Reliable amortized variational inference with physics-based latent distribution correction

Ali Siahkoohi [1] Gabrio Rizzuti [2] Rafael Orozco [3] Felix J. Herrmann [3]

[1]Formerly Georgia Institute of Technology; Presently Rice University

[2]Utrecht University

[3]Georgia Institute of Technology

SLIM
Georgia Institute of Technology

Ali Siahkoohi, Gabrio Rizzuti, Rafael Orozco, and Felix J. Herrmann. “Reliable amortized variational inference with physics-based latent distribution correction”. July 2022. URL: https://arxiv.org/abs/2207.11640.

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 $\sigma^2\mathbf{I}$

Bayesian inverse problems

Represent the solution as a distribution over the model space

i.e., posterior distribution

Albert Tarantola. *Inverse problem theory and methods for model parameter estimation*. SIAM, 2005. ISBN: 978-0-89871-572-9. DOI: 10.1137/1.9780898717921.

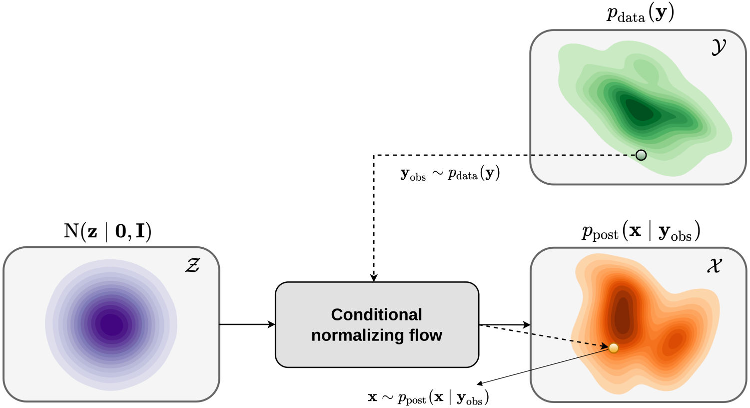
Challenges of solving Bayesian inverse problems

Choosing a prior distribution

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

Computational cost

- costly forward operator
- high dimensional sampling/integration



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.

Amortized variational inference w/ normalizing flows

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

$$= \arg\min_{\phi} \mathbb{E}_{\mathbf{x},\mathbf{y}\sim p(\mathbf{x},\mathbf{y})}\left[\underbrace{\frac{1}{2}\left\|f_\phi(\mathbf{x};\mathbf{y})\right\|_2^2}_{\text{normalizes the input}} - \underbrace{\log\left|\det \nabla_{\mathbf{x}} f_\phi(\mathbf{x};\mathbf{y})\right|}_{\substack{\text{entropy regularization} \\ \text{e.g., avoids } f_\phi(\mathbf{x};\mathbf{y}) \equiv \mathbf{0}}}\right]$$

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

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.

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

---

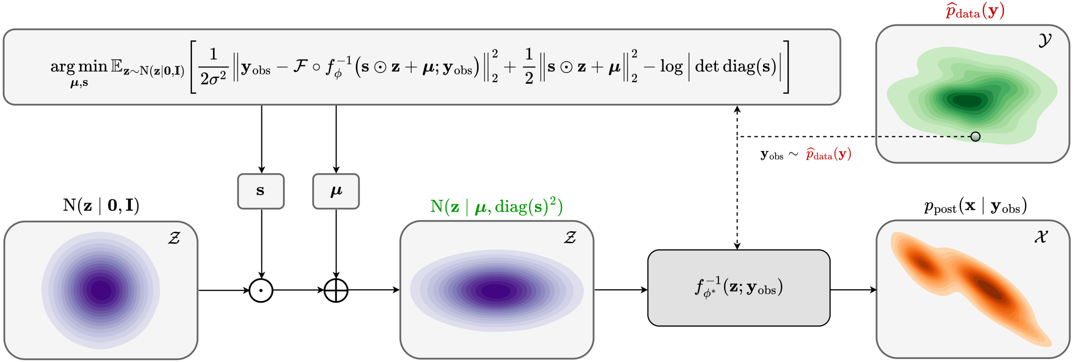
Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. "Invertible generative models for inverse problems: mitigating representation error and dataset bias". In: *Proceedings of the 37th International Conference on Machine Learning*. 2020, pp. 399–409.

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`.

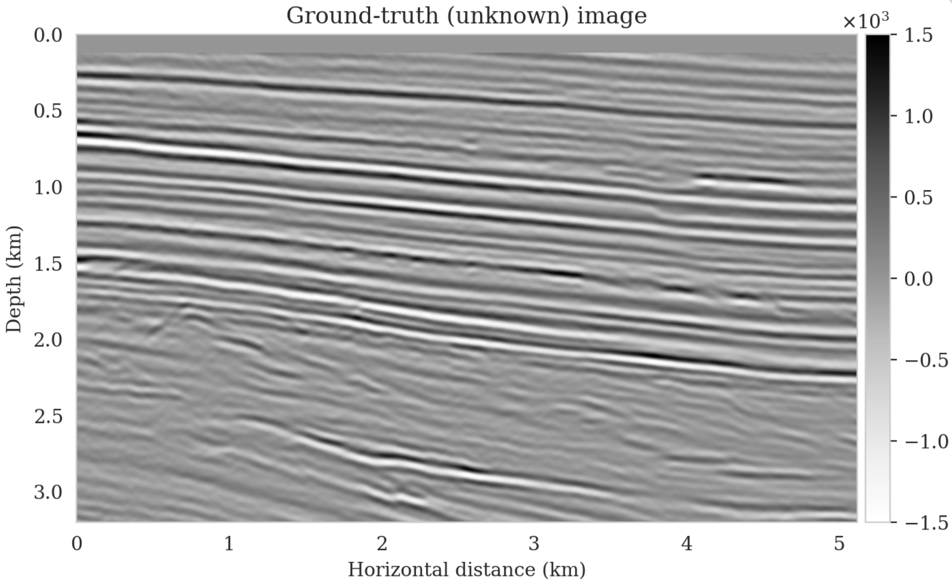
AmirEhsan Khorashadizadeh et al. "Conditional Injective Flows for Bayesian Imaging". In: *arXiv preprint arXiv:2204.07664* (2022).

