Full-waveform inversion from compressively recovered updates

Xiang Li and Felix J. Herrmann

SLIM
University of British Columbia

SLIM

Motivation

Curse of dimensionality for d>2

- *Exponentially* increasing *data volumes*
- *Helmholtz* requires *implicit* solvers to address *bandwidth*
- Computational complexity grows *linearly* with # RHS's
- Makes *computation* of the misfit functional & gradients prohibitively *expensive*

Wish list

An *inversion* technology that

- is based on a *time-harmonic* PDE solver, which is easily *parallelizable*, and *scalable* to 3D
- does *not* require *multiple* iterations with *all* data
- removes the *linearly* increasing costs of *implicit* solvers for increasing numbers of frequencies & RHS's
- produces *high-resolution* inversion results

Key technologies

Simultaneous sources & phase encoding [Beasley, '98, Berkhout, '08] [Morton, '98, Romero, '00]

- supershots [Krebs et.al., '09, Operto et. al., '09, Herrmann et.al., '08-10']

Stochastic optimization & machine learning [Bertsekas, '96]

- stochastic gradient decent

Compressive sensing [Candès et.al, Donoho, '06]

- *sparse* recovery & *randomized* subsampling

[Nemeth et. al. '99]

Imaging

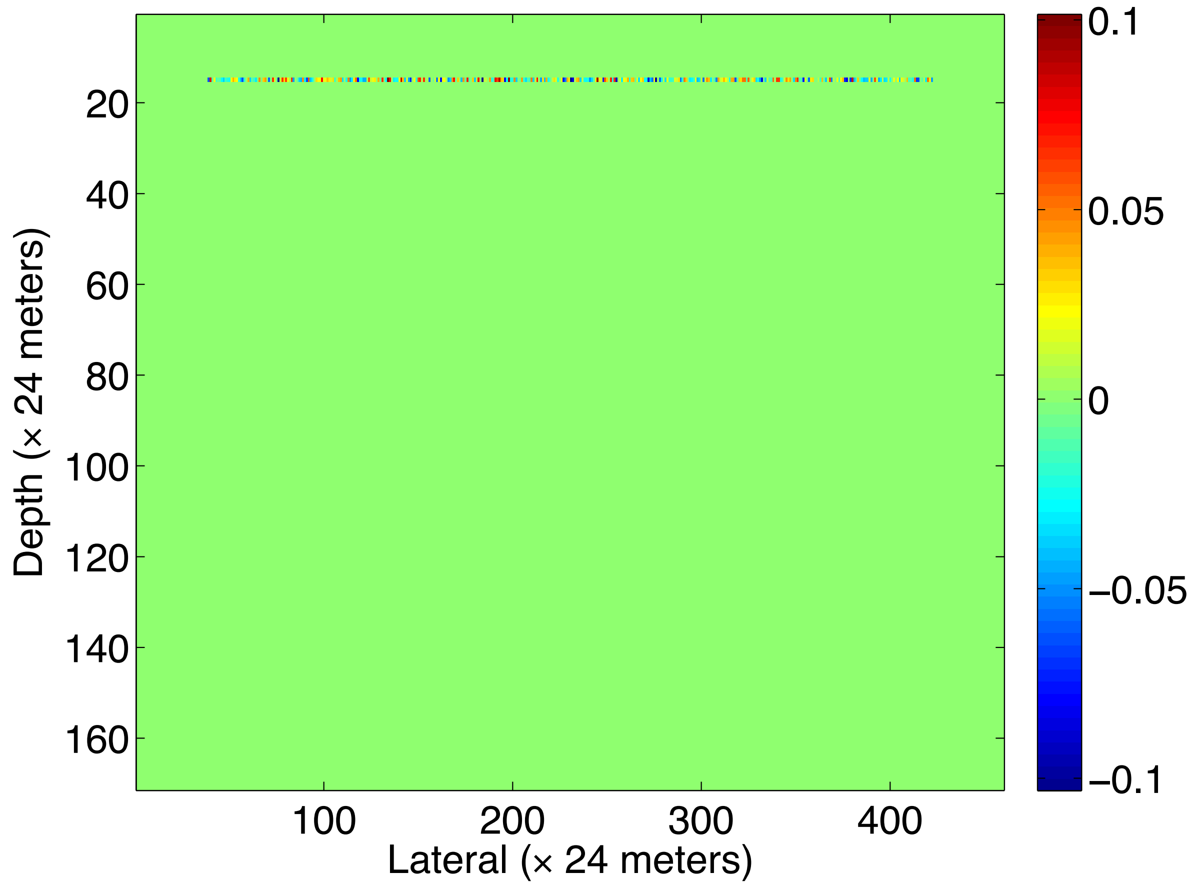
Least-squares migration:

$$\delta\widetilde{\mathbf{m}} = \arg\min_{\delta\mathbf{m}} \frac{1}{2}\|\delta\mathbf{d} - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}_0;\mathbf{Q}]\delta\mathbf{m}\|_2^2$$

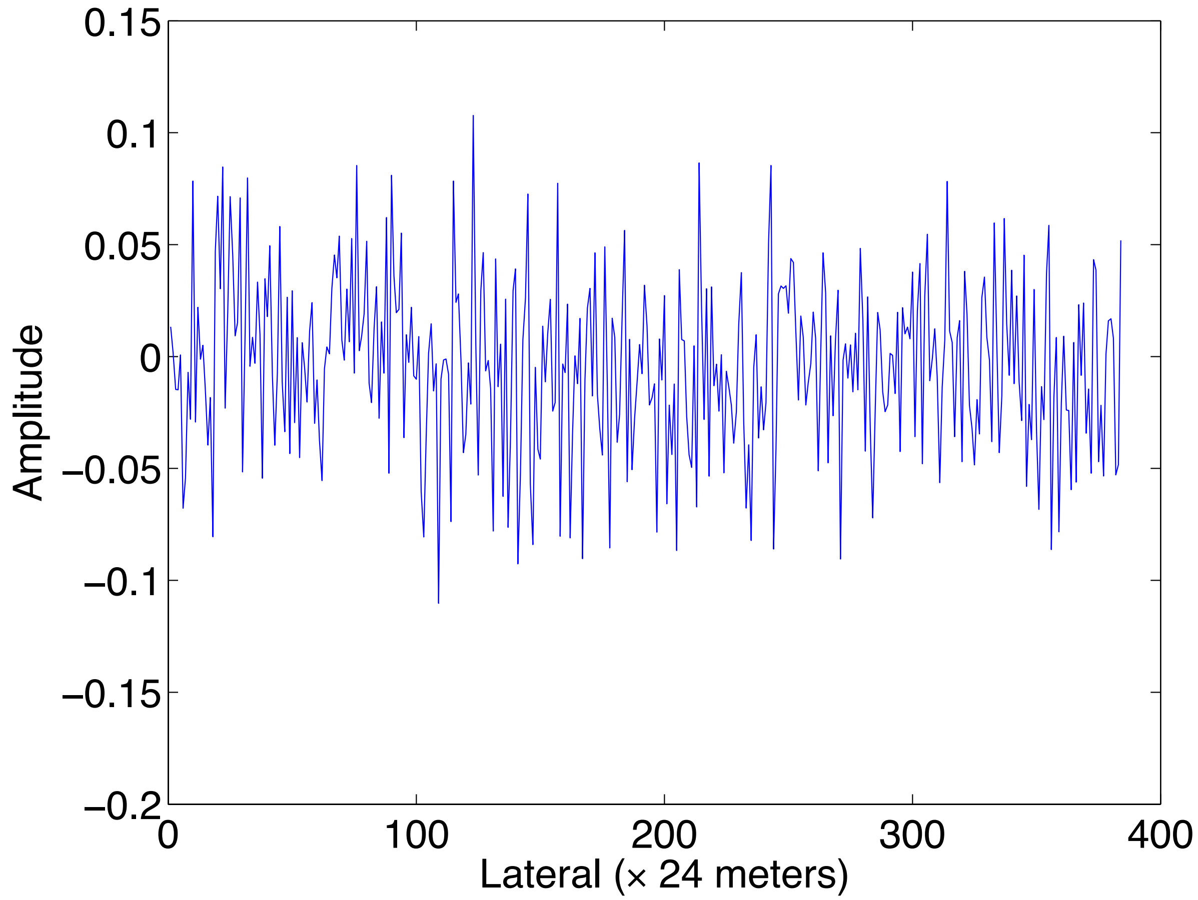
$$\begin{aligned}
\delta\mathbf{d} &= \text{Multi-source multi-frequency data residue} \\
\boldsymbol{\nabla}\boldsymbol{\mathcal{F}}[\mathbf{m}_0;\mathbf{Q}] &= \text{Linearized Born-scattering operator} \\
\mathbf{m}_0 &= \text{Background velocity model} \\
\mathbf{Q} &= \text{Sources} \\
\delta\tilde{\mathbf{m}} &= \text{image}
\end{aligned}$$

Phase encoding

Simultaneous source



Randomized amplitudes along the shot line



Create *supershot* via *superposition*

[Morton, '98, Romero, '00]

