Stable seismic data recovery

Felix J. Herrmann*

joint work with

Peyman Moghaddam*, Gilles Hennenfent* & Chris Stolk (Universiteit Twente)

*Seismic Laboratory for Imaging and Modeling

slim.eos.ubc.ca

AIP 2007, Vancouver, June 26

SLIM Seismic Laboratory for Imaging and Modeling

Combinations of ***parsimonious*** *signal representations with nonlinear* ***sparsity*** *promoting programs hold the* ***key*** *to the next-generation of seismic inversion algorithms ...*

Since they allow for formulations that are ***stable*** *w.r.t.*

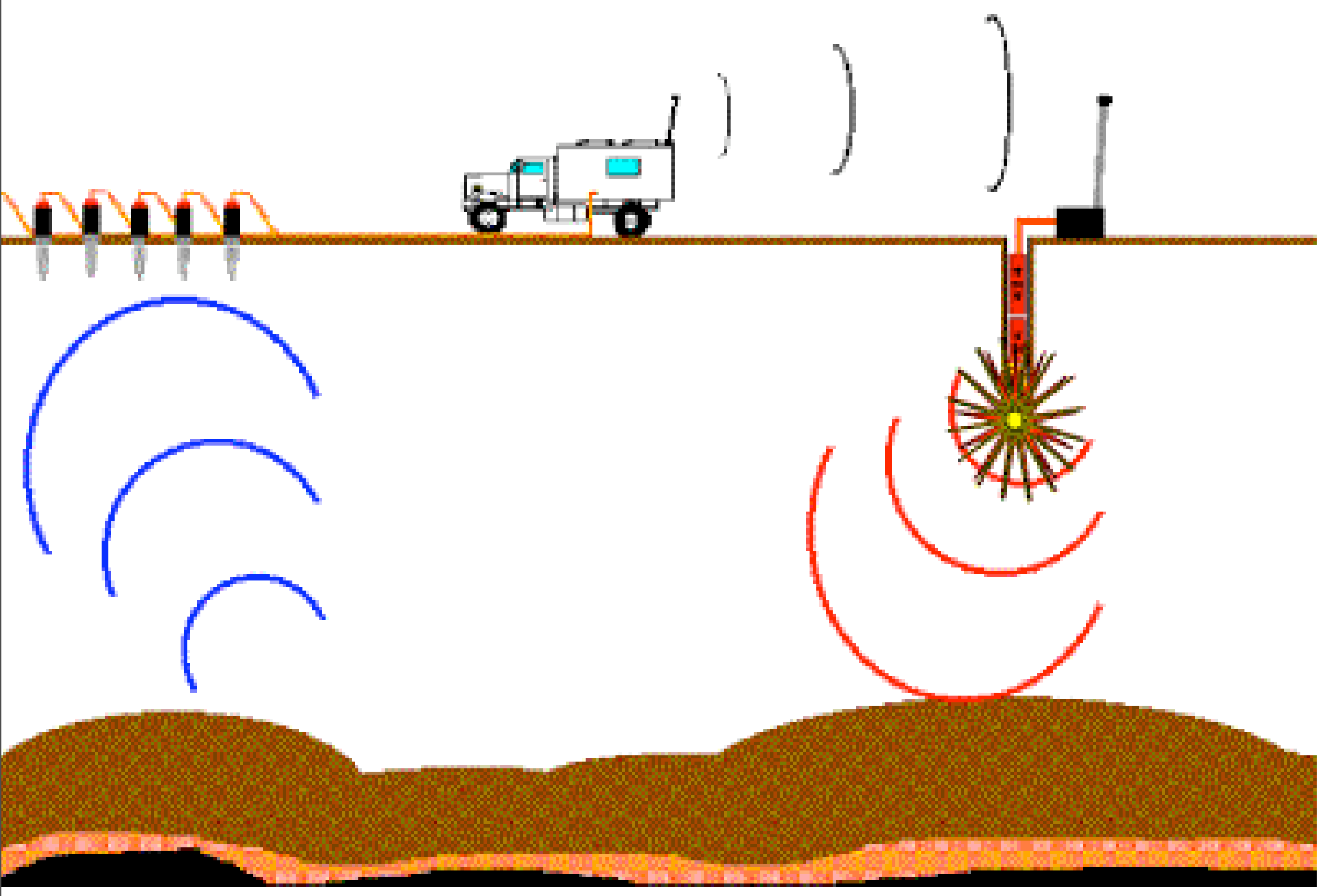
- *noise*
- *incomplete data*
- *moderate phase rotations and amplitude errors*

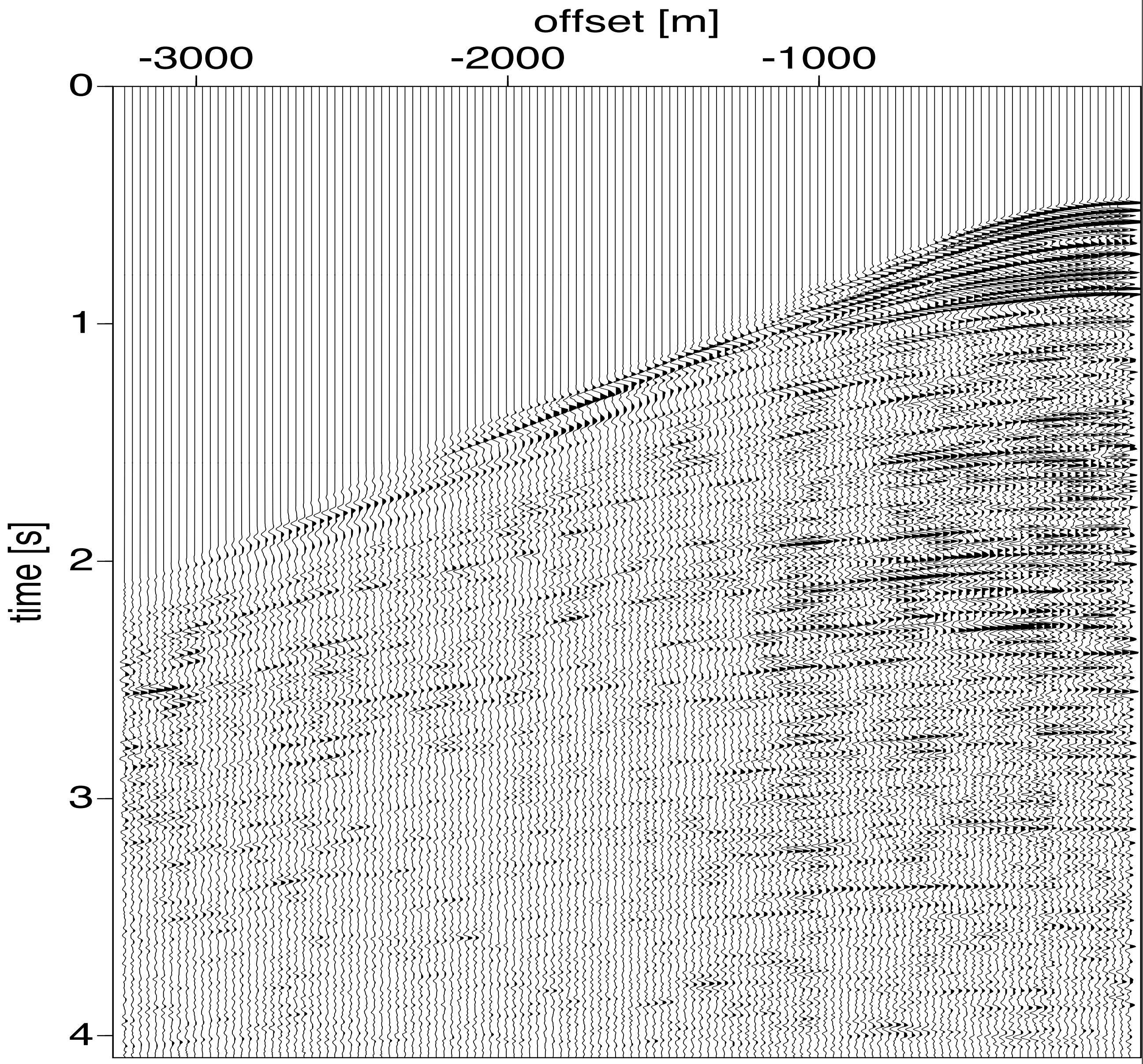
Finding a **sparse** representation for seismic data & images is complicated because of

- wavefronts & reflectors are multiscale & multi-directional
- the presence of caustics, faults and pinchouts
- the presence of operators (FIO’s & PsDO’s)

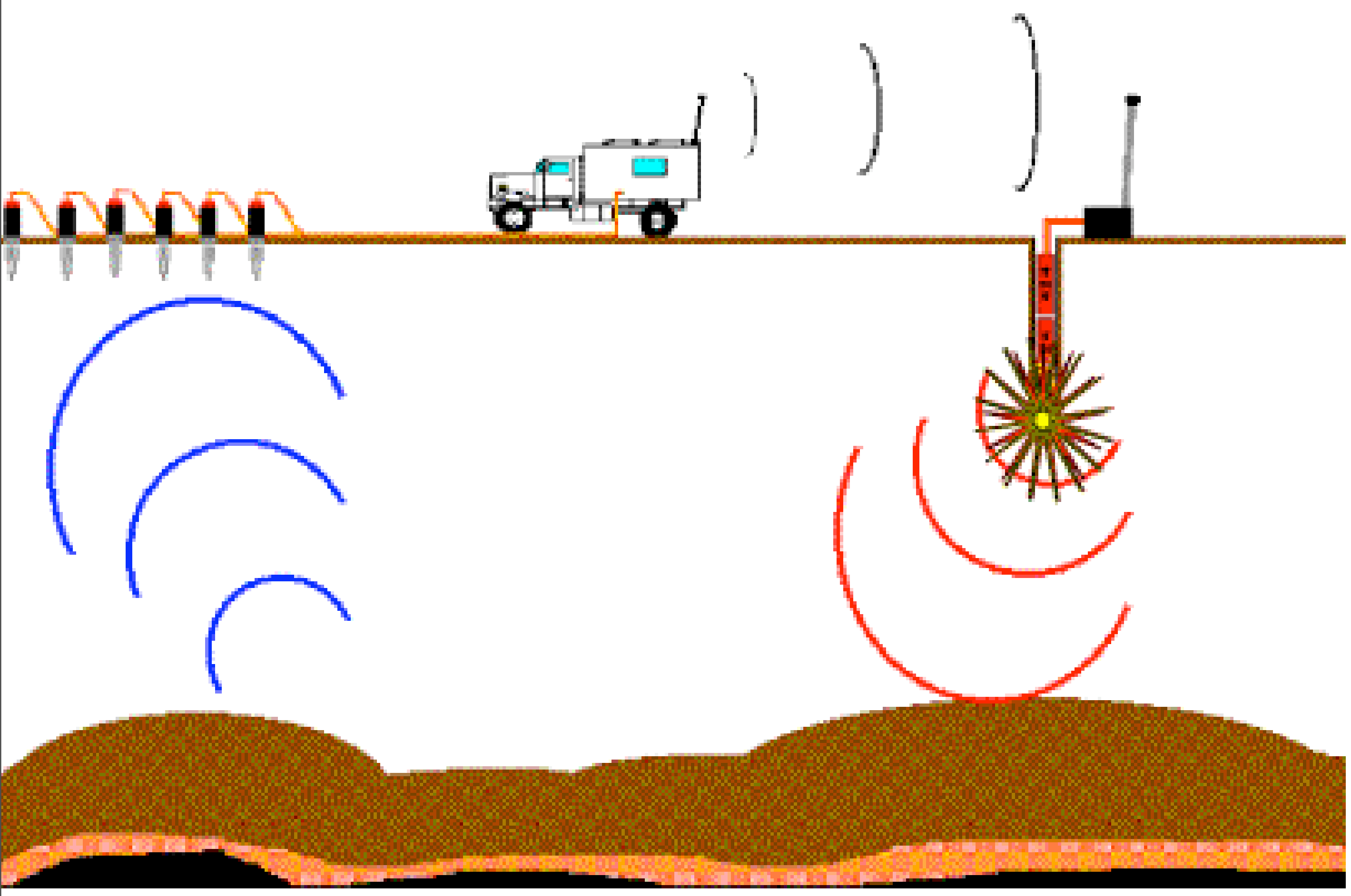
The seismic method

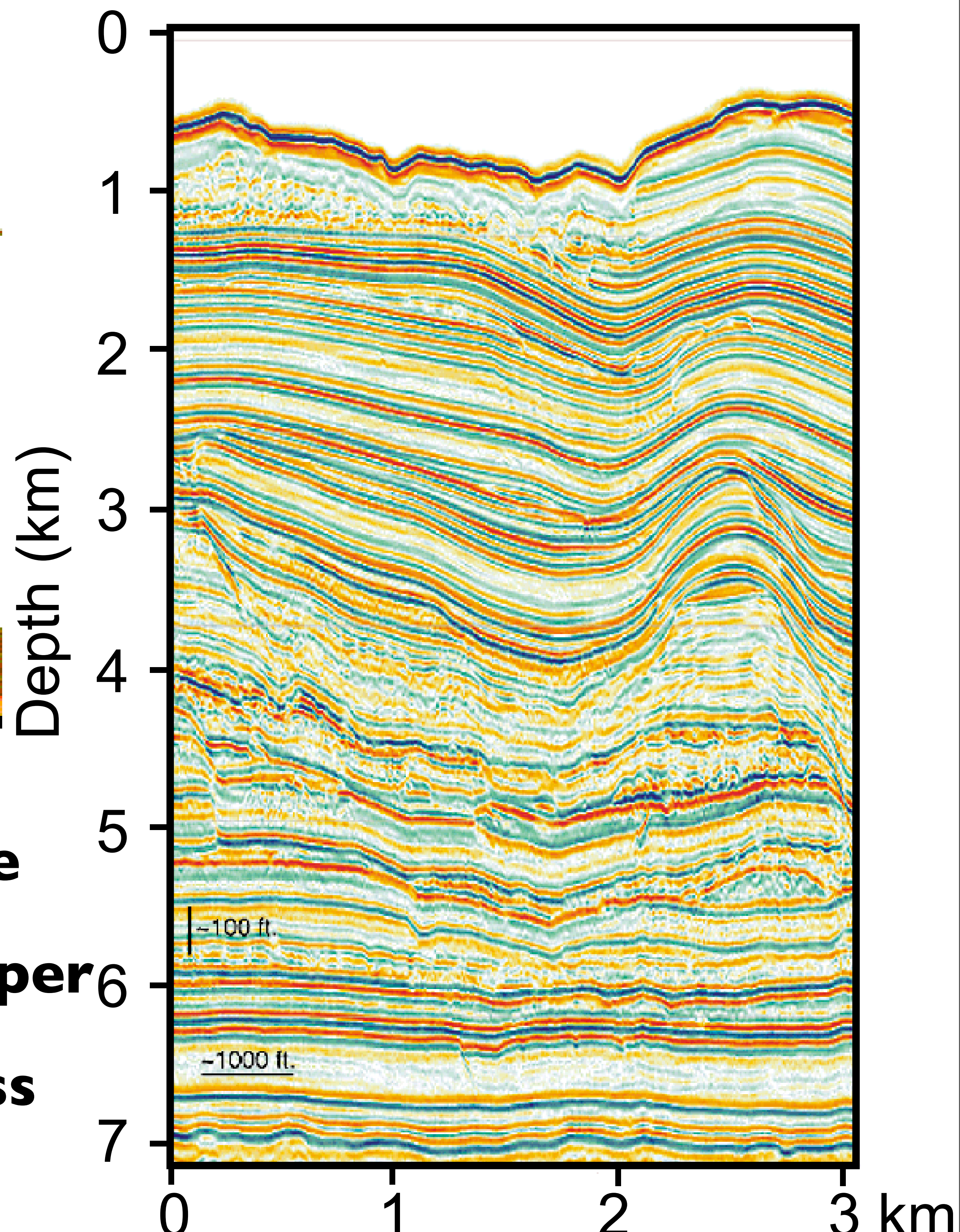
Seismic data acquisition





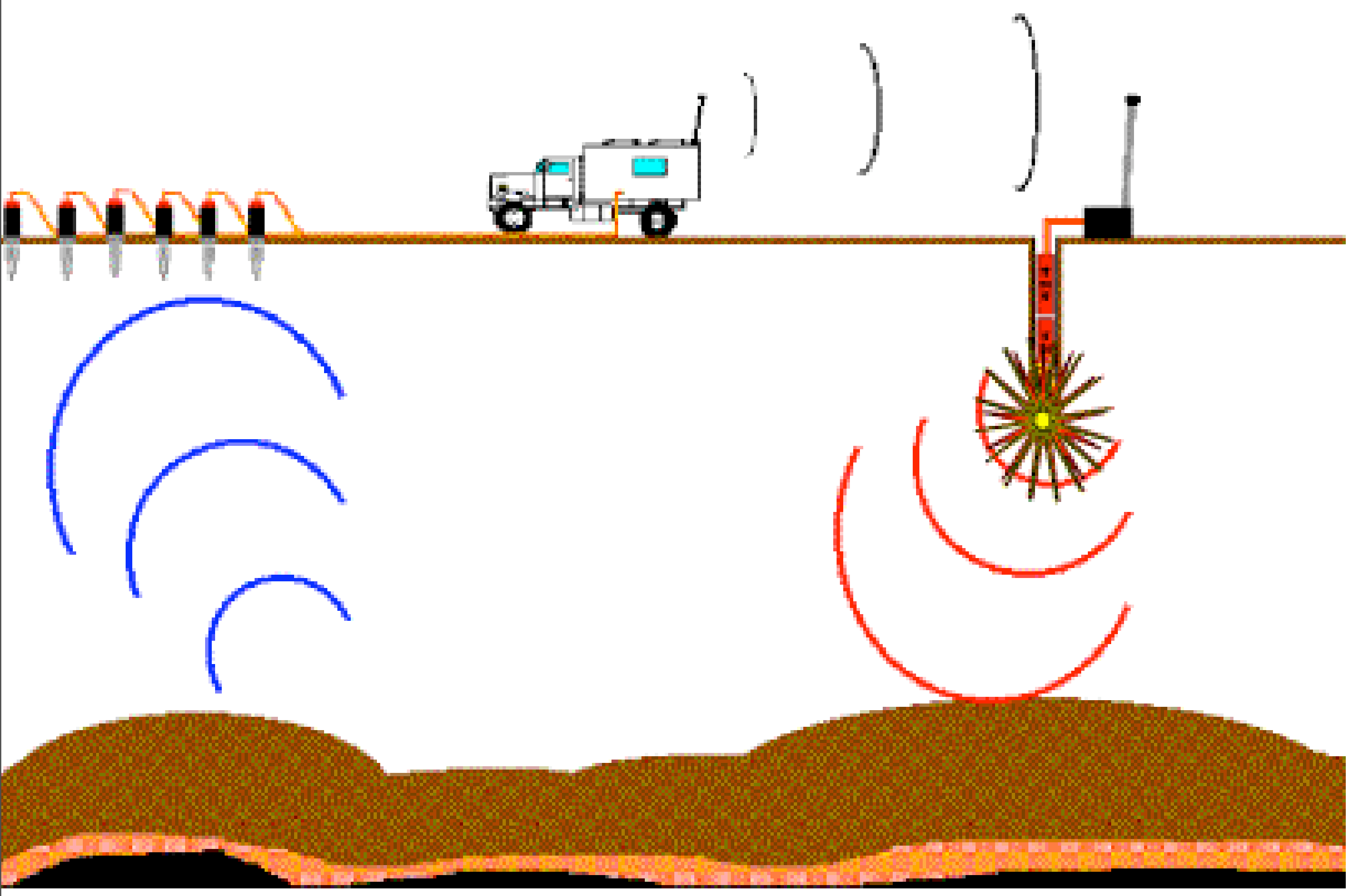
Exploration seismology

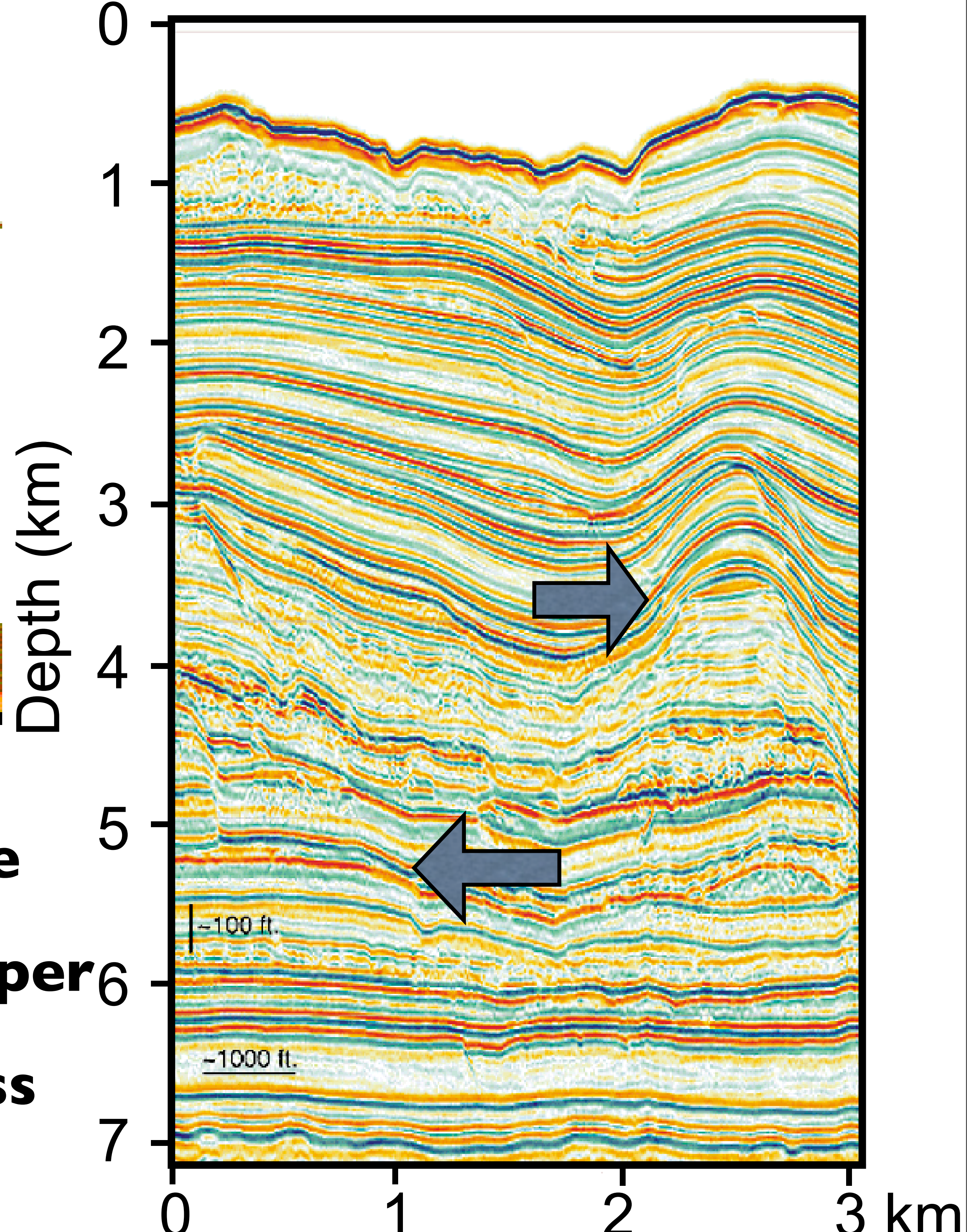




- **create images of the subsurface**
- **need for higher resolution/deeper**
- **clutter and data incompleteness are problems**

