PDE-free Gauss-Newton Hessian for Wavefield Reconstruction Inversion

Zhilong Fang* and Felix J. Herrmann*
*Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia

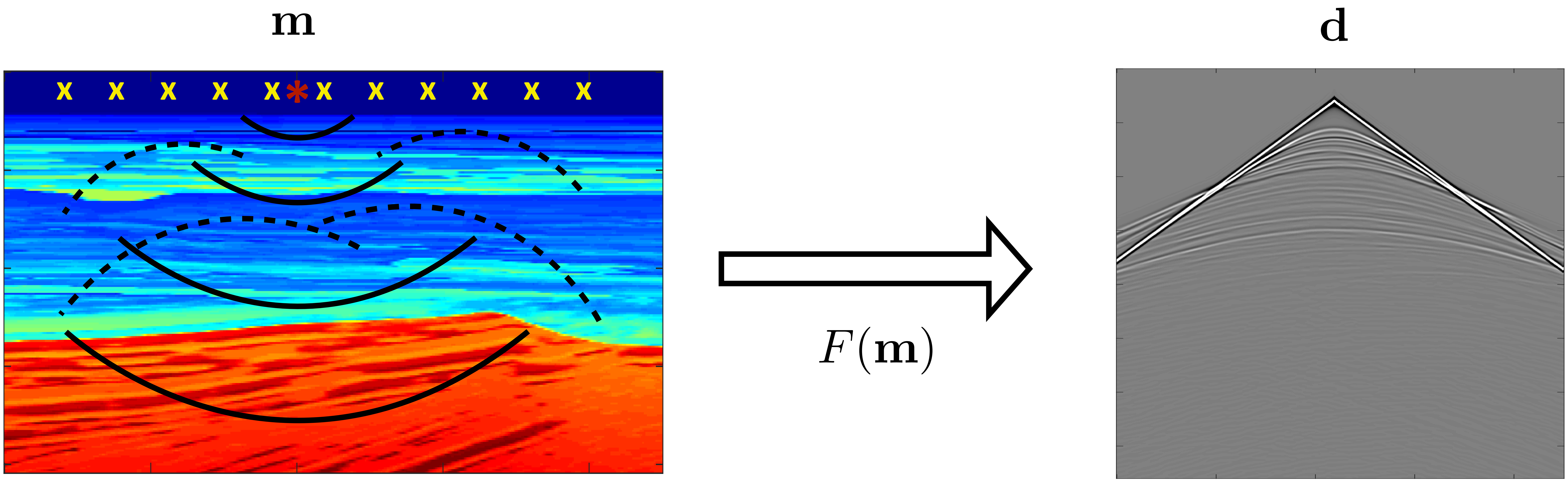
2017/10/04

SLIM
University of British Columbia

SLIM

Motivation

Forward problem



Motivation

Forward map $F(\mathbf{m})$:

$$\mathbf{A}(\mathbf{m})\mathbf{u} = \mathbf{q},\ \mathbf{d} = \mathbf{P}\mathbf{u} \longrightarrow \text{wavefields}$$

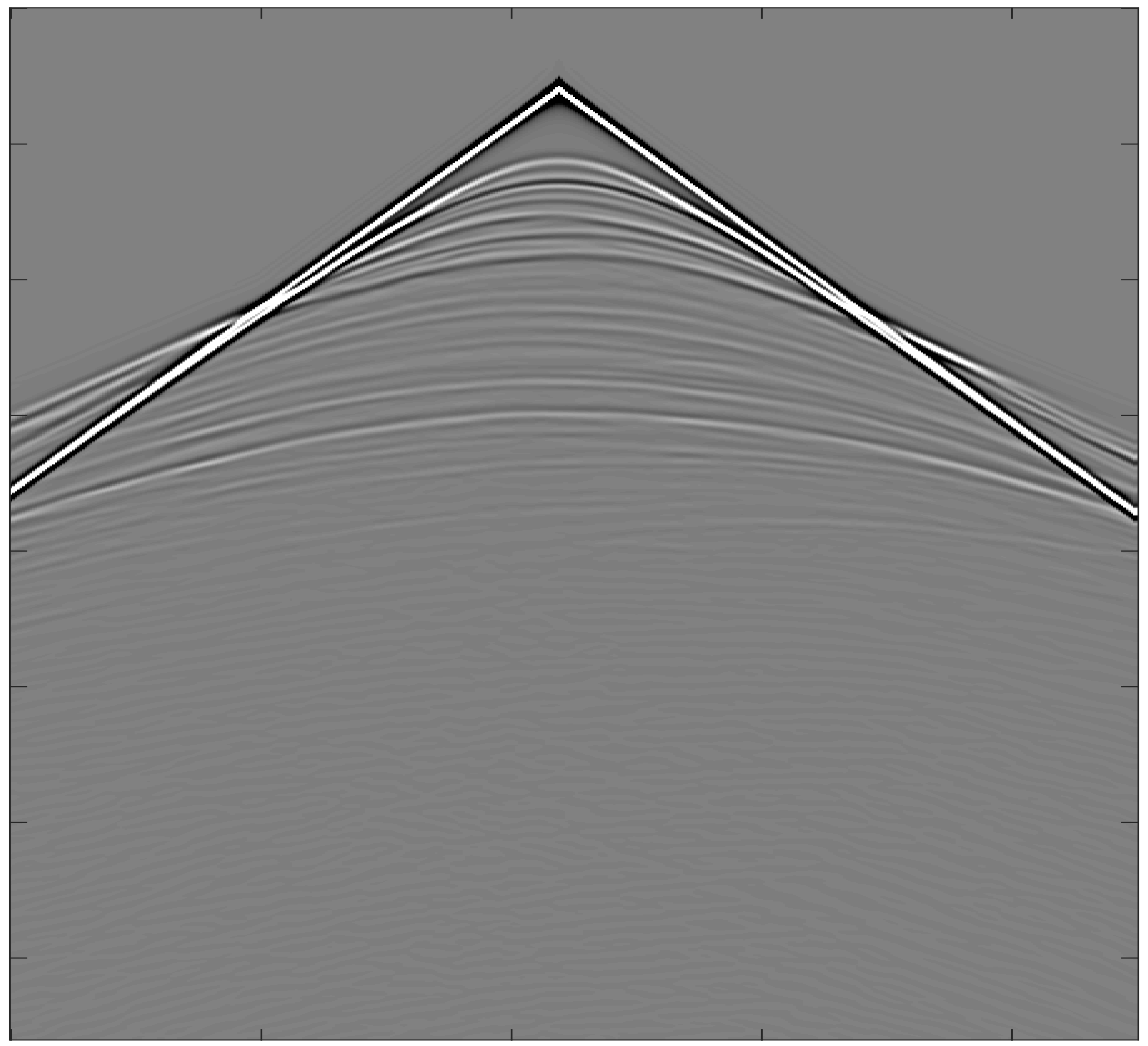
$\mathbf{A}(\mathbf{m})$ → $\Delta + \omega^2\mathbf{m}$ (ω: frequency, $\mathbf{m}$: squared-slowness)

$\mathbf{q}$ → source

$\mathbf{P}$ → projection operator

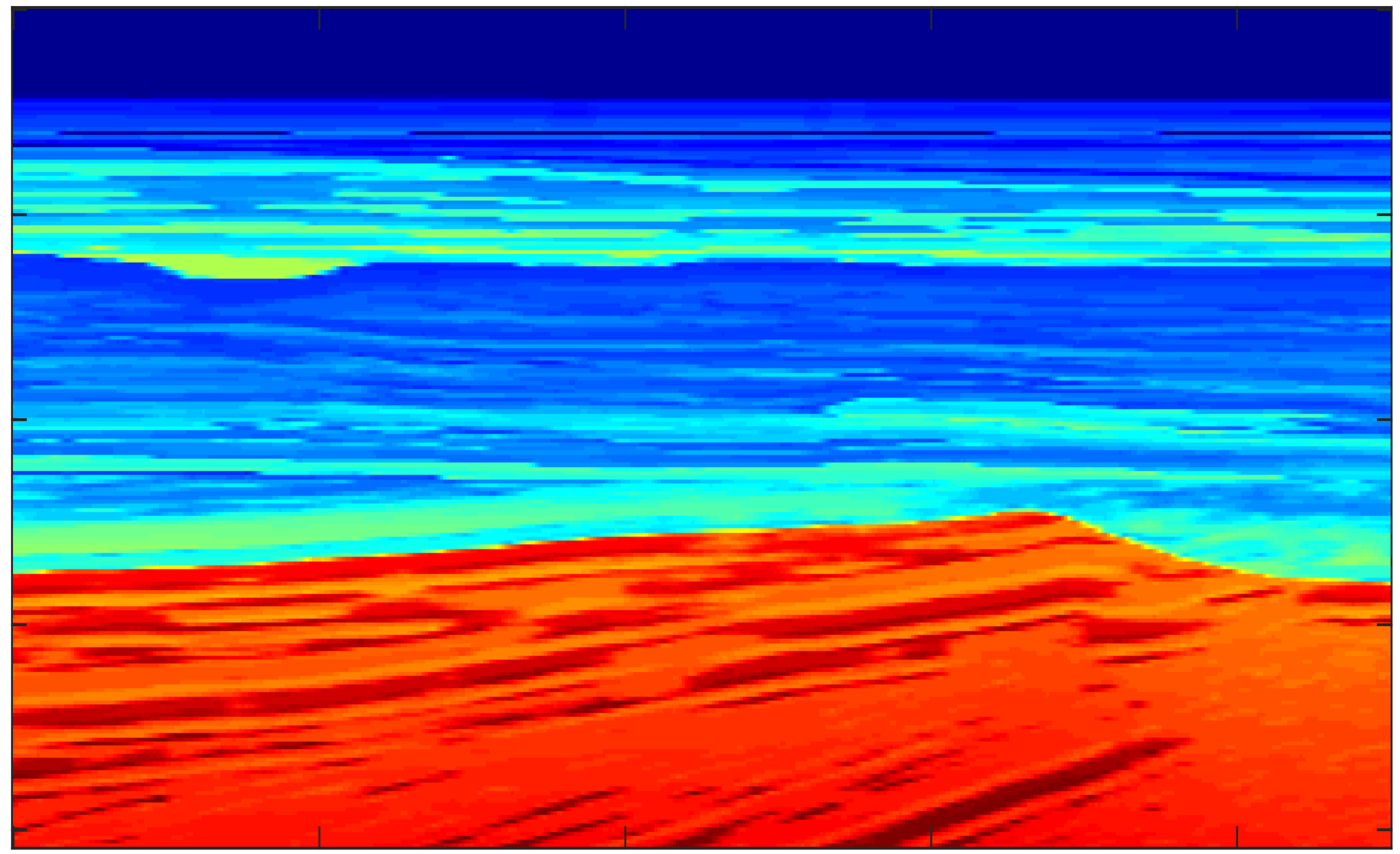
Motivation

Inverse problem



$\mathbf{d}$

$$F^{-1}(\mathbf{d})$$



$\mathbf{m}$

SLIM

Motivation

Objective:

$$\min_{\mathbf{m}} f(\mathbf{m}) = \frac{1}{2}\|F(\mathbf{m}) - \mathbf{d}\|_2^2$$

Motivation

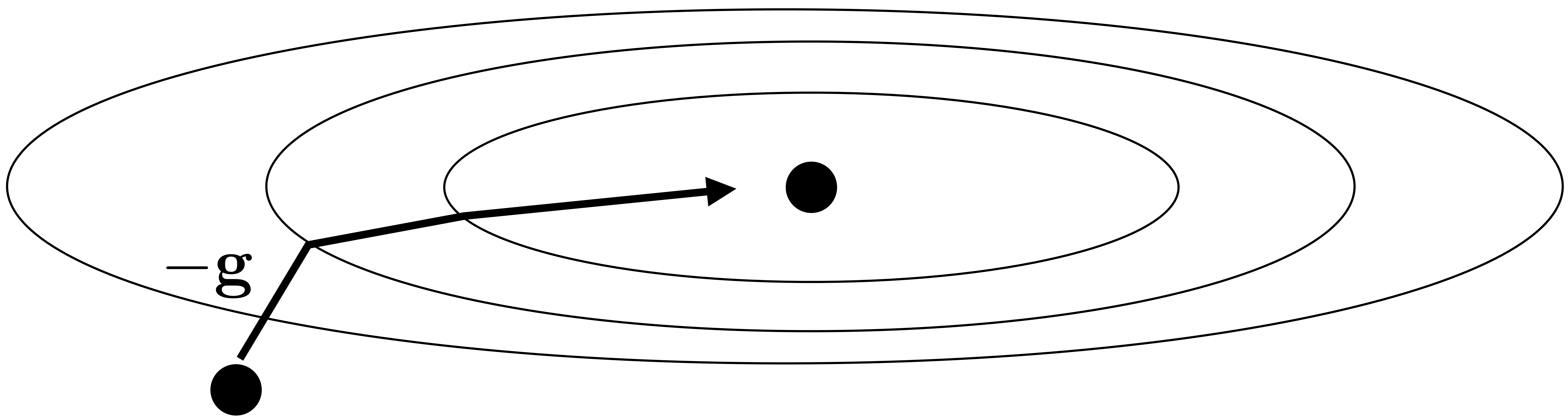
Objective:

$$\min_{\mathbf{m}} f(\mathbf{m}) = \frac{1}{2}\|F(\mathbf{m}) - \mathbf{d}\|_2^2$$

First order method w/ gradient $\mathbf{g}(\mathbf{m}_k)$:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha\mathbf{g}(\mathbf{m}_k)$$

Motivation



SLIM

Motivation

Objective:

$$\min_{\mathbf{m}} f(\mathbf{m}) = \frac{1}{2}\|F(\mathbf{m}) - \mathbf{d}\|_2^2$$

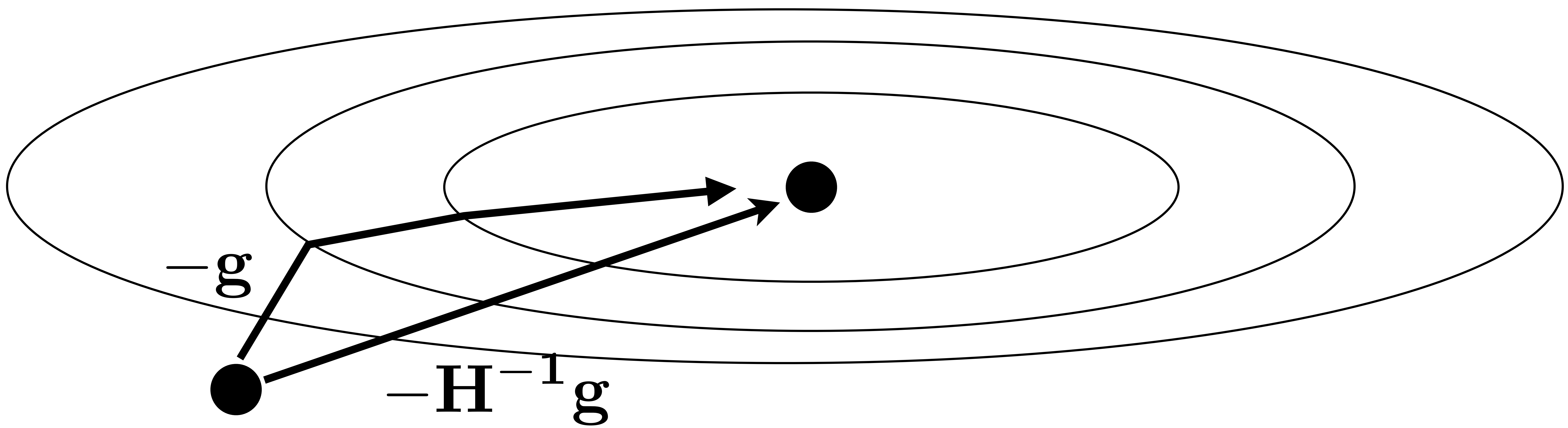
First order method w/ gradient $\mathbf{g}(\mathbf{m}_k)$:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \mathbf{g}(\mathbf{m}_k)$$

Second order method w/ Hessian $\mathbf{H}(\mathbf{m}_k)$:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha \mathbf{H}(\mathbf{m}_k)^{-1} \mathbf{g}(\mathbf{m}_k)$$

Motivation



Motivation

Challenges associated w/ Hessian:

- Storage cost – $n_g \times n_g$
- Computational cost – 4 PDE solves per each shot and frequency to compute one matrix-vector product

Motivation

Challenges associated w/ Hessian:

- Storage cost – $n_g \times n_g$
- Computational cost – 4 PDE solves per each shot and frequency to compute one matrix-vector product

Goal:

- Fast method to compute the matrix-vector product with the Hessian

[Tarantola, 1984]

[J. Virieux and S. Operto, 2009]

Full-waveform inversion (FWI)

PDE-constrained optimization problem:

$$\min_{\mathbf{u},\mathbf{m}} \frac{1}{2N} \sum_{i=1}^{n_s} \sum_{l=1}^{n_f} \|\mathbf{P}\mathbf{u}_{i,l} - \mathbf{d}_{i,l}\|_2^2$$

$$\text{subject to } \mathbf{A}_{i,l}(\mathbf{m})\mathbf{u}_{i,l} = \mathbf{q}_{i,l}$$

where $N = n_s \times n_f$

[Tarantola, 1984]

[J. Virieux and S. Operto, 2009]

SLIM

FWI

Reduced/adjoint-state method:

$$\min_{\mathbf{m}} \frac{1}{2N} \sum_{i=1}^{n_s} \sum_{l=1}^{n_f} \|\mathbf{P}\mathbf{A}_{i,l}(\mathbf{m})^{-1}\mathbf{q}_{i,l} - \mathbf{d}_{i,l}\|_2^2$$

with the gradient given by

$$\mathbf{g} = \frac{1}{N} \sum_{i=1}^{n_s} \sum_{l=1}^{n_f} \mathbf{u}_{i,l}^{\top} \frac{\partial \mathbf{A}_{i,l}^{\top}}{\partial \mathbf{m}} \mathbf{v}_{i,l}$$

$$\mathbf{u}_{i,l} = \mathbf{A}_{i,l}(\mathbf{m})^{-1}\mathbf{q}_{i,l}$$

$$\mathbf{v}_{i,l} = -\mathbf{A}_{i,l}^{-\top}(\mathbf{m})\mathbf{P}^{\top}\mathbf{r}_{i,l}$$

$$\mathbf{r}_{i,l} = \mathbf{P}\mathbf{A}_{i,l}(\mathbf{m})^{-1}\mathbf{q}_{i,l} - \mathbf{d}_{i,l}$$

2 PDE solves are required !

[van Leeuwen, T and Herrmann, F J , 2013]

[Peters, B, Herrmann, F J and van Leeuwen, T, 2014]

[Golub, G H and Pereyra, V , 2003]

Wavefield-reconstruction inversion (WRI)

Penalty method:

$$\min_{\mathbf{u},\mathbf{m}} \frac{1}{2N} \sum_{i=1}^{n_s} \sum_{l=1}^{n_f} \|\mathbf{P}\mathbf{u}_{i,l} - \mathbf{d}_{i,l}\|_2^2 + \lambda^2 \|\mathbf{A}_{i,l}(\mathbf{m})\mathbf{u}_{i,l} - \mathbf{q}_{i,l}\|_2^2$$

Eliminating $\mathbf{u}$ w/ variable projection:

$$\overline{\mathbf{u}} = \arg\min_{\mathbf{u}} \frac{1}{2N} \sum_{i=1}^{n_s} \sum_{l=1}^{n_f} \|\mathbf{P}\mathbf{u}_{i,l} - \mathbf{d}_{i,l}\|_2^2 + \lambda^2 \|\mathbf{A}_{i,l}(\mathbf{m})\mathbf{u}_{i,l} - \mathbf{q}_{i,l}\|_2^2$$

[van Leeuwen, T and Herrmann, F J , 2013]

[Golub, G and Pereyra, V, 2003]

SLIM

WRI

Corresponds to solving the following augmented system:

$$\begin{pmatrix} \lambda \mathbf{A}_{i,l} \\ \mathbf{P} \end{pmatrix} \overline{\mathbf{u}}_{i,l} = \begin{pmatrix} \lambda \mathbf{q}_{i,l} \\ \mathbf{d}_{i,l} \end{pmatrix}$$

with the gradient

1 augmented system solves is required !

