Dimensionality Reduction in FWI

Felix J. Herrmann

SLIM

Seismic Laboratory for Imaging and Modeling
the University of British Columbia

Dimensionality Reduction in FWI

Lessons learned from a realistic complex blind synthetic...

Felix J. Herrmann
SLIM team

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

Wave-equation based *inversion*

Industry has difficulty *replenishing* produced *resources*

- have to find oil & gas in *complex* geological areas
- *low* hit rates (1 out of 10) w/ drill costs $.25 B / well
- reduce *risks* of *subsalt* production (BP horizon)

Big drive for *transformative wave*-equation based technology

- *PDE* constrained optimization or full-waveform inversion (FWI)

Challenges

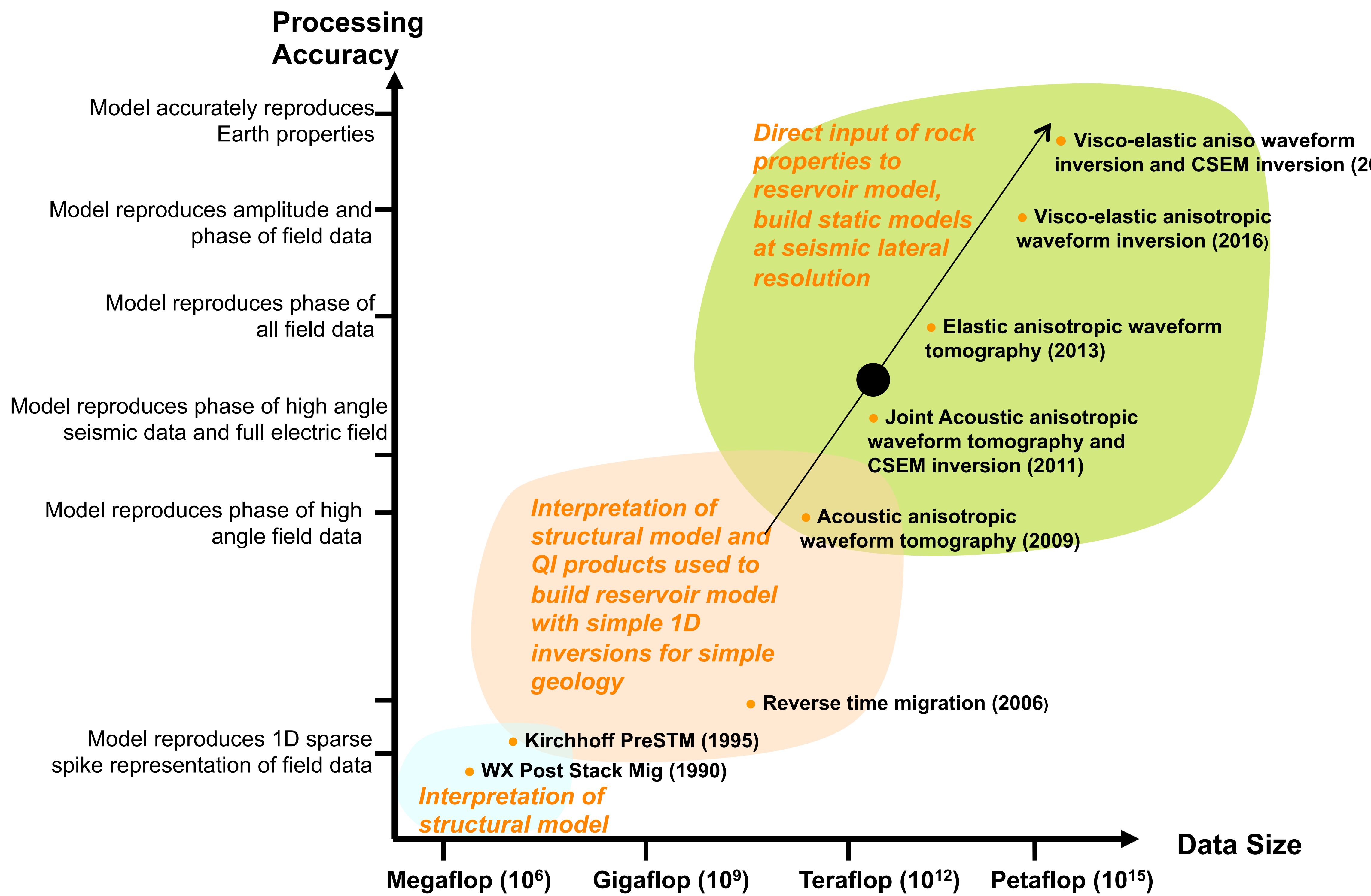
Computational costs

- proportional w/ # of sources
- sky rocket (X 1000) when move to elastic

FWI:

- *nonconvex*, i.e., *local* minima
  (cycle skipping)
- *incomplete* data
  (missing low frequencies, 'Marine' acquisition)

Computational costs



Today's agenda

Brief overview of dimensionality *reduction*

- *randomized* source aggregates/subsets
- *batching* for Quasi-Newton (l-BFGS)
- *CS & AMP* for Gauss-Newton (SPG)

Comparison on *realistic* blind GOM synthetic

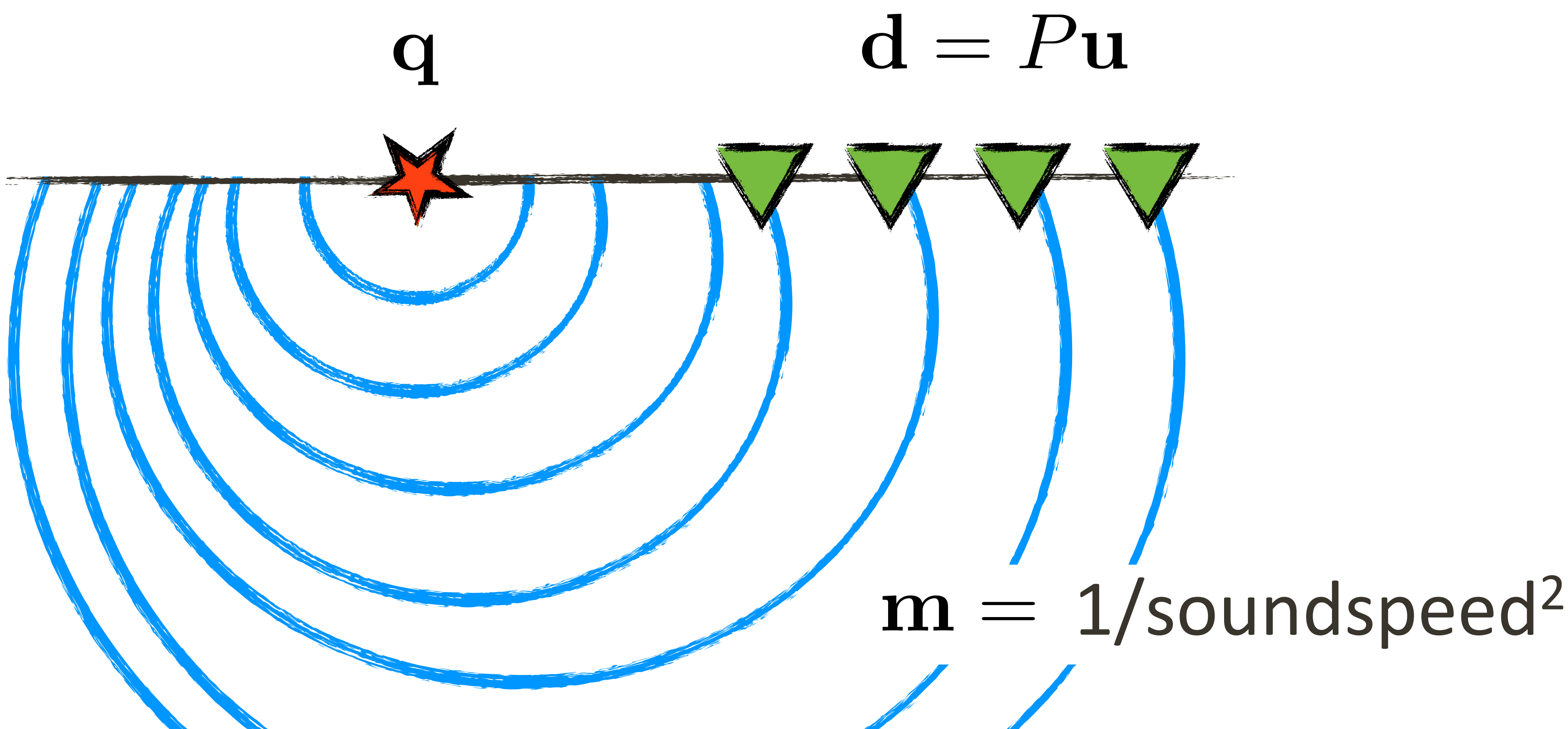
- *data*-side *regularization* & *linear* gradient *smoothings*
- *model*-side *regularization* & *nonlinear* gradient *approximations*

Strategies

- *separable* structure
  (*linearity* w.r.t. *sources*)
- use data *redundancy* to speed up *progress* towards *solution*
  (form *independent* randomized source *subsets*)
- *convex-composite* structure
  (*convex* on the outside & *smooth* within)
- *multiscale* structure
  (*curvelet*-domain *sparsity* GN updates)

Wave-equation based inversion

We model the data in the *acoustic* approximation $(\omega^2\mathbf{m} + \nabla^2)\mathbf{u} = \mathbf{q}$



SLIM

Adjoint-state/ *reduced* formulation

non-linear least-squares problem:

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) = \sum_{i=1}^{M} ||\mathbf{d}_i - P_i \mathbf{u}_i||_2^2$$

gradient:

$$\frac{\partial \Phi}{\partial m_k} = \sum_{i=1}^{M} \mathbf{u}_i^H \left(\frac{\partial H(\mathbf{m})}{\partial m_k} \right)^H \mathbf{v}_i$$

where:

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

$$H(\mathbf{m})^H \mathbf{v}_i = P_i^T (\mathbf{d}_i - P_i \mathbf{u}_i)$$

Inversion of very large sparse linear systems

Batched optimization

$$\min_{\mathbf{m}} \Phi[\mathbf{m}] = \frac{1}{K}\sum_{i=1}^{K} \phi_i[\mathbf{m}]$$

Quasi-Newton approach

$$\mathbf{s}_k = -B_k \nabla\Phi[\mathbf{m}_k]$$
$$\mathbf{m}_{k+1} = \mathbf{m}_k + \lambda_k \mathbf{s}_k$$

But: evaluation of *full* misfit and gradient is very expensive.

Batched optimization

The gradient is the *average*

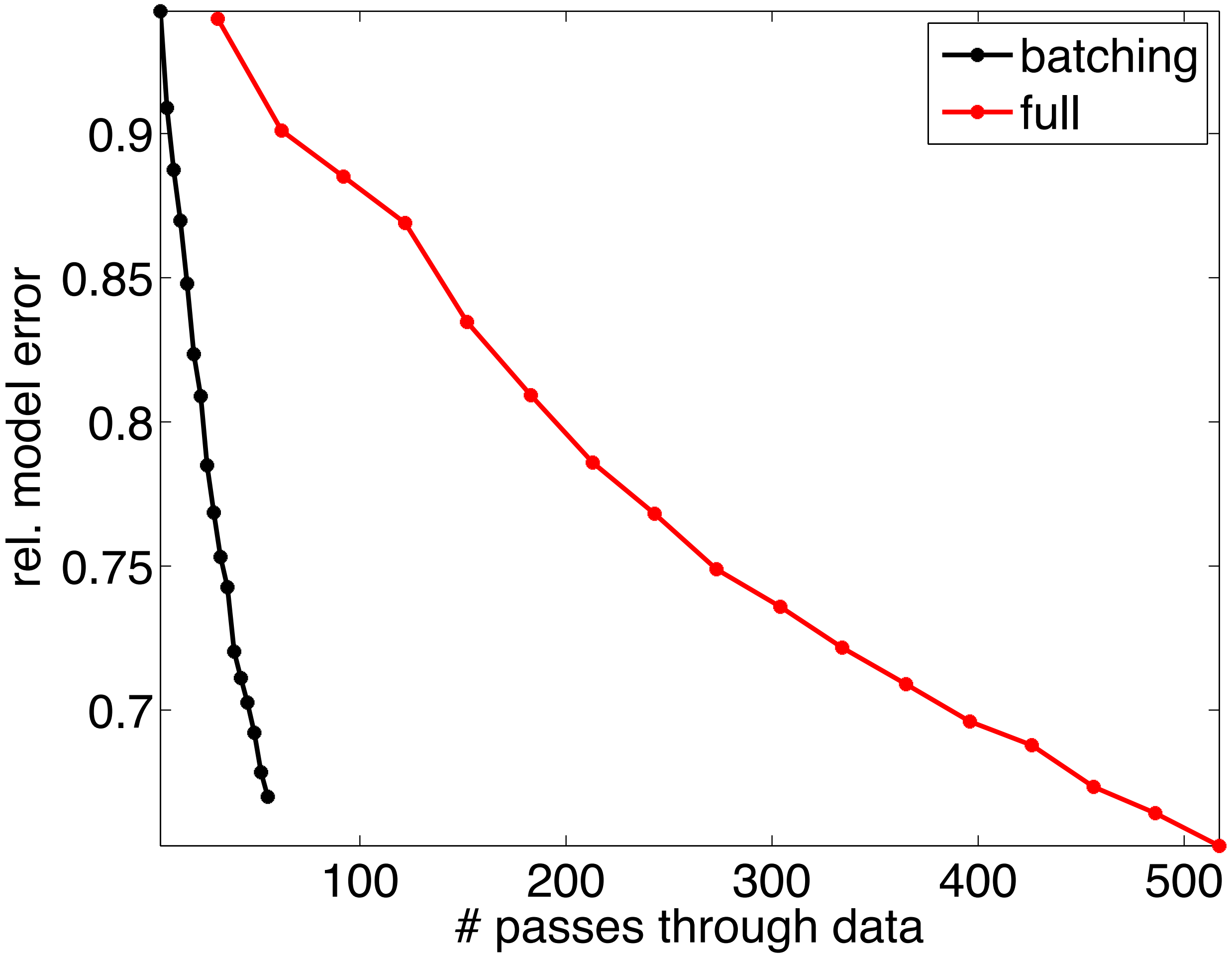
$$\nabla\Phi = \frac{1}{K}\sum_{i=1}^{K}\nabla\phi_i$$

