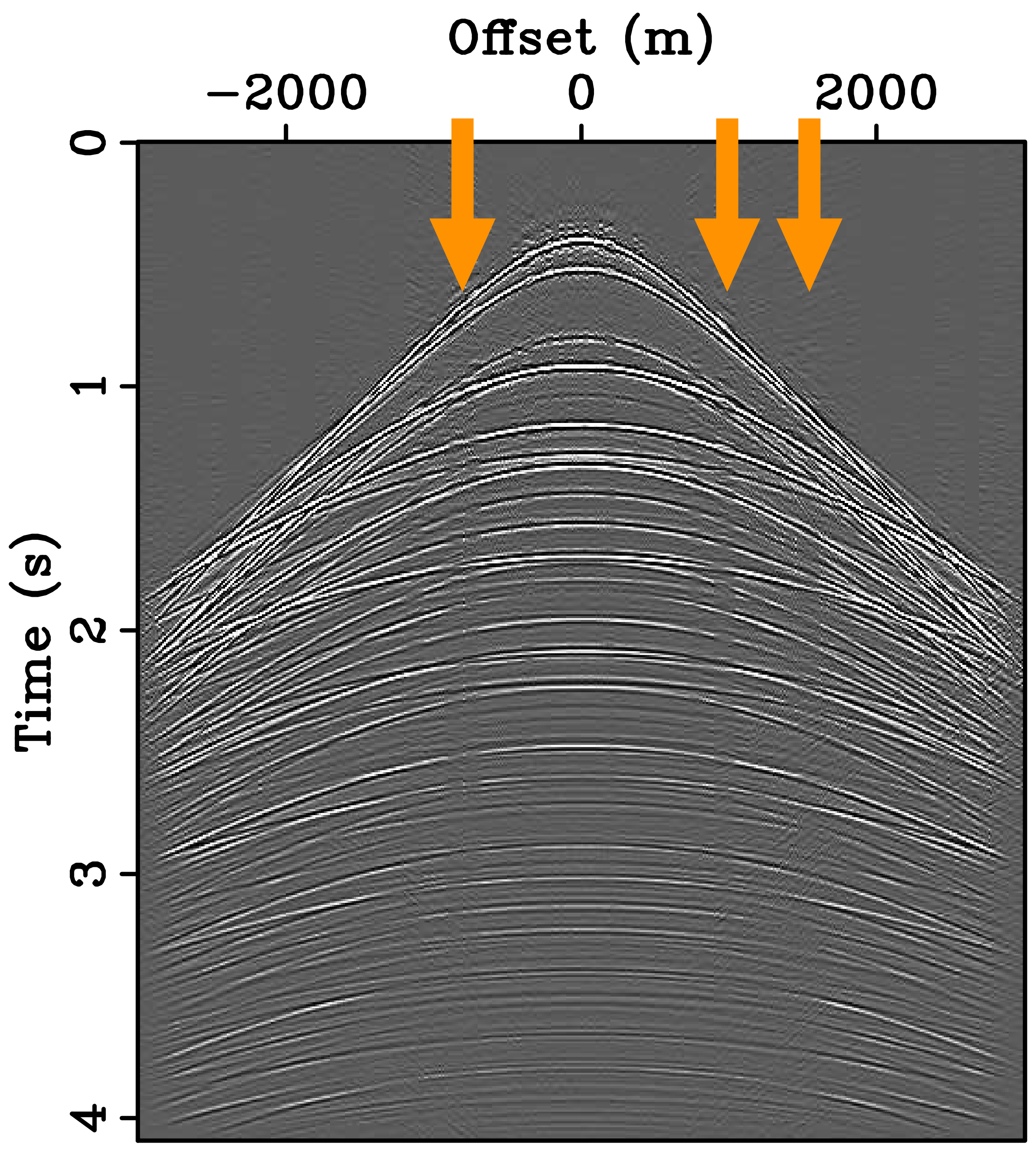
recovered

Uniform random sampling

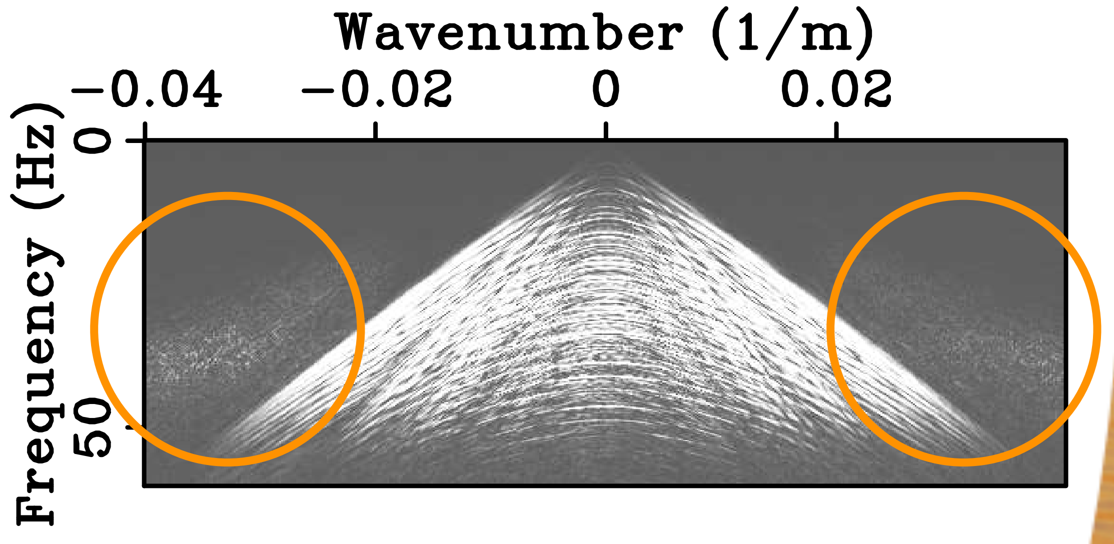




3-fold undersampled

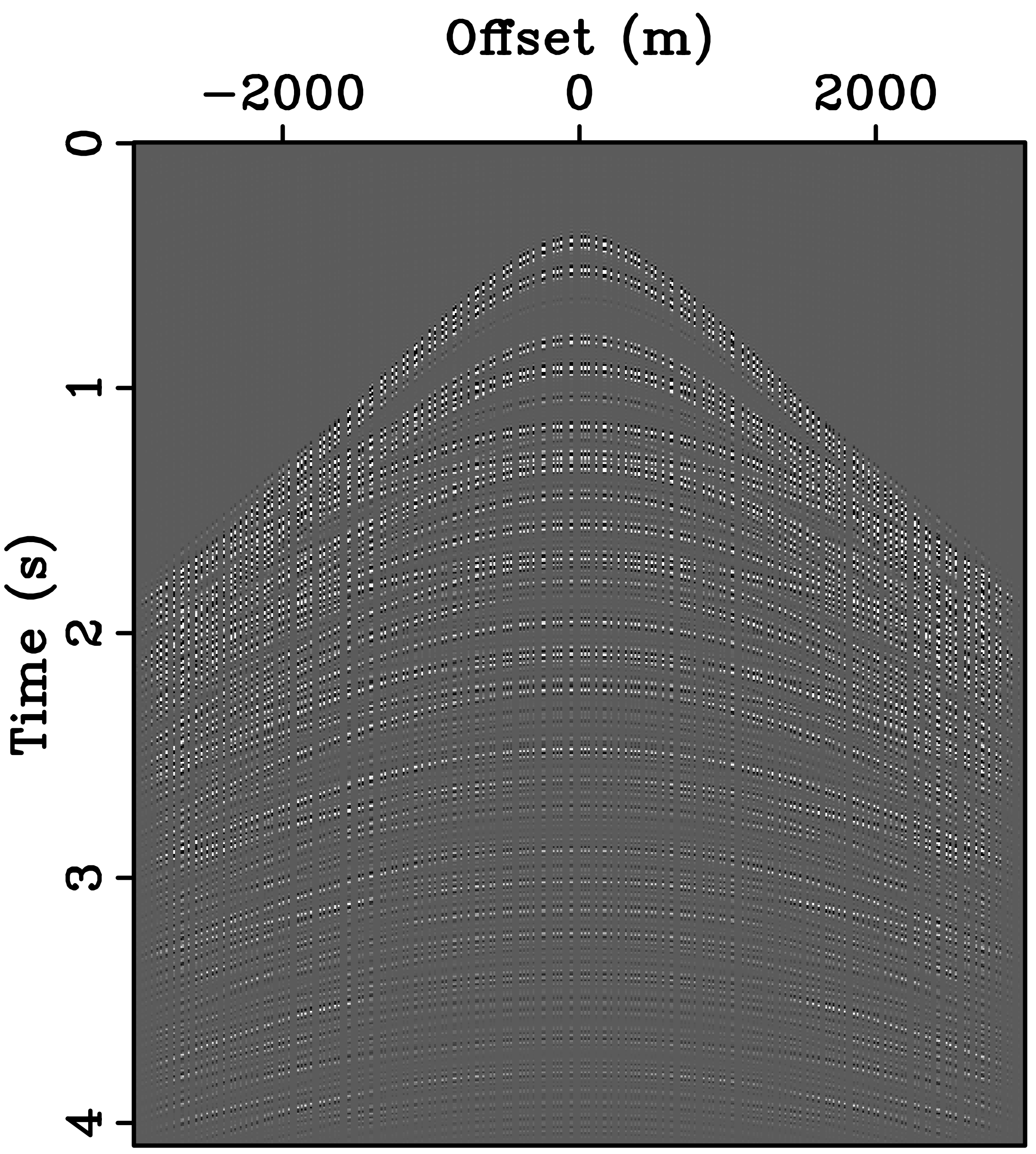


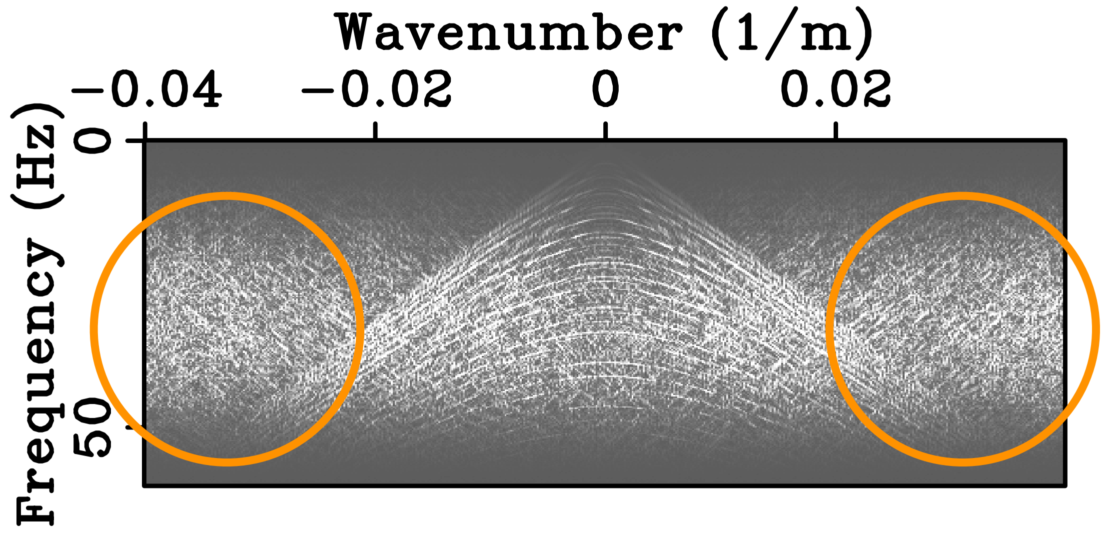
SNR = 9.72 dB



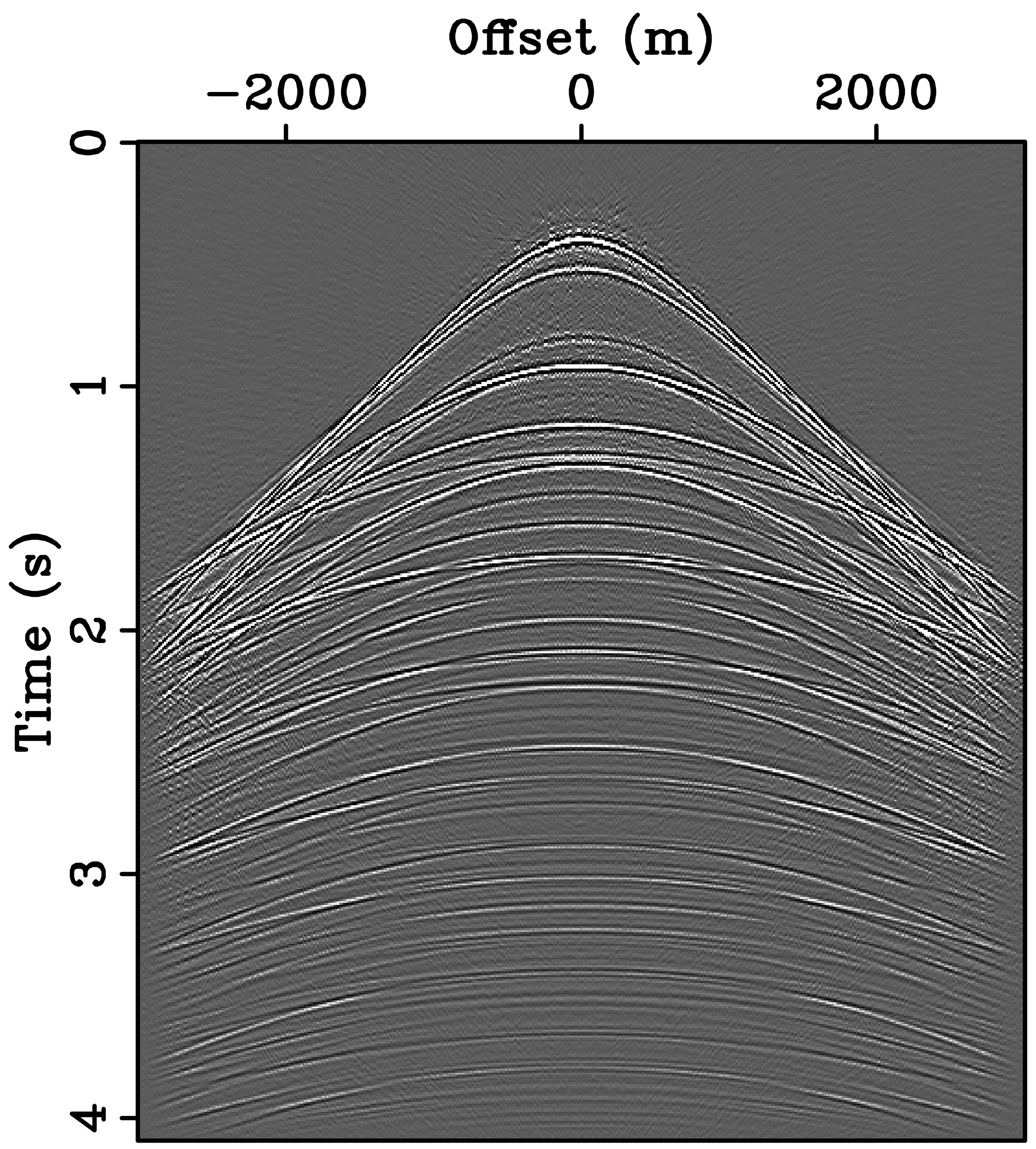
recovered

Jittered sampling

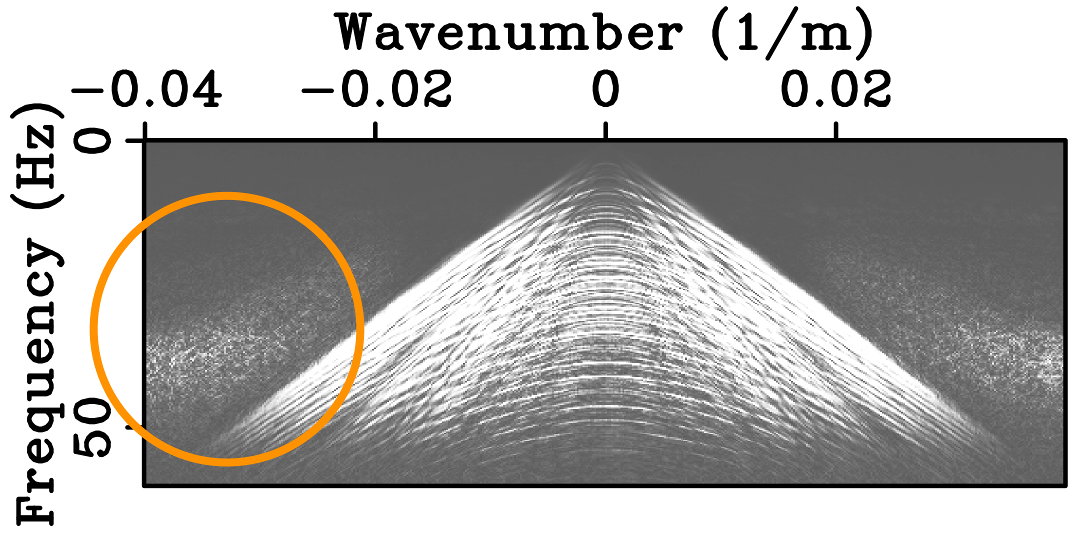




3-fold undersampled



SNR = 10.42 dB



recovered

Application to marine acquisition

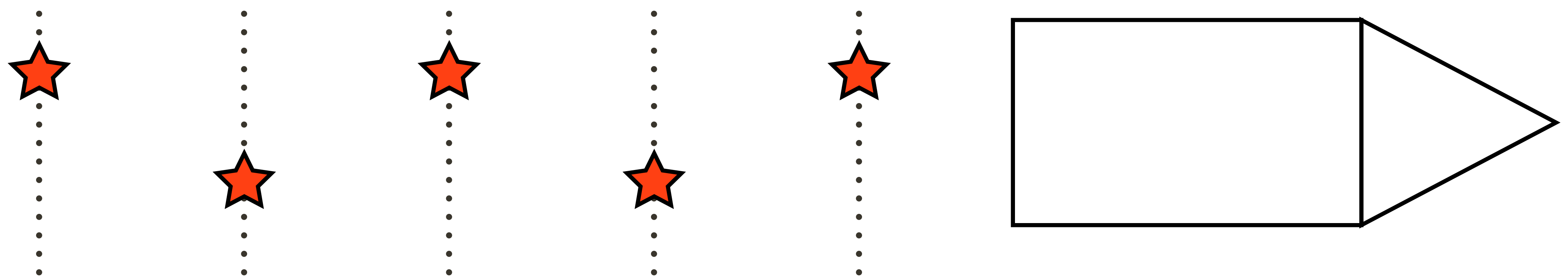
Haneet Wason & Rajiv Kumar


SLIM

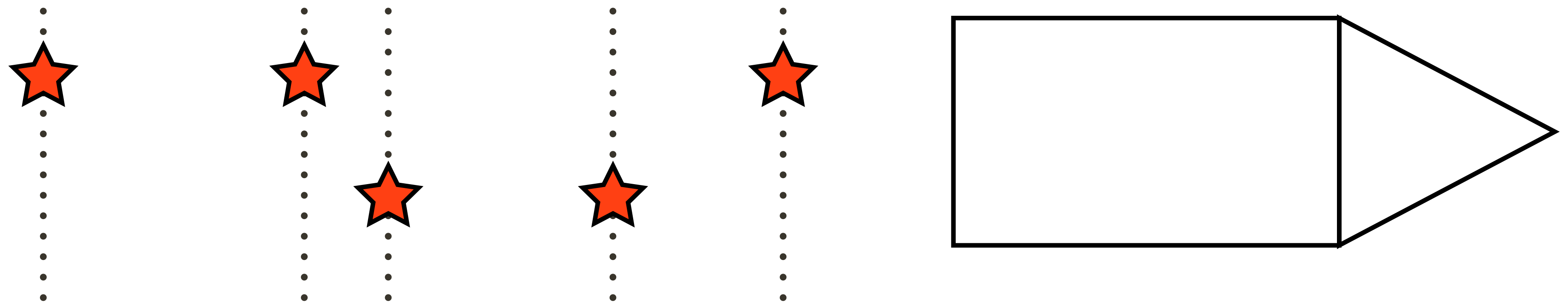
University of British Columbia

Periodic vs. jittered in marine
– continuous recording w/ OBC/OBN

regular **periodically** sampled spatial grid



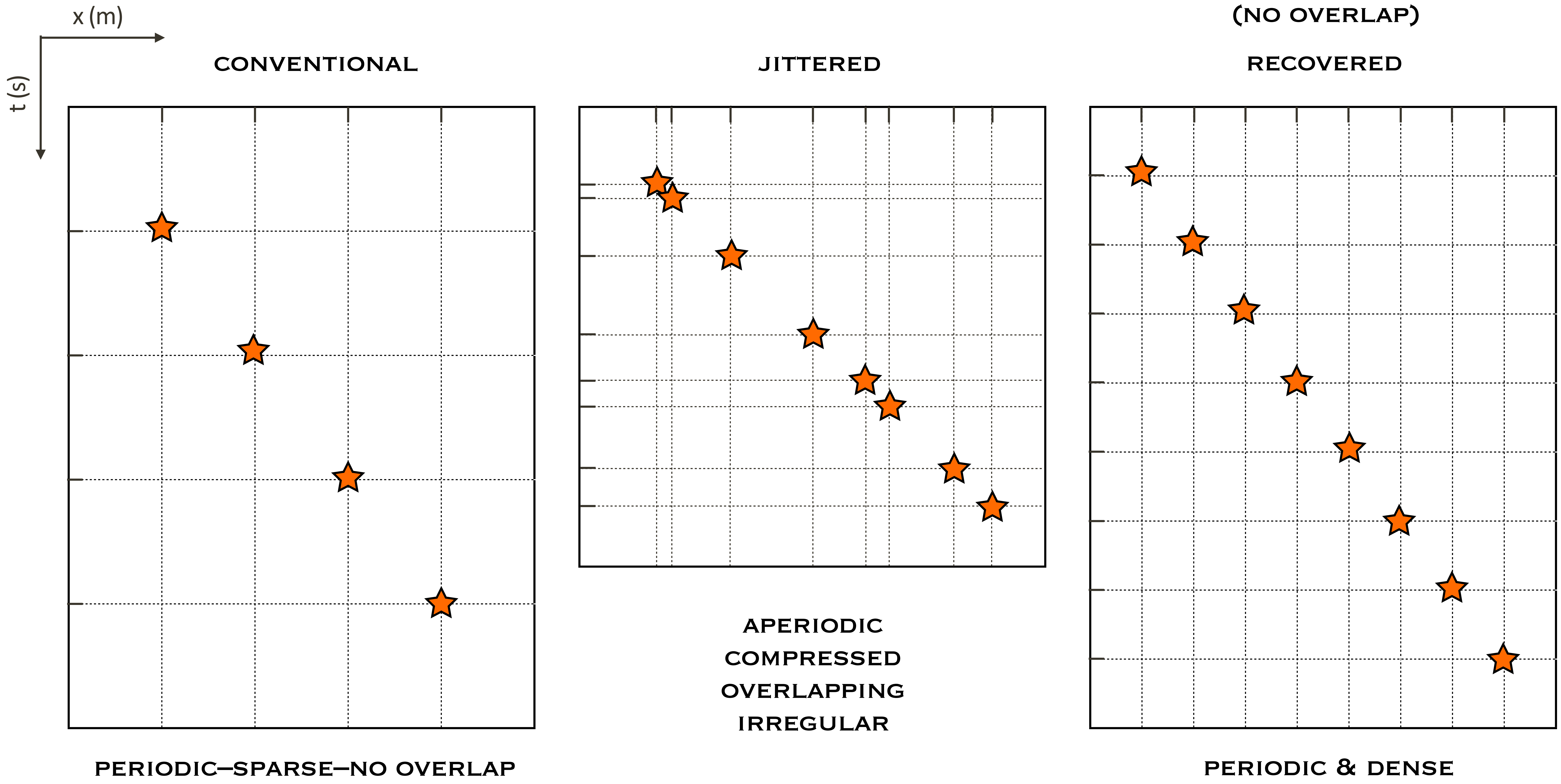
randomly jittered sampled spatial grid



Haneet Wason and Felix J. Herrmann, "**Time-jittered ocean bottom seismic acquisition**", *SEG*, 2013
Hassan Mansour, Haneet Wason, Tim T.Y. Lin, and Felix J. Herrmann, "**Randomized marine acquisition with compressive sampling matrices**", *Geophysical Prospecting*, vol. 60, p. 648-662, 2012

Randomized jitter sampling in marine

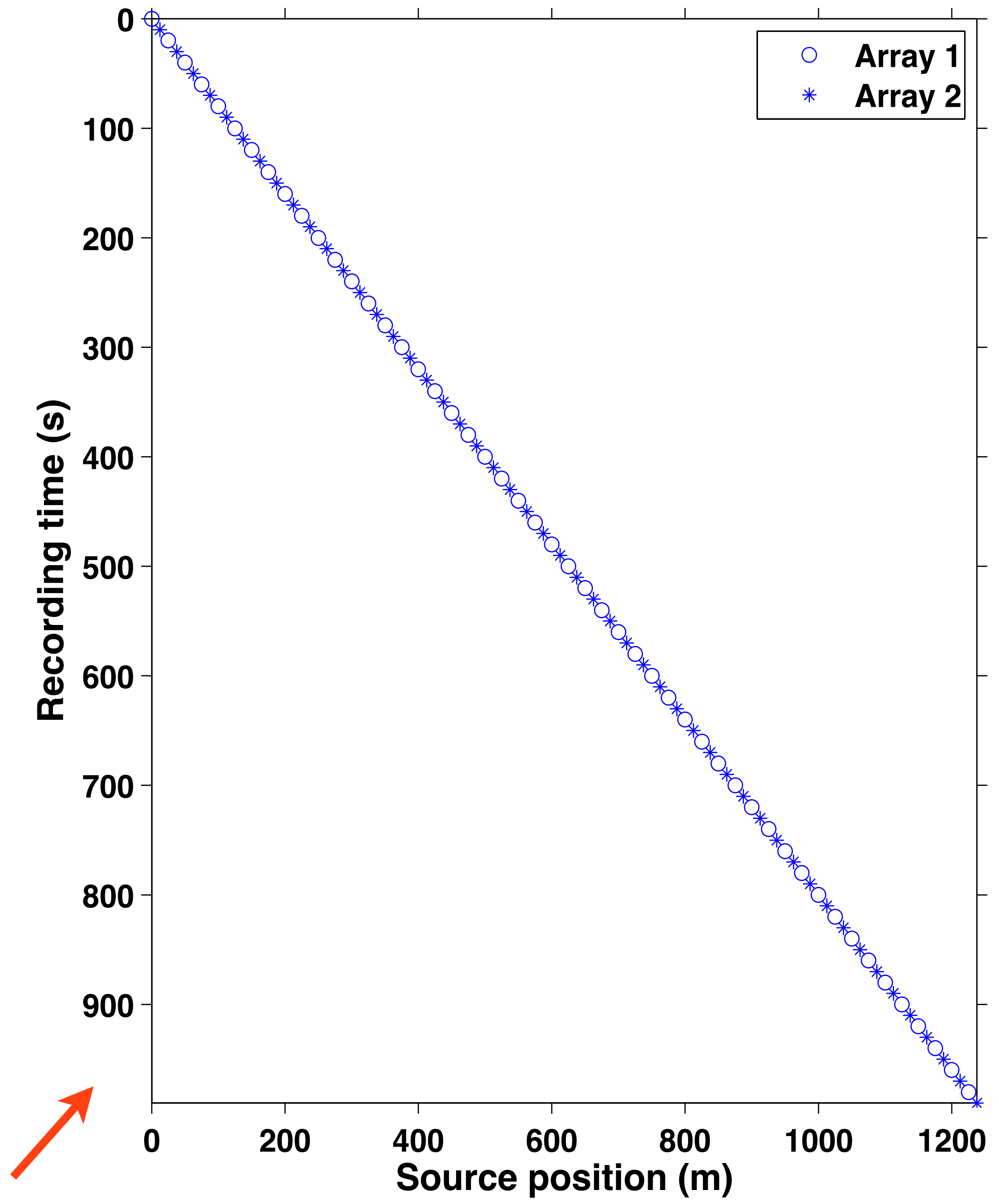
– continuous recording w/ OBC/OBN



Conventional vs. time-jittered sources

– subsampling ratio = 2, 2 source arrays

conventional

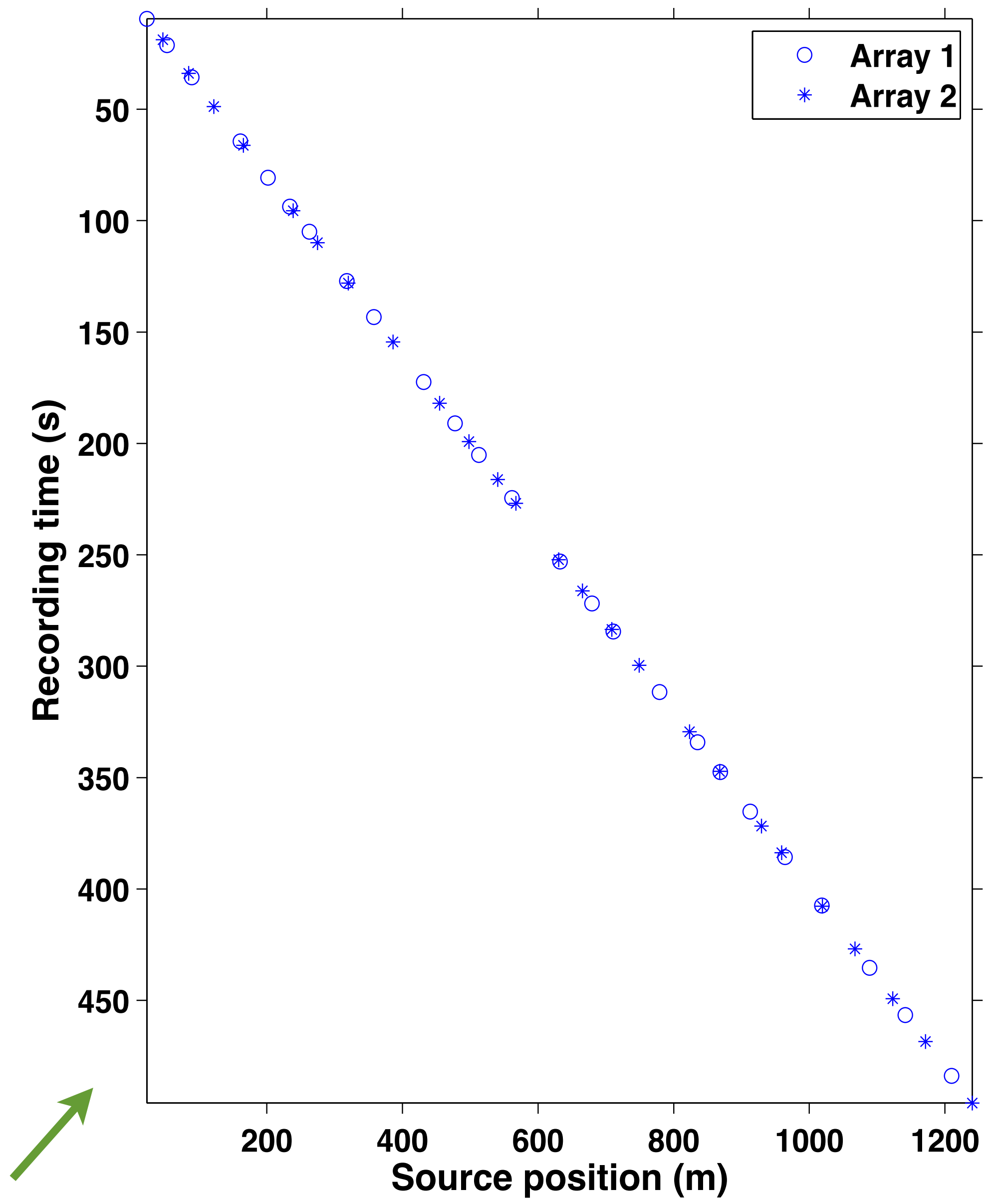


"unblended" **shot gathers**
number of shots = **100** (per array)
shot record length: 10.0 s
spatial sampling: **12.5 m**
vessel speed: **1.25 m/s**
recording time = 100 x 10.0 = **1000.0 s**

[BLENDING & SUBSAMPLING]
spatial subsampling factor = 2

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

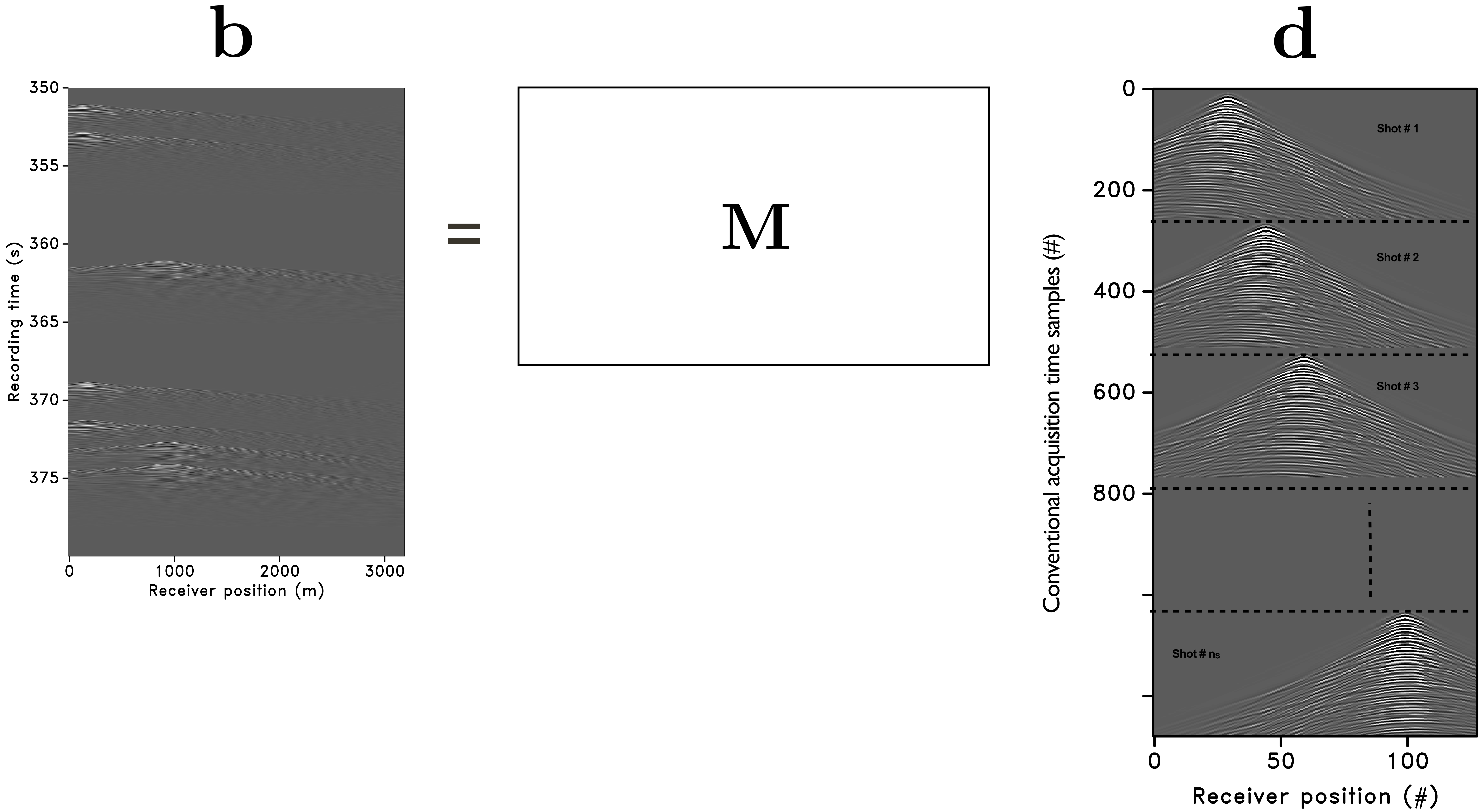
jittered acquisition



"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**

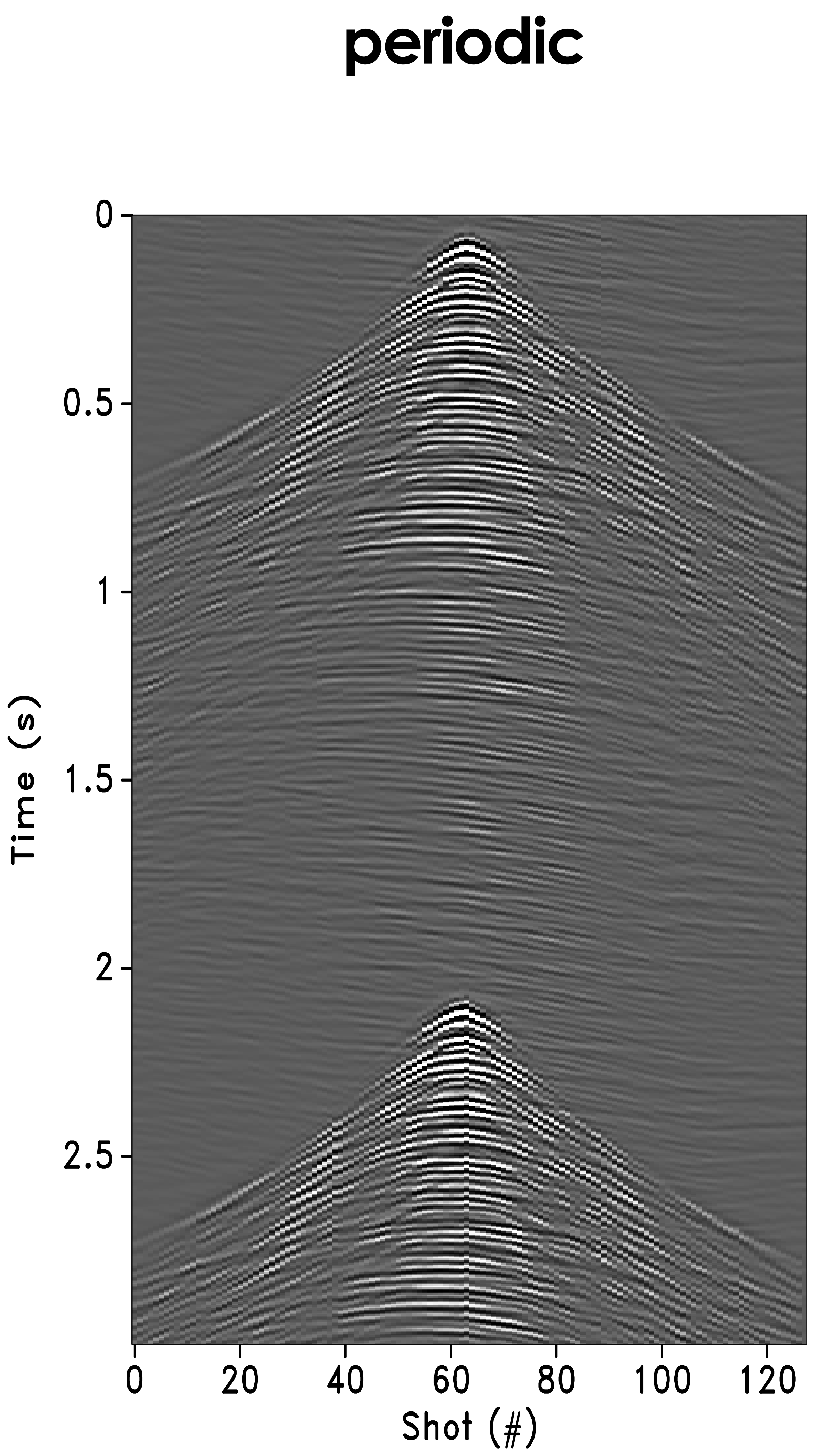
acquire in the field on *irregular* grid (*subsampled* shots *w/ overlap* between shot records)

