Low-cost uncertainty quantification for large-scale inverse problems

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.

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.

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 $\sigma^2 \mathbf{I}$

Amortized variational inference w/ normalizing flows

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

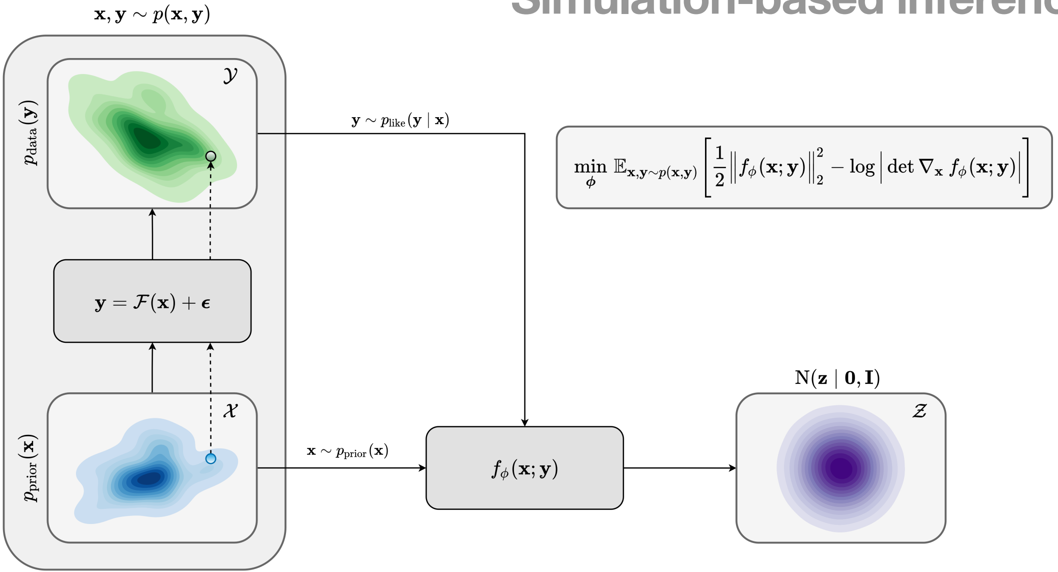
$$= \underset{\phi}{\arg\min}\ \mathbb{E}_{\mathbf{x},\mathbf{y}\sim p(\mathbf{x},\mathbf{y})}\left[\ \underbrace{\frac{1}{2}\left\|f_\phi(\mathbf{x};\mathbf{y})\right\|_2^2}_{\text{normalizes the input}} - \underbrace{\log\left|\det\nabla_{\mathbf{x}} f_\phi(\mathbf{x};\mathbf{y})\right|}_{\substack{\text{entropy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

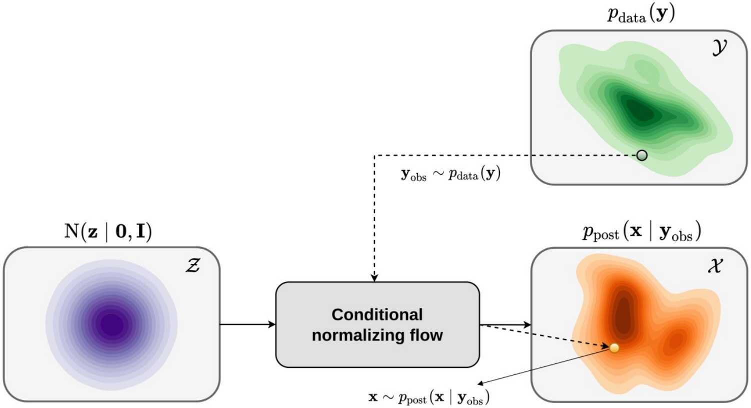
$f_\phi(\,\cdot\,;\mathbf{y}) : \mathcal{X}\to\mathcal{Z}$ an invertible neural net

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.

Simulation-based inference



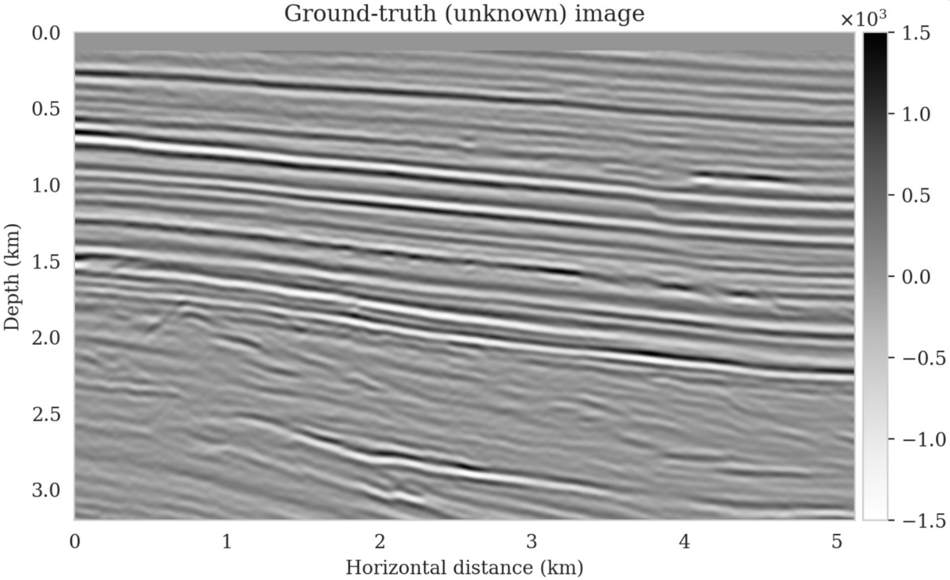


Jakob Kruse, Gianluca Detommaso, Robert Scheichl, and Ullrich Köthe. "HINT: Hierarchical Invertible Neural Transport for Density Estimation and Bayesian Inference". In: *Proceedings of AAAI-2021* (2021). URL: https://arxiv.org/pdf/1905.10687.pdf.

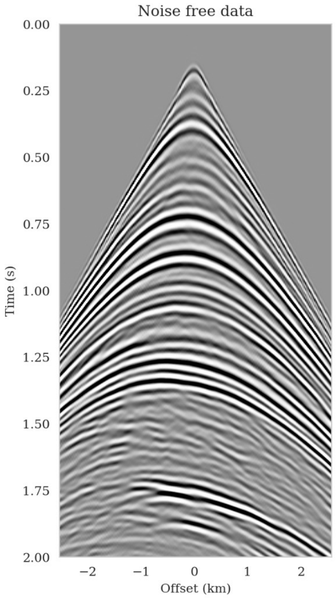
Ali Siahkoohi and Felix J. Herrmann. "Learning by example: fast reliability-aware seismic imaging with normalizing flows". Apr. 2021. URL: https://arxiv.org/pdf/2104.06255.pdf.

