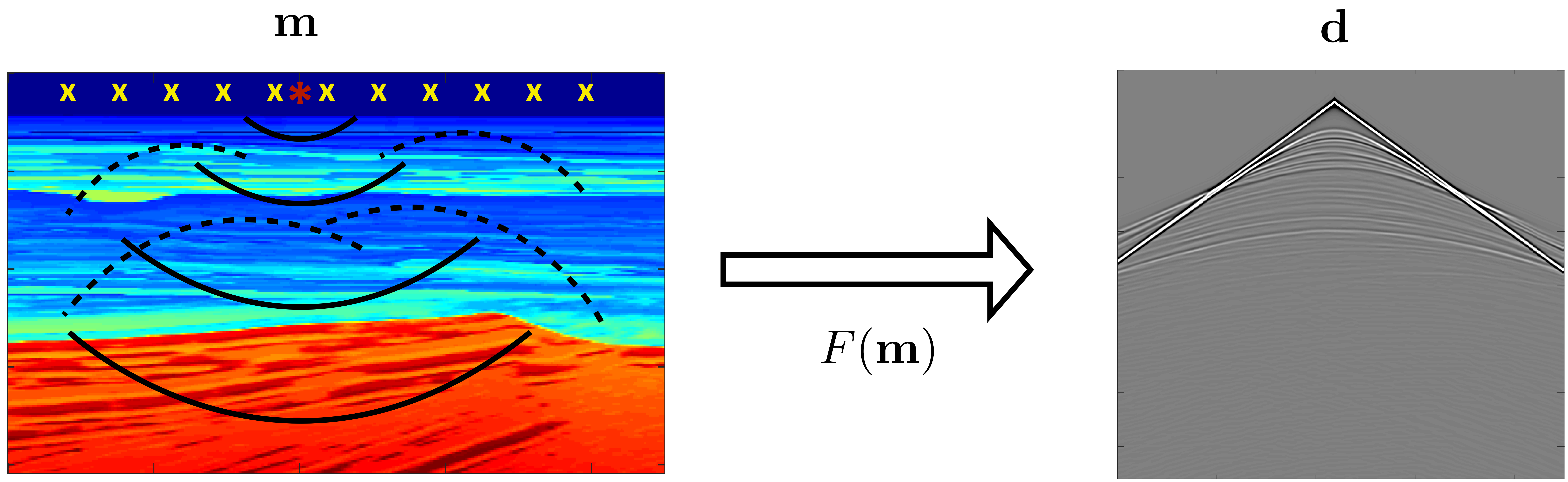
Uncertainty quantification for inverse problems with a weak wave-equation constraint

Zhilong Fang*, Curt Da Silva*, Rachel Kuske** and Felix J. Herrmann*
*Seismic Laboratory for Imaging and Modeling (SLIM), University of British Columbia
**Department of Mathematics, Georgia Institute of Technology

SLIM
University of British Columbia

Motivation

Forward problem



Motivation

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

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

$\mathbf{A}(\mathbf{m})$ → $\Delta + \omega^2 \mathbf{m}$

$\mathbf{q}$ → source

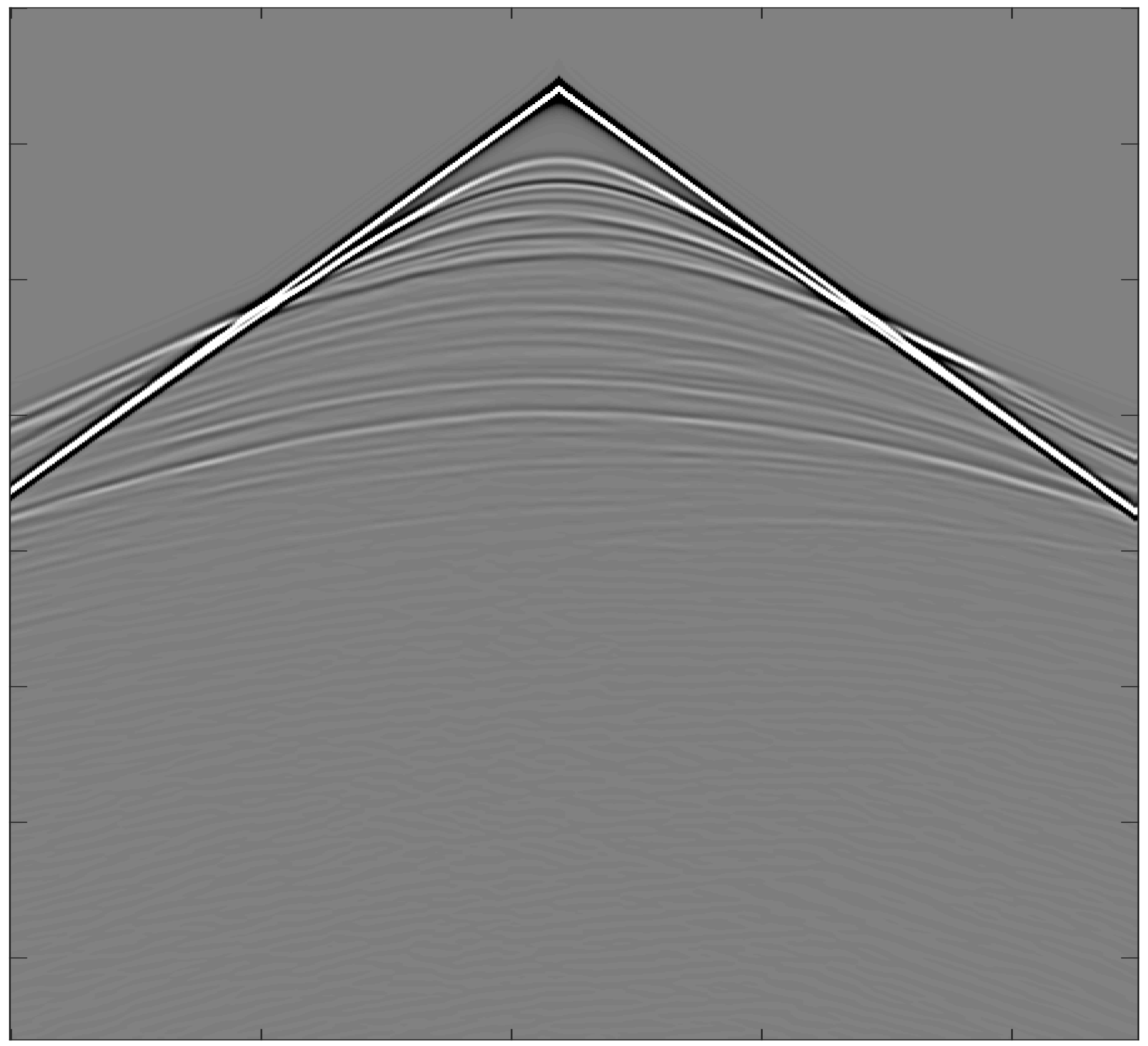
$\mathbf{P}$ → projection operator

ω → frequency

$\mathbf{m}$ → squared-slowness

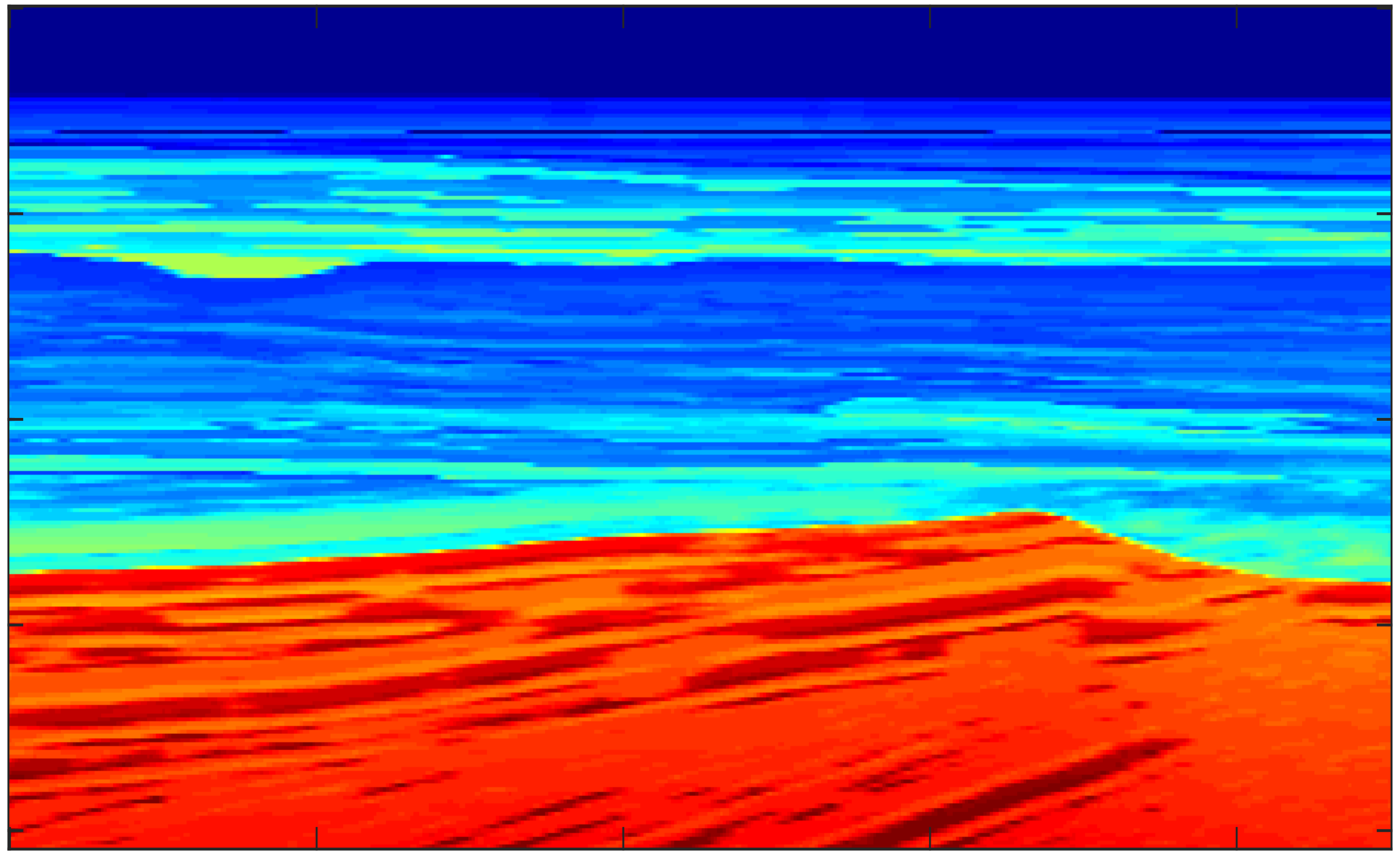
Motivation

Inverse problem



$\mathbf{d}$

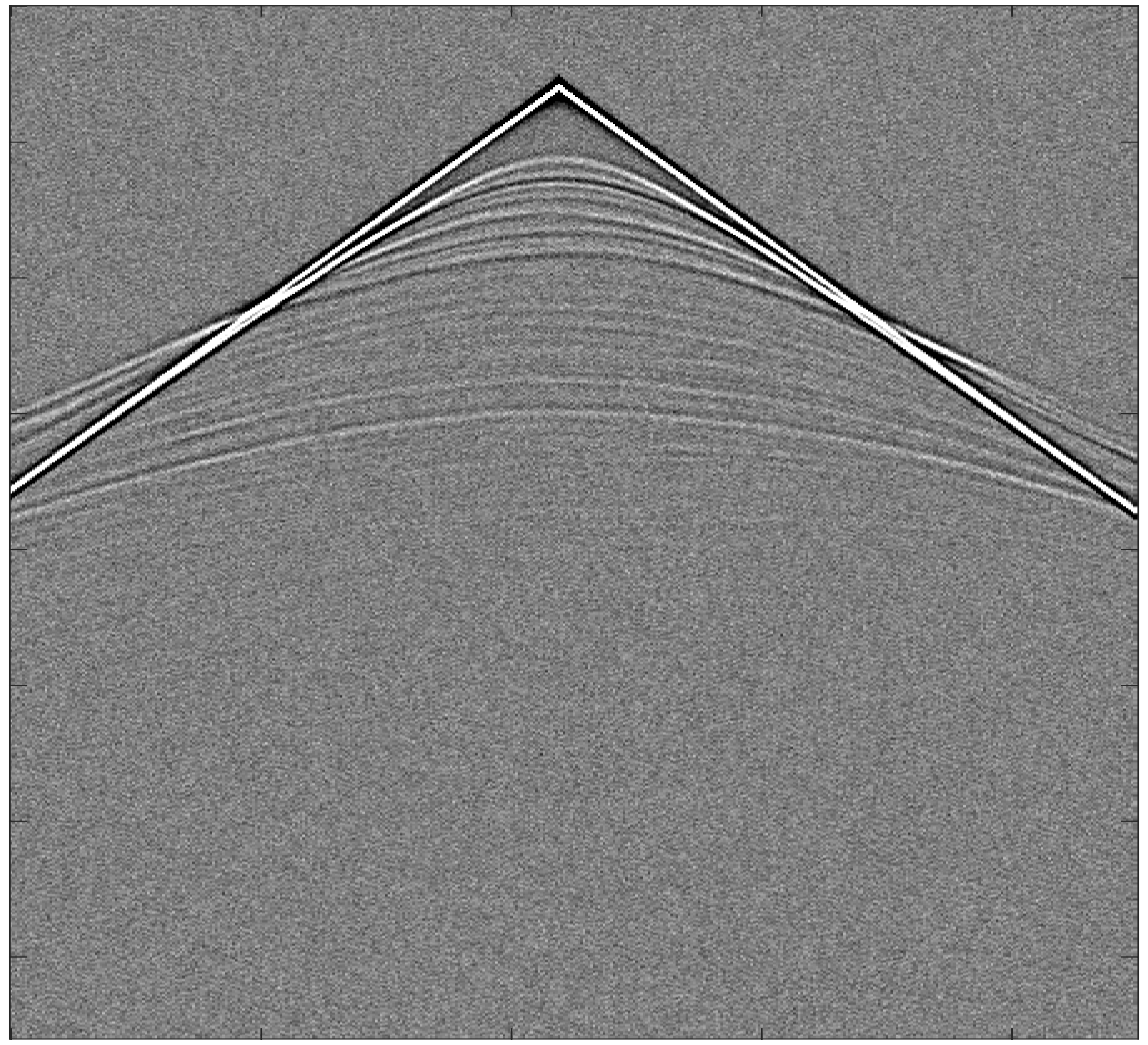
$\xrightarrow{\quad}$ $F^{-1}(\mathbf{d})$