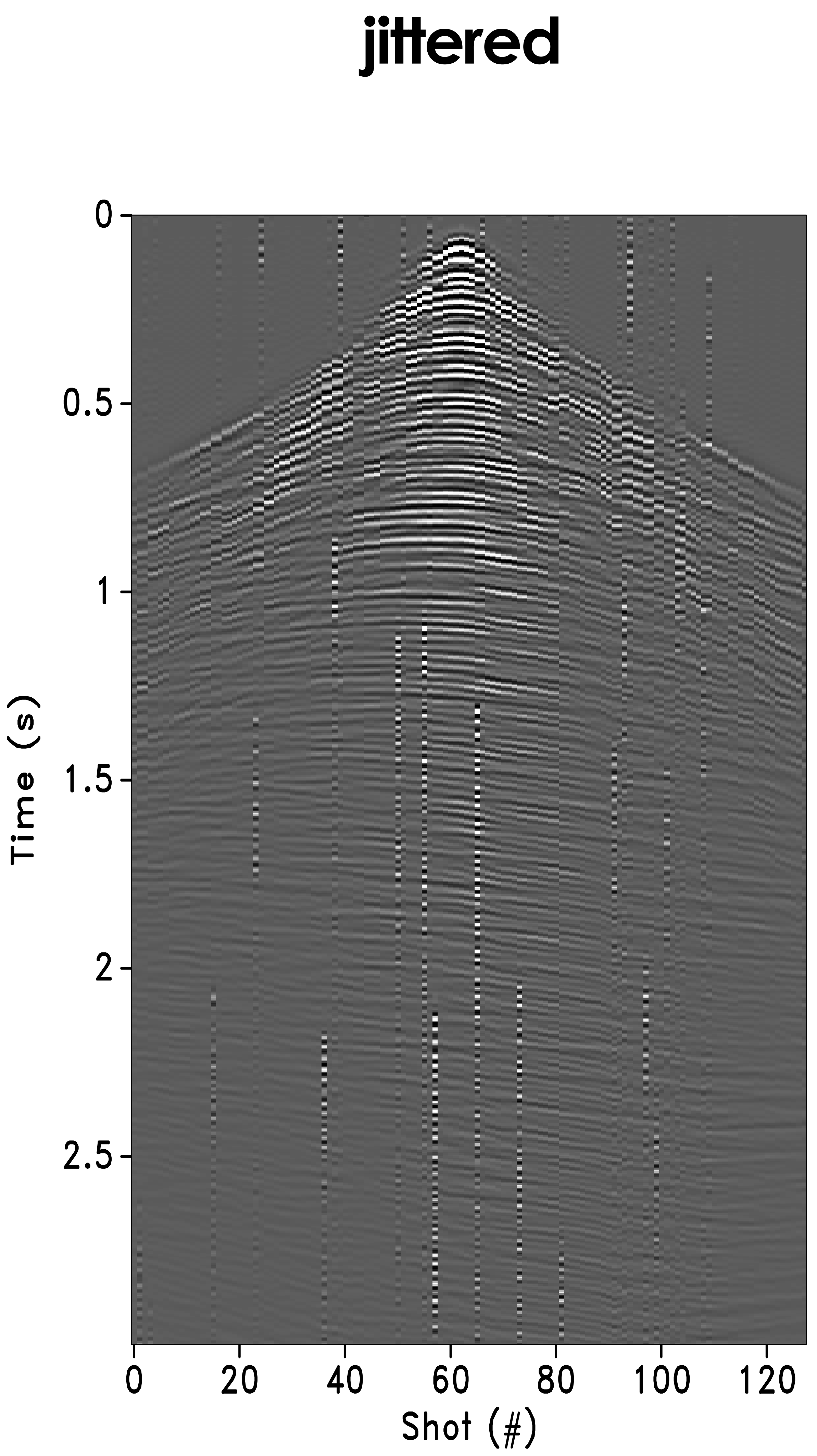
would like to have on *regular* grid (*all* shots *w/o overlaps* between shot records)



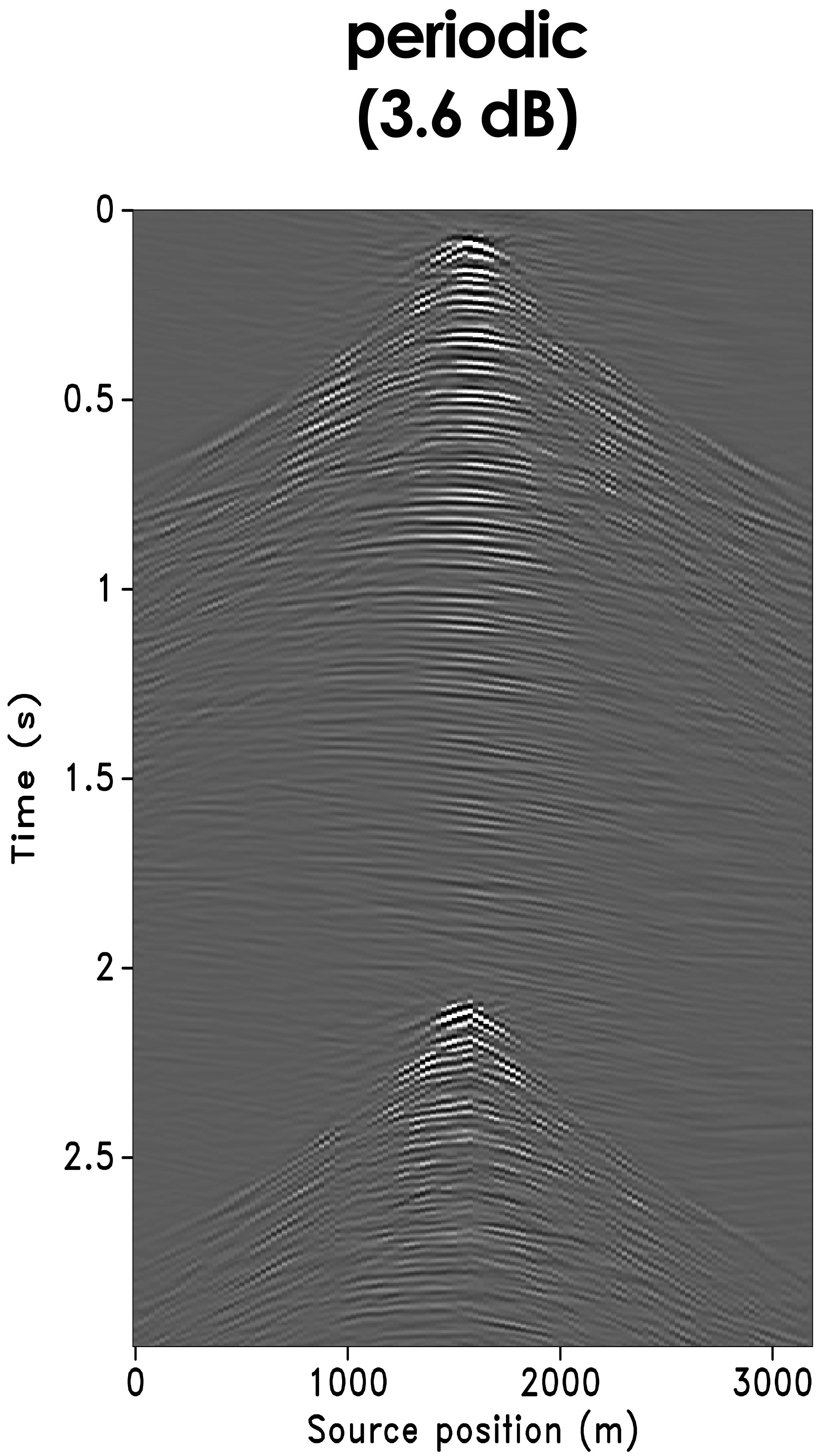
Interferences

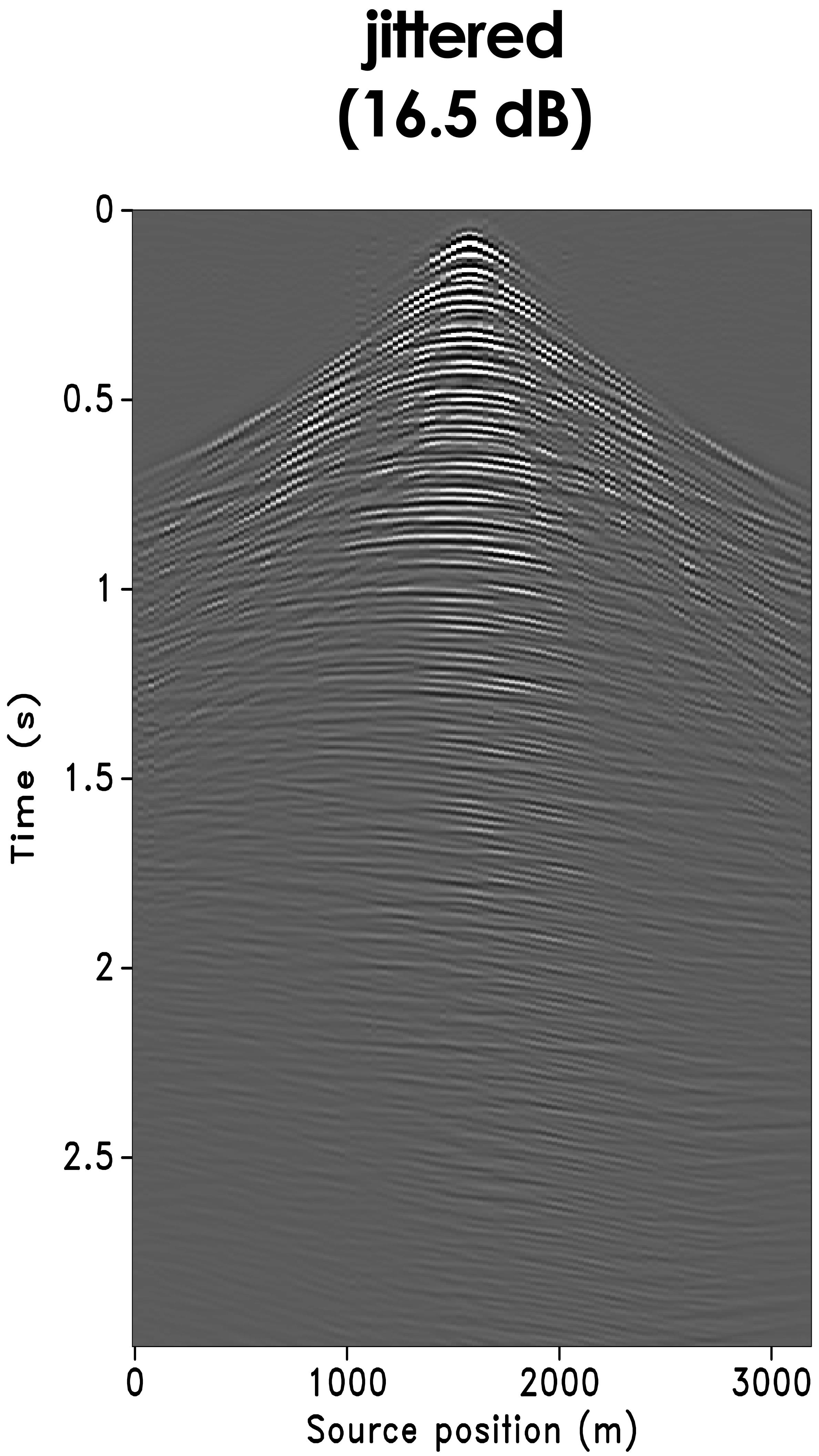
– source-crosstalk for common receiver





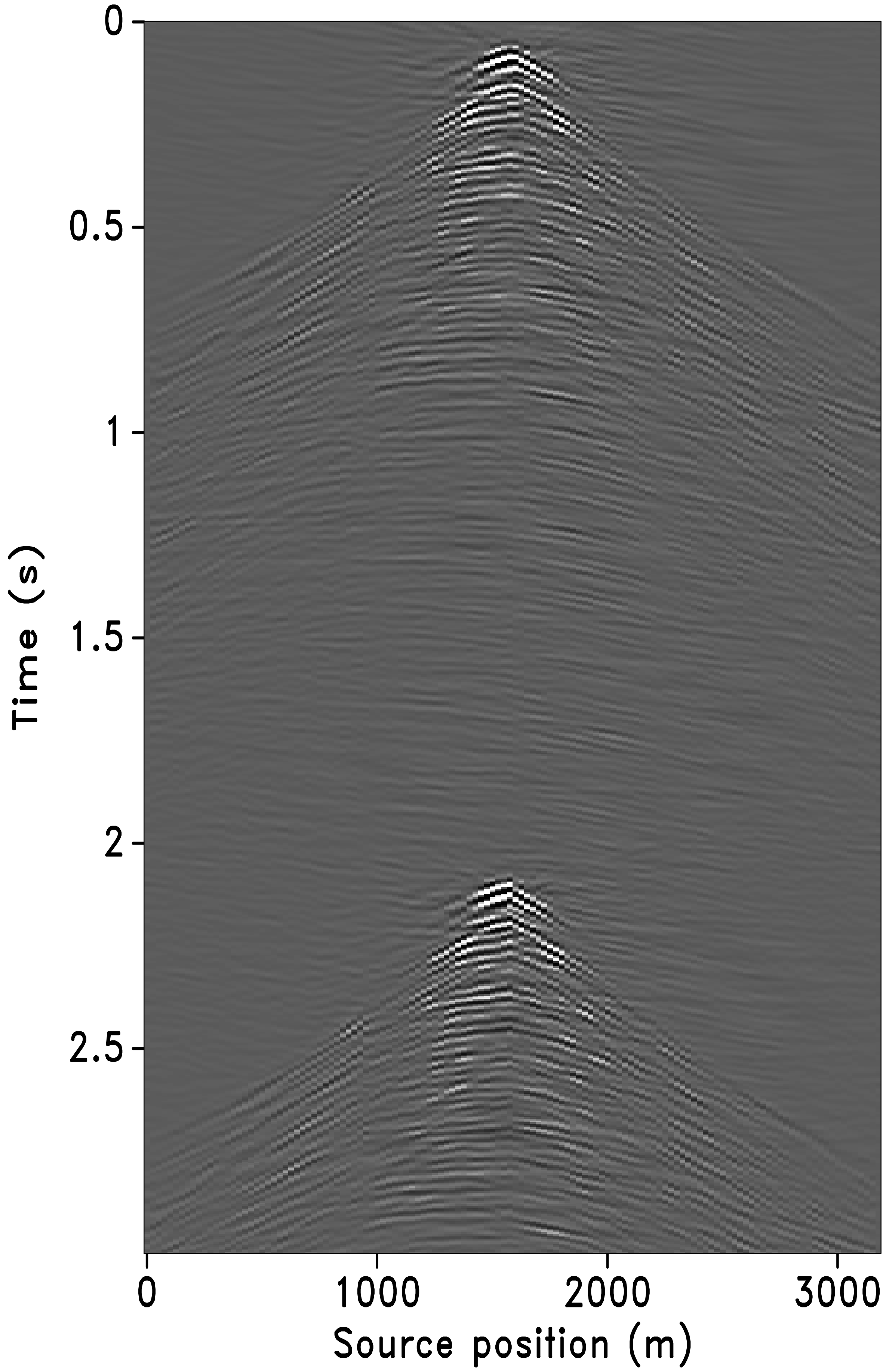
Recovery
– via sparsity promotion



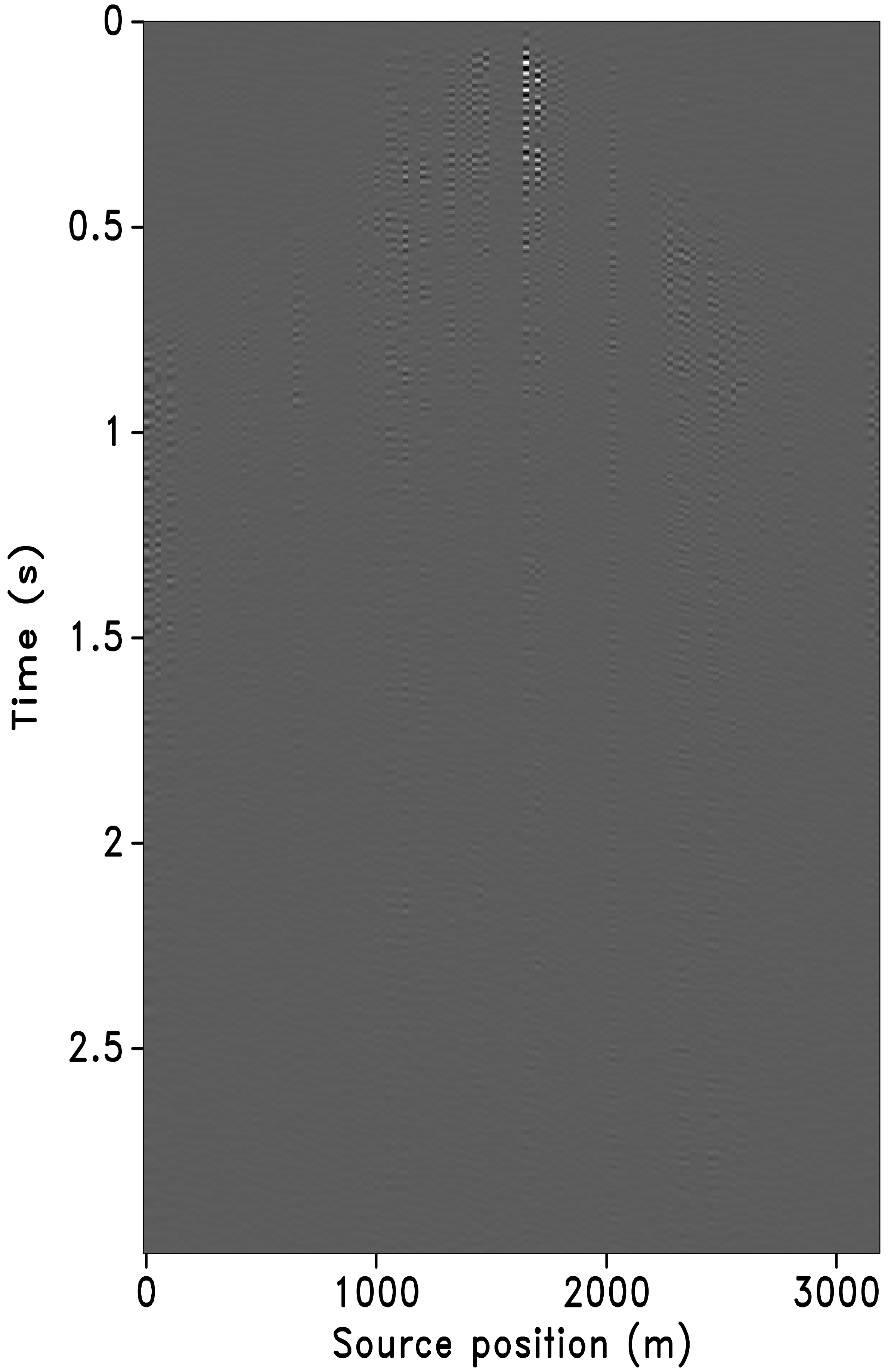


Difference

periodic

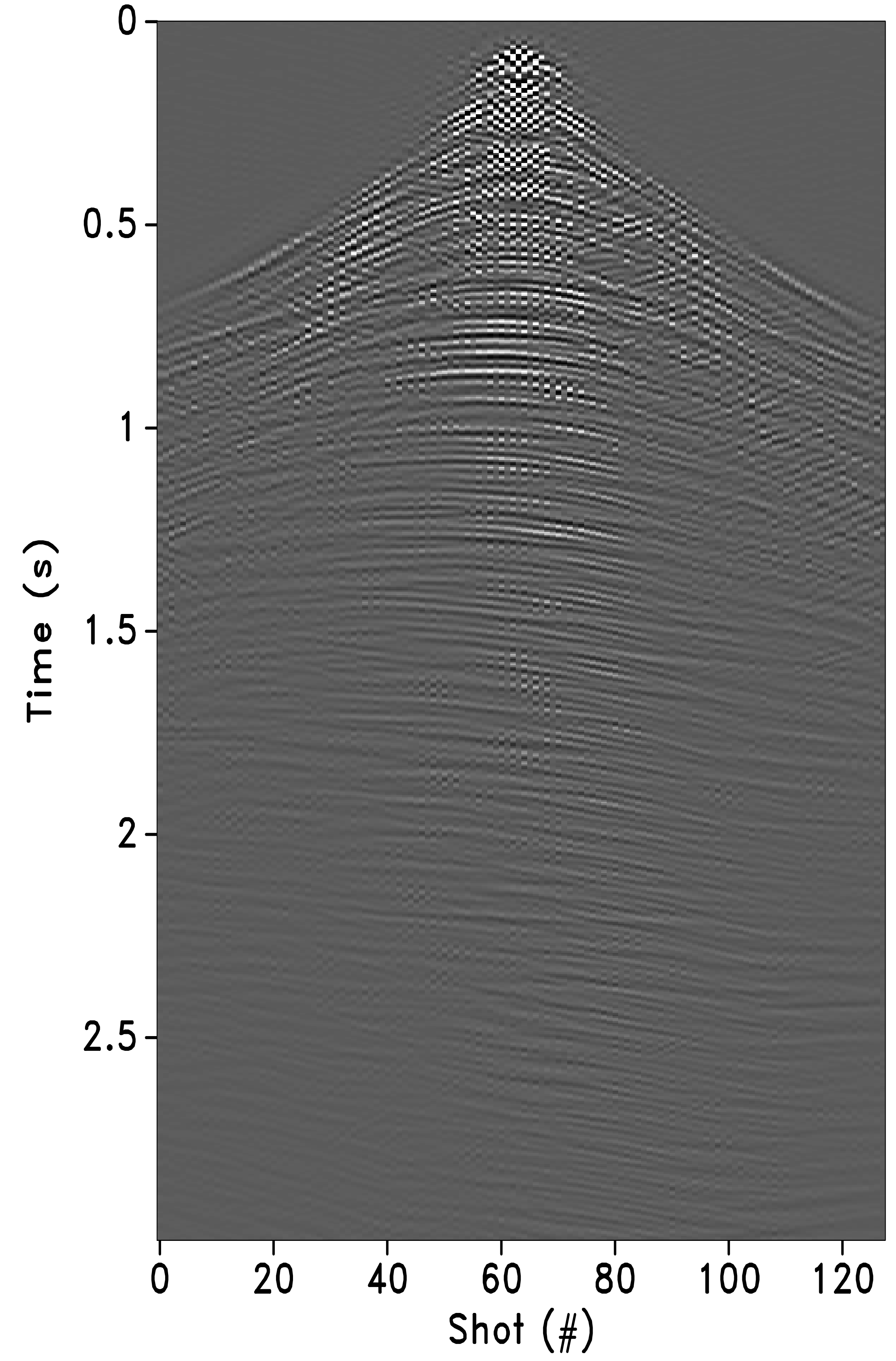


jittered

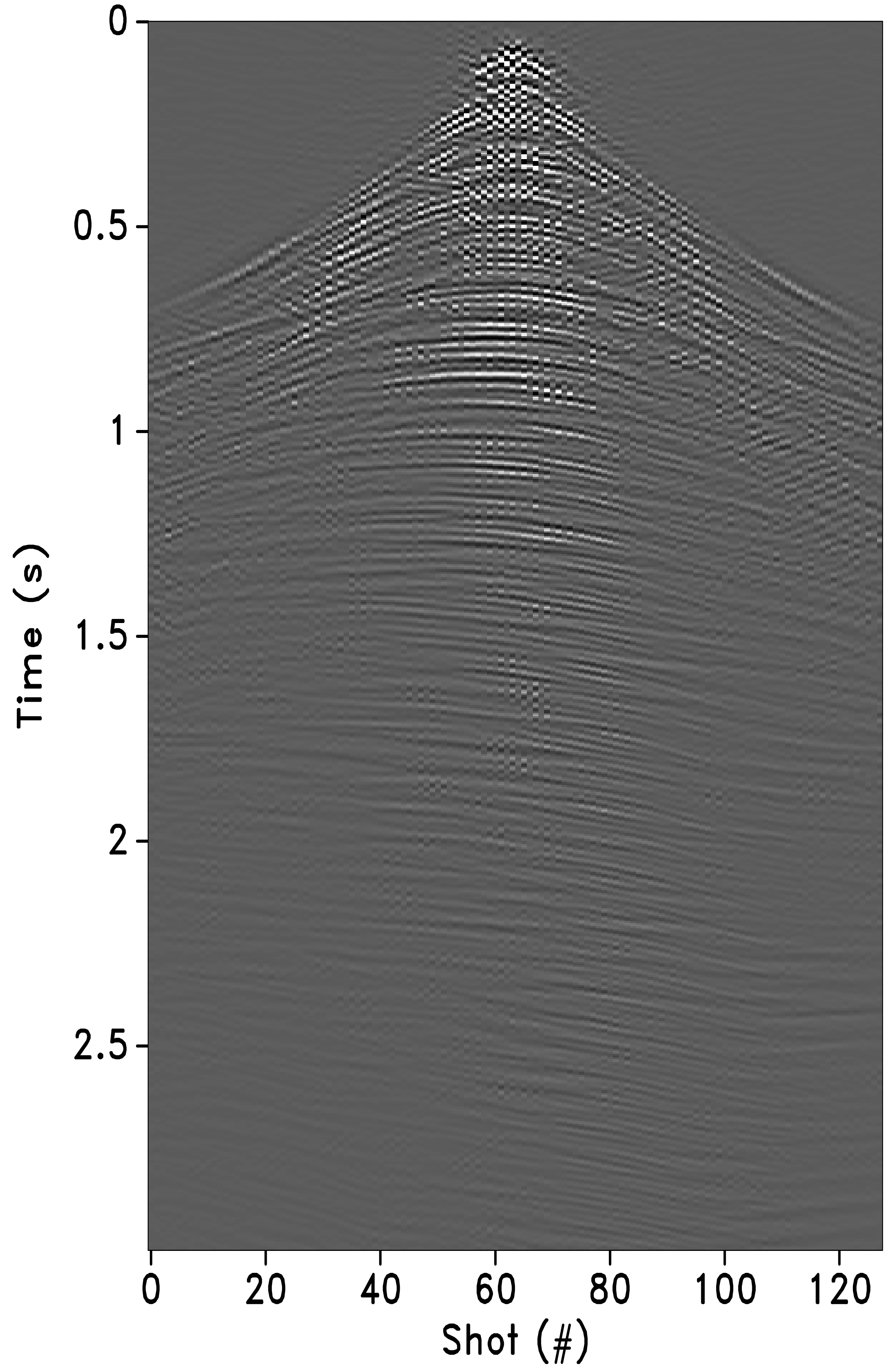


Recovery

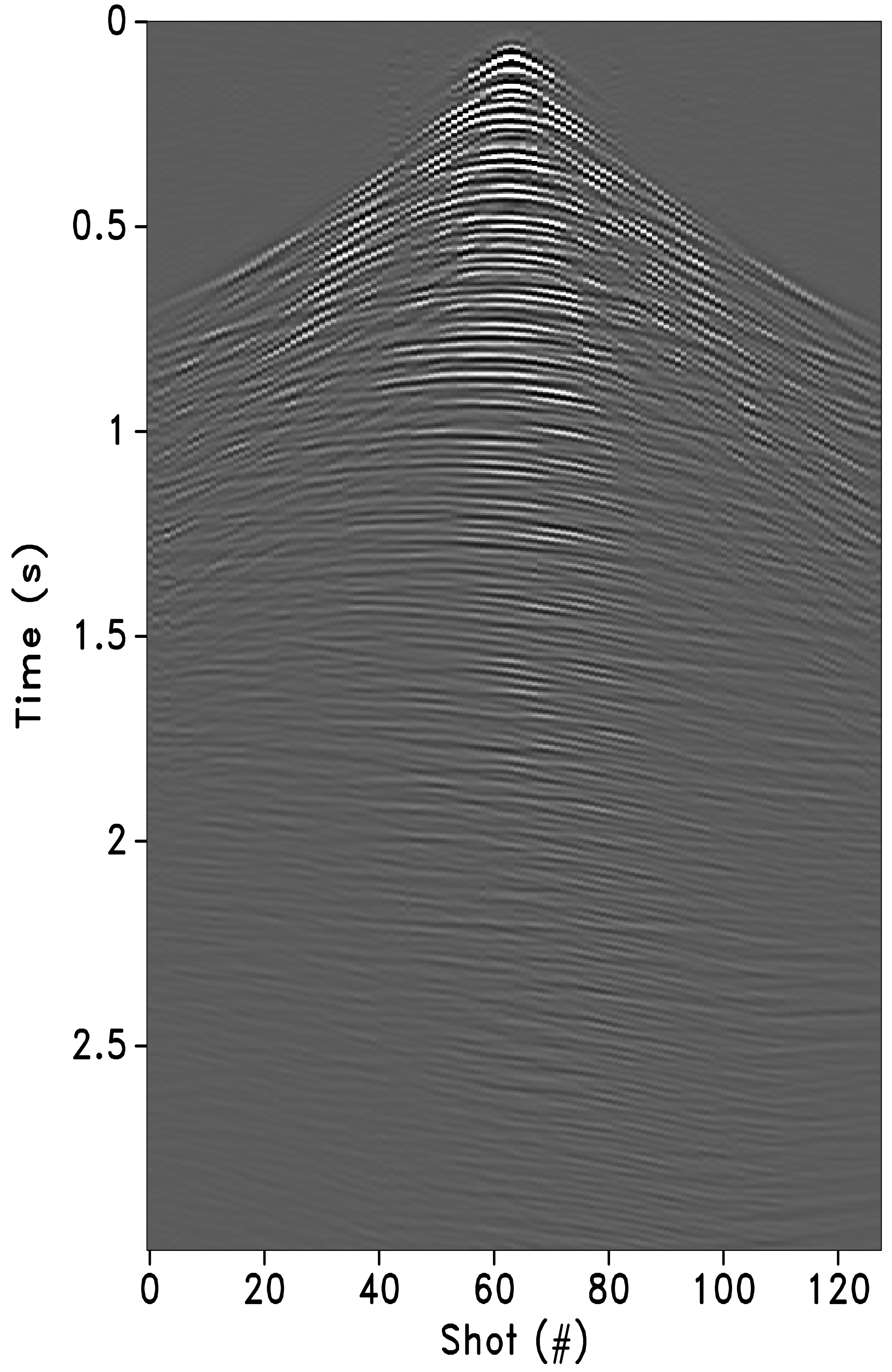
periodic



low-jitter variability



high-jitter variability



Observations

Transform-based CS works well for large variability

- limited to static geometries such as OBC / OBN

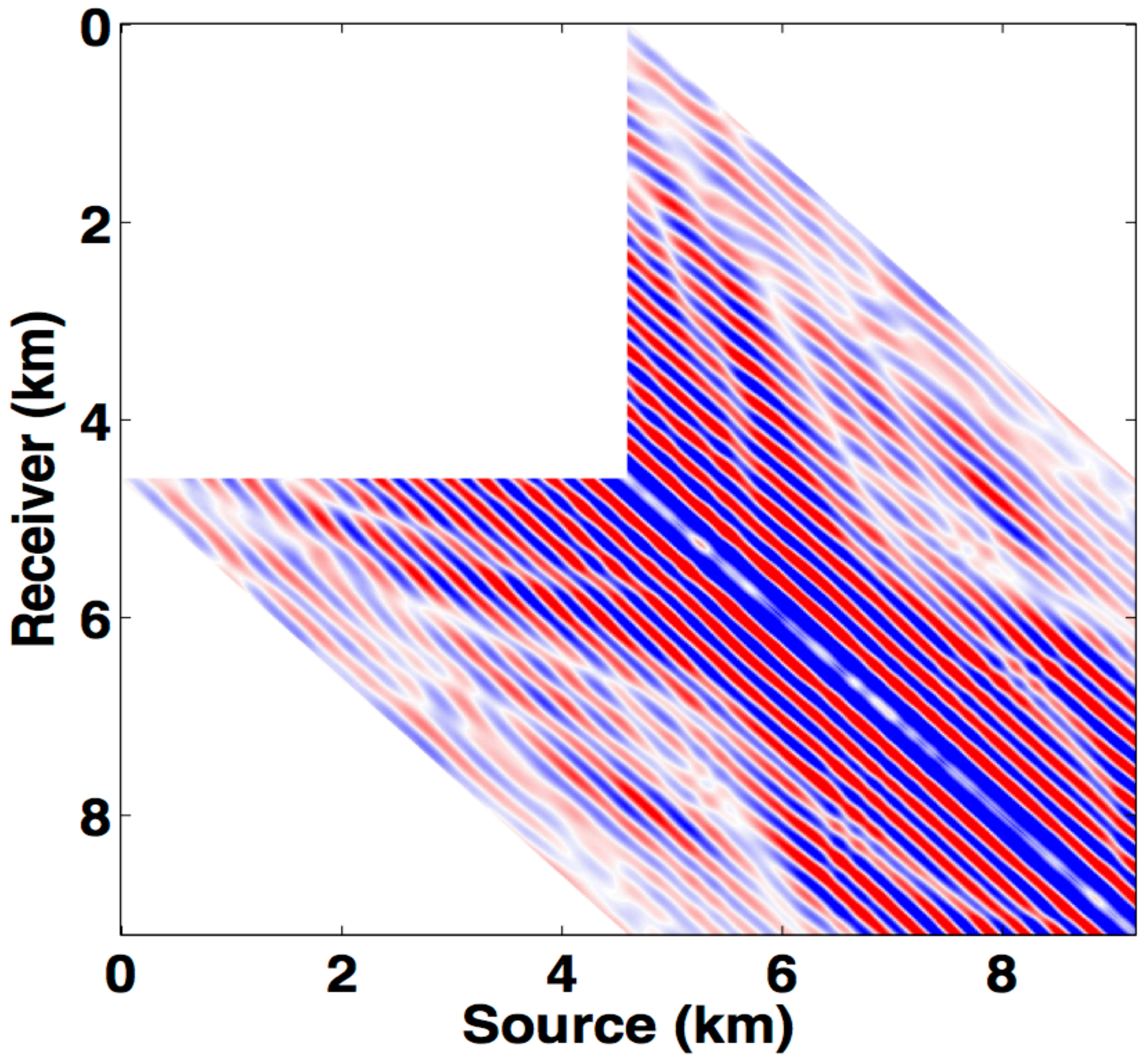
Can we relax requirement of large variability?

- enabler for dynamic geometries such as towed arrays
- over-under w/ random delays < 1S
- shot-by-shot source-separation

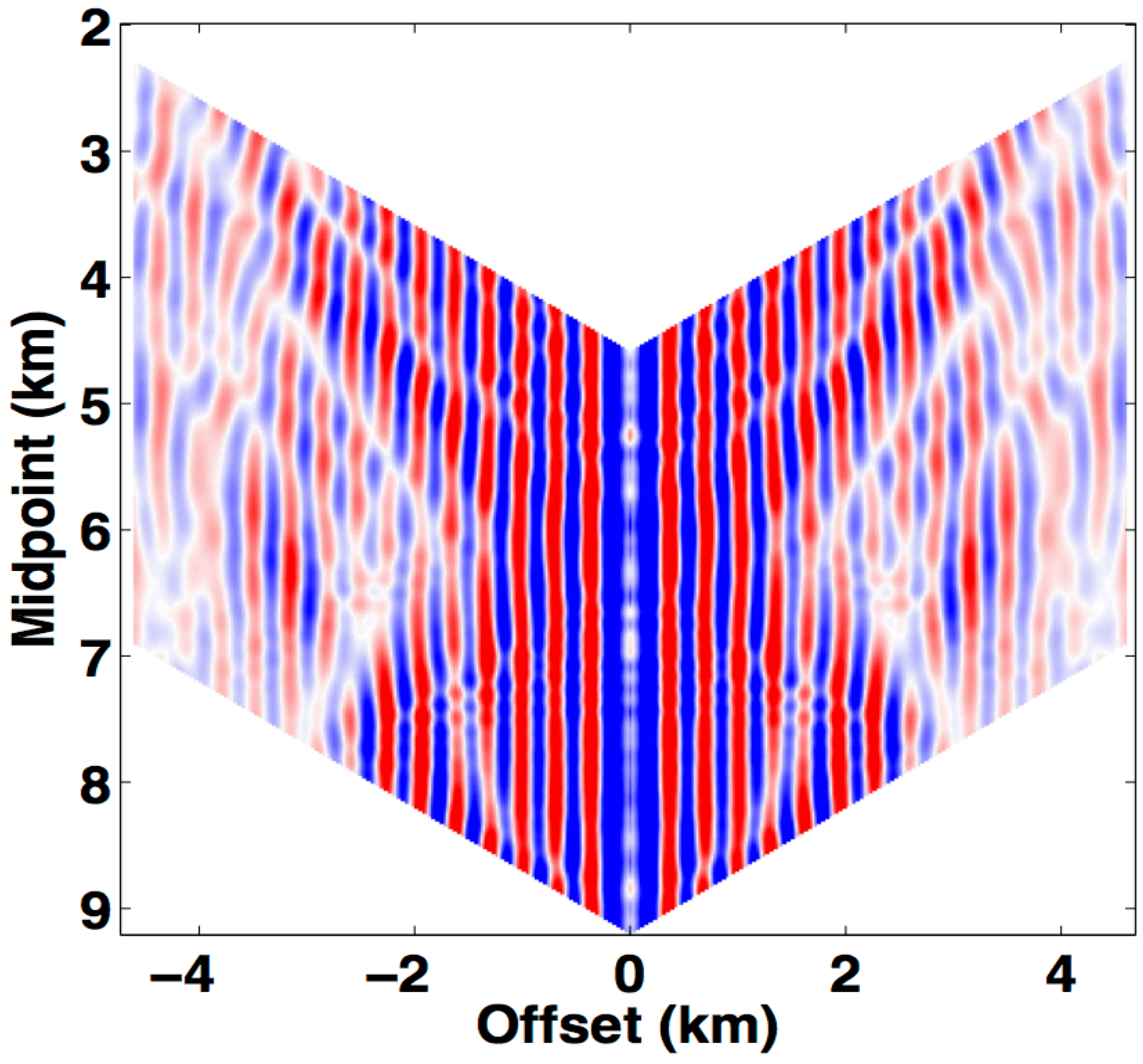
Low-rank structure in which domain?