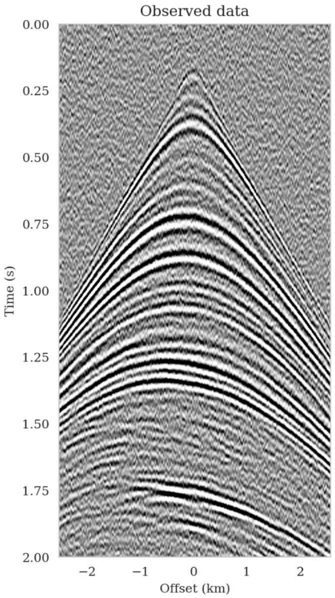
Seismic imaging example

in-distribution amortized posterior sampling

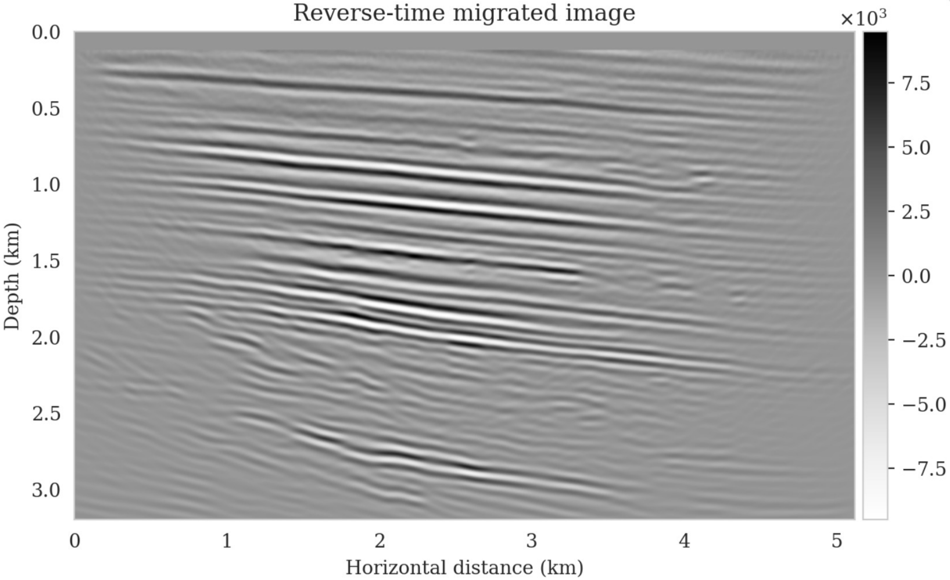


previously unseen (test) seismic image



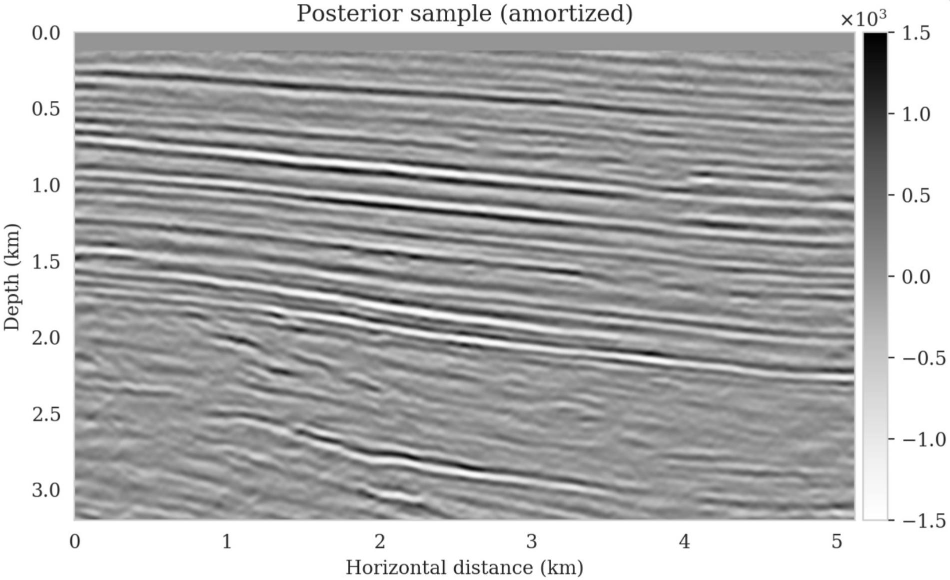


(left) noise free data (right) noisy data

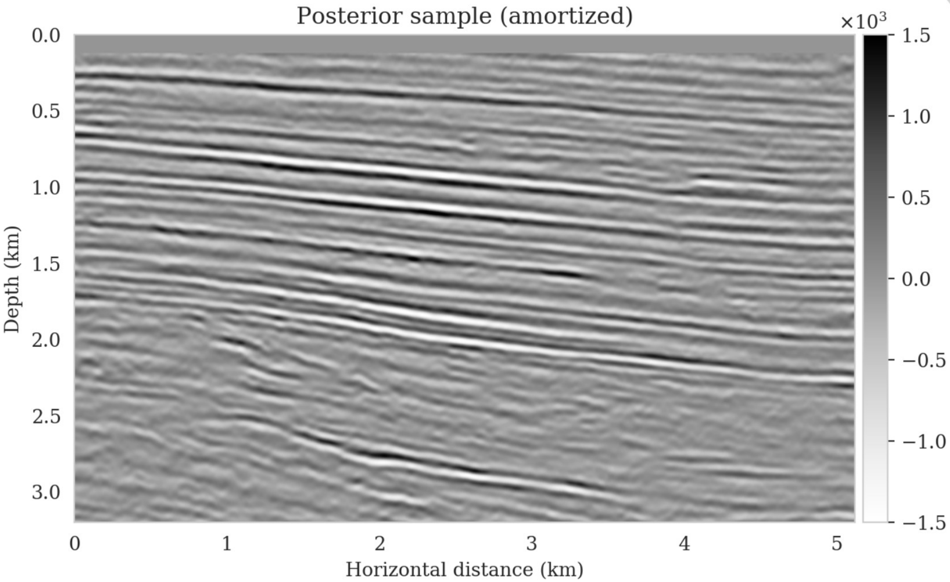


Reverse-time migrated image
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
×10^3
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5

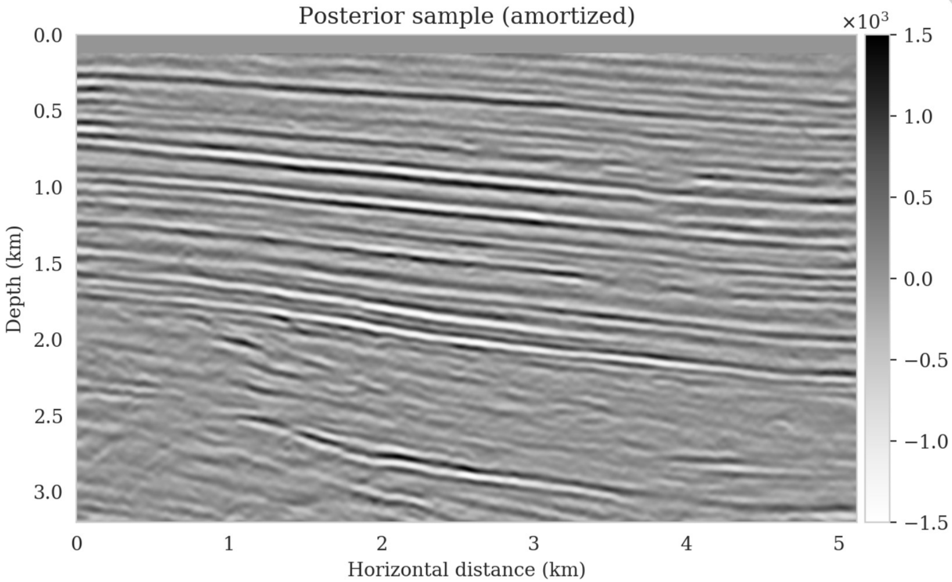
SNR −12.17 dB


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)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

