Compressive Sensing and Sparse Recovery in Exploration Seismology

Felix J. Herrmann

Felix J. Herrmann

SLIM
Seismic Laboratory for Imaging and Modeling
the University of British Columbia

Drivers

Our incessant

- demand for *hydrocarbons* while we are *no* longer finding oil...
- desire to understand the Earth's inner workings

Push for improved *seismic inversion* to

- create *more high-resolution* information
- from *noisier* and *incomplete* data

Controversial statements

Size of our *discretizations* is *dictated* by

- a *far* too *pessimistic Nyquist-sampling criterion* compounded by the *curse of dimensionality*
- our *insistence* to sample *periodically* and/or *sequentially*

Our desire to work with *all* data

- leads to "over emphasis" on data *collection* & *full-data processing*
- prohibits *inversion* that requires *multiple* passes through *data*

Wish list

Acquisition & *inversion* costs determined by *structure of data* & *complexity* of the *subsurface*

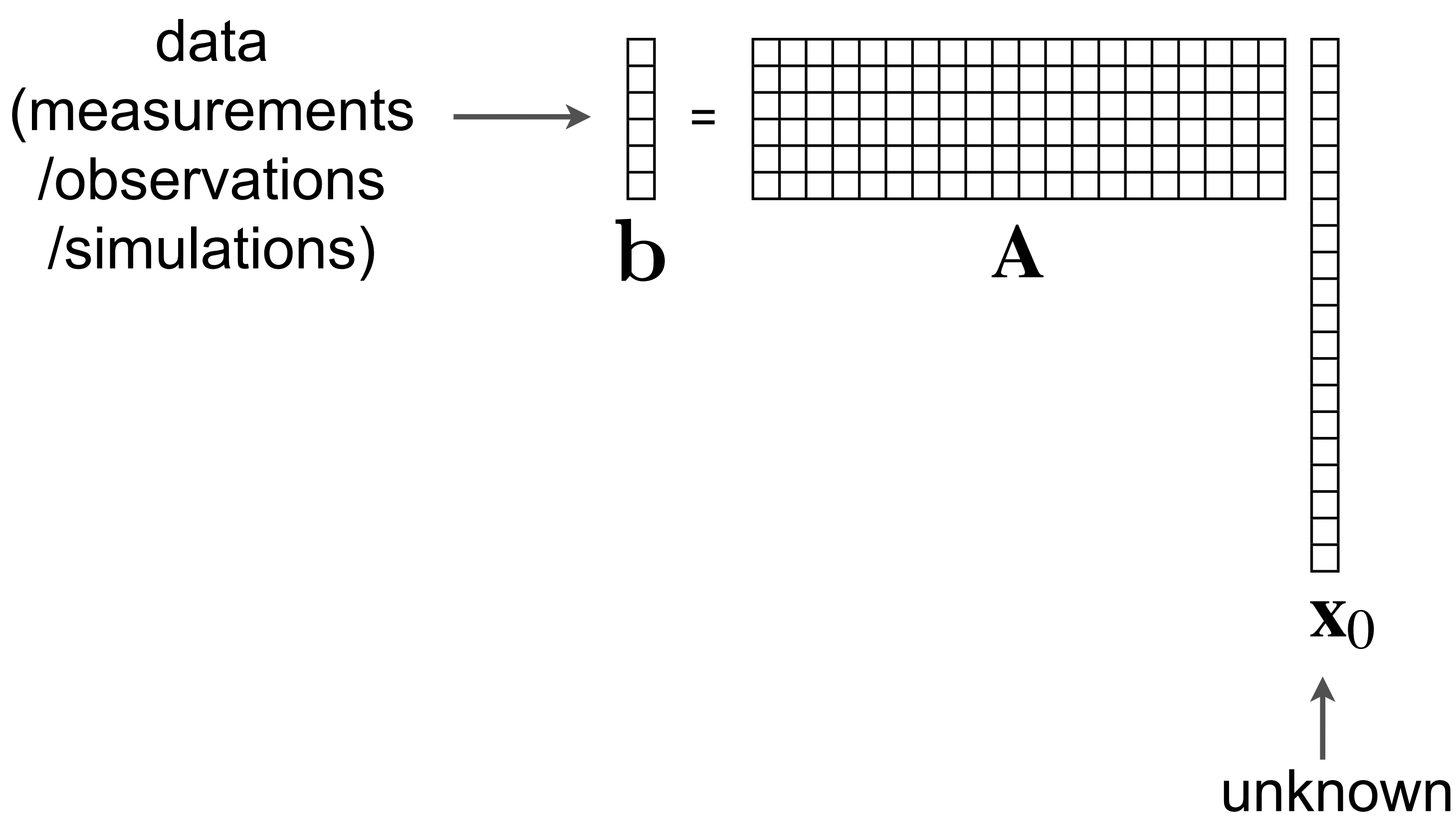
- sampling *criteria* that are *dictated by* transform-domain *sparsity* and *not* by the *size* of the *discretization*

Controllable error that depends on

- degree of *subsampling* / *dimensionality* reduction
- available computational resources

Problem statement

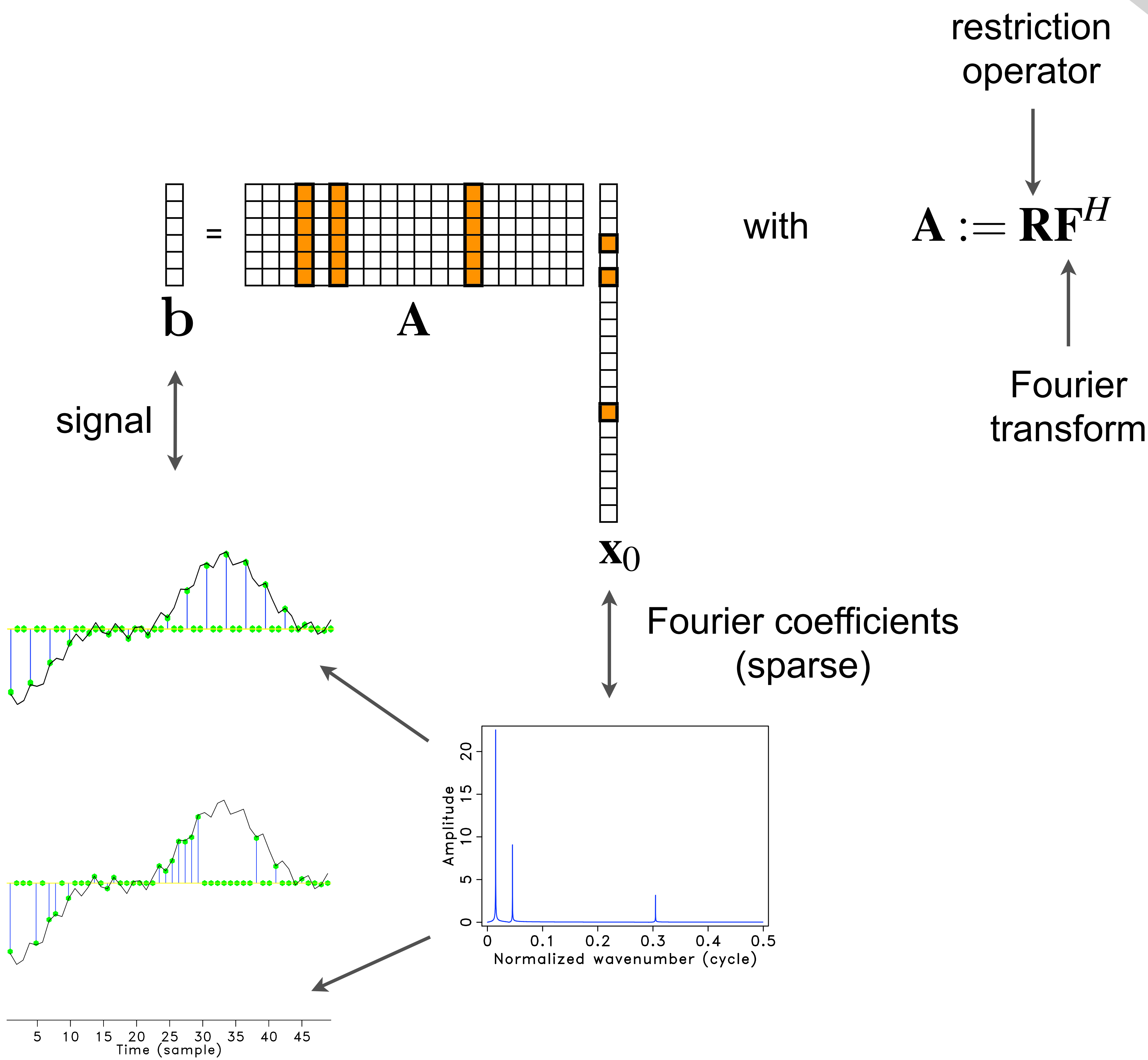
Consider the following (severely) *underdetermined* system of *linear* equations:



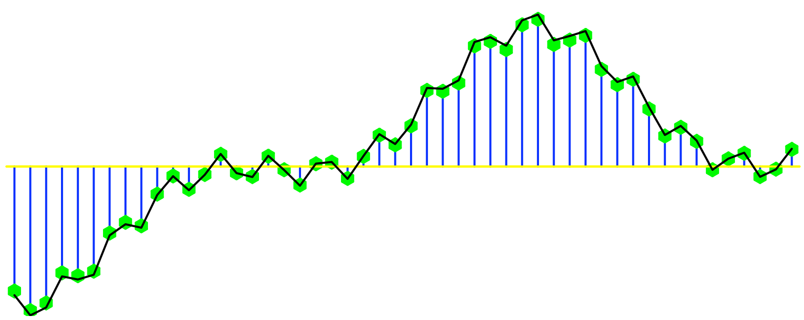
Is it possible to recover $\mathbf{x}_0$ accurately from $\mathbf{b}$?

The new field of *Compressive Sensing* attempts to answer this.

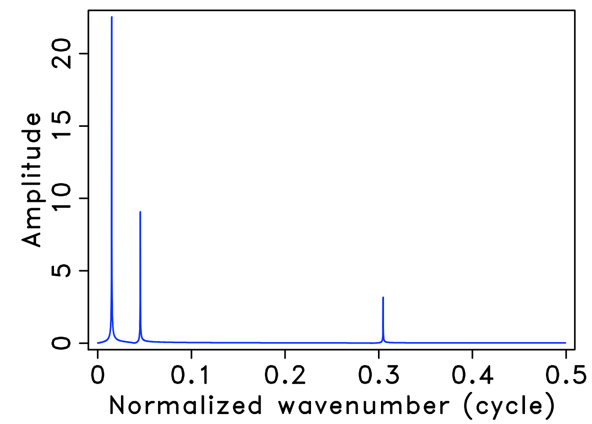
Sparse recovery



Coarse sampling schemes

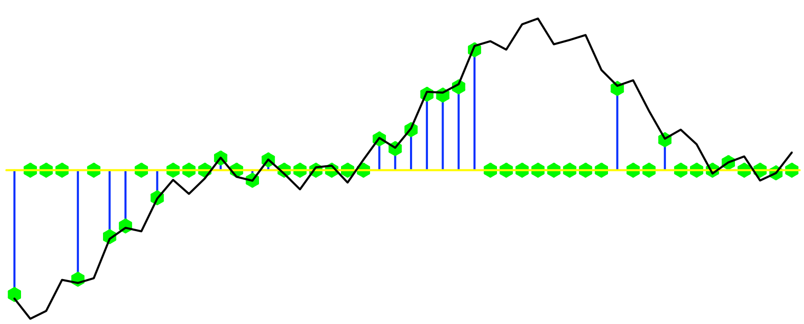


Fourier → transform

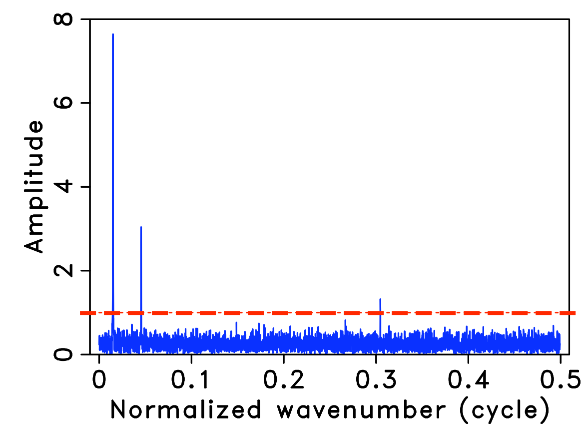


few significant coefficients

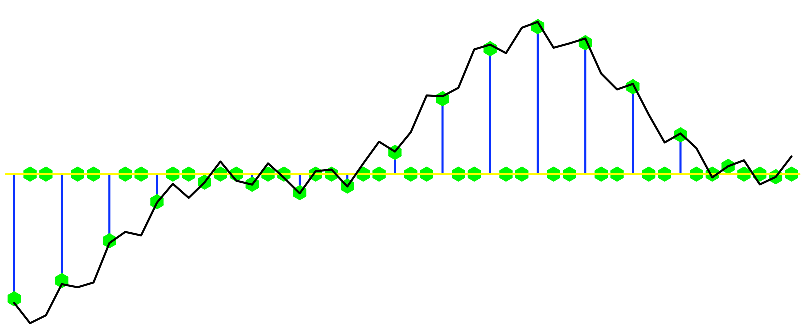
3-fold under-sampling



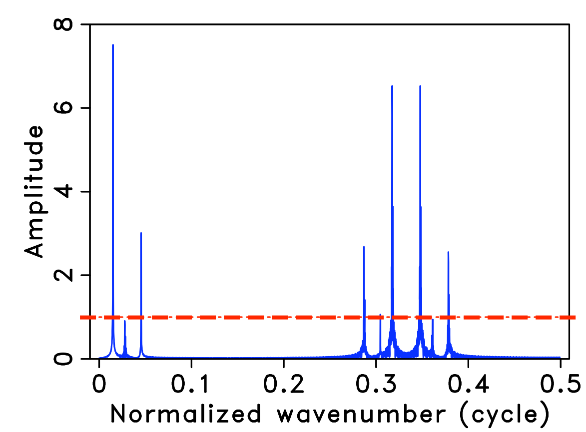
Fourier → transform



✓ significant coefficients detected



Fourier → transform



✗ ambiguity

[Hennenfent & Herrmann, ’08]

Sparse one-norm recovery

Signal model

$$\mathbf{b} = \mathbf{A}\mathbf{x}_0 \quad \text{where} \quad \mathbf{b} \in \mathbb{R}^n$$

and $\mathbf{x}_0$ *k* sparse

Sparse one-norm recovery

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \stackrel{\text{def}}{=} \sum_{i=1}^{N} |x[i]| \quad \text{subject to} \quad \mathbf{b} = \mathbf{A}\mathbf{x}$$

with $\mathrm{n} \ll N$ where N is the *ambient dimension*

Study recovery as a function of

- the subsampling ratio n/N
- “over sampling” ratio k/n

[Sacchi ’98]
[Candès et.al, Donoho, ’06]

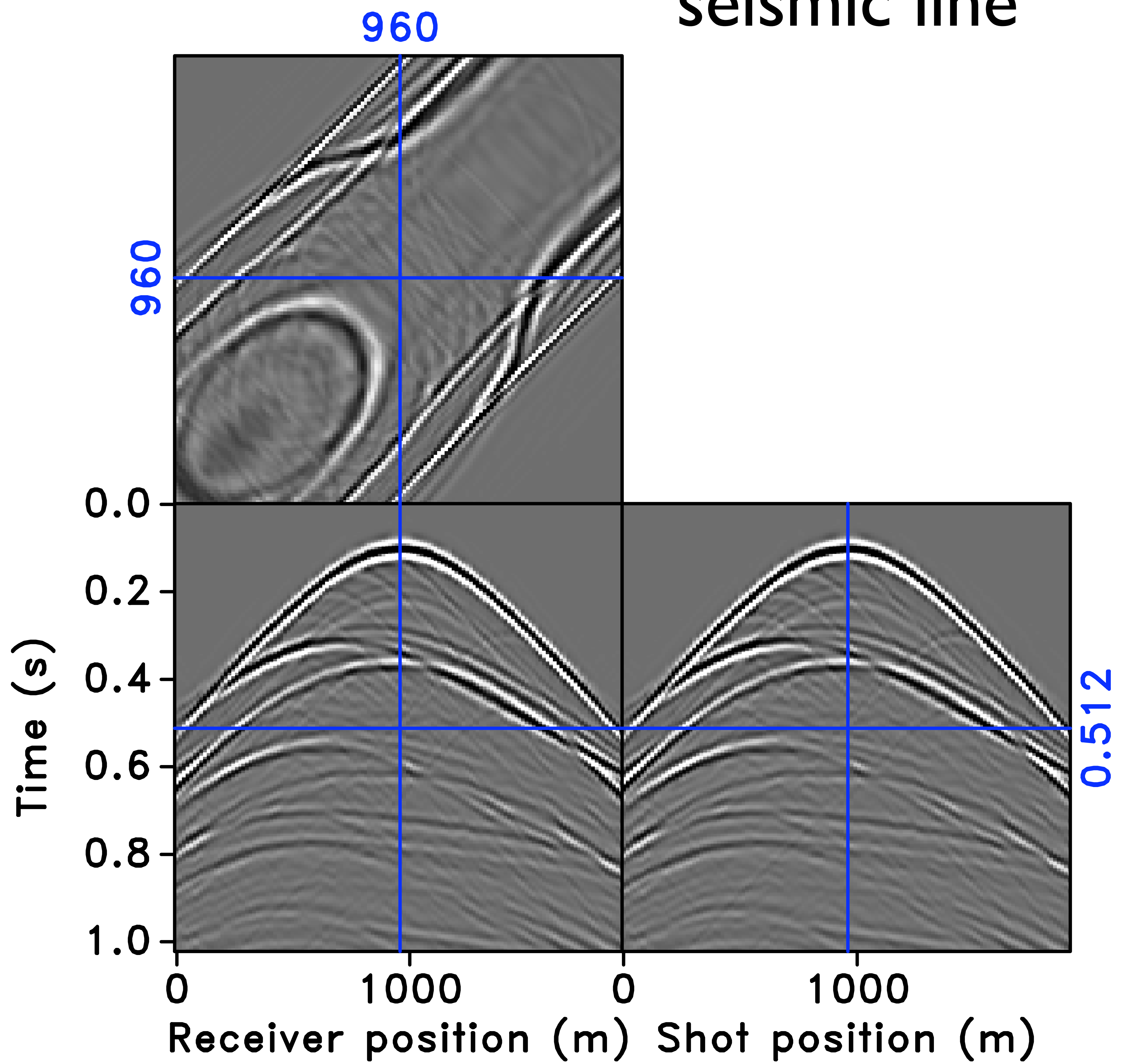
Case study I

Acquisition design according to Compressive Sensing

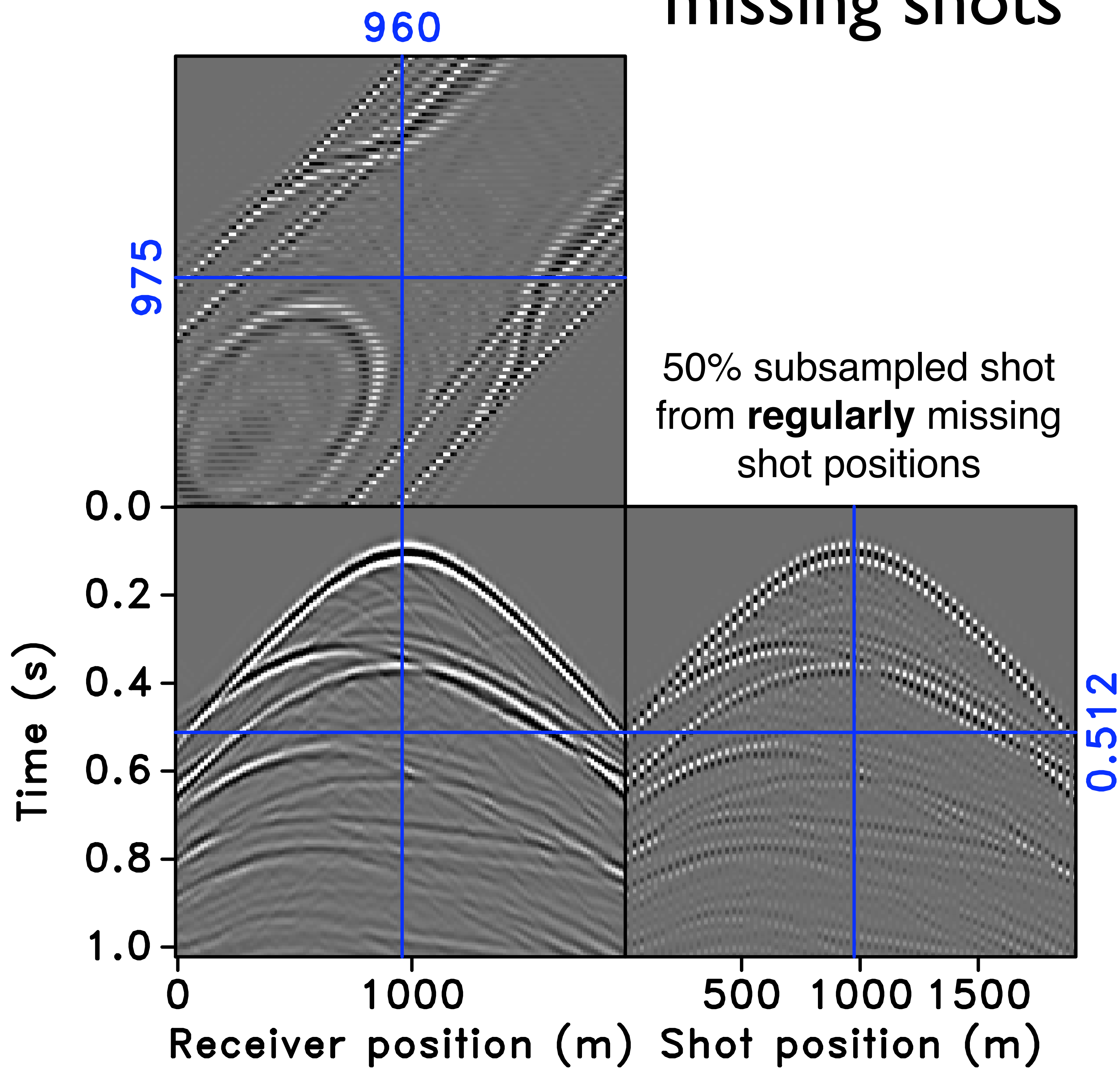
- ***Periodic*** *subsampling* vs ***randomized jittered*** sampling of ***sequential*** sources
- *Subsampling with randomized jittered* ***sequential*** sources vs *randomized phase-encoded* ***simultaneous*** sources

[Hennenfent & Herrmann, '08]