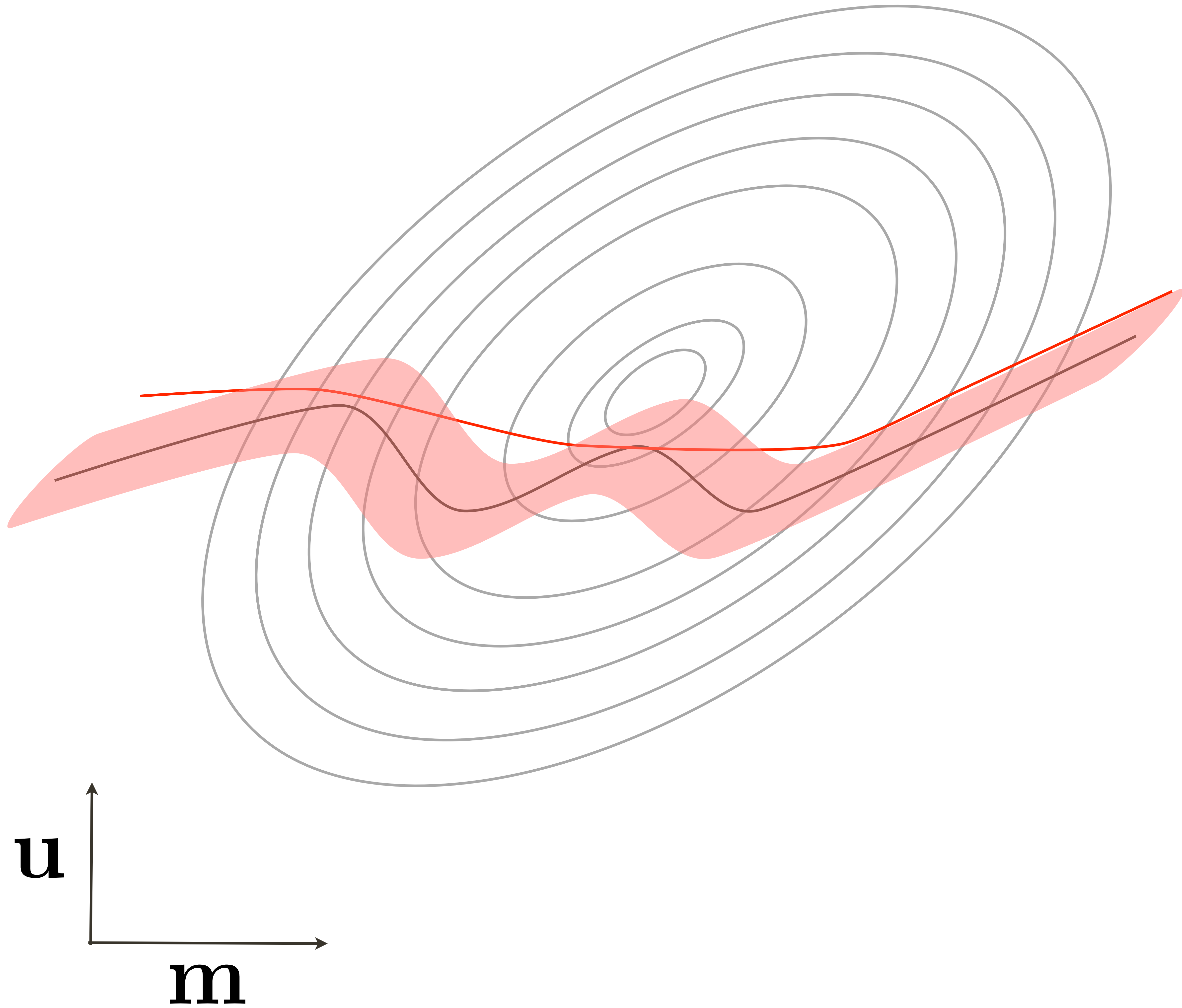
$$\mathbf{g} = \frac{1}{N} \sum_{i=1}^{n_s} \sum_{l=1}^{n_f} \overline{\mathbf{u}}_{i,l}^{\top} \frac{\partial \mathbf{A}_{i,l}^{\top}}{\partial \mathbf{m}} \overline{\mathbf{v}}_{i,l}$$

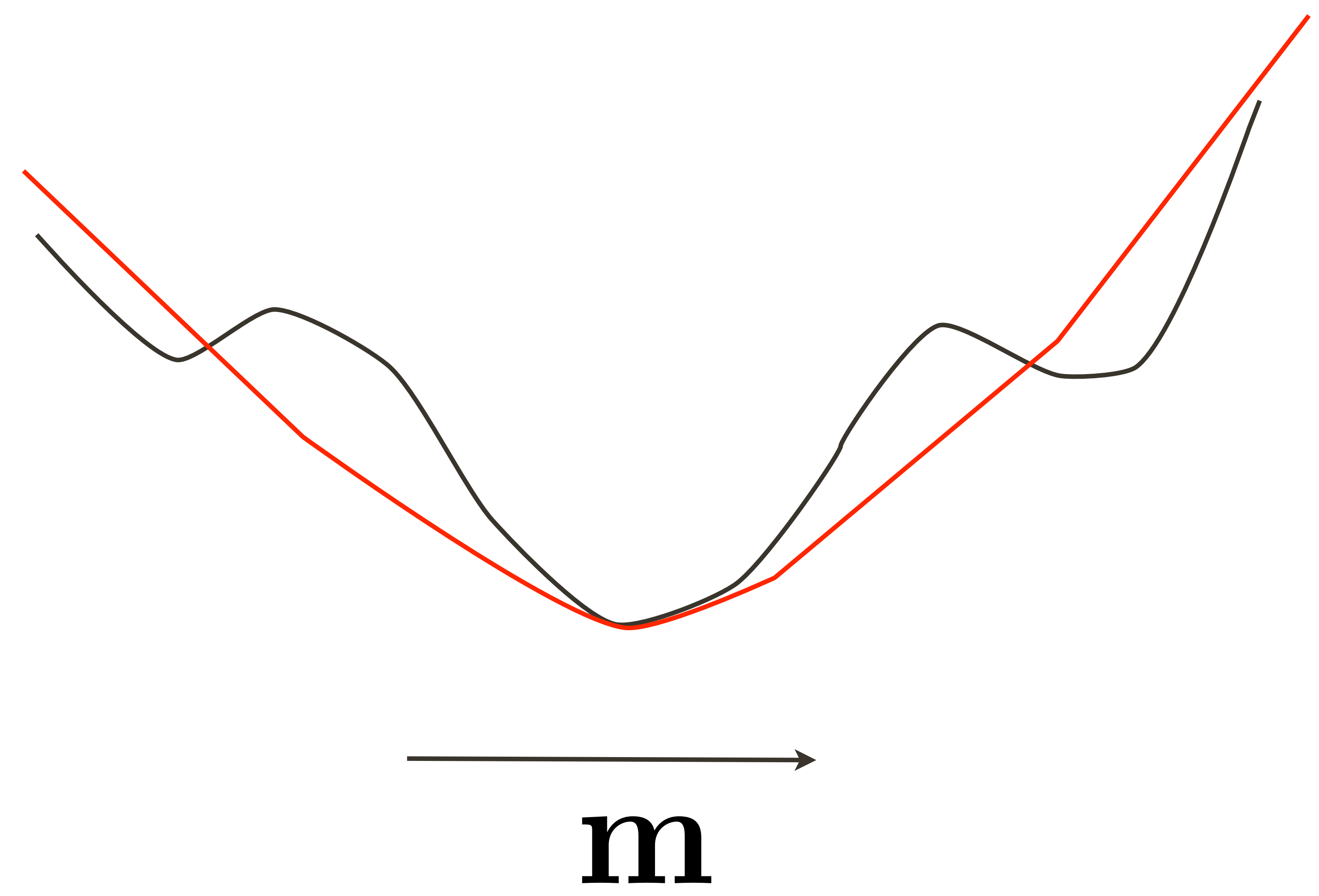
$$\overline{\mathbf{v}}_{i,l} = \mathbf{A}_{i,l}(\mathbf{m}) \overline{\mathbf{u}}_{i,l} - \mathbf{q}_{i,l}$$

WRI vs. FWI

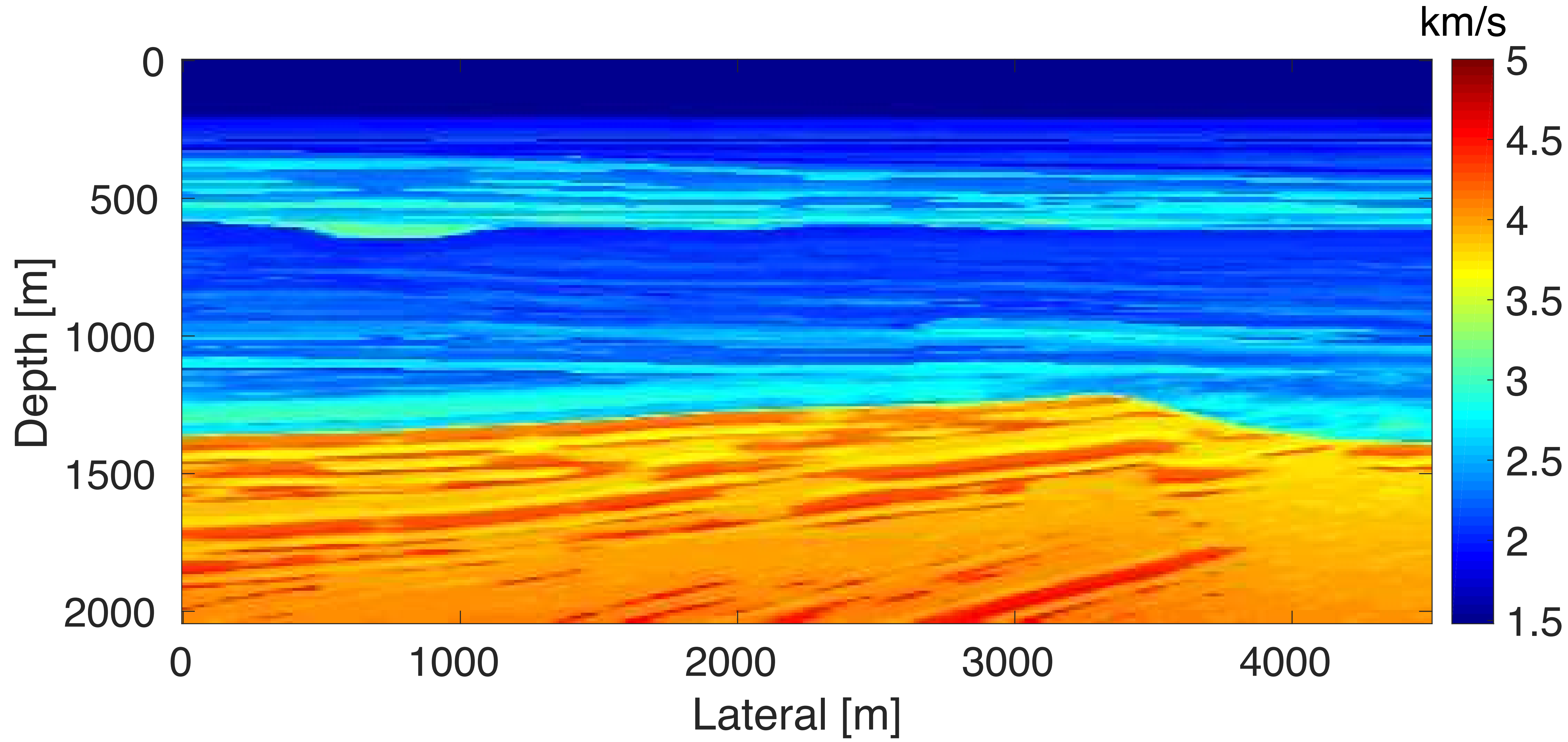
[van Leeuwen, T and Herrmann, F J , 2013]

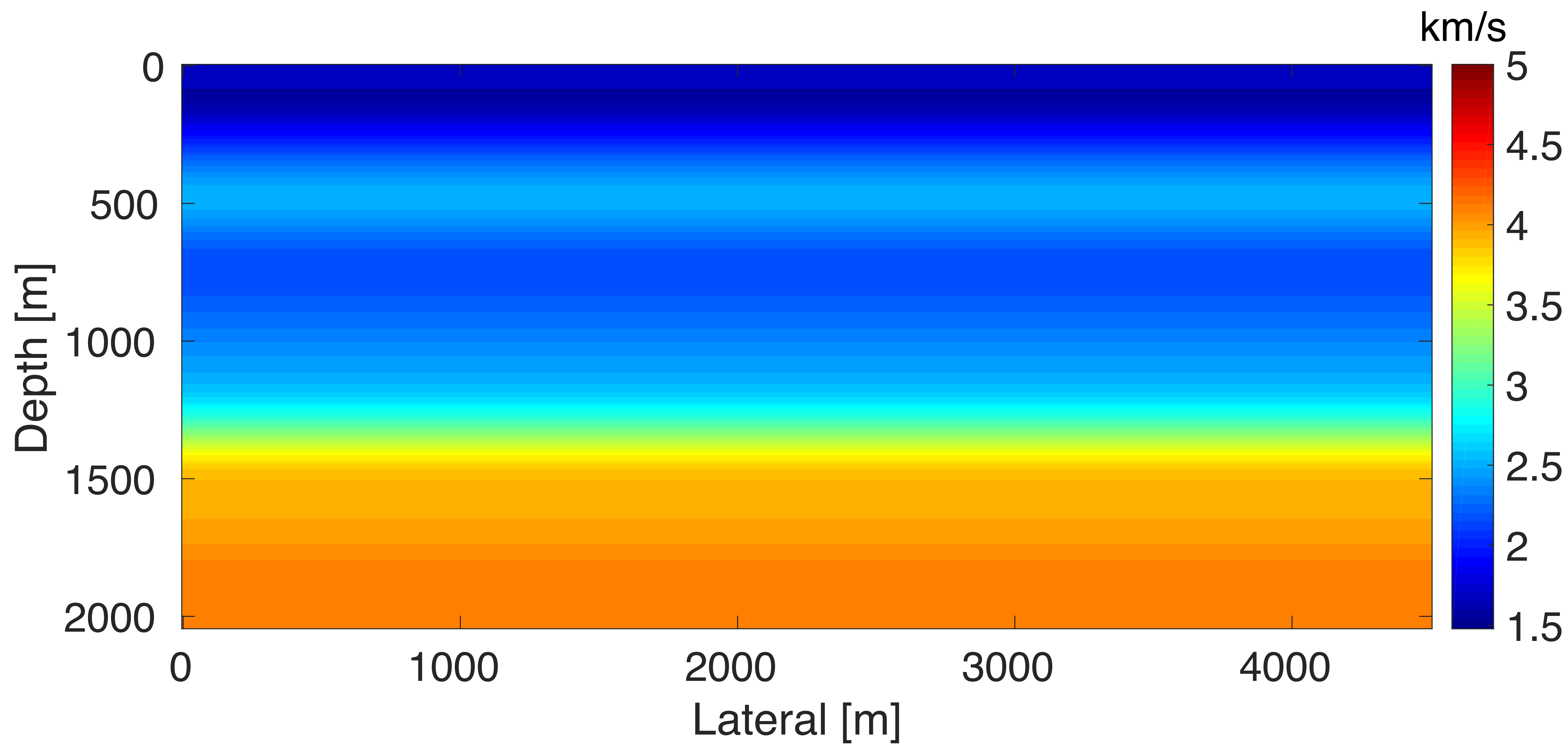
[Peters, B, Herrmann, F J and van Leeuwen, T, 2014]



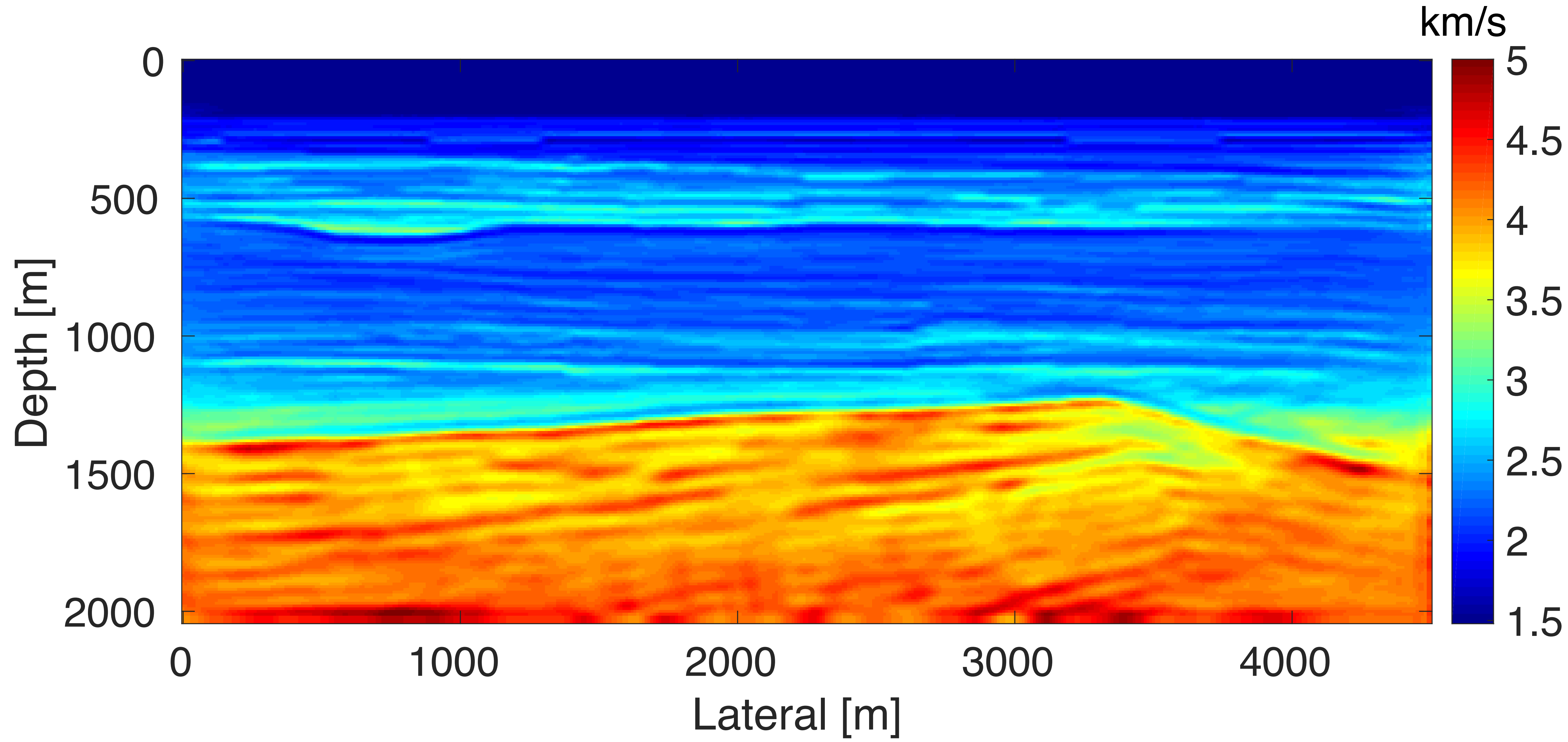


True & initial model

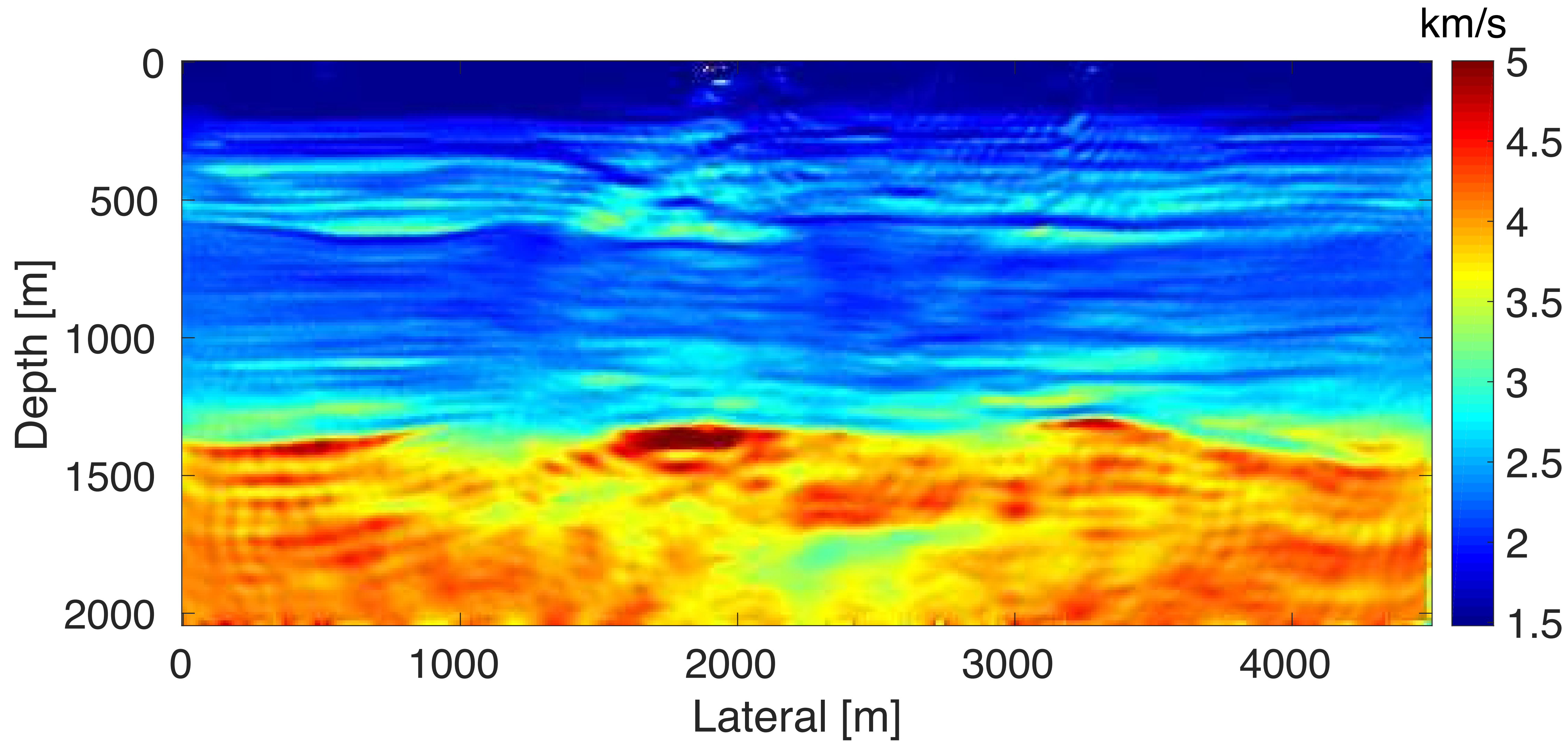




FWI vs WRI



WRI result



FWI result

Gauss-Newton Hessian of WRI

GN Hessian:

$$\mathbf{H}_{\text{GN}} = \frac{1}{N}\sum_{i=1}^{n_s}\sum_{l=1}^{n_f}\mathbf{G}_{i,l}^{\top}\mathbf{A}_{i,l}^{-\top}\mathbf{P}^{\top}\underbrace{(\mathbf{I}+\frac{1}{\lambda^2}\mathbf{P}\mathbf{A}_{i,l}^{-1}\mathbf{A}_{i,l}^{-\top}\mathbf{P}^{\top})^{-1}}_{n_r^2}\underbrace{\mathbf{P}\mathbf{A}_{i,l}^{-1}}_{n_r n_g}\underbrace{\mathbf{G}_{i,l}}_{\text{diagonal}}$$

where

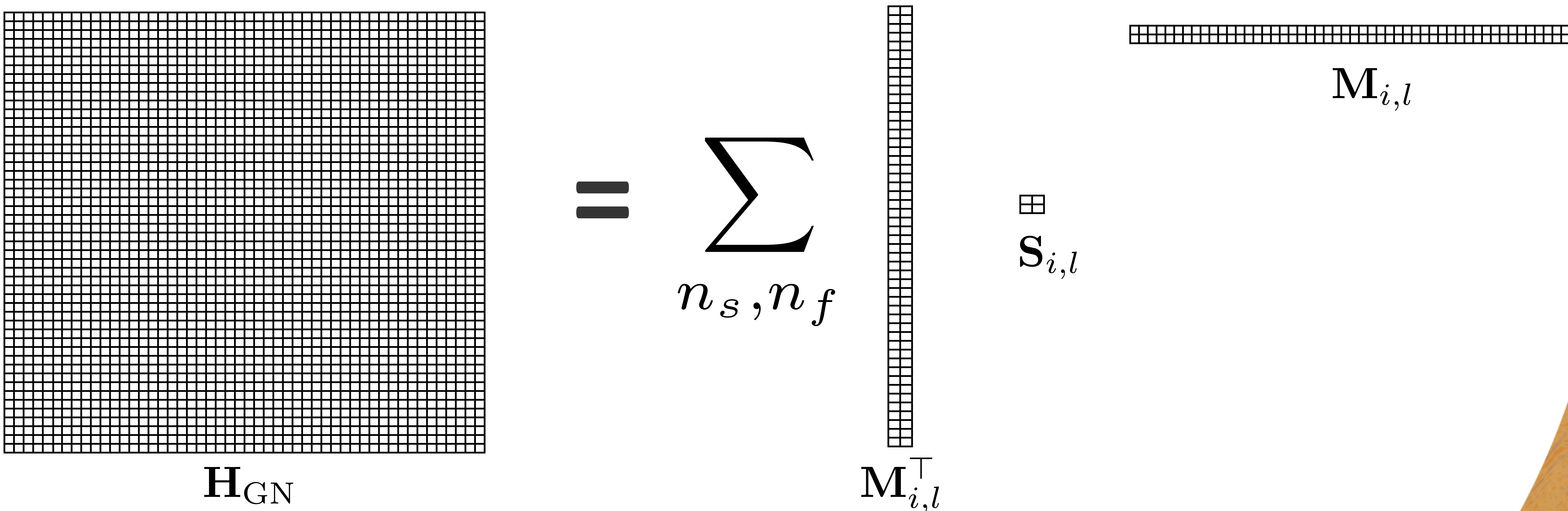
$$\mathbf{G}_{i,l} = \frac{\partial \mathbf{A}_{i,l}}{\partial \mathbf{m}}\mathbf{u}_{i,j}$$