Physics-based latent distribution correction

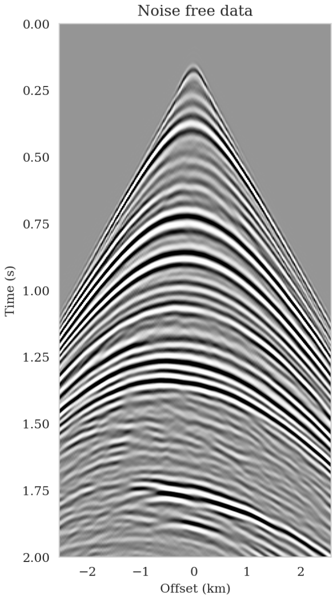


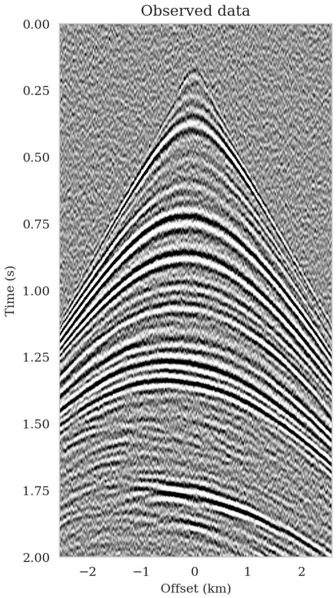
Seismic imaging example

in-distribution amortized posterior sampling

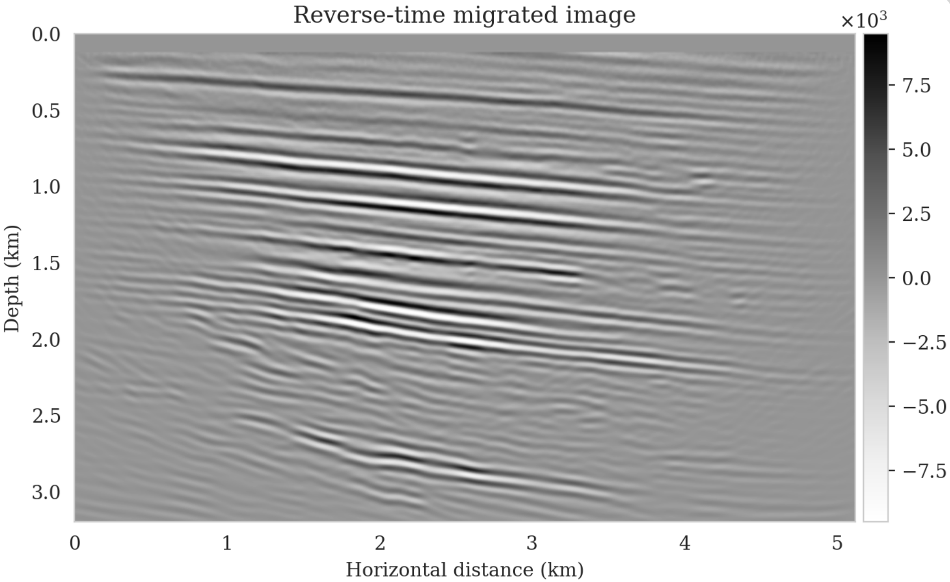


previously unseen (test) seismic image

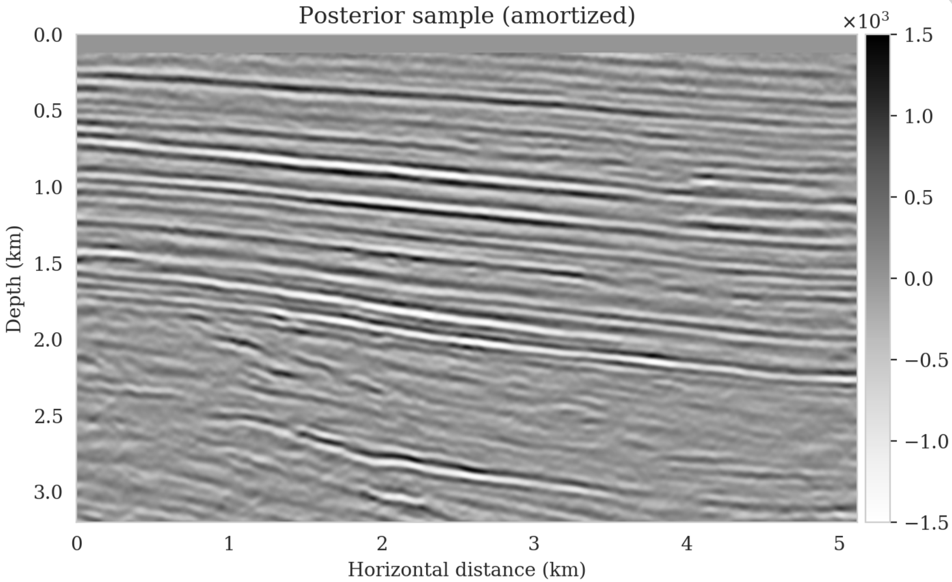




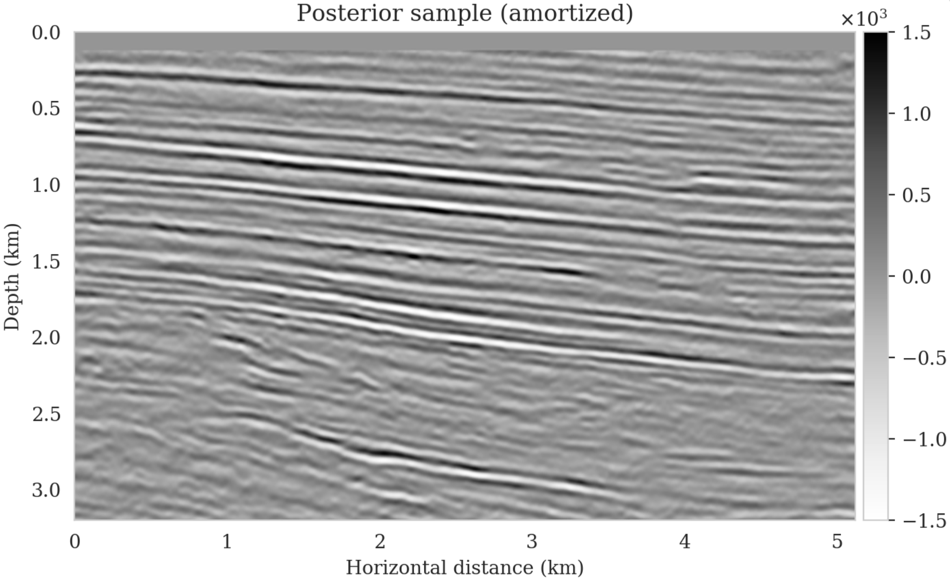
(left) noise free data (right) noisy data


Reverse-time migrated image
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
×10^3
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5

SNR -12.17 dB

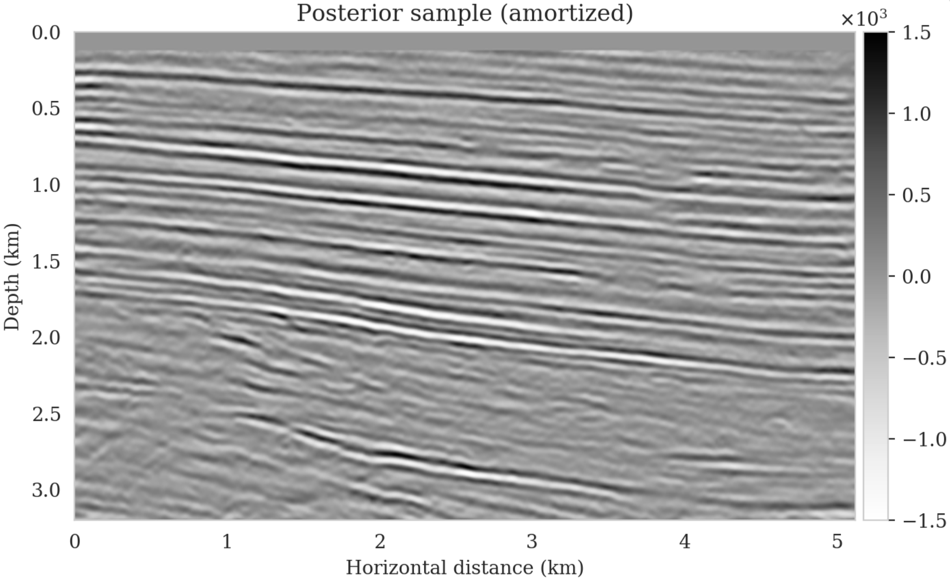

Posterior sample (amortized)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 8.54 dB