Exploration seismology





- **create images of the subsurface**
- **need for higher resolution/deeper**
- **clutter and data incompleteness are problems**

Forward problem

$$F[c]u := \left(\frac{1}{c^2(x)} \cdot \frac{\partial^2}{\partial t^2} - \sum_{i=1}^{d} \frac{\partial^2}{\partial x_1^2} \right) \mathbf{u}(x,t) = f(x,t)$$

- second order hyperbolic PDE
- interested in the singularities of

$$m = c - \bar{c}$$

Inverse problem

Minimization:

$$\widetilde{m} = \arg\min_{m} \|d - F[m]\|_2^2$$

After linearization (Born app.) forward model with noise:

$$d(x_s, x_r, t) = \big(Km\big)(x_s, x_r, t) + n(x_s, x_r, t)$$

Conventional imaging:

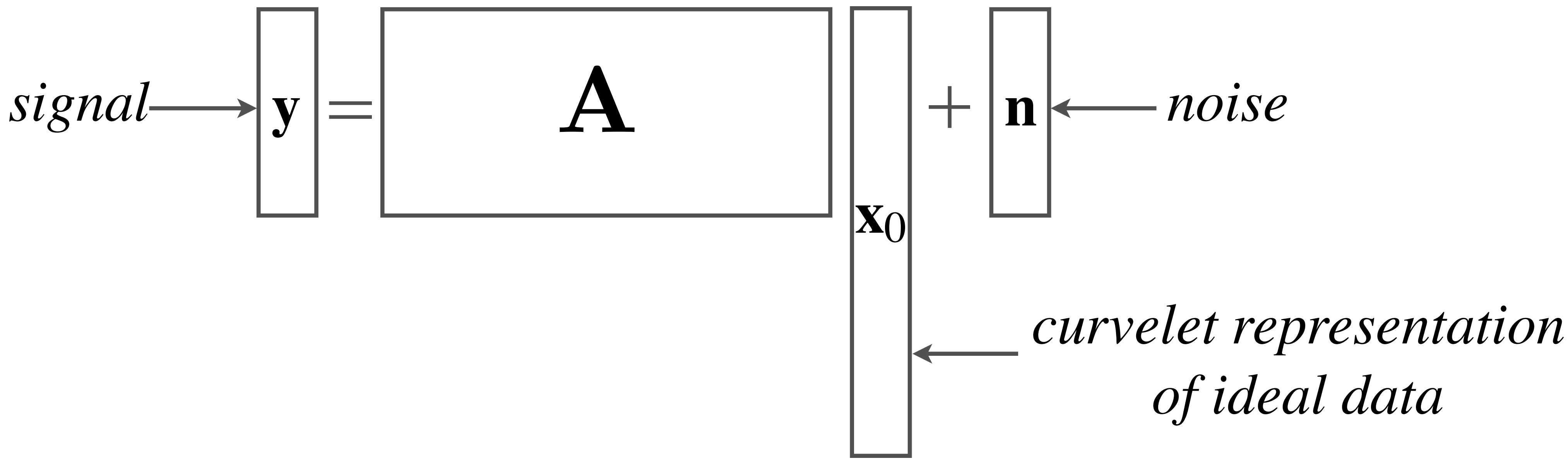
$$\begin{aligned}\big(K^T d\big)(x) &= \big(K^T K m\big)(x) + \big(K^T n\big)(x)\\ y(x) &= \big(\mathbf{\Psi} m\big)(x) + e(x)\end{aligned}$$

Ψ **is prohibitively expensive to invert**

requires regular sampling ...

Sparsity promoting inversion

Formulate as inverse problem



$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon$$

sparsity enhancement (↑ $\|\mathbf{x}\|_1$) *data misfit* (↑ $\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2$)

When a traveler reaches a fork in the road, the l1 -norm tells him to take either one way or the other, but the l2 -norm instructs him to head off into the bushes.

John F. Claerbout and Francis Muir, 1973

New field “compressive sampling”: D. Donoho, E. Candes et. al., M. Elad etc.

Preceded by others in geophysics: M. Sacchi & T. Ulrych and co-workers etc.

Sparsity promoting inversion

$\mathbf{x}_0$ can be recovered by solving

$$\mathbf{P}_\epsilon : \quad \begin{cases} \tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \\ \tilde{\mathbf{f}} = \mathbf{S}^T \tilde{\mathbf{x}} \end{cases}$$

with

$$\begin{aligned} \mathbf{y} &= \text{(incomplete) data} \\ \mathbf{A} &= \text{modeling matrix, e.g. } \mathbf{A} = \mathbf{R}\mathbf{S}^T \\ \tilde{\mathbf{x}} &= \text{recovered sparsity vector} \\ \epsilon &= \text{a number dependent on the noise level} \\ \mathbf{S}^T &= \text{the synthesis matrix} \\ \tilde{\mathbf{f}} &= \text{the recovered function } \mathbf{f} \end{aligned}$$

Crux lies in finding the sparse representation!

SLIM
Seismic Laboratory for Imaging and Modeling

Curvelets & seismology

Wish list

Transform that is parsimonious

- detects the wavefronts
- localized in space and frequency (phase space)
- some invariance under "wave propagation"

Events correspond to curved singularities with conflicting dips

- caustics
- faults & pinch outs

Need a transform that is

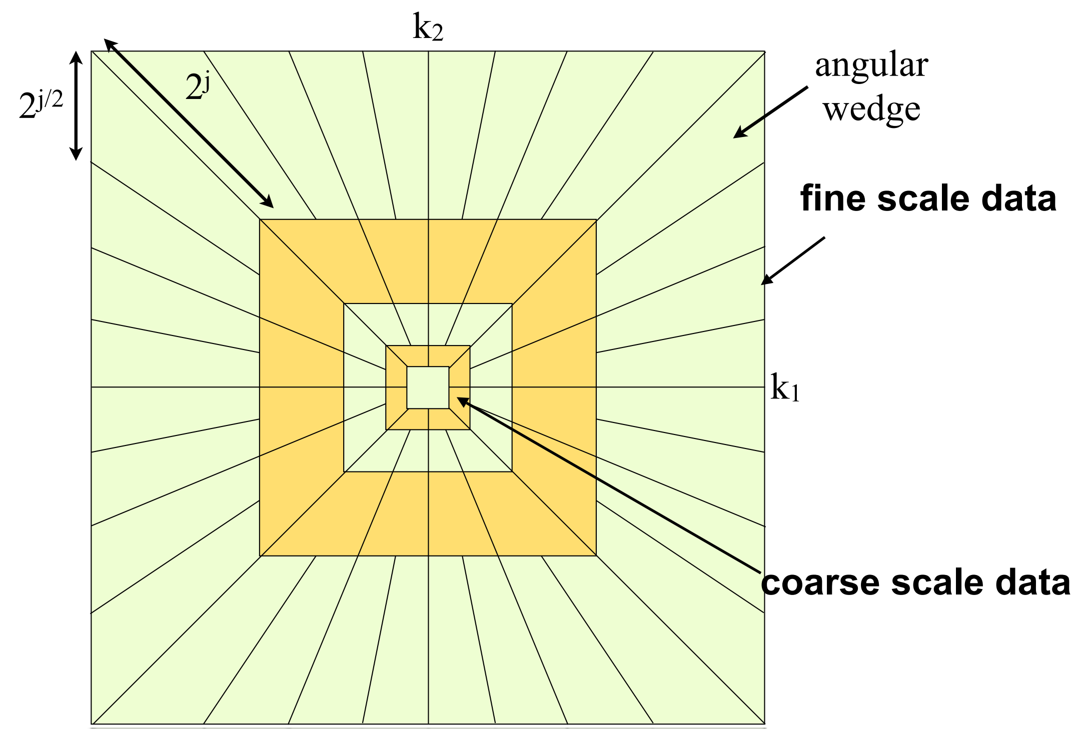
- multiscale
- multidirectional
- exactly reconstructs

Representations for seismic data

Transform	Underlying assumption
FK	plane waves
linear/parabolic Radon transform	linear/parabolic events
wavelet transform	point-like events (1D singularities)
curvelet transform	**curve-like events (2D singularities)**

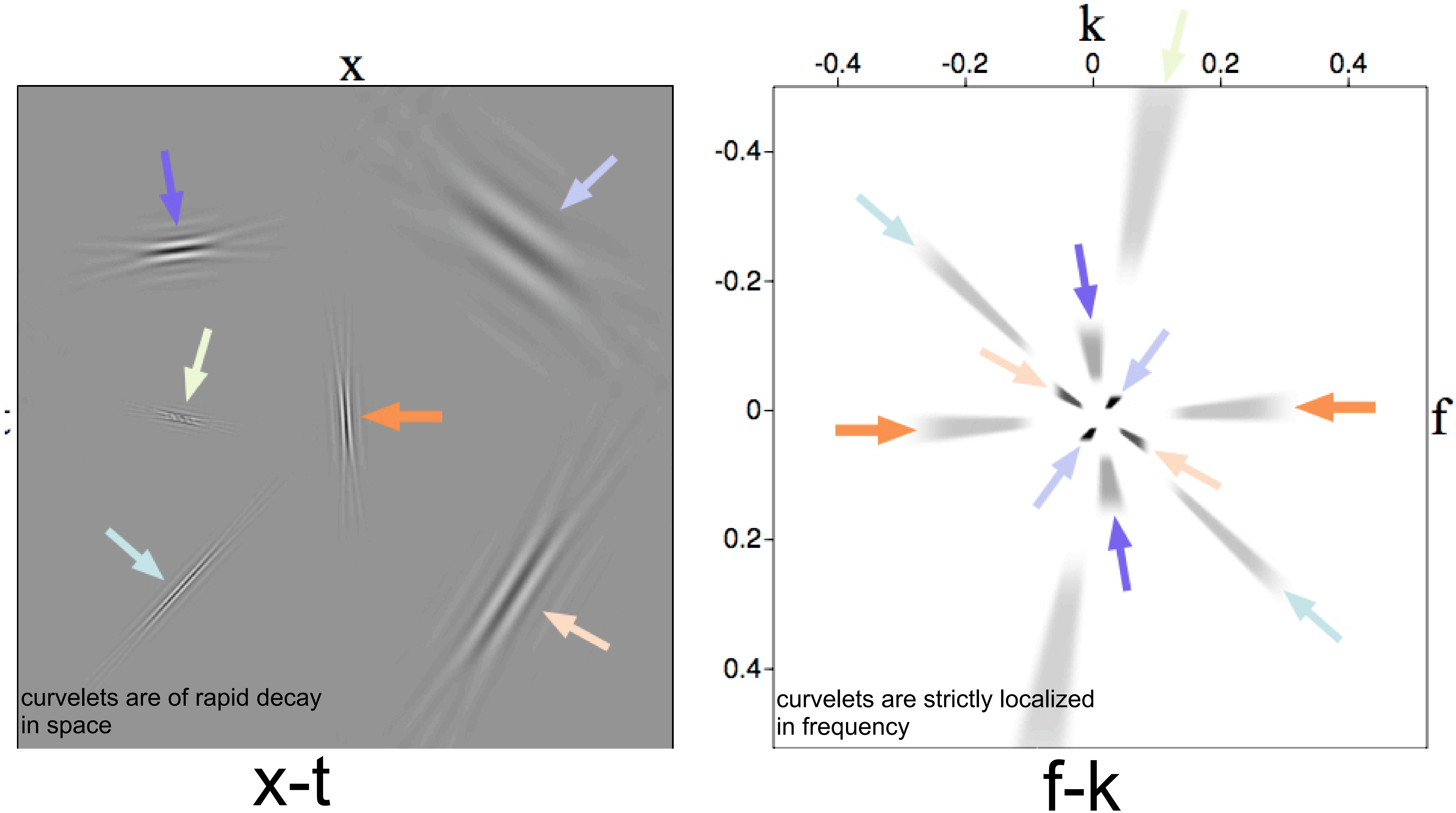
Properties curvelet transform:

- **multiscale:** tiling of the FK domain into dyadic coronae
- **multi-directional:** coronae sub-partitioned into angular wedges, # of angle doubles every other scale
- **anisotropic:** parabolic scaling principle
- **Rapid decay space**
- **Strictly localized in Fourier**
- **Frame with moderate redundancy**



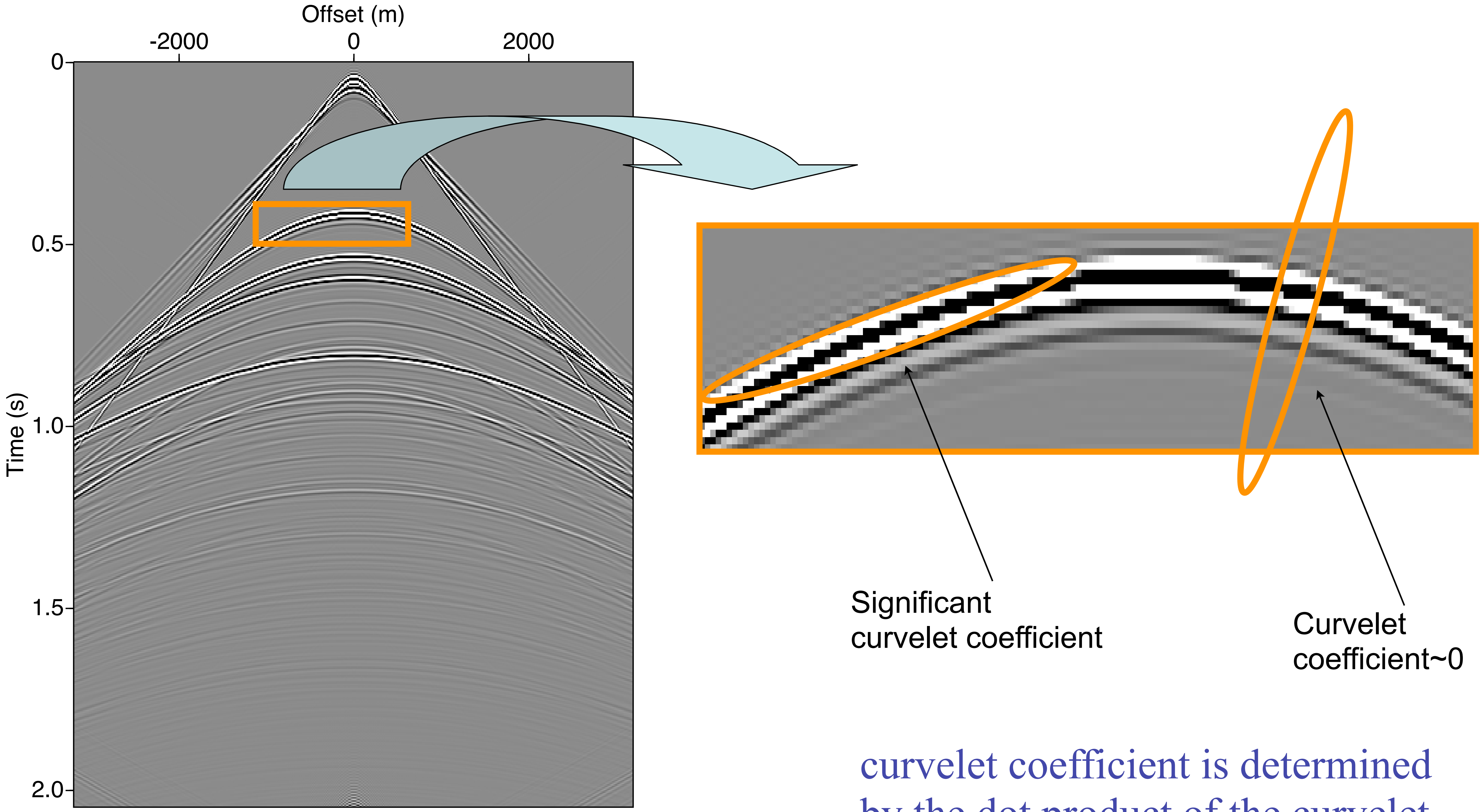
2-D curvelets

[Candes, Donoho, Demanet, Ying]



Oscillatory in one direction and smooth in the others!

Wavefront detection



curvelet coefficient is determined by the dot product of the curvelet function with the data

SLIM
Seismic Laboratory for Imaging and Modeling

Compression

Interested in functions discontinuous along a piecewise smooth (C^2) interface, and otherwise smooth (C^2).

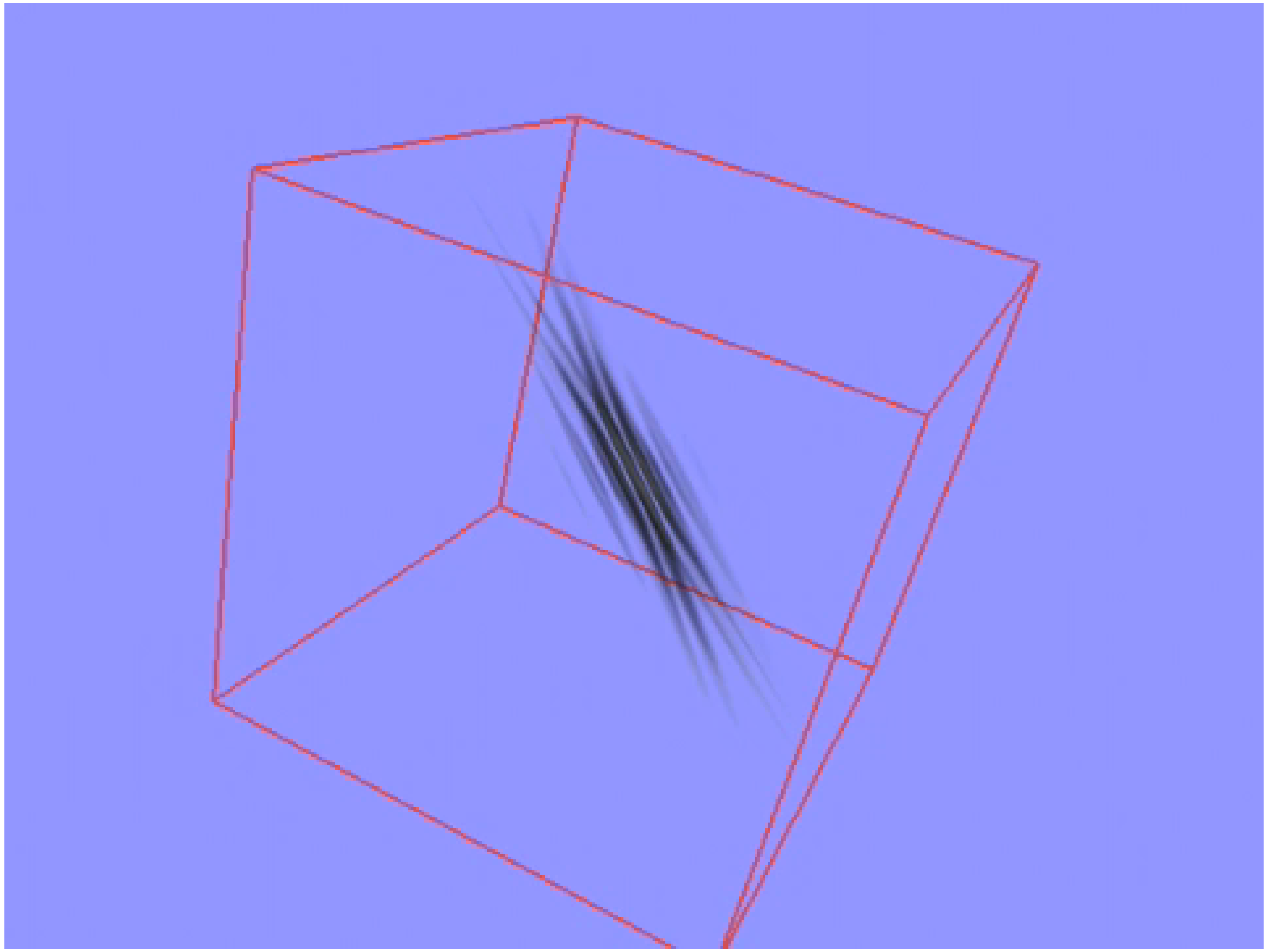
Theorem (Candès, Donoho). For such a model f, the best m-term curvelet expansion f_m obeys

$$||f - f_m||^2 \leq C m^{-2} (\log m)^3.$$

Note: wavelets would give $O(m^{-1})$, so do ridgelets (Candès).