Gauss-Newton Hessian of WRI

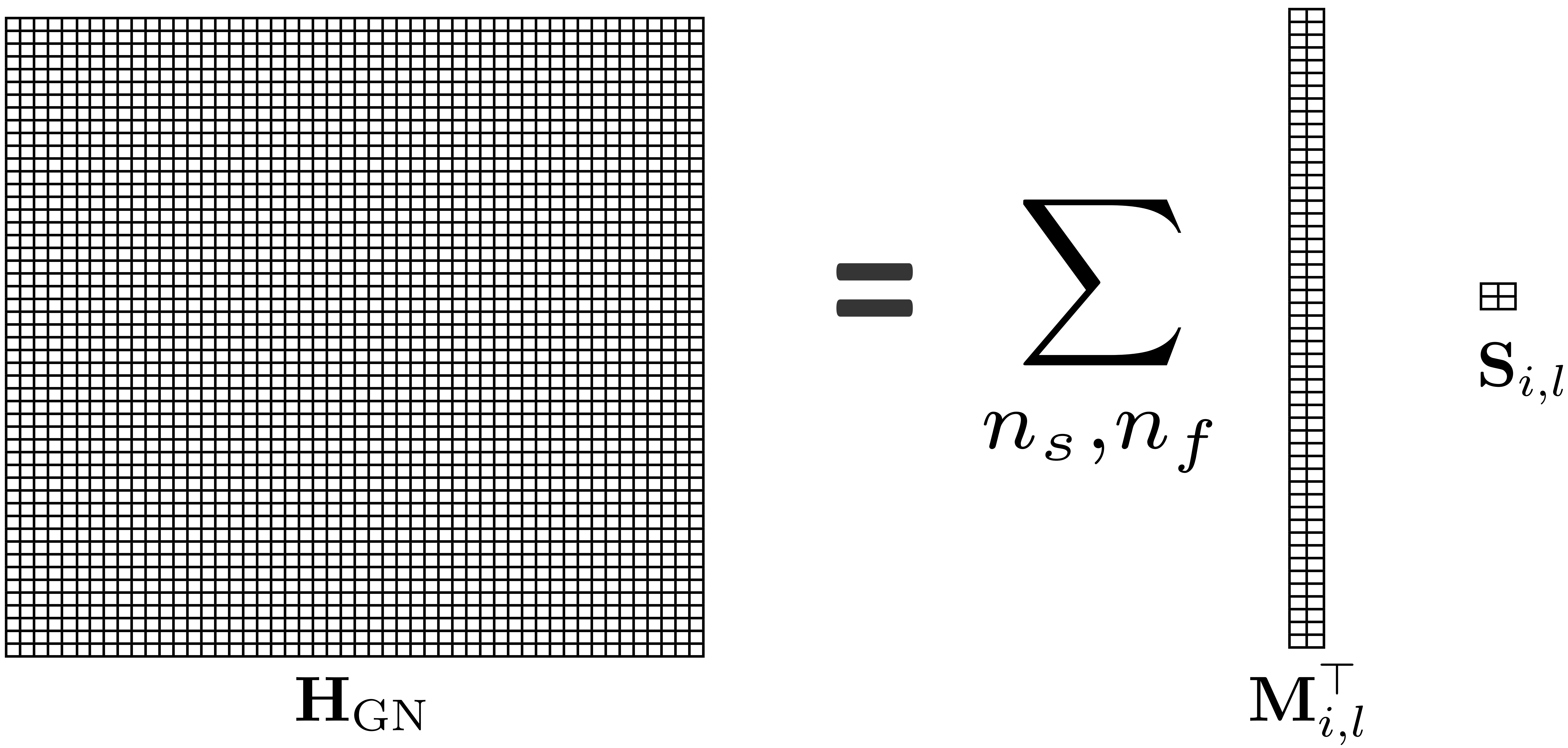
GN Hessian:

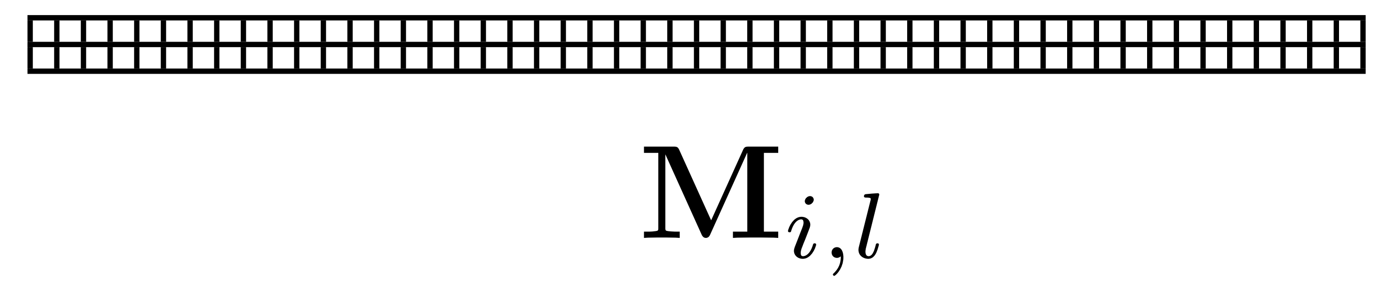
$$\mathbf{H}_{\mathrm{GN}} = \frac{1}{N}\sum_{i=1}^{n_s}\sum_{l=1}^{n_f}\mathbf{G}_{i,l}^{\top}\mathbf{A}_{i,l}^{-\top}\mathbf{P}^{\top}\underbrace{(\mathbf{I}+\frac{1}{\lambda^2}\mathbf{P}\mathbf{A}_{i,l}^{-1}\mathbf{A}_{i,l}^{-\top}\mathbf{P}^{\top})^{-1}}_{\mathbf{S}_{i,l}}\underbrace{\mathbf{P}\mathbf{A}_{i,l}^{-1}\mathbf{G}_{i,l}}_{\mathbf{M}_{i,l}}$$

GN Hessian of WRI



GN Hessian of WRI





Computational cost:

$n_f * (n_s + n_r)$

Storage cost:

$n_f * n_g * (n_s + n_r)$

GN Hessian of WRI

Diagonal part of the Hessian:

$$\mathbf{H}_{\mathrm{GN}} = \frac{1}{N}\sum_{i=1}^{n_s}\sum_{l=1}^{n_f}\mathbf{M}_{i,l}^{\top}\mathbf{S}_{i,l}\mathbf{M}_{i,l}$$

$$= \frac{1}{N}\sum_{i=1}^{n_s}\sum_{l=1}^{n_f}\mathbf{B}_{i,l}^{\top}\mathbf{B}_{i,l}$$

where

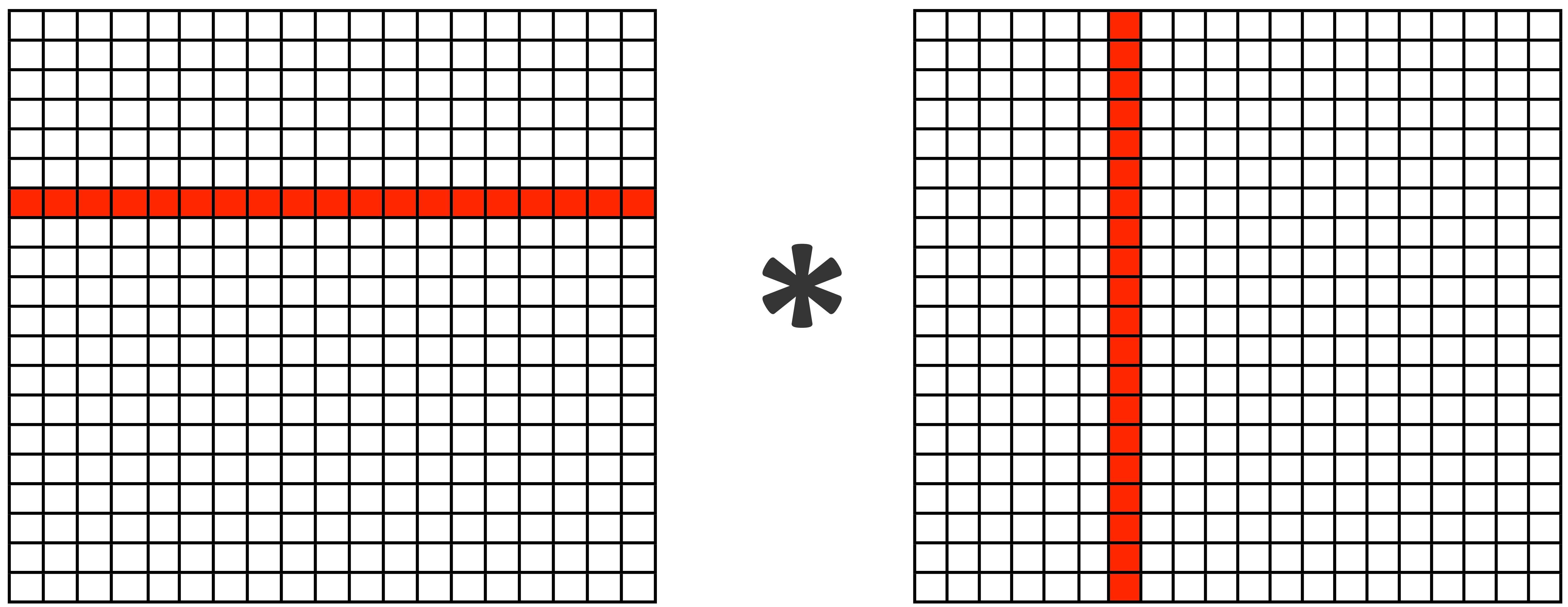
$$\mathbf{B}_{i,l} = \mathbf{S}_{i,l}^{1/2}\mathbf{M}_{i,l}$$

GN Hessian of WRI

Diagonal part of the Hessian:

$$h_{j,j} = \mathbf{B}_{i,l}(:,j)^\top * \mathbf{B}_{i,l}(:,j)$$

$h_{j,j} =$



$\mathbf{B}_{i,l}^\top$ $\mathbf{B}_{i,l}$

Dimensional reduction

Full objective:

$$f(\mathbf{u},\mathbf{m}) = \frac{1}{2N}\sum_{i=1}^{n_s}\sum_{l=1}^{n_f}\|\mathbf{P}\mathbf{u}_{i,l}-\mathbf{d}_{i,l}\|_2^2+\lambda^2\|\mathbf{A}_{i,l}(\mathbf{m})\mathbf{u}_{i,l}-\mathbf{q}_{i,l}\|_2^2$$

Dimensional reduction

Simultaneous shots:

$$\mathbf{Q}_l = [\mathbf{q}_{1,l}, \mathbf{q}_{2,l}, \ldots, \mathbf{q}_{n_s,l}]$$
$$\mathbf{D}_l = [\mathbf{d}_{1,l}, \mathbf{d}_{2,l}, \ldots, \mathbf{d}_{n_s,l}]$$
$$\overline{\mathbf{Q}}_l = \mathbf{Q}_l\mathbf{W} = [\overline{\mathbf{q}}_{1,l}, \overline{\mathbf{q}}_{2,l}, \ldots, \overline{\mathbf{q}}_{\tilde{n}_s,l}]$$
$$\overline{\mathbf{D}}_l = \mathbf{D}_l\mathbf{W} = [\overline{\mathbf{d}}_{1,l}, \overline{\mathbf{d}}_{2,l}, \ldots, \overline{\mathbf{d}}_{\tilde{n}_s,l}]$$
$$\mathbf{W} \in \mathcal{R}^{n_s \times \tilde{n}_s}$$

Simultaneous receivers:

$$\overline{\mathbf{P}} = \tilde{\mathbf{W}}\mathbf{P},\ \tilde{\mathbf{W}} \in \mathcal{R}^{\tilde{n}_r \times n_r}$$

Dimensional reduction

Full objective:

$$f(\mathbf{u},\mathbf{m}) = \frac{1}{2N}\sum_{i=1}^{n_s}\sum_{l=1}^{n_f}\|\mathbf{P}\mathbf{u}_{i,l}-\mathbf{d}_{i,l}\|_2^2+\lambda^2\|\mathbf{A}_{i,l}(\mathbf{m})\mathbf{u}_{i,l}-\mathbf{q}_{i,l}\|_2^2$$

Stochastic objective:

$$\overline{f}(\mathbf{u},\mathbf{m}) = \frac{1}{2N}\sum_{i=1}^{\tilde{n}_s}\sum_{l=1}^{n_f}\|\tilde{\mathbf{W}}\mathbf{P}\mathbf{u}_{i,l}-\tilde{\mathbf{W}}\overline{\mathbf{d}}_{i,l}\|_2^2+\lambda^2\|\mathbf{A}_{i,l}(\mathbf{m})\mathbf{u}_{i,l}-\overline{\mathbf{q}}_{i,l}\|_2^2$$

Dimensional reduction

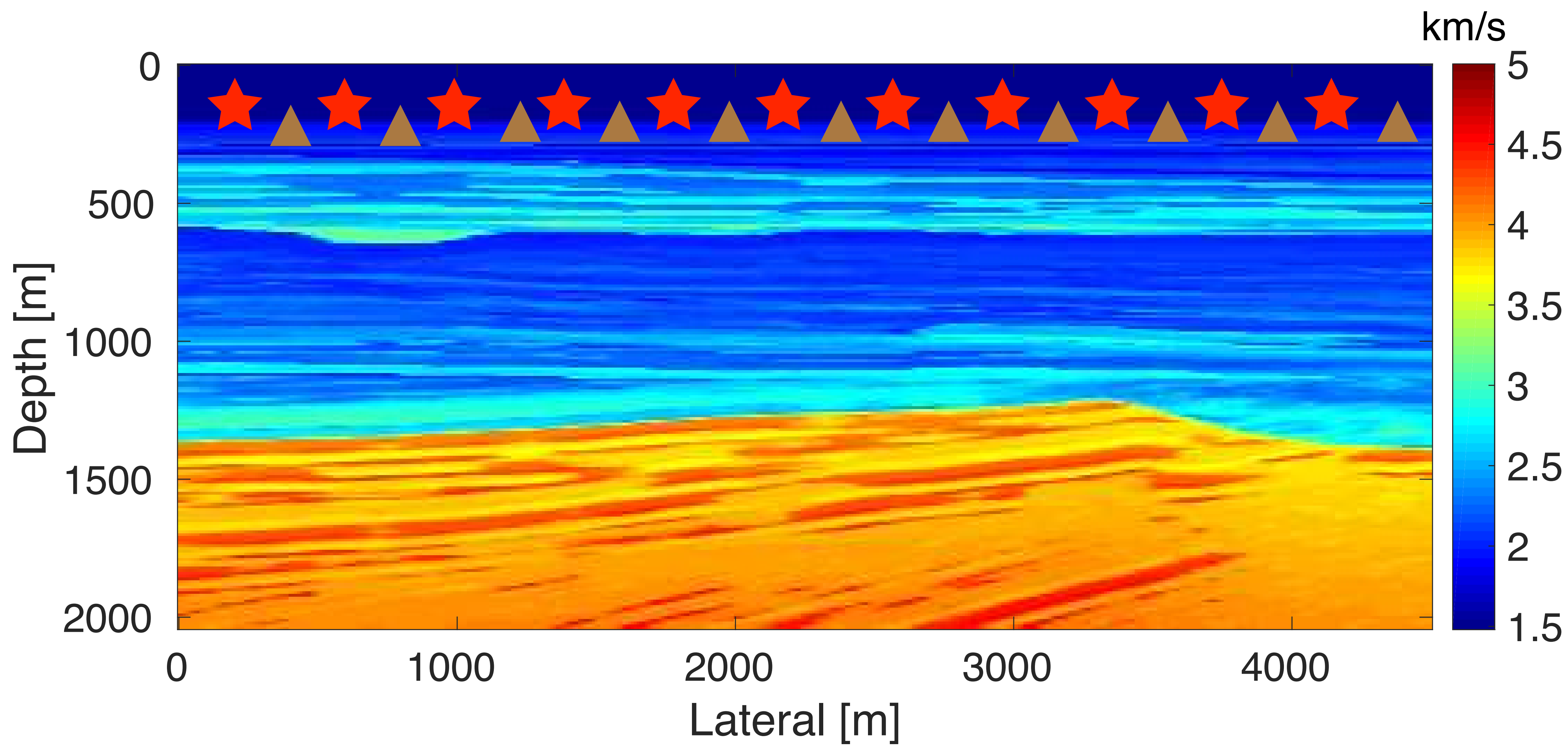
Computational cost:

$$\mathcal{O}(n_f * (n_s + n_r)) \rightarrow \mathcal{O}(n_f * (\tilde{n}_s + \tilde{n}_r))$$

Storage cost:

$$\mathcal{O}(n_f * n_g * (n_s + n_r)) \rightarrow \mathcal{O}(n_f * n_g * (\tilde{n}_s + \tilde{n}_r))$$

BG model



True model

Modeling information:

Model size: 2000m x 4500m

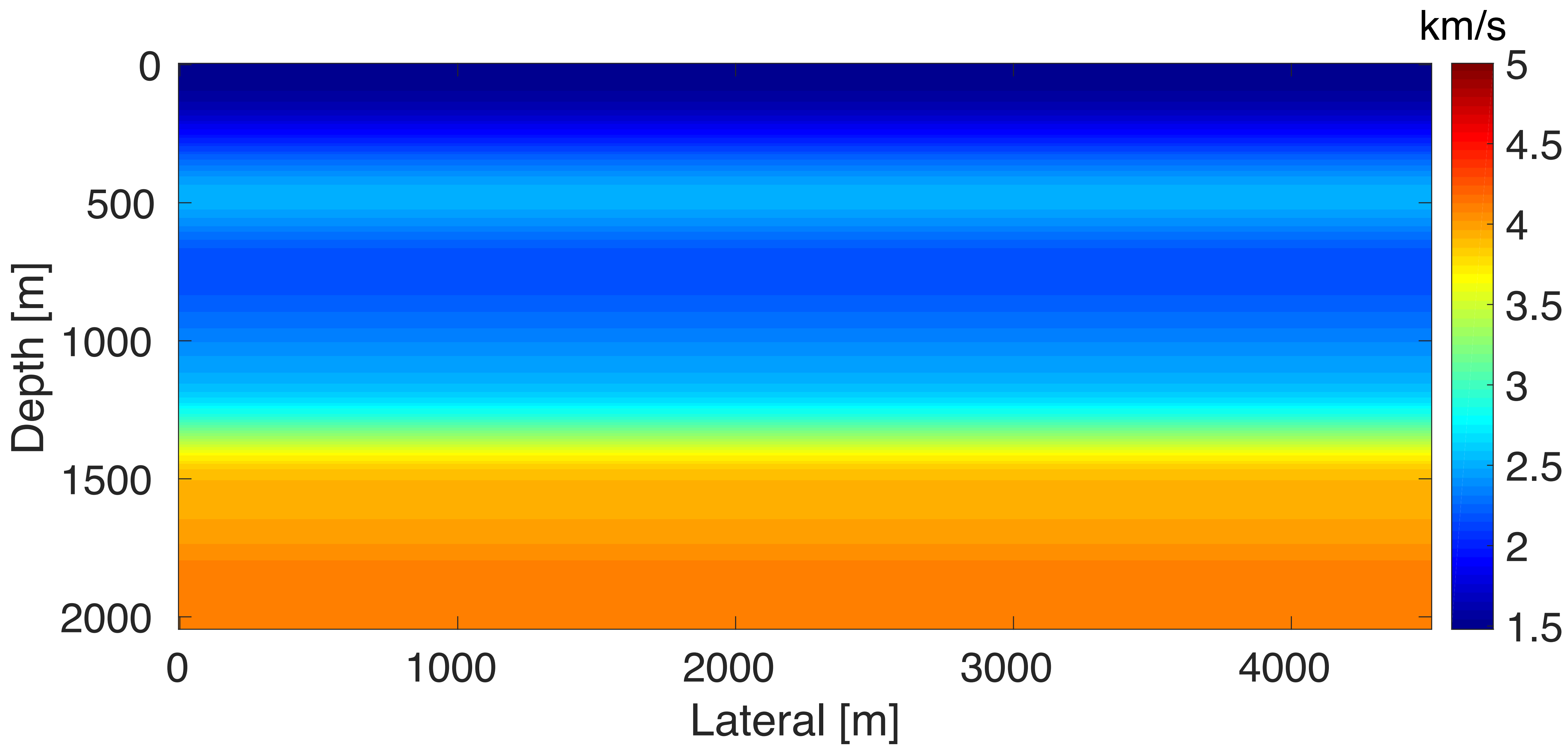
Source spacing: 50m

Receiver spacing: 10m

Fixed spread 4.5km

Frequency : 2~31 Hz

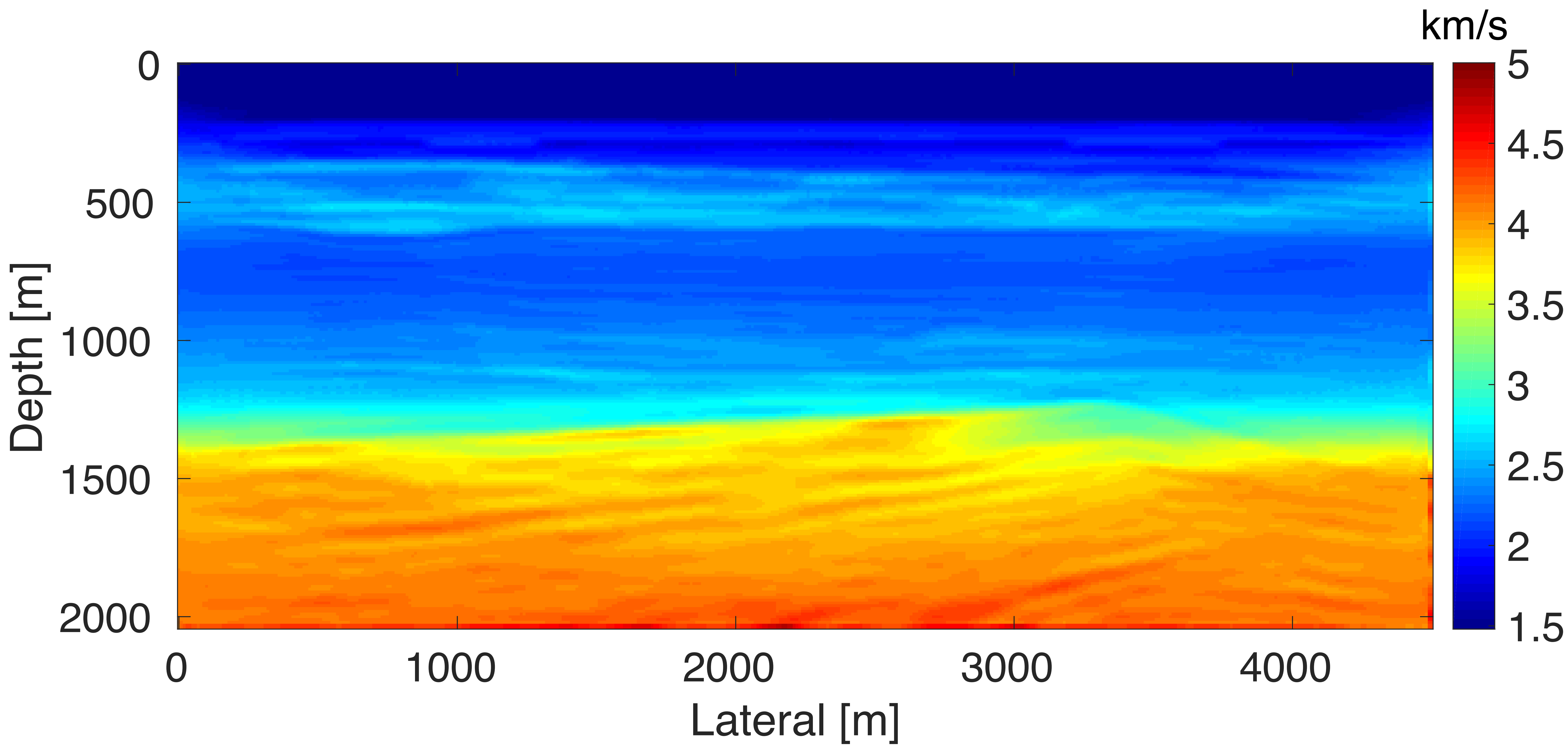
Inversion setting



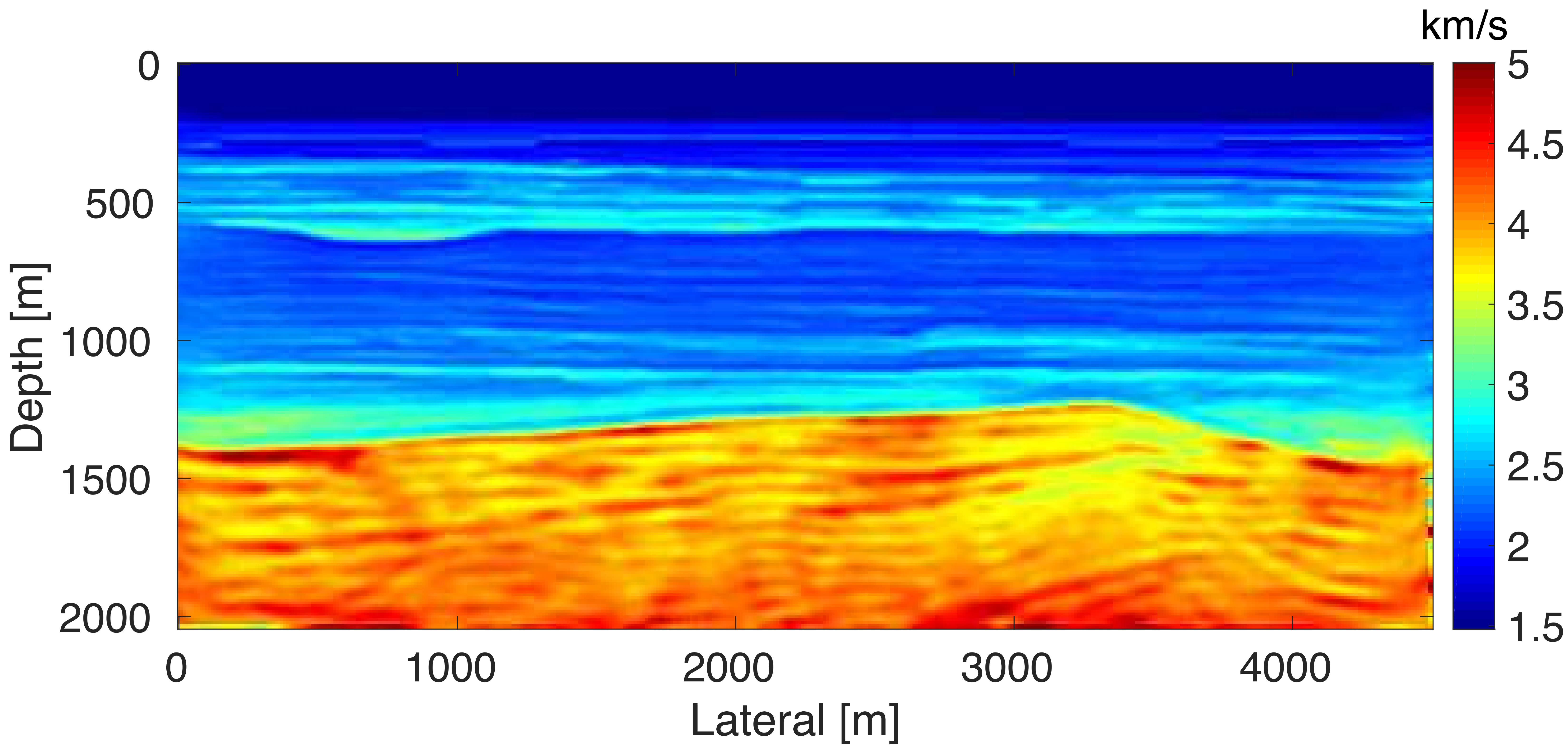
Initial model

1. **GN**
   - 15 simultaneous shots and 76 simultaneous receivers
   - 15 iterations per each frequency band
2. **l-BFGS**
   - all shots and receivers
   - 15 iterations per each frequency band
3. **Gradient-descent (GD)**
   - all shots and receivers
   - 15 iterations per each frequency band

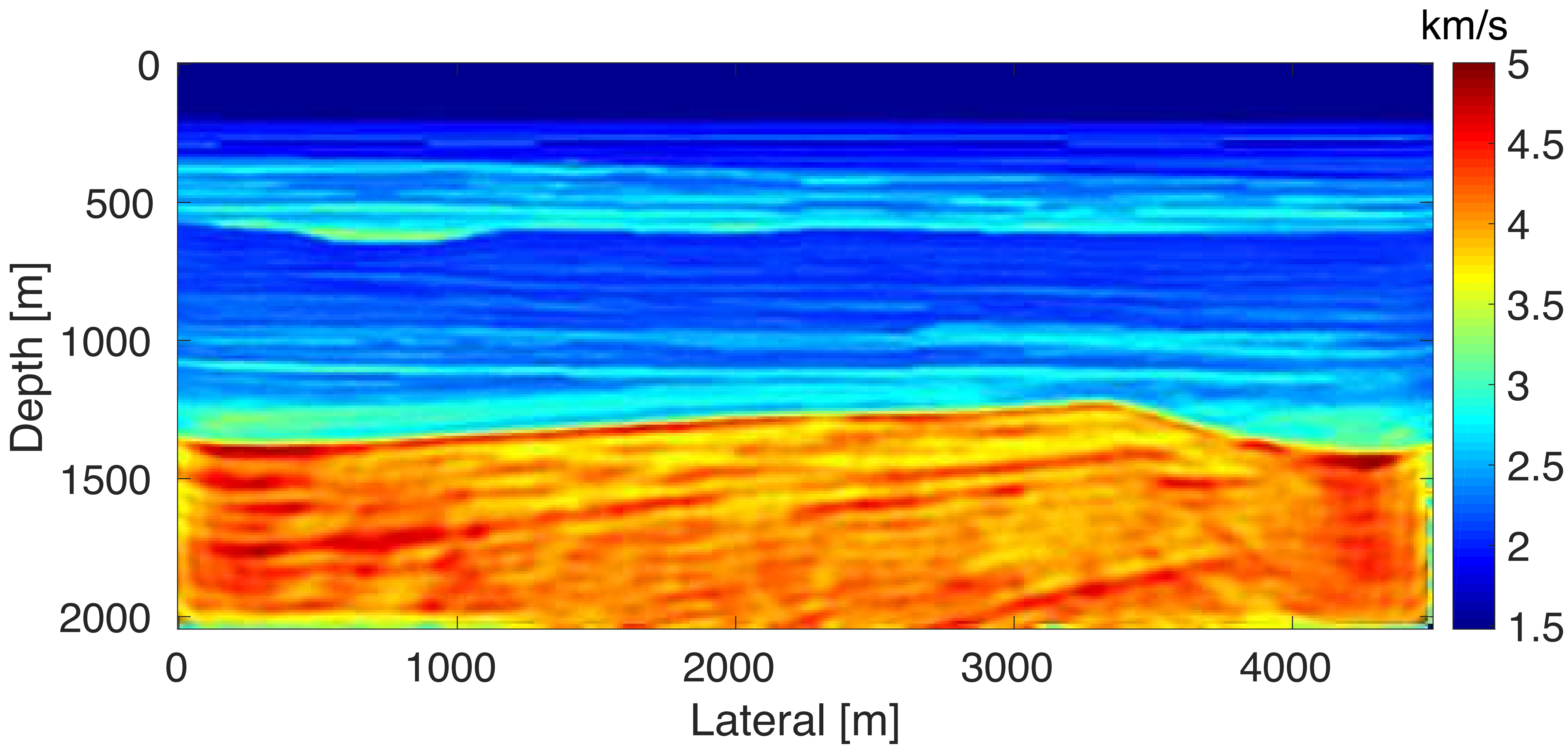
GD result


km/s
5
4.5
4
3.5
3
2.5
2
1.5
0
500
1000
1500
2000
Depth [m]
0
1000
2000
3000
4000
Lateral [m]

l-BFGS result

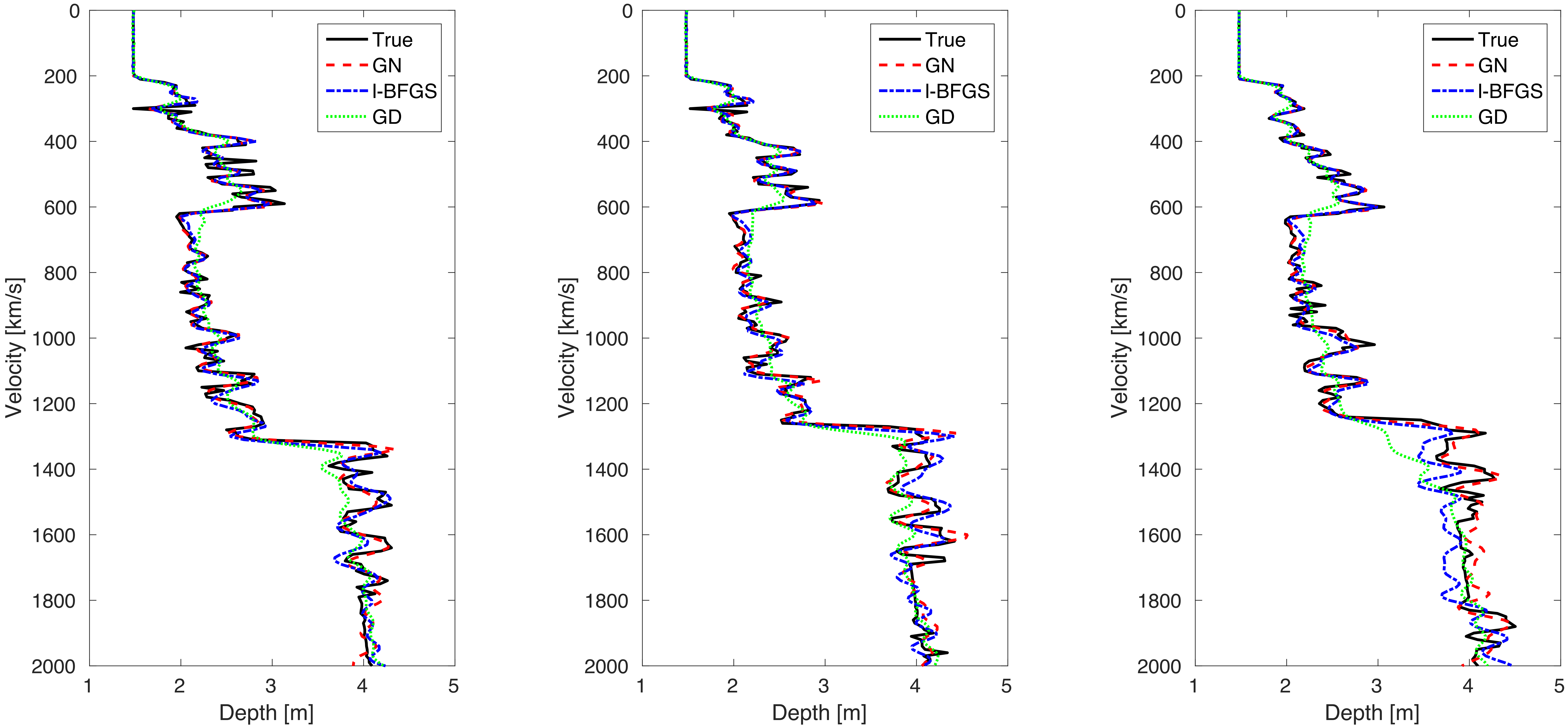


GN result



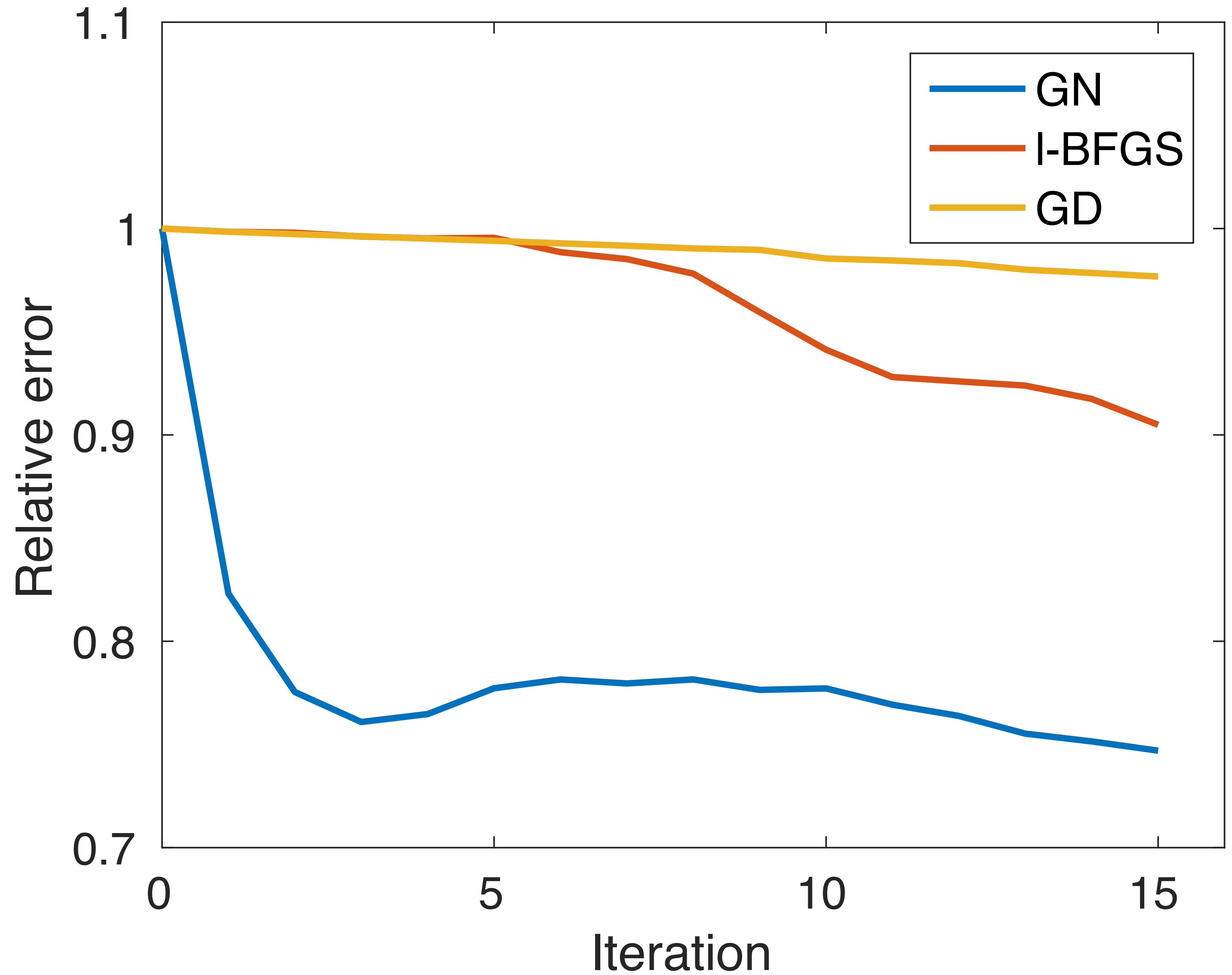
km/s
0
500
1000
1500
2000
Depth [m]
0
1000
2000
3000
4000
Lateral [m]
5
4.5
4
3.5
3
2.5
2
1.5

Vertical profiles

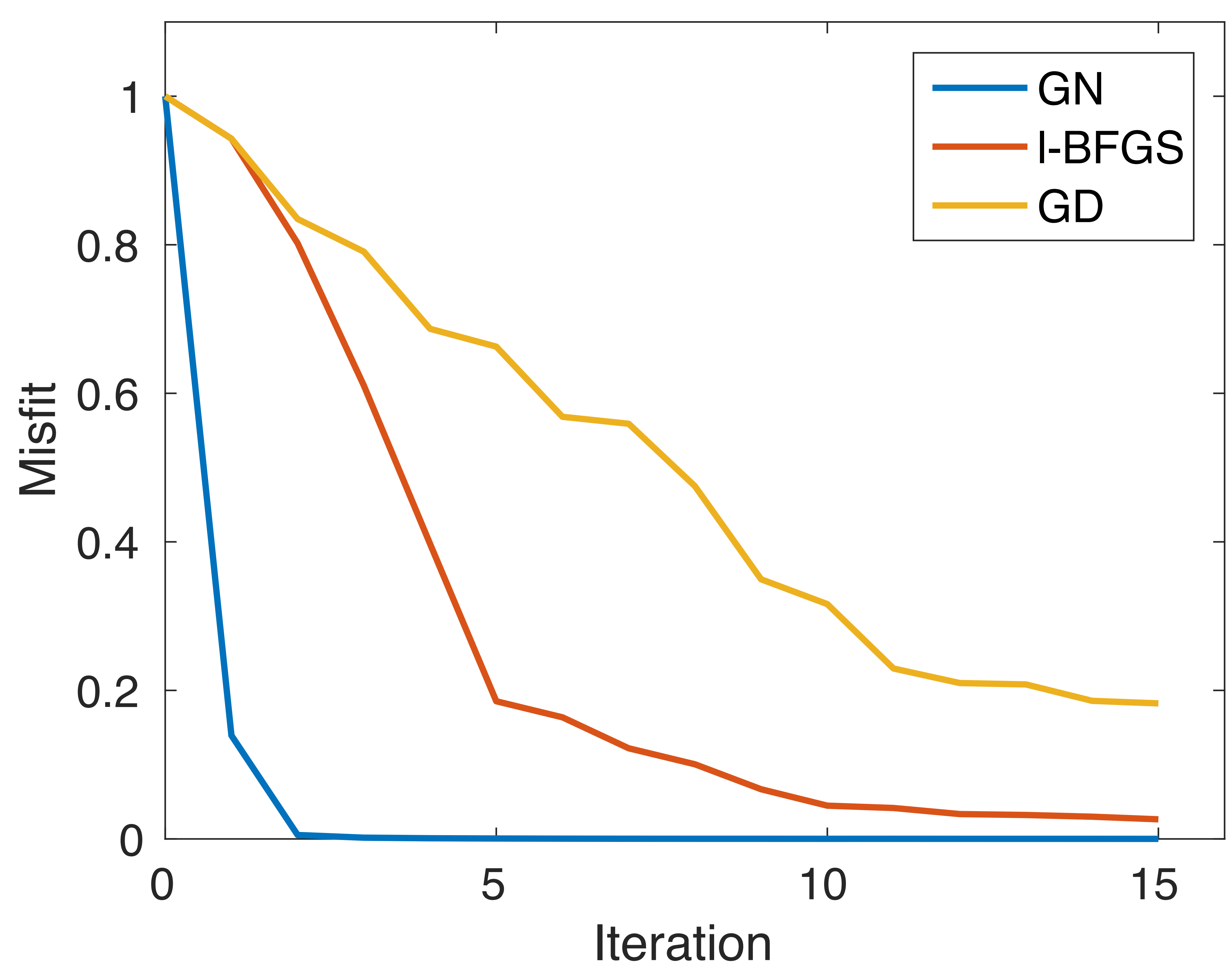


x = 1500 m **x = 2500 m** **x = 3500 m**

Model error and misfit



Model error



Misfit

Application to uncertainty quantification

[A. Tarantola and B. Valette, 1982]
[J. Kaipio and E. Somersalo, 2004]

SLIM

Bayesian inference

Prior probability density function (PDF):