Simultaneous shot at 5 Hz

Sequential-source wavefield

Simultaneous-source wavefield

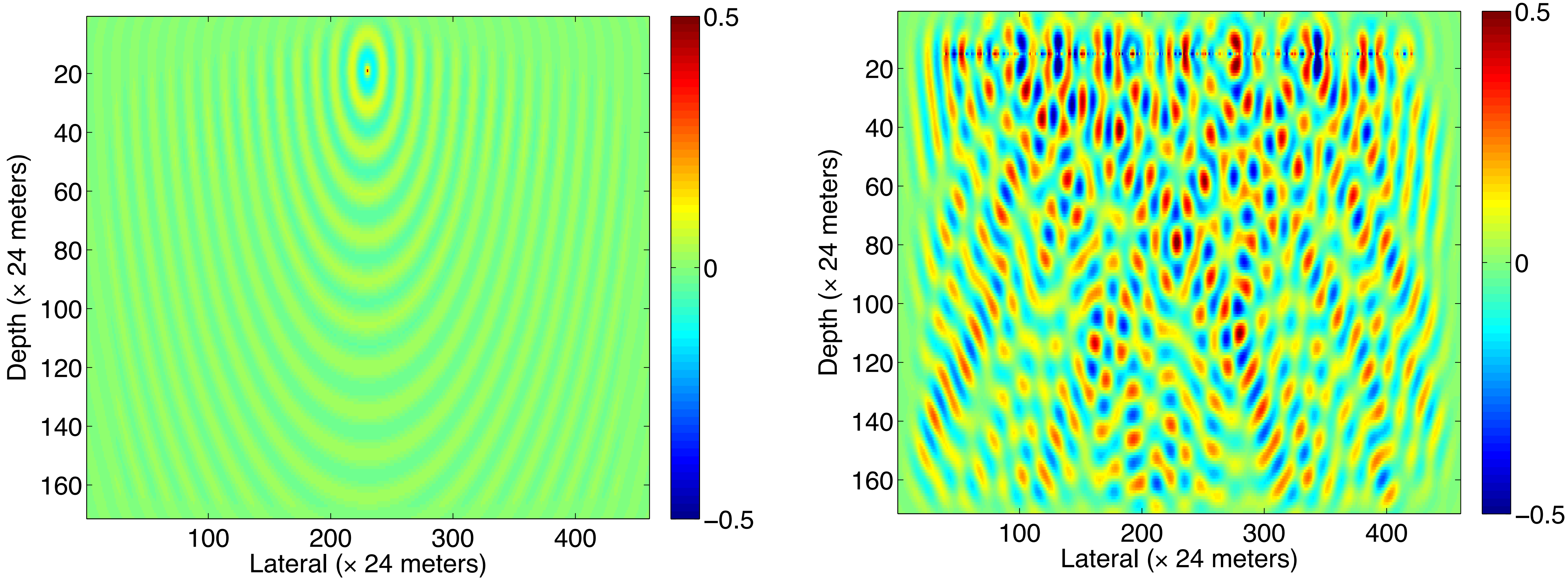
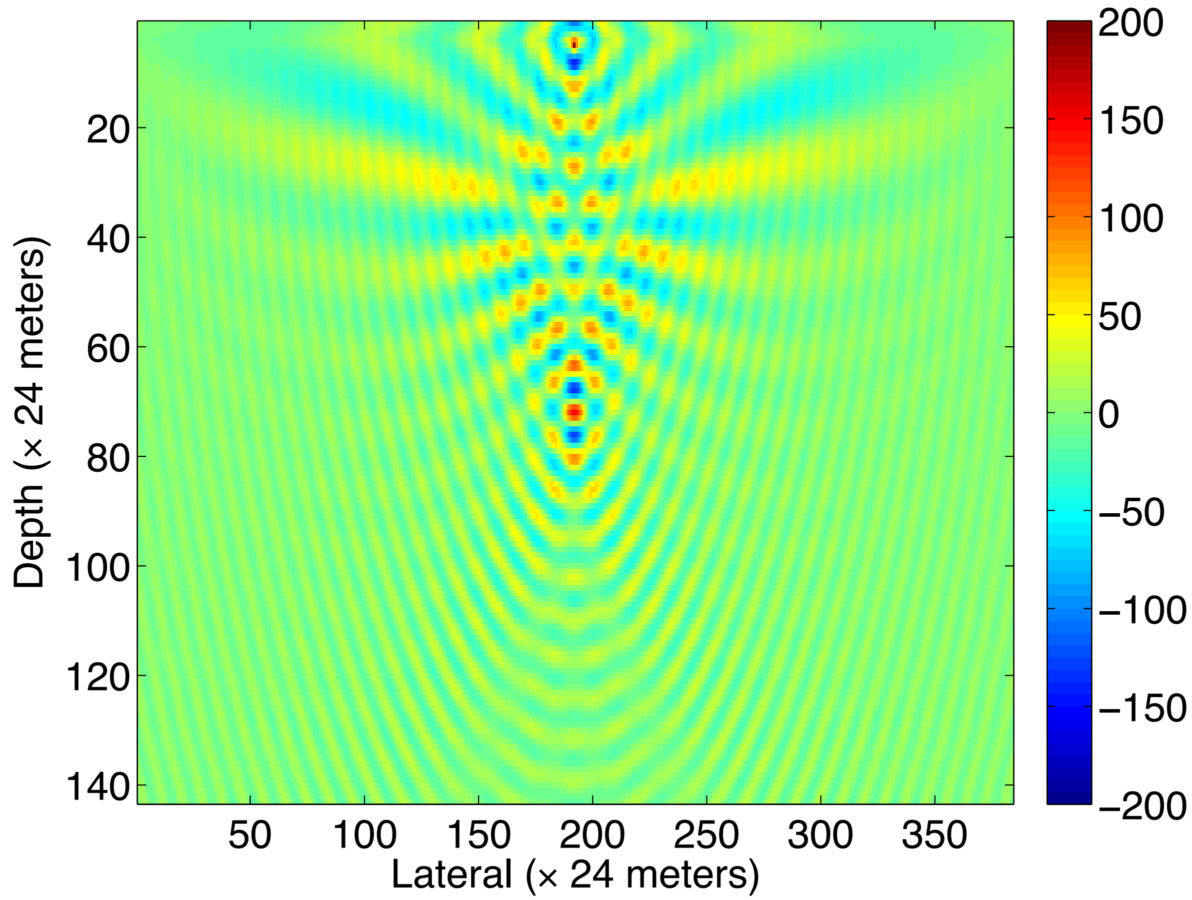
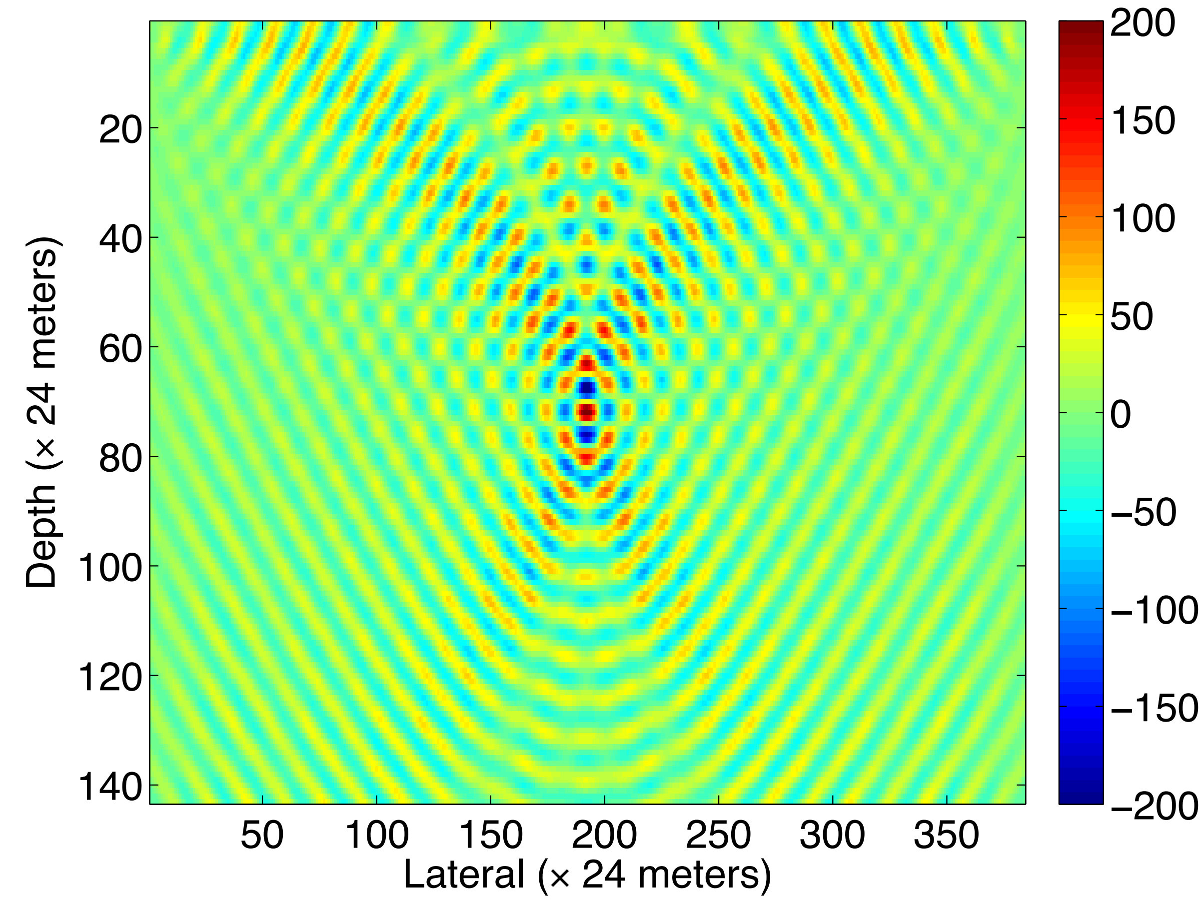