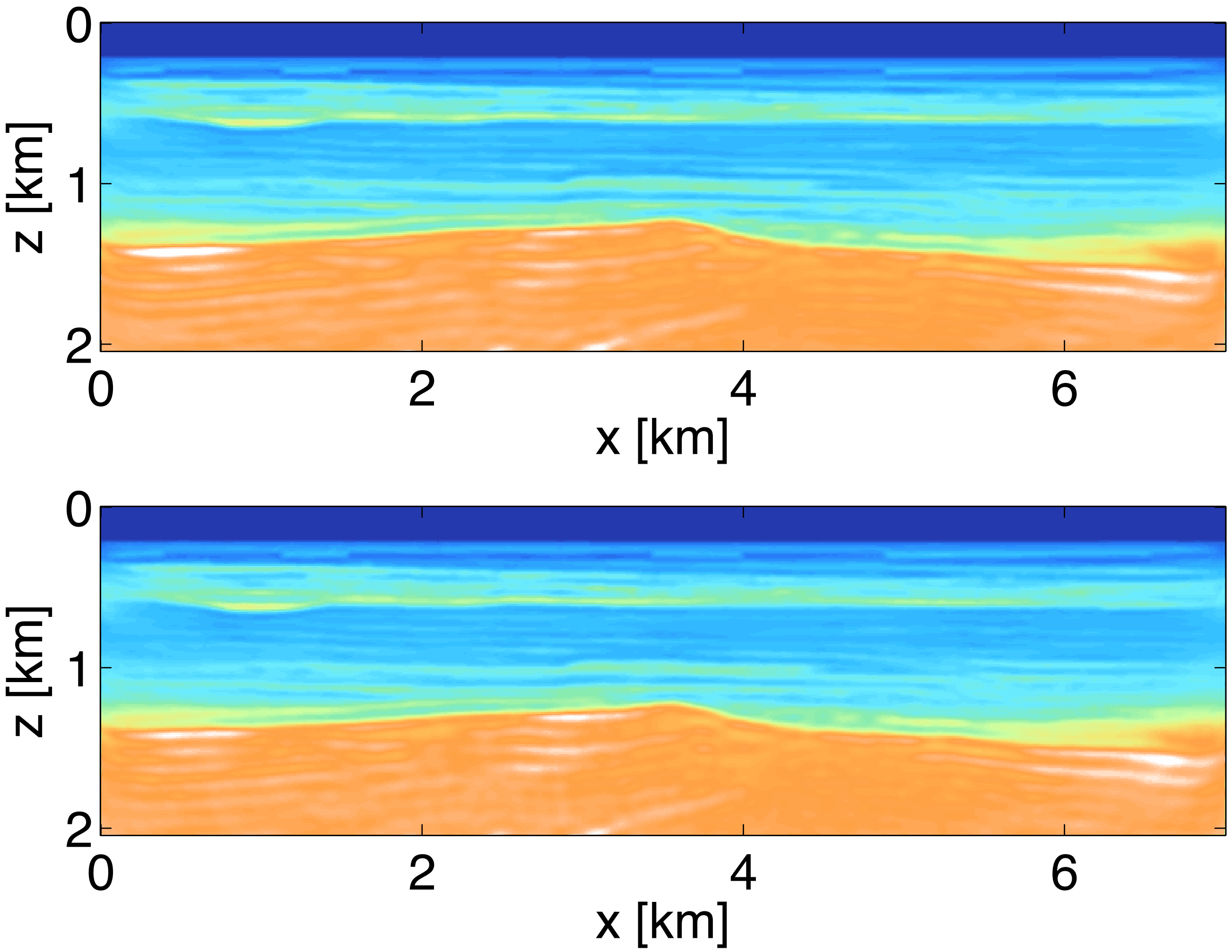
which we can approximate by

$$\nabla\Phi \approx \nabla\widetilde{\Phi} = \frac{1}{|\mathcal{I}|}\sum_{i\in\mathcal{I}}\nabla\phi_i$$

Optimization

10 x speedup





[van Leeuwen & FJH ’11]

Observations

Our *batching* strategy controls *sampling* and/or *simulation* errors

- by *growing* the *batch* size in *accordance* w/ *convergence* rate
- best of both worlds: *stochastic* versus *deterministic*
- removes noise *sensitivity* of *stochastic* gradients

Can we exploit *sparse* structure of gradient *updates*

- *dimensionality* reduction w/ Compressive Sensing
- *acceleration* w/ Approximate Message Passing

Convex composite structure [Burke & Ferris, '95.]

FWI:

$$\min_{\mathbf{m}} \quad \phi(\mathbf{m}) := \underbrace{\frac{1}{2}\| \overbrace{\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}]}^{\text{smooth}} \|_F^2}_{\text{convex}}$$

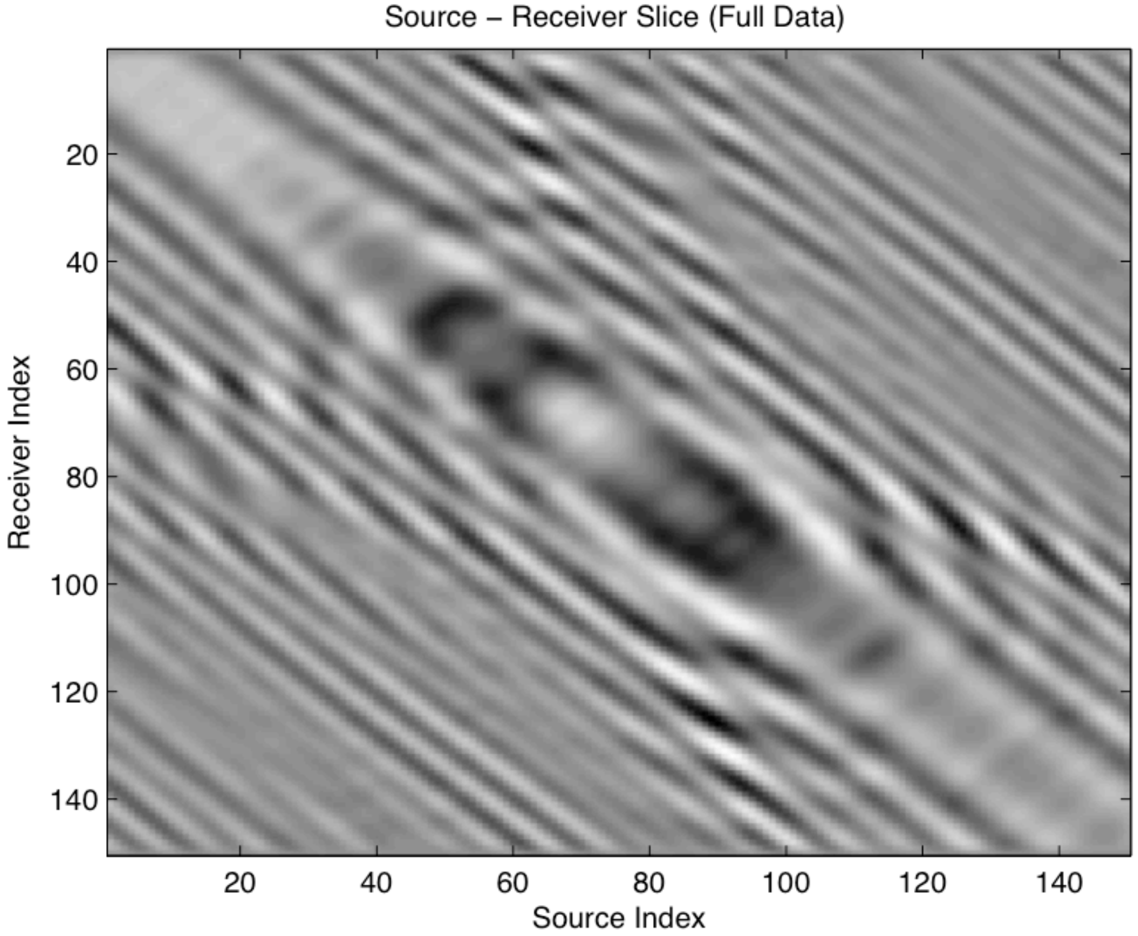
- exploit *convexity* by linearizing *within*

$$\min_{\boldsymbol{\delta}\mathbf{m}} \quad \phi(\boldsymbol{\delta}\mathbf{m}) := \frac{1}{2}\|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}] - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}]\boldsymbol{\delta}\mathbf{m}\|_F^2$$

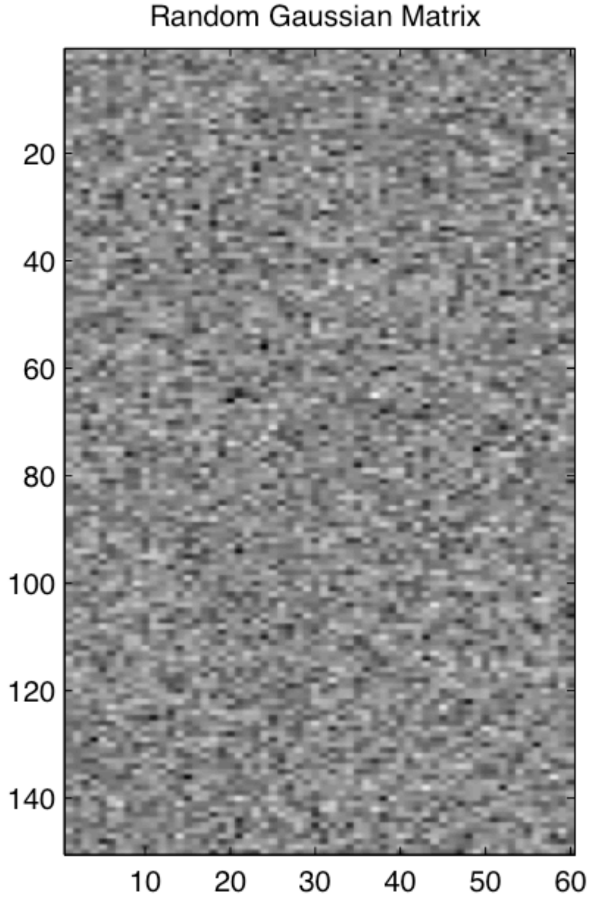
- control the norm of the updates to *guarantee* convergence

Randomized source aggregates

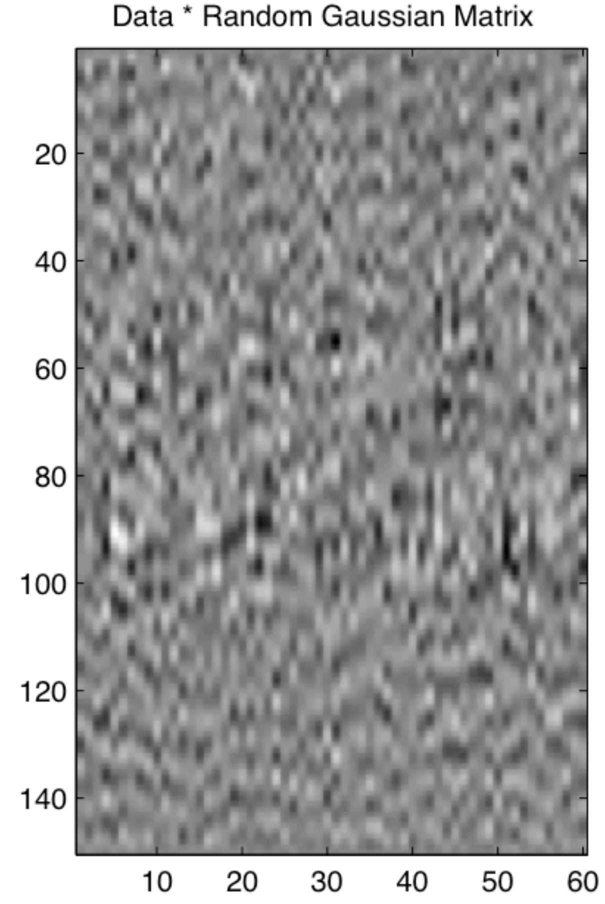
$$\mathbf{D}$$



$$\color{red}{\mathbf{W}}$$



$$\color{red}{\underline{\mathbf{D}}} = \mathbf{D}\color{red}{\mathbf{W}}$$



Convex optimization

[p=2 or p=1]

Linearized inversion with *randomized supershots:*

$$\delta\widetilde{\mathbf{m}} = \mathbf{S}^* \operatorname*{arg\,min}_{\delta\mathbf{x}} \|\delta\mathbf{x}\|_{\ell_p} \quad \text{subject to} \quad \| \underbrace{\delta\underline{\mathbf{d}}}_{\mathbf{b}} - \underbrace{\overbrace{\nabla\boldsymbol{\mathcal{F}}[\mathbf{m}_0;\underline{\mathbf{Q}}]}^{\text{Jacobian}}\mathbf{S}^*}_{\mathbf{A}} \delta\mathbf{x}\|_2 \leq \sigma$$

$$\begin{aligned}
\delta\mathbf{x} &= \text{Sparse curvelet-coefficient vector}\\
\mathbf{S}^* &= \text{Curvelet synthesis}\\
\underline{\mathbf{Q}} &= \text{Simultaneous sources}\\
\delta\underline{\mathbf{d}} &= \text{Super shots}
\end{aligned}$$

Fast Gauss-Newton step
[*via* stochastic optimization]

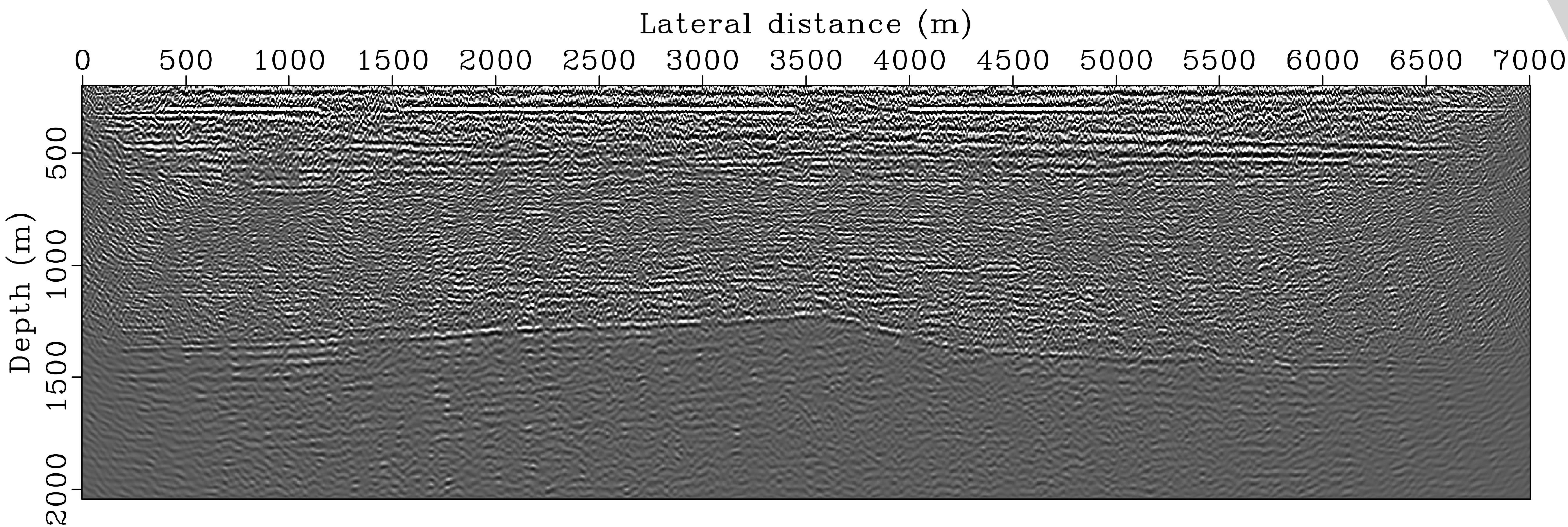
*E*xploit multi-experiment *redundancy* of seismic data volumes by *re*randomized sampling

- *regularly* draw *independent* source *aggregates*
- cancels *crosstalk/interference* by *rerandomization*

