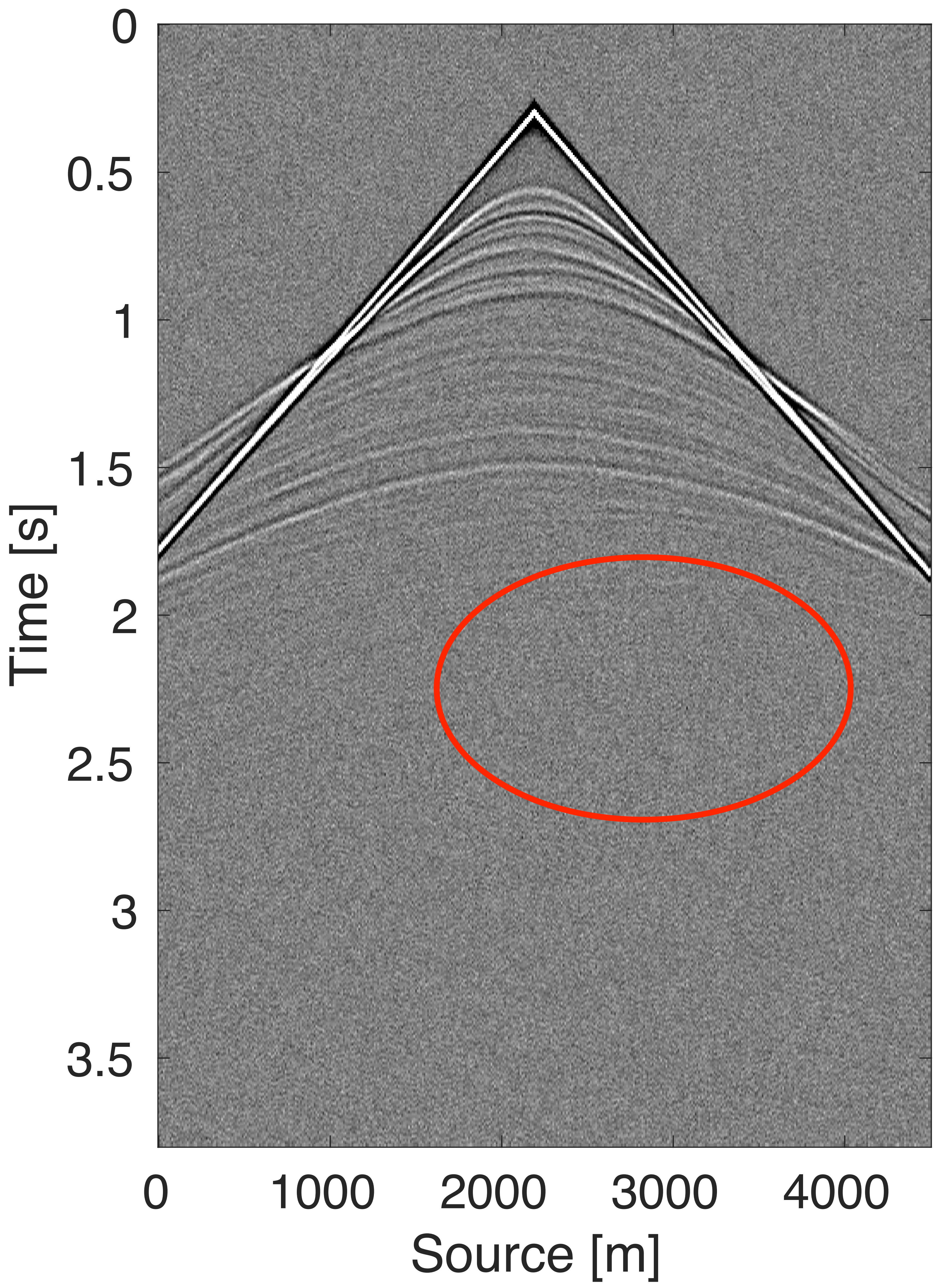
# Uncertainty quantification for Wavefield Reconstruction Inversion using a PDE free semi-definite Hessian and randomize-then-optimize method

Zhilong Fang*, Chia Ying Lee†, Curt Da Silva*, Tristan van Leeuwen** and Felix J. Herrmann*
*Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia
†Department of Mathematics, University of British Columbia
**Mathematical Institute, Utrecht University

2016/10/17

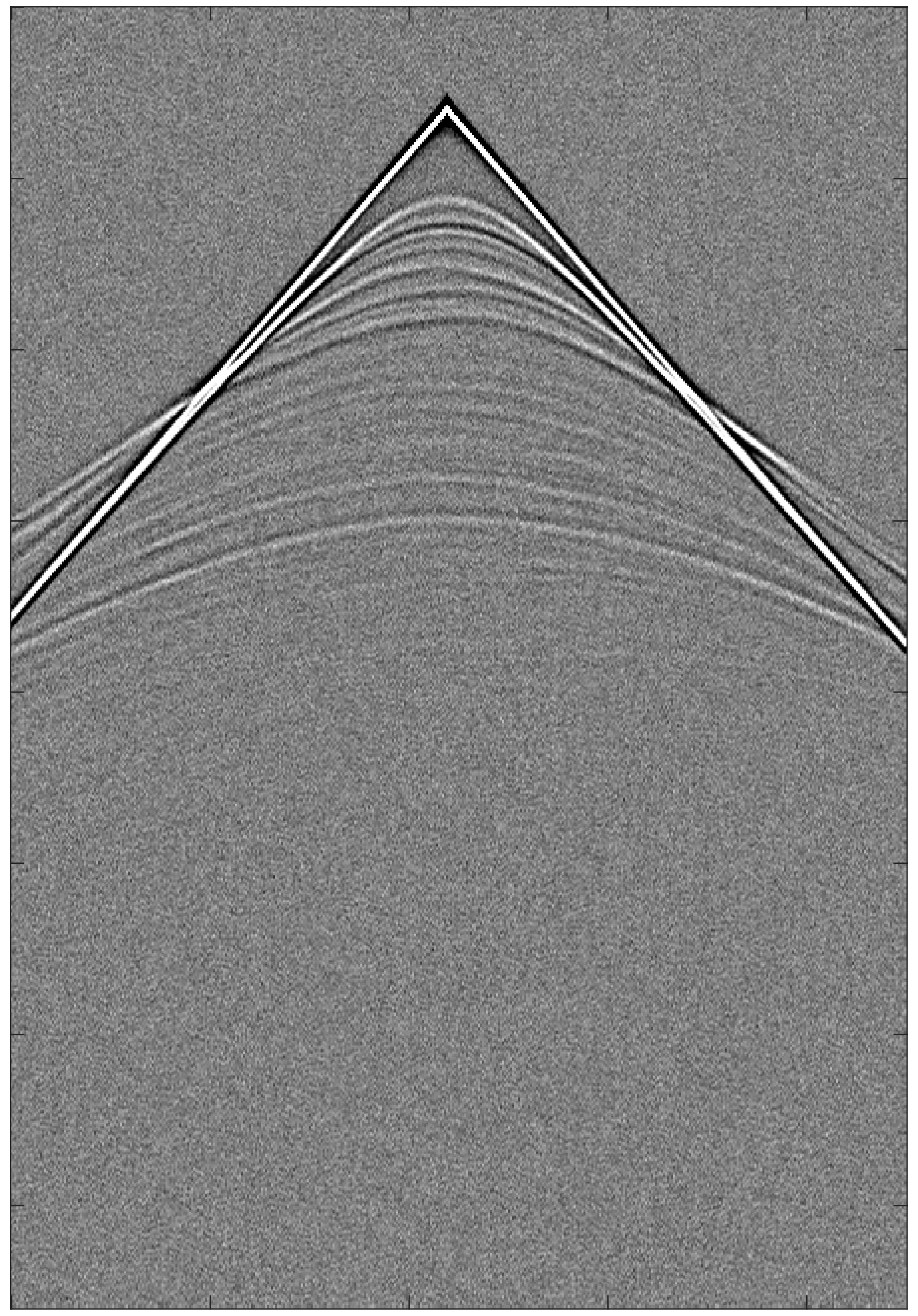
SLIM
University of British Columbia

# Motivation

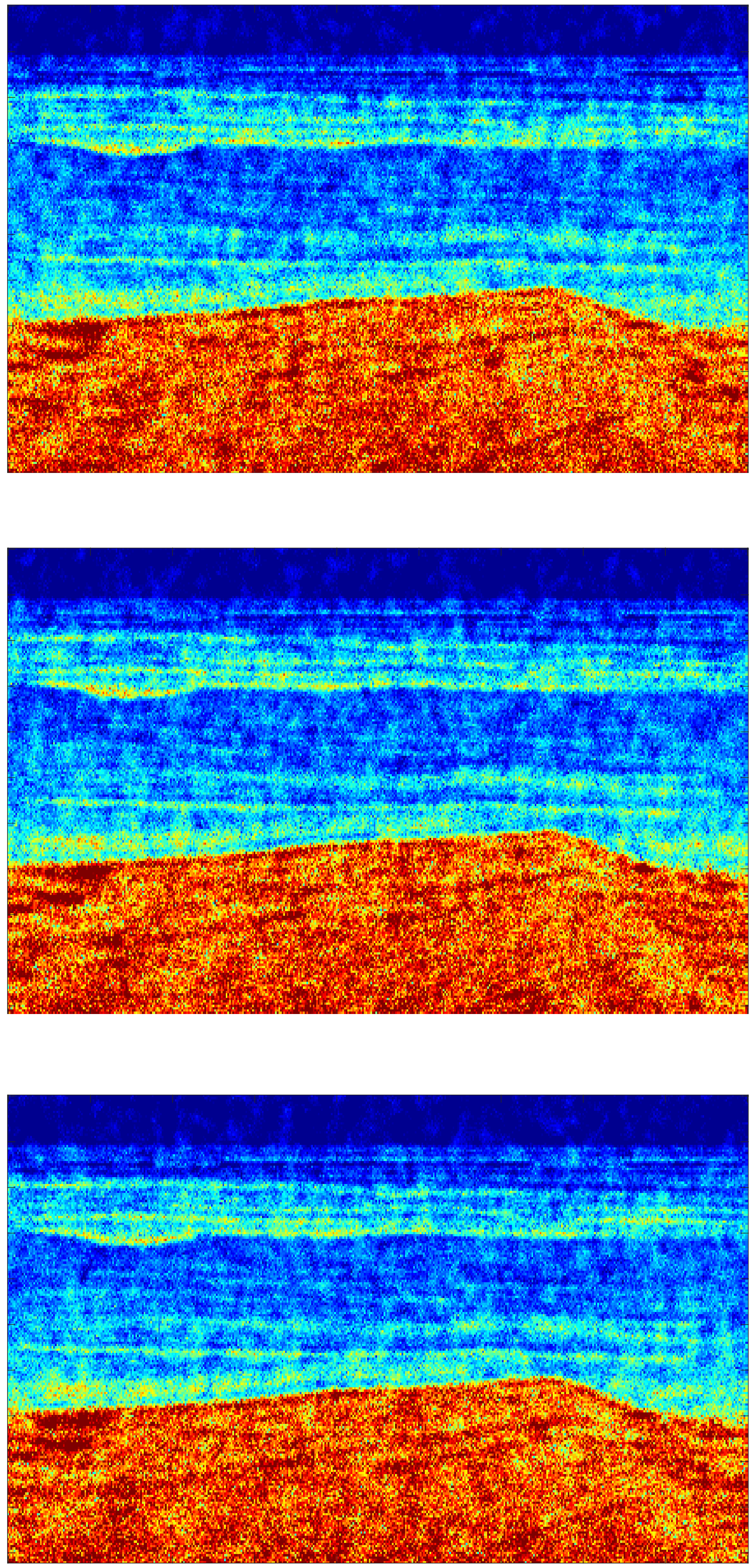



Noisy acquired data

# Motivation



Uncertainty in data

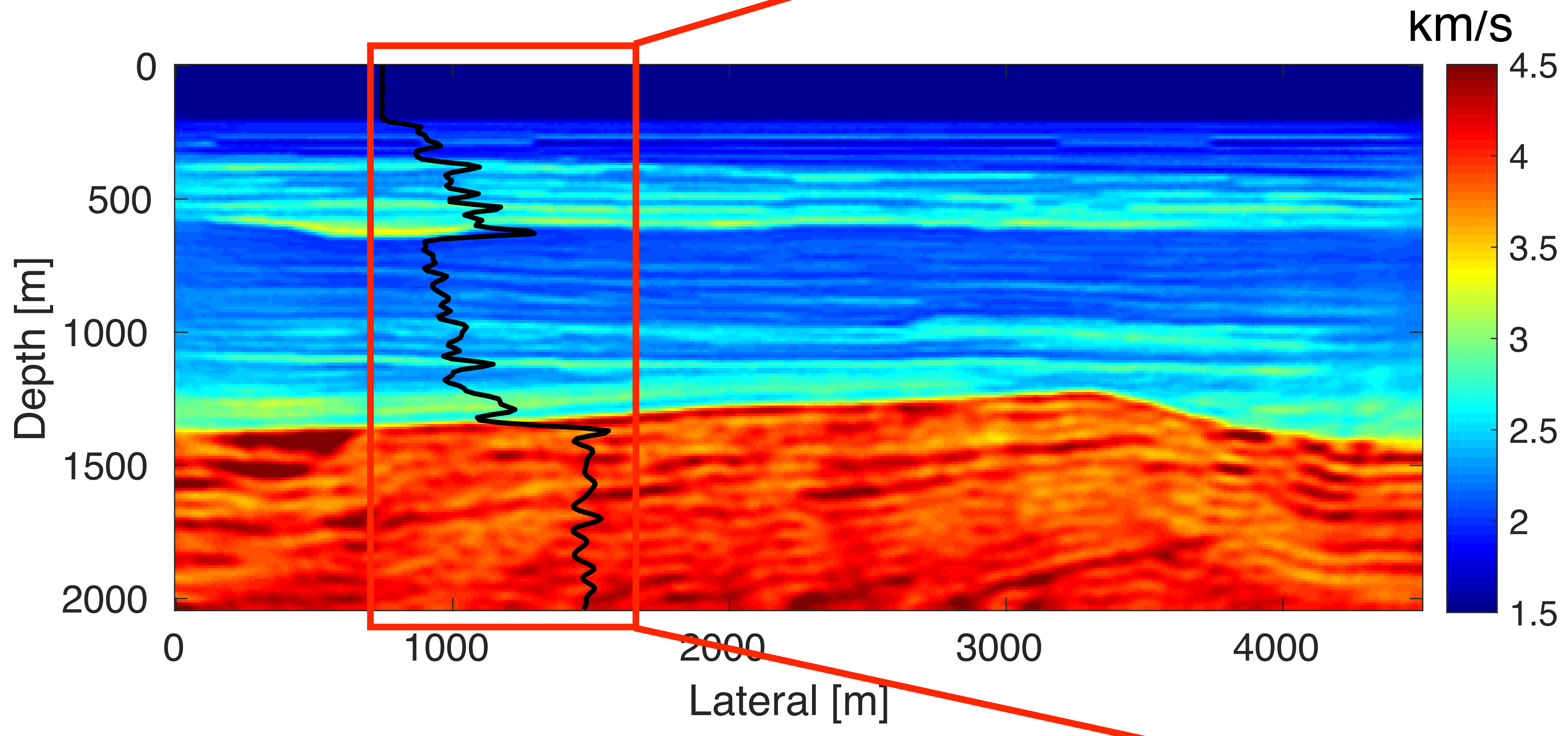


Possible velocity model

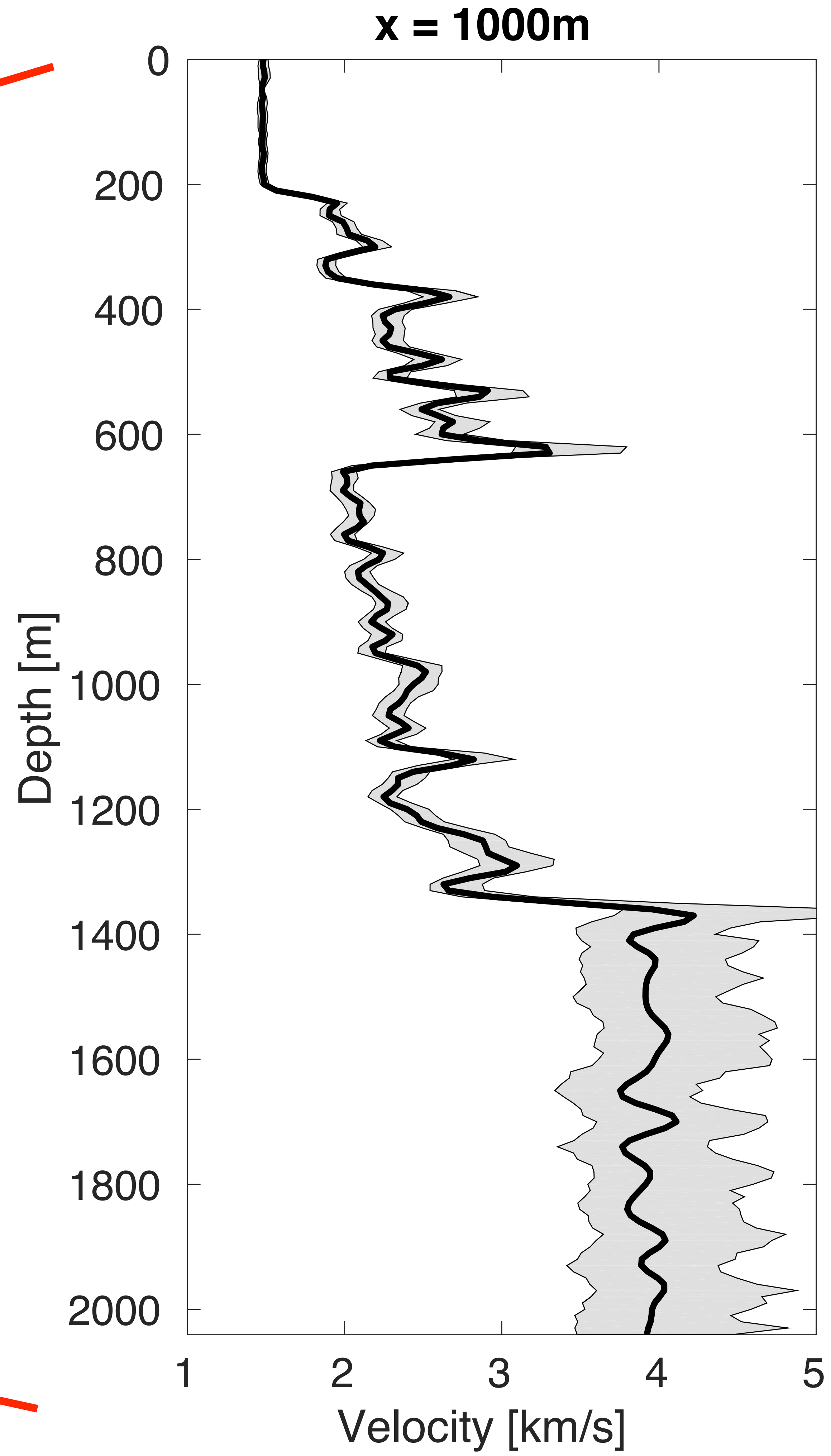
......