seismic line

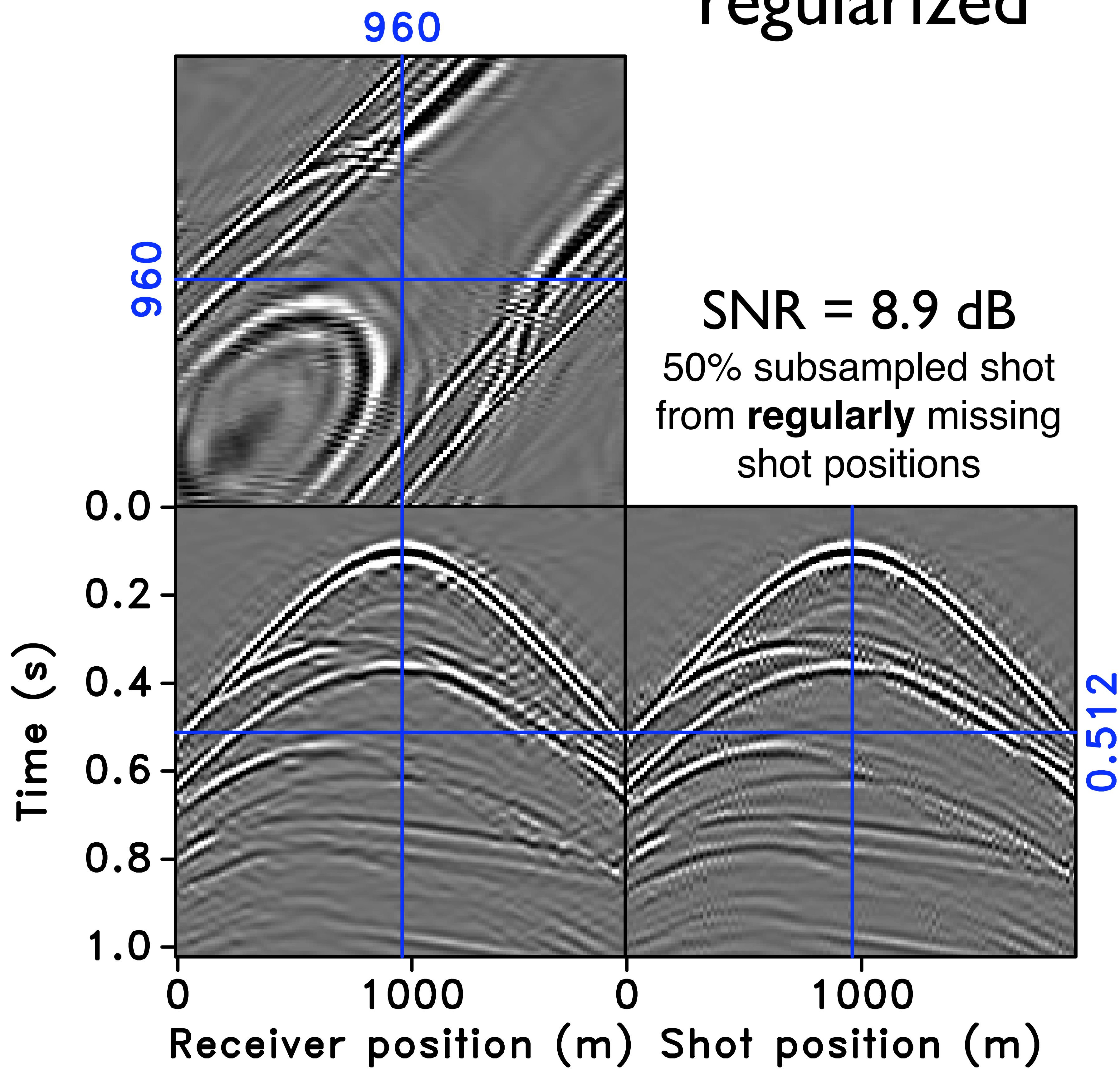


missing shots



50% subsampled shot from **regularly** missing shot positions

regularized



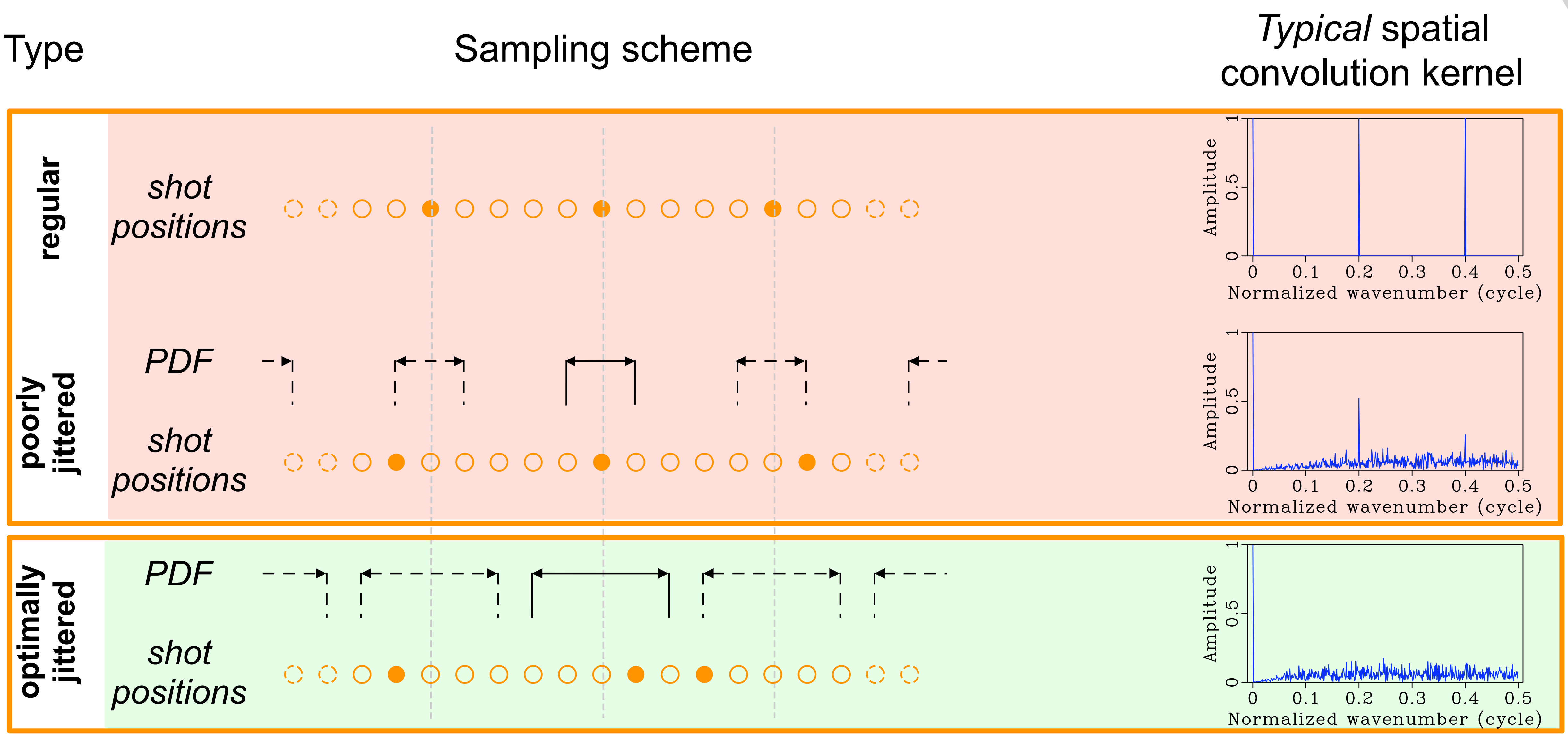
SNR = 8.9 dB

50% subsampled shot from **regularly** missing shot positions

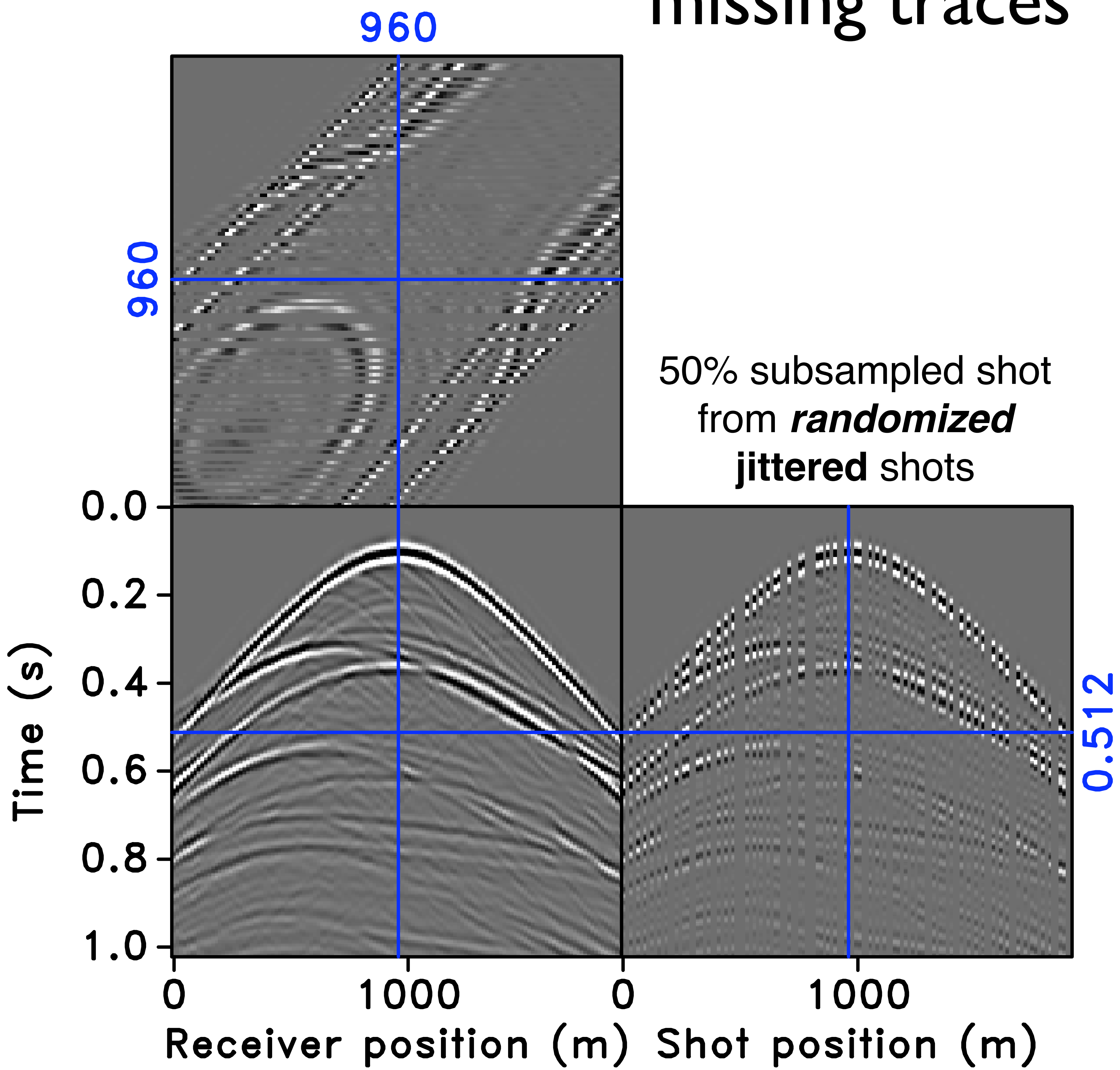
[Hennenfent & FJH, '08]
[Gang et.al., '09]

SLIM

Jittered sampling

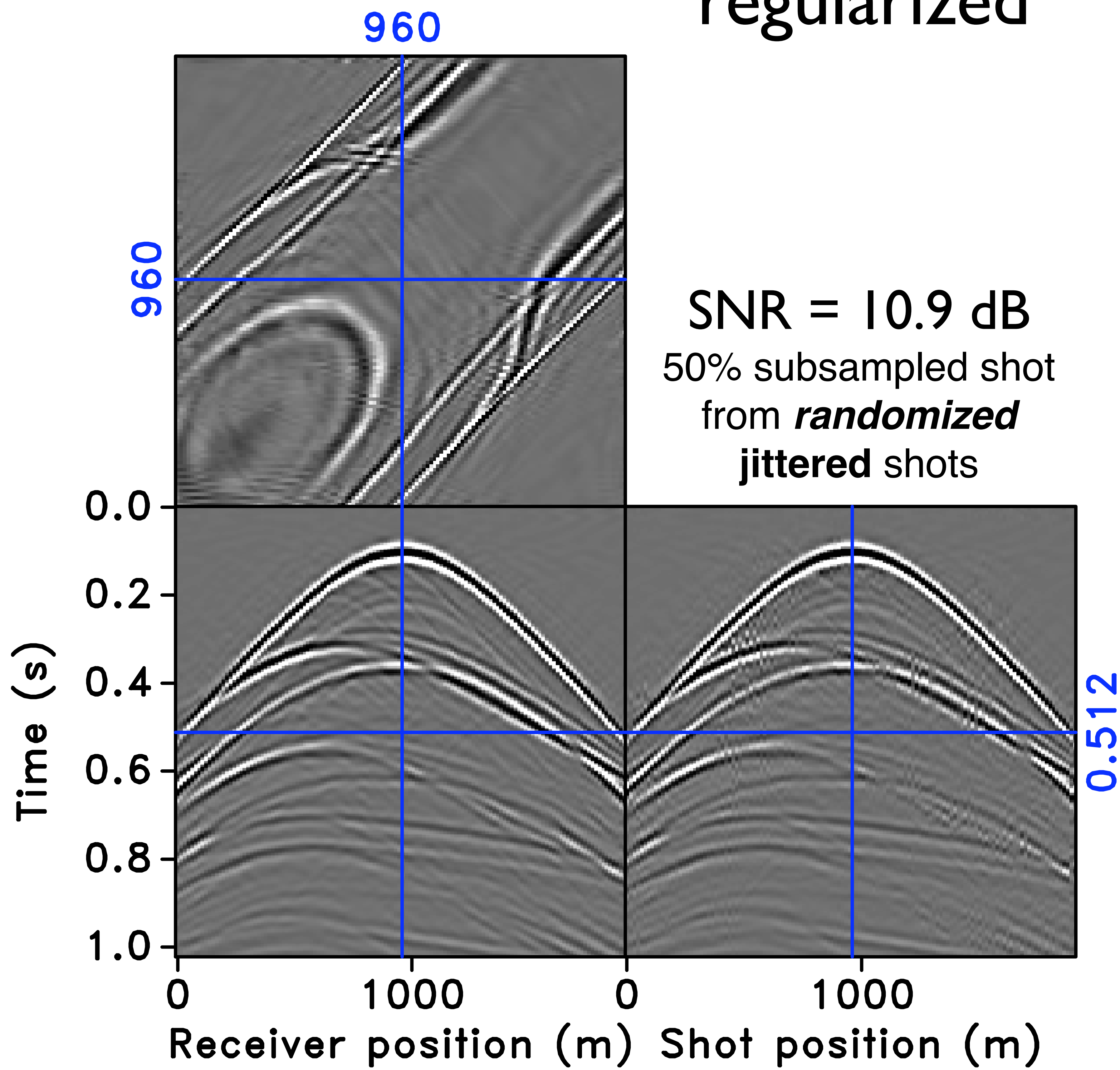


missing traces



50% subsampled shot from ***randomized*** **jittered** shots

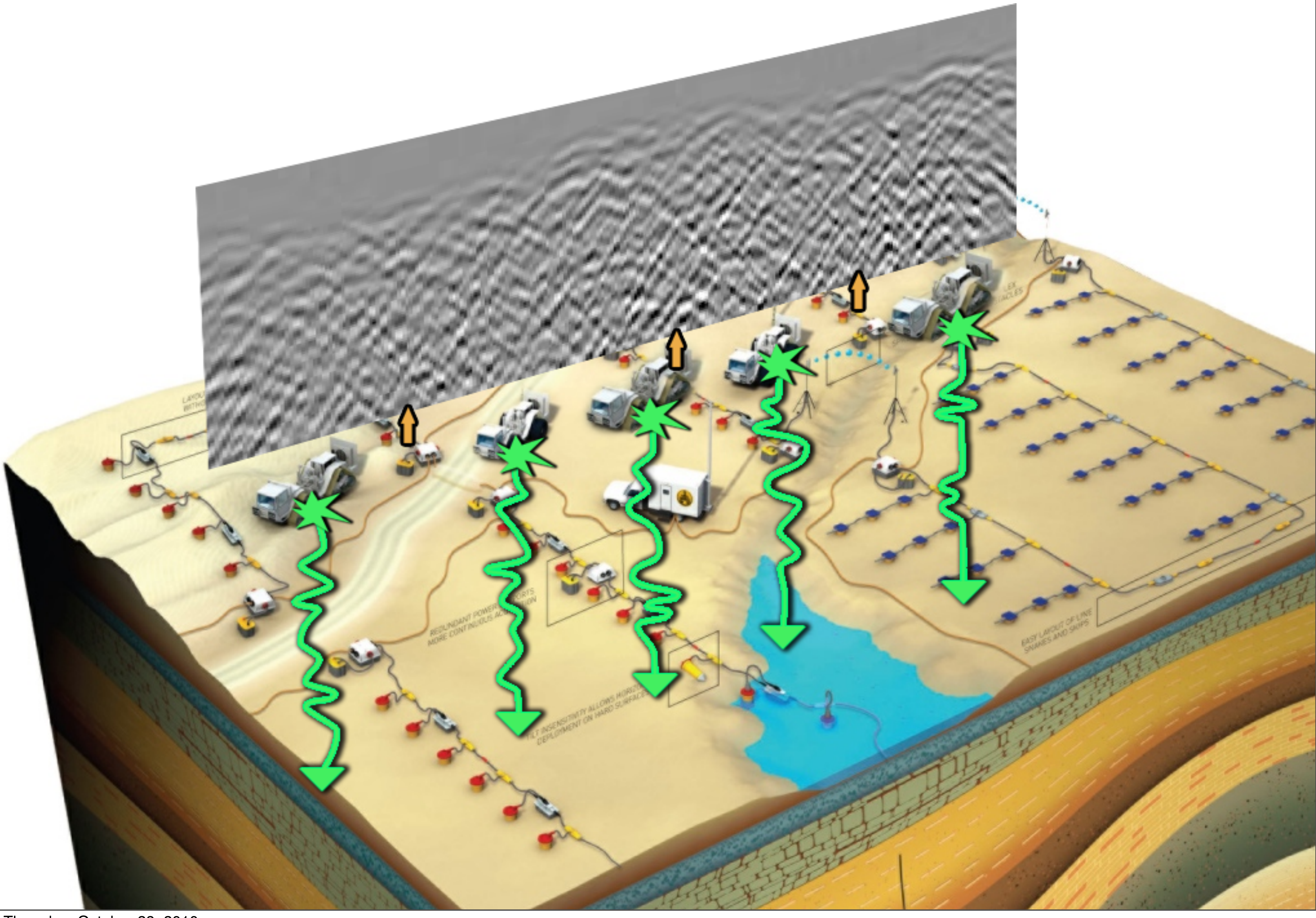
regularized



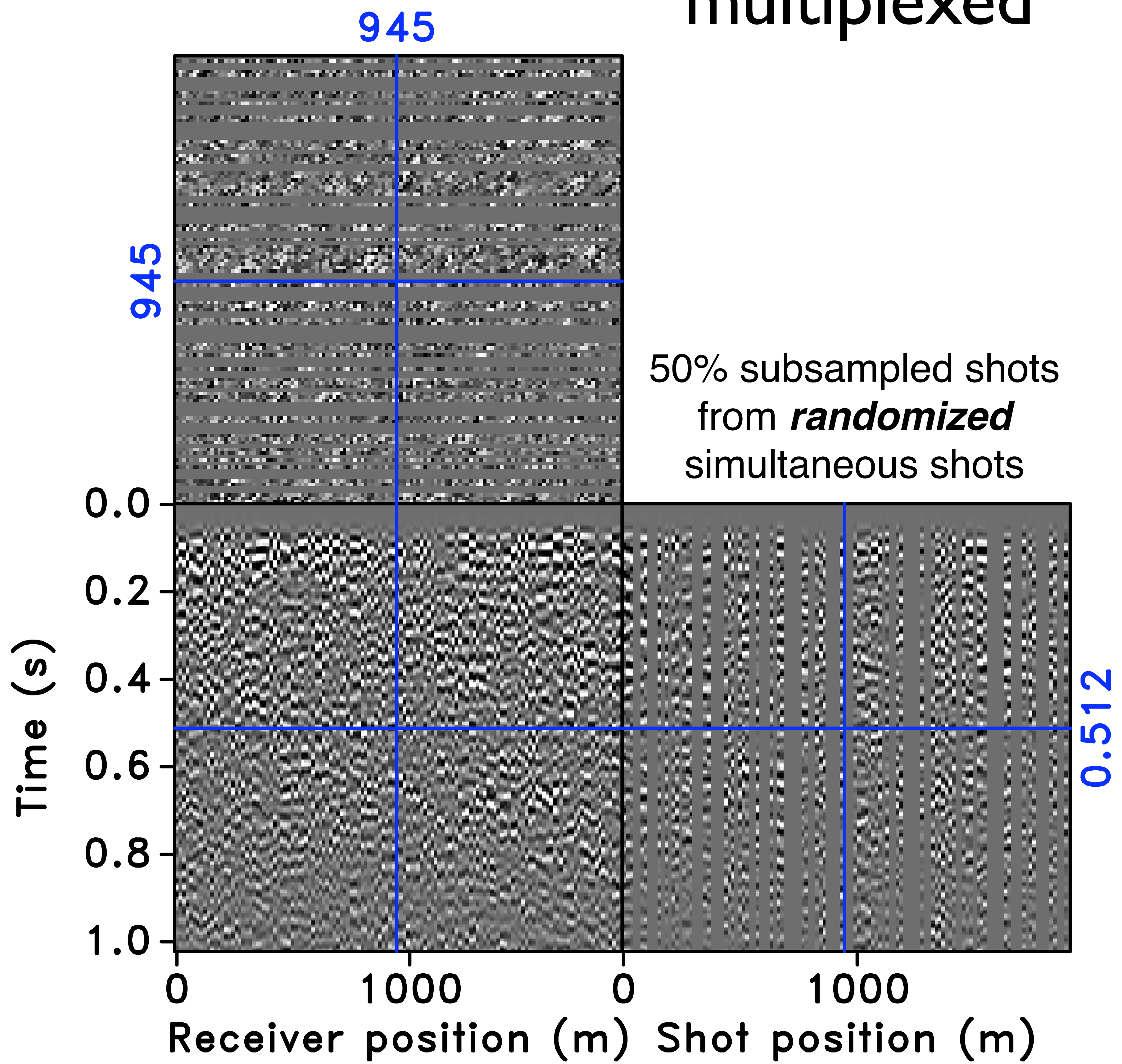
SNR = 10.9 dB

50% subsampled shot from ***randomized*** **jittered** shots

Simultaneous & incoherent sources

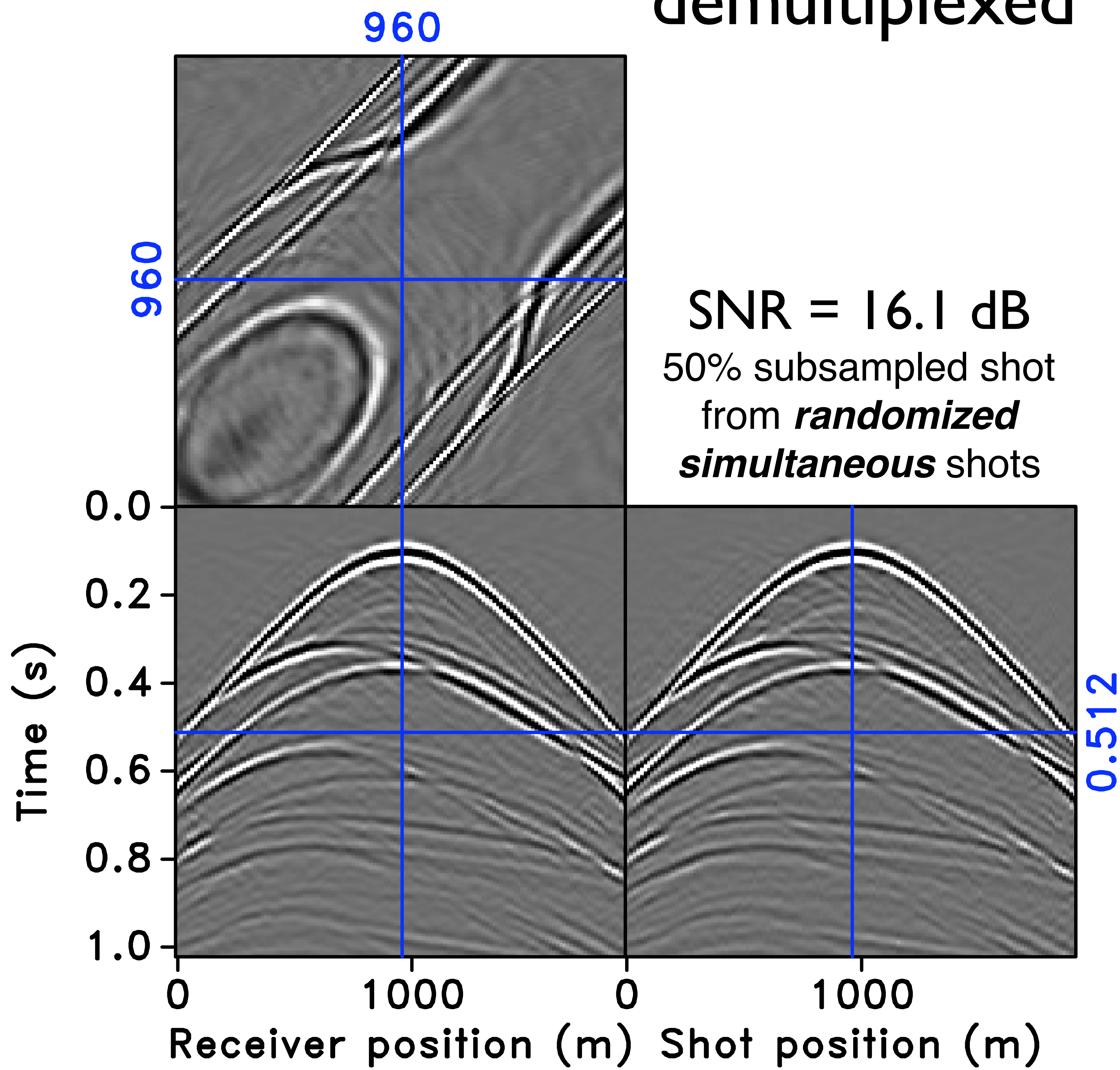


multiplexed



50% subsampled shots from ***randomized*** simultaneous shots

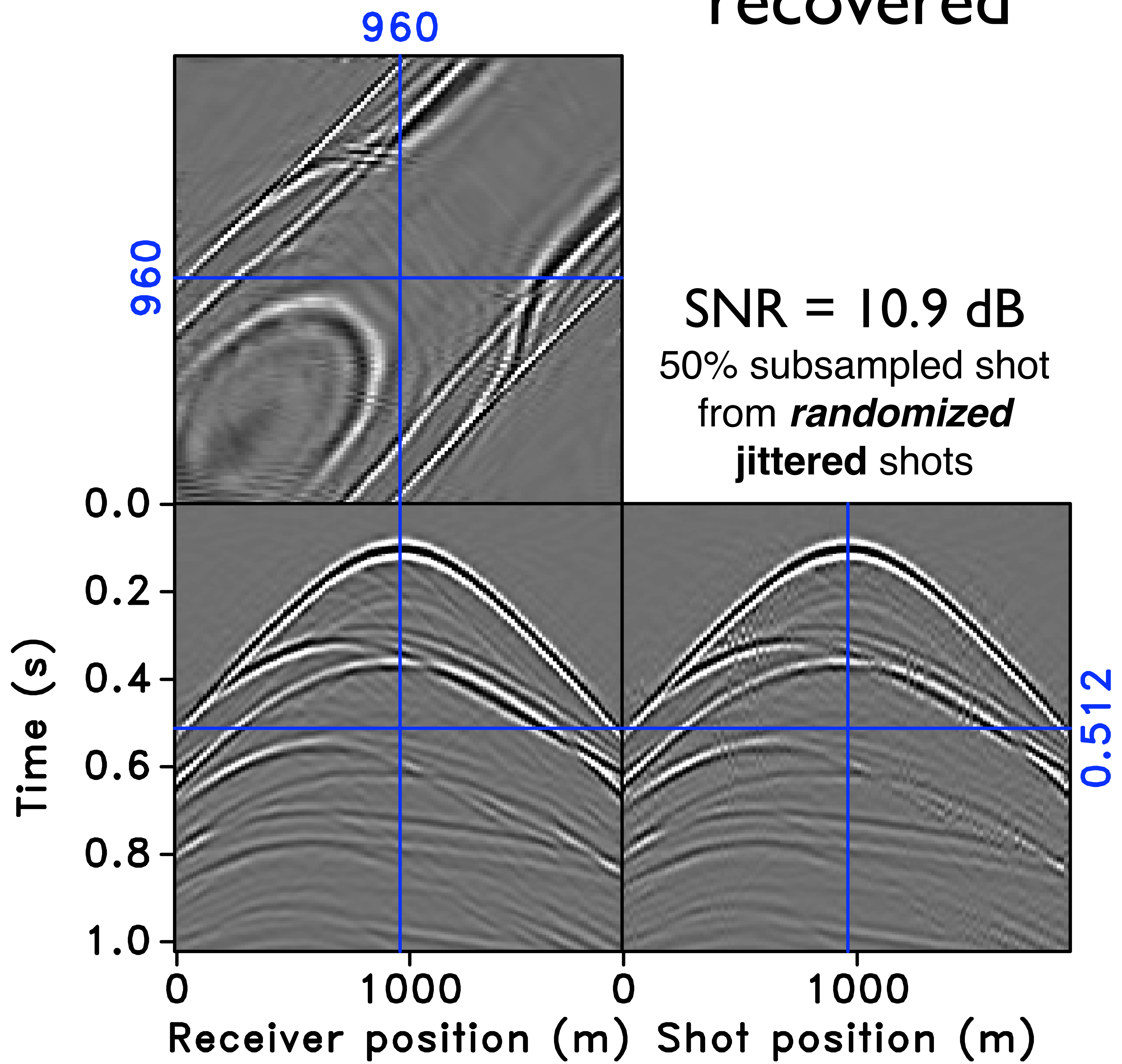
demultiplexed



SNR = 16.1 dB

50% subsampled shot from ***randomized simultaneous*** shots

recovered



SNR = 10.9 dB

50% subsampled shot from ***randomized*** **jittered** shots

The math of Compressive Sensing [Candès et.al, '06]

Recovery is *possible* & *stable* as long as each *subset S* of *k* columns of $\mathbf{A} \in \mathbb{R}^{n \times N}$ with $k \leq N$ the # of nonzeros *approximately* behaves as an *orthogonal* basis.

In that case, we have

$$(1 - \hat{\delta}_k)\|\mathbf{x}_S\|_{\ell_2}^2 \leq \|\mathbf{A}_S\mathbf{x}_S\|_{\ell_2}^2 \leq (1 + \hat{\delta}_k)\|\mathbf{x}_S\|_{\ell_2}^2,$$

where *S* runs over all sets with cardinality $\leq k$

- the smaller the *restricted isometry constant* (RIP) $\hat{\delta}_k$ the more *energy* is captured and the more *stable* the *inversion* of **A**
- determined by the *mutual coherence* of the cols in **A**

Math cont'ed

[Candès et.al, '06]

RIP constant is bounded by

$$\hat{\delta}_k \leq (k-1)\mu$$

where

$$\mu = \max_{1 \leq i \neq j \leq N} |\mathbf{a}_i^H \mathbf{a}_j|$$

Matrices with small $\hat{\delta}_k$ contain subsets of *k* *incoherent* columns.

Gaussian random matrices with *i.i.d.* entries have this property.

