Wave-equation-based inversion with amortized variational Bayesian inference

Ali Siahkoohi [1] Gabrio Rizzuti [2] Rafael Orozco [1] Felix J. Herrmann [1]

[1]Georgia Institute of Technology

[2]Utrecht University

SLIM

Georgia Institute of Technology

Importance of uncertainty quantification

Seismic imaging is challenged by

- noisy data and linearization errors
- bandwidth and aperture limitations
- presence of shadow zones

Failure to assess uncertainty may have implications on downstream tasks

- uncertainty associated with tasks: horizon tracking, semantic segmentation, etc
- challenged by
  - high-dimensionality of seismic images
  - expensive-to-evaluate migration/demigration operator

Ali Siahkoohi, Gabrio Rizzuti, and Felix J Herrmann. "Deep Bayesian inference for seismic imaging with tasks". In: *arXiv preprint arXiv:2110.04825* (2021).

Challenges of solving Bayesian inverse problems

Choosing a prior distribution

- encode prior knowledge
- avoid unwanted bias due to overly simplifying priors

Computational cost

- costly forward operator
- high dimensional sampling/integration

Deep generative networks for solving inverse problems

learned prior and posterior distributions

fast conditional sampling

tractable density estimation

rely on access to high-quality training data

negative bias induced by distribution shifts during inference

Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. "Invertible generative models for inverse problems: mitigating representation error and dataset bias". In: *Proceedings of the 37th International Conference on Machine Learning*. 2020, pp. 399–409.

Ali Siahkoohi, Gabrio Rizzuti, Mathias Louboutin, Philipp Witte, and Felix J. Herrmann. "Preconditioned training of normalizing flows for variational inference in inverse problems". In: *3rd Symposium on Advances in Approximate Bayesian Inference*. Jan. 2021. URL: https://openreview.net/pdf?id=P9m1sMaNQ8T.

Ali Siahkoohi and Felix J. Herrmann. "Learning by example: fast reliability-aware seismic imaging with normalizing flows". Apr. 2021. URL: https://arxiv.org/pdf/2104.06255.pdf.

AmirEhsan Khorashadizadeh et al. "Conditional Injective Flows for Bayesian Imaging". In: *arXiv preprint arXiv:2204.07664* (2022).

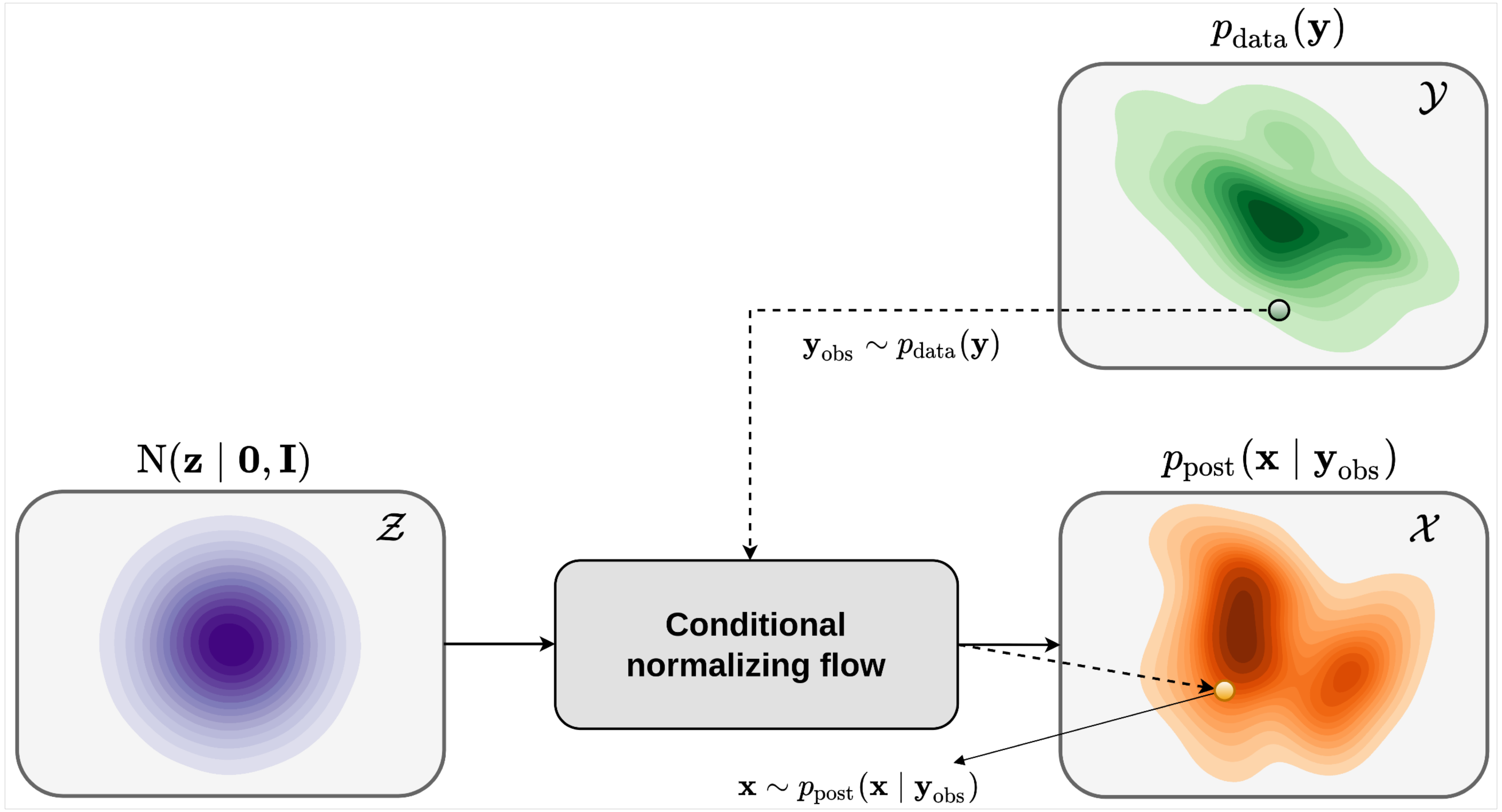
Proposed approach

learning prior and amortized posterior distributions with conditional normalizing flows

data-specific (non-amortized), low-cost, physics-based latent distribution correction

- cheap and unlimited posterior samples
- directly informed by data and physics
- minimizes the negative bias of distribution shifts during inference
- feasible in domains with limited access to training data

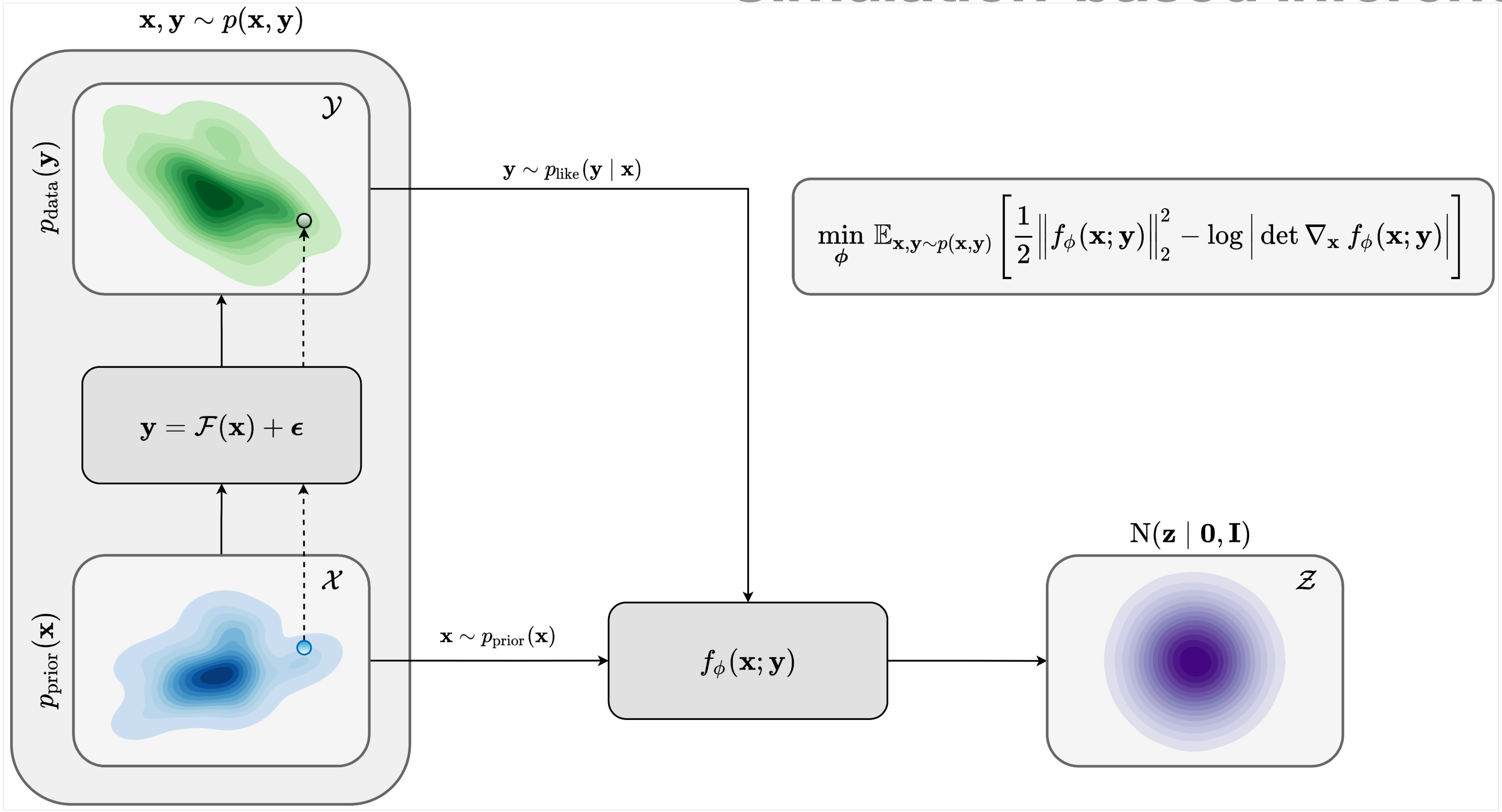
Amortized posterior sampling w/ normalizing flows



Jakob Kruse, Gianluca Detommaso, Robert Scheichl, and Ullrich Köthe. “HINT: Hierarchical Invertible Neural Transport for Density Estimation and Bayesian Inference”. In: *Proceedings of AAAI-2021* (2021). URL: https://arxiv.org/pdf/1905.10687.pdf.

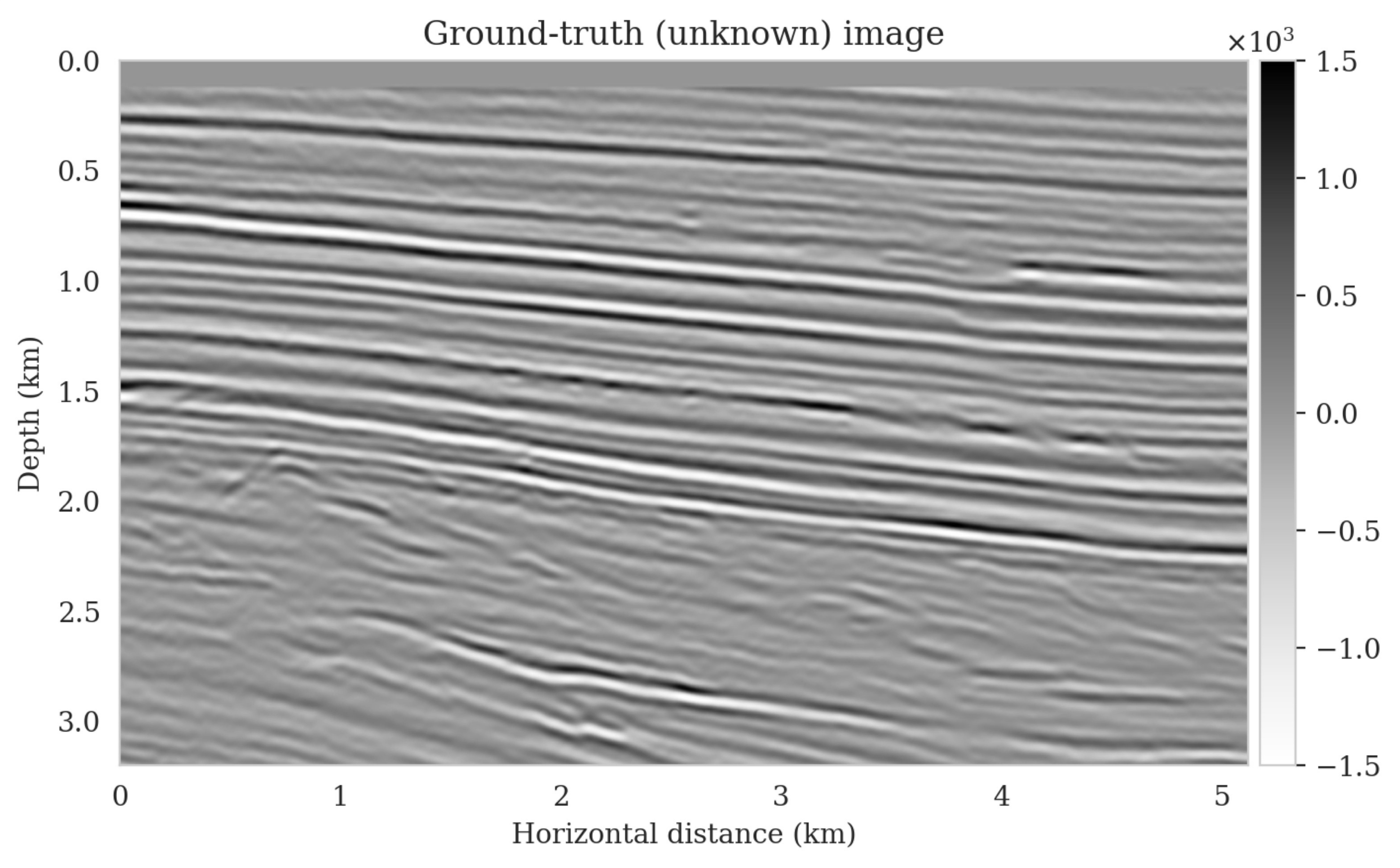
Ali Siahkoohi and Felix J. Herrmann. “Learning by example: fast reliability-aware seismic imaging with normalizing flows”. Apr. 2021. URL: https://arxiv.org/pdf/2104.06255.pdf.