Risk in oil and gas volume

# Motivation




**Inversion result**




**Confidence Interval**

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

# Bayesian inference

Prior probability density function (PDF):

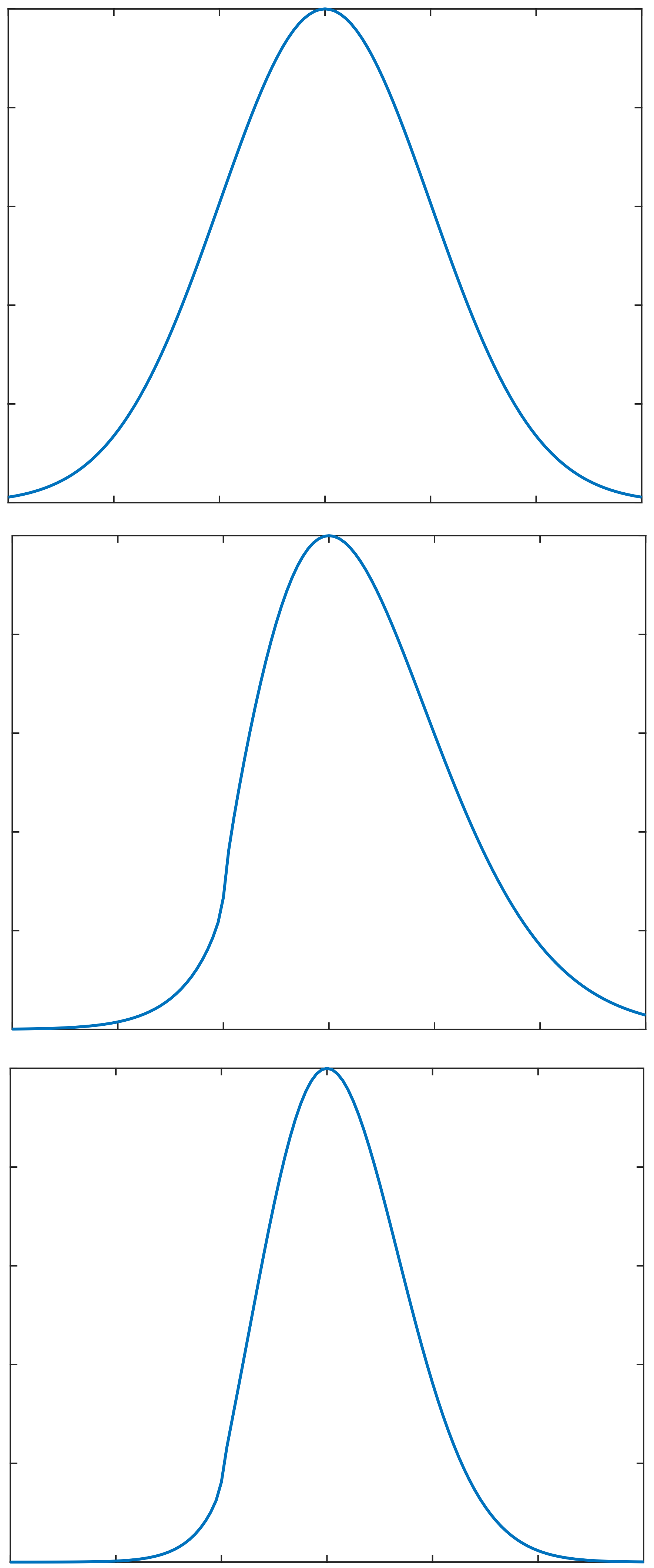
$$\mathbf{m} \longrightarrow \rho_{\text{prior}}(\mathbf{m})$$

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

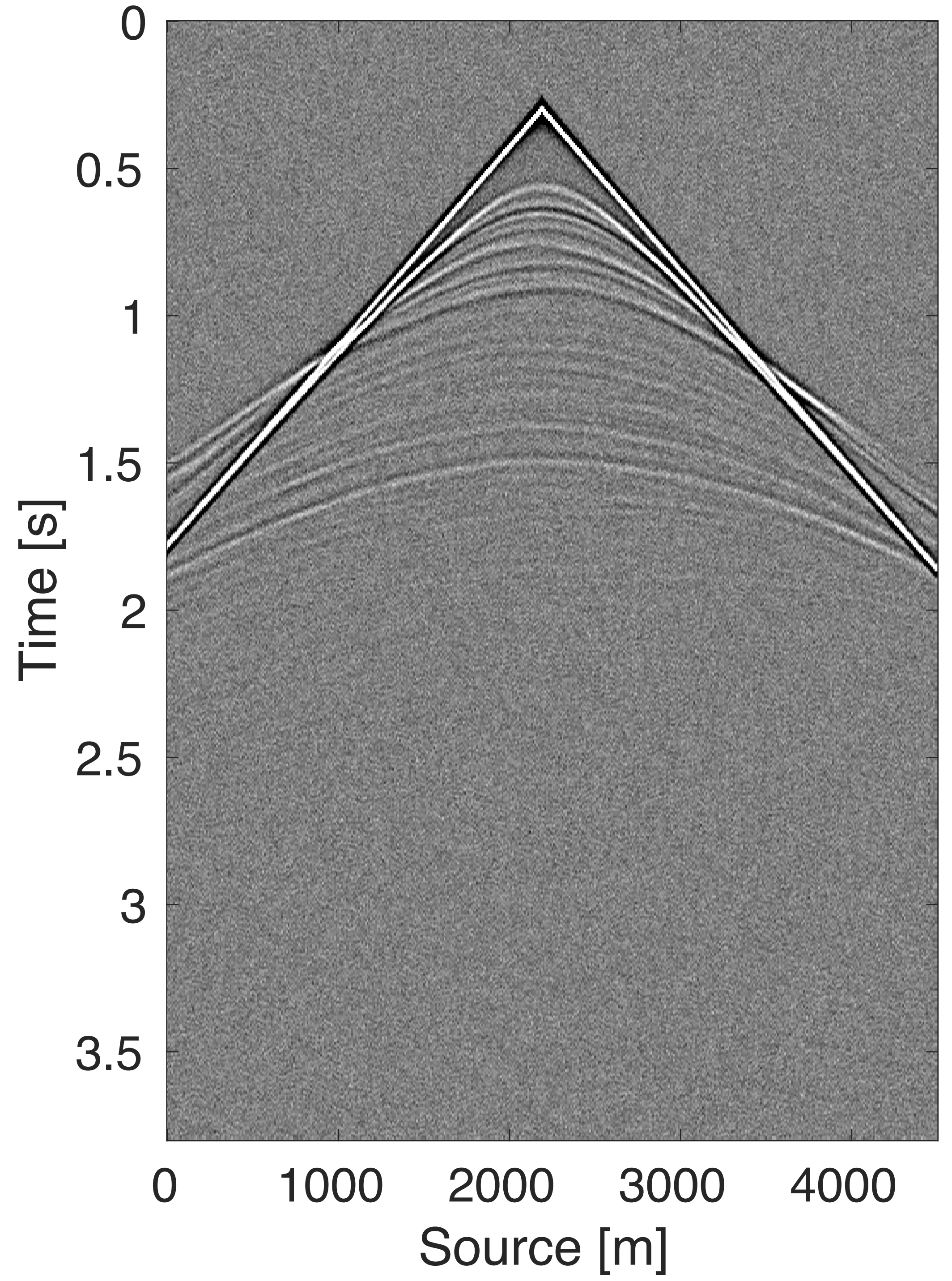
$$\mathbf{m} \longrightarrow \rho_{\text{like}}(\mathbf{d}|\mathbf{m})$$

Posterior PDF (Bayes' rule):

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

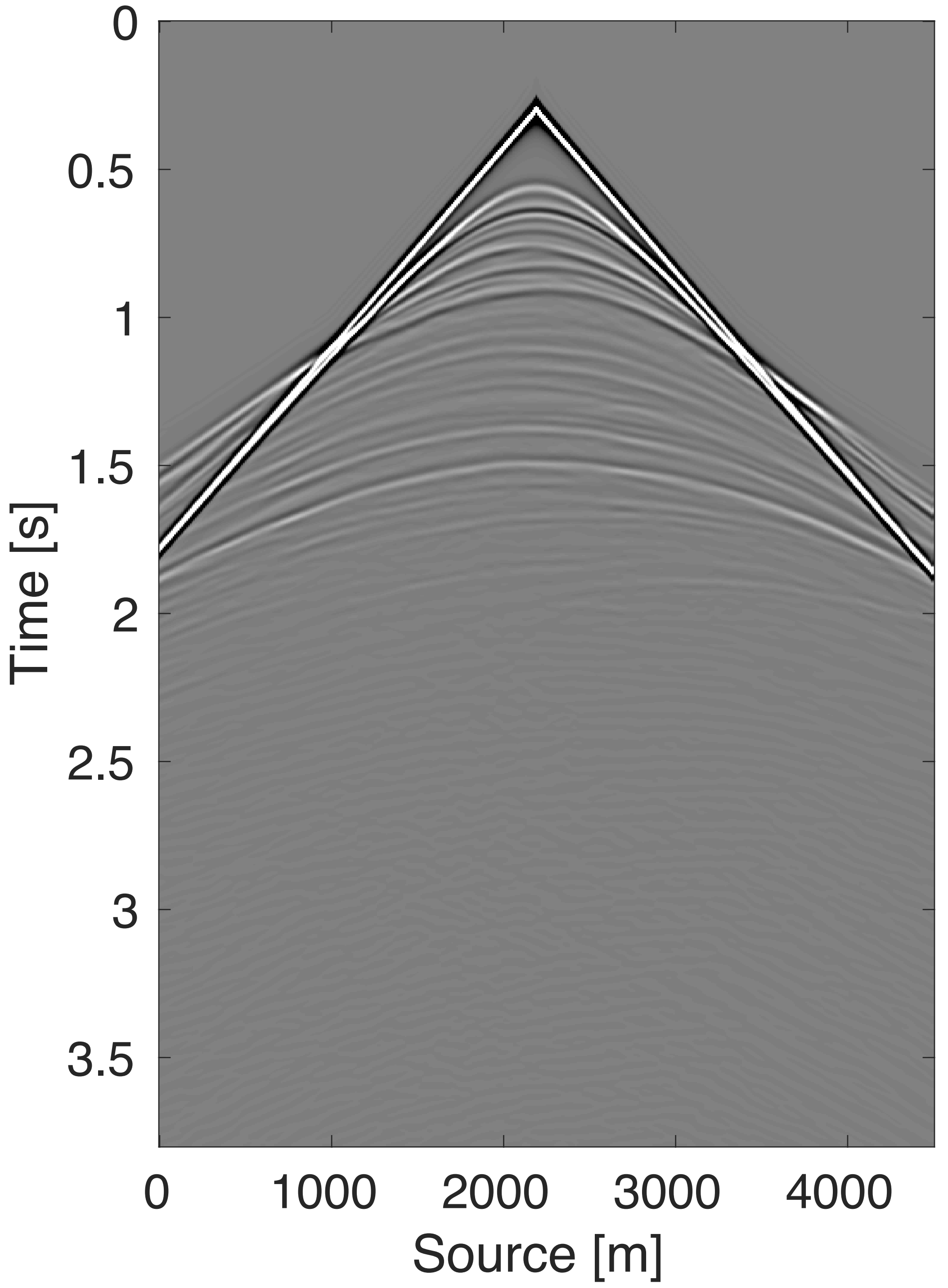


# Bayesian w/ FWI



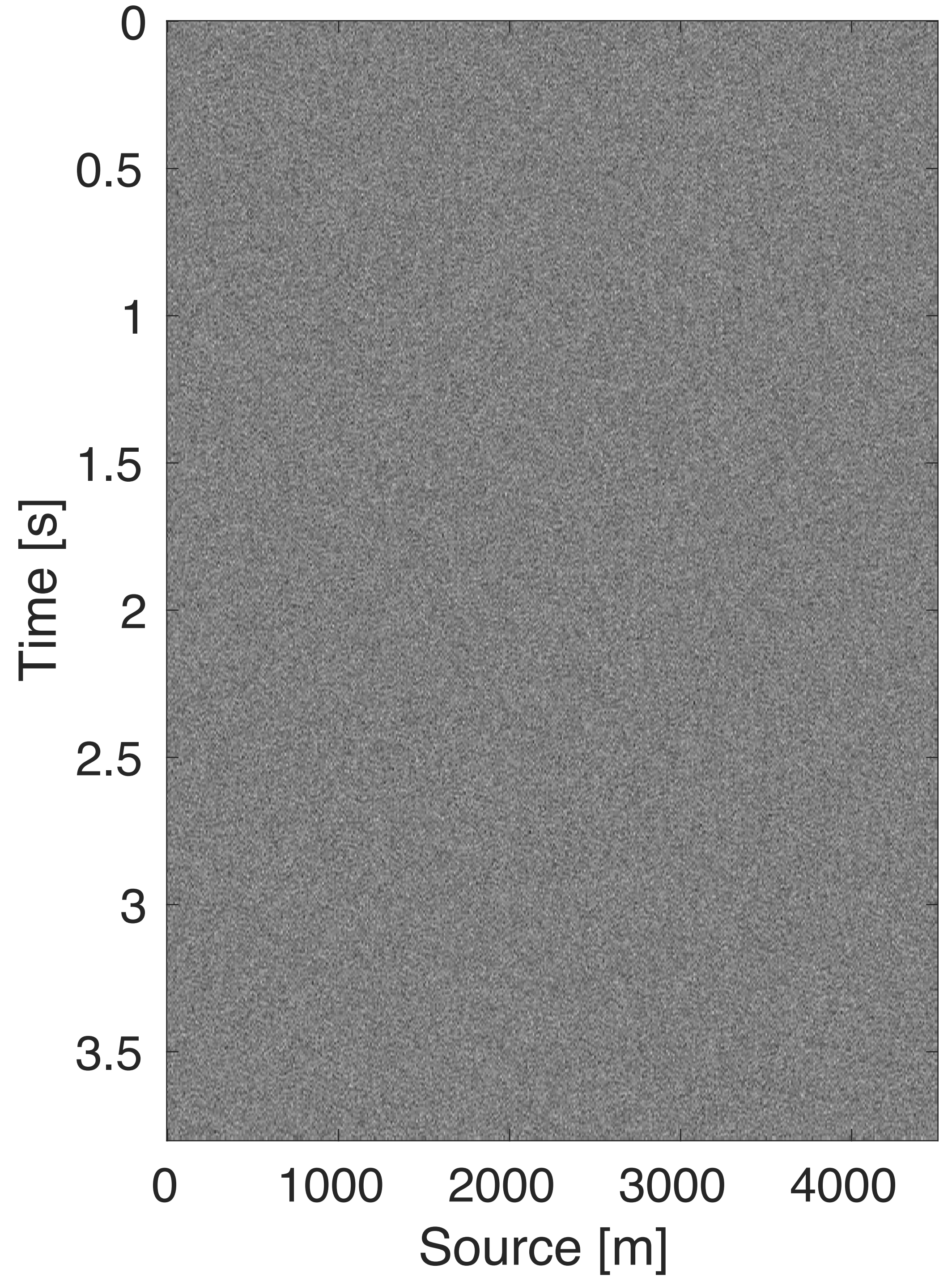

$\mathbf{d}_{\mathrm{obs}}$

$=$




$F(\mathbf{m})$

$+$




$\epsilon \sim \mathcal{N}(0, \Sigma_{\mathrm{noise}})$

# Bayesian w/ FWI

**Reduced formulation in the frequency domain:**

$$F(\mathbf{m}) = \mathbf{P}\mathbf{A}^{-1}\mathbf{q},$$

$$\mathbf{A} = \Delta + \omega^2\mathbf{m}$$

$\mathbf{m}$ : Squared-slowness

$\mathbf{q}$ : Source

$\omega$ : Frequency

$\Delta$ : Laplacian operator

$\mathbf{P}$ : Projection onto receivers

[J. Virieux and S. Operto, 2009]