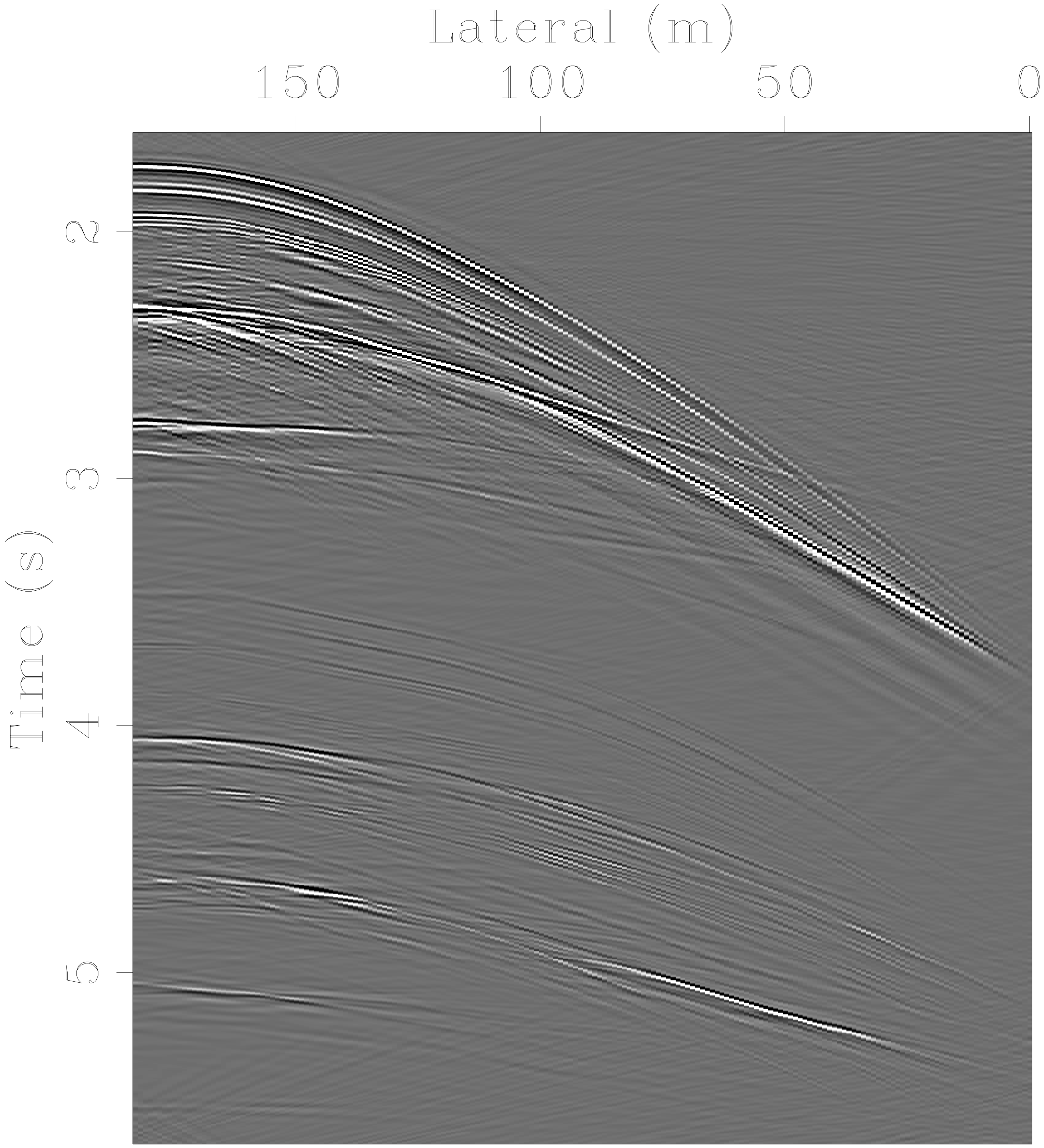
[From Demanet '05]

3-D curvelets



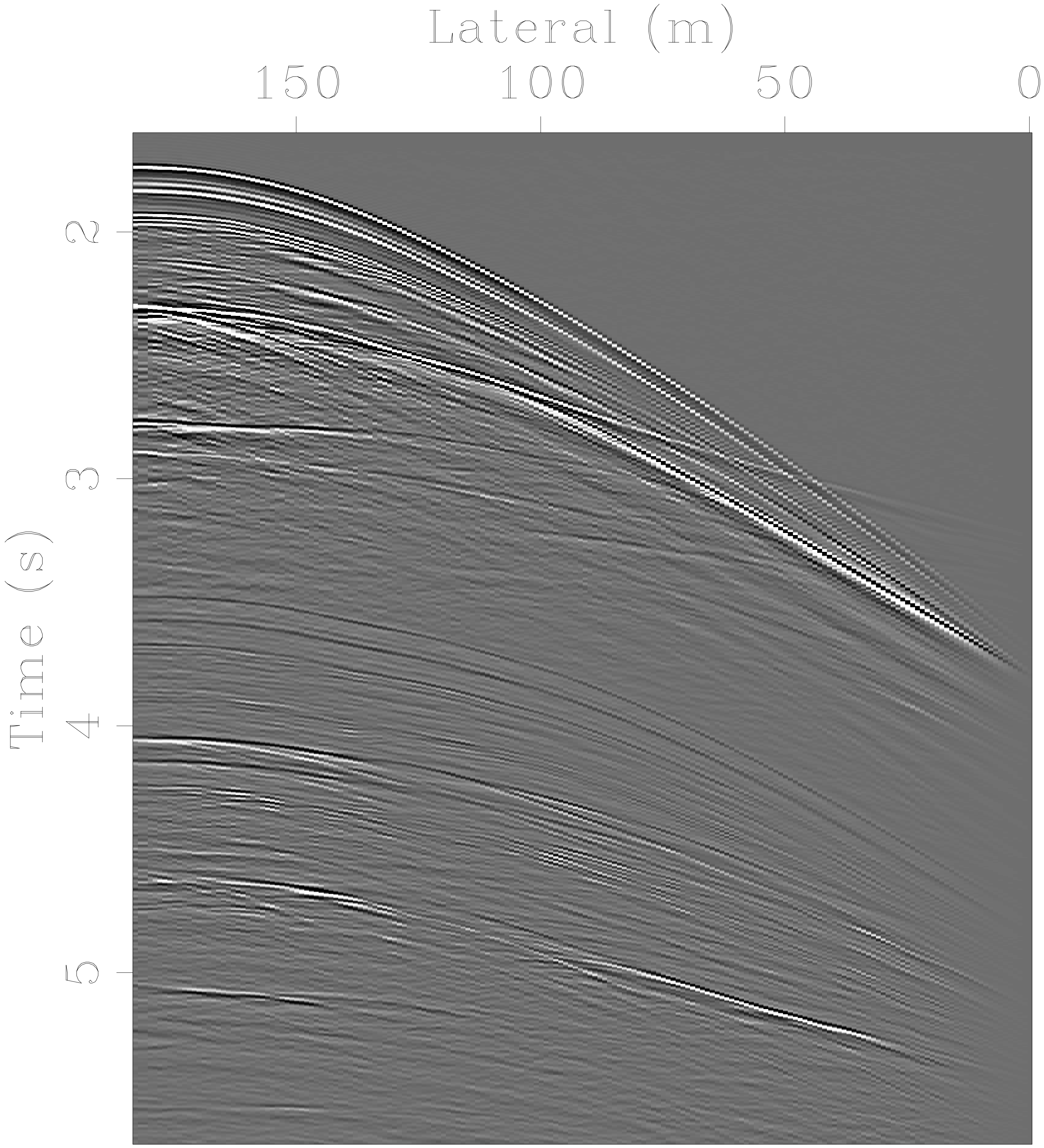
Curvelets live in wedges in the 3 D Fourier plane...

Nonlinear approximation



reconstructed data with p=5

Nonlinear approximation



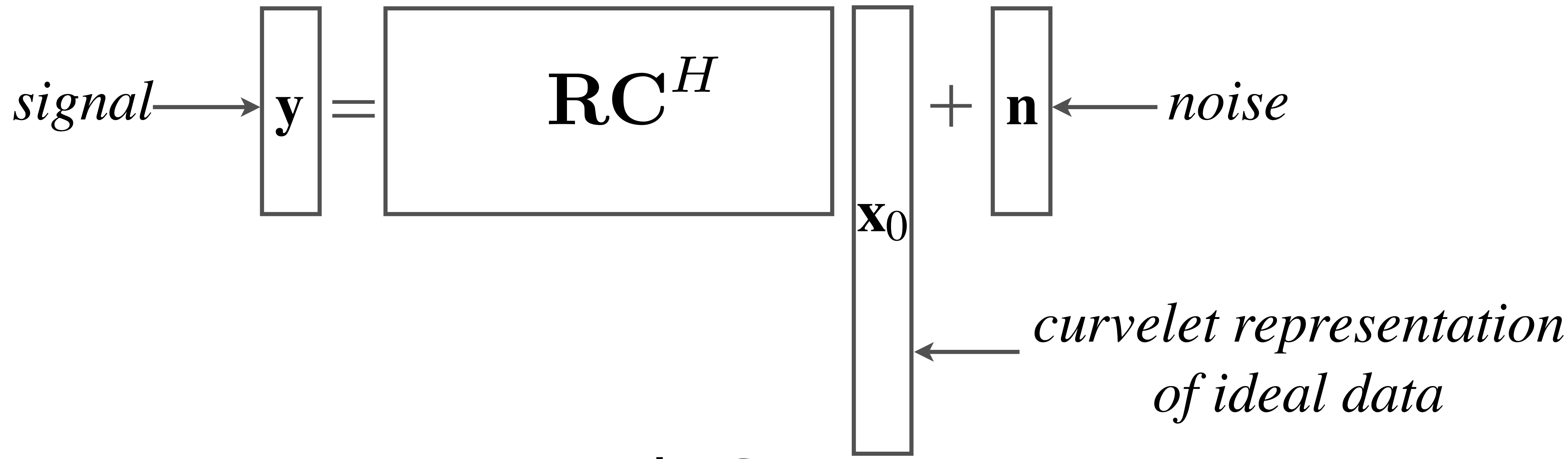
reconstructed data with p=99

Curvelet-based seismic data recovery

joint work with Gilles Hennenfent

Sparsity-promoting inversion*

Reformulation of the problem



Curvelet Reconstruction with Sparsity-promoting Inversion (CRSI)

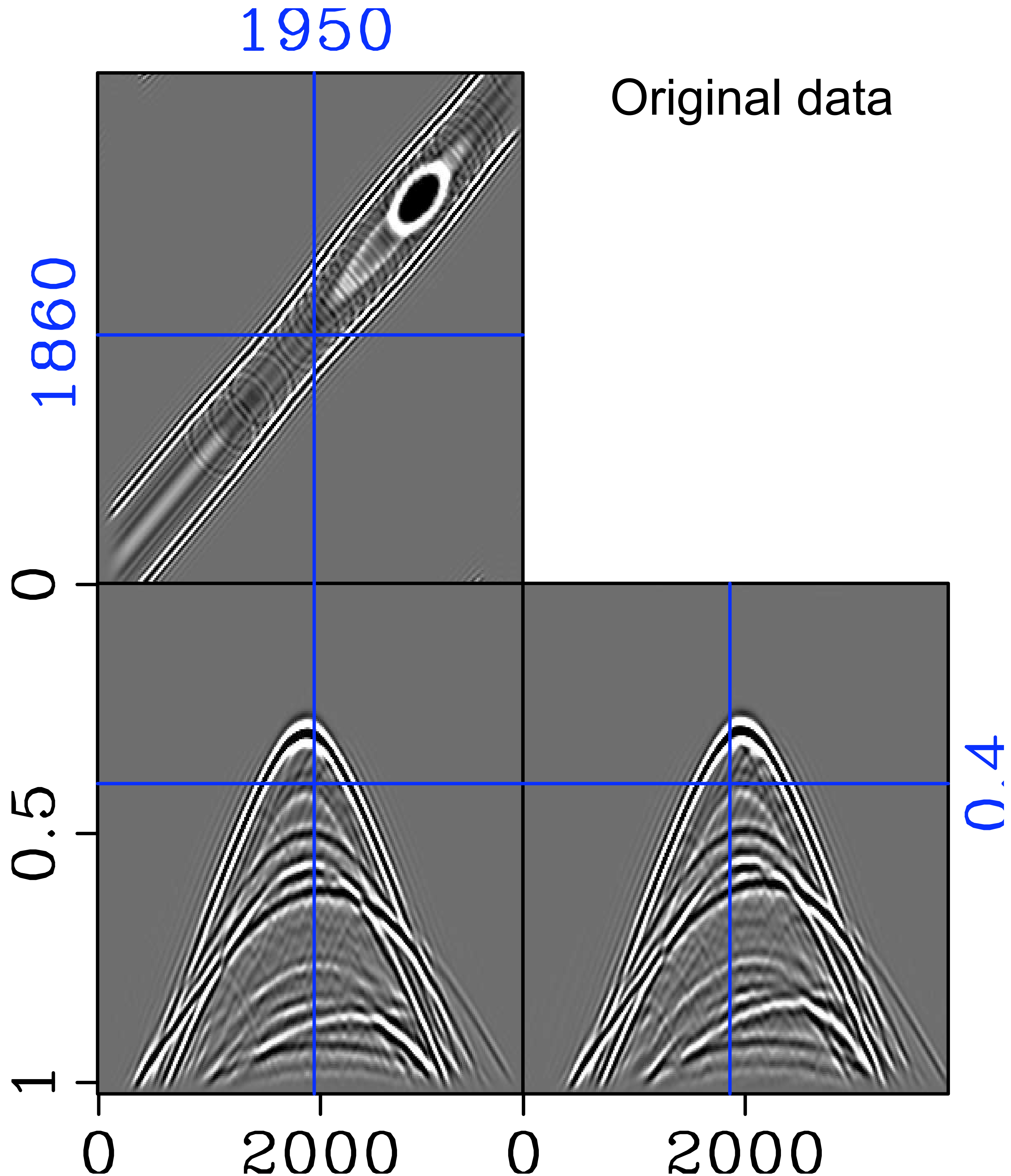
- look for the **sparsest/most compressible**, **physical** solution

KEY POINT OF THE RECOVERY

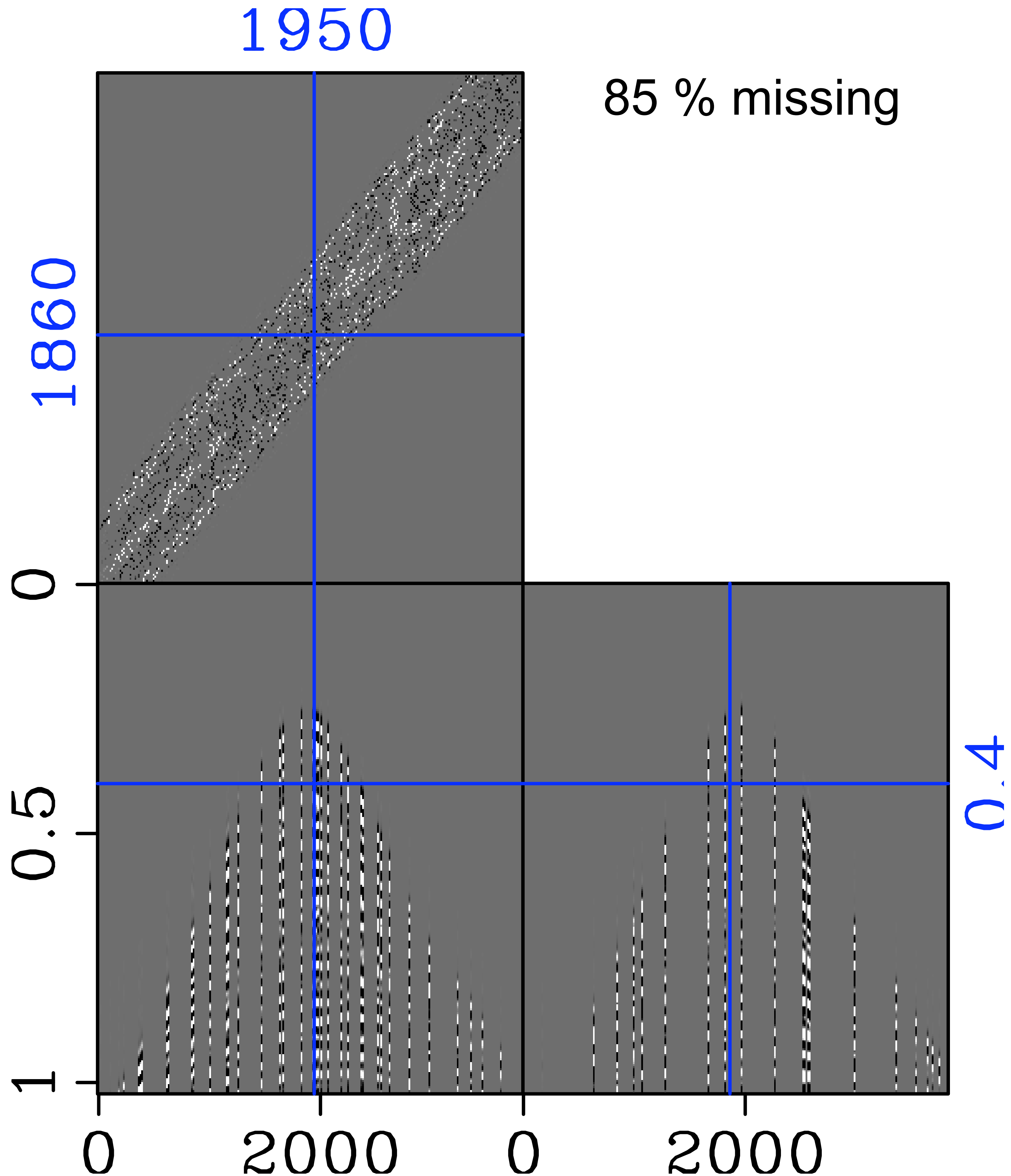
$$\mathbf{P}_\epsilon : \begin{cases} \tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \overbrace{\|\mathbf{W}\mathbf{x}\|_1}^{\text{sparsity constraint}} \quad \text{s.t.} \quad \overbrace{\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2}^{\text{data misfit}} \leq \epsilon \\ \tilde{\mathbf{f}} = \mathbf{C}^T\tilde{\mathbf{x}} \end{cases}$$

* inspired by Stable Signal Recovery (SSR) theory by E. Candès, J. Romberg, T. Tao, Compressed sensing by D. Donoho & Fourier Reconstruction with Sparse Inversion (FRSI) by P. Zwartjes

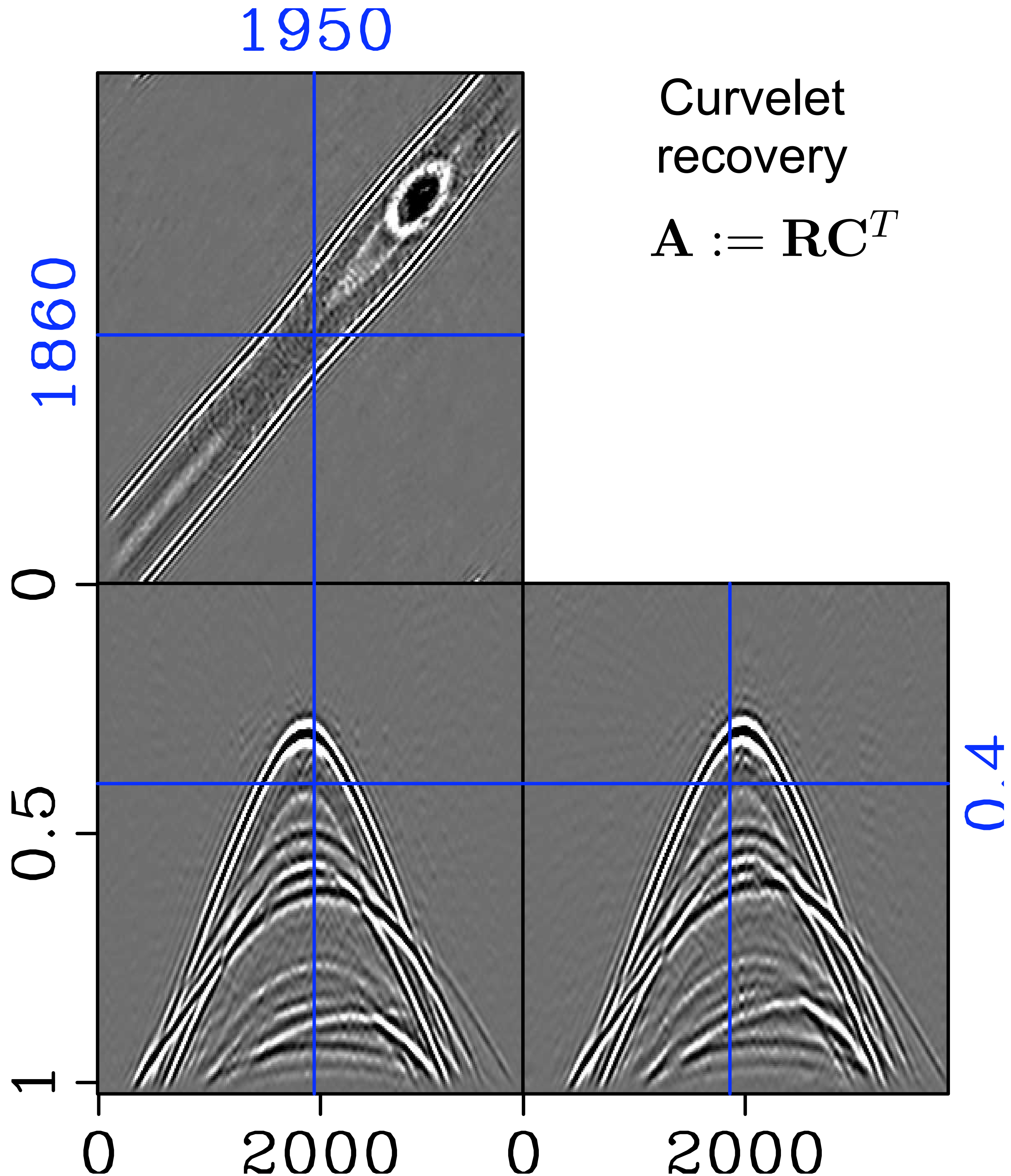
SLIM Seismic Laboratory for Imaging and Modeling



1950
Original data
1860
0
0.5
1
0.4
0
2000
0
2000



1950
85 % missing
1860
0
0.5
1
0.4
0
2000
0
2000



1950
1860
Curvelet
recovery
$\mathbf{A} := \mathbf{R}\mathbf{C}^T$
0
0.5
1
0.4
0
2000
0
2000

Observations

Inverted a rectangular matrix

- worked because the curvelet transform is sparse
- exploits the higher dimensional geometry of seismic wavefields
- curvelets are incoherent with the Dirac measurement basis

Data is recovered for large percentages of traces missing

Is an example of an inverse problem with incomplete data

Can these ideas be extended to recover migration amplitudes?

