Wavefield-reconstruction inversion

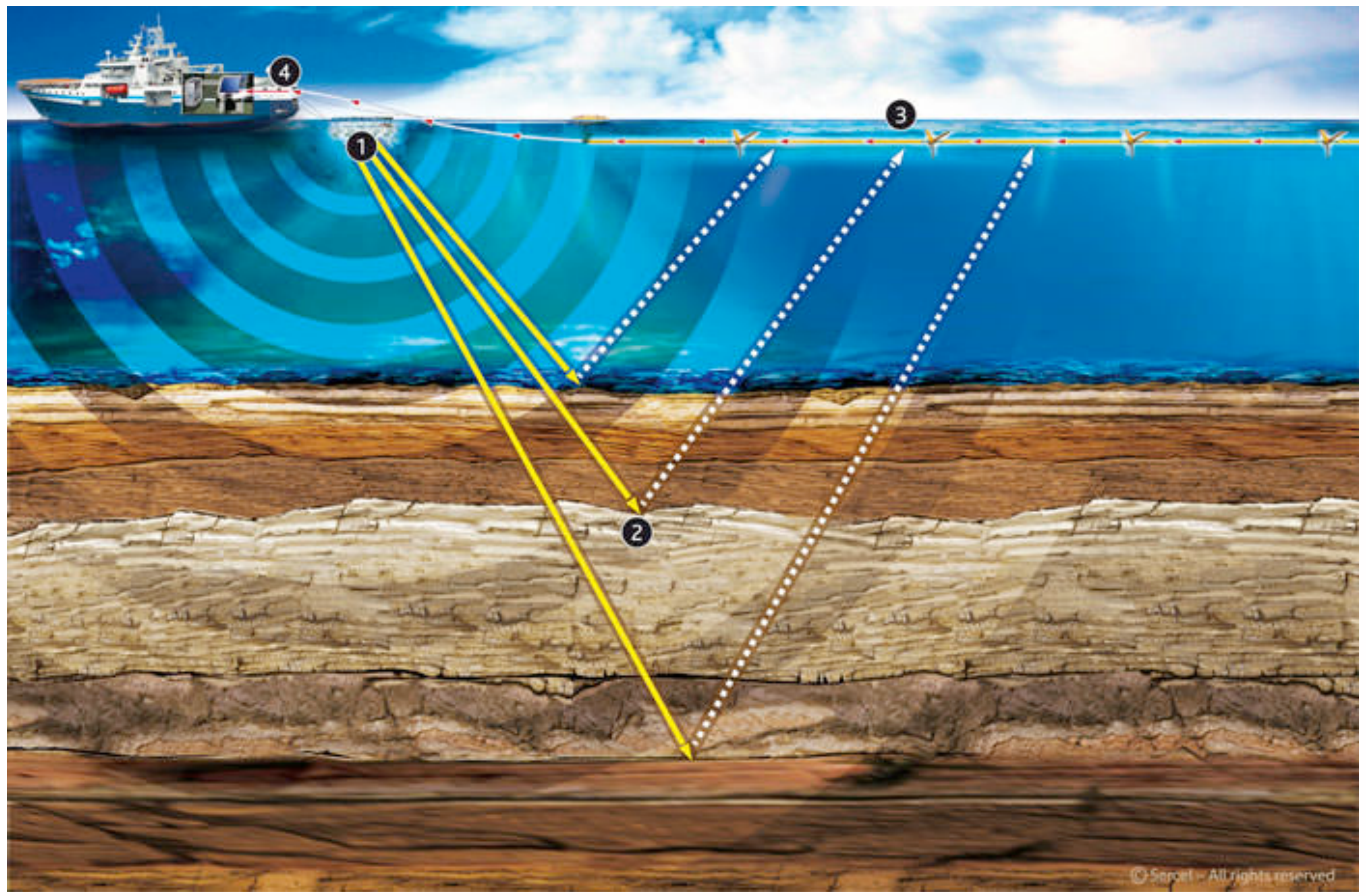
Bas Peters
Joint work with: Felix J. Herrmann, Tristan van Leeuwen

Applied Inverse Problems 2015
May 28

SLIM
University of British Columbia

PDE-constrained optimization

This talk is about parameter estimation with wavefields.



[from:http://www.sercel.com/about/Pages/what-is-geophysics.aspx]

PDE-constrained optimization

This talk is about parameter estimation with the Helmholtz equation.

Challenging because:

- data and predicted fields are both oscillatory
- non-convex
- local minimizers often unacceptable
- 1 PDE: ~ [1e6 – 1e9] grid points
- working with multiple [10 – 1000] PDE's simultaneously is very challenging

PDE-constrained optimization

known:

- source/receiver locations
- source function (sometimes)
- the PDE (usually simplified physics)

unknown:

- PDE-coefficients (acoustic velocity)

notation:

- fields ('state variables')
- medium parameters ('control variables')

PDE-constrained optimization

Use the 'discretize-then-optimize' framework:

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u} - \mathbf{d}\|_2^2 \quad \text{s.t.} \quad H(\mathbf{m})\mathbf{u} = \mathbf{q}$$

$H(\mathbf{m}) \in \mathbb{C}^{N\times N}$ discrete PDE

$\mathbf{m} \in \mathbb{R}^N$ medium parameters

$P \in \mathbb{R}^{m\times N}$ selects field at receivers

$\mathbf{u} \in \mathbb{C}^N$ field

$\mathbf{d} \in \mathbb{C}^m$ observed data

$\mathbf{q} \in \mathbb{C}^N$ source

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 \quad \text{s.t.} \quad H(\mathbf{m})\mathbf{u}=\mathbf{q}$$

$$\mathcal{L}(\mathbf{m},\mathbf{u},\boldsymbol{\gamma}) = \frac{1}{2}\|P\mathbf{u}-\mathbf{d}\|_2^2 + \boldsymbol{\gamma}^*\left(H(\mathbf{m})\mathbf{u}-\mathbf{q}\right)$$

eliminate field variables

$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2$$

reduced space

[A Tarantola, 1984; E Haber et al., 2000; I Epanomeritakis et al., 2008]

- storage as low as two fields at a time
- highly nonlinear function value computation is
  - expensive
  - inexact when sub-problems are solved iteratively
- dense reduced-Hessian
- requires extra safeguards/accuracy control

[T. van Leeuwen & F.J. Herrmann, 2014]

Reduced-space PDE-constrained optimization

Example 1 (easy):

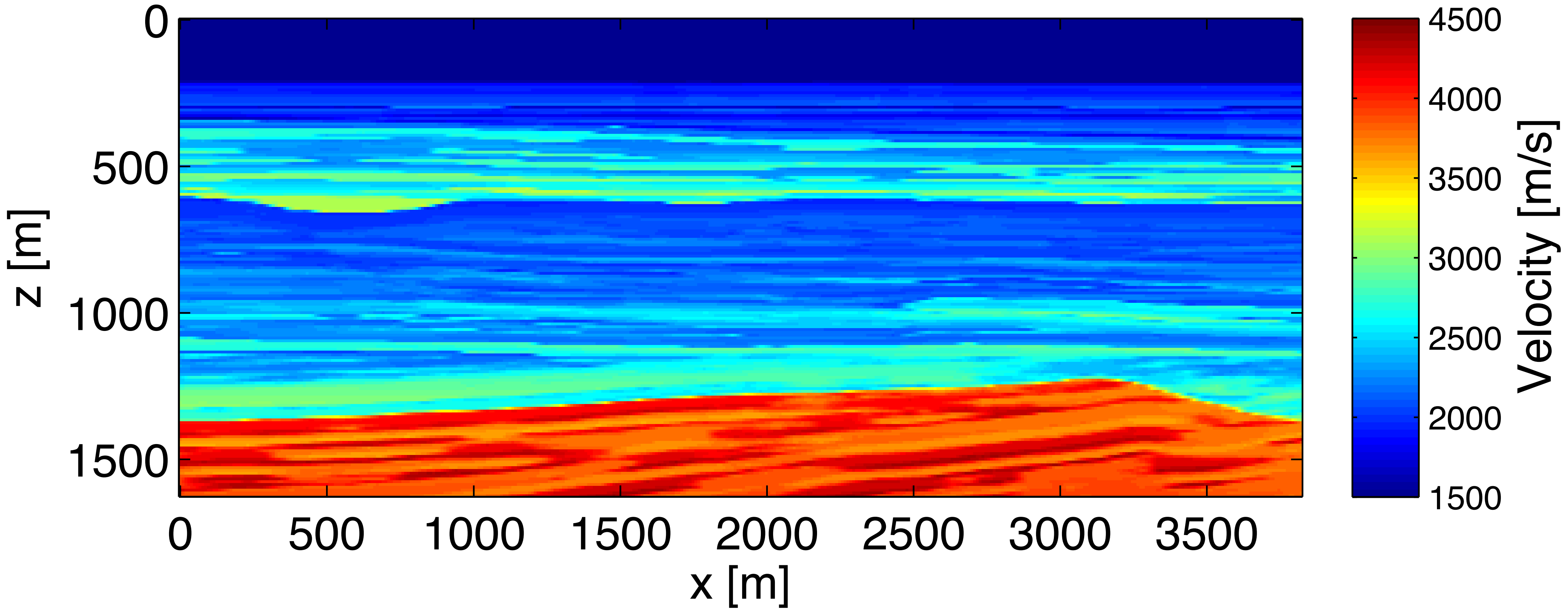
- Two-metric-projection with L-BFGS Hessian and bound constraints

[M. Schmidt et. al., 2012]