Image at 5 Hz

Sequential-source image



Simultaneous-source image

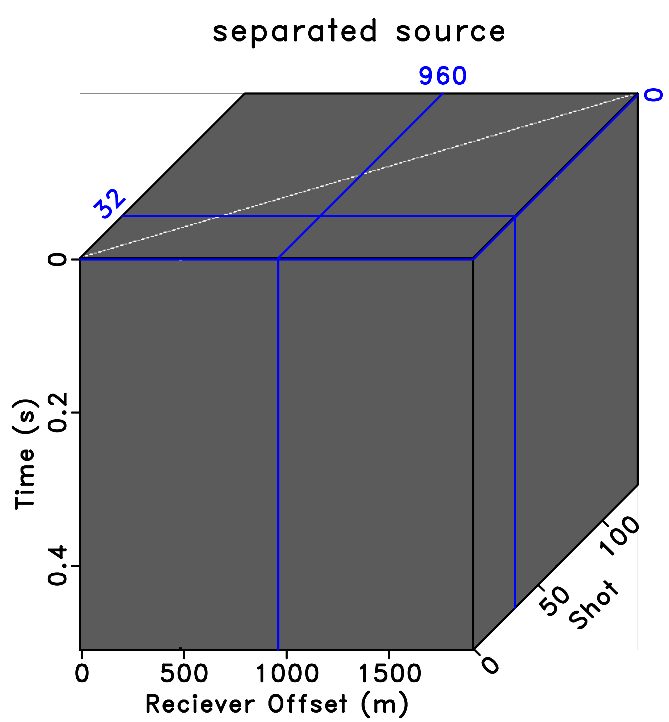


[Morton, ’98, Romero, ’00]

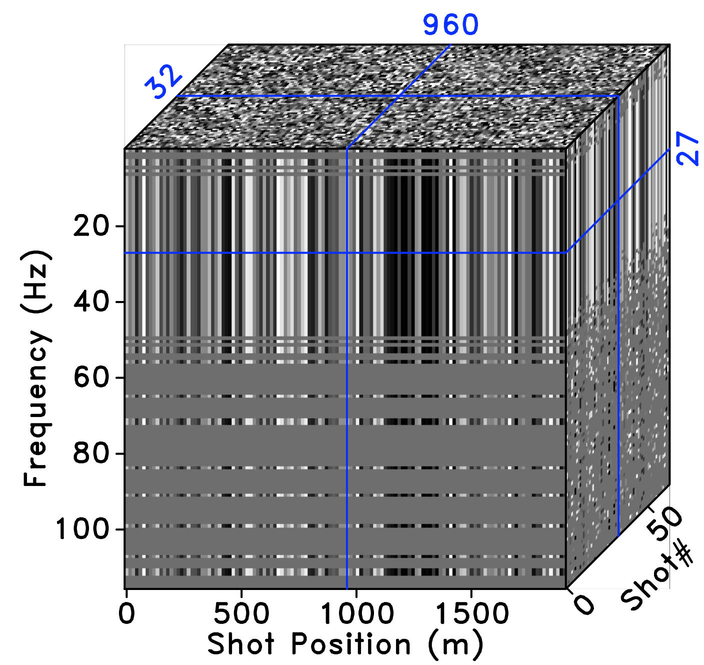
[Herrmann et. al. ’08-’10]

Supershot

adapted from Herrmann et. al. ,09



$\mathbf{Q}$



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

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

Phase encoding

Least-squares migration:

$$\delta\widetilde{\mathbf{m}} = \arg\min_{\delta\mathbf{m}} \frac{1}{2}\|\delta\underline{\mathbf{d}} - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}_0;\underline{\mathbf{Q}}]\delta\mathbf{m}\|_2^2$$

$\delta\underline{\mathbf{d}}$ = Simultaneous-source data residue

$\underline{\mathbf{Q}}$ = Simultaneous sources

[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

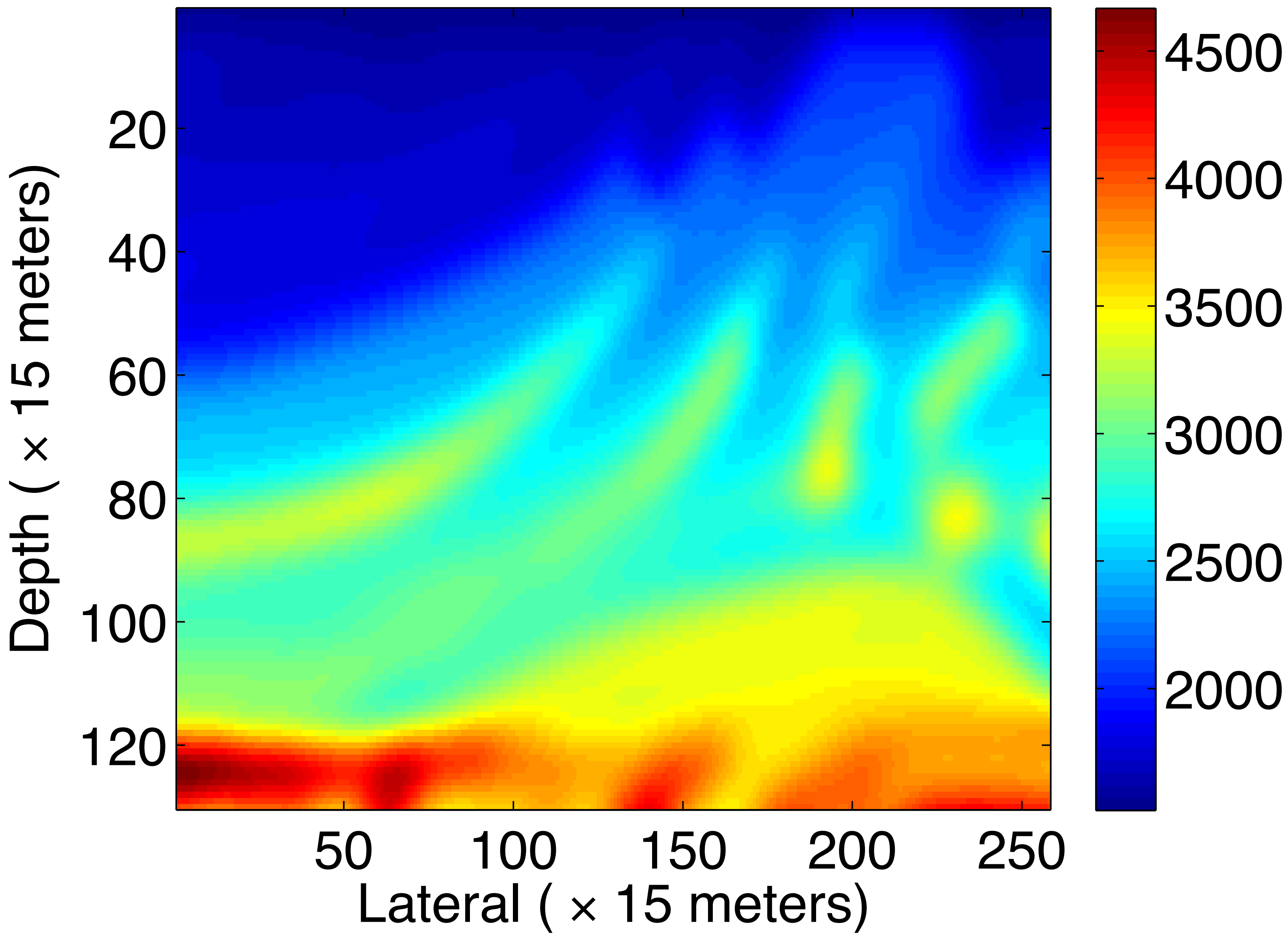
$$n_{PDE}^{\ell_1} \times K \ll n_{PDE}^{\ell_2} \times n_f \times n_s$$