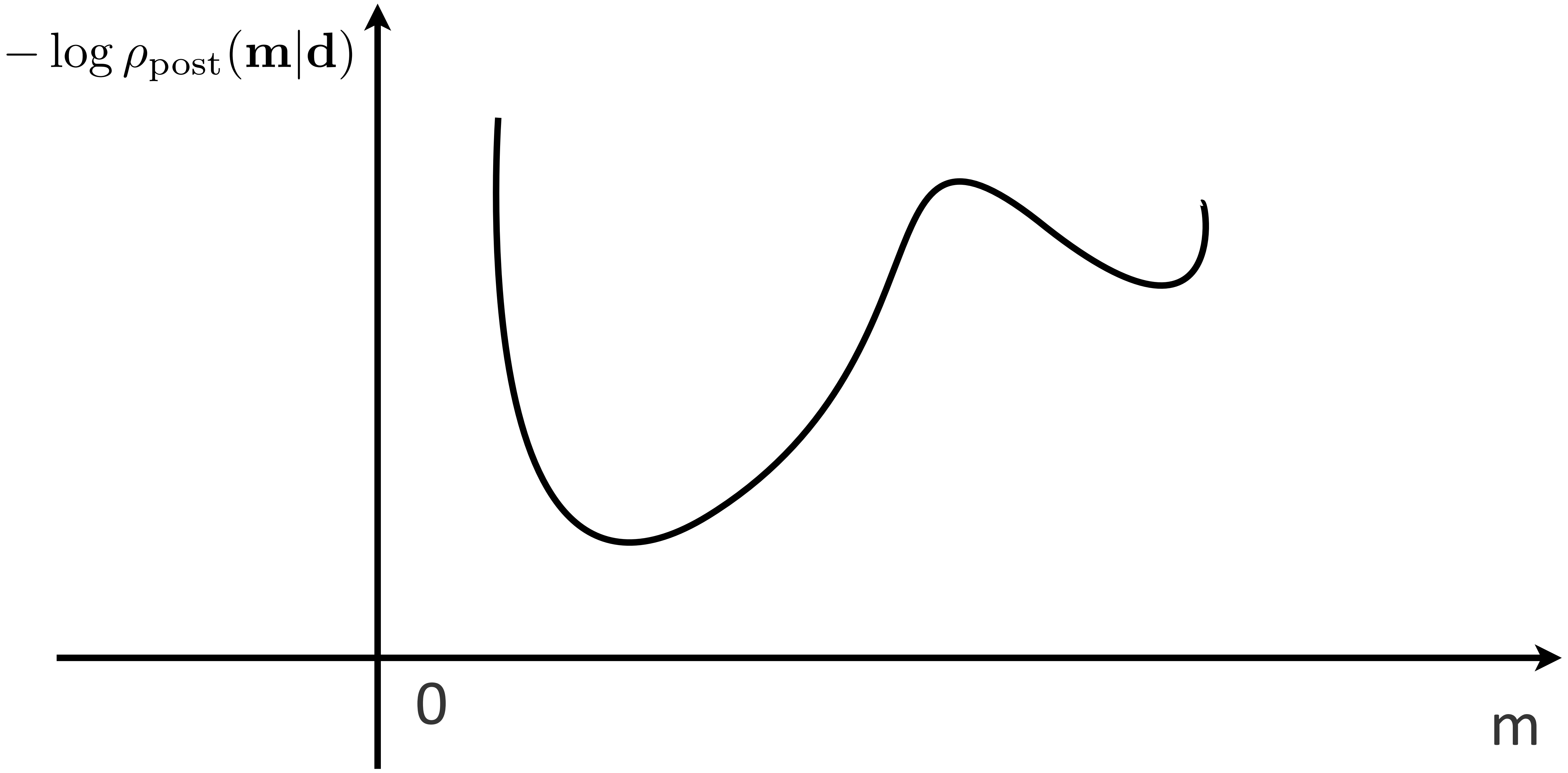
# Bayes w/ FWI

## Posterior PDF of FWI:

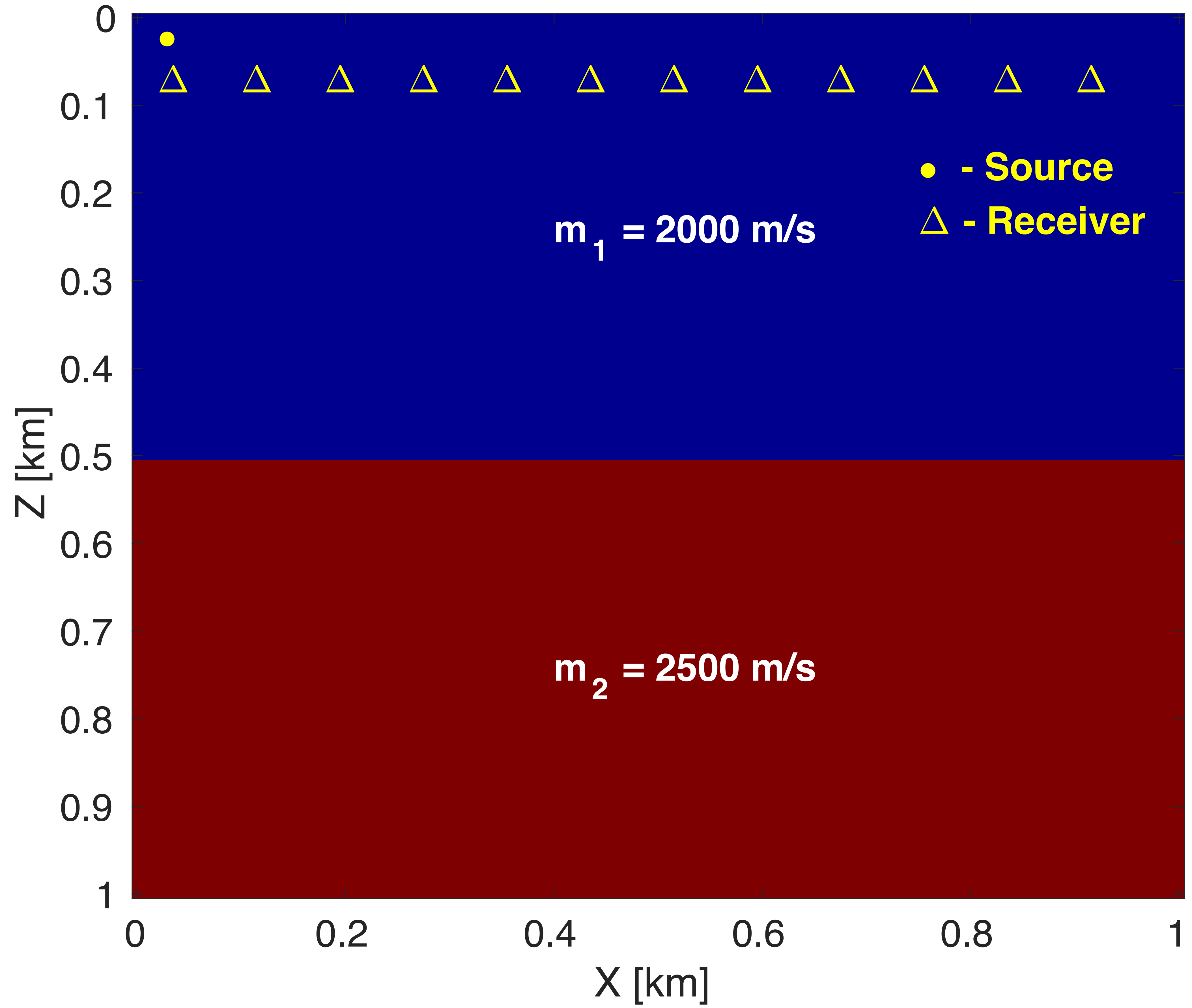
$$\rho_{\text{post}}(\mathbf{m}|\mathbf{d}) \propto \exp\left(-\frac{1}{2}\|\mathbf{P}\mathbf{A}(\mathbf{m})^{-1}\mathbf{q}-\mathbf{d}\|^2_{\mathbf{\Sigma}^{-1}_{\text{noise}}} - \frac{1}{2}\|\mathbf{m}-\mathbf{m}_{\text{prior}}\|^2_{\mathbf{\Sigma}^{-1}_{\text{prior}}}\right)$$

Strong nonlinearity

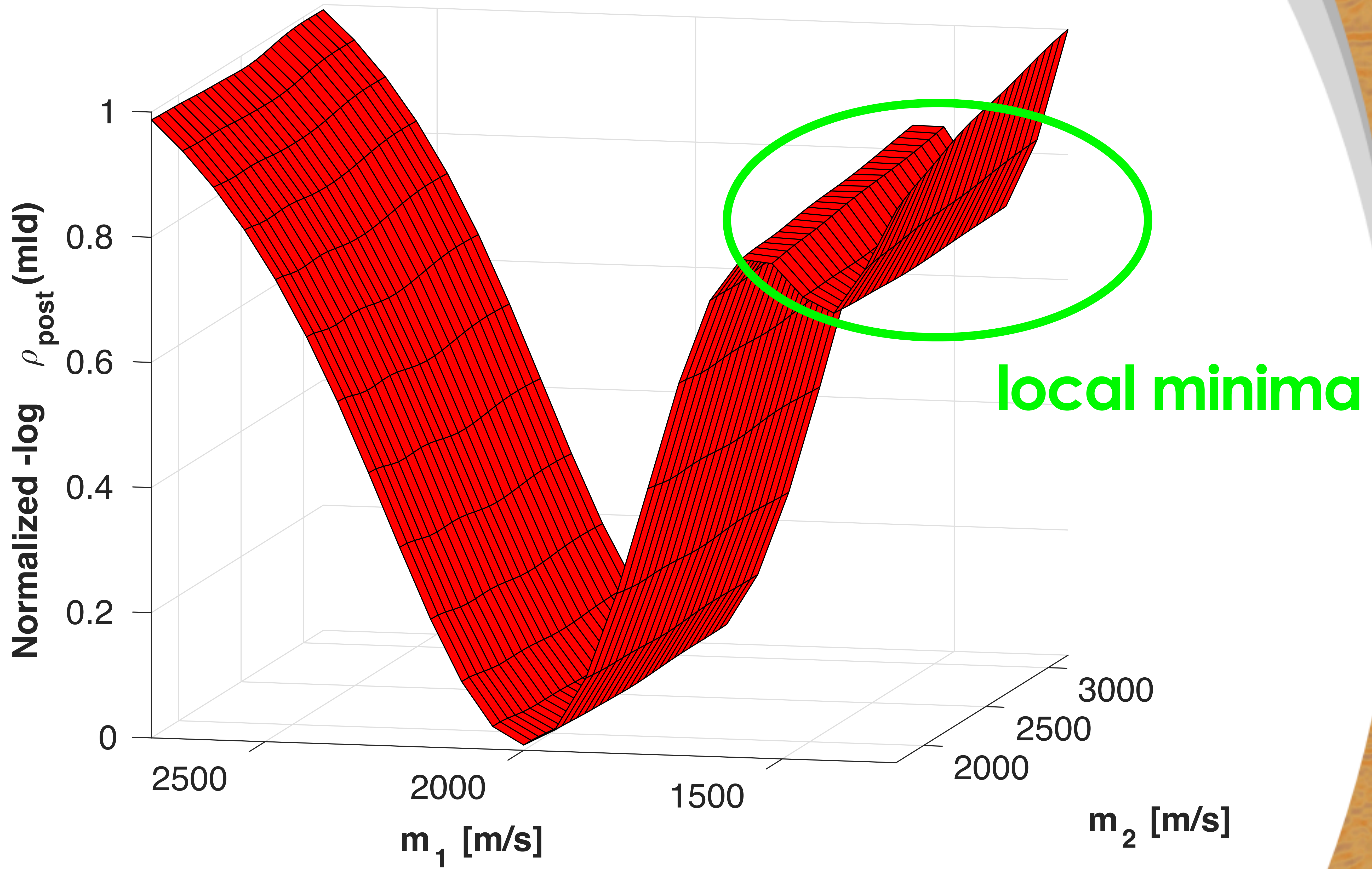
Many local minima



# Two layer example – FWI




**Model**




**FWI** $-\log \rho_{\text{post}}(\mathbf{m}|\mathbf{d})$

# Bayesian w/ FWI

**Local minima:**

- slow down McMC convergence
- render the posterior PDF “less Gaussian“

[T. van Leeuwen and F. J. Herrmann , 2013]
[T. van Leeuwen and F. J. Herrmann , 2015]

# Wavefield Reconstruction Inversion – WRI

## Penalty formulation:

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

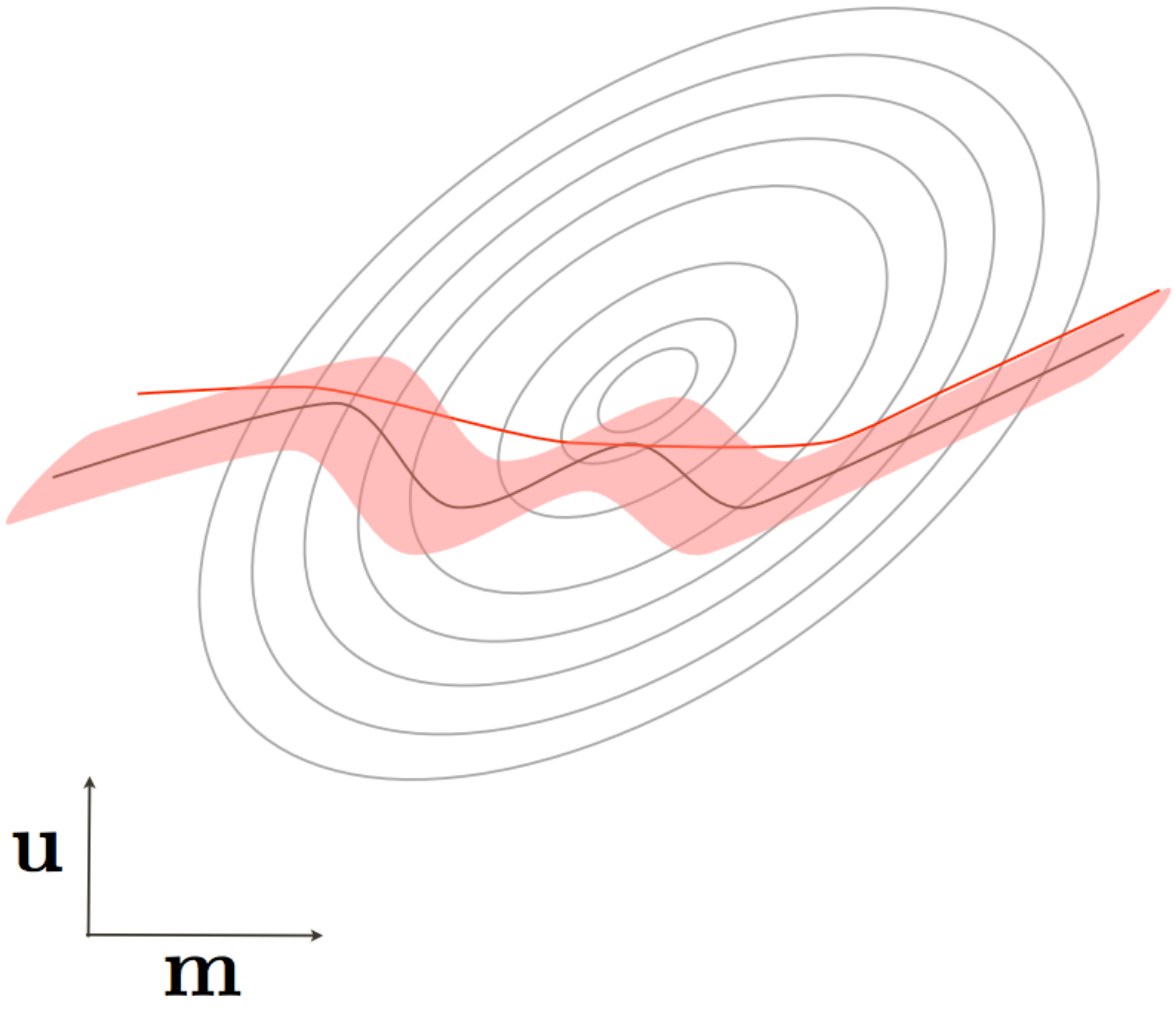
## Properties:

- bi-linear w/ respect to $\mathbf{u}$ and $\mathbf{m}$
- larger searching space

# WRI vs FWI

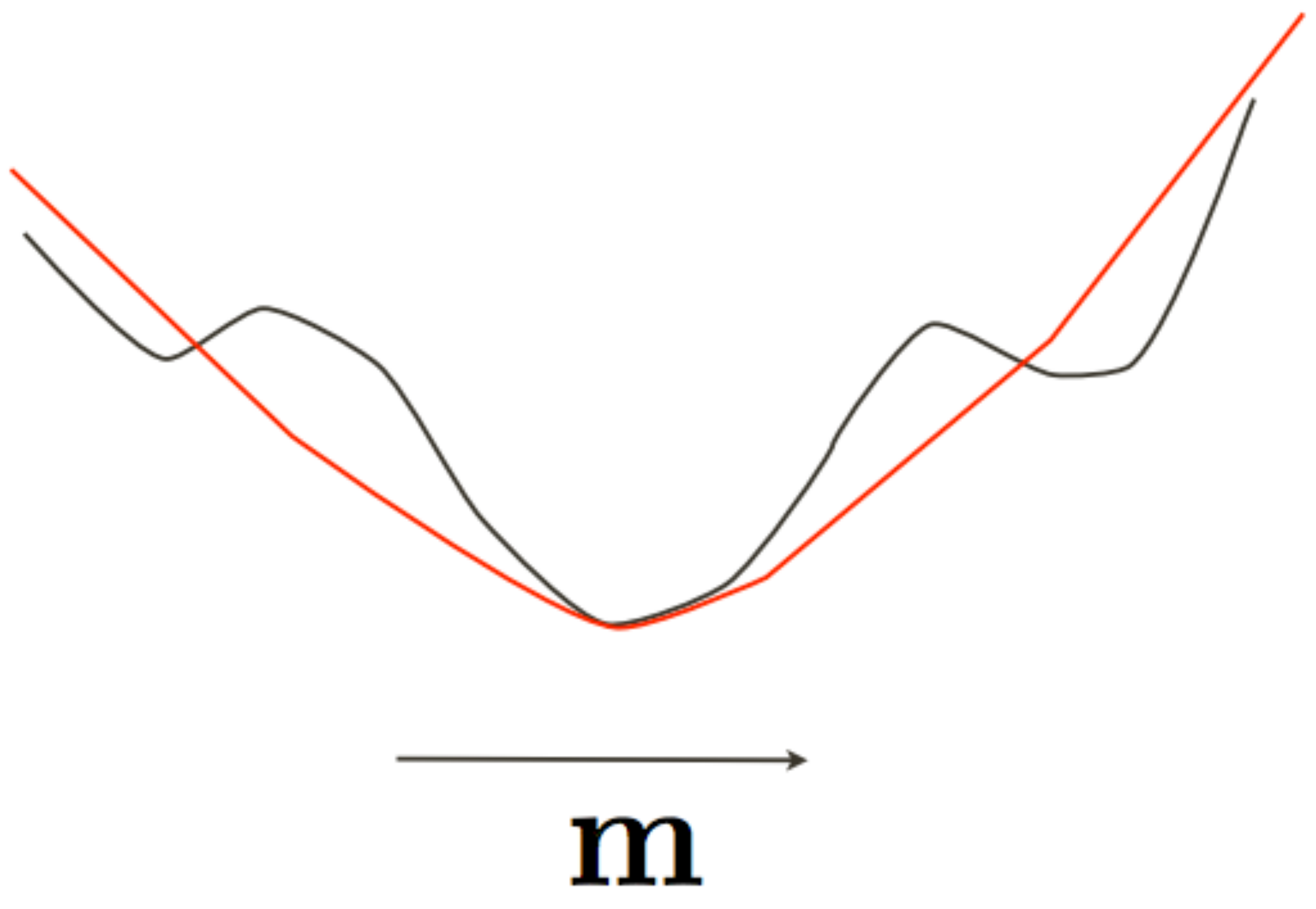
[van Leeuwen, T and Herrmann, F J , 2013]

Larger # of degrees of freedom




“more convex”

less local minima



[T. van Leeuwen and F. J. Herrmann , 2013]
[T. van Leeuwen and F. J. Herrmann , 2015]

SLIM

# Solving WRI

**Variable projection:**

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

where

$$\overline{\mathbf{u}}(\mathbf{m}) = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{q}\|_2^2$$

[Z. Fang *et al*, 2015]

# Bayes w/ WRI

**Posterior PDF of WRI:**

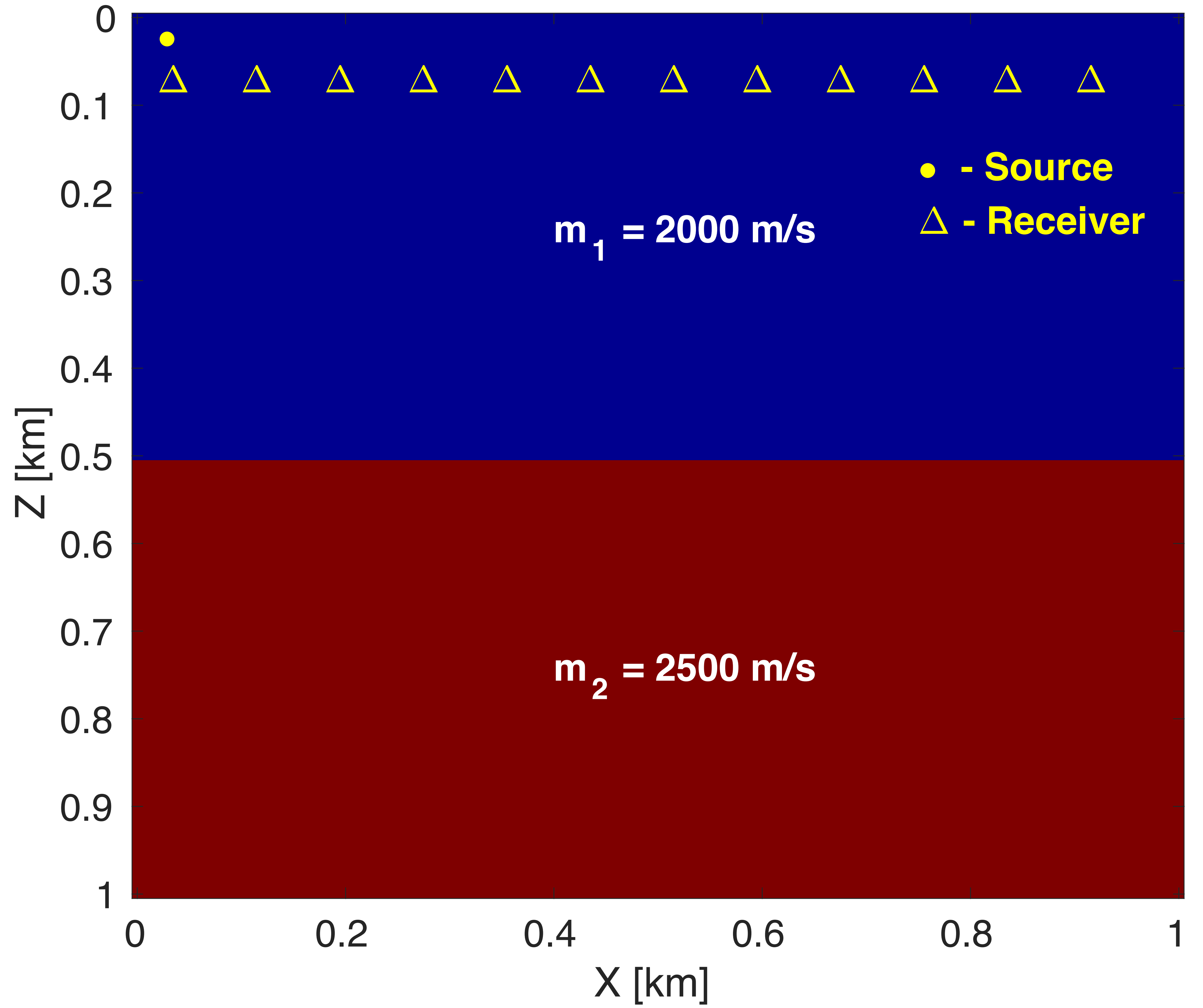
$$\rho_{\text{post}}(\mathbf{m}|\mathbf{d}) \propto$$

$$\exp\left(\underbrace{-\frac{1}{2}\|\mathbf{P}\bar{\mathbf{u}}(\mathbf{m})-\mathbf{d}\|^2_{\mathbf{\Sigma}^{-1}_{\text{noise}}} - \frac{\lambda^2}{2}\|\mathbf{A}(\mathbf{m})\bar{\mathbf{u}}(\mathbf{m})-\mathbf{q}\|^2}_{\text{Likelihood}} \underbrace{- \frac{1}{2}\|\mathbf{m}-\mathbf{m}_p\|^2_{\mathbf{\Sigma}^{-1}_{\text{prior}}}}_{\text{Prior}}\right),$$