- **approximately invert a PsDO**
- **diagonalize zero-order PsDO’s**

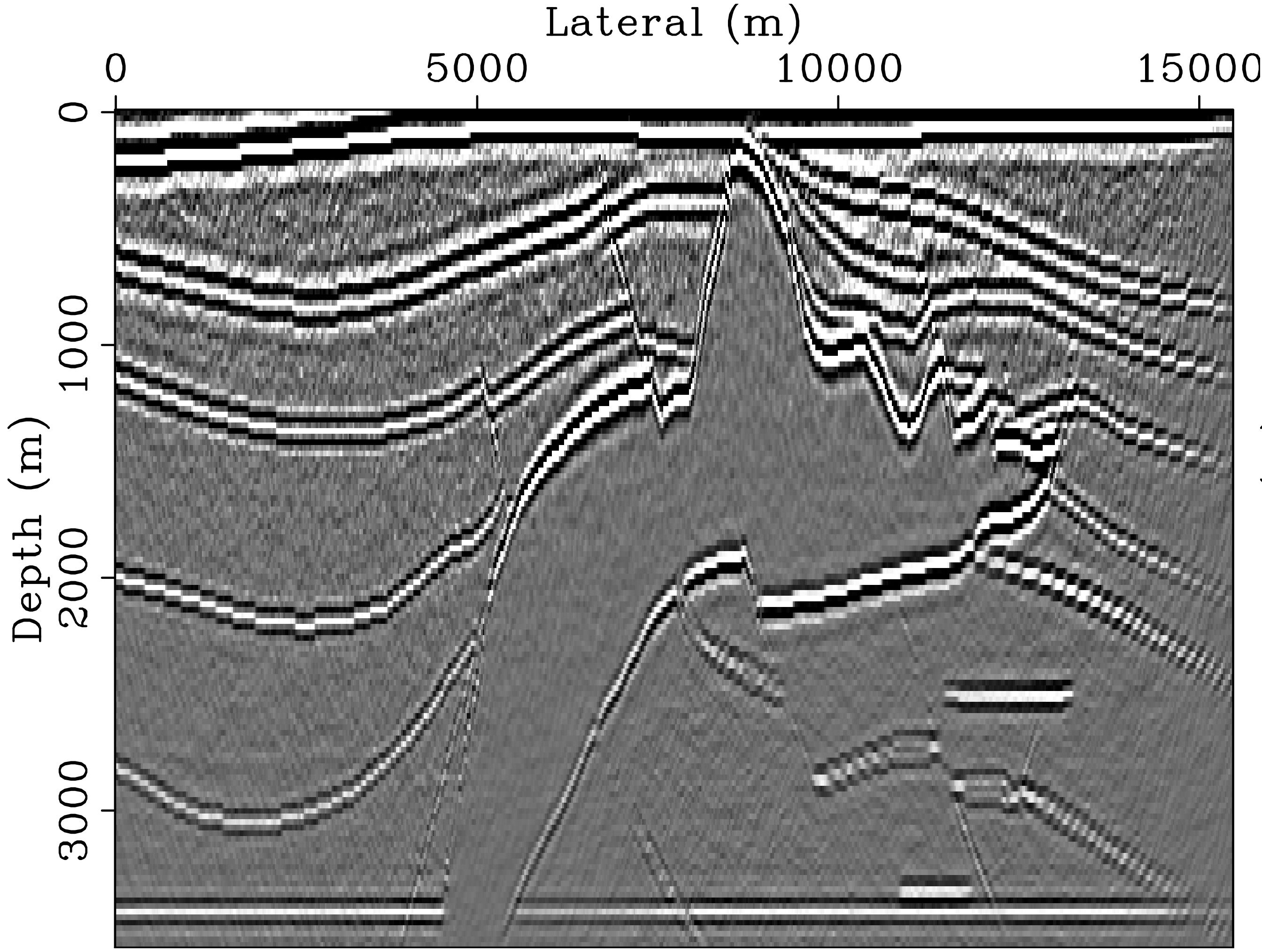
Stable seismic amplitude recovery

"Sparsity- and continuity-promoting seismic image recovery with curvelet frames"

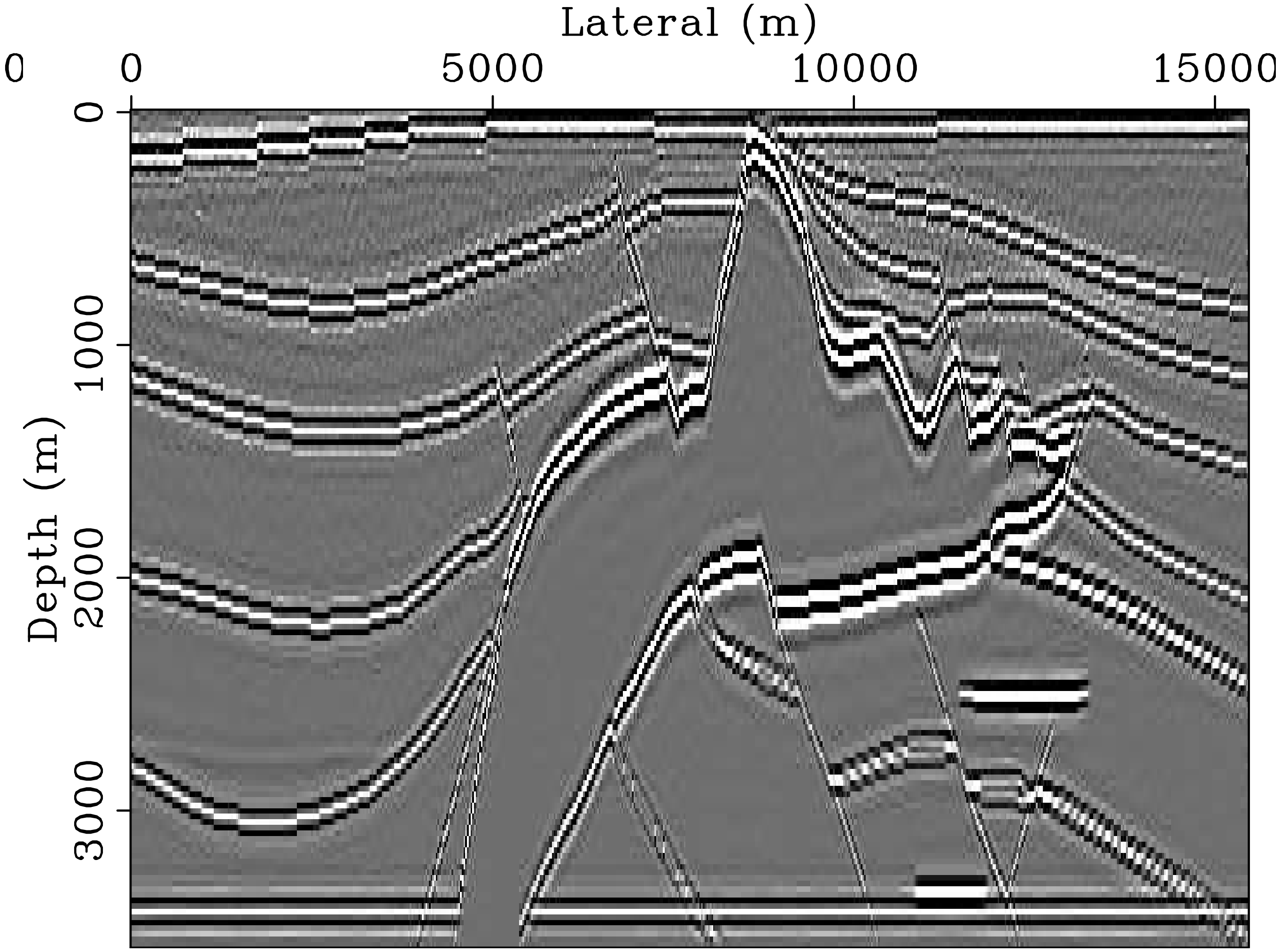
by

F.H, P. Moghaddam & C. Stolk

to appear in special issue on imaging in ACHA



Migrated data



Amplitude-corrected & denoised migrated data

Existing scaling methods

Methods are based on a diagonal approximation of Ψ

- Illumination-based normalization (Rickett '02)
- Amplitude preserved migration (Plessix & Mulder '04)
- Amplitude corrections (Guitton '04)
- Amplitude scaling (Symes '07)

We are interested in an 'Operator and image adaptive' scaling method which

- estimates the action of Ψ from a *reference* vector close to the actual image
- assumes a *smooth* symbol of Ψ in space and angle
- does not require the reflectors to be conormal <=> allows for conflicting dips
- stably inverts the diagonal

Our approach

"Forward" model:

$$\begin{aligned} \mathbf{y} &= \mathbf{K}^T\mathbf{K}\mathbf{m} + \boldsymbol{\varepsilon} \\ &\approx \mathbf{A}\mathbf{x}_0 + \boldsymbol{\varepsilon} \end{aligned}$$

with

$$\begin{aligned} \mathbf{y} &= \text{migrated data} \\ \mathbf{A} &:= \mathbf{C}^T\boldsymbol{\Gamma} \\ \mathbf{A}\mathbf{A}^T\mathbf{r} &\approx \mathbf{K}^T\mathbf{K}\mathbf{r} \\ \mathbf{K} &= \text{the demigration operator} \\ \boldsymbol{\epsilon} &= \text{migrated noise.} \end{aligned}$$

- diagonal approximation of the demigration-migration operator
- costs one demigration-migration to estimate the diagonal weighting

Solution

Solve

$$\mathbf{P}: \quad \begin{cases} \min_{\mathbf{x}} J(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \le \epsilon \\ \\ \tilde{\mathbf{m}} = (\mathbf{A}^{\mathbf{H}})^{\dagger}\tilde{\mathbf{x}} \end{cases}$$

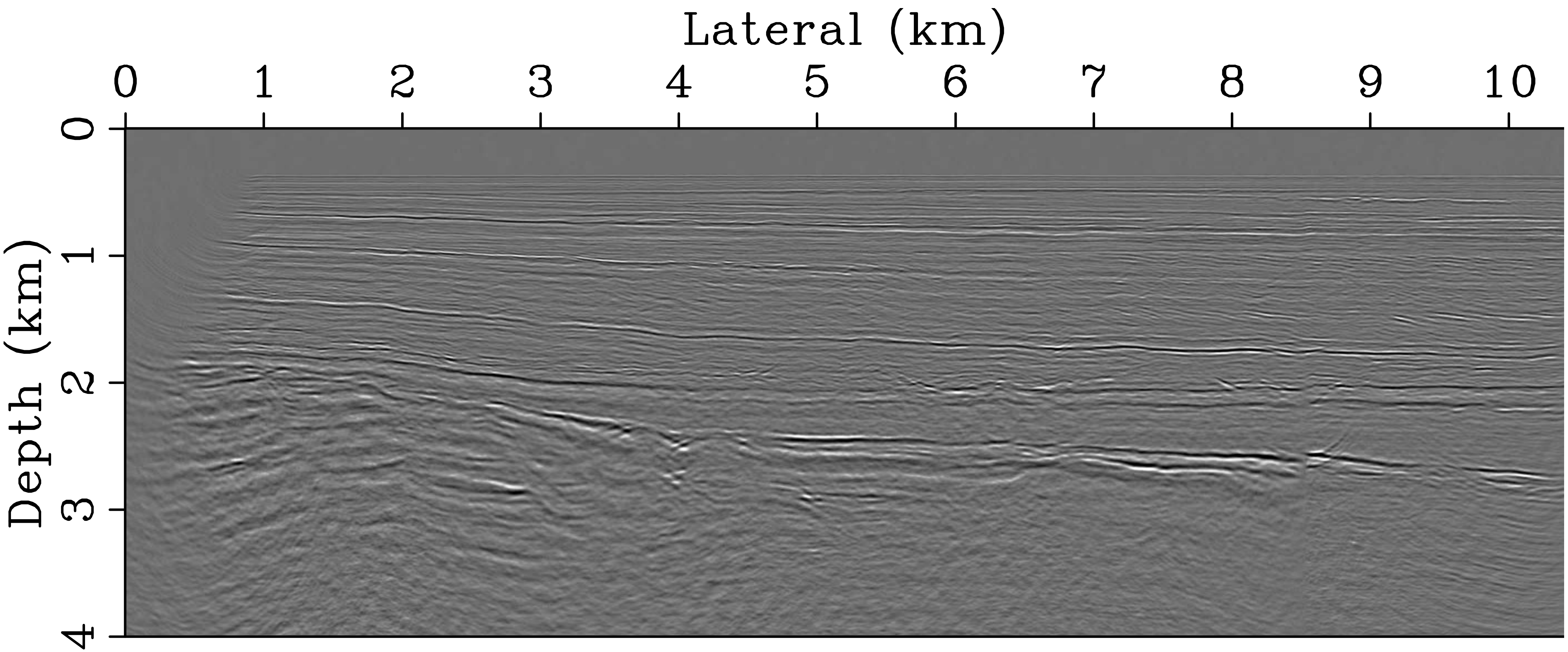
with

$$J(\mathbf{x}) = \overbrace{\alpha\|\mathbf{x}\|_1}^{\text{sparsity}} + \beta \underbrace{\left\|\mathbf{\Lambda}^{1/2}\left(\mathbf{A}^{H}\right)^{\dagger}\mathbf{x}\right\|_p}_{\text{continuity}}.$$

- need sparsity on the model
- invariance under the normal operator

Nonlinear approximation

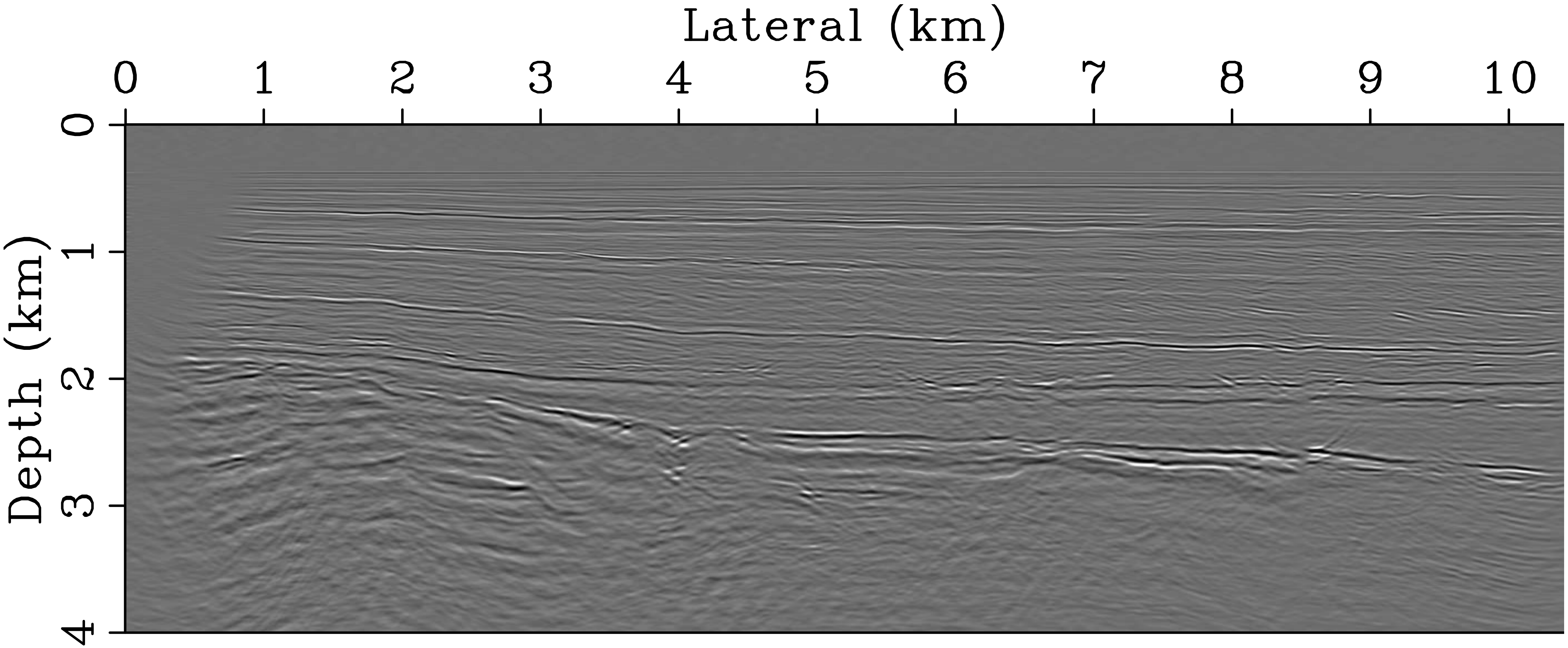
Migrated mobil data set



reconstructed data with p=99

Nonlinear approximation

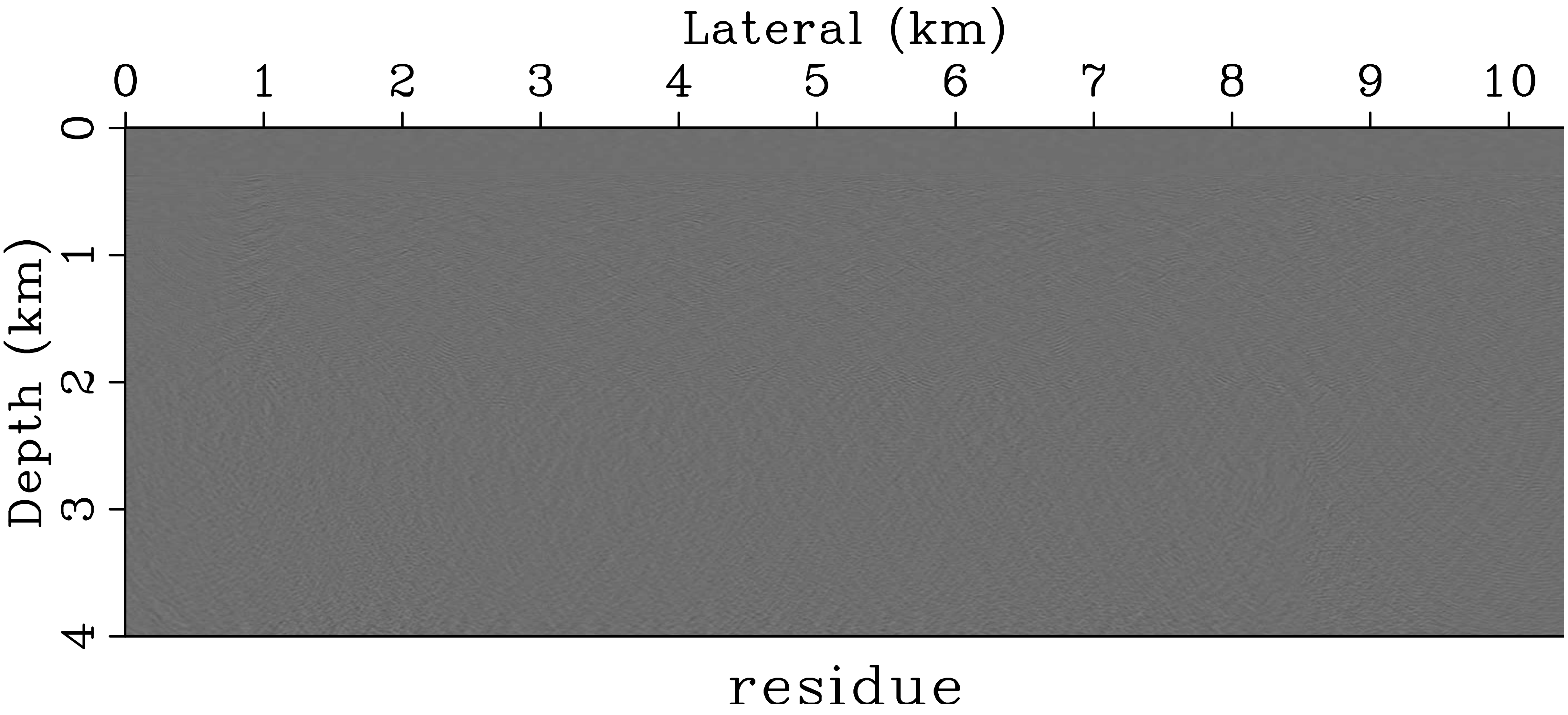
Recovery from largest 3 %



reconstructed data with p=3

Nonlinear approximation

Difference



Diagonal approximation of the Hessian

Normal/Gramm operator

[Stolk 2002, ten Kroode 1997, de Hoop 2000, 2003]