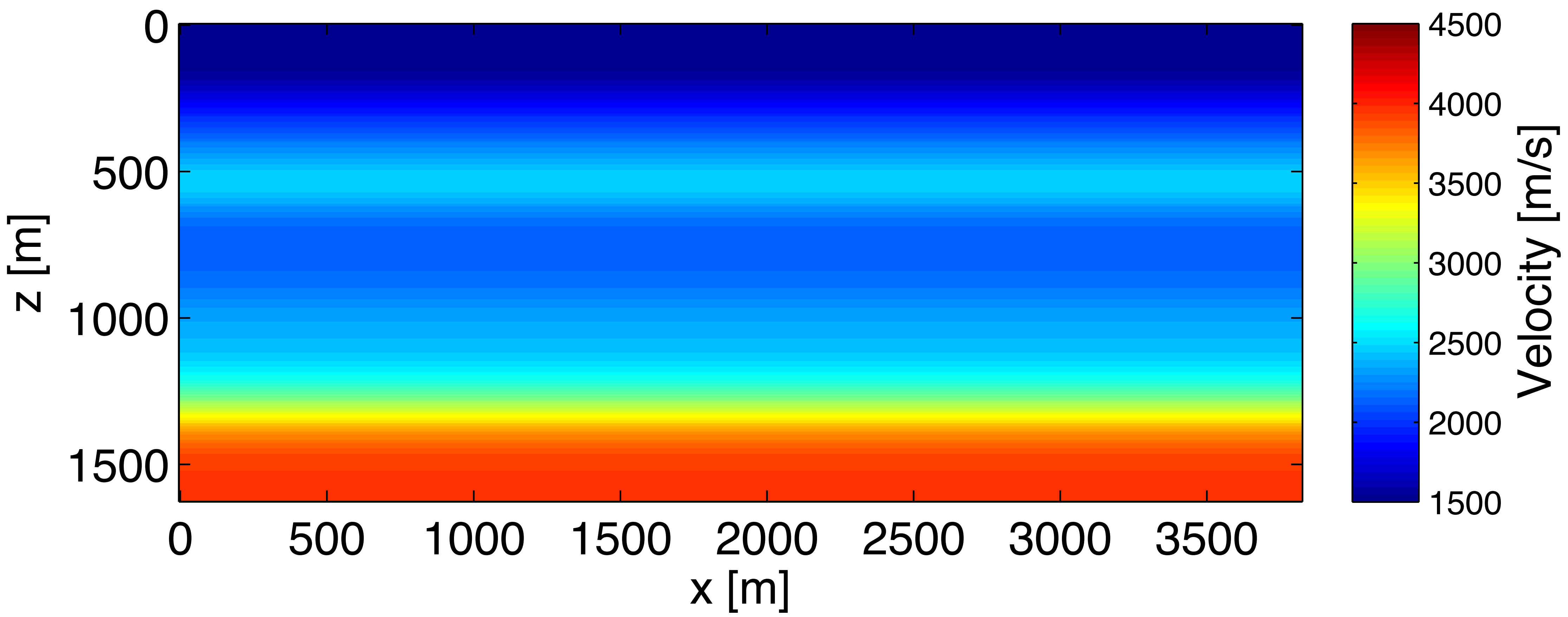
- 64 equally distributed sources and receivers near the surface
- 18 frequency batches: {2 3}, {3 4} ,... ,{19 20} Hertz
- No noise
- No regularization