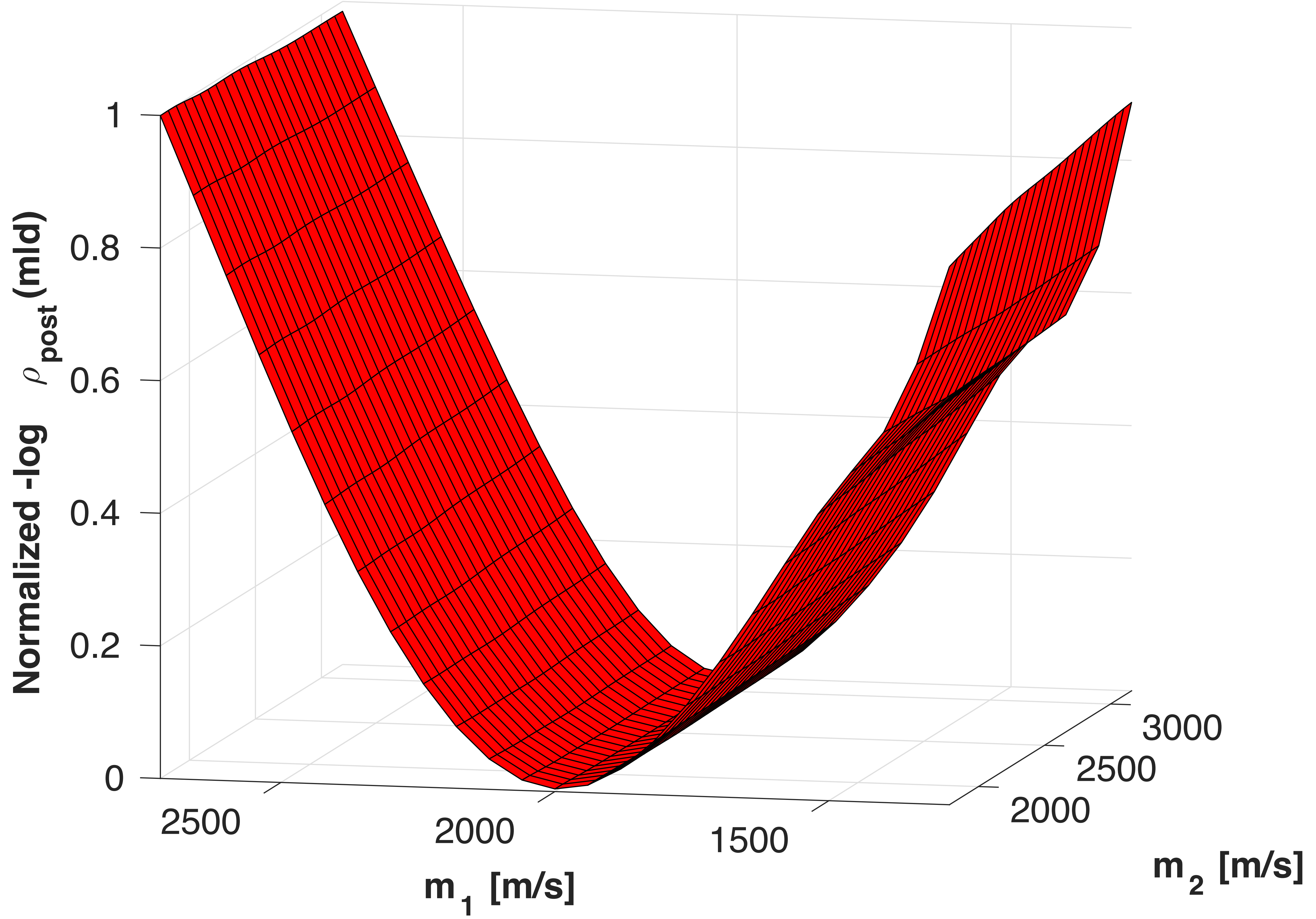
where

$$\bar{\mathbf{u}}(\mathbf{m}) = \arg\min_{\mathbf{m}} \frac{1}{2}\|\mathbf{P}\mathbf{u}-\mathbf{d}\|^2_{\mathbf{\Sigma}^{-1}_{\text{noise}}} + \frac{\lambda^2}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u}-\mathbf{q}\|^2.$$

# Two layer example – WRI




**Model**




WRI $-\log \rho_{\text{post}}(\mathbf{m}|\mathbf{d})$

## Challenges for large-scale UQ

Evaluations of the posterior PDF

- many PDE solves to evaluate PDF
- expensive PDE solves

High-dimensional space to explore

- numerical integration to expensive
- McMC based methods are impractical
  - too many iterations
  - converge too slow

# UQ for large-scale problems

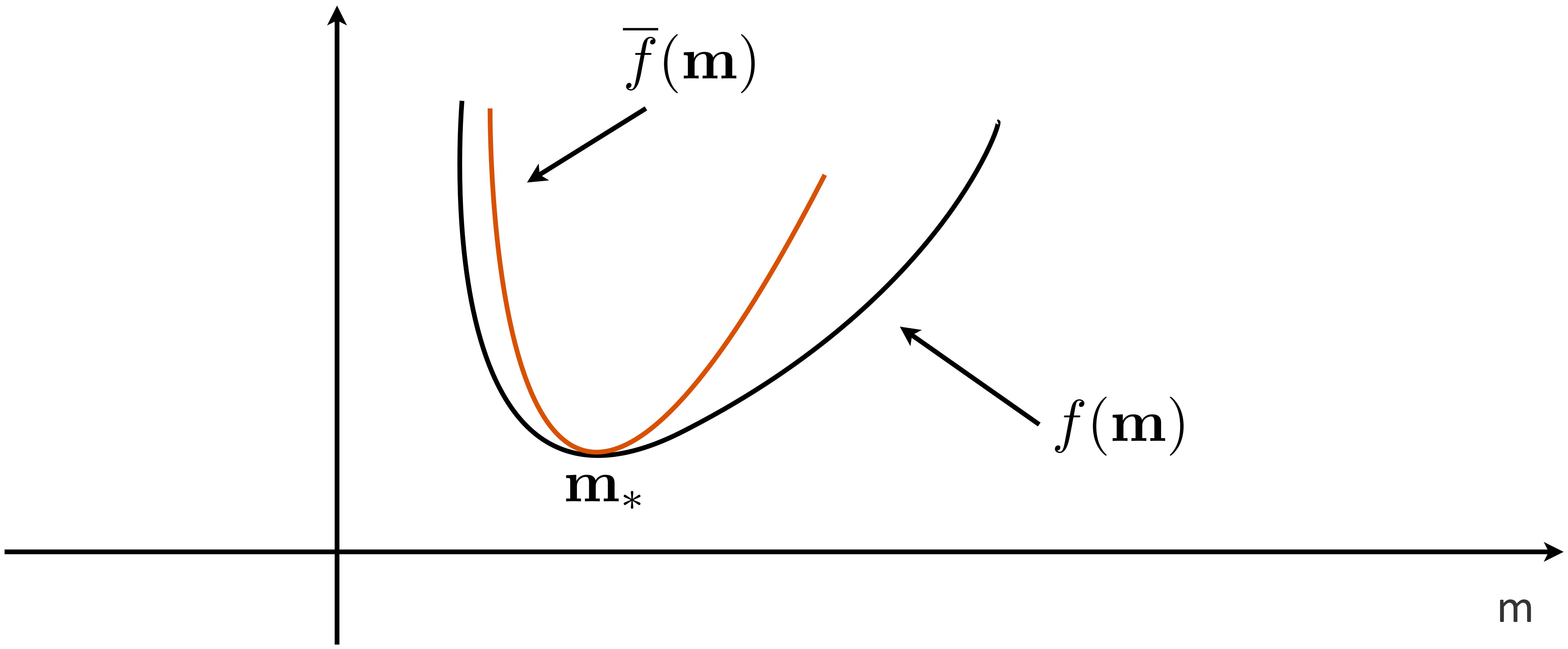
Approximate posterior PDF by Gaussians

Sample the Gaussians w/ randomize then optimize (RTO) method

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

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

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



# Approximate posterior PDF

**Gaussian approximation:**

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

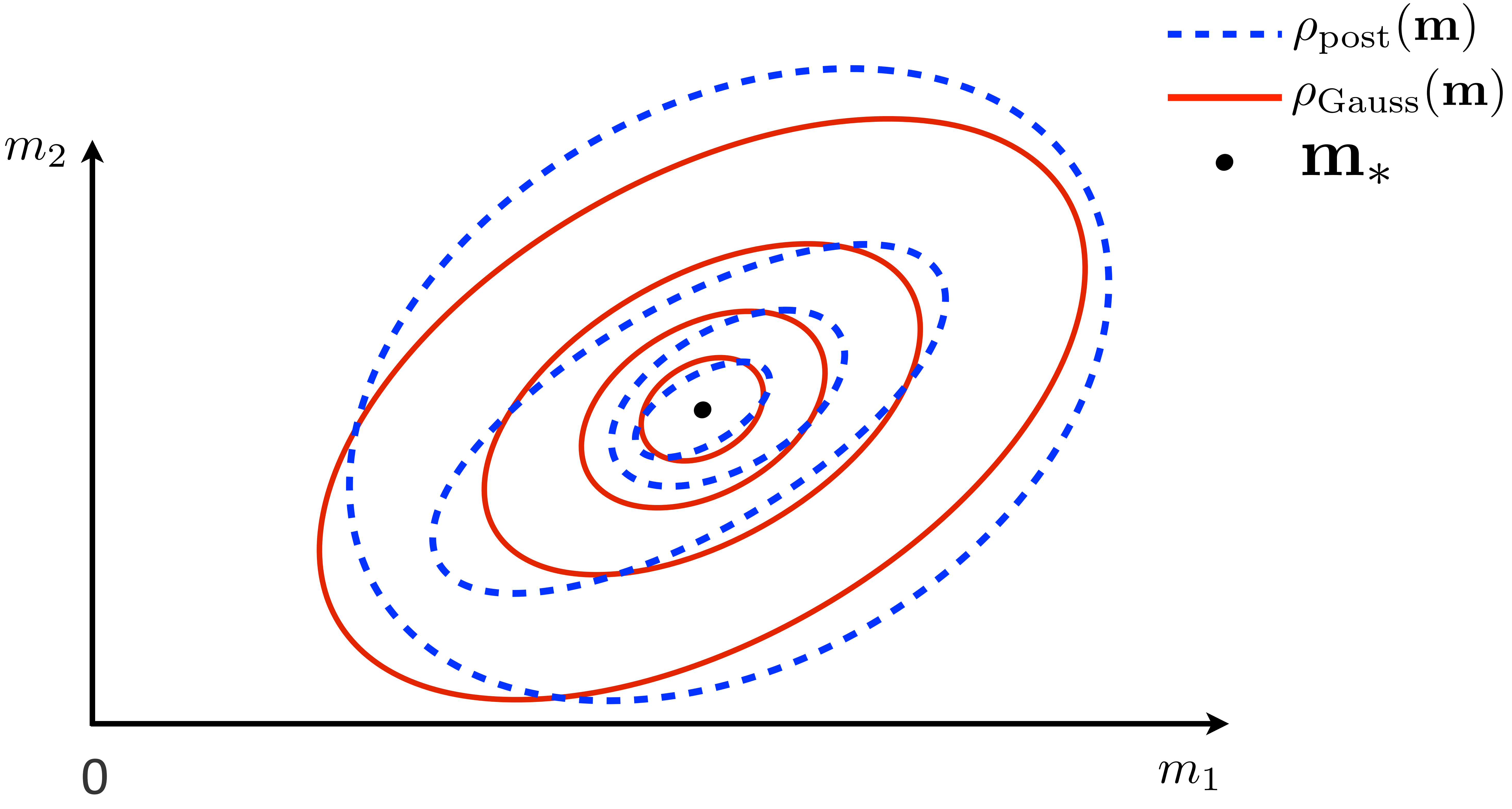
where

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

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

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

# Approximate posterior PDF



SLIM

# Form Hessian

## Gauss-Newton Hessian:

$$\mathbf{H}_{\mathrm{l}} = \mathbf{G}^{\top}\mathbf{A}^{-\top}\mathbf{P}^{\top}(\mathbf{I} + \frac{1}{\lambda^2\sigma^2}\mathbf{P}\mathbf{A}^{-1}\mathbf{A}^{-\top}\mathbf{P}^{\top})^{-1}\mathbf{P}\mathbf{A}^{-1}\mathbf{G},$$

where $\Sigma_{\mathrm{noise}} = \sigma^2\mathbf{I}$ and $\mathbf{G} = \dfrac{\partial\mathbf{A}\bar{\mathbf{u}}}{\partial\mathbf{m}}$ .

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

# Form Hessian

## Gauss-Newton Hessian:

$$\mathbf{H}_{\mathrm{l}} = \underbrace{\boxed{\mathbf{G}^\top}\boxed{\mathbf{A}^{-\top}\mathbf{P}^\top}}_{\mathbf{W}^\top}\underbrace{\left(\mathbf{I} + \frac{1}{\lambda^2\sigma^2}\mathbf{P}\mathbf{A}^{-1}\mathbf{A}^{-\top}\mathbf{P}^\top\right)^{-1}}_{\mathbf{S}}\mathbf{P}\mathbf{A}^{-1}\mathbf{G},$$

where $\Sigma_{\mathrm{noise}} = \sigma^2\mathbf{I}$ and $\mathbf{G} = \dfrac{\partial \mathbf{A}\bar{\mathbf{u}}}{\partial \mathbf{m}}$ .

# GN Hessian of WRI

$$\mathbf{H}_{\mathrm{l}} = \mathbf{W}^\top \mathbf{S} \mathbf{W}$$

**Computational cost:**

$n_{\mathrm{freq}} \times (n_{\mathrm{src}} + n_{\mathrm{rcv}})$

**storage cost:**

$n_{\mathrm{freq}} \times n_{\mathrm{grid}} \times (n_{\mathrm{src}} + n_{\mathrm{rcv}})$

**in parallel !!**

## UQ for large-scale problems

Approximate posterior PDF by Gaussians

Sample the Gaussians w/ randomize then optimize (RTO) method

[Rue, 2001]

SLIM

# Conventional method

## Sample Gaussian distribution:

$$\rho_{\text{Gauss}}(\mathbf{m}) \propto \exp(-\frac{1}{2}(\mathbf{m}-\mathbf{m}_*)^\top \mathbf{H}(\mathbf{m}-\mathbf{m}_*)).$$

## Cholesky factorization:

$$\mathbf{H} = \mathbf{L}^\top \mathbf{L},$$

$$\mathbf{m}_{\text{s}} = \mathbf{m}_* + \mathbf{L}^{-1}\mathbf{r}, \mathbf{r} \sim \mathcal{N}(0, \mathcal{I}_{n_{\text{grid}} \times n_{\text{grid}}}).$$

$\mathbf{H}$ should be an explicit matrix,

Computational cost is $\mathcal{O}(n_{\text{grid}}^3)$.

# RTO method

Re-formulate the posterior distribution with:

$$\mathbf{H} = \mathbf{H}_{\mathrm{l}} + \mathbf{H}_{\mathrm{p}}, \ \mathbf{H}_{\mathrm{l}} = \mathbf{L}_{\mathrm{l}}^{\top}\mathbf{L}_{\mathrm{l}}, \text{ and } \ \mathbf{H}_{\mathrm{p}} = \mathbf{L}_{\mathrm{p}}^{\top}\mathbf{L}_{\mathrm{p}},$$

then

$$\begin{aligned}\rho_{\text{Gauss}}(\mathbf{m}) \propto \exp(&-\frac{1}{2}(\mathbf{L}_{\mathrm{l}}\mathbf{m} - \mathbf{L}_{\mathrm{l}}\mathbf{m}_*)^{\top}(\mathbf{L}_{\mathrm{l}}\mathbf{m} - \mathbf{L}_{\mathrm{l}}\mathbf{m}_*) \\ &-\frac{1}{2}(\mathbf{L}_{\mathrm{p}}\mathbf{m} - \mathbf{L}_{\mathrm{p}}\mathbf{m}_*)^{\top}(\mathbf{L}_{\mathrm{p}}\mathbf{m} - \mathbf{L}_{\mathrm{p}}\mathbf{m}_*))\end{aligned}$$

[J. M. Bardsley *et al*, 2014]

## RTO method

Generate a sample by solving the optimization problem:

$$\mathbf{m}_{\mathrm{s}} = \arg\min_{\mathbf{m}} \left\| \mathbf{L}_{\mathrm{l}}\mathbf{m} - (\mathbf{L}_{\mathrm{l}}\mathbf{m}_* + \mathbf{r}_{\mathrm{l}}) \right\|^2 + \left\| \mathbf{L}_{\mathrm{p}}\mathbf{m} - (\mathbf{L}_{\mathrm{p}}\mathbf{m}_* + \mathbf{r}_{\mathrm{p}}) \right\|^2,$$

where

$$\mathbf{r}_{\mathrm{l}} \sim \mathcal{N}(0, \mathcal{I}_{n_{\mathrm{rcv}} \times n_{\mathrm{rcv}}}),$$

$$\mathbf{r}_{\mathrm{p}} \sim \mathcal{N}(0, \mathcal{I}_{n_{\mathrm{grid}} \times n_{\mathrm{grid}}}).$$

## RTO method

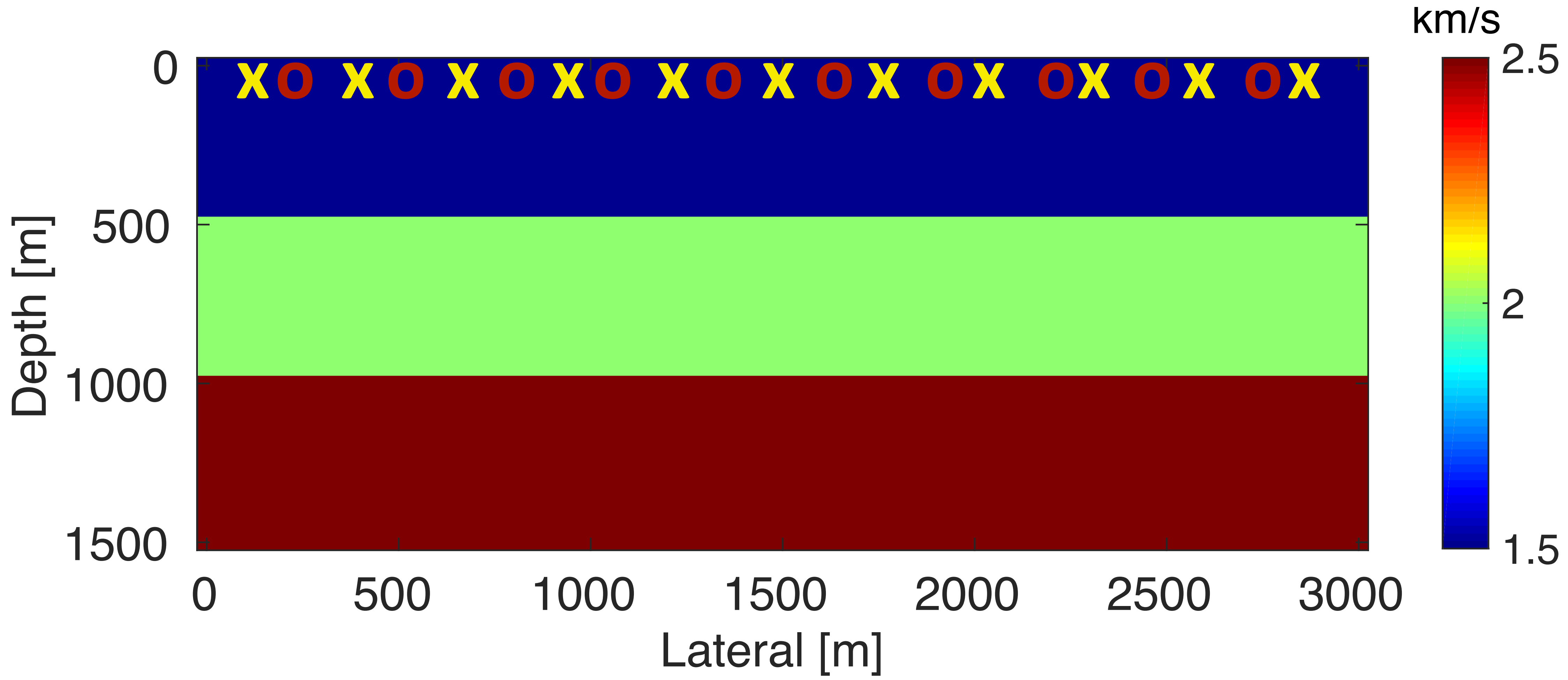
Factorization of $\mathbf{H}_{\mathrm{l}}$:

$$\mathbf{H}_{\mathrm{l}} = \mathbf{L}_{\mathrm{l}}^{\top}\mathbf{L}_{\mathrm{l}},$$