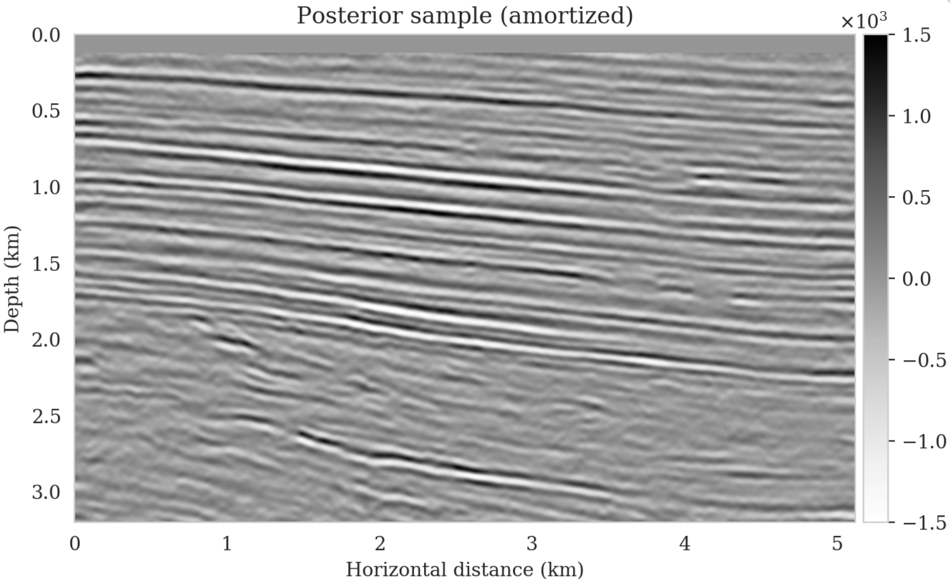
Posterior sample (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 8.49 dB



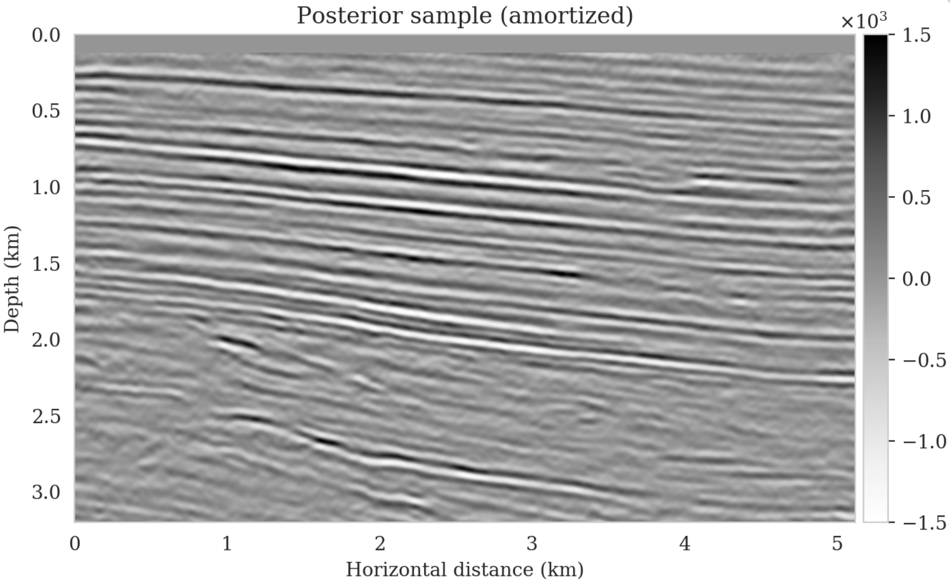
Posterior sample (amortized)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 8.42 dB

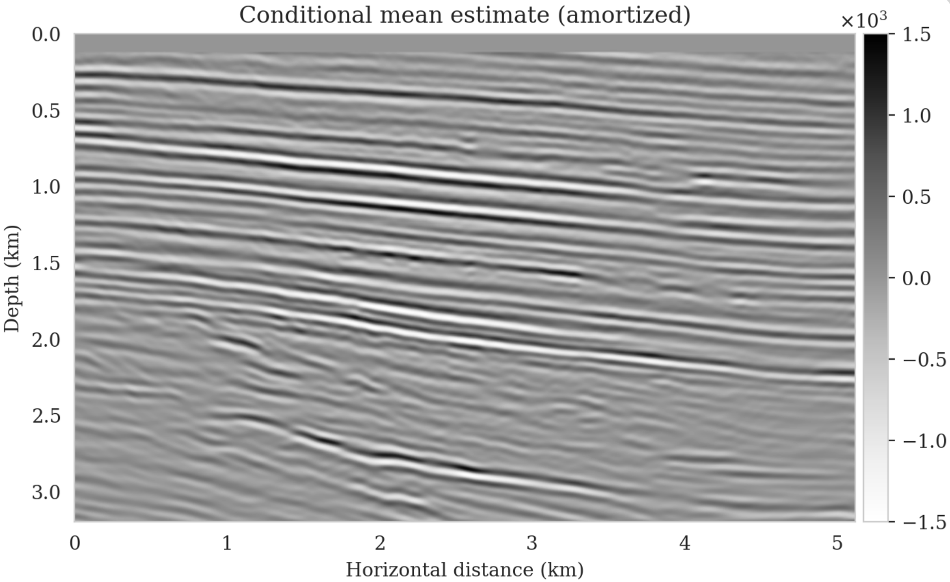


Posterior sample (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

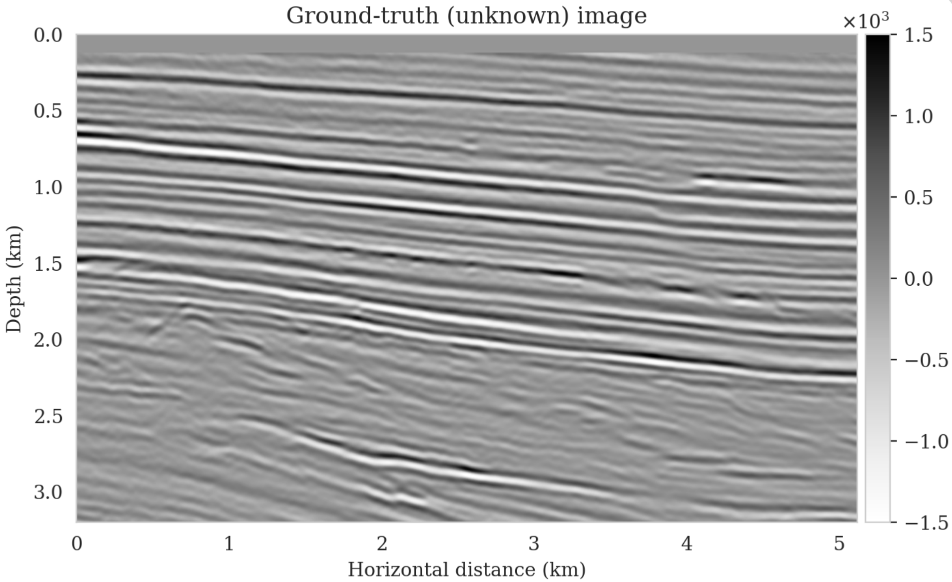
SNR 8.59 dB


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

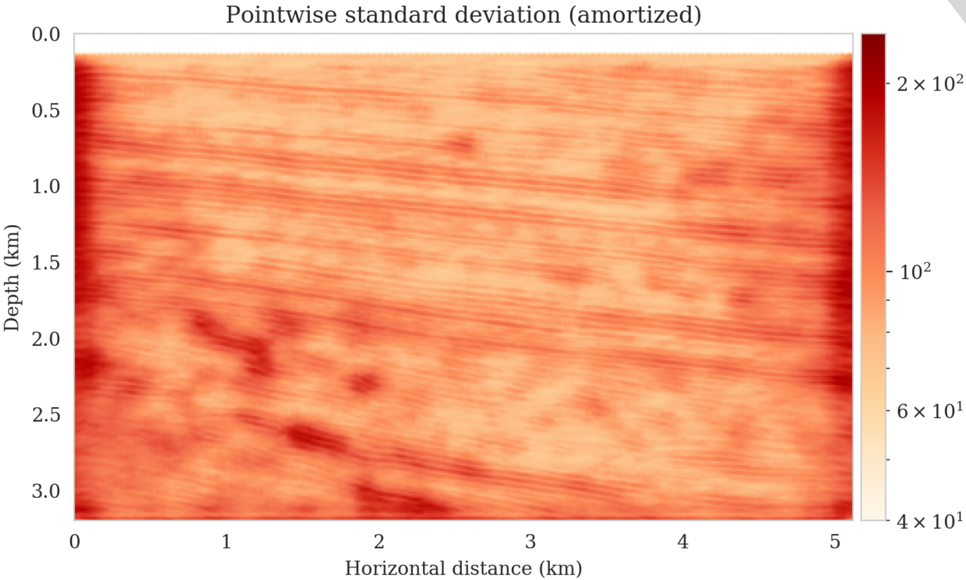
SNR 8.54 dB


Conditional mean estimate (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 11.24 dB



previously unseen (test) seismic image



Pointwise standard deviation (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)
2×10^2
10^2
6×10^1
4×10^1