One-norm solvers recover $\mathbf{x}_0$ as long it is *k* sparse and

$$k \leq C \cdot \frac{n}{\log_2(N/n)},$$

yields an *oversampling ratio* of

$$n/k \approx C \cdot \log_2 N$$

Key elements

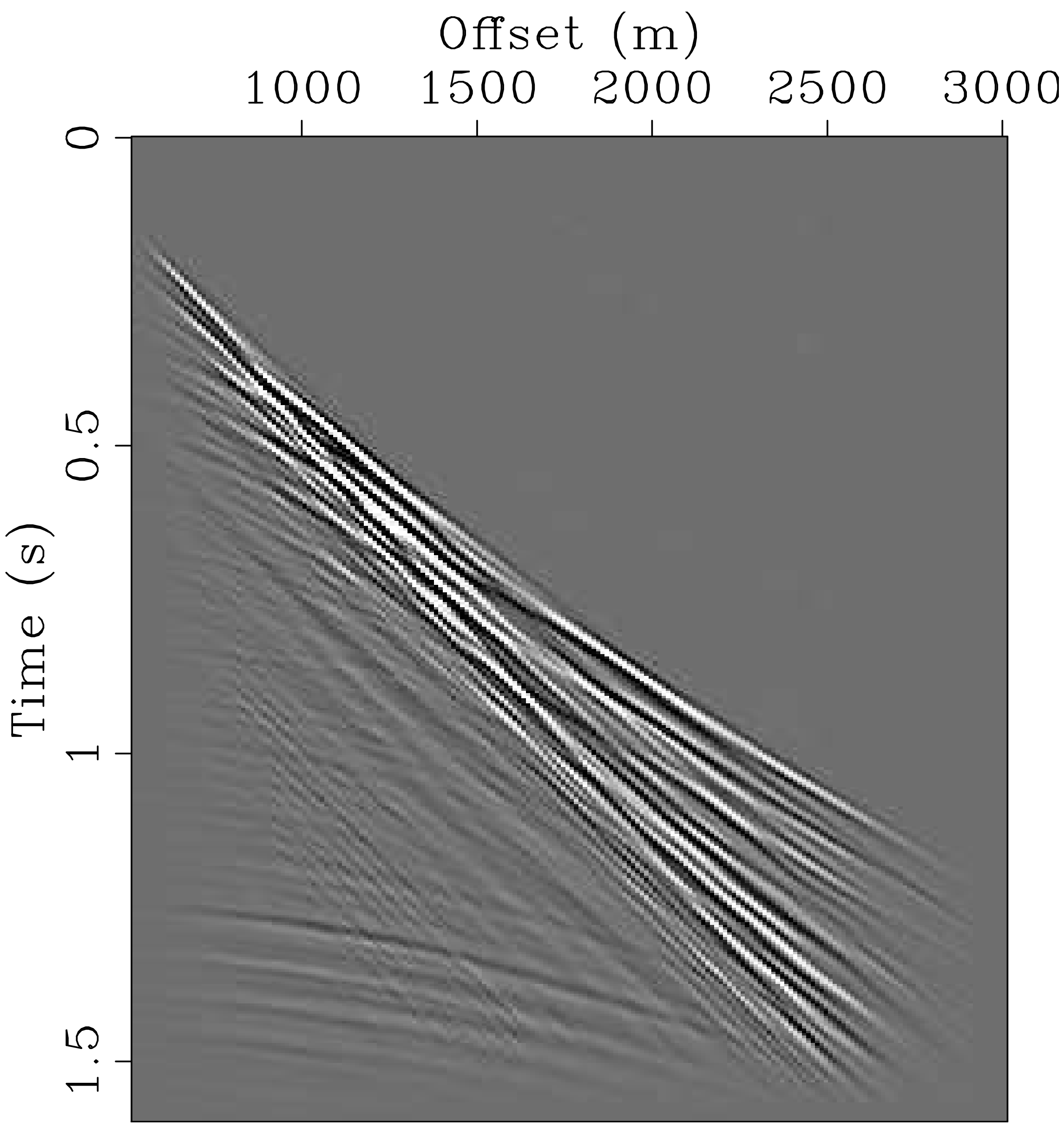
SLIM

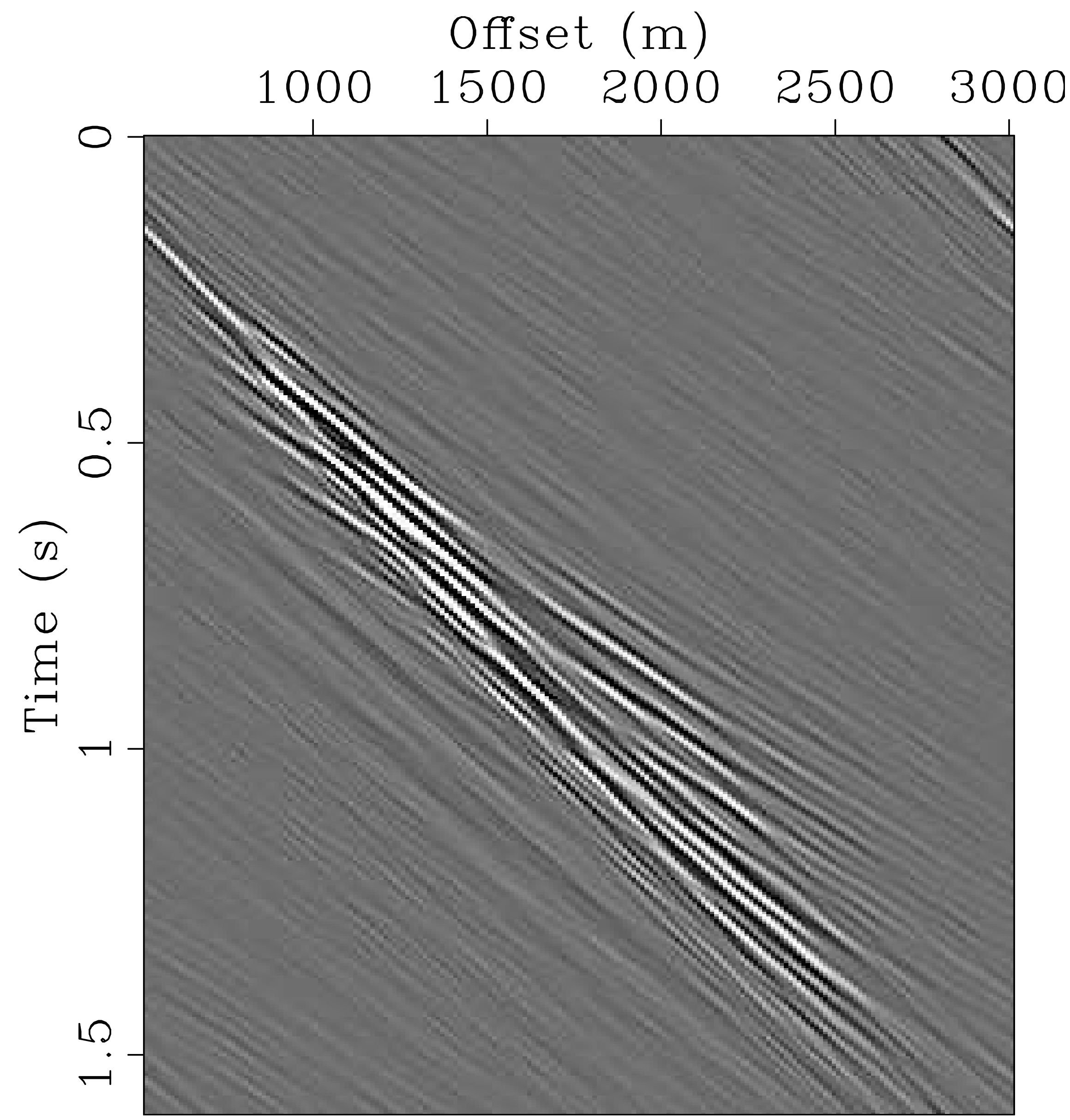
- *sparsifying transform*
  - typically **localized** in the time-space domain to handle the complexity of seismic data

- *advantageous coarse randomized sampling*
  - generates incoherent random undersampling "noise" in the sparsifying domain

- *sparsity-promoting solver*
  - requires few matrix-vector multiplications

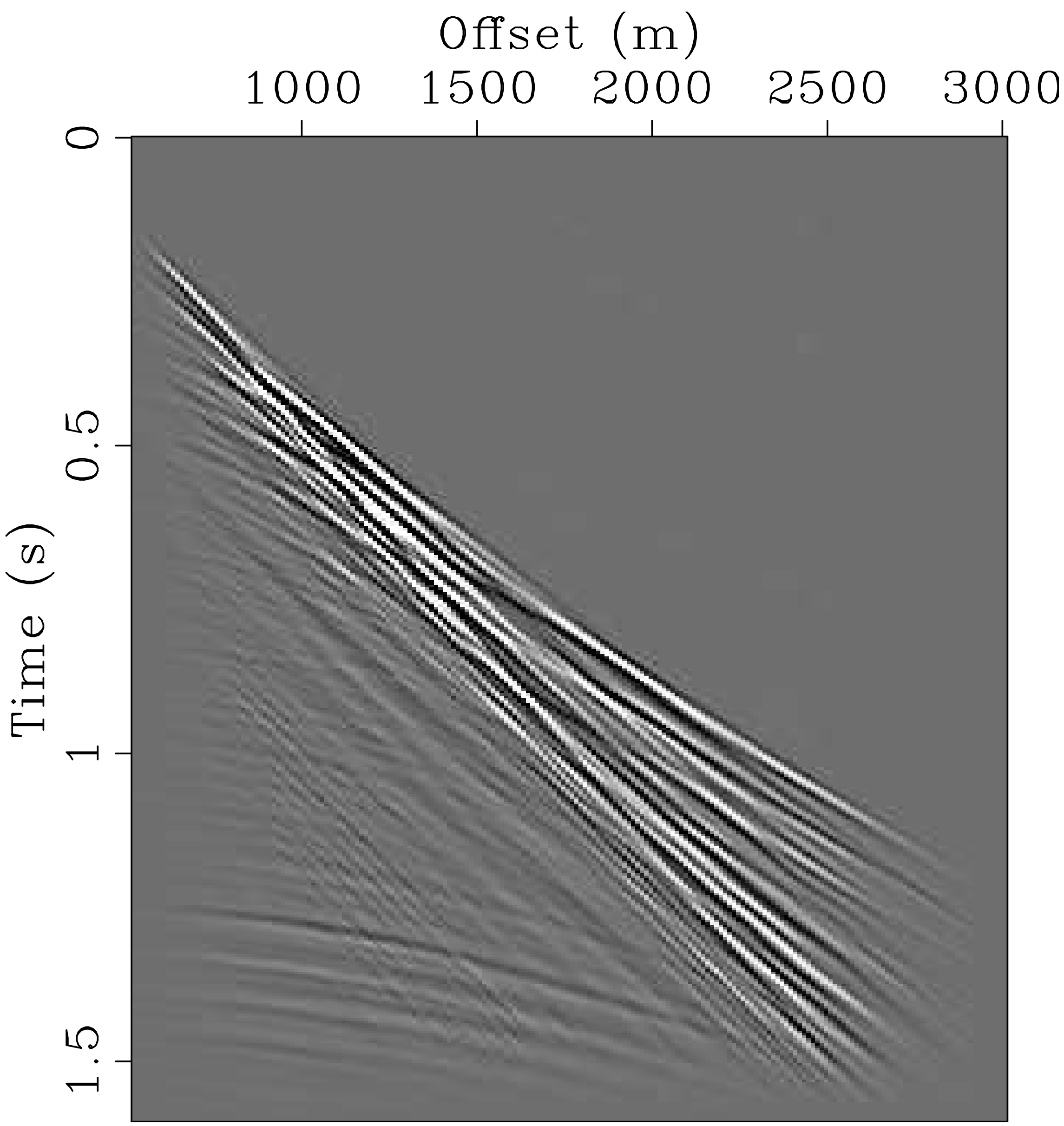
Fourier reconstruction

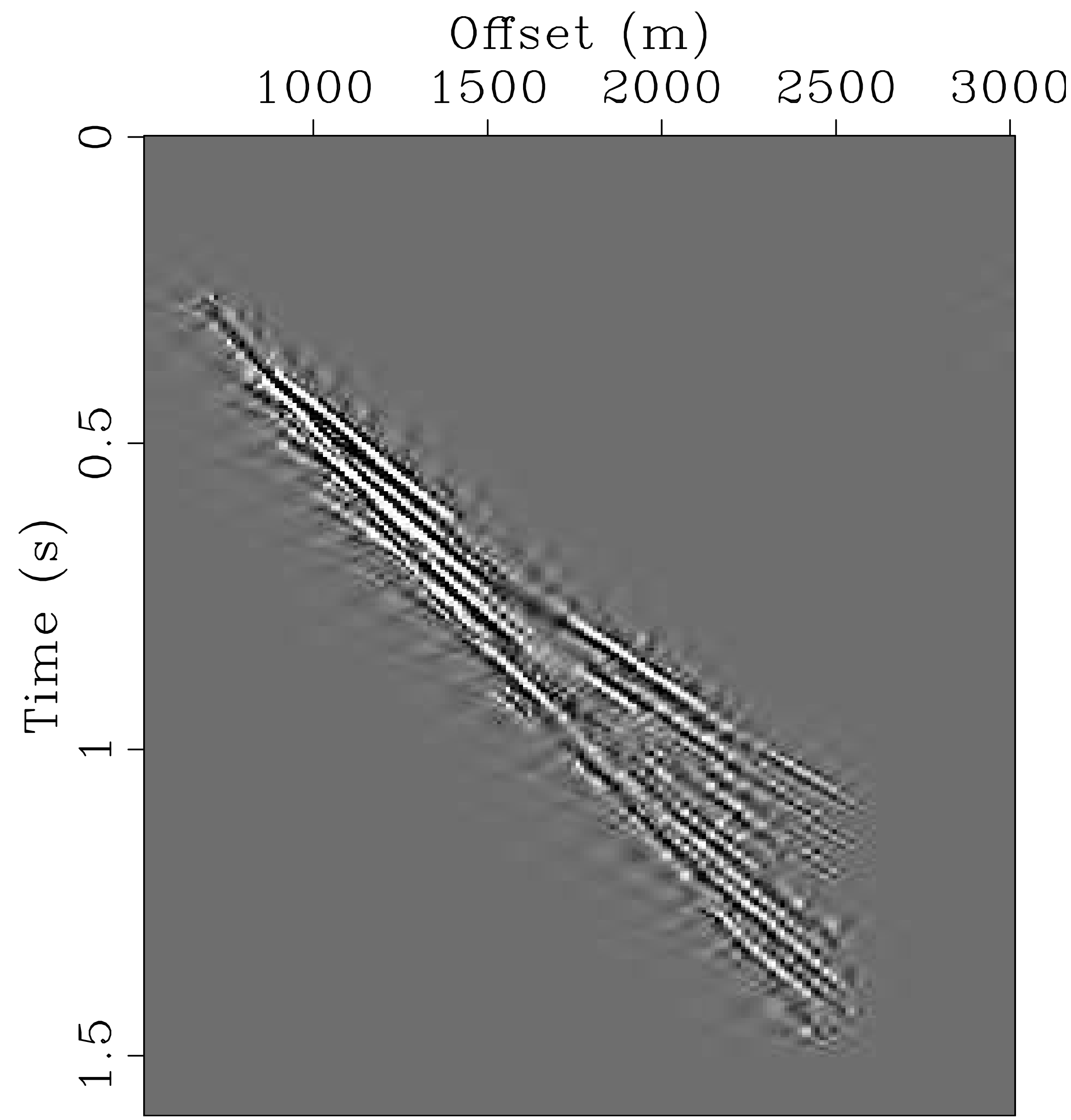




1 % of coefficients

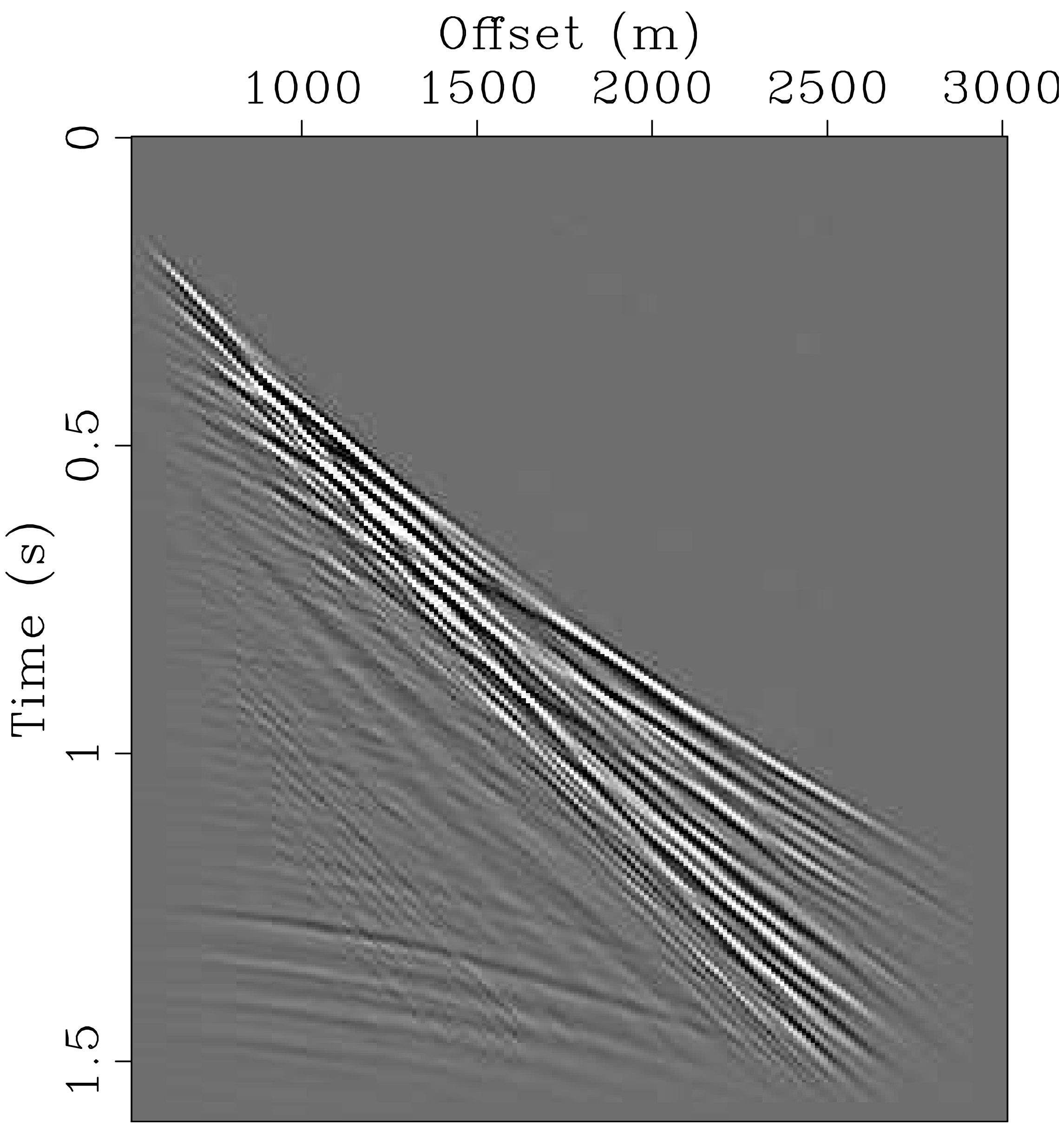
Wavelet reconstruction

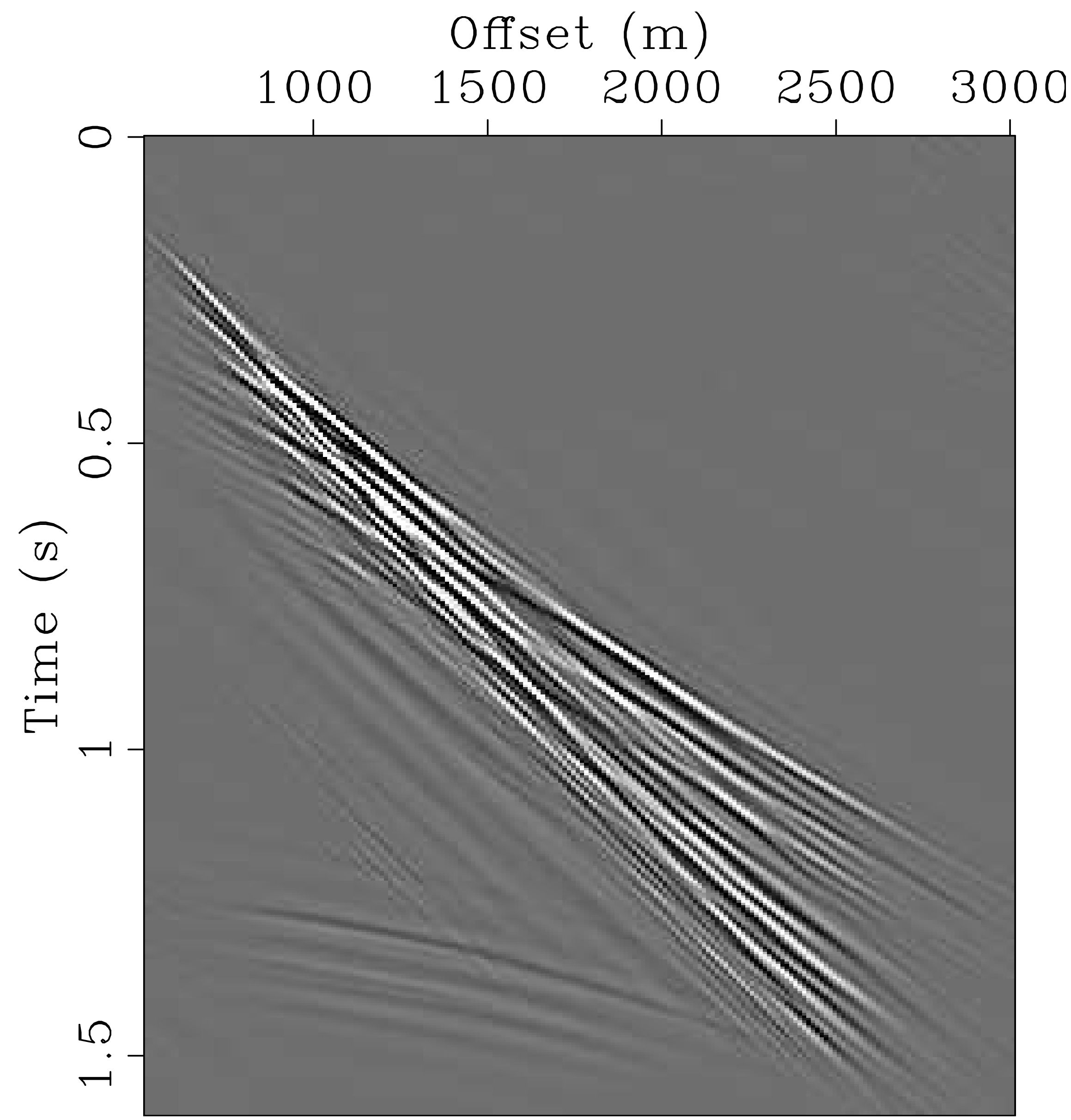




1 % of coefficients

Curvelet reconstruction



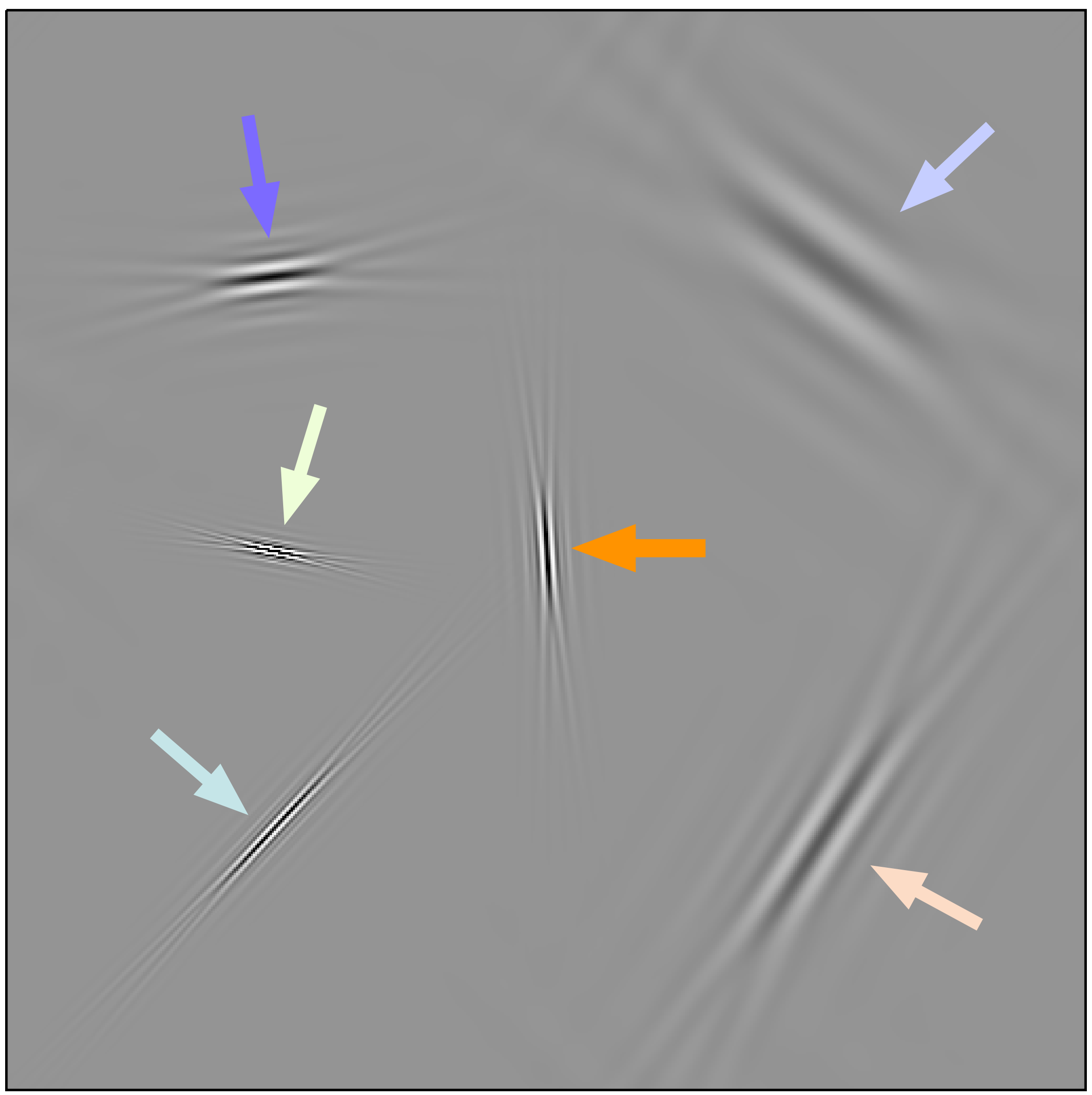


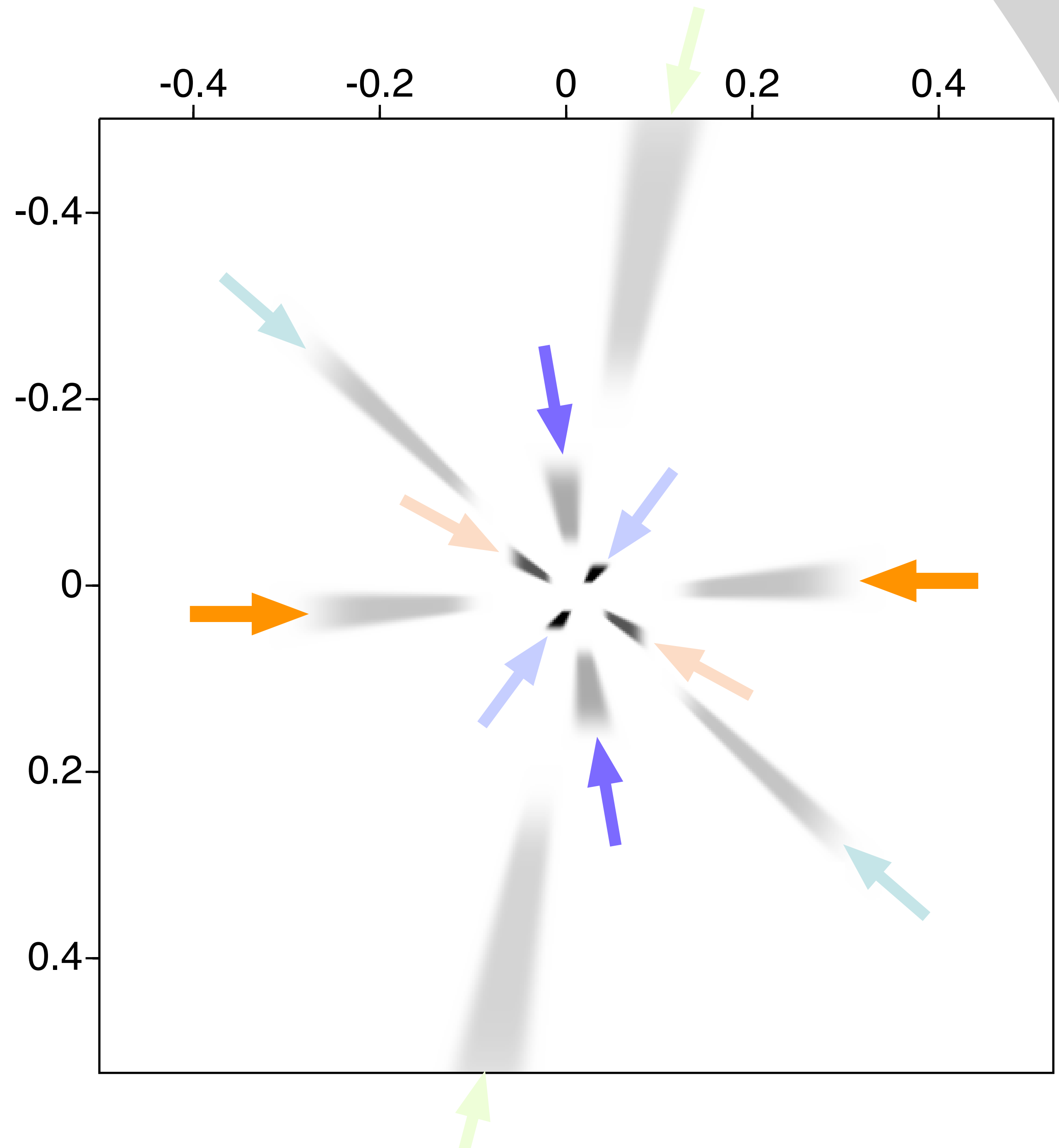
1 % of coefficients

SLIM

[Demanet et. al., '06]

Curvelets





Extension

Extend CS framework:

$$\mathbf{A} := \mathbf{R}\mathbf{M}\mathbf{S}^H$$

restriction matrix — measurement matrix — sparsity matrix

Expected to perform well when

$$\mu = \max_{1 \leq i \neq j \leq N} \left| \left(\mathbf{R}\mathbf{M}\mathbf{s}^i\right)^H \mathbf{R}\mathbf{M}\mathbf{s}^j \right|$$

Generalizes to *redundant* transforms for cases where

- max of RIP constants for **M**, **S** are small [Rauhut et.al, '06]
- or $\mathbf{S}\mathbf{S}^H\mathbf{x}$ remains sparse for **x** sparse [Candès et.al, '10]

Open research topic...