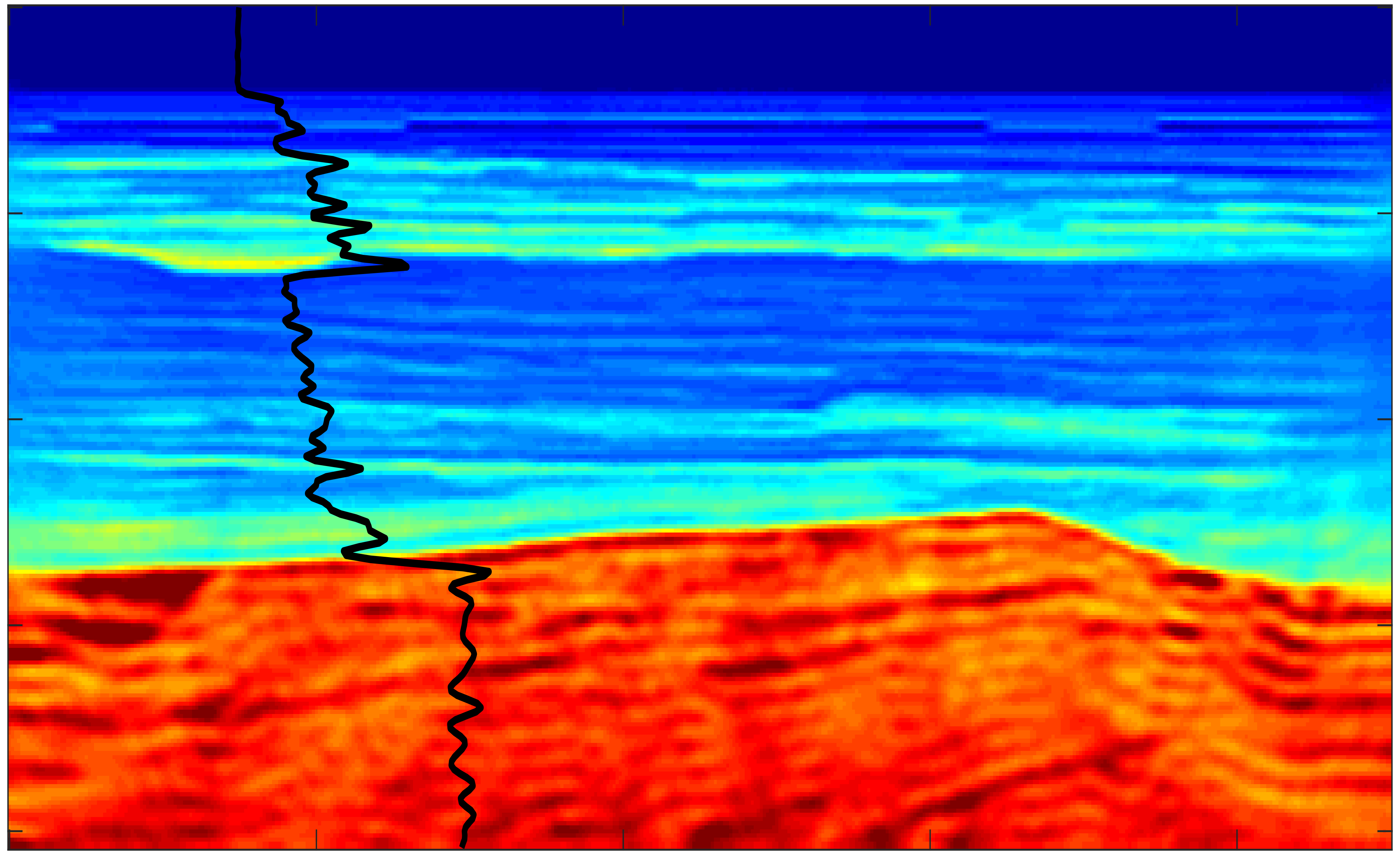
$\mathbf{m}$

Motivation

Statistical inverse problem



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



$\mathbf{d}$

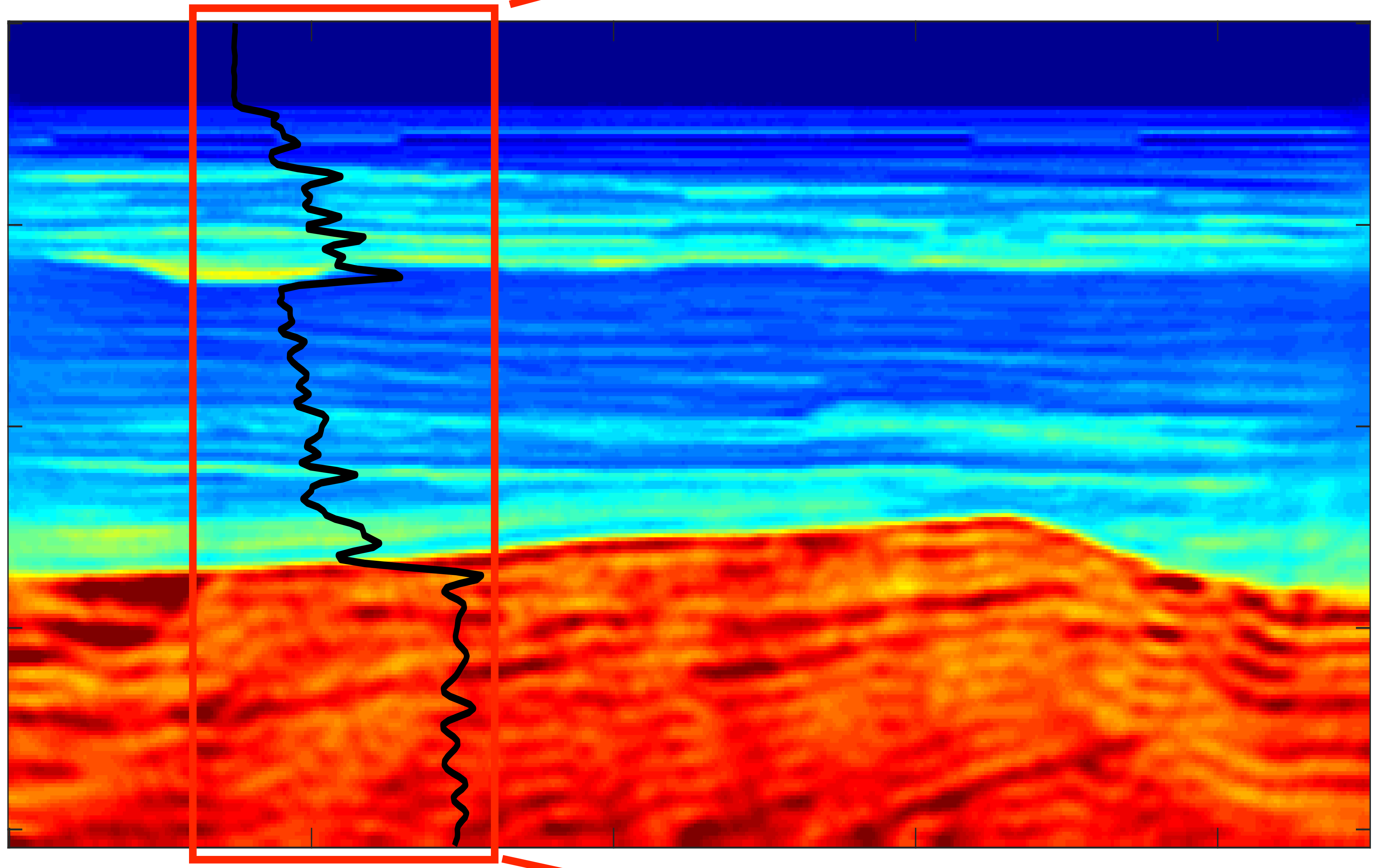
noisy

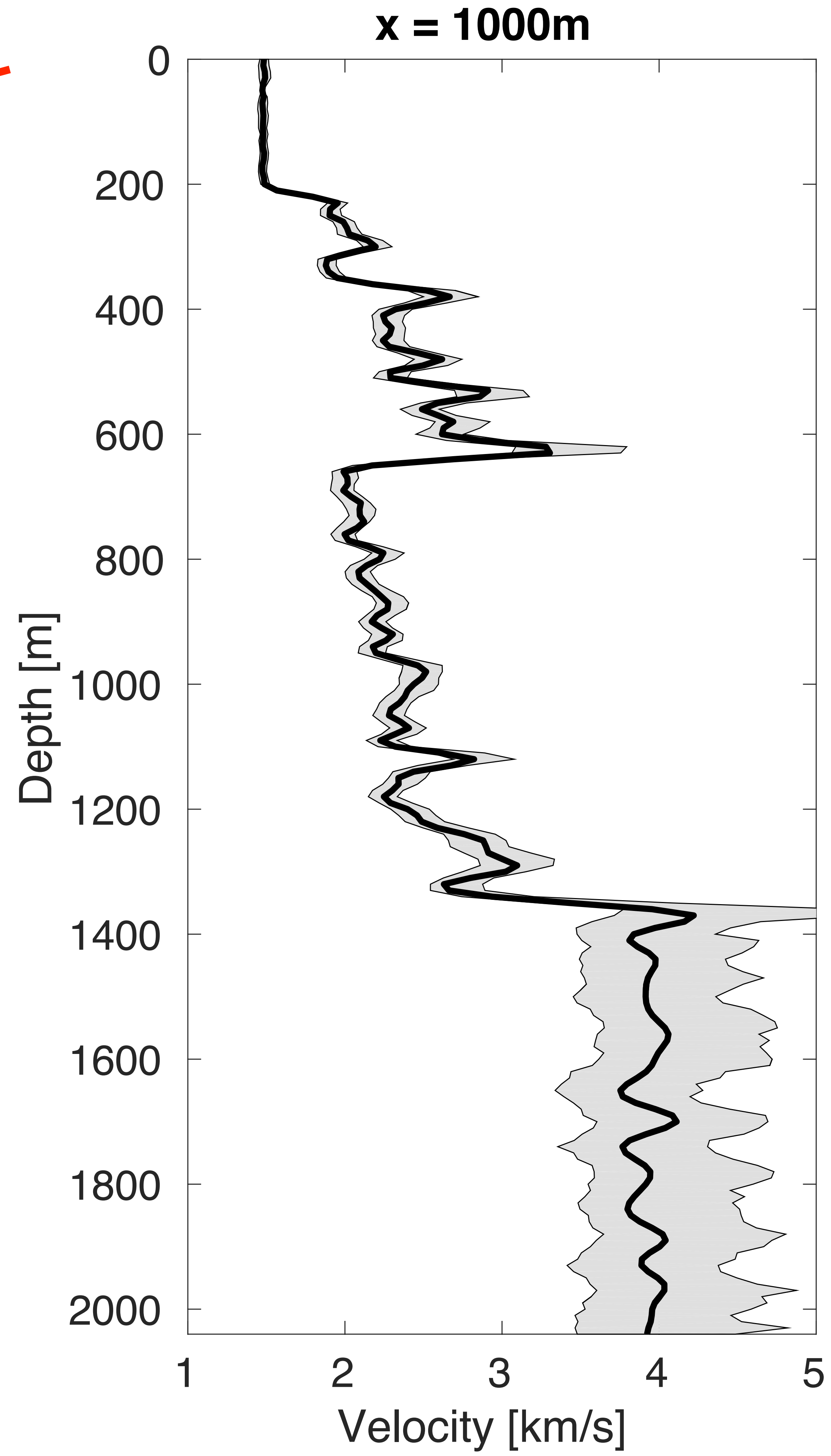
$\mathbf{m}$

noisy

Motivation

Statistical inverse problem





Confidence Interval

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

SLIM

Bayesian inference

Prior probability density function (PDF):

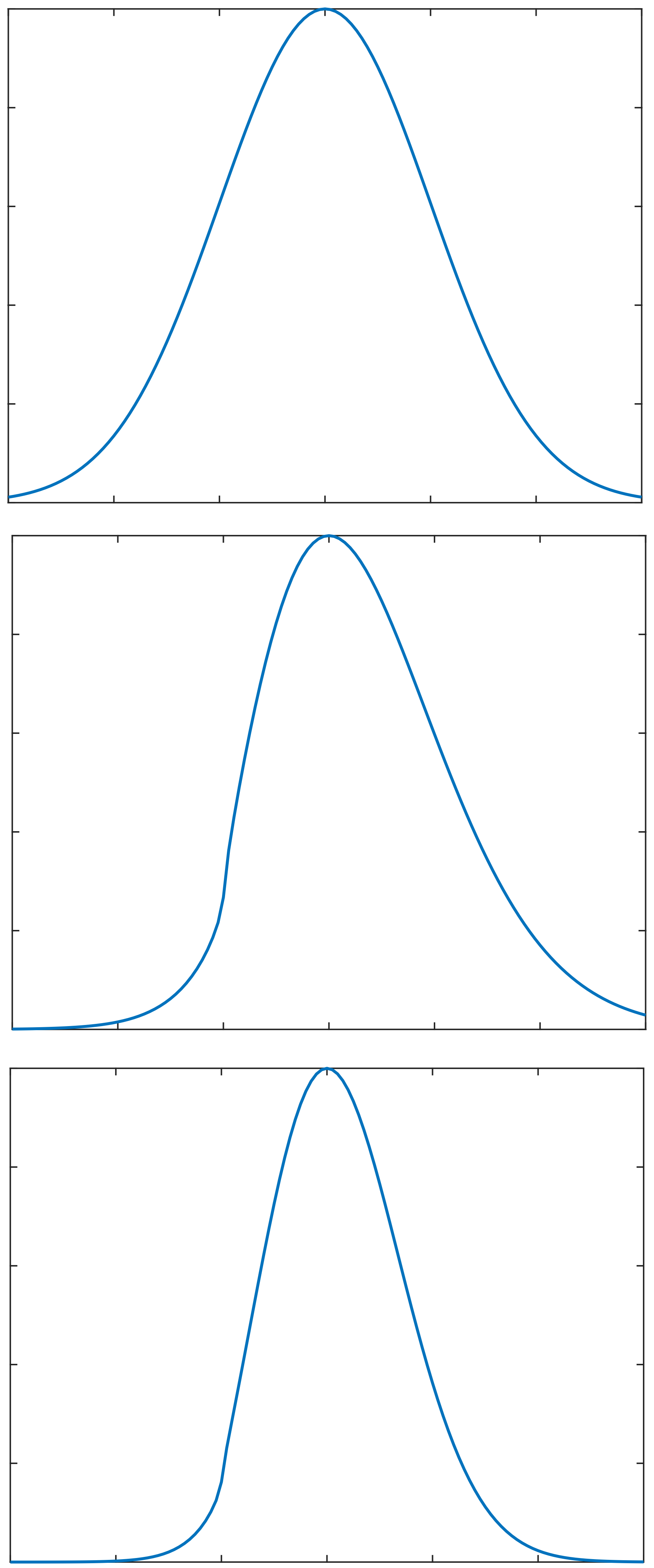
$$\mathbf{m} \longrightarrow \rho_{\text{prior}}(\mathbf{m})$$

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

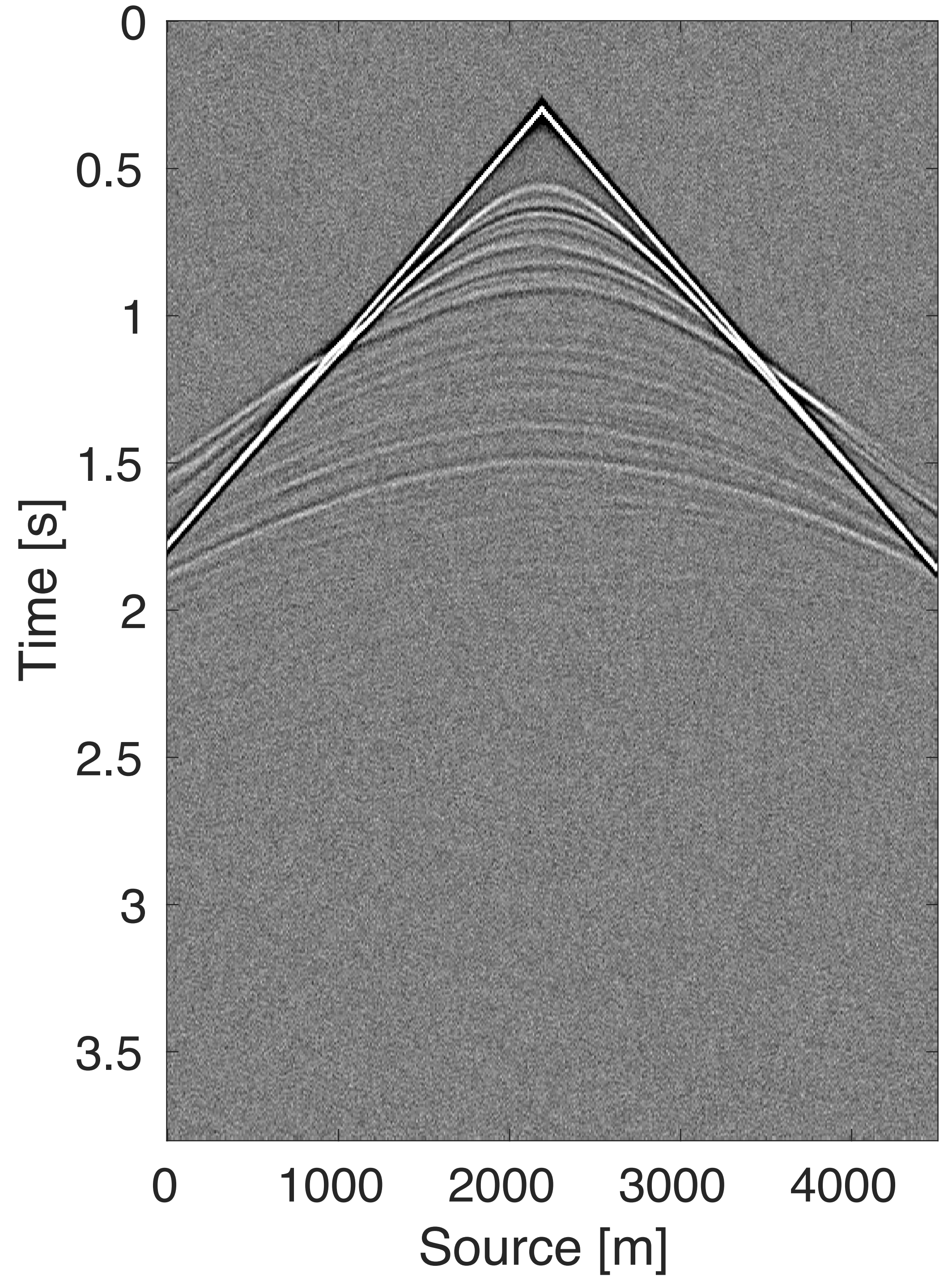
$$\mathbf{m} \longrightarrow \rho_{\text{like}}(\mathbf{d}|\mathbf{m})$$

Posterior PDF (Bayes' rule):

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

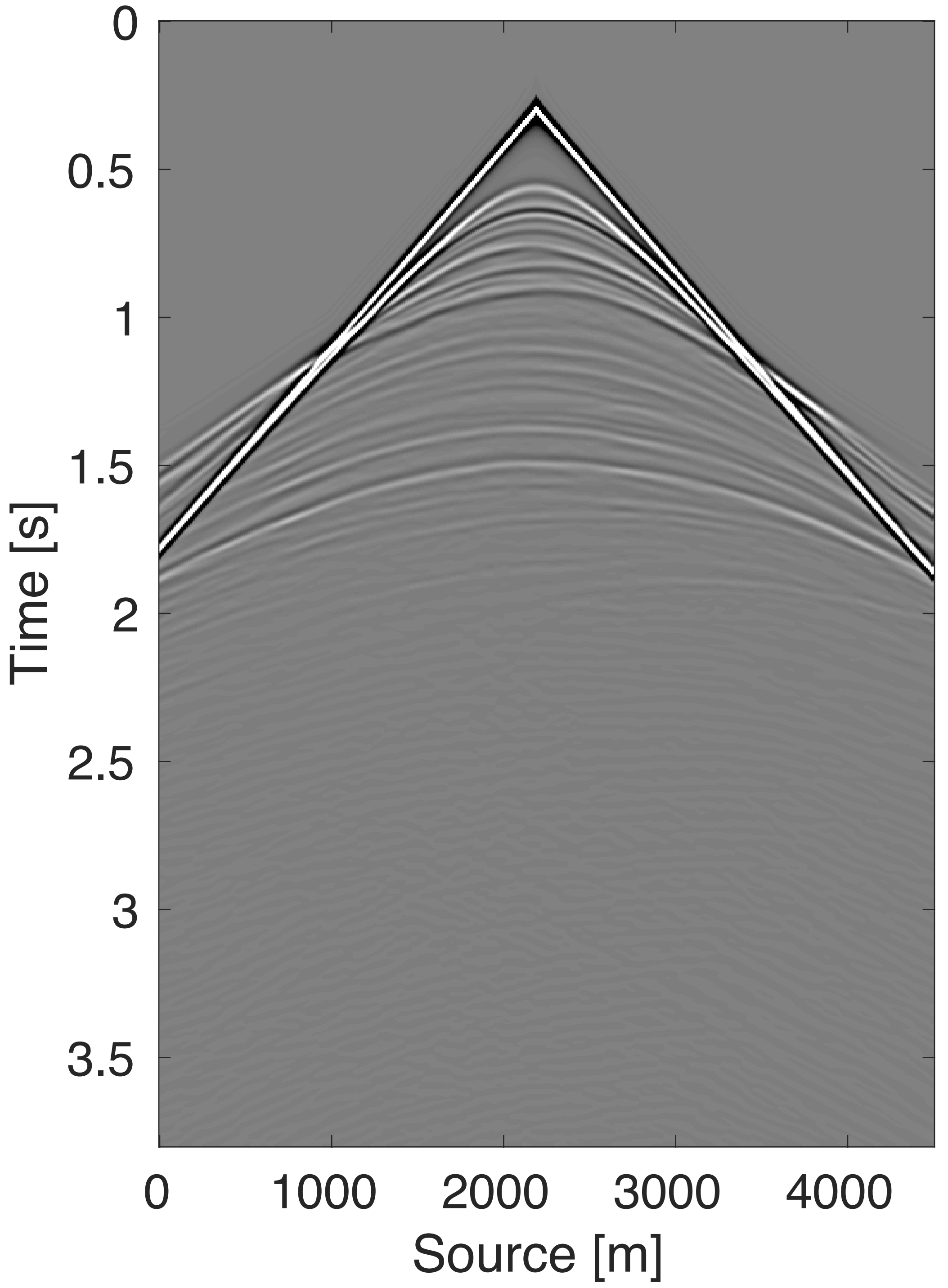


Bayes w/ strong PDE constraints



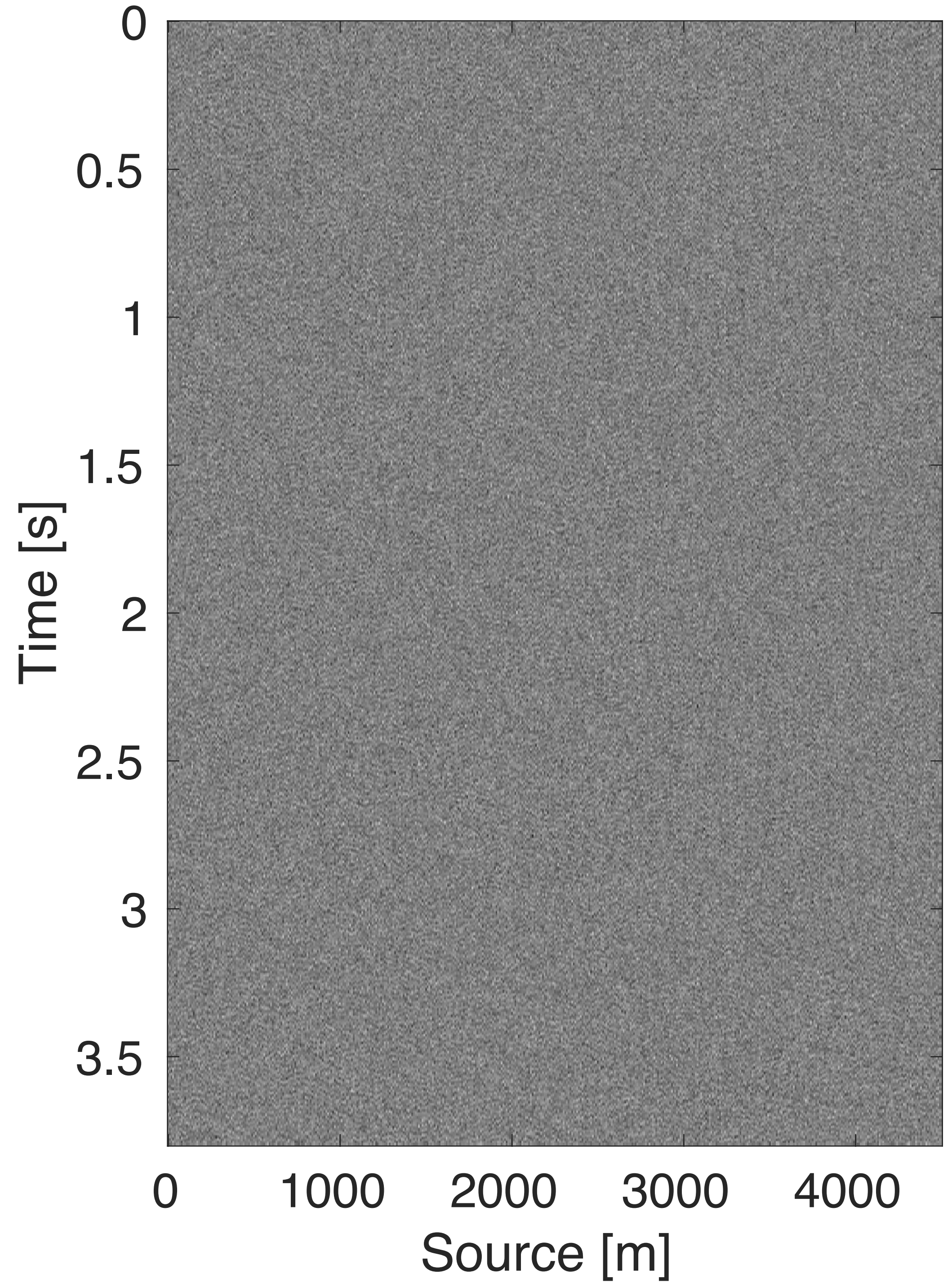
$\mathbf{d}$

=



$F(\mathbf{m})$

+



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

Bayes w/ strong PDE constraints

Posterior PDF w/ strong PDE constraints:

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

Challenges:

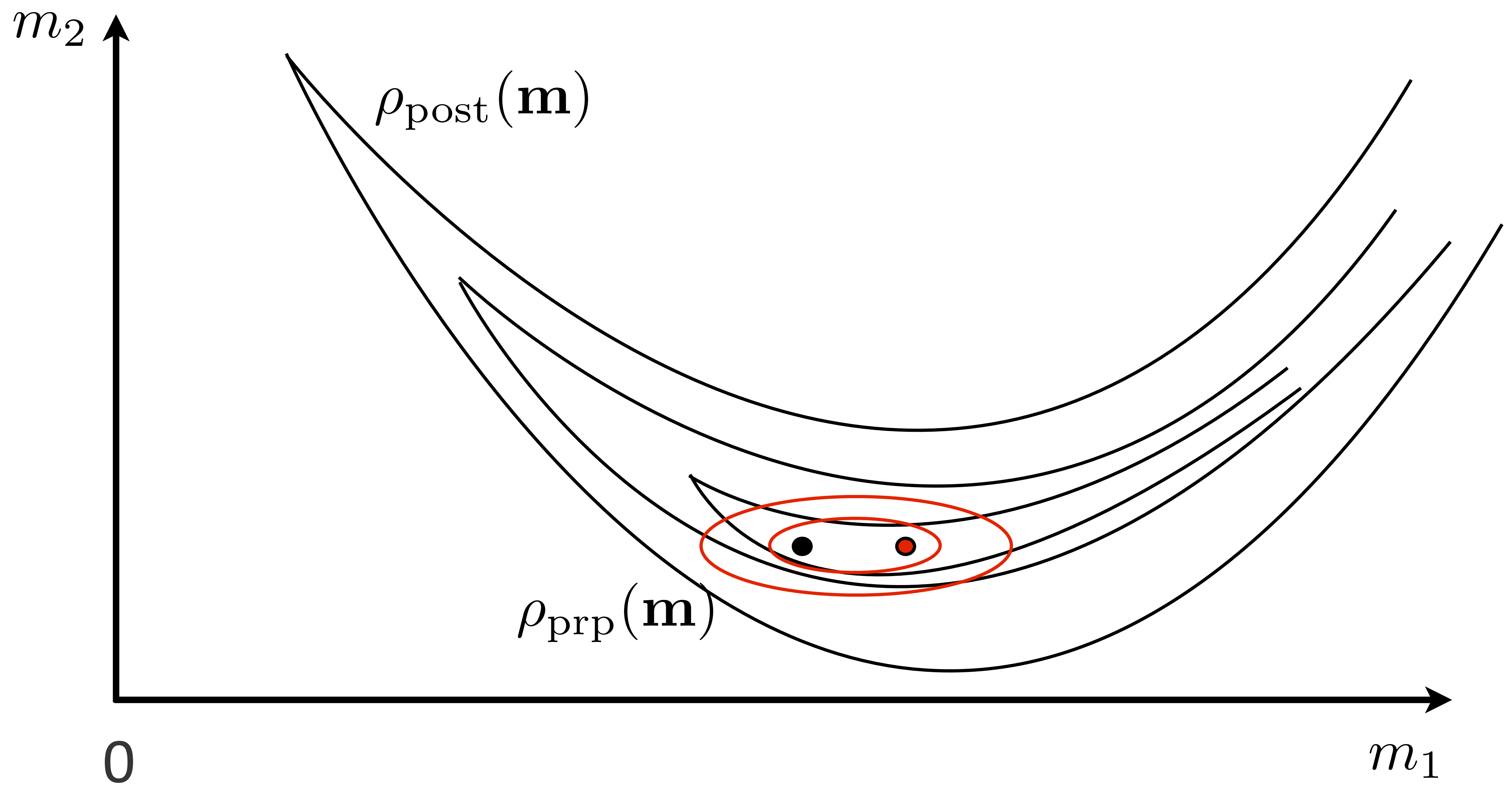
- high-dimensional model space and data space
- non-linear and expensive model-to-data map
- no closed form solution

Bayes w/ strong PDE constraints

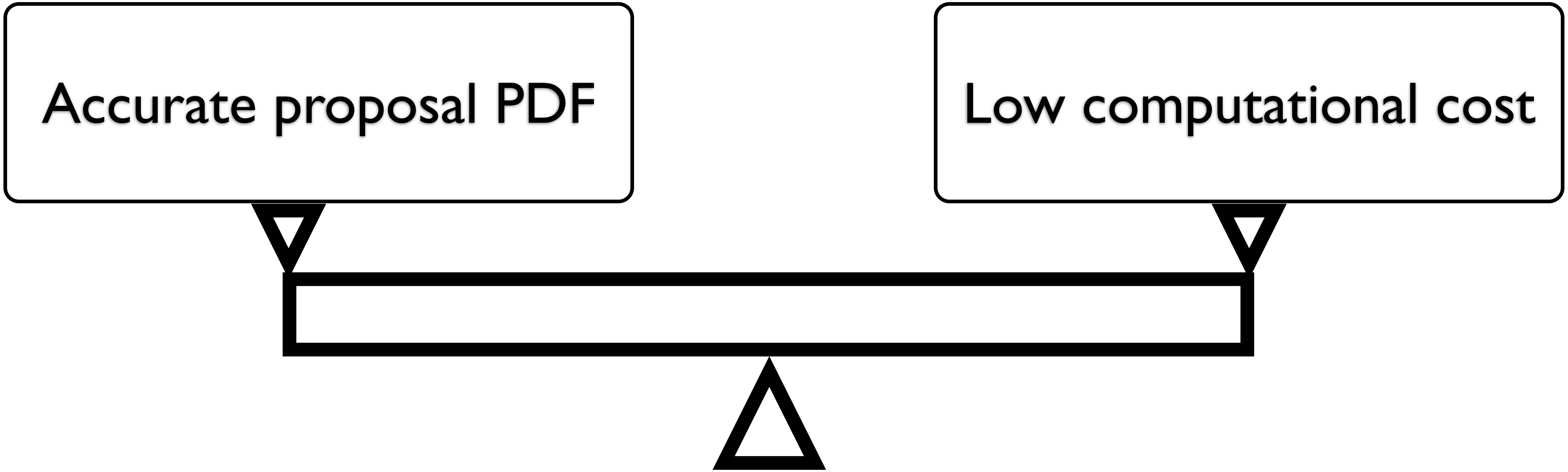
[J. Kaipio and E. Somersalo, 2004]
[A. M. Stuart *et al.*, 2004]
[J. Martin *et al.*, 2012]

McMC type methods:

- Metropolis-Hasting method
  - draw samples with a proposal PDF
- Langevin method
  - construct the proposal PDF with a preconditioning matrix
- Newton type McMC method
  - construct the proposal PDF with local Hessian matrix



Bayes w/ strong PDE constraints



Bayes w/ strong PDE constraints

[J. Kaipio and E. Somersalo, 2004]
[A. M. Stuart *et al.*, 2004]
[J. Martin *et al.*, 2012]
[R. G. Pratt, 1999]

McMC type methods:

- Metropolis-Hasting method
  - draw samples with a proposal PDF
- Langevin method
  - construct the proposal PDF with a preconditioning matrix
- Newton type McMC method
  - construct the proposal PDF with local Hessian matrix
  - Hessian matrix and GN Hessian are dense and require additional PDE solves
    e.g. Gauss-Newton Hessian of $-\log \rho_{\text{like}}(\mathbf{d}|\mathbf{m})$:

$$\mathbf{H}_{\text{GN}}(\mathbf{m}) = \omega^4 \text{diag}\big(\text{conj}(\mathbf{u})\big) \underline{\mathbf{A}(\mathbf{m})^{-\top}\mathbf{P}^{\top}} \mathbf{\Sigma}_{\text{noise}}^{-1} \mathbf{P}\mathbf{A}(\mathbf{m})^{-1} \text{diag}(\mathbf{u})$$

additional PDE solves

PDE constrained optimization problems

[M. Fisher *et al*, 2004]
[T. van Leeuwen and F. J. Herrmann, 2013]
[T. van Leeuwen and F. J. Herrmann, 2015]

SLIM

Original problem:

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

Adjoint-state method – *strong constraint*:

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

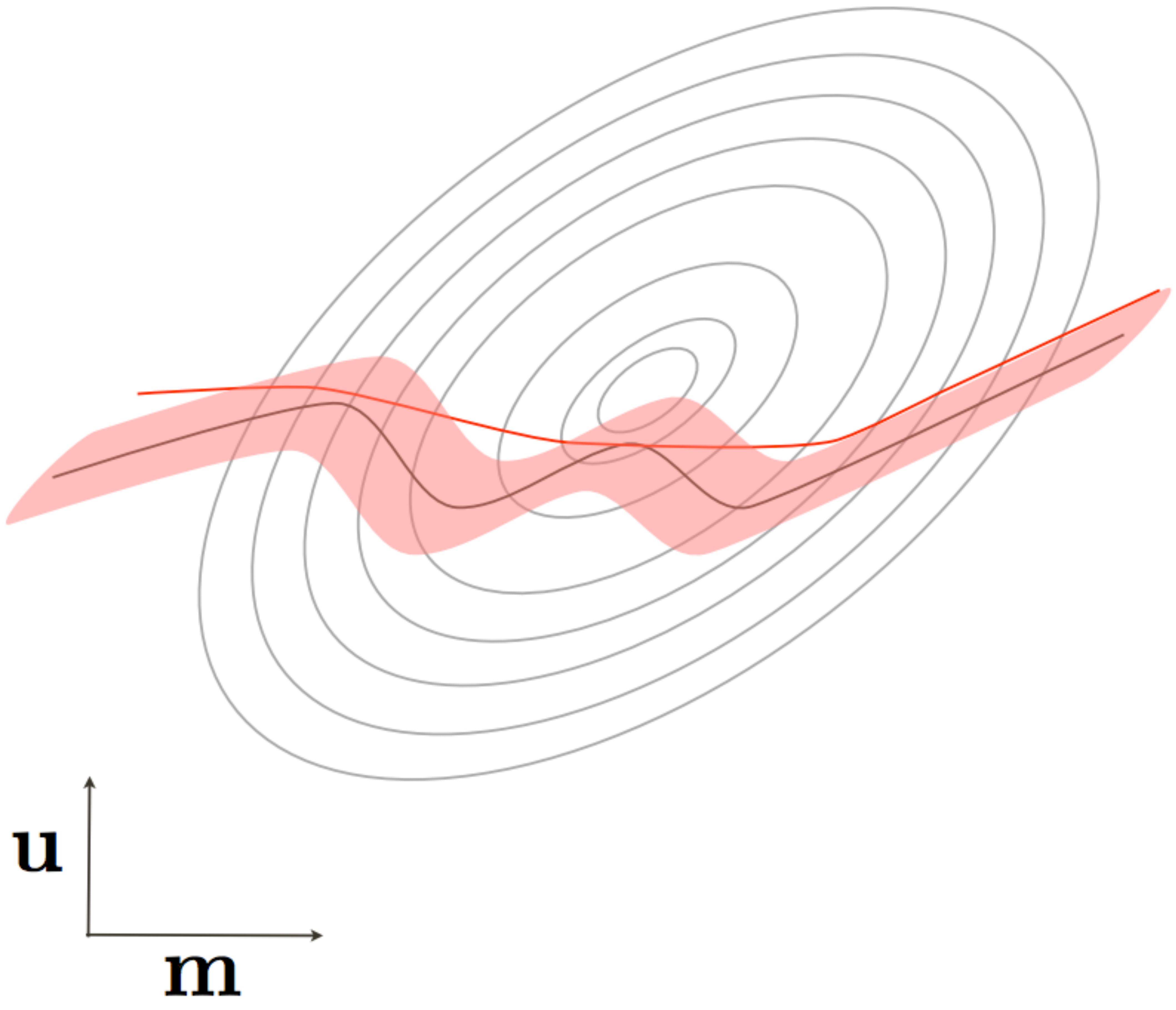
Penalty method – *weak constraint*:

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

Extend the search space

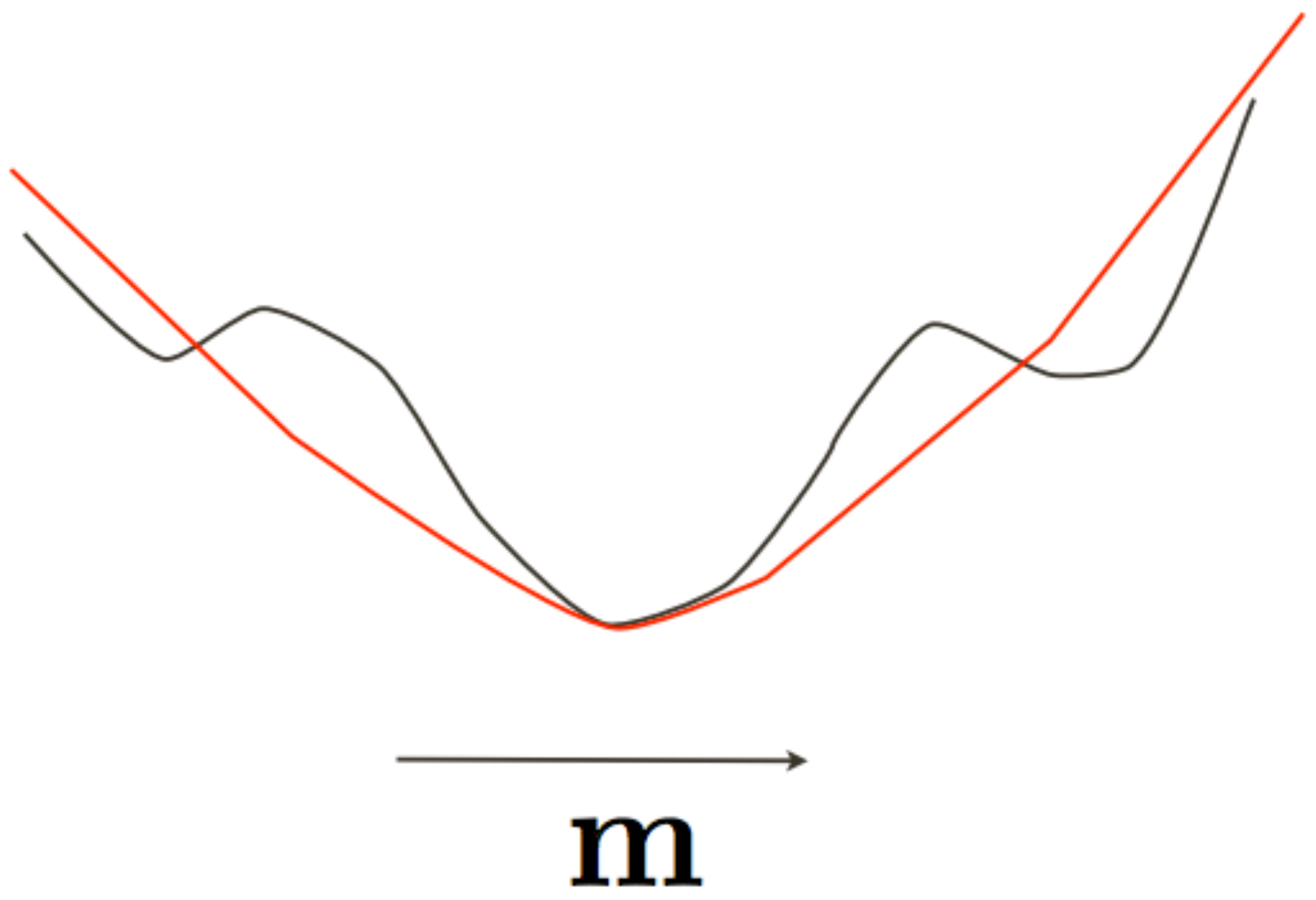
[T. van Leeuwen and F. J. Herrmann, 2013]
[T. van Leeuwen and F. J. Herrmann, 2015]

Larger # of degrees of freedom



"more convex"

less local minima



Extend the search space

Introduce auxiliary variable:

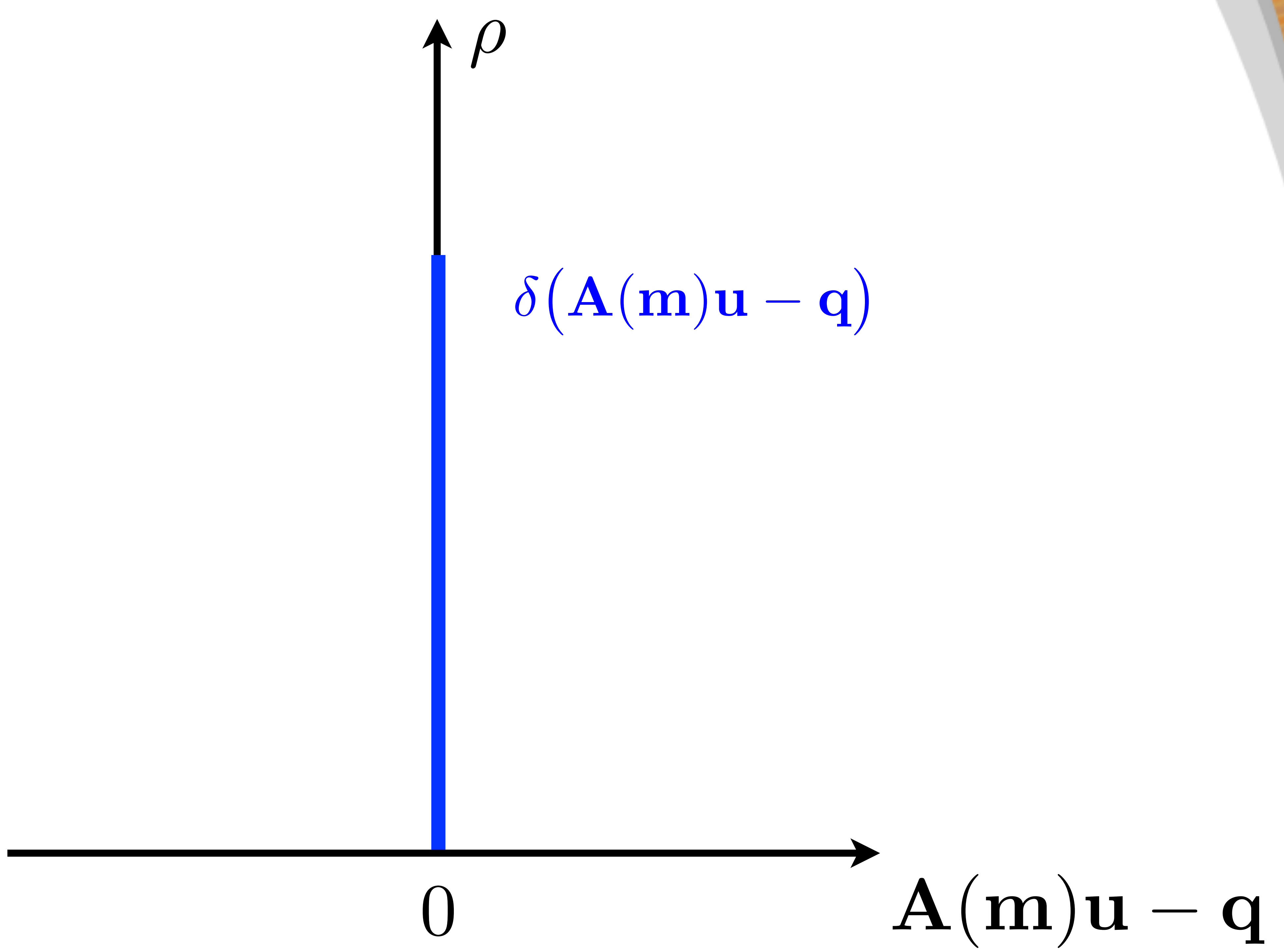
$$\rho_{\text{post}}(\mathbf{u},\mathbf{m}|\mathbf{d}) \propto \rho(\mathbf{d}|\mathbf{u},\mathbf{m})\rho(\mathbf{u},\mathbf{m})$$

$$\rho(\mathbf{u},\mathbf{m}) = \rho(\mathbf{u}|\mathbf{m})\rho_{\text{prior}}(\mathbf{m})$$

Strong PDE constraints:

$$\rho(\mathbf{d}|\mathbf{u},\mathbf{m}) \propto \exp\left(-\frac{1}{2}\|\mathbf{P}\mathbf{u}-\mathbf{d}\|^2_{\mathbf{\Sigma}^{-1}_{\text{noise}}}\right)$$

$$\rho(\mathbf{u}|\mathbf{m}) = \delta\big(\mathbf{A}\left(\mathbf{m}\right)\mathbf{u}-\mathbf{q}\big)$$