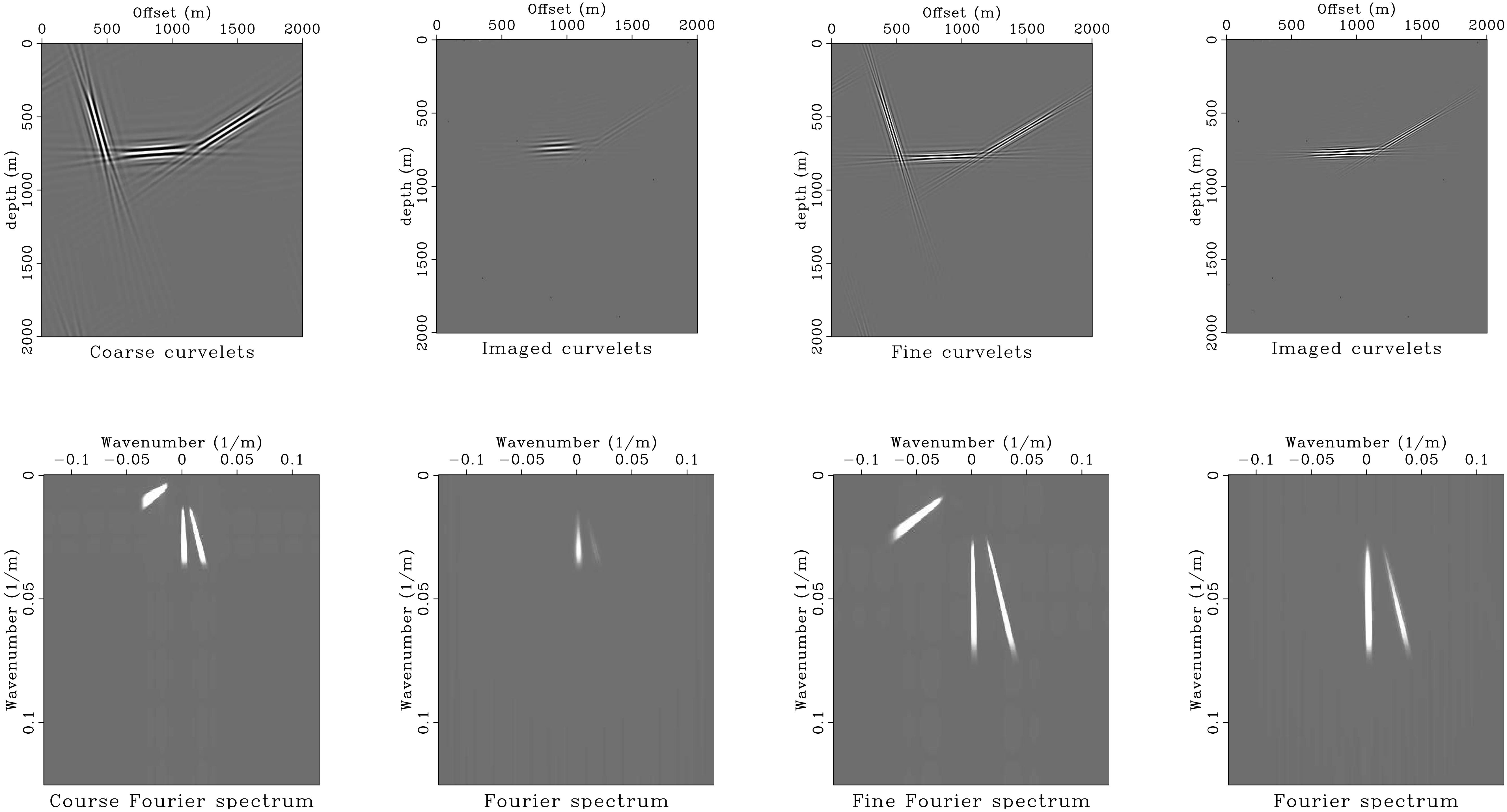
In high-frequency limit Ψ is a PsDO

$$\big(\Psi f\big)(x) = \int_{\mathbb{R}^d} e^{-ix\cdot\xi} a(x,\xi)\hat{f}(\xi)\mathrm{d}\xi$$

- pseudolocal
- singularities are preserved

Inversion corrects for the ‘Hessian’

Invariance under Gramm matrix



- curvelets remain invariant
- approximation improves for higher frequencies

Approximation

So let $\mathbf{\Psi} = \mathbf{\Psi}(x, D)$ be a pseudodifferential operator of order 0, with homogeneous principal symbol $a(x, \xi)$.

$K \mapsto K(-\Delta)^{-1/2}$ **or** $K \mapsto \partial_t^{-1/2} K$

$m \mapsto (-\Delta)^{1/2} m$ **with** $((-\Delta)^{\alpha} f)^{\wedge}(\xi) = |\xi|^{2\alpha} \cdot \hat{f}(\xi)$.

Lemma 1. *With C' some constant, the following holds*

$$\|(\Psi(x, D) - a(x_\nu, \xi_\nu))\varphi_\nu\|_{L^2(\mathbb{R}^n)} \leq C' 2^{-|\nu|/2}. \tag{14}$$

To approximate Ψ, we define the sequence $\mathbf{u} := (u_\mu)_{\mu \in \mathcal{M}} = a(x_\mu, \xi_\mu)$. Let $\mathbf{D}_\Psi$ be the diagonal matrix with entries given by $\mathbf{u}$. Next we state our result on the approximation of Ψ by $C^T \mathbf{D}_\Psi C$.

Approximation

Theorem 1. *The following estimate for the error holds*

$$\|(\Psi(x,D) - C^T \mathbf{D}_\Psi C)\varphi_\mu\|_{L^2(\mathbb{R}^n)} \leq C'' 2^{-|\mu|/2},$$

where C'' is a constant depending on Ψ.

Allows for the decomposition

$$\begin{aligned} \big(\Psi\varphi_\mu\big)(x) &\simeq \big(C^T \mathbf{D}_\Psi C \varphi_\mu\big)(x) \\ &= \big(AA^T \varphi_\mu\big)(x) \end{aligned}$$

with $A := \sqrt{\mathbf{D}_\Psi} C$ and $A^T := C^T \sqrt{\mathbf{D}_\Psi}$.

Approximation

$$\begin{aligned} y(x) &= (\Psi m)(x) + e(x) \\ &\simeq (AA^T m)(x) + e(x) \\ &= Ax_0 + e, \end{aligned}$$

Wavelet-vagulette like

Amenable to nonlinear recovery

SLIM
Seismic Laboratory for Imaging and Modeling

Estimation of the diagonal scaling

Diagonal estimation

Define a reference vector (say conventional image).

Calculate 'data'

$$\mathbf{b} = \mathbf{\Psi r}$$

Define the matrix

$$\mathbf{P} := \mathbf{C}^T \operatorname{diag}(\mathbf{v}) \qquad \text{with} \qquad \mathbf{v} = \mathbf{Cr}$$

Invert

$$\tilde{\mathbf{u}} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{b} - \mathbf{Pu}\|_2^2 + \eta^2\|\mathbf{Lu}\|_2^2$$

Diagonal estimation

Impose smoothness in phase space

$$\mathbf{L} = [\mathbf{D}_1 \quad \mathbf{D}_2 \quad \mathbf{D}_\theta]$$

Calculate: $\mathbf{b} = \mathbf{\Psi r}$ and $\mathbf{v} = \mathbf{Cr}$.

Set: $\eta = \eta_{min}$;

while $\exists\,(\tilde{u}_\mu)_{\mu \in \mathcal{M}} < 0$ **do**

Solve

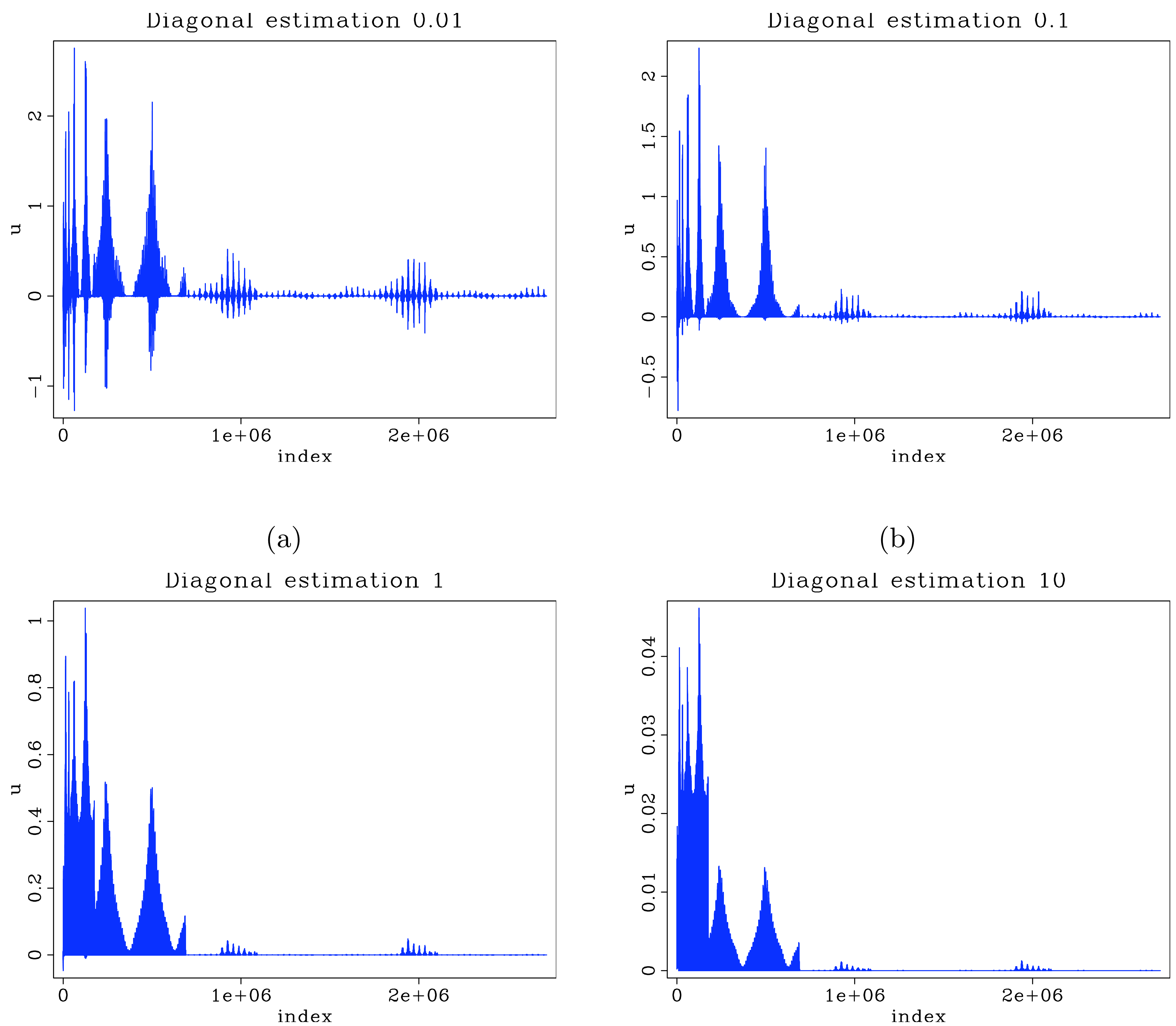
$\widetilde{\mathbf{u}} = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{b} - \mathbf{Pu}\|_2^2 + \eta^2\|\mathbf{Lu}\|_2^2$

Increase the Lagrange multiplier

$\lambda = \eta + \Delta\eta$

end while

Diagonal estimation



Seismic amplitude recovery

Recovery

Final form

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \boldsymbol{\varepsilon}$$

with $\mathbf{x}_0 = \mathbf{\Gamma}\mathbf{C}\mathbf{m}$ and $\boldsymbol{\epsilon} = \mathbf{A}\mathbf{e}$.

Solve

$$\mathbf{P}: \quad \begin{cases} \min_{\mathbf{x}} J(\mathbf{x}) \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \\ \\ \tilde{\mathbf{m}} = (\mathbf{A}^{\mathbf{H}})^{\dagger}\tilde{\mathbf{x}} \end{cases}$$

with

$$J(\mathbf{x}) = \overbrace{\alpha\|\mathbf{x}\|_1}^{\text{sparsity}} + \beta \underbrace{\|\mathbf{\Lambda}^{1/2}\left(\mathbf{A}^{H}\right)^{\dagger}\mathbf{x}\|_p}_{\text{continuity}}.$$

Image recovery
anisotropic diffusion

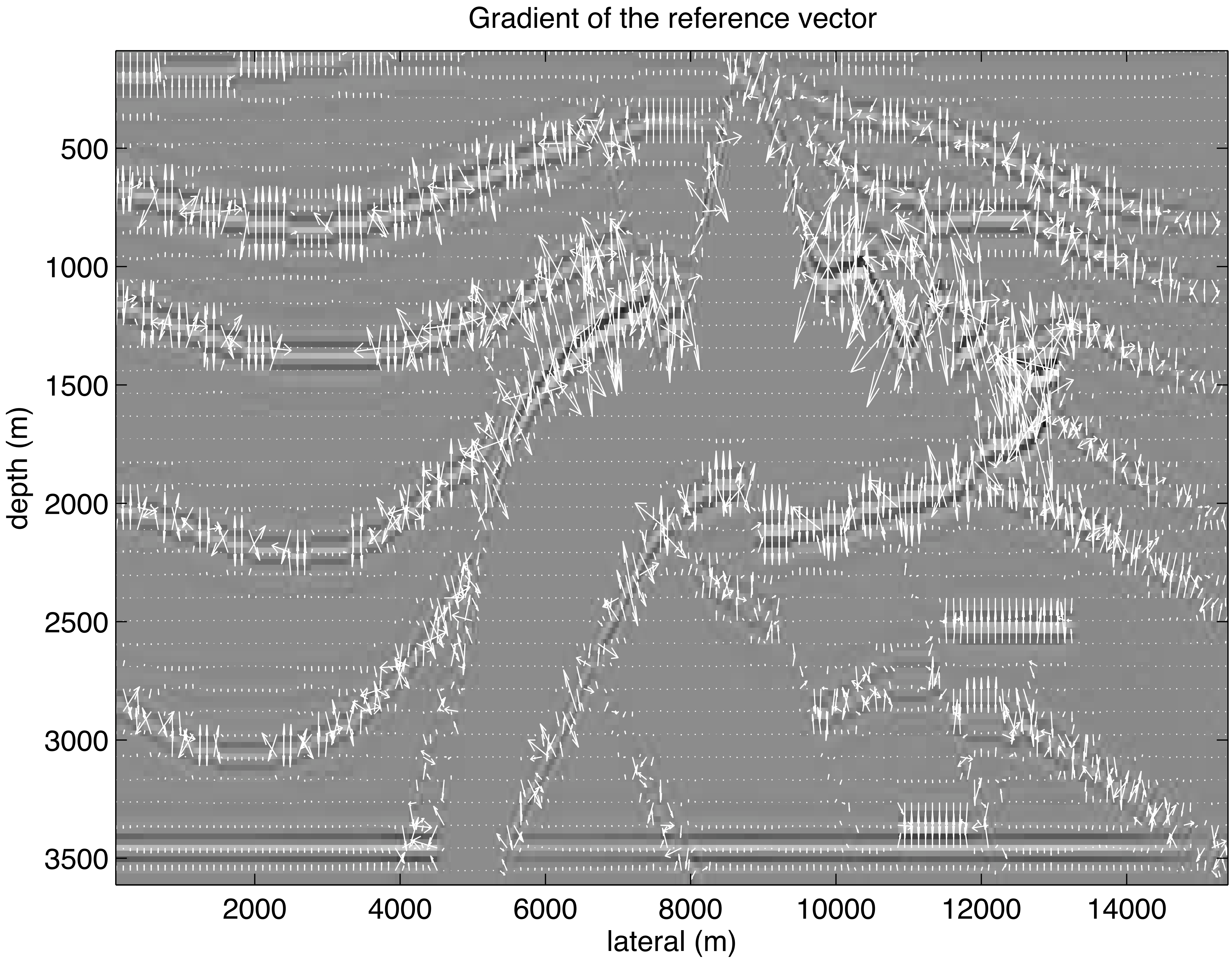
[Black et. al '98, Fehmers et. al. '03 and Shertzer '03]

Define

$$J_c(\mathbf{m}) = \|\mathbf{\Lambda}^{1/2}\boldsymbol{\nabla}\mathbf{m}\|_p$$

with p=2

$$\mathbf{\Lambda}[\mathbf{r}] = \frac{1}{\|\boldsymbol{\nabla}\mathbf{r}\|_2^2 + 2\upsilon}\left\{\begin{pmatrix} +\mathbf{D}_2\mathbf{r} \\ \\ -\mathbf{D}_1\mathbf{r}\end{pmatrix}\begin{pmatrix} +\mathbf{D}_2\mathbf{r} & -\mathbf{D}_1\mathbf{r}\end{pmatrix} + \upsilon\mathbf{Id}\right\}$$



Gradient of the reference vector
depth (m)
500
1000
1500
2000
2500
3000
3500
lateral (m)
2000
4000
6000
8000
10000
12000
14000

Recovery

Step 1: Update of the Jacobian of $\frac{1}{2}\|\mathbf{y}-\mathbf{Ax}\|_2^2$:

$$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{A}^T\left(\mathbf{y}-\mathbf{Ax}\right);$$

Step 2: projection onto the ℓ_1 ball $S = \{\|\mathbf{x}\|_1 \leq \|\mathbf{x}_0\|_1\}$ by soft thresholding

$$\mathbf{x} \leftarrow T_{\lambda\mathbf{w}}(\mathbf{x});$$

Step 3: projection onto the anisotropic diffusion ball $C = \{\mathbf{x} : J(\mathbf{x}) \leq J(\mathbf{x}_0)\}$ by

$$\mathbf{x} \leftarrow \mathbf{x} - \kappa \boldsymbol{\nabla}_{\mathbf{x}} J_c(\mathbf{x})$$

Initialize:

$\mathsf{m} = 0$;

$\mathbf{x}^0 = \mathbf{0}$;

$\mathbf{y} = \mathbf{K}^T\mathbf{d}$;

Choose:

M and L

$\|\mathbf{A}^T\mathbf{y}\|_\infty > \lambda_1 > \lambda_2 > \cdots$

while $\|\mathbf{y} - \mathbf{A}\widetilde{\mathbf{x}}\|_2 > \epsilon$ **do**

$\mathsf{m} = \mathsf{m} + 1$;

$\mathbf{x}^{\mathsf{m}} = \mathbf{x}^{\mathsf{m}-1}$;

for $l = 1$ to L **do**

$\mathbf{x}^{\mathsf{m}} = T_{\lambda_{\mathsf{m}}}\big(\mathbf{x}^{\mathsf{m}} + \mathbf{A}^T\left(\mathbf{y} - \mathbf{x}^{\mathsf{m}}\right)\big)$ {Iterative thresholding}

end for

Anisotropic descent update;