SLIM

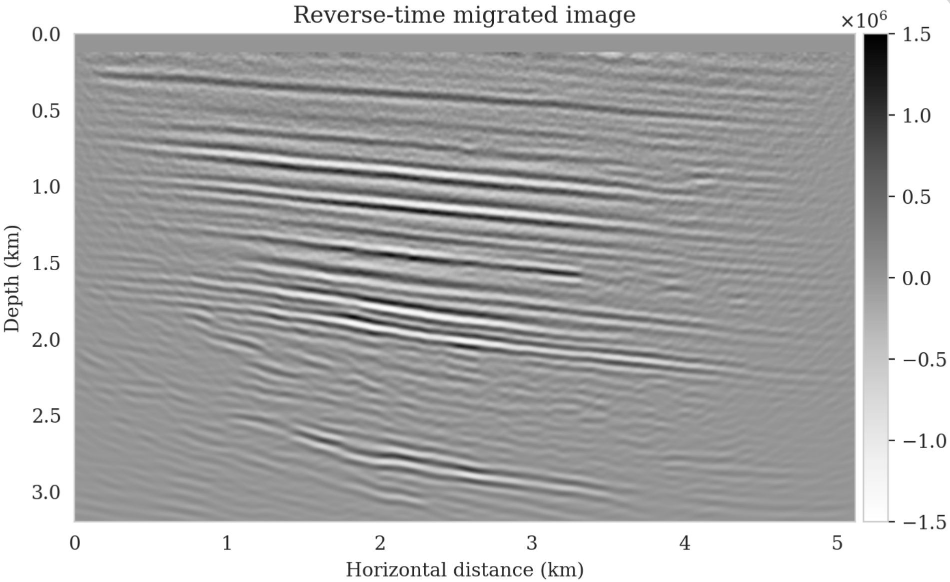
Introducing distribution shifts

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

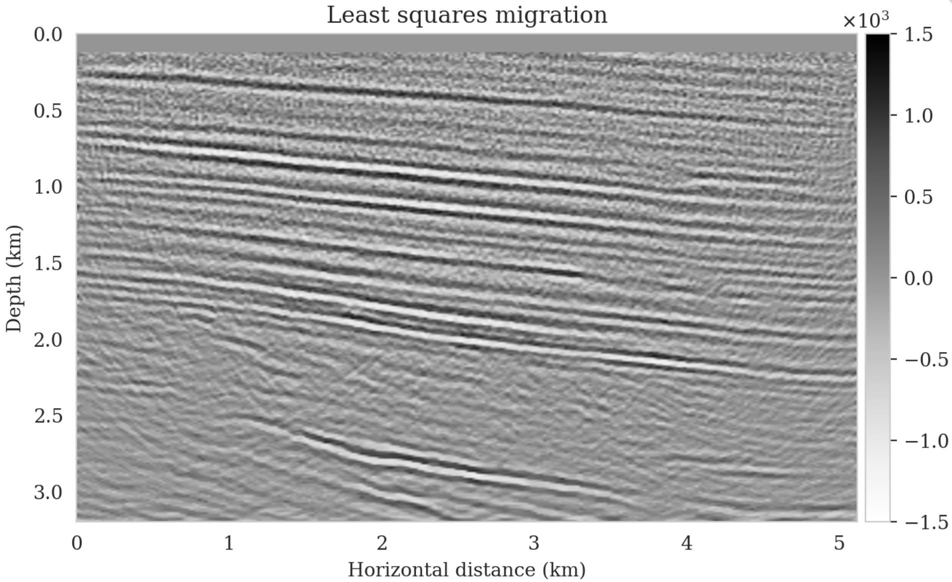
$4\times$ less sources

Physics-based latent distribution correction

computational cost: approximately $5\times$ RTMs

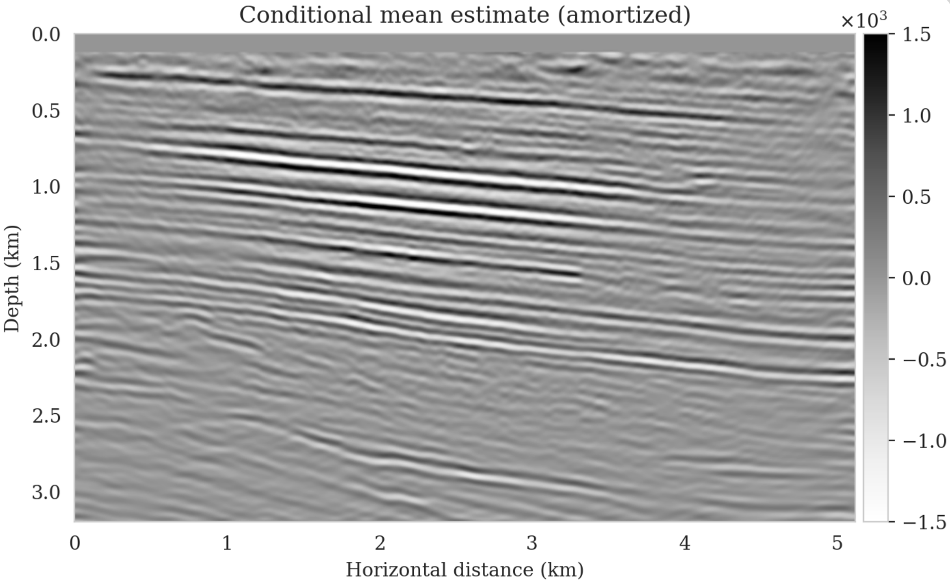


SNR $-8.22\,\mathrm{dB}$



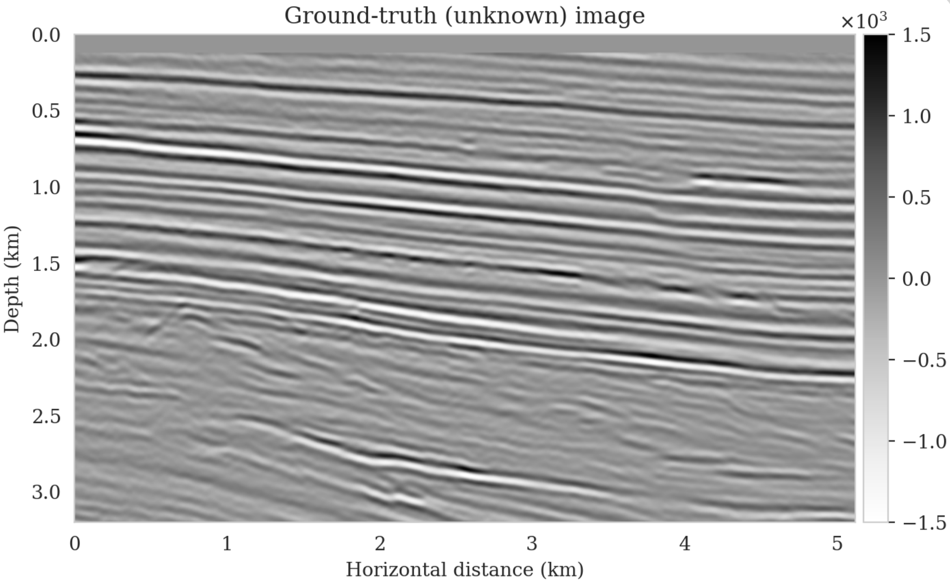
Least squares migration
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 6.90 dB