Heuristic of *current* phase-encoding/*dimensionality* reduction for imaging/FWI

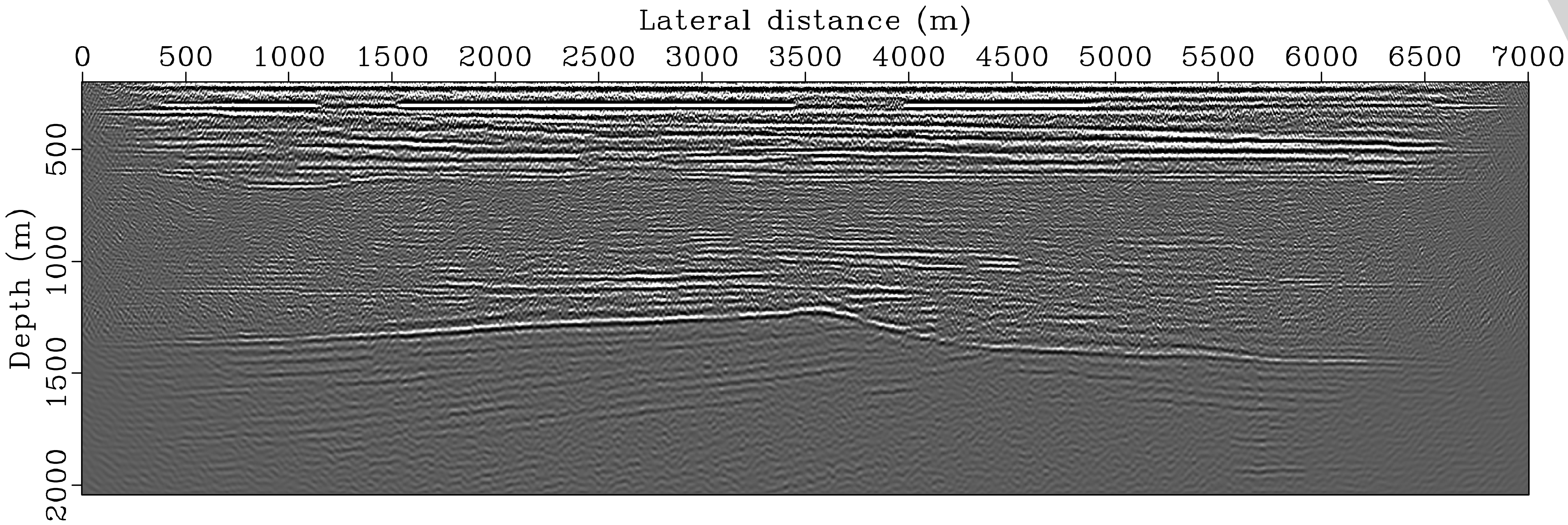
Fast Gauss-Newton step

[ℓ_2 *w/o rerandomization 3 super shots*]



Fast Gauss-Newton step

[ℓ_2 *w/ rerandomization* 3 *super* shots]



Fast Gauss-Newton step

[*via* compressive sensing]

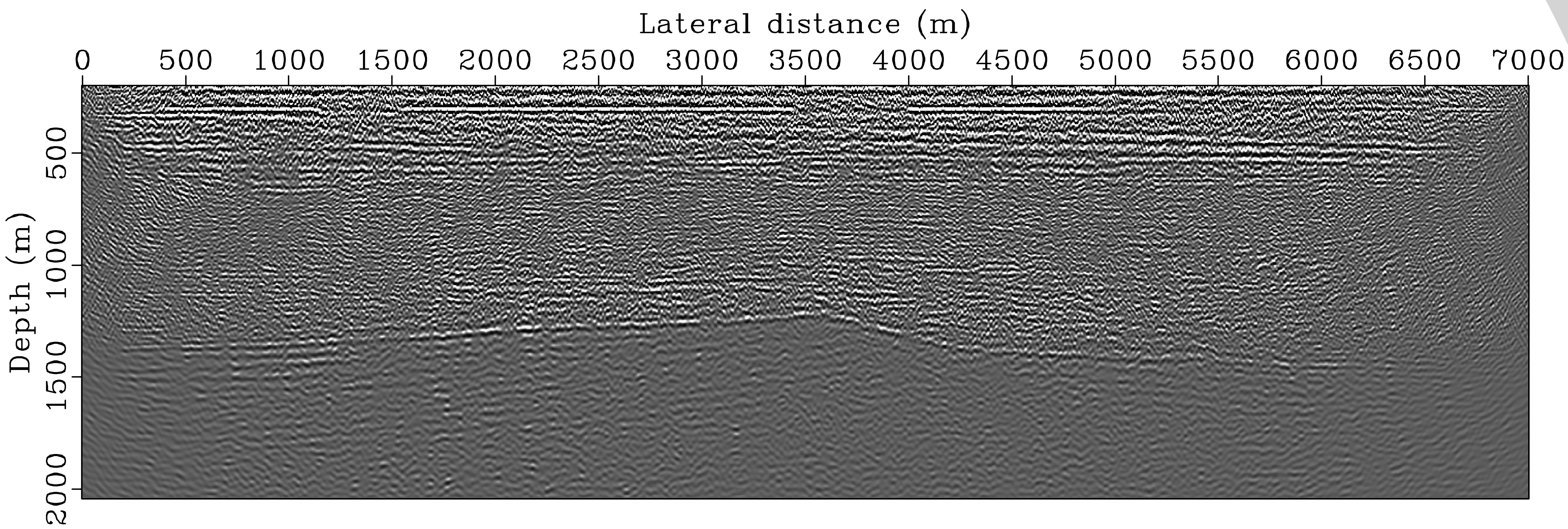
Randomized sampling turns *coherent* source crosstalk/ interferences into ***non-sparse*** *incoherent* noise.

Exploits transform-domain *structure* exhibited *within* GN *updates*

- leverage *curvelet*-domain ***sparsity*** promotion
- *map* "noisy" crosstalk/interferences to *coherent* reflectors

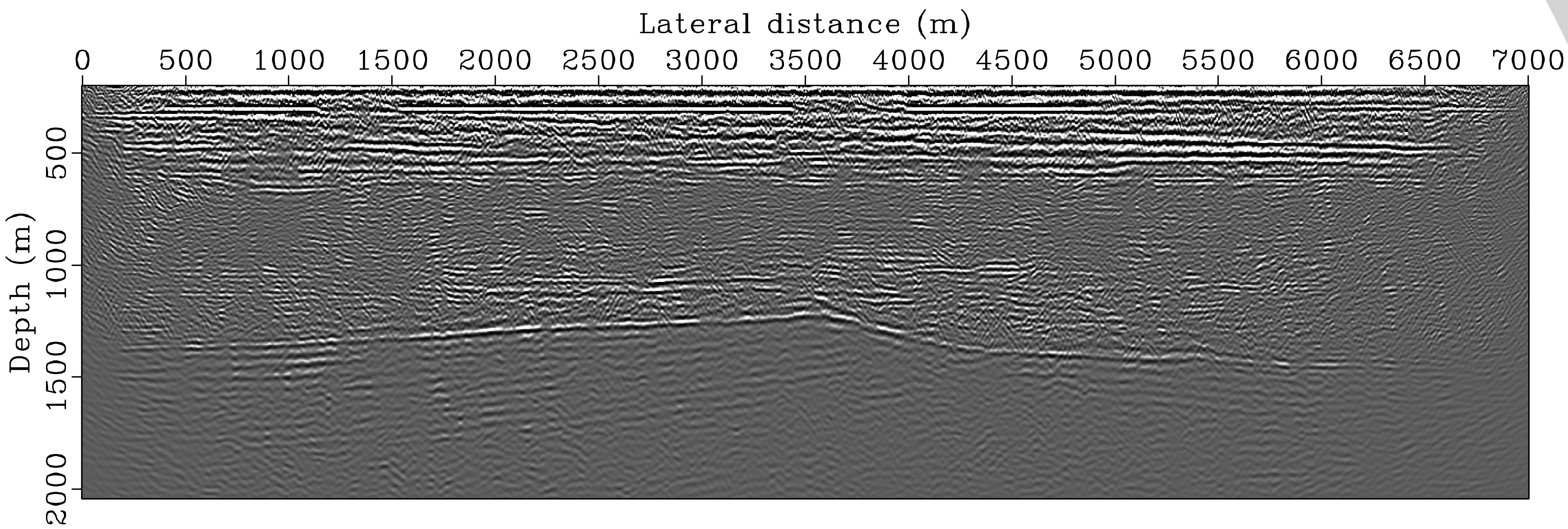
Fast Gauss-Newton step

[ℓ_2 3 *super* shots]



Fast Gauss-Newton step

[ℓ_1 3 *super* shots]



Observations
[w/ reasonable PDE solve budget]

Rerandomization and *curvelet*-domain *sparsity* promotion:

- *partly* eliminate "*noisy*" *crosstalk*
- *fail* to *remove* "small" *incoherent* crosstalk

Can we somehow combine these two methods?

- *continuation* method for large-scale *convex* optimization
- use *insights* from *approximate* message passing

Supercooling

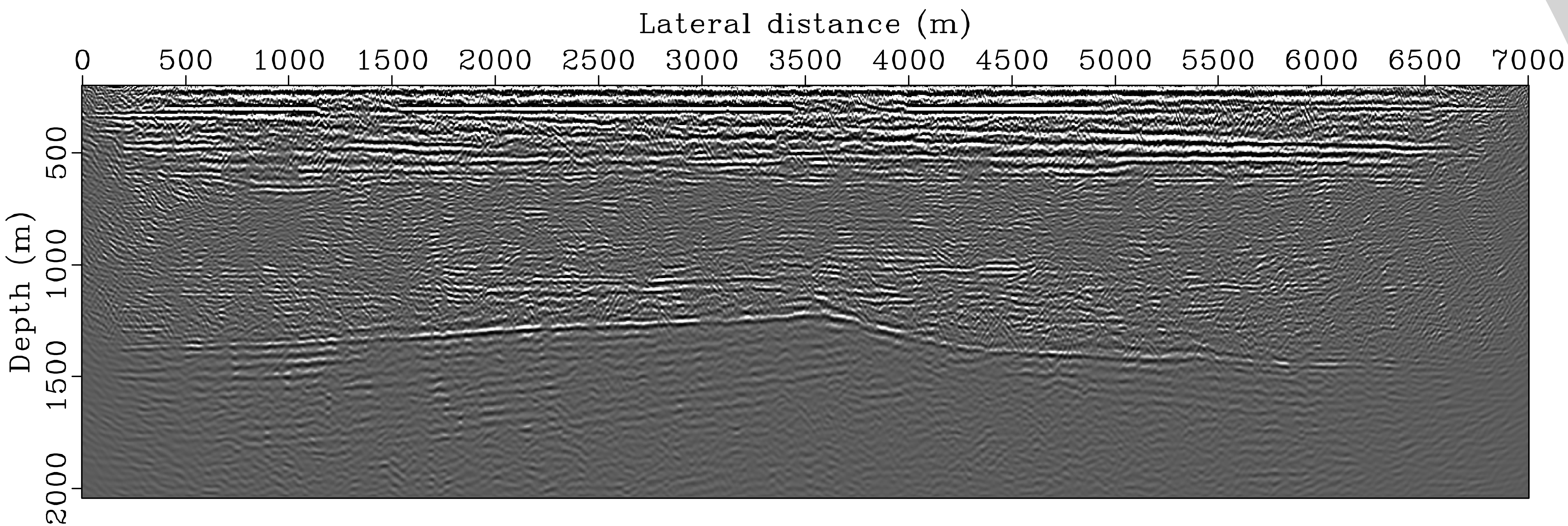
Break *correlations* between the model *iterate* and matrix **A** by *rerandomization*

- draw new *independent* $\{\mathbf{b}_t, \mathbf{A}_t\}$ after each subproblem is solved
- brings in "*extra*" information *without* growing the *system*
- ***minimal*** *extra* computational & memory cost

Progress one-norm solvers *no* longer *stalled*...

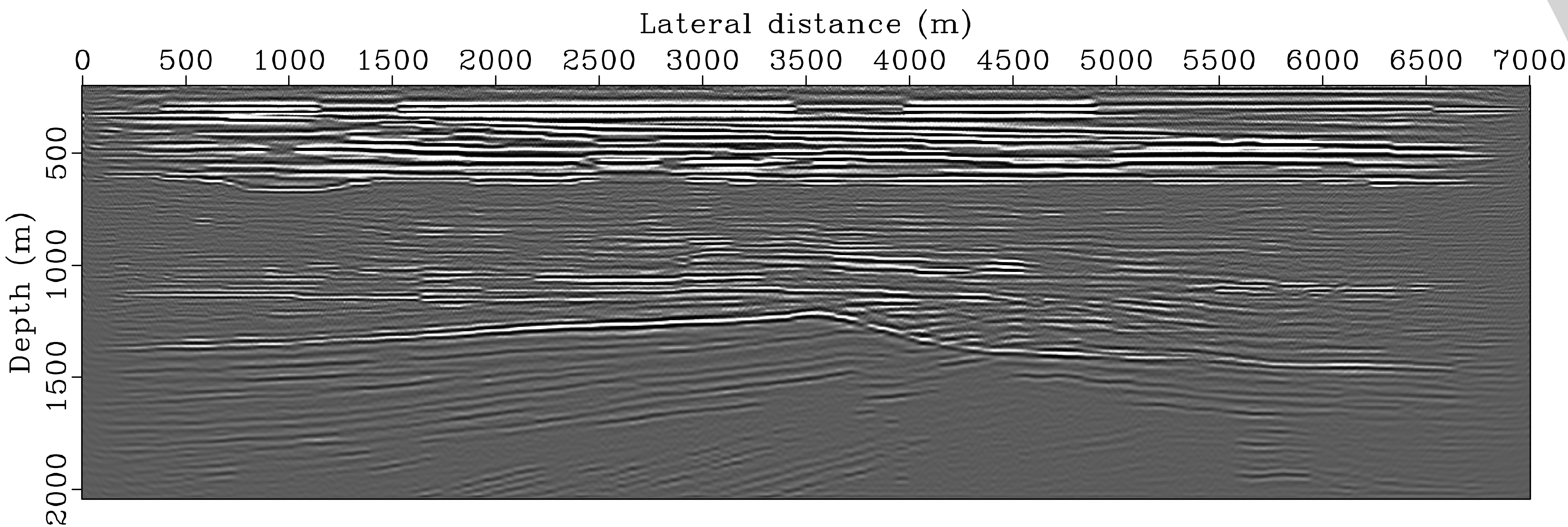
Fast Gauss-Newton step

[ℓ_1 *w/o* rerandomization 3 *super* shots]



Fast Gauss-Newton step

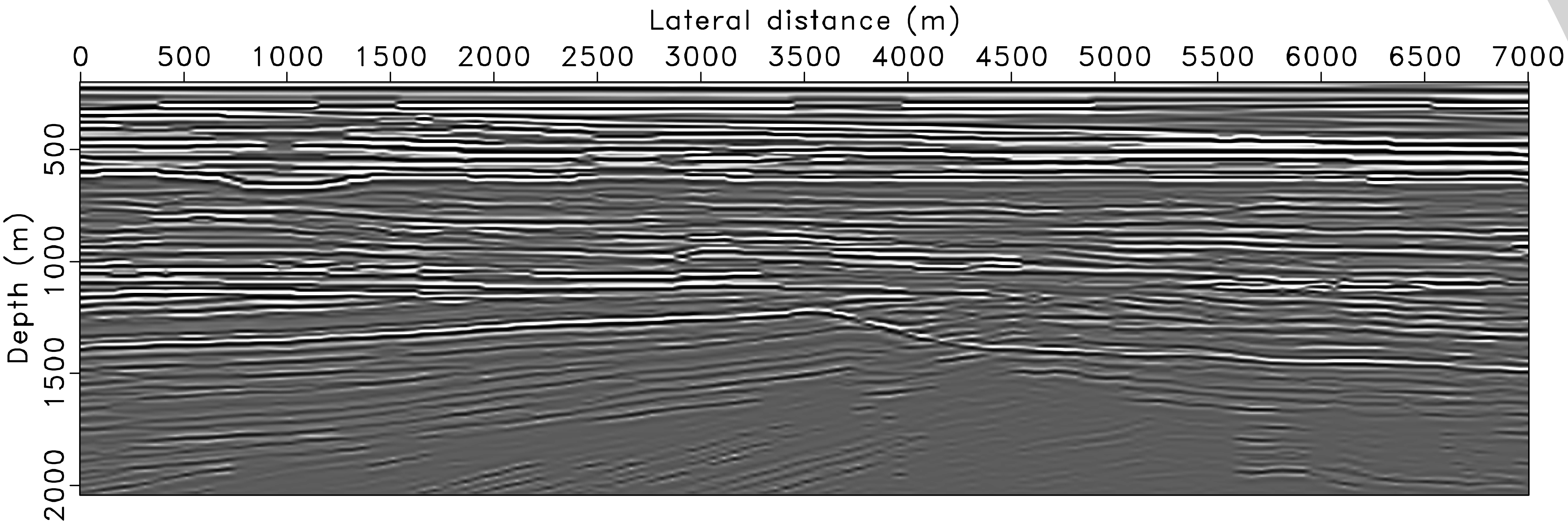
[ℓ_1 *w/* rerandomization 3 *super* shots]



cost of 1/2 *gradient* update w/ *all* data

Fast Gauss-Newton step

[*true* update]