- frequency slice at 5 Hz / over-under acquisition

source-receiver domain
(w/ reciprocity)

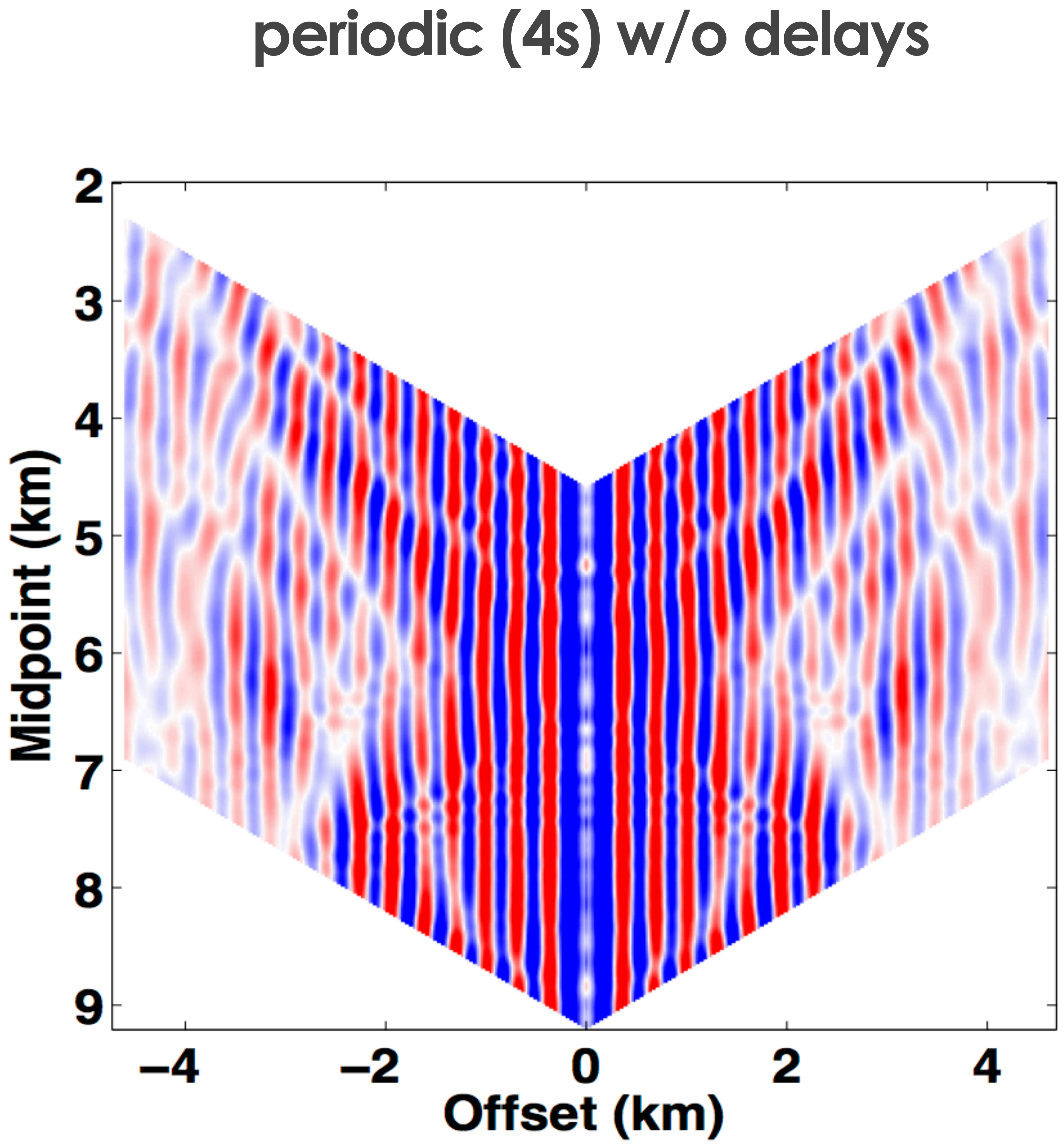


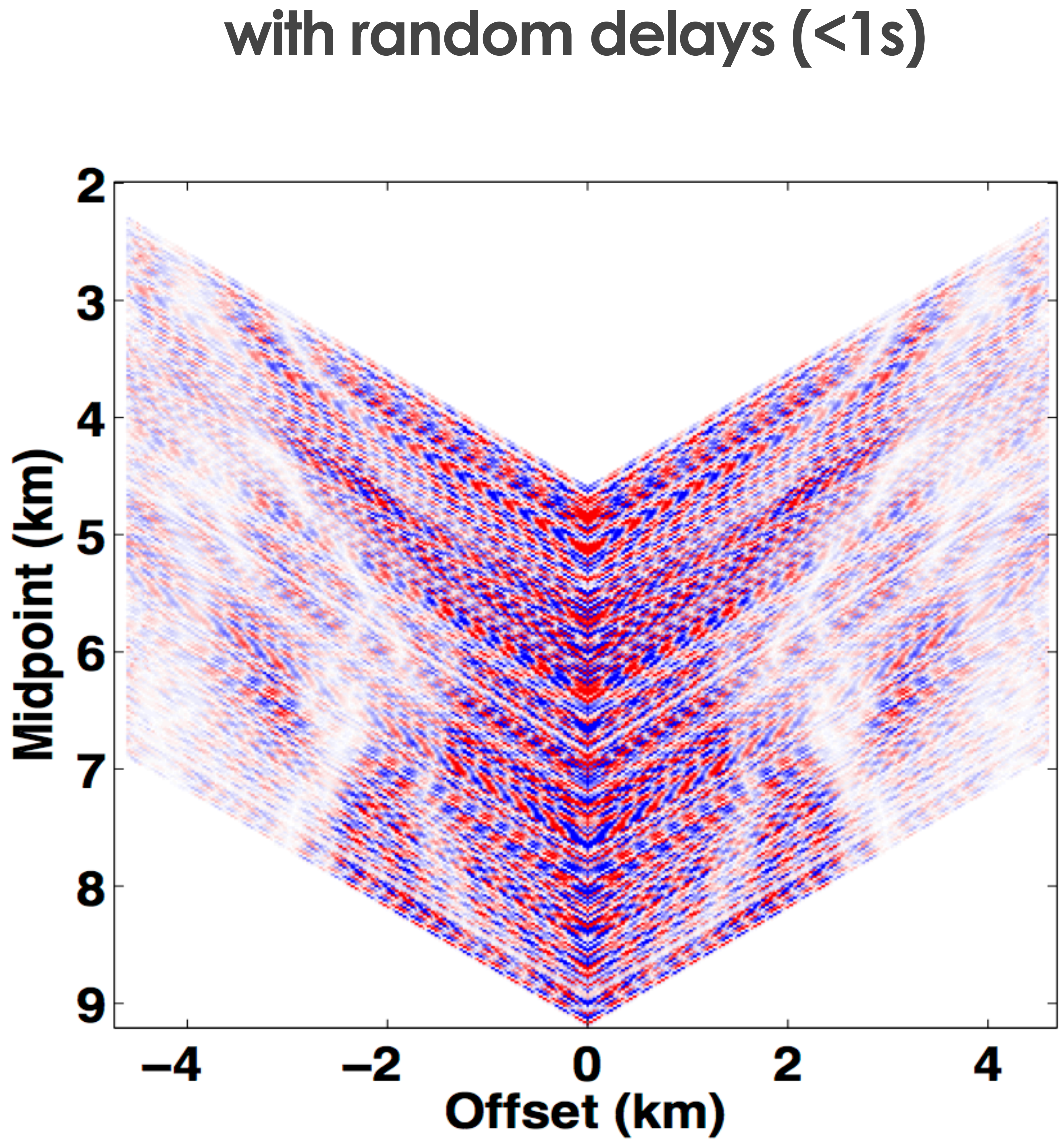
midpoint-offset domain
(w/ reciprocity)



How to destroy the structure?

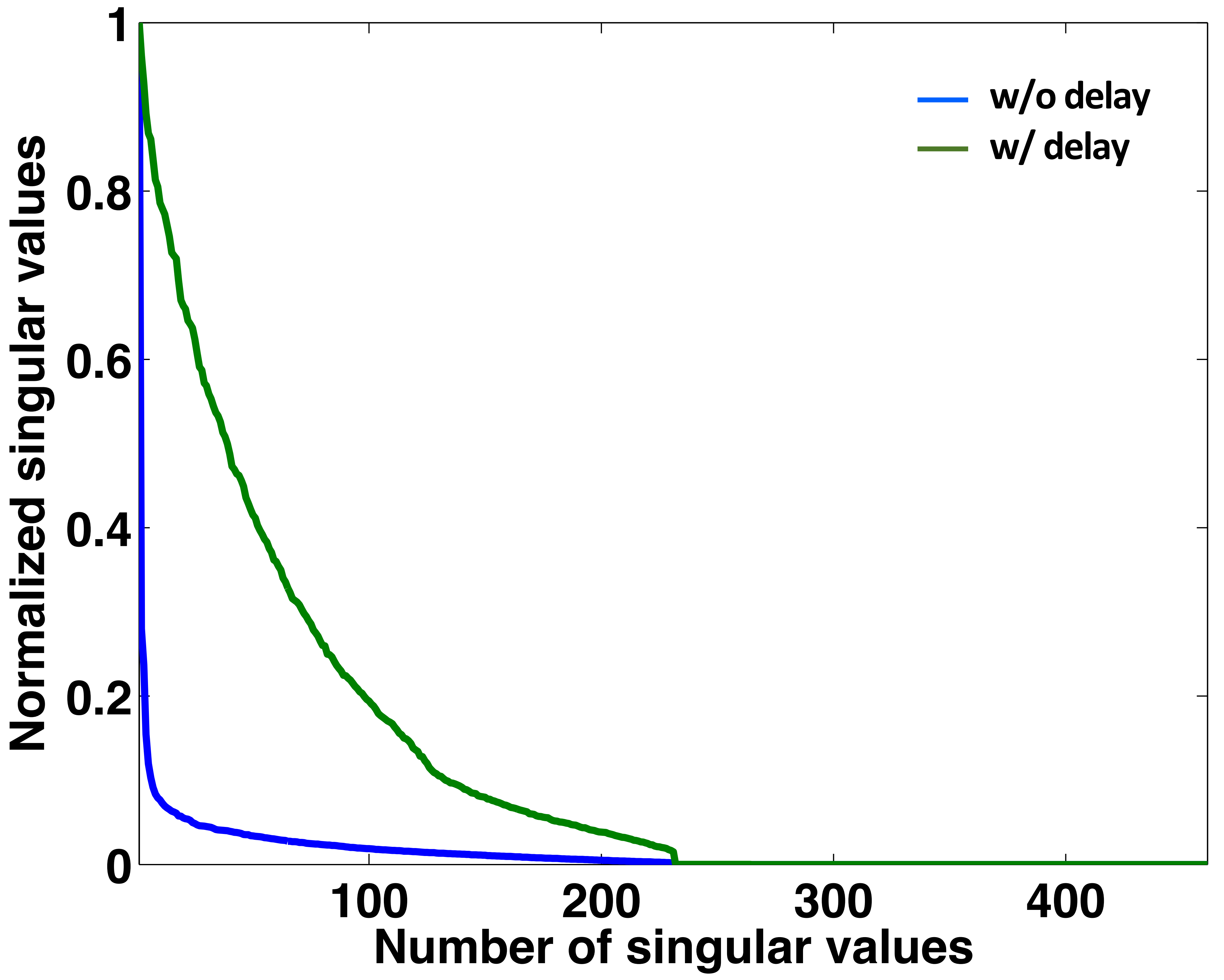
- add random time delays (<1s)



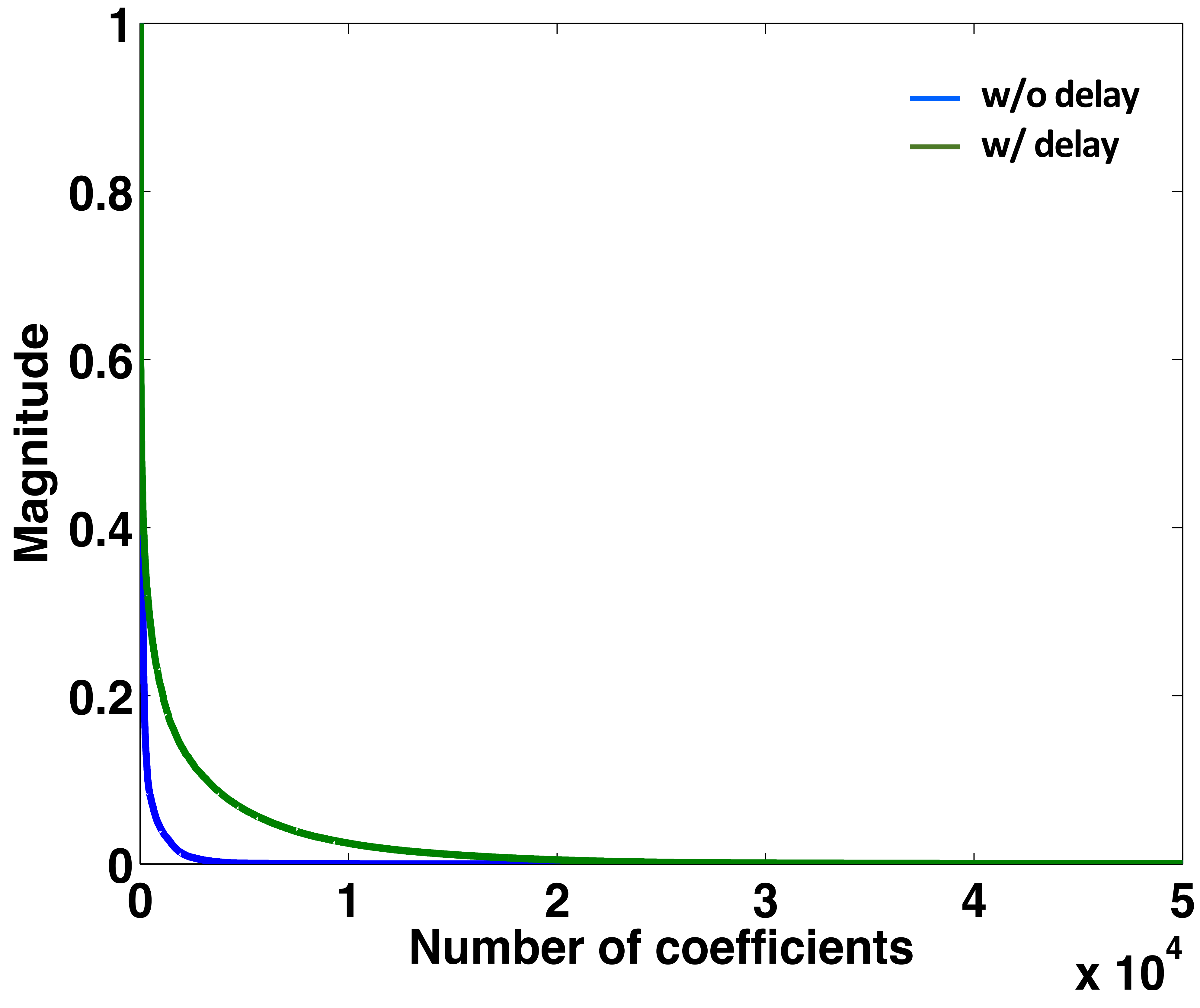


Rank vs. sparsity

rank-minimization
(midpoint-offset domain)



sparsity-promotion
(source-receiver domain)



Rank-minimization

$$\min_{\mathbf{X}} \quad \underbrace{\operatorname{rank}(\mathbf{X})} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

number of singular values of $\mathbf{X}$

Rank-minimization

$$\min_{\mathbf{X}} \quad \underbrace{\text{rank}(\mathbf{X})} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

number of singular values of $\mathbf{X}$

for blended acquisition:

$\mathbf{b}$: blended data

$$\mathcal{A} := \left[\mathbf{MS^H} \quad \mathbf{MTS^H}\right]$$

↑ time delay matrix

unblended data matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$$

$\mathbf{X}_1$ ⟵ source 1

$\mathbf{X}_2$ ⟵ source 2

Rank-minimization

expensive
(search over all possible values of rank)

$$\min_{\mathbf{X}} \quad \underbrace{\operatorname{rank}(\mathbf{X})} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

number of singular values of $\mathbf{X}$

Rank-minimization

expensive
(search over all possible values of rank)

$$\min_{\mathbf{X}} \quad \underbrace{\mathrm{rank}(\mathbf{X})} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

number of singular values of $\mathbf{X}$

Nuclear norm-minimization

convex relaxation of rank-minimization

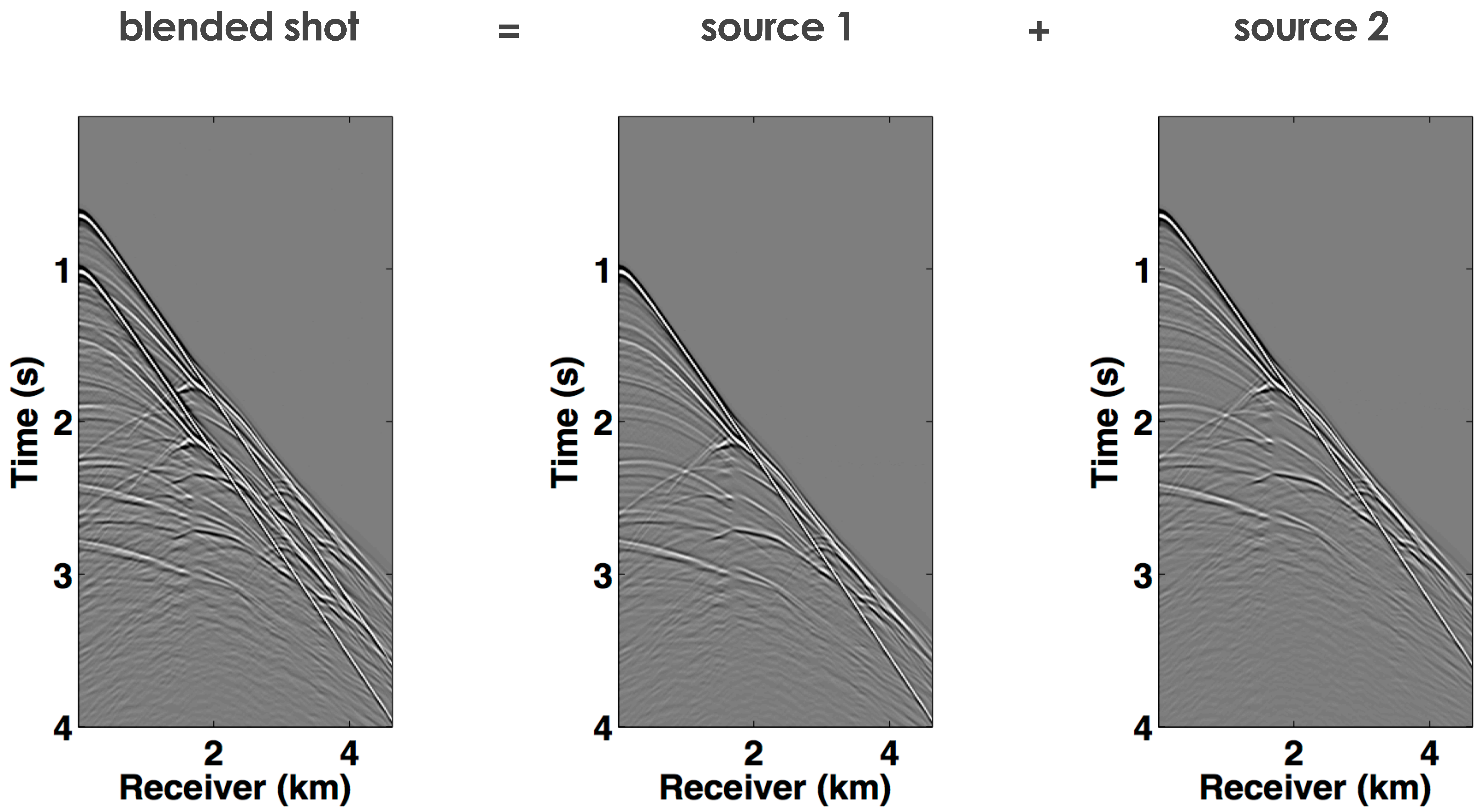
[Recht, et. al., 2010]

$$\min_{\mathbf{X}} \quad \underbrace{||\mathbf{X}||_*} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

sum of singular values of $\mathbf{X}$

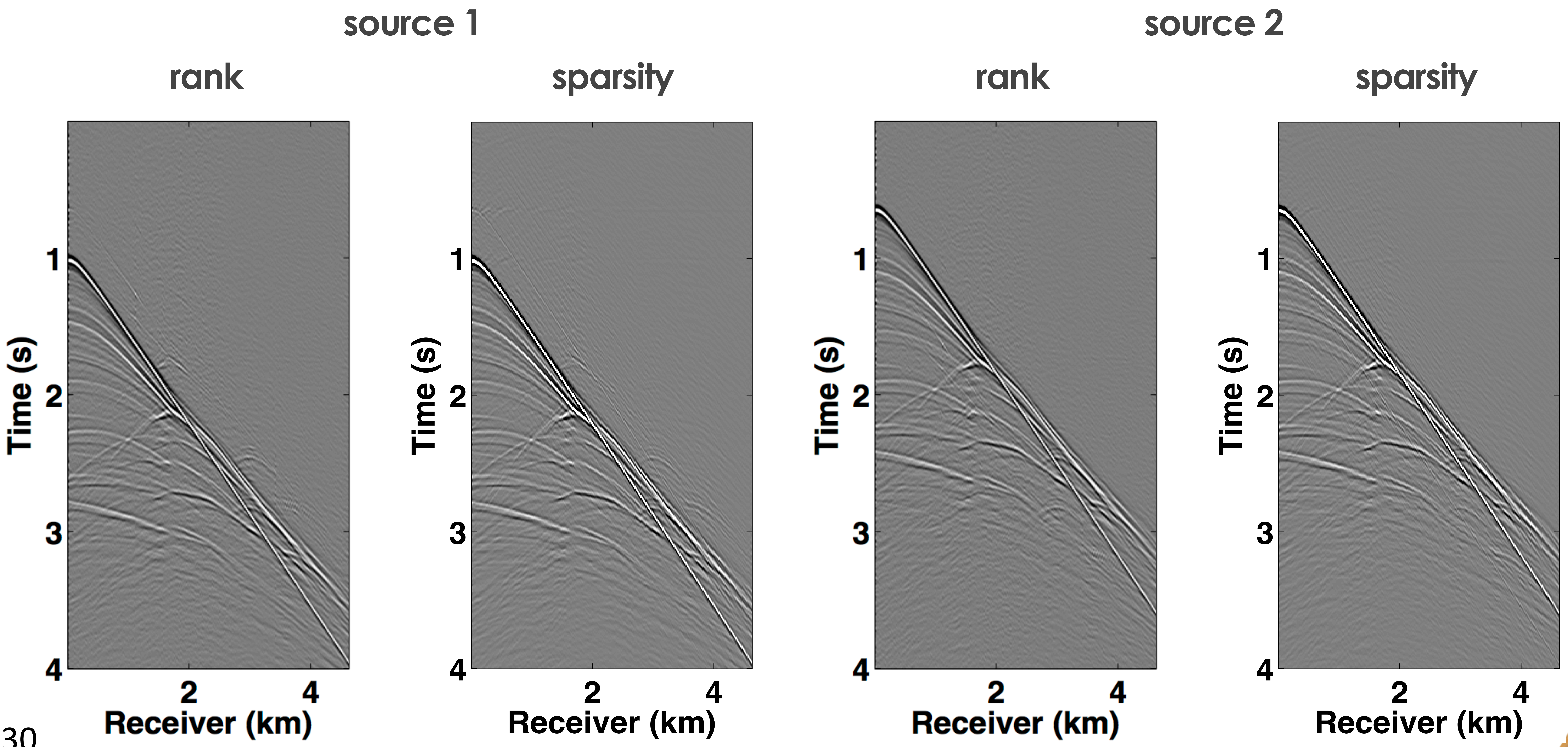
Blended data (w/ delay)

- random time delays (< 1 sec) applied to both sources



Source separation - *rank vs. sparsity*

computation time = **5 vs. 62 hours**; memory usage = **2.8 vs. 7.0 GB**; average SNR = **15.7 dB** (for both)



Randomized seismic-date acquisition by industry

thanks to Chuck Mosher & Nick Moldoveanu

SLIM
University of British Columbia

[Moldoveanu, 2010 –]

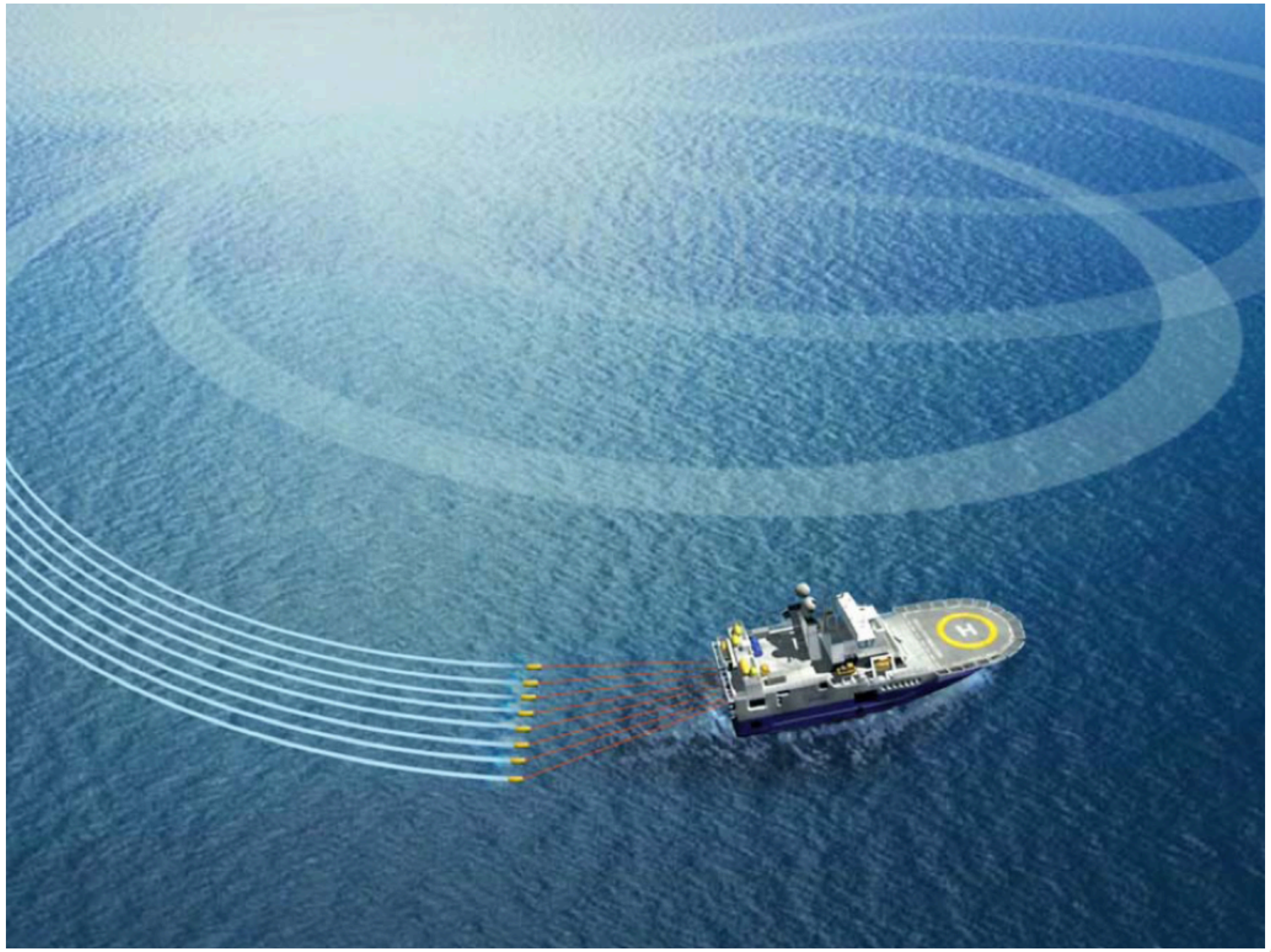
Randomized acquisition
– examples from industry (WesternGeco)

Randomized sim-source coil:

- long-offset wide azimuth
- incoherent (noisy) artifacts in image space
- high fold & diverse illumination stacks out imaging & sampling related artifacts

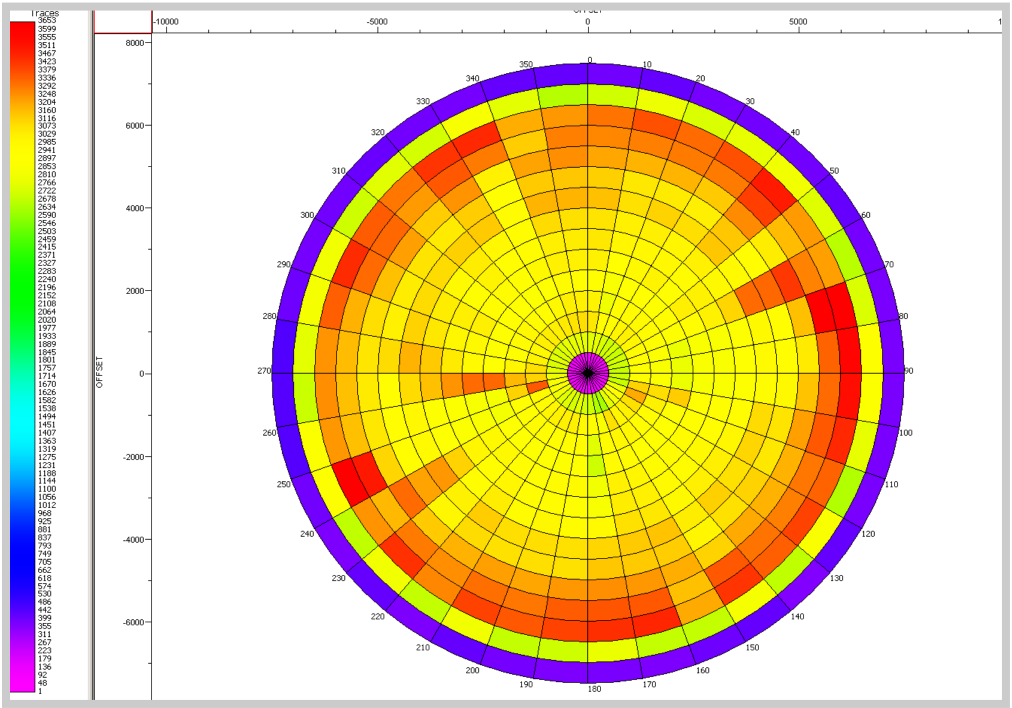
Output:

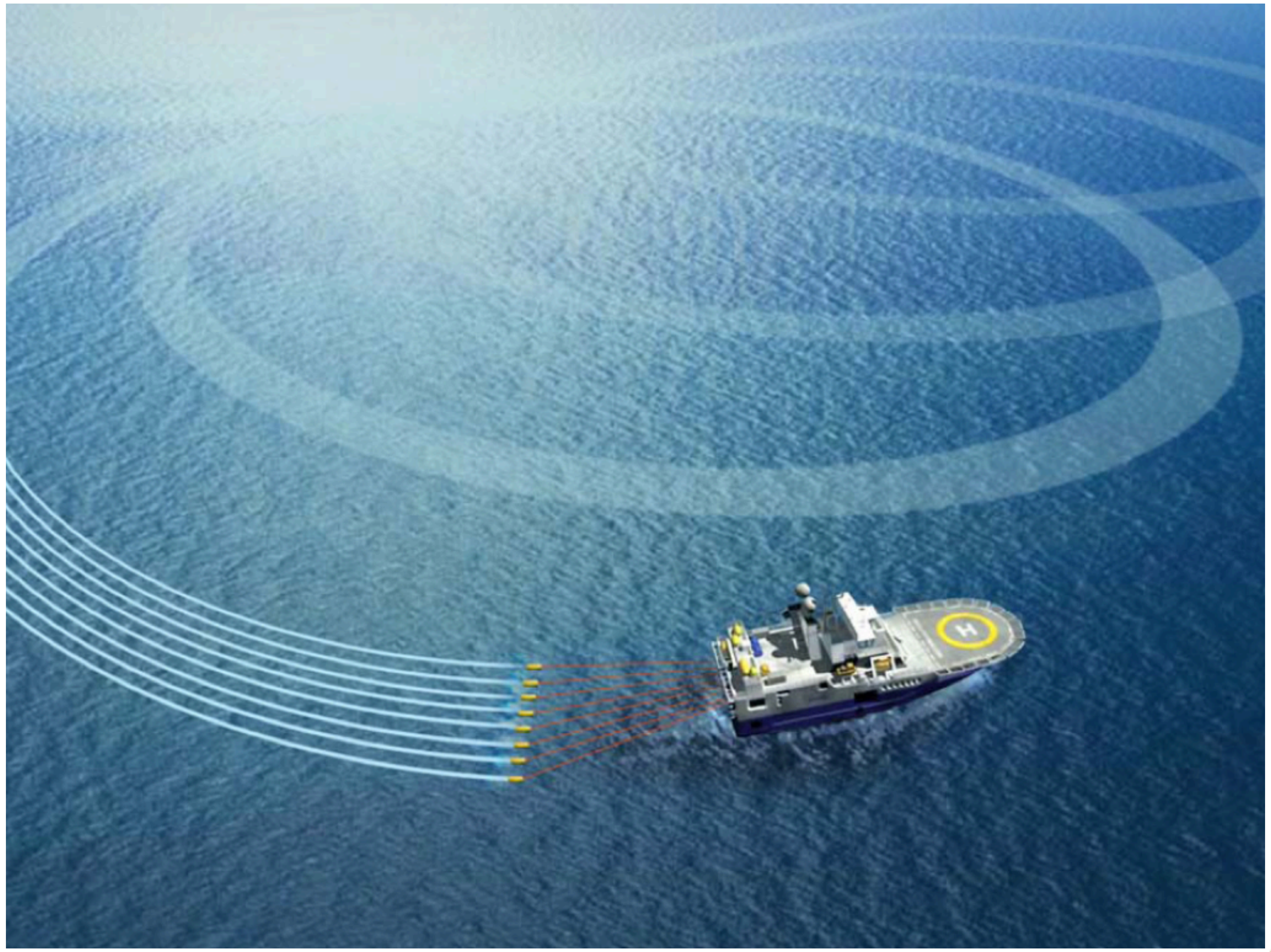
- high-fi & high-resolution images



[Moldoveanu, 2010 –]

Randomized acquisition
– examples from industry (WesternGeco)





Randomized sim-source coil:

- long-offset wide azimuth
- incoherent (noisy) artifacts in image space
- high fold & diverse illumination stacks out imaging & sampling related artifacts

Output:

- high-fi & high-resolution images

Mosher, C. C., Keskula, E., Kaplan, S. T., Keys, R. G., Li, C., Ata, E. Z., ... & Sood, S. (2012, November). Compressive Seismic Imaging. In *2012 SEG Annual Meeting*. Society of Exploration Geophysicists.