Empirical performance analysis

Selection of the appropriate sparsifying transform

- nonlinear approximation error

$$\mathrm{SNR}(\rho) = -20\log\frac{\|\mathbf{f}-\mathbf{f}_\rho\|}{\|\mathbf{f}\|} \quad \text{with} \quad \rho = k/P$$

- recovery error

$$\mathrm{SNR}(\delta) = -20\log\frac{\|\mathbf{f}-\tilde{\mathbf{f}}_\delta\|}{\|\mathbf{f}\|} \quad \text{with} \quad \delta = n/N$$

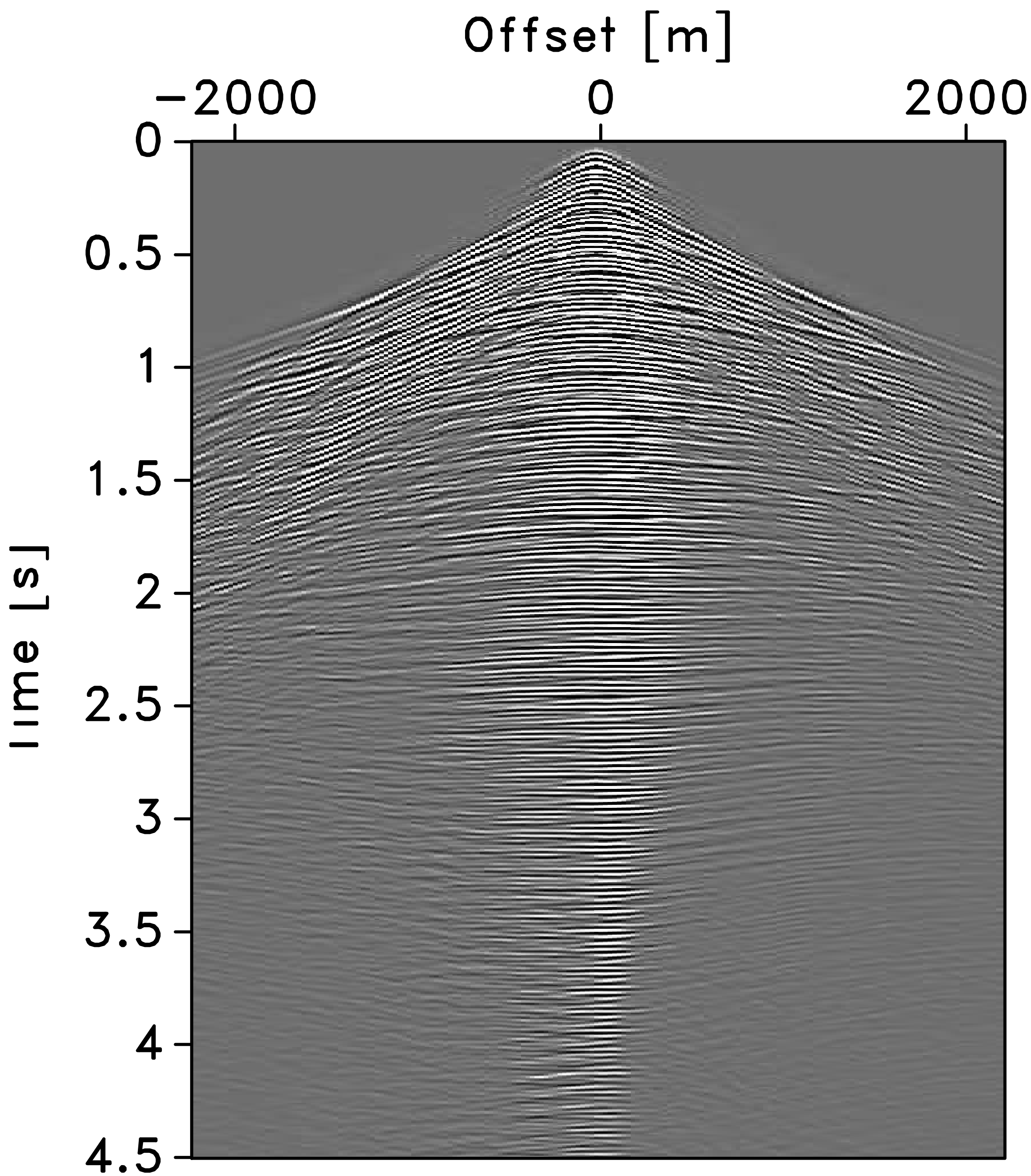
- oversampling ratio

$$\delta/\rho \quad \text{with} \quad \rho = \inf\{\tilde{\rho} : \quad \overline{\mathrm{SNR}}(\delta) \leq \mathrm{SNR}(\tilde{\rho})\}$$

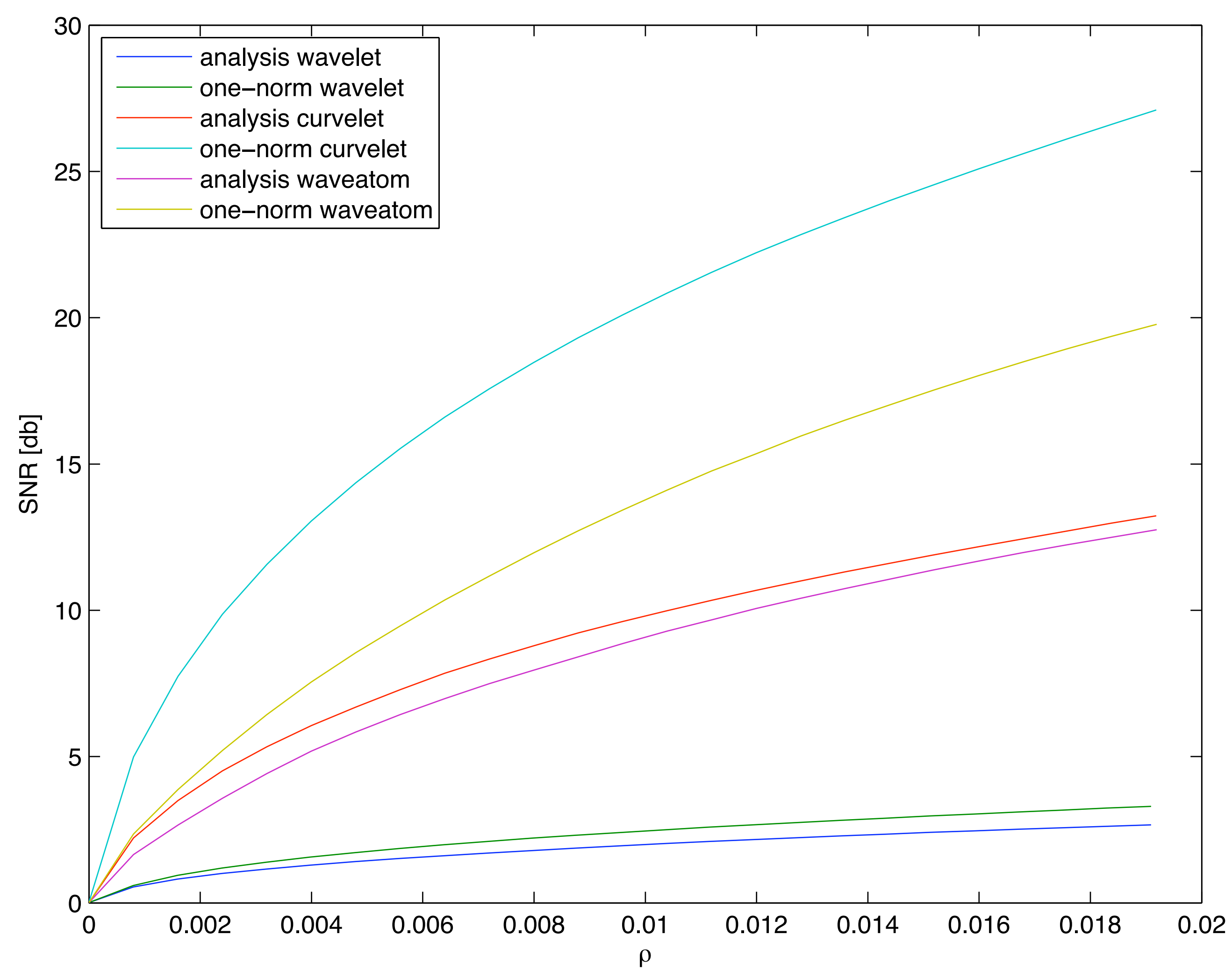
[FJH, '10]

Nonlinear approximation error

common receiver gather



recovery error



[FJH, '10]

Key elements

☑ *sparsifying transform*

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☐ *advantageous coarse sampling*

- generates incoherent random undersampling "noise" in the sparsifying domain

☐ *sparsity-promoting solver*

- requires few matrix-vector multiplications

Key elements

☑ *sparsifying transform*

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☑ *advantageous coarse sampling*

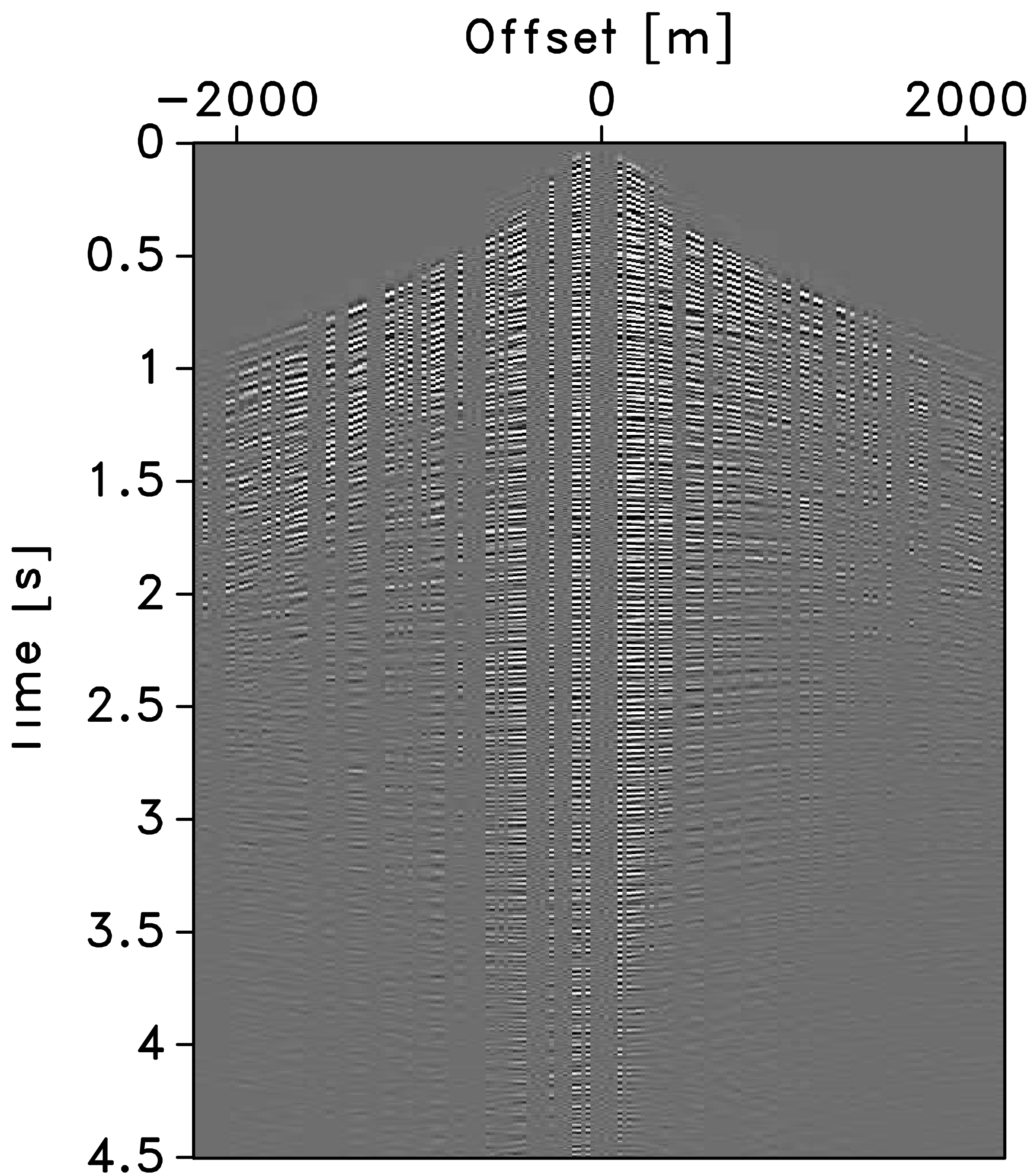
- generates incoherent random undersampling "noise" in the sparsifying domain
- does not create large gaps for measurement in the physical domain
- does not create coherent interferences in simultaneous acquisition

☐ *sparsity-promoting solver*

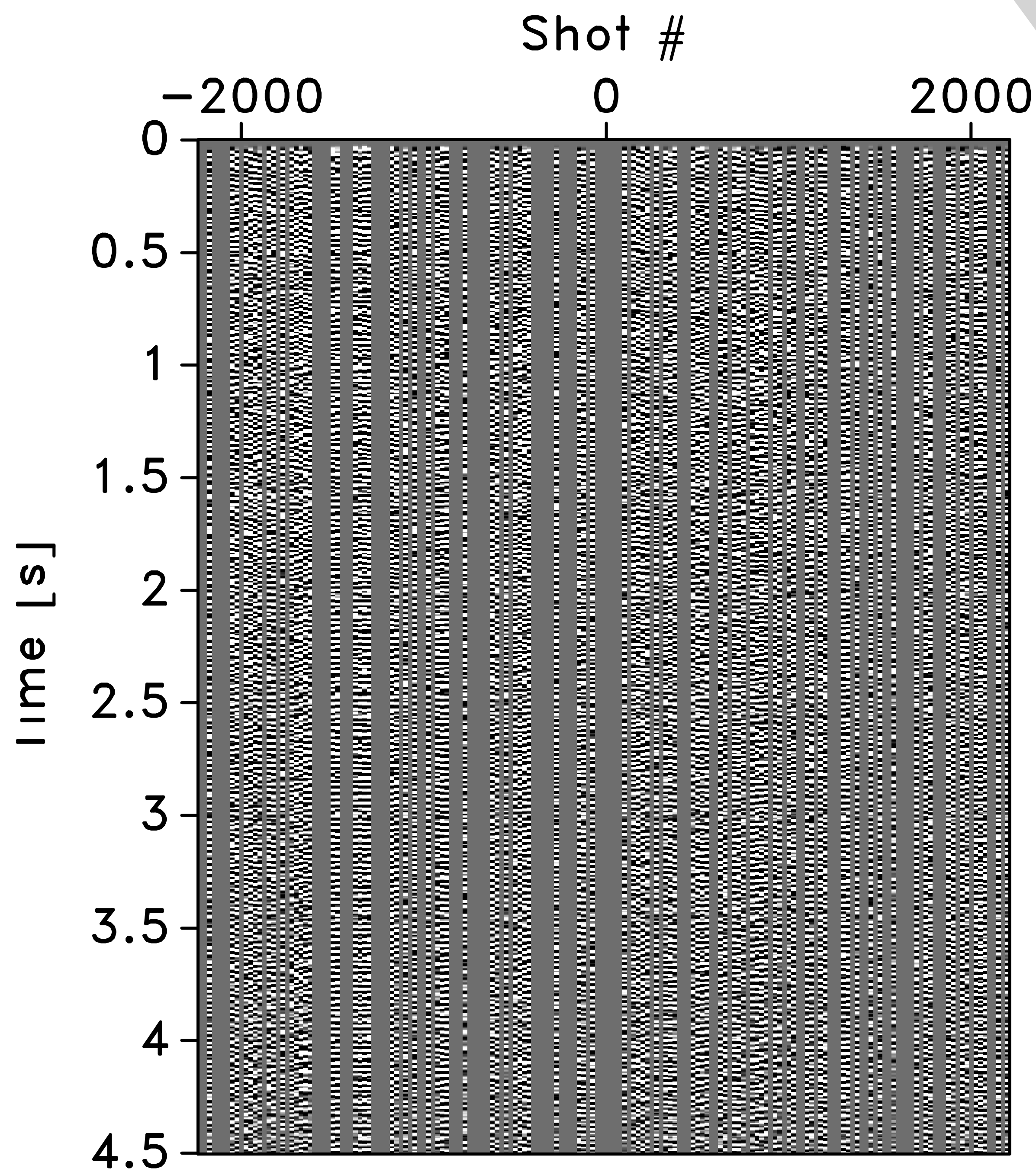
- requires few matrix-vector multiplications

Data

missing shots

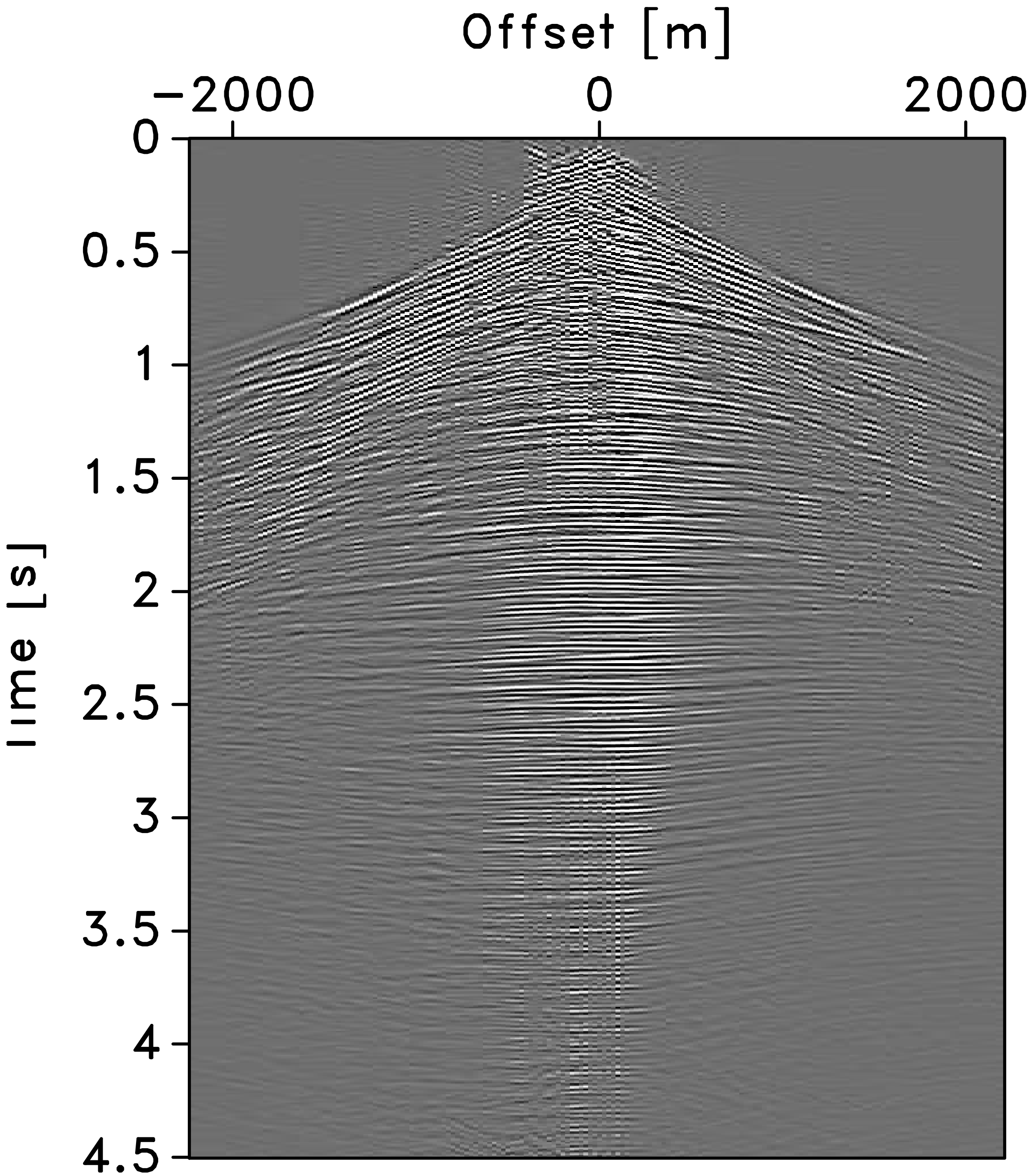


sim. shots

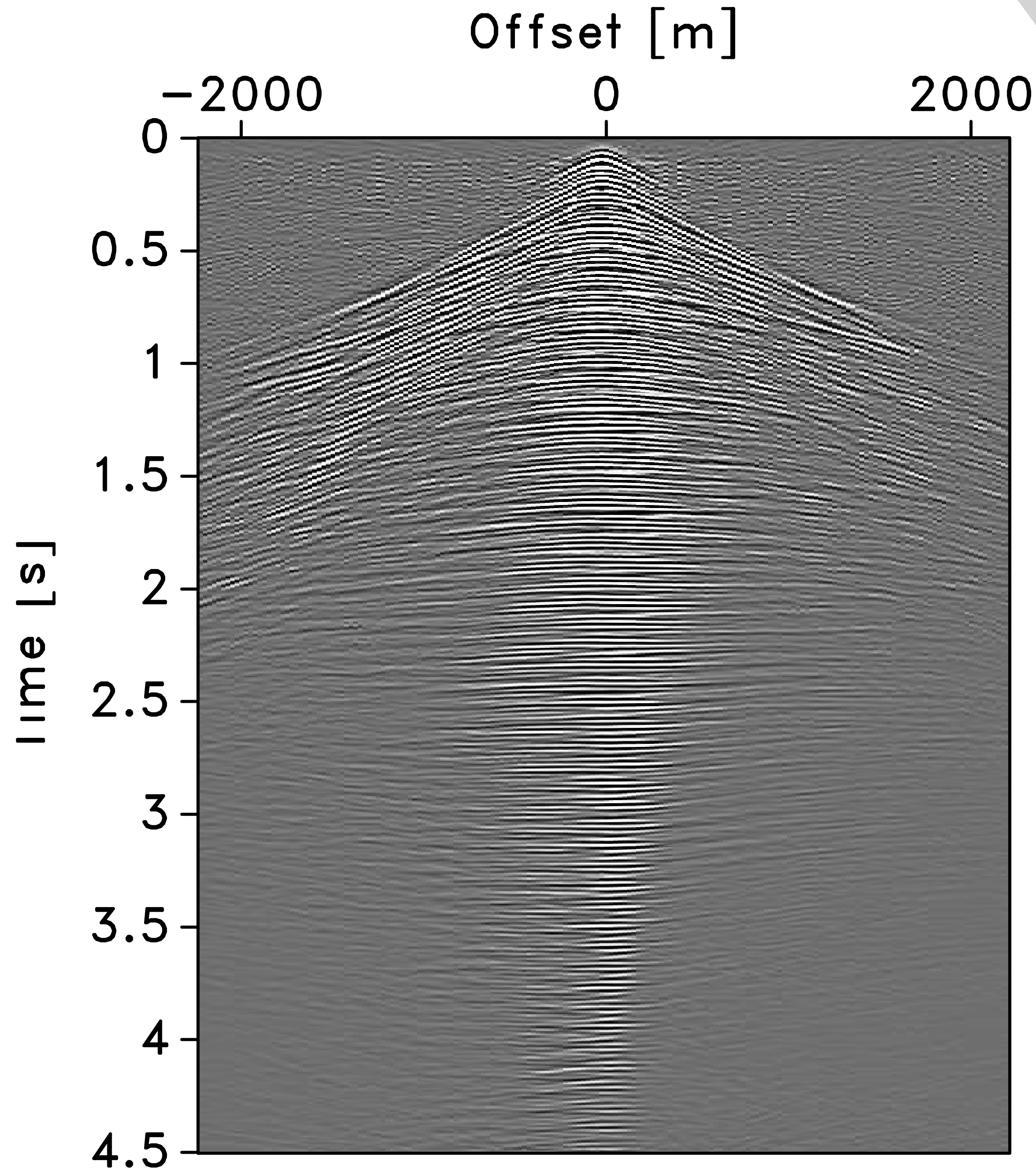


Sparse recovery

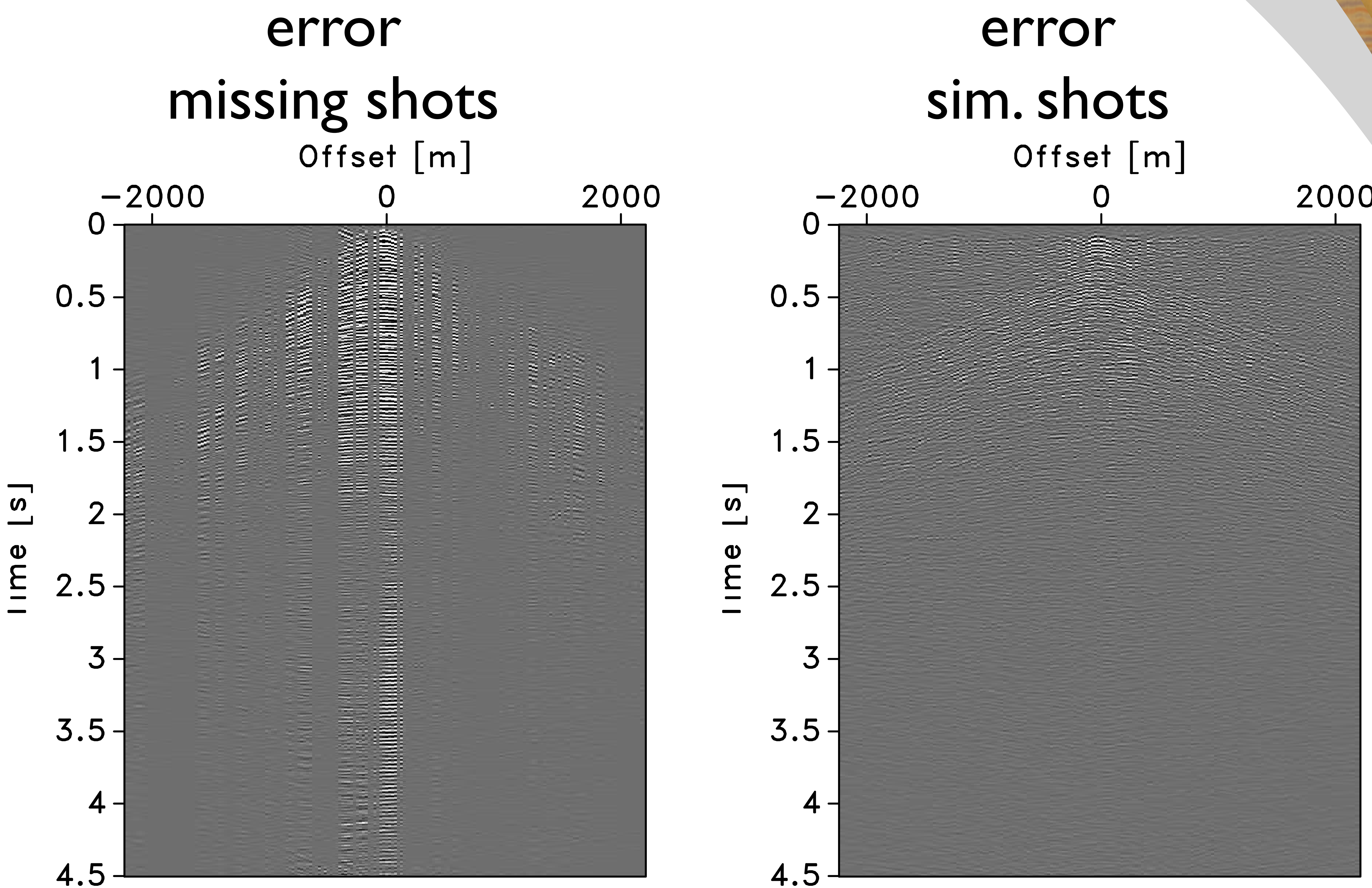
recovery
missing shots



recovery
sim. shots



Sparse recovery error



Empirical performance analysis

Selection of the appropriate sparsifying transform

- nonlinear approximation error

$$\mathrm{SNR}(\rho) = -20\log\frac{\|\mathbf{f}-\mathbf{f}_\rho\|}{\|\mathbf{f}\|} \quad \text{with} \quad \rho = k/P$$