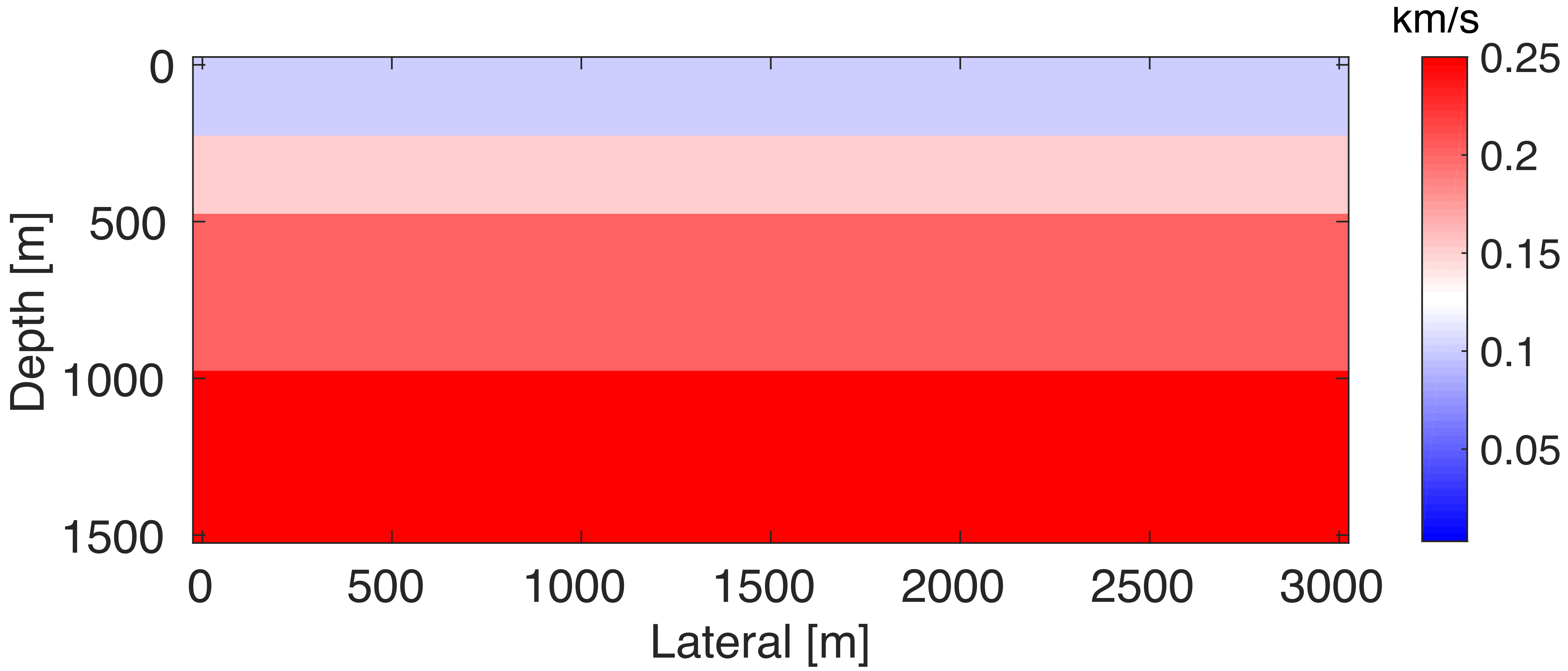
$$\mathbf{L}_{\mathrm{l}} = \underbrace{(\mathbf{I} + \frac{1}{\lambda^2\sigma^2}\mathbf{P}\mathbf{A}^{-1}\mathbf{A}^{-\top}\mathbf{P}^{\top})^{-\frac{1}{2}}}_{n_{\mathrm{rcv}} \times n_{\mathrm{rcv}}}\mathbf{P}\mathbf{A}^{-1}\mathbf{G}.$$

# Numerical Experiment – Layer model

Depth of sources and receivers: 50 m
Number of sources and receivers: 61
Frequency: 2,3 and 4 Hz
Lambda: 4e4
sigma: 10




(a) True model and prior model




(b) STD of the prior distribution

SLIM

# Randomized Maximum Likelihood - RML

Generate independent samples from $\rho_{\text{post}}(\mathbf{m})$ by solving:

$$\min_{\mathbf{m}} \frac{1}{2}(\sigma^{-2}\|\mathbf{P}\overline{\mathbf{u}}(\mathbf{m}) - \mathbf{d} - \mathbf{r}_{\text{d}}\|^2 + \lambda^2\|\mathbf{A}(\mathbf{m})\overline{\mathbf{u}}(\mathbf{m}) - \mathbf{q} - \mathbf{r}_{\text{s}}\|^2) + \frac{1}{2}\|\mathbf{m} - \mathbf{m}_{\text{p}} - \mathbf{r}_{\text{p}}\|^2_{\Sigma^{-1}_{\text{prior}}},$$

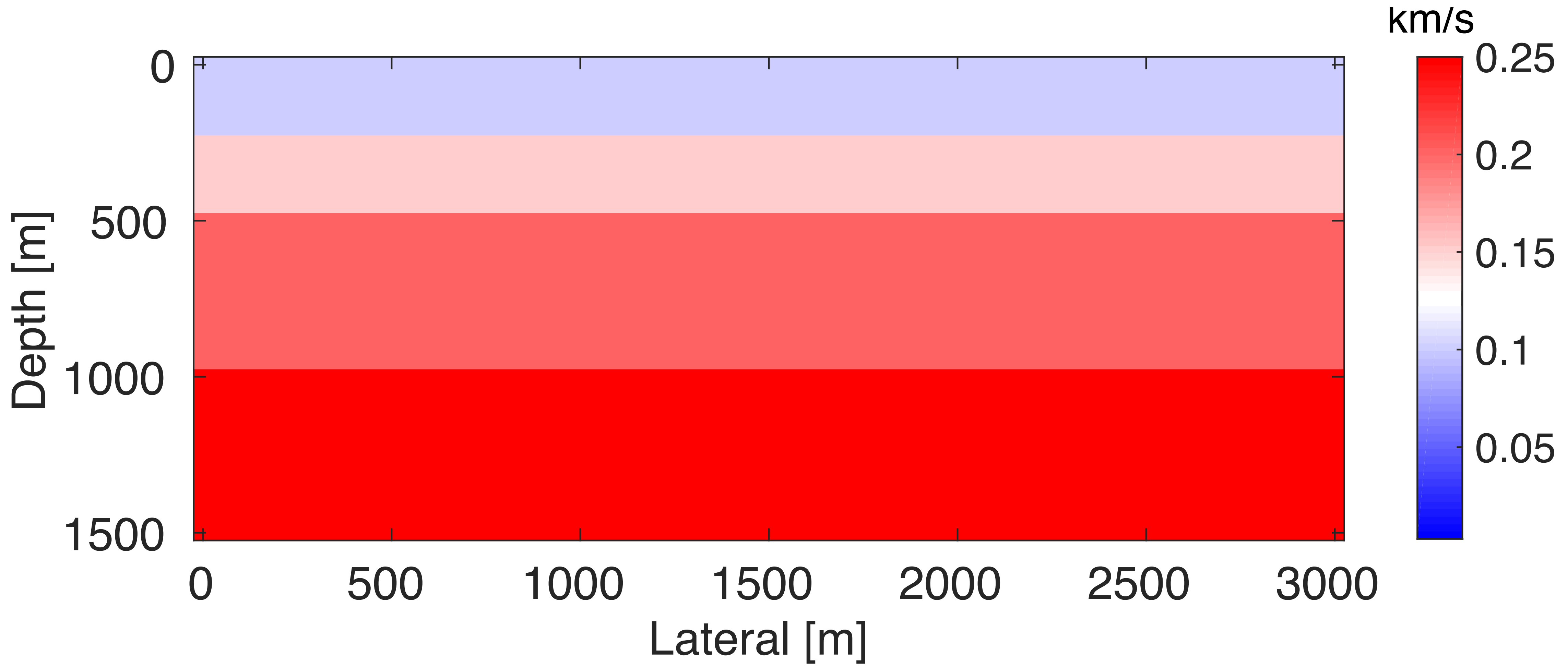
where

$$\mathbf{r}_{\text{d}} \sim \mathcal{N}(0, \sigma^2 \mathcal{I}_{n_{\text{rcv}} \times n_{\text{rcv}}}),$$

$$\mathbf{r}_{\text{s}} \sim \mathcal{N}(0, \lambda^{-2} \mathcal{I}_{n_{\text{grid}} \times n_{\text{grid}}}),$$

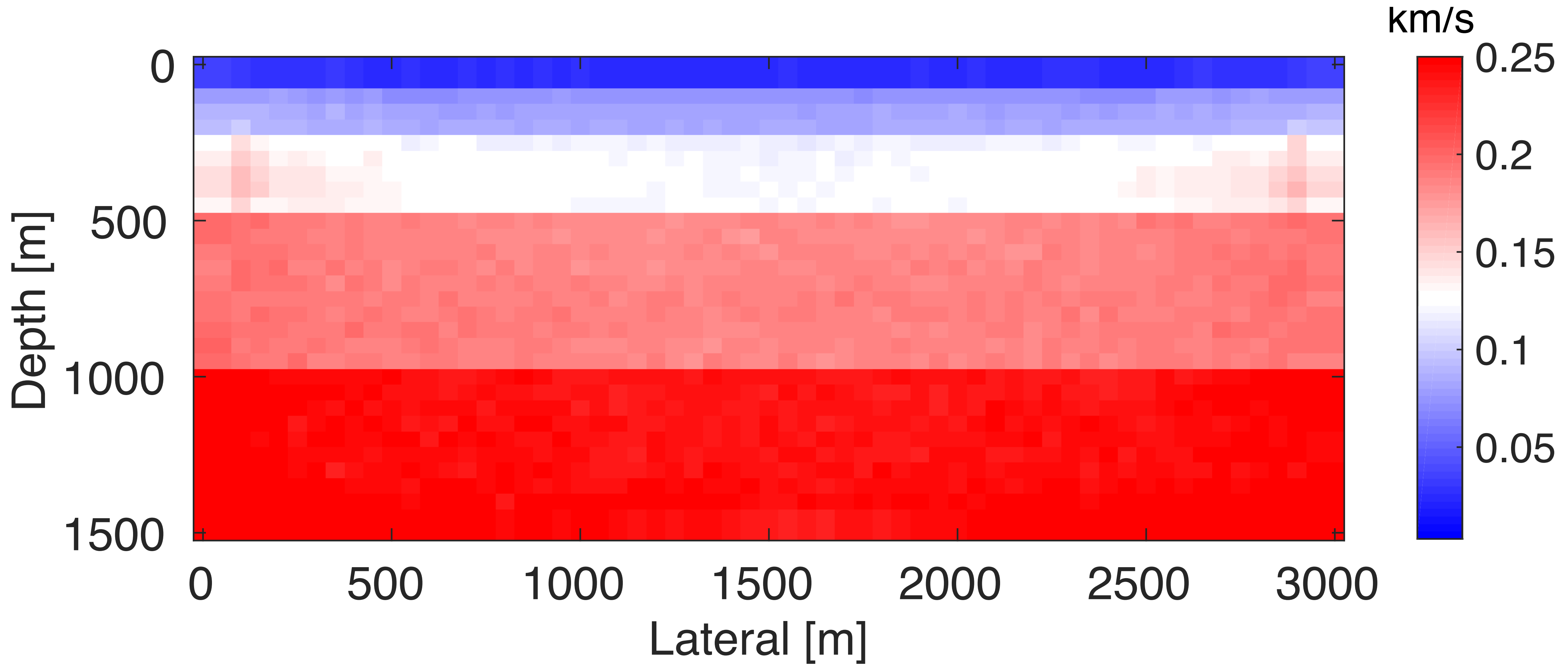
$$\mathbf{r}_{\text{p}} \sim \mathcal{N}(0, \Sigma_{\text{prior}}).$$