Experiment

Linearized *sparsity promoting* least-squares migration

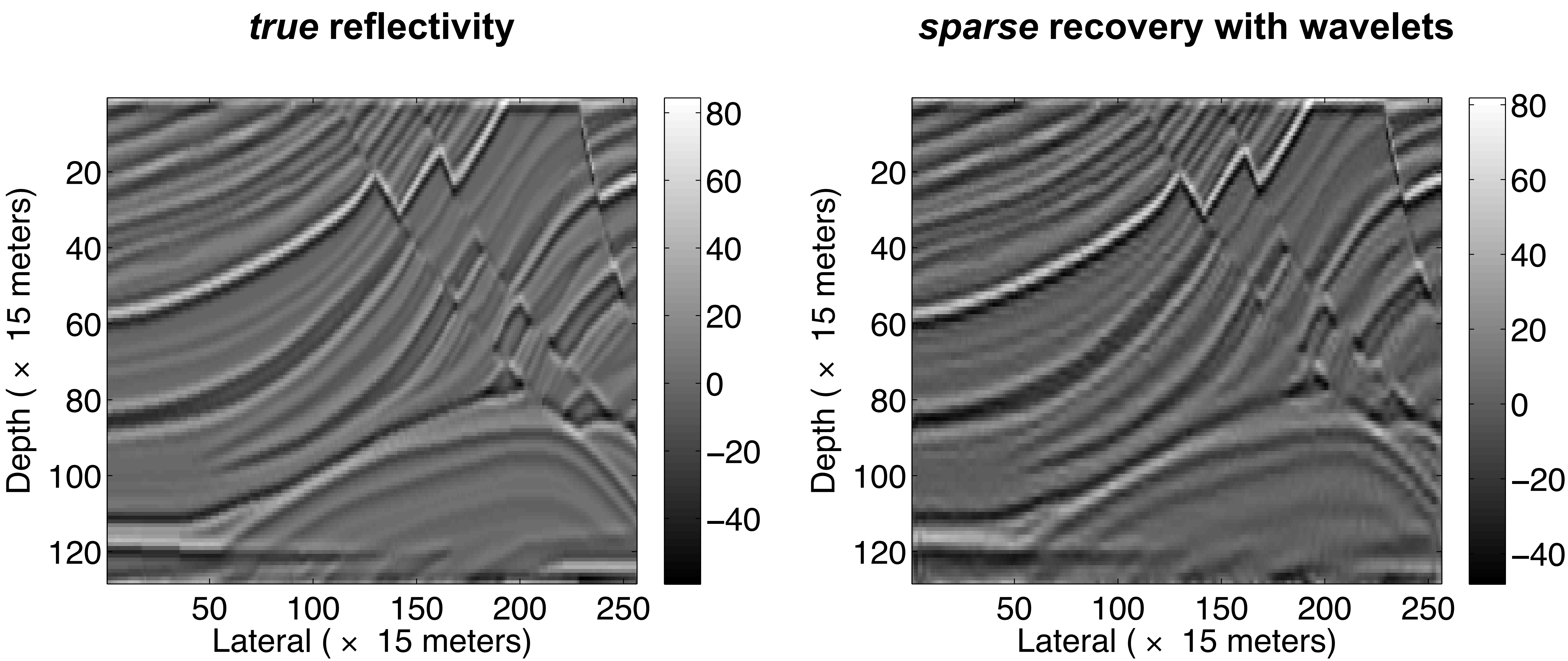
- Marmousi model (128x256) with grid size 15 m
- use different
  - # of simultaneous shots (50, 20,10)
  - # of frequencies (10, 10, 5)

Initial model



Linearized sparse inversion

30 simultaneous shots 10 random frequencies

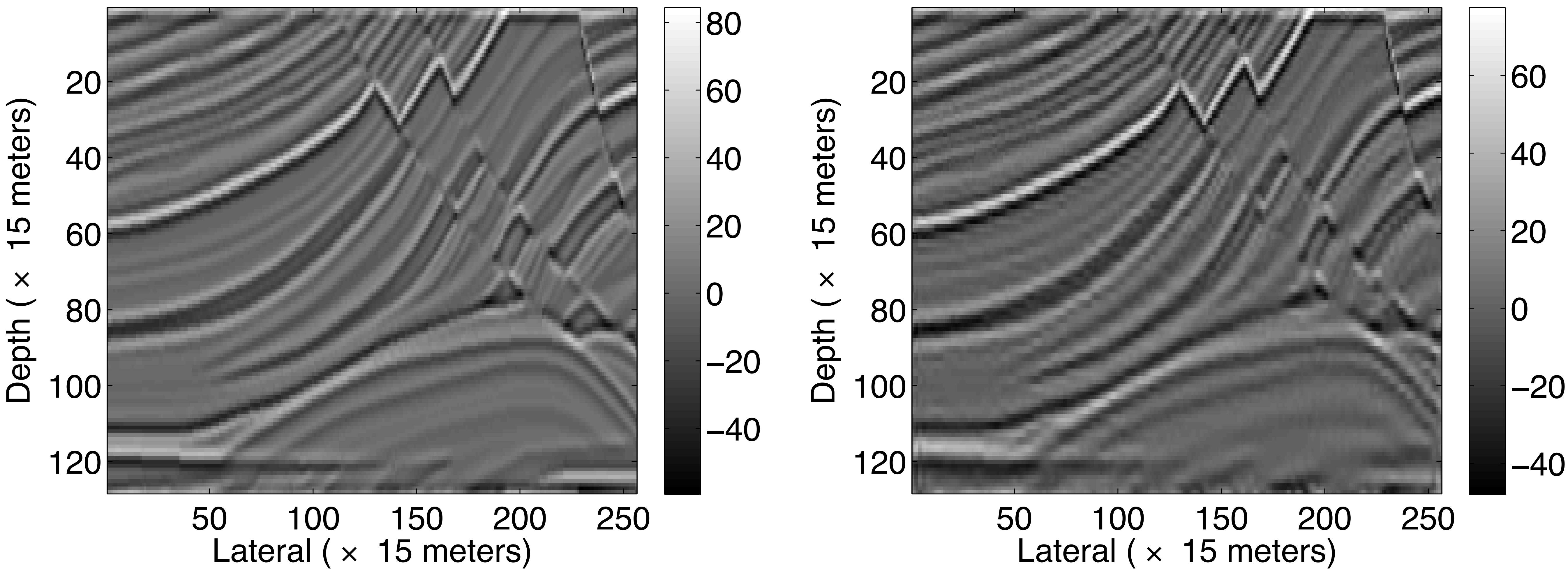


Speed up: X 86

Linearized sparse inversion

20 simultaneous shots 10 random frequencies

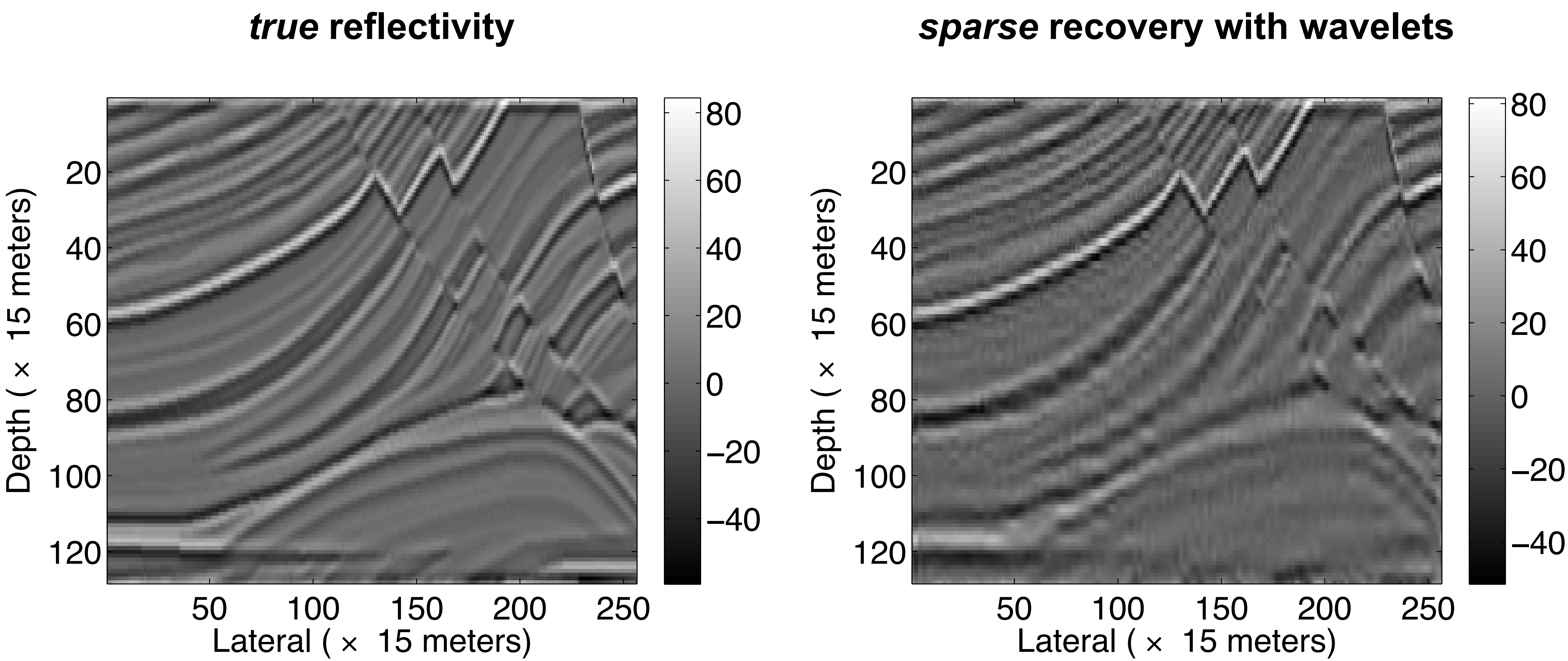
***true* reflectivity** | ***sparse* recovery with wavelets**



Speed up: x129

Linearized sparse inversion

10 simultaneous shots 5 random frequencies



Speed up: x517

Linearized sparse inversion

Subsample ratio	0.015	0.006	0.002
n'_f/n'_s	recovery error (dB)		
5	**17.44** (1.32)	**11.66** (0.78)	**6.83** (-0.14)
1	**17.53** (1.59)	**11.89** (1.05)	**7.19** (0.15)
0.2	**18.22** (1.68)	**12.11** (1.32)	**7.46** (0.27)
Speed up (×)	66	166	500

SNRs for "migration" in parentheses

Observations

Reconstruct model updates

- from *randomized* subsamplings
- with correct amplitudes
  (like Gauss-Newton updates)

Recovery quality depends on *degree* of *subsampling*

Significant speedups attainable...