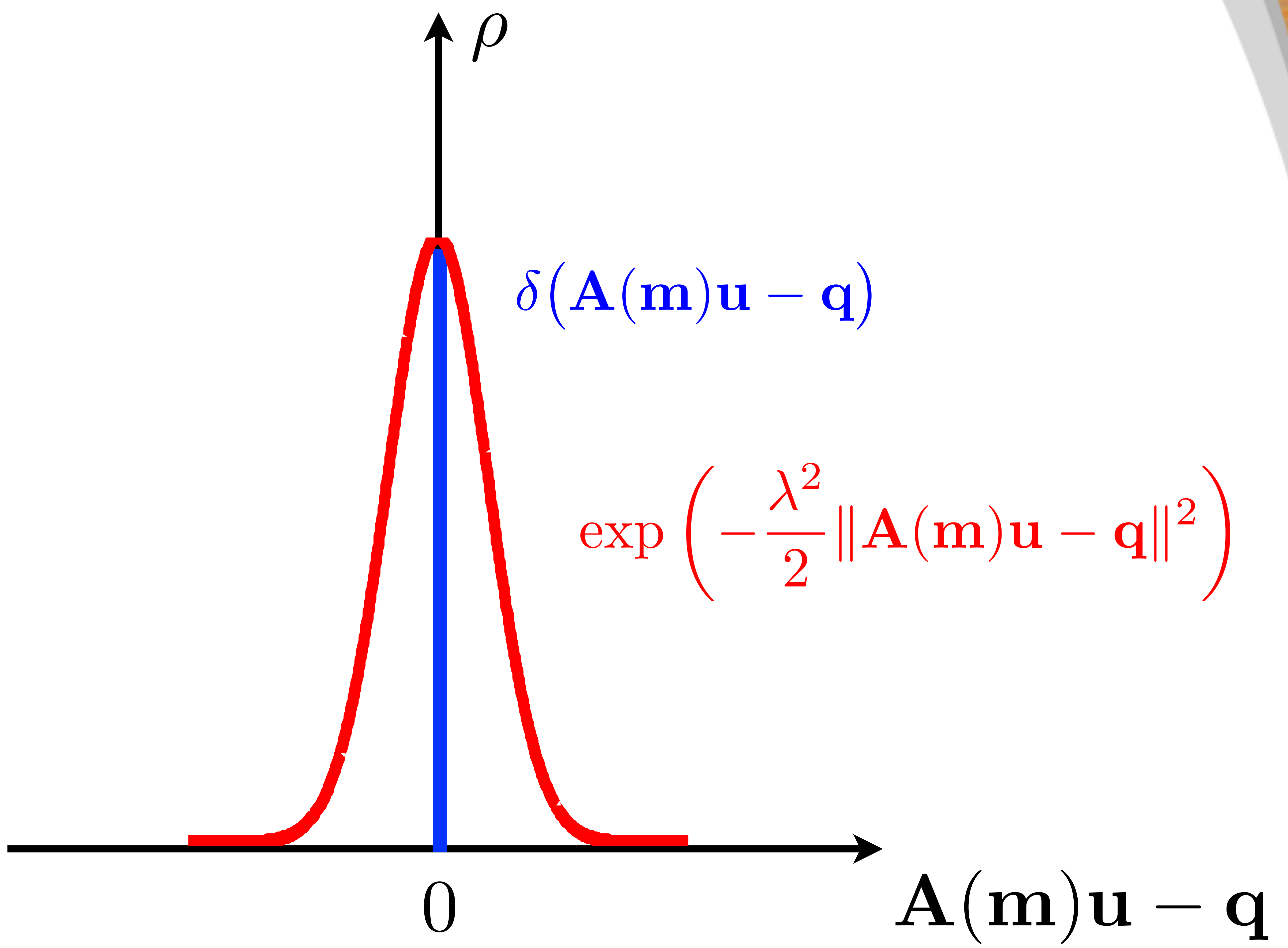
Extend the search space

Weaken PDE constraints:

$$\rho(\mathbf{u}|\mathbf{m}) = \delta\big(\mathbf{A}\left(\mathbf{m}\right)\mathbf{u} - \mathbf{q}\big)$$

↓

$$\rho(\mathbf{u}|\mathbf{m}) \propto \exp\left(-\frac{\lambda^2}{2}\|\mathbf{A}(\mathbf{m})\mathbf{u} - \mathbf{q}\|^2\right)$$



Bayes w/ weak PDE constraints

Joint posterior PDF:

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

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

This is a bi-Gaussian PDF!!

Bayes w/ weak PDE constraints

Joint posterior PDF:

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

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

linear operators

- **For fixed m**

SLIM

Conditional PDF

For fixed $\mathbf{m}$:

$$\mathbf{u} \sim \mathcal{N}(\overline{\mathbf{u}}, \mathbf{H}_{\mathrm{u}}^{-1})$$

with

$$\overline{\mathbf{u}} = \mathbf{H}_{\mathrm{u}}^{-1}\left(\mathbf{P}^{\top}\mathbf{\Sigma}_{\mathrm{noise}}^{-1}\mathbf{d} + \lambda^2\mathbf{A}^{\top}(\mathbf{m})\mathbf{q}\right)$$

$$\mathbf{H}_{\mathrm{u}} = \mathbf{P}^{\top}\mathbf{\Sigma}_{\mathrm{noise}}^{-1}\mathbf{P} + \lambda^2\mathbf{A}^{\top}(\mathbf{m})\mathbf{A}(\mathbf{m})$$

SLIM

Bayes w/ weak PDE constraints

Joint posterior PDF:

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

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

- **For fixed m**
- **For fixed u**

$$\boxed{\omega^2 \text{diag}(\mathbf{m})\mathbf{u} + \Delta\mathbf{u}}$$

$$\boxed{\omega^2 \text{diag}(\mathbf{u})\mathbf{m} + \Delta\mathbf{u}}$$

linear in m

SLIM

Conditional PDF

For fixed $\mathbf{u}$:

$$\mathbf{m} \sim \mathcal{N}(\overline{\mathbf{m}}, \mathbf{H}_{\mathrm{m}}^{-1})$$

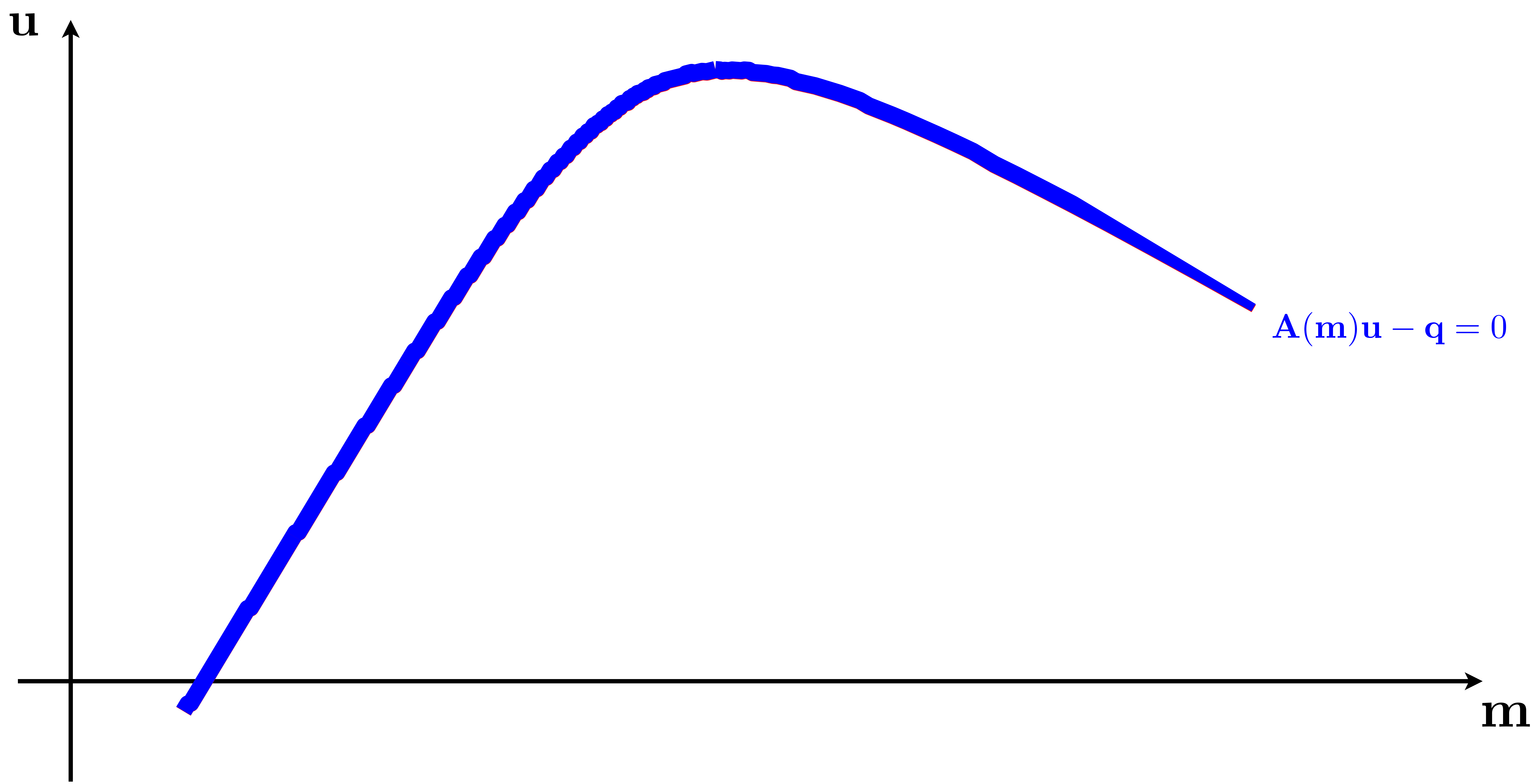
with

$$\overline{\mathbf{m}} = \mathbf{H}_{\mathrm{m}}^{-1}\Big(\lambda^2\omega^2\mathrm{diag}\big(\mathrm{conj}(\mathbf{u})\big)(\mathbf{q} - \Delta\mathbf{u}) + \mathbf{\Sigma}_{\mathrm{prior}}^{-1}\mathbf{m}_{\mathrm{prior}}\Big)$$

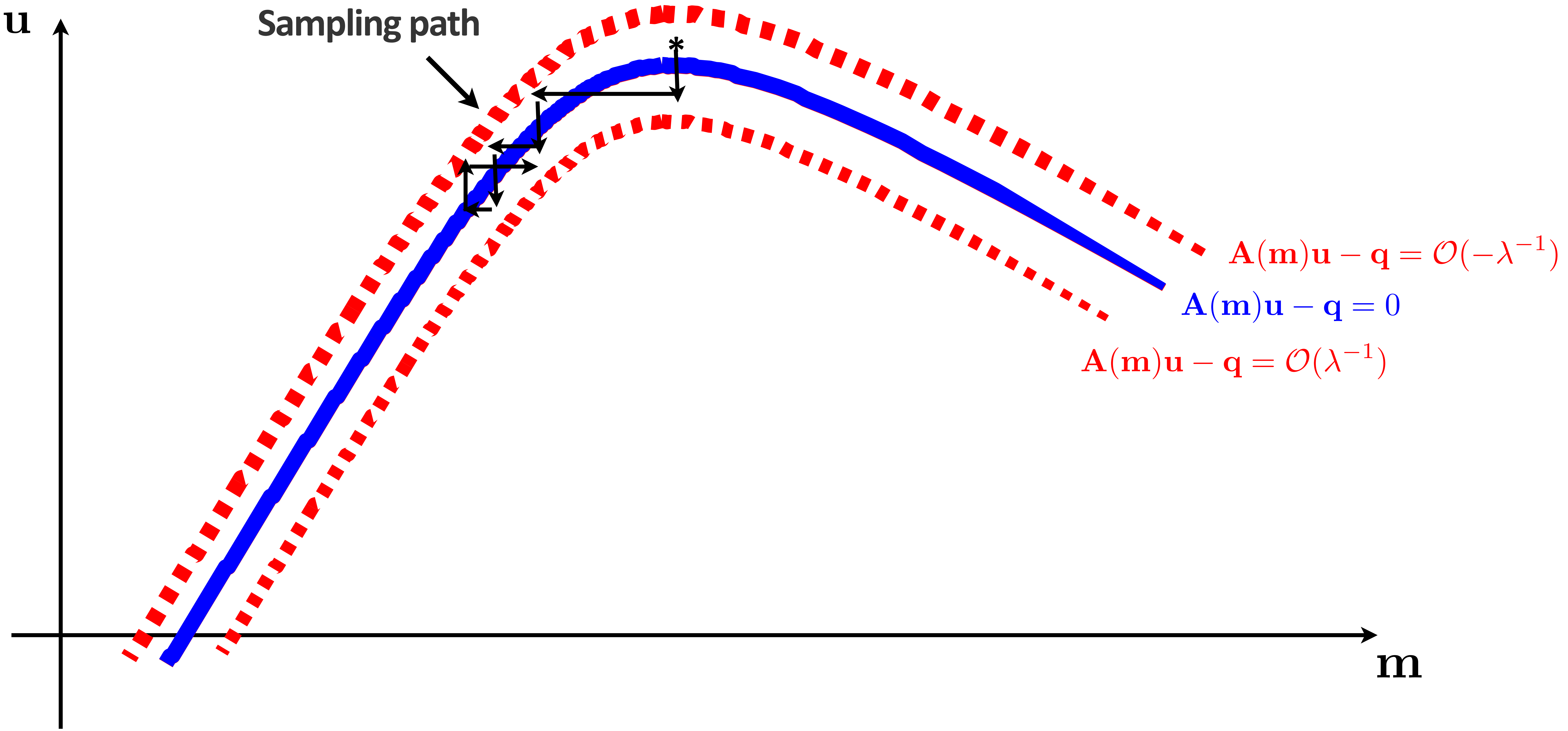
$$\mathbf{H}_{\mathrm{m}} = \boxed{\lambda^2\omega^4\mathrm{diag}\big(\mathrm{conj}(\mathbf{u})\big)\mathrm{diag}(\mathbf{u})} + \mathbf{\Sigma}_{\mathrm{prior}}^{-1}$$

Diagonal matrix

Gibbs sampling



Gibbs sampling



Bayes w/ weak PDE constraints

Gibbs sampling:

1. start with $\mathbf{m}_0$
2. for k = 1 : n_{smp}
3. compute $\overline{\mathbf{u}}(\mathbf{m}_k)$ and $\mathbf{H}_{\mathrm{u}}(\mathbf{m}_k)$
4. draw $\mathbf{u}_k$ from $\mathcal{N}\left(\overline{\mathbf{u}}(\mathbf{m}_k), \mathbf{H}_{\mathrm{u}}^{-1}(\mathbf{m}_k)\right)$
5. compute $\overline{\mathbf{m}}(\mathbf{u}_k)$ and $\mathbf{H}_{\mathrm{m}}(\mathbf{u}_k)$
6. draw $\mathbf{m}_{k+1}$ from $\mathcal{N}\left(\overline{\mathbf{m}}(\mathbf{u}_k), \mathbf{H}_{\mathrm{m}}^{-1}(\mathbf{u}_k)\right)$
7. end

Main computational cost

Computational cost

Drawing $\mathbf{u}_k \sim \mathcal{N}\left(\overline{\mathbf{u}}(\mathbf{m}_k), \mathbf{H}_{\mathrm{u}}^{-1}(\mathbf{m}_k)\right)$**:**

Step 1	$\mathbf{H}_{\mathrm{u}}(\mathbf{m}_k) = \mathbf{R}^\top \mathbf{R}$	$\mathcal{O}(n_{\mathrm{grid}}^3)$
Step 2	$\overline{\mathbf{u}}(\mathbf{m}_k) = \mathbf{R}^{-1}\mathbf{R}^{-\top}\left(\mathbf{P}^\top \mathbf{\Sigma}_{\mathrm{noise}}^{-1/2}\mathbf{d} + \lambda^2 \mathbf{A}^\top(\mathbf{m}_k)\mathbf{q}\right)$	$\mathcal{O}(n_{\mathrm{grid}}^2)$
Step 3	$\mathbf{u}_k = \overline{\mathbf{u}}(\mathbf{m}_k) + \mathbf{R}^{-1}\epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I})$	$\mathcal{O}(n_{\mathrm{grid}}^2)$

Cost per each source = solving one PDE

Weak constraint v.s. Strong constraint

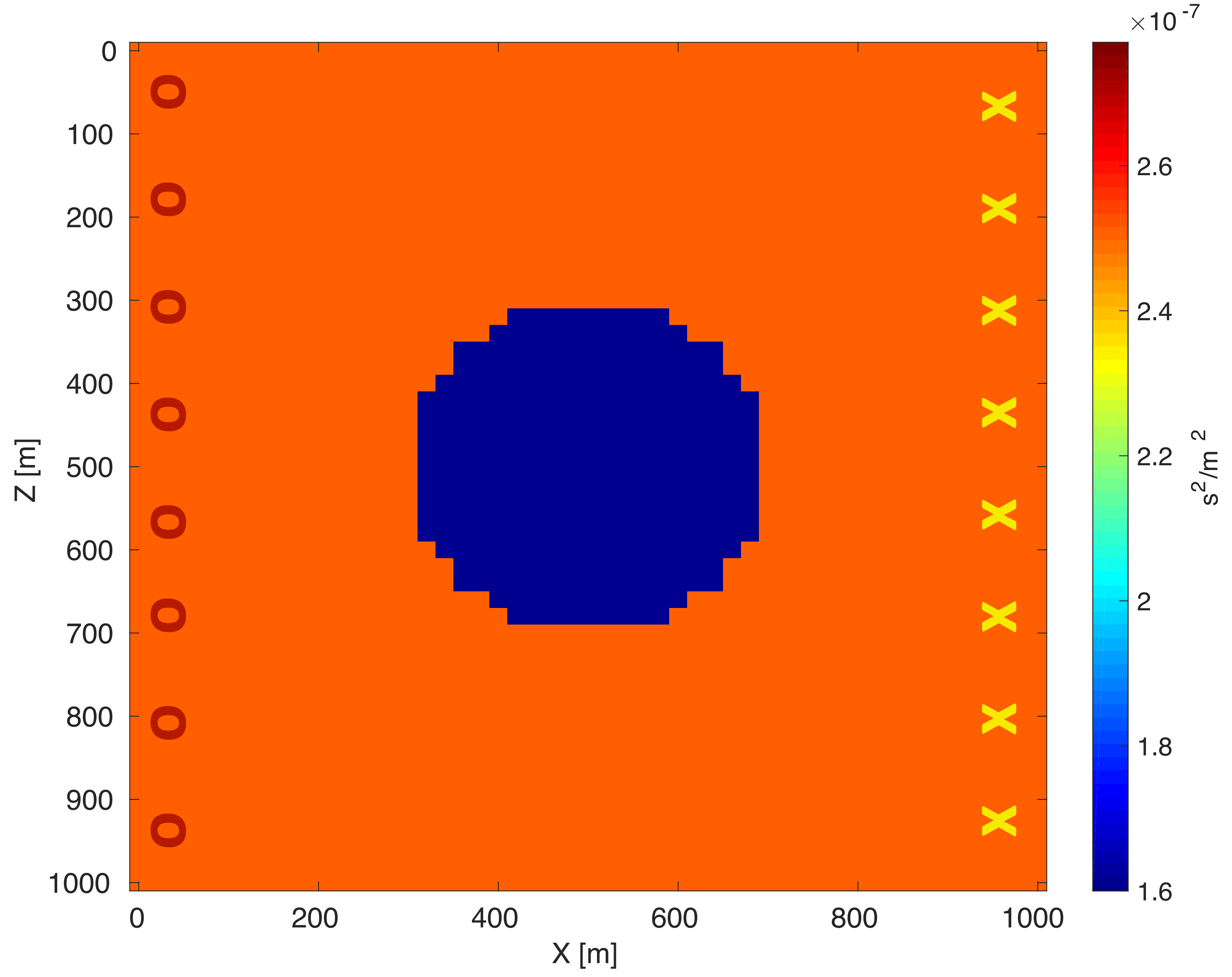
Weak constraint w/ Gibbs sampling method

- joint bi-Gaussian distribution
- sparse Hessian matrix
- constructing Hessian matrix does not require additional PDE solves
- draw samples in a straight-forward manner from the Gaussian conditional distributions

Strong constraint w/ M-H type method

- no special structure
- dense Hessian matrix
- constructing Hessian matrix requires additional PDE solves
- draw samples from certain proposal distribution with accept-reject criteria
- constructing the proposal distribution may require more PDE solves