True, initial and final models

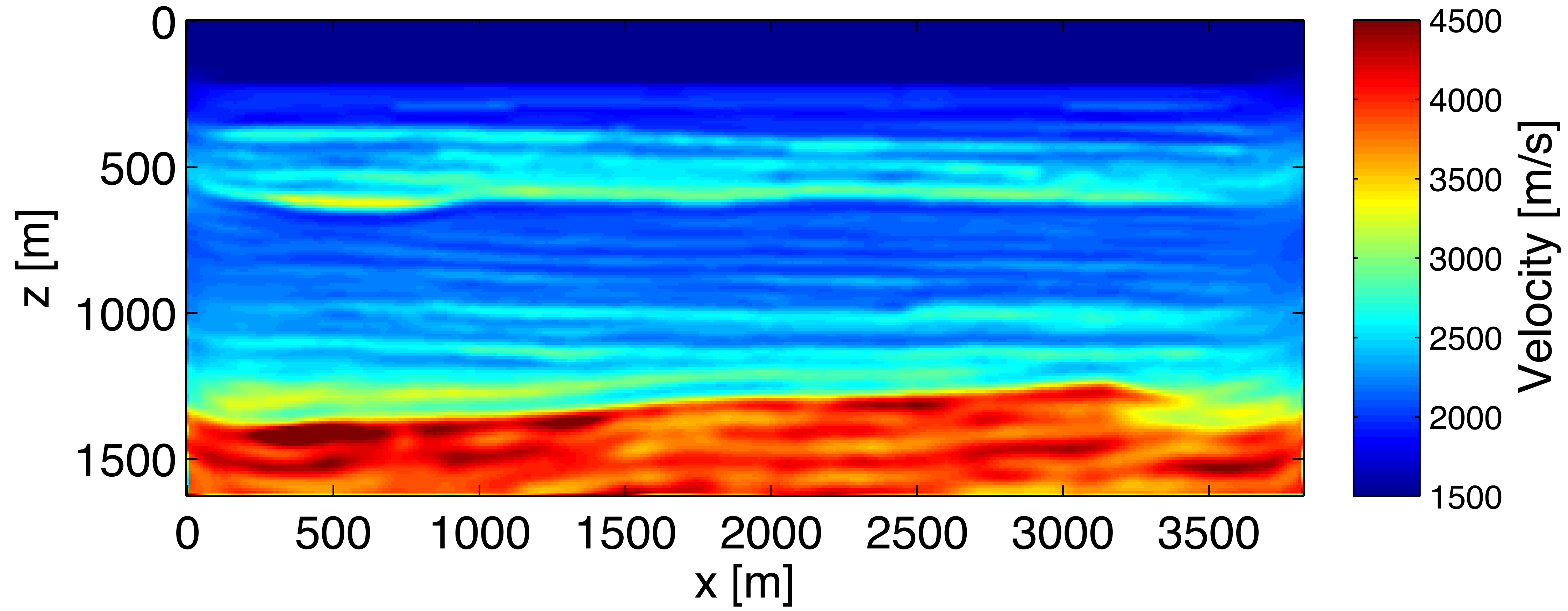
True model



Initial model



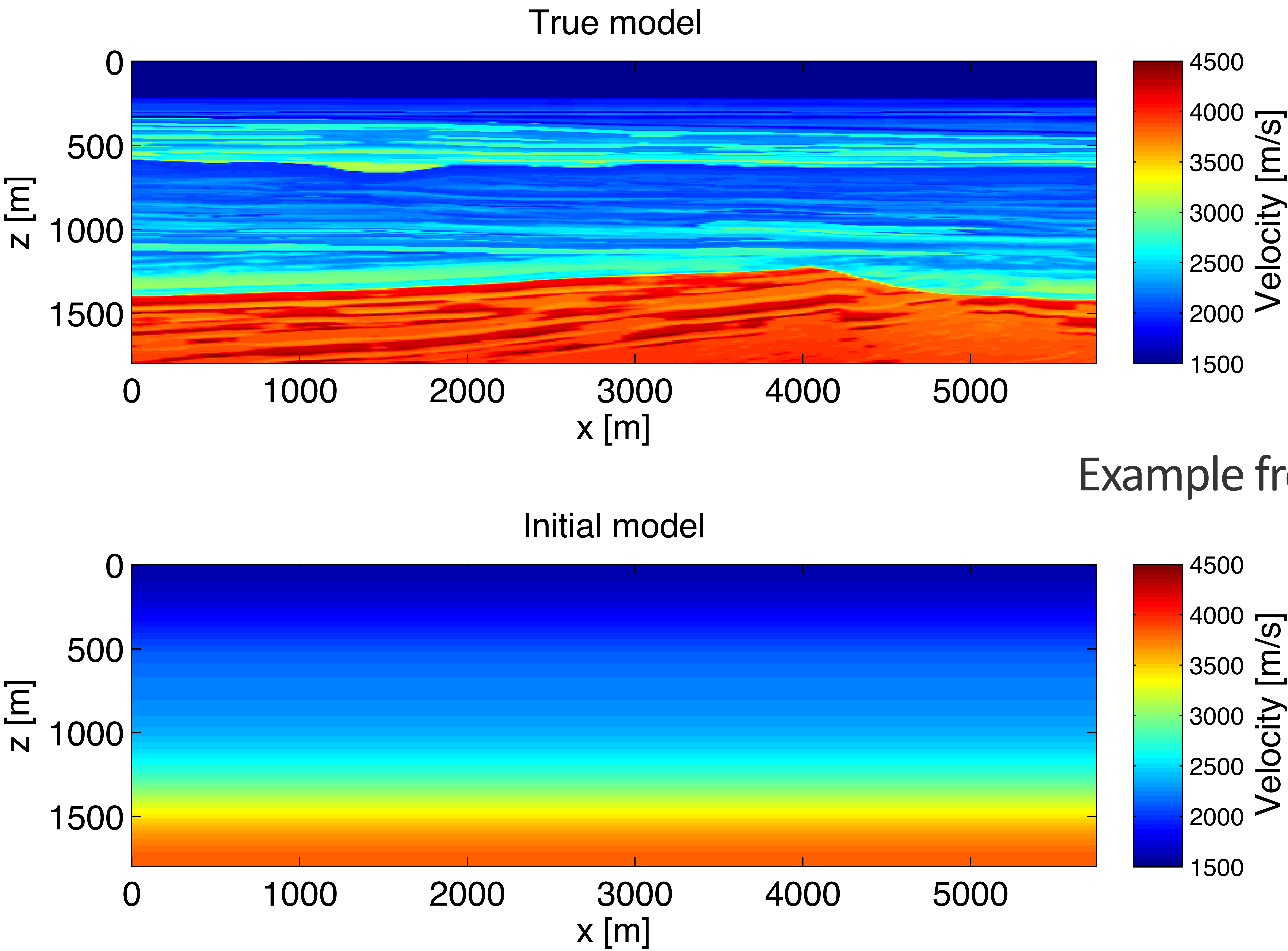
Result reduced Lagrangian



Reduced-space PDE-constrained optimization

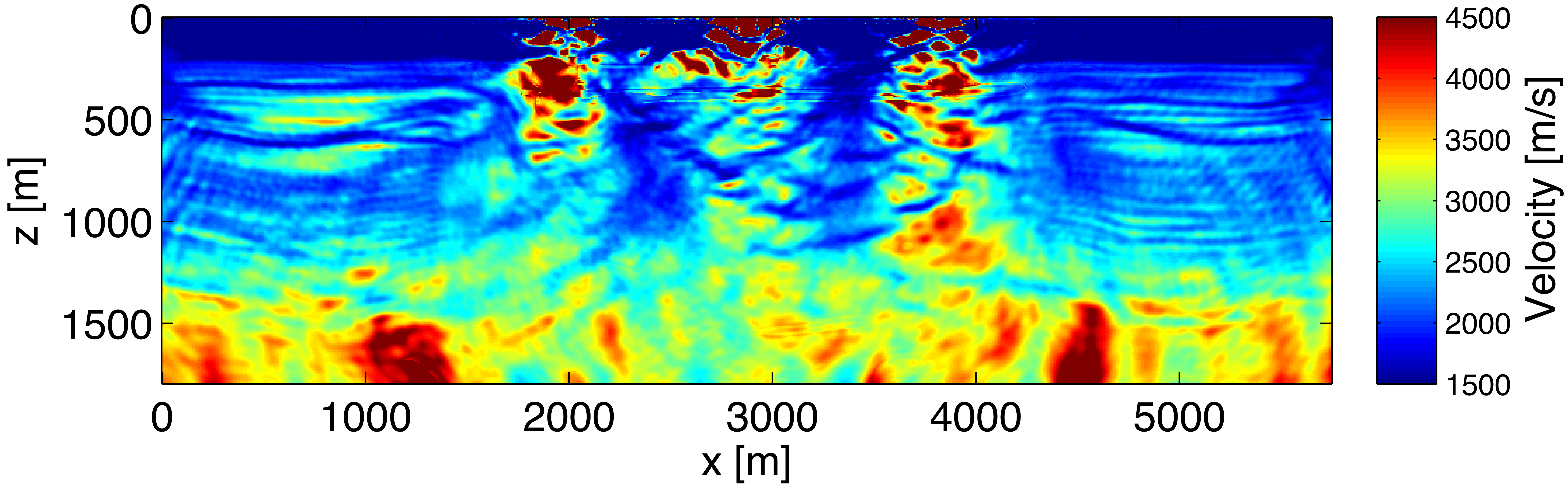
Example 2 (difficult):

- Frequency band is {5-28} Hz instead of {2-20} Hz
- Less accurate initial model



Example from **[Peters et al. 2014]**

$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|_2^2$$



Reduced-space PDE-constrained optimization

- Eliminating PDE-constraints implies PDE (with incorrect coefficients) is always satisfied.
- Minimizing the difference between oscillatory observed and predicted data only works if they are within a ‘cycle’ of each other.
- Known as the ‘cycle-skipping’ problem.

- This type of parameter estimation works well for accurate starting models.

Flipping objective and constraints

If the observed and predicted data need to be close, why not solve:

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|H(\mathbf{m})\mathbf{u} - \mathbf{q}\|_2^2 \quad \text{s.t.} \quad P\mathbf{u} = \mathbf{d}$$

instead of the common choice

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u} - \mathbf{d}\|_2^2 \quad \text{s.t.} \quad H(\mathbf{m})\mathbf{u} = \mathbf{q}$$

Flipping objective and constraints

If the observed and predicted data need to be close, why not solve (with noise):

$$\min_{\mathbf{m},\mathbf{u}} \|H(\mathbf{m})\mathbf{u} - \mathbf{q}\|_2^2 \quad \text{s.t.} \quad \|P\mathbf{u} - \mathbf{d}\|_2^2 \leq \sigma$$

Can solve this using a quadratic-penalty method:

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\mathbf{u} - \mathbf{q}\|_2^2$$

few algorithms are based on the quadratic-penalty form:

[R.E. Kleinman & P.M.van den Berg, 1992 ; T. van Leeuwen & F.J. Herrmann, 2013]

$$\min_{\mathbf{m},\mathbf{u}} \|H(\mathbf{m})\mathbf{u} - \mathbf{q}\|_2^2 \quad \text{s.t.} \quad \|P\mathbf{u} - \mathbf{d}\|_2^2 \leq \sigma$$

$$\downarrow$$

$$\min_{\mathbf{m},\mathbf{u}} \frac{1}{2}\|P\mathbf{u} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\mathbf{u} - \mathbf{q}\|_2^2$$

$$\downarrow$$

eliminate field variables: solve $\nabla_{\mathbf{u}}\phi(\mathbf{m}, \bar{\mathbf{u}}, \lambda) = 0$

[T. van Leeuwen & F.J. Herrmann, 2013]

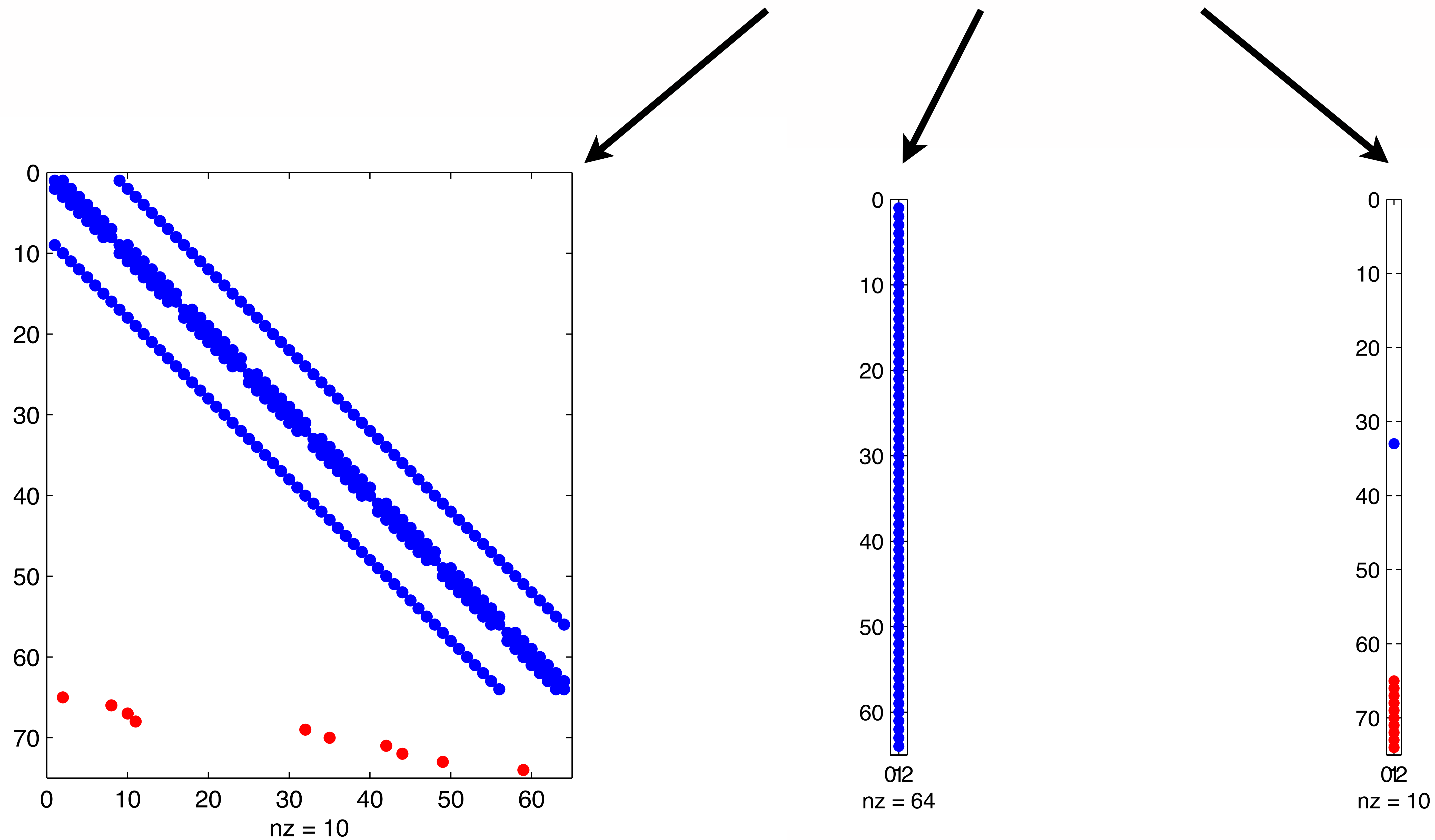
$$\downarrow$$

$$\min_{\mathbf{m}} \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$$

reduced quadratic-penalty

reduced quadratic-penalty: $\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda) = \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$

$$\nabla_{\mathbf{u}}\phi(\mathbf{m}, \bar{\mathbf{u}}, \lambda) = 0 \longleftrightarrow \bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$



Workflow

- choose σ in: $\min_{\mathbf{u}} \|H(\mathbf{m}_0)\mathbf{u} - \mathbf{q}\|_2^2 \quad \text{s.t.} \quad \|P\mathbf{u} - \mathbf{d}\|_2^2 \leq \sigma$

- Find λ corresponding to σ for the initial model, based on a small data sample. [W. Gander, 1980; A. Bjork, 1996]
- keep λ fixed

- solve: $\min_{\mathbf{m}} \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$

SLIM

[T. van Leeuwen & F.J. Herrmann, 2013]

A reduced-space quadratic-penalty method

To minimize: $\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda) = \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$

at every iteration:

- compute $\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$
- evaluate $\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda)$ & $\nabla_{\mathbf{m}}\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda)$
- update $\mathbf{m}$

SLIM

[T. van Leeuwen & F.J. Herrmann, 2013]

A reduced-space quadratic-penalty method

To minimize: $\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda) = \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$

at every iteration:

- compute $\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$
- evaluate $\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda)$ & $\nabla_{\mathbf{m}}\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda)$
- update $\mathbf{m}$