Randomized acquisition
– examples from industry (ConocoPhilips)

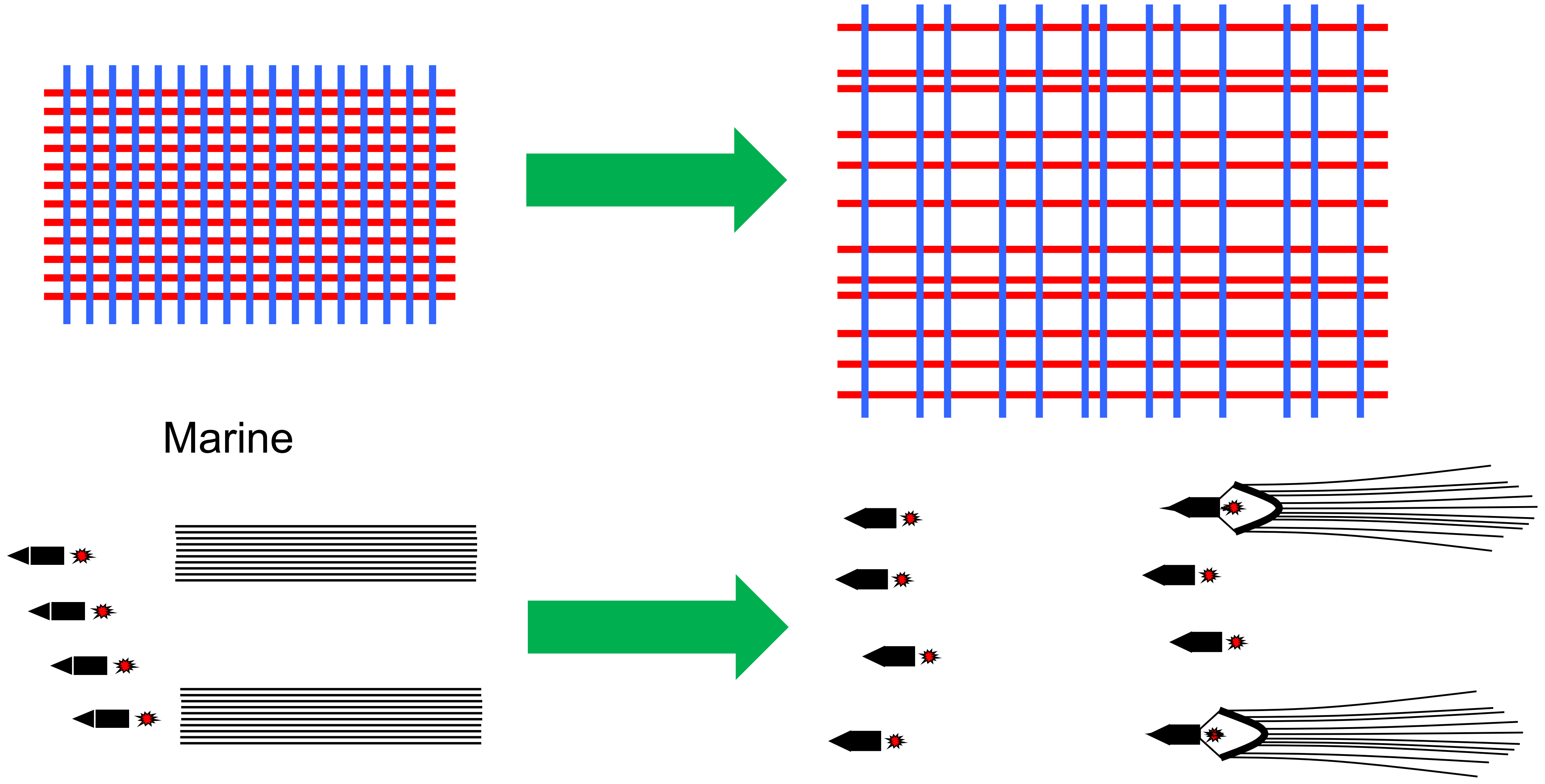
Deliberate & natural randomness in acquisition

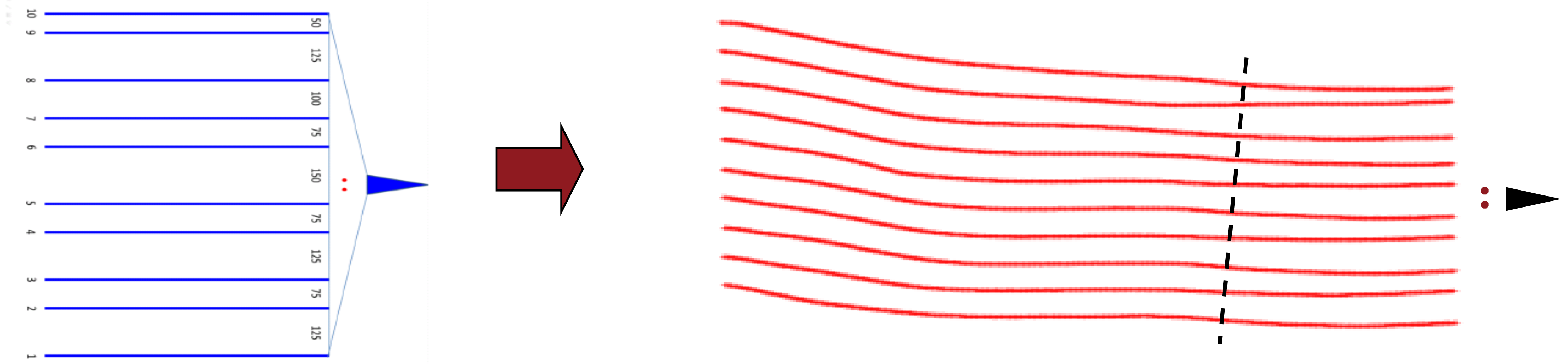
(thanks to Chuck Mosher)

Compressive Sensing = Acquisition Efficiency

Land / Node

Marine





Designed

Actually Acquired

Bottom line

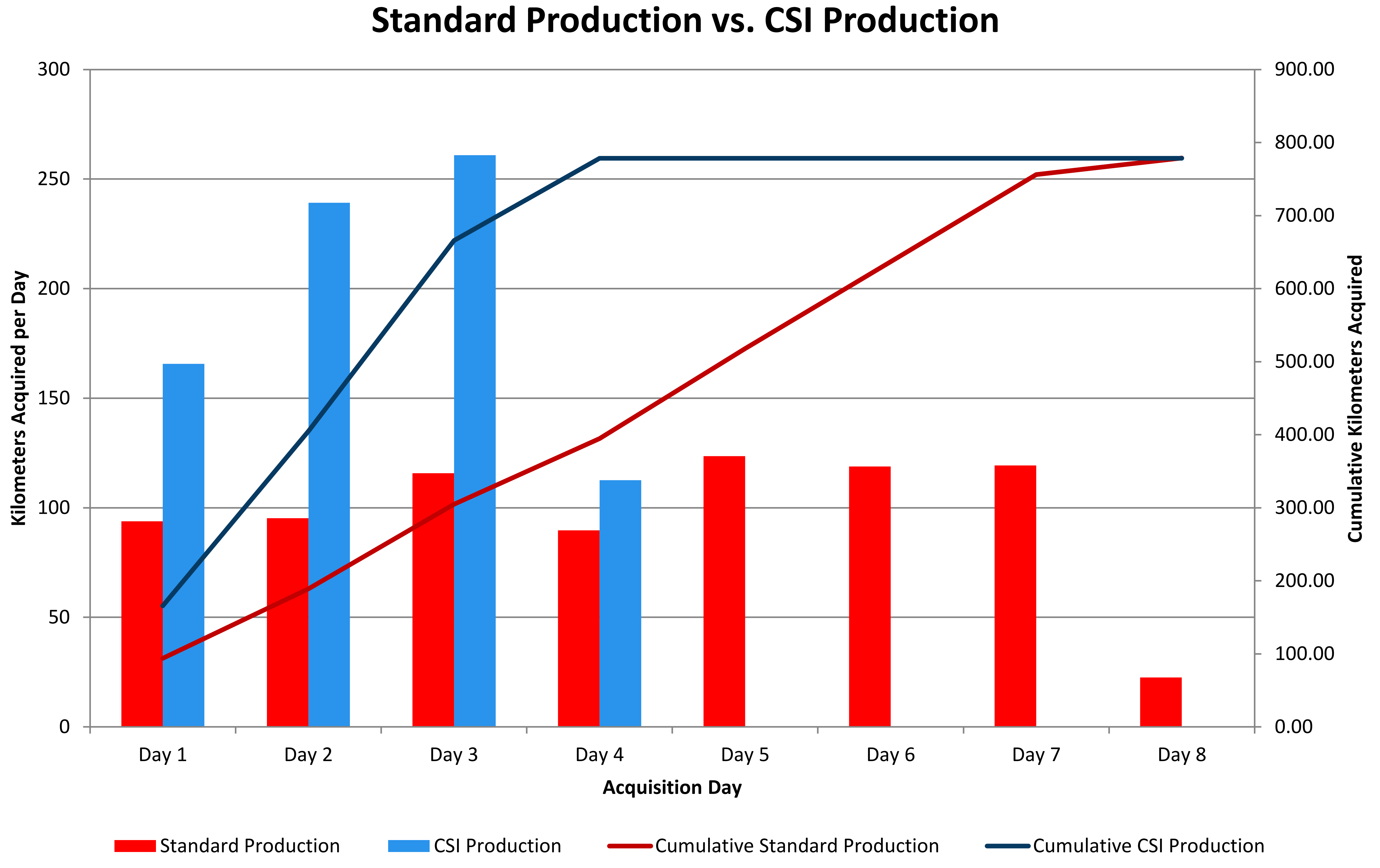
– examples from industry (ConocoPhilips)

Randomized subsampling:

- exploits (natural) randomness & structure in seismic
- economic subsampled data
- recovers dense data via structure-promoting inversion

Output:

- improved quality artifact-fact free long-offset wide azimuth data
- 5 X – 10 X cost & environmental impact reduction



SLIM

Observations

CS corresponds to an acquisition design problem

- validated in the field
- dense surveys (static & dynamic geometries) from economic randomly subsampled data

Creates a data deluge

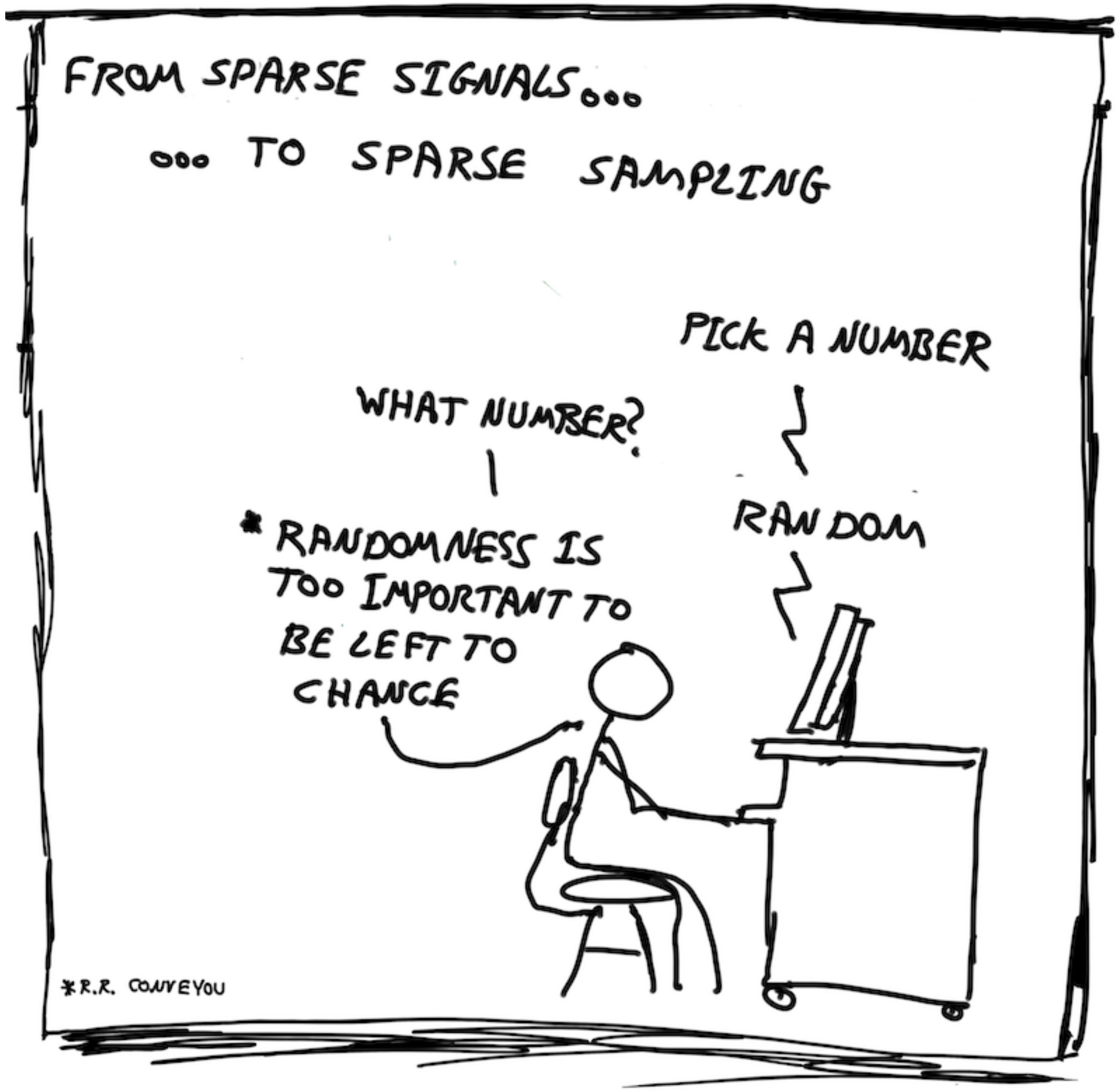
- leads to excessive demands on compute to image increased data volumes
- strains turn-around times wave-equation based inversions

Bottom line: Randomized sampling reduces

- acquisition costs (5 X – 10 X)
- environmental imprint
- improved data quality

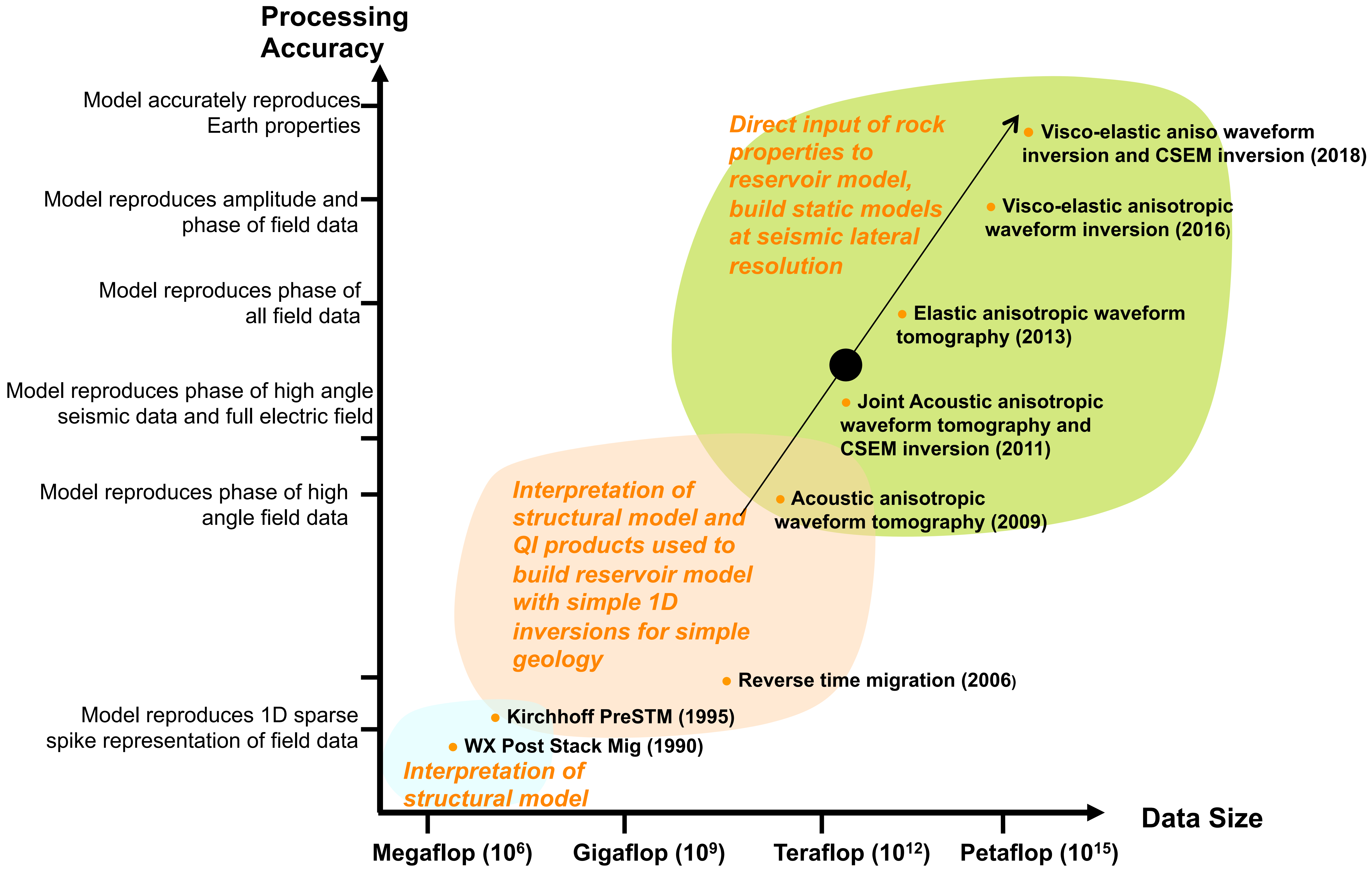
SLIM

From Michael Lustig



http://www.eecs.berkeley.edu/~mlustig/comics/MRM_17.png

Computational costs



courtesy. BG Group

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

Batched randomized computations

Tristan van Leeuwen, Sasha Aravkin, and Michael Friedlander



SLIM
University of British Columbia

Tristan van Leeuwen and Felix J. Herrmann, "3D frequency-domain seismic inversion with controlled sloppiness", *SIAM Journal on Scientific Computing*, vol. 36, p. S192-S217, 2014
Felix J. Herrmann, Andrew J. Calvert, Ian Hanlon, Mostafa Javanmehri, Rajiv Kumar, Tristan van Leeuwen, Xiang Li, Brendan Smithyman, Eric Takam Takougang, and Haneet Wason, "Frugal full-waveform inversion: from theory to a practical algorithm", *The Leading Edge*, vol. 32, p. 1082-1092, 2013

Frugal FWI

Strategy:

- reduce costs by working w/ random subsets of sources
- allow for inaccurate physics (e.g., approximate PDE solves)
- convergence guarantees via dynamic accuracy control via dynamic increase size subsets & accuracy PDE solves

Outcome:

- computationally affordable for 2-D & 3-D FWI

Tristan van Leeuwen and Felix J. Herrmann, “Fast waveform inversion without source encoding”, *Geophysical Prospecting*, vol. 61, p. 10-19, 2013.
Aleksandr Y. Aravkin, Michael P. Friedlander, Felix J. Herrmann, and Tristan van Leeuwen, “Robust inversion, dimensionality reduction, and randomized sampling”, *Mathematical Programming*, vol. 134, p. 101-125, 2012

Frugal FWI
– separable structure

misfit per source

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) = \frac{1}{M}\sum_{i=1}^{M}\phi_i(\mathbf{m})$$

solution with *steepest descent*

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \lambda_k \nabla\Phi(\mathbf{m}_k)$$

requires evaluation of *full* misfit and is *very* expensive.

Frugal FWI

– with errors

Allow errors in gradients–i.e.,

$$\nabla\widetilde{\Phi}(\mathbf{m}_k) = \nabla\Phi(\mathbf{m}_k) + \mathbf{e}_k$$

error in gradient

- draw independent source aggregates (supershots) or subsets of sources after each model update
- stochastic/incremental gradients

Leads to sublinear convergence & to instabilities due to noise.

Aleksandr Y. Aravkin, Tristan van Leeuwen and Felix J. Herrmann, “3D frequency-domain seismic inversion with controlled sloppiness”, *SIAM Journal on Scientific Computing*, vol. 36, p. S192-S217, 2014
Michael P. Friedlander, Felix J. Herrmann, and Tristan van Leeuwen, “Robust inversion, dimensionality reduction, and randomized sampling”, *Mathematical Programming*, vol. 134, p. 101-125, 2012
Michael P. Friedlander and Mark Schmidt, “Hybrid deterministic-stochastic methods for data fitting”, *SIAM Journal on Scientific Computing*, vol. 34, p. A1380-A1405, 2012.

Frugal FWI

– with error control

Approximate gradients by sample averages–i.e,

batch size

$$\nabla\Phi(\mathbf{m}_k) \approx \nabla\widetilde{\Phi}(\mathbf{m}_k) = \frac{1}{|\mathcal{I}_k|}\sum_{i\in\mathcal{I}_k}\nabla\phi_i(\mathbf{m}_k), \quad \mathcal{I}_k \subseteq \{1,2,\ldots,M\},\ |\mathcal{I}_k| = B_k$$

Guarantee convergence by bounding the errors $e_k = \|\mathbf{e}_k\|_2^2$ by increasing the batch size to

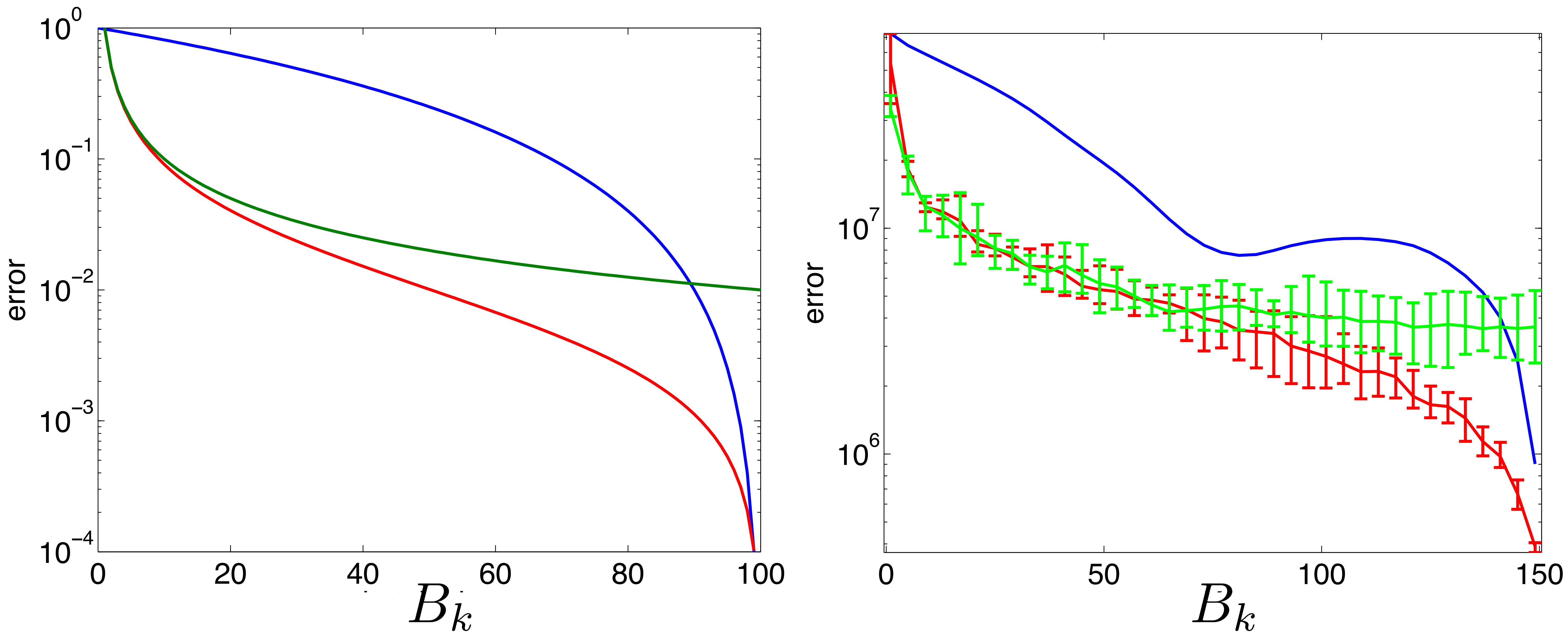
$$B_k \sim \min\{(e_k + M^{-1})^{-1}, M\}$$

at the k^{th} iteration.

Frugal FWI
– increasing the batch size

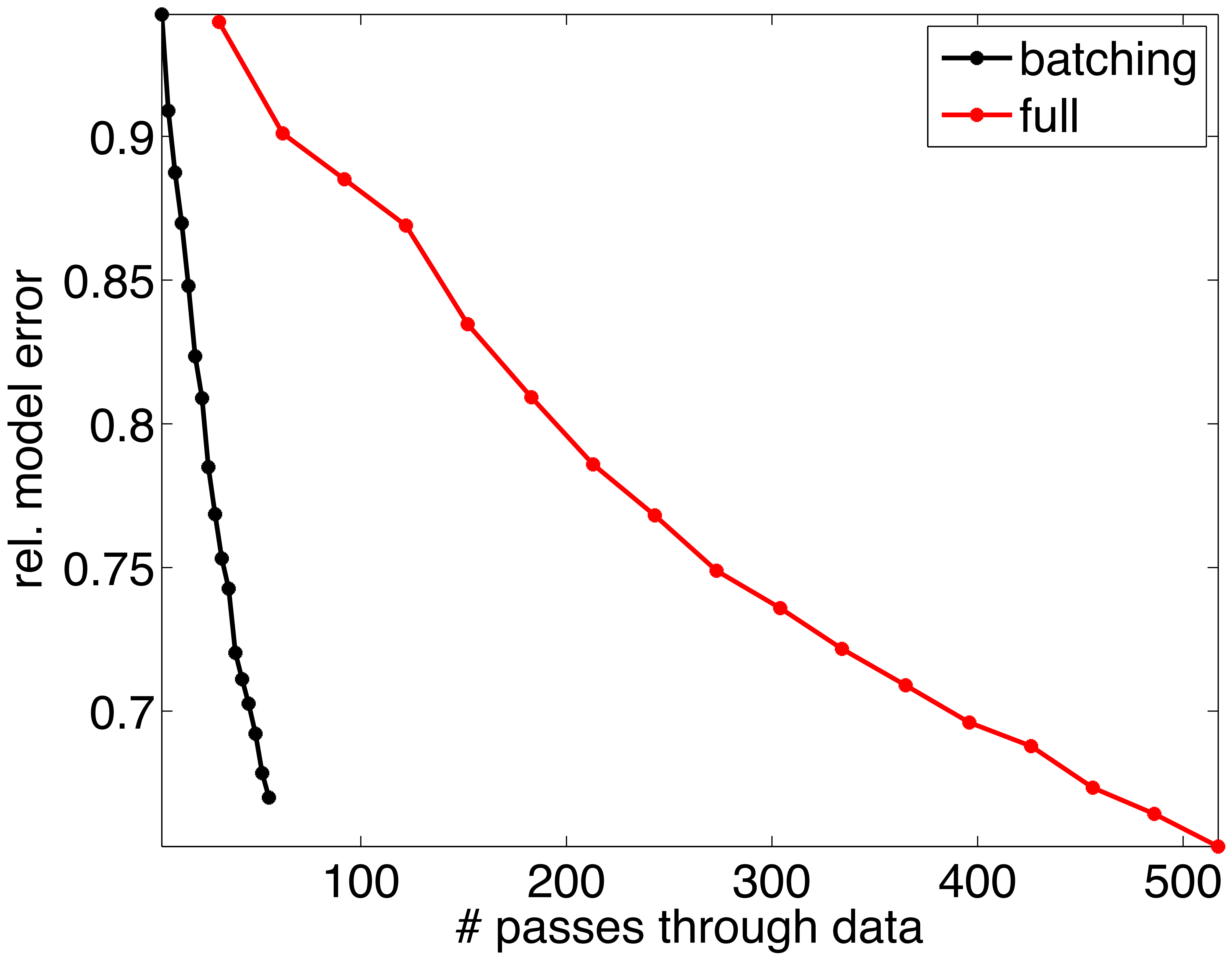
Select sources

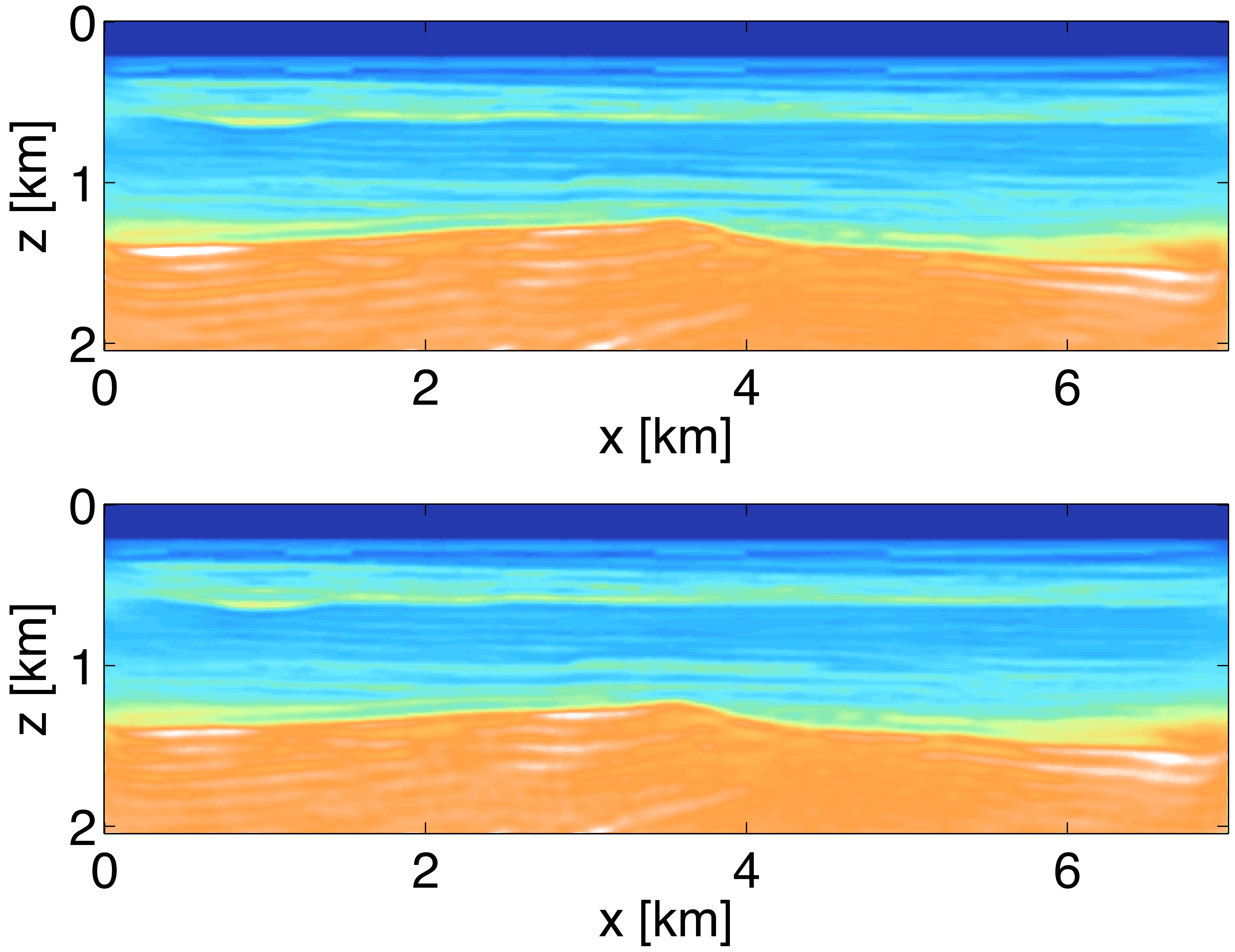
- in a pre-scribed order
- random *without* replacement
- random-*amplitude* source encoding



Tristan van Leeuwen and Felix J. Herrmann, "**Fast waveform inversion without source encoding**", *Geophysical Prospecting*, vol. 61, p. 10-19, 2013.