# STD result comparison
## – 2-4 Hz

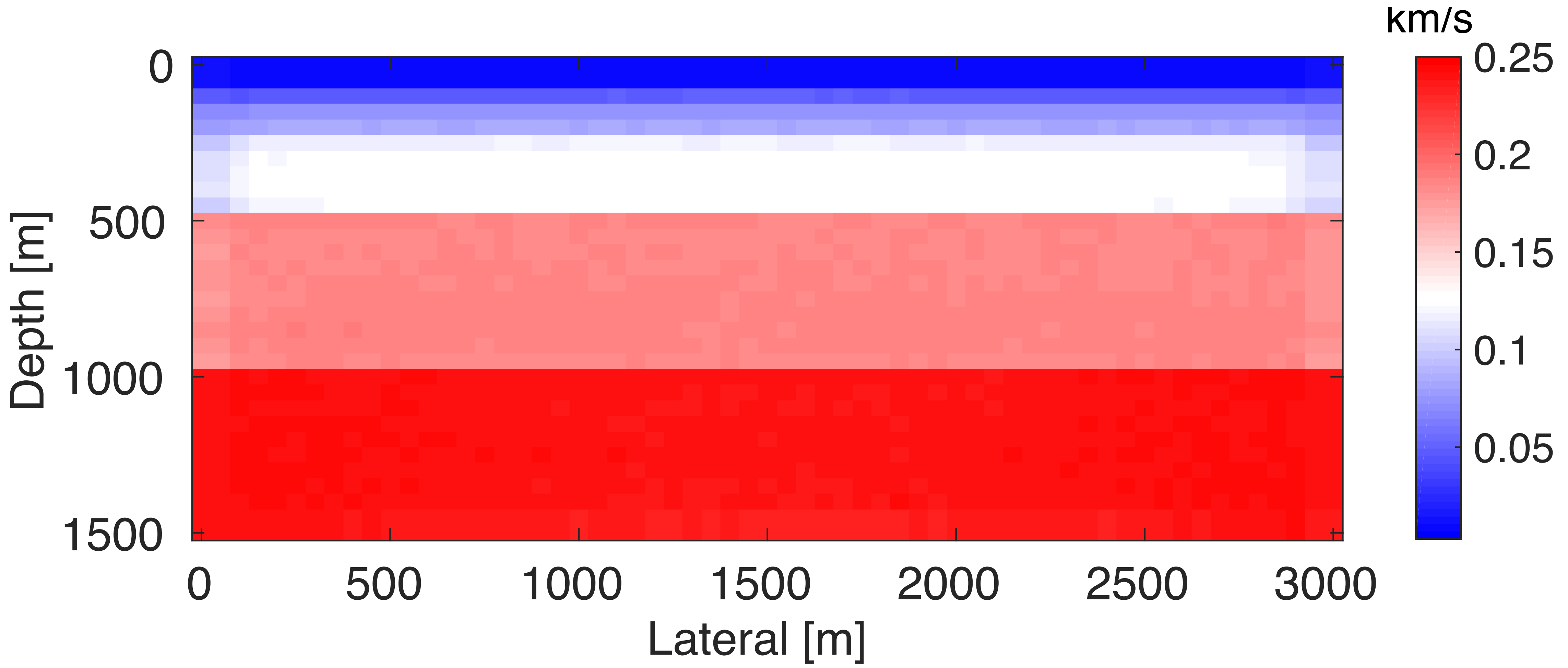



(a) STD of prior distribution

10000 realizations for both methods

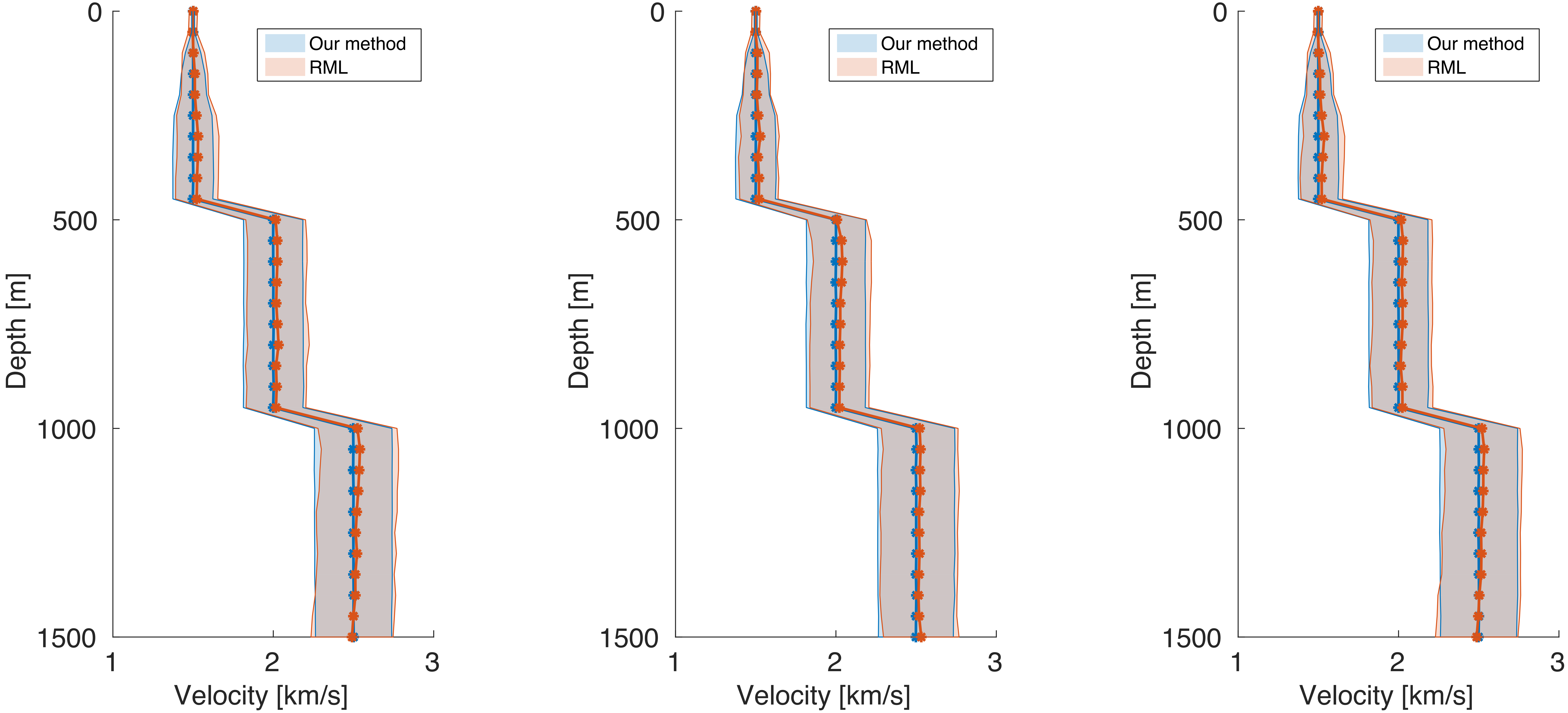



(b) STD obtained by RML



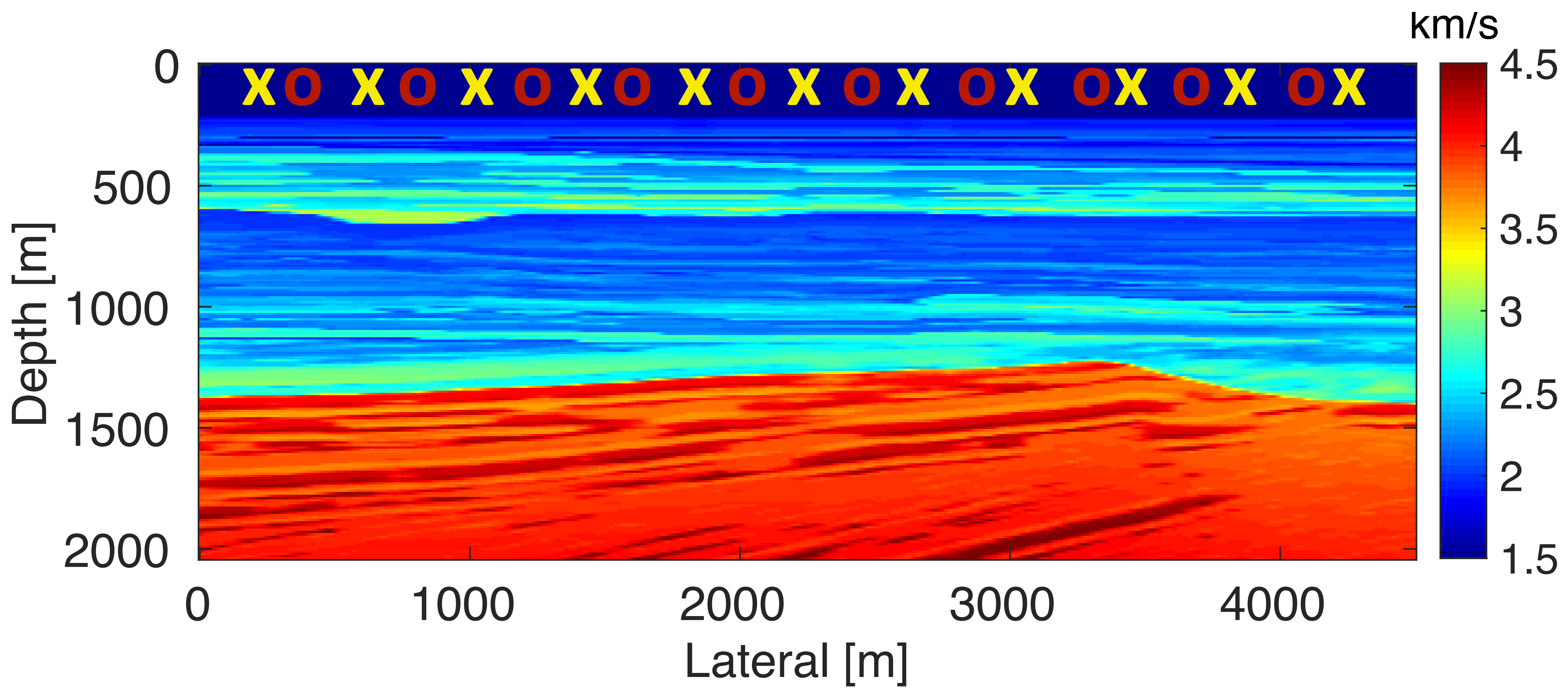

(c) STD of our method

# Cross section comparison



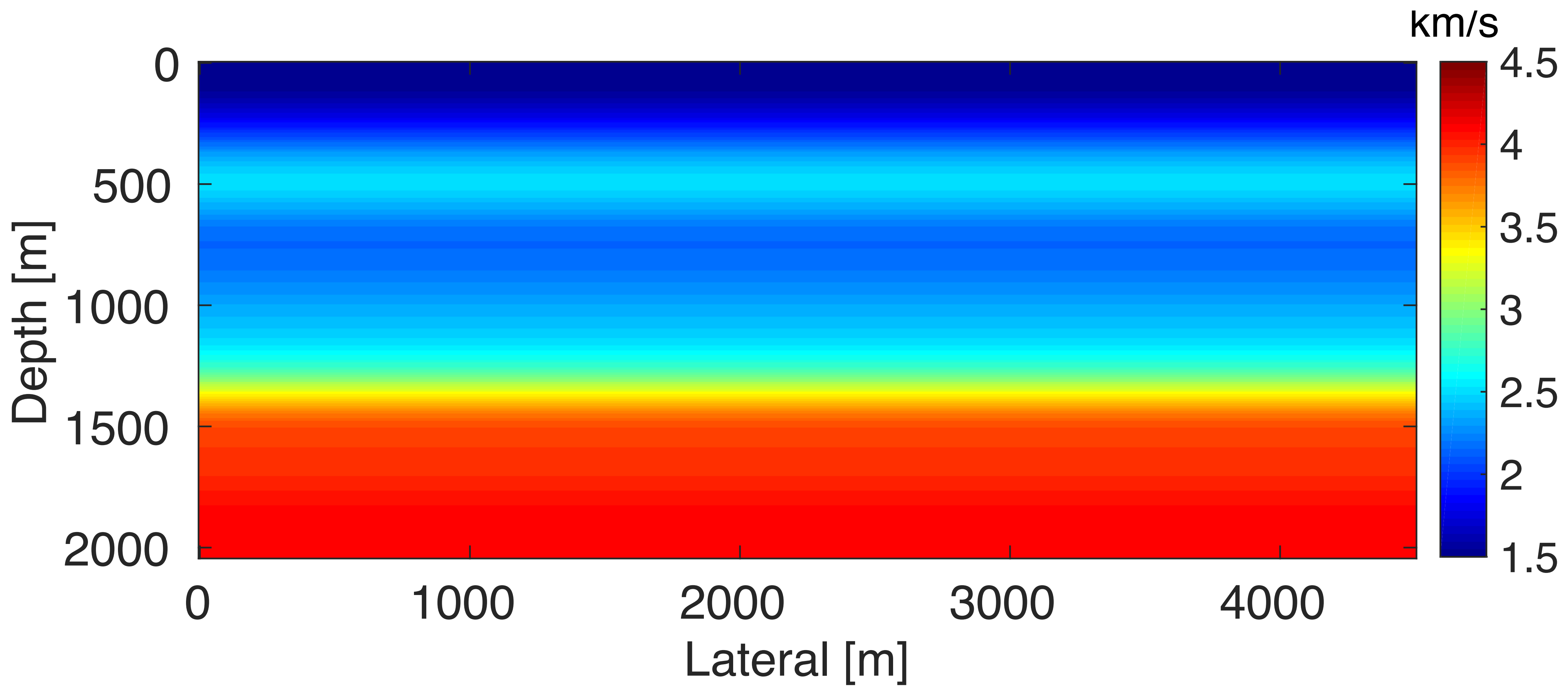

(a) x = 500 m (b) x = 1500 m (c) x = 2500 m

# BG Compass model

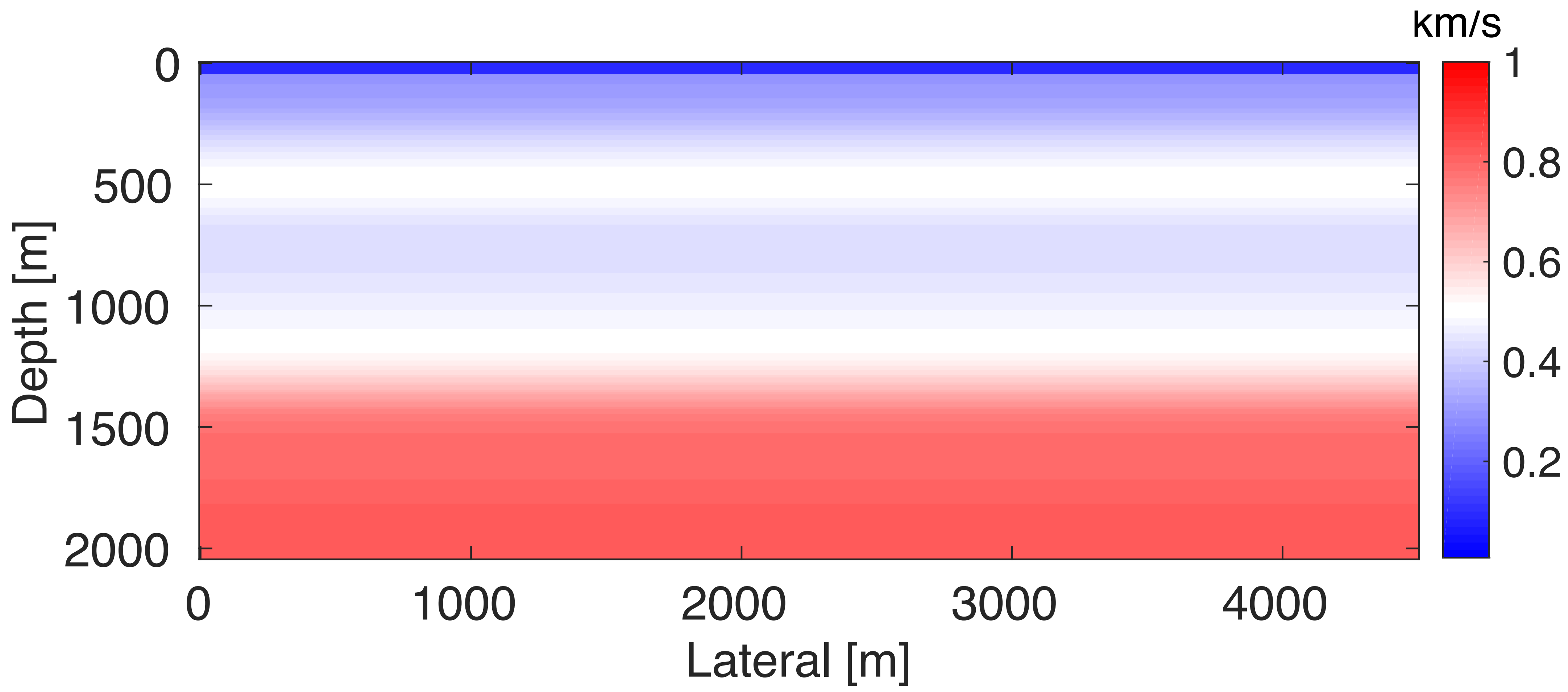



(a) True model

**Depth of sources and receivers: 50**
**Number of sources and receivers: 91 / 451**
**Central frequency: 15 Hz**
**Frequency: 2-31 Hz**
**Lambda: 1e4**