Numerical example - transmission case



True model

Model size: 1000m x 1000m

Grid size: 20m x 20m

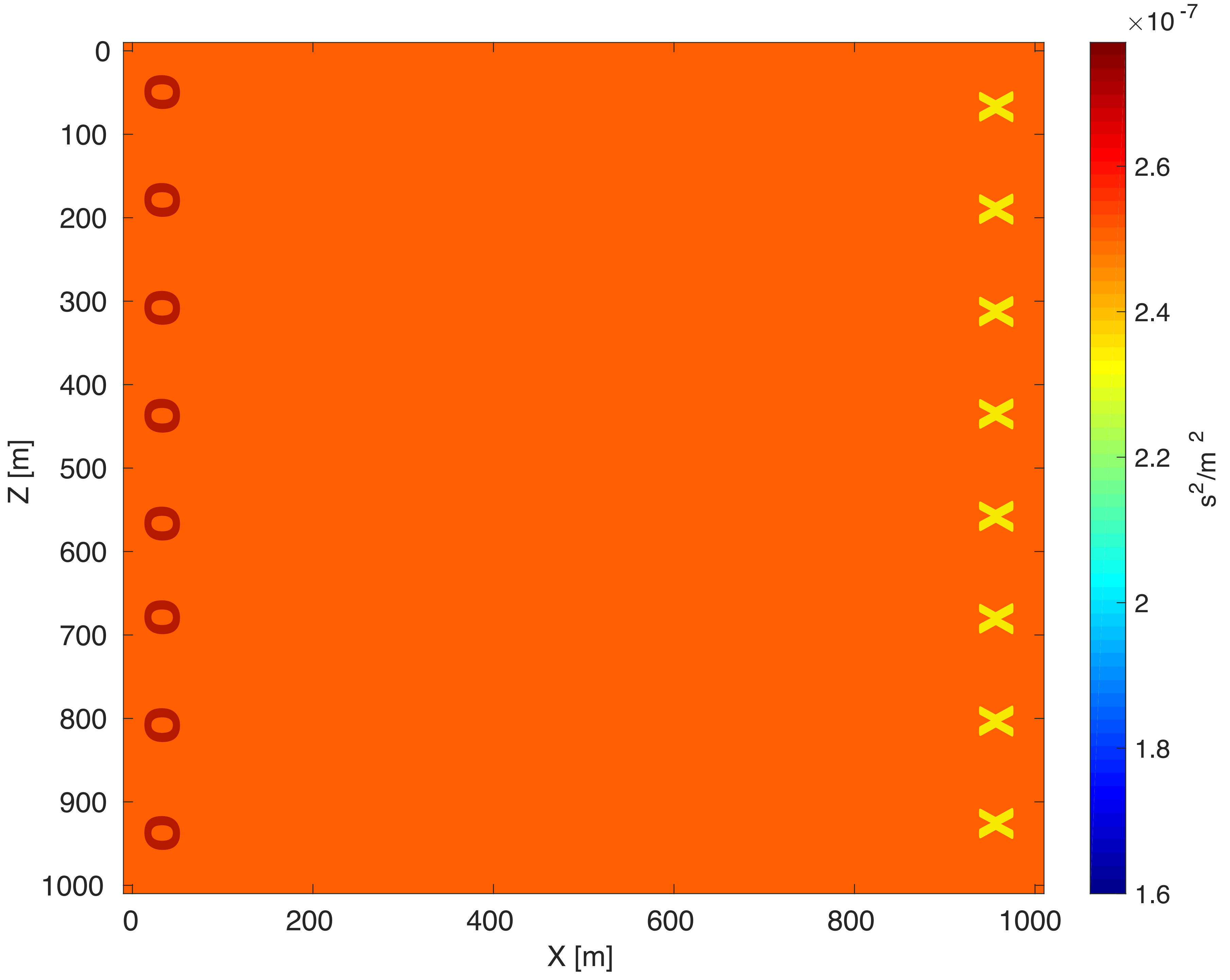
Frequency: [2, 6, 10, 14] Hz

Number of sources: 51

Number of receivers: 51

Number of samplers: 1e6

Numerical example - transmission case



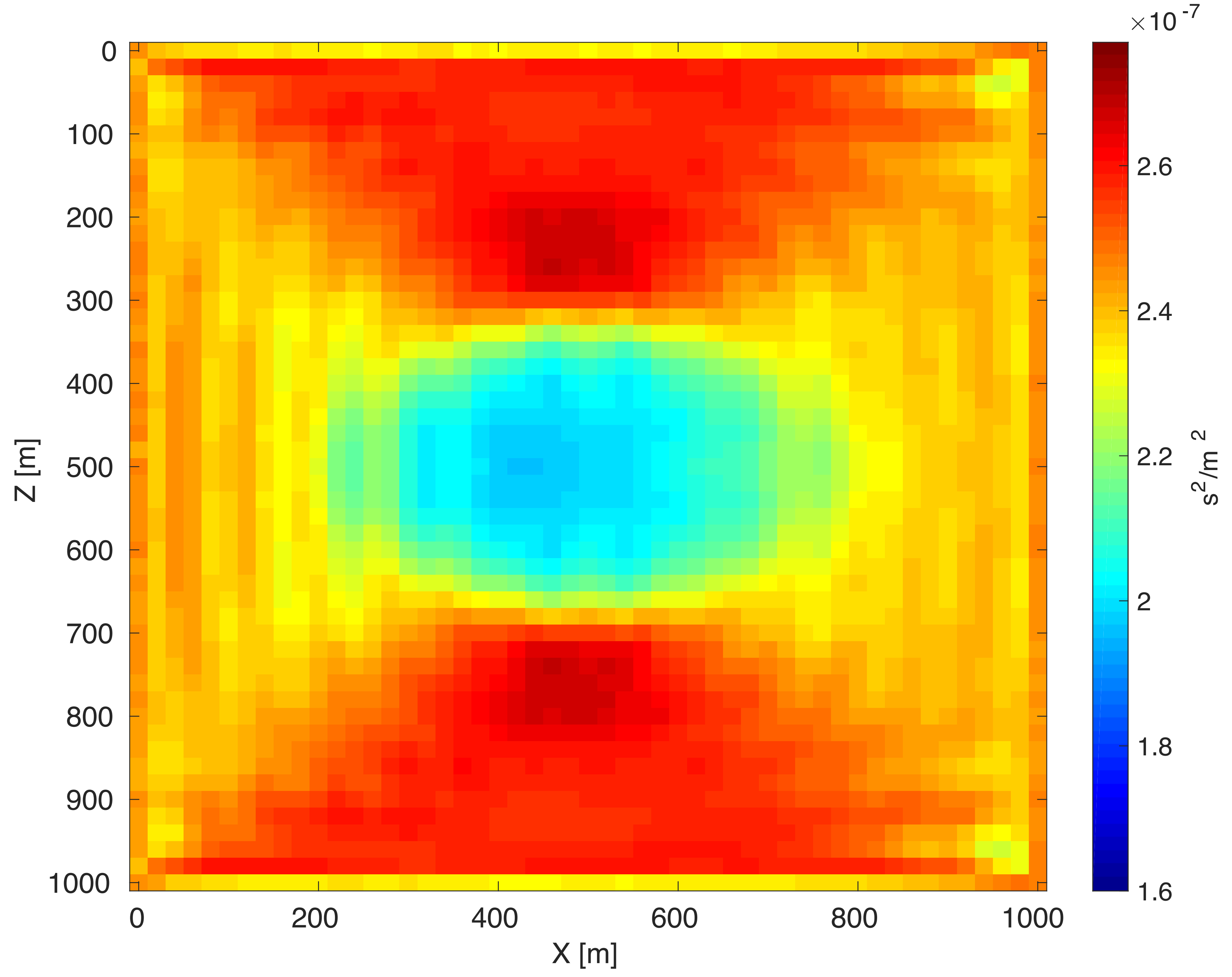
Prior mean model

$$\mathbf{\Sigma}_{\text{prior}} = \sigma^2_{\text{prior}}\mathbf{I}, \sigma_{\text{prior}} = 1e\text{-}8\text{s}^2/\text{m}^2$$

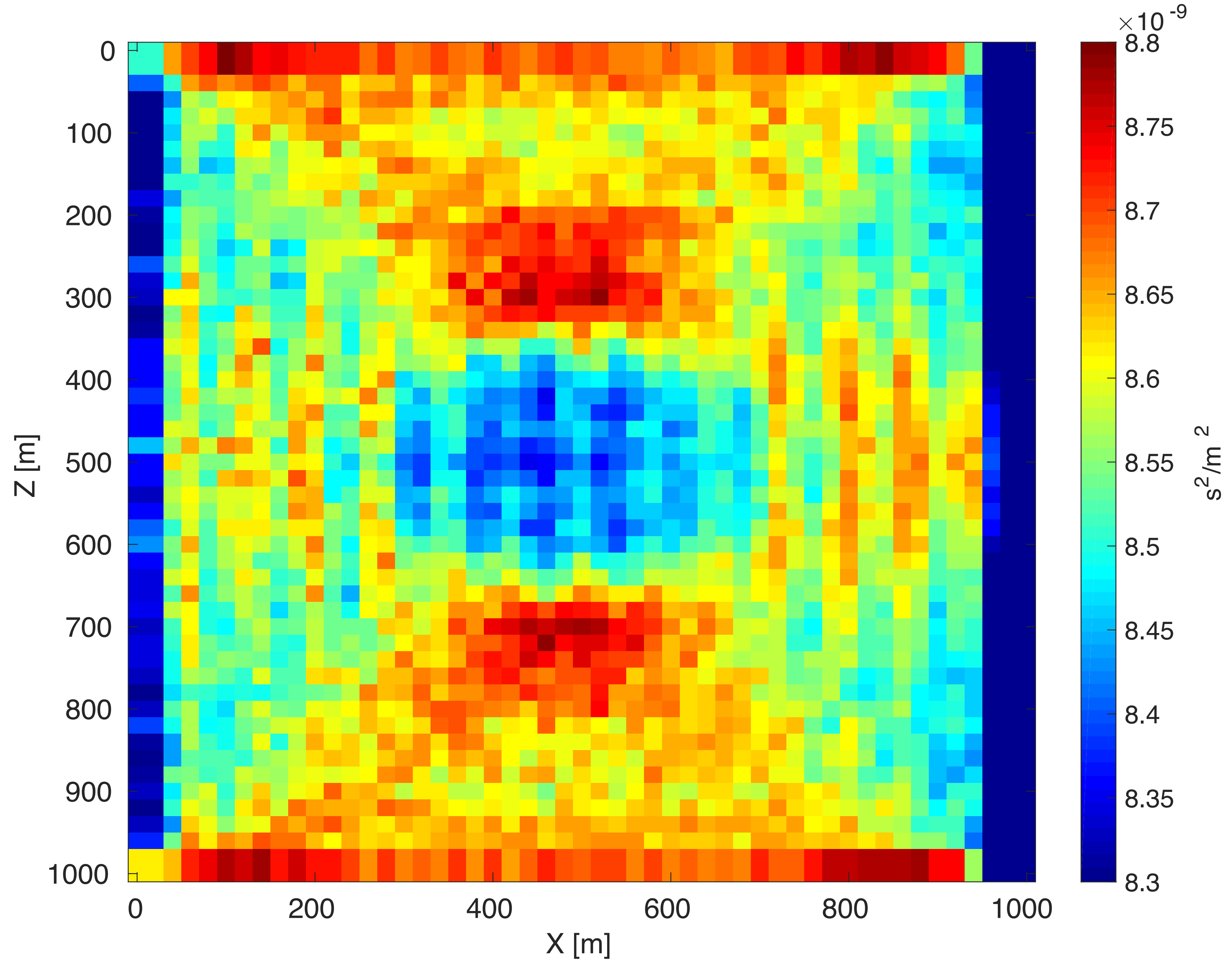
$$\mathbf{\Sigma}_{\text{noise}} = \sigma^2_{\text{noise}}\mathbf{I}, \sigma_{\text{noise}} = 1e0$$