Simulation-based inference

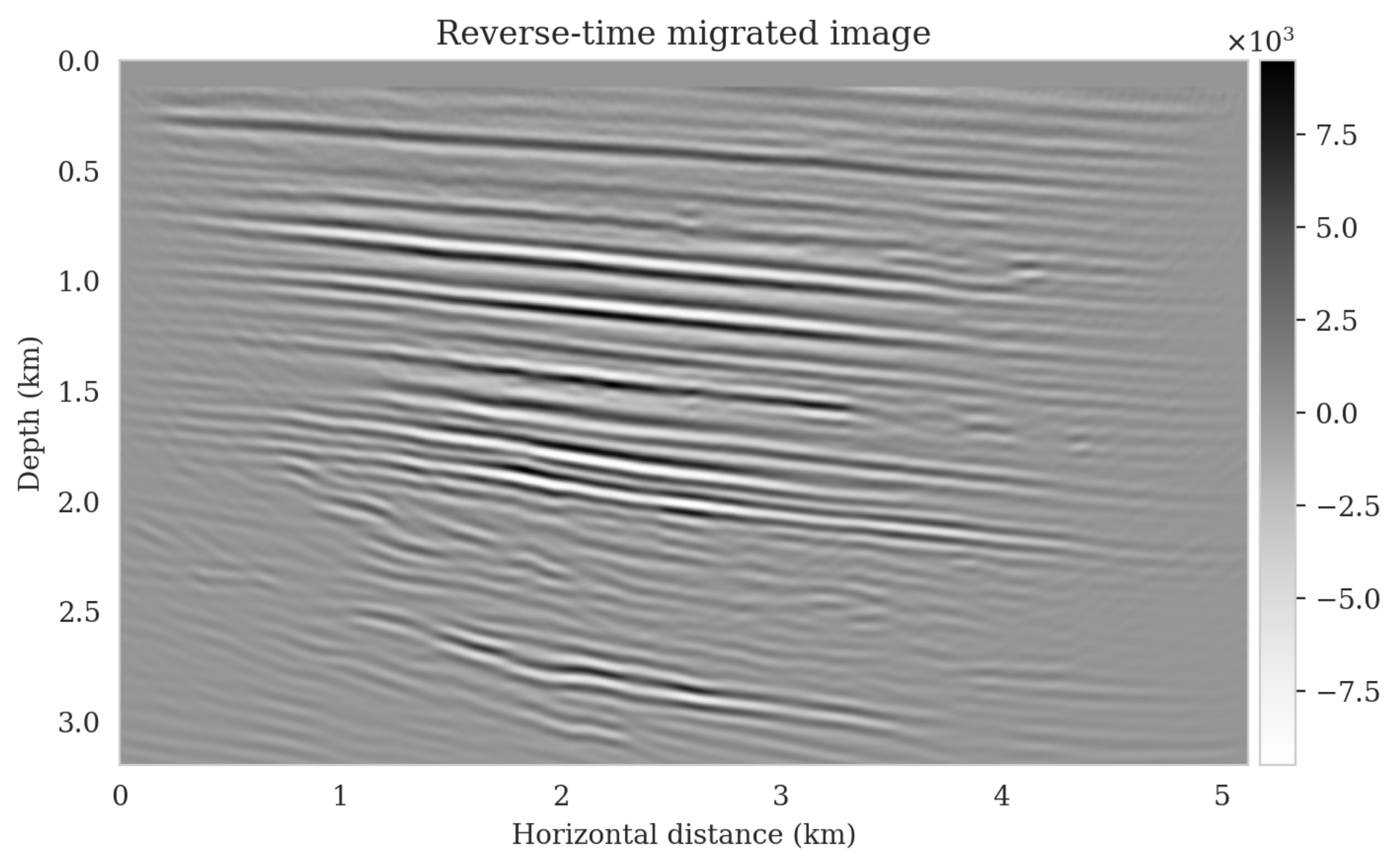


Seismic imaging example

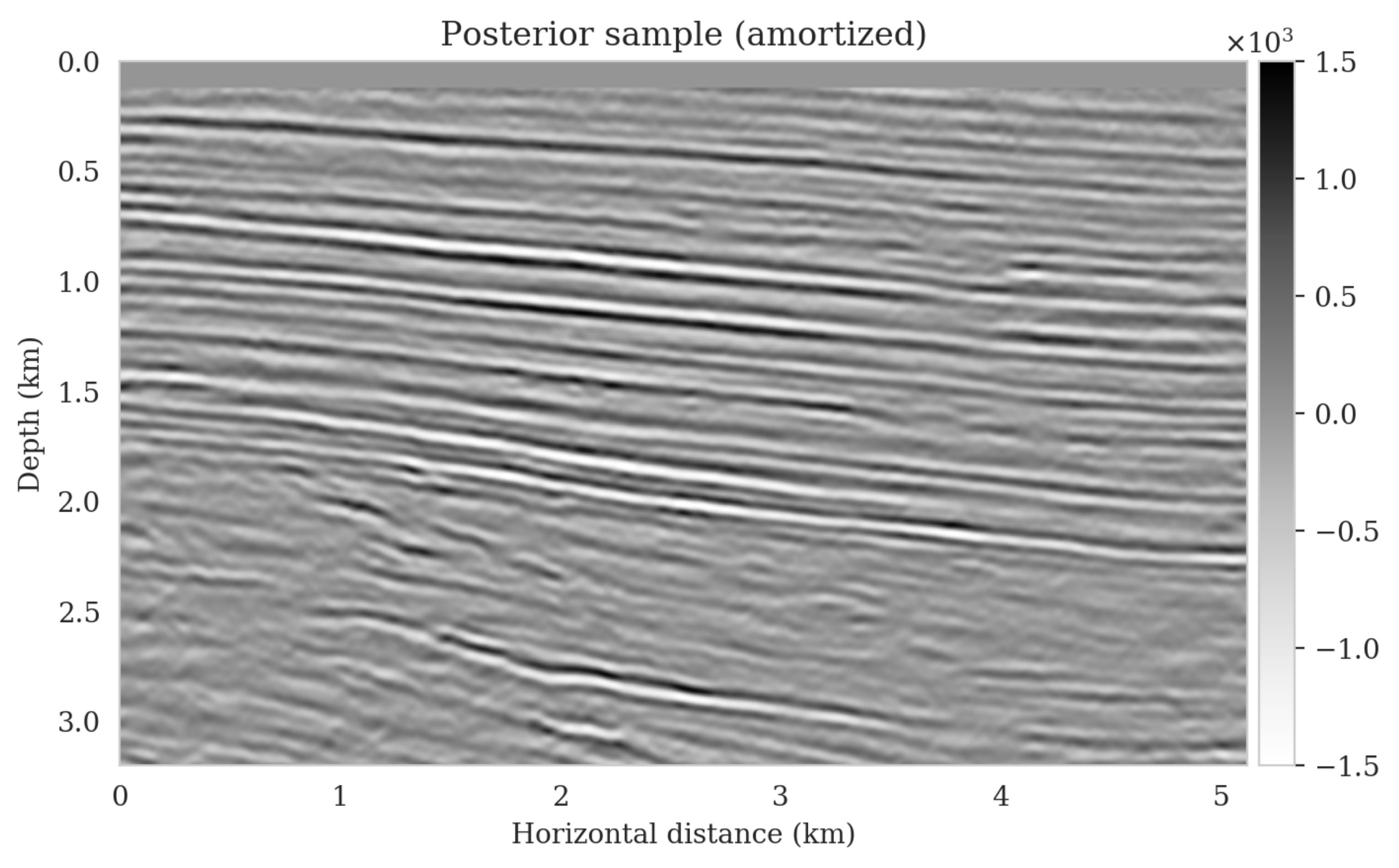
in-distribution amortized posterior sampling



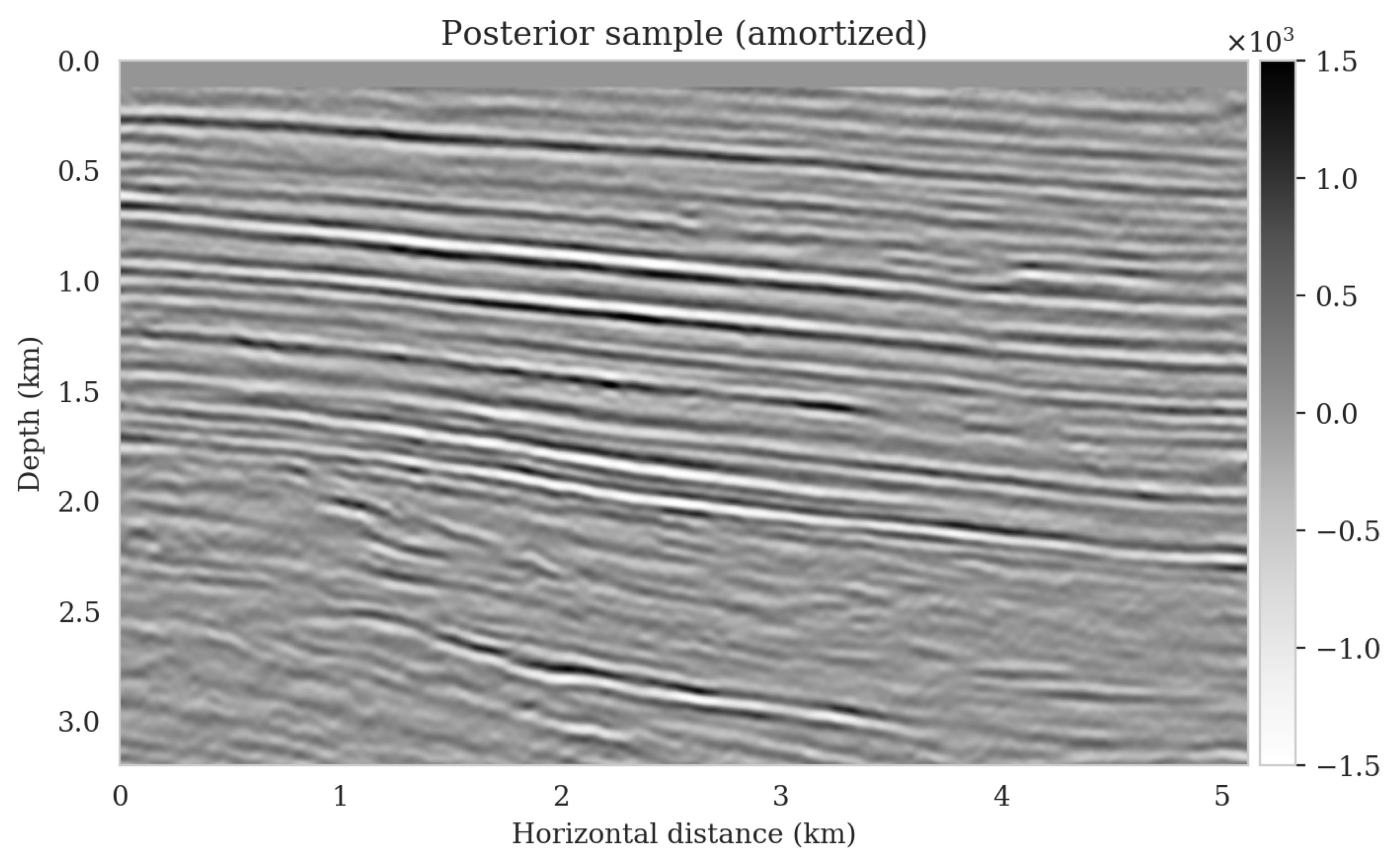
previously unseen (test) seismic image


Reverse-time migrated image
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10³
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5