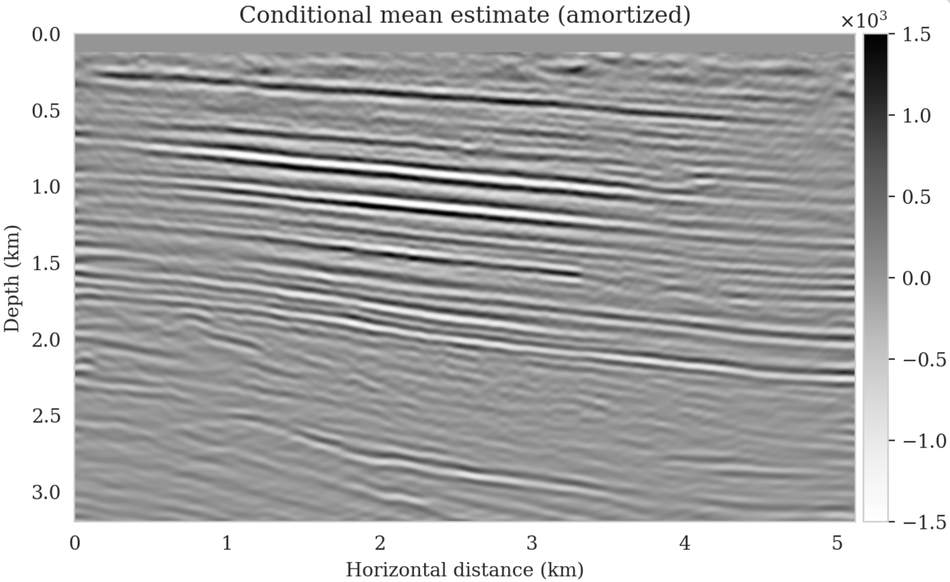


Conditional mean estimate (amortized)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

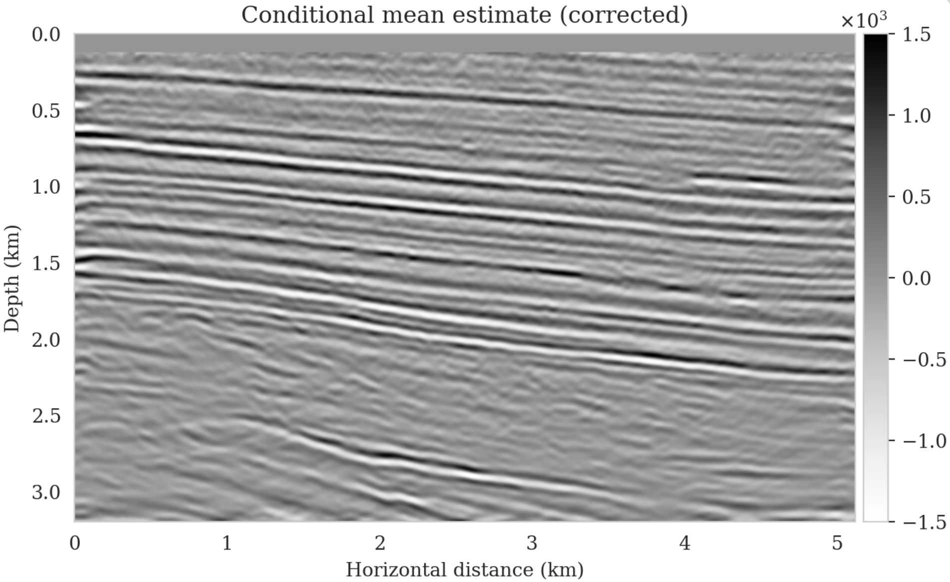
SNR 6.29 dB



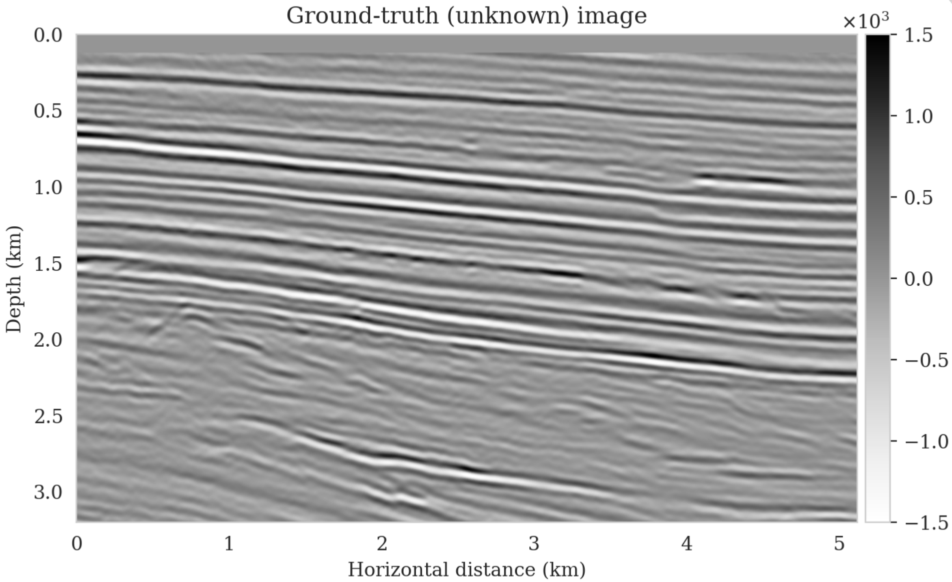
previously unseen (test) seismic image


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

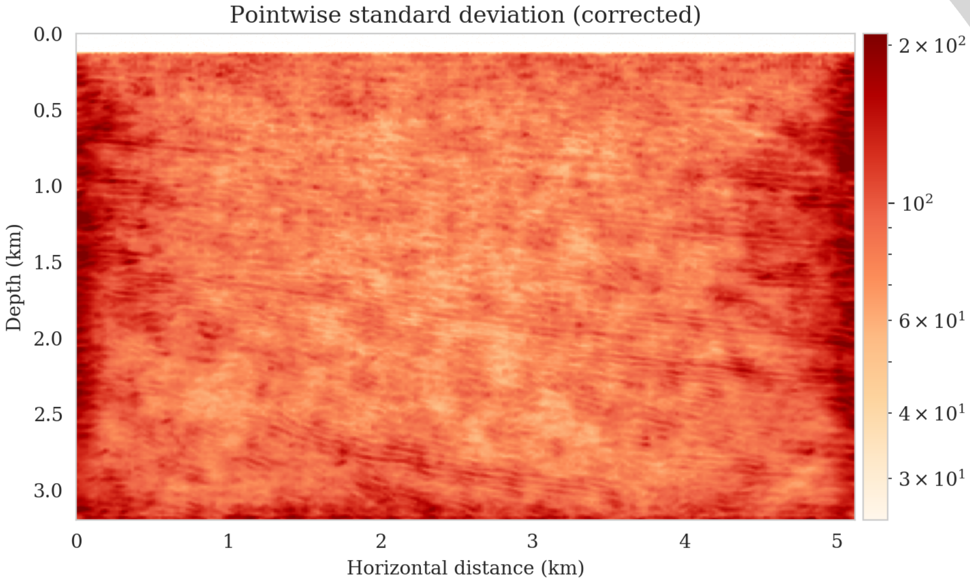
SNR 6.29 dB


Conditional mean estimate (corrected)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
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)