$$\lambda = 1e4$$

Posterior mean and STD

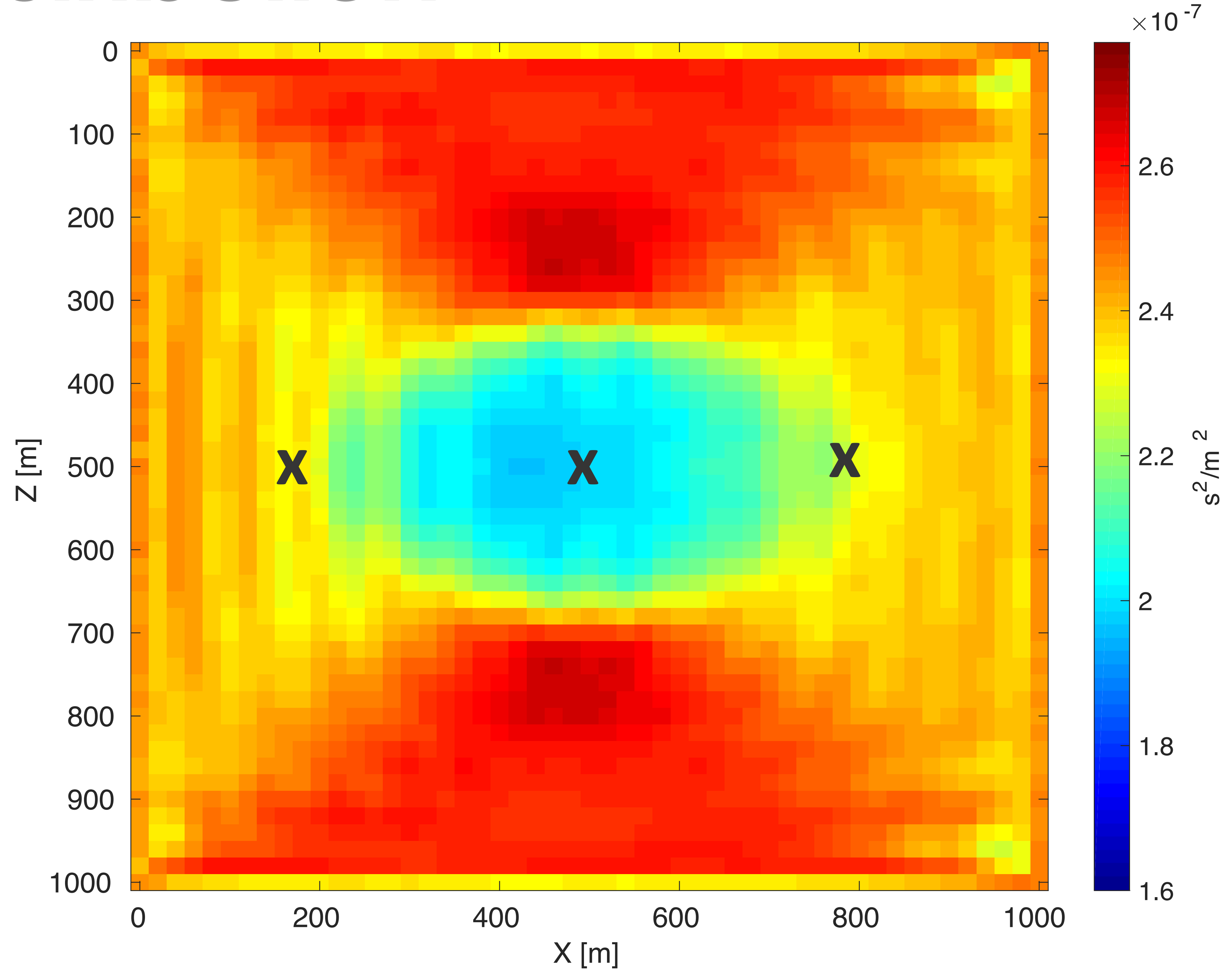


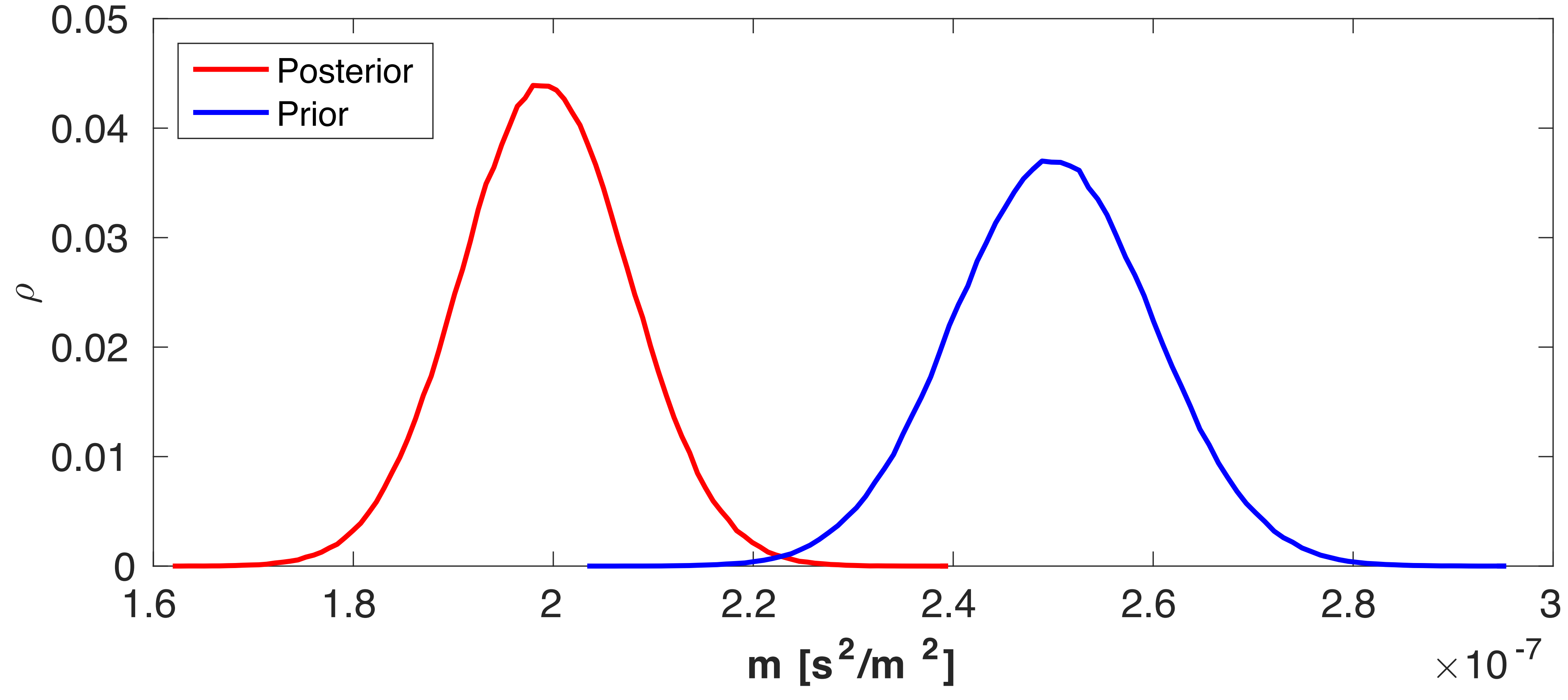
Posterior mean model



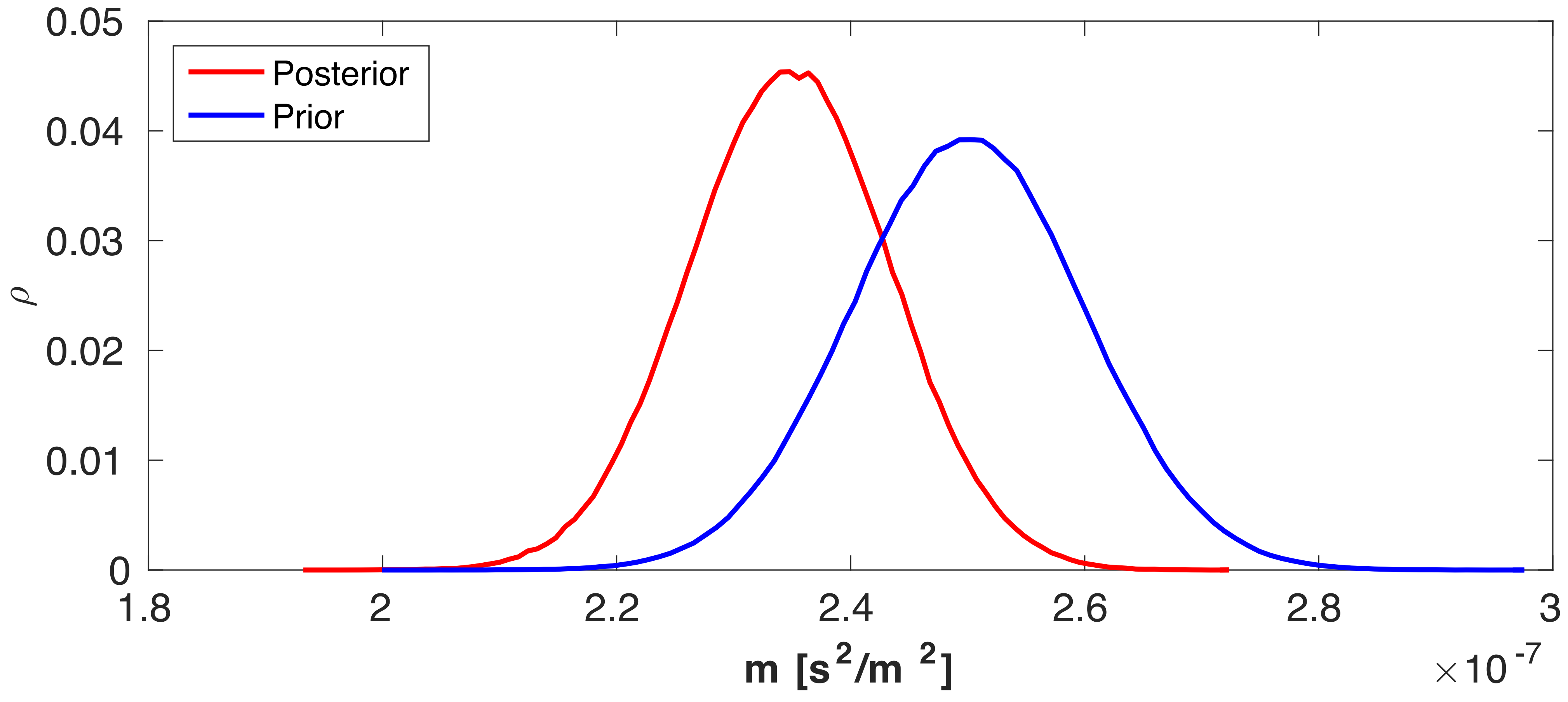
Posterior STD

Distribution

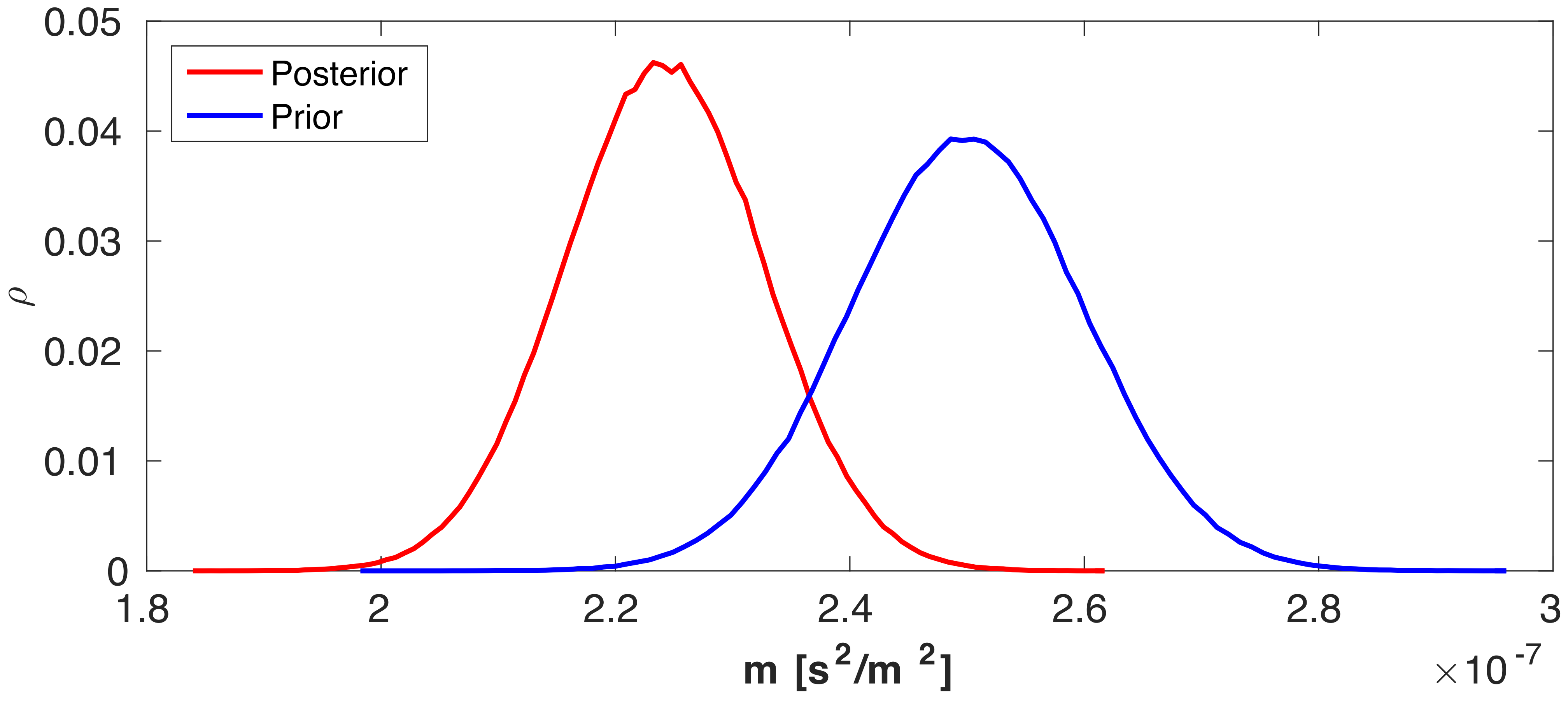




x = 500m z = 500m

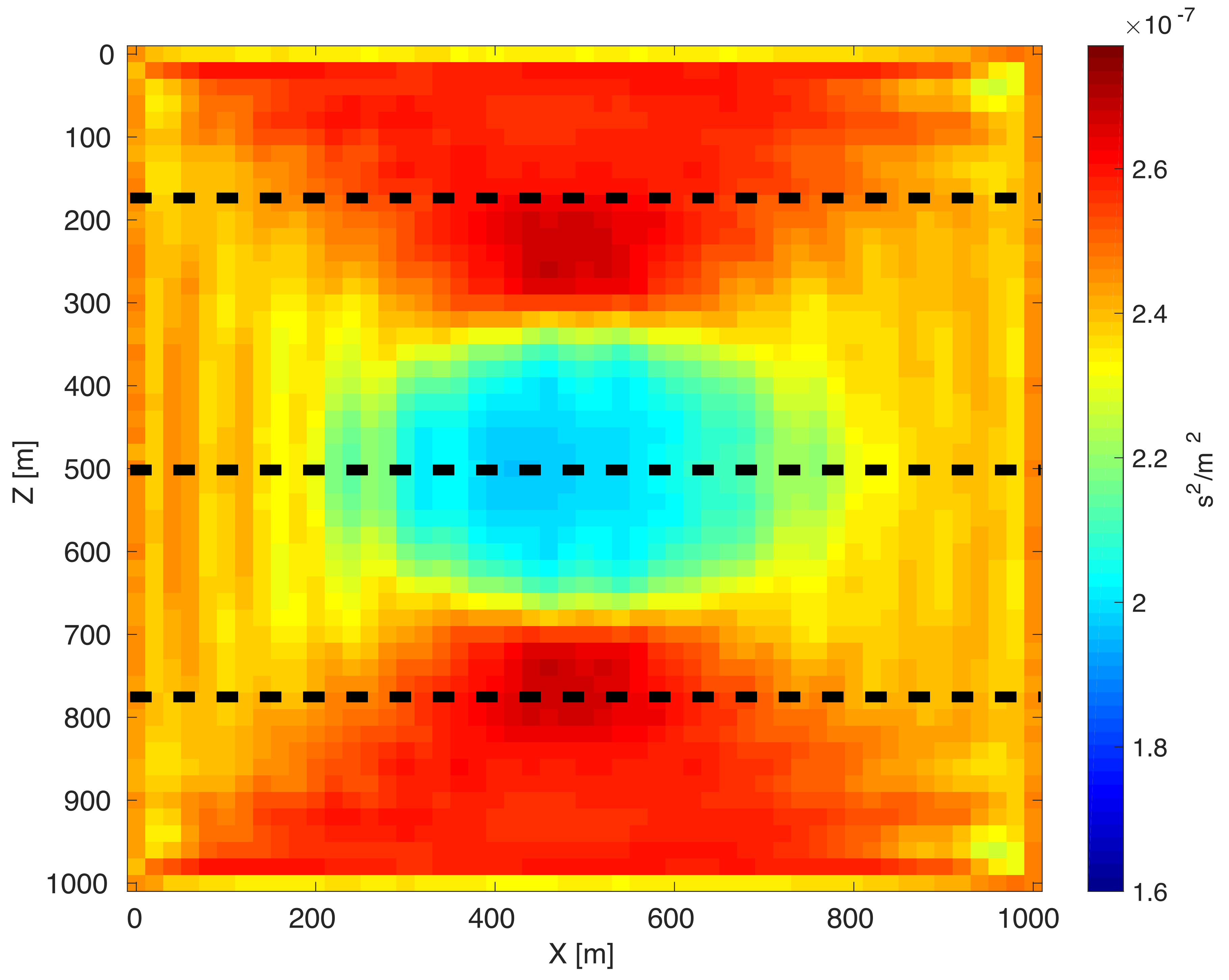


x = 180m z = 500m

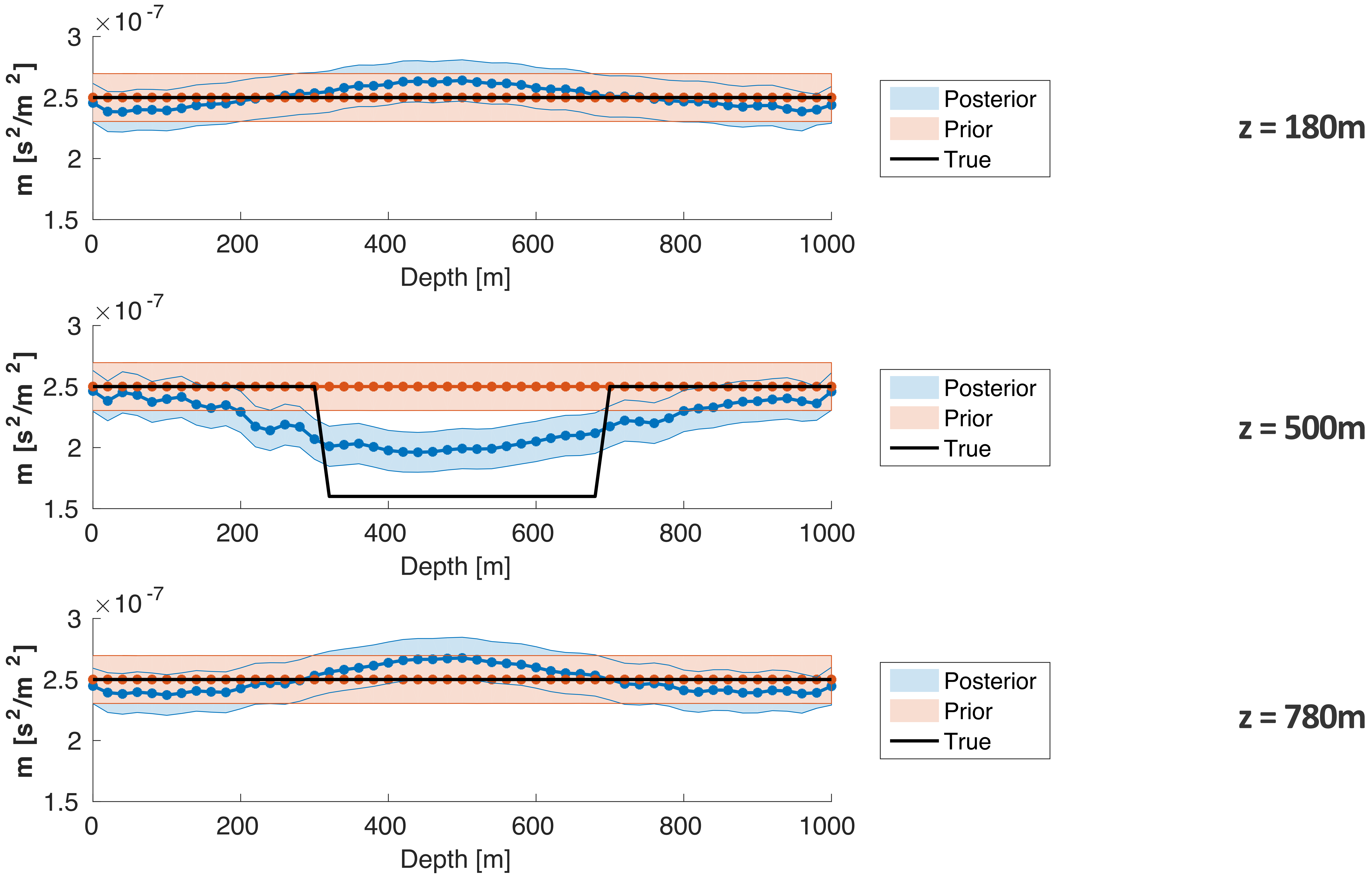


x = 780m z = 500m

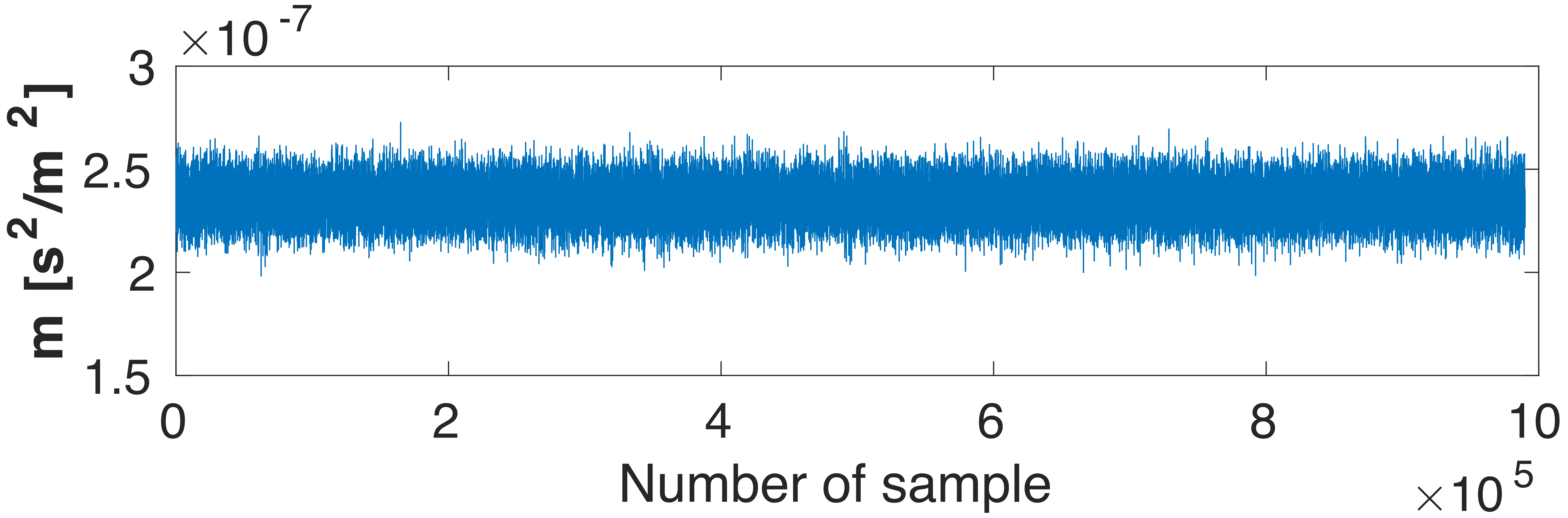
95% Confidence interval



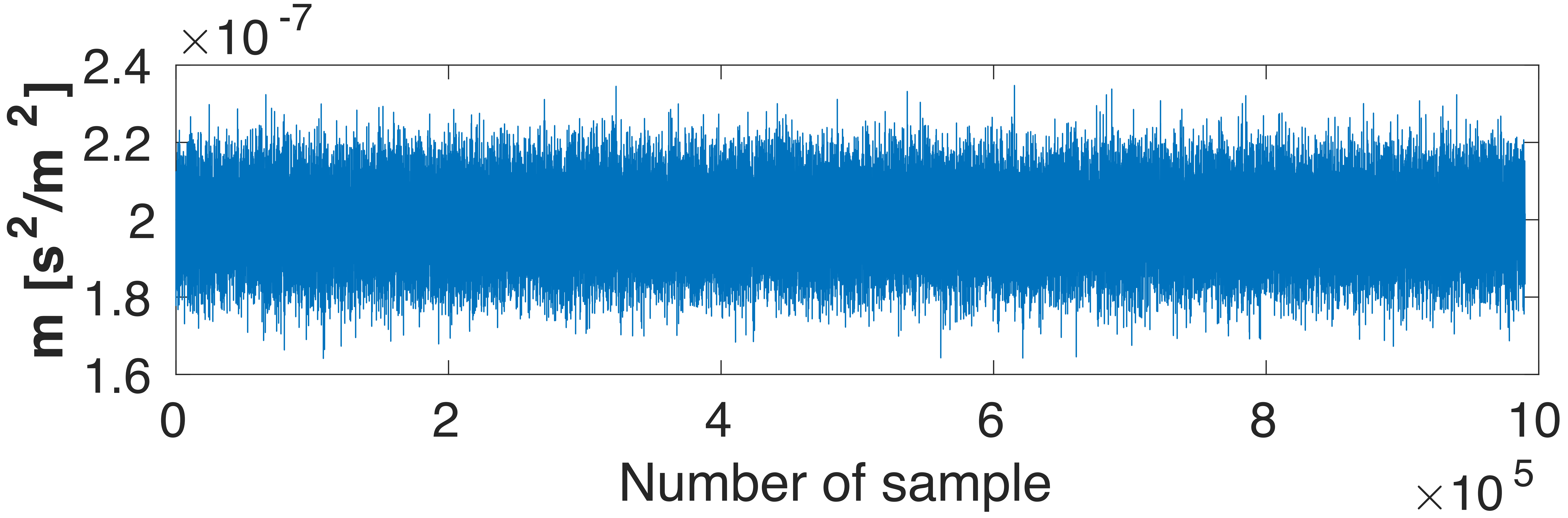
95% Confidence interval



Samples

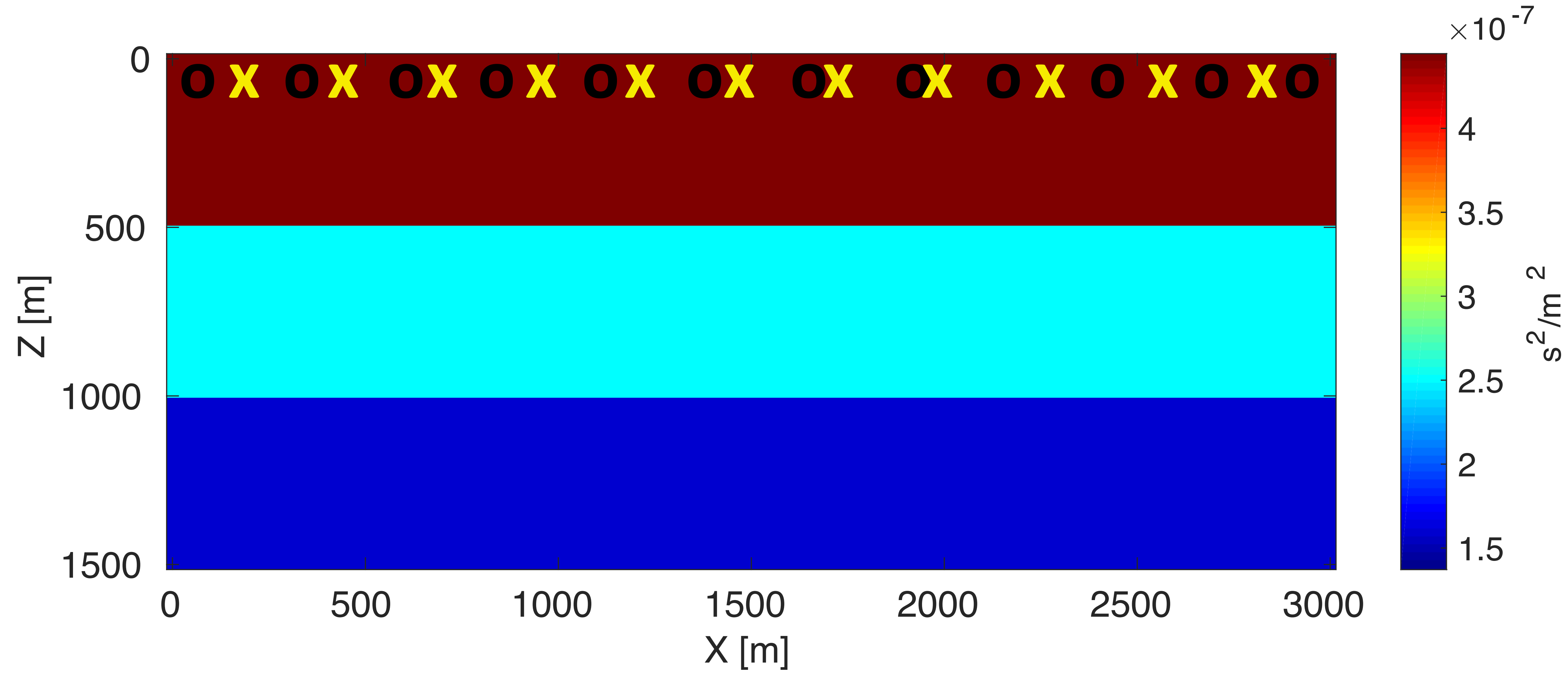


x = 180m z = 500m



x = 500m z = 500m