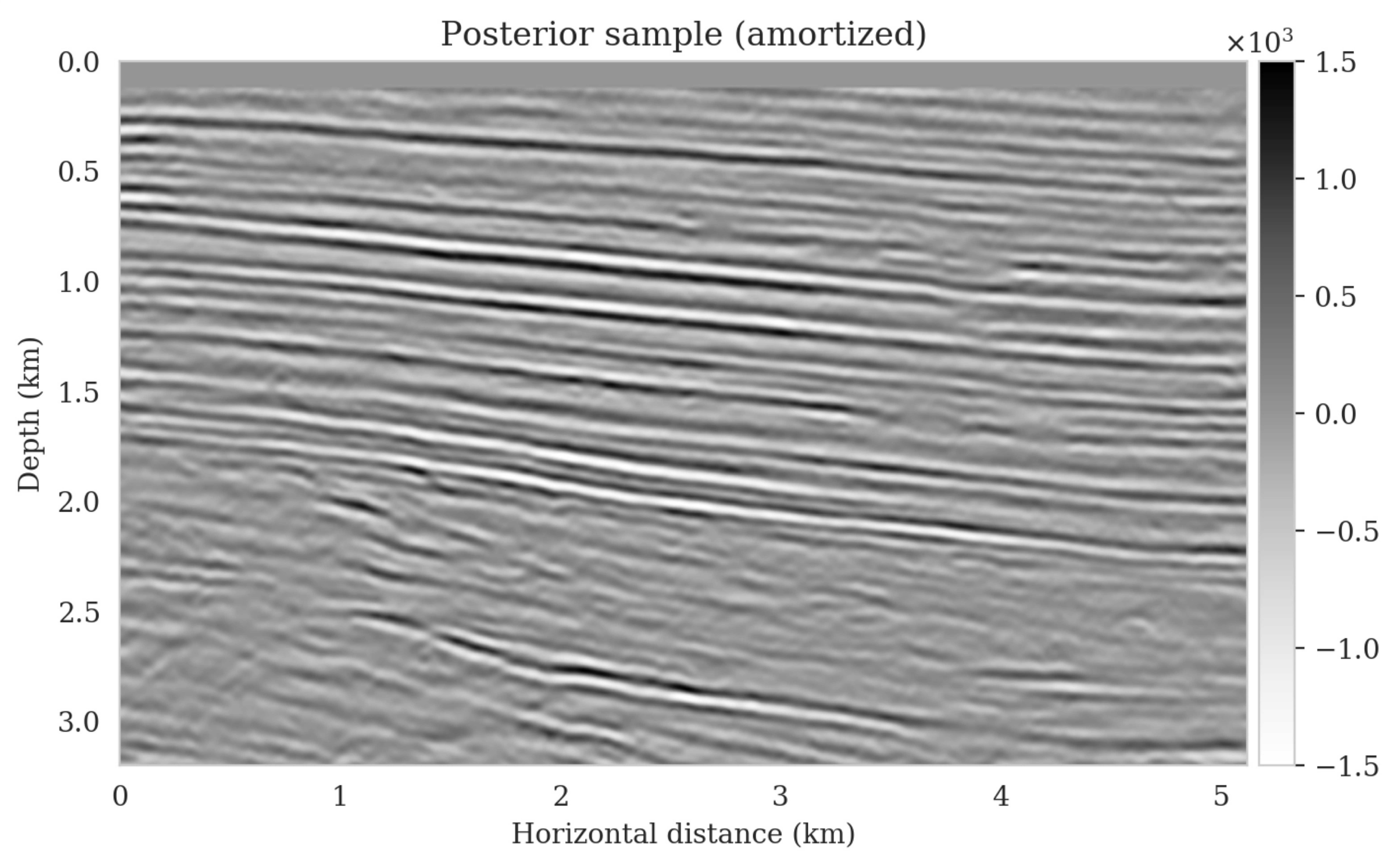
SNR −12.17 dB


Posterior sample (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 8.54 dB

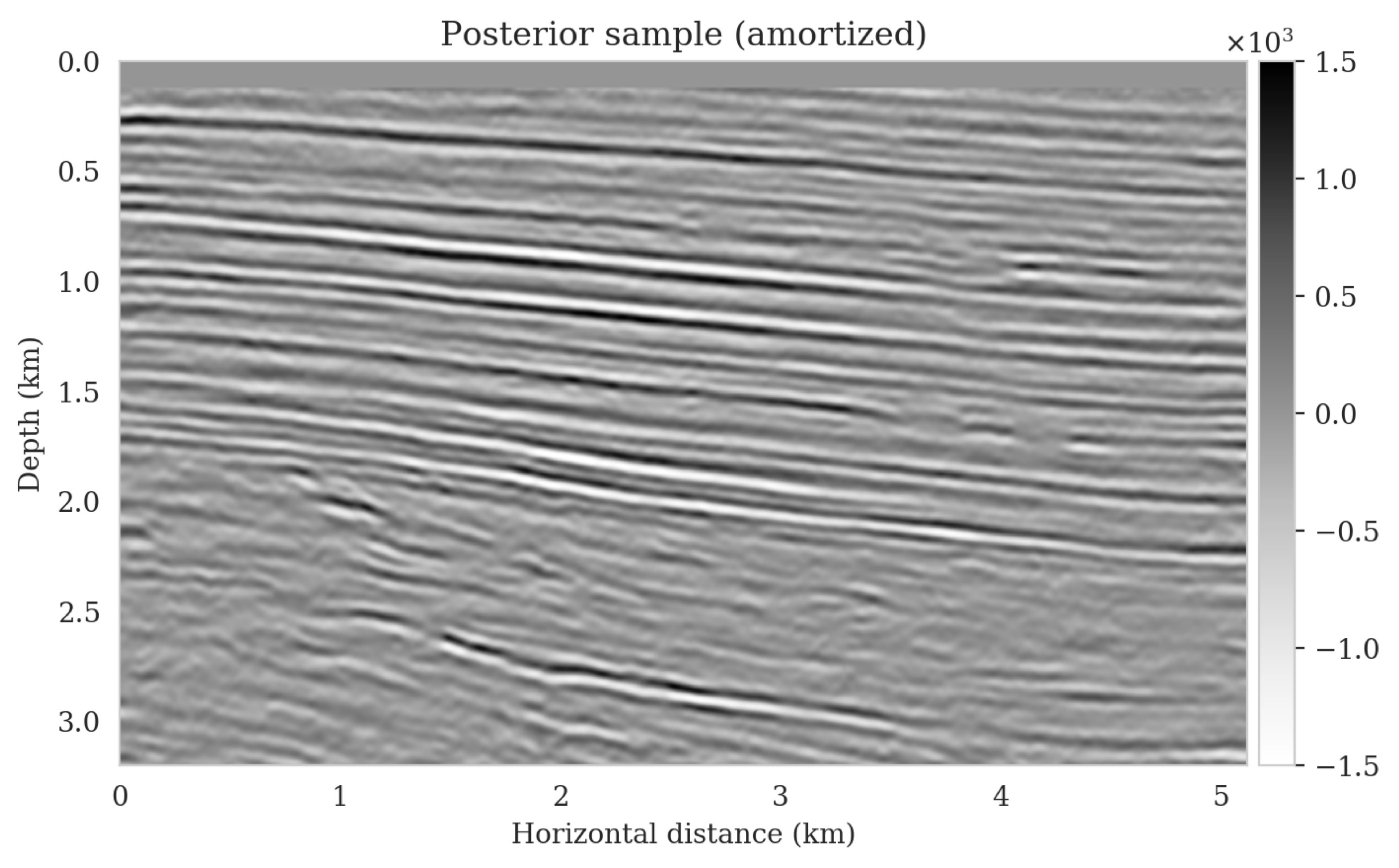

Posterior sample (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 8.49 dB

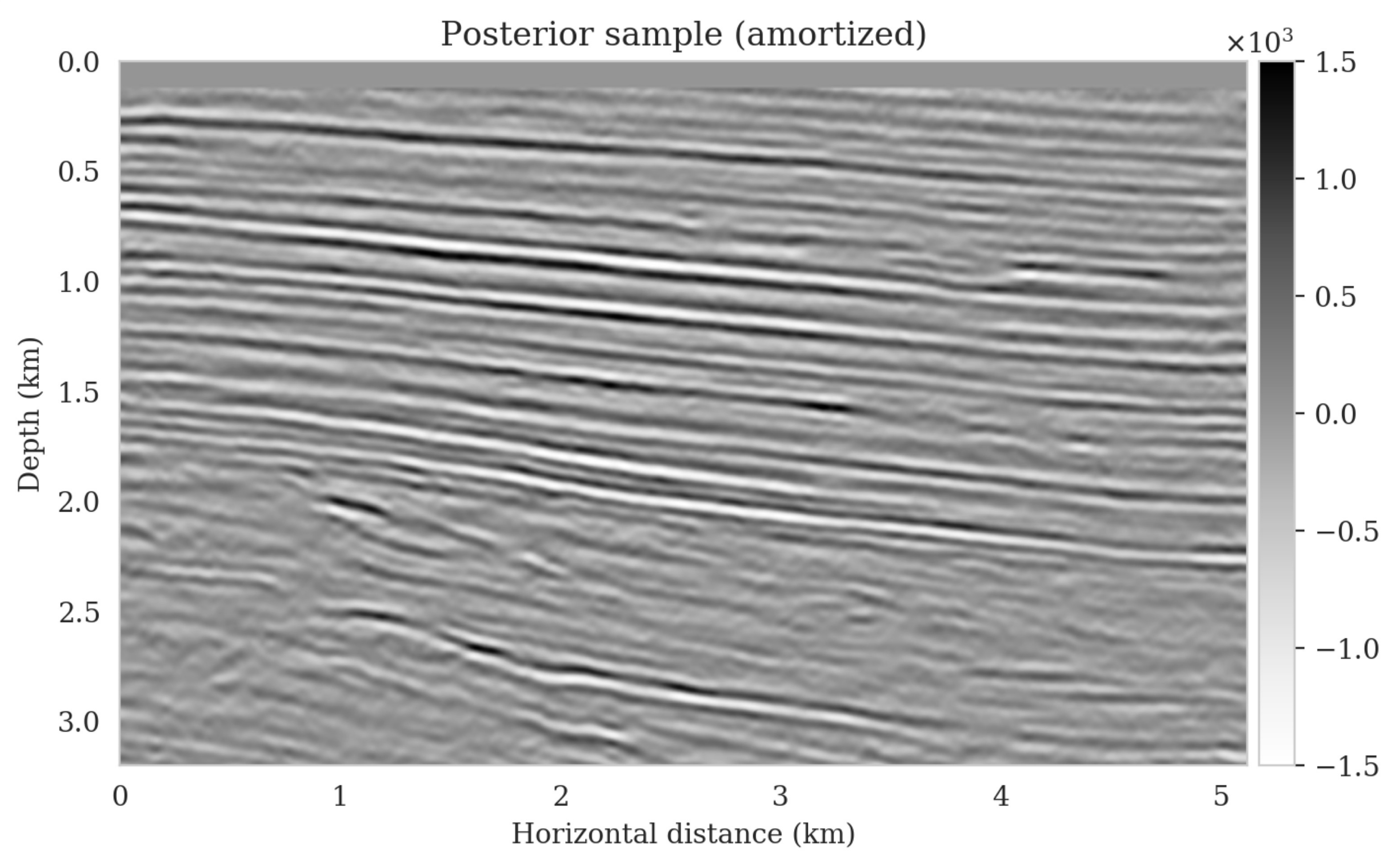


Posterior sample (amortized)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