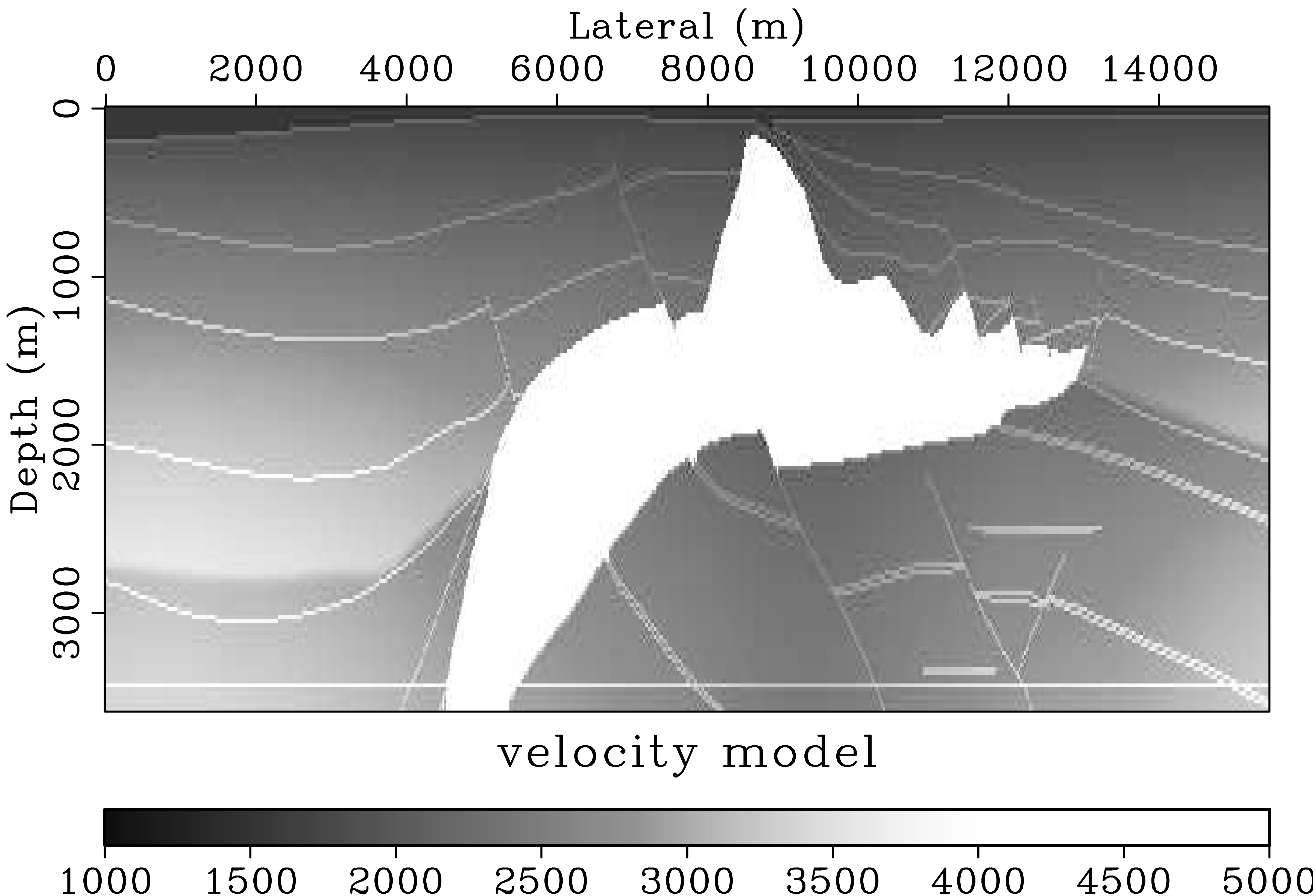
$\mathbf{x}^{\mathsf{m}} = \mathbf{x}^{\mathsf{m}} - \beta\boldsymbol{\nabla}_{\mathbf{x}^{\mathsf{m}}}\ J_c(\mathbf{x}^{\mathsf{m}})$;

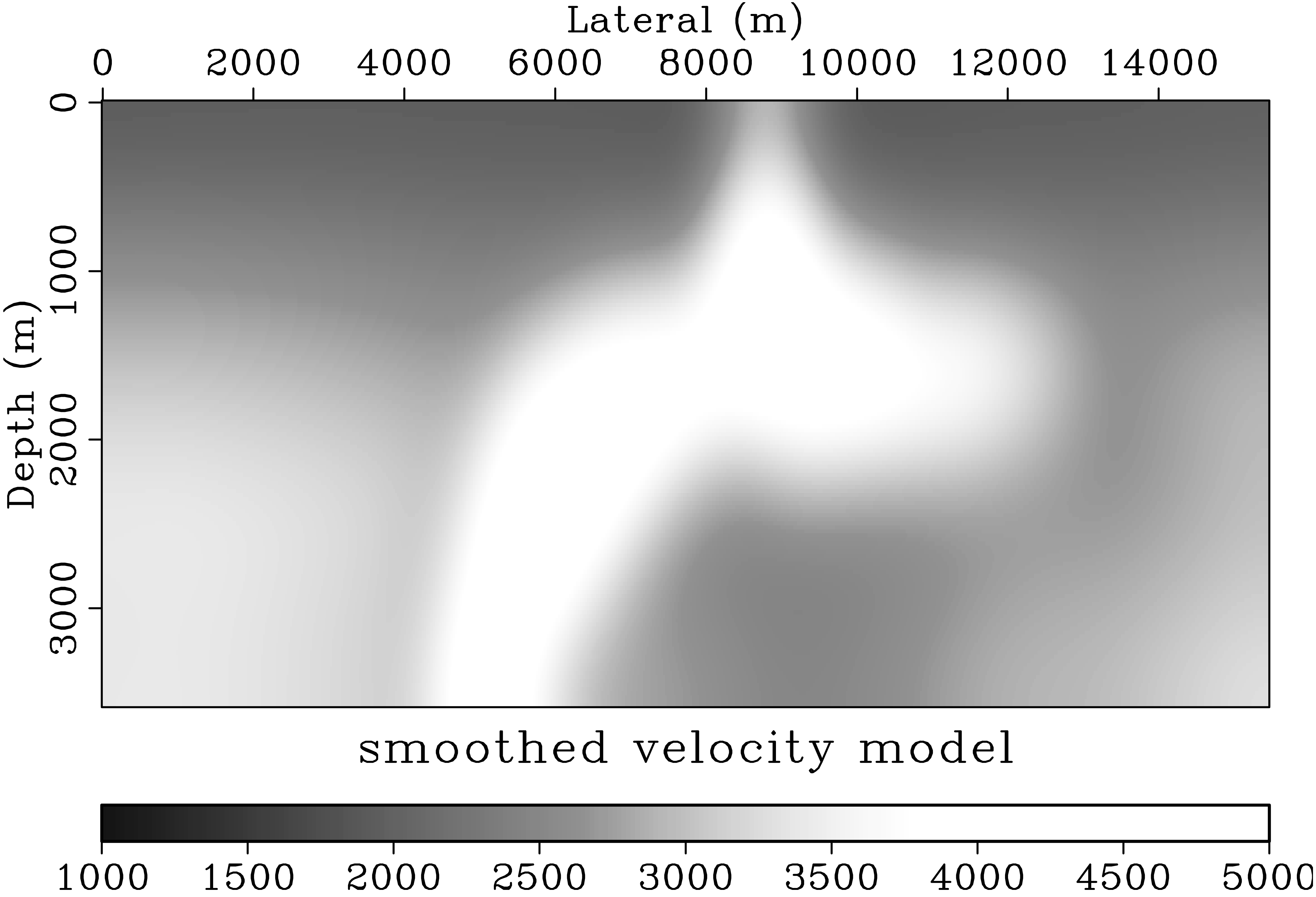
end while

$\widetilde{\mathbf{x}} = \mathbf{x}_{\mathsf{m}}$; $\widetilde{\mathbf{m}} = \left(\mathbf{A}^T\right)^{\dagger}\widetilde{\mathbf{x}}$.

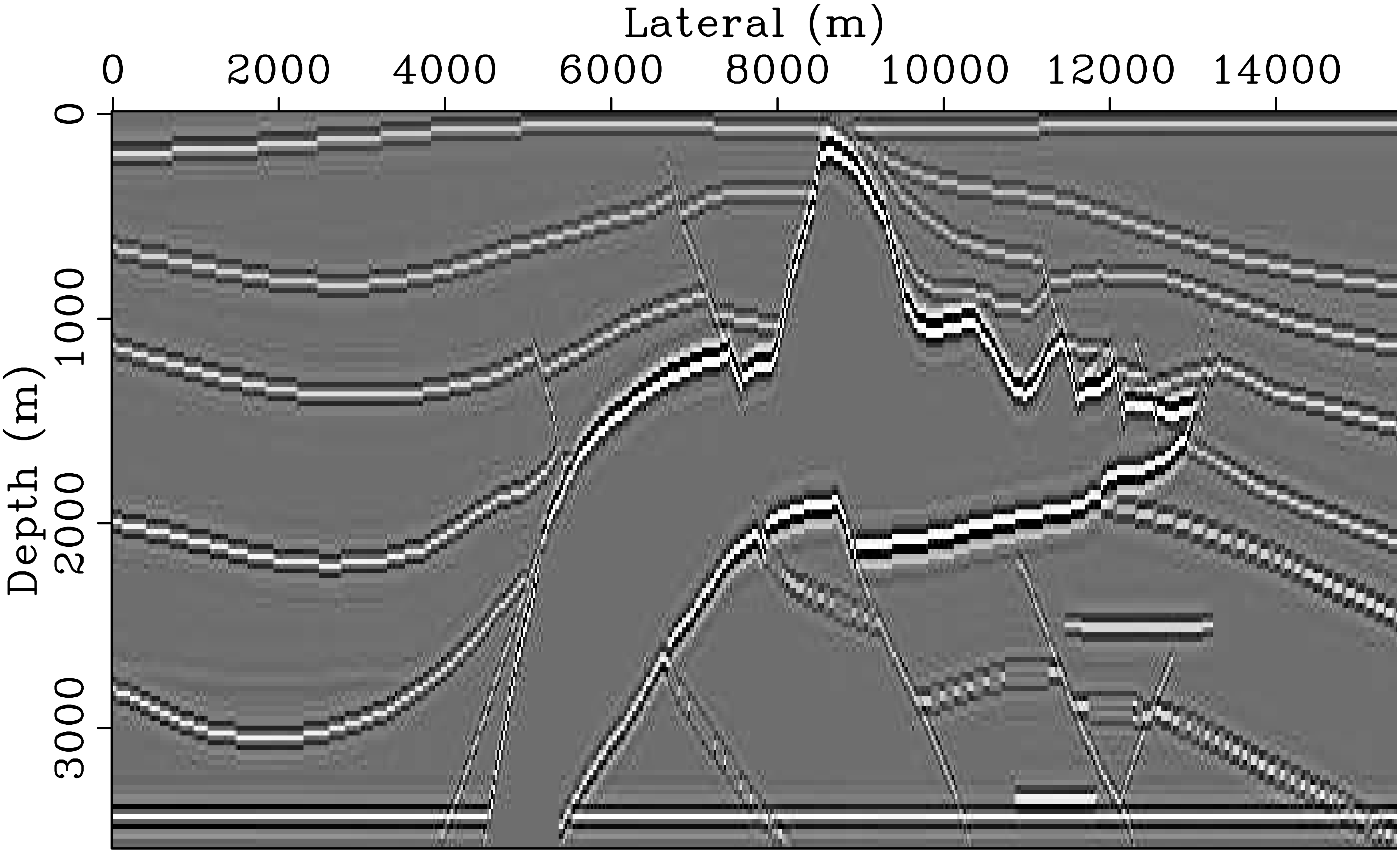
Table 2: Sparsity-and continuity-enhancing recovery of seismic amplitudes.

Application to the SEG AA' model


Lateral (m)
0
2000
4000
6000
8000
10000
12000
14000
Depth (m)
0
1000
2000
3000
velocity model
1000
1500
2000
2500
3000
3500
4000
4500
5000



smoothed velocity model



bandpass−filtered reflectivity

Example

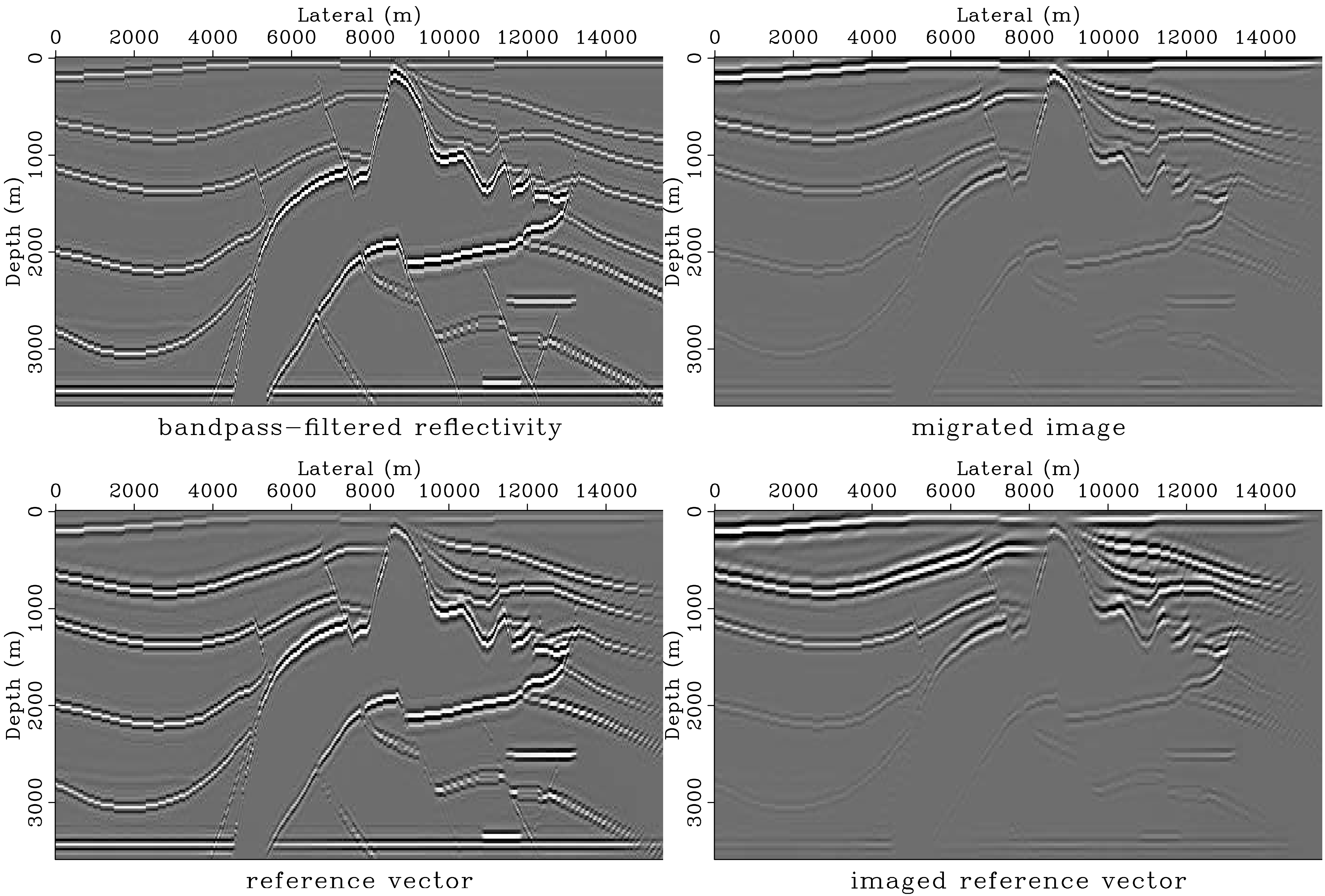
SEGAA' data:

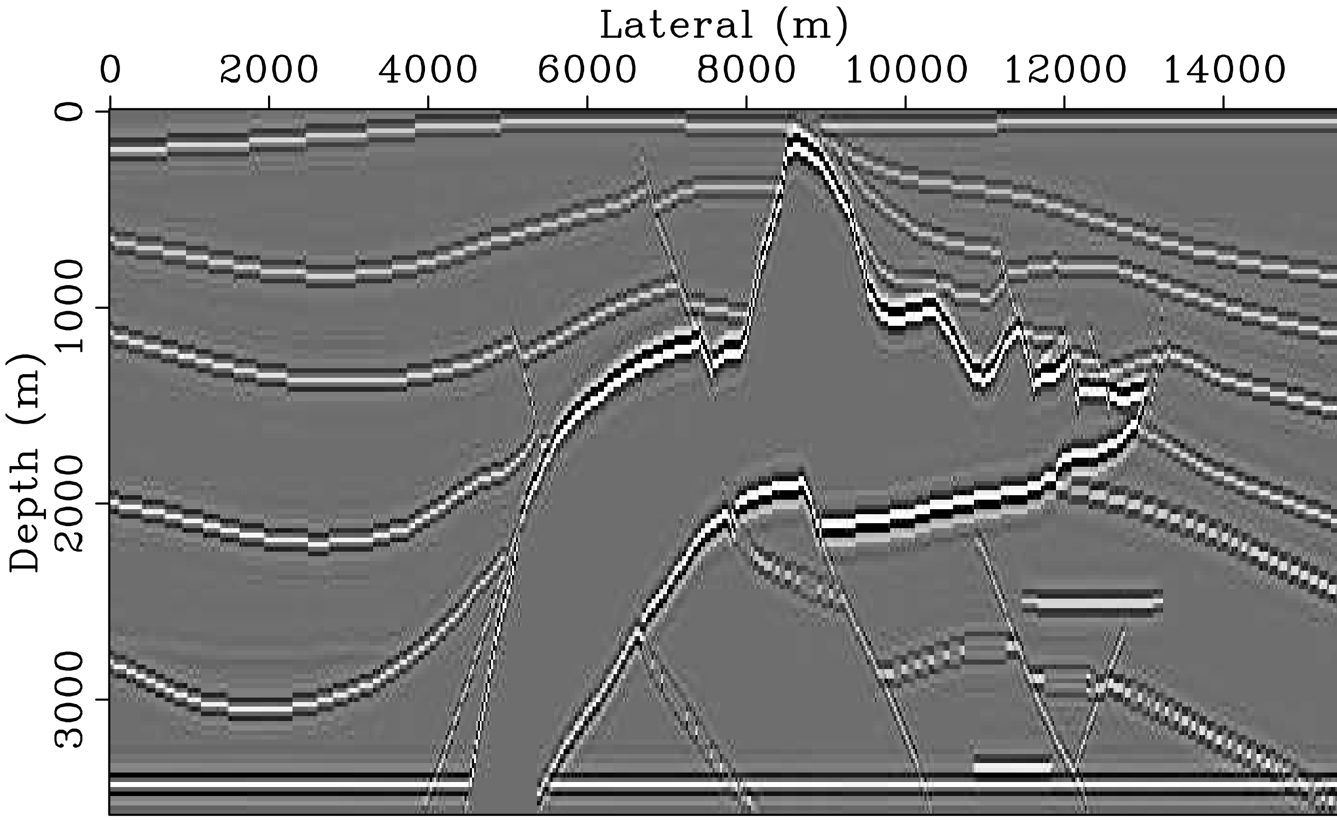
- "broad-band" half-integrated wavelet [5-60 Hz]
- 324 shots, 176 receivers, shot at 48 m
- 5 s of data

Modeling operator

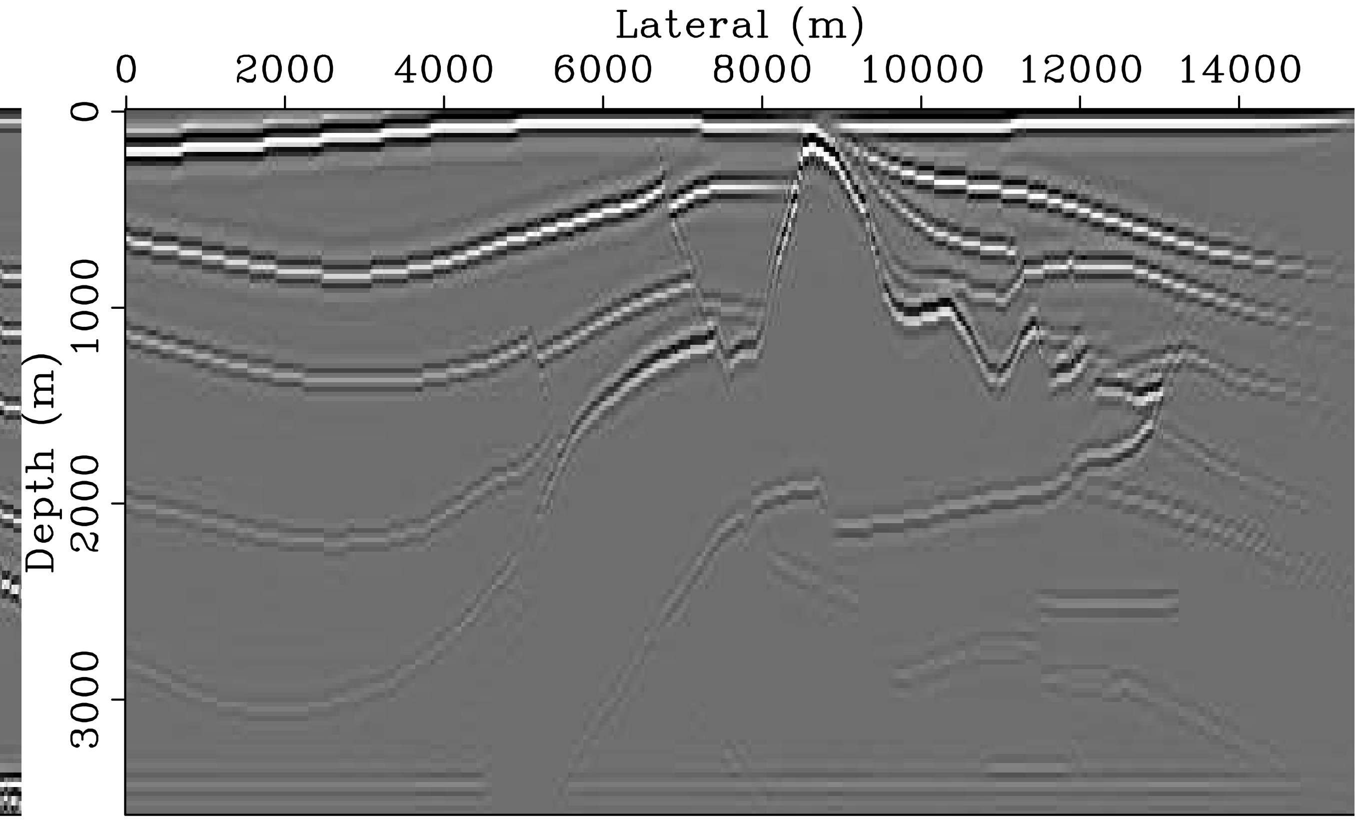
- Reverse-time migration with optimal check pointing (Symes '07)
- 8000 time steps
- **linearized** modeling 64, and migration 294 minutes on 68 CPU's