Frugal FWI
– 10 X speedup





Batched randomized computations by industry

thanks to Denes Vigh & Nick Moldoveanu

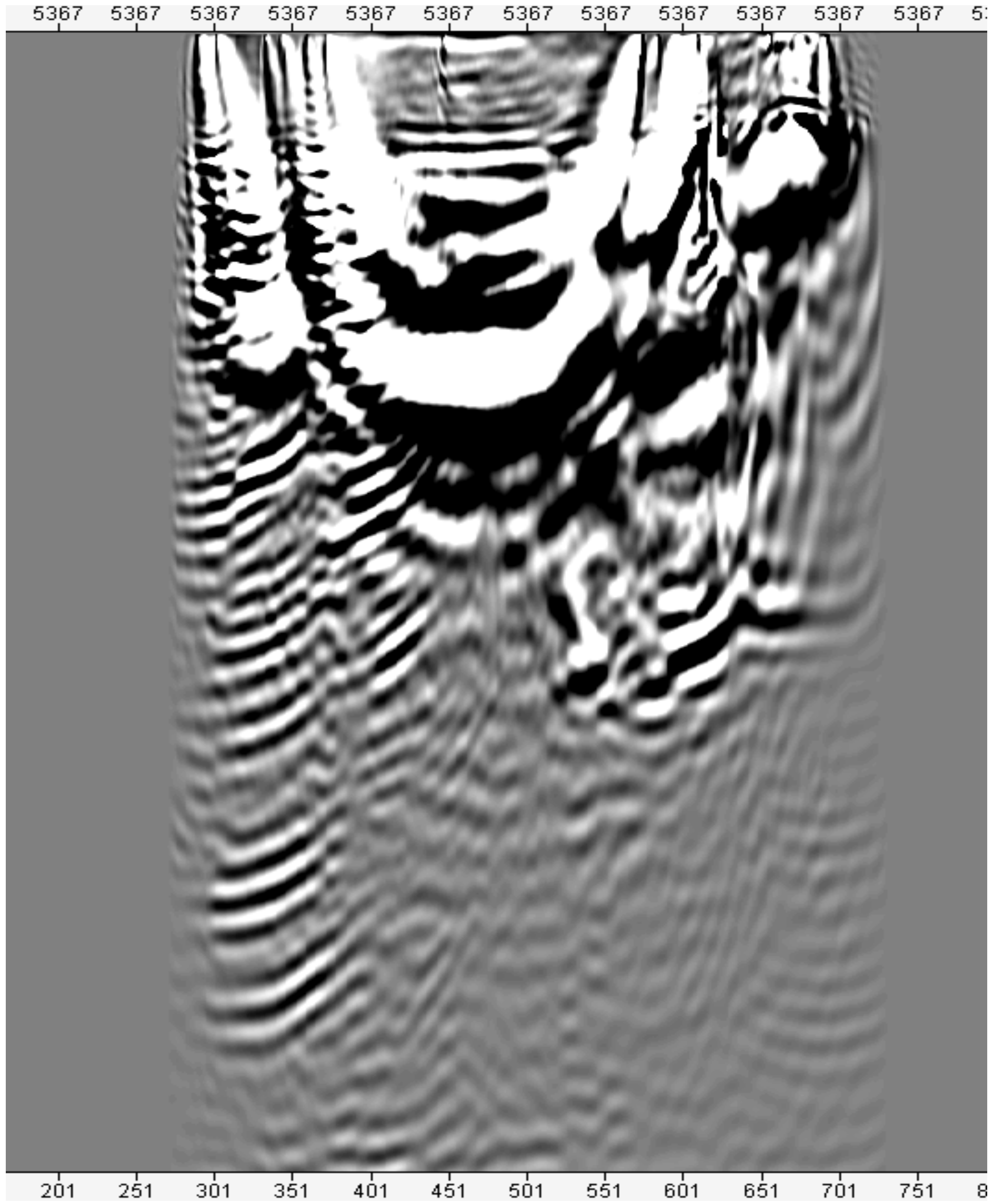
SLIM
University of British Columbia

All Shots

Total NO. Shots = 1749

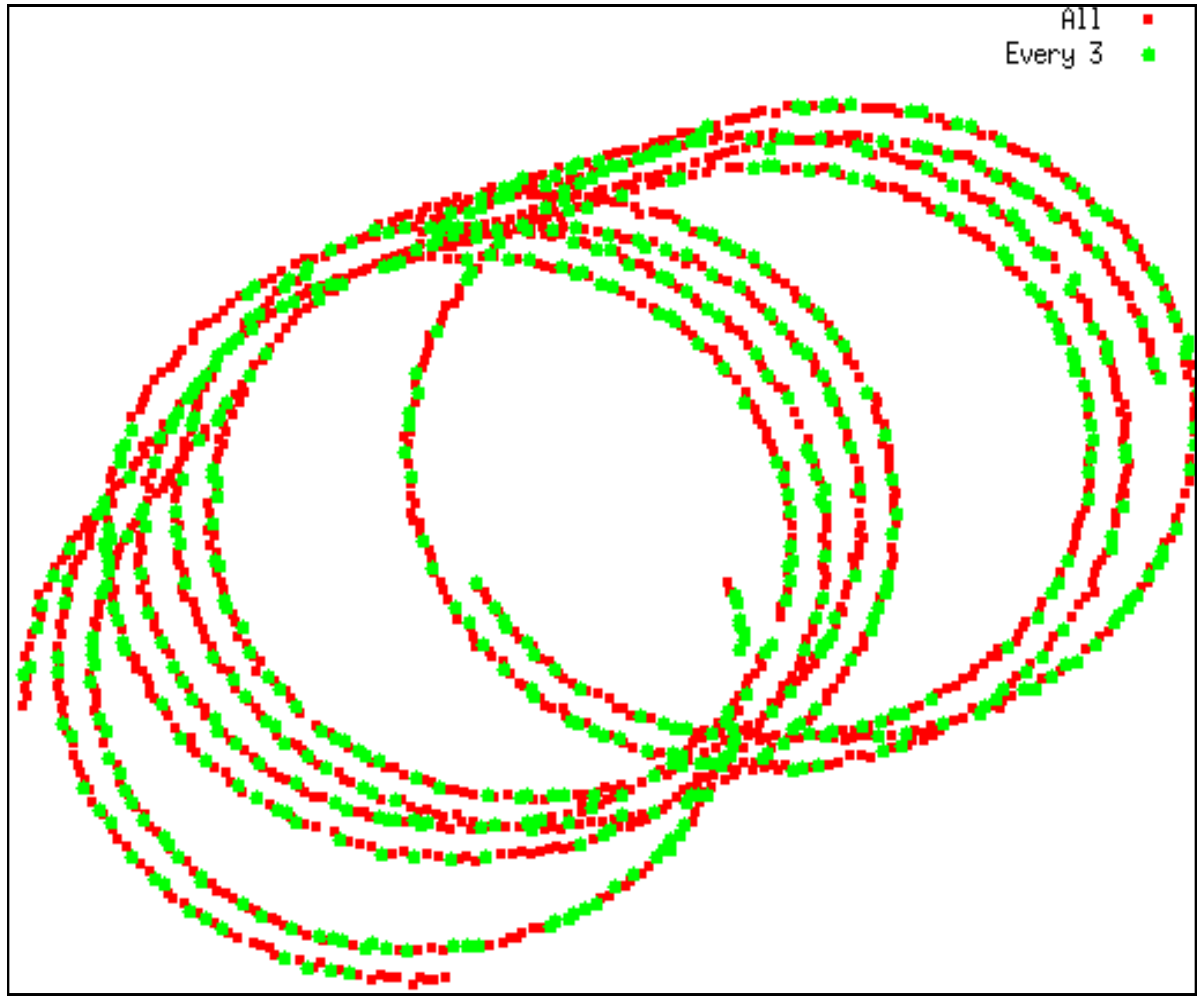
All shots selected

Gradient

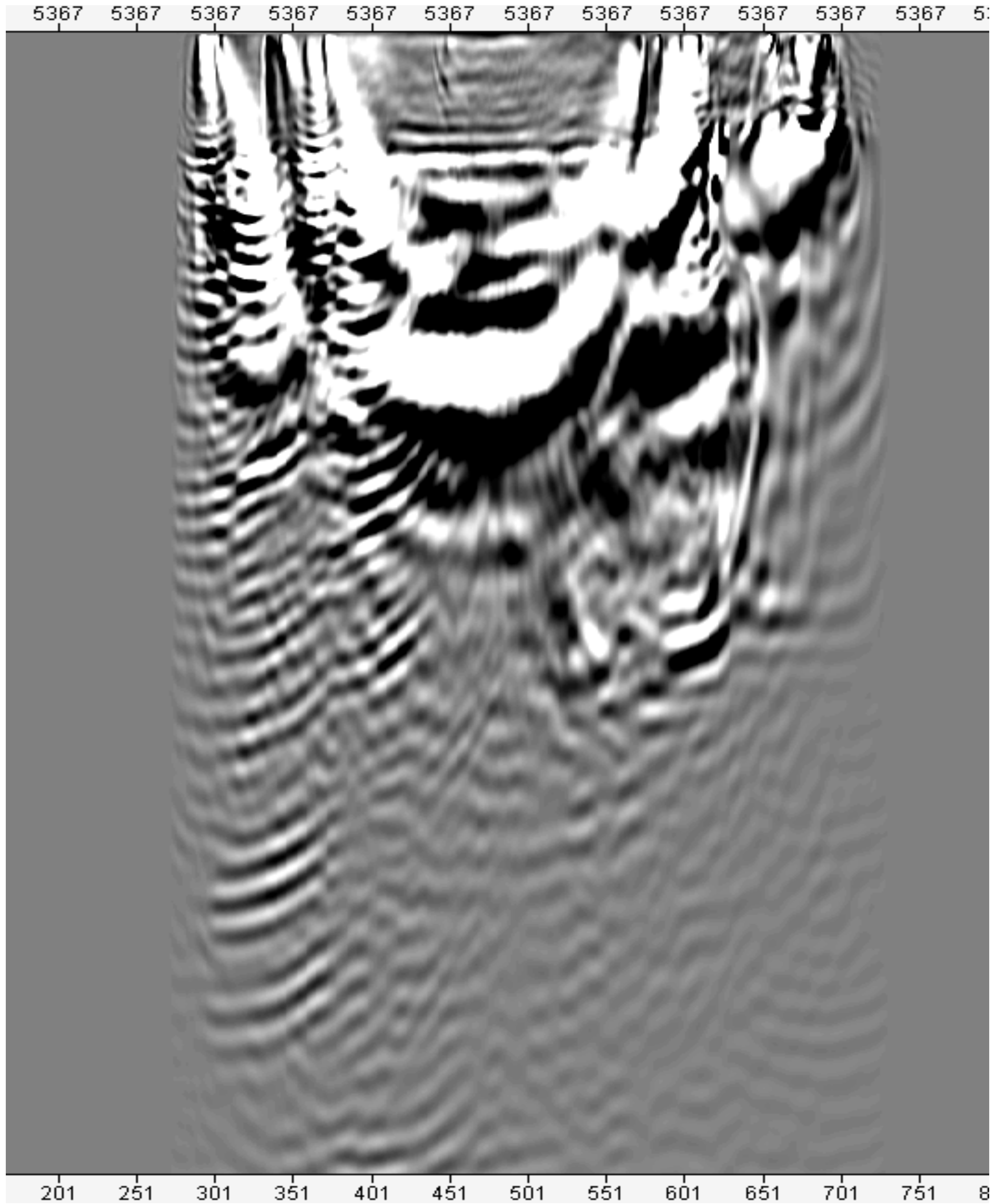


Fixed-increment sampling

Total NO. Shots = 1749
Periodic w/ inc 3 NO. Shots =584

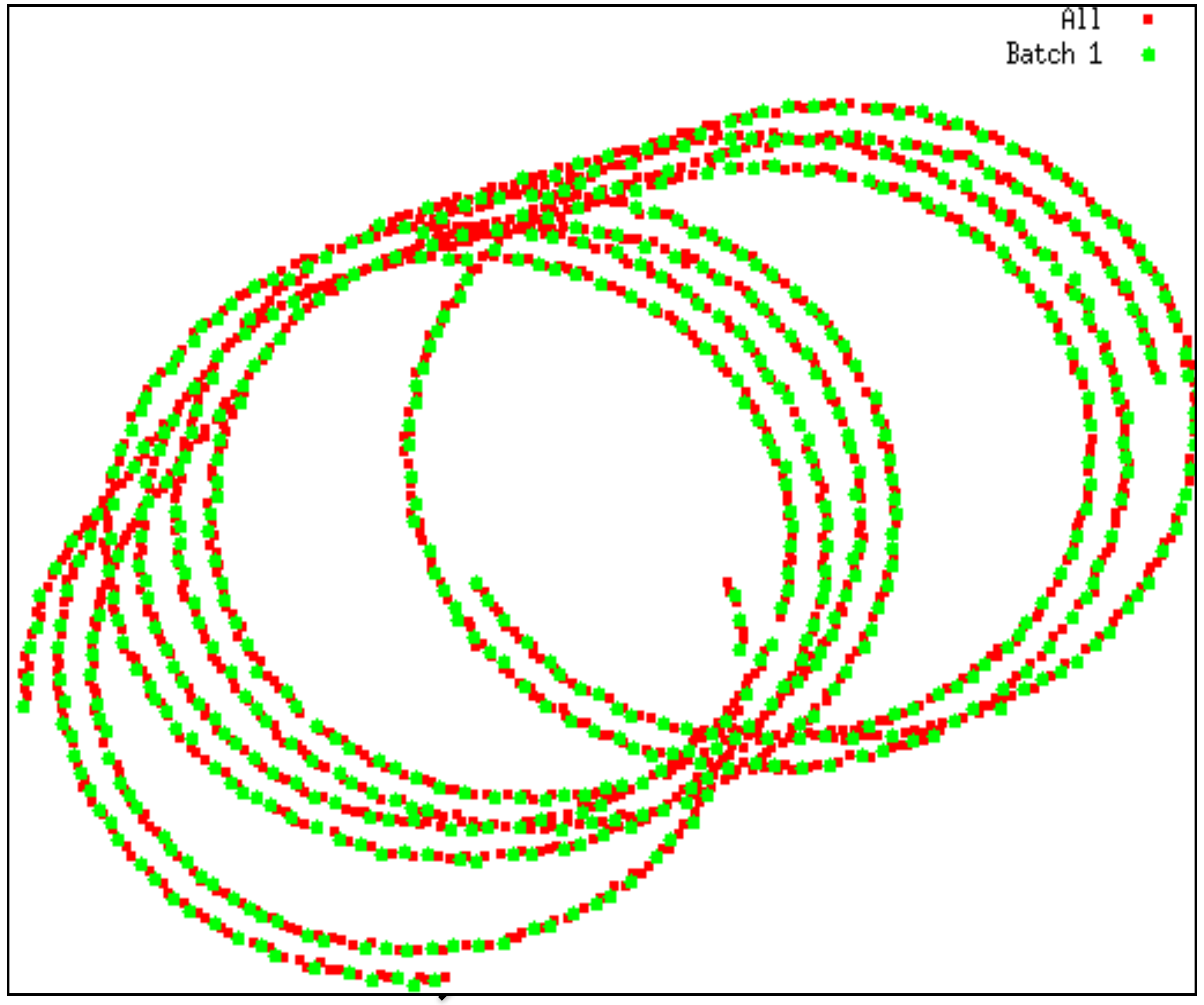


Gradient

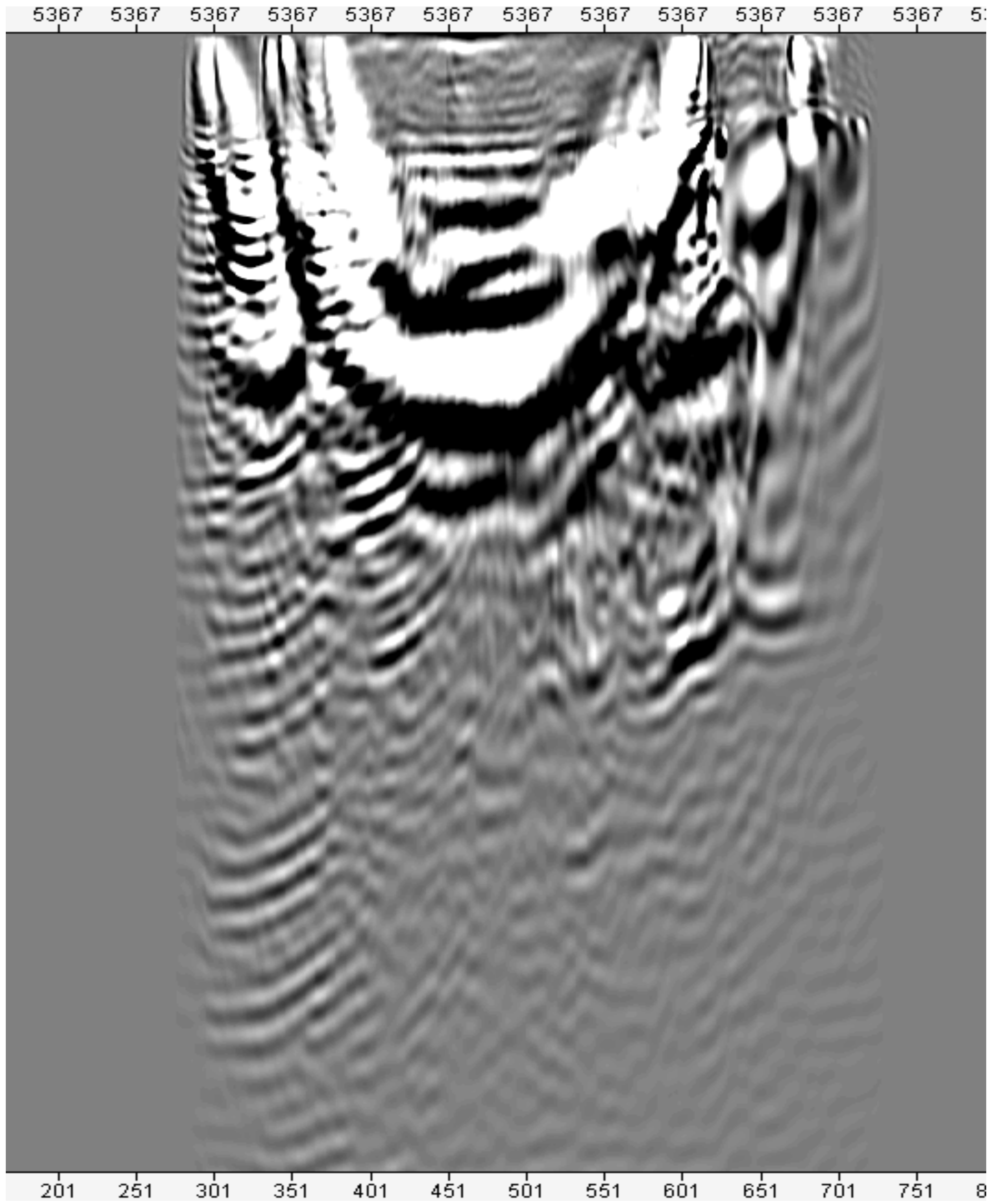


Random-batch sampling 1

Total NO. Shots = 1749
Batch 1 NO. Shots =612

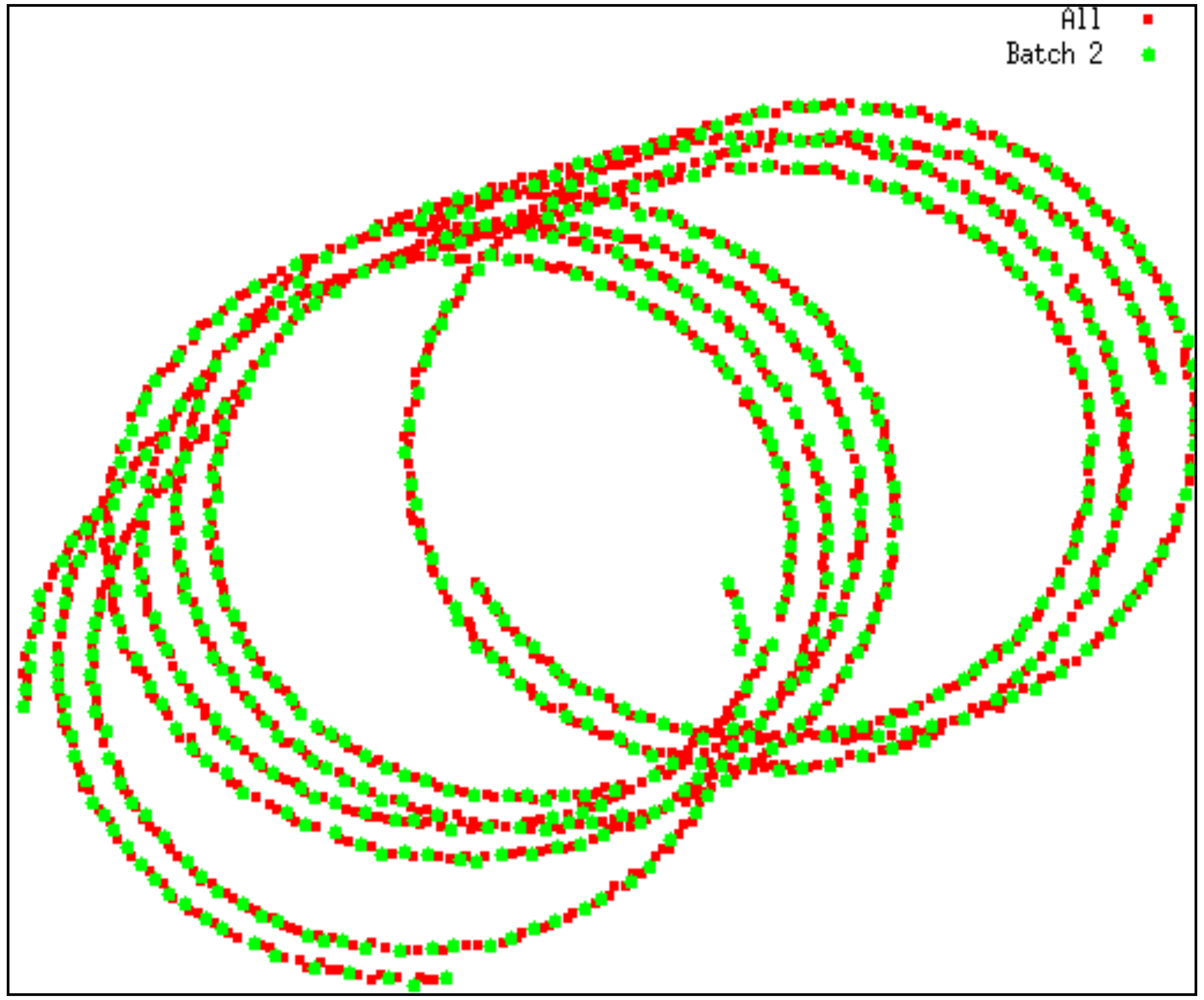


Gradient

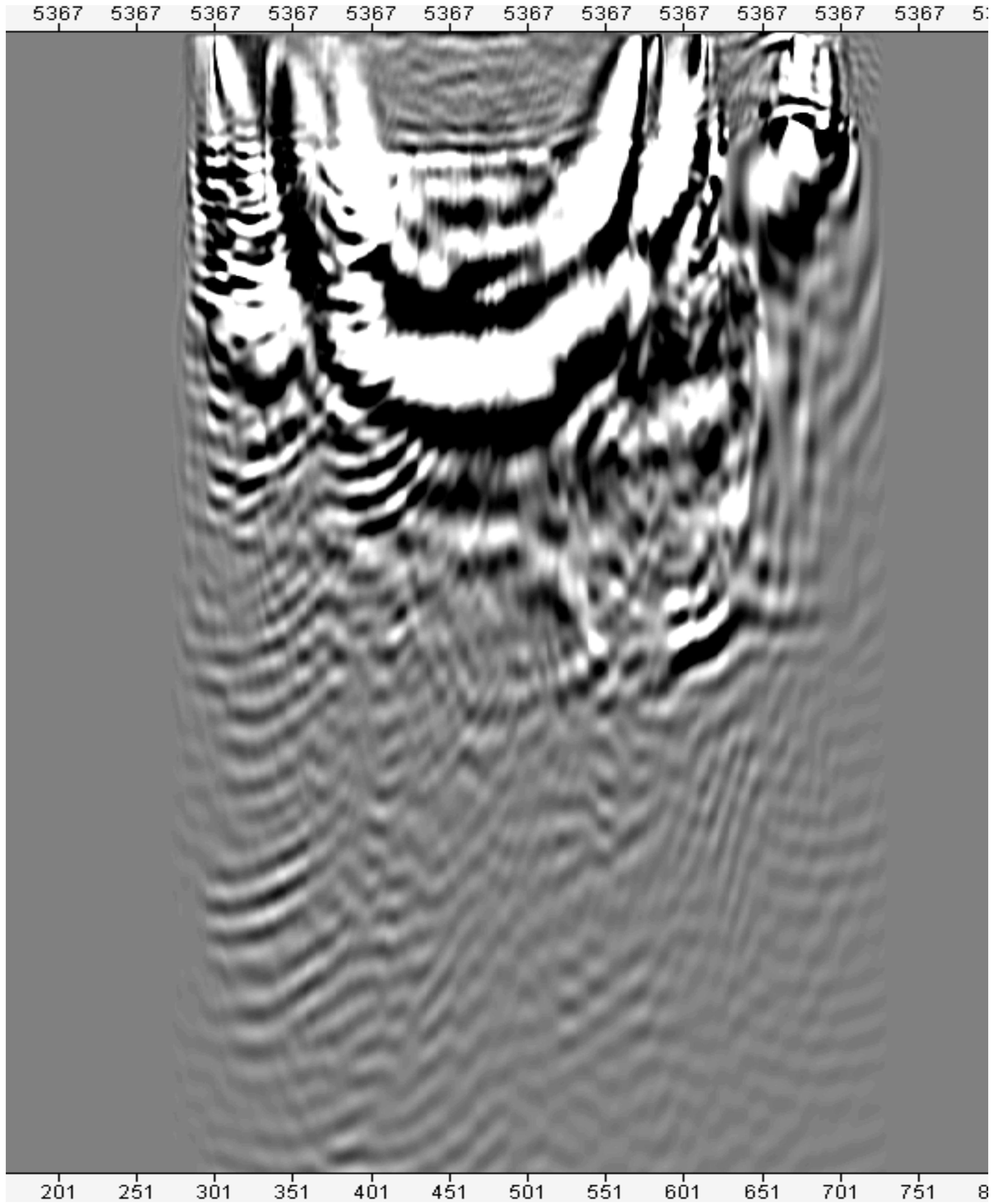


Random-batch sampling 2

Total NO. Shots = 1749
Batch 2 NO. Shots =604



Gradient



Tristan van Leeuwen, Aleksandr Y. Aravkin, and Felix J. Herrmann, "Seismic waveform inversion by stochastic optimization", *International Journal of Geophysics*, vol. 2011, 2011

Observations

Working w/ random source aggregates or subsets leads to

- significant speedups & improved inversion results
- but can be prone to (noise-related) instabilities unless error is controlled
- e.g., by increasing the batch size

Bottom line: selection of random subsets of shots leads to

- 5 X – 7 X reduction in computational costs for FWI w/ coil shooting
- makes FWI economically viable while still using information from all data

Does this approach extend to high-frequency RTM?

Compressive imaging

Xiang Li & Ning Tu

SLIM
University of British Columbia

Strategy & challenges

Compressive Sensing = randomized dimensionality reduction

Exploit structure:

- e.g. via the curvelet transform

Break structure & reduce computations:

- e.g. via randomly selected sources / via randomized composite sources

Restore structure:

- e.g. via curvelet-domain ℓ_1 or via Total-variation norm minimization

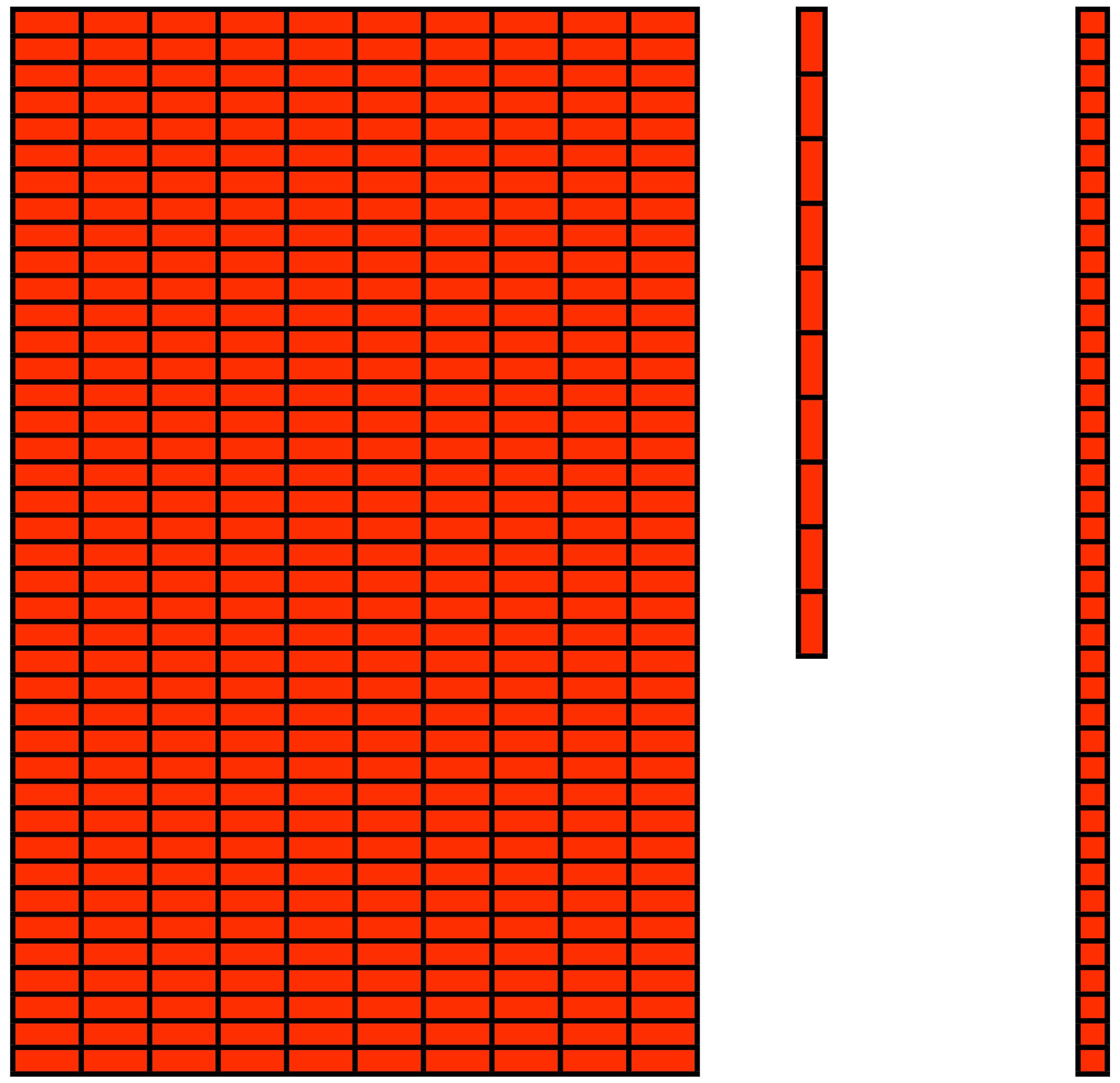
Challenges:

- algorithmic complexity
- ability to work w/ random subsets of data in parallel

Current imaging paradigm

Linear forward model: **A** **x** = **b**

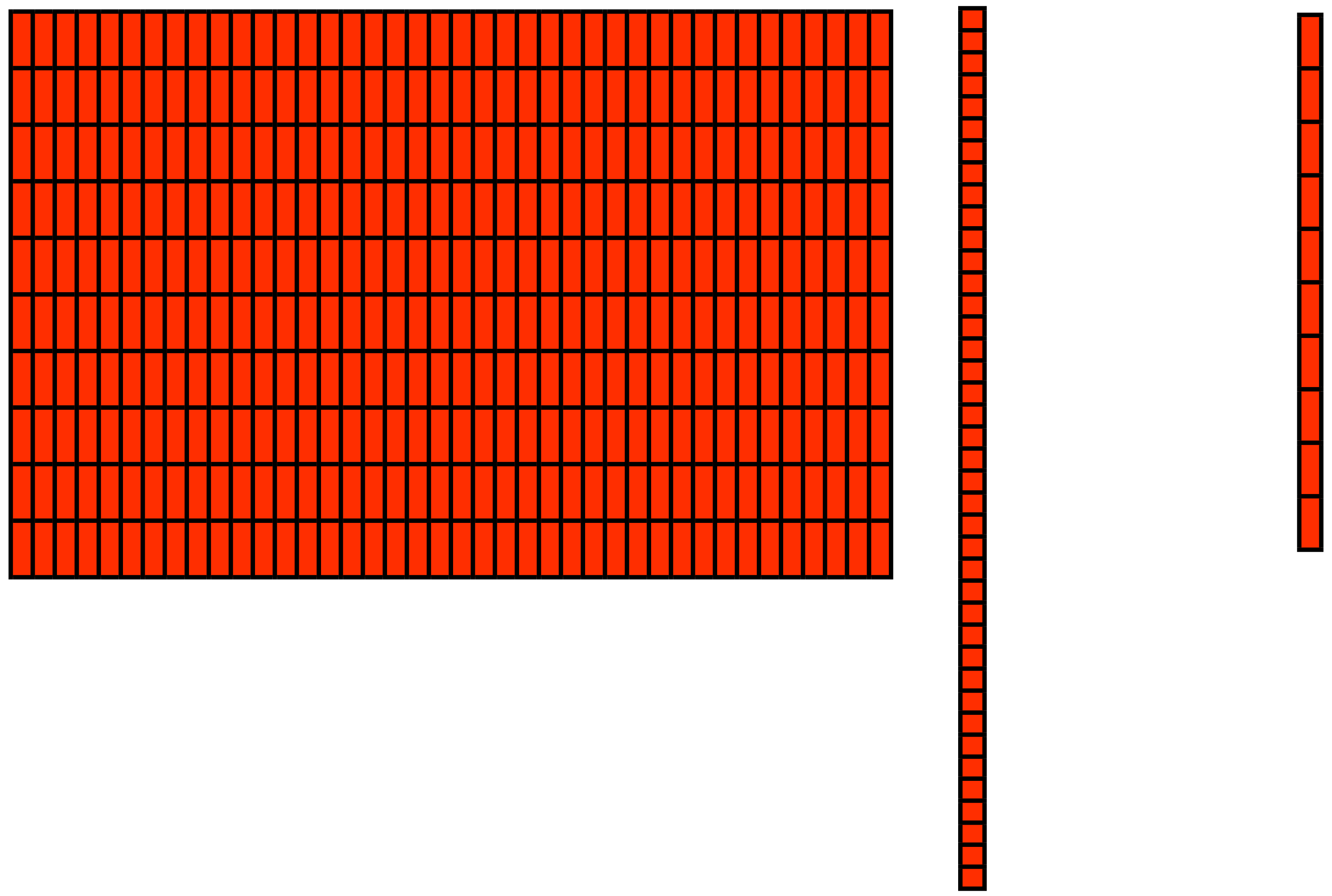
Tall matrix
(all data)



Current imaging paradigm

Migration: $\mathbf{A}^H \quad \mathbf{b} \quad = \quad \mathbf{x}_{\text{migrated}}$

adjoint of tall matrix
(all data)



Costs = # of PDE solves

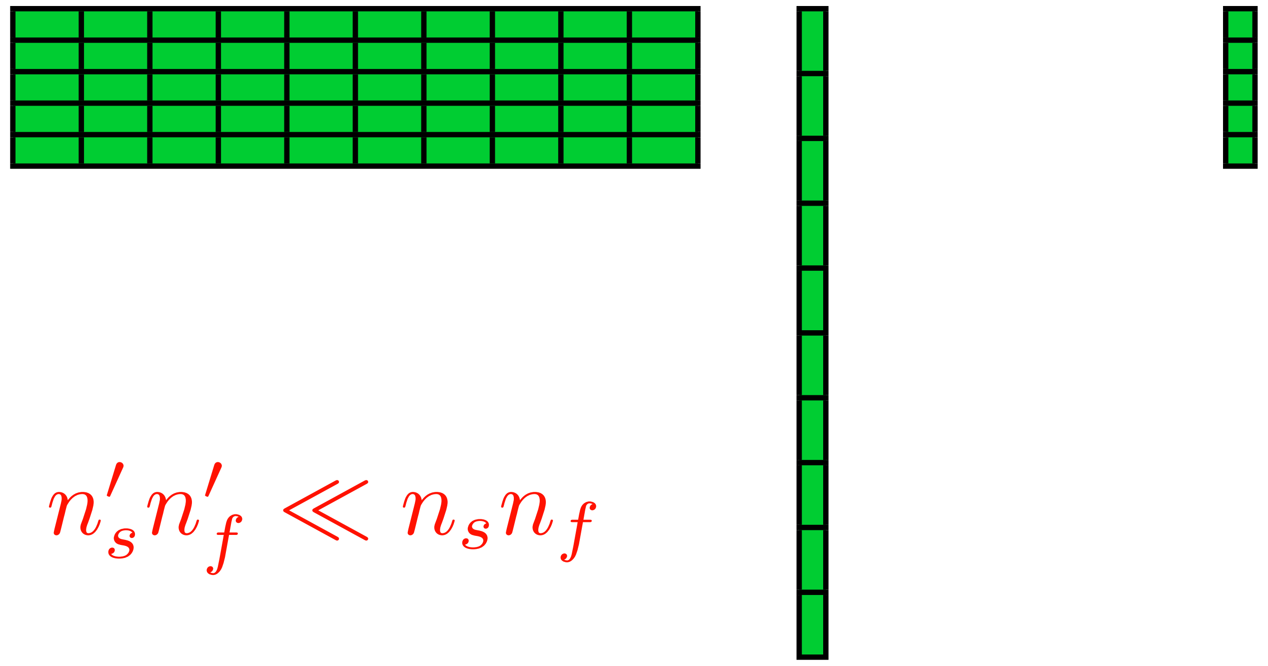
Adjoint $\neq$ Inverse

Ning Tu and Felix J. Herrmann, “Fast imaging with surface-related multiples by sparse inversion”. 2015.
Felix J. Herrmann and Xiang Li, “Efficient least-squares imaging with sparsity promotion and compressive sensing”, *Geophysical Prospecting*, vol. 60, p. 696-712, 2012.

New CS paradigm

Invert underdetermined systems: **A** **x** = **b**

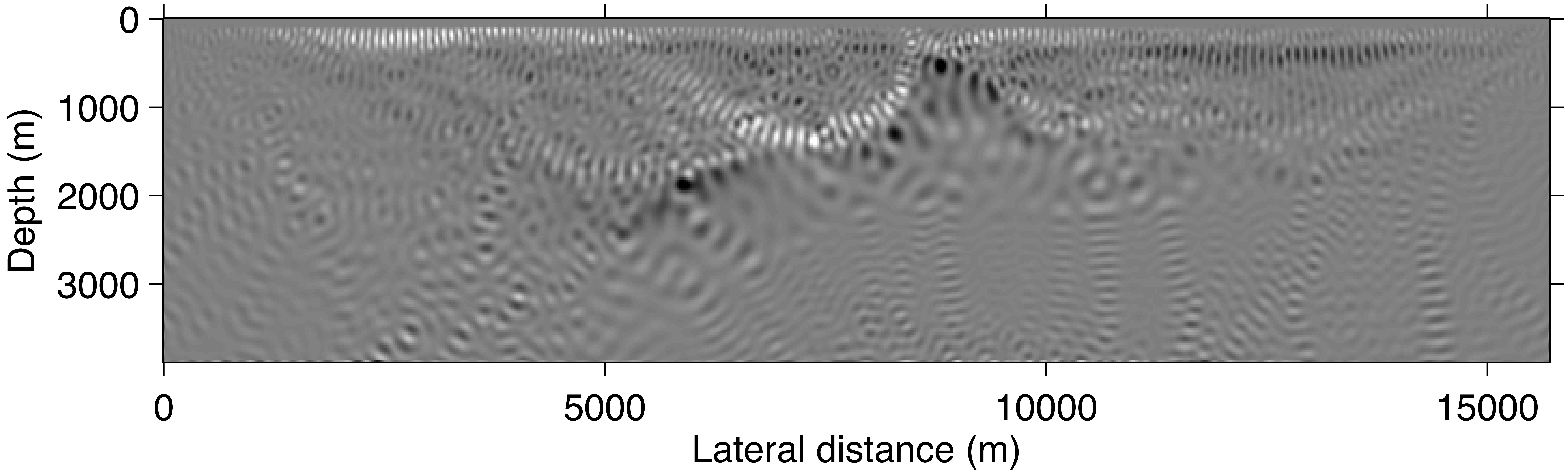
wide matrix
(random subsets)



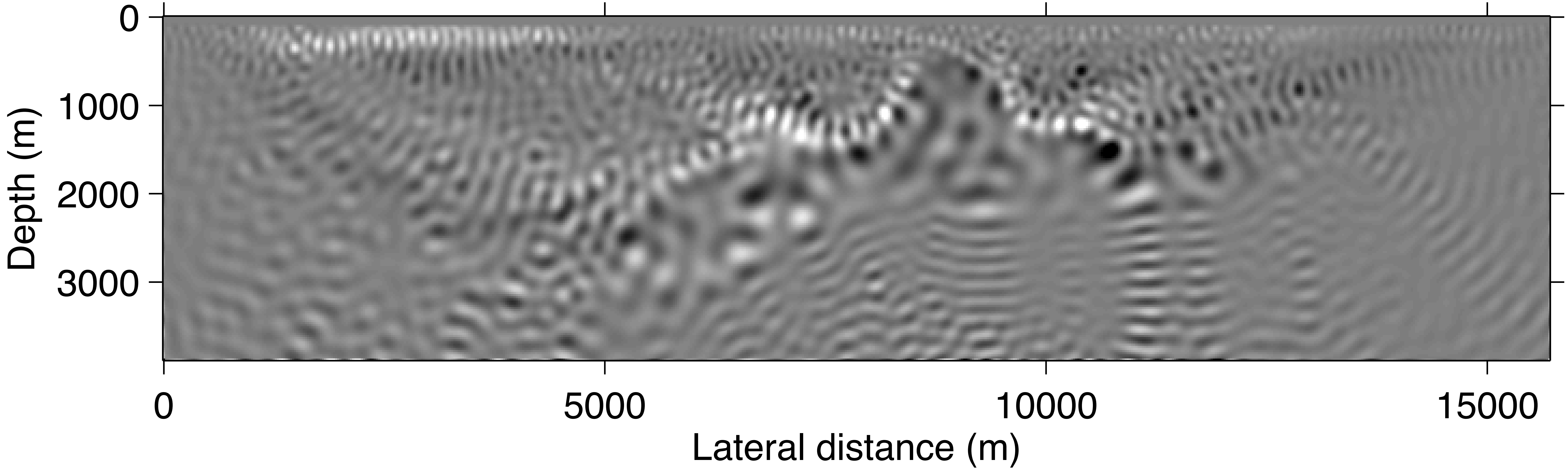
$n'_s n'_f \ll n_s n_f$

Subsets create artifacts...