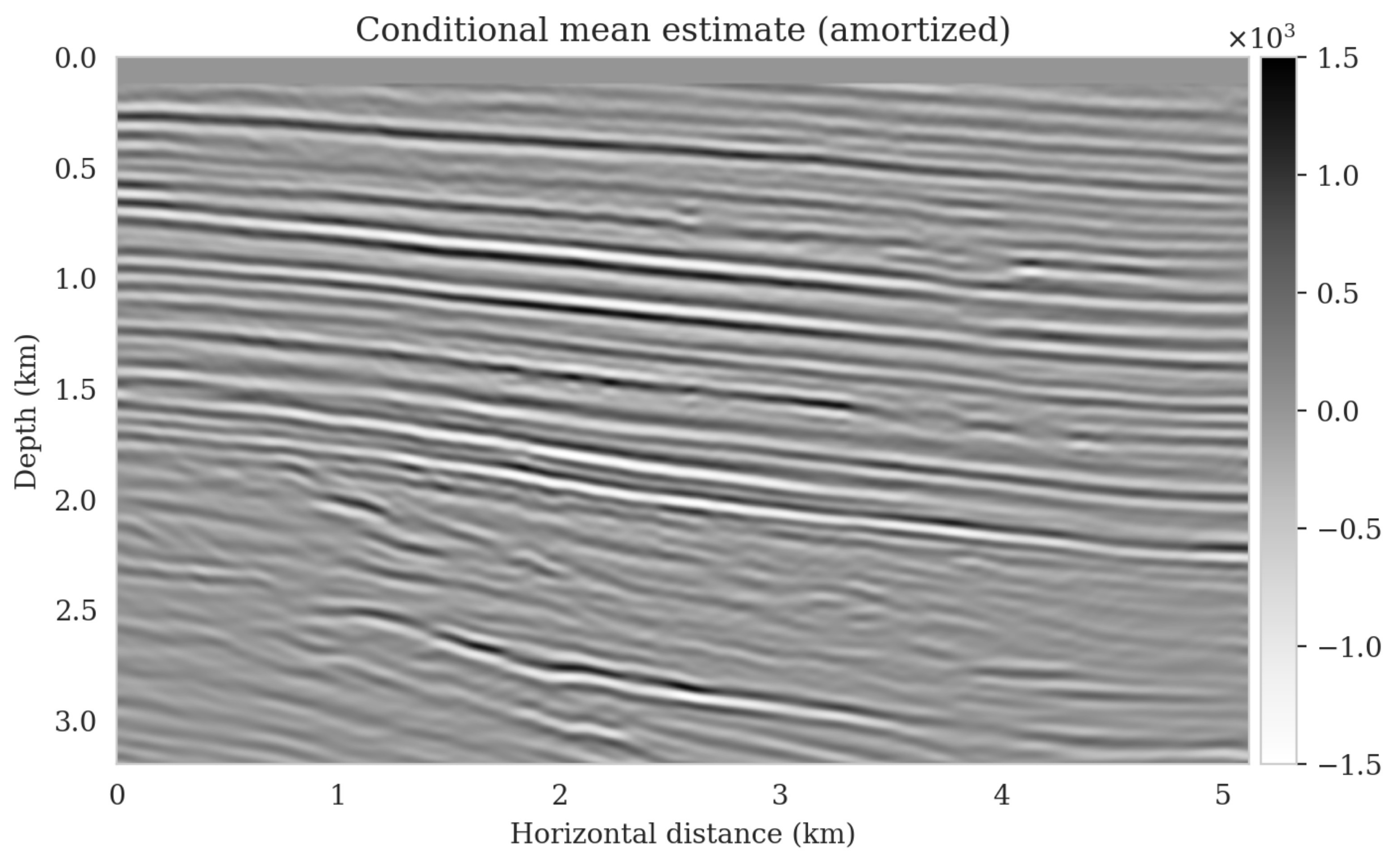
SNR 8.42 dB


Posterior sample (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

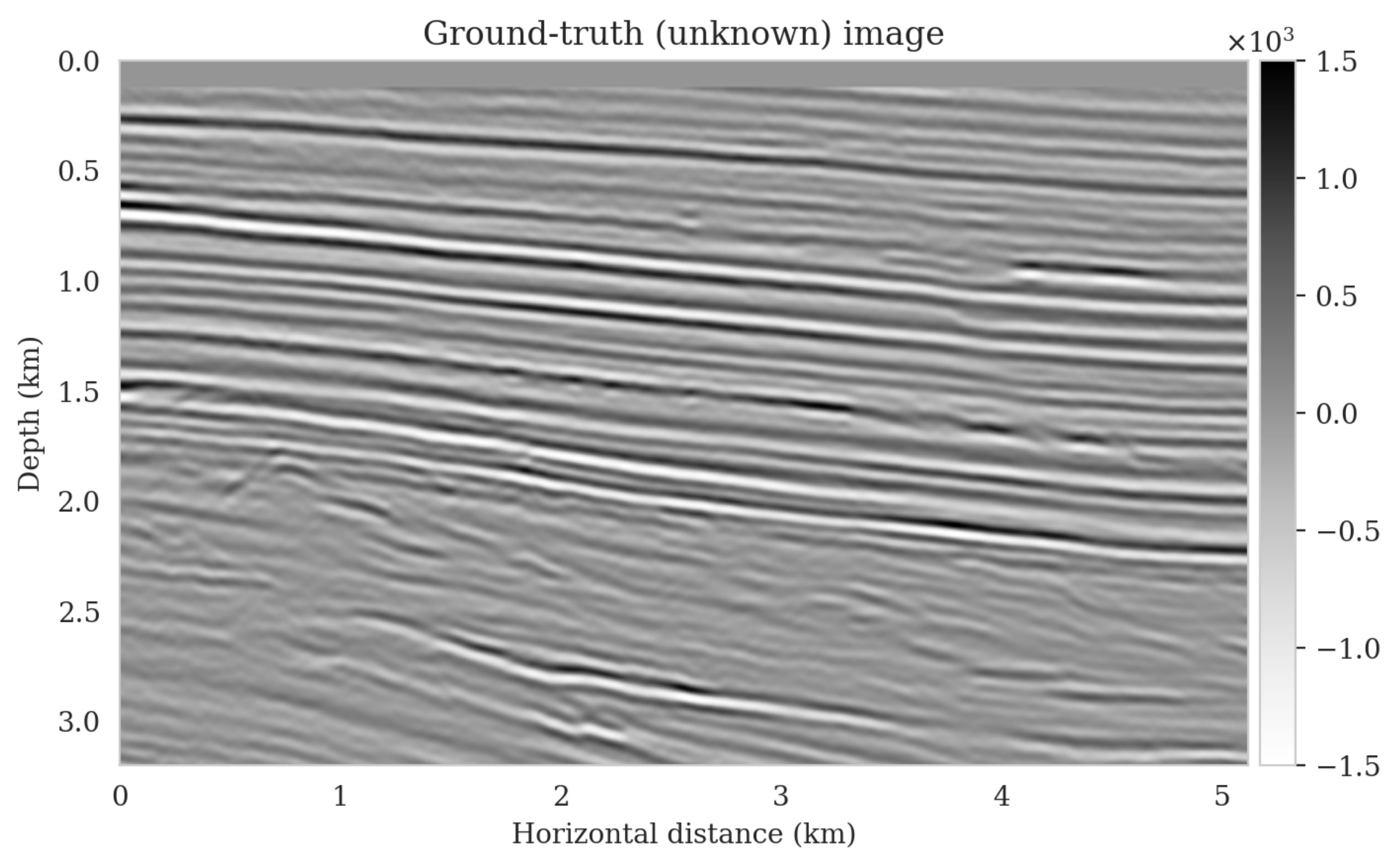
SNR 8.59 dB


Posterior sample (amortized)
×10³
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

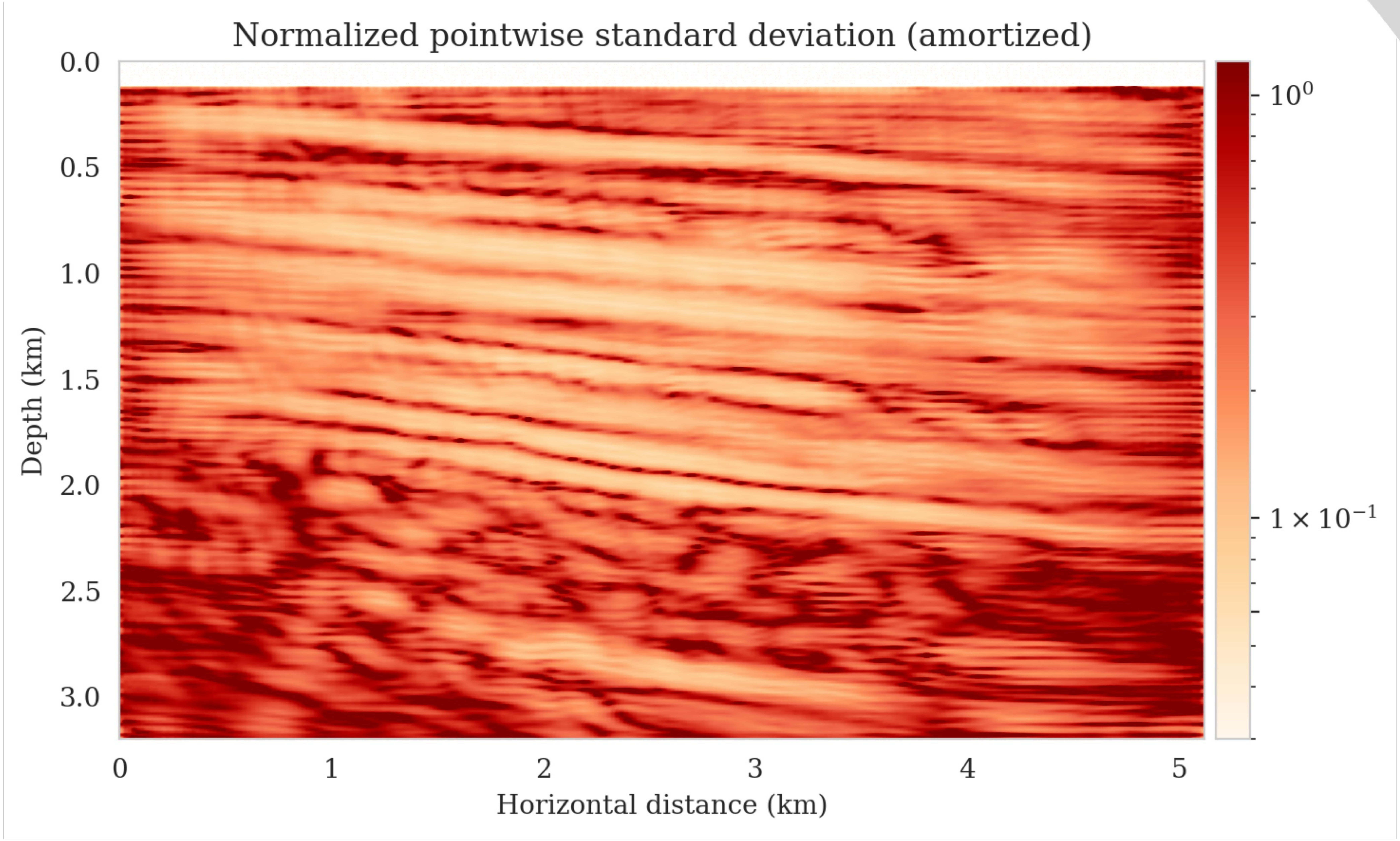
SNR 8.54 dB


Conditional mean estimate (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 11.24 dB