FWI formulation

Multiexperiment unconstrained optimization problem:

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

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

[Tarantola, 84; Pratt, '98; Plessix, 06]
[Haber, Chung, and Herrmann, '10]

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**

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

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

$k \longleftarrow k+1$;

end

FWI with phase encoding

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
- exploits *linearity* in the *sources* & *block-diagonal* structure of the *Helmholtz system*
- uses *randomized* frequency *selection* and *phase encoding*

[Krebs et.al., '09, Operto et. al., '09 ; Herrmann et. al. '08-'10]

SLIM

Renewals

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

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

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

- motivated by *stochastic approximation* [Nemirovski, '09]
- related to Kaczmarz ('37) method applied by Natterer, '01
- *supersedes ad hoc* approach by Krebs *et.al.*, 2009

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**

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

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

$k \longleftarrow k + 1$;

end

Observations

Stochastic optimization

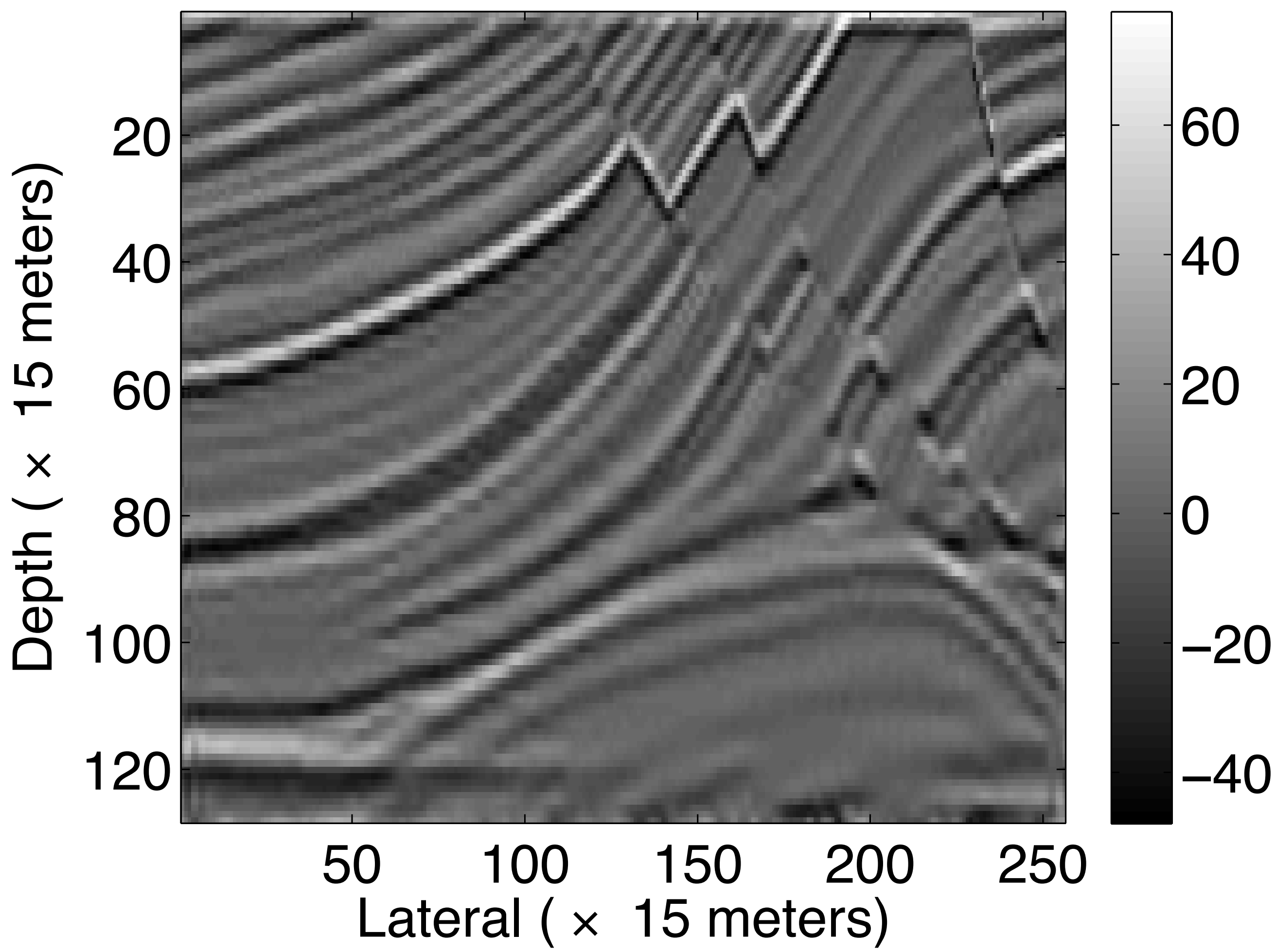
- introduces noisy search directions
- interferences go down *slowly* as batch size *increases*
- requires *averaging* over *previous* model *updates*

Formulation does not exploit *sparsity* on the *model*

[Bertsekas, '96]

[Krebs et.al, '09]

Sparse Linearized inversion



Suggests that *sparsity promotion* recovers search directions *accurately* from *randomized source encoding*

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 the Pareto curve

[Candes et al., ’06; Donoho, ’06]

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

SLIM

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**

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

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

$k \longleftarrow k+1$;

end

[van den Berg & Friedlander, '08]

SLIM

Example

Marmousi model:

- 128x384 with a mesh size of 24 meters
- 384 co-located shots and receivers with offset = 3 X depth
- 2.4s recording time

Explicit Time-harmonic Helmholtz solver

- 9-point finite difference
- Absorbing boundary condition

Example

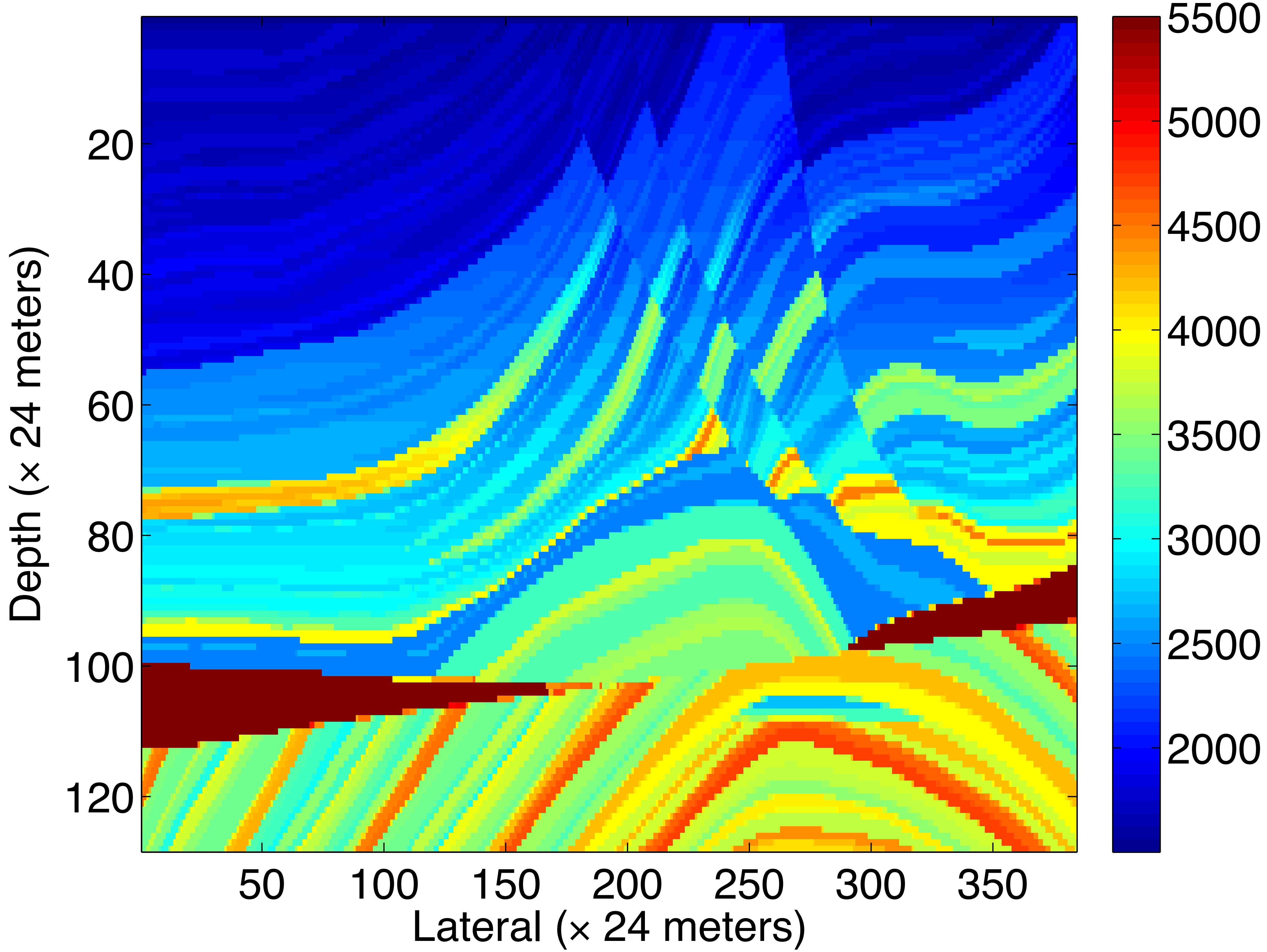
FWI specs:

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

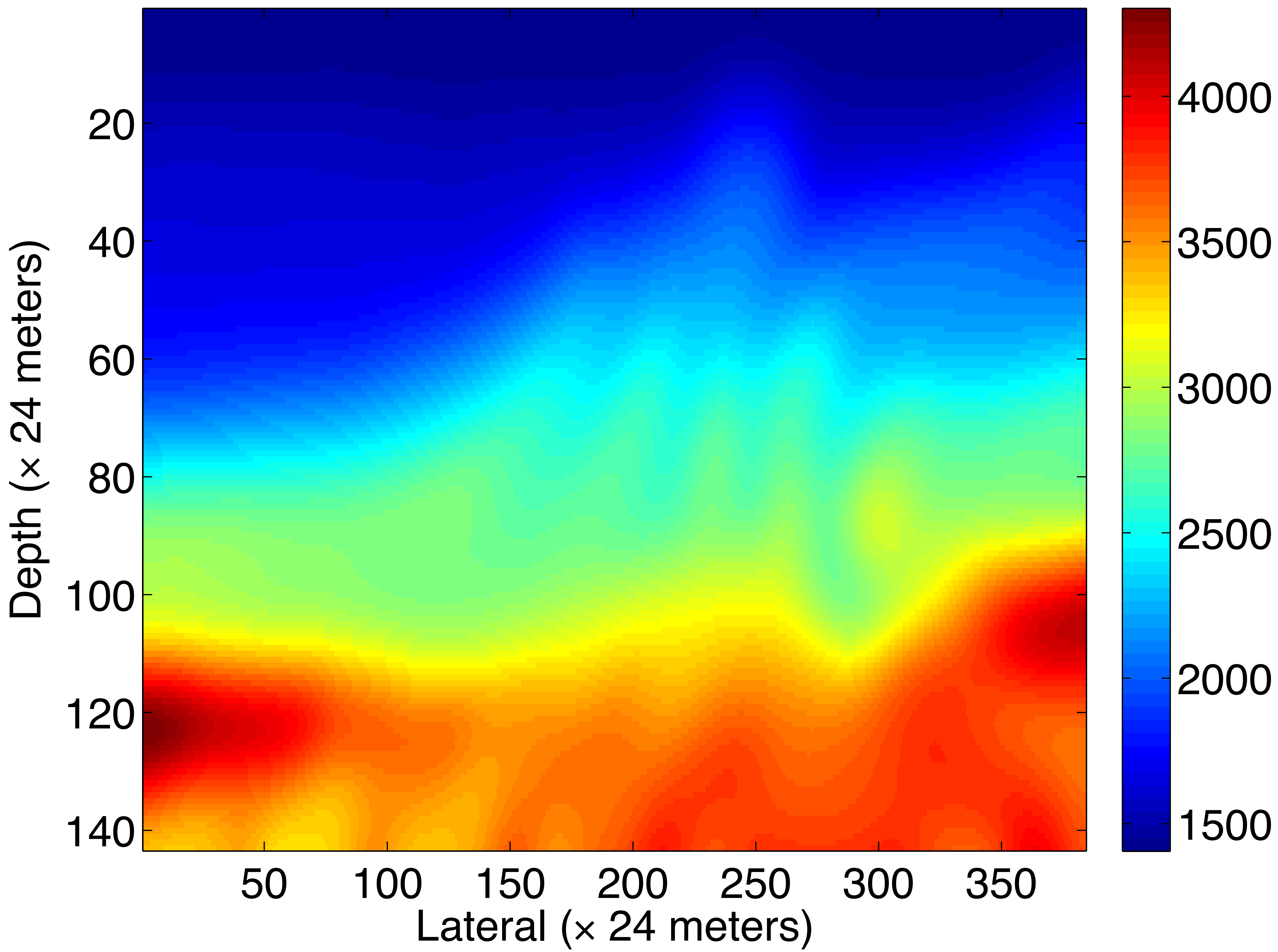
$$K = 10 \times 15 \ll 100 \times 384$$

SLIM

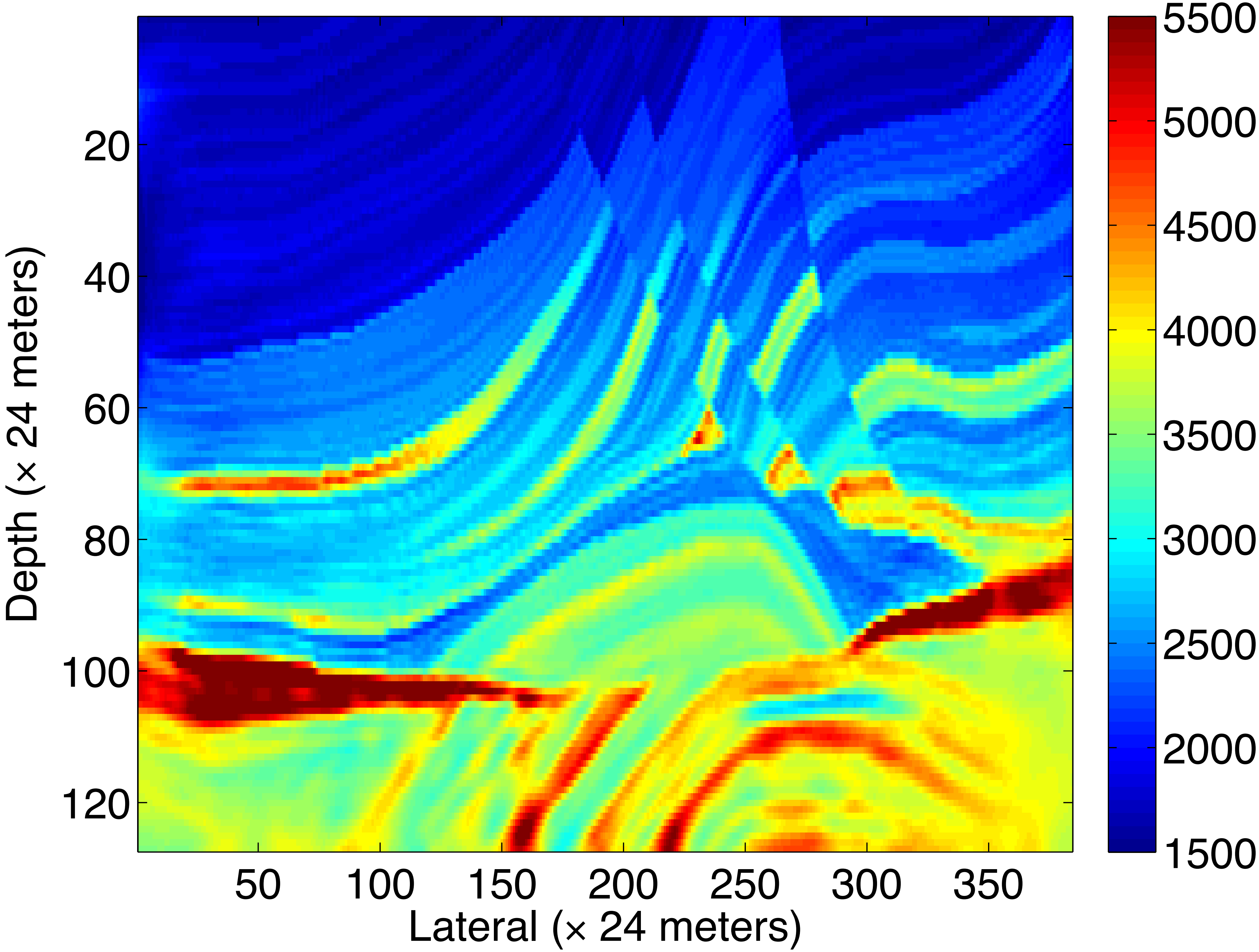
True model



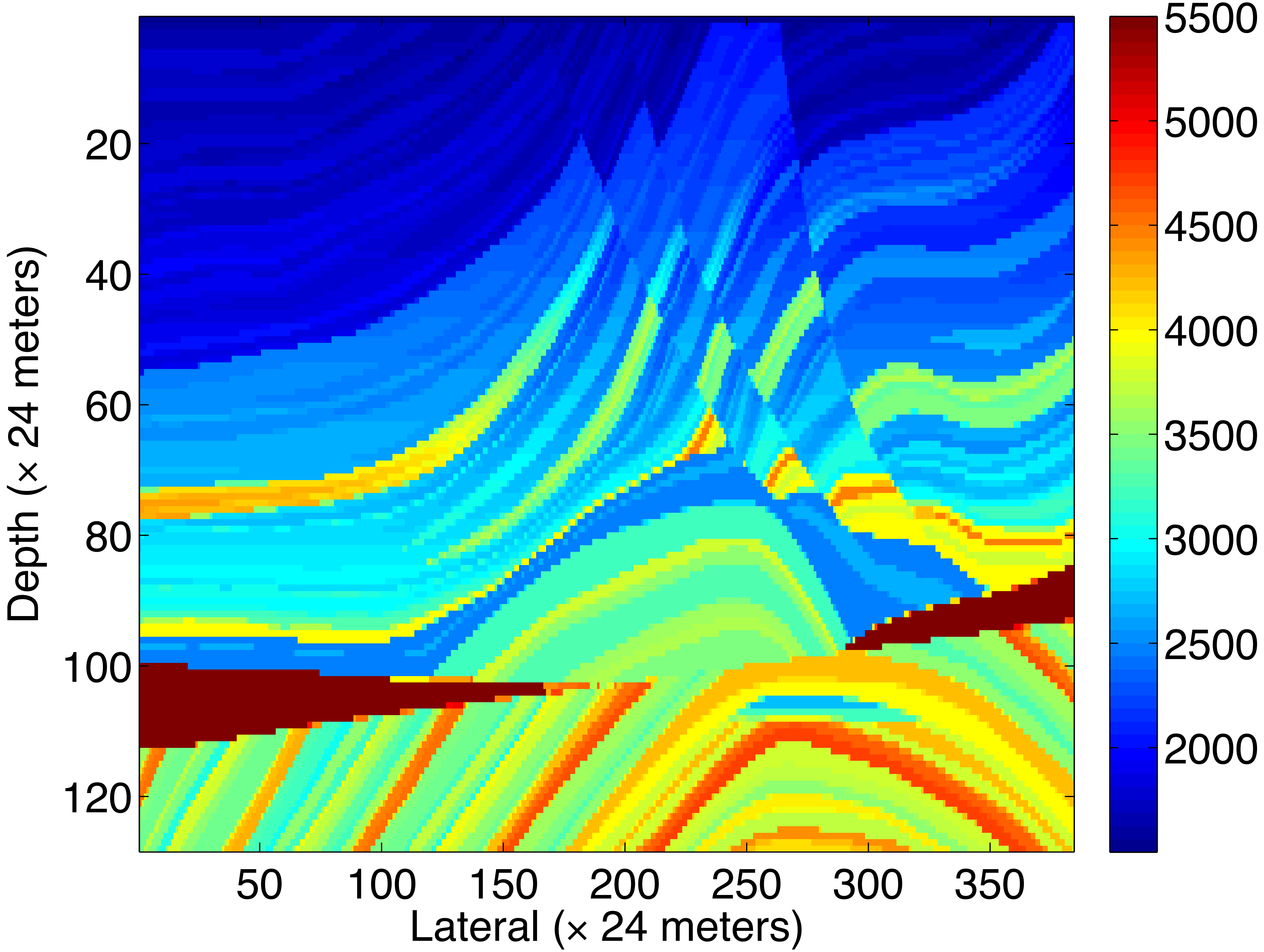
Initial model