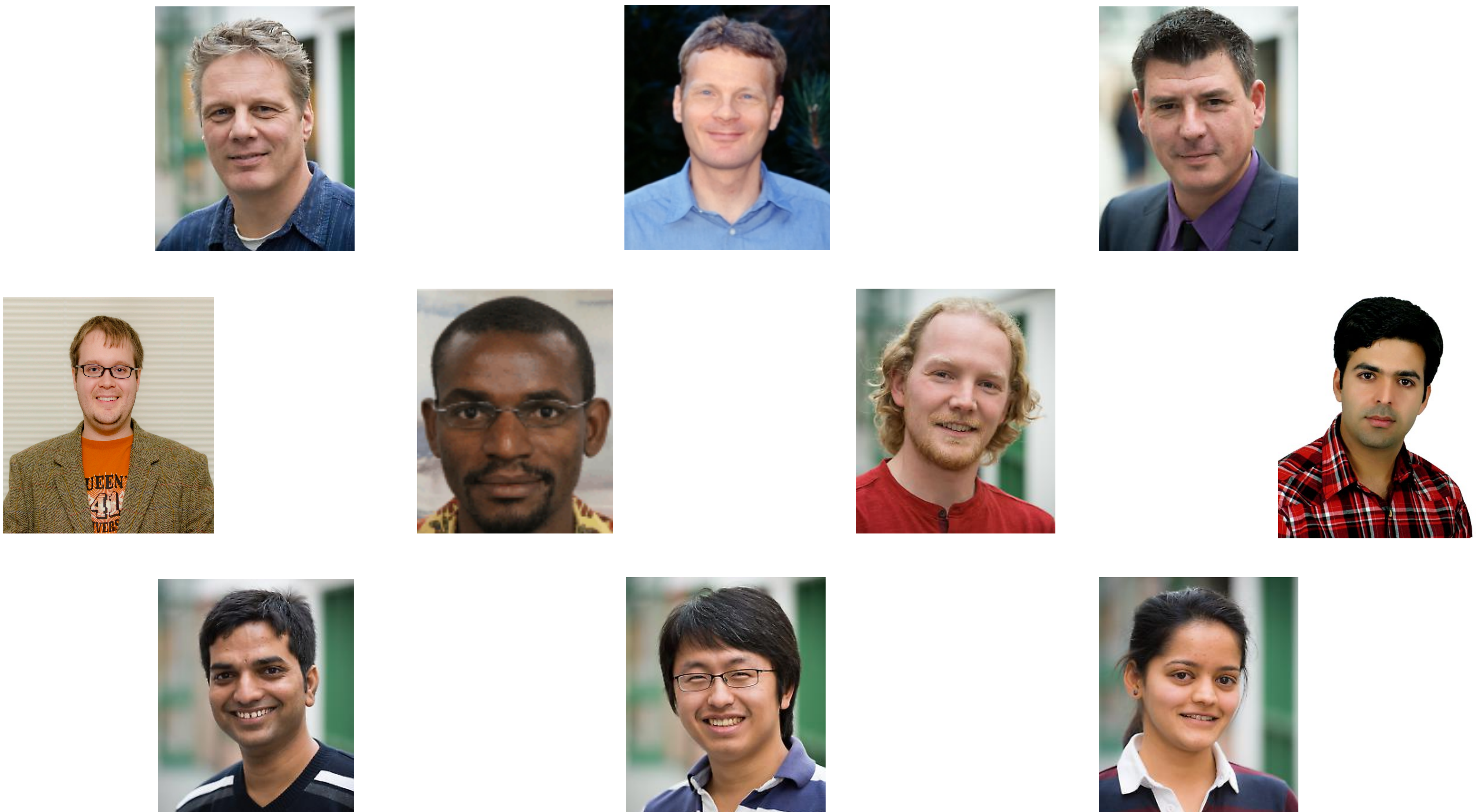


Blind GOM subsalt case study

Elastic synthetic data (3200 shots) for an *unknown* earth model was generated by a team from Chevron, ExxonMobil, and Schlumberger

Several contractor companies worked w/ large teams for weeks/months to get results w/ a lot of "hand holding"...

Dream team



Comparison

- nonlinear conjugate gradients
  - ***linear*** *anisotropic* smoothing of *gradients*
- *modified* Gauss-Newton
  - ***nonlinear*** "smoothing"/ *approximations* of *gradients* with *anisotropic* curvelets

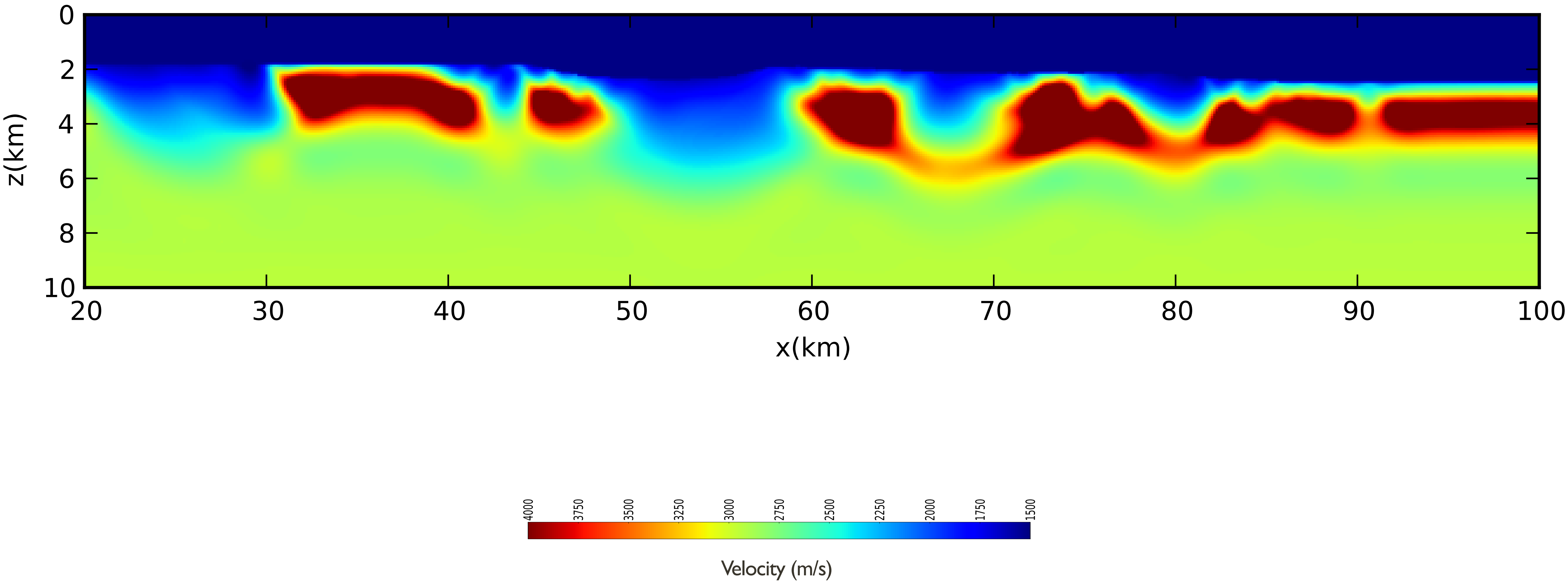
Acoustic FWI workflows with *minimal* parameters & user *intervention*...

Preprocessing

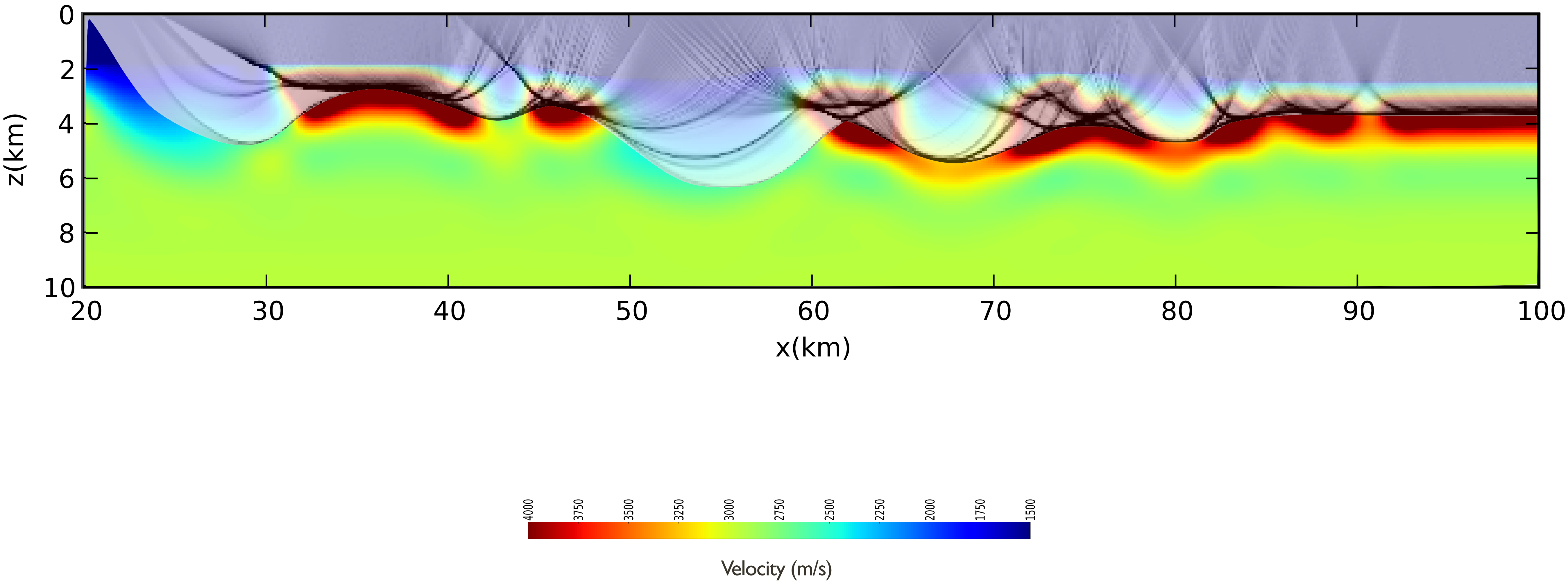
Travel-time *tomography* based on hand-picked *first* breaks ⟶ *initial velocity model*

Curvelet-denoising at selected *low*-frequencies ⟶ *improve SNR* for FWI

Initial model
[ray-based tomography]



Ray paths
[RMS traveltime misfit 11ms]



Denoising

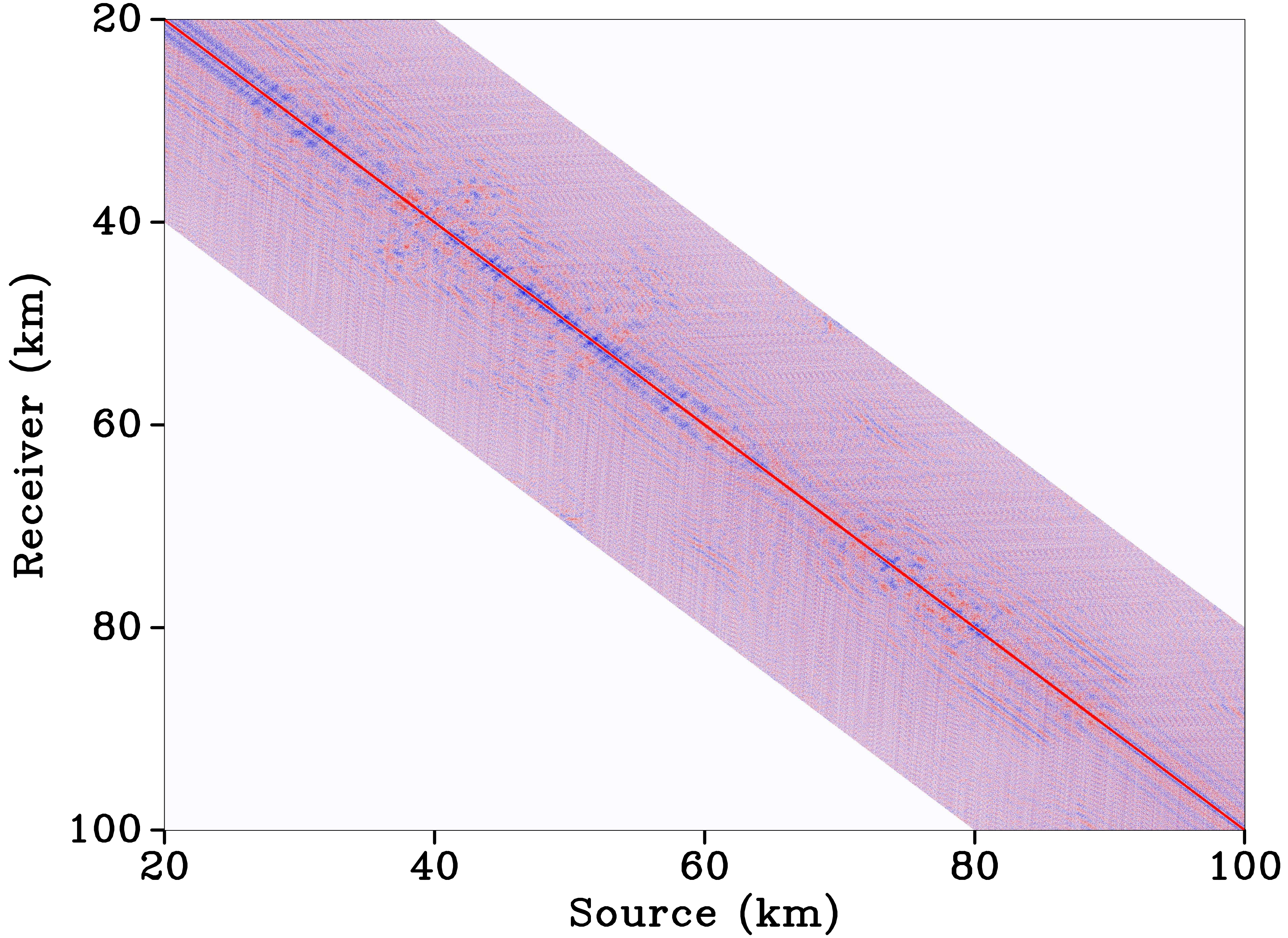
Work on “frequency slices” in the *source/ receiver* plane

Find *support* by *hand*-selected *thresholding* of *syntheses* curvelet coefficients

Followed by *debiasing* to restore *amplitudes*...