Numerical example - reflection case



True model

Model size: 1500m x 3000m

Grid size: 30m x 30m

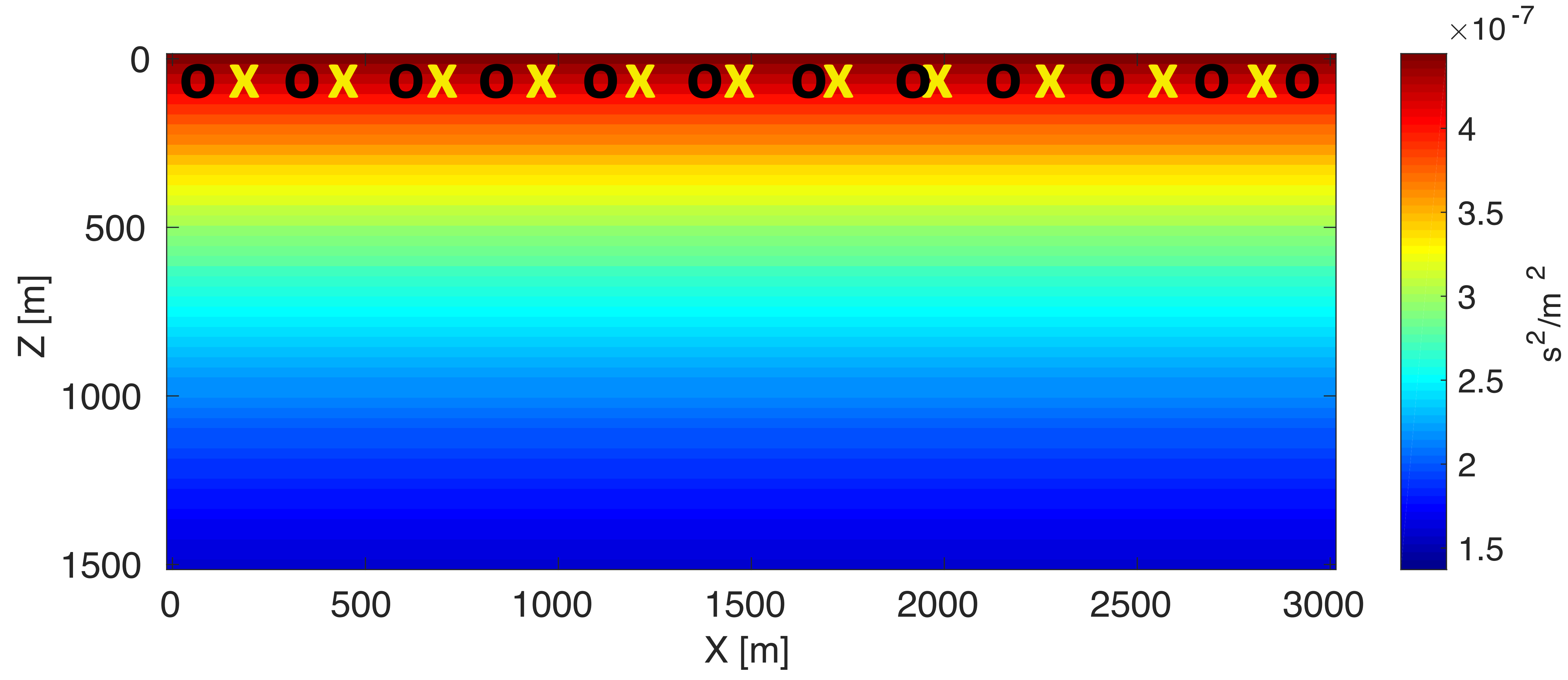
Frequency: [2, 4, 6, 8] Hz

Number of sources: 21

Number of receivers: 101

Number of samplers: 5e5

Numerical example - reflection case



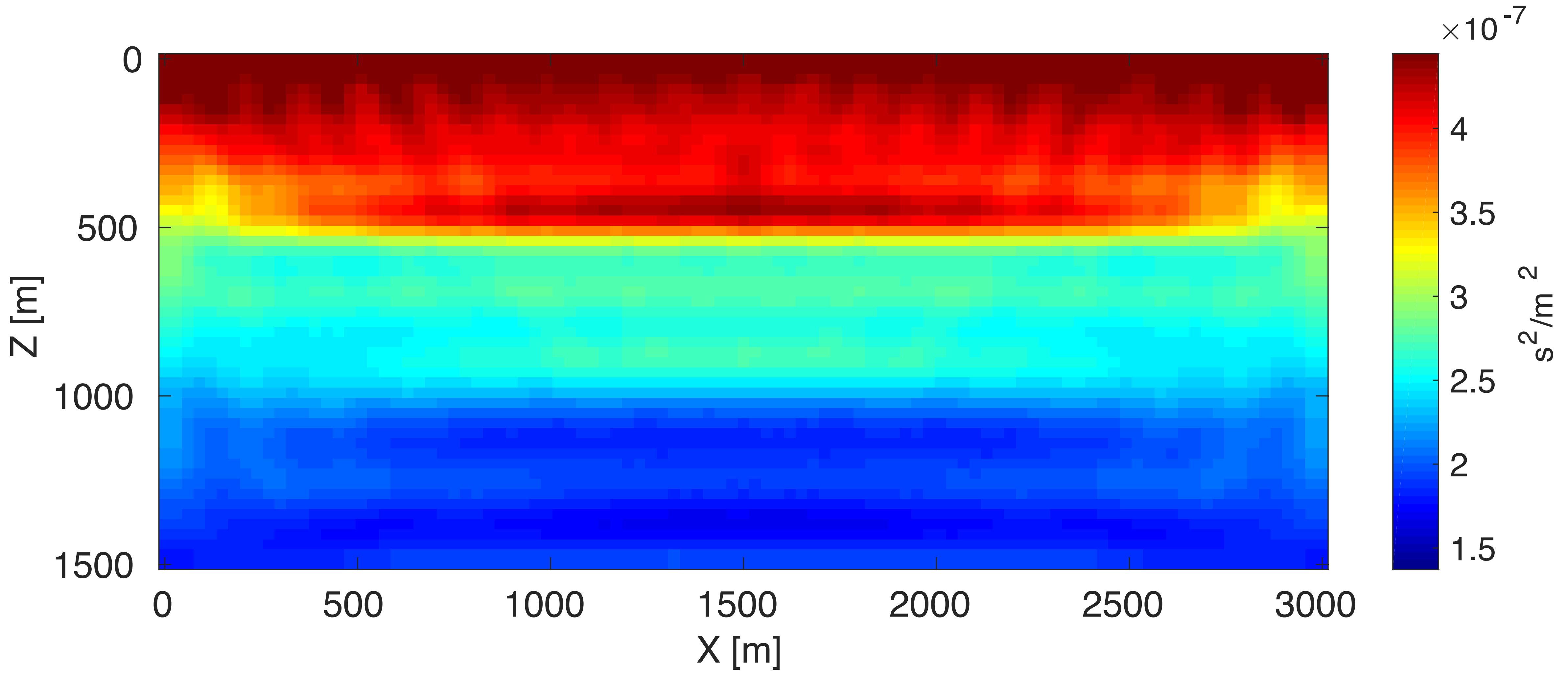
Prior mean model

$$\boldsymbol{\Sigma}_{\text{prior}} = \sigma^2_{\text{prior}}\mathbf{I}, \sigma_{\text{prior}} = 3e\text{-}8\text{s}^2/\text{m}^2$$

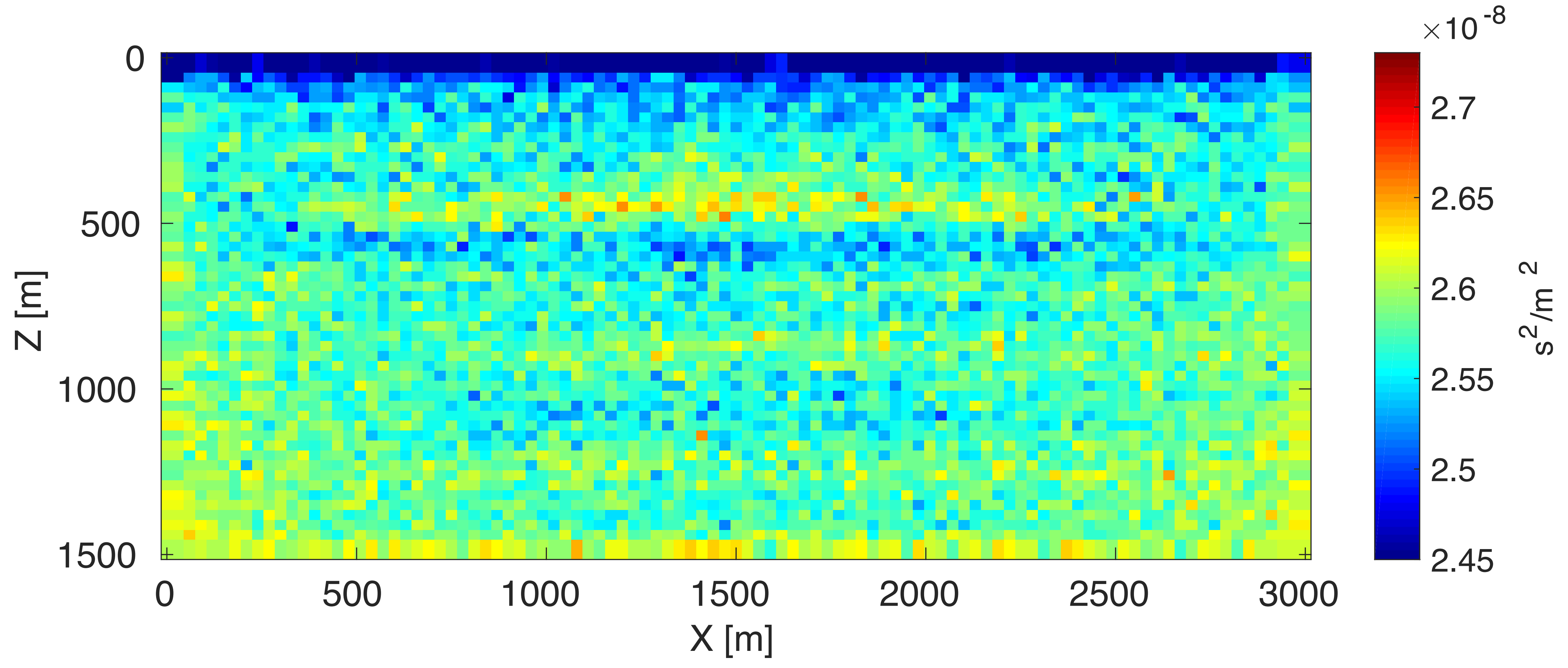
$$\boldsymbol{\Sigma}_{\text{noise}} = \sigma^2_{\text{noise}}\mathbf{I}, \sigma_{\text{noise}} = 1e1$$