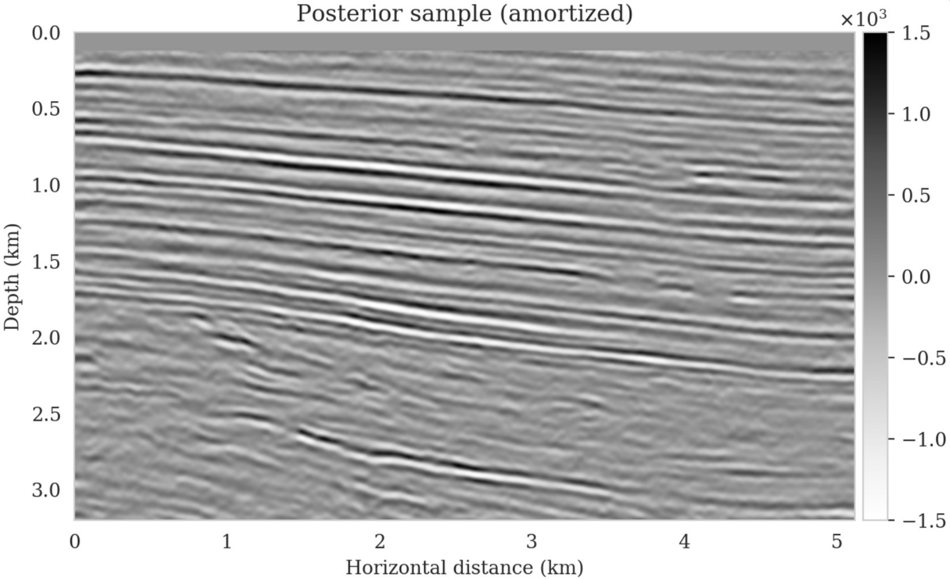
SNR 8.54 dB


Posterior sample (amortized)
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
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 8.49 dB

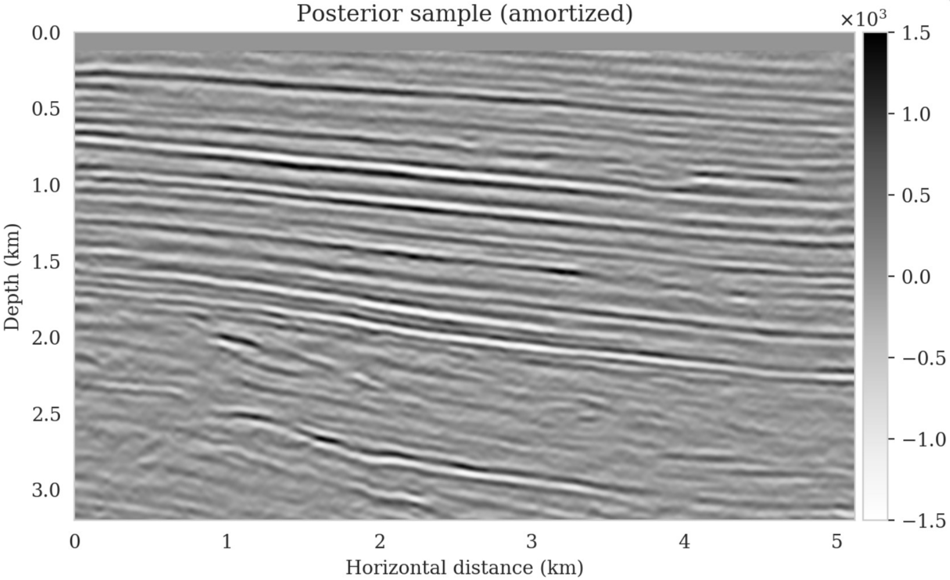

Posterior sample (amortized)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 8.42 dB

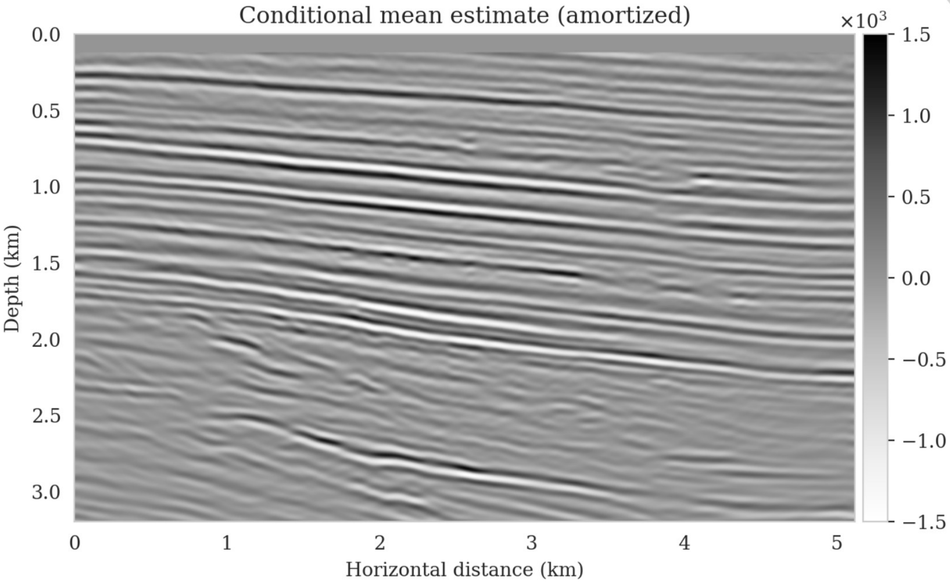


Posterior sample (amortized)
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)

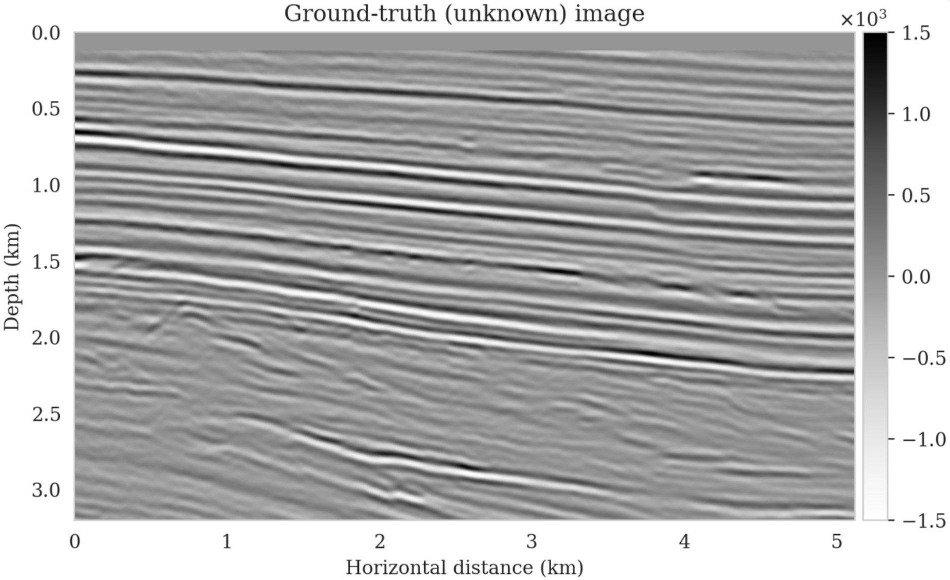
SNR 8.59 dB


Posterior sample (amortized)
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
$\times10^3$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