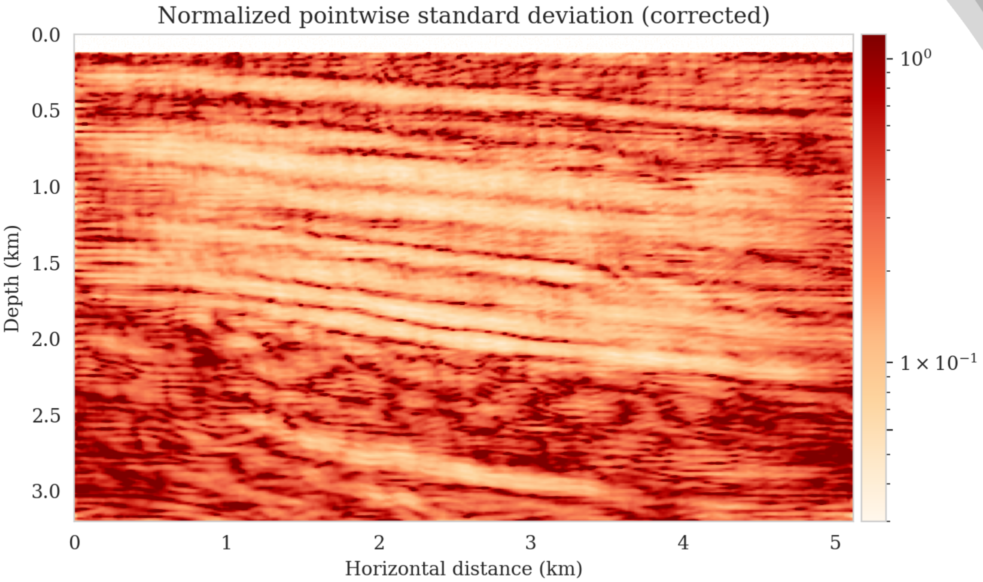
SNR 10.36 dB



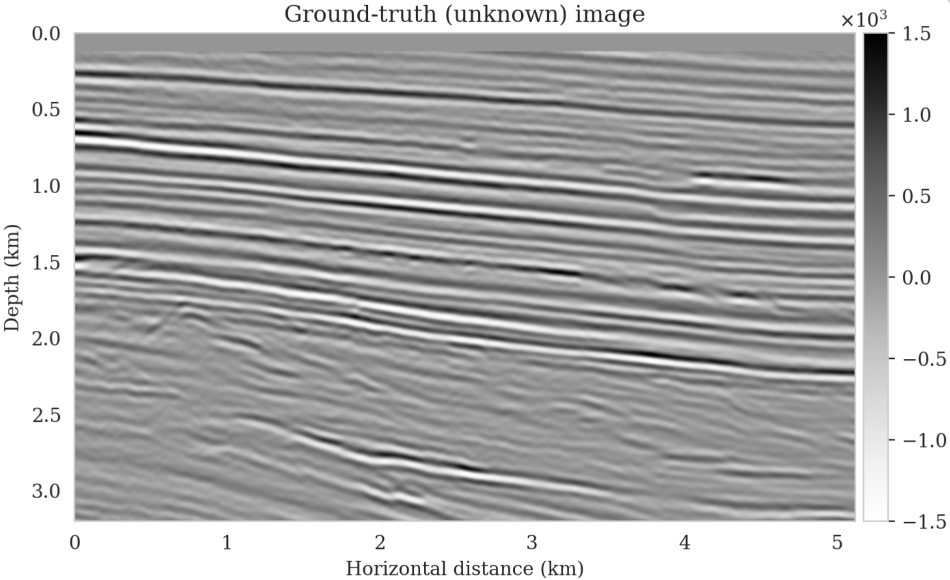
previously unseen (test) seismic image



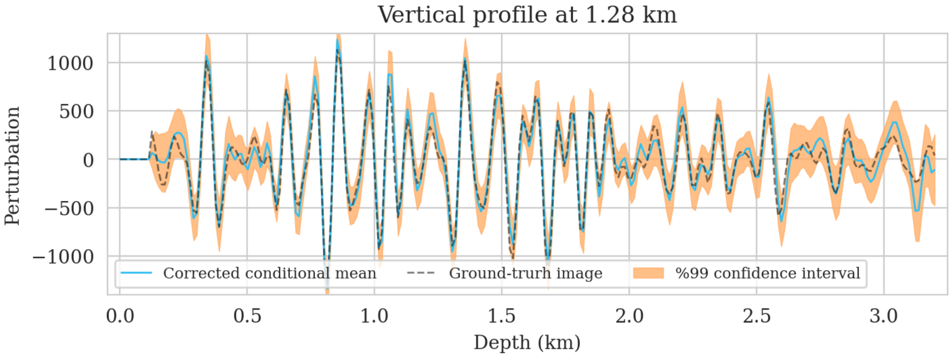
Pointwise standard deviation (corrected)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
2×10^2
10^2
6×10^1
4×10^1
3×10^1
SLIM



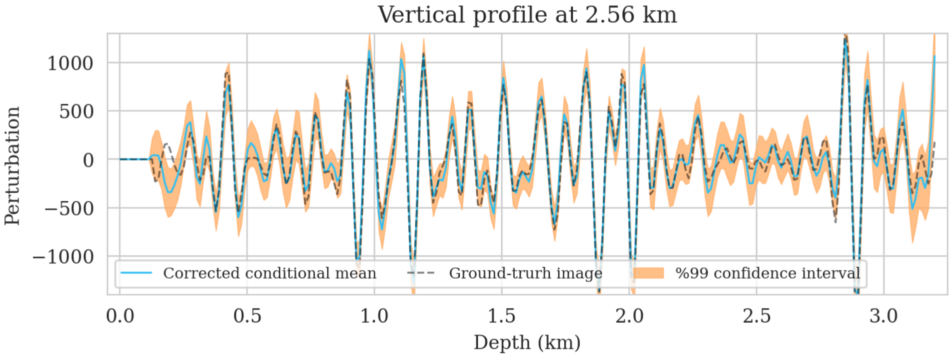
normalized by the the envelope of the conditional mean



previously unseen (test) seismic image



Vertical profile at 1.28 km
Perturbation
1000
500
0
−500
−1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
Corrected conditional mean
Ground-truth image
%99 confidence interval



Vertical profile at 2.56 km
Perturbation
1000
500
0
−500
−1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
Corrected conditional mean
Ground-truth image
%99 confidence interval

Data distribution shifts

inaccurate normalization through the conditional normalizing flow

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

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 relaxation

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

normalizing flows are invertible

- guarantees the existence of a latent correction that fits the posterior
- correction can be also learned by a normalizing flow

Konik Kothari, AmirEhsan Khorashadizadeh, Maarten de Hoop, and Ivan Dokmanić. "Trumpets: Injective flows for inference and inverse problems". In: *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*. Vol. 161. Proceedings of Machine Learning Research. PMLR, 2021, pp. 1269–1278.

Jay Whang, Erik Lindgren, and Alex Dimakis. "Composing normalizing flows for inverse problems". In: *International Conference on Machine Learning*. PMLR. 2021, pp. 11158–11169.