Scaling required 1 extra migration-demigration

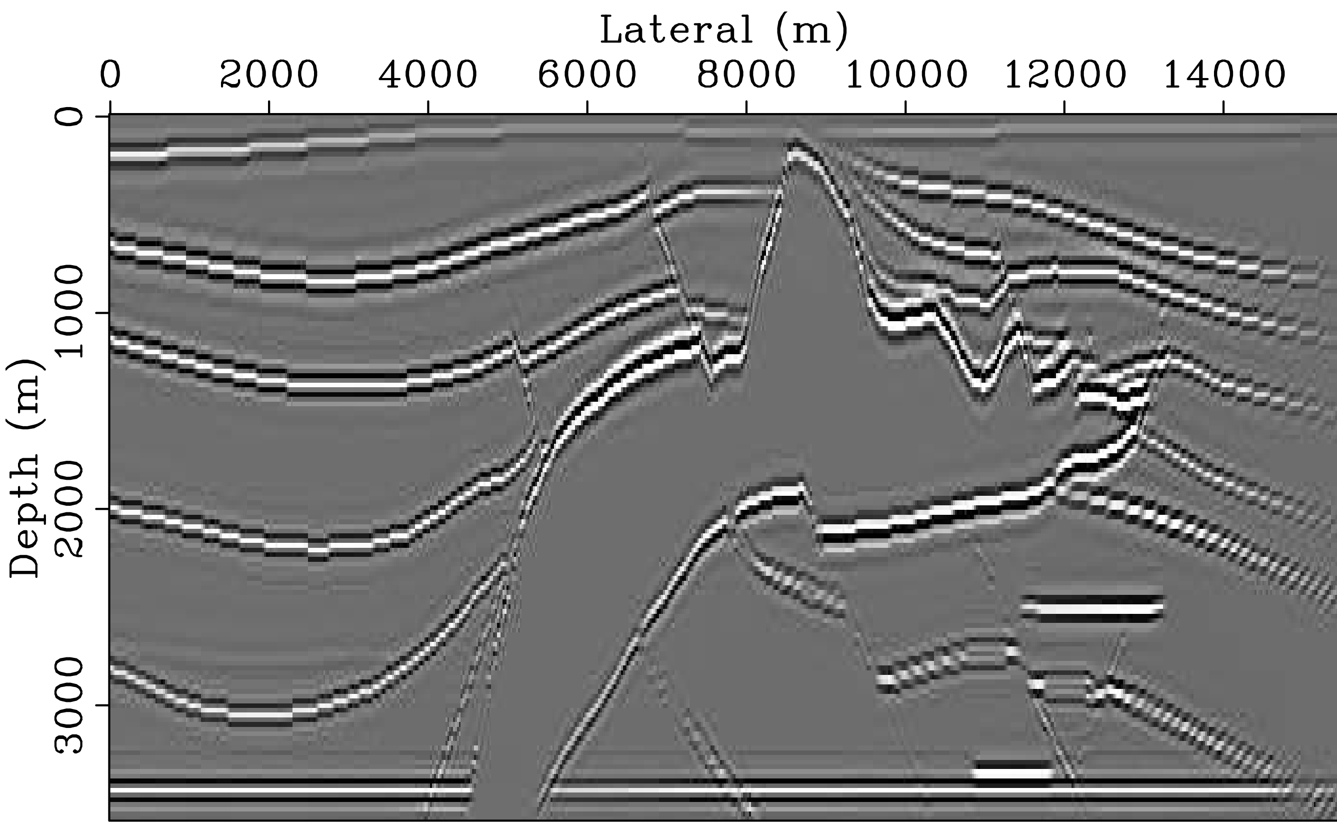

Lateral (m)
0
2000
4000
6000
8000
10000
12000
14000
Depth (m)
0
1000
2000
3000
bandpass-filtered reflectivity
migrated image
reference vector
imaged reference vector



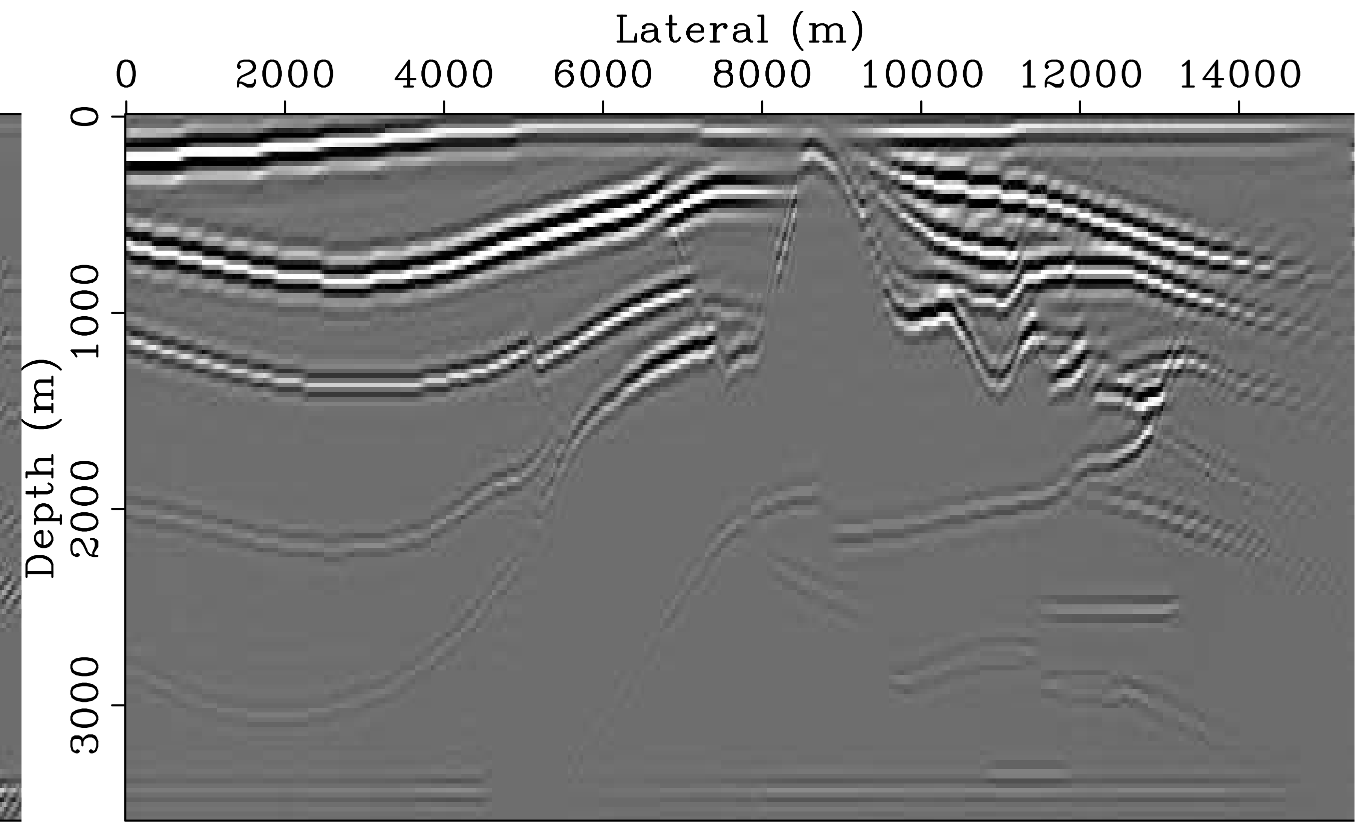
bandpass−filtered reflectivity



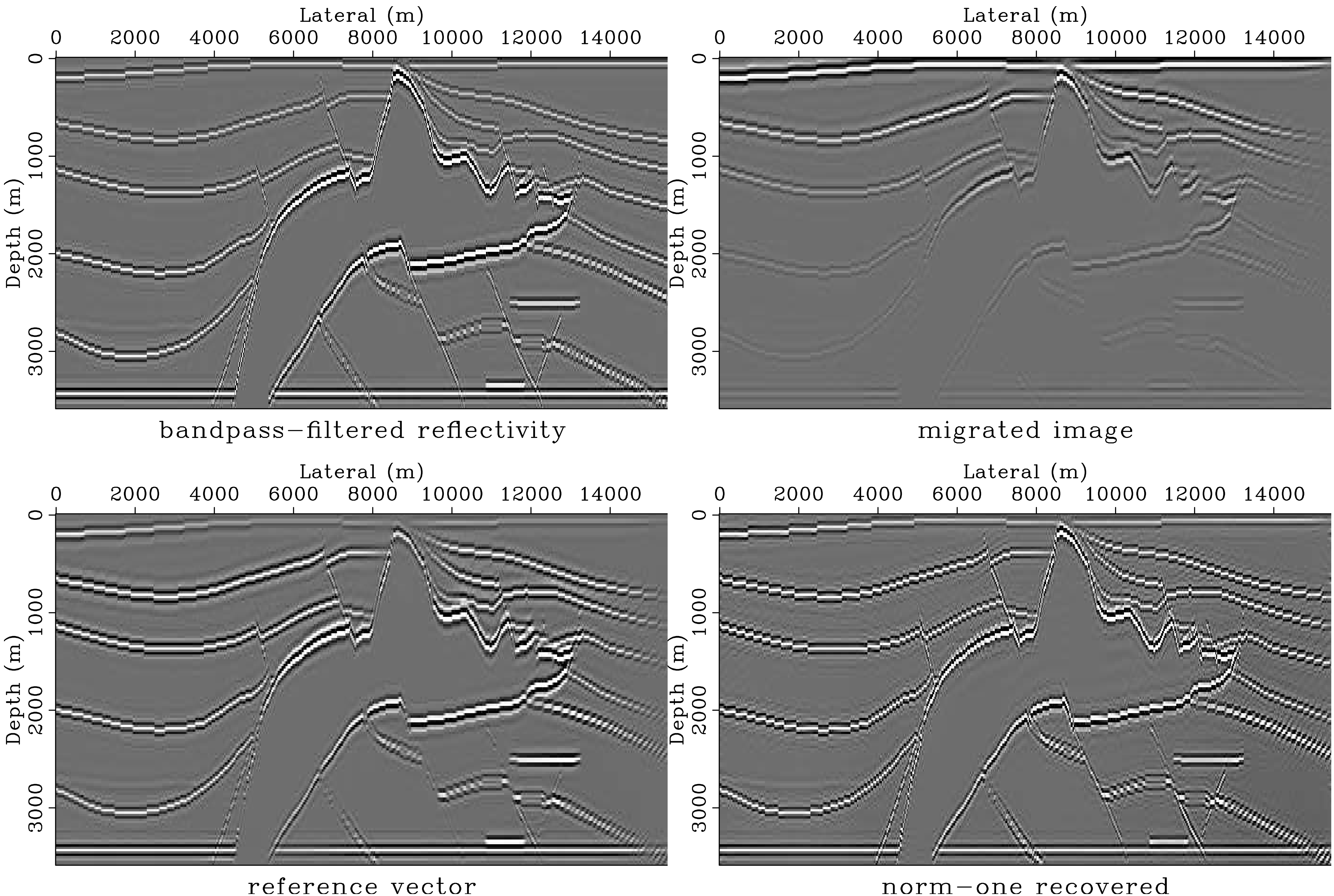
migrated image

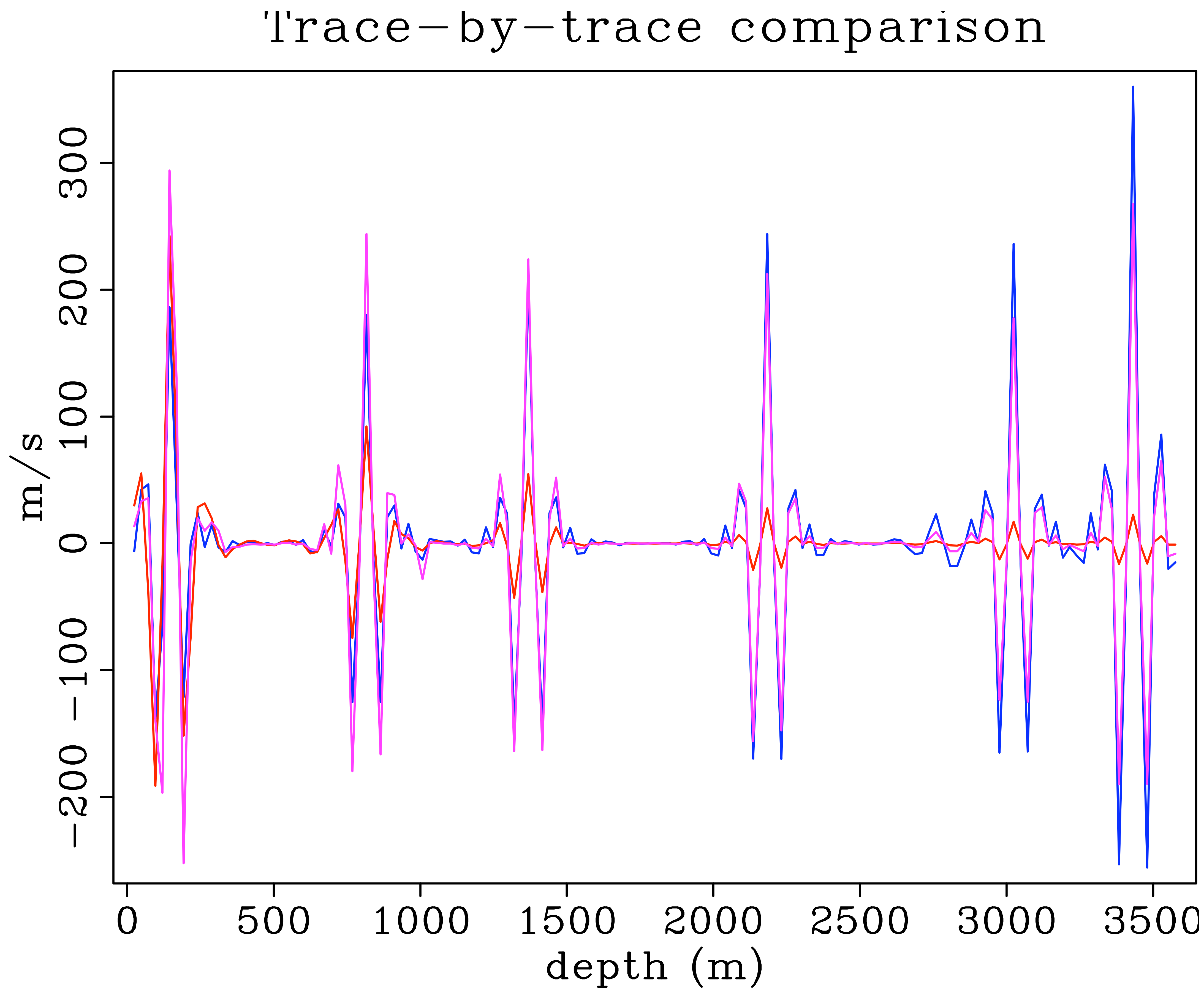


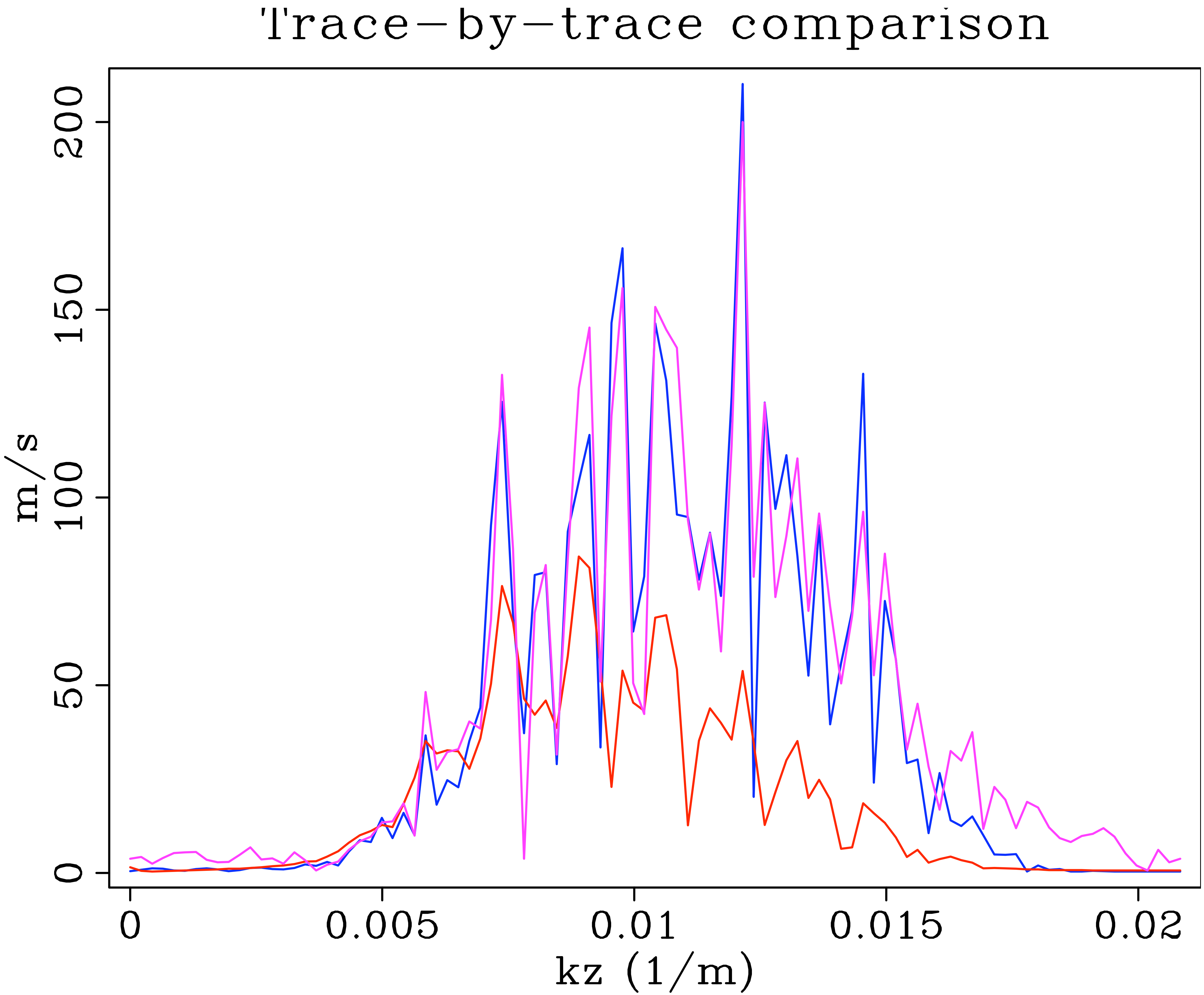
reference vector



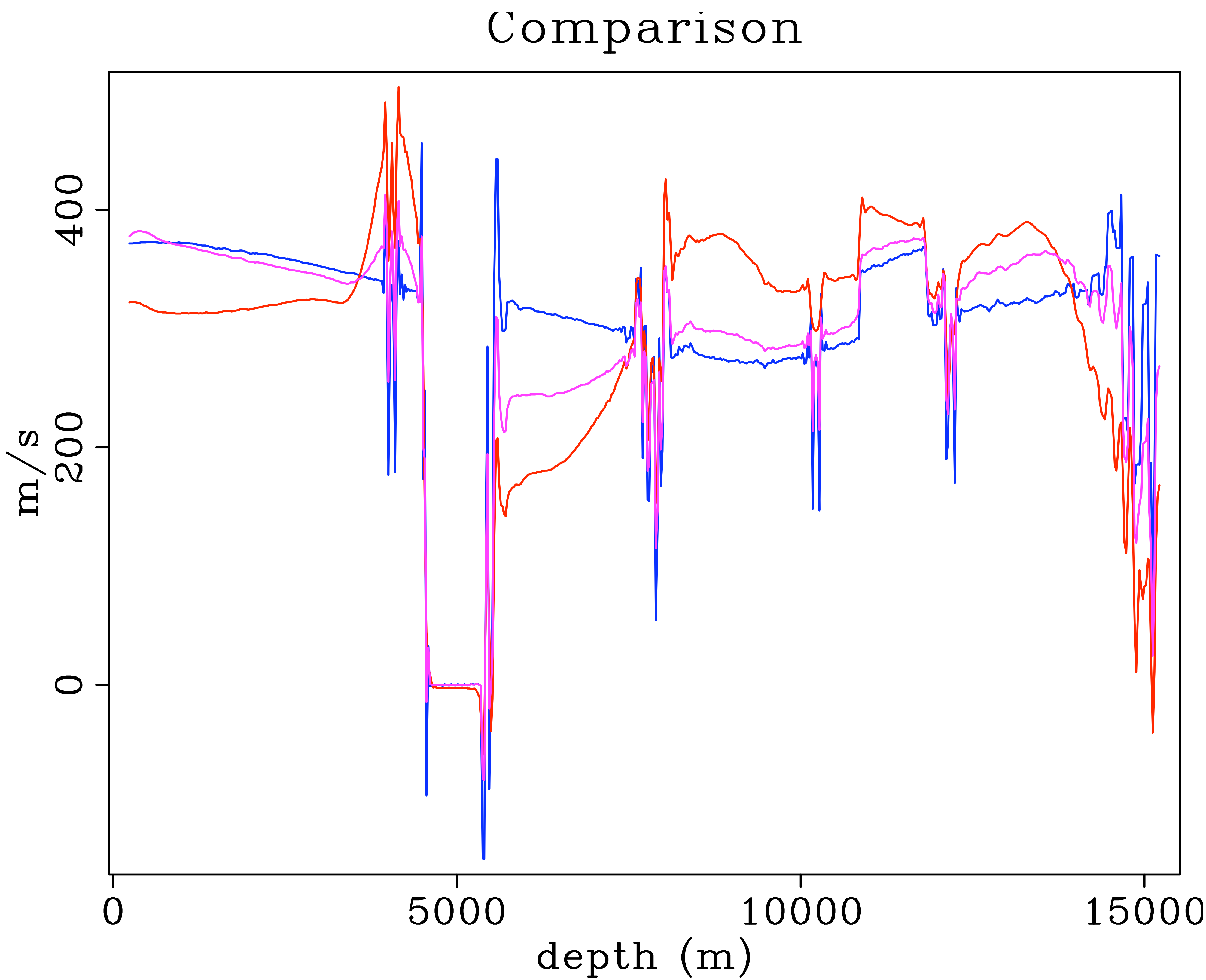
diagonal approximation


Lateral (m)
0 2000 4000 6000 8000 10000 12000 14000
Depth (m)
0 1000 2000 3000
bandpass−filtered reflectivity
Lateral (m)
0 2000 4000 6000 8000 10000 12000 14000
Depth (m)
0 1000 2000 3000
migrated image
Lateral (m)
0 2000 4000 6000 8000 10000 12000 14000
Depth (m)
0 1000 2000 3000
reference vector
Lateral (m)
0 2000 4000 6000 8000 10000 12000 14000
Depth (m)
0 1000 2000 3000
norm−one recovered