➡ recovery error

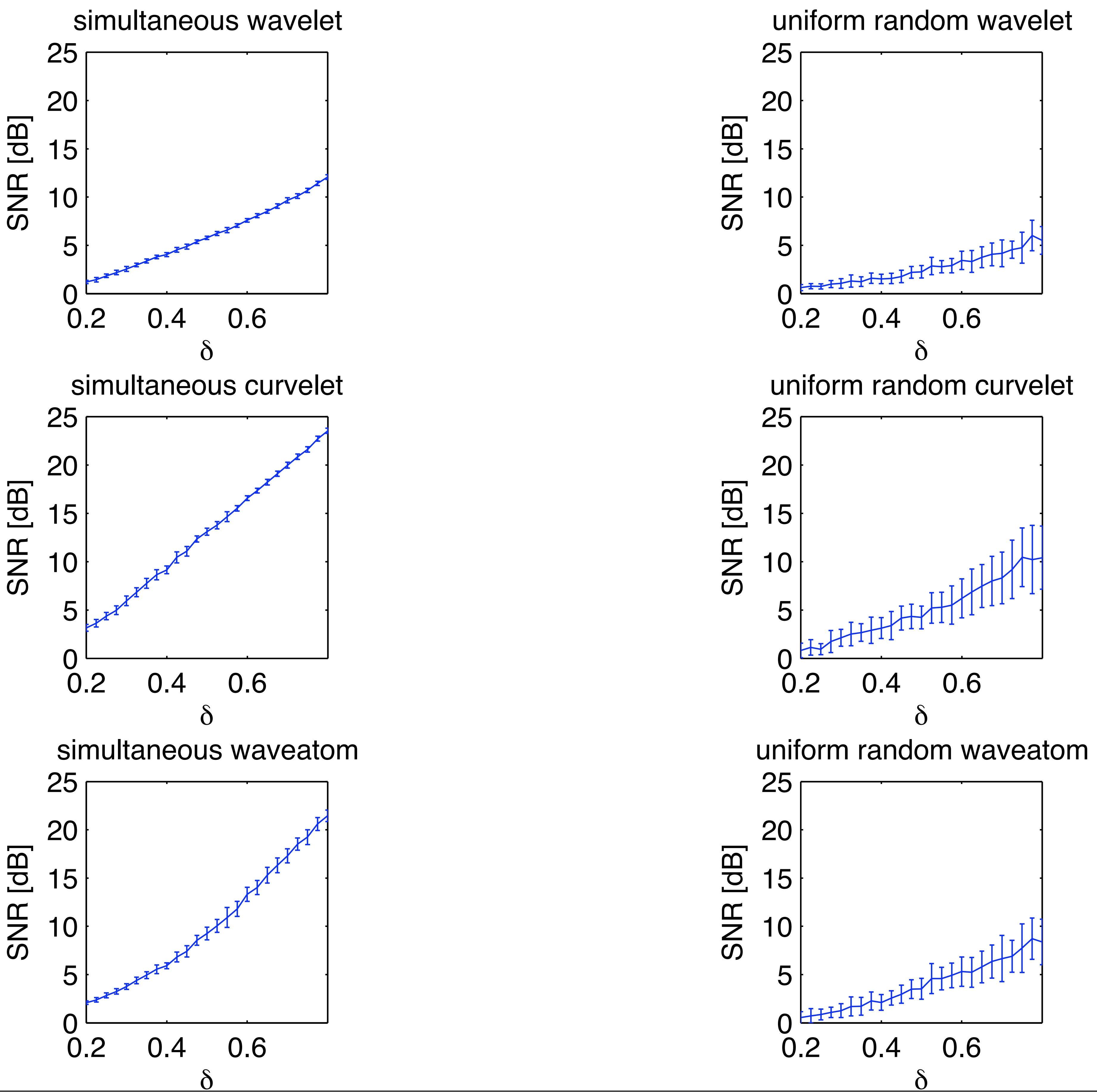
$$\mathrm{SNR}(\delta) = -20\log\frac{\|\mathbf{f}-\tilde{\mathbf{f}}_\delta\|}{\|\mathbf{f}\|} \quad \text{with} \quad \delta = n/N$$

- oversampling ratio

$$\delta/\rho \quad \text{with} \quad \rho = \inf\{\tilde{\rho} : \quad \overline{\mathrm{SNR}}(\delta) \leq \mathrm{SNR}(\tilde{\rho})\}$$

[FJH, '10]

Multiple experiments



Empirical performance analysis

Selection of the appropriate sparsifying transform

- nonlinear approximation error

$$\mathrm{SNR}(\rho) = -20 \log \frac{\|\mathbf{f} - \mathbf{f}_\rho\|}{\|\mathbf{f}\|} \quad \text{with} \quad \rho = k/P$$

- recovery error

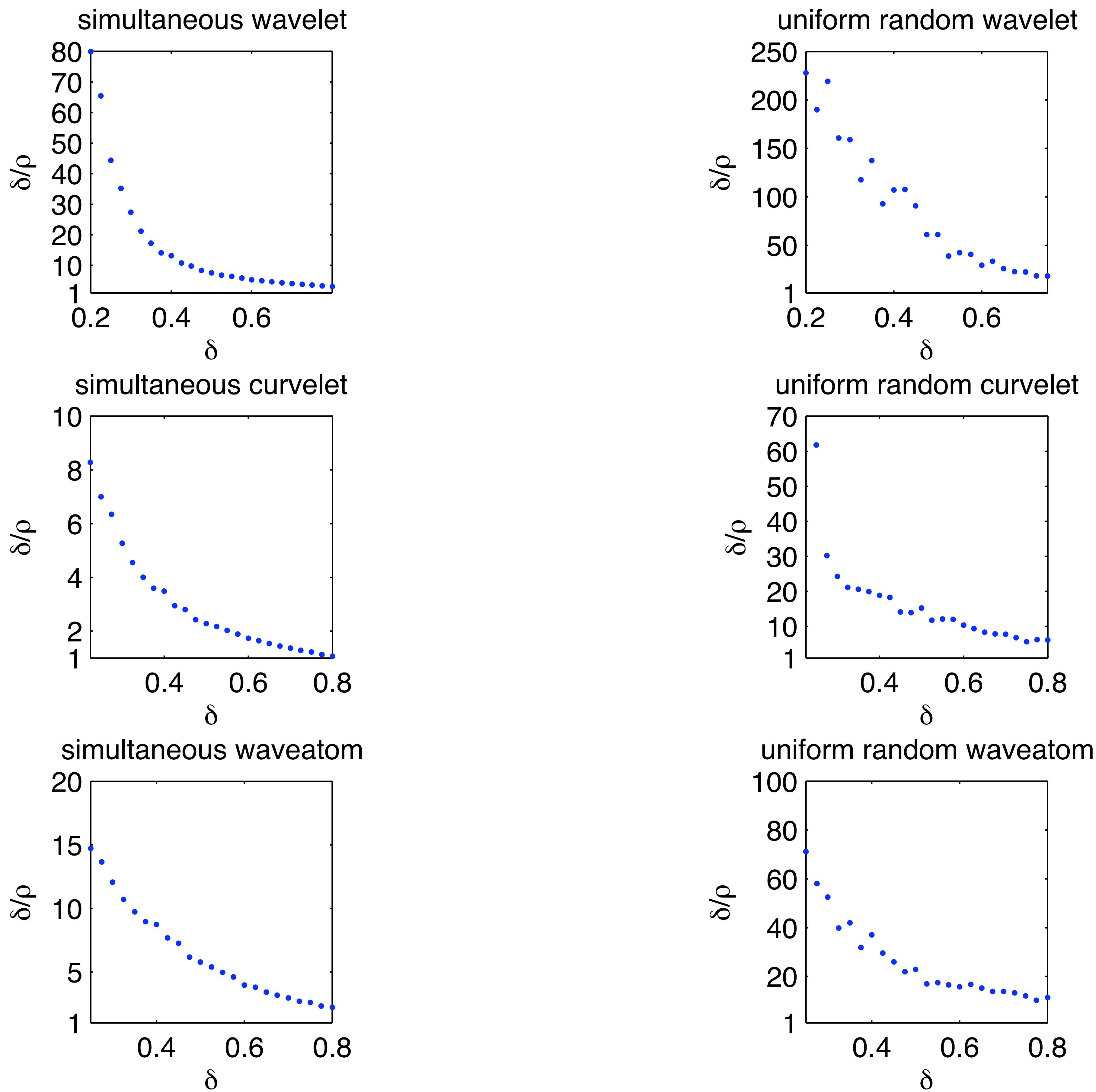
$$\mathrm{SNR}(\delta) = -20 \log \frac{\|\mathbf{f} - \tilde{\mathbf{f}}_\delta\|}{\|\mathbf{f}\|} \quad \text{with} \quad \delta = n/N$$

➡ oversampling ratio

$$\delta/\rho \quad \text{with} \quad \rho = \inf\{\tilde{\rho} : \quad \overline{\mathrm{SNR}}(\delta) \leq \mathrm{SNR}(\tilde{\rho})\}$$

[FJH, '10]

Oversampling ratios



Key elements

☑ *sparsifying transform*

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☑ *advantageous coarse sampling (mixing)*

- generates incoherent random undersampling "noise" in the sparsifying domain
- does not create large gaps for measurement in the physical domain
- does not create coherent interferences in simultaneous acquisition

☐ *sparsity-promoting solver*

- requires few matrix-vector multiplications

Reality check

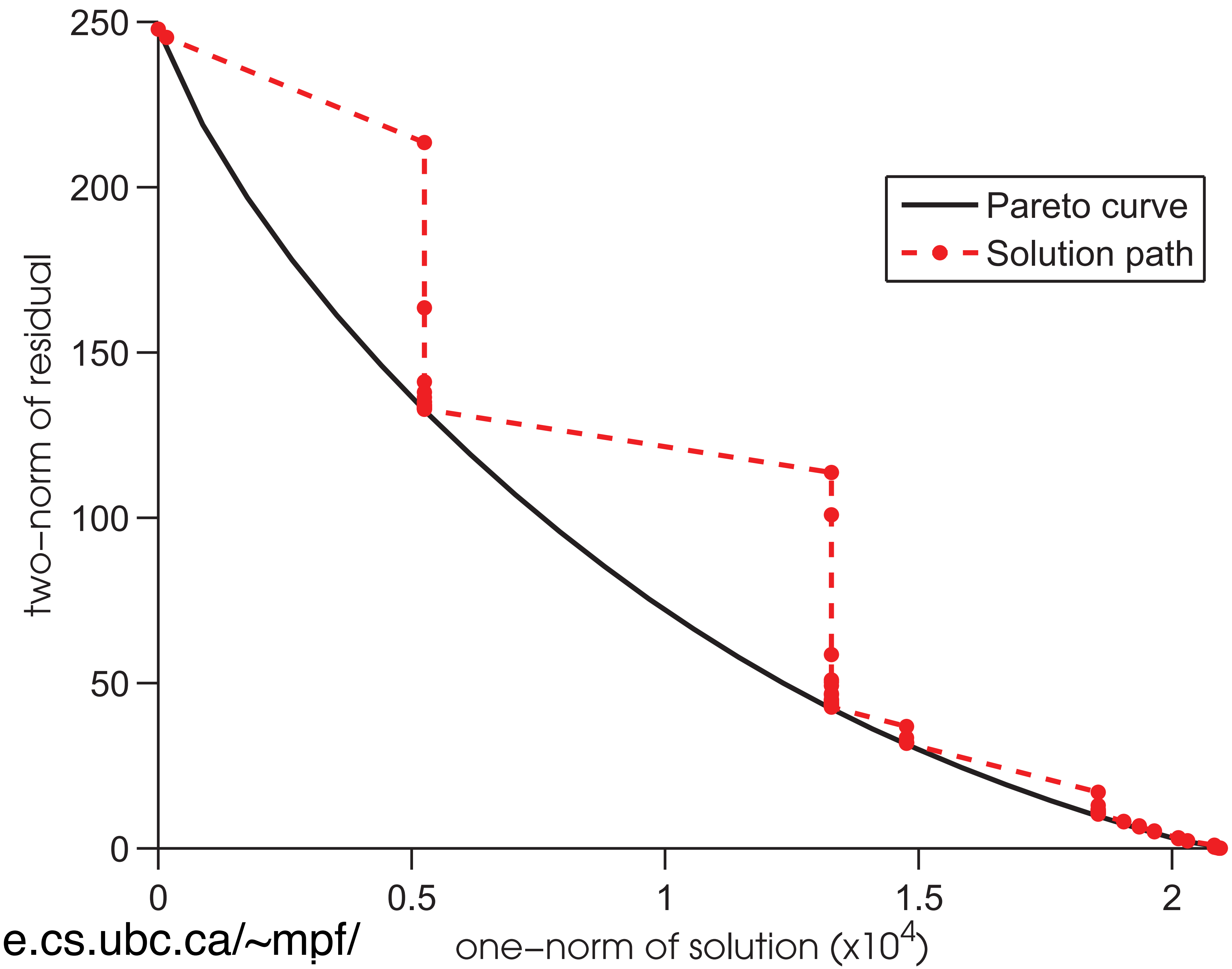
"When a traveler reaches a fork in the road, the l_1-norm tells him to take either one way or the other, but the l_2 -norm instructs him to head off into the bushes."

John F. Claerbout and Francis Muir, 1973

[van den Berg & Friedlander, '08]
[Hennenfent, FJH, et. al, '08]

SLIM

One-norm solver



from http://people.cs.ubc.ca/~mpf/

Key elements

SLIM

☑ *sparsifying transform*

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☑ *advantageous coarse sampling (mixing)*

- generates incoherent random undersampling "noise" in the sparsifying domain
- does not create large gaps for measurement in the physical domain
- does not create coherent interferences in simultaneous acquisition

☑ *sparsity-promoting solver*

- requires few matrix-vector multiplications

Recent results

Recovery of seismic lines based

- on “separable” *sparsifying* transform

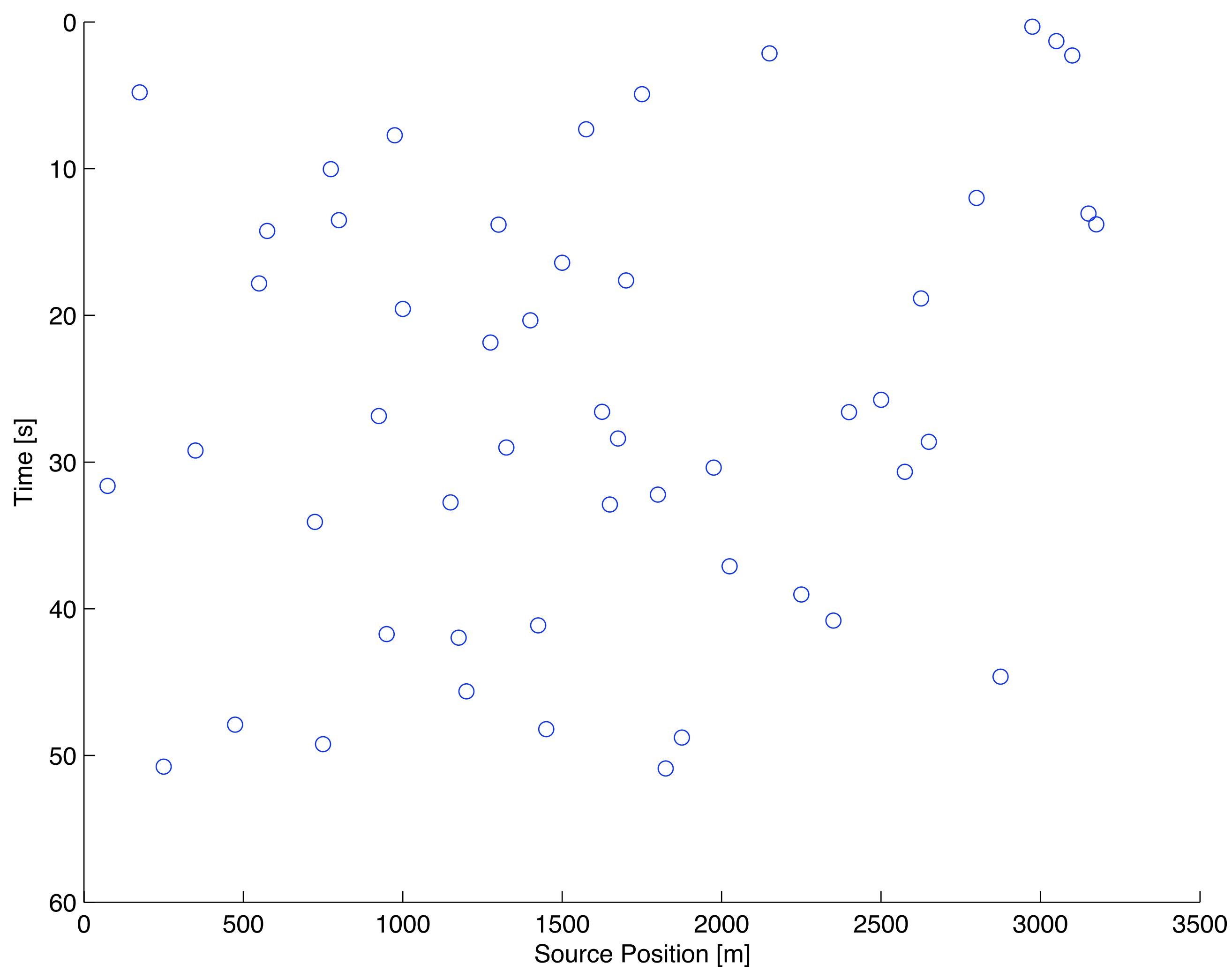
$$\mathbf{S} = \mathbf{C} \otimes \mathbf{W}$$

- *favorable* simultaneous acquisition

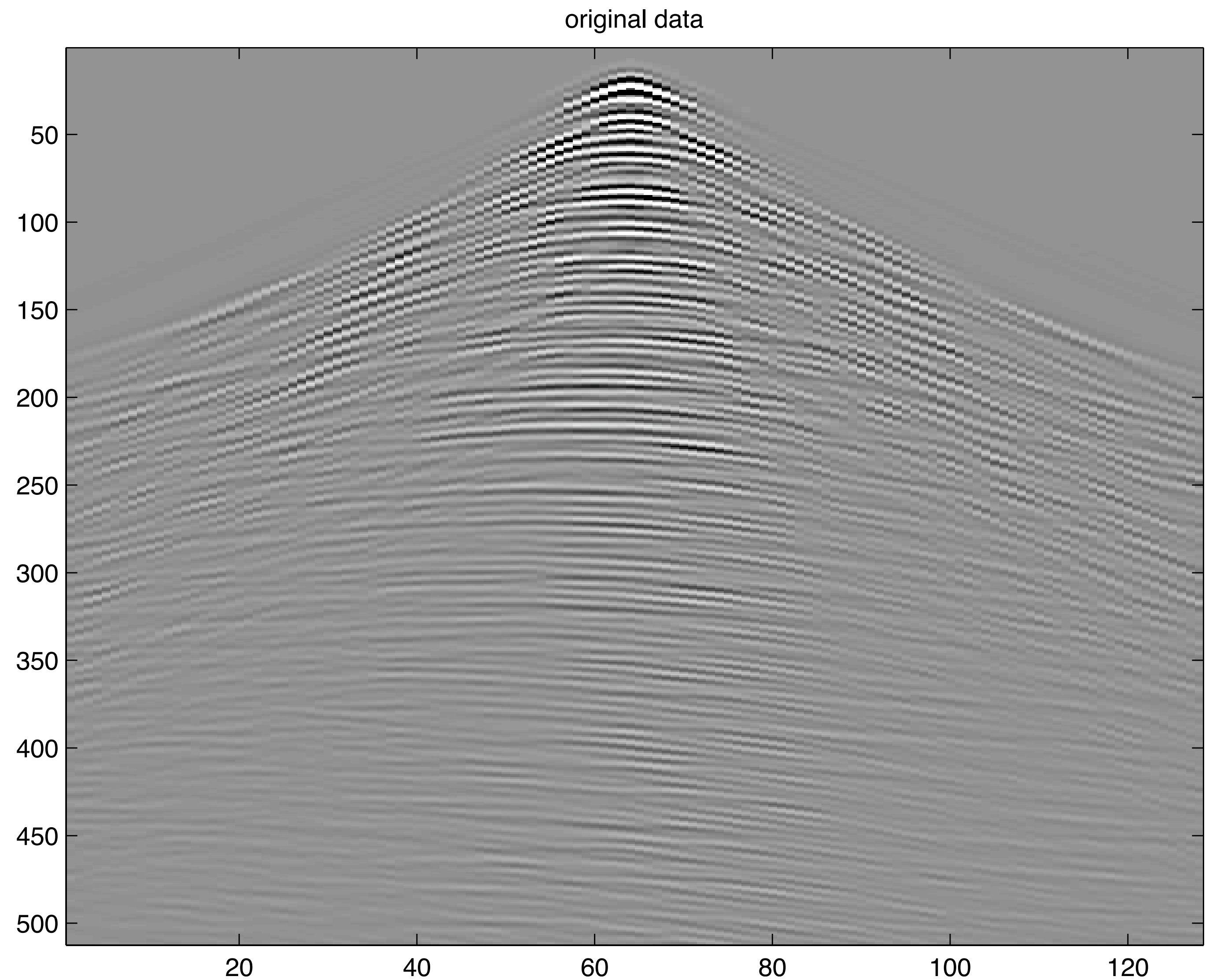
Consider “Marine” case

Simultaneous sources

Marine case

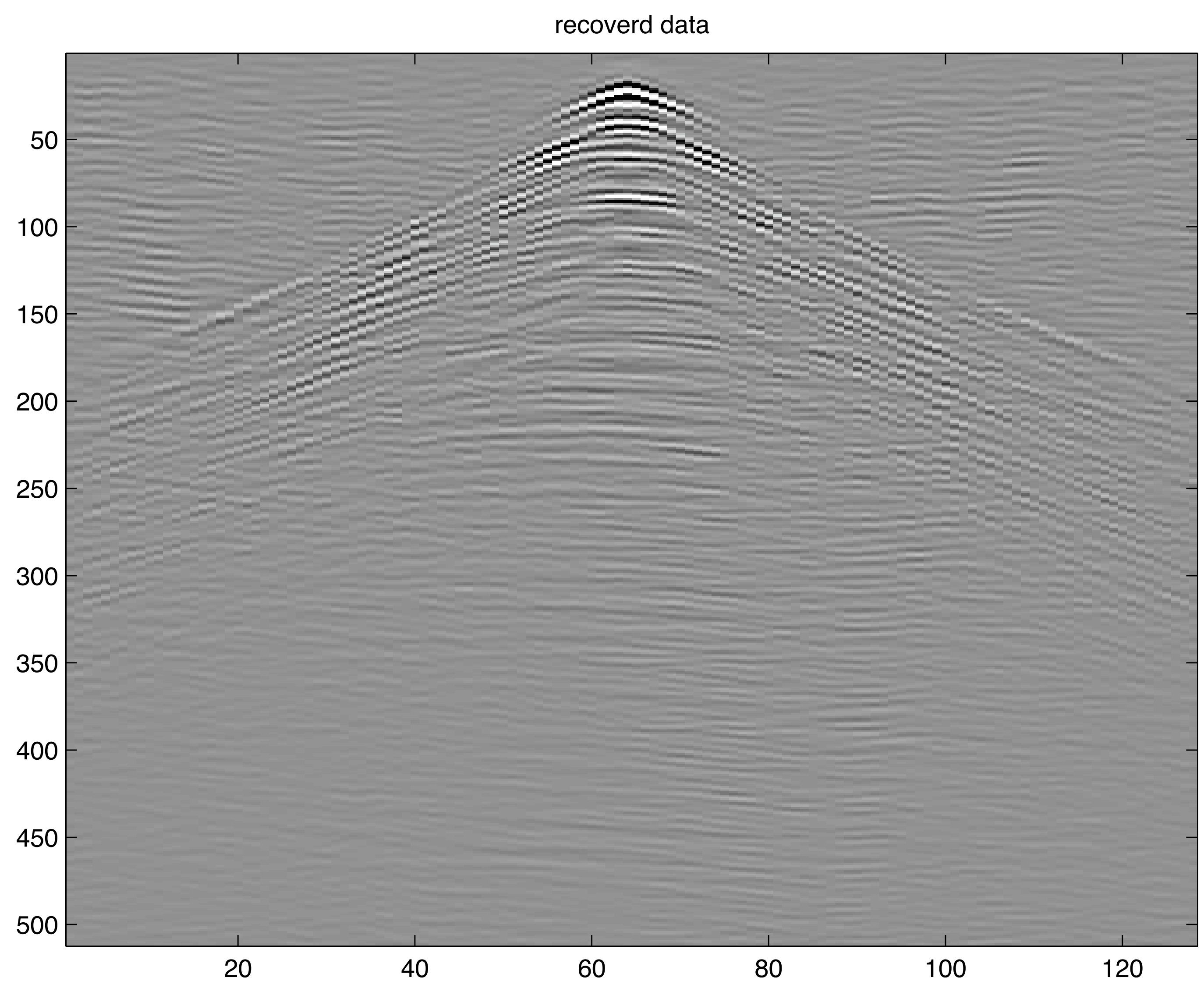


Original data



Recovered data

40 % of shots in 20 % of recording time



SLIM

Observations

Controllable error for reconstruction from *randomized* subsamplings

Oversampling compared to *conventional compression* is small

Combination of *sampling* & *encoding* into a *single* ***linear*** step has profound implications