For the previously unseen out-of-distribution data $\mathbf{y}_{\text{obs}} \sim \widehat{p}_{\text{data}}(\mathbf{y})$

$$\min_{\boldsymbol{\mu},\mathbf{s}} \mathbb{KL}\left(\mathrm{N}\left(\mathbf{z} \mid \boldsymbol{\mu}, \mathrm{diag}(\mathbf{s})^2\right) \,||\, p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}})\right)$$

with

$$-\log p_\phi(\mathbf{z} \mid \mathbf{y}_{\text{obs}}) = \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left\|\mathbf{y}_{\text{obs},i} - \mathcal{F}_i \circ f_\phi(\mathbf{z}; \mathbf{y}_{\text{obs}})\right\|_2^2 + \frac{1}{2}\left\|\mathbf{z}\right\|_2^2 + \text{const.}$$

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

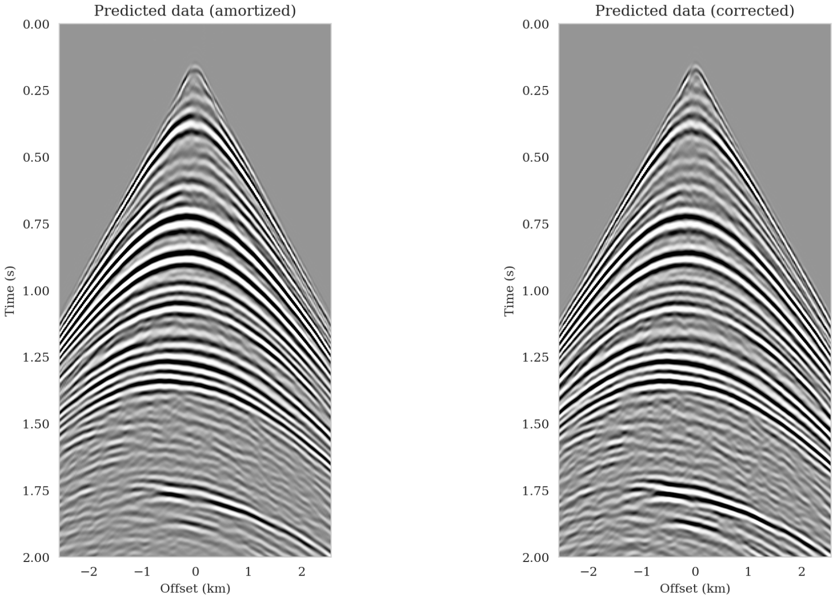
initializing with $\boldsymbol{\mu} = \mathbf{0}$ and $\operatorname{diag}(\mathbf{s})^2 = \mathbf{I}$

initialization acts as a **warm-start** and an implicit regularization

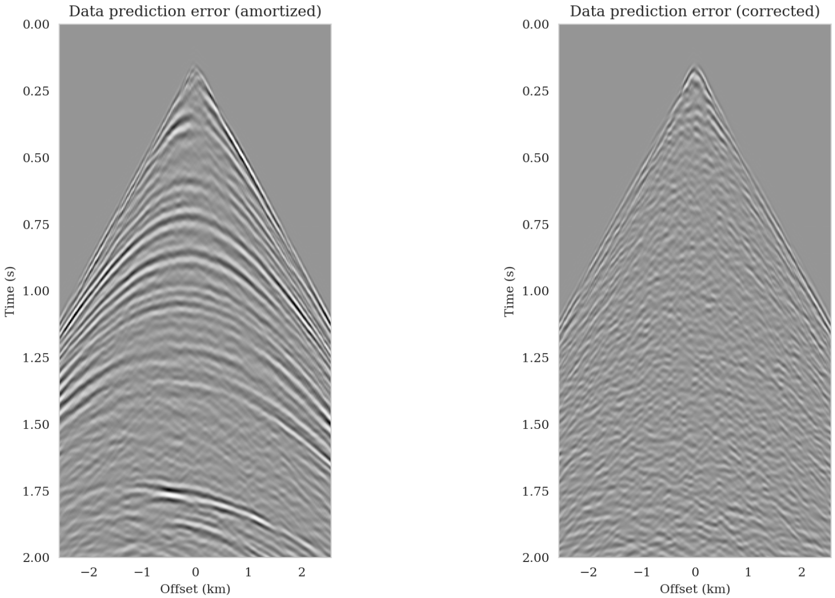
the pretrained f_{ϕ}^{-1} acts as a **nonlinear preconditioner** for the optimization

expected to be solved relatively cheaply due to the amortization of f_{ϕ}

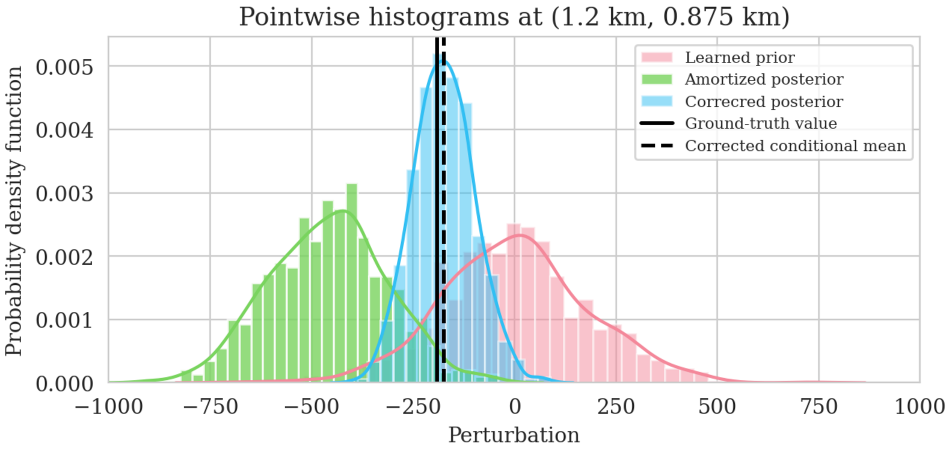
non-amortized, i.e., specific to one set of observations $\mathbf{y}_{\mathrm{obs}} \sim \widehat{p}_{\mathrm{data}}(\mathbf{y})$

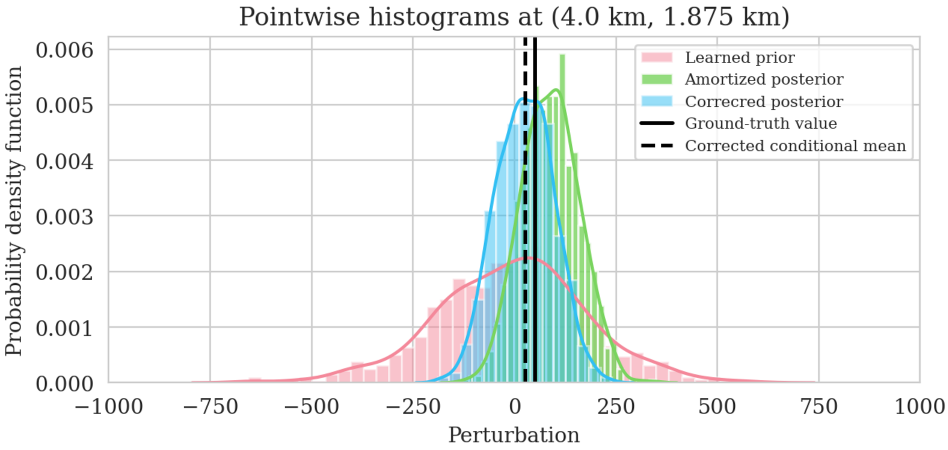


predicted data (left) amortized, SNR 11.62 dB (right) corrected, SNR 16.57 dB



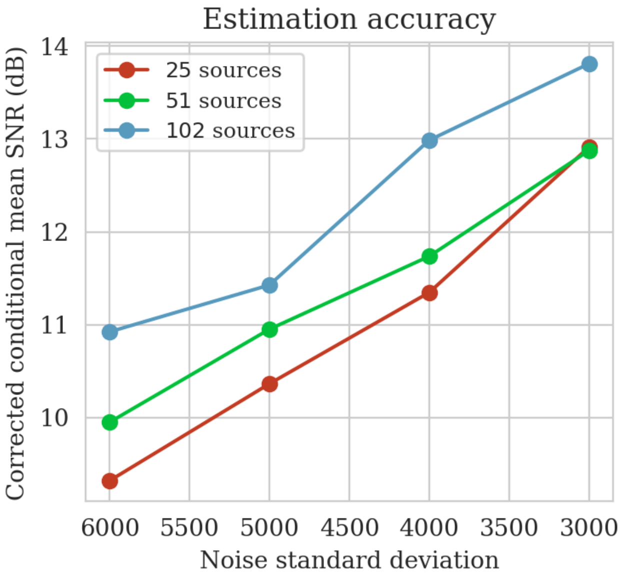
data residual of (left) amortized (right) corrected