previously unseen (test) seismic image



normalized by the the envelope of the conditional mean

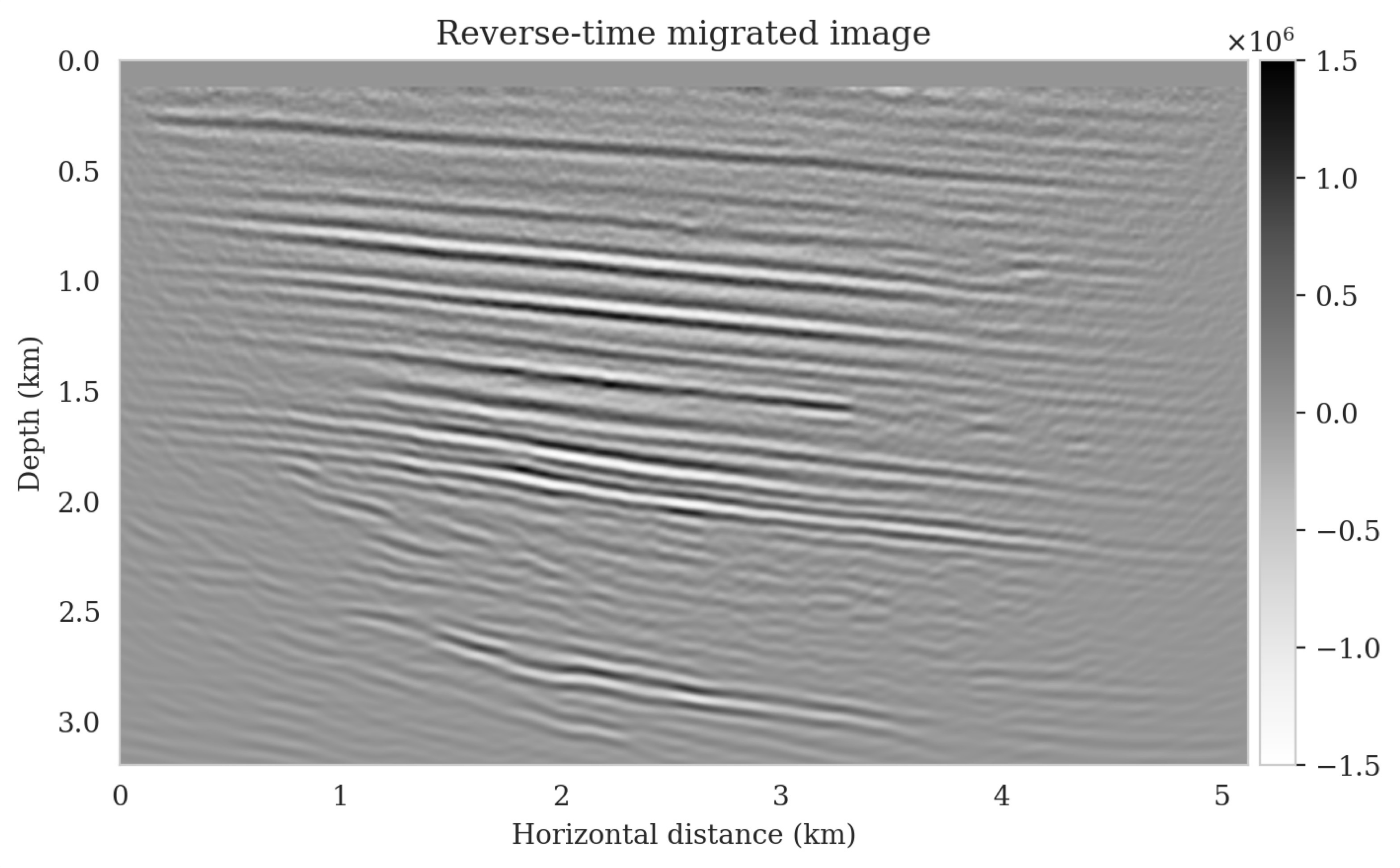
Introducing distribution shifts

band-limited noise with $6.25\times$ larger variance

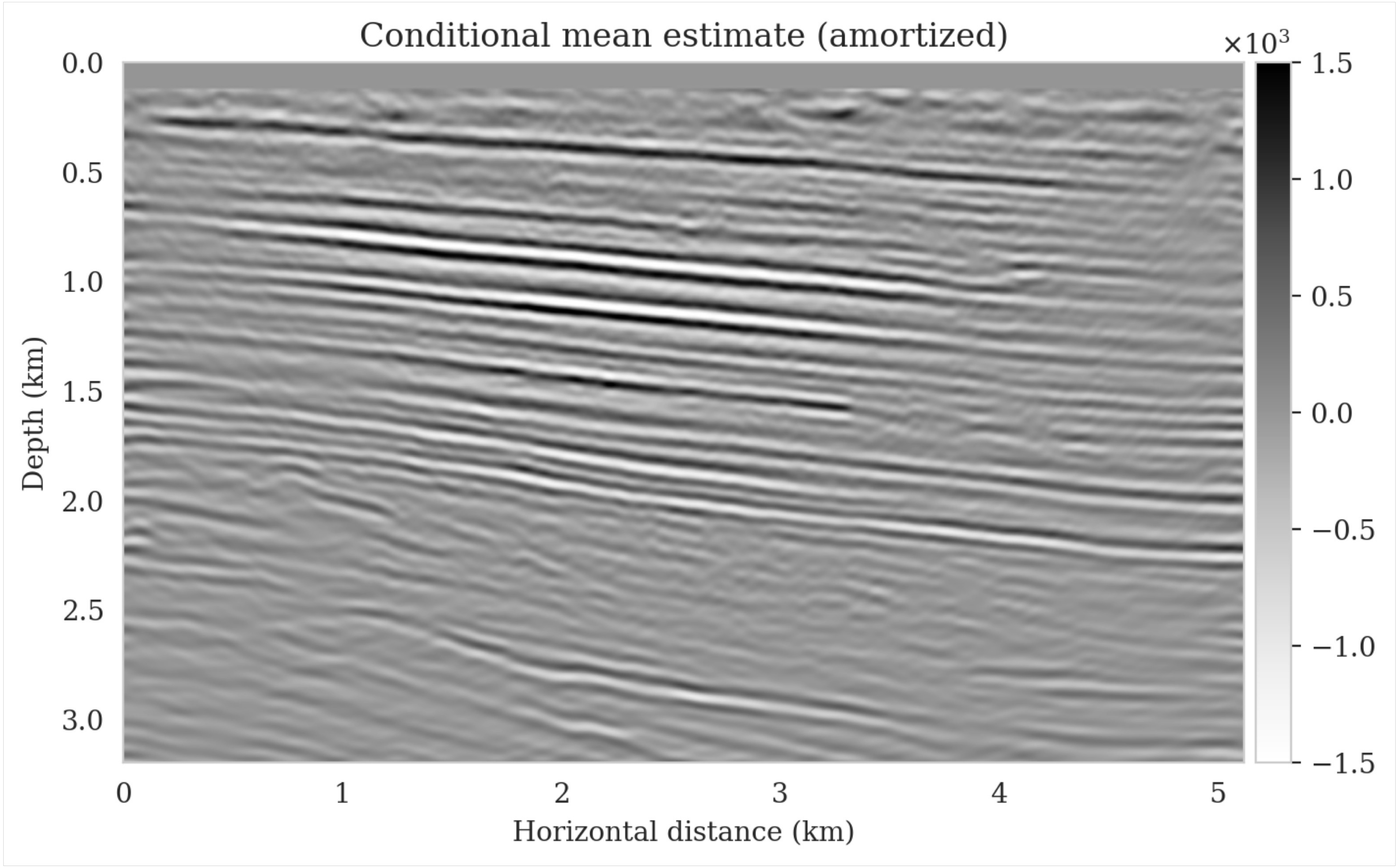
$4\times$ less sources

Physics-based latent distribution correction

computational cost: approximately $5\times$ RTMs

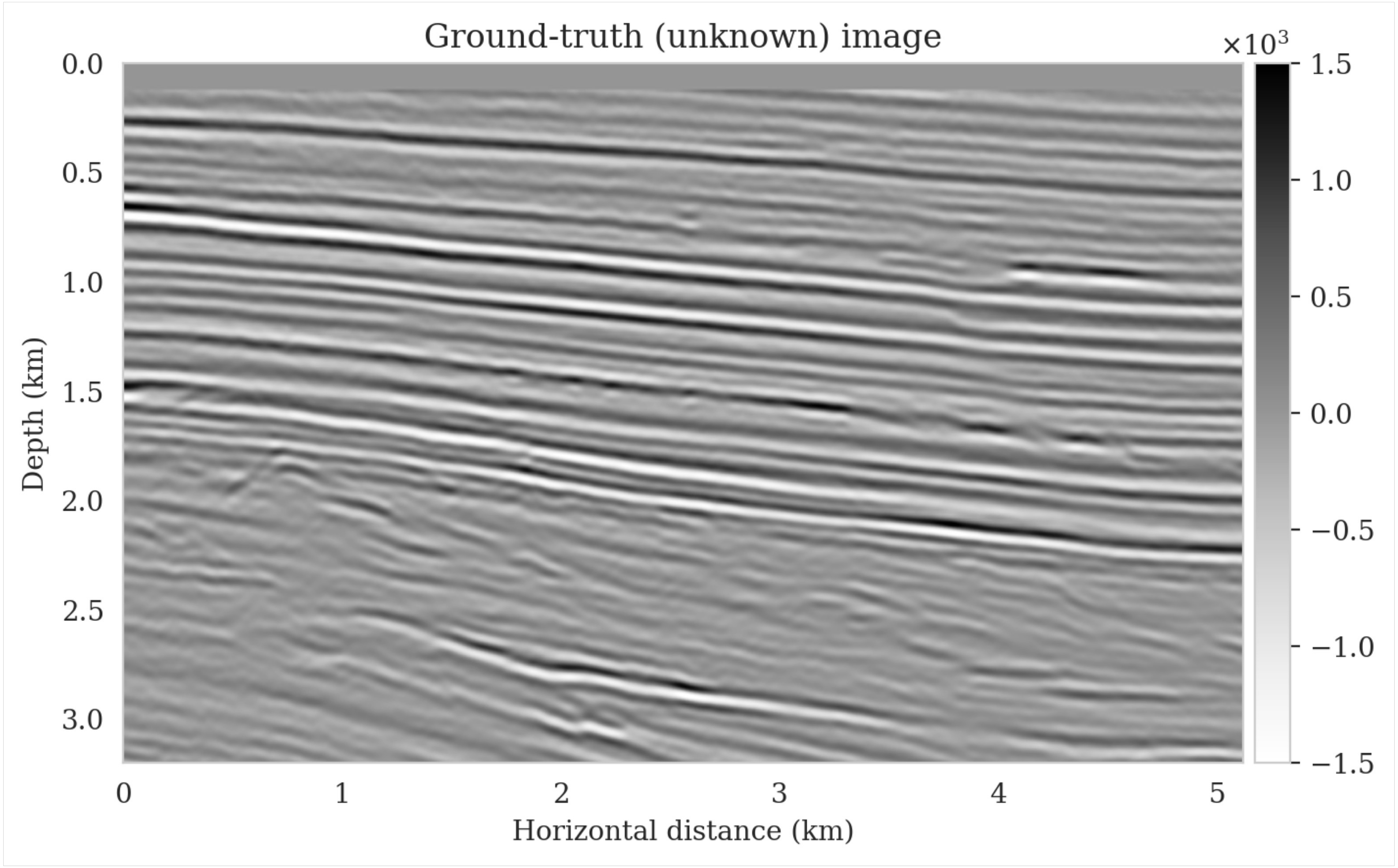

Reverse-time migrated image
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10^6
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR -8.22 dB



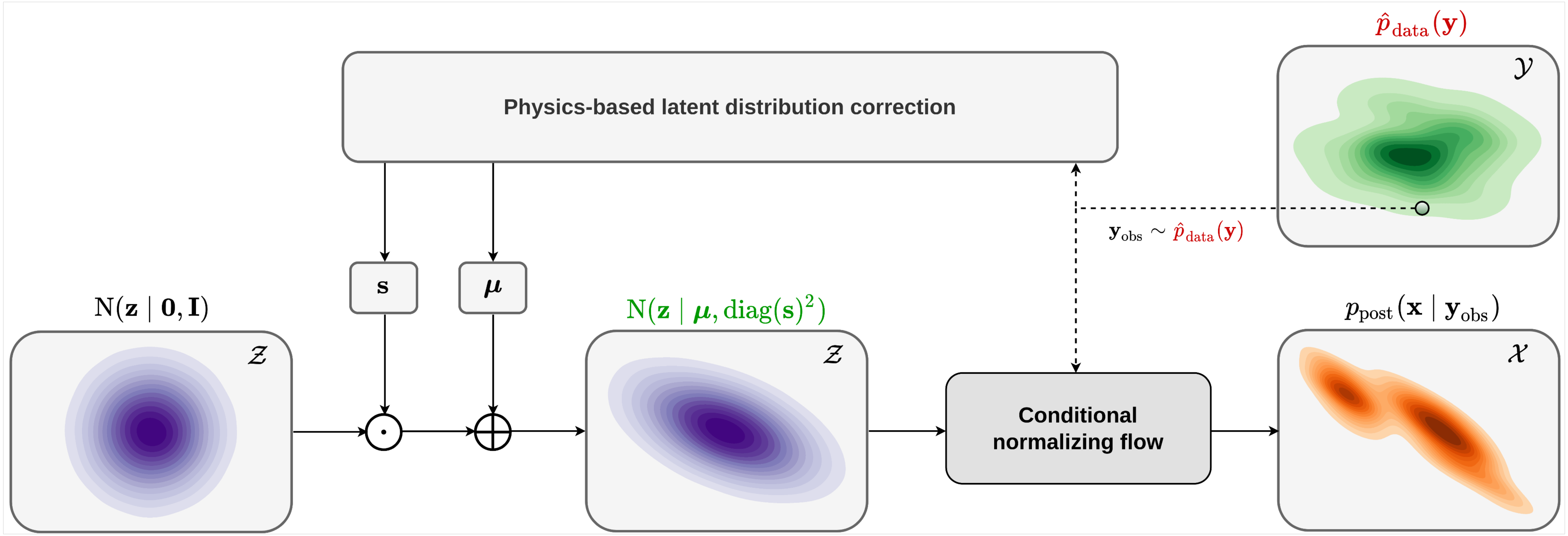
Conditional mean estimate (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 6.29 dB