- *acquisition* costs ***no*** longer determined by *resolution & size*
- *but* by *transform-domain sparsity & recovery error*

3-D Curvelets and simultaneous acquisition perform the best

Extensions

Include more "physics" in the formulation *via*

- *discretization* of *integral* equations of the *second* kind
- *prediction* of *surface-related* multiples [Lin & FJH, 09-10]
- *linearized-scattering* operator [Lin et. al.,'10]

Incorporate *dimensionality reductions* in *full-waveform* inversion

- via creation of *supershots*

FWI formulation

Multiexperiment unconstrained optimization problem:

$$\min_{\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{D}-\mathcal{F}[\mathbf{m};\mathbf{Q}]\|_{2,2}^2 \quad \text{with} \quad \mathcal{F}[\mathbf{m};\mathbf{Q}] := \mathbf{P}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{Q}$$

- requires large number of PDE solves
- linear in the sources
- apply *randomized* dimensionality reduction

[Tarantola, 84; Pratt, '98; Plessix, '06]

[FJH et.al., '08-10', Krebs et.al., '09, Operto et. al., '09]
[Haber, Chung, and FJH, '10]

SLIM

Reduced FWI formulation

Multiexperiment unconstrained optimization problem:

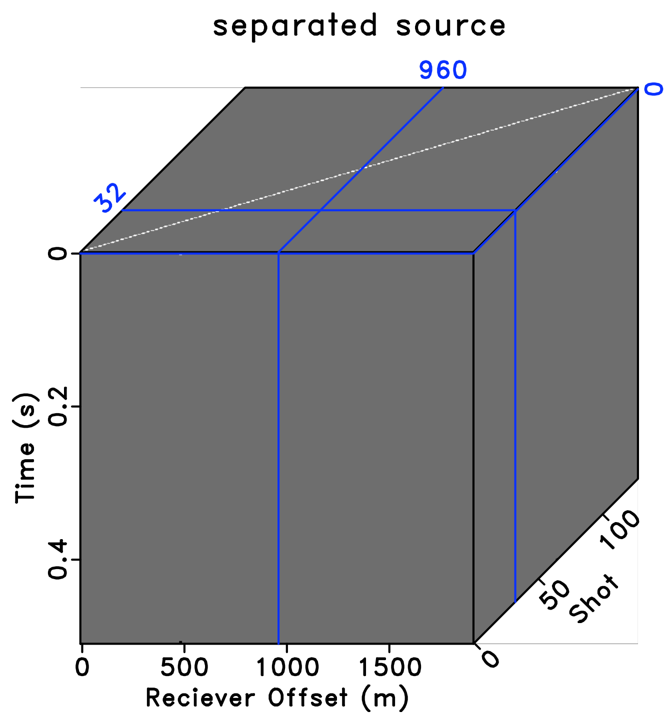
$$\min_{\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\underline{\mathbf{D}} - \mathcal{F}[\mathbf{m};\underline{\mathbf{Q}}]\|_{2,2}^2 \quad \text{with} \quad \mathcal{F}[\mathbf{m};\underline{\mathbf{Q}}] := \mathbf{P}\underline{\mathbf{H}}^{-1}\underline{\mathbf{Q}}$$

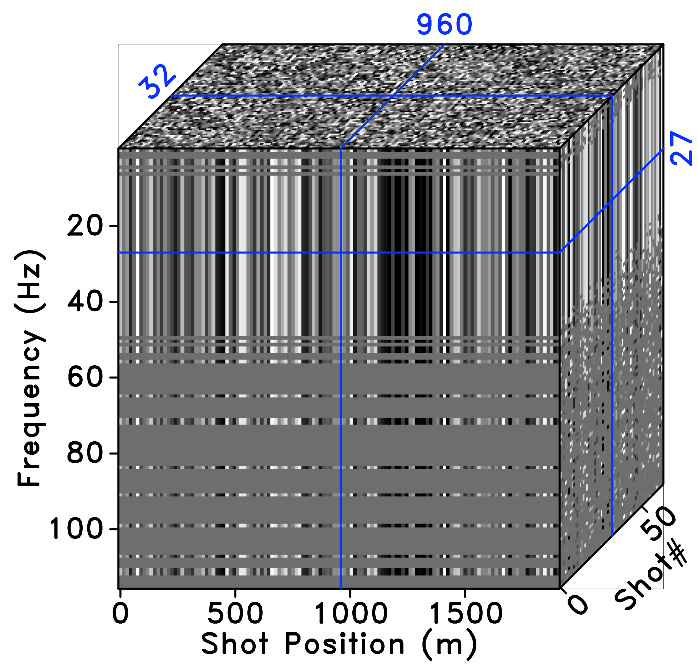
- requires *smaller* number of PDE solves
- explores *linearity* in the *sources* & *block-diagonal* structure of the *Helmholtz system*
- uses *randomized* frequency *selection* and *phase encoding*

[FJH et. al. '08-'10]

Batch/mini experiment

adapted from FJH et. al. ,09





$\mathbf{Q}$ $\qquad$ $\underline{\mathbf{Q}} = \mathbf{RMQ}$

Collection of K *simultaneous-source* experiments with *batch* size $K \ll n_f \times n_s$

Math [Romberg, '07]

Compressive-sampling operator

$$\mathbf{RM} = (\mathbf{R}^{\boldsymbol{\Sigma}}\mathbf{M}^{\boldsymbol{\Sigma}} \otimes \mathbf{I} \otimes \mathbf{R}^{\boldsymbol{\Omega}})\mathbf{F}_3$$

with

$$\mathbf{M}^{\boldsymbol{\Sigma}} = \mathrm{sign}(\eta) \odot \mathbf{F}_1^H e^{j\theta}$$

where $\theta \in \mathrm{Uniform}(-\pi,\, \pi]$, and $\eta \in \mathrm{Normal}(0, 1)$

Interpretations

Consider *randomized* dimensionality reduction as instances of

- *stochastic optimization* & *machine learning*
- *compressive sensing* [FJH et. al, '08-'10]

Stochastic optimization

Replace *deterministic*-optimization problem

$$\min_{\mathbf{m}\in\boldsymbol{\mathcal{M}}} f(\mathbf{m}) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\|\mathbf{d}_i - \mathcal{F}[\mathbf{m};\mathbf{q}_i]\|_2^2$$

with *sum* cycling over *different sources & corresponding shot records*
(columns of D & Q)

[Natterer, '01]

SLIM

Stochastic *average* approximation

[Haber, Chung, and FJH, '10]

by *a stochastic-optimization* problem

$$\min_{\mathbf{m}\in\boldsymbol{\mathcal{M}}} \mathbf{E}_{\mathbf{w}}\{f(\mathbf{m},\mathbf{w}) = \frac{1}{2}\|\mathbf{D}\mathbf{w} - \mathcal{F}[\mathbf{m};\mathbf{Q}\mathbf{w}]\|_2^2\}$$

$$\approx \frac{1}{K}\sum_{j=1}^{K}\frac{1}{2}\|\underline{\mathbf{d}}_j - \mathcal{F}[\mathbf{m};\underline{\mathbf{q}}_j]\|_2^2$$

with $\mathbf{w} \in N(0,1)$ and $\mathbf{E}_{\mathbf{w}}\{\mathbf{w}\mathbf{w}^H\} = \mathbf{I}$

and $\underline{\mathbf{d}}_j = \mathbf{D}\mathbf{w}_j$, $\underline{\mathbf{q}}_j = \mathbf{Q}\mathbf{w}_j$

Stochastic *average* approximation

In the *limit* $K \to \infty$, *stochastic* & *deterministic* formulations are *identical*

We *gain* as long as $K \ll N$ ...

Since the error in *Monte-Carlo* methods decays only slowly $(\mathcal{O}(K^{-1/2}))$

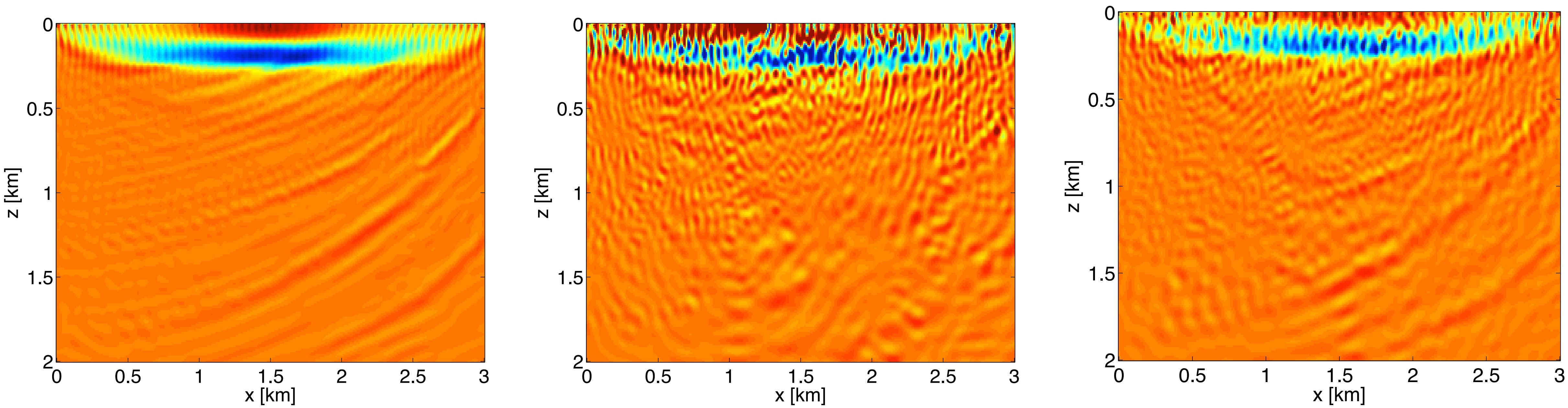
this approach may be problematic...

However, the location for the *minimum* of the *misfit* may be relatively *robust*...

Stylized example

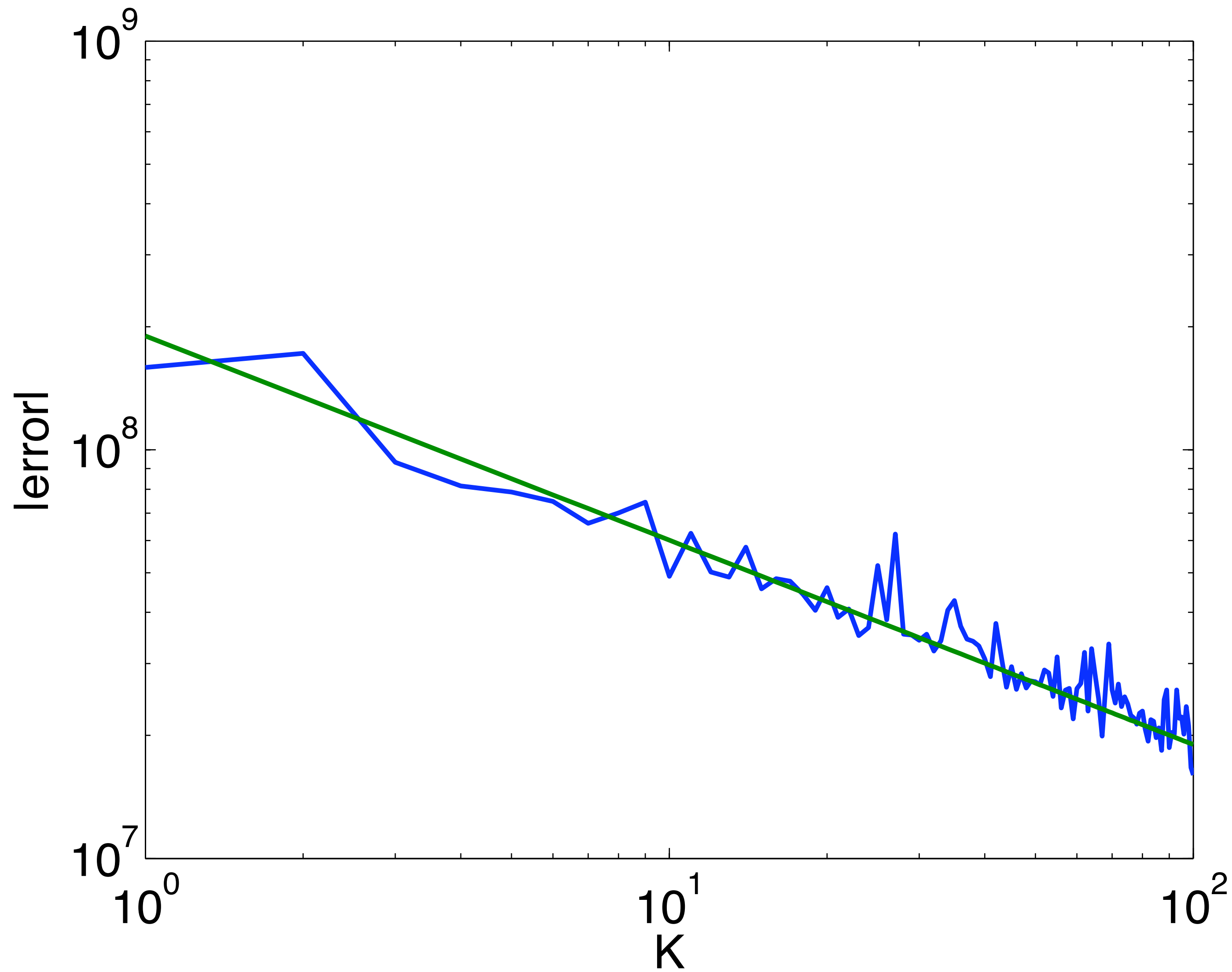
Search direction for batch size K:

$$\mathbf{g}_K \approx \frac{1}{K}\sum_{j=1}^{K} \nabla\mathcal{F}^*[\mathbf{m};\underline{\mathbf{q}}_j]\delta\underline{\mathbf{d}}_j$$



full K=1 K=5

Decay

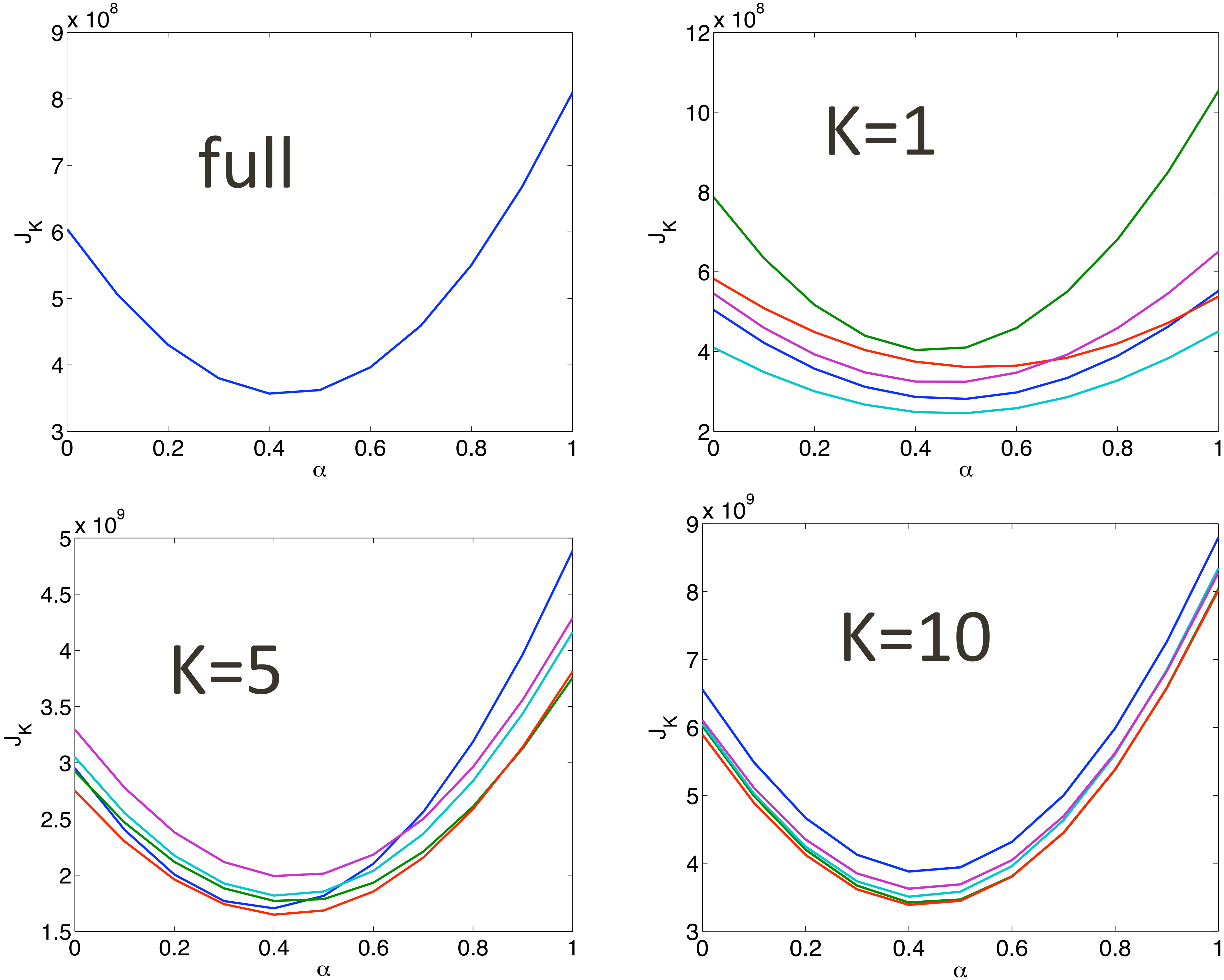


error between full and sampled gradient

Misfit functional

[adapted from Haber, Chung, and FJH, '10]

$$f_K(\mathbf{g}_K) = \frac{1}{K}\sum_{j=1}^{K}\frac{1}{2}\|\underline{\mathbf{d}}_j - \mathcal{F}[\mathbf{m} + \alpha\mathbf{g}_K; \underline{\mathbf{q}}_j]\|_2^2$$



Stochastic approximation [Bertsekas,' '96; Nemirovski, '09]

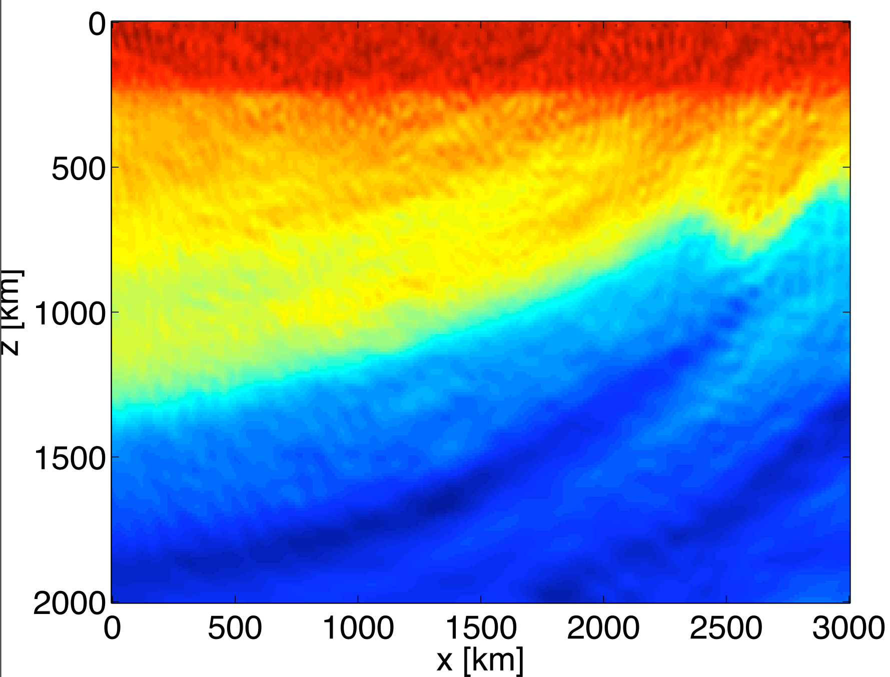
Use *different* simultaneous shots for each *subproblem*, i.e.,

$$\underline{\mathbf{Q}} \quad \mapsto \quad \underline{\mathbf{Q}}^{k}$$

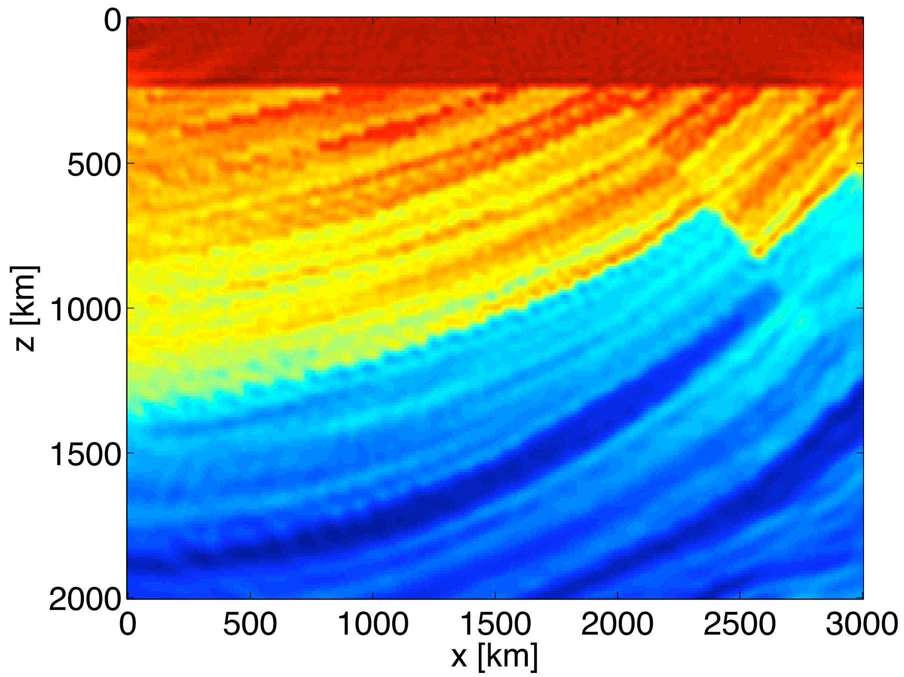
Requires *fewer* PDE solves for each GN *subproblem*...