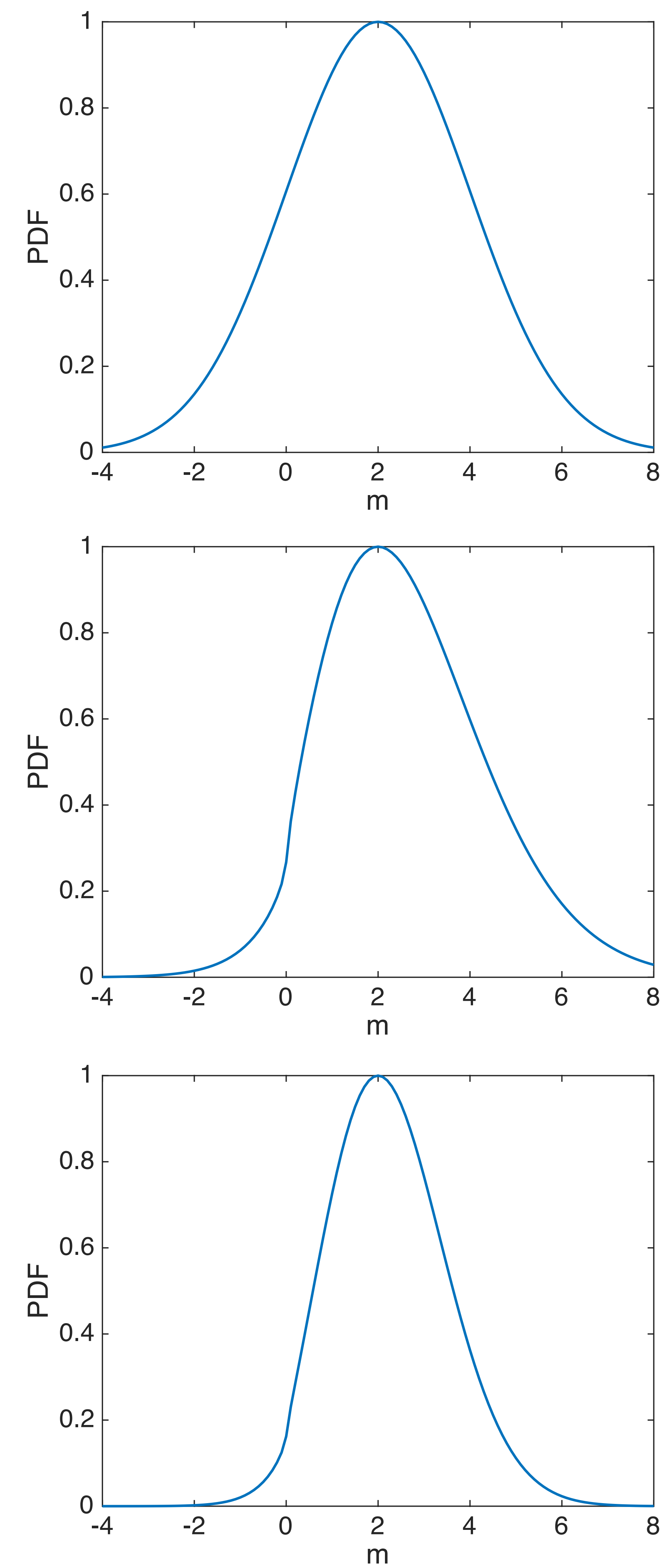
$$\mathbf{m} \longrightarrow \rho_{\text{prior}}(\mathbf{m})$$

Likelihood PDF: given data $\mathbf{d}$

$$\mathbf{m} \longrightarrow \rho_{\text{like}}(\mathbf{d}|\mathbf{m})$$

Posterior PDF (Bayes' rule):

$$\rho_{\text{post}}(\mathbf{m}|\mathbf{d}) = \rho_{\text{like}}(\mathbf{d}|\mathbf{m})\rho_{\text{prior}}(\mathbf{m})$$



Bayesian inference

Mean value of the model:

$$\mathbb{E}(\mathbf{m}) = \int \mathbf{m}\rho_{\text{post}}(\mathbf{m})d\mathbf{m},$$

Covariance matrix:

$$C_{i,j} = \mathbb{E}(m_i m_j) - \mathbb{E}(m_i)\mathbb{E}(m_j),$$

Marginal distribution:

$$\rho_{\text{M}}(m_i) = \int \cdots \int \rho_{\text{post}}(\mathbf{m}|\mathbf{d}) \prod_{j=1, j\neq i}^{n_{\text{grid}}} dm_j \quad .$$

SLIM

Bayesian w/ WRI

Posterior PDF of WRI:

$$\rho_{\text{post}}(\mathbf{m}|\mathbf{d}) \propto \rho_{\text{like}}(\mathbf{d}|\mathbf{m})\rho_{\text{prior}}(\mathbf{m})$$

$$\rho_{\text{like}}(\mathbf{d}|\mathbf{m}) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n_s}\sum_{l=1}^{n_f}\left(\|\mathbf{P}\overline{\mathbf{u}}_{i,l} - \mathbf{d}_{i,l}\|^2_{\mathbf{\Sigma}^{-1}_{\text{noise}}} + \lambda^2\|\mathbf{A}_{i,l}\overline{\mathbf{u}}_{i,l} - \mathbf{q}_{i,l}\|^2\right)\right)$$

$$\rho_{\text{prior}}(\mathbf{m}) \propto \exp\left(-\frac{1}{2}\|\mathbf{m} - \mathbf{m}_p\|^2_{\mathbf{\Sigma}^{-1}_{\text{prior}}}\right)$$

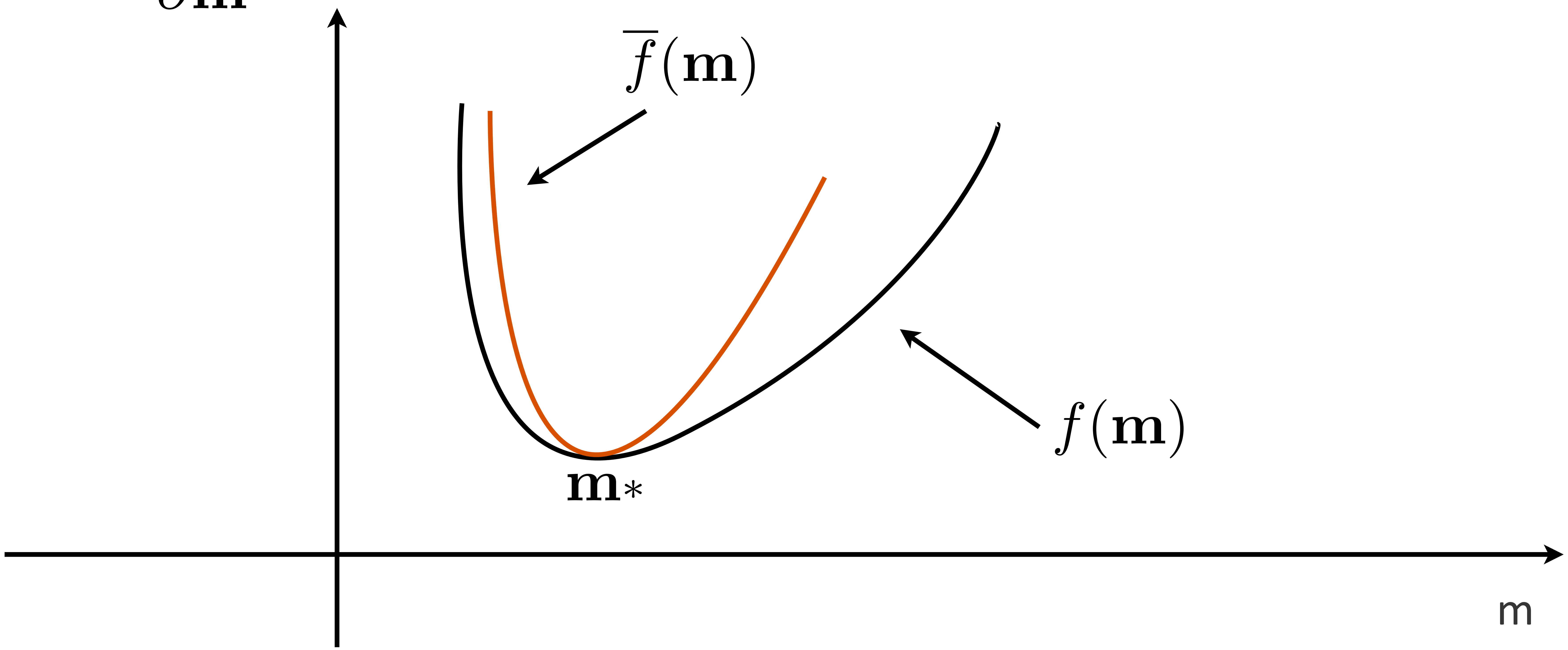
where,

$$\begin{pmatrix}\lambda\mathbf{A}_{i,l}\\ \mathbf{\Sigma}^{-1/2}_{\text{noise}}\mathbf{P}\end{pmatrix}\overline{\mathbf{u}}_{i,l} = \begin{pmatrix}\lambda\mathbf{q}_{i,l}\\ \mathbf{\Sigma}^{-1/2}_{\text{noise}}\mathbf{d}_{i,l}\end{pmatrix}.$$

Quadratic approximation of $-\log \rho_{\text{post}}(\mathbf{m})$

$$-\log \rho_{\text{post}}(\mathbf{m}) = f(\mathbf{m})$$
$$\approx f(\mathbf{m}_*) + \frac{1}{2}(\mathbf{m} - \mathbf{m}_*)^\top \mathbf{H}(\mathbf{m} - \mathbf{m}_*) := \overline{f}(\mathbf{m})$$

where, $\mathbf{H} = \dfrac{\partial^2 f}{\partial \mathbf{m}^2}$.



Approximate posterior PDF

Gaussian approximation:

$$\rho_{\text{post}}(\mathbf{m}) \approx \rho_{\text{Gauss}}(\mathbf{m}) = \mathcal{N}(\mathbf{m}_*, \mathbf{H}^{-1})$$

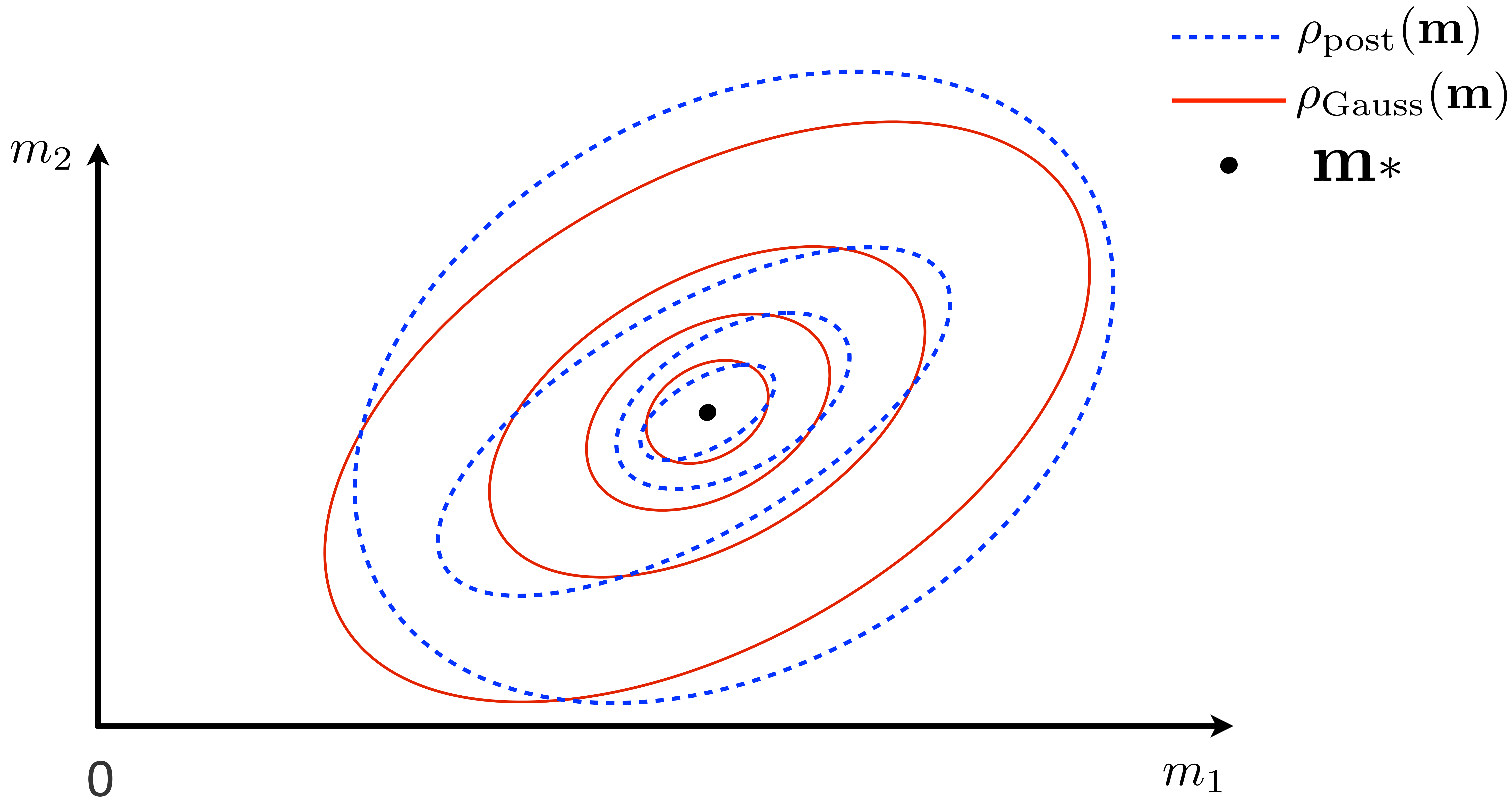
where

$$\mathbf{H} = \mathbf{H}_{\text{l}} + \mathbf{H}_{\text{p}},$$

$$\mathbf{H}_{\text{l}} = \frac{\partial^2 f_{\text{l}}(\mathbf{m})}{\partial \mathbf{m}^2}, \quad f_{\text{l}}(\mathbf{m}) = \text{-log}\rho_{\text{like}}(\mathbf{d}|\mathbf{m}),$$

$$\mathbf{H}_{\text{p}} = \frac{\partial^2 f_{\text{p}}(\mathbf{m})}{\partial \mathbf{m}^2}, \quad f_{\text{p}}(\mathbf{m}) = \text{-log}\rho_{\text{prior}}(\mathbf{m}).$$

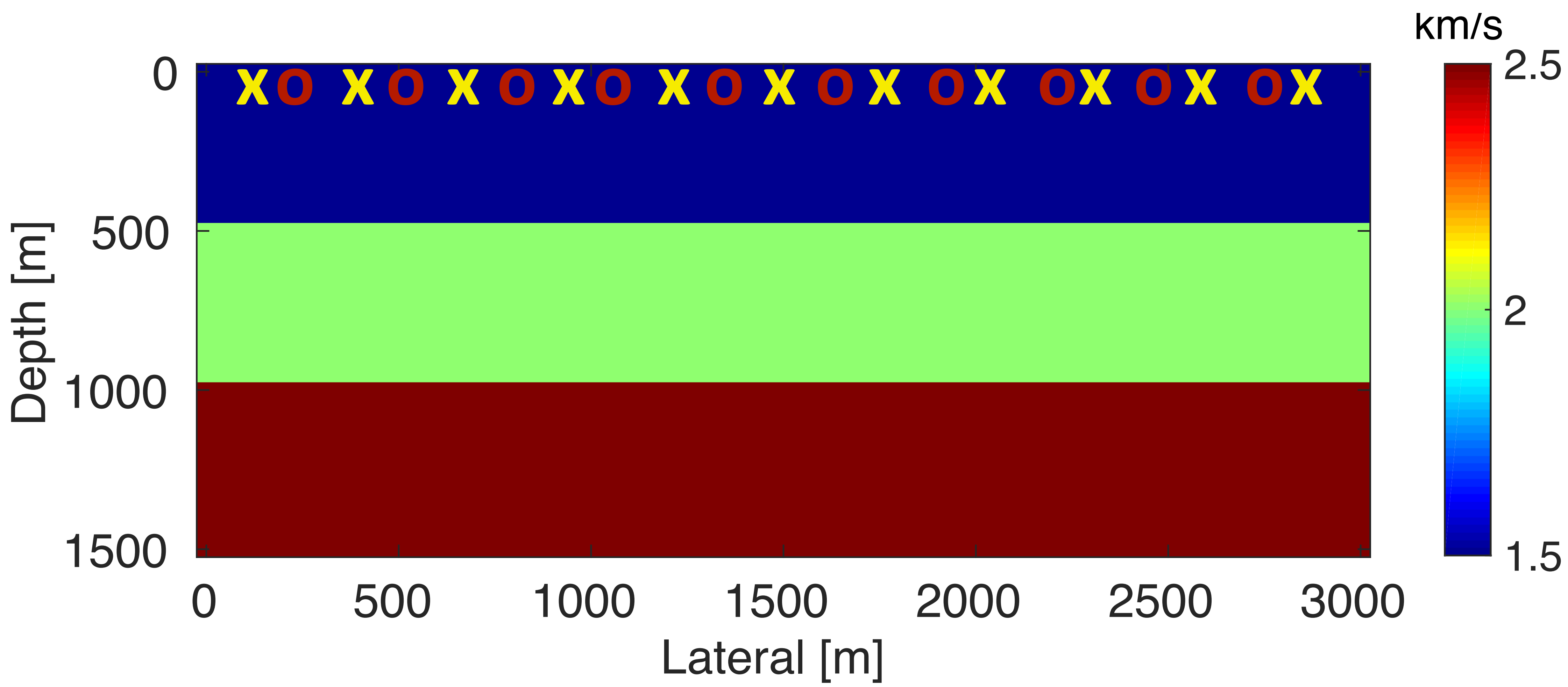
Approximate posterior PDF



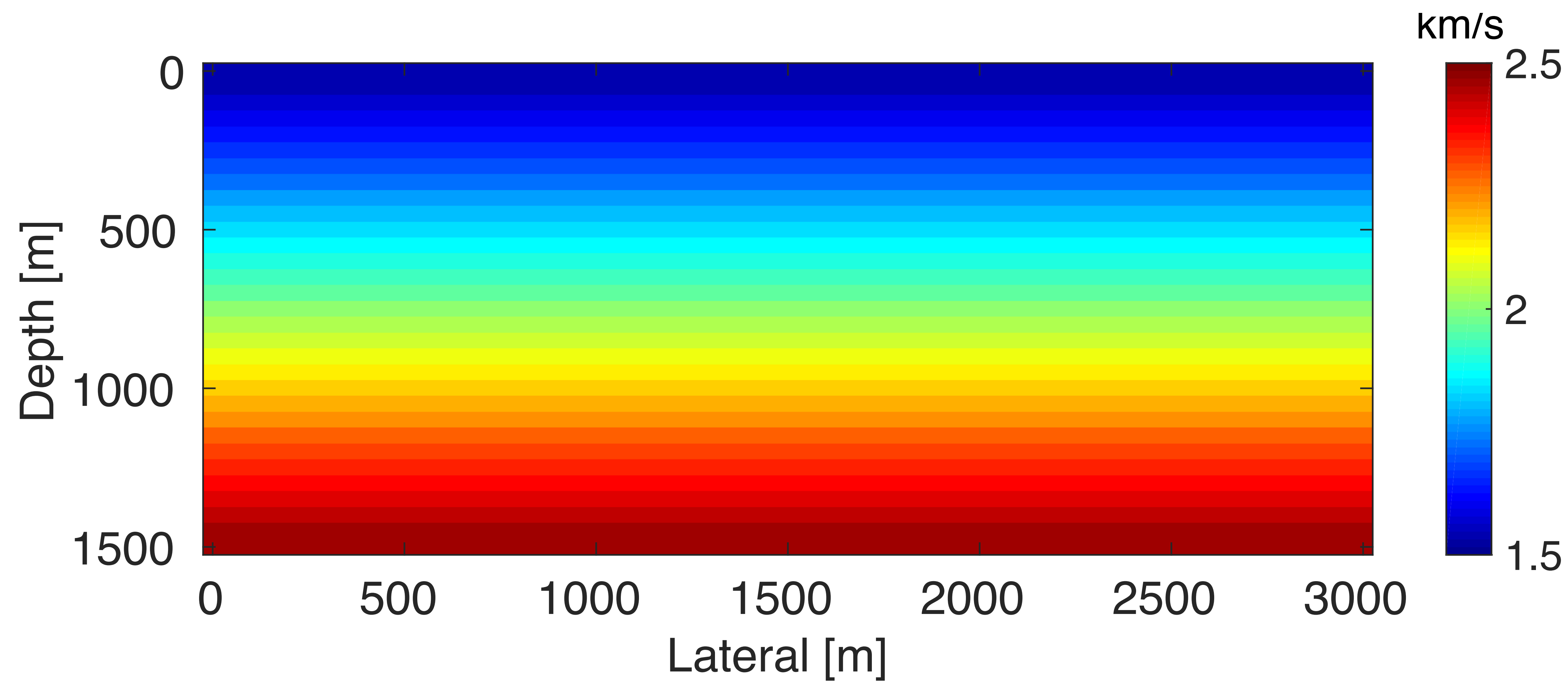
[1] [van Leeuwen, T and Herrmann, F J , 2015]

Numerical Experiment – Layer model

Depth of sources and receivers: 50 m
Number of sources and receivers: 61
Frequency: 5,6 and 7 Hz
Lambda: computed according to [1]
sigma: 10
15% Gaussian noise