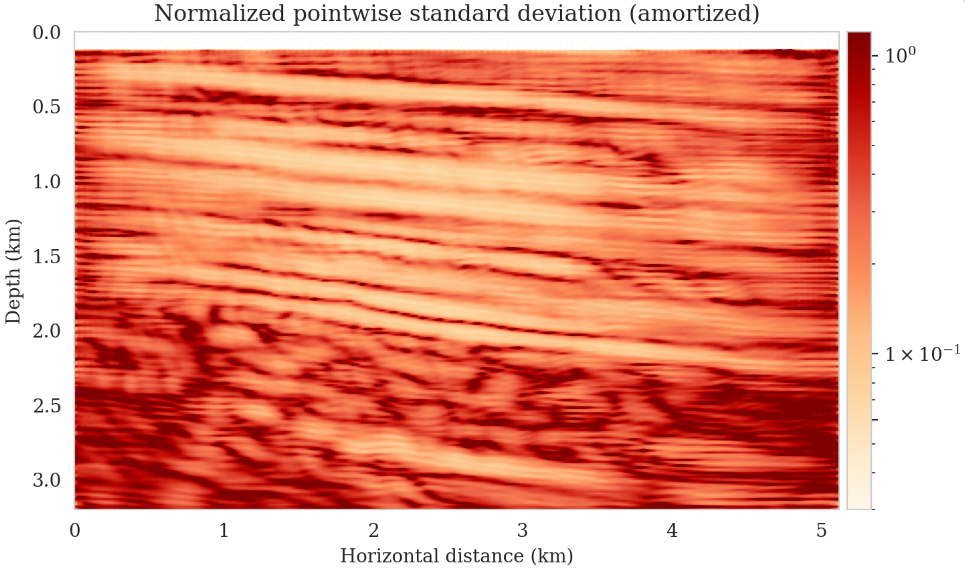
SNR 8.54 dB


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



normalized by the the envelope of the conditional mean

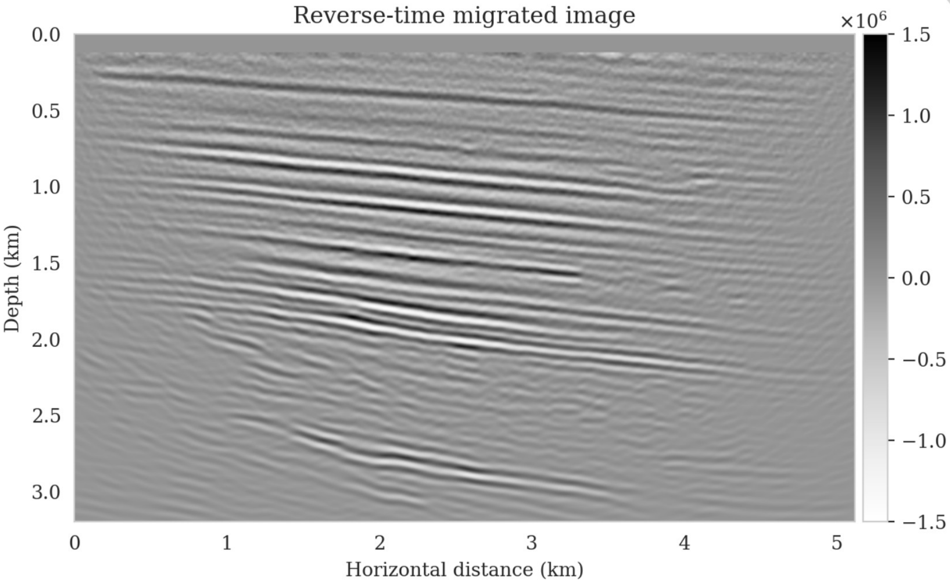
Introducing distribution shifts

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

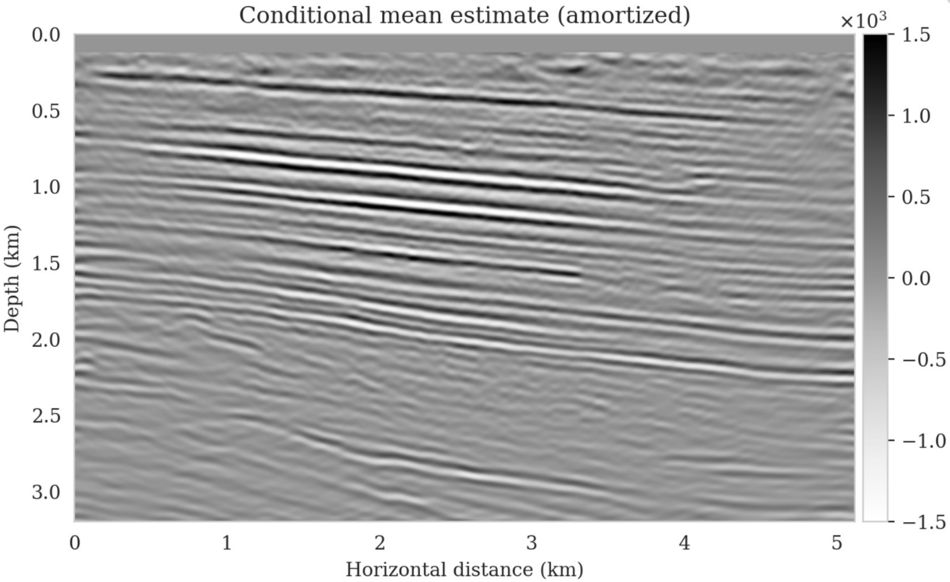
$4\times$ less sources

Physics-based latent distribution correction

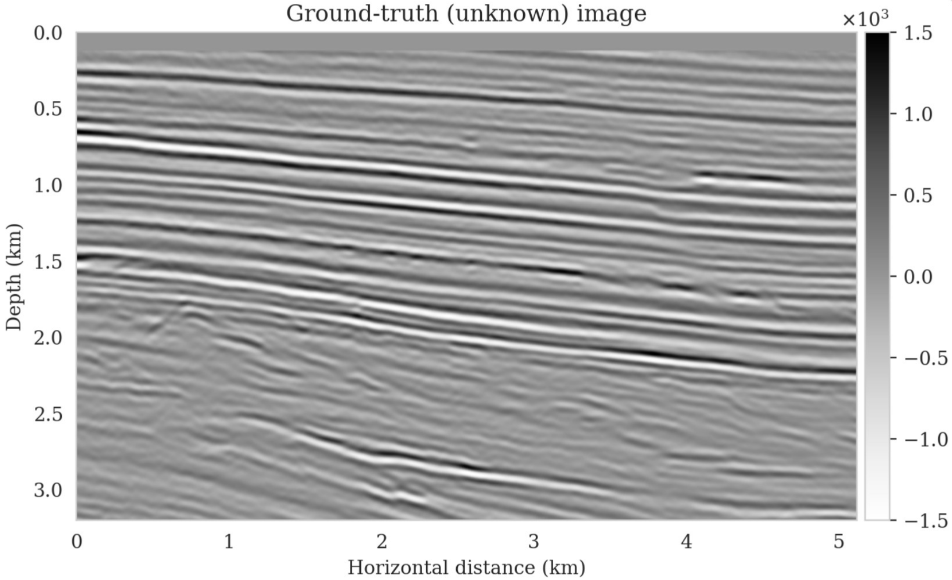
computational cost: approximately $5\times$ RTMs



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


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 6.29 dB



previously unseen (test) seismic image

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

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

with

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

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

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

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

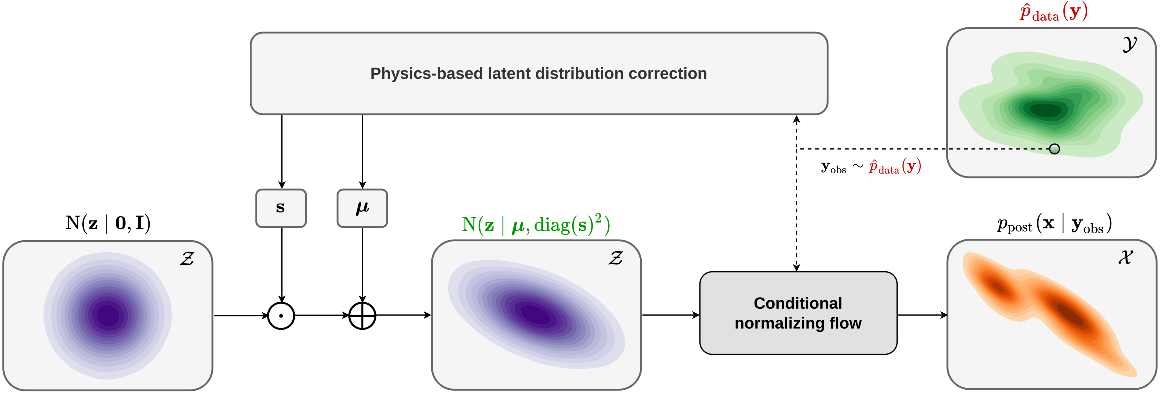
initialization acts as a **warm-start** and an implicit regularization