Inverted model

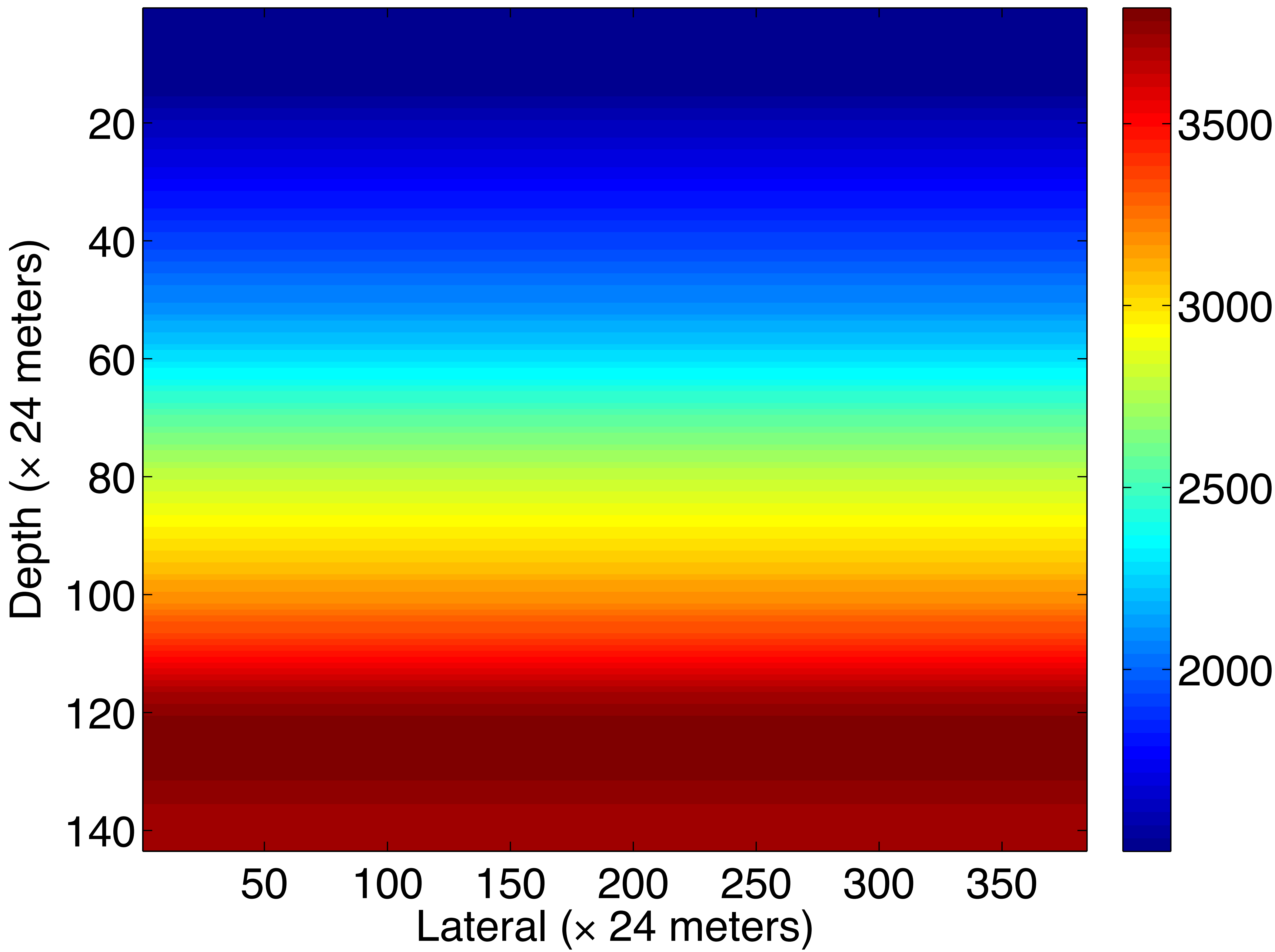


True model

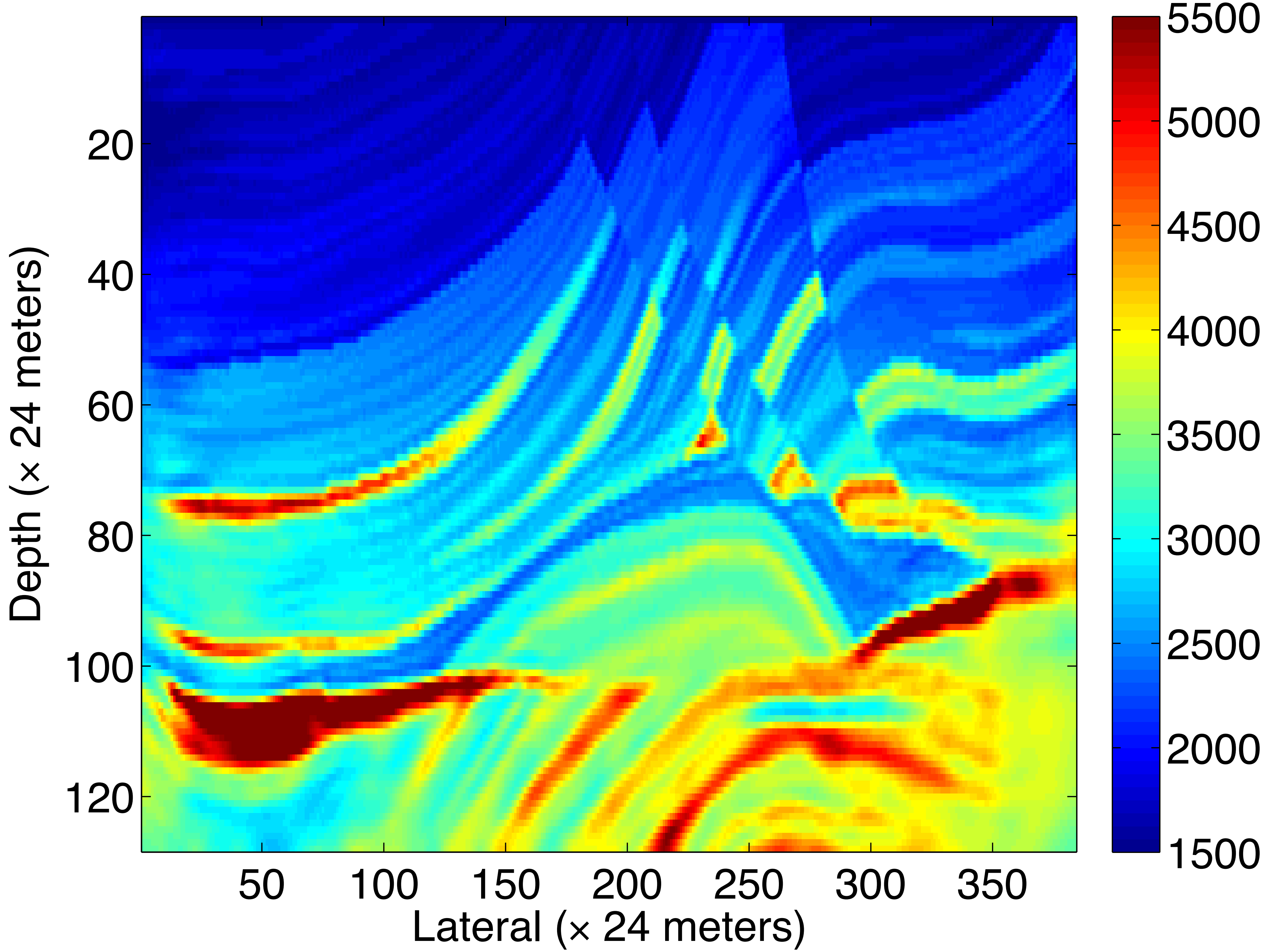


Depth (× 24 meters)
20
40
60
80
100
120
50
100
150
200
250
300
350
Lateral (× 24 meters)
5500
5000
4500
4000
3500
3000
2500
2000