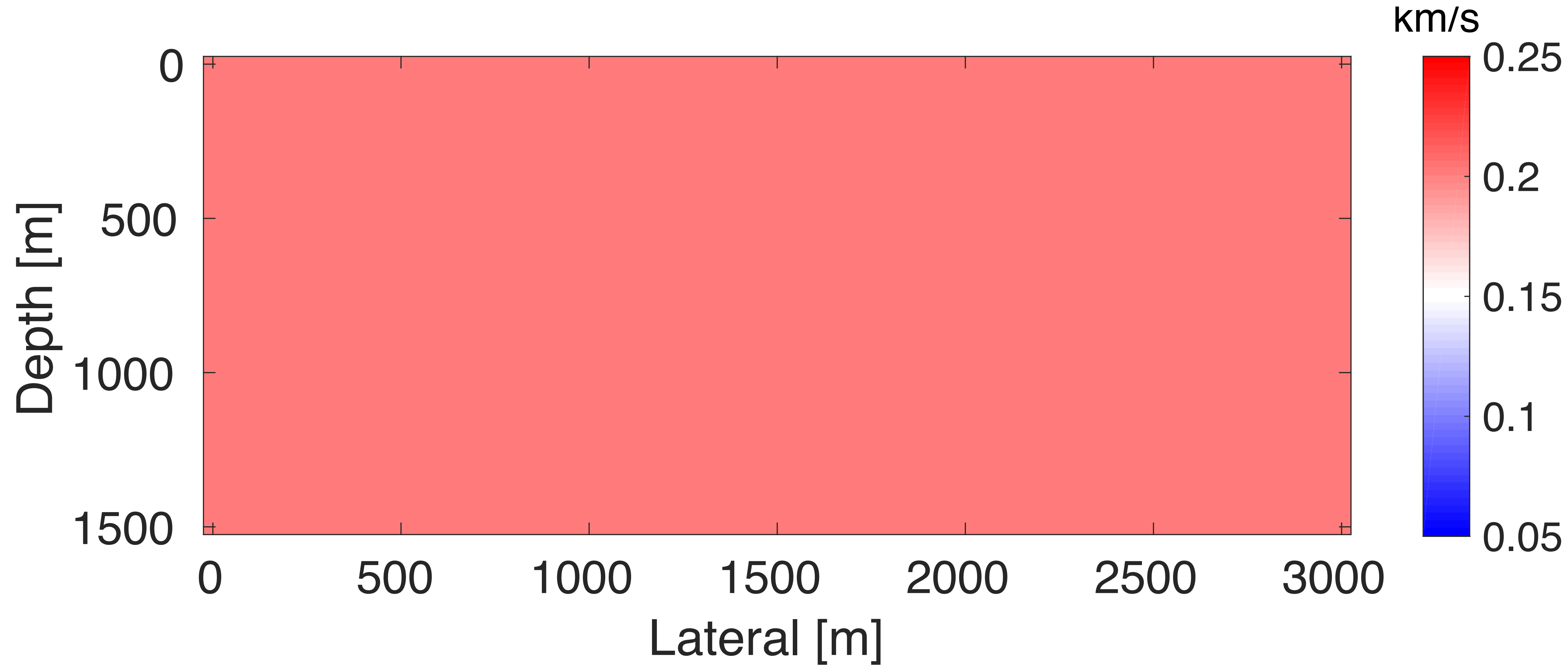


(a) True model

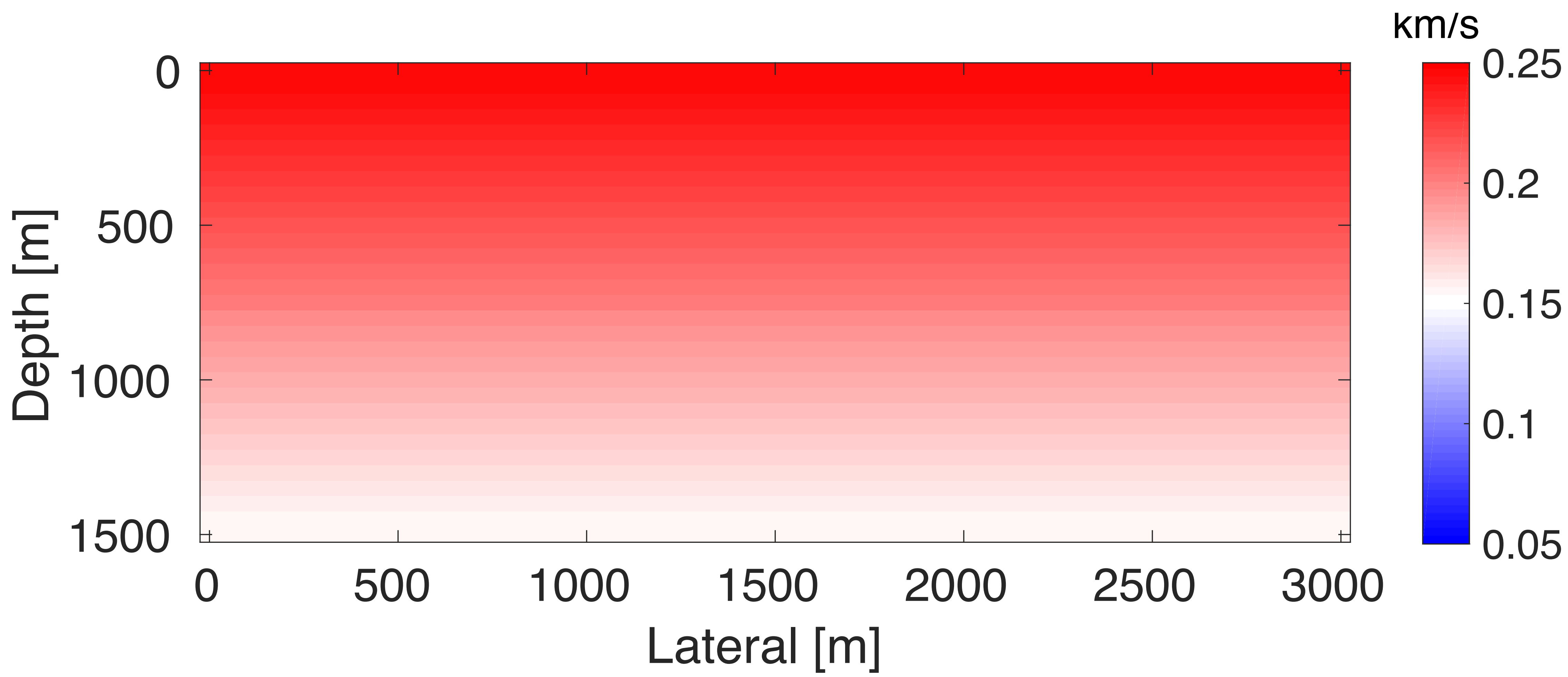


(b) Prior model

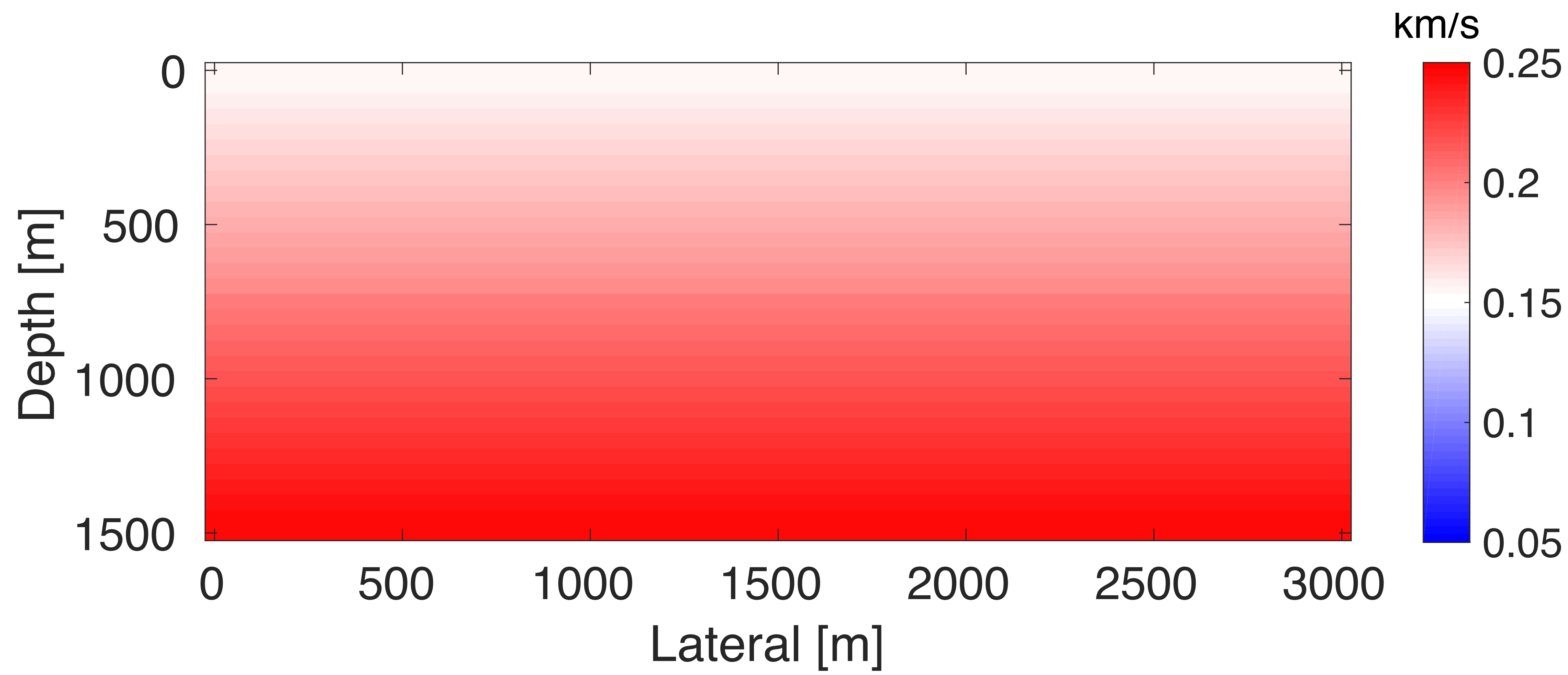
Prior standard deviations



(a) Standard deviation 1

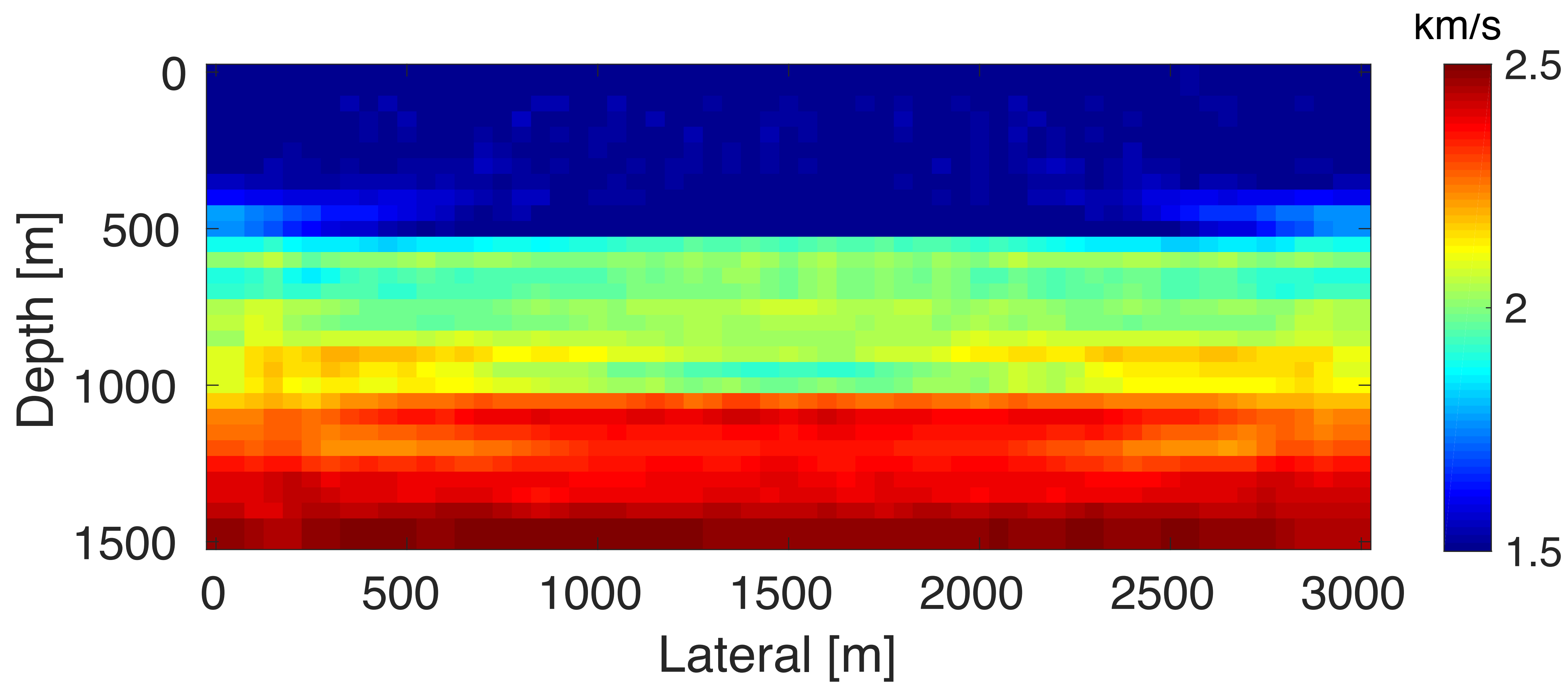


(b) Standard deviation 2

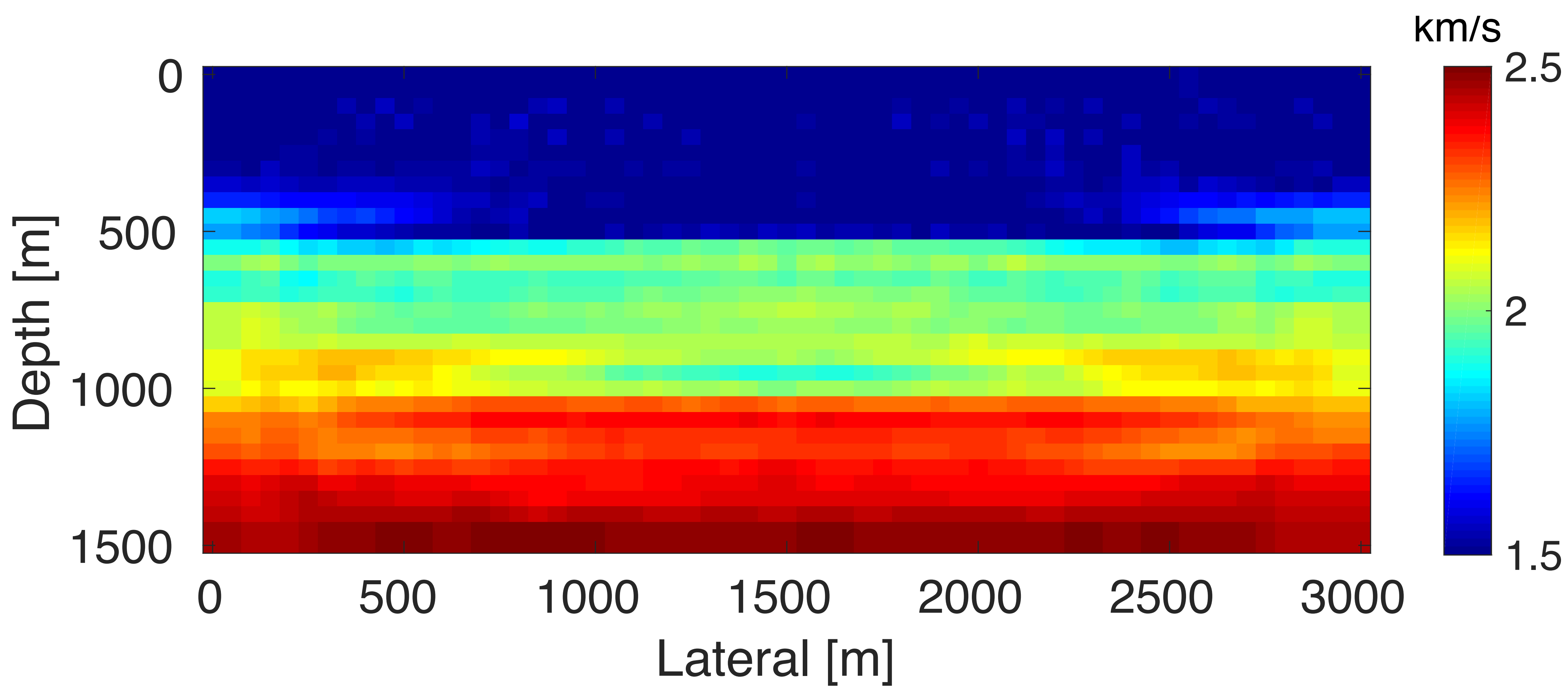


(c) Standard deviation 3

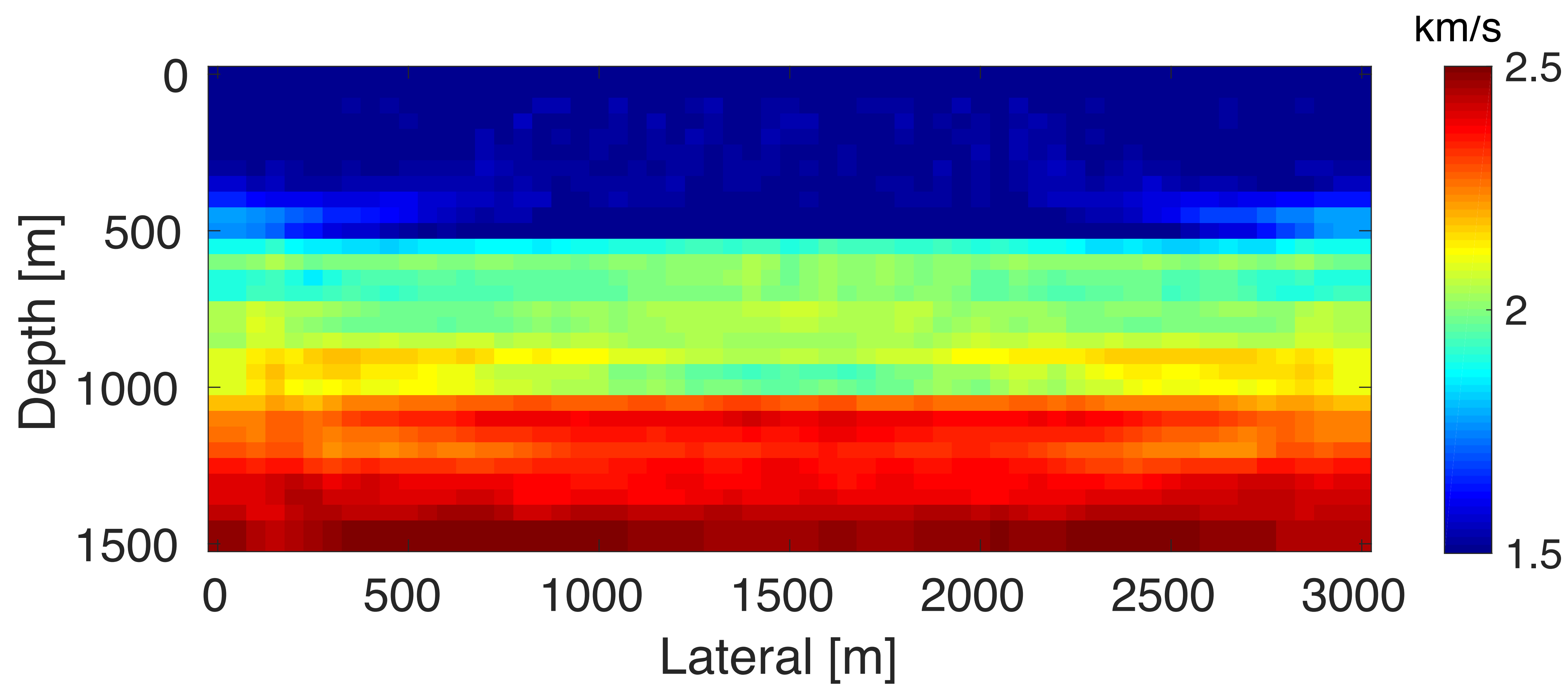
MAPs



(a) MAP1

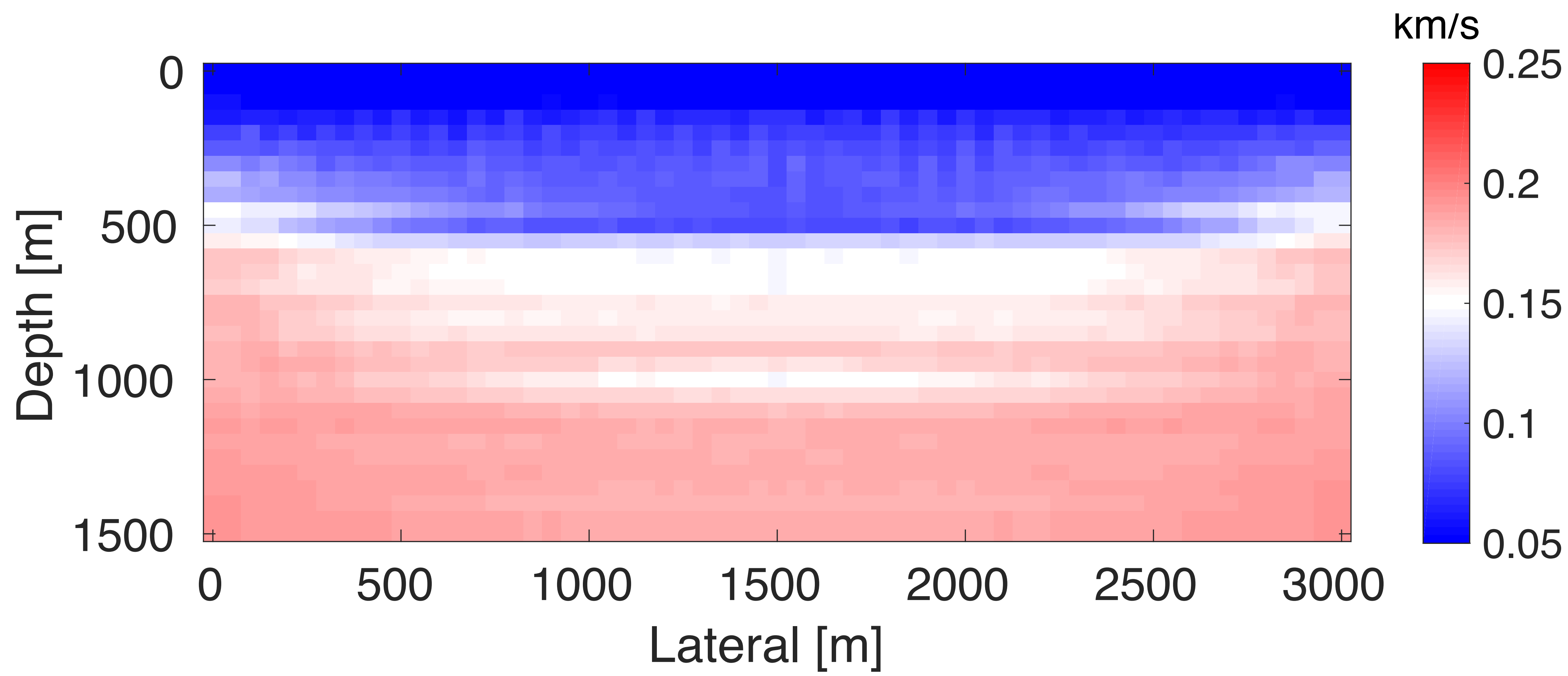


(b) MAP 2

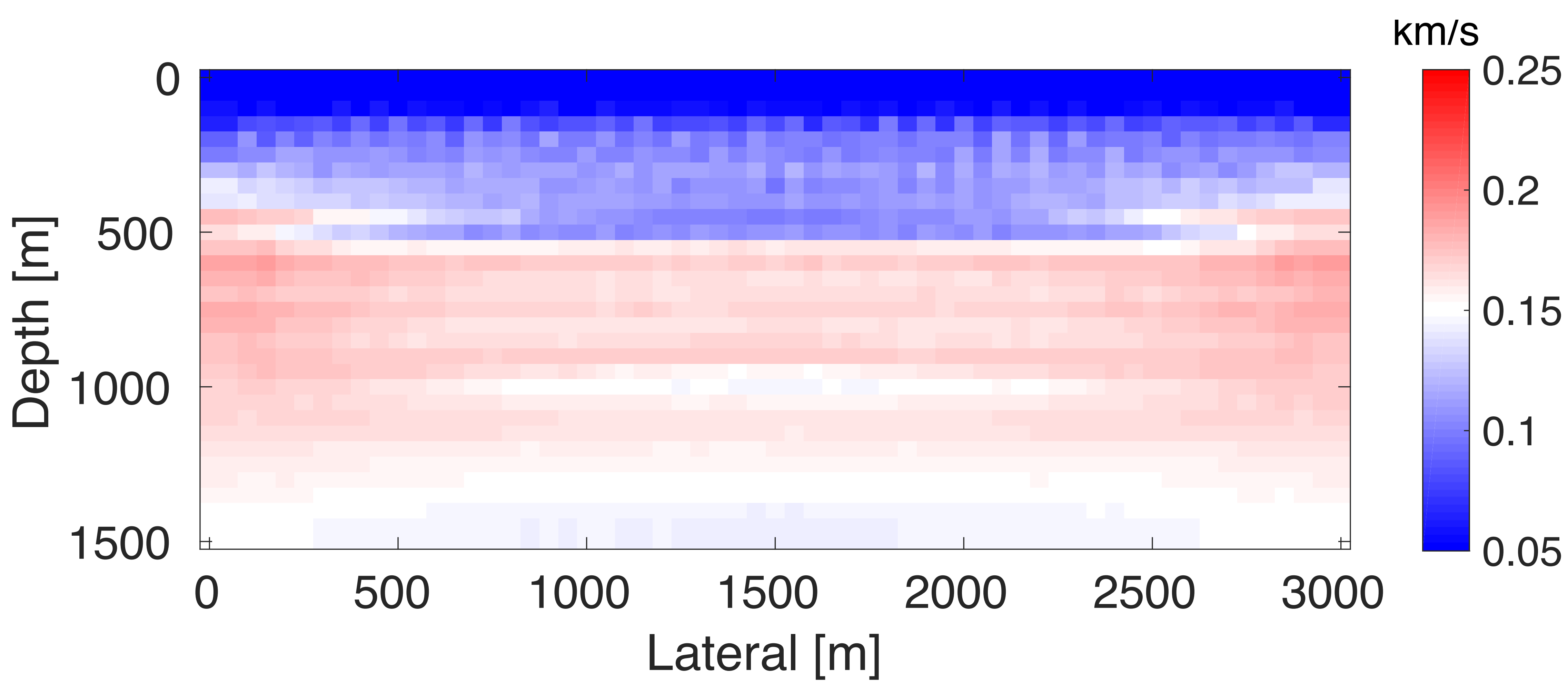


(c) MAP 3

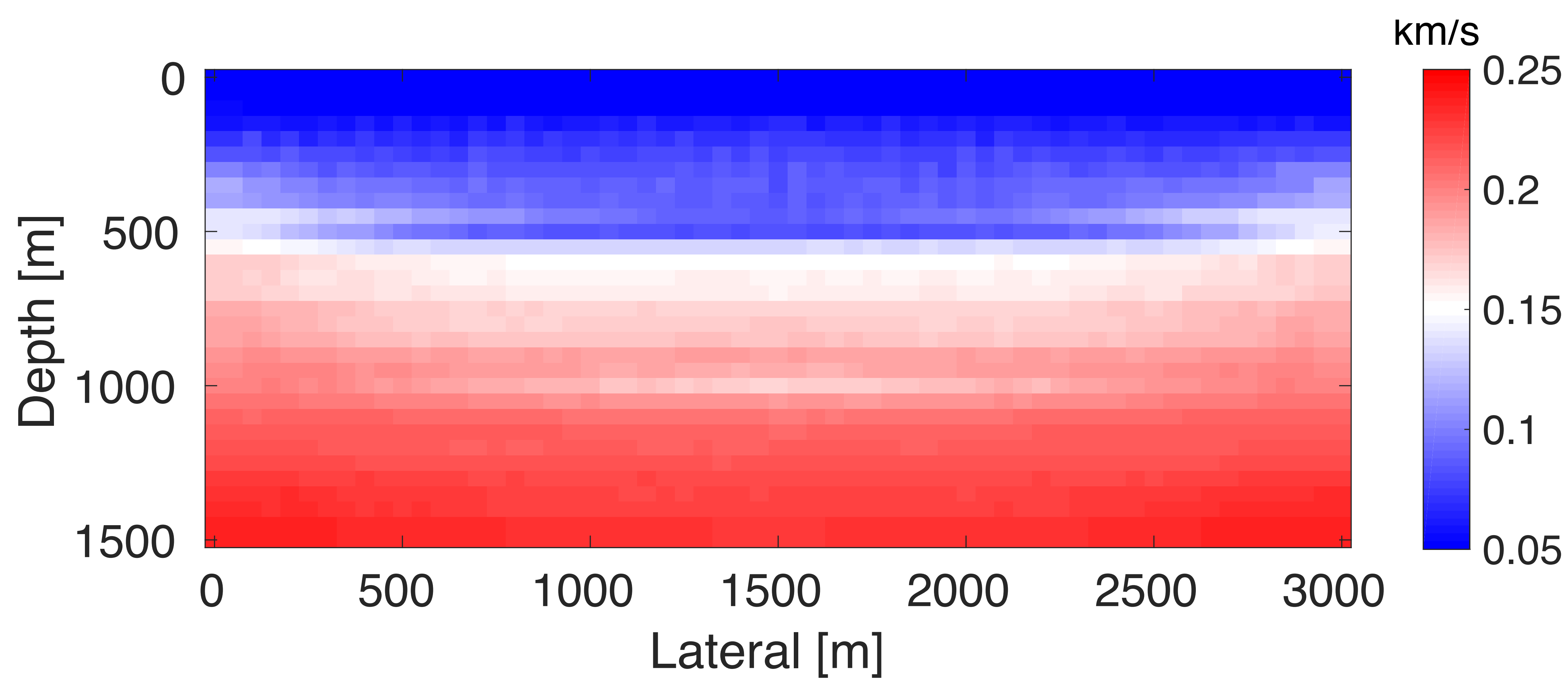
Posterior standard deviations



(a) Standard deviation 1