$$\lambda = 1e3$$

Posterior mean and STD

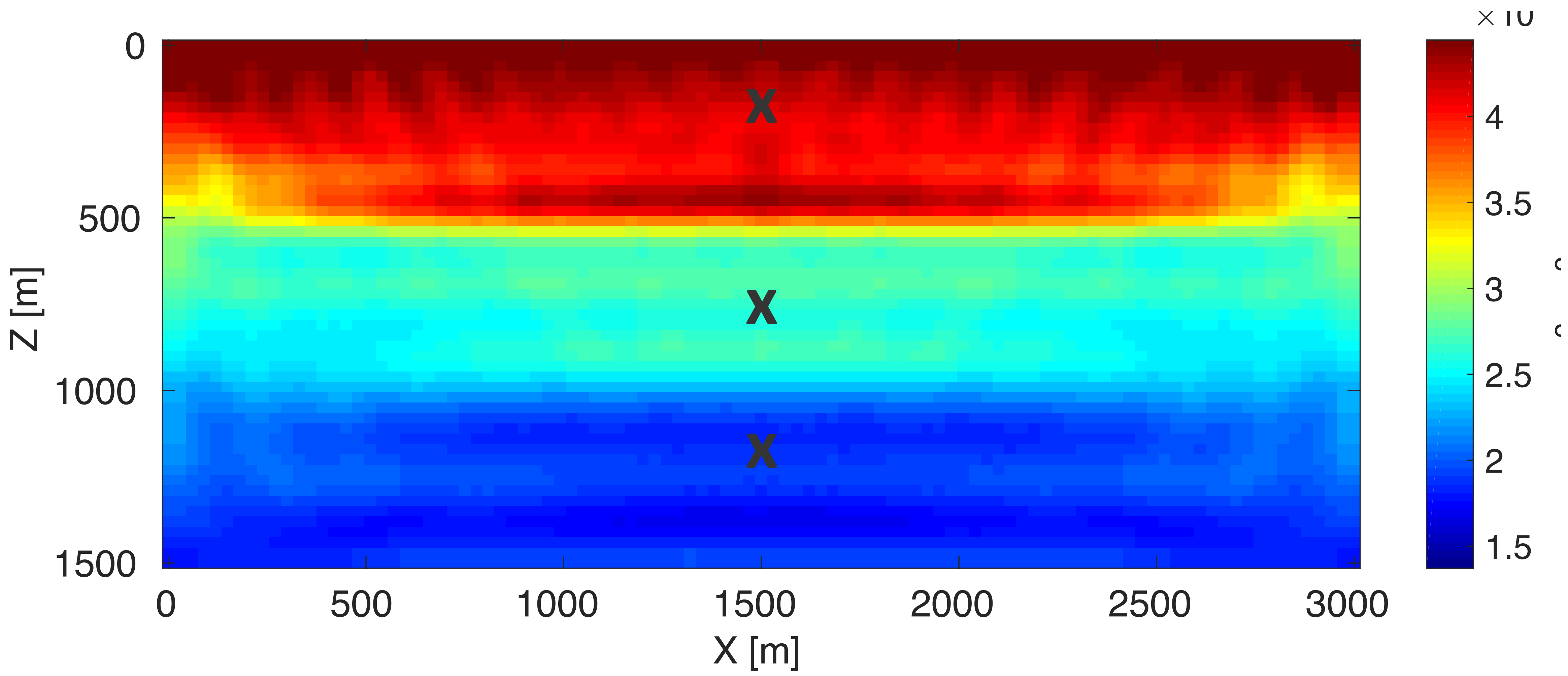


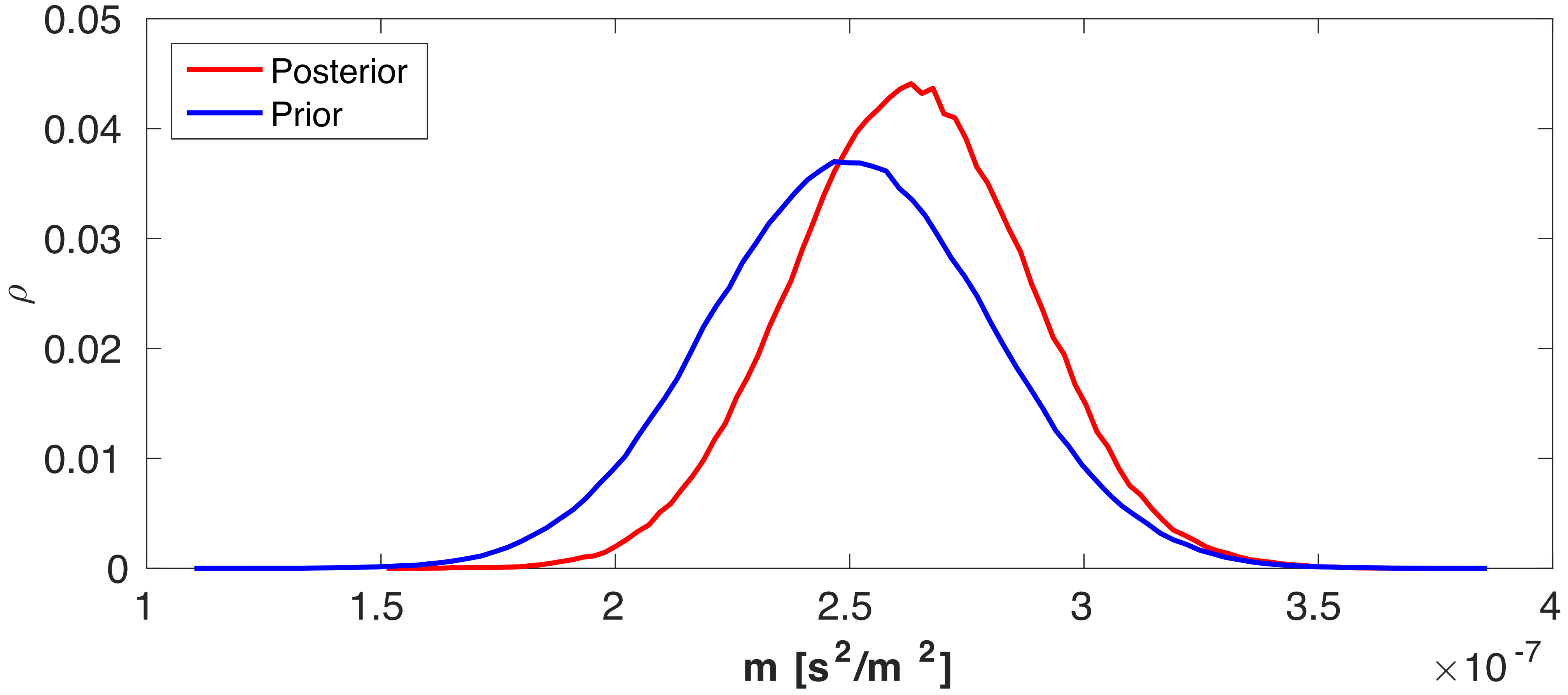
Posterior mean model



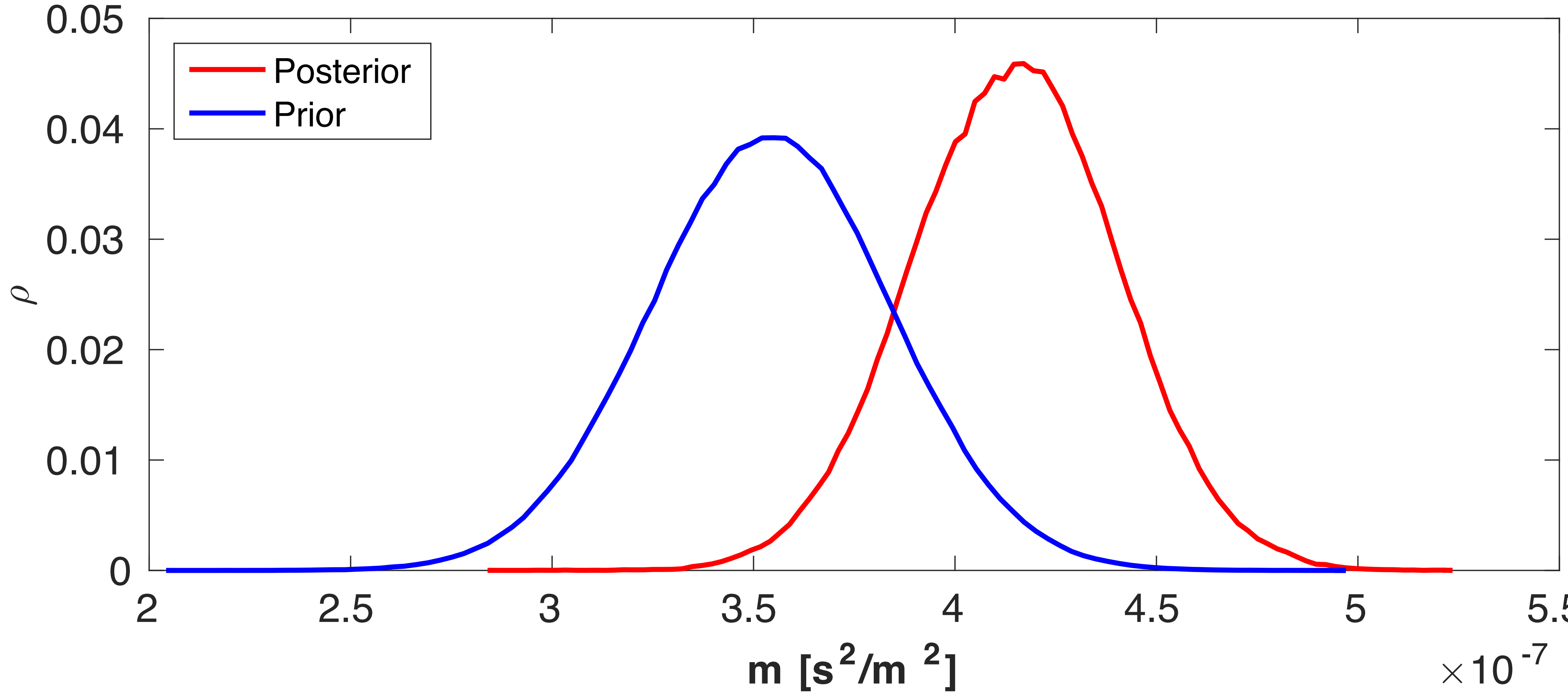
Posterior STD

Distribution

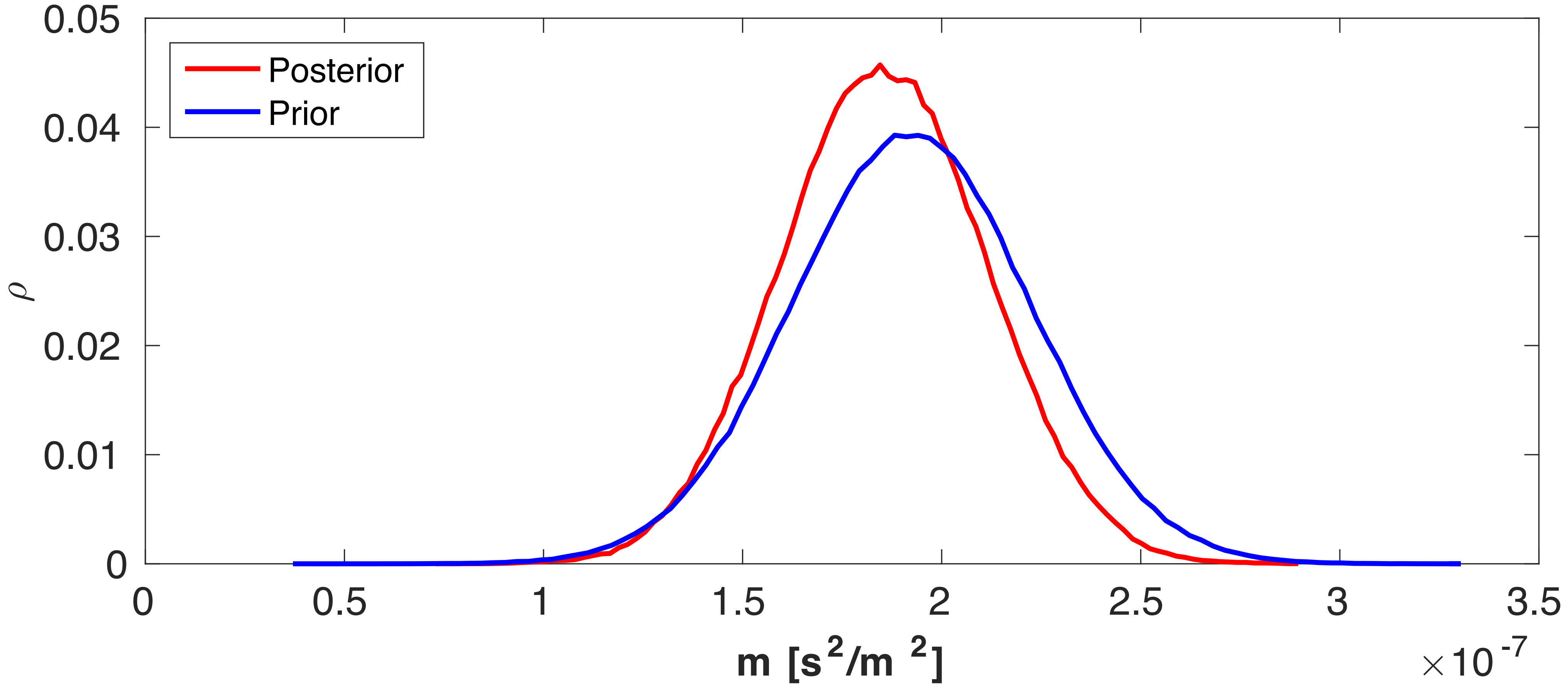




x = 1500m z = 750m

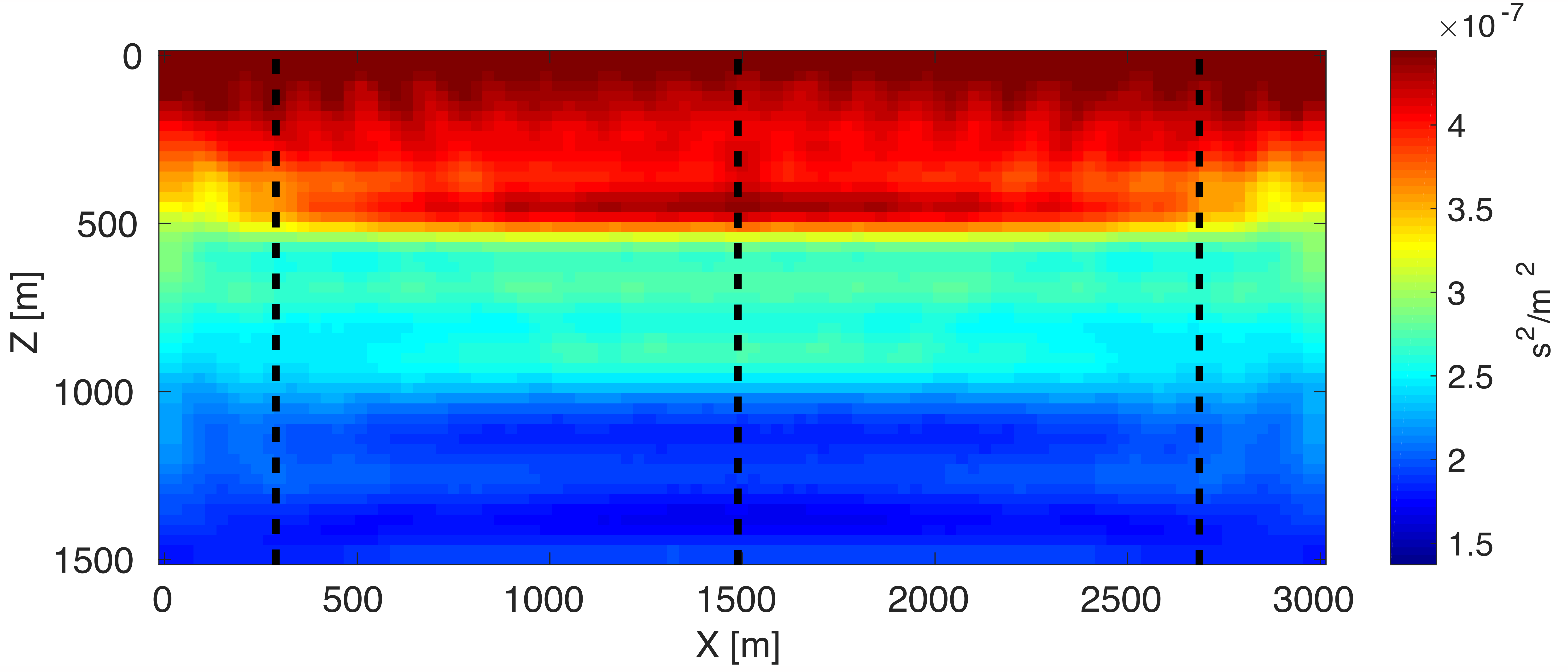


x = 1500m z = 270m

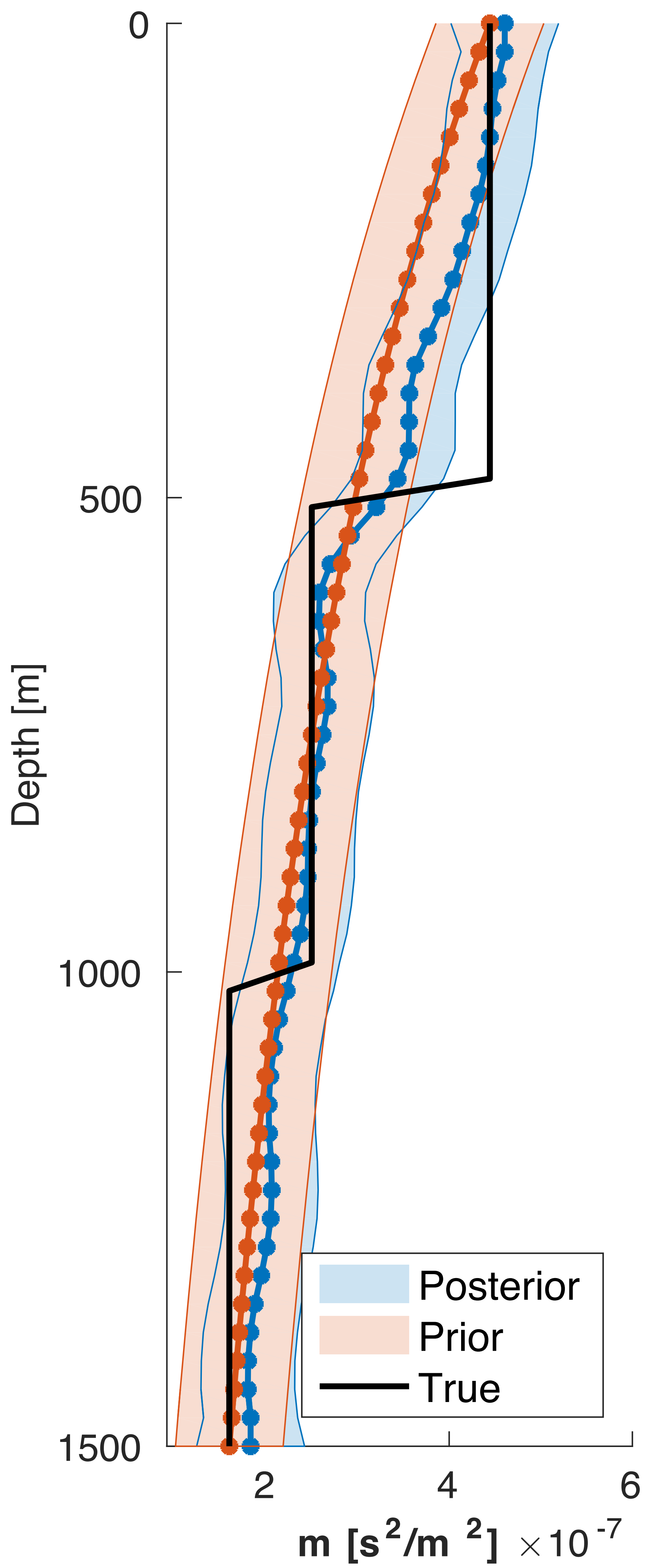


x = 1500m z = 1170m

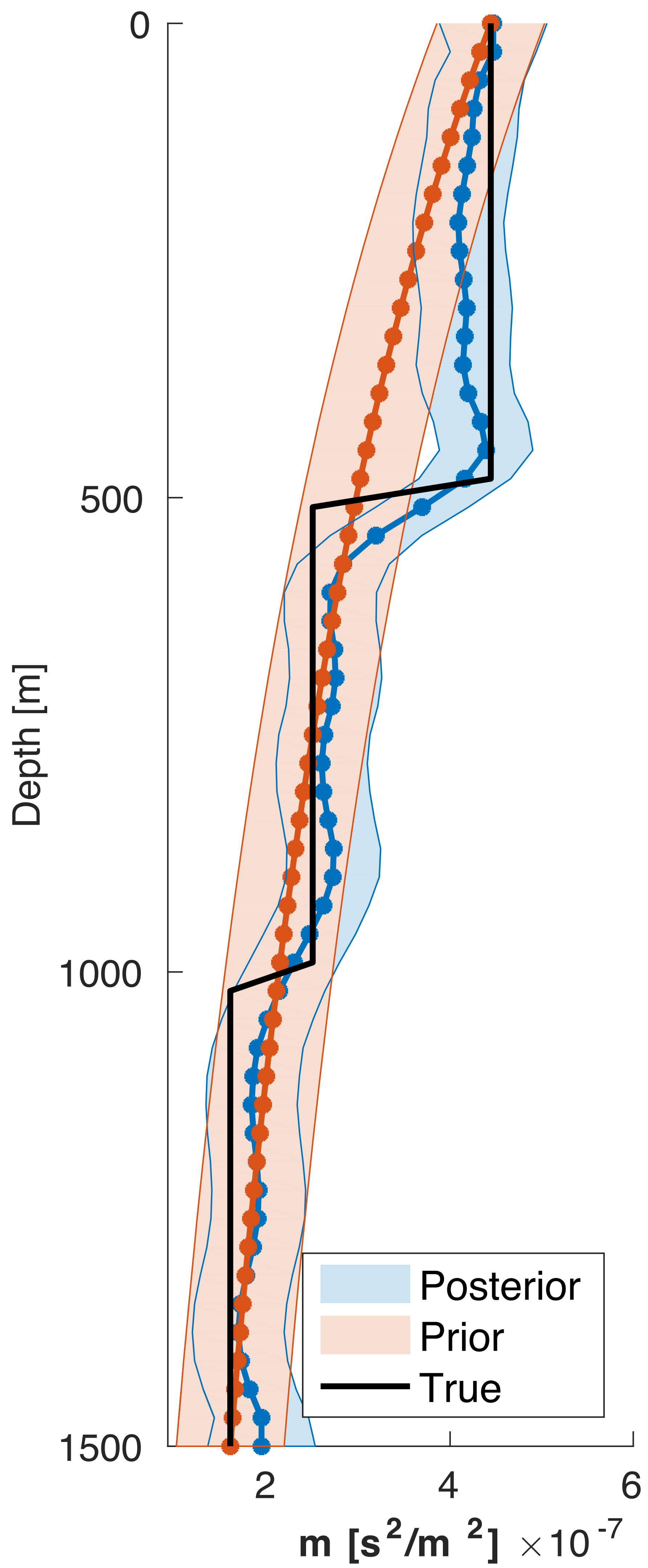
95% Confidence interval



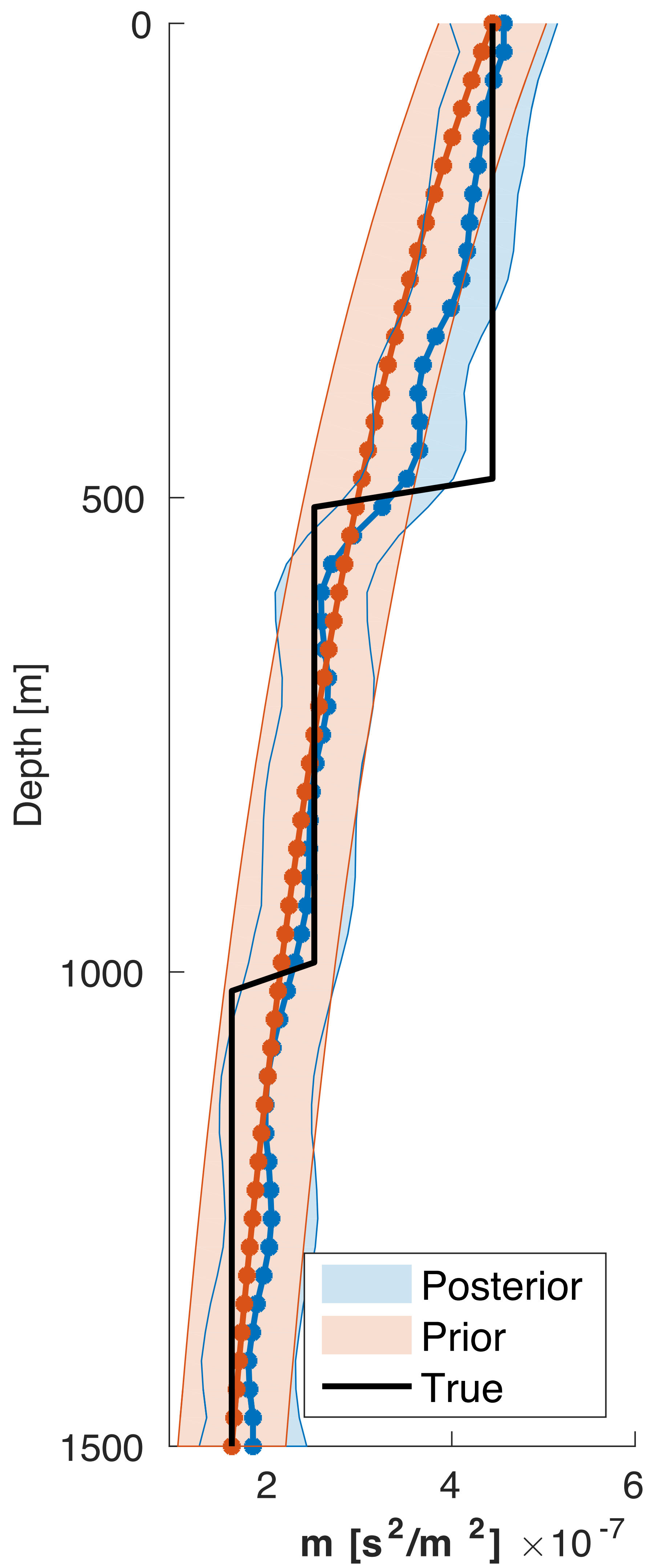
95% Confidence interval



x = 270m

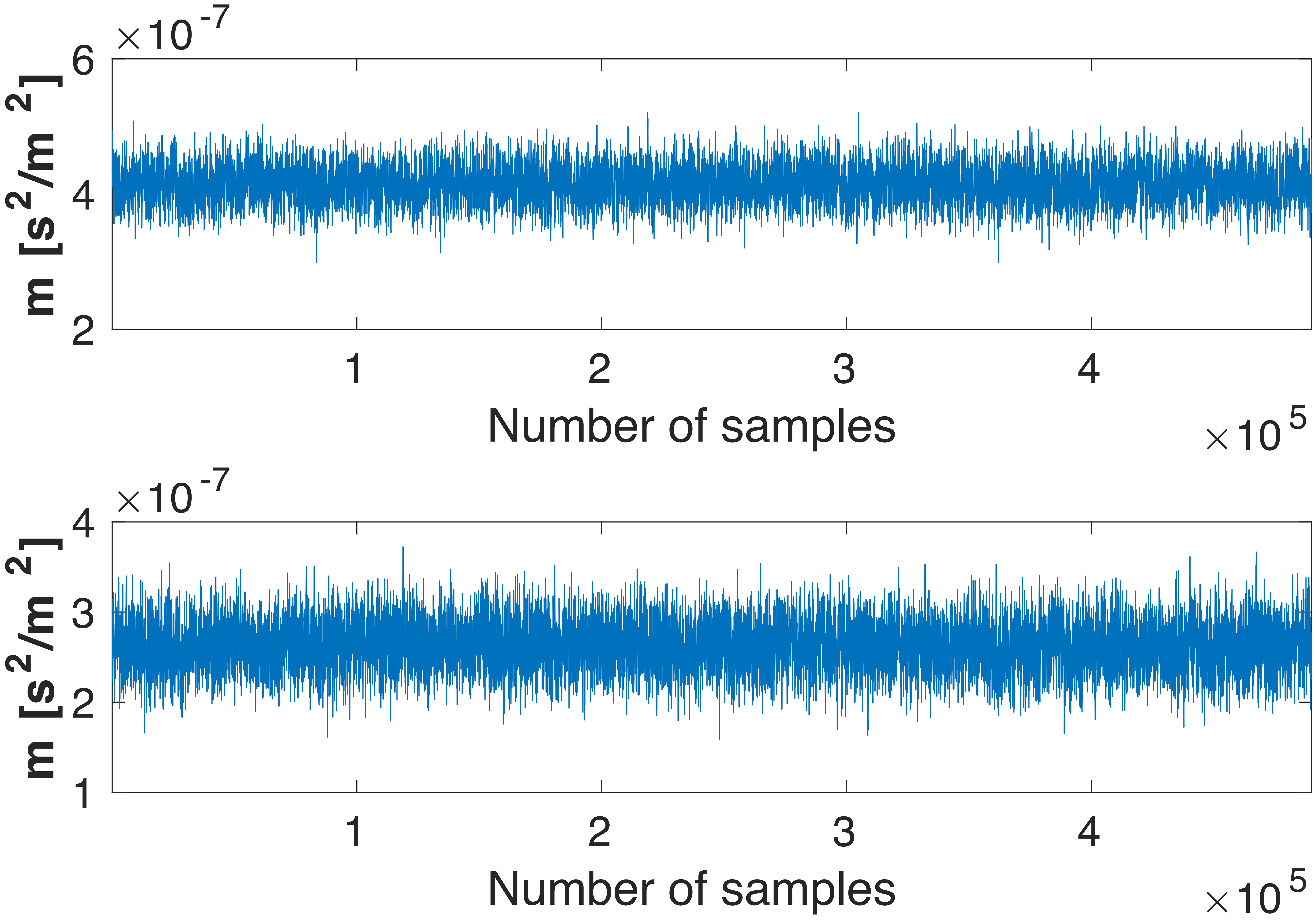


x = 1500m



x = 2670m

Samples



x = 1500m z = 270m

x = 1500m z = 750m

Conclusion

Posterior distribution for weak PDE constraints inverse problems:

- joint distribution with respect to the model parameters and wavefields
- conditional distributions – Gaussian distributions with sparse covariance matrices
- Gibbs sampling method – sample model parameters and wavefields from the corresponding conditional distributions alternatively
- Computational cost – one PDE solve per each source per each iteration

Reference

Albert Tarantola, and Bernard Valette. "Inverse problems= quest for information." *J. geophys* 50.3 (1982): 150-170.

Andrew M. Stuart, Jochen Voss, and Petter Wilberg. "Conditional path sampling of SDEs and the Langevin MCMC method." *Communications in Mathematical Sciences* 2.4 (2004): 685-697.

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.

Jean Virieux and Stéphane Operto. An overview of full-waveform inversion in exploration geophysics. *GEOPHYSICS*, 74(6):WCC1–WCC26, 2009.

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 2013.

Tristan van Leeuwen and Felix J. Herrmann. A penalty method for PDE-constrained optimization in inverse problems. *Inverse Problems*, 32(1):015007, 12 2015.

Michael Fisher, Mand Leutbecher, and G. A. Kelly. "On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation." *Quarterly Journal of the Royal Meteorological Society* 131.613 (2005): 3235-3246.

Acknowledgements

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