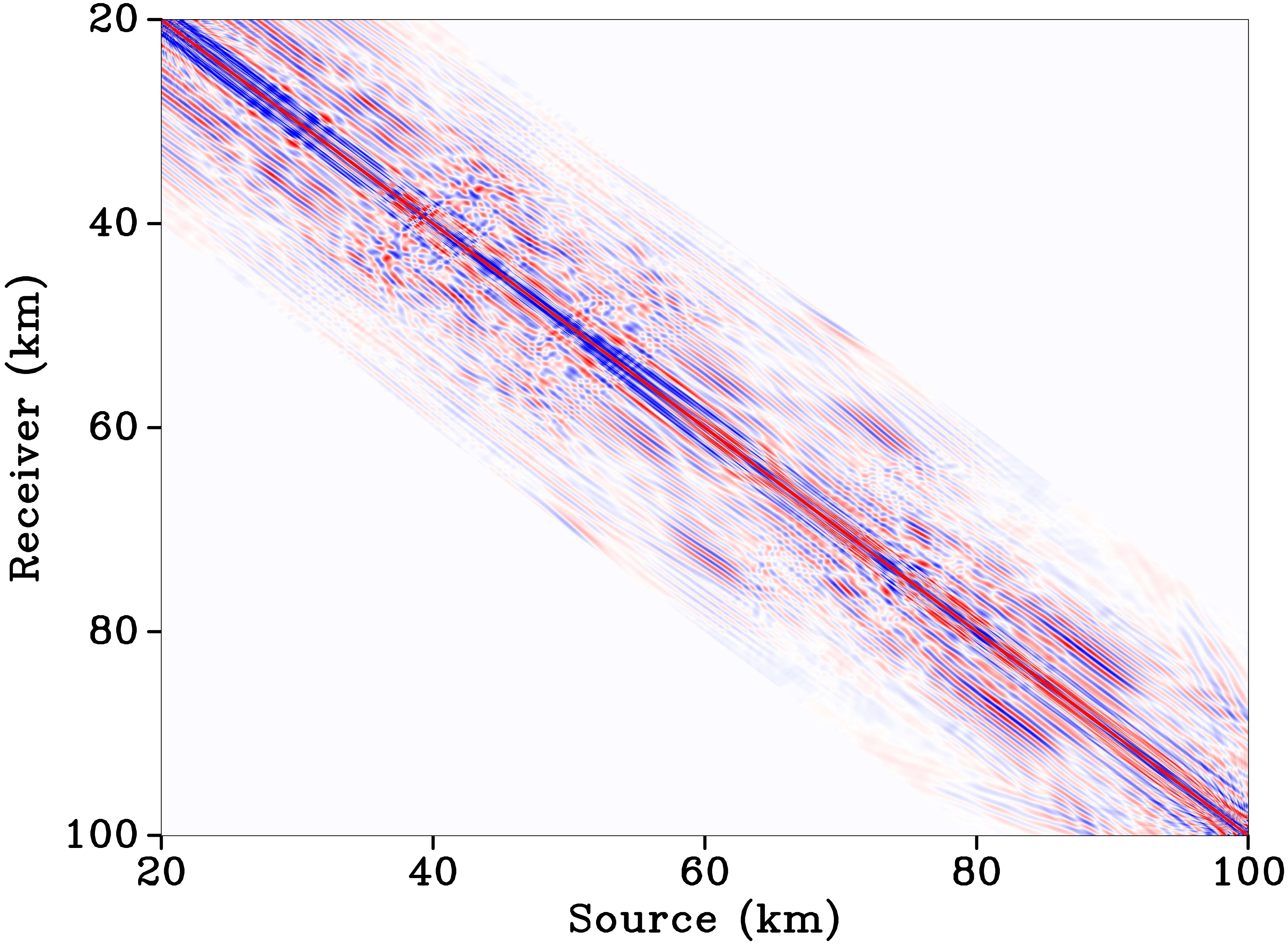
Before denoising

[real part at 2.0 Hz]



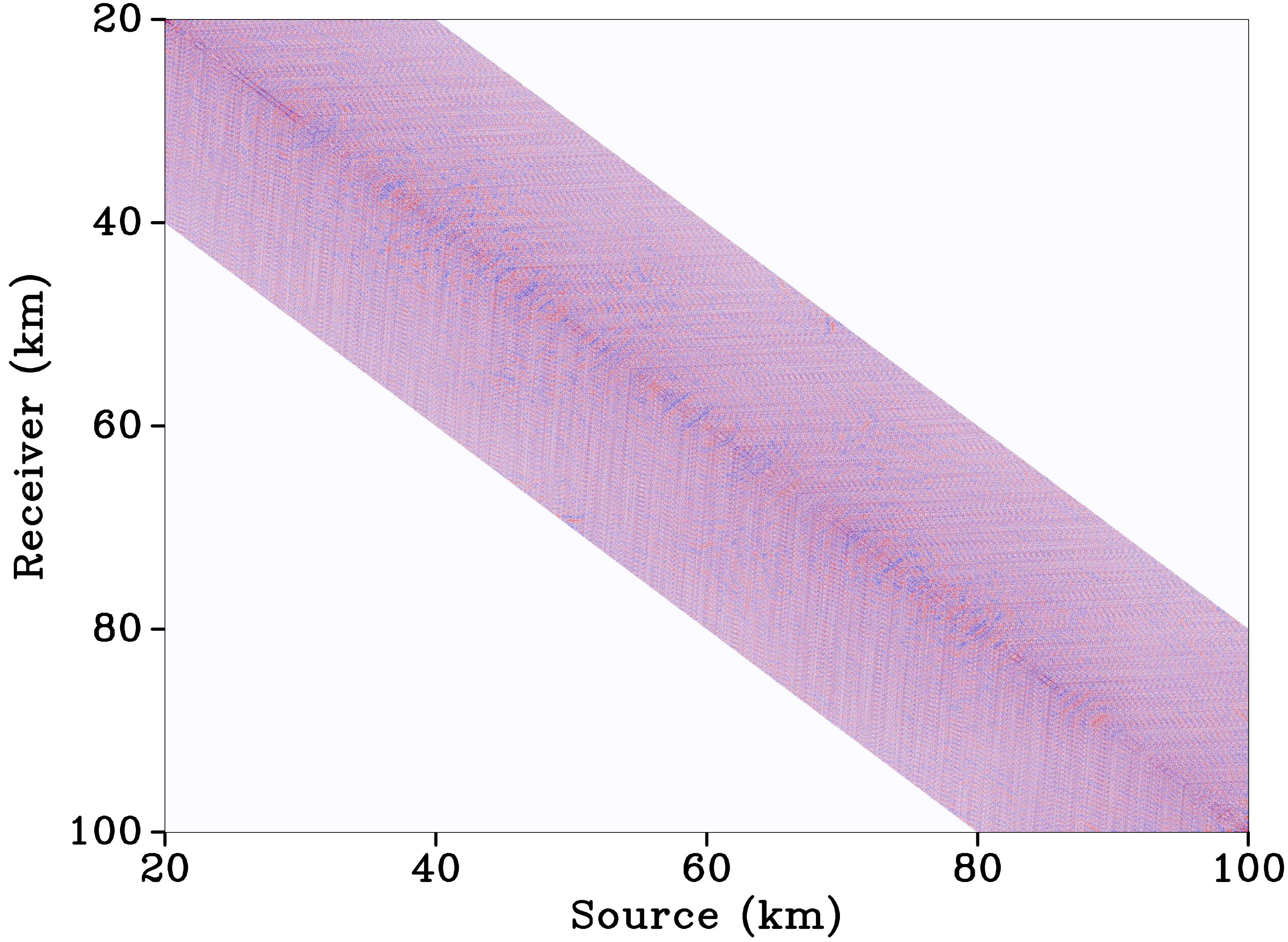
After denoising

[real part at 2.0 Hz]



Difference

[real part at 2.0 Hz]

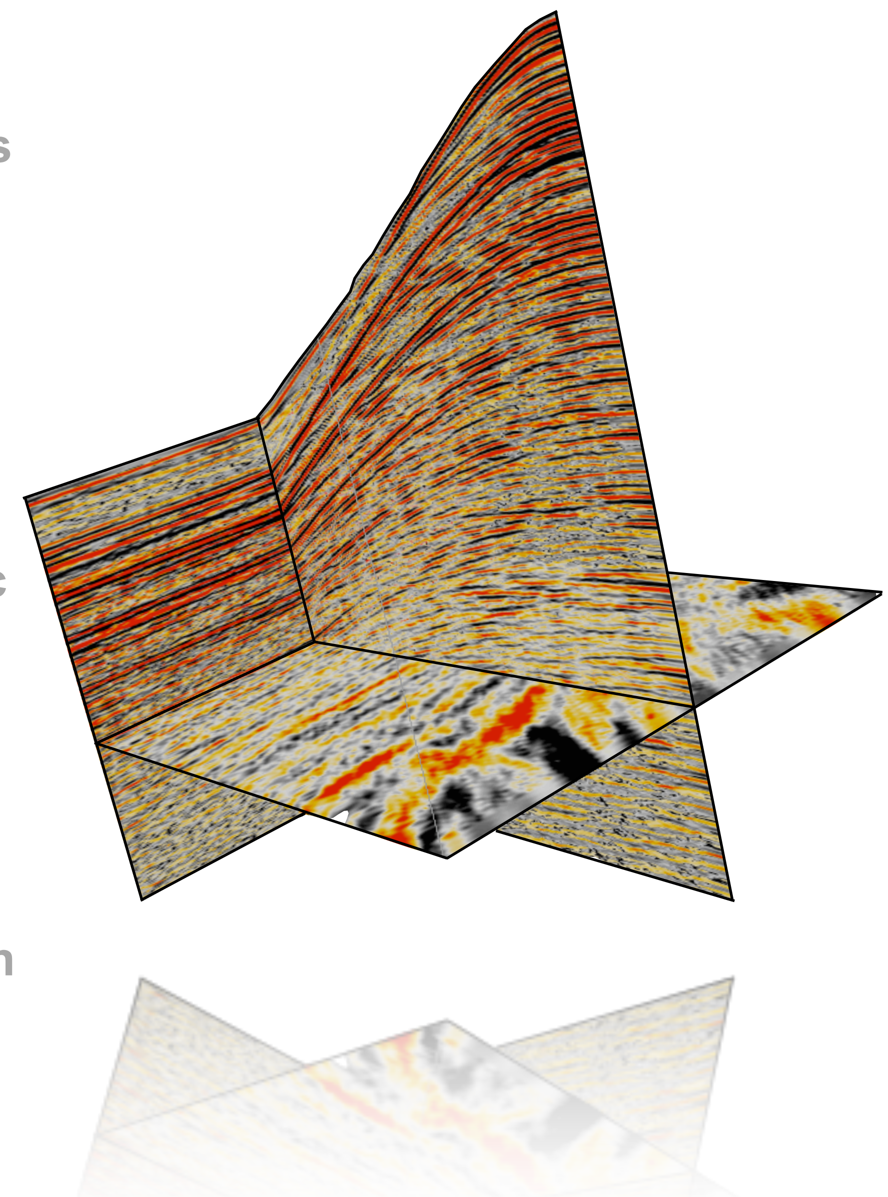


Nonlinear Conjugate Gradients

Pratt, R.G., Shin, Changsoo and Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency domain seismic waveform inversion. Geophysical Journal International, 133, 341-362.

Brenders, A. J. and Pratt, R. G., 2007. Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model, Geophysical Journal International, 168, 133-151

Takam Takougang, E. M. and Calvert, A. J., 2011. Application of waveform tomography to marine seismic reflection data from the Queen Charlotte Basin of western Canada, Geophysics,76, B1-B16



Methodology

Non-linear conjugate gradients (first order)

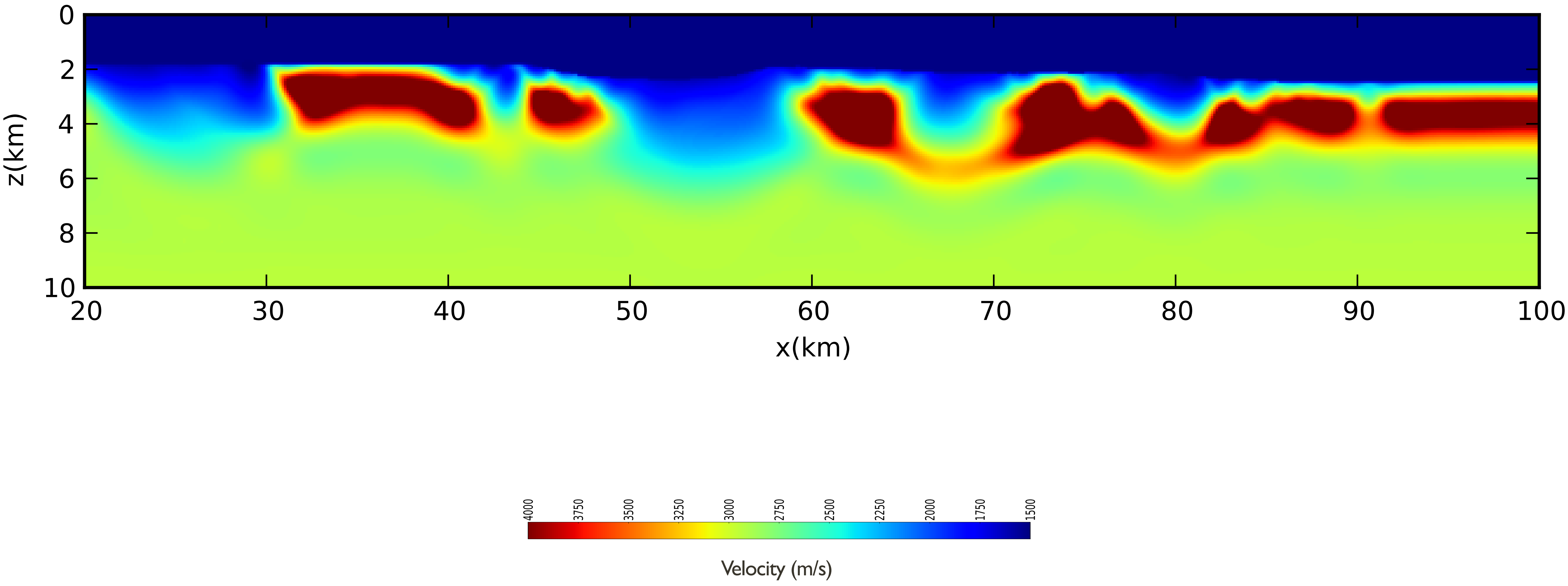
Scaling by *RMS* from (initial) *forward* model

Layer stripping with gradient & offset weights

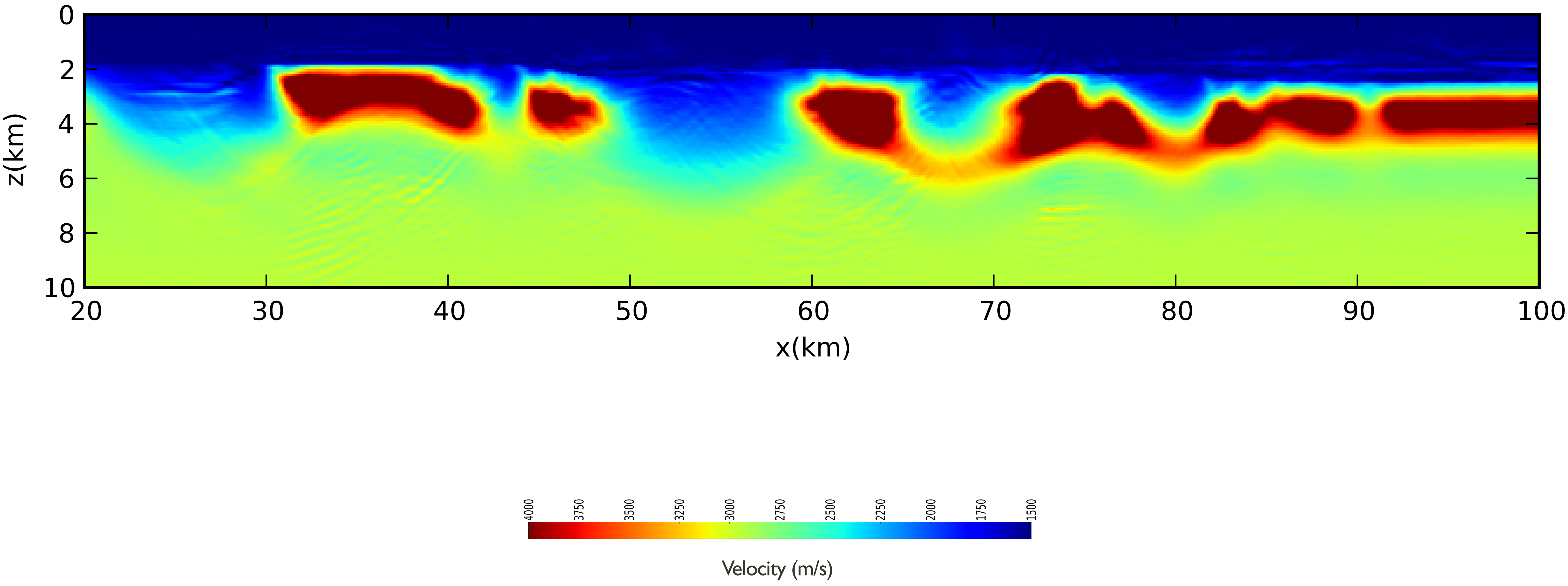
- *phase* only 2–7 Hz

Anisotropic *smoothing* on *gradients*...