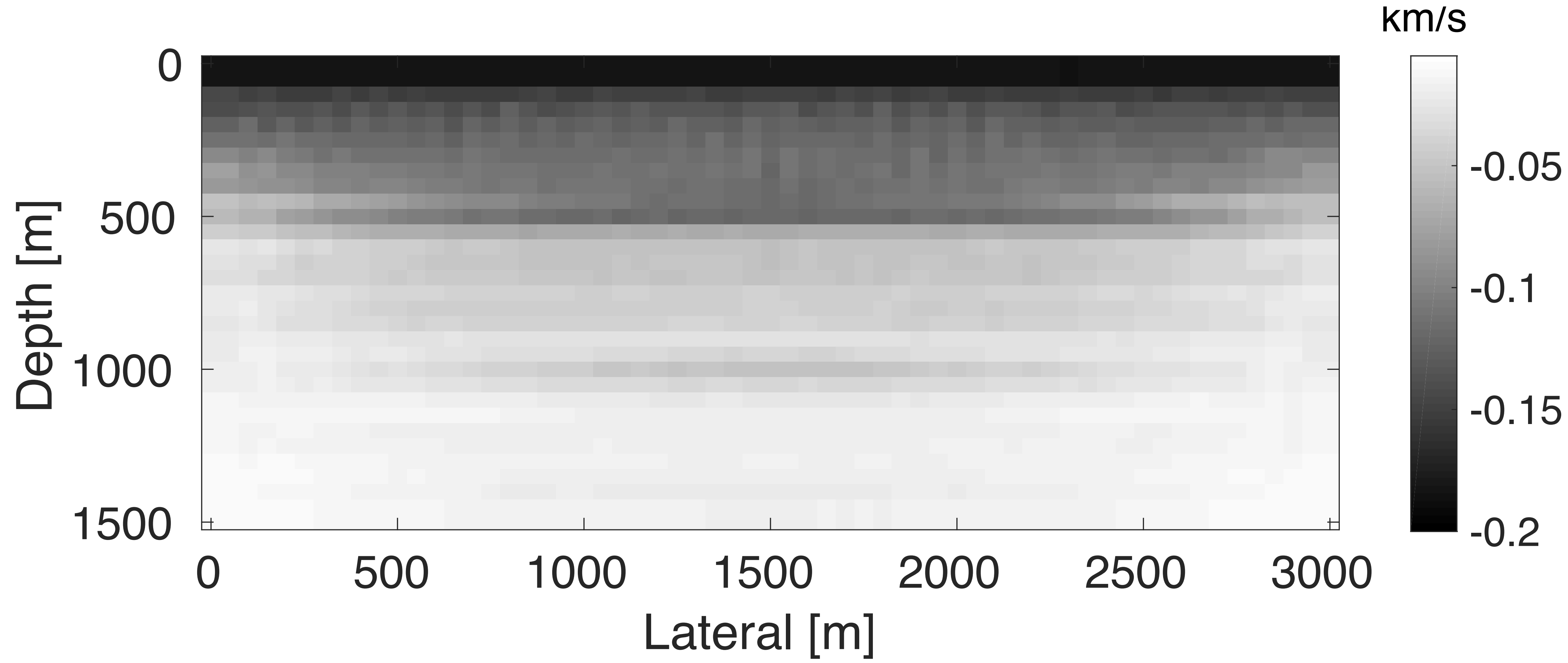


(b) Standard deviation 2

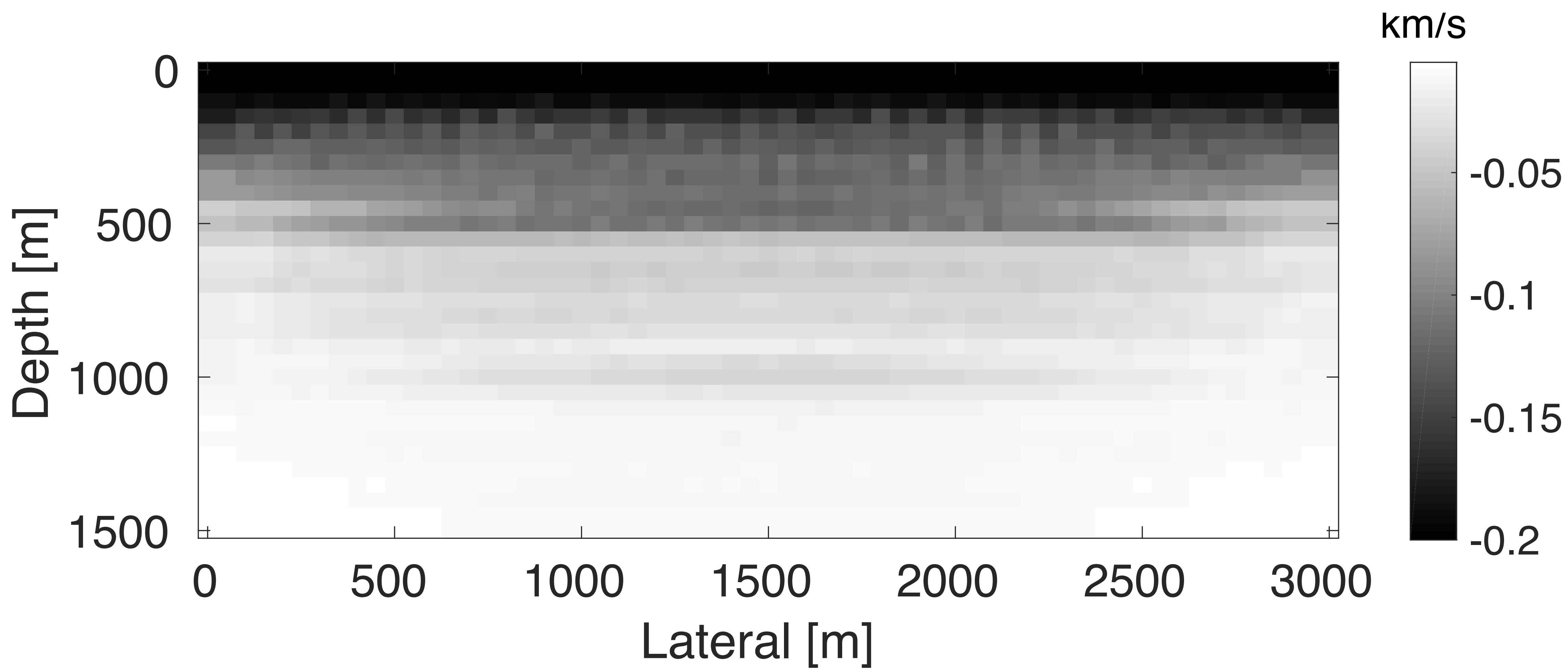


(c) Standard deviation 3

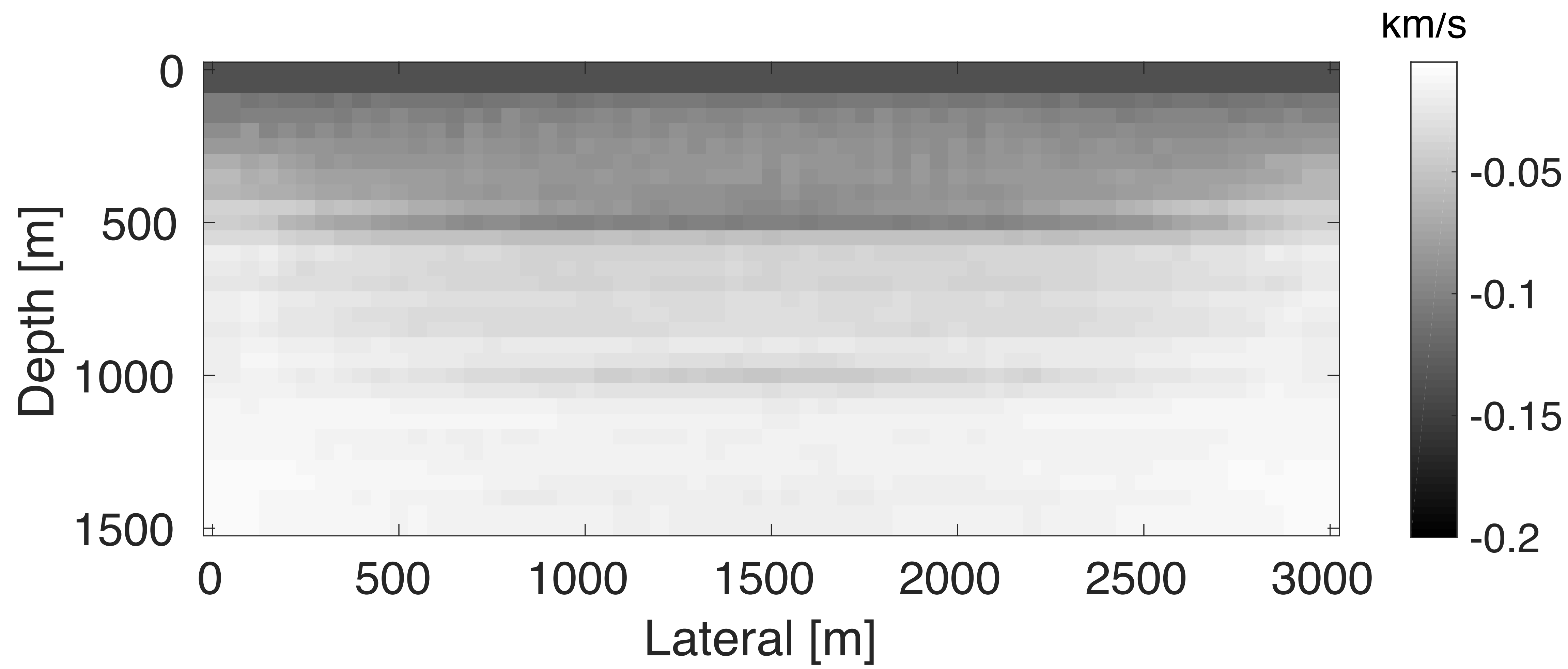
Differences between posterior and prior



(a) Difference 1



(b) Difference 2

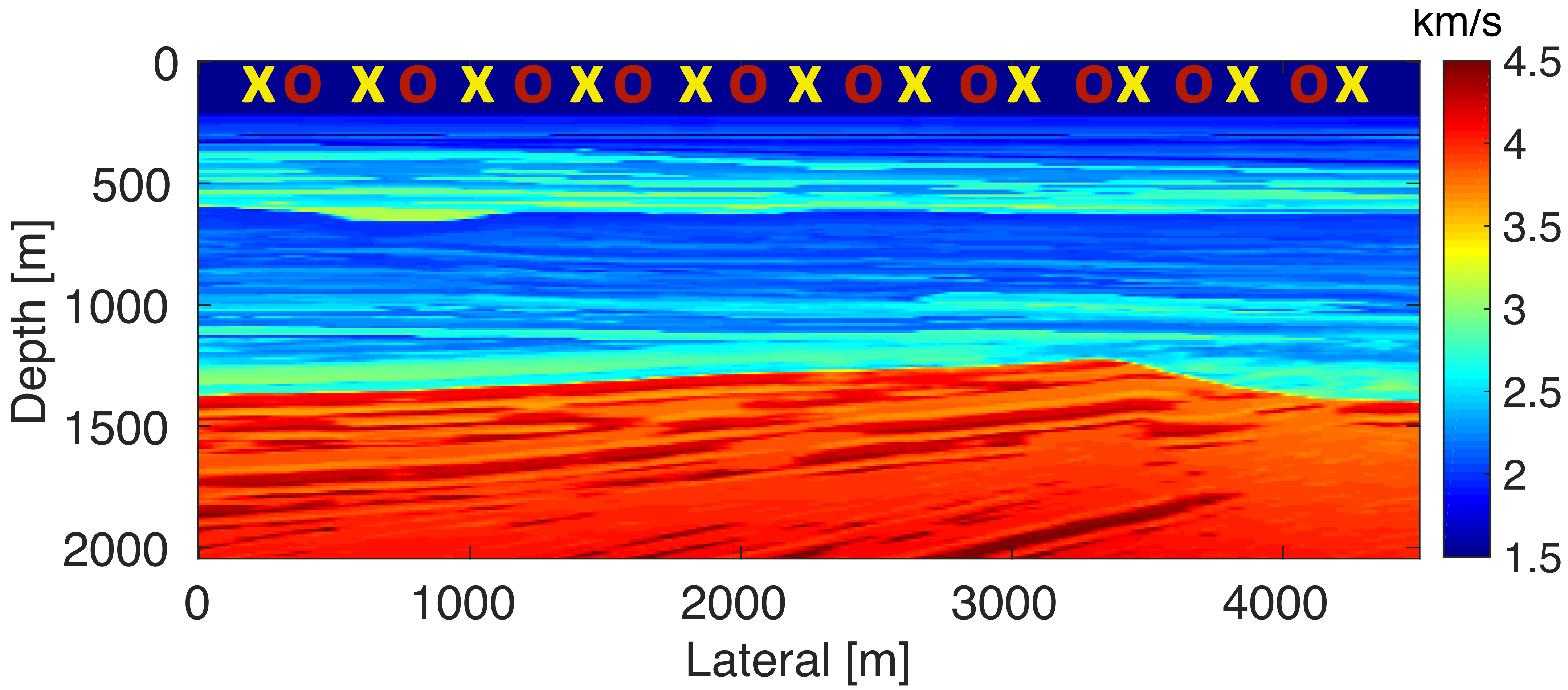


(c) Difference 3

[1] [van Leeuwen, T and Herrmann, F J , 2015]

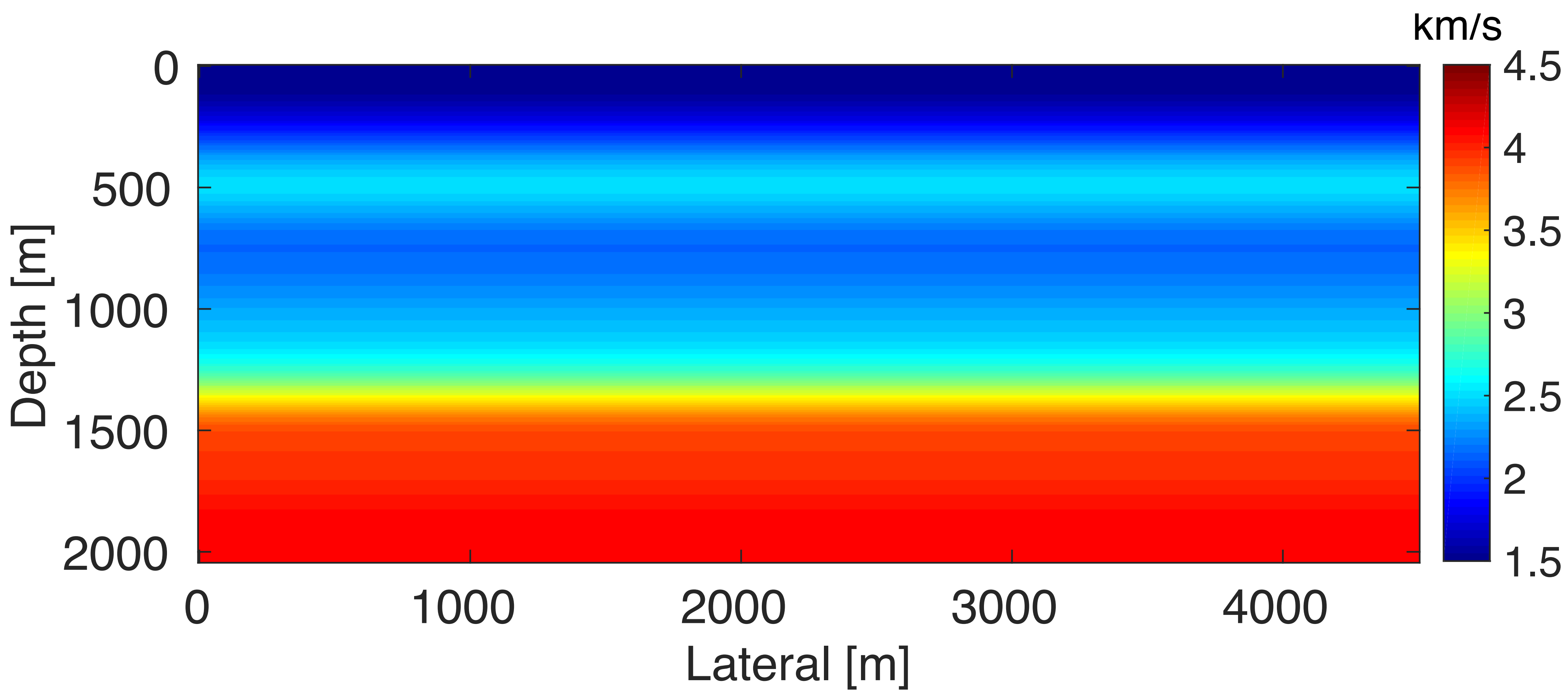
SLIM

BG Compass model

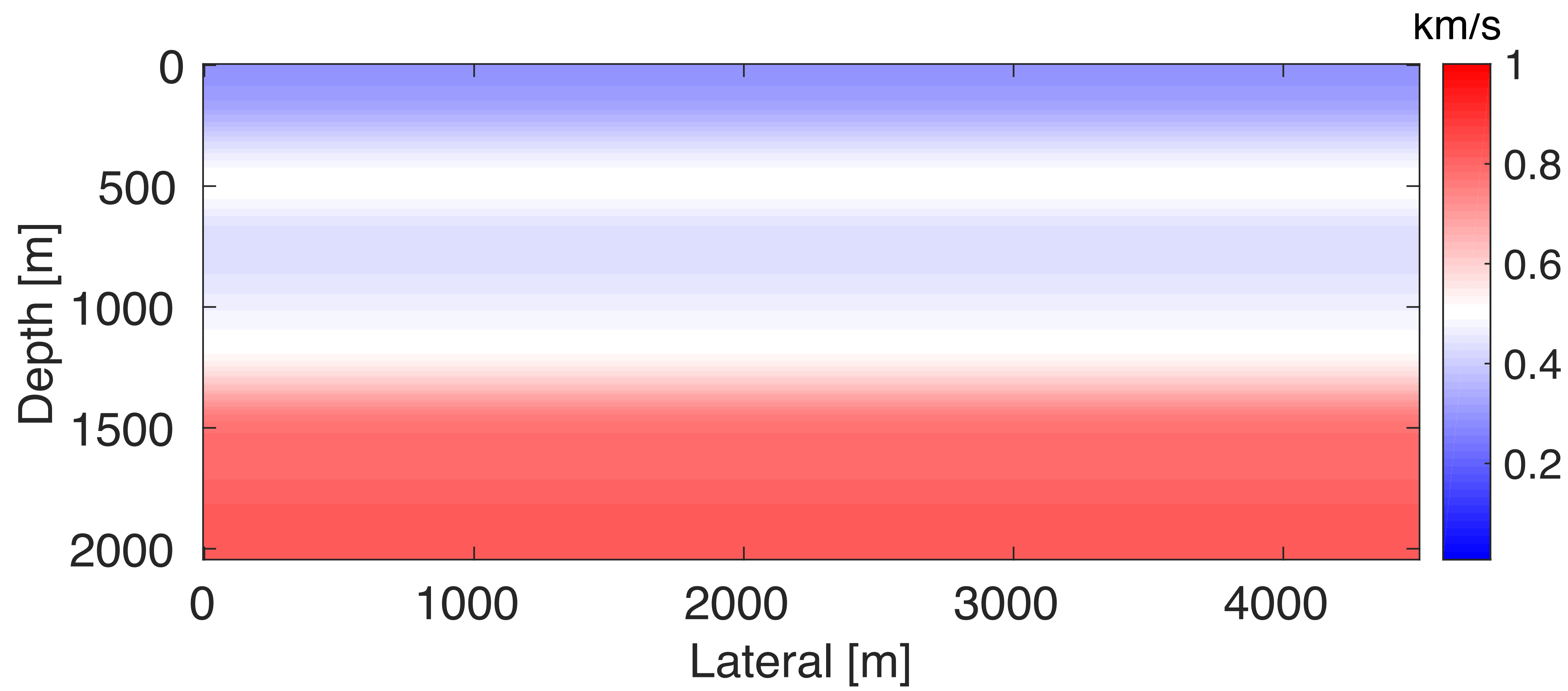


(a) True model

Depth of sources and receivers: 50
Number of sources and receivers: 91 / 451
Central frequency: 15 Hz
Frequency: 2-31 Hz
Lambda: computed according to [1]