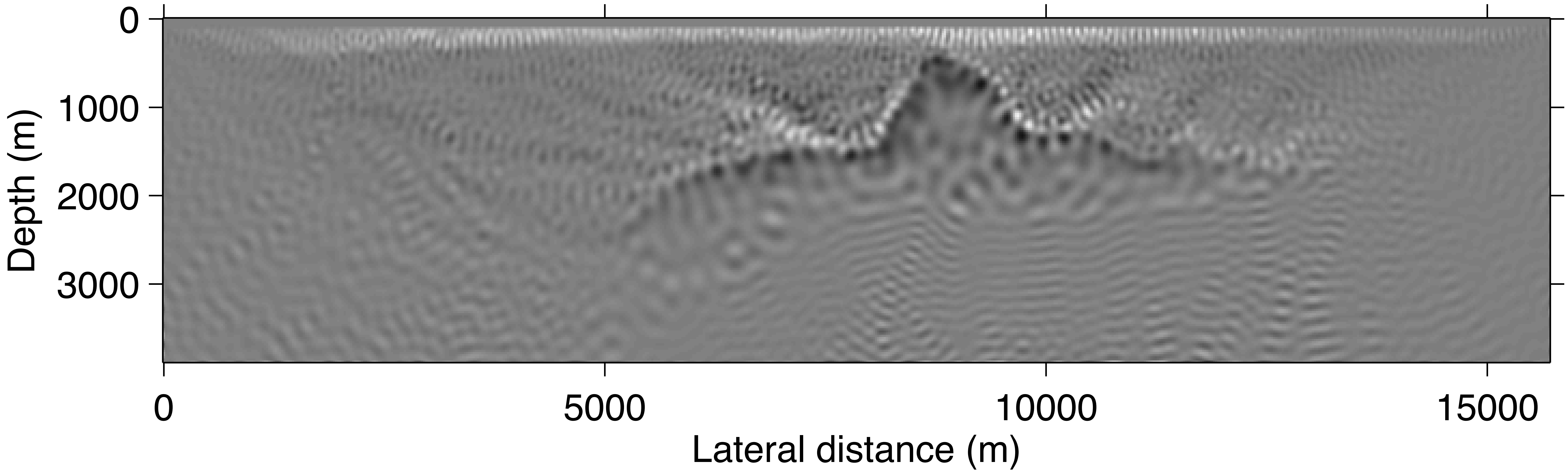
Migration
– 1 sim. shot & 1 frequencies



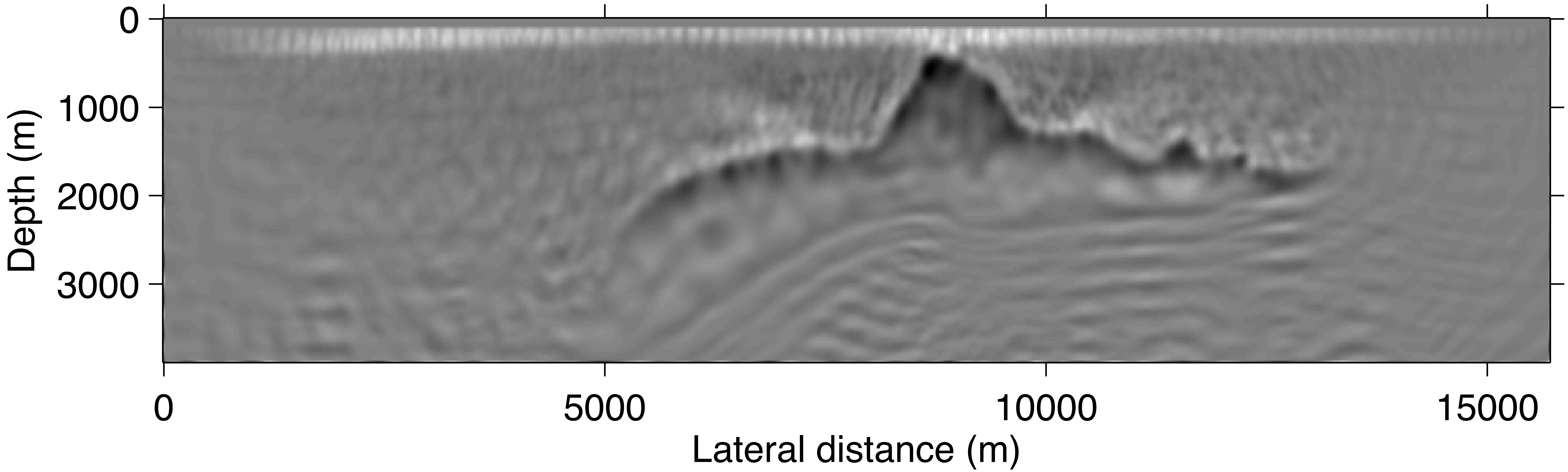
Migration
– 2 sim. shots & 2 frequencies



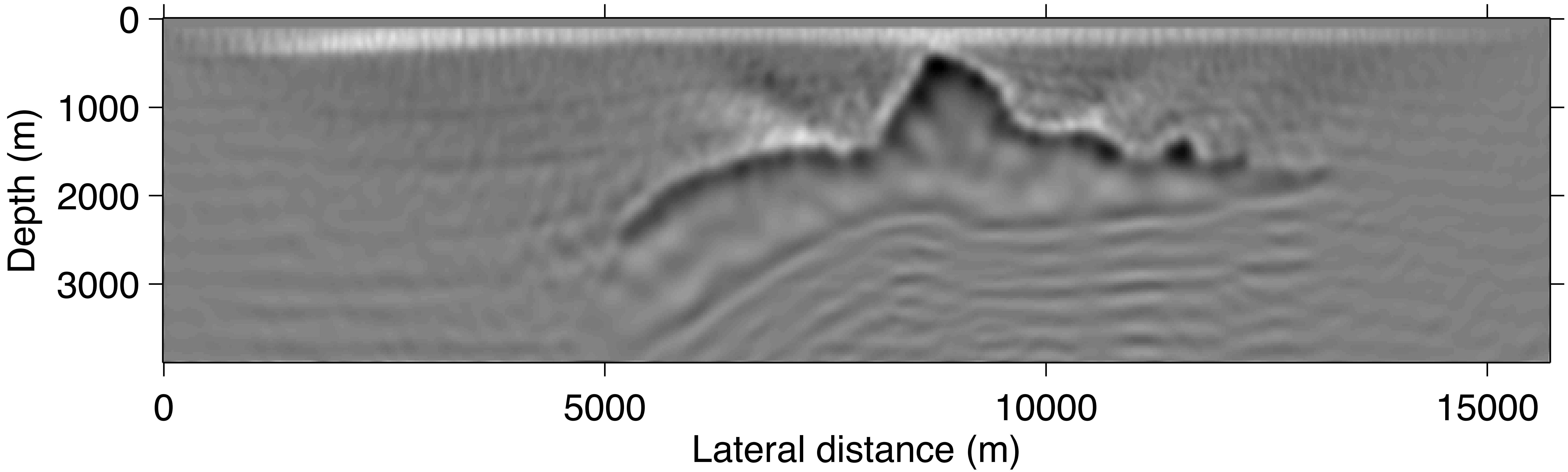
Migration
– 4 sim. shots & 4 frequencies



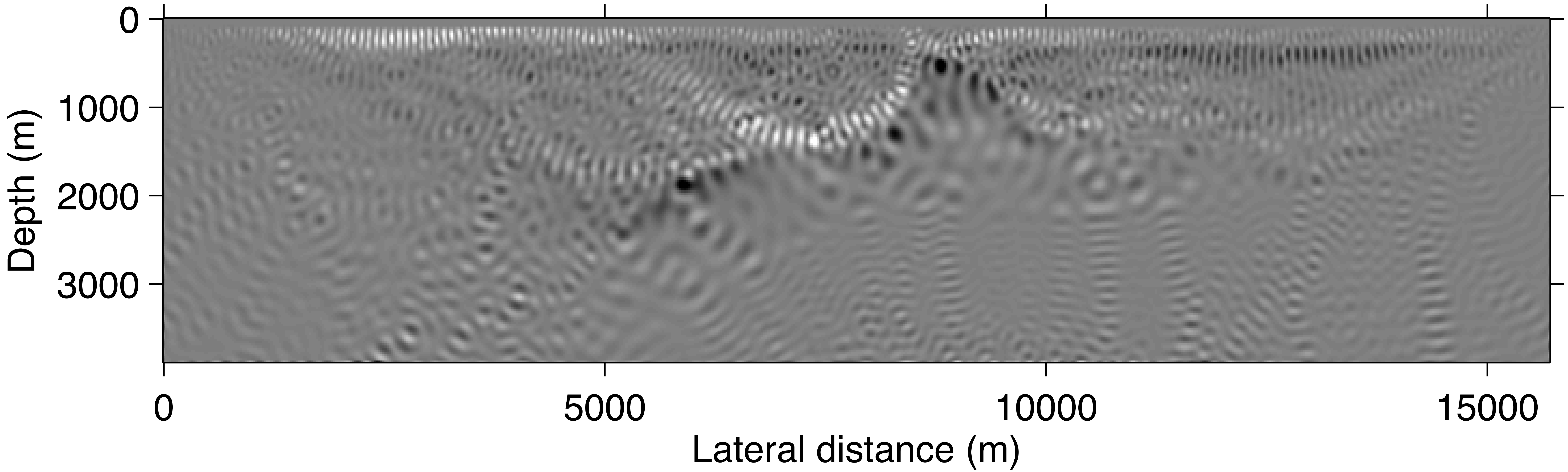
Migration
– 8 sim. shots & 8 frequencies



Migration
– 16 sim. shots & 16 frequencies

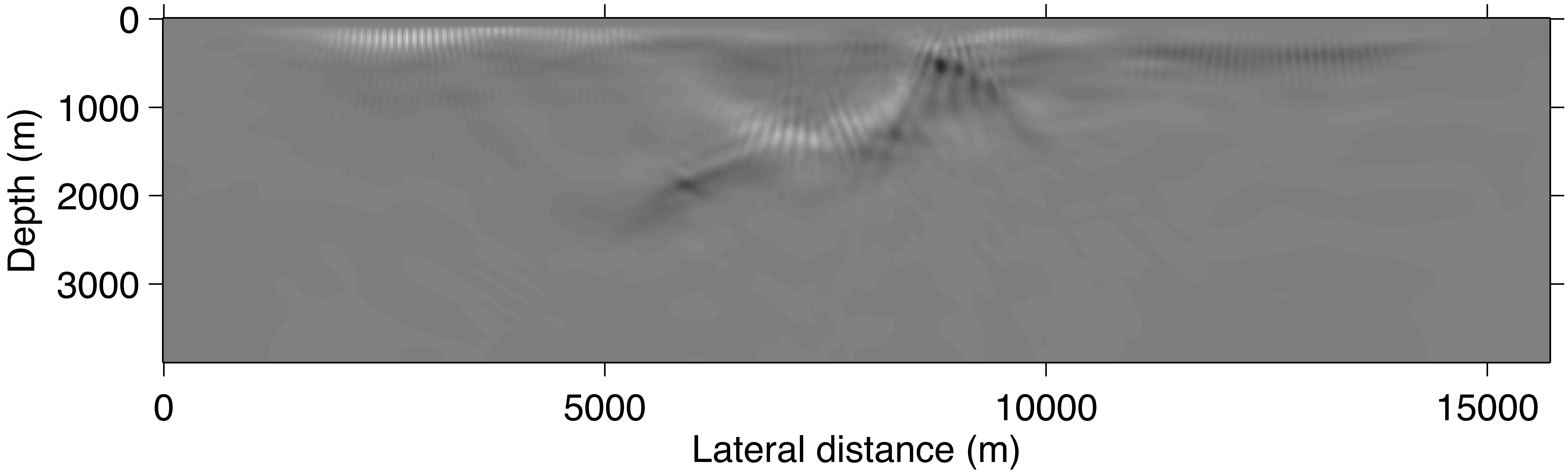


Migration
– 1 sim. shot & 1 frequencies

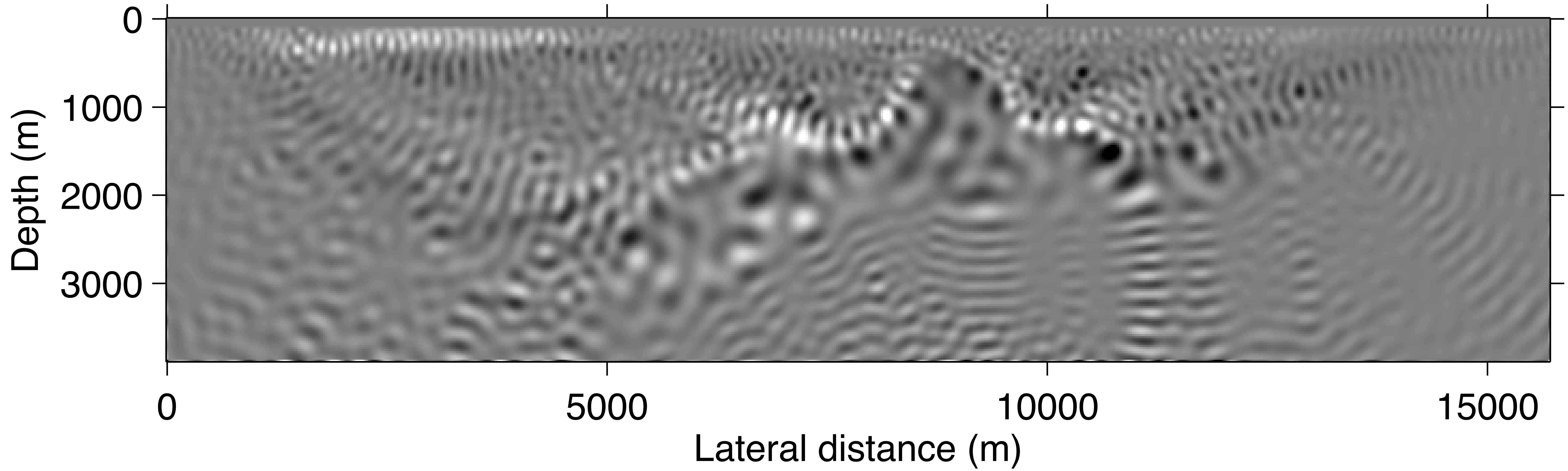


Migration + threshold

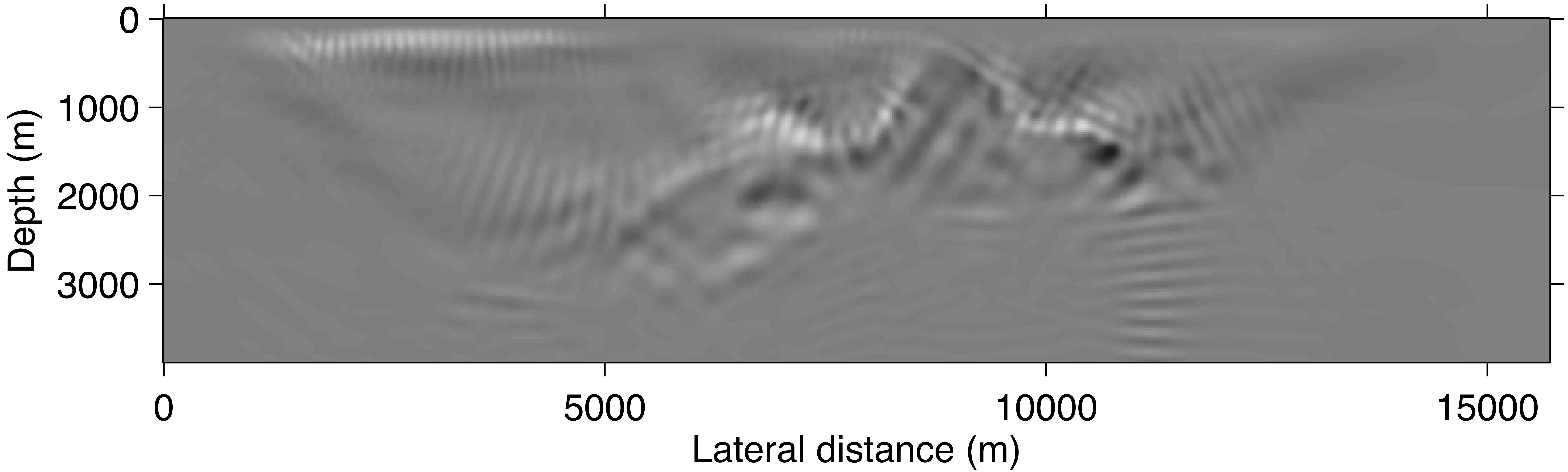
– 1 sim. shot & 1 frequencies



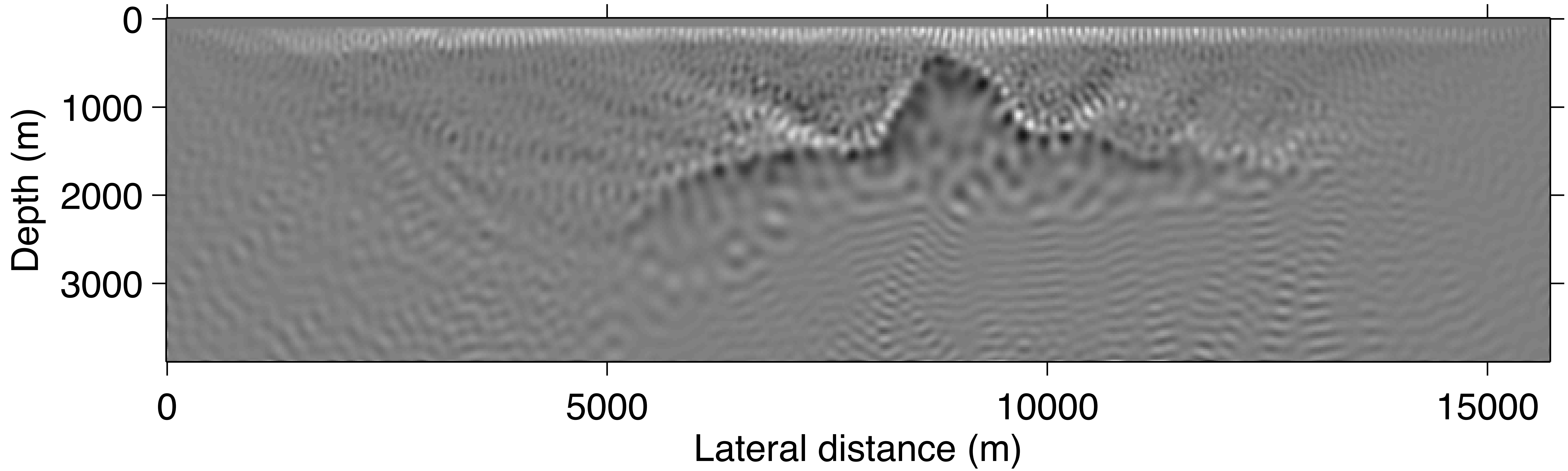
Migration
– 2 sim. shots & 2 frequencies



Migration + threshold
– 2 sim. shots & 2 frequencies

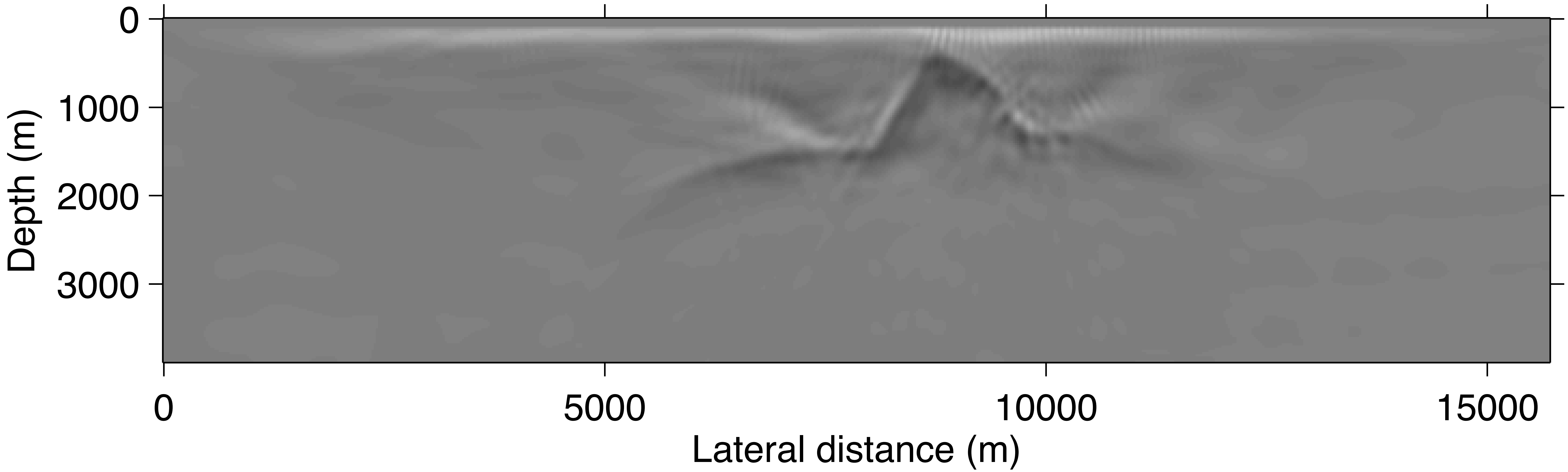


Migration
– 4 sim. shots & 4 frequencies

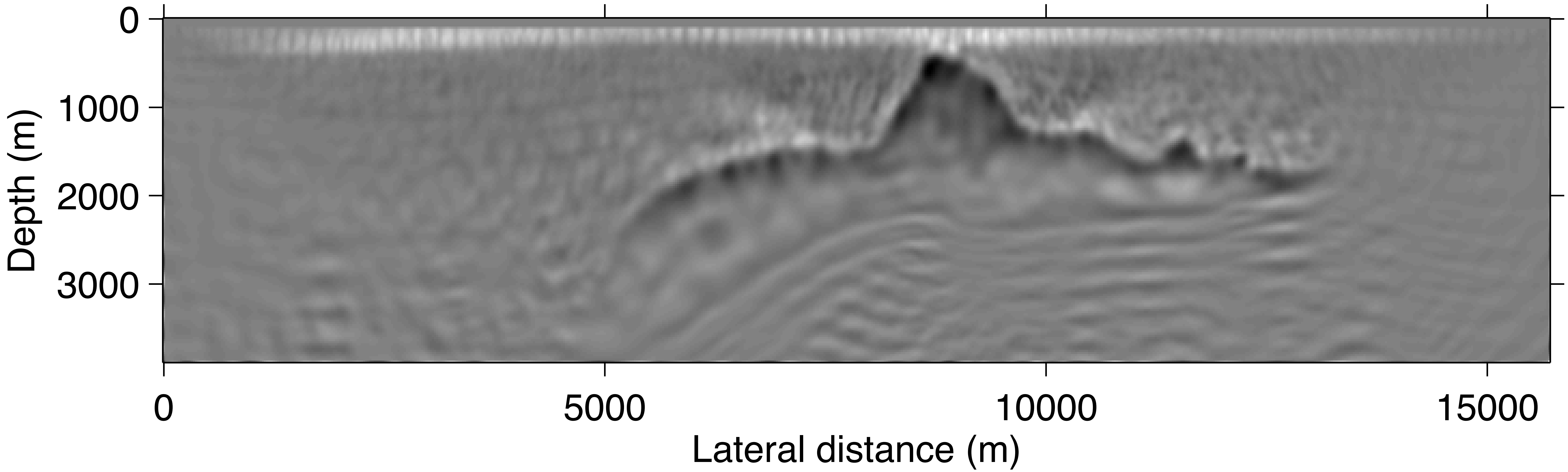


Migration + threshold

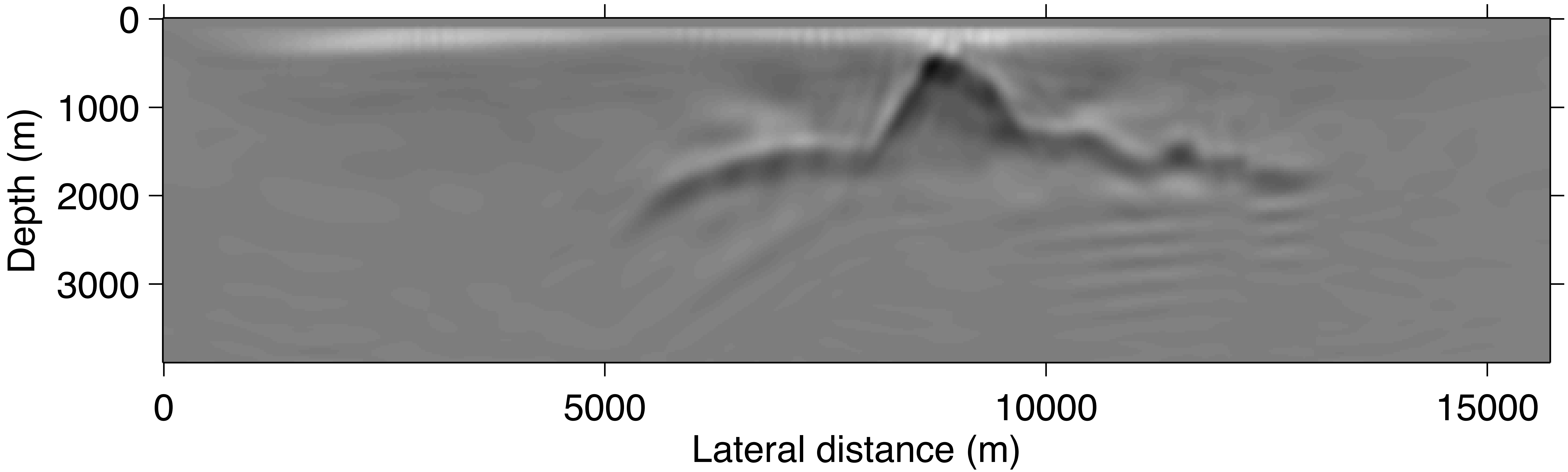
– 4 sim. shots & 4 frequencies



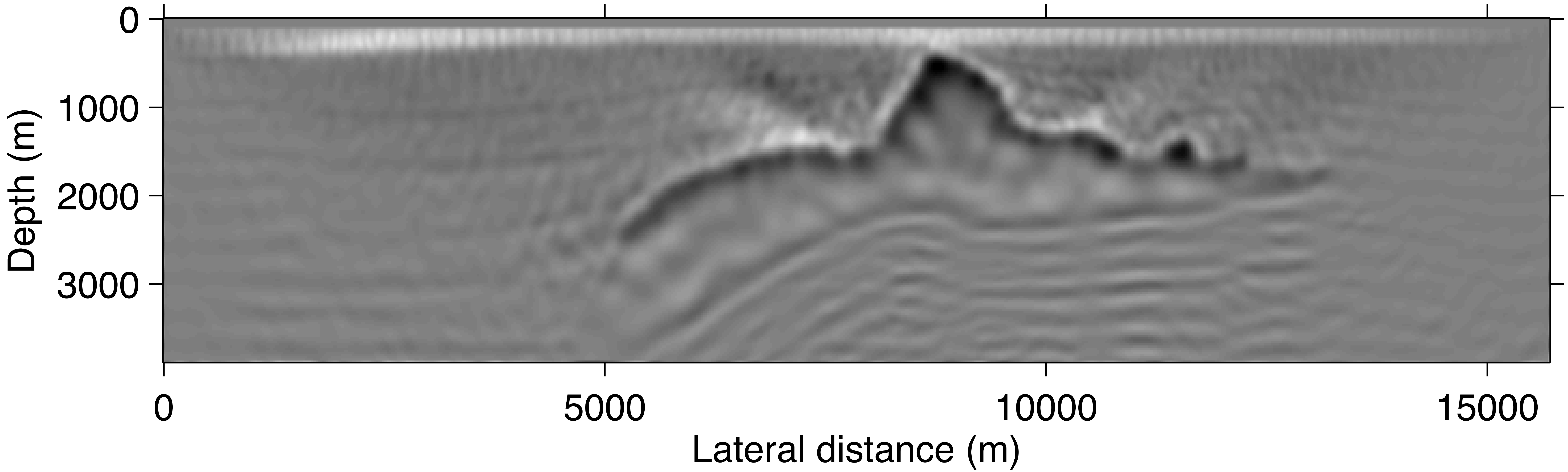
Migration
– 8 sim. shots & 8 frequencies



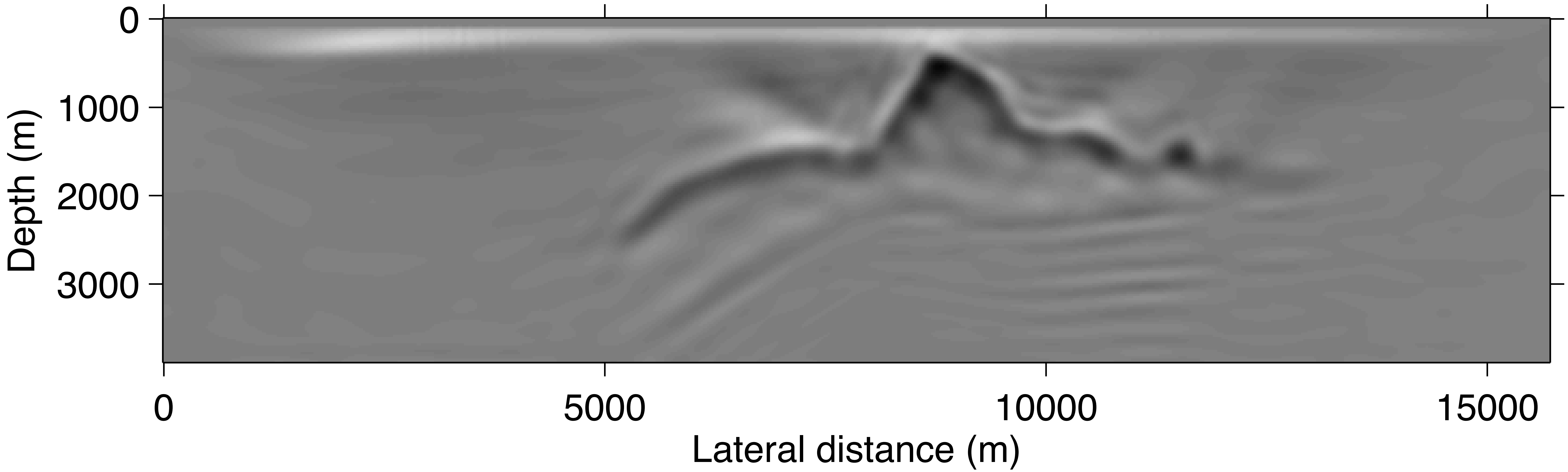
Migration + threshold
– 8 sim. shots & 8 frequencies



Migration
– 16 sim. shots & 16 frequencies



Migration + threshold
– 16 sim. shots & 16 frequencies



Felix J. Herrmann, Ning Tu, and Ernie Esser, "Fast "online" migration with Compressive Sensing". 2015

Approach

Use linearized Bregman Projections (LBP) to solve

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}\|_2 \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \end{aligned}$$

For λ large enough,

- converges to the solution of the ℓ_1 problem
- amenable to working with small subsets (combine with randomized Kaczmarz)

Dirk A Lorenz et. al., preprint, 2014

SLIM

Fast compressive imaging

Leverage sparse randomized block-Kaczmarz solver:

obtain submatrix $\underline{\mathbf{A}}^k$ and data $\underline{\mathbf{b}}^k$ at the k^{th} iteration

compute residual of the previous step $\mathbf{r}^k = \underline{\mathbf{A}}^k \mathbf{x}^k - \underline{\mathbf{b}}$

compute gradient $\mathbf{g}^k = \underline{\mathbf{A}}^{k'} \mathbf{r}^k$

compute steplength $t^k = \frac{\|\mathbf{r}^k\|_2^2}{\|\mathbf{g}^k\|_2^2}$

gradient descent $\mathbf{z}^{k+1} = \mathbf{z}^k - t^k \mathbf{g}^k$

soft thresholding $\mathbf{x}^{k+1} = \mathcal{S}_\lambda(\mathbf{z}^{k+1})$

LBP vs. SPGl1: computer codes

LBP vs. SPGl1: computer codes

```
r = b_sub - A_sub*x;
g = A_sub'*r;
rnorm = norm(r,2);
gnorm = norm(g,2);
sl = (rnorm/gnorm)^2;
z = z+sl*g;
x = sign(x).*max(0,abs(x)-lambda);
```

LBP

E. van den Berg and M. Friedlander, 2008

LBP vs. SPGl1: computer codes

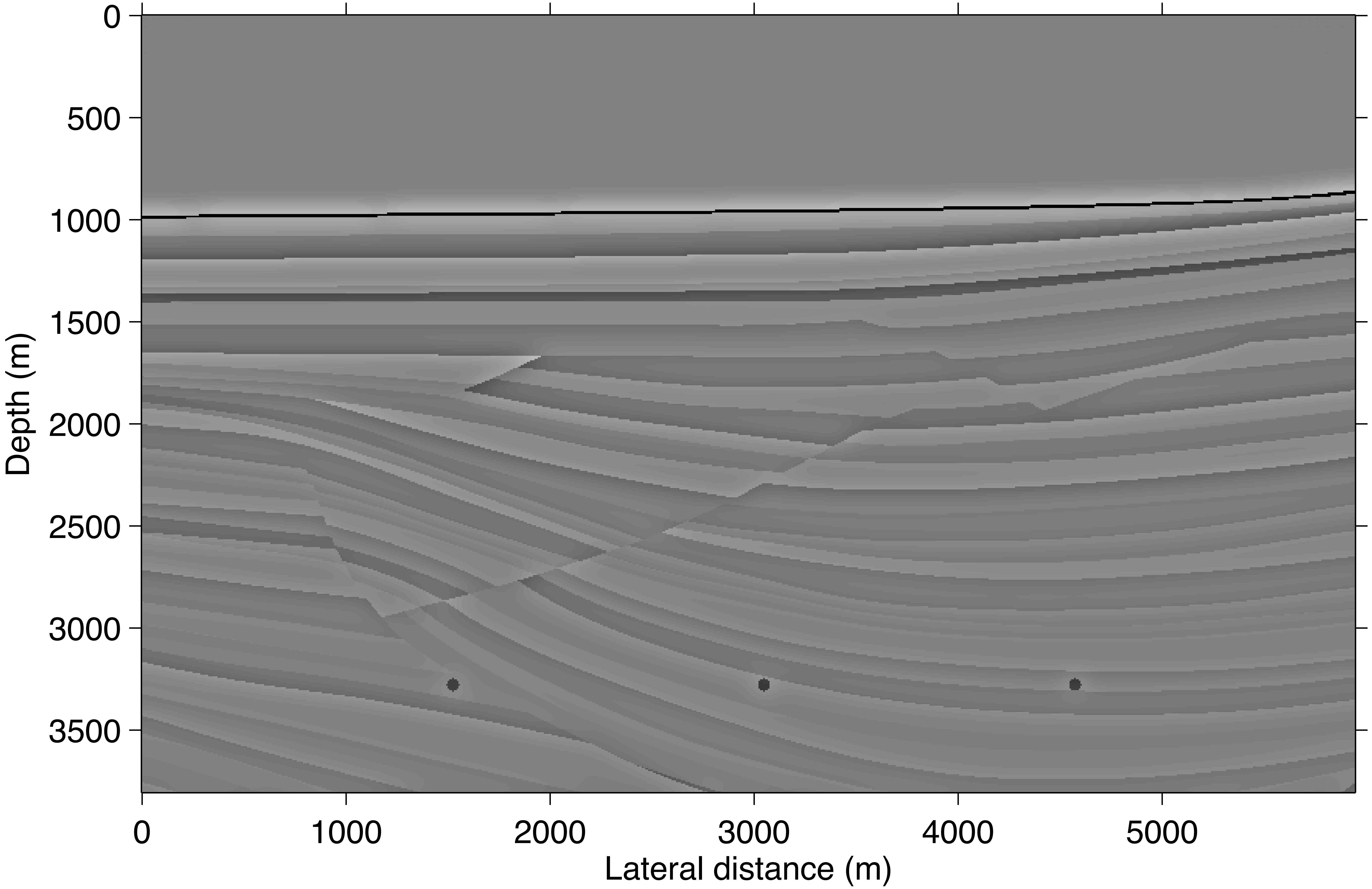
```
r = b_sub - A_sub*x;
g = A_sub'*r;
rnorm = norm(r,2);
gnorm = norm(g,2);
sl = (rnorm/gnorm)^2;
z = z+sl*g;
x = sign(x).*max(0,abs(x)-lambda);
```

LBP

```
function [x,r,g,info] = spgl1( A, b, tau, sigma, x, options )

m = length(b);

%----------------------------------------------------------------------
% Check arguments.
%----------------------------------------------------------------------
if ~exist('options','var'), options = []; end
if ~exist('x','var'), x = []; end
if ~exist('sigma','var'), sigma = []; end
if ~exist('tau','var'), tau = []; end

if nargin < 2 || isempty(b) || isempty(A)
   error('At least two arguments are required');
elseif isempty(tau) && isempty(sigma)
   tau = 0;
   sigma = 0;
   singleTau = false;
elseif isempty(sigma) % && ~isempty(tau)  <-- implied
   singleTau = true;
else
   if isempty(tau)
      tau = 0;
   end
   singleTau = false;
end

%----------------------------------------------------------------------
% Grab input options and set defaults where needed.
%----------------------------------------------------------------------
defaultopts = spgSetParms(...
'fid'        ,      1 , ... % File ID for output
'verbosity'  ,      2 , ... % Verbosity level
'iterations' ,   10*m , ... % Max number of iterations
'nPrevVals'  ,      3 , ... % Number previous func values for linesearch
'bpTol'      ,  1e-06 , ... % Tolerance for basis pursuit solution
'optTol'     ,  1e-04 , ... % Optimality tolerance
'decTol'     ,  1e-04 , ... % Req'd rel. change in primal obj. for Newton
'stepMin'    ,  1e-16 , ... % Minimum spectral step
'stepMax'    ,  1e+05 , ... % Maximum spectral step
'rootMethod' ,      2 , ... % Root finding method: 2=quad,1=linear (not used).
'activeSetIt',    Inf , ... % Exit with EXIT_ACTIVE_SET if nnz same for # its.
'subspaceMin',      0 , ... % Use subspace minimization
'iscomplex'  ,    NaN , ... % Flag set to indicate complex problem
'maxMatvec'  ,    Inf , ... % Maximum matrix-vector multiplies allowed
'weights'    ,      1 , ... % Weights W in ||Wx||_1
'Kaczmarz'   ,      0 , ... % Toggles whether Kaczmarz mode is on (experimental)
'KaczScale'  ,      1 , ... % Scaling factor for Tau when using Kaczmarz-type submatrices
'quitPareto' ,      0 , ... % Exits when pareto curve is reached
'minPareto'  ,      3 , ... % If quitPareto is on, the minimum number of iterations before checking for quitPareto conditions
'lineSrchIt' ,      1 , ... % Maximum number of line search iterations for spgLineCurvy, originally 10 ...
'feasSrchIt' ,  10000 , ... % Maximum number of feasible direction line search iteraitons, originally 10 ...
'ignorePErr' ,      0 , ... % Ignores projections error by issuing a warning instead of an error ...
'project'    , @NormL1_project , ...
'primal_norm', @NormL1_primal  , ...
'dual_norm'  , @NormL1_dual      ...
   );
options = spgSetParms(defaultopts, options);

fid           = options.fid;
logLevel      = options.verbosity;
maxIts        = options.iterations;
nPrevVals     = options.nPrevVals;
bpTol         = options.bpTol;
optTol        = options.optTol;
decTol        = options.decTol;
stepMin       = options.stepMin;
stepMax       = options.stepMax;
activeSetIt   = options.activeSetIt;
subspaceMin   = options.subspaceMin;
maxMatvec     = max(3,options.maxMatvec);
weights       = options.weights;
quitPareto    = options.quitPareto;
minPareto     = options.minPareto;
Kaczmarz      = options.Kaczmarz;
lineSrchIt    = options.lineSrchIt;
feasSrchIt    = options.feasSrchIt;
ignorePErr    = options.ignorePErr;

% maxLineErrors TEMPORARILY DISABLED to prevent very lar[illegible] issue[illegible] algorithm
maxLineErrors = Inf;     % Maximum number of line-search failures.
pivTol        = 1e-12;  % Threshold for significant Newton step.

%----------------------------------------------------------------------
% Initialize local variables.
%----------------------------------------------------------------------
iter          = 0;  itnTotLSQR = 0; % Total SPGL1 and LSQR iterations.
nProdA        = 0;  nProdAt    = 0;
lastFv        = -inf(nPrevVals,1);  % Last m function values.
nLineTot      = 0;                  % Total no. of linesearch steps.
printTau      = false;
nNewton       = 0;
bNorm         = norm(b,2);
stat          = false;
timeProject   = 0;
timeMatProd   = 0;
nnzIter       = 0;                  % No. of its with fixed pattern.
nnzIdx        = [];                 % Active-set indicator.
subspace      = false;              % Flag if did subspace min in current itn.
stepG         = 1;                  % Step length for projected gradient.
testUpdateTau = 0;                  % Previous step did not update tau

% Determine initial x, vector length n, and see if problem is complex
explicit = ~(isa(A,'function_handle'));
```