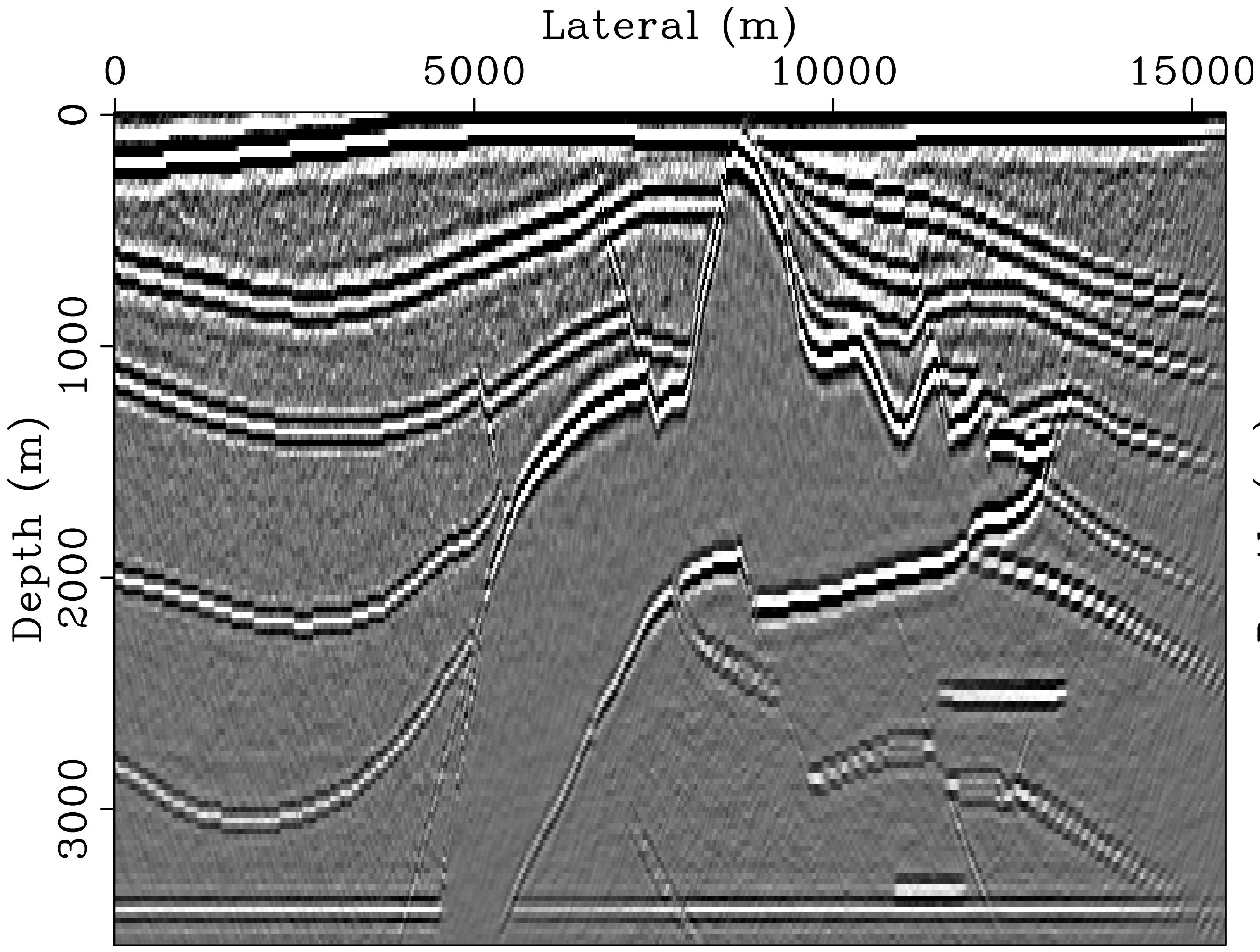

Trace-by-trace comparison
m/s
300
200
100
0
−100
−200
0
500
1000
1500
2000
2500
3000
3500
depth (m)



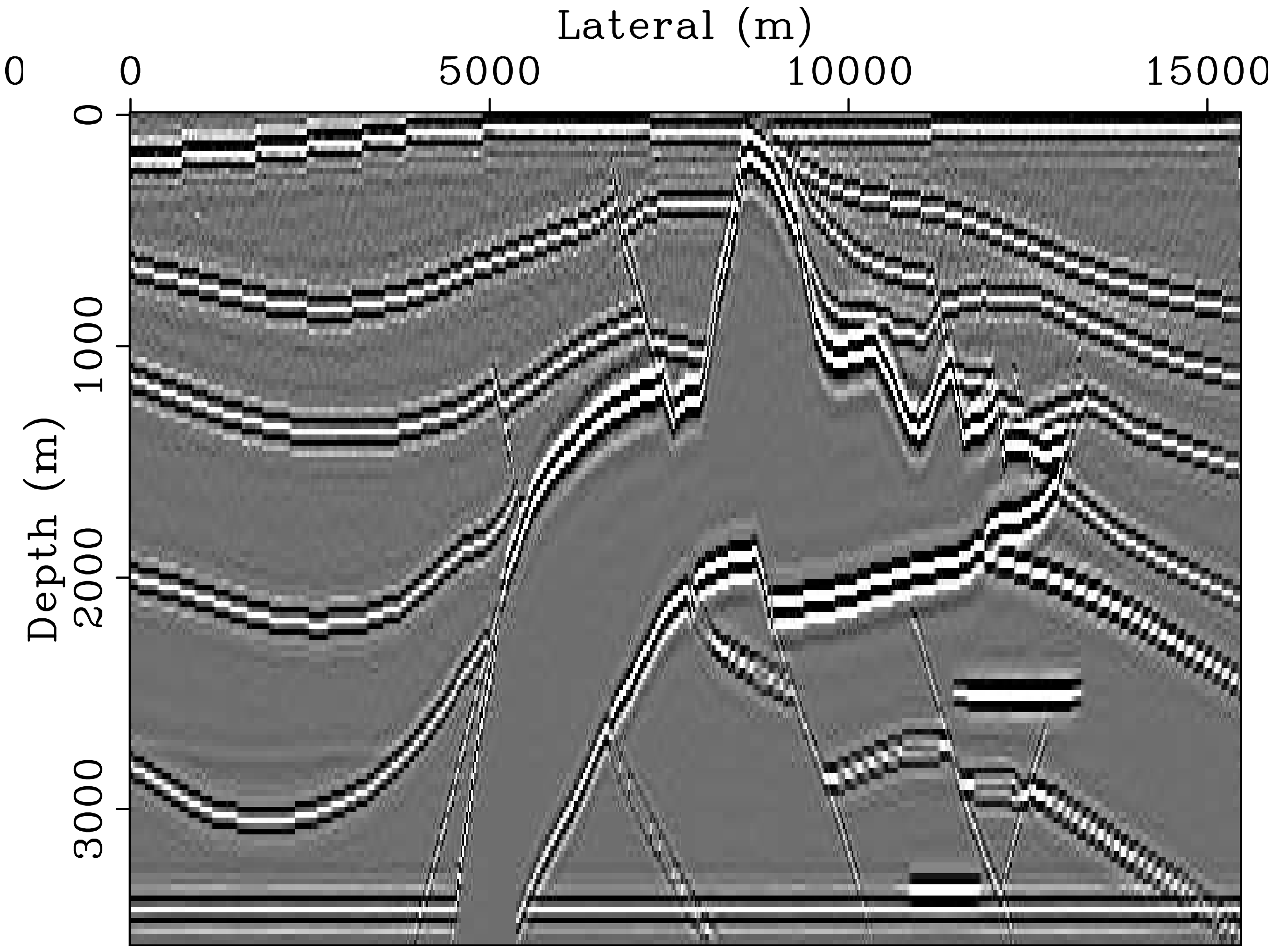
Trace-by-trace comparison
m/s
200
150
100
50
0
0
0.005
0.01
0.015
0.02
kz (1/m)



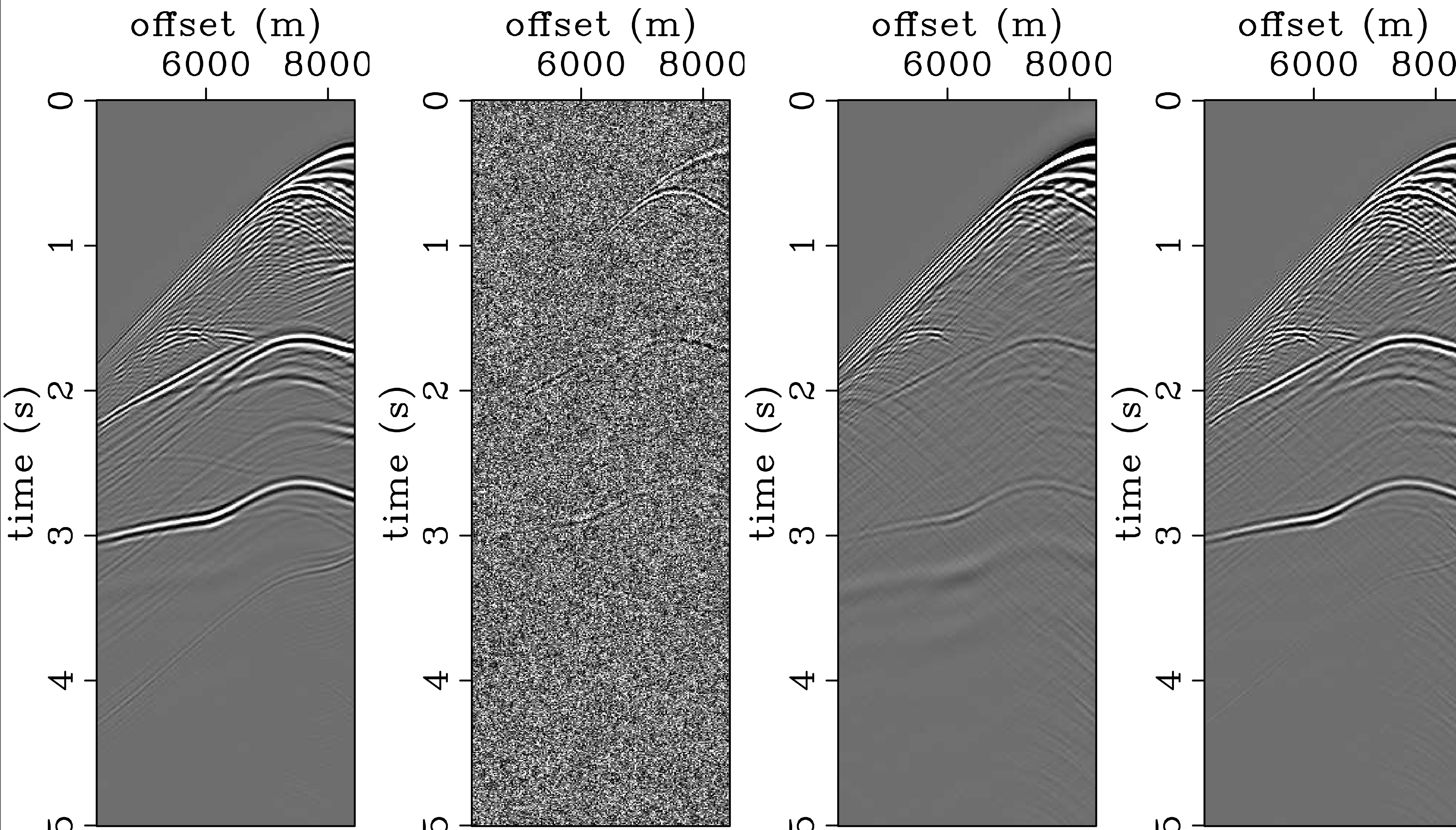
Comparison
m/s
400
200
0
0
5000
10000
15000
depth (m)



Migrated data



Amplitude-corrected & denoised migrated data



Noise-free data

Noisy data (3 dB)

Data from migrated image

Data from amplitude-corrected & denoised migrated image

Example

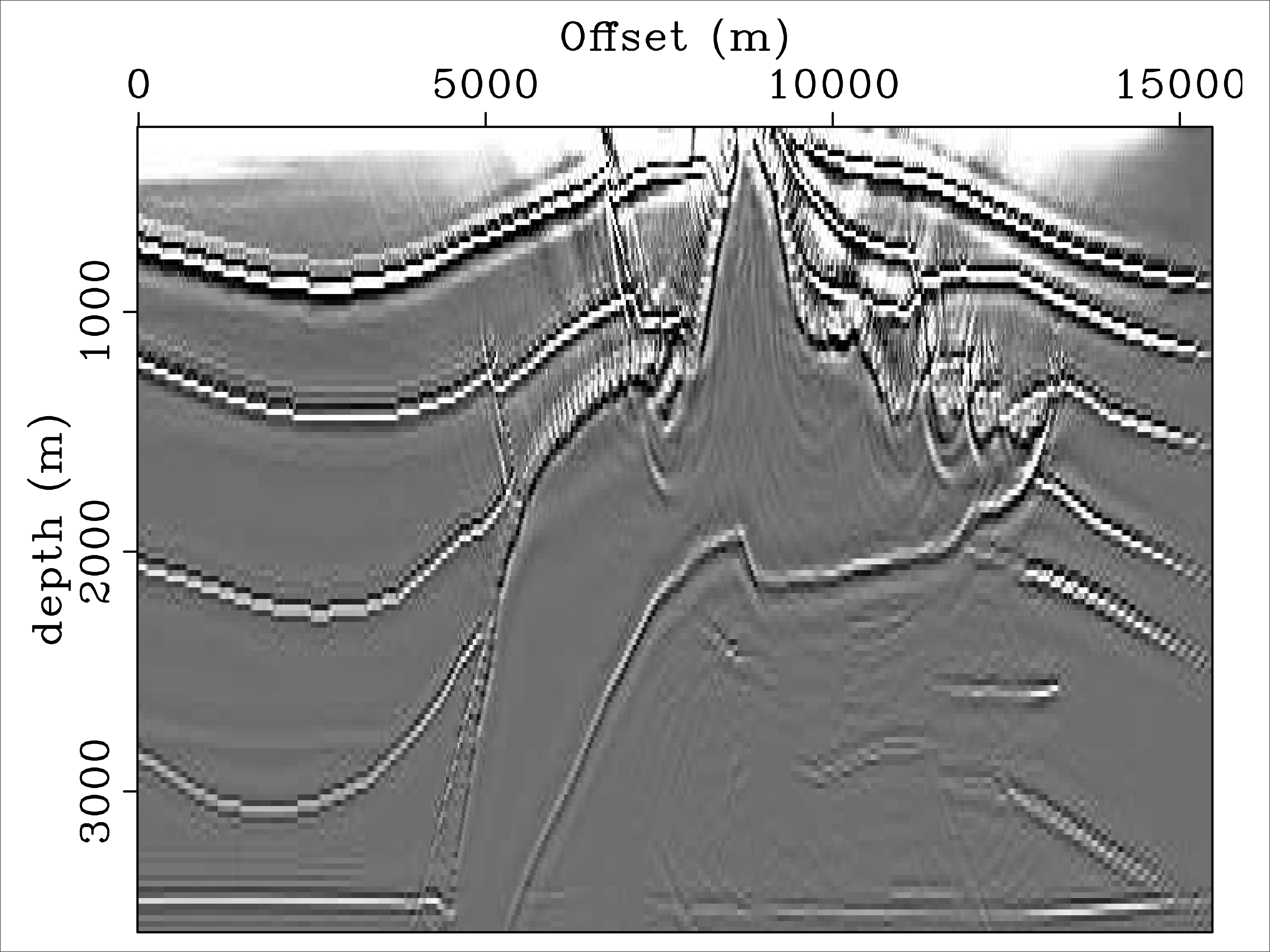
SEGAA' data:

- "broad-band" half-integrated wavelet [5-60 Hz]
- 324 shots, 176 receivers, shot at 48 m
- 5 s of data

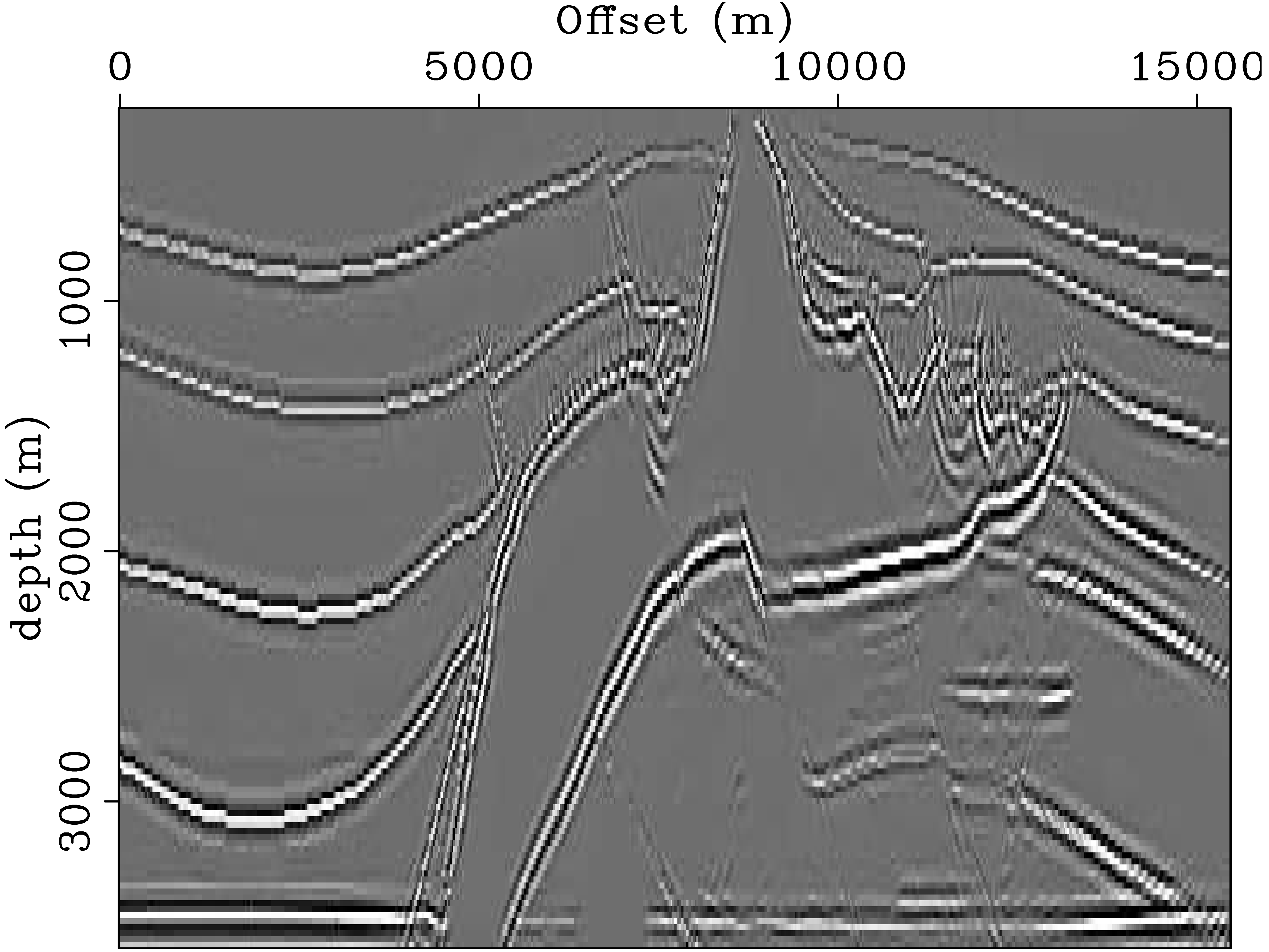
Modeling operator

- Reverse-time migration with optimal check pointing (Symes '07)
- 8000 time steps
- **full** modeling

Scaling required 1 extra migration-demigration



Offset (m)
0
5000
10000
15000
depth (m)
1000
2000
3000


Offset (m)
0
5000
10000
15000
depth (m)
1000
2000
3000

Conclusions

Curvelet-domain scaling

- handles conflicting dips (conormality assumption)
- exploits invariance under the PsDO
- robust w.r.t. noise

Diagonal approximation

- exploits smoothness of the symbol
- uses “neighbor” structure of the curvelet transform

Results on the SEG AA’ show

- recovery of amplitudes beneath the Salt
- successful recovery of clutter
- improvement of the continuity

SLIM Seismic Laboratory for Imaging and Modeling

Acknowledgments

The authors of CurveLab (Demanet, Ying, Candes, Donoho)

Dr. W. W. Symes for his reverse-time migration code

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 (334810-05) of F.J.H. This research was carried out as part of the SINBAD project with support, secured through ITF (the Industry Technology Facilitator), from the following organizations: BG Group, BP, Chevron, ExxonMobil and Shell.