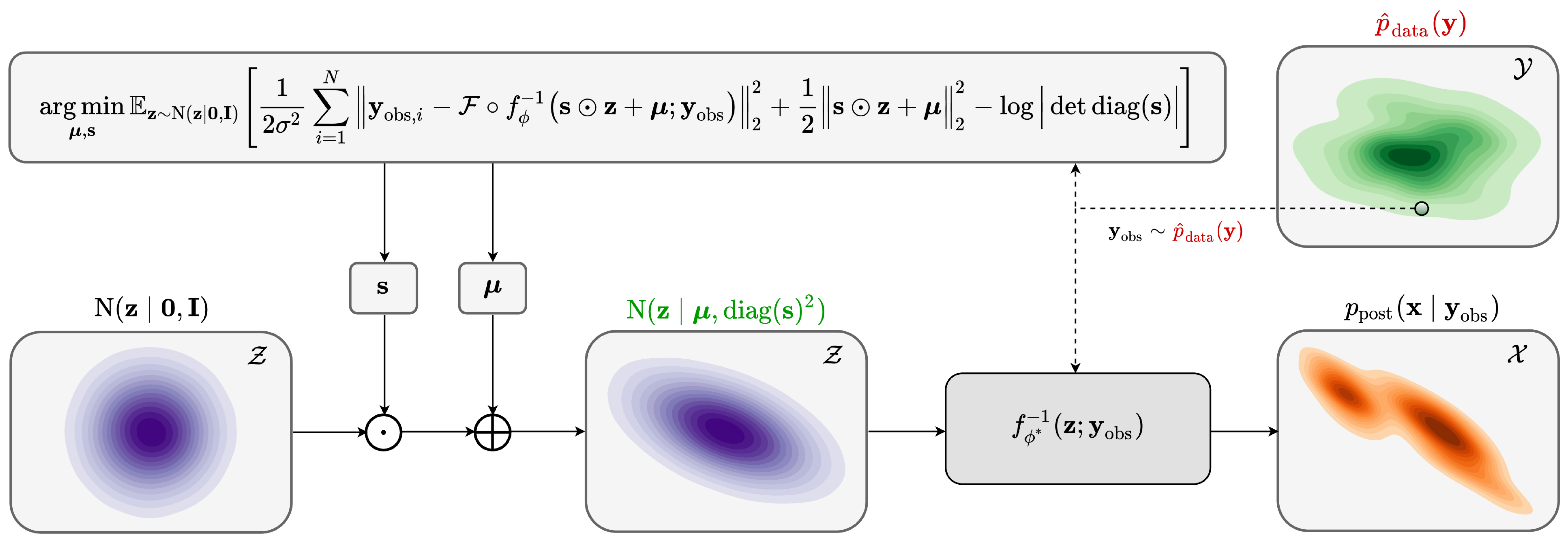


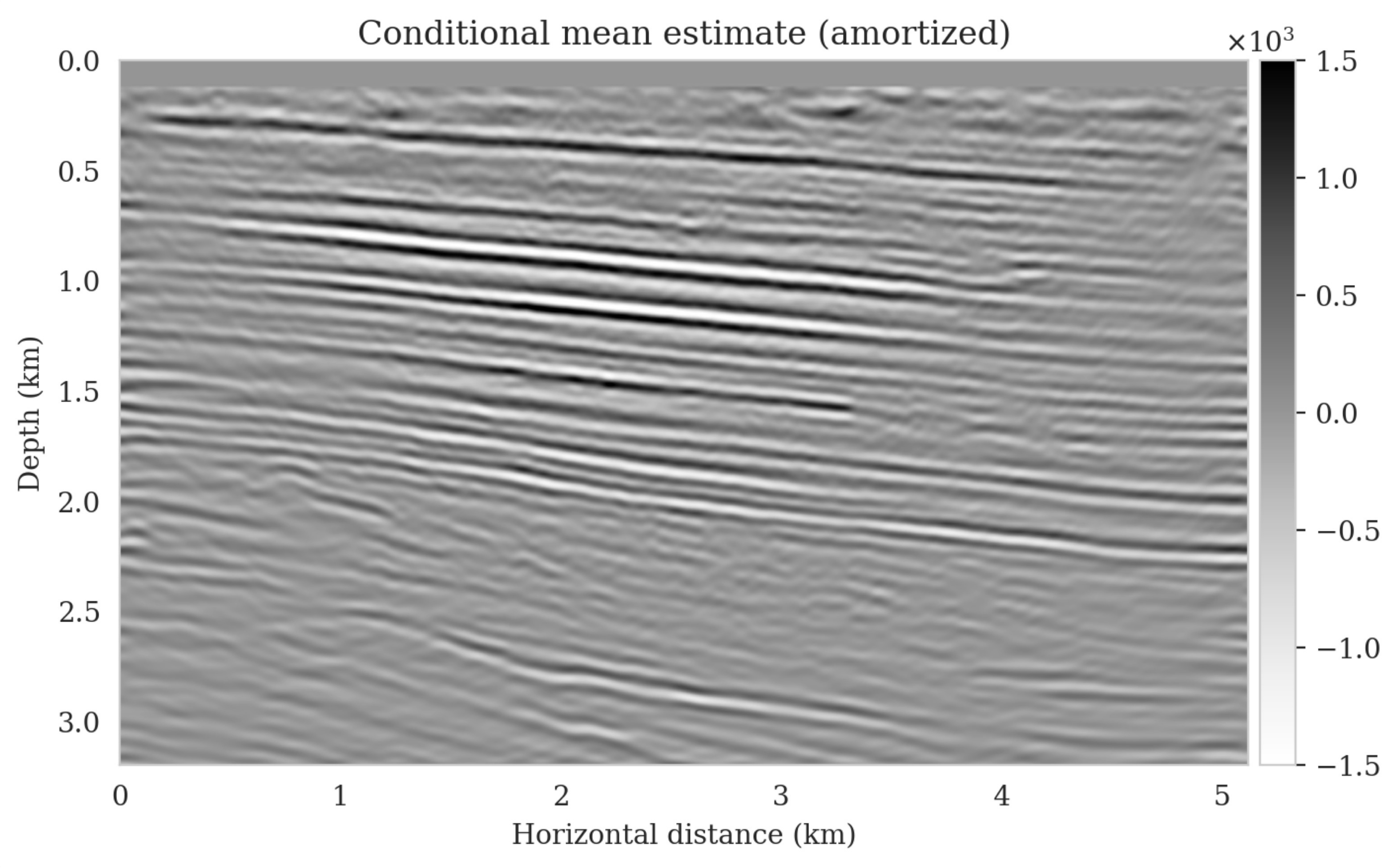
previously unseen (test) seismic image

Physics-based latent distribution correction

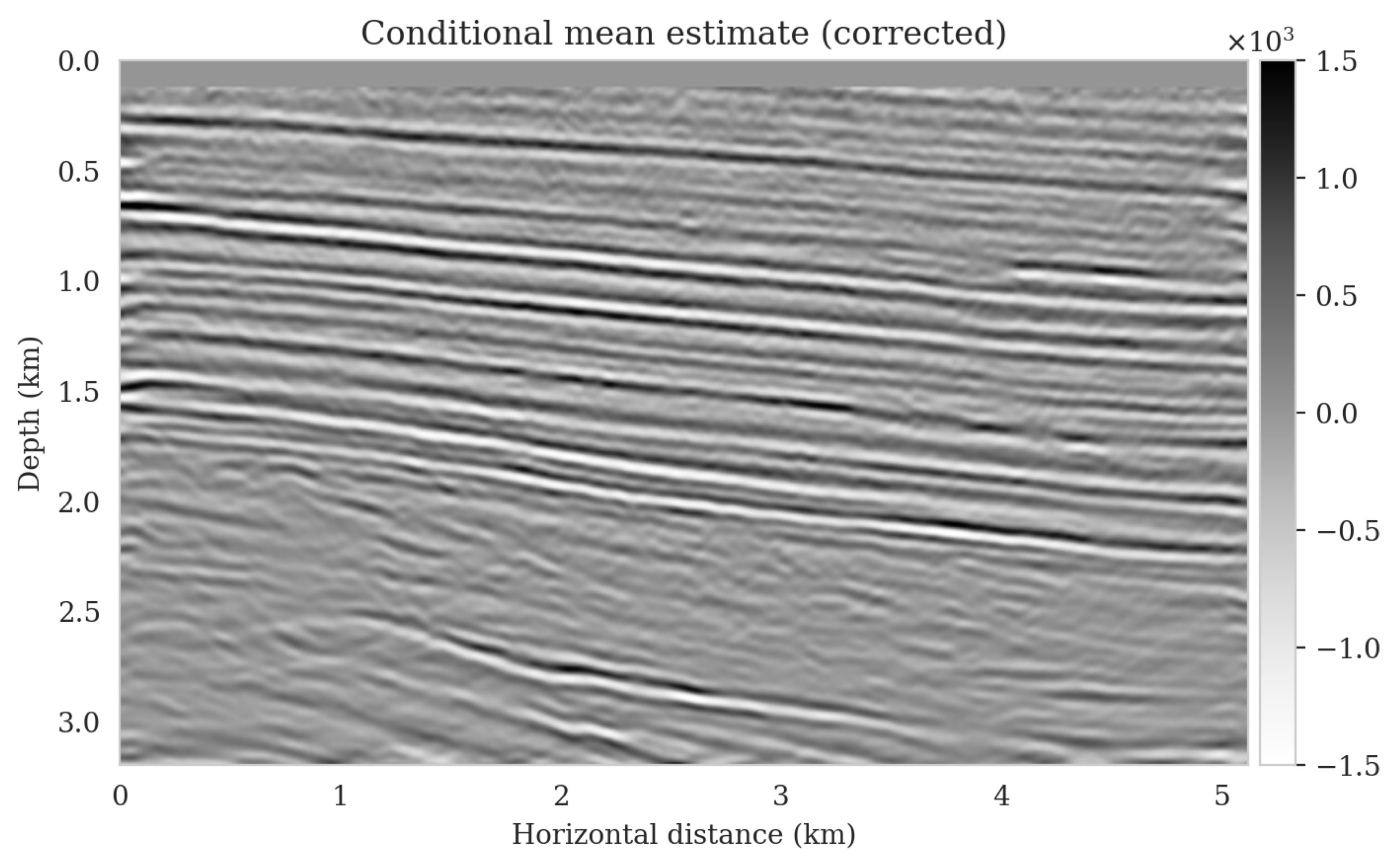


Physics-based latent distribution correction

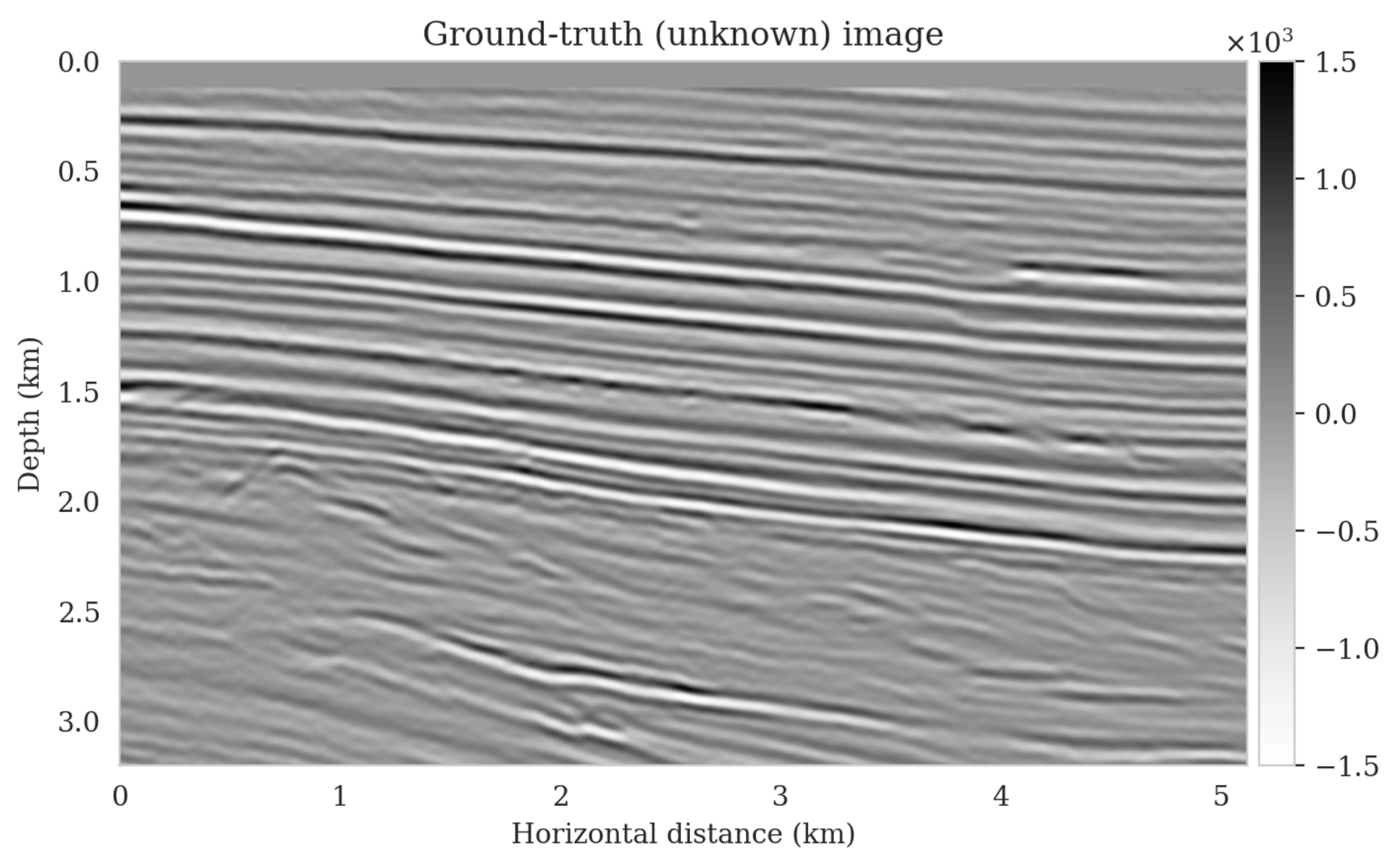



Conditional mean estimate (amortized)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

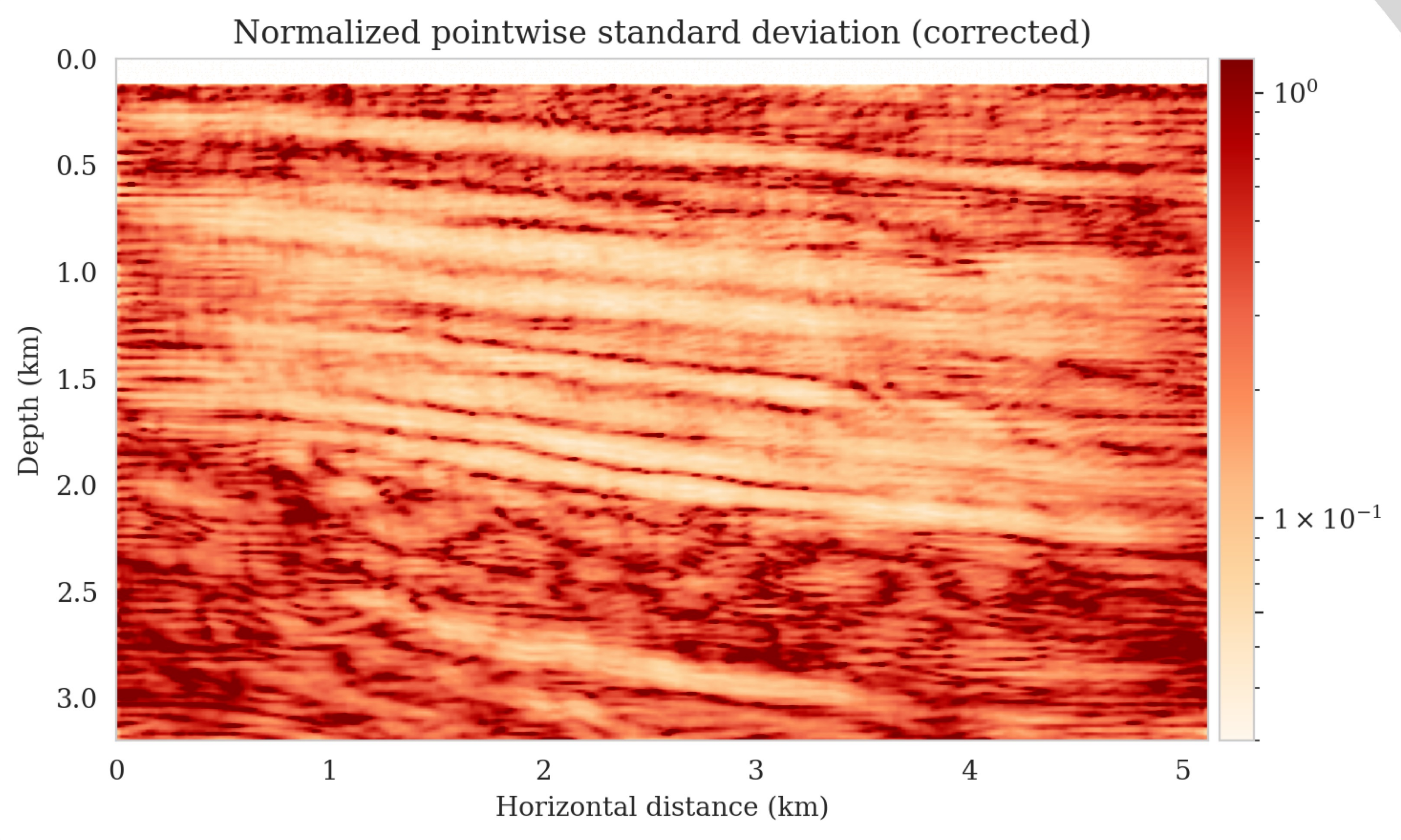
SNR 6.29 dB


Conditional mean estimate (corrected)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)