+ Trust-region / line-search

A reduced-space quadratic-penalty method

$$\bar{\phi}(\mathbf{m}, \bar{\mathbf{u}}, \lambda) = \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$$

for true fields: $P\mathbf{u}_* = \mathbf{d}$

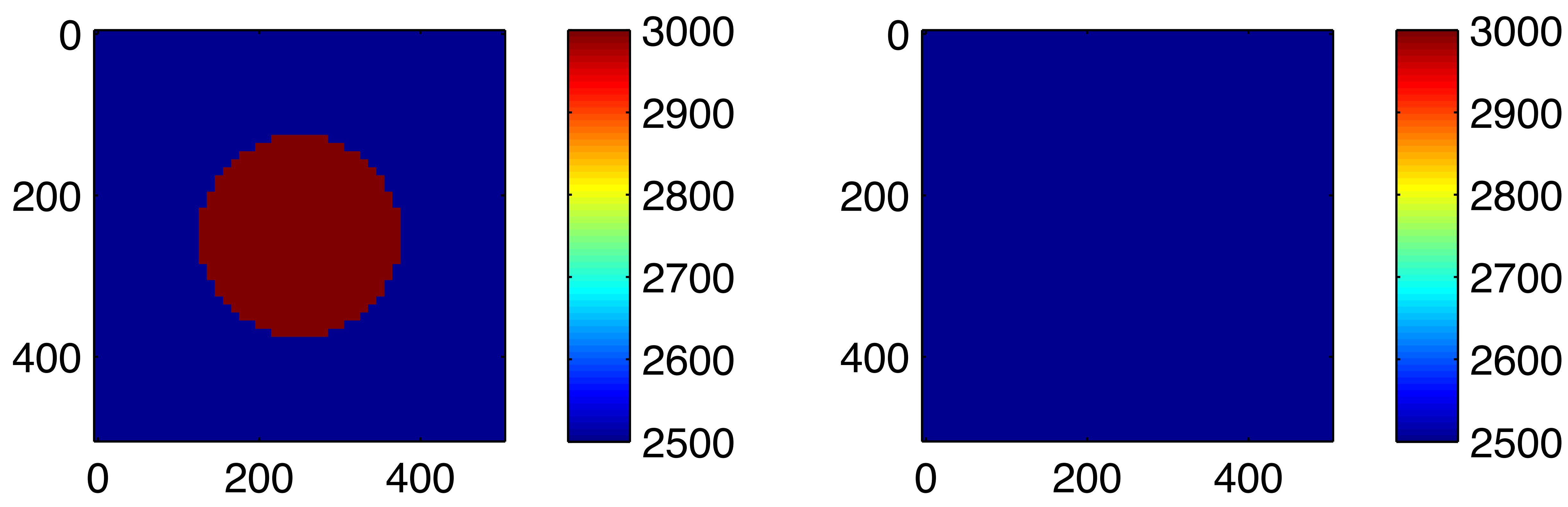
for true medium parameters: $H(\mathbf{m}_*)\mathbf{u}_* = \mathbf{q}$

➡ $\bar{\phi}(\mathbf{m}_*, \mathbf{u}_*, \lambda) = 0$ for any λ

Suggests no continuation strategy of λ is required in practice.

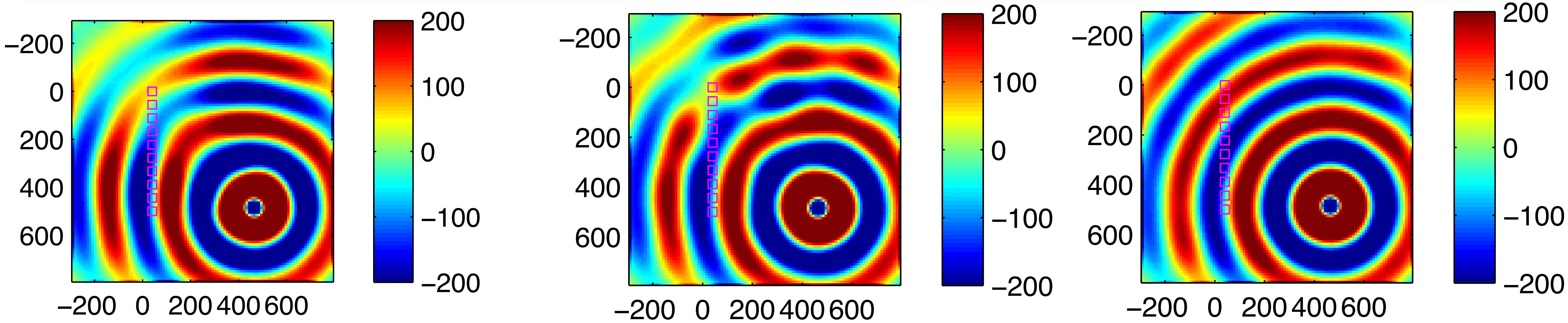
numerical tests support this

2D example



$$\mathbf{u} = H(\mathbf{m}_*)^{-1}\mathbf{q}$$

$$\mathbf{u} = H(\mathbf{m}_0)^{-1}\mathbf{q}$$

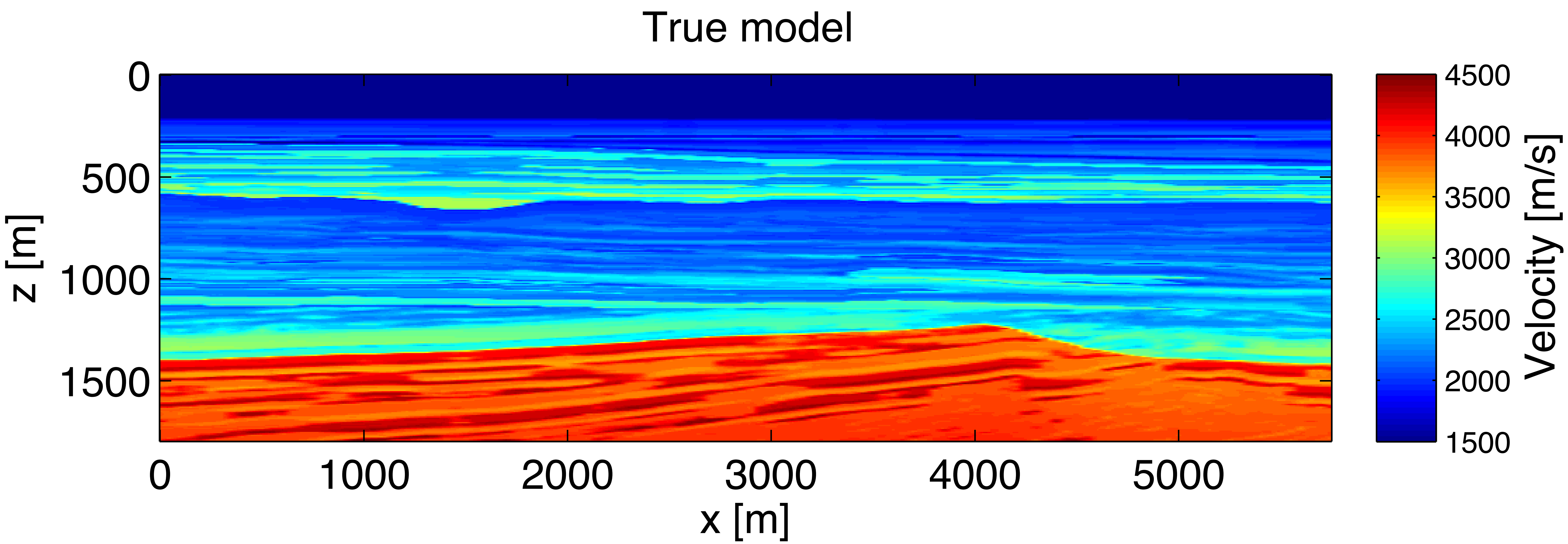


$$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}_0) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$

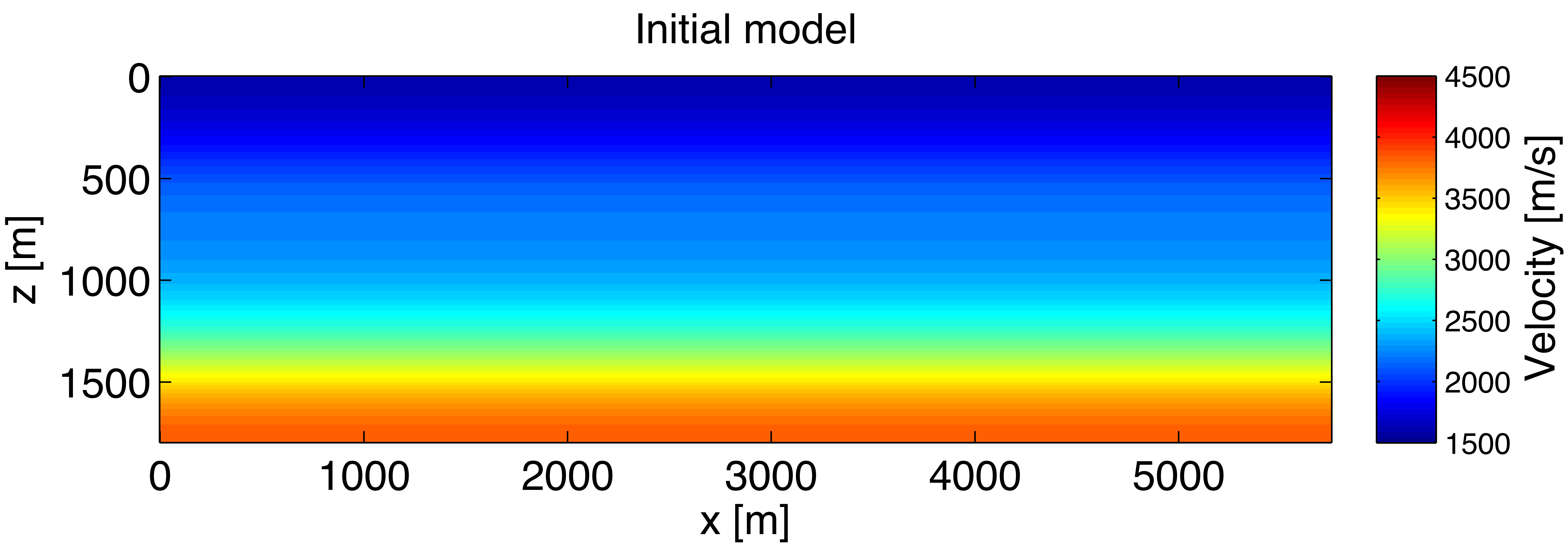
Reduced-space PDE-constrained optimization

Example 2, revisited:

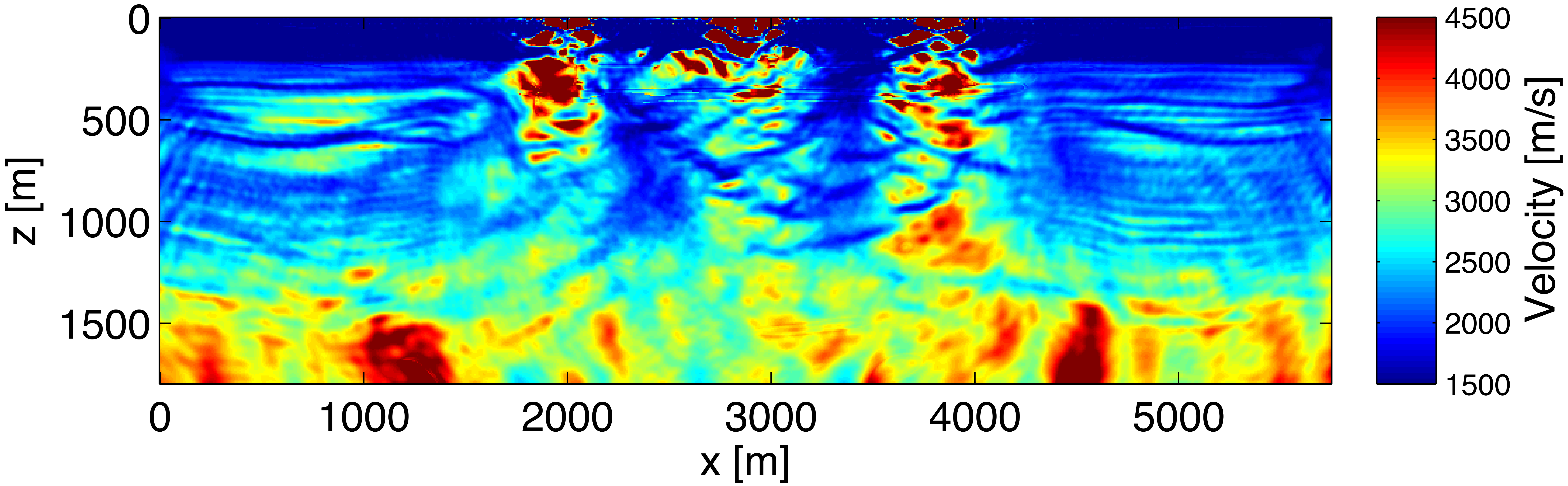
- Frequency band is {5-28} Hz instead of {2-20} Hz
- Less accurate initial model



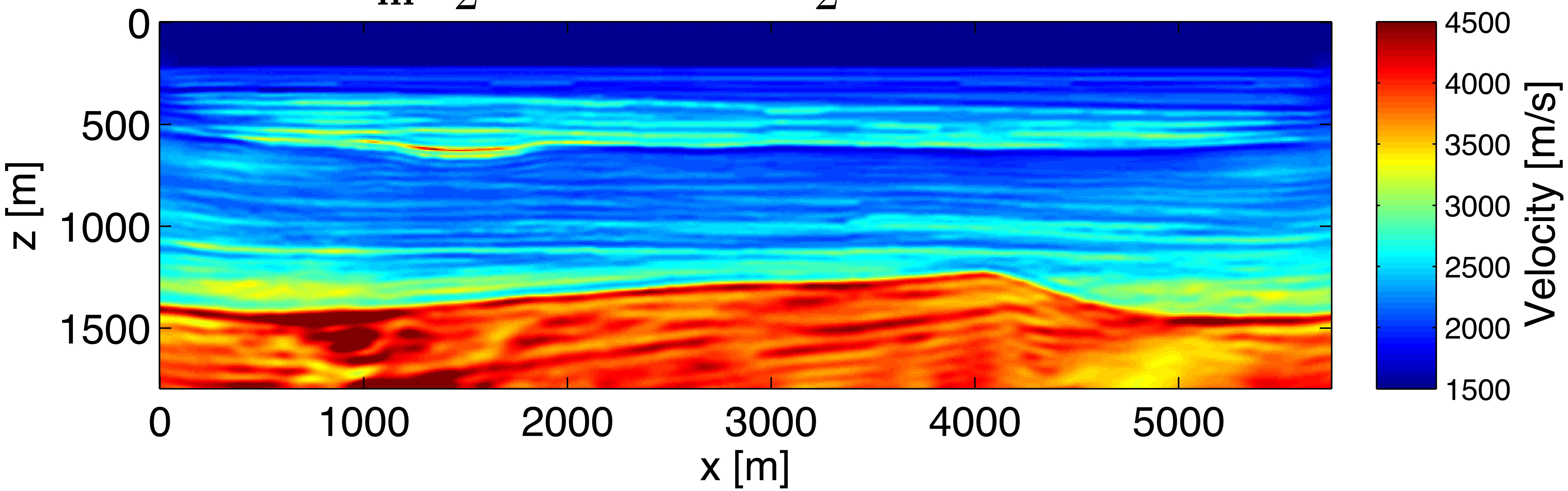
Example from **[Peters et al. 2014]**



$$\min_{\mathbf{m}} \frac{1}{2}\|PH(\mathbf{m})^{-1}\mathbf{q} - \mathbf{d}\|_2^2$$



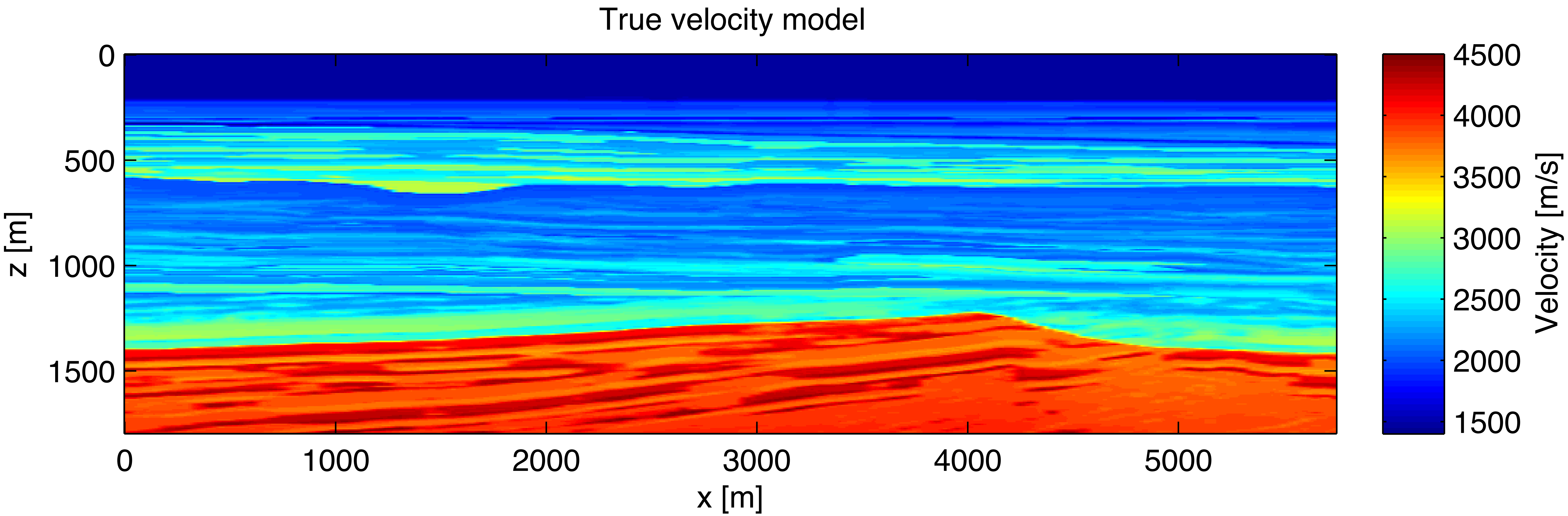
$$\min_{\mathbf{m}} \frac{1}{2}\|P\bar{\mathbf{u}} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|H(\mathbf{m})\bar{\mathbf{u}} - \mathbf{q}\|_2^2$$ (small λ)