- corresponds to *stochastic approximation* [Nemirovski, '09]
- related to Kaczmarz ('37) method applied by Natterer, '01
- *supersedes ad hoc* approach by Krebs *et.al.*, '09

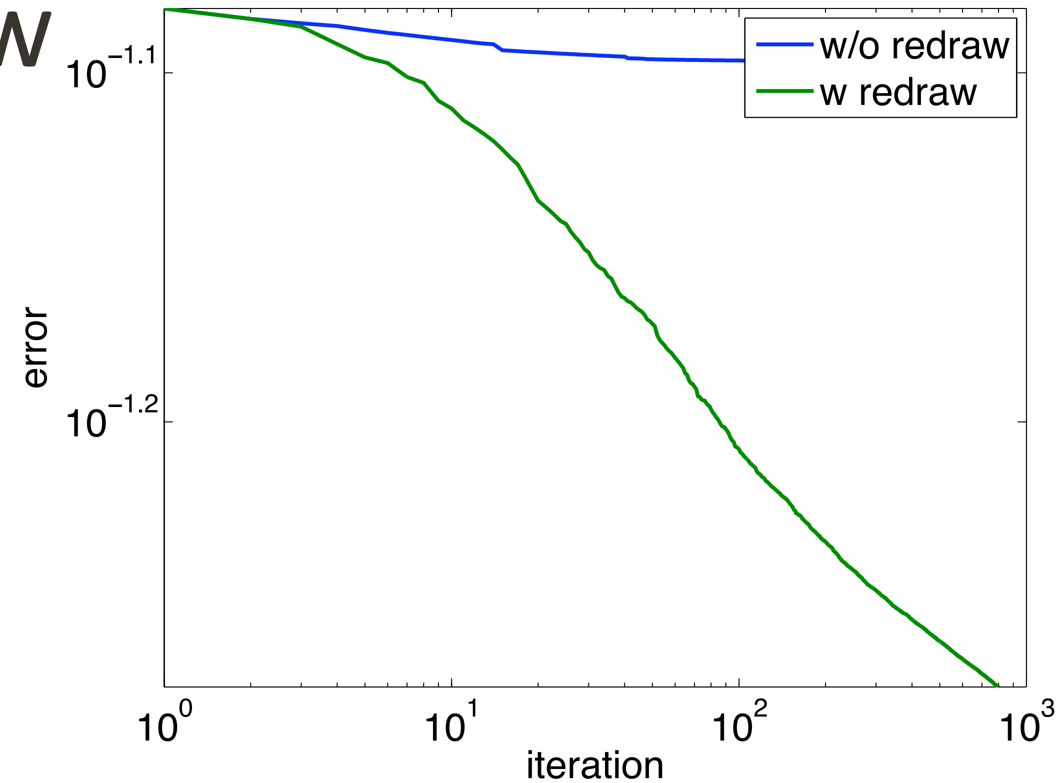
K=1 w and w/o redraw [noise-free case]



w/o redraw



w redraw



model error K=1, no averaging

Known issues

Renewals introduce *stochasticity* in the *gradients*

May lead to

- lack of convergence
- sensitivity to noise in data [Krebs, ’09-’10]

Solutions

- increase the batch size
- average over the past model updates

Stochastic approximation

Algorithm 1: Stochastic gradient descent

Result: Output estimate for the model $\mathbf{m}$

$\mathbf{m} \longleftarrow \mathbf{m}_0$; $k \longleftarrow 0$; `// initial model`

while not converged **do**

$\mathbf{g}^k \longleftarrow \frac{1}{K}\sum_{j=1}^{K} \nabla\boldsymbol{\mathcal{F}}^*[\mathbf{m}^{k-1}, \underline{\mathbf{q}}_j^k](\underline{\mathbf{d}}_j^k - \boldsymbol{\mathcal{F}}[\mathbf{m}^{k-1}, \underline{\mathbf{q}}_j^k])$; `// gradient`

$\underline{\mathbf{m}}^{k+1} \longleftarrow \mathbf{m}^k - \gamma^k \mathbf{g}^k$; `// update with linesearch`

$\mathbf{m}^{k+1} = \frac{1}{k+1}\left(\sum_{i=1}^{k} \mathbf{m}^i + \underline{\mathbf{m}}^{k+1}\right)$; `// average`

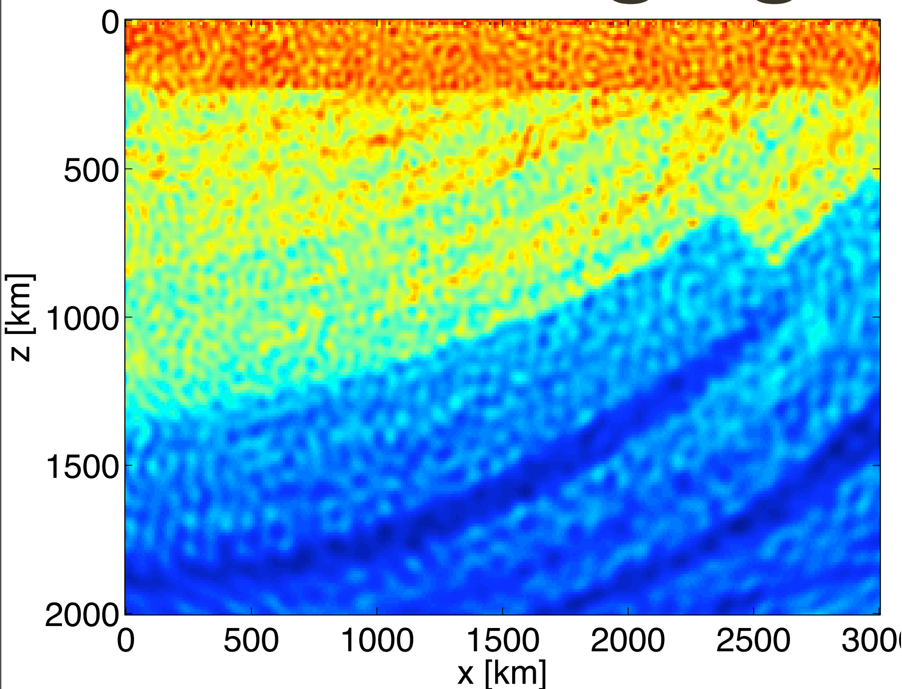
$k \longleftarrow k+1$;

end

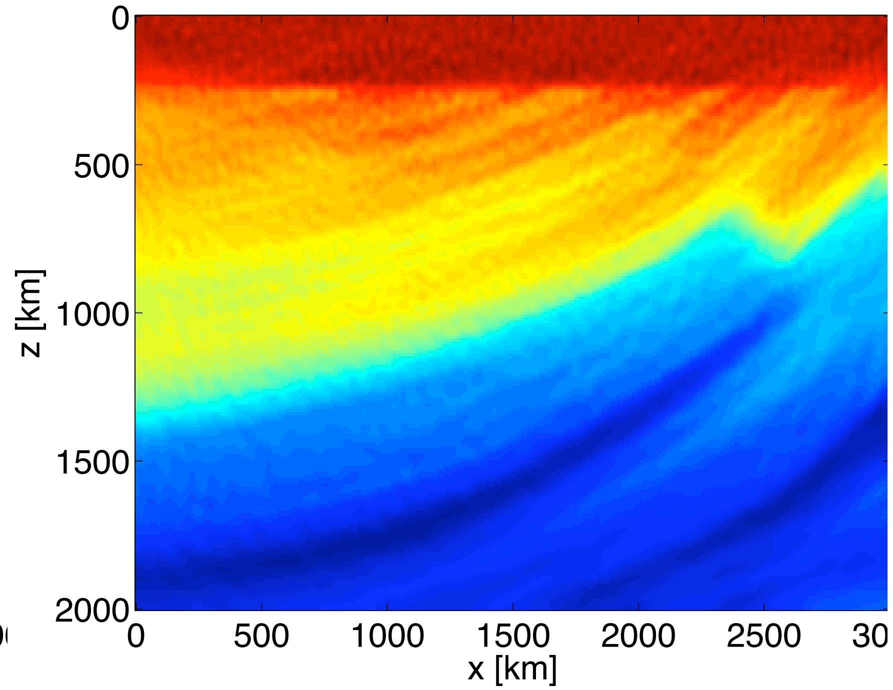
[Bertsekas, '96; Haber, Chung, and FJH, '10]

K=1 [noisy case]

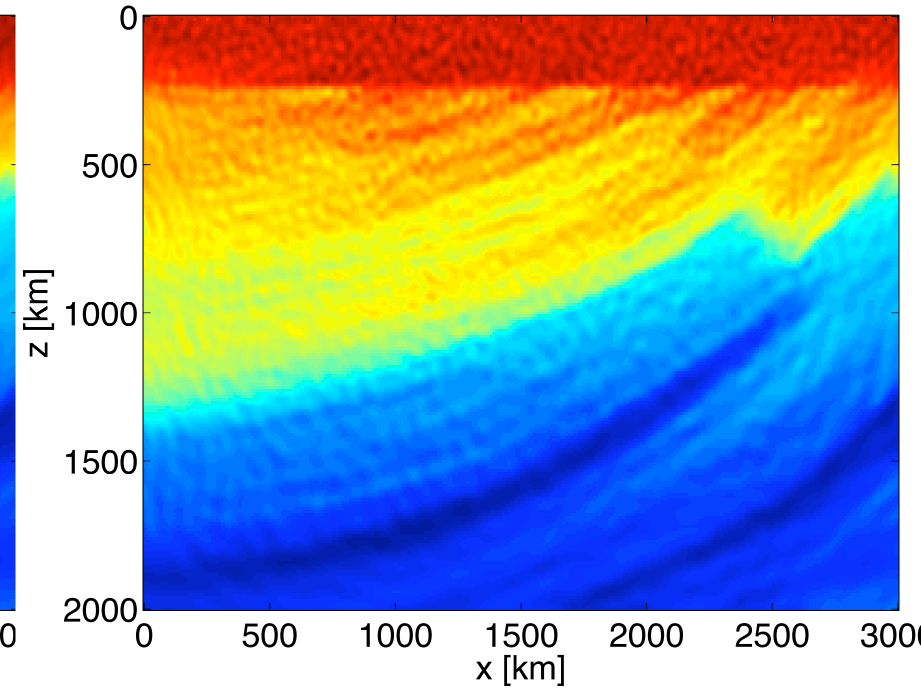
w/o averaging

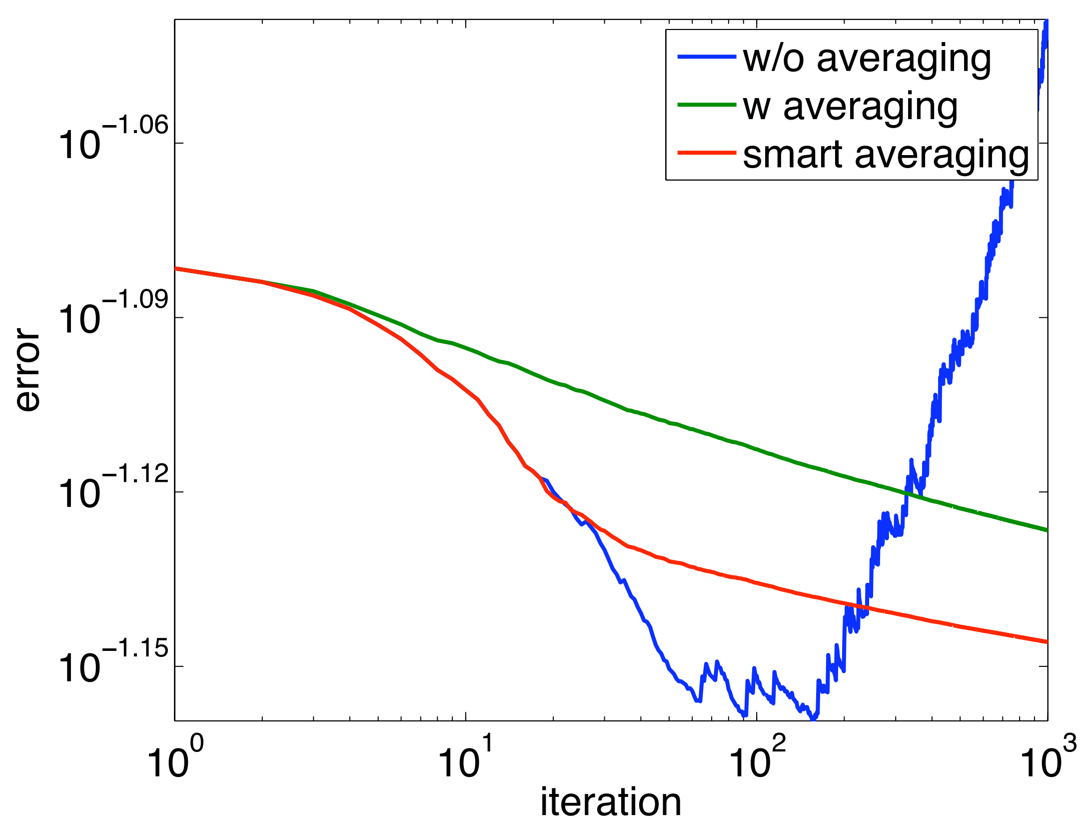


w averaging



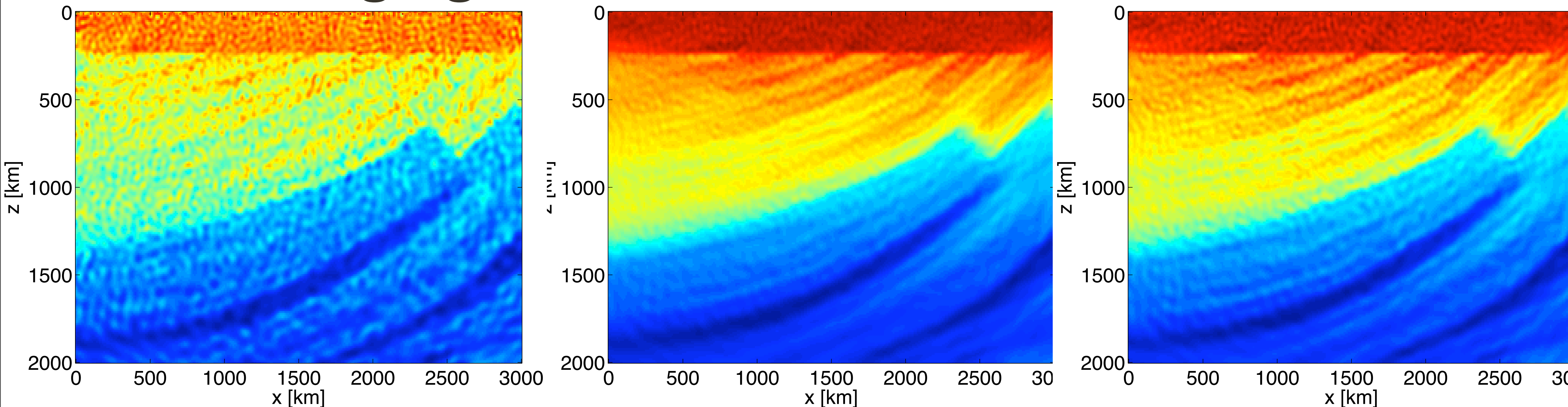
smart averaging

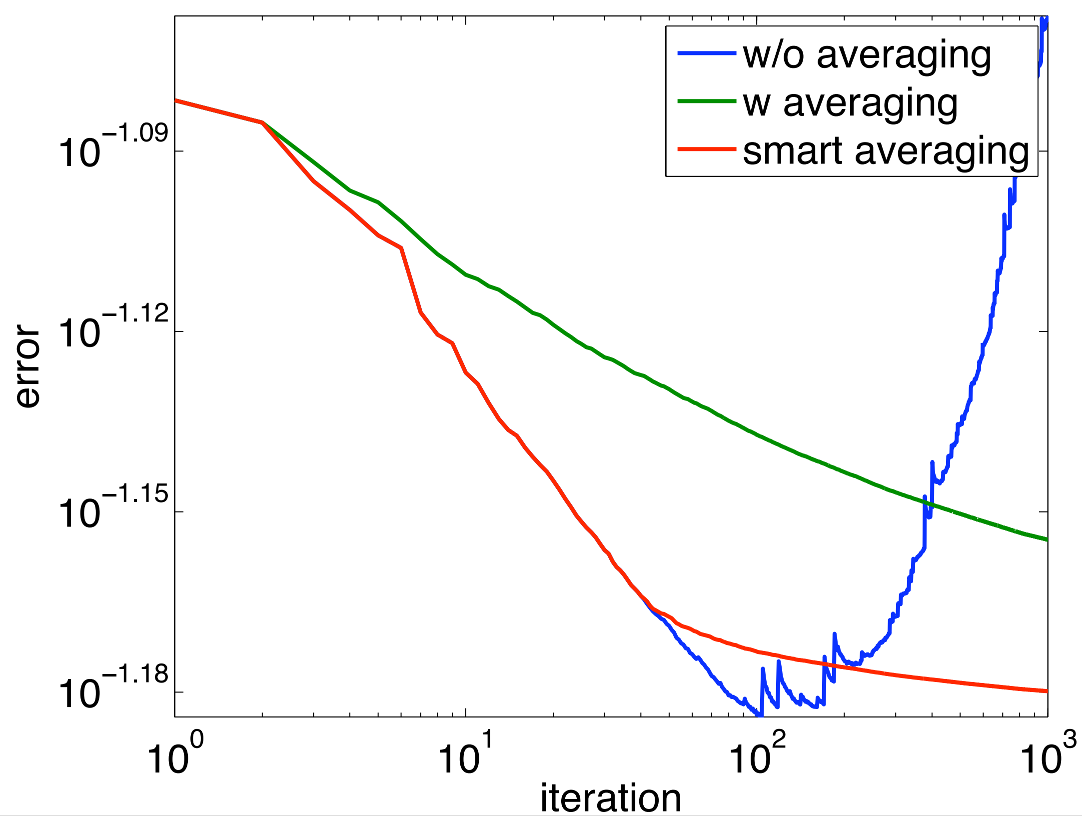




K=5 [noisy case]

w/o averaging | w averaging | smart averaging





SLIM

Observations

Renewals improve convergence *significantly*

Averaging removes noise *instability* but is *detrimental* to the *convergence*

Smart averaging over *limited history* improves *convergence*

Increasing the *batch size* in *combination* with *smart averaging* leads to *superior convergence*

Second-order methods *ad hoc* & not well understood

Produces *noisy* updates ... Sounds *familiar?*

Our approach

Leverage findings from *sparse recovery* & *compressive sensing*

- consider each *phase-encoded* Gauss-Newton update as separate *compressive-sensing* experiment
- remove *interferences* by *curvelet-domain sparsity* promotion
- exploit properties of Pareto curves

[Candes et al., ’06; Donoho, ’06]

[Demanet et. al. ’07; Herrmann & Li, ’08-’09]

Rationale

Wavefields are *compressible* in curvelet *frames*

- *correlations* between *source* & *residual wavefields* are *compressible*
- *velocity* distributions of *sedimentary basins* are *also compressible*

Linearized subproblems are *convex*

Assume proximity Pareto *curves* for *successive linearizations*

Gauss-Newton

Algorithm 1: Gauss Newton

Result: Output estimate for the model $\mathbf{m}$

$\mathbf{m} \longleftarrow \mathbf{m}_0$; $k \longleftarrow 0$; `// initial model`

while not converged **do**

$\quad \mathbf{p}^k \longleftarrow \arg\min_{\mathbf{p}} \frac{1}{2}\|\delta\mathbf{d} - \nabla\mathcal{F}[\mathbf{m}^k;\mathbf{Q}]\mathbf{p}\|_2^2 + \lambda^k\|\mathbf{p}\|_2^2$; `// search dir.`

$\quad \mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k\mathbf{p}^k$; `// update with linesearch`

$\quad k \longleftarrow k+1$;

end

Phase encoding

Algorithm 1: Gauss Newton with renewed phase encodings

Result: Output estimate for the model $\mathbf{m}$

$\mathbf{m} \longleftarrow \mathbf{m}_0$; $k \longleftarrow 0$; `// initial model`

while not converged **do**

$\mathbf{p}^k \longleftarrow \arg\min_{\mathbf{p}} \frac{1}{2}\|\delta\underline{\mathbf{d}}^k - \nabla\mathcal{F}[\mathbf{m}^k;\underline{\mathbf{Q}}^k]\mathbf{p}\|_2^2 + \lambda^k\|\mathbf{p}\|_2^2$; `// search dir.`

$\mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k\mathbf{p}^k$; `// update with linesearch`

$k \longleftarrow k+1$;

end

[Wang & Sacchi, '07]

Sparse recovery

Least-squares migration with *sparsity* promotion

$$\delta\widetilde{\mathbf{m}} = \mathbf{S}^* \arg\min_{\delta\mathbf{x}} \frac{1}{2}\|\delta\mathbf{x}\|_{\ell_1} \quad \text{subject to} \quad \|\delta\underline{\mathbf{d}} - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}_0;\underline{\mathbf{Q}}]\mathbf{S}^*\delta\mathbf{x}\|_2 \leq \sigma$$

$$\begin{aligned} \delta\mathbf{x} &= \text{Sparse curvelet-coefficient vector} \\ \mathbf{S}^* &= \text{Curvelet synthesis} \end{aligned}$$

leads to *significant* speedup as long as

$$n^{\ell_1}_{PDE} \times K \ll n^{\ell_2}_{PDE} \times n_f \times n_s$$

Compressive updates

Algorithm 1: Gauss Newton with sparse updates

Result: Output estimate for the model $\mathbf{m}$

$\mathbf{m} \longleftarrow \mathbf{m}_0$; $k \longleftarrow 0$; `// initial model`

while not converged **do**

$\quad \mathbf{p}^k \longleftarrow \mathbf{S}^* \arg\min_{\mathbf{x}} \frac{1}{2}\|\delta\underline{\mathbf{d}}^k - \nabla\mathcal{F}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\mathbf{S}^*\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau^k$

$\quad \mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k \mathbf{p}^k$; `// update with linesearch`

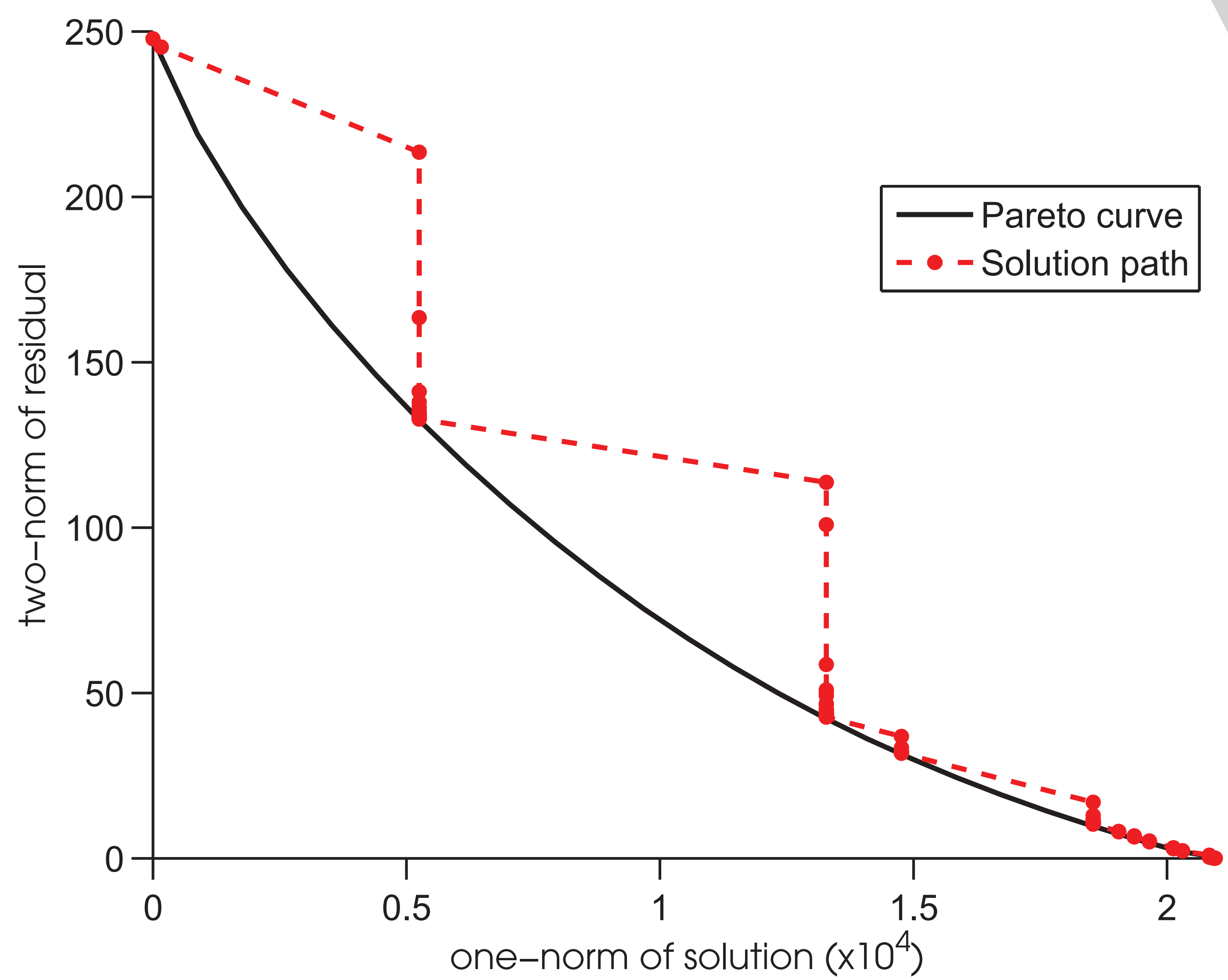
$\quad k \longleftarrow k + 1$;

end

[van den Berg & Friedlander, '08]

Solution strategy

- Draw *new* CS *experiment* when *Pareto curve* is reached
- Do *new linearization*
- *Sweep* from low to hight frequencies



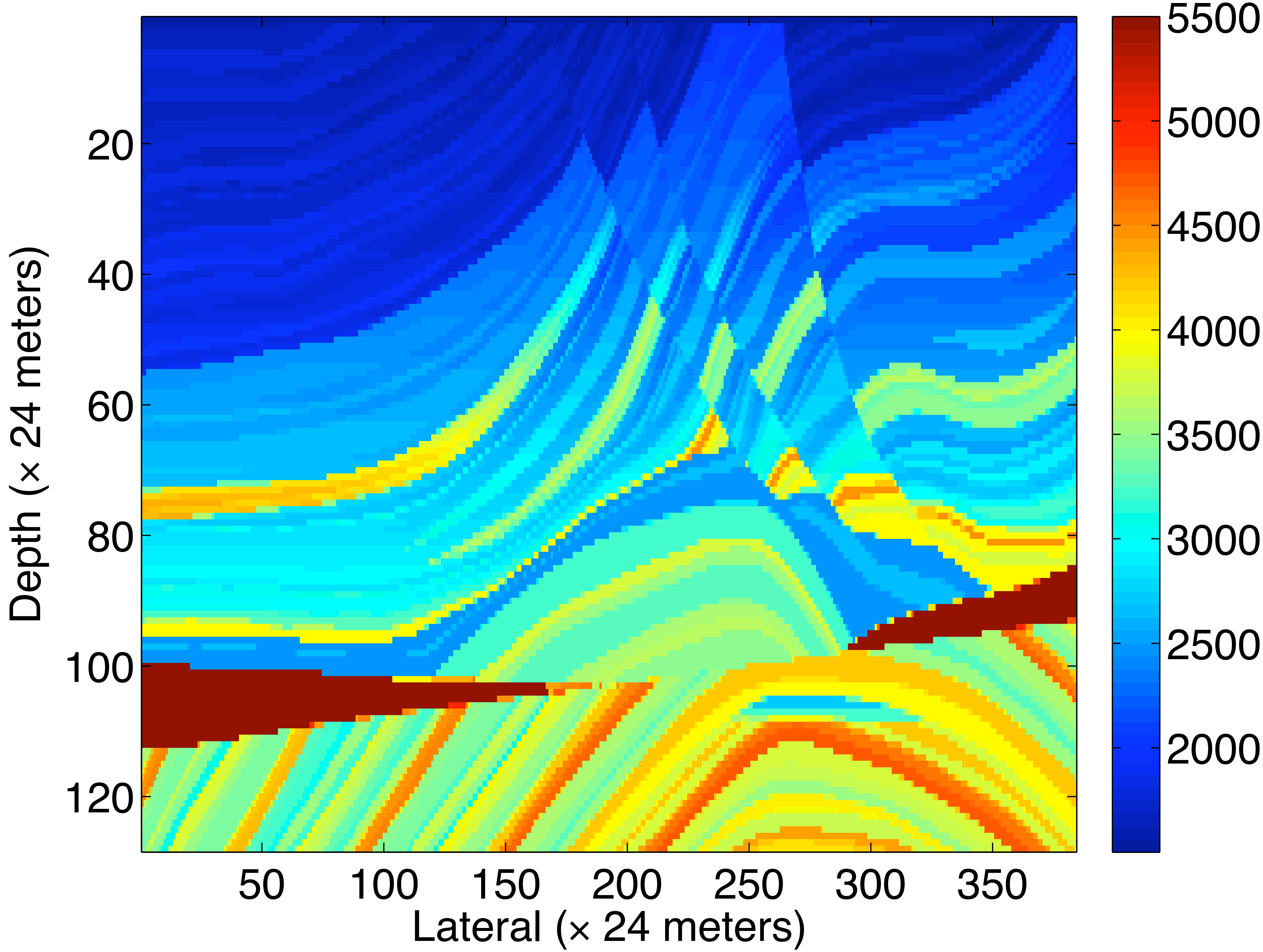
Example

FWI specs:

- Committed *inversion crime*
- Frequency continuation over 10 bands
- 15 *simultaneous* shots with 10 *frequencies* each