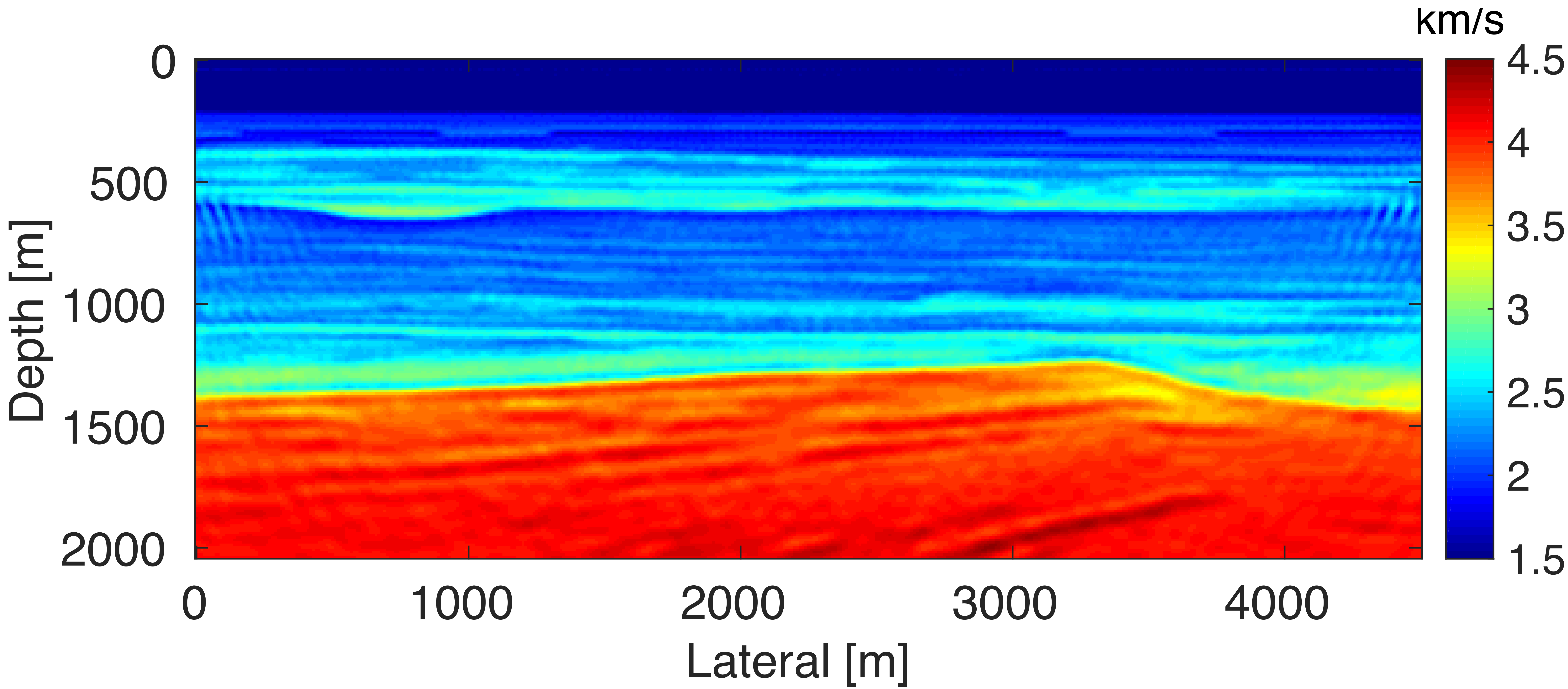


(b) Prior model



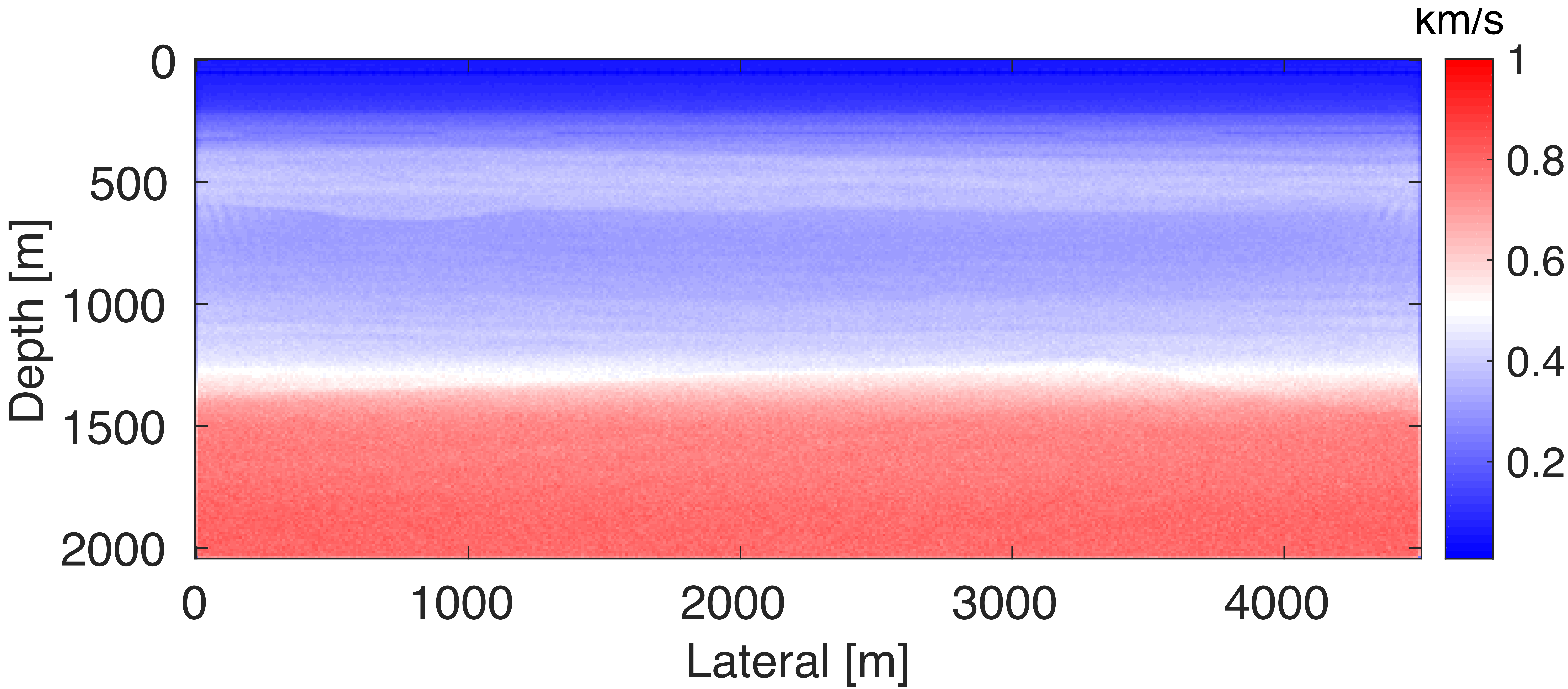
(c) STD of prior distribution

MAP

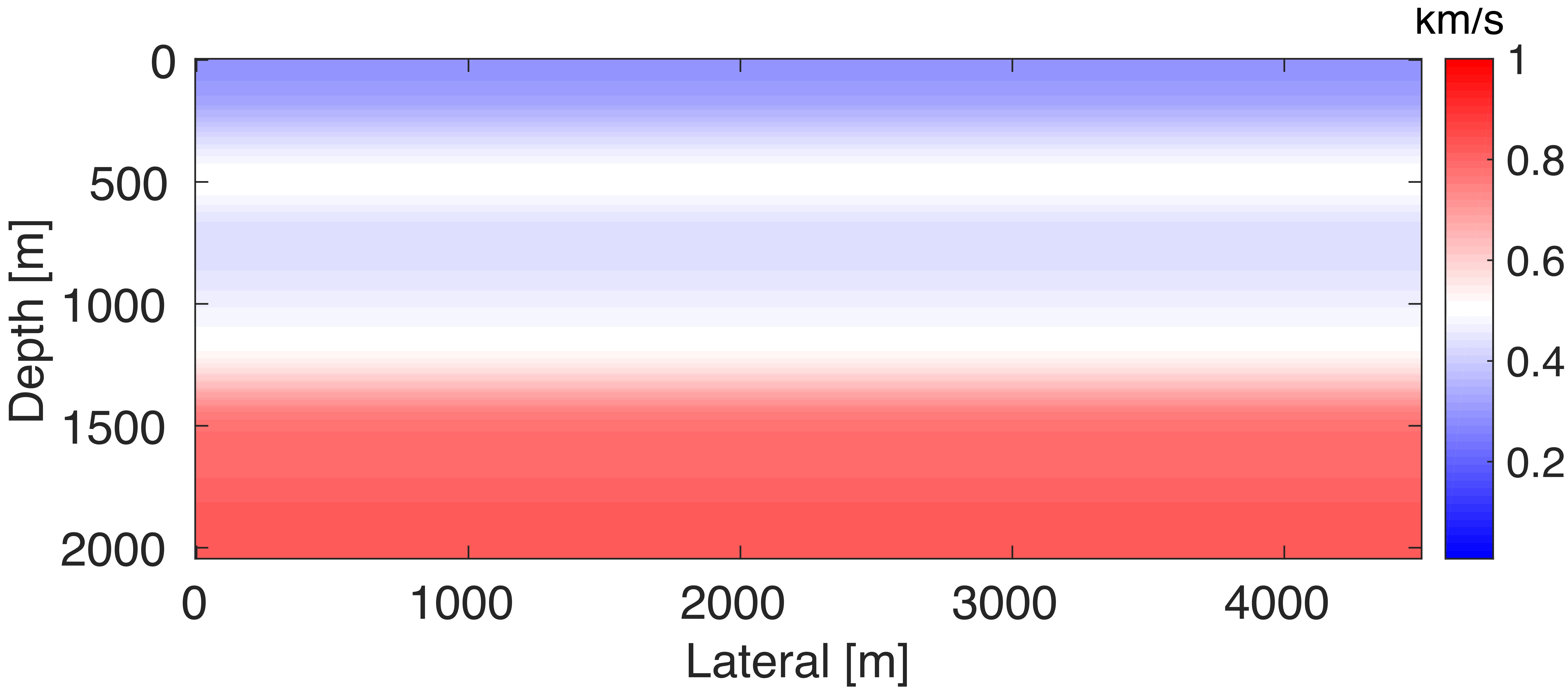


km/s
0
500
1000
1500
2000
Depth [m]
0
1000
2000
3000
4000
Lateral [m]
4.5
4
3.5
3
2.5
2
1.5

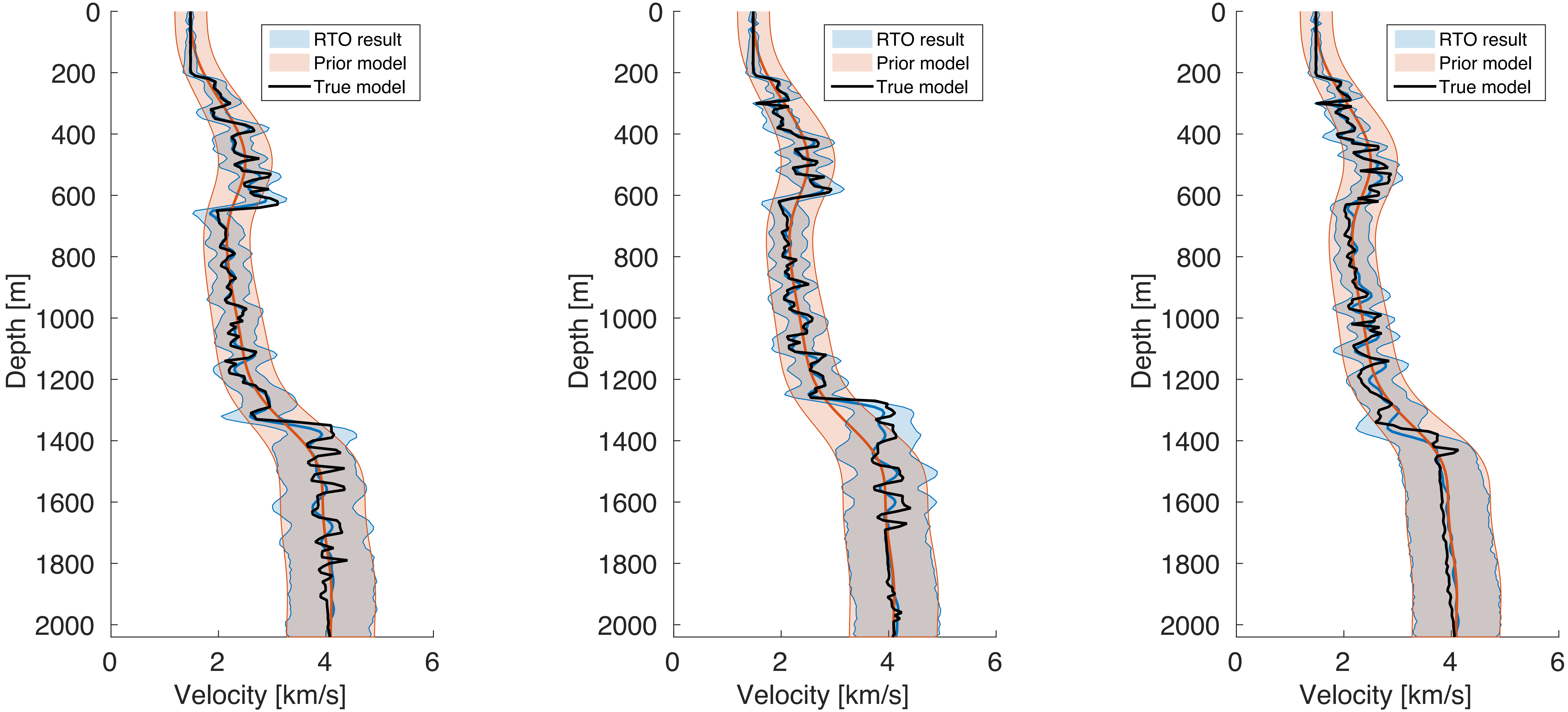
Posterior standard deviation



Prior standard deviation

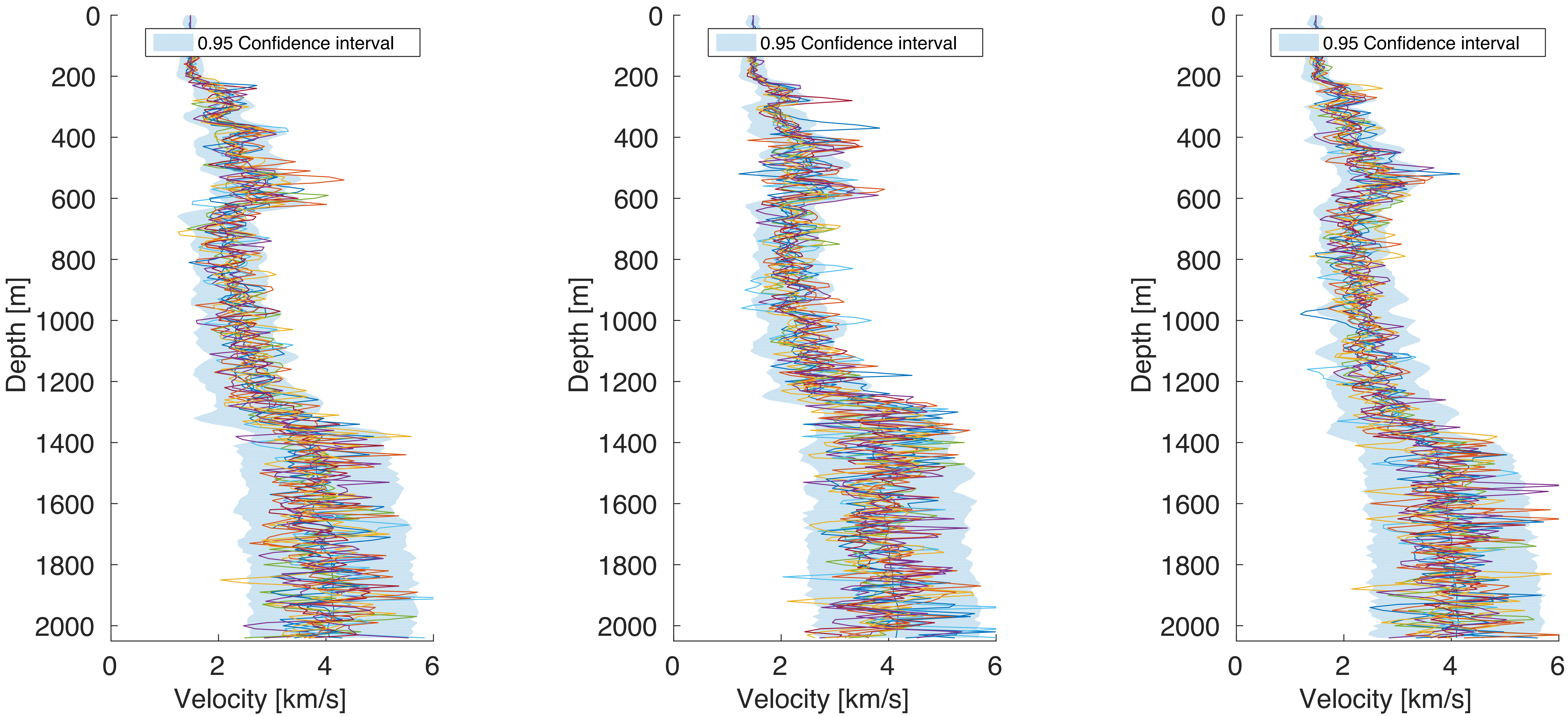


Cross section comparison
– posterior vs prior



(a) x = 1000 m (b) x = 2500 m (c) x = 4000 m

Cross section comparison
– 95% confidence interval vs 10 realizations by RML



(a) x = 1000 m **(b) x = 2500 m** **(c) x = 4000 m**

Conclusions

The proposed PDE-free GN Hessian operator can compute the matrix-vector product without additional PDE solves.

With the same PDE solves, GN method with the proposed GN Hessian operator can converge faster than gradient descent and l-BFGS methods.

The proposed GN Hessian operator can also benefit the quantification of the uncertainty in the WRI inversion results.

Reference

1. Johnathan M Bardsley, Antti Solonen, Heikki Haario, and Marko Laine. Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems. *SIAM Journal on Scientific Computing*, 36(4):A1895–A1910, 2014.
2. Yan Chen and Dean S Oliver. Ensemble randomized maximum likelihood method as an iterative ensemble smoother. *Mathematical Geosciences*, 44(1):1–26, 2012.
3. Knud Skou Cordua, Thomas Mejer Hansen, and Klaus Mosegaard. Monte carlo full-waveform inversion of crosshole gpr data using multiple-point geostatistical a priori information. *Geophysics*, 77(2):H19–H31, 2012.
4. Zhilong Fang, C Lee, C Silva, F Herrmann, and Rachel Kuske. Uncertainty quantification for wavefield reconstruction inversion. In *77th EAGE Conference and Exhibition 2015*, 2015.
5. James Martin, Lucas C. Wilcox, Carsten Burstedde, and Omar Ghattas. A Stochastic Newton MCMC Method for Large-scale Statistical Inverse Problems with Application to Seismic Inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487, 2012.
6. Tristan van Leeuwen and Felix J. Herrmann. Mitigating local minima in full-waveform inversion by expanding the search space. *Geophysical Journal International*, 195:661–667, 10 2013b.
7. Tristan van Leeuwen and Felix J. Herrmann. A penalty method for PDE-constrained optimization in inverse problems. *Inverse Problems*, 32(1):015007, 12 2015.
8. Håvard Rue. Fast sampling of gaussian markov random fields. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(2):325–338, 2001.
9. Albert Tarantola and Bernard Valette. Inverse problems = quest for information. Journal of Geophysics, 50: 159–170, 1982

Acknowledgements

This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.