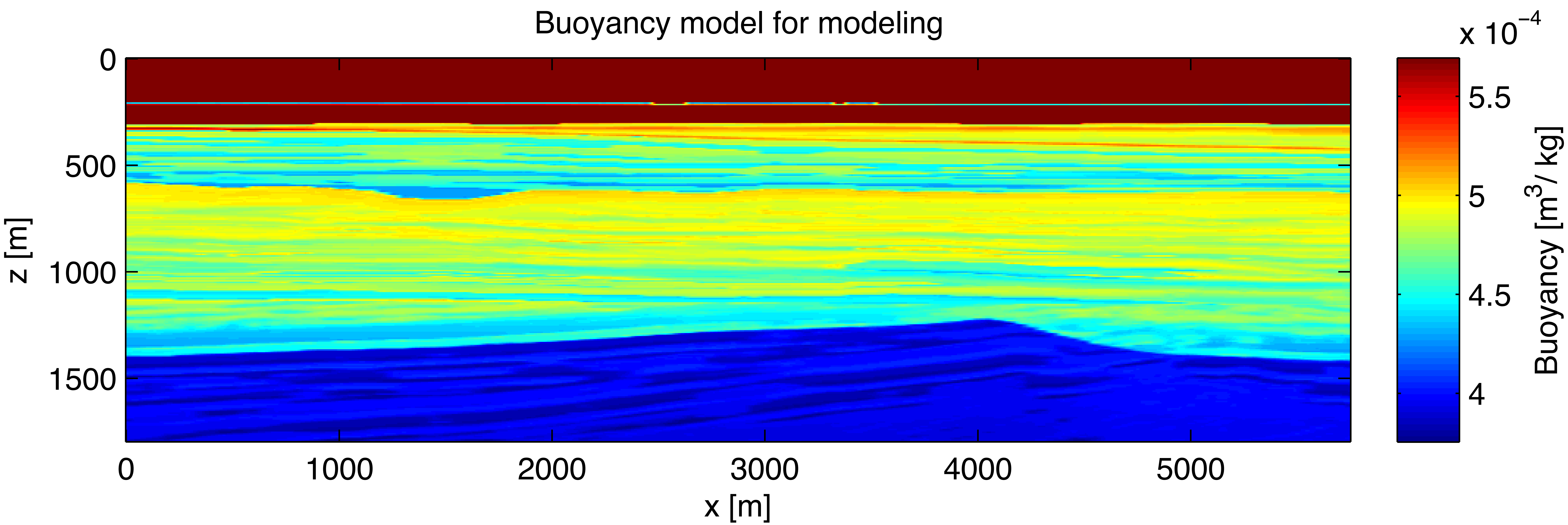


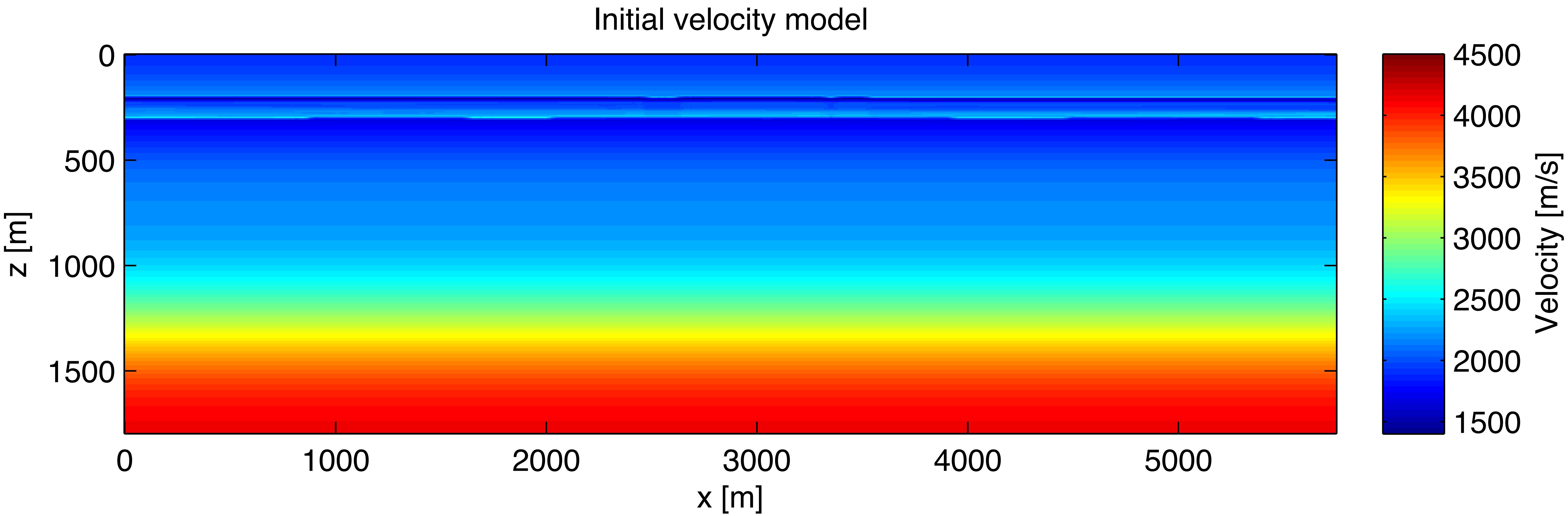
Example – BG Compass model no inverse crime

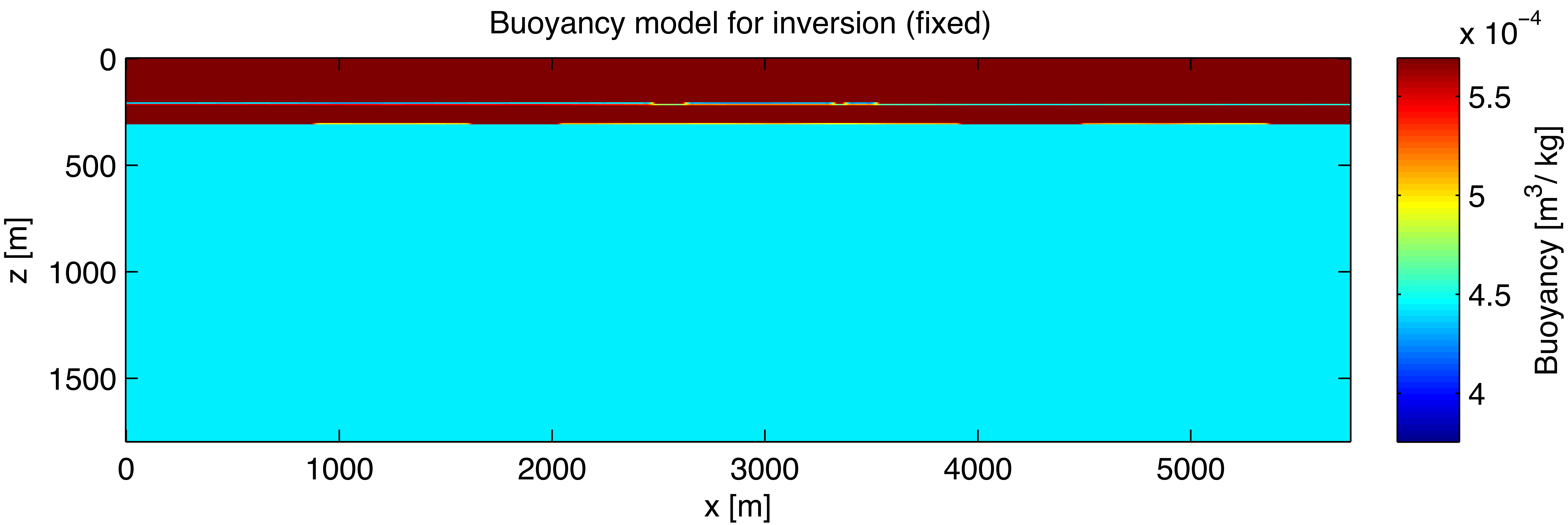
- Generate 'observed' data using a compressibility and buoyancy model.
- Invert for compressibility, fixed and inaccurate buoyancy.
- Obtain velocity model from inverted compressibility and fixed inaccurate buoyancy.

- 15 frequency batches {5 6} ,{6 7},... ,{19 20} Hertz. Each interval contains 5 frequencies.

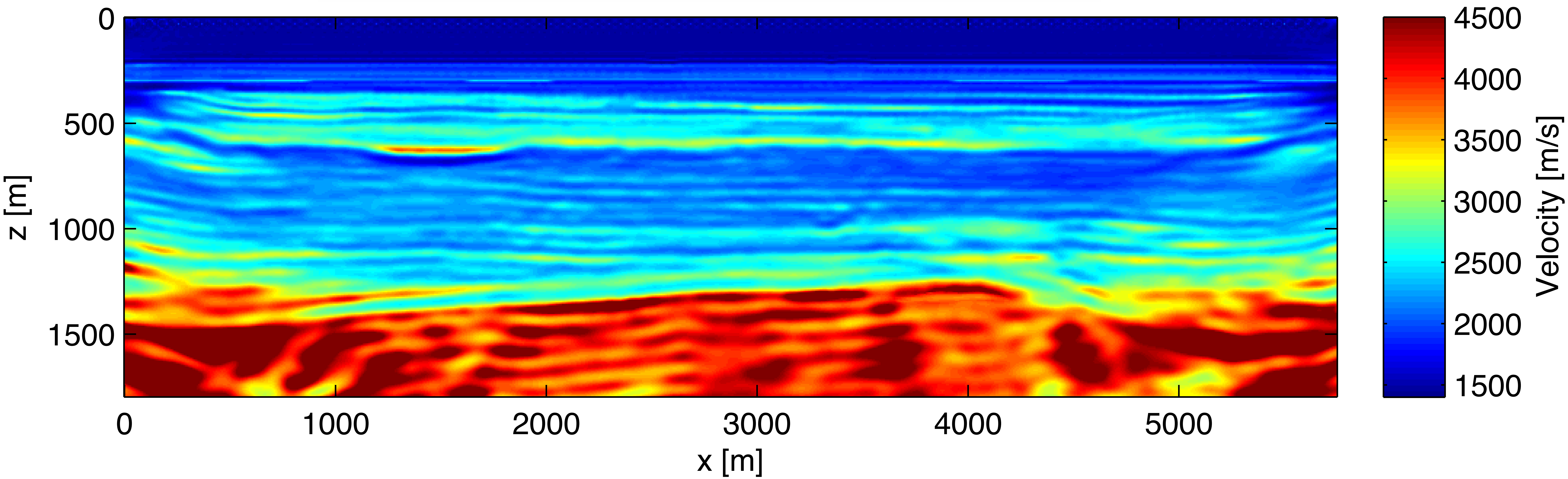

True velocity model
z [m]
0
500
1000
1500
x [m]
0
1000
2000
3000
4000
5000
Velocity [m/s]
4500
4000
3500
3000
2500
2000
1500


Buoyancy model for modeling
z [m]
0
500
1000
1500
x [m]
0
1000
2000
3000
4000
5000
x 10^{-4}
Buoyancy [m^3/ kg]
5.5
5
4.5
4