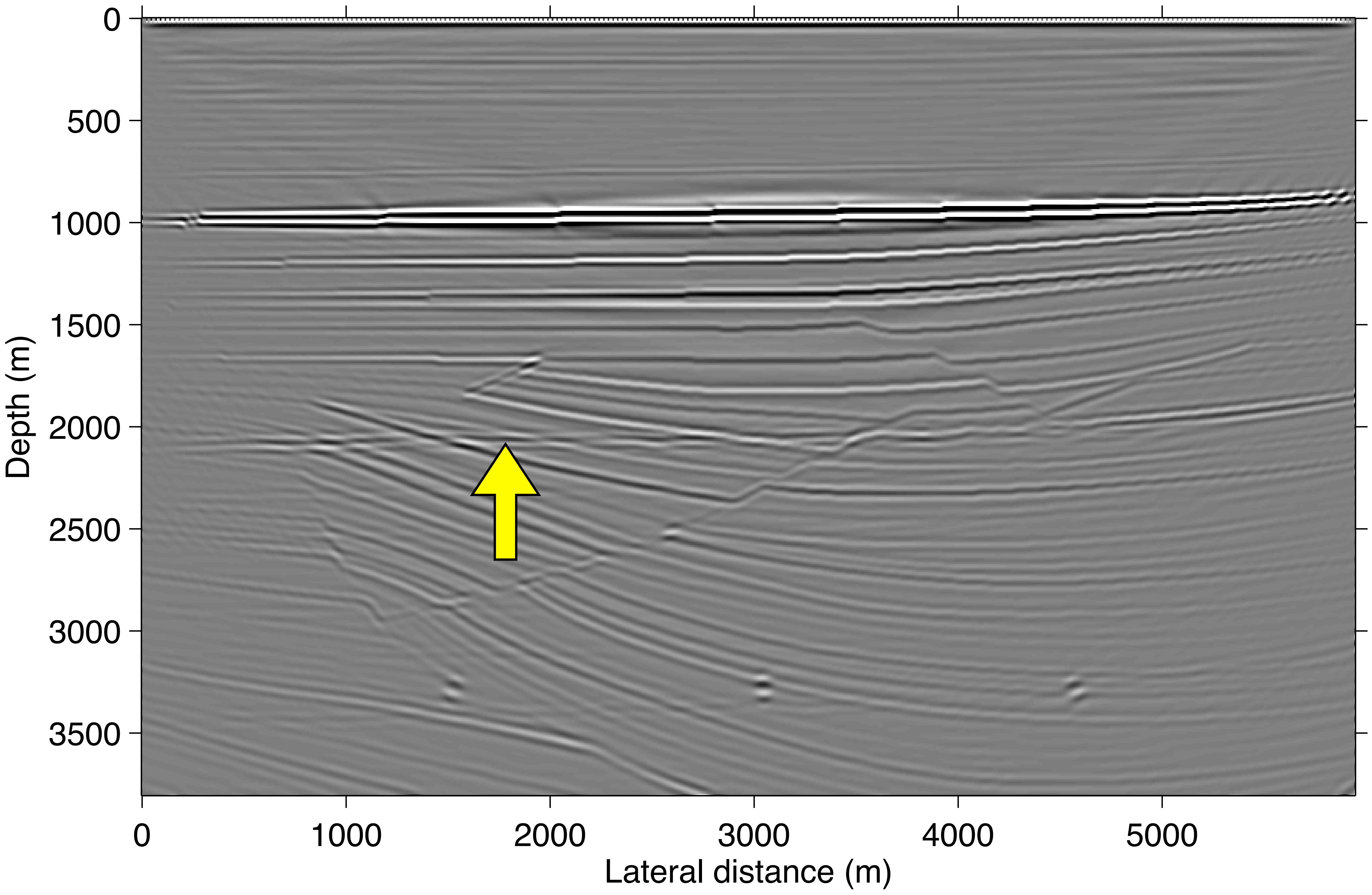
SPGl1

Fast imaging with multiples: true image

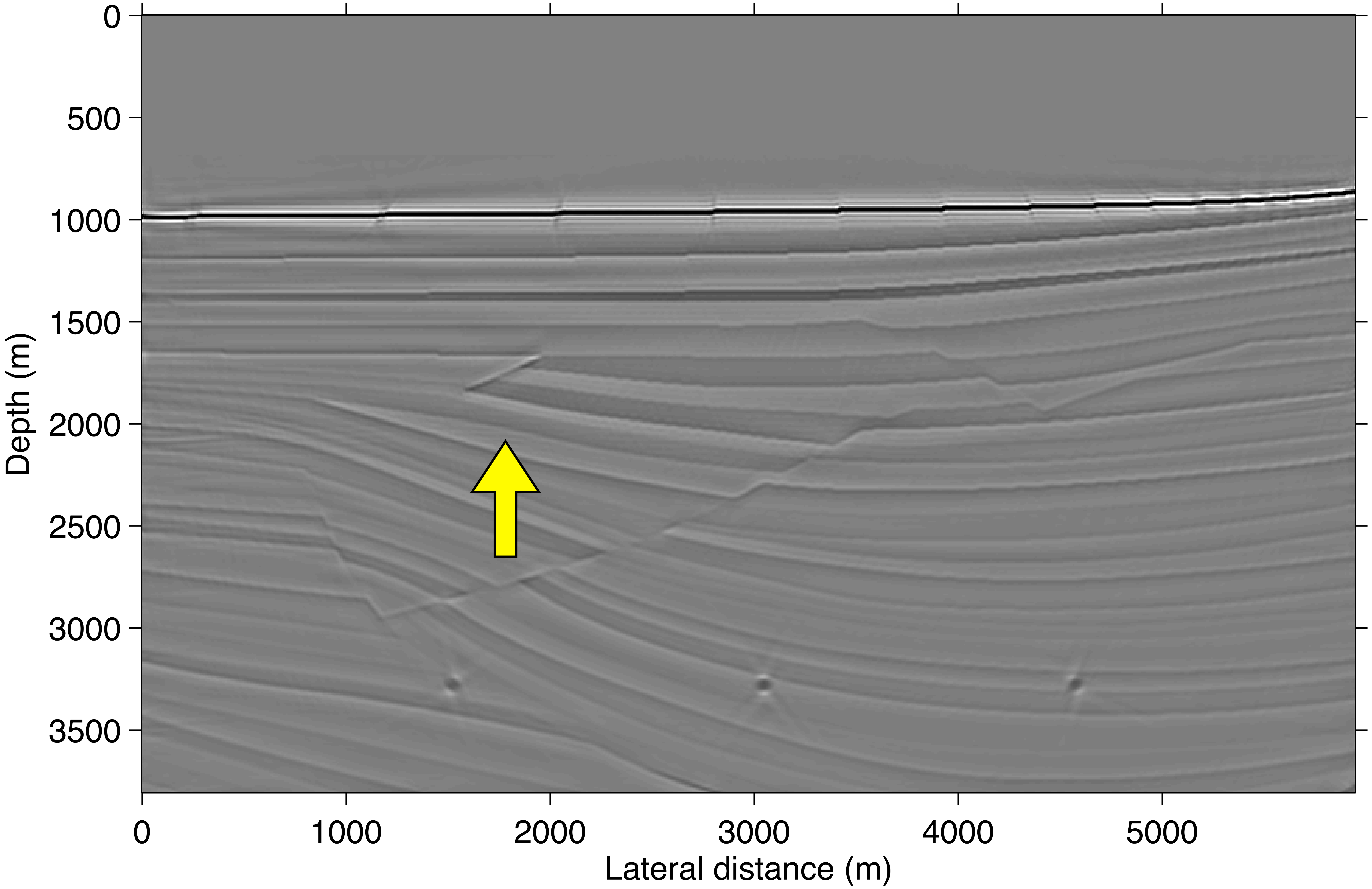


Ning Tu and Felix J. Herrmann, "**Fast imaging with surface-related multiples by sparse inversion**". 2015.

RTM w/ multiples



Fast imaging with multiples & **LBP**, final image



Simulation cost **~2 RTMs** w/ all data

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

But, requires

- densely sampled data
- good velocity models...

SLIM

Motivation

Full-waveform inversion is plagued with local minima

Derive an alternative extended formulation

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

http://fc00.deviantart.net/fs70/i/2012/066/c/d/nc__elephant_in_the_room_by_bthomas64-d4s05d9.jpg

Wavefield Reconstruction Inversion

Tristan van Leeuwen, Bas Peters, and Ernie Esser



SLIM
University of British Columbia

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

SLIM

Waveform inversion

Adjoint-state/reduced-space methods:

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

Full-space or all-at-once methods:

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

Waveform inversion

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

"known" physics → $A(\mathbf{m})$; "known" source → $\mathbf{q}_i$

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

"unknown" wavefield → $\mathbf{u}_i$

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

[Richter, '81]

SLIM

Equation error approach

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

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

via

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

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

SLIM

WRI – Wavefield Reconstruction Inversion

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

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

and wave-equations

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

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

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

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

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

wave-equation X wavefield = source

versus

$$\left(\frac{\text{wave-equation}}{\color{red}{\text{sampling operator}}}\right) \times \text{wavefield} = \left(\frac{\text{source}}{\color{red}{\text{data}}}\right)$$

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

Wavefield Reconstruction Inversion

WRI method

for each source i

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

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

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

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

end

diagonal Hessian = pseudo Hessian

Conventional method

for each source i

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

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

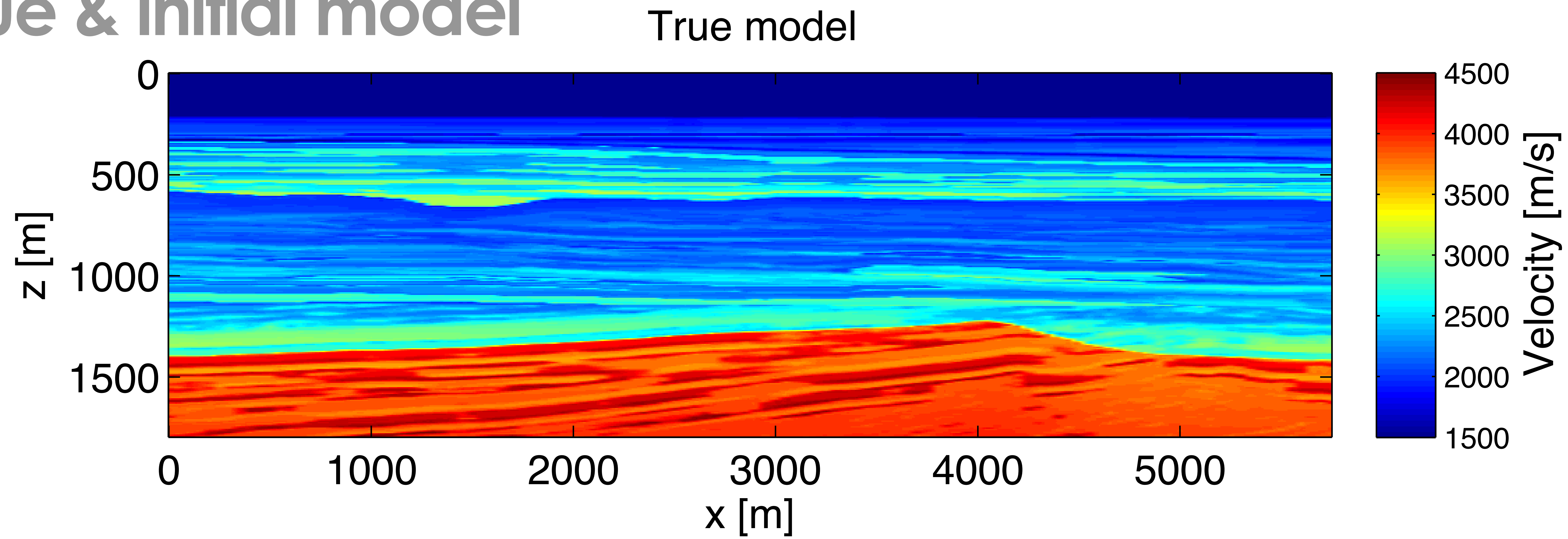
$\mathbf{g} = \mathbf{g} + \omega^2 \text{diag}(\mathbf{u}_i)^*\mathbf{v}_i$

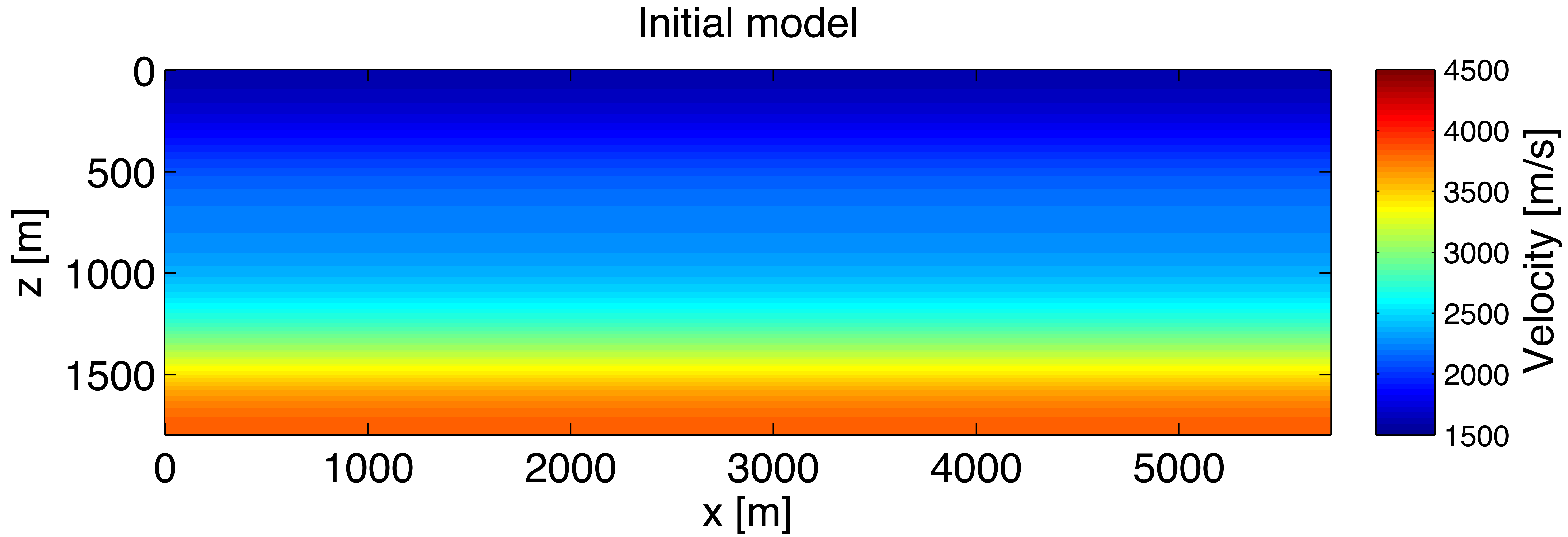
$\mathbf{m} = \mathbf{m} - \alpha\mathbf{g}$

end

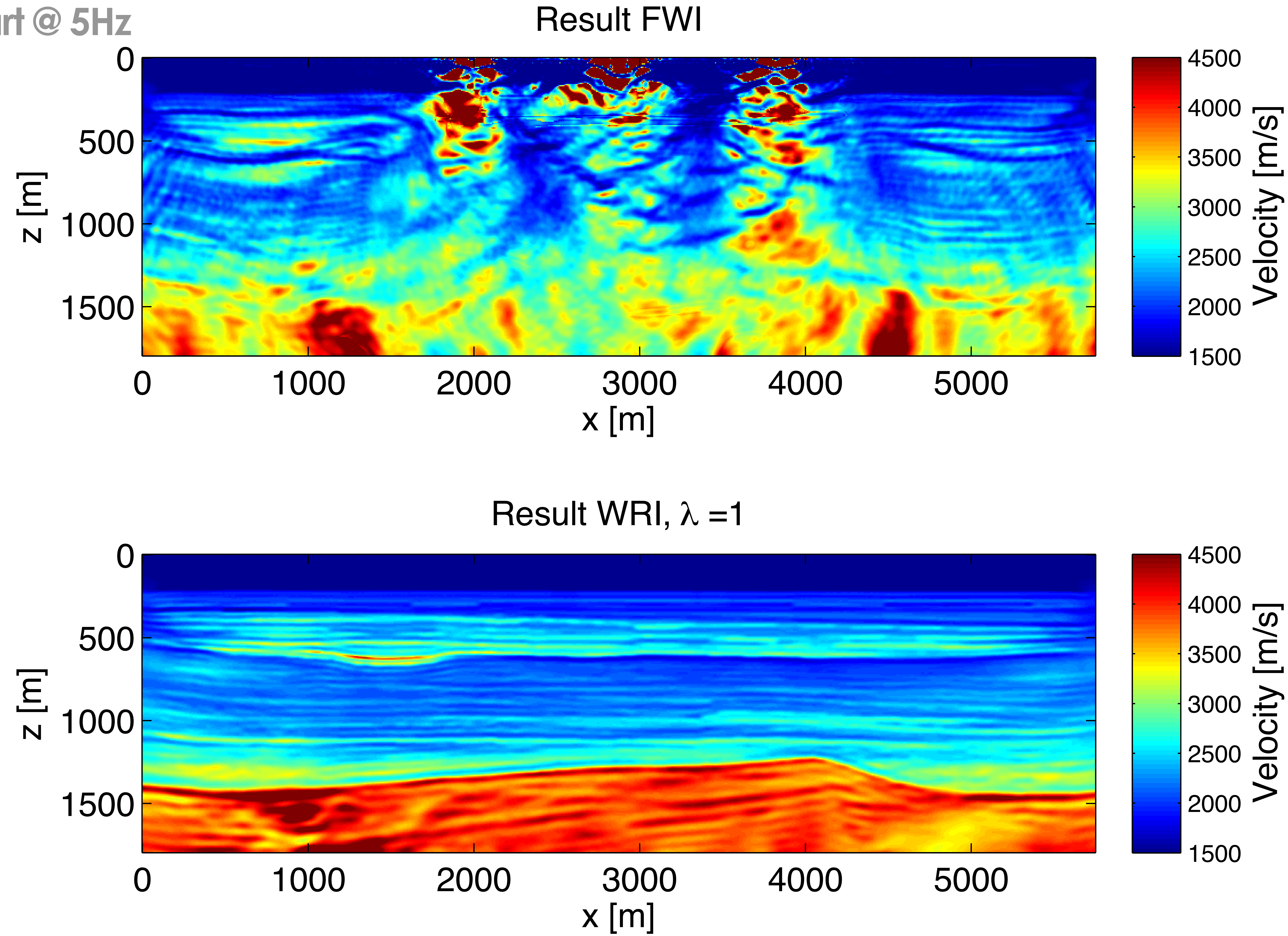
dense Hessian & too expensive

True & initial model

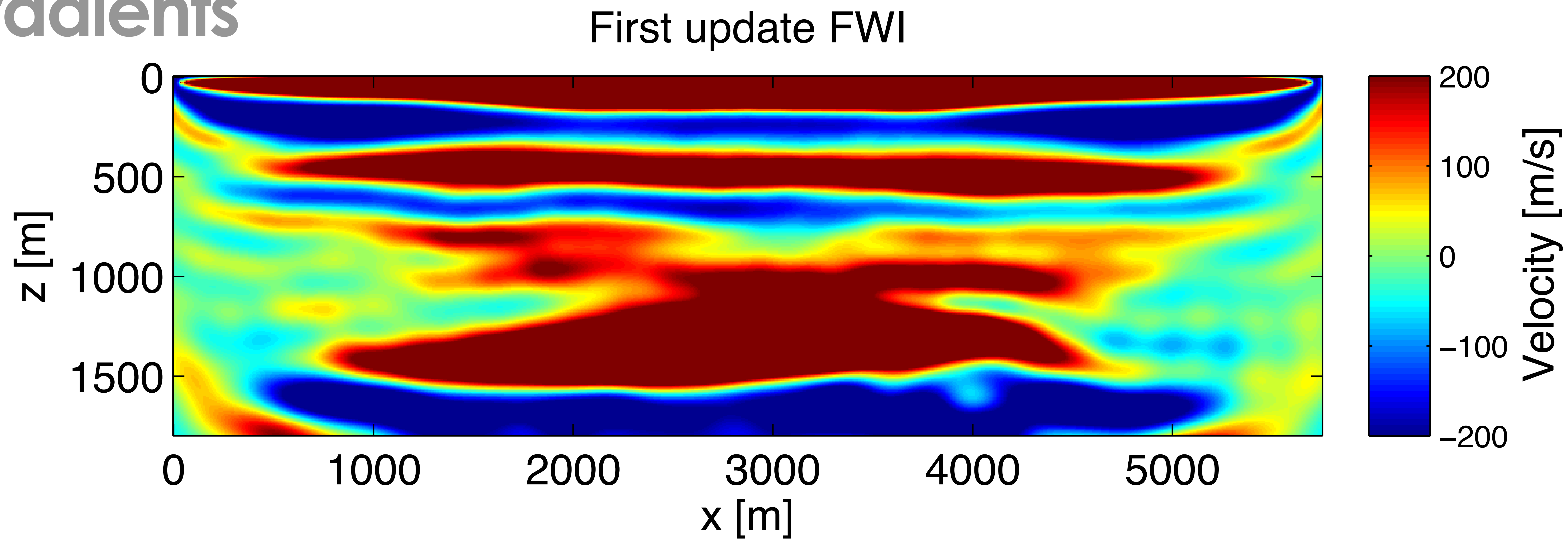




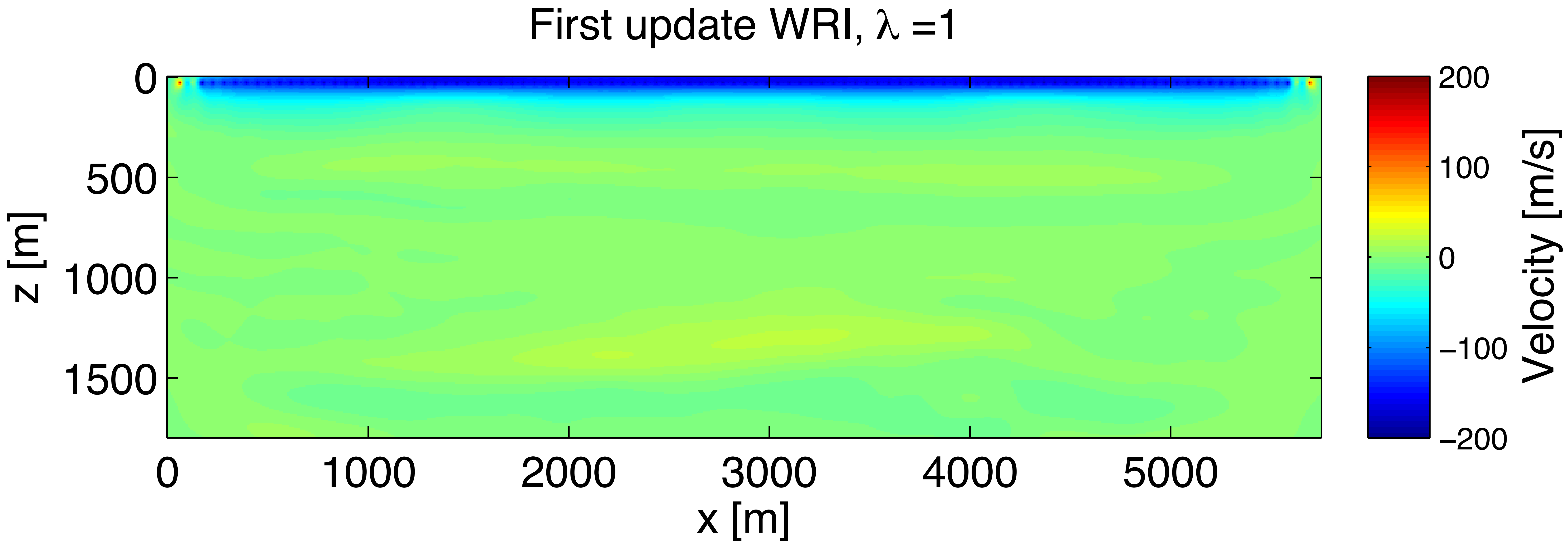
FWI vs WRI
– start @ 5Hz



Gradients

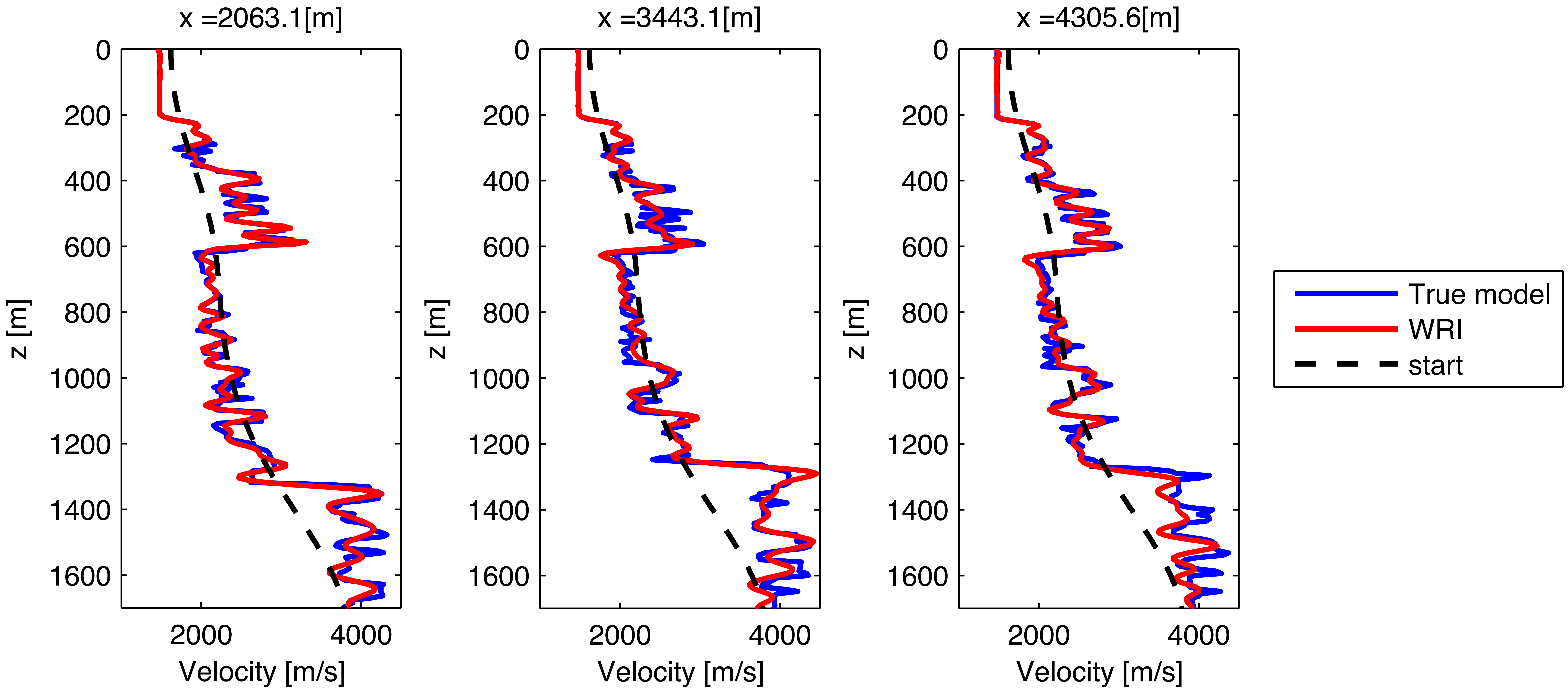


First update FWI
0
500
1000
1500
z [m]
0
1000
2000
3000
4000
5000
x [m]
200
100
0
−100
−200
Velocity [m/s]



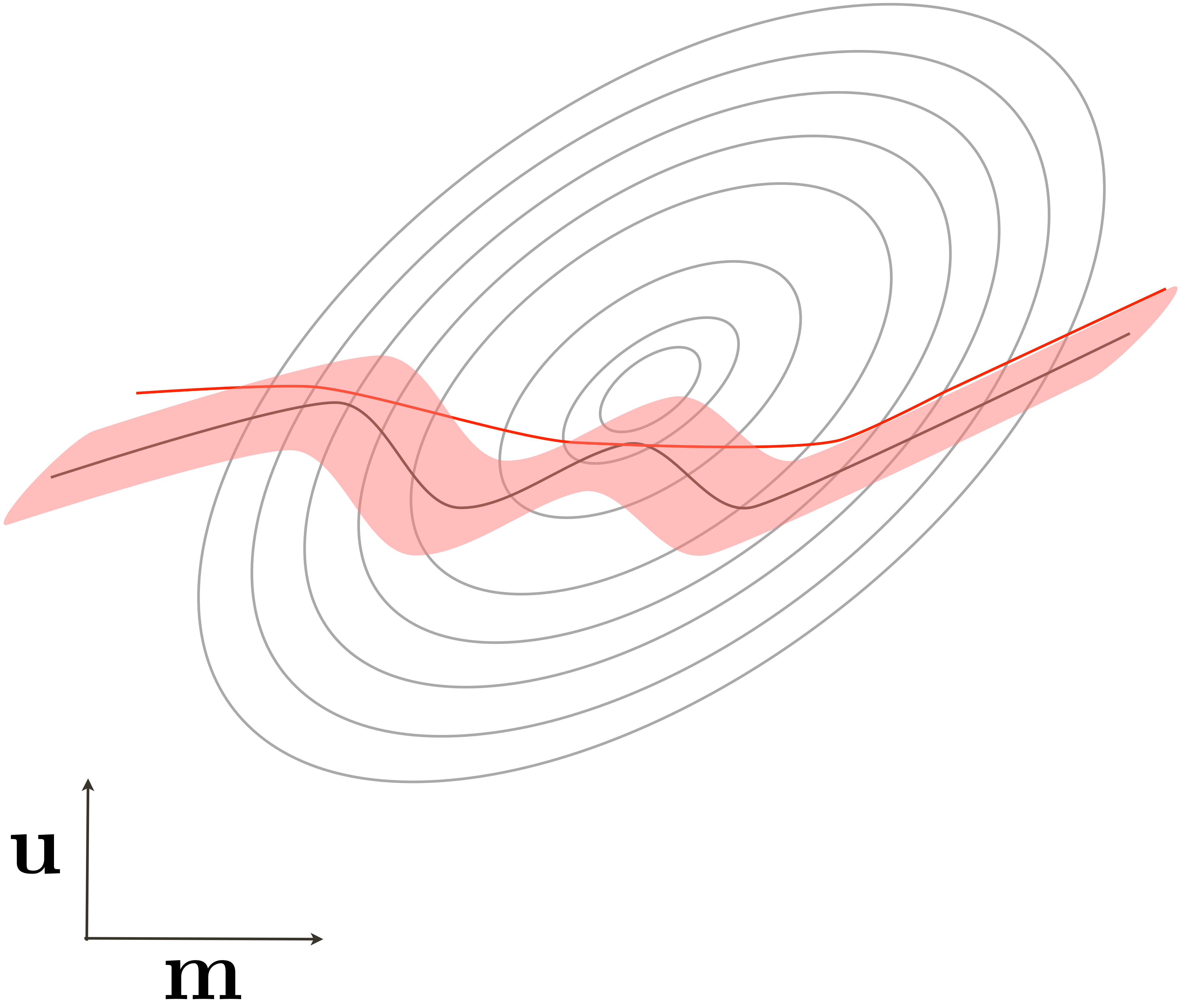
First update WRI, λ =1
0
500
1000
1500
z [m]
0
1000
2000
3000
4000
5000
x [m]
200
100
0
−100
−200
Velocity [m/s]

Cross sections

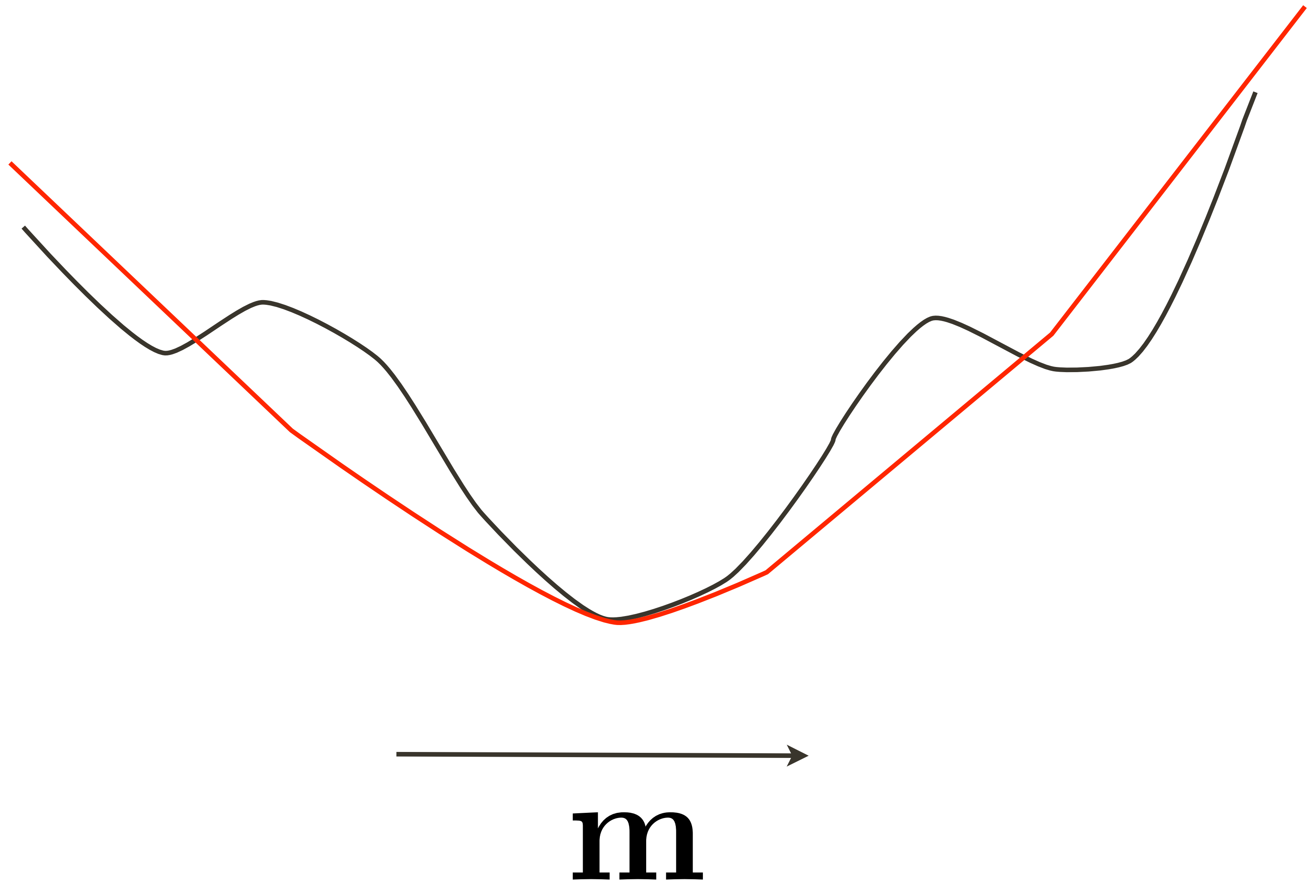


WRI vs. FWI

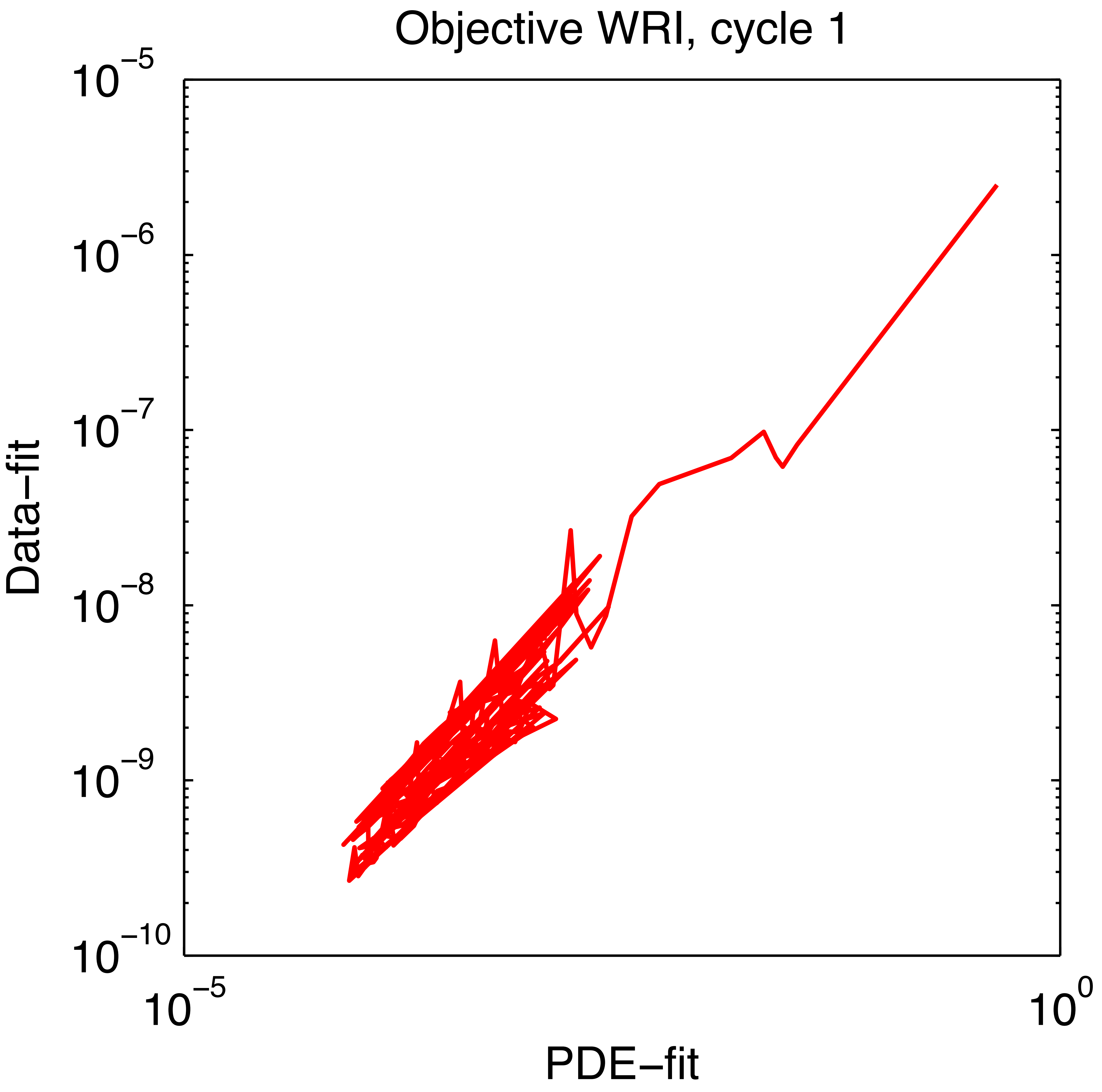
Larger # of degrees of freedom

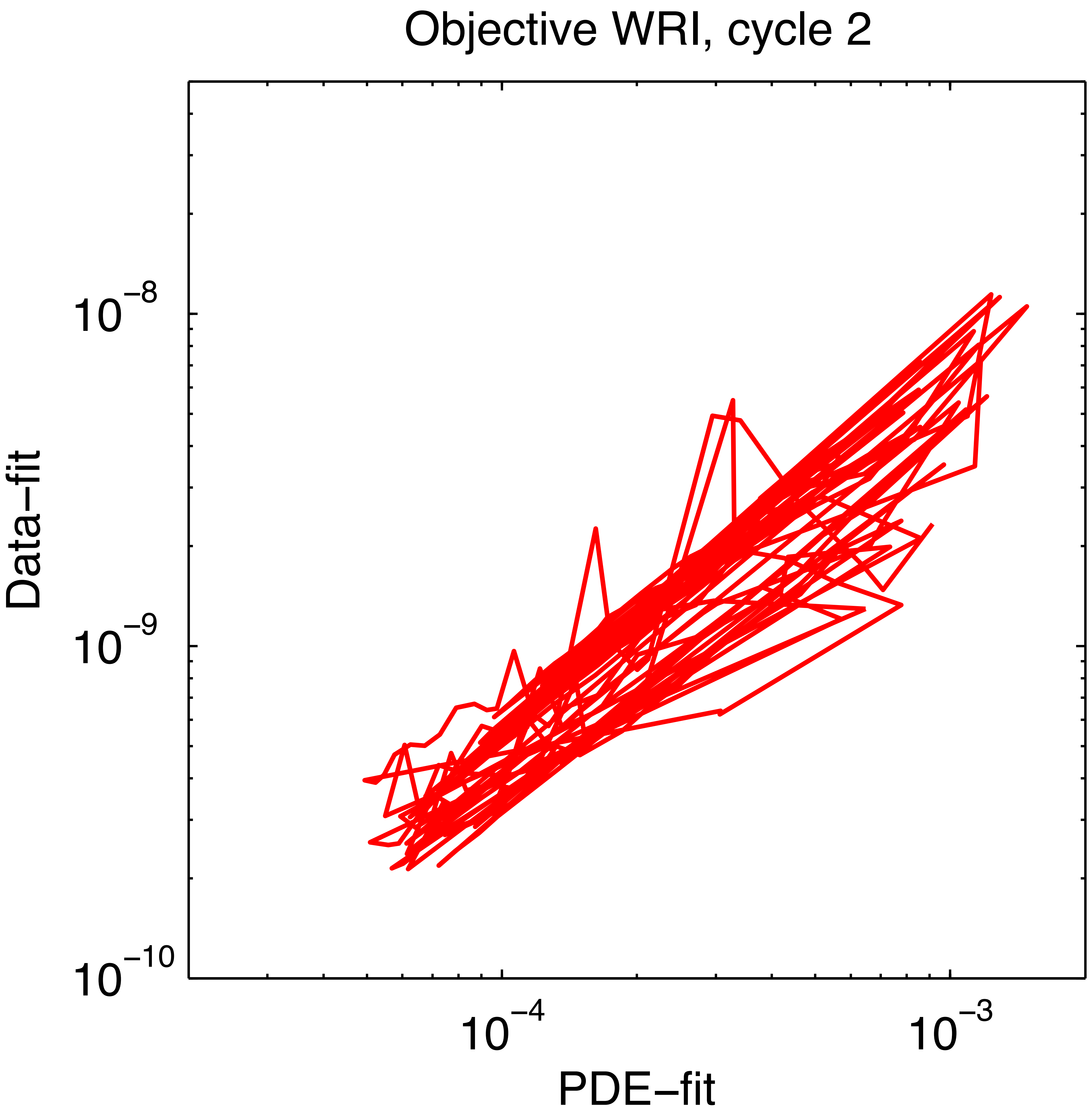


"more convex"