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

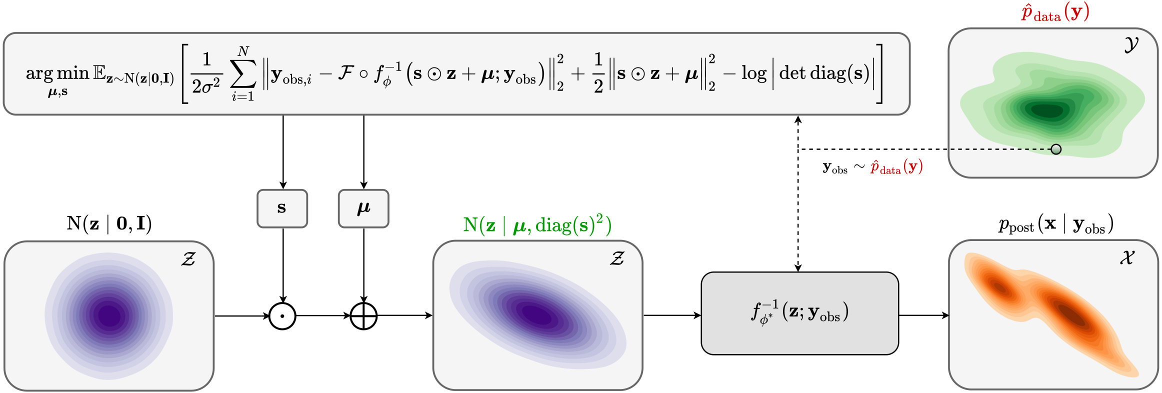
expected to be solved relatively cheaply due to the amortization of f_{ϕ}

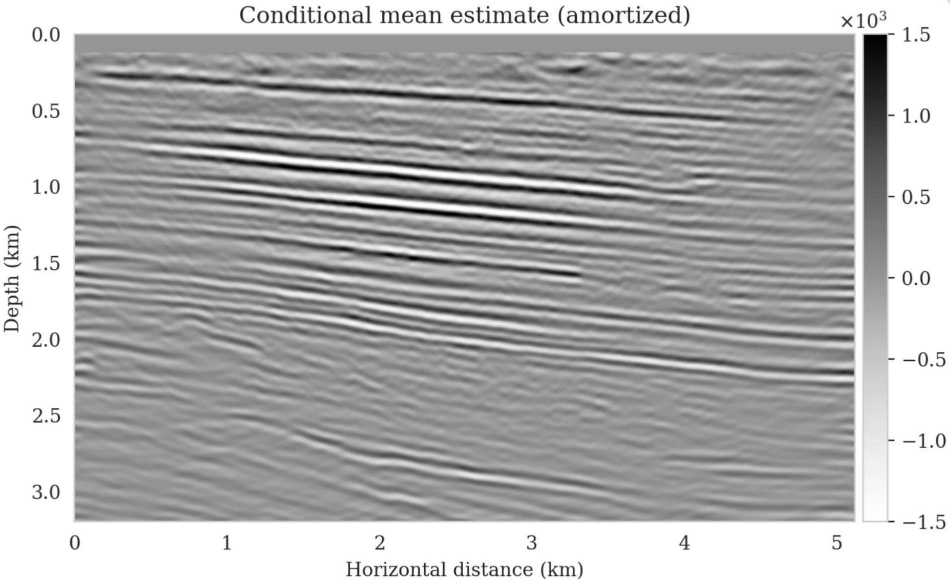
non-amortized, i.e., specific to one set of observations $\mathbf{y}_{\mathsf{obs}} \sim \widehat{p}_{\mathsf{data}}(\mathbf{y})$

Physics-based latent distribution correction

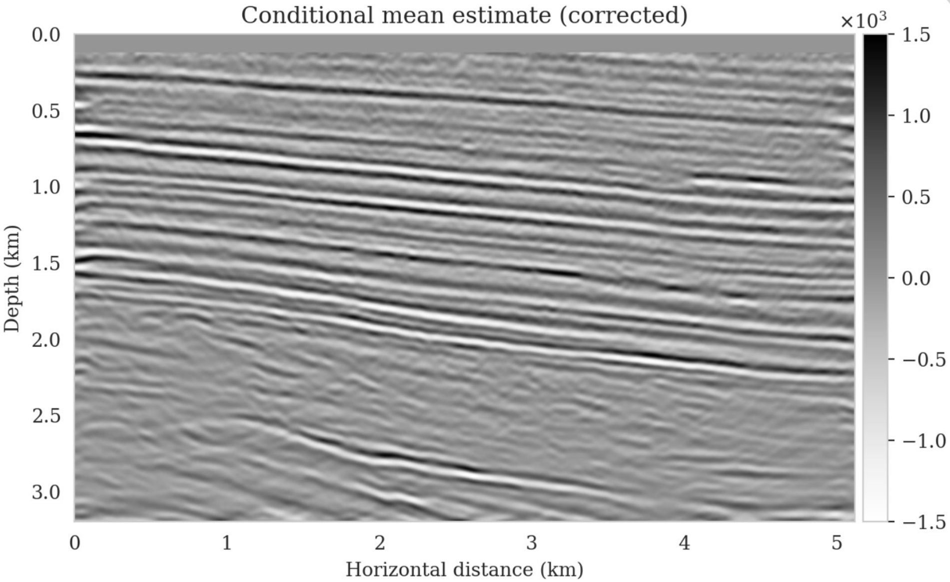


Physics-based latent distribution correction

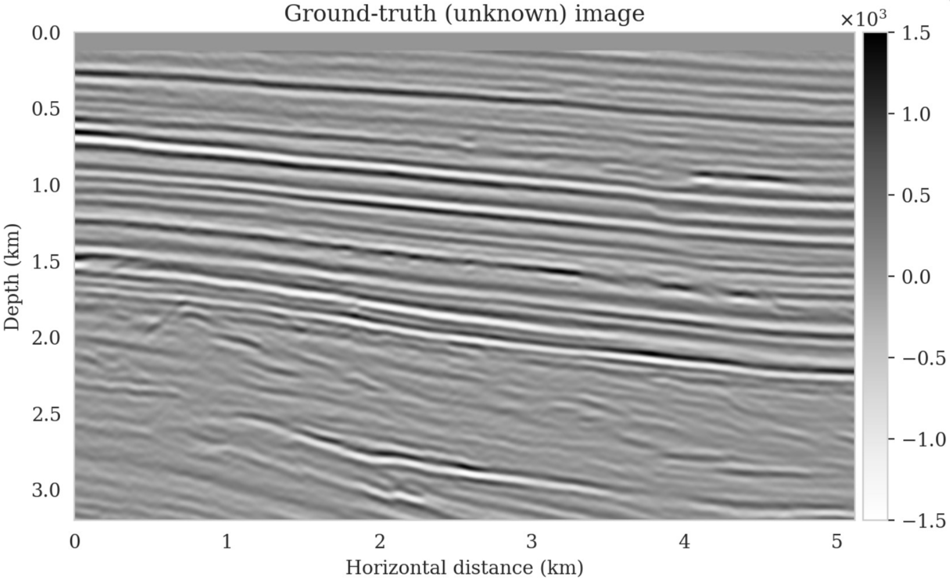



Conditional mean estimate (amortized)
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