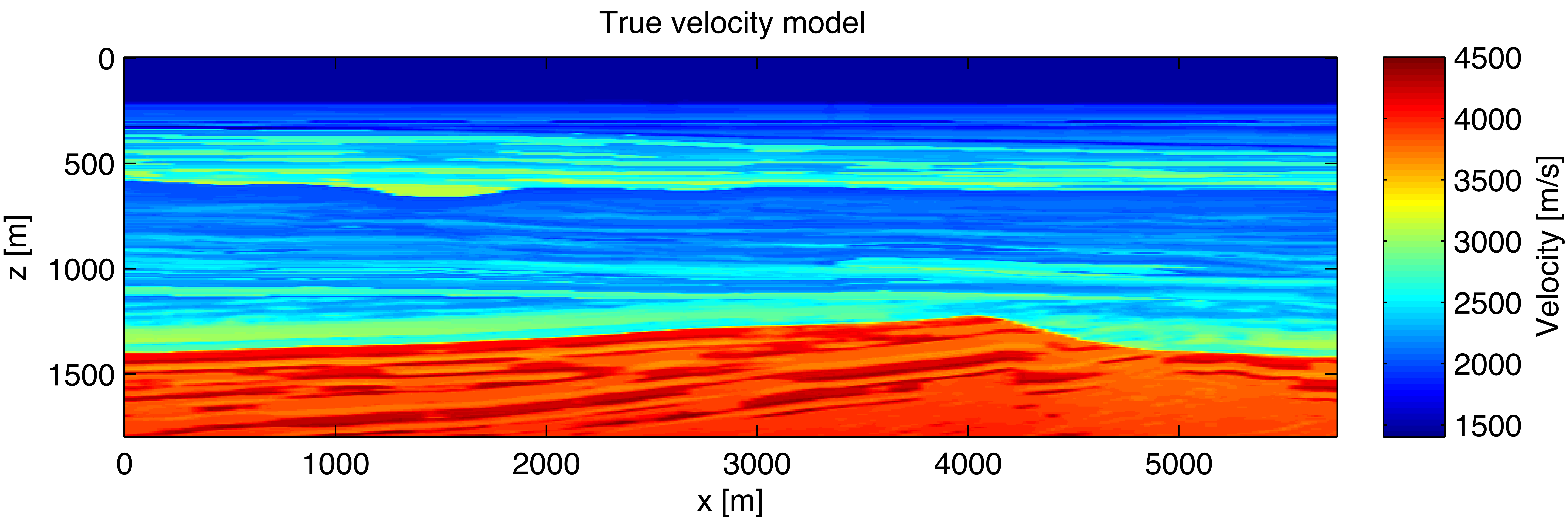

Initial velocity model
z [m]
x [m]
Velocity [m/s]
0
500
1000
1500
0
1000
2000
3000
4000
5000
4500
4000
3500
3000
2500
2000
1500


Buoyancy model for inversion (fixed)
z [m]
x [m]
Buoyancy [m^3/ kg]
x 10^-4
0
500
1000
1500
0
1000
2000
3000
4000
5000
5.5
5
4.5
4

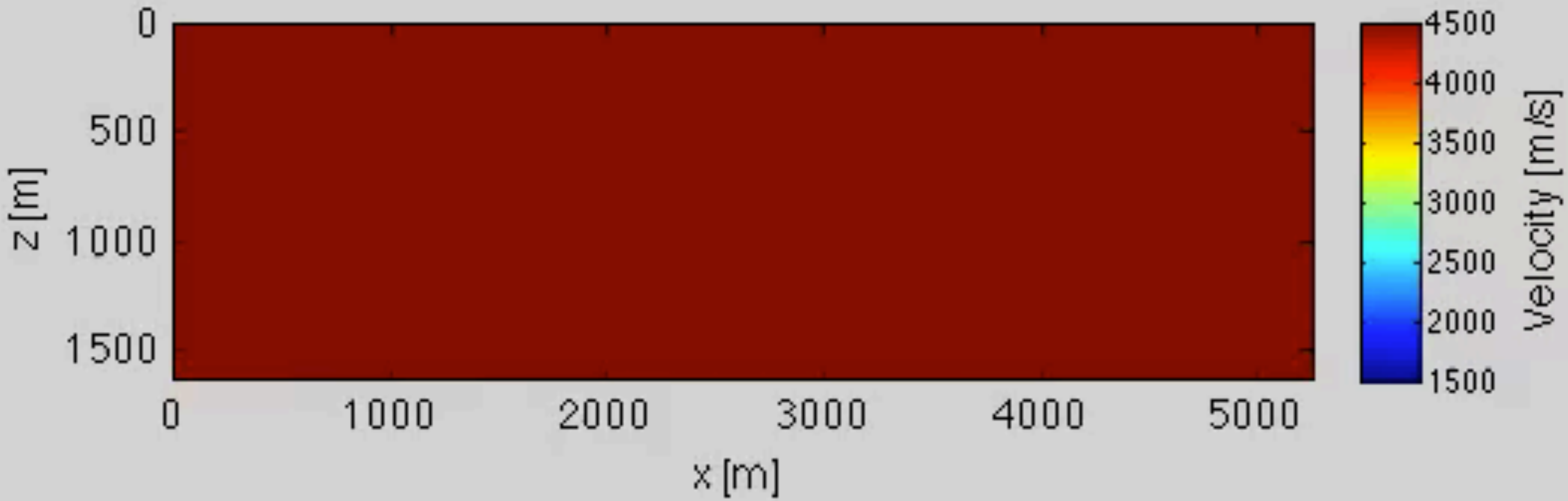
Final velocity estimate



True velocity

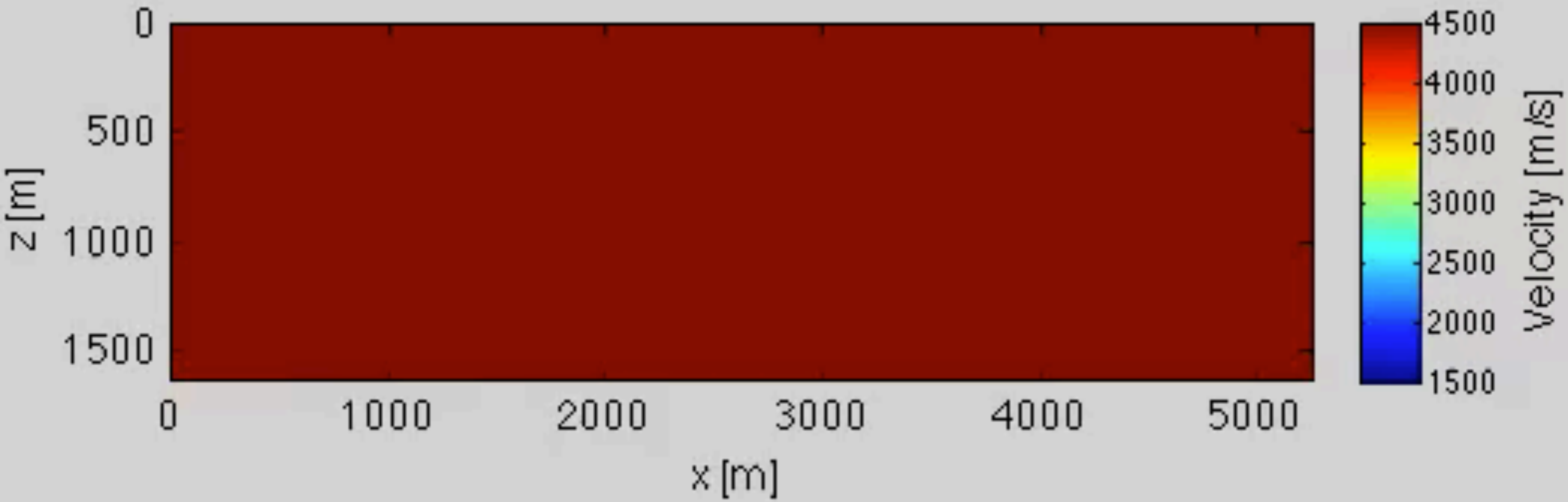


Cycle 1

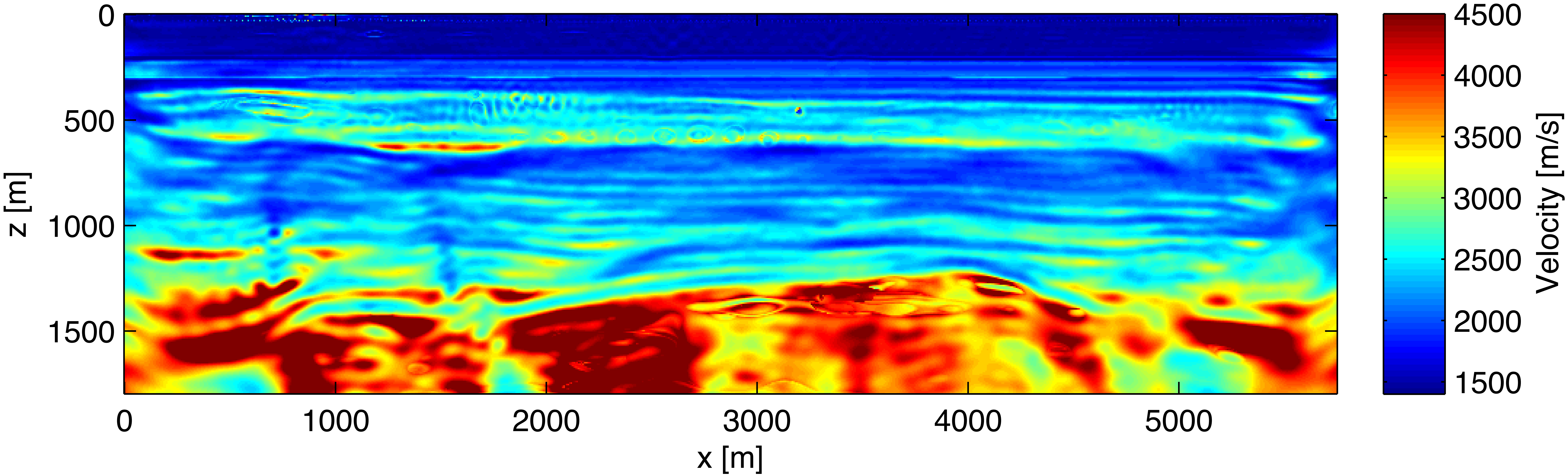


0
500
1000
1500
z [m]
0
1000
2000
3000
4000
5000
x [m]
4500
4000
3500
3000
2500
2000
1500
Velocity [m/s]