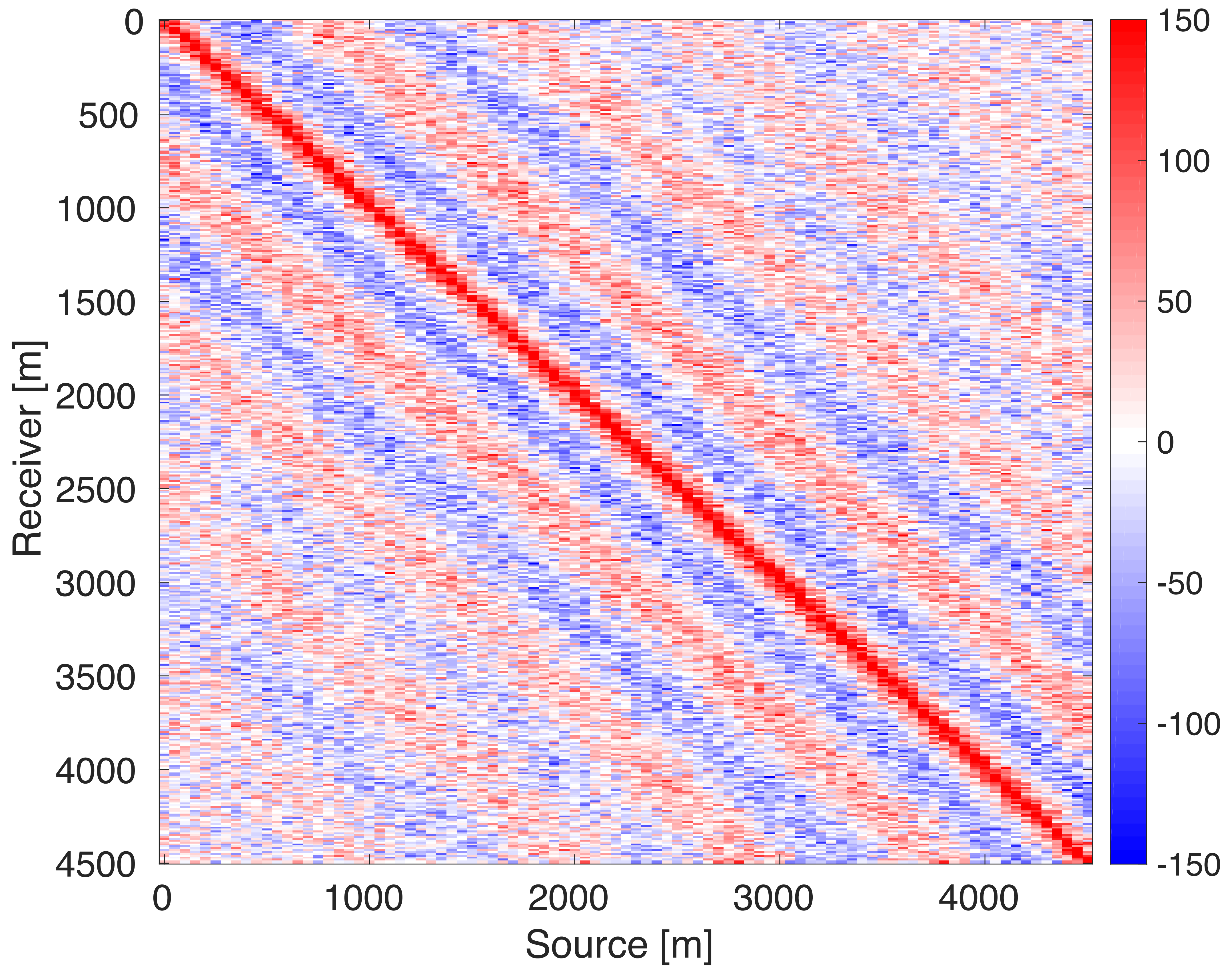

(b) Prior model



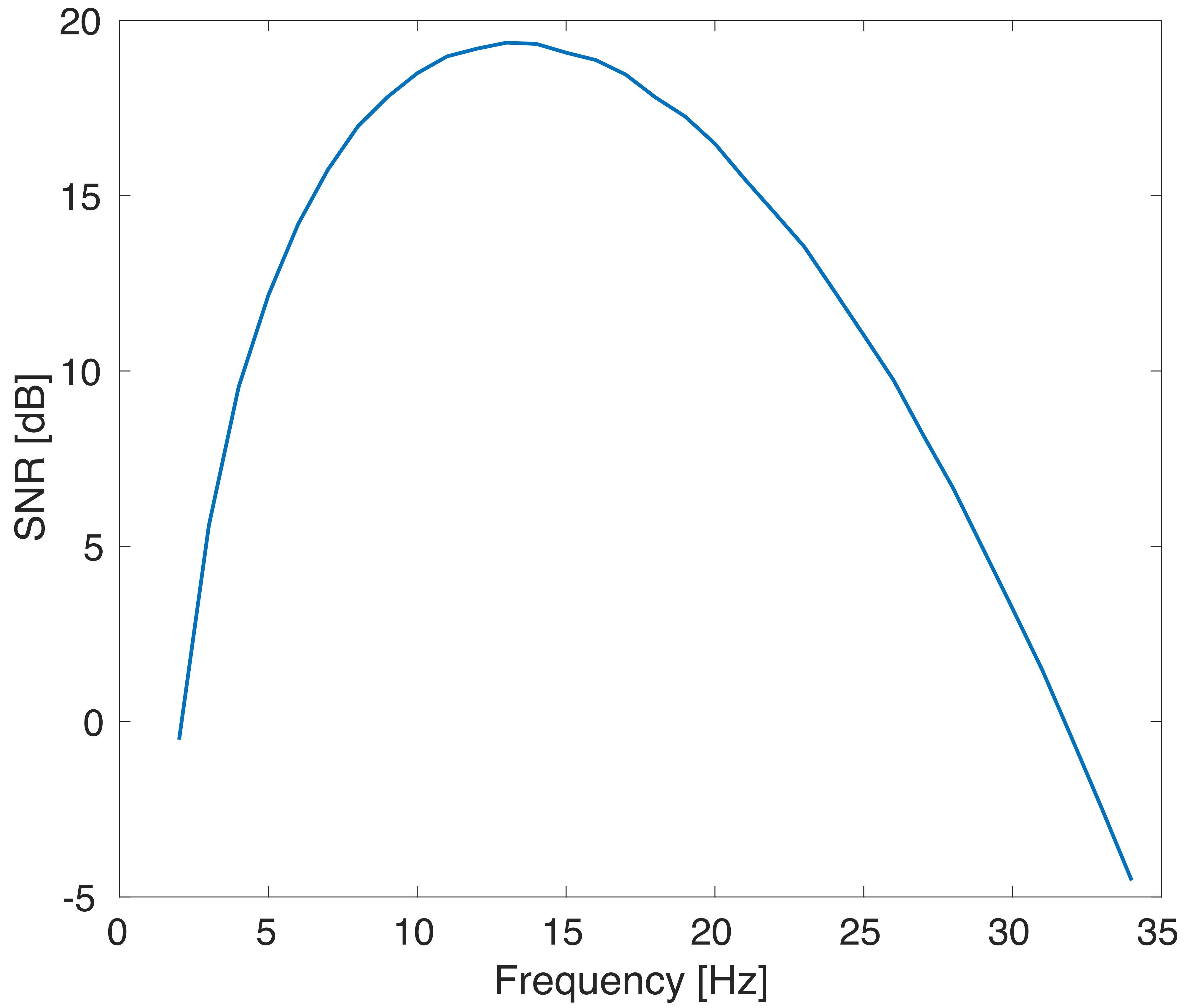

(c) STD of prior distribution

# Data

$\sigma = 34$

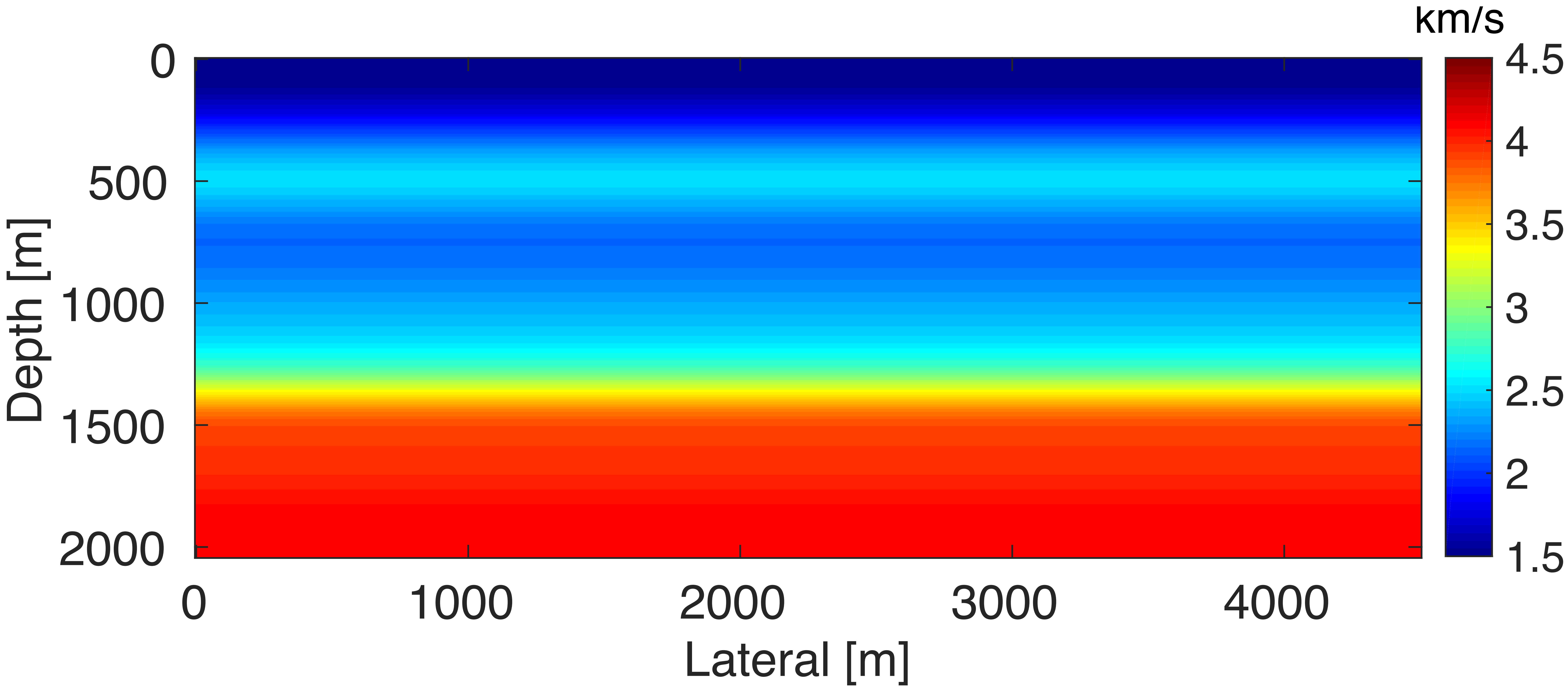



(a) Data at 2Hz

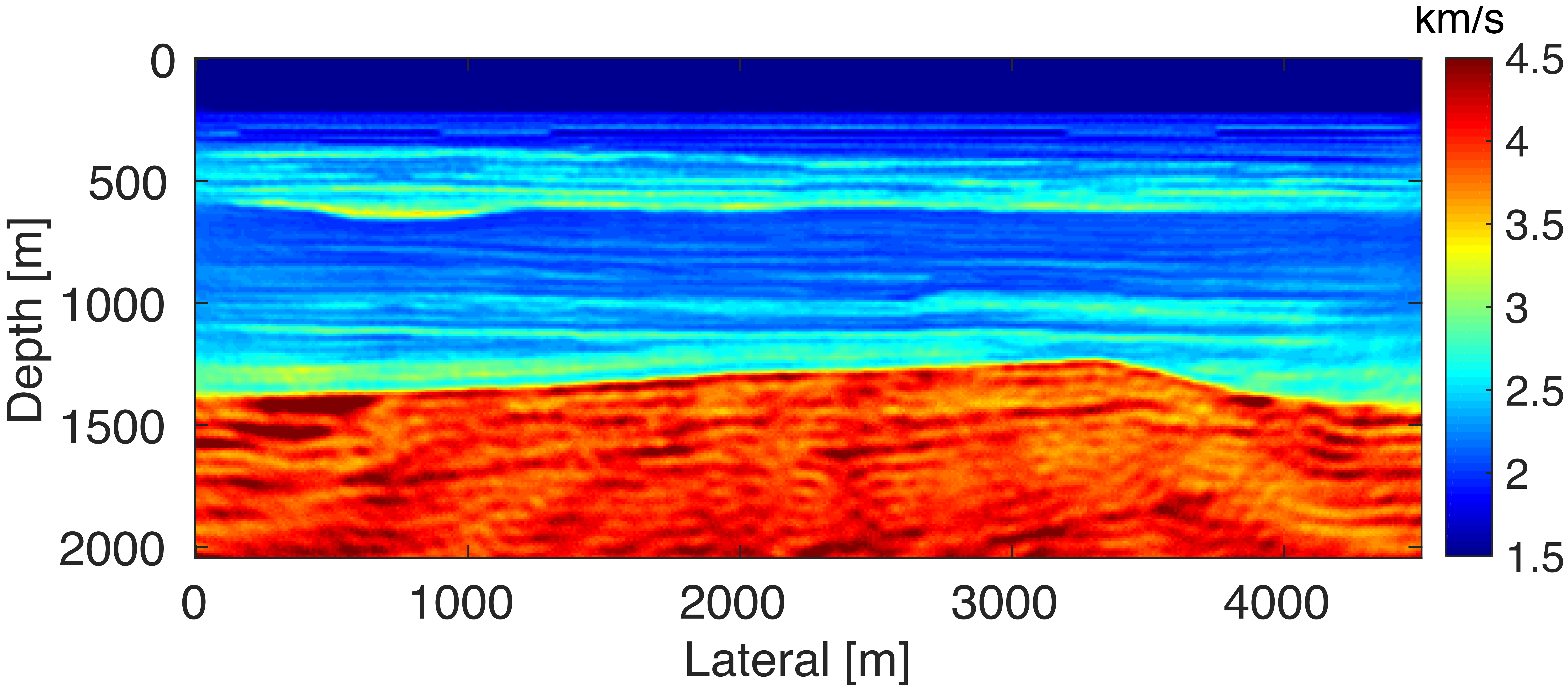



(b) Signal to noise ratio

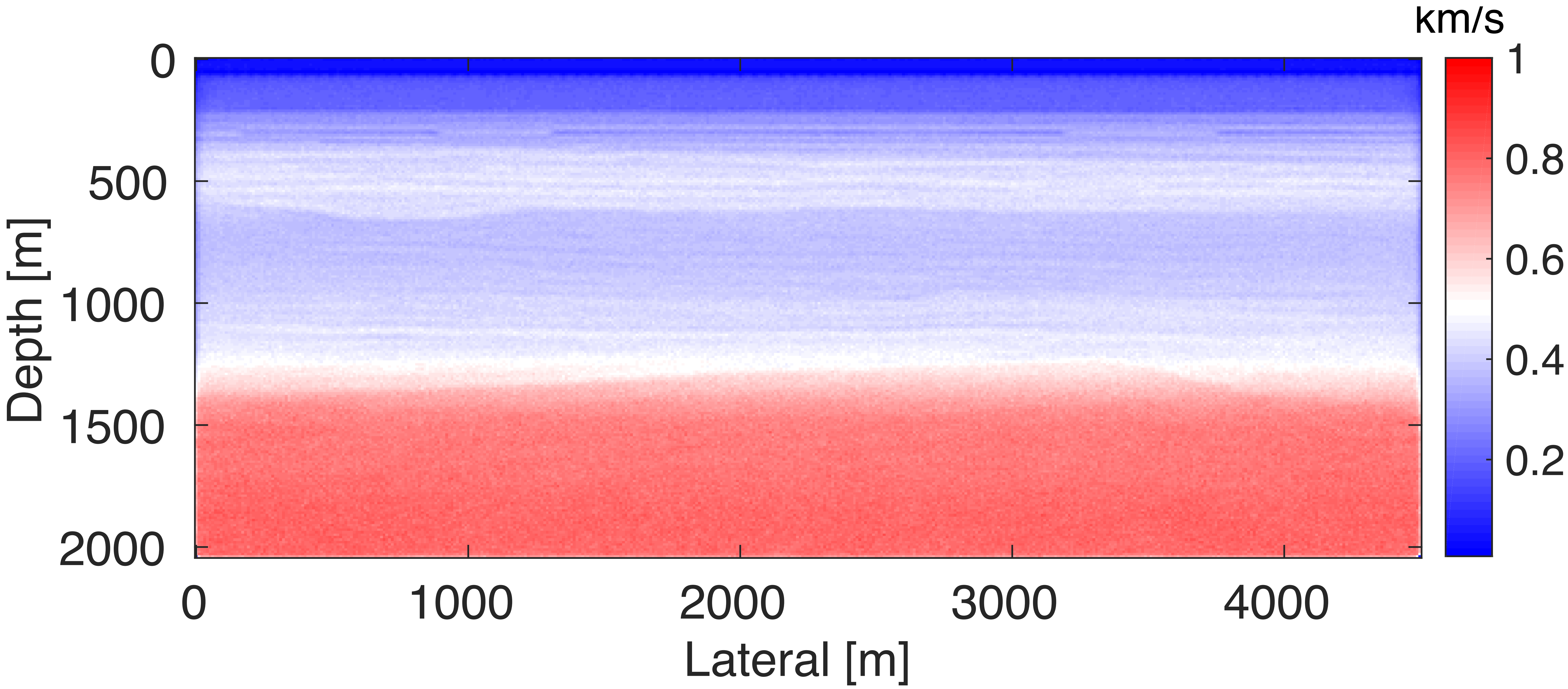
# Prior and initial model



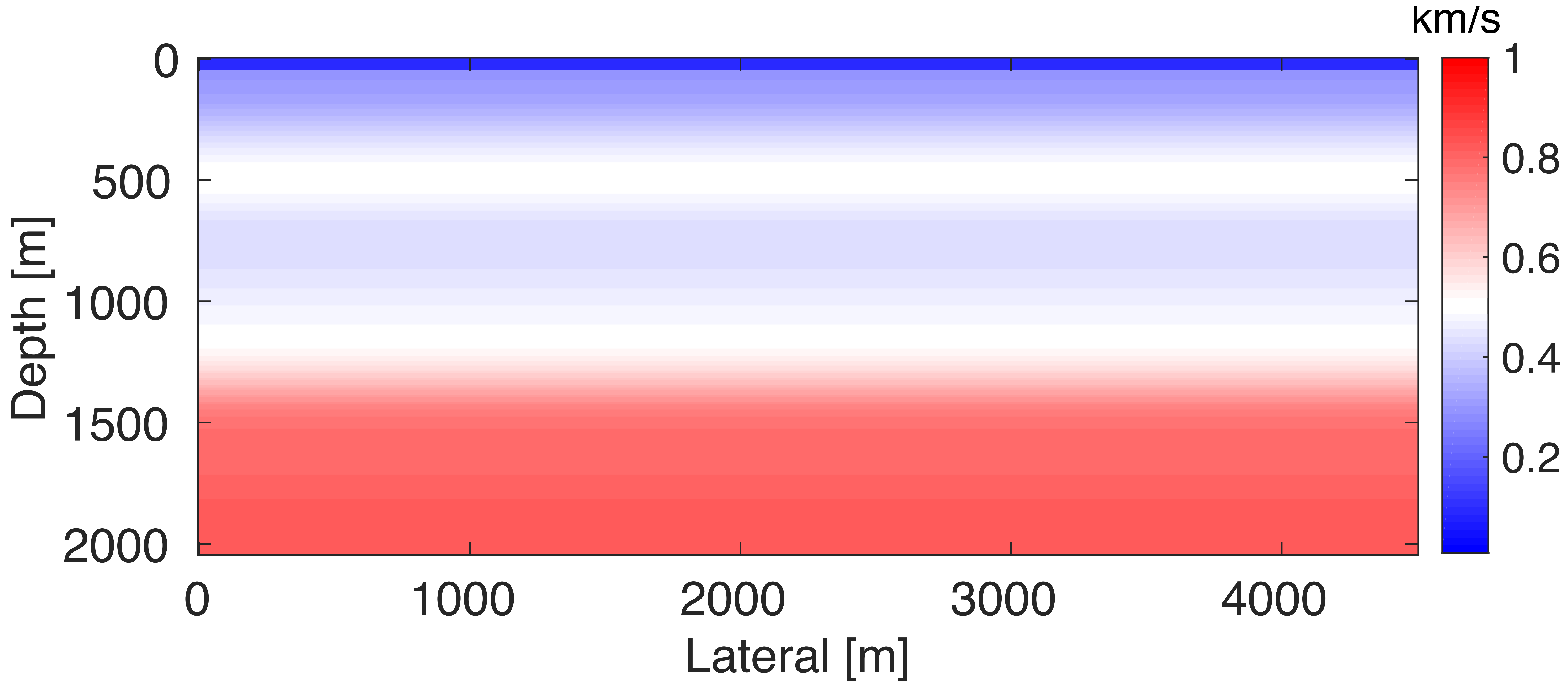
# MAP estimate



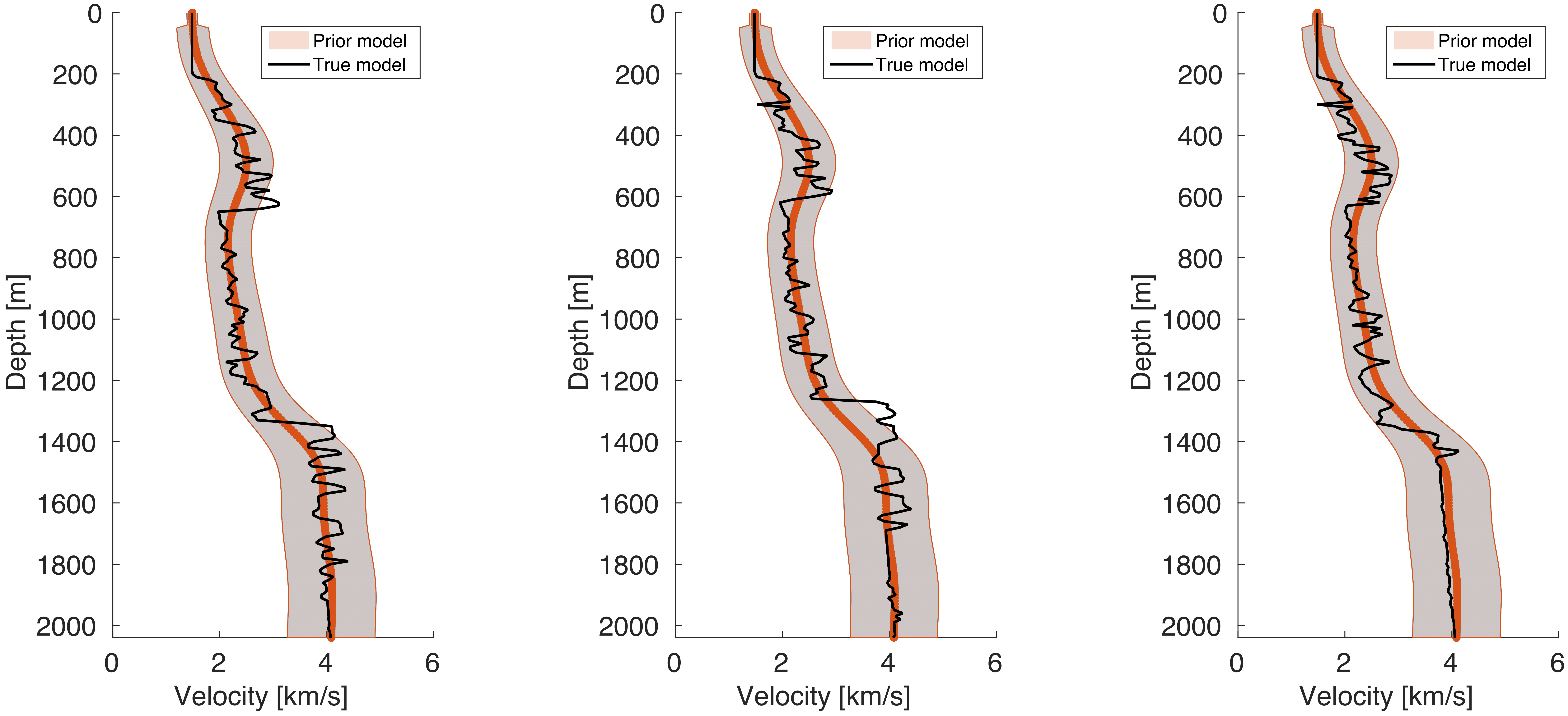
# Posterior STD



# Prior STD


km/s
1
0.8
0.6
0.4
0.2
0
500
1000
1500
2000
Depth [m]
0
1000
2000
3000
4000
Lateral [m]