Initial model
[ray-based tomography]



Final result
[w/o denoising]



Batched FWI with nuisance parameter estimation

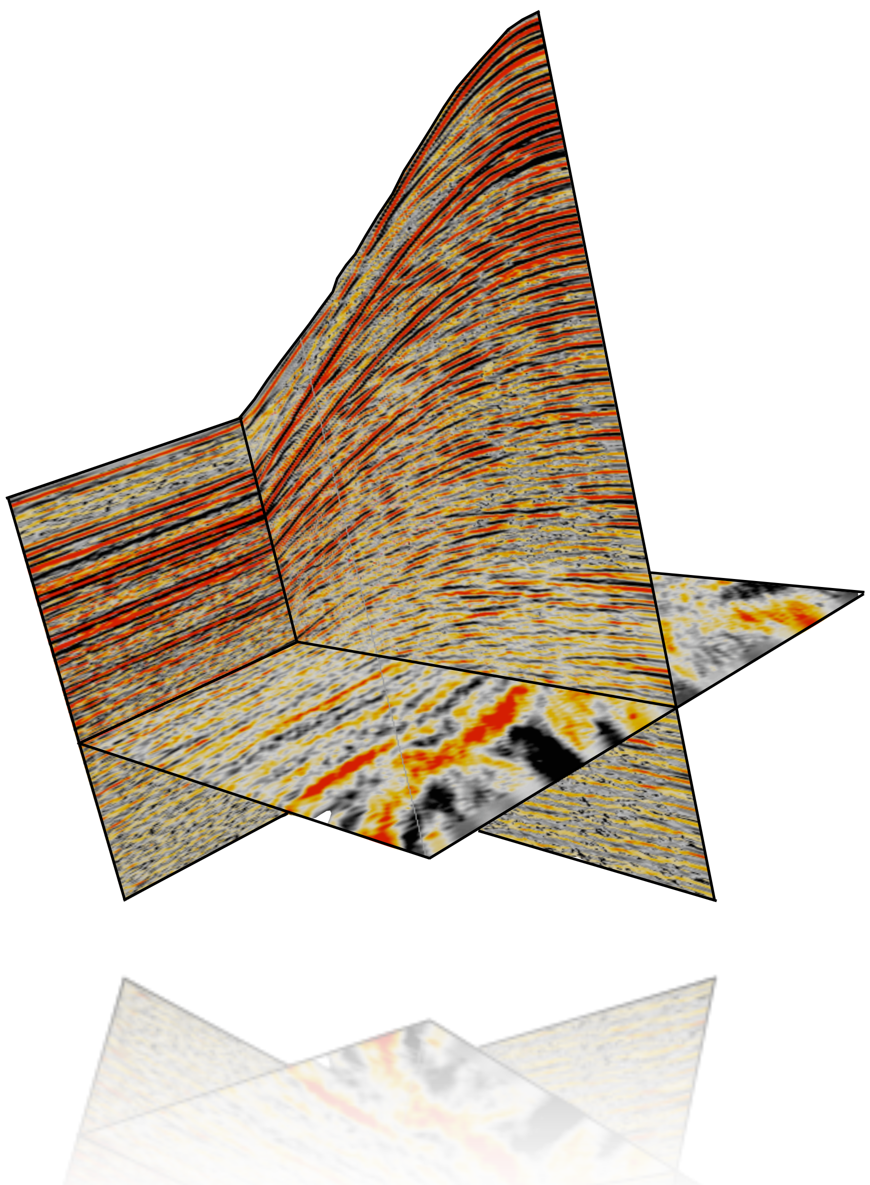
A.Y. Aravkin and T. van Leeuwen - Estimating nuisance parameters in inverse problems. Inverse Problems, 2012.

T. van Leeuwen and F.J. Herrmann - Fast waveform inversion without source encoding. Geophysical Prospecting, 2012

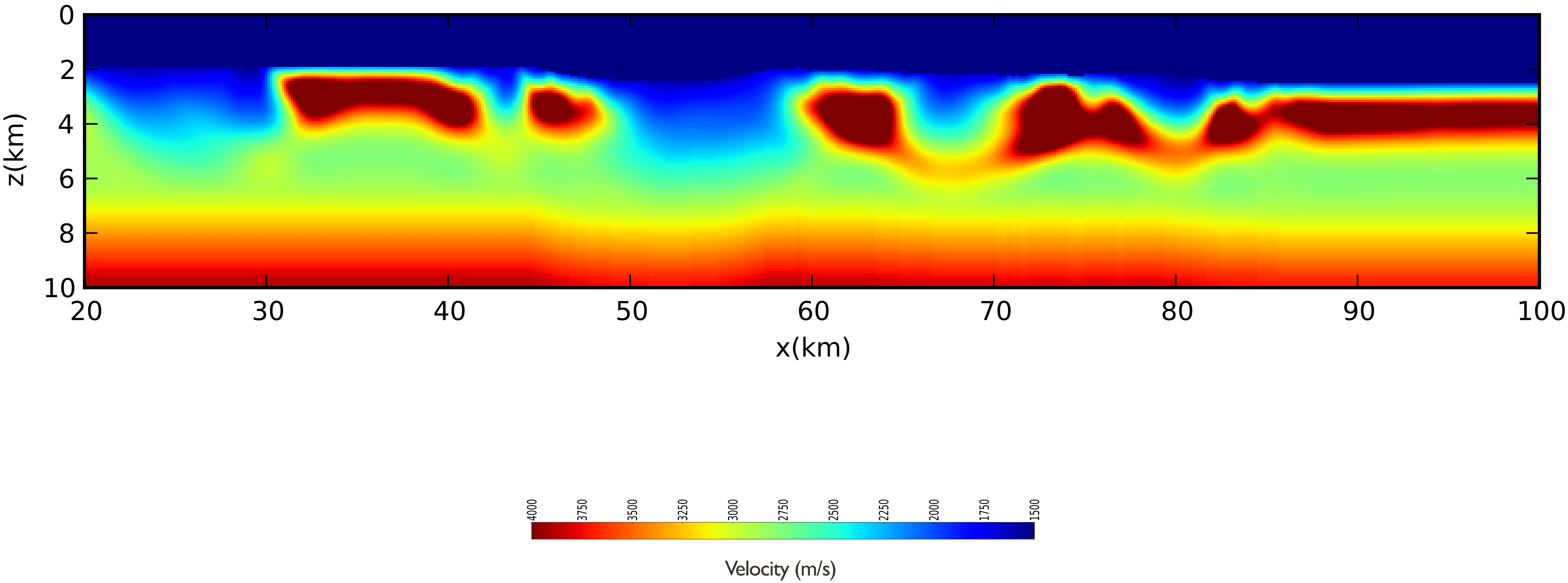
A. Aravkin, M.P. Friedlander, F.J. Herrmann and T. van Leeuwen - Robust inversion, dimensionality reduction and randomized sampling. Mathematical Programming, 2012.

Tristan van Leeuwen, “A parallel matrix-free framework for frequency-domain seismic modelling, imaging and inversion in Matlab”. 2012.

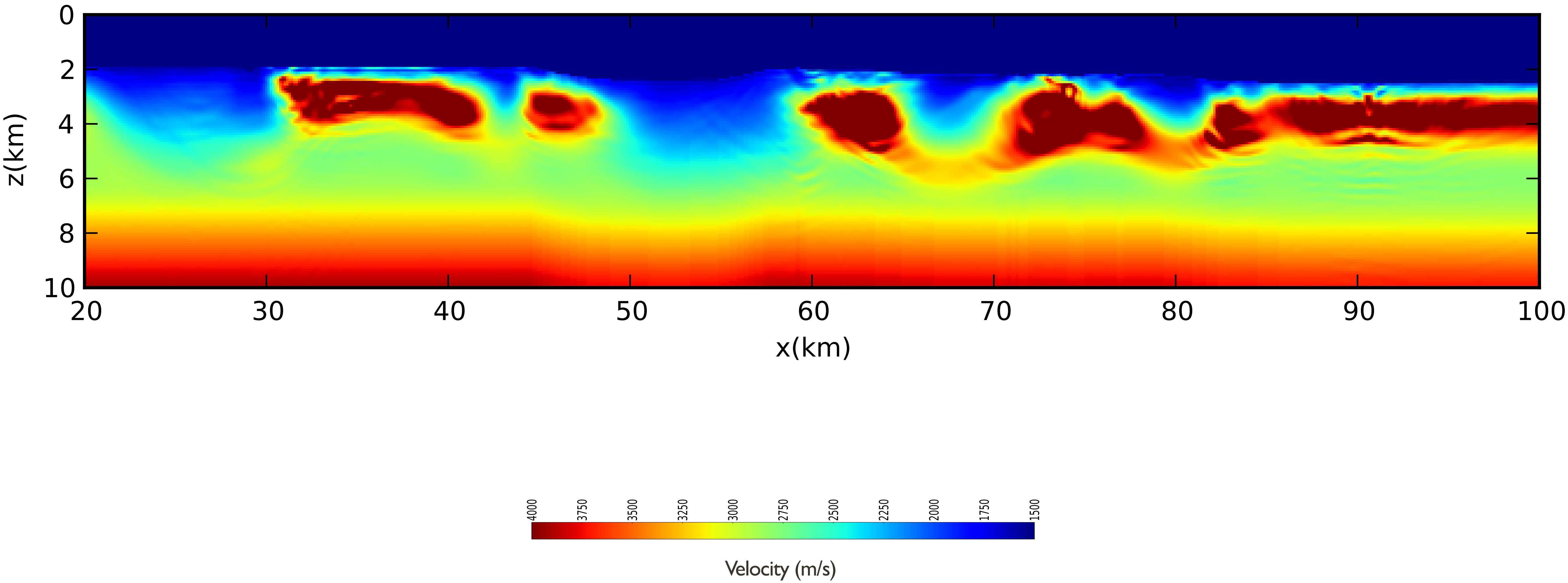
Tristan van Leeuwen and Felix J. Herrmann, “A parallel, object-oriented framework for frequency-domain wavefield imaging and inversion.”. 2012.



Input Model
[ray-based tomography + NMO]



Final result
[w/o denoising]

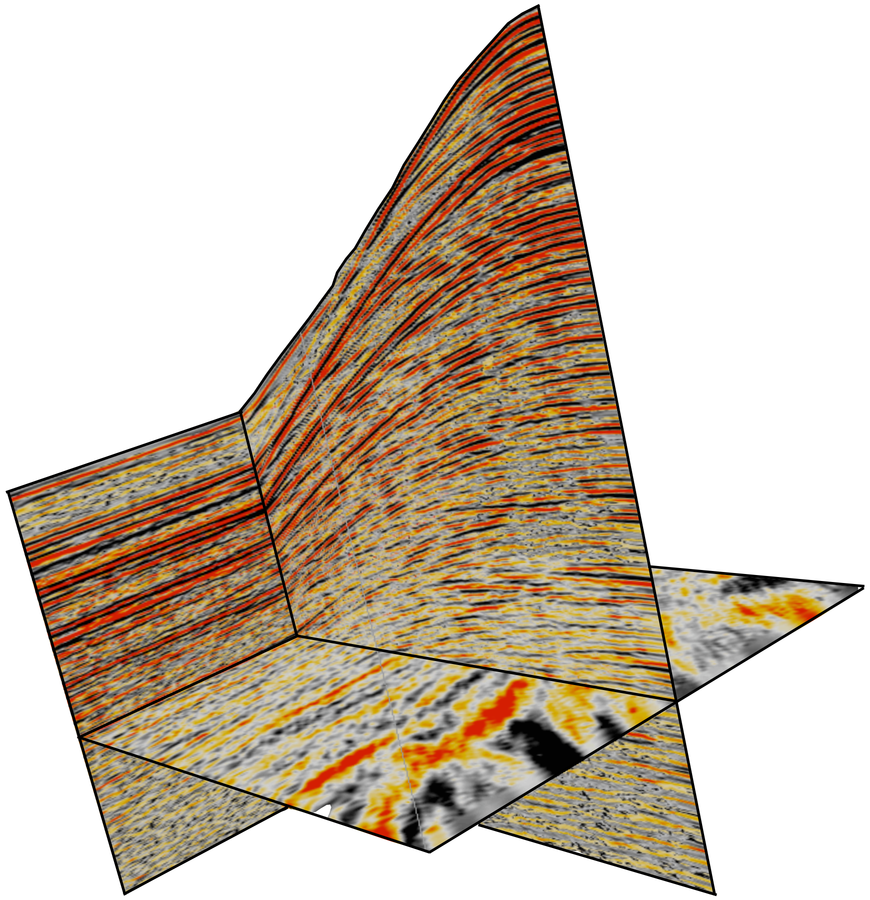


Modified Gauss-Newton

F. J. Herrmann, X. Li, A. Y. Aravkin, and T. van Leeuwen, “A modified, sparsity promoting, Gauss-Newton algorithm for seismic waveform inversion”, in Proc. SPIE, 2011, vol. 2011.

X. Li, F.J. Herrmann, A.Y. Aravkin and T. van Leeuwen - Fast randomized full waveform inversion with compressed sensing. Geophysics 77 (A13), 2012.