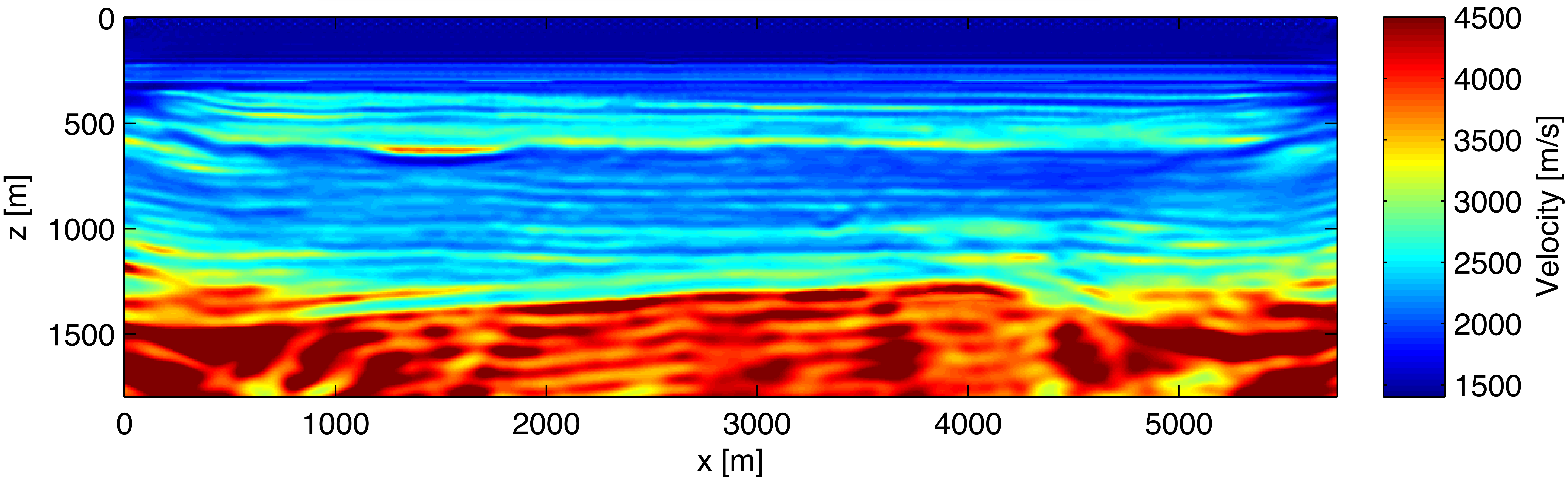
Cycle 2



0
500
1000
1500
z [m]
0
1000
2000
3000
4000
5000
x [m]
4500
4000
3500
3000
2500
2000
1500
Velocity [m/s]

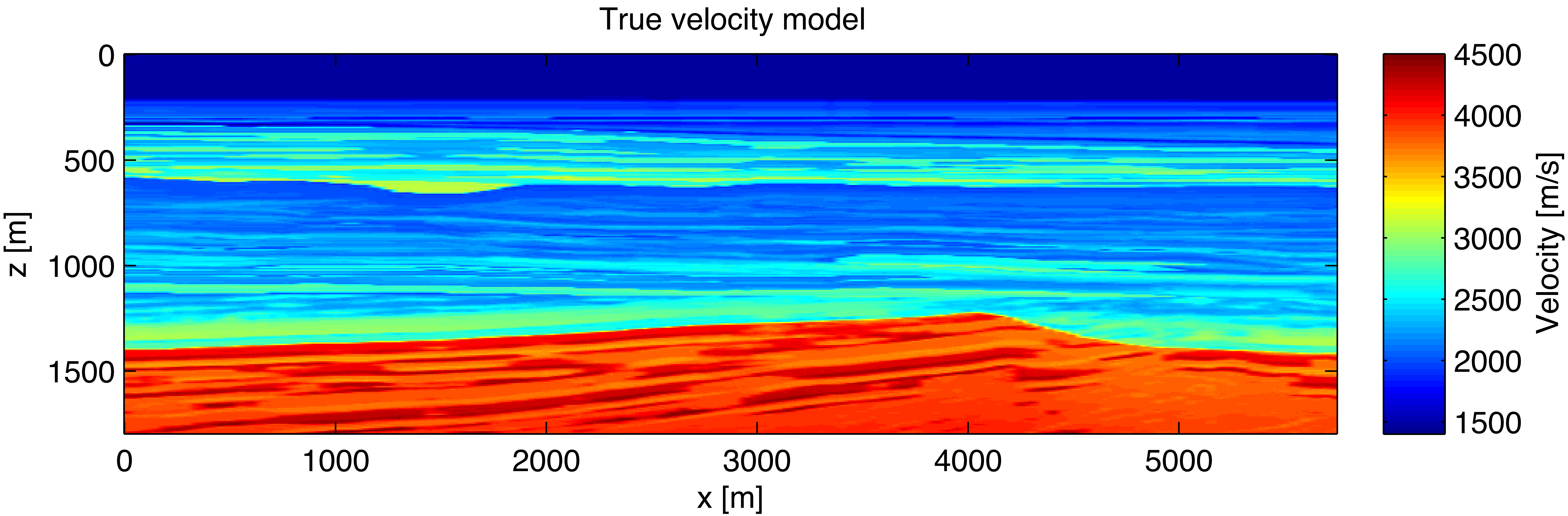


λ large -> does not fit data at the start



λ small -> fits data approximately at the start

True velocity



Solution of the sub-problem

Main challenge: solve $$\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \left\| \begin{pmatrix} \lambda H(\mathbf{m}) \\ P \end{pmatrix} \mathbf{u} - \begin{pmatrix} \lambda \mathbf{q} \\ \mathbf{d} \end{pmatrix} \right\|_2$$

- 2D: direct factorization **[L. M. Delves & I. Barrodale, 1979 ; T. A. Davis, 2011]**

- 3D: currently in development
- iteratively & matrix-free
- no QR or LU factorizations
- 1 PDE solve per source location + 1 PDE solve per receiver location

Including prior information

Solve constrained problem: $$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

$\mathcal{C}_1$ and $\mathcal{C}_2$ are closed convex, $\mathcal{C}_1 \bigcap \mathcal{C}_2 \neq \varnothing$

Including prior information

Solve constrained problem: $$\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$$

$\mathcal{C}_1$ and $\mathcal{C}_2$ are closed convex, $\mathcal{C}_1 \bigcap \mathcal{C}_2 \neq \varnothing$

examples of useful sets:

- bound constraints
- minimum smoothness

Including prior information

Solve constrained problem: $\min_{\mathbf{m}} f(\mathbf{m}) \quad \text{s.t.} \quad \mathbf{m} \in \mathcal{C}_1 \bigcap \mathcal{C}_2$

Possible algorithms:

- Projected gradient
- Projected Quasi-Newton **[M. Schmidt et. al., 2009]**
- Projected Newton using a (easy to invert) Hessian approximation

assumption: projections are cheap compared to PDE-solves

Conclusions

- Fitting the data at the start seems key.
- Minimizing the PDE-residual shows promising results in case of inaccurate initial models.
- Still limits on missing low-frequency data/inaccuracy of start model.
- Quadratic-penalty formulation leads to a low-memory reduced-space algorithm.
- Computational cost comparable with methods which eliminate the PDE-constraints.

Acknowledgements

PhD students and Postdocs at SLIM

SINBAD
NSERC CRSNG

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada via the Collaborative Research and Development Grant DNOISEII (375142--08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, DownUnder GeoSolutions, Hess, ION, Petrobras, PGS, Schlumberger, Statoil, Sub Salt Solutions, Total SA, and Woodside.

References

1. A. Bjorck and T. Elfving, Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations, BIT, 19 (1979), pp. 145–163.
2. D. Gordon and R. Gordon, CARP-CG: A robust and efficient parallel solver for linear systems, applied to strongly convection dominated pdes, Parallel Computing, 36 (2010), pp. 495– 515.
3. Tristan van Leeuwen and Felix J. Herrmann, frequency-domain seismic inversion with controlled sloppiness, SIAM Journal on Scientific Computing, 36 (2014), pp. S192–S217.
4. M.J. Grote, J. Huber, and O. Schenk, Interior point methods for the inverse medium problem on massively parallel architectures, Procedia Computer Science, 4 (2011), pp. 1466 – 1474. Proceedings of the International Conference on Computational Science, {ICCS} 2011.
5. Eldad Haber, Uri M Ascher, and Doug Oldenburg, On optimization techniques for solving nonlinear inverse problems, Inverse Problems, 16 (2000), pp. 1263–1280.
6. E Haber and U M Ascher, Preconditioned all-at-once methods for large, sparse parameter estimation problems, Inverse Problems, 17 (2001), p. 1847.
7. I Epanomeritakis, V Akcelik, O Ghattas, and J Bielak. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. Inverse Problems, 24(3):034015, June 2008.
8. George Biros and Omar Ghattas, Parallel lagrange–newton–krylov– schur methods for pde-constrained optimization. part i: The krylov–schur solver, SIAM Journal on Scientific Computing, 27 (2005), pp. 687–713.
9. R.E. Kleinman and P.M.van den Berg, A modified gradient method for two- dimensional problems in tomography, Journal of Computational and Applied Mathematics, 42 (1992), pp. 17 – 35.

References (2)

10. B Peters, F.J. Herrmann, T. van Leeuwen. Wave-equation Based Inversion with the Penalty Method-Adjoint-state Versus Wavefield-reconstruction Inversion. 76th EAGE Conference, 2014.
11. Calandra, H., Gratton, S., Pinel, X. and Vasseur, X. [2013] An improved two-grid preconditioner for the solution of three-dimensional Helmholtz problems in heterogeneous media. Numerical Linear Algebra with Applications.
12. Erlangga, Y.A. [2008] Advances in iterative methods and preconditioners for the Helmholtz equation. Archives of Computational Methods in Engineering, 15, 37–66.
13. Delves, L. M., and I. Barrodale. "A fast direct method for the least squares solution of slightly overdetermined sets of linear equations." IMA Journal of Applied Mathematics 24.2 (1979): 149-156.
14. Rafael Lago, Felix J. Herrmann. Towards a robust geometric multigrid scheme for {Helmholtz} equation, Tech Report, UBC, 2015.
15. Timothy A. Davis, Algorithm 915, suitesparseqr: Multifrontal multithreaded rank-revealing sparse qr factorization, ACM Trans. Math. Softw., 38 (2011), pp. 8:1–8:22.
16. Schmidt, M., van den Berg, E., Friedlander, M. and Murphy, K. [2009] Optimizing costly functions with simple constraints: A limited-memory projected quasi-newton algorithm. JMLR, vol. 5, 456–463.
17. Ake Bjork, Numerical methods for least squares problems. siam, 1996.
18. Walter Gander, Least squares with a quadratic constraint, Numerische Mathematik, 36 (1980), pp. 291–307.
19. Albert Tarantola, Inversion of seismic reflection data in the acoustic approximation, Geophysics, 49 (1984), pp. 1259–1266.
20. Mark Schmidt and Dongmin Kim and Suvrit Sra. Projected Newton-type Methods in Machine Learning. Book chapter in Optimization for Machine Learning, 2012. MIT Press.