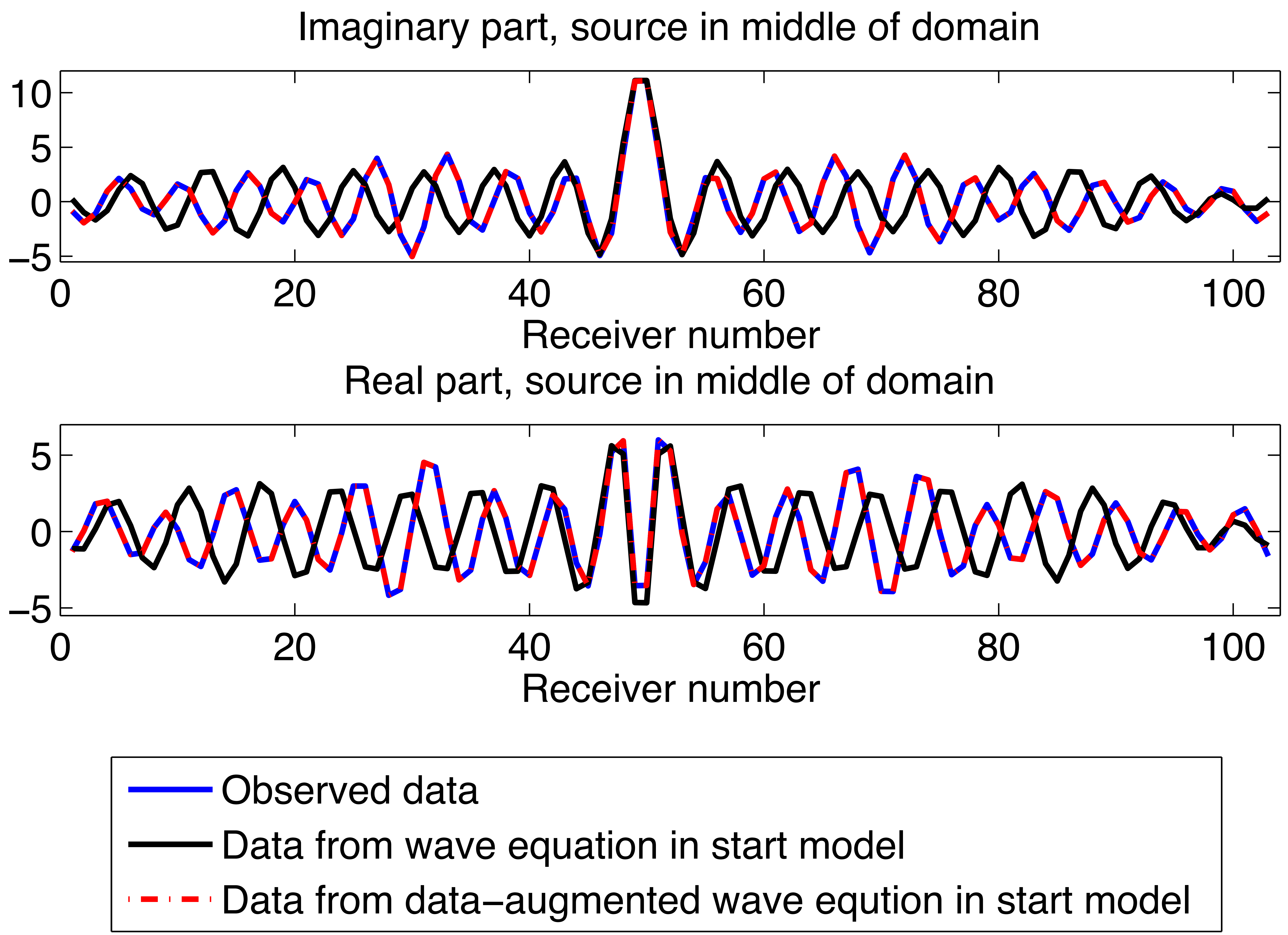
Objective function value





Data fit increases at some iterates

Data fit


Imaginary part, source in middle of domain
10
5
0
−5
0
20
40
60
80
100
Receiver number
Real part, source in middle of domain
5
0
−5
0
20
40
60
80
100
Receiver number
Observed data
Data from wave equation in start model
Data from data−augmented wave eqution in start model

Scaled-gradient projections
– w/ Total Variation & Bound Constraints

Solve

$$\min_{\mathbf{m}} \mathbf{g}(\mathbf{m}) \quad \text{subject to} \quad \begin{cases} m_i \in [B_1, B_2] \\ \|\mathbf{m}\|_{TV} \leq \tau \end{cases}$$

with $\|\mathbf{m}\|_{TV} = \sum_{ij} \frac{1}{h} \left\| \begin{bmatrix} (m_{i,j+1} - m_{i,j}) \\ (m_{i+1,j} - m_{i,j}) \end{bmatrix} \right\|$

by iterating (*outer* loop)

diagonal

$$\Delta\mathbf{m} = \arg\min_{\Delta\mathbf{m}} \Delta\mathbf{m}^T \nabla \mathbf{g}(\mathbf{m}) + \frac{1}{2} \Delta\mathbf{m}^T H_{GN} \Delta\mathbf{m} + c\Delta\mathbf{m}^T \Delta\mathbf{m}$$

$$\text{subject to} \quad \mathbf{m} + \Delta m_i \in [B_1, B_2] \text{ and } \|\mathbf{m} + \Delta\mathbf{m}\|_{TV} \leq \tau$$

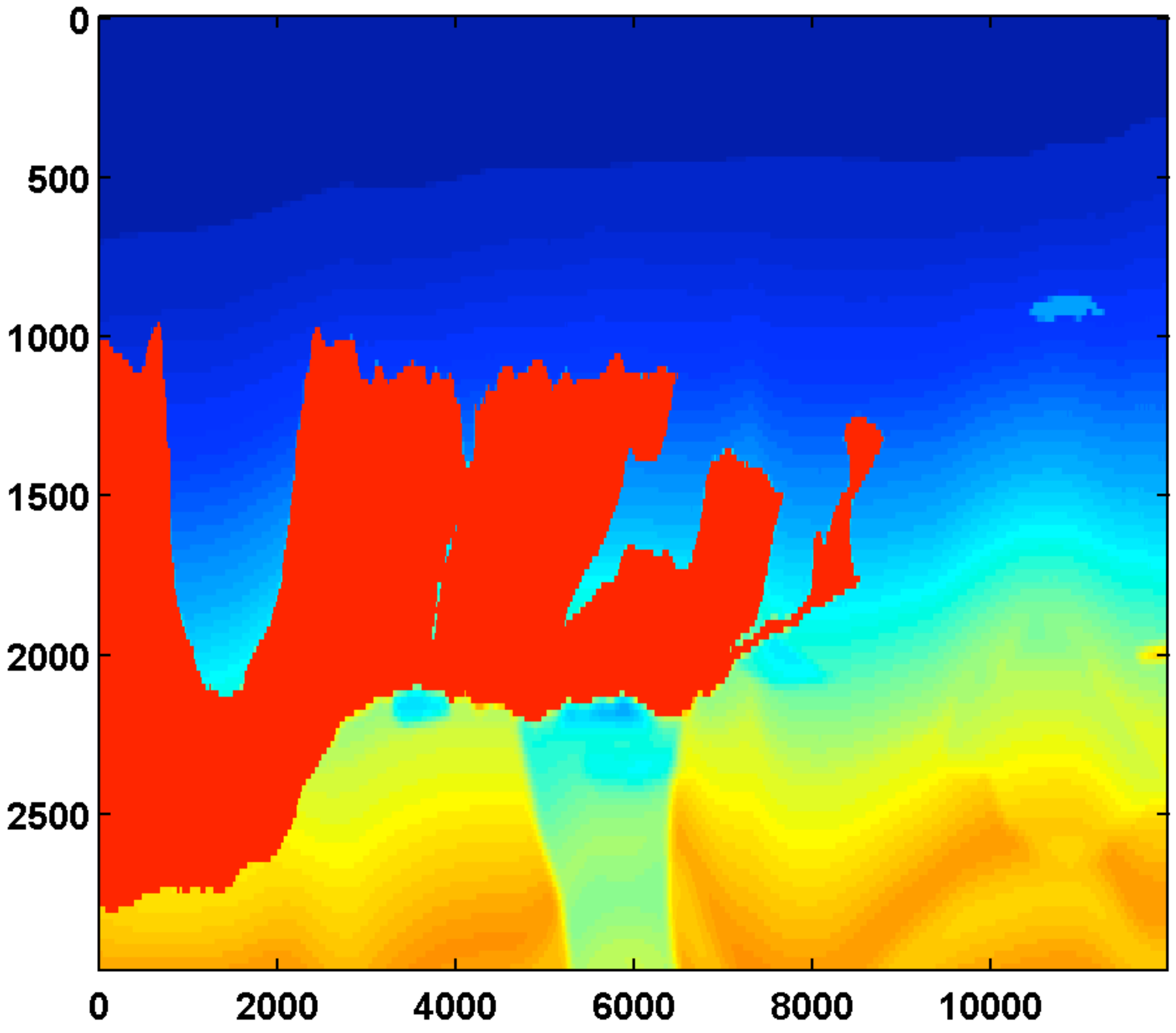
$$\mathbf{m} = \mathbf{m} + \Delta\mathbf{m}$$

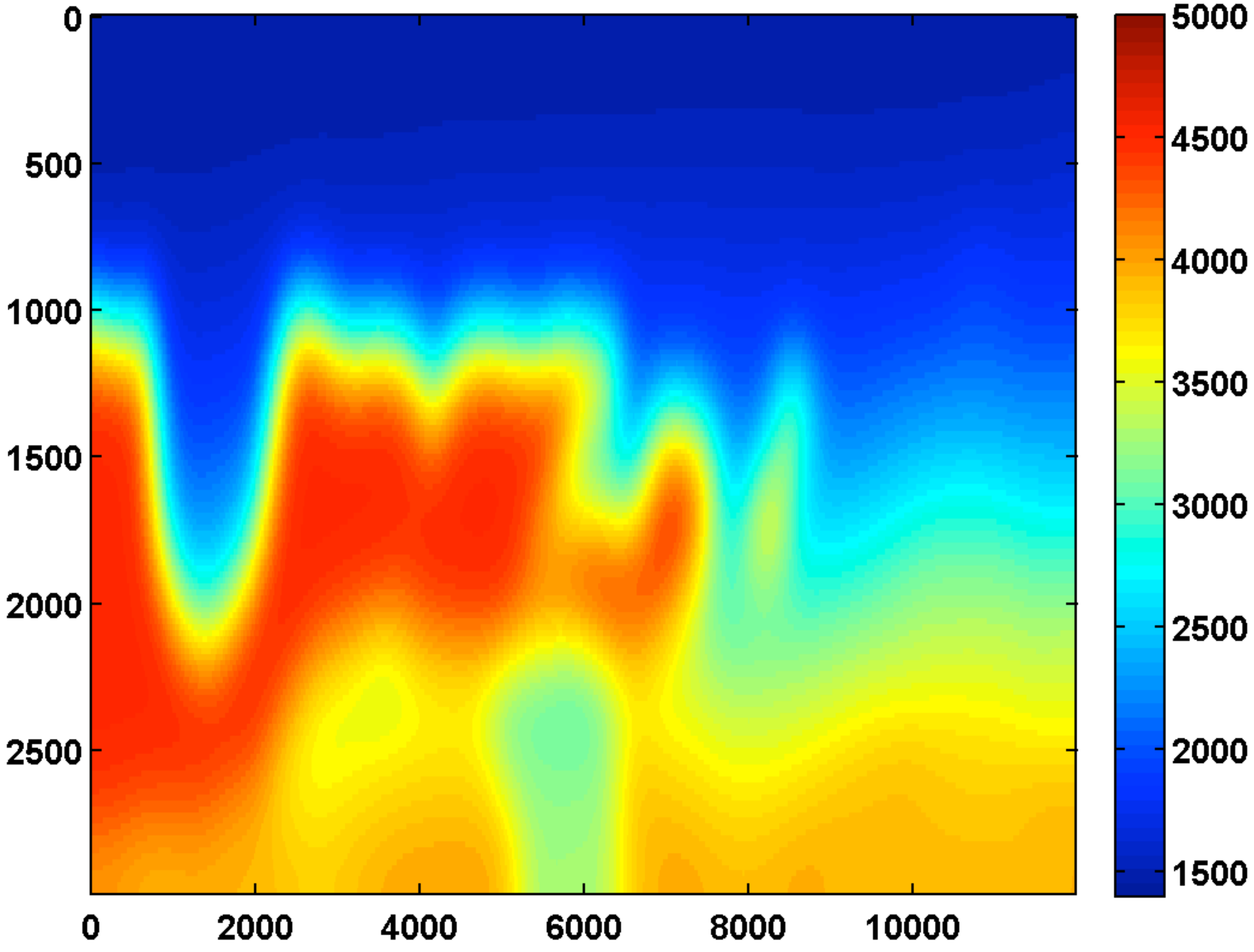
SLIM

BP model

- number of sources: 126 (starting 1000m in from boundary)
- number of receivers: 299
- frequency range: 3-20Hz in overlapping batches of 2
- maximum number of outer iterations per frequency batch: 25
- maximum number of inner iterations for convex subproblems: 2000
- known Ricker wavelet sources with 15Hz peak frequency
- two simultaneous shots with Gaussian weights w/ redraws
- no added noise

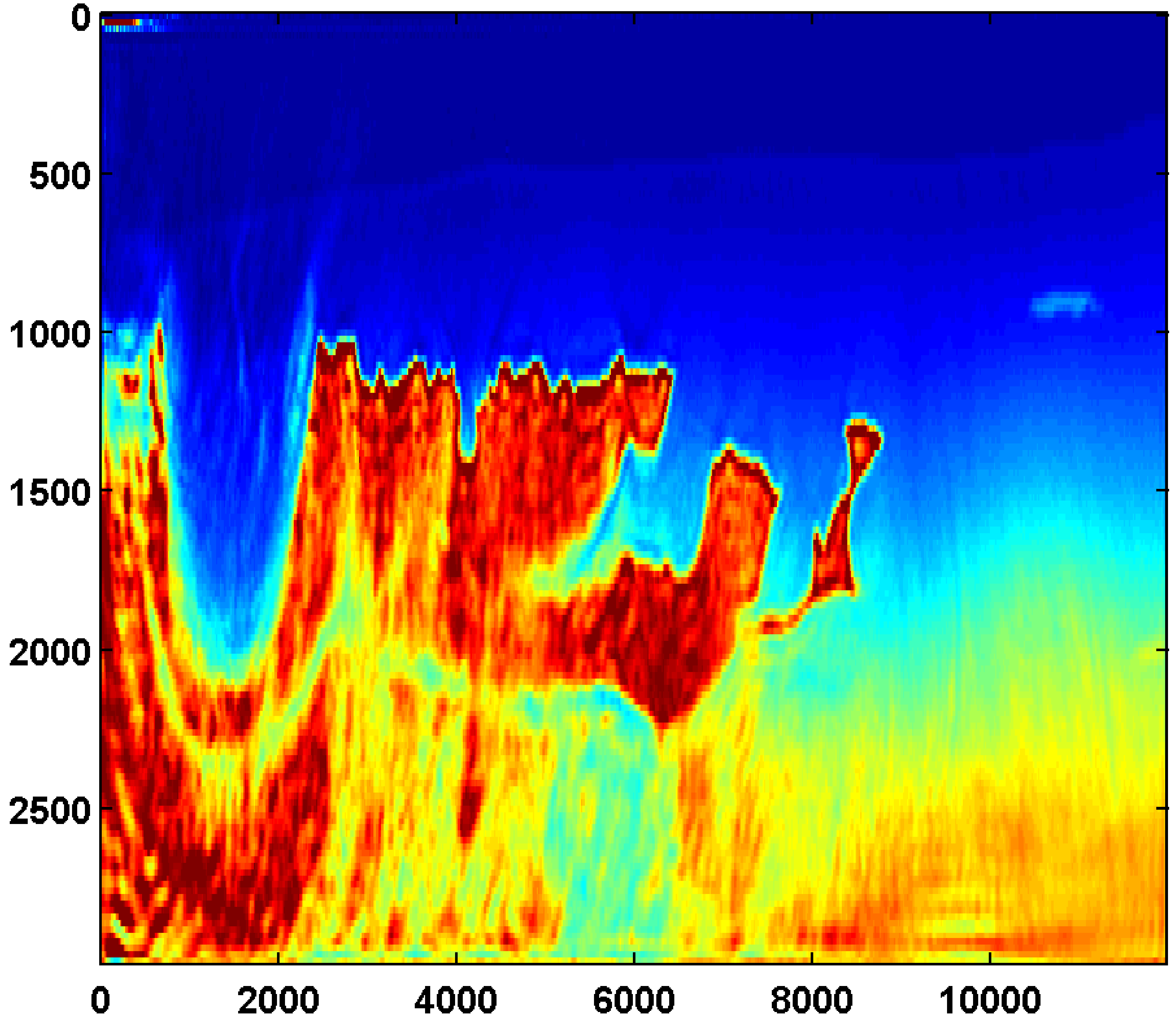
True velocity & good starting model



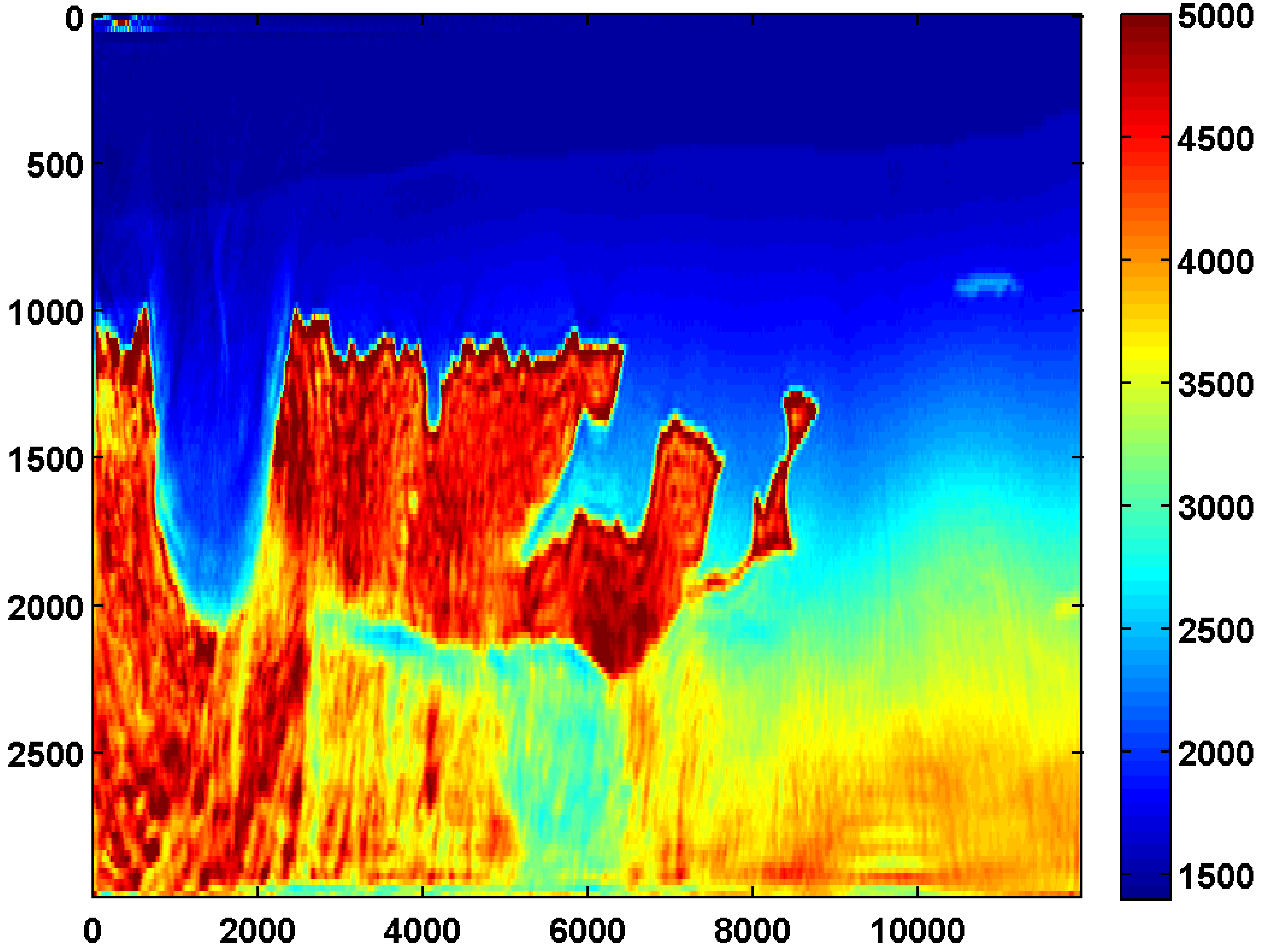


Results w/o TV

After one cycle through the frequencies

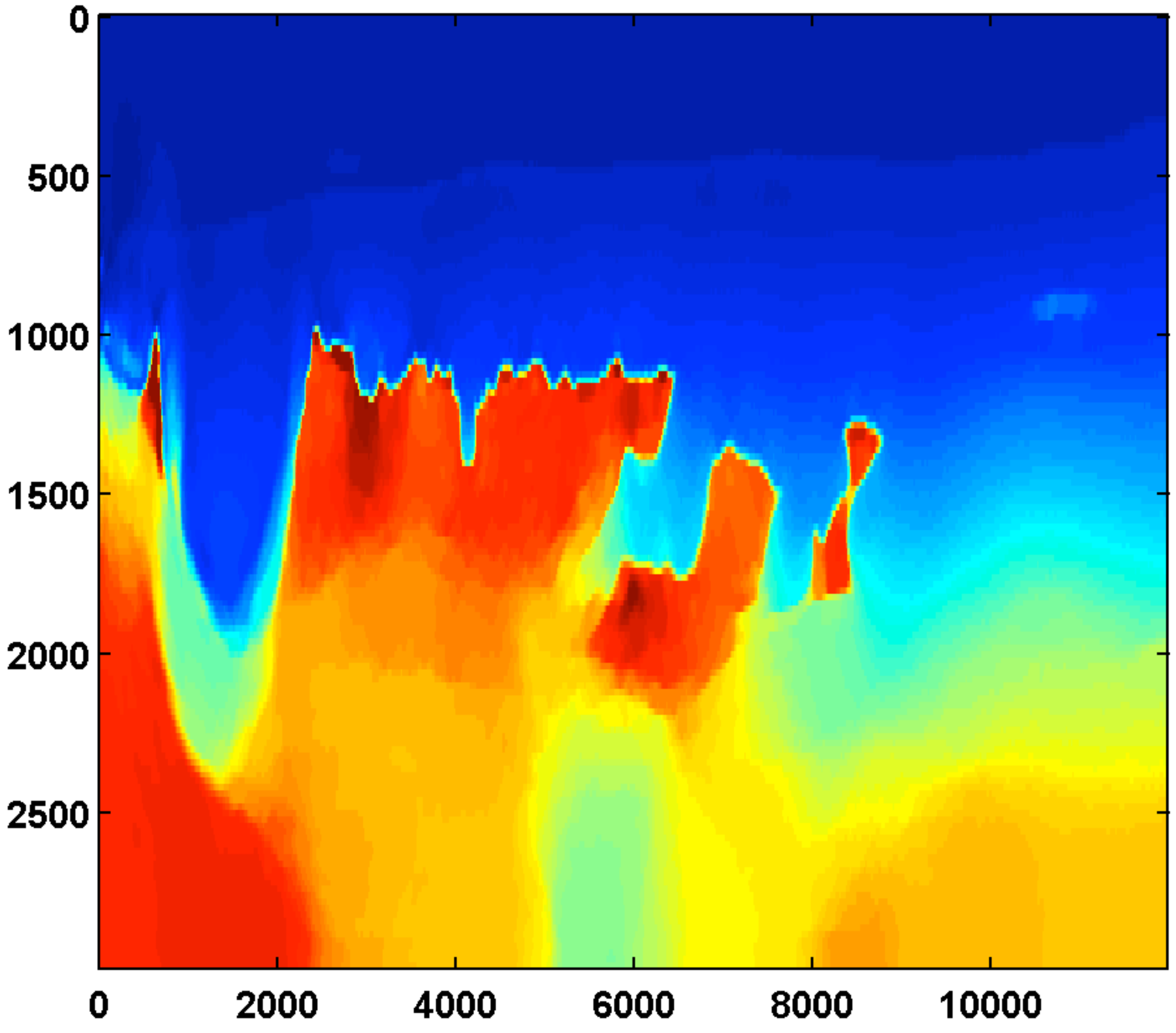


After two cycles through the frequencies

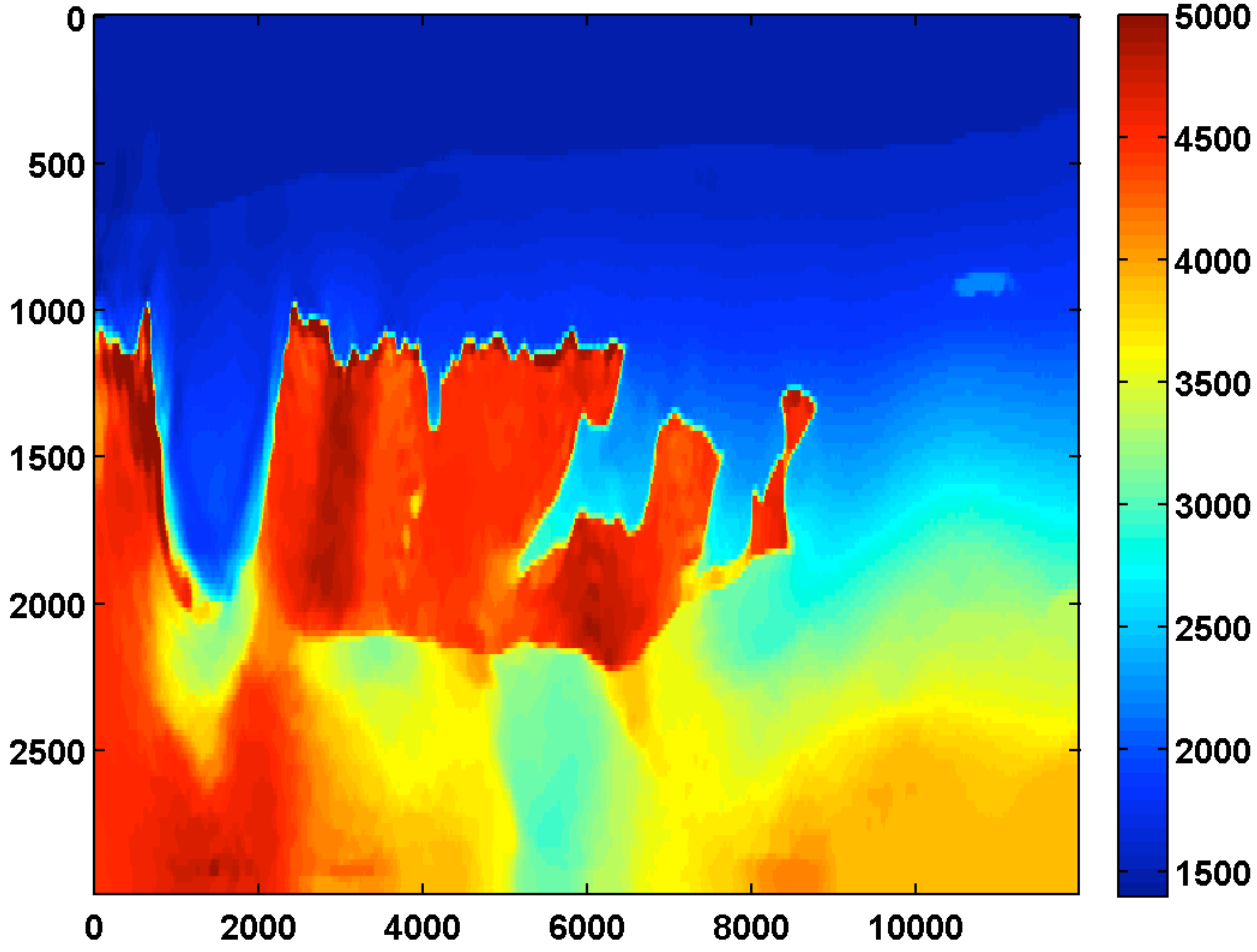


Results w/ TV

After one cycle through the frequencies

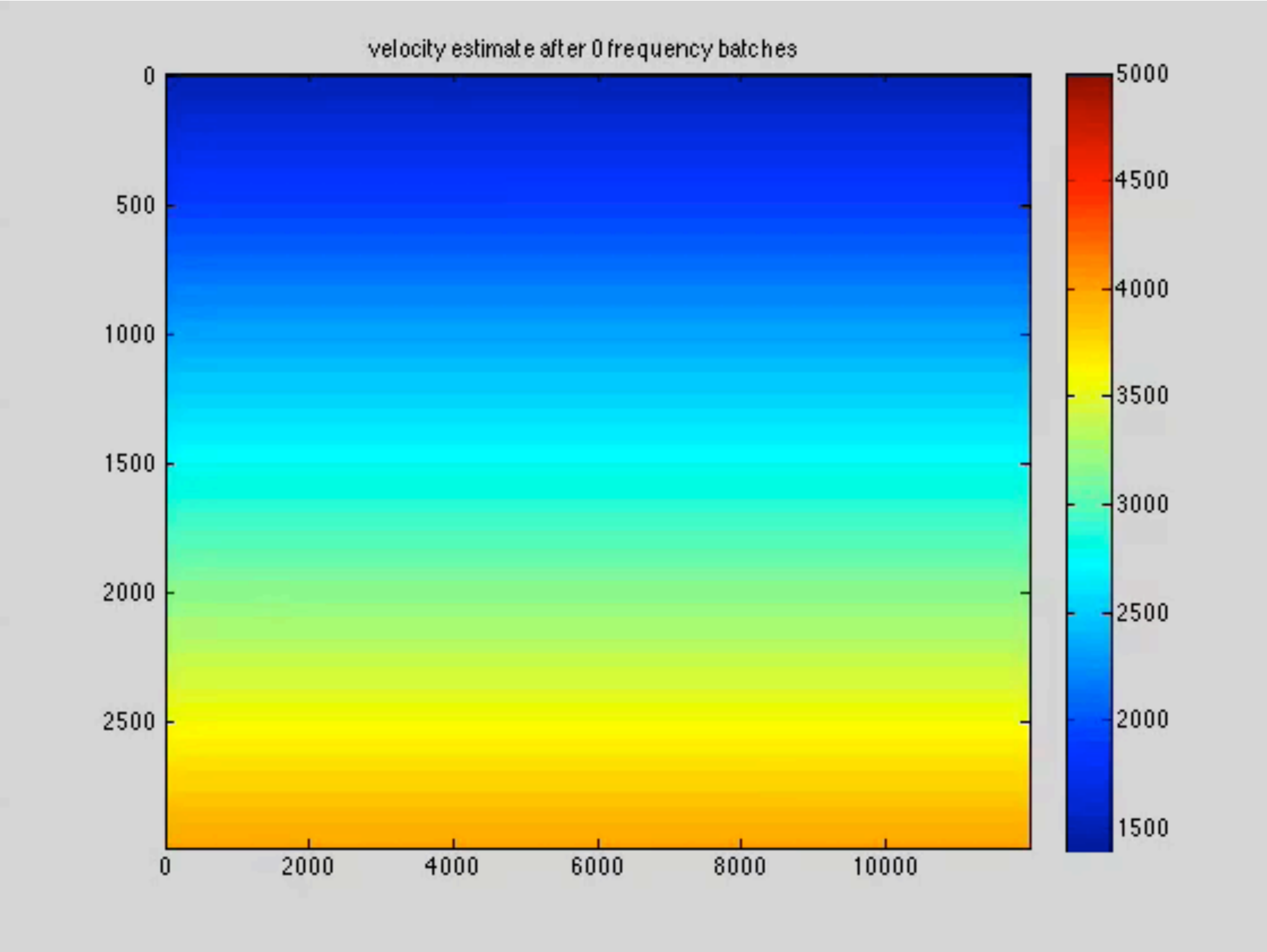


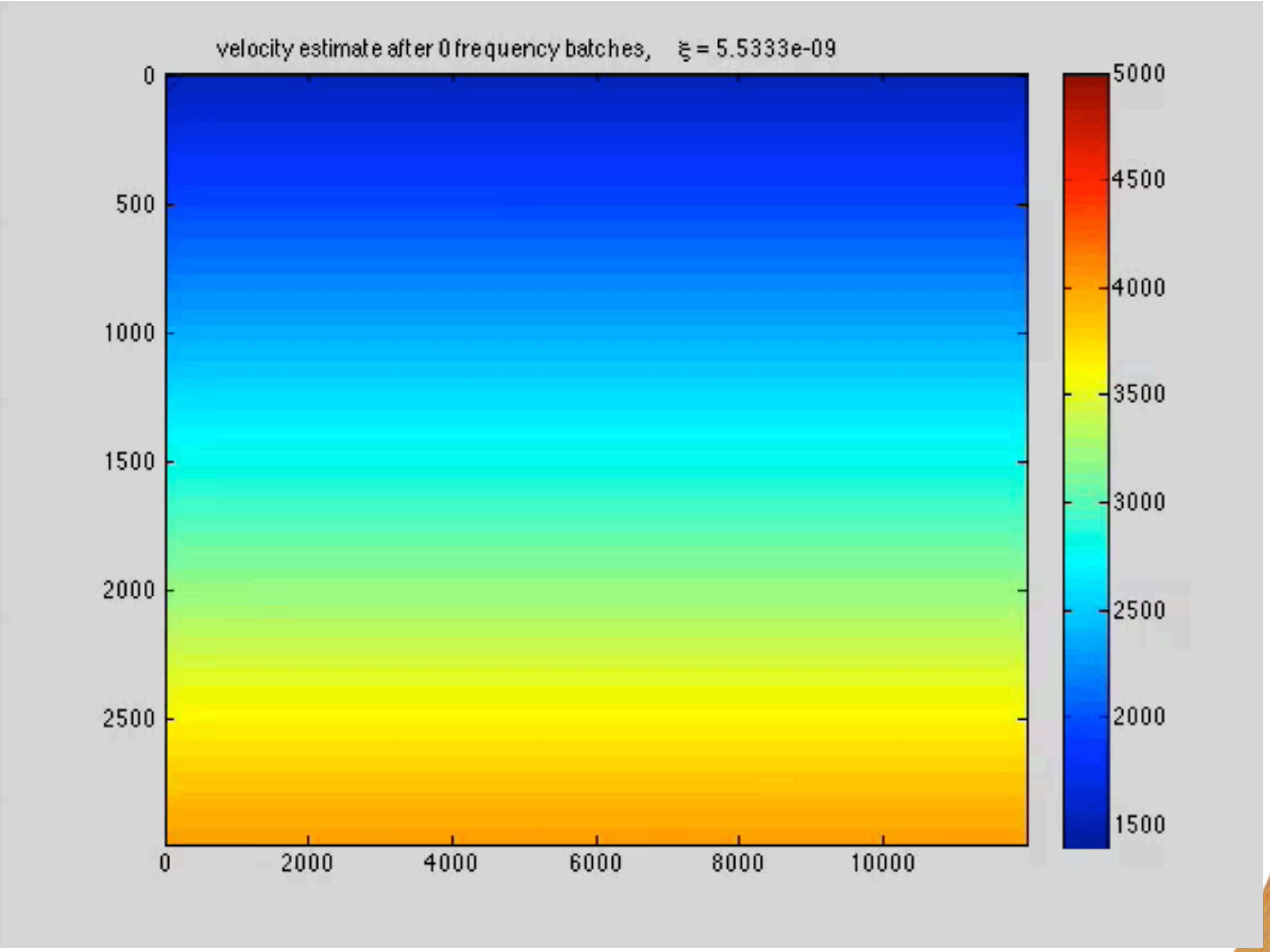
After two cycles through the frequencies



WRI

w/ or w/o TV-norm projections & bad starting model





Observations

WRI

- removes (some) reliance accurate starting models by extending the search space
- Total-variation projections create reflectors that detect top & bottom of Salt
- automatic Salt flooding

Randomization makes WRI

- computationally affordable

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

SLIM

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.