F.J. Herrmann, X. Li, A.Y. Aravkin, and T. van Leeuwen - A modified, sparsity-promoting, Gauss-Newton algorithm for seismic waveform inversion. Proc. SPIE 8138, 81380V (2011



Methodology

Modified Gauss-Newton:

- frequency continuation (7 bands 4 freqs, 2-5Hz)
- *rerandomized subsets* with *only* 600 shots
- 6 GN *iterations* per frequency *band*
- *preconditioning* of *Jacobian* by depth *weighting*
- *projection* of water *layer*

One-norm curvelet *regularized* gradients...

Algorithm
modified Gauss-Newton

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

$k \longleftarrow 0; \mathbf{m}^k \longleftarrow \mathbf{m}_0$

while not converged **do**

$\{\underline{\mathbf{D}}^k, \underline{\mathbf{Q}}^k\} \longleftarrow \{\mathbf{D}\mathbf{W}^k, \mathbf{Q}\mathbf{W}^k\}$ with $\mathbf{W}^k \subset [\mathbf{e}_1, \cdots, \mathbf{e}_{n_s}]$

$\underline{\delta\mathbf{D}}^k \longleftarrow \underline{\mathbf{D}}^k - \boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]$

$\tau^k \longleftarrow \|\underline{\delta\mathbf{D}}^k\|_F \,/\, \|\mathbf{C}_2 \nabla\boldsymbol{\mathcal{F}}^*[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\underline{\delta\mathbf{D}}^k\|_\infty$

$\delta\mathbf{x} \leftarrow \arg\min_{||\mathbf{x}||_1 \leq \tau_k} \|\underline{\delta\mathbf{D}}^k - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\mathbf{C}_2^H \mathbf{x}\|_F^2$

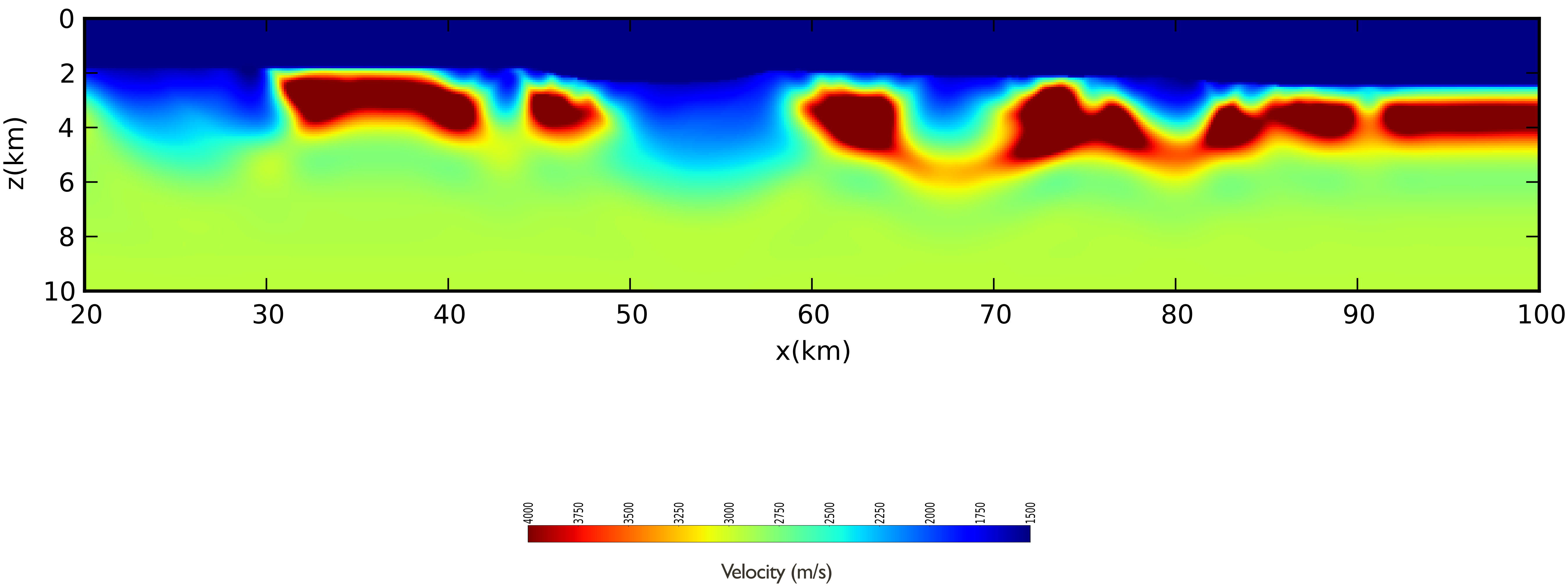
$\mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k \mathbf{C}_2^H \delta\mathbf{x}$

$k \longleftarrow k + 1;$

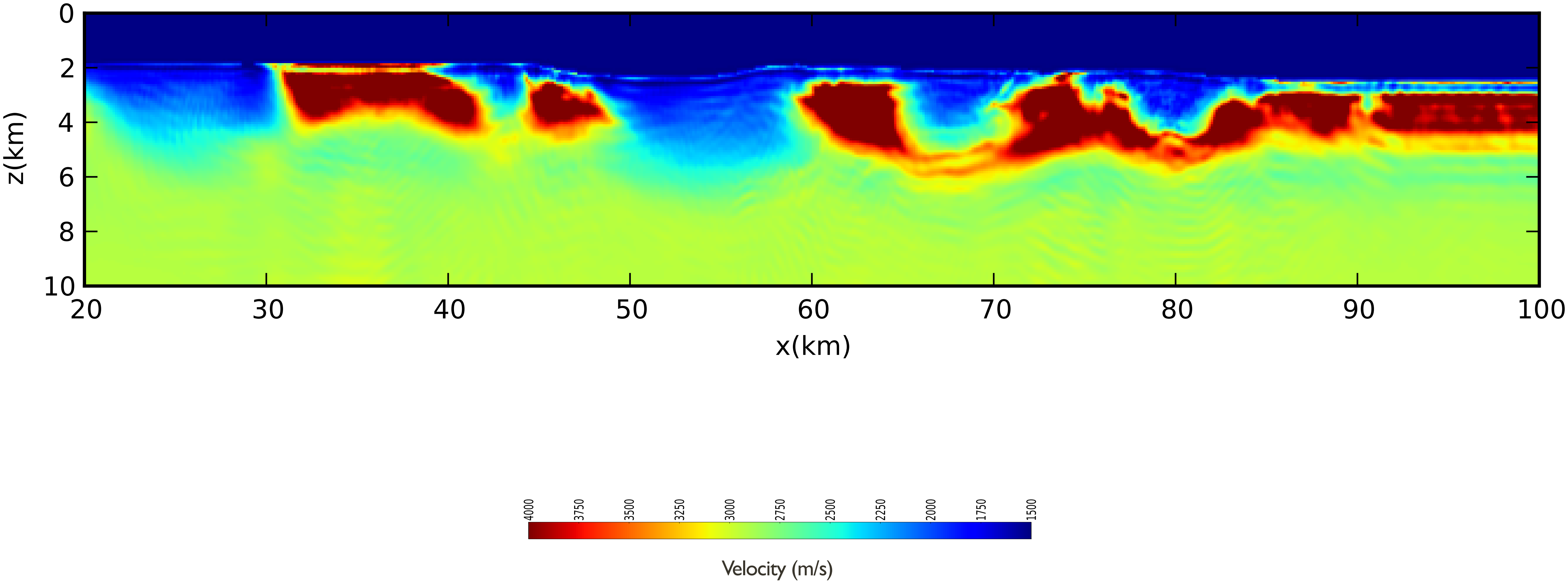
end

Algorithm 1: modified Gauss Newton with sparsity promotion

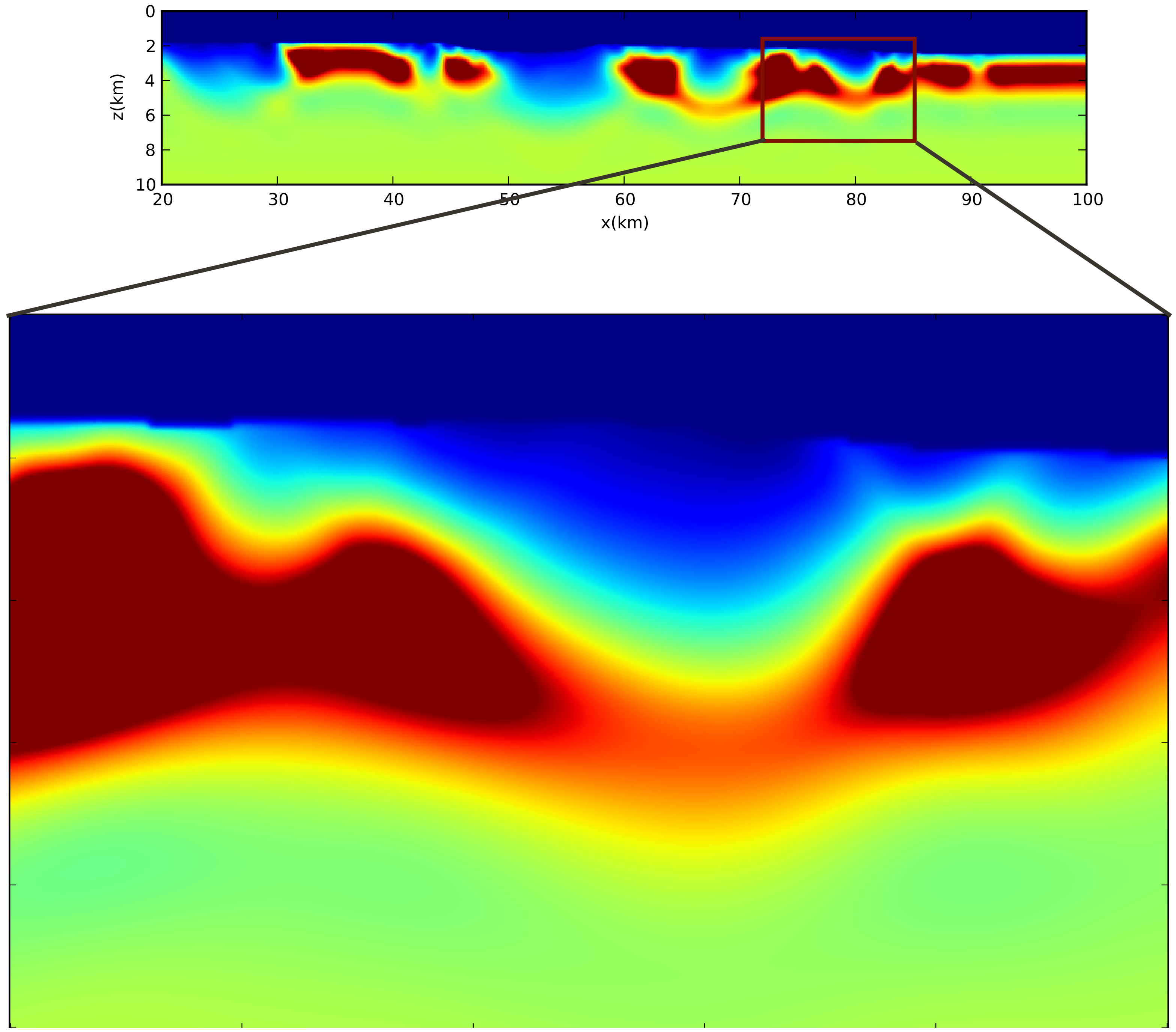
Initial model
[ray-based tomography]