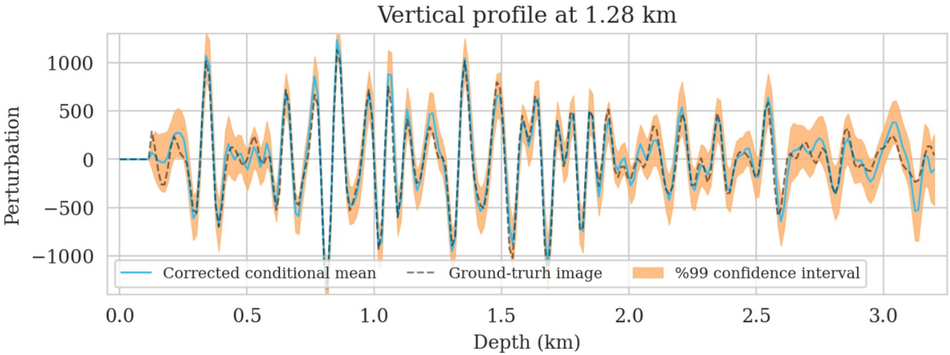
SNR 6.29 dB

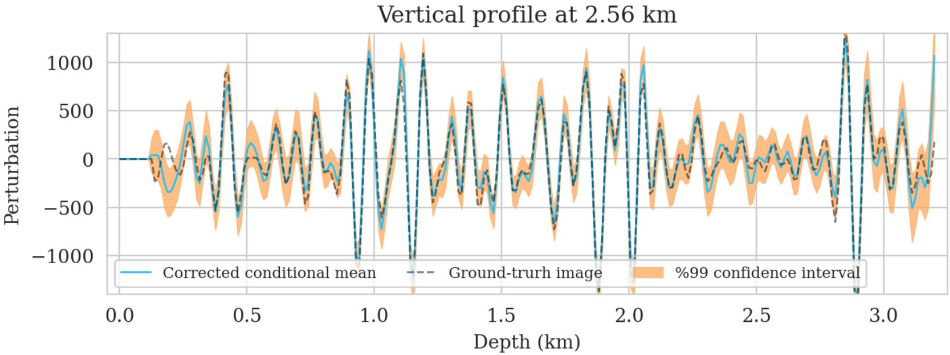

Conditional mean estimate (corrected)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)

SNR 10.36 dB

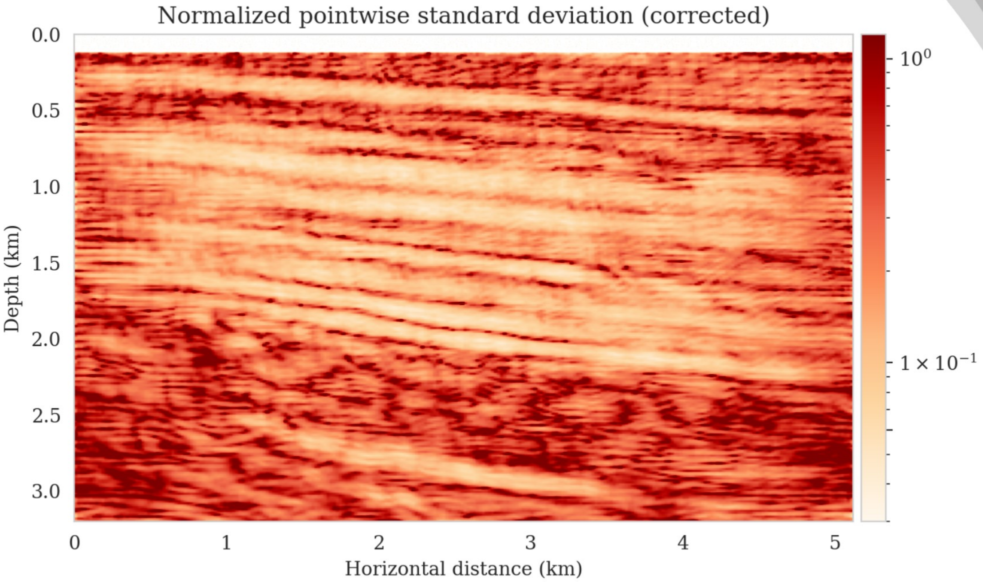


previously unseen (test) seismic image

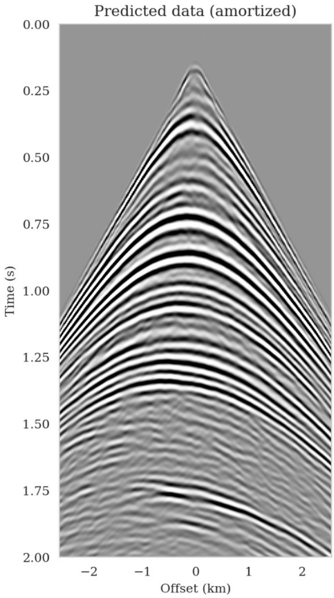

Vertical profile at 1.28 km
Perturbation
Depth (km)
1000
500
0
−500
−1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Corrected conditional mean
Ground-trurh image
%99 confidence interval

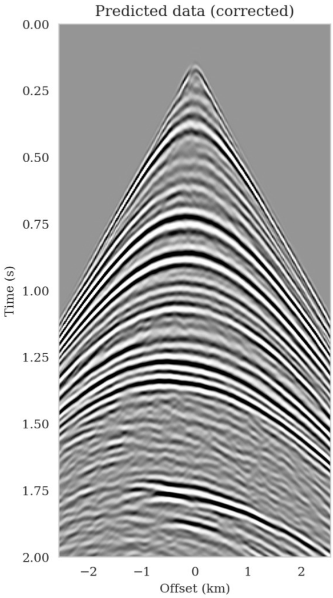


Vertical profile at 2.56 km
Perturbation
1000
500
0
−500
−1000
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
Corrected conditional mean
Ground-truth image
%99 confidence interval