SNR 10.36 dB



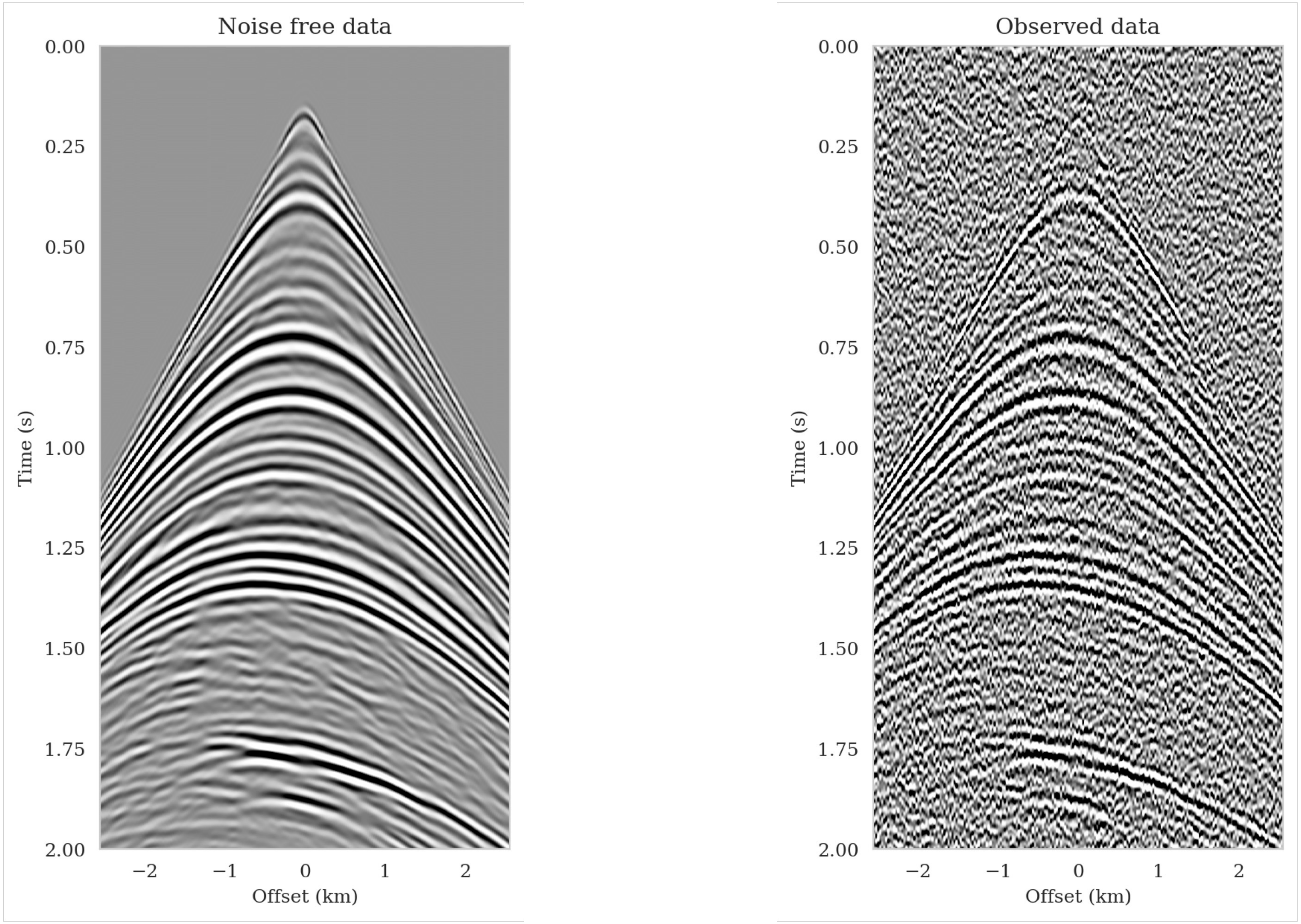
previously unseen (test) seismic image



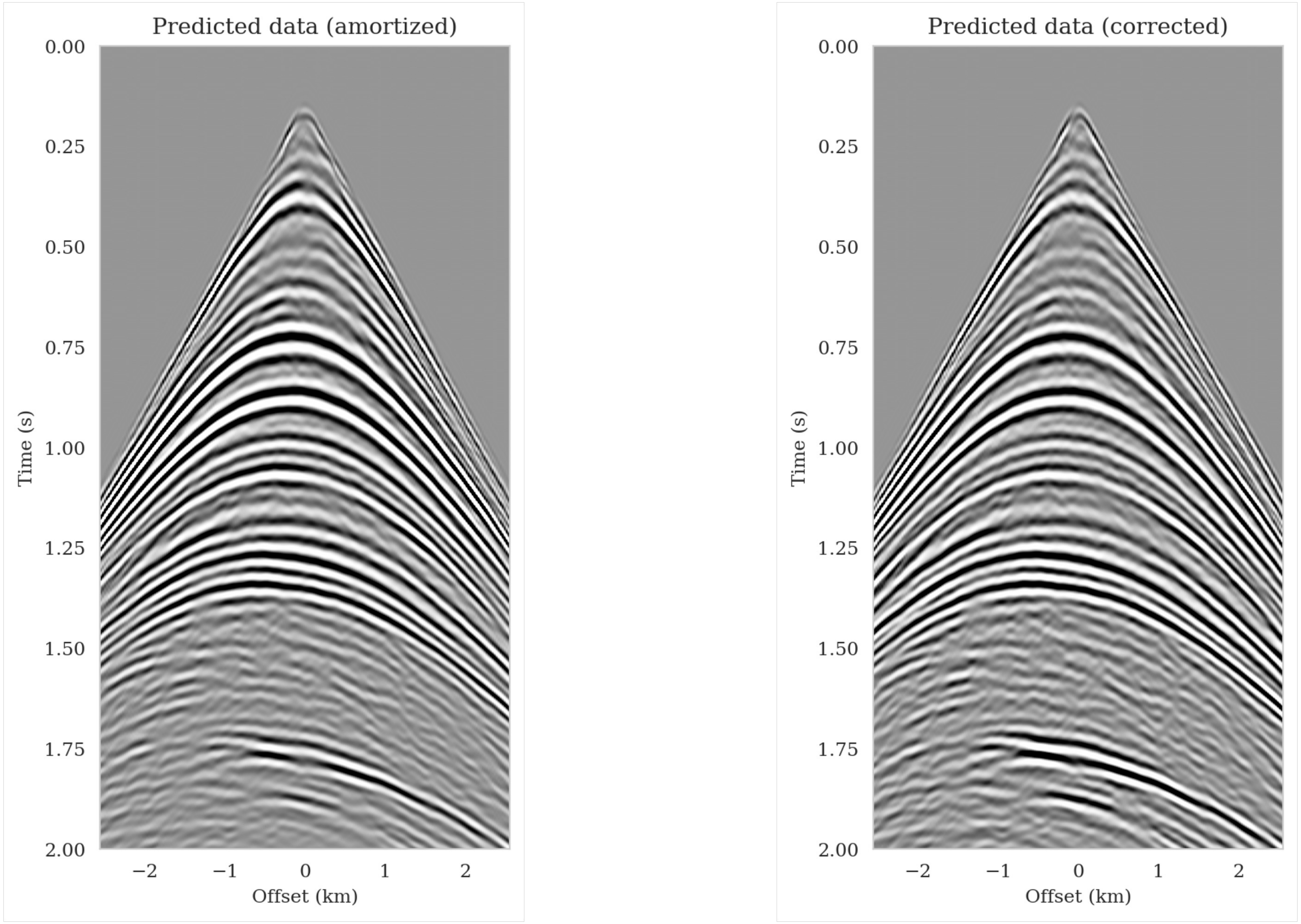
normalized by the the envelope of the conditional mean

Data space QC

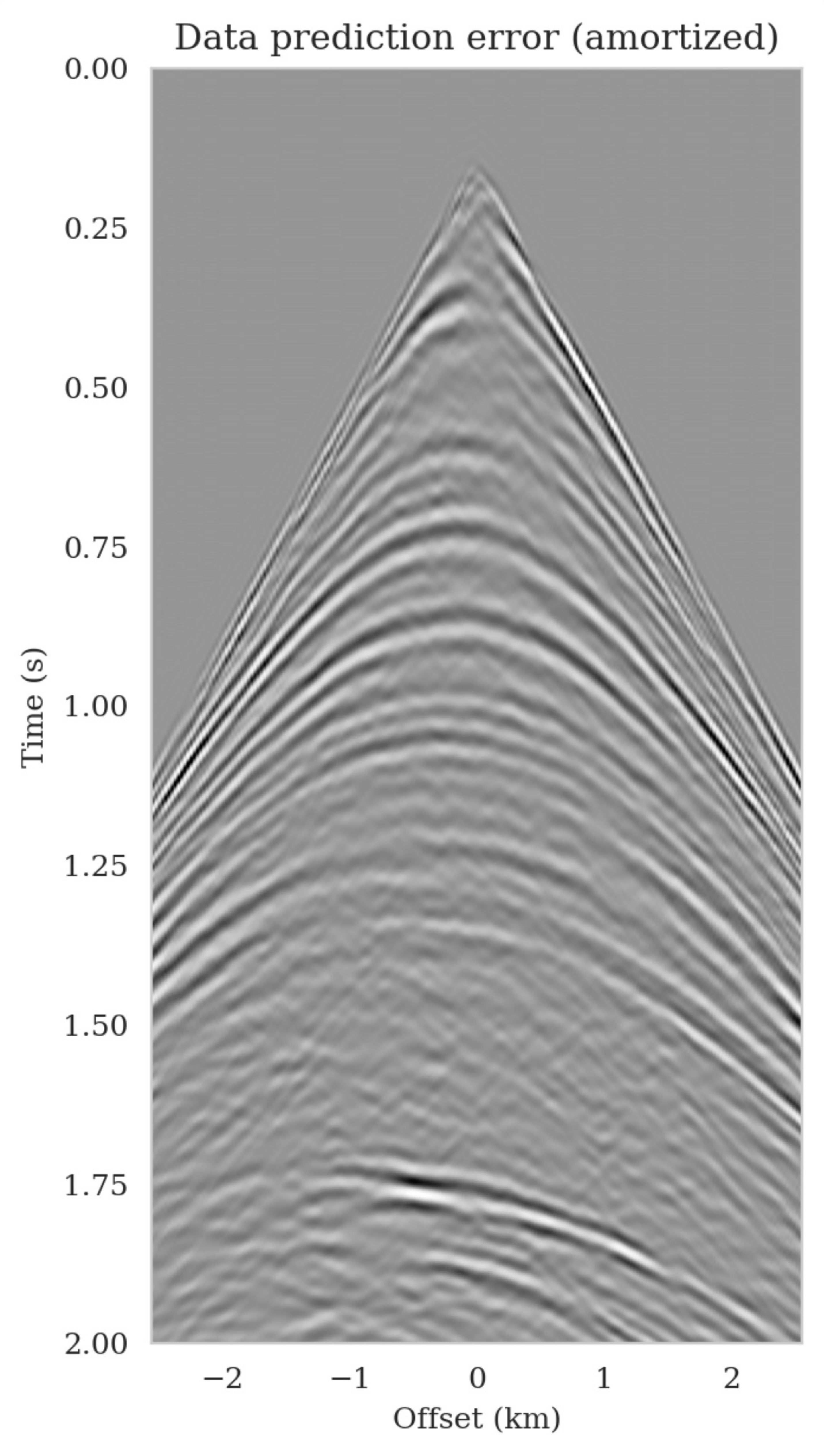
inspect data residual before and after latent distribution correction