# Cross section comparison
## – prior

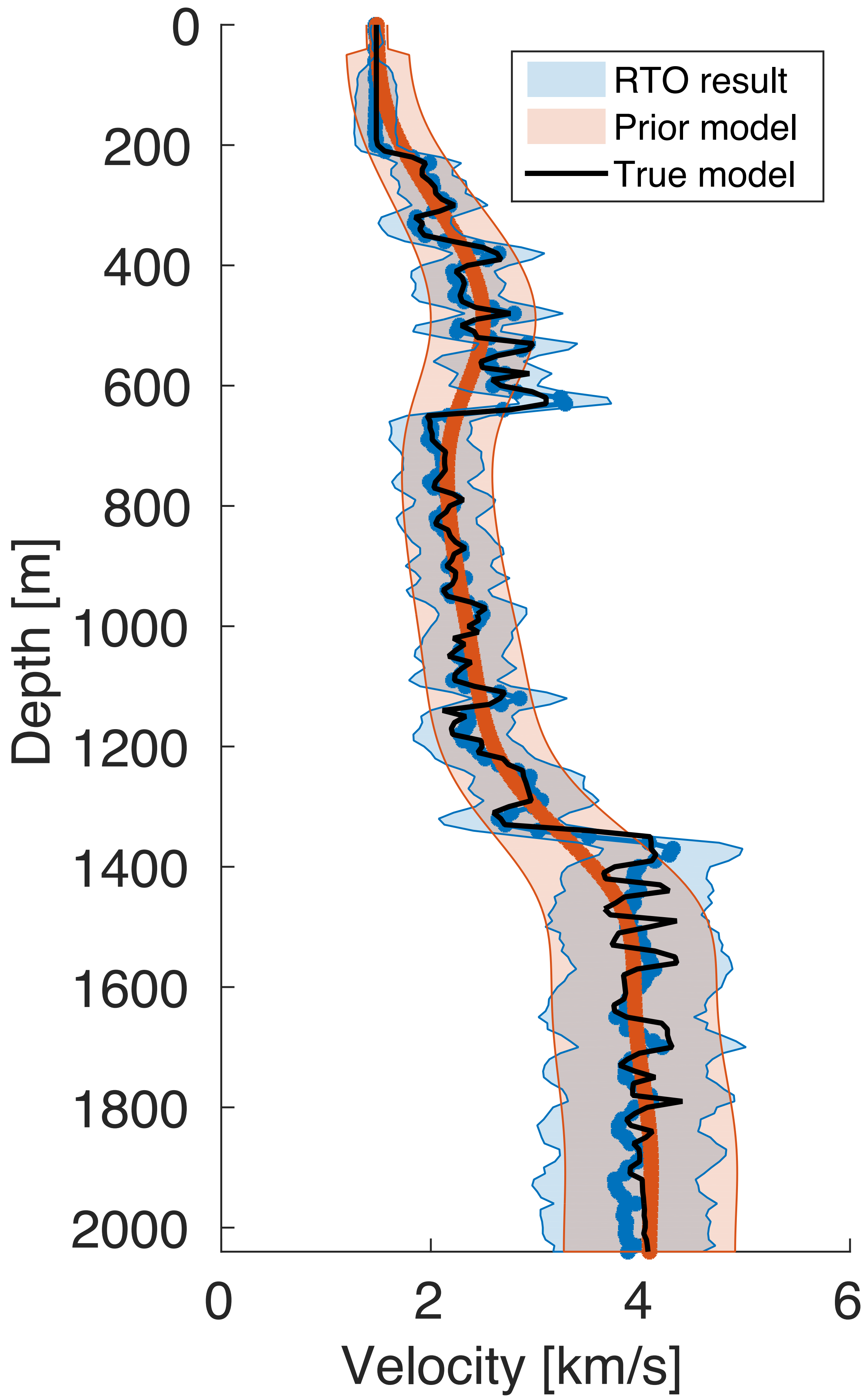



(a) x = 1000 m

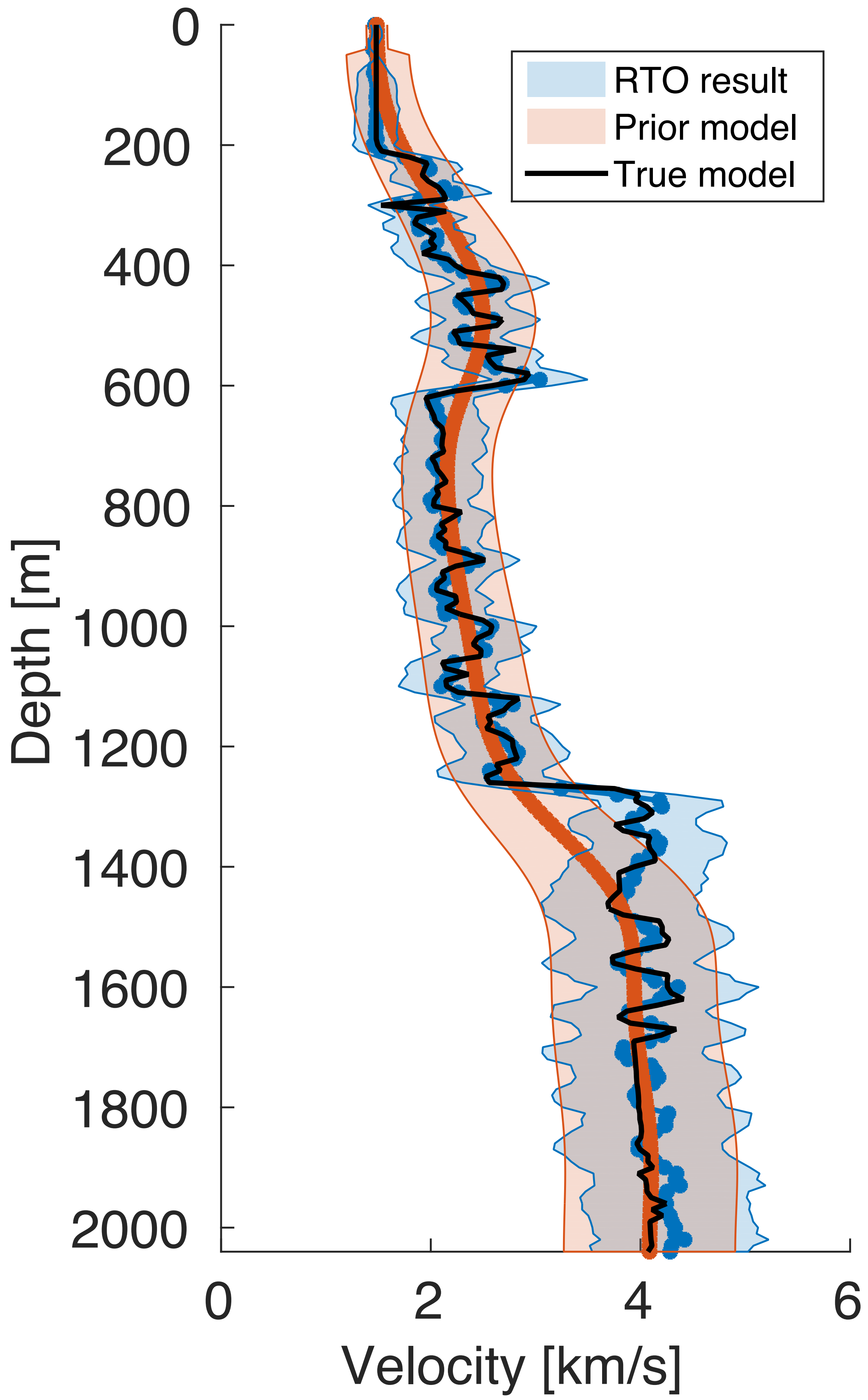
(b) x = 2500 m

(c) x = 4000 m

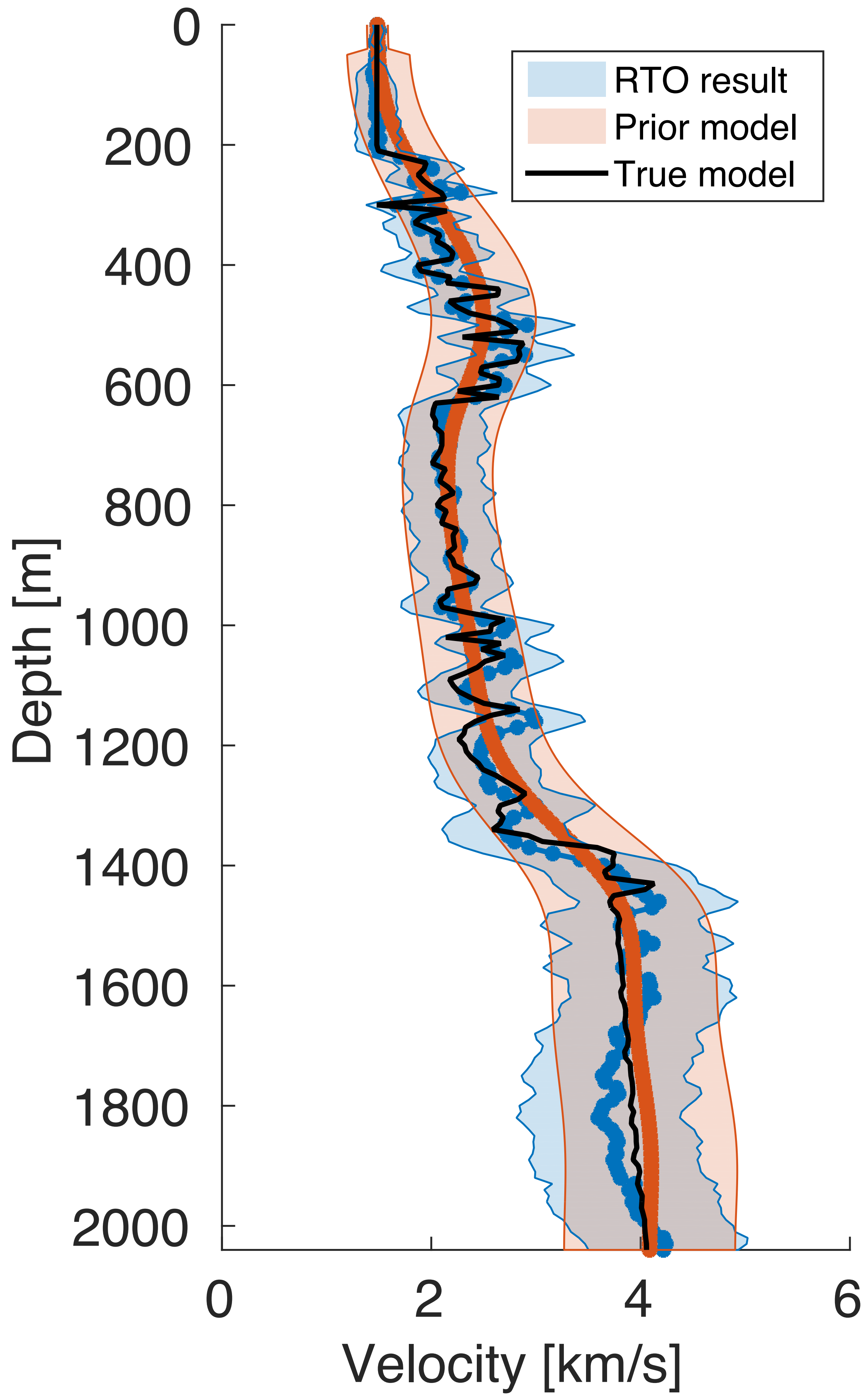
# Cross section comparison
## – posterior vs prior




(a) x = 1000 m

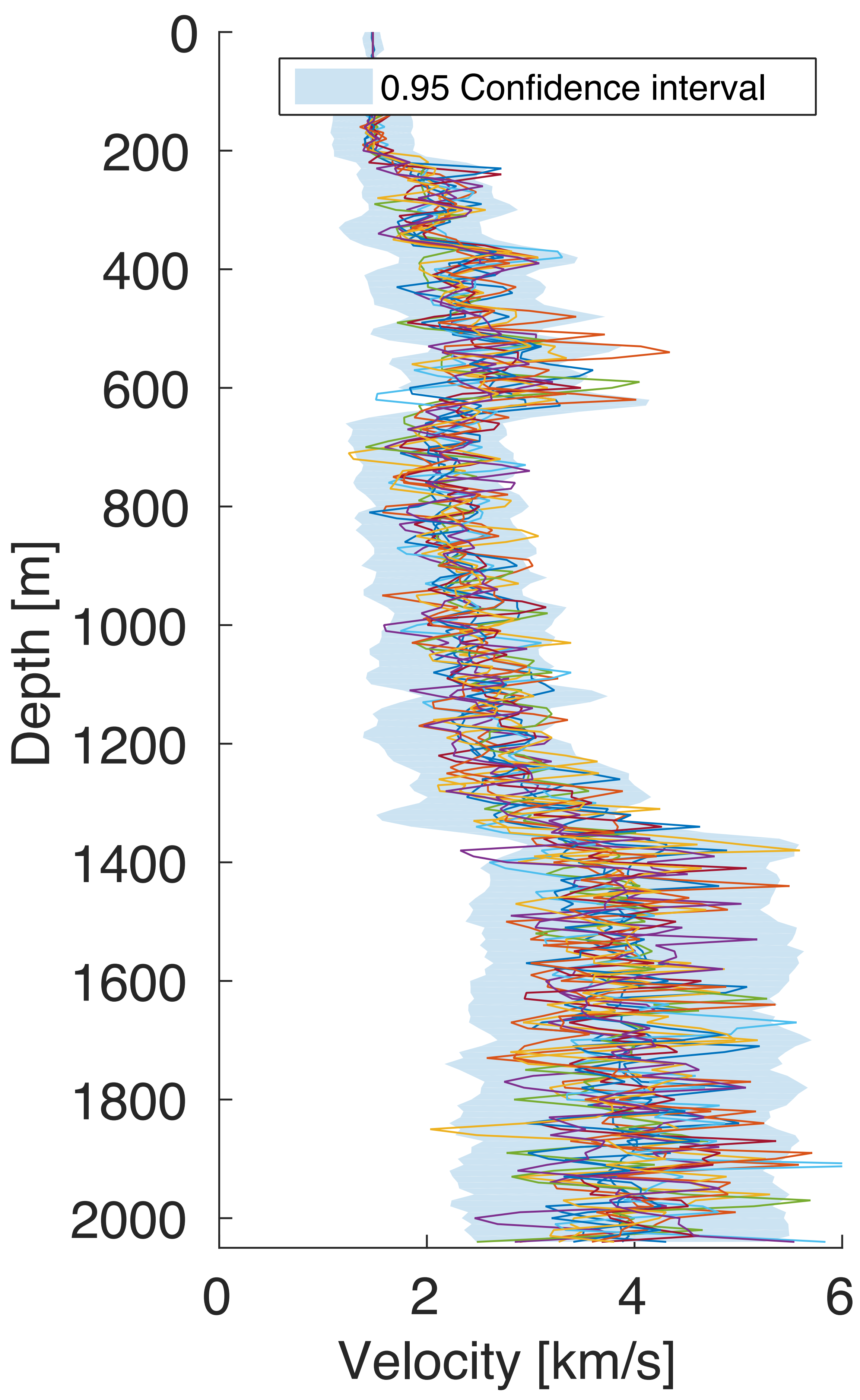



(b) x = 2500 m

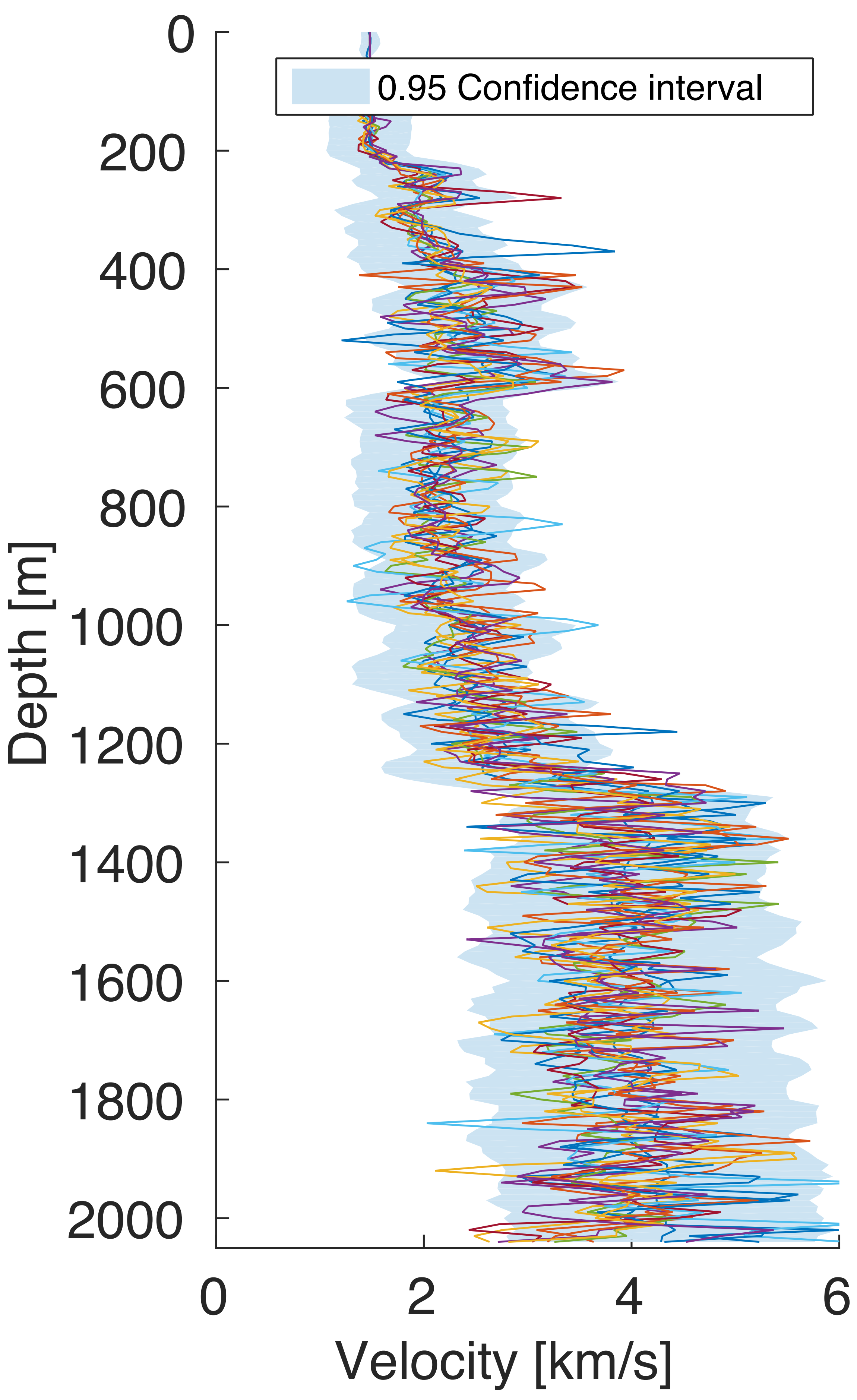



(c) x = 4000 m

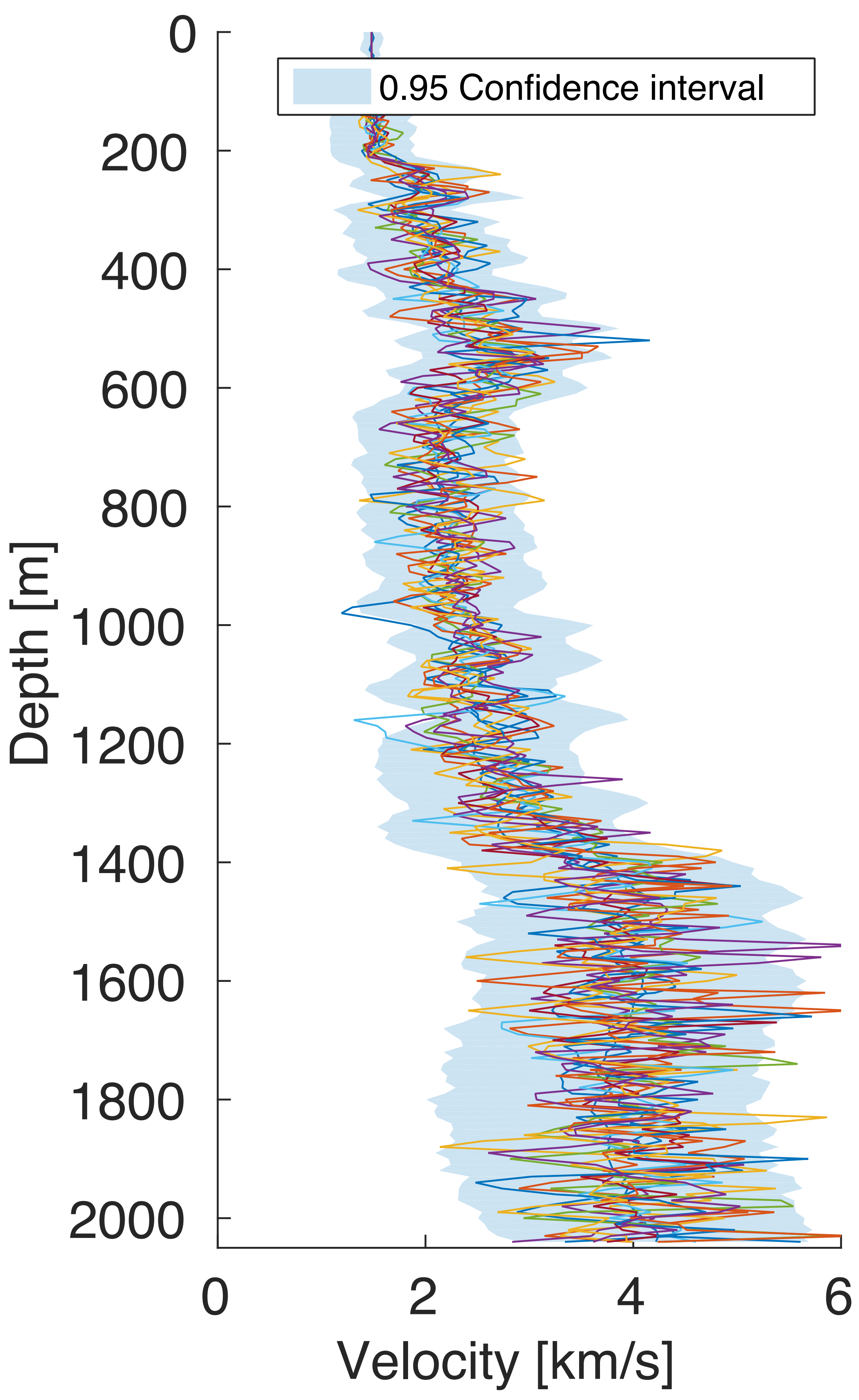
# Cross section comparison
## – 95% confidence interval vs 10 realizations by RML




**(a) x = 1000 m**




**(b) x = 2500 m**




**(c) x = 4000 m**

# Conclusion

The penalty method provides a formulation of posterior PDF with less local maxima, which has a better Gaussian approximation than the reduced formulation.

The implicit formulation of the Gauss-Newton Hessian for WRI provides an efficient way to compute the matrix-vector products.

The RTO method can sample the Gaussian PDF efficiently without the requirement of the explicit Hessian matrix. Numerical experiments show the effectiveness of our method.

# 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.

## Acknowledgements

The authors wish to acknowledge the SENAI CIMATEC Supercomputing Center for Industrial Innovation, with support from BG Brasil, Shell, and the Brazilian Authority for Oil, Gas and Biofuels (ANP), for the provision and operation of computational facilities and the commitment to invest in Research & Development.