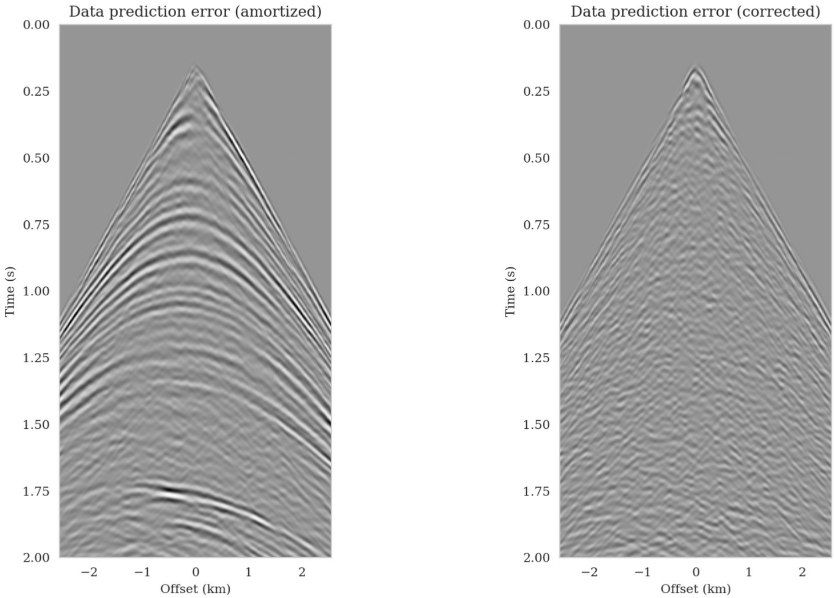


normalized by the the envelope of the conditional mean

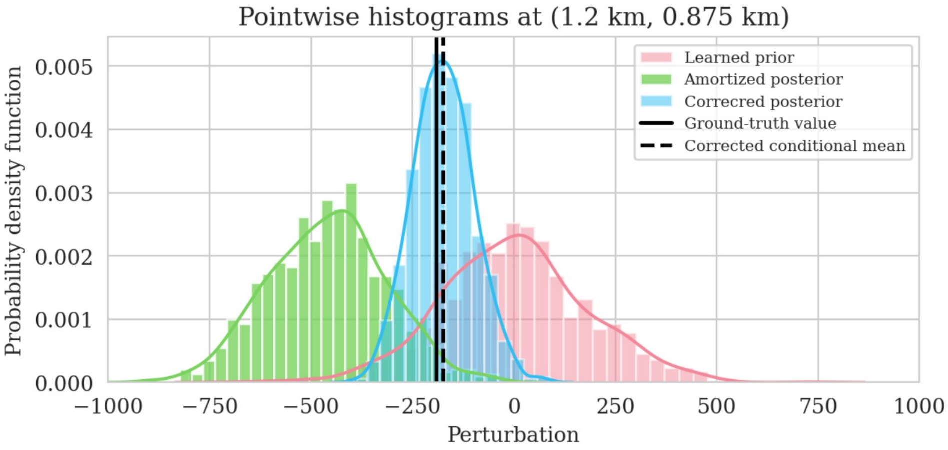




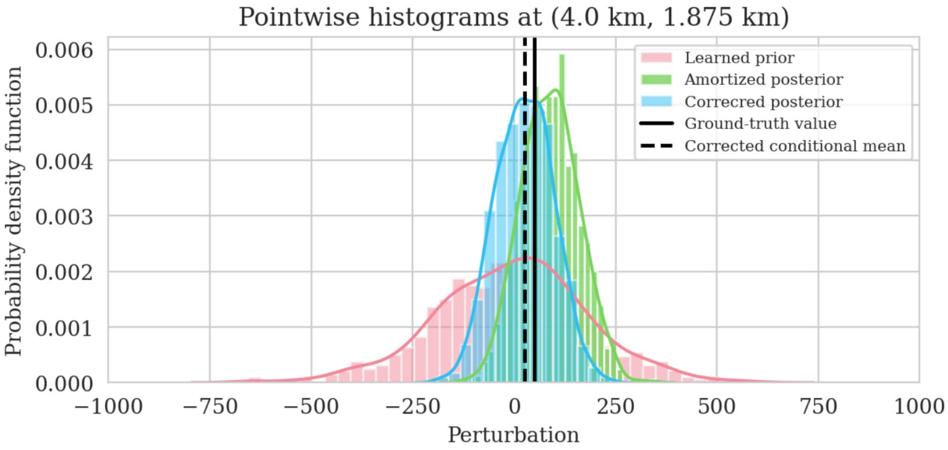
predicted data (left) amortized, SNR 11.62 dB (right) corrected, SNR 16.57 dB



data residual of (left) amortized (right) corrected



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


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

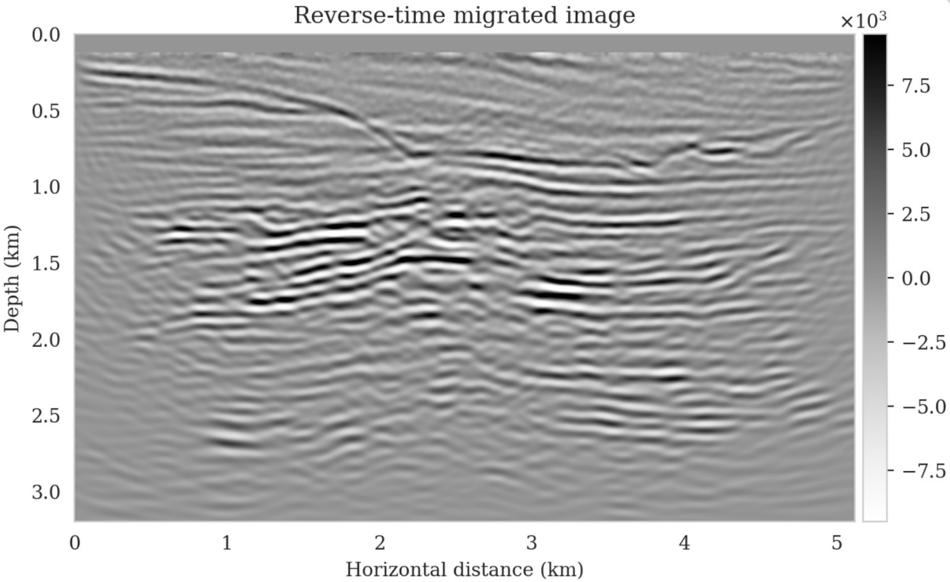
Data distribution shift

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

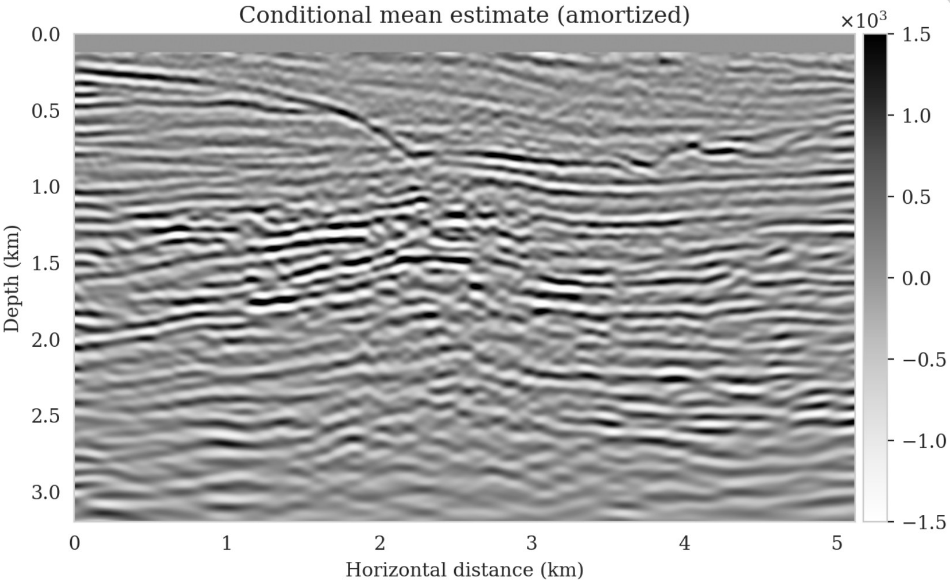
$4\times$ less sources

imaging deeper sections of Parihaka dataset

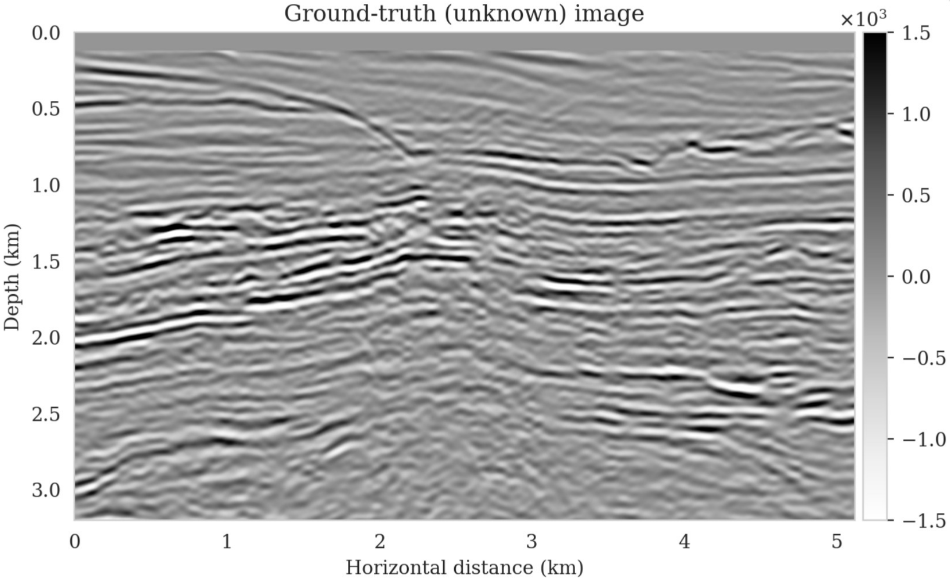
$10\,\mathrm{Hz}$ lower frequency source wavelet


Reverse-time migrated image
×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)
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5