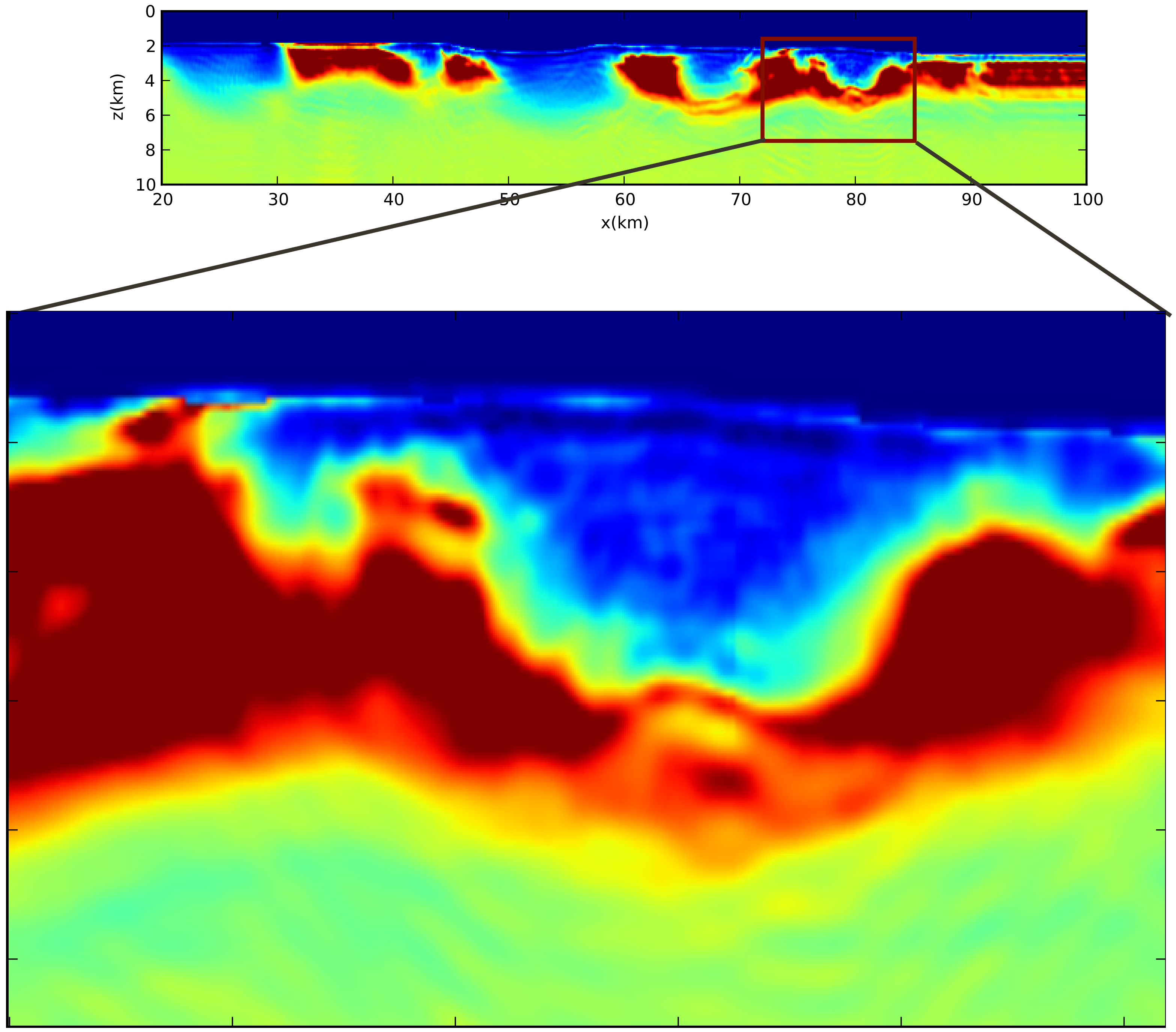
Final result



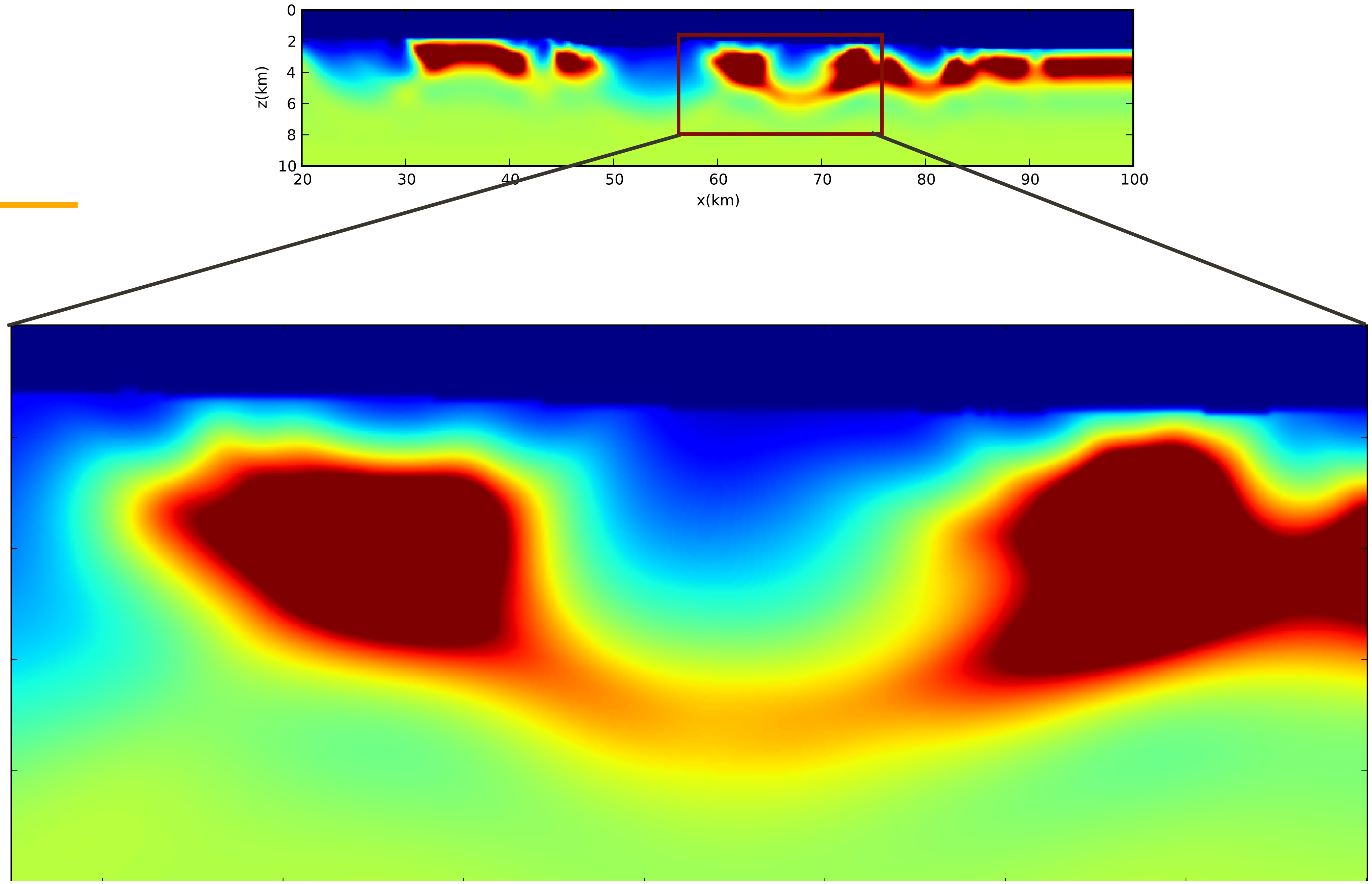
Initial Model



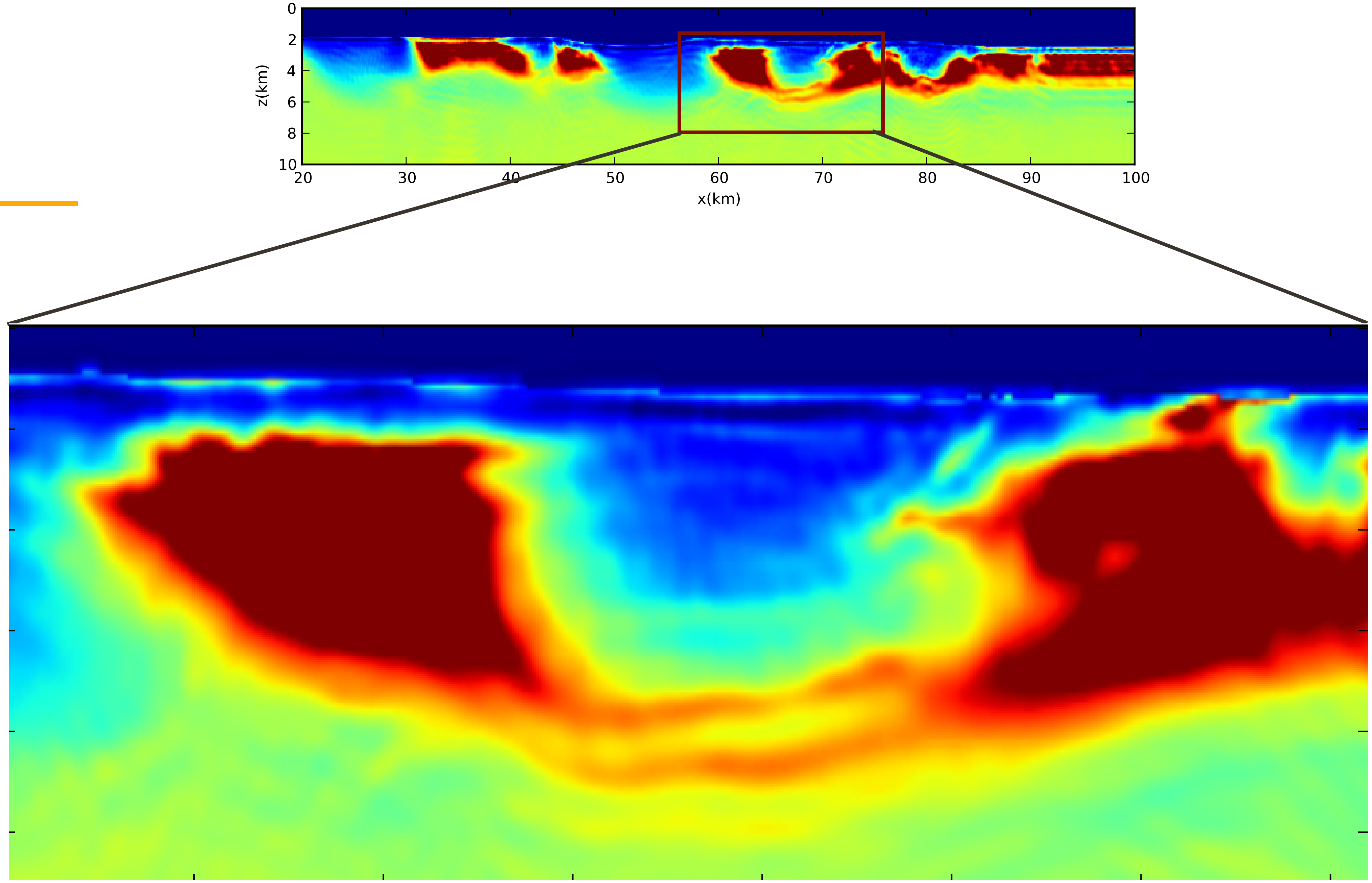
Final result
[w/ denoising & complex velocity]



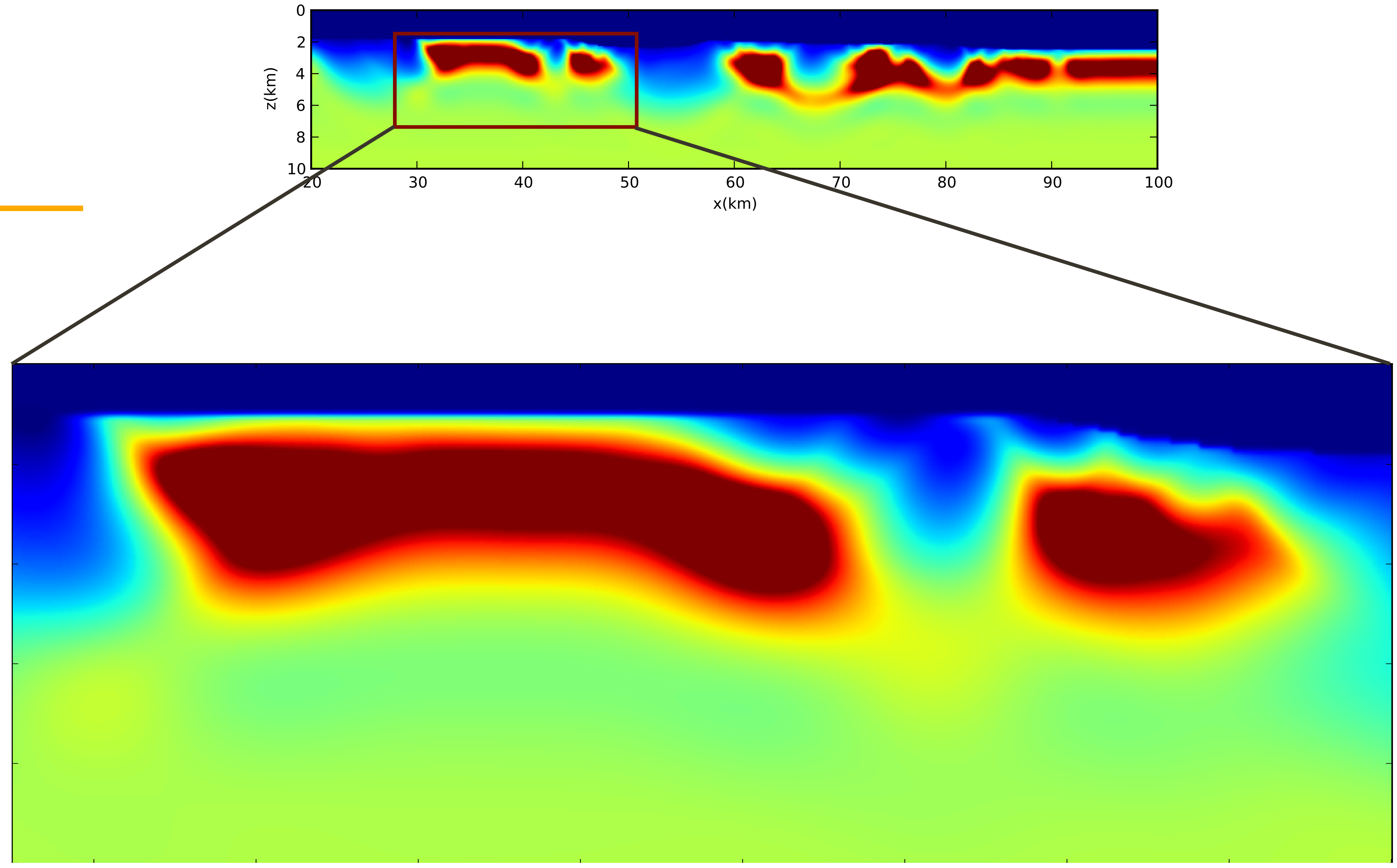
Initial Model



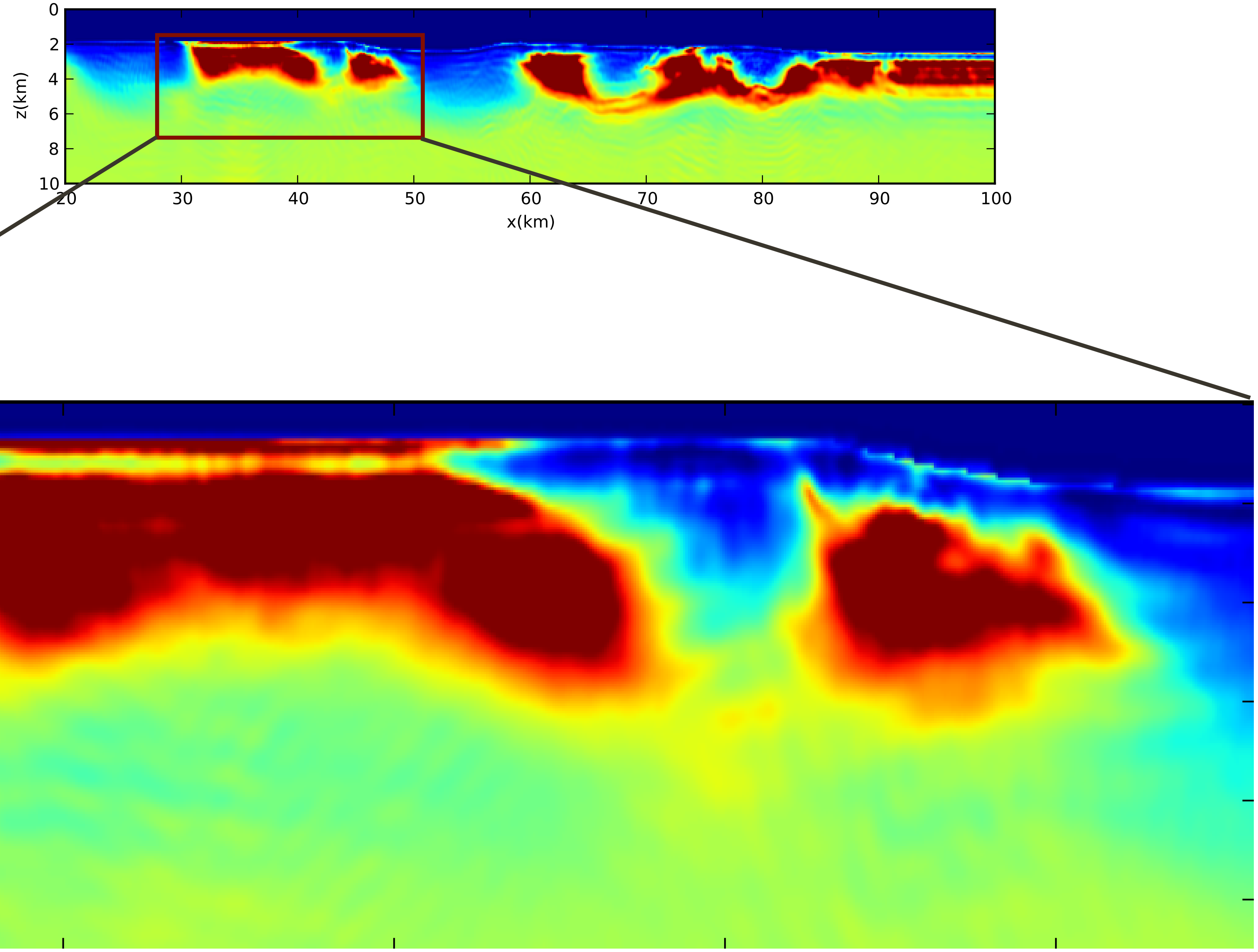
Final result
[w/ denoising & complex velocity]



Initial Model



Final result
[w/ denoising & complex velocity]



Observations

	Nonlin. CG [phase]	Modified GN [amp. & phase]
Acoustic	mixed results	encouraging
Noise	?	important
Conditioning	involved	“simple”
Quality	smooth	“sharp”

Conclusions

Dimensionality reduction techniques allow us to impose *sparsity* on the GN updates

GOM data set is *extremely* challenging

- *data*-space *regularization* is *insufficient*
- *model*-space regularization w/ *curvelets* is encouraging

Results *really* depend on *details*...

SLIM

Acknowledgments

We would like to thank Charles Jones from BG and Dimitri Bevc for providing for providing synthetic models. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada 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, BP, BGP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

NSERC
CRSNG

Thank you

www.slim.eos.ubc.ca