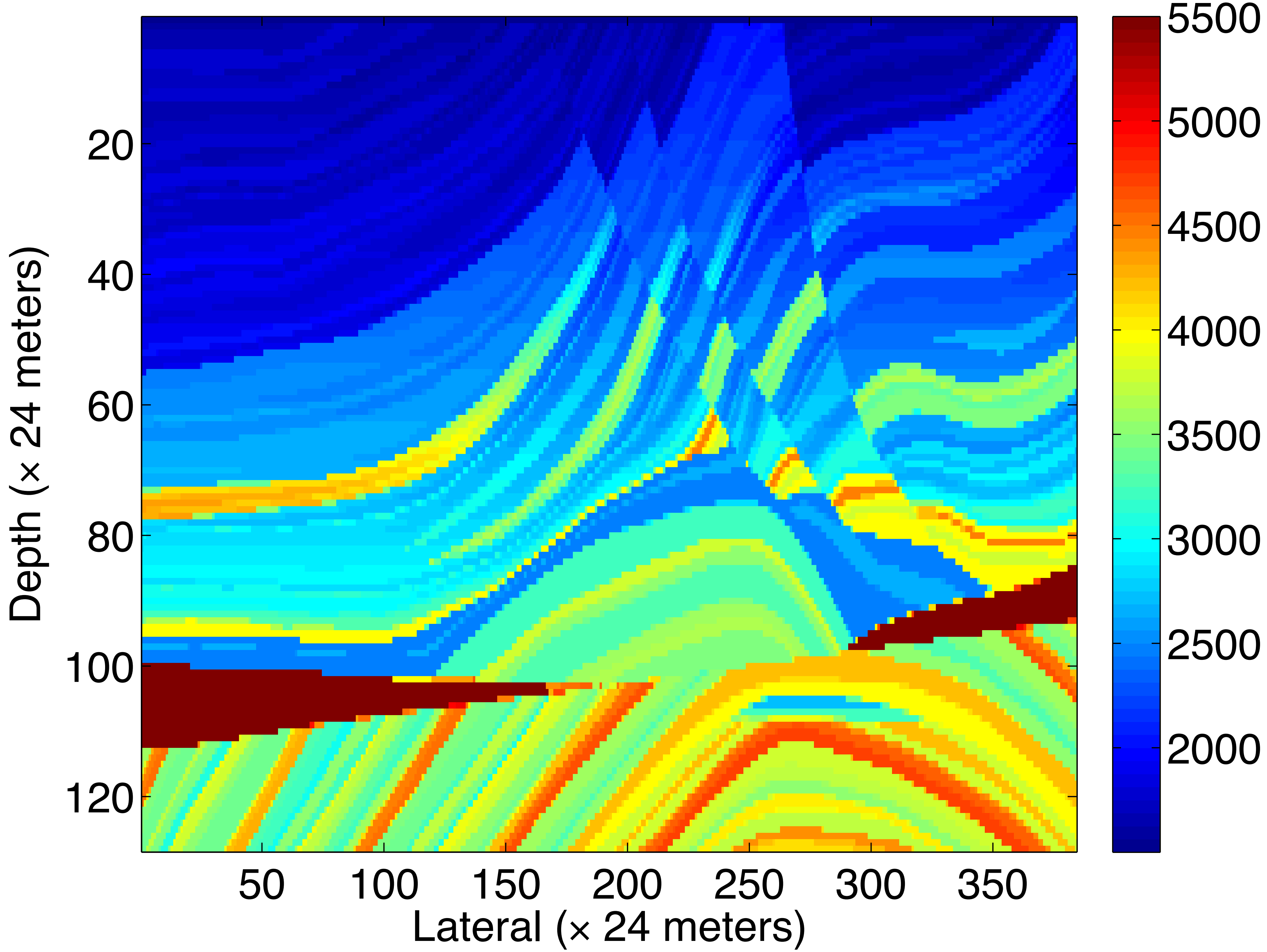
Initial model



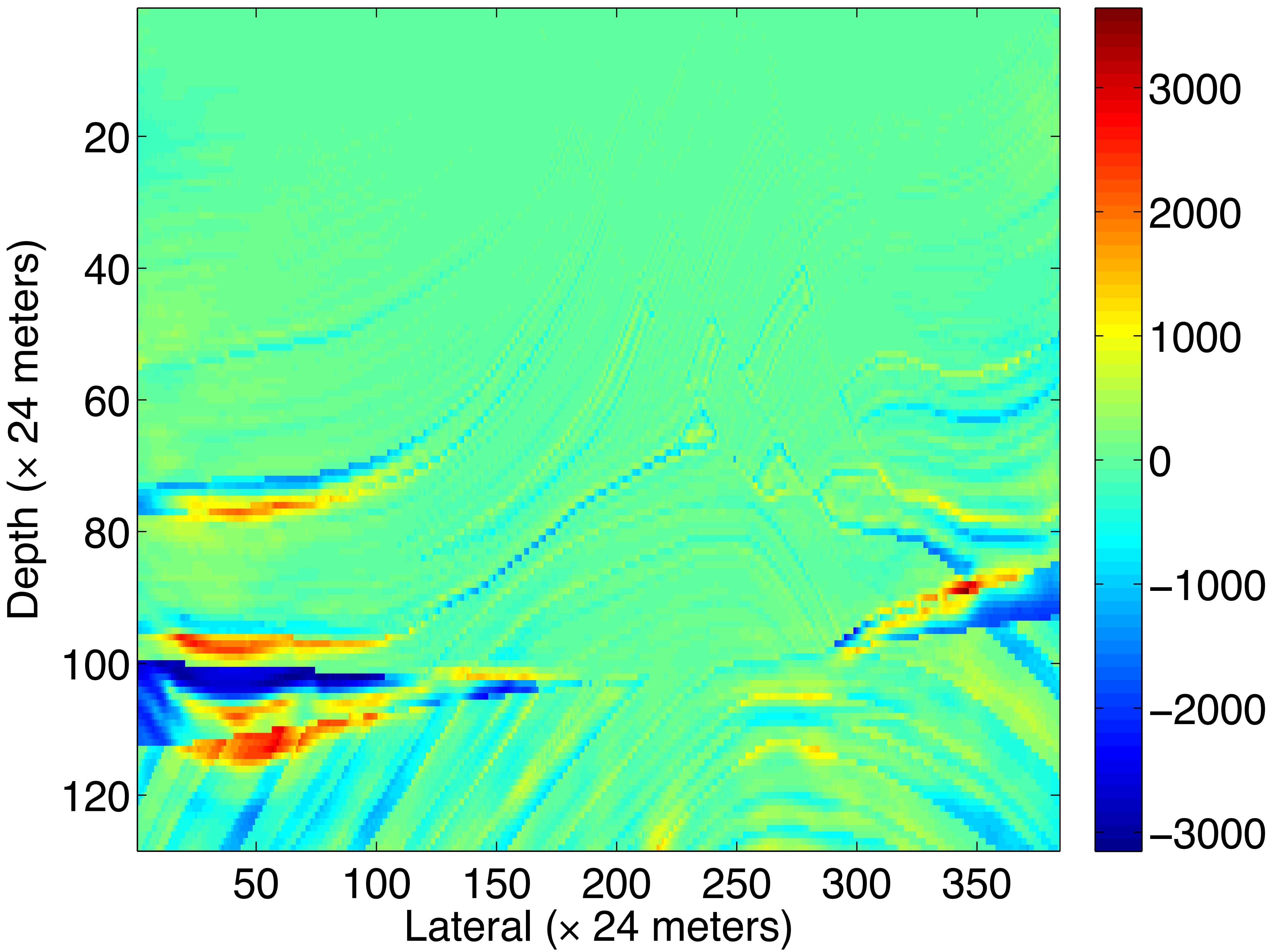
Inverted model



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

Difference


20
40
60
80
100
120
Depth (× 24 meters)
50
100
150
200
250
300
350
Lateral (× 24 meters)
3000
2000
1000
0
−1000
−2000
−3000

Performance

Remember per *subproblem*

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

$$\begin{aligned} n^{\ell_1}_{PDE} &\approx 200 \\ K &= 150 \end{aligned} \quad \textbf{\textit{versus}} \quad \begin{aligned} n^{\ell_2}_{PDE} &\approx 10 \\ K &= 38400 \end{aligned}$$

SPEEDUP of 13 X

Conclusions

Because Compressive Sensing does *not* rely on *averaging* but on *sparsity*, our *approach is a viable* alternative to the *stochastic approximation*

Sparse recoveries offset *random interferences* due to *source encoding*

Hight-quality & high-resolution inversions have been *achieved with significant accelerations*

No need for additional *migration step*

Improvements come from *sparsity promotion & curvelets*

Indications that the *curse of dimensionality* can be removed...

Future plans

Investigate

- *Noise sensitivity*
- *continuation* with batch size (ref latest paper with Haber)
- explore multiscale structure of curvelets
- incomplete data
- extension to 3D

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 (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

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
- Curvelets and Wave Atoms for Mirror-Extended Images by L. Demanet, L. Ying, 07.

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.
- *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 and Recursive Algorithms and Applications* by Kushner and Lin
- *Stochastic Approximation approach to Stochastic Programming* by Nemirovski
- *An effective method for parameter estimation with PDE constraints with multiple right hand sides. by* Eldad Haber, Matthias Chung, and Felix J. Herrmann. '10