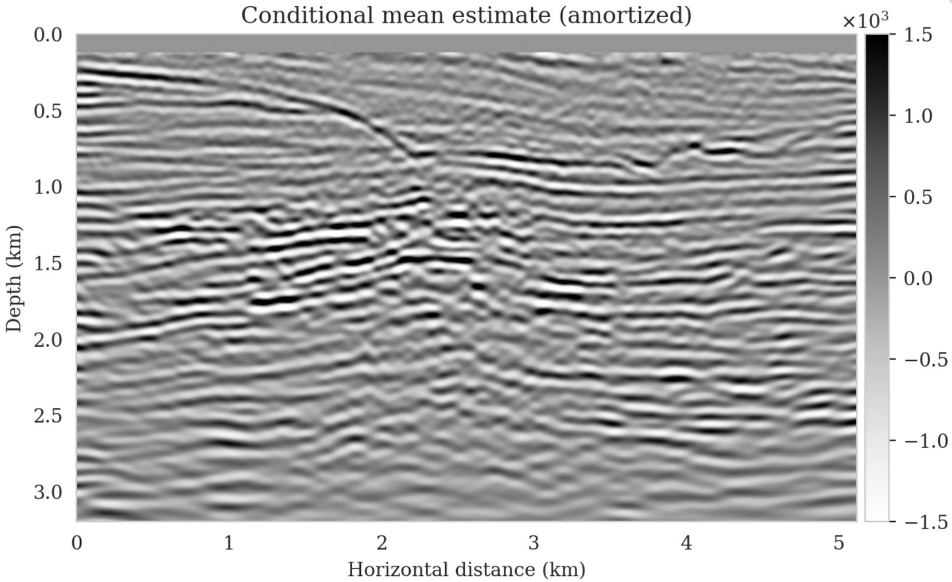
SNR −12.49 dB


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

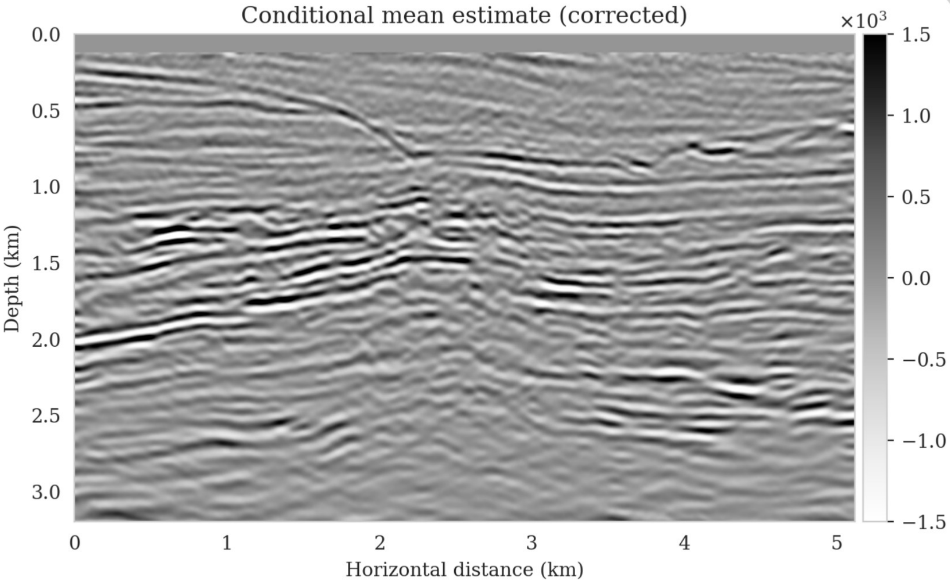
SNR 5.01 dB



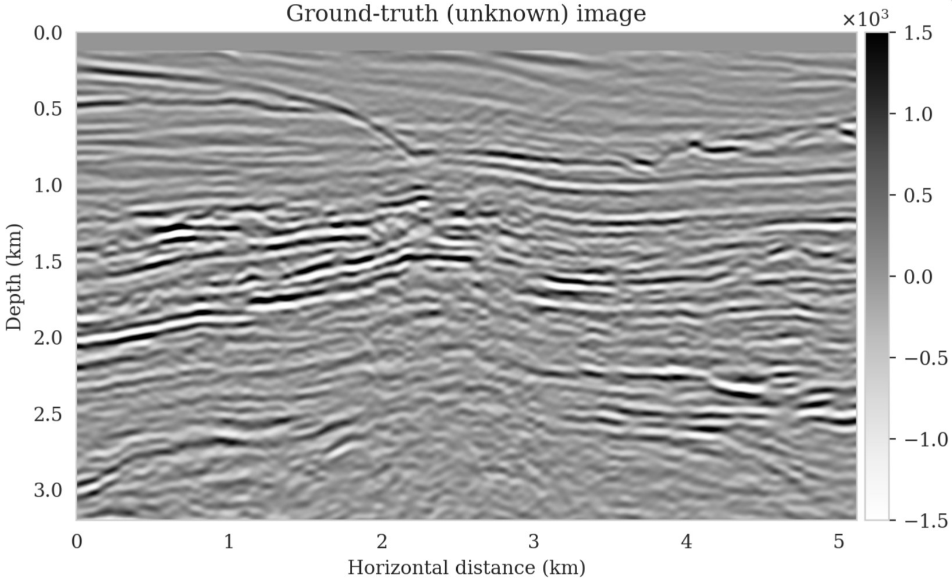
previously unseen (test) seismic image


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

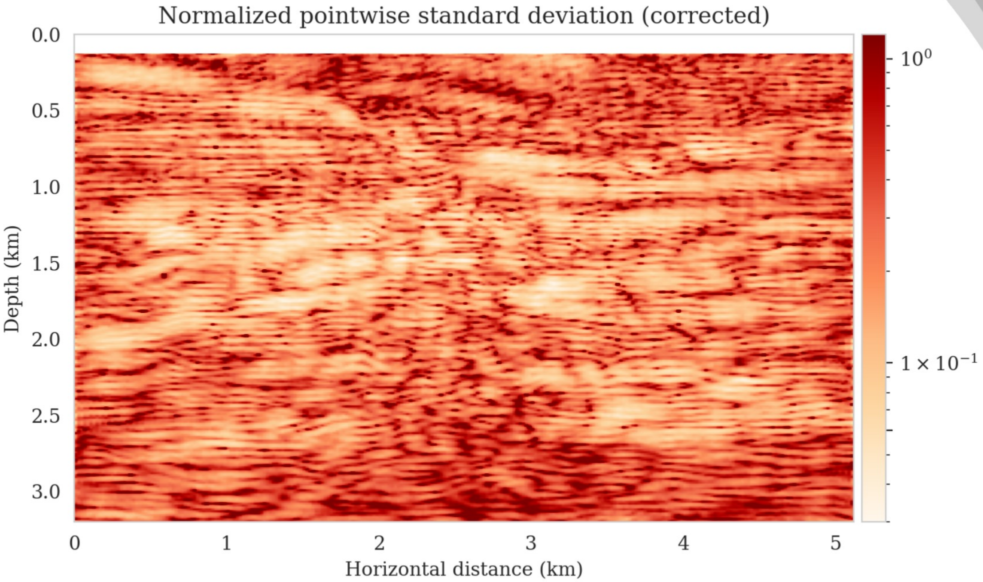
SNR 5.01 dB


Conditional mean estimate (corrected)
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
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SNR 8.96 dB



previously unseen (test) seismic image



normalized by the the envelope of the conditional mean

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

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.jl

Acknowledgement

This research was carried out with the support of Georgia Research Alliance and partners of the ML4Seismic Center.

Thank the Diamond sponsors (Equinor ASA, Microsoft Corporation, Occidental Petroleum) and Platinum sponsors (Chevron Corporation, ExxonMobil Corporation, Shell plc, S&P Global Inc) for their support.