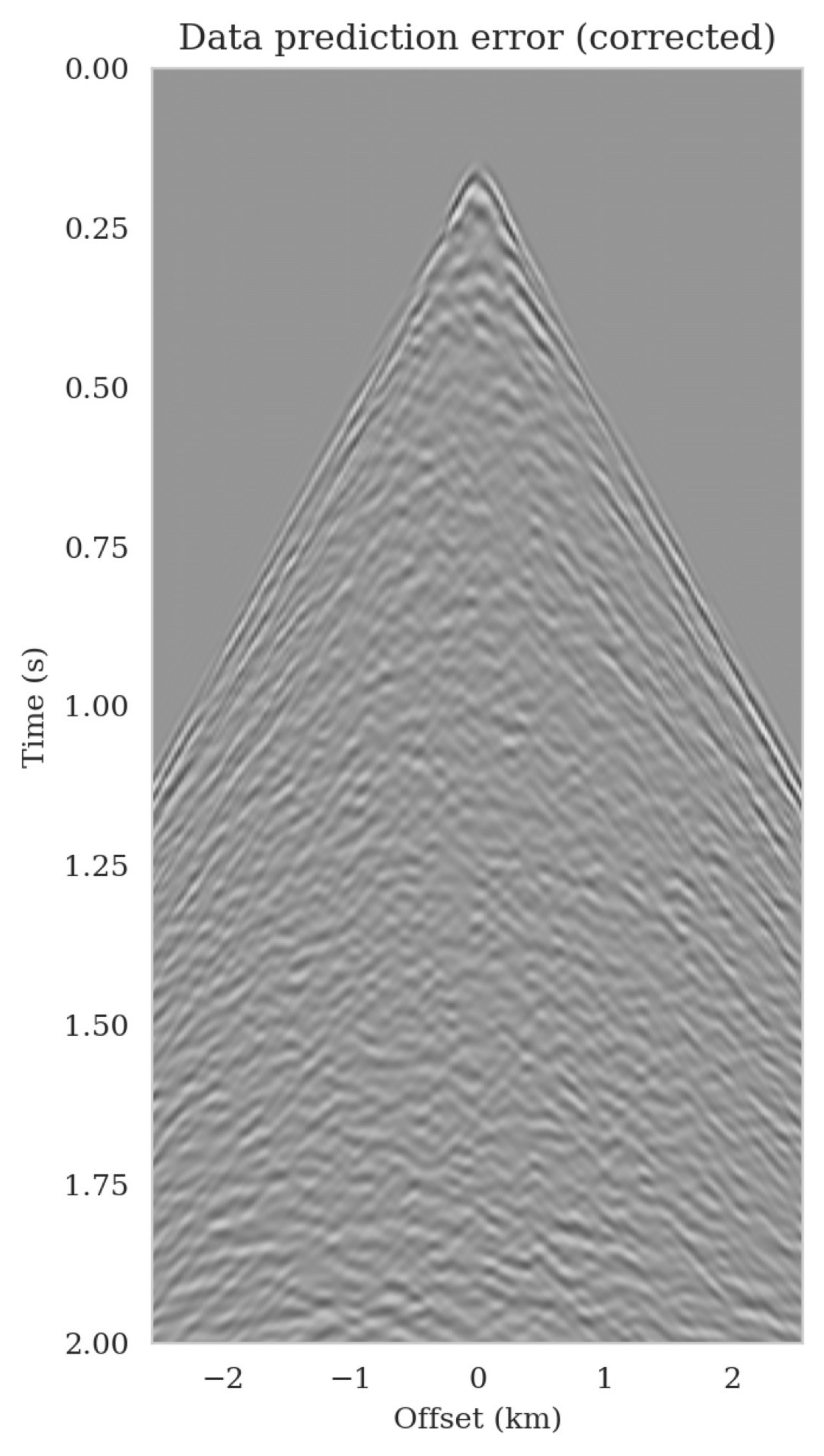


(left) noise free data (right) noisy data with SNR $-2.78\,\mathrm{dB}$



predicted data (left) amortized, SNR 11.62 dB (right) corrected, SNR 16.57 dB





data residual of (left) amortized (right) corrected

Multi-experiment inverse problems

Find $\mathbf{x}$ such that

$$\mathbf{y}_i = \mathcal{F}_i(\mathbf{x}) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}), \quad i = 1, \ldots, N$$

- observed data $\mathbf{y} = \{\mathbf{y}_i\}_{i=1}^{N}, \ \mathbf{y}_i \in \mathcal{Y}$
- unknown quantity $\mathbf{x} \in \mathcal{X}$
- expensive-to-evaluate forward operator $\mathcal{F}_i : \mathcal{X} \to \mathcal{Y}$
- noise and/or modeling error $\boldsymbol{\epsilon}_i$
- noise covariance $\sigma^2\mathbf{I}$

Negative-log posterior

$$-\log p_{\text{post}}(\mathbf{x} \mid \mathbf{y}) = -\sum_{i=1}^{N} \log p_{\text{like}}(\mathbf{y}_i \mid \mathbf{x}) - \log p_{\text{prior}}(\mathbf{x}) + \log p_{\text{data}}(\mathbf{y})$$

$$= \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left\| \mathbf{y}_i - \mathcal{F}_i(\mathbf{x}) \right\|_2^2 - \log p_{\text{prior}}(\mathbf{x}) + \text{const.}$$

likelihood function following Gaussian noise assumption

prior distribution $p_{\text{prior}}(\mathbf{x})$

constant term (data density) independent of $\mathbf{x}$

Amortized variational inference with normalizing flows

$$\phi^* = \arg\min_{\phi} \mathbb{E}_{\mathbf{y}\sim p_{\text{data}}(\mathbf{y})}\left[\mathbb{KL}\Big(p_{\text{post}}(\mathbf{x} \mid \mathbf{y}) \,||\, p_\phi(\mathbf{x} \mid \mathbf{y})\Big)\right]$$

$$= \arg\min_{\phi} \mathbb{E}_{\mathbf{x},\mathbf{y}\sim p(\mathbf{x},\mathbf{y})}\left[\underbrace{\frac{1}{2}\left\|f_\phi(\mathbf{x};\mathbf{y})\right\|_2^2}_{\text{normalizes the input}} - \underbrace{\log\left|\det \nabla_{\mathbf{x}} f_\phi(\mathbf{x};\mathbf{y})\right|}_{\text{entropy regularization}}\right]$$

entropy regularization e.g., avoids $f_\phi(\mathbf{x};\mathbf{y}) \equiv \mathbf{0}$

$f_\phi(\,\cdot\,;\mathbf{y}) : \mathcal{X} \to \mathcal{Z}$ an invertible neural net

approximate expectation with available model and data pairs $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^n \sim p(\mathbf{x}, \mathbf{y})$

negligible computational cost of $\det \nabla_{\mathbf{x}} f_\phi(\mathbf{x};\mathbf{y})$'s gradient due to f_ϕ's architecture

Jakob Kruse, Gianluca Detommaso, Robert Scheichl, and Ullrich Köthe. "HINT: Hierarchical Invertible Neural Transport for Density Estimation and Bayesian Inference". In: *Proceedings of AAAI-2021* (2021). URL: https://arxiv.org/pdf/1905.10687.pdf.

Amortized variational inference—sampling

For previously unseen data $\mathbf{y}_{\text{obs}} \sim p_{\text{data}}(\mathbf{y})$

$$f_{\phi^*}^{-1}(\mathbf{z}; \mathbf{y}_{\text{obs}}) \sim p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}}), \quad \mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$$

learned prior and posterior distributions

fast and cheap amortized posterior sampling given new observation $\mathbf{y}_{\text{obs}} \sim p_{\text{data}}(\mathbf{y})$

no optimization or MCMC sampling for new observations

Jakob Kruse, Gianluca Detommaso, Robert Scheichl, and Ullrich Köthe. "HINT: Hierarchical Invertible Neural Transport for Density Estimation and Bayesian Inference". In: *Proceedings of AAAI-2021* (2021). URL: https://arxiv.org/pdf/1905.10687.pdf.

Ali Siahkoohi and Felix J. Herrmann. "Learning by example: fast reliability-aware seismic imaging with normalizing flows". Apr. 2021. URL: https://arxiv.org/pdf/2104.06255.pdf.

Implications of imperfect pretraining

SLIM

Inaccurate normalization through the conditional normalizing flow

- $p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}}) \neq \mathrm{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$
- $p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}})$ distribution of $\mathbf{z} = f_\phi(\mathbf{x}; \mathbf{y}_{\text{obs}})$ for $\mathbf{x} \sim p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}}),\ \mathbf{y}_{\text{obs}} \sim p_{\text{data}}(\mathbf{y})$

Inaccurate posterior sampling given $\mathbf{z} \sim \mathrm{N}(\mathbf{z} \mid \mathbf{0}, \mathbf{I})$ as input

- $p_\phi(\mathbf{x} \mid \mathbf{y}_{\text{obs}}) \neq p_{\text{post}}(\mathbf{x} \mid \mathbf{y}_{\text{obs}})$
- $p_\phi(\mathbf{x} \mid \mathbf{y}_{\text{obs}})$ predicted posterior distribution

---

Stefan T Radev, Ulf K Mertens, Andreas Voss, Lynton Ardizzone, and Ullrich Köthe. "BayesFlow: Learning complex stochastic models with invertible neural networks". In: *IEEE transactions on neural networks and learning systems* (2020).

Latent distribution correction

$$\mathbf{z} \sim \mathrm{N}\big(\mathbf{z} \mid \mathbf{0}, \mathbf{I}\big) \quad \longrightarrow \quad \mathbf{z} \sim \mathrm{N}\big(\mathbf{z} \mid \boldsymbol{\mu}, \mathrm{diag}(\mathbf{s})^2\big)$$

For certain classes of distribution shifts

- $p_\phi(\mathbf{z} \mid \mathbf{y})$ not too far from normal distribution
- improving the initial approximation $p_\phi(\mathbf{z} \mid \mathbf{y}) \approx \mathrm{N}\big(\mathbf{z} \mid \mathbf{0}, \mathbf{I}\big)$

Physics-based latent distribution correction

$$\min_{\boldsymbol{\mu},\mathbf{s}} \mathbb{E}_{\mathbf{z}\sim \mathrm{N}(\mathbf{0},\mathbf{I})} \left[\frac{1}{2\sigma^2} \sum_{i=1}^{N} \left\| \mathbf{y}_{\text{obs},i} - \mathcal{F}_i \circ f_{\phi}^{-1}(\mathbf{s} \odot \mathbf{z} + \boldsymbol{\mu}; \mathbf{y}_{\text{obs}}) \right\|_2^2 \right.$$

$$\left. + \frac{1}{2} \left\| \mathbf{s} \odot \mathbf{z} + \boldsymbol{\mu} \right\|_2^2 - \log \left| \det \mathrm{diag}(\mathbf{s}) \right| \right]$$

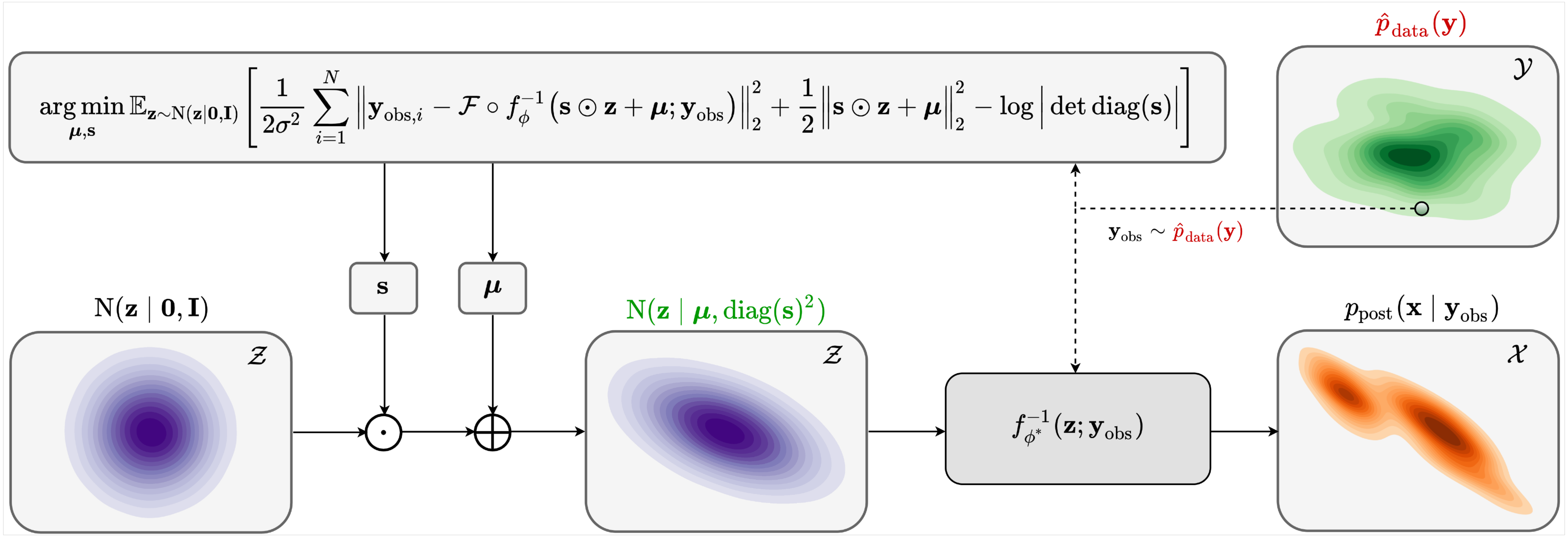
initializing with $\boldsymbol{\mu} = \mathbf{0}$ and $\mathrm{diag}(\mathbf{s})^2 = \mathbf{I}$

initialization acts as a warm-start and an implicit regularization

expected to be solved relatively cheaply due to the amortization of f_ϕ

non-amortized, i.e., specific to one set of observations $\mathbf{y}_{\text{obs}} \sim \widehat{p}_{\text{data}}(\mathbf{y})$

Non-amortized latent distribution correction



Conclusions

Uncertainty quantification is rendered impractical when

- the forward operators are expensive to evaluate
- the problem is high dimensional

Amortized variational inference with physics-based latent distribution correction

- can lead to orders of magnitude computational improvements compared to MCMC and traditional variational inference methods
- limits the adverse affects of data distribution shifts
- provides fast (same cost as 5 RTMs) and reliable posterior inference

Code

https://github.com/slimgroup/InvertibleNetworks.jl

https://github.com/slimgroup/ConditionalNFs4Imaging.jl

Acknowledgment

This research was carried out with the support of

Georgia Research Alliance and partners of the ML4Seismic Center