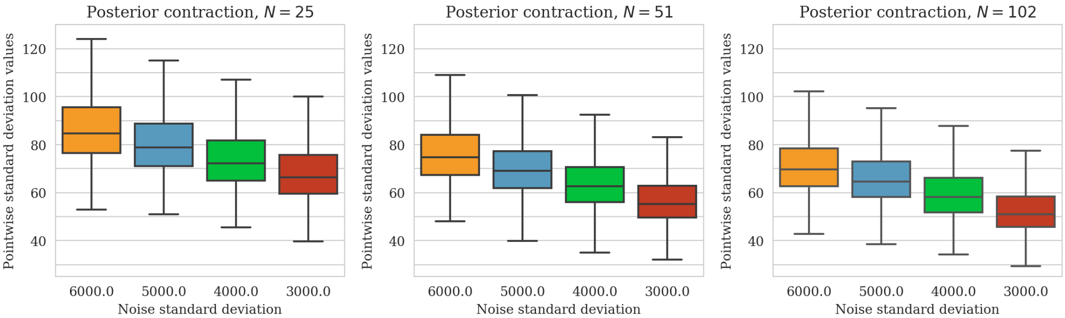

Pointwise histograms at (1.2 km, 0.875 km)
Probability density function
0.005
0.004
0.003
0.002
0.001
0.000
−1000
−750
−500
−250
0
250
500
750
1000
Perturbation
Learned prior
Amortized posterior
Correcred posterior
Ground-truth value
Corrected conditional mean
SLIM



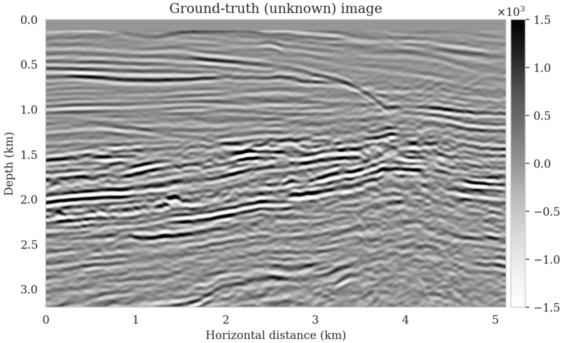
Pointwise histograms at (4.0 km, 1.875 km)
Probability density function
0.006
0.005
0.004
0.003
0.002
0.001
0.000
−1000
−750
−500
−250
0
250
500
750
1000
Perturbation
Learned prior
Amortized posterior
Correcred posterior
Ground-truth value
Corrected conditional mean

SLIM

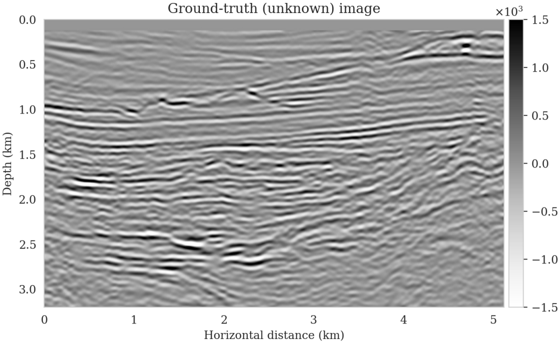

Estimation accuracy
25 sources
51 sources
102 sources
Corrected conditional mean SNR (dB)
14
13
12
11
10
6000
5500
5000
4500
4000
3500
3000
Noise standard deviation



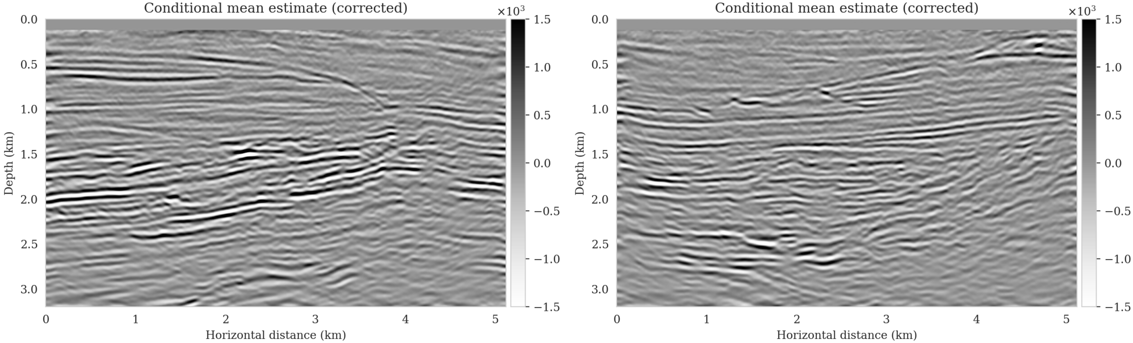
(left) $N = 25$, (middle) $N = 51$, and (right) $N = 102$.



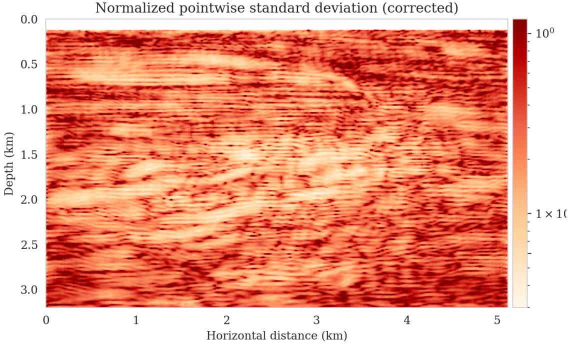
Ground-truth (unknown) image
×10³
Depth (km)
Horizontal distance (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5



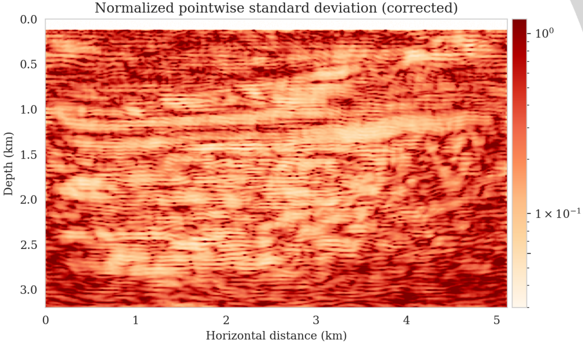
Ground-truth (unknown) image
×10³
Depth (km)
Horizontal distance (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5



SNR (left) 9.40 dB and (right) 9.11 dB



Normalized pointwise standard deviation (corrected)
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)
10^0
$1 \times 1($



Normalized pointwise standard deviation (corrected)
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)
10^0
1×10^{-1}

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

Acknowledgment

This research was carried out with the support of

Georgia Research Alliance and partners of the ML4Seismic Center

Contributions

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

https://github.com/slimgroup/ReliableAVI.jll

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