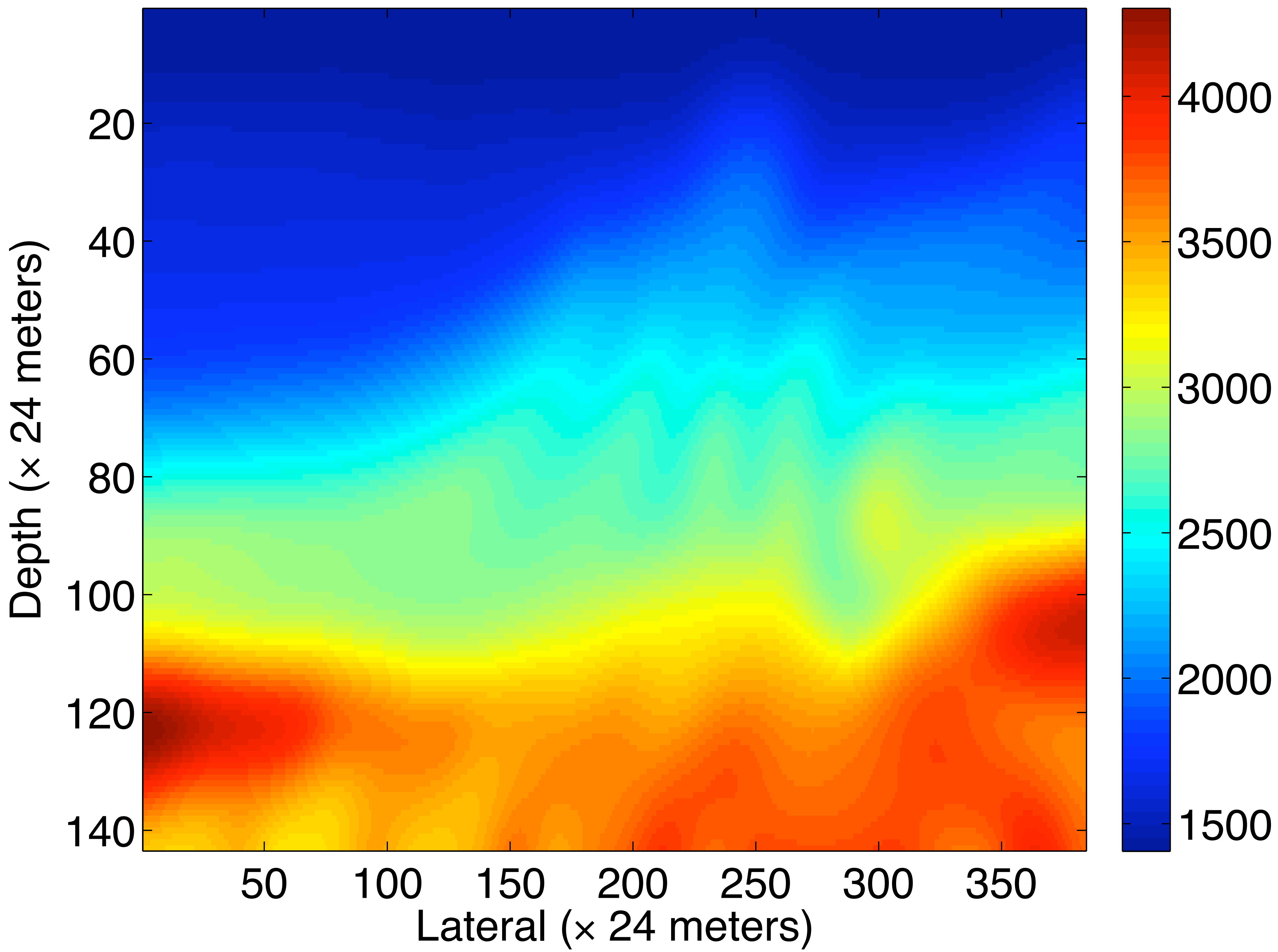
$$K = 10 \times 15 \ll 100 \times 384$$

True model

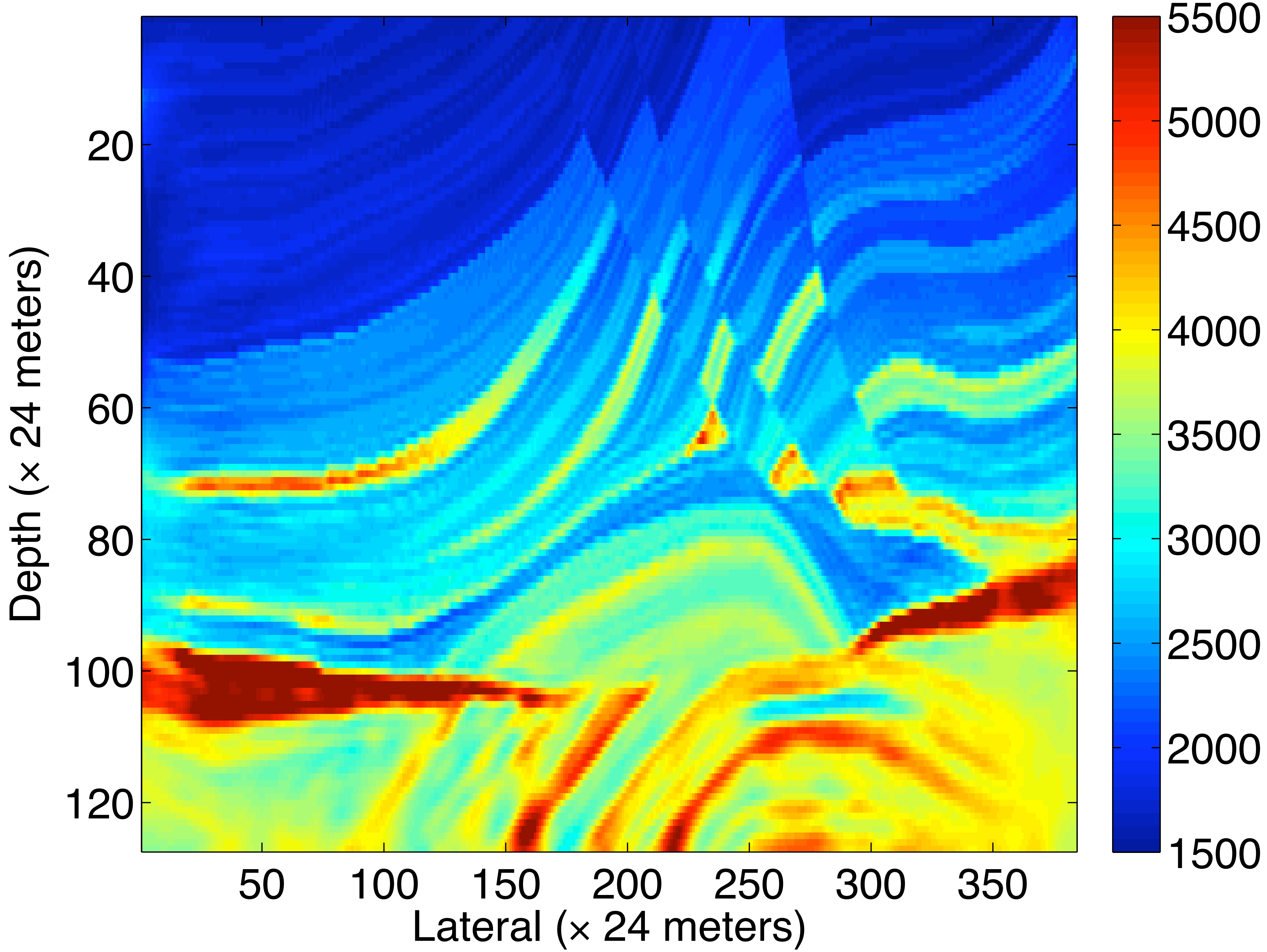


Depth (× 24 meters)
20
40
60
80
100
120
50
100
150
200
250
300
350
Lateral (× 24 meters)
5500
5000
4500
4000
3500
3000
2500
2000

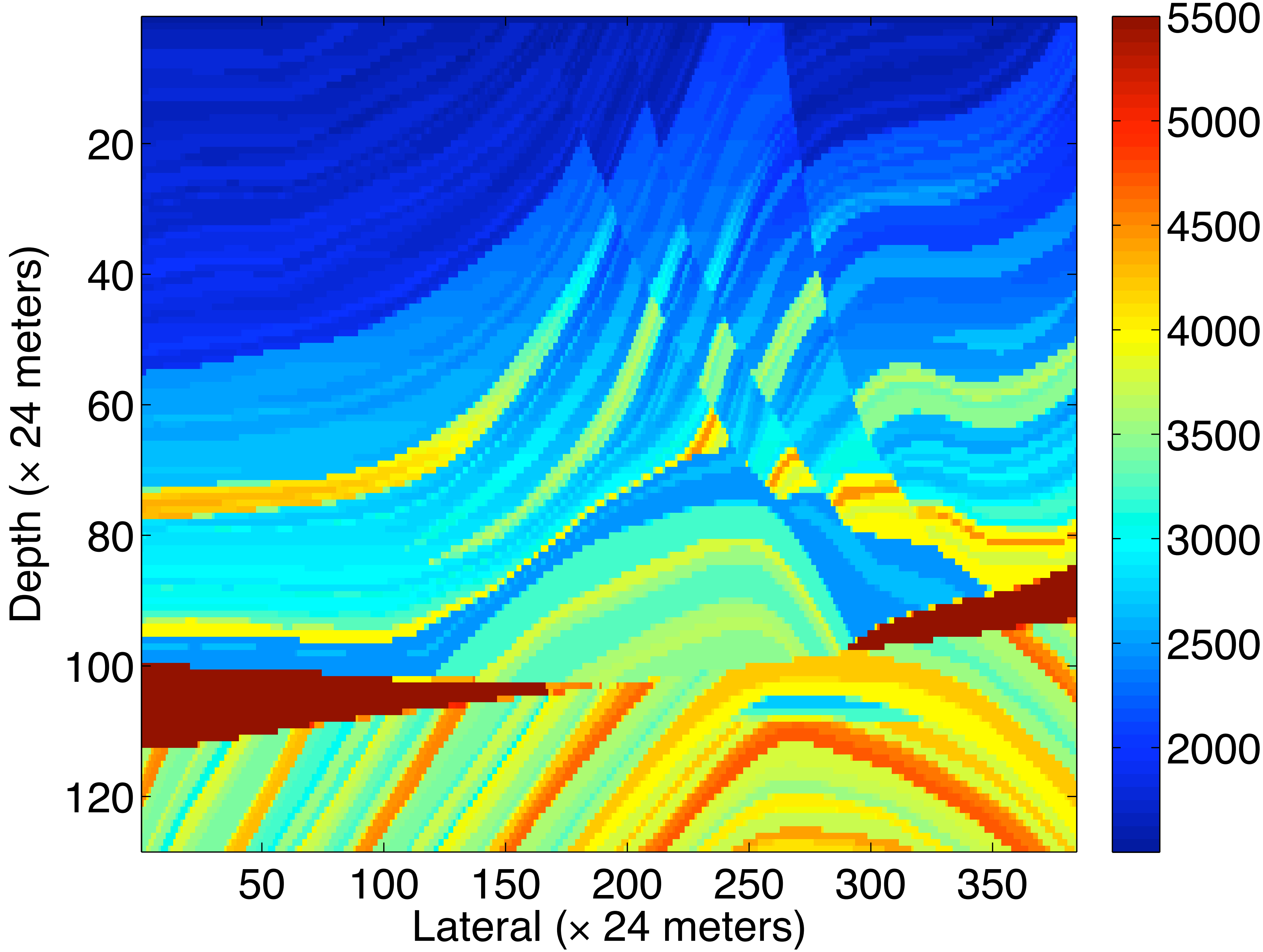
Initial model



Inverted model

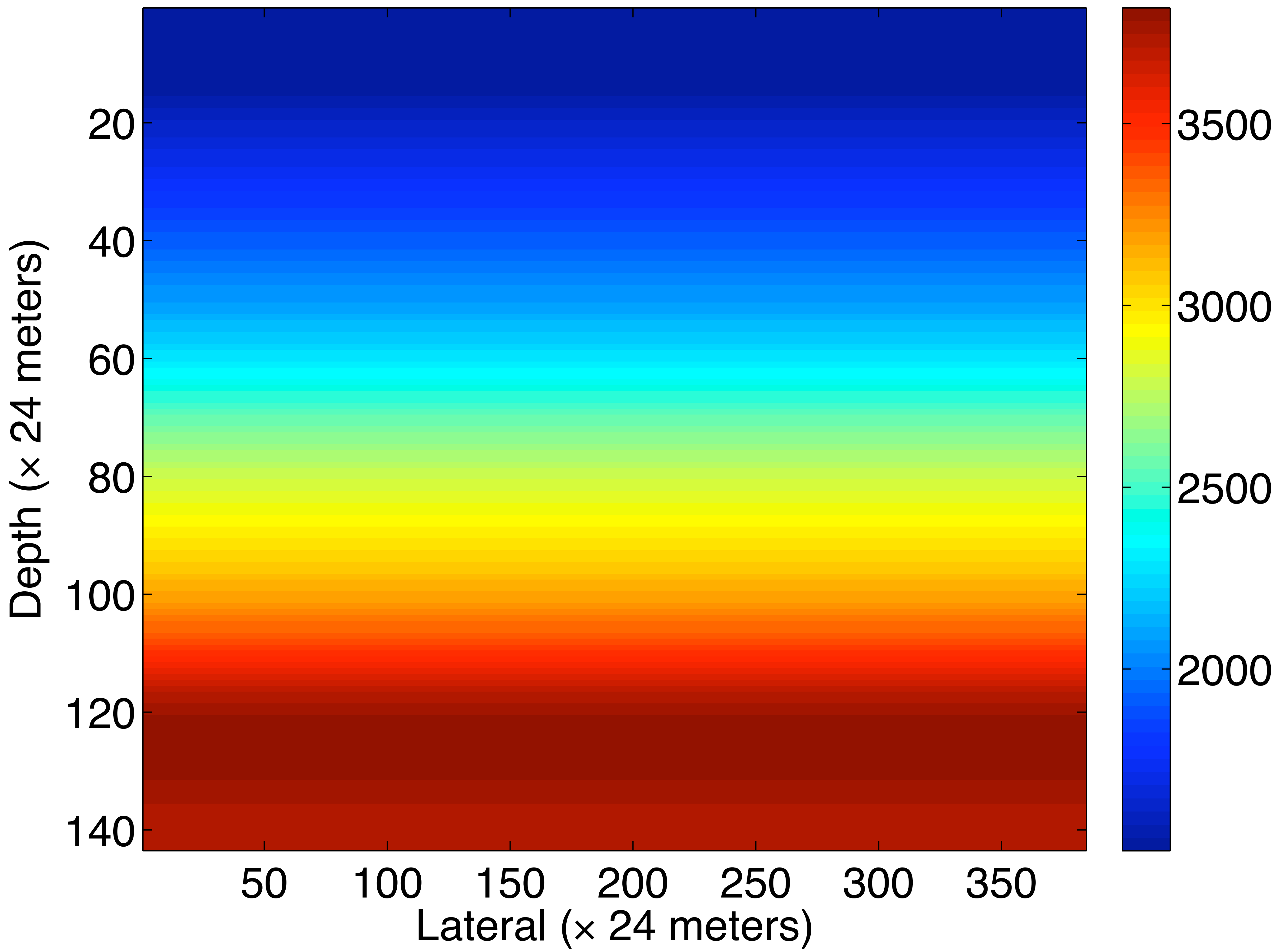


True model

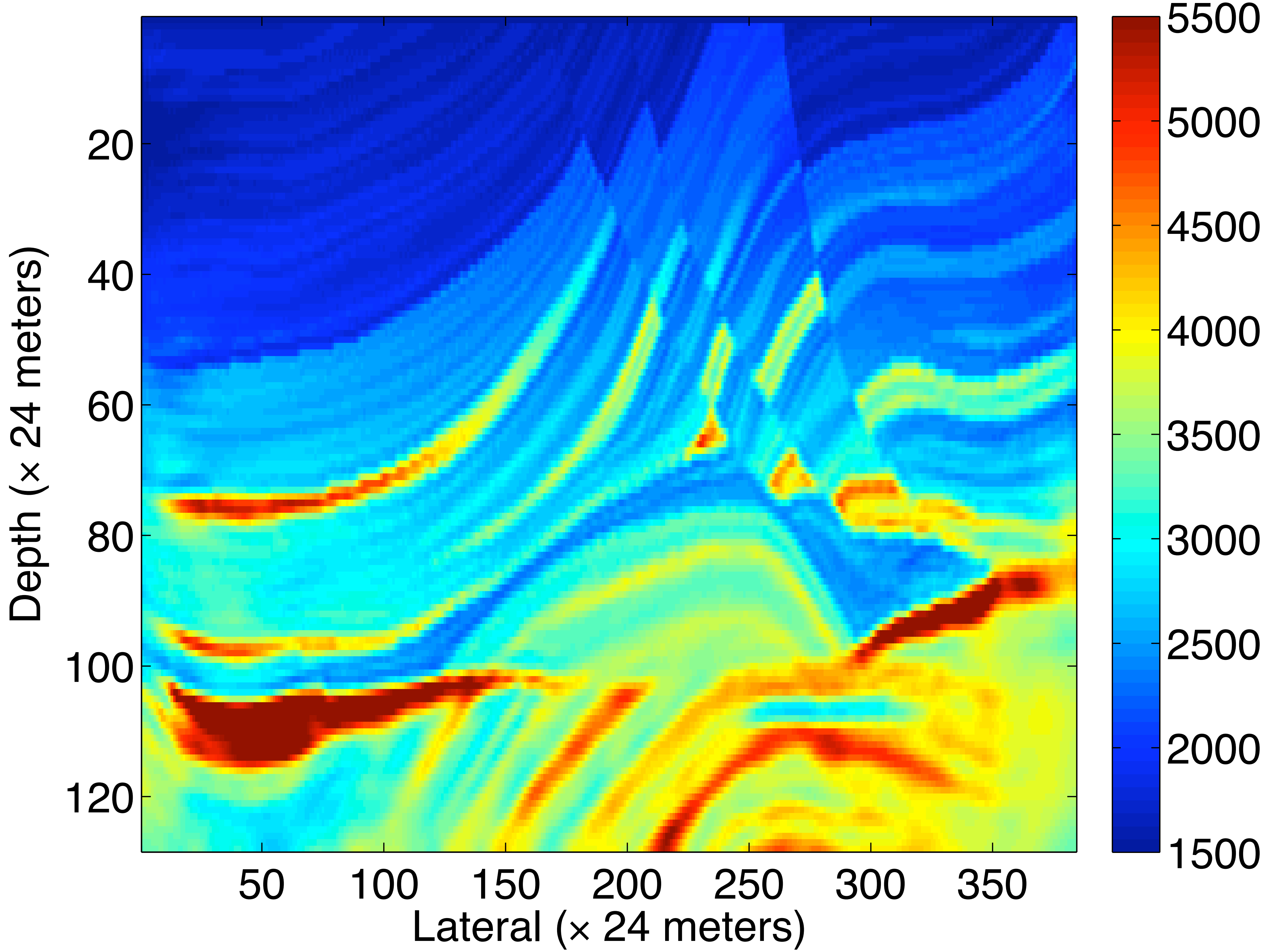


Depth (× 24 meters)
20
40
60
80
100
120
50
100
150
200
250
300
350
Lateral (× 24 meters)
5500
5000
4500
4000
3500
3000
2500
2000

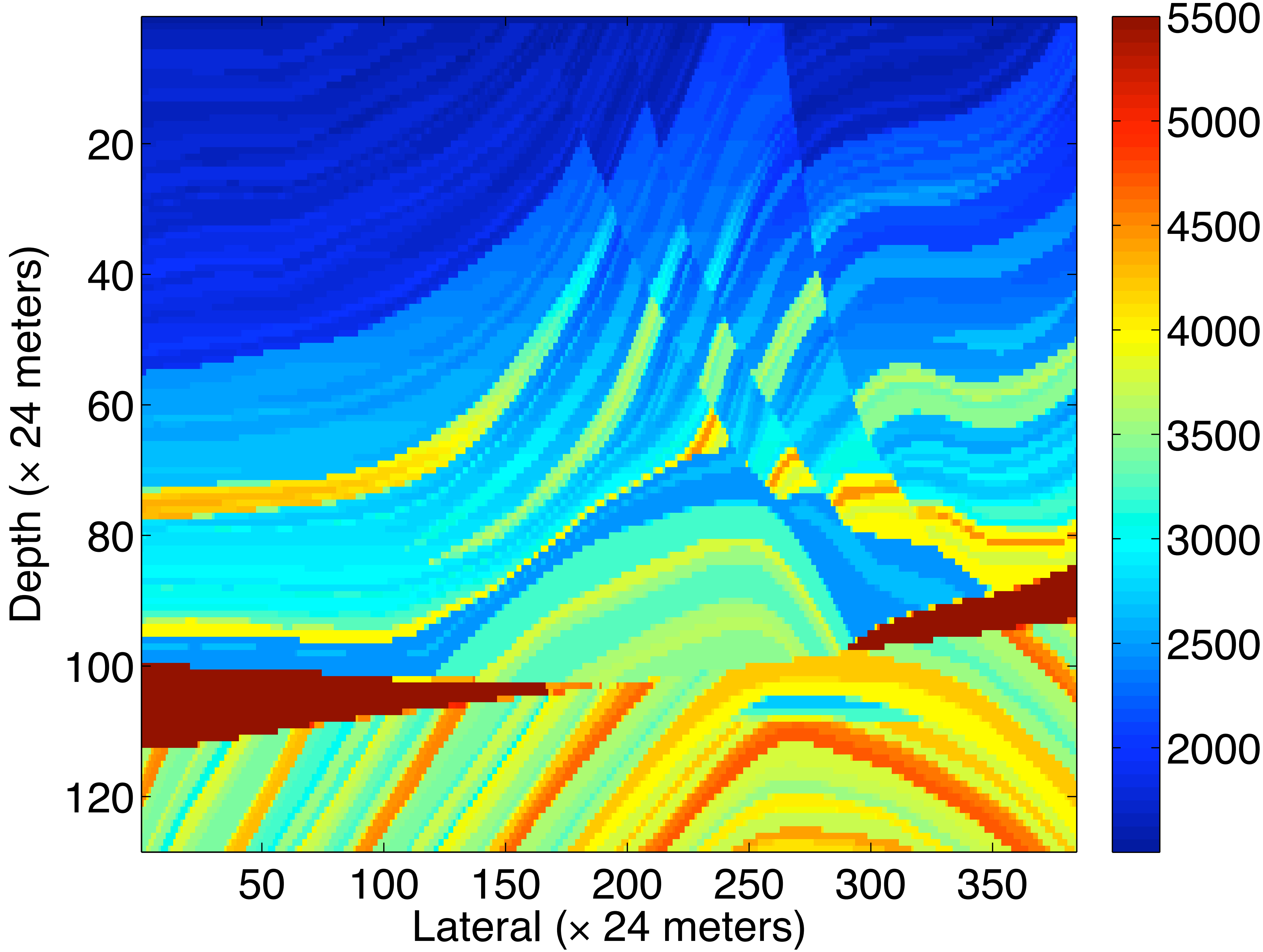
Initial model



Inverted model



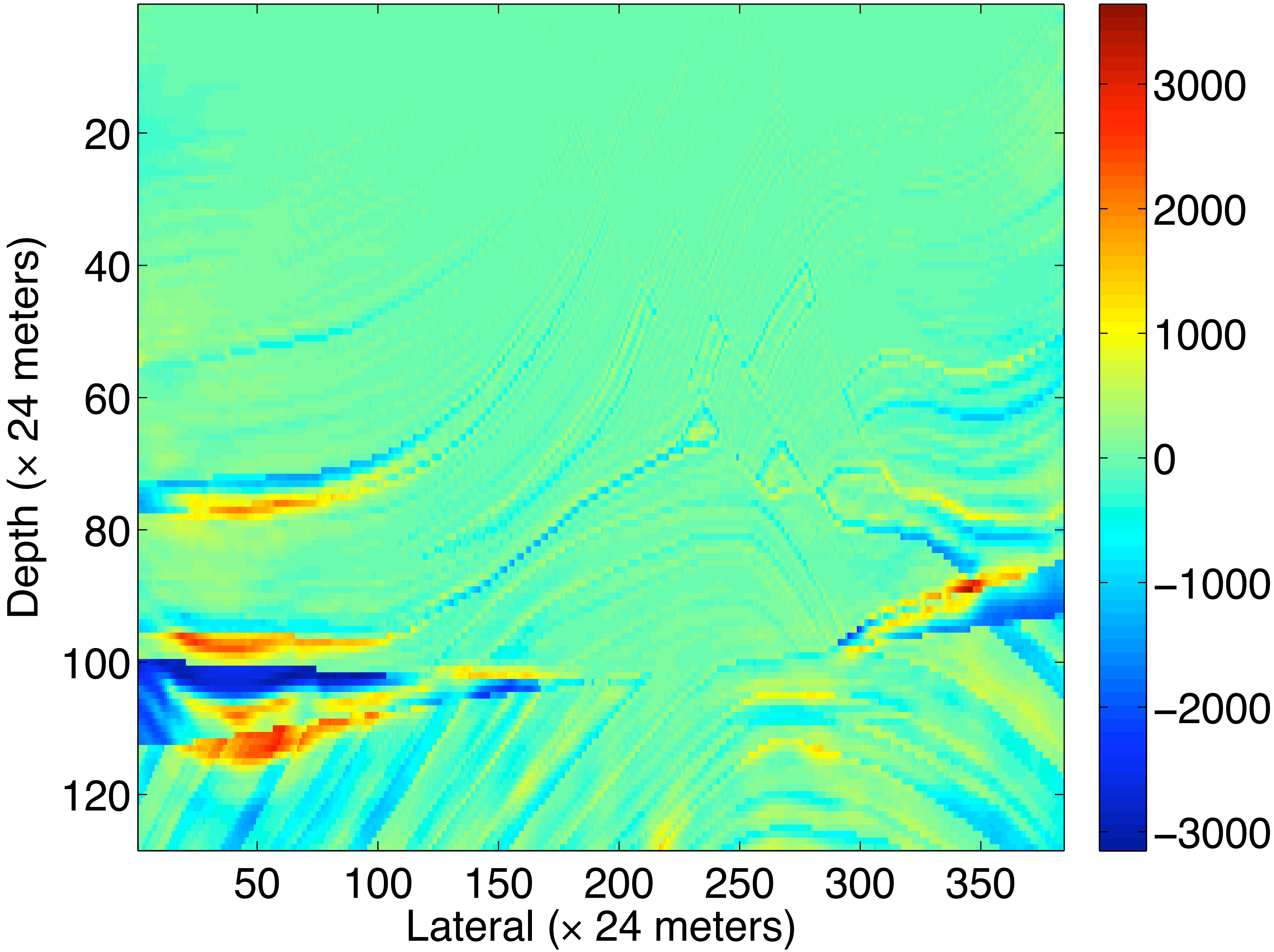
True model



Depth (× 24 meters)
20
40
60
80
100
120
50
100
150
200
250
300
350
Lateral (× 24 meters)
5500
5000
4500
4000
3500
3000
2500
2000

SLIM

Difference



Performance

Remember per *subproblem*

$$n_{PDE}^{\ell_1} \times K \ll n_{PDE}^{\ell_2} \times n_f \times n_s$$

$$n_{PDE}^{\ell_1} \approx 200 \qquad K = 150$$

versus

$$n_{PDE}^{\ell_2} \approx 10 \qquad K = 38400$$

SPEEDUP of 13 X

SLIM

Conclusions

Dimensionality reduction will *revolutionize* our field

- *reduction* of *acquisition* costs
- *less reliance* on *full sampling*
- *decrease* in *processing* time
- *high-resolution* inversions that are otherwise *infeasible* with fully-sample (Nyquist-based) methods

Non uniqueness & missing low frequencies remain *fundamental* problems...

SLIM

Thank you

slim.eos.ubc.ca

Further reading

Compressive sensing

- *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information by Candes, 06*.
- *Compressed Sensing* by D. Donoho, '06

Simultaneous acquisition

- *A new look at simultaneous sources* by Beasley et. al., '98.
- *Changing the mindset in seismic data acquisition* by Berkhout '08.

Simultaneous simulations, imaging, and full-wave inversion:

- *Faster shot-record depth migrations using phase encoding* by Morton & Ober, '98.
- *Phase encoding of shot records in prestack migration* by Romero et. al., '00.
- *High-resolution wave-equation amplitude-variation-with-ray-parameter (AVP) imaging with sparseness constraints* by Wang & Sacchi, '07
- Efficient Seismic Forward Modeling using Simultaneous Random Sources and Sparsity by N. Neelamani et. al., '08.
- *Compressive simultaneous full-waveform simulation* by FJH et. al., '09.
- *Fast full-wavefield seismic inversion using encoded sources* by Krebs et. al., '09
- *Randomized dimensionality reduction for full-waveform inversion* by FJH & X. Li, '10

Stochastic optimization and machine learning:

- *A Stochastic Approximation Method* by Robbins and Monro, 1951
- *Neuro-Dynamic Programming* by Bertsekas, '96
- *Robust stochastic approximation approach to stochastic programming* by Nemirovski et. al., '09
- *Stochastic Approximation approach to Stochastic Programming* by Nemirovski
- *Randomized dimensionality reduction for full-waveform inversion* by FJH & X. Li, '10
- *An effective method for parameter estimation with PDE constraints with multiple right hand sides. by* Eldad Haber, Matthias Chung, and Felix J. Herrmann. '10

Acknowledgments

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).

We also would like to thank the authors of CurveLab.

This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.

NSERC
CRSNG