Fast and reliability-aware seismic imaging with conditional normalizing flows

Ali Siahkoohi

Georgia Institute of Technology

SLIM

Georgia Institute of Technology

Fast and reliability-aware seismic imaging with conditional normalizing flows

Ali Siahkoohi [1] Rafael Orozco [1] Gabrio Rizzuti [2] Philipp A. Witte [3]
Mathias Louboutin [1] Felix J. Herrmann [1]

[1]Georgia Institute of Technology

[2]Utrecht University

[3]Microsoft

SLIM
Georgia Institute of Technology

Problem setup

Find $\mathbf{x}$ such that:

$$\mathbf{y} = F(\mathbf{x}) + \boldsymbol{\epsilon}$$

observed data $\mathbf{y}$

(non-Gaussian) noise $\boldsymbol{\epsilon}$

expensive (nonlinear) forward operator F

Stochastic gradient Langevin dynamics

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \frac{\alpha_k}{2}\mathbf{M}_k \nabla_{\mathbf{w}} \left(\frac{n_s}{2\sigma^2} \left\| \mathbf{d}_i - \mathbf{J}[\mathbf{m}_0, \mathbf{q}_i] g(\mathbf{z}, \mathbf{w}_k) \right\|_2^2 + \frac{\lambda^2}{2} \left\| \mathbf{w}_k \right\|_2^2 \right) + \boldsymbol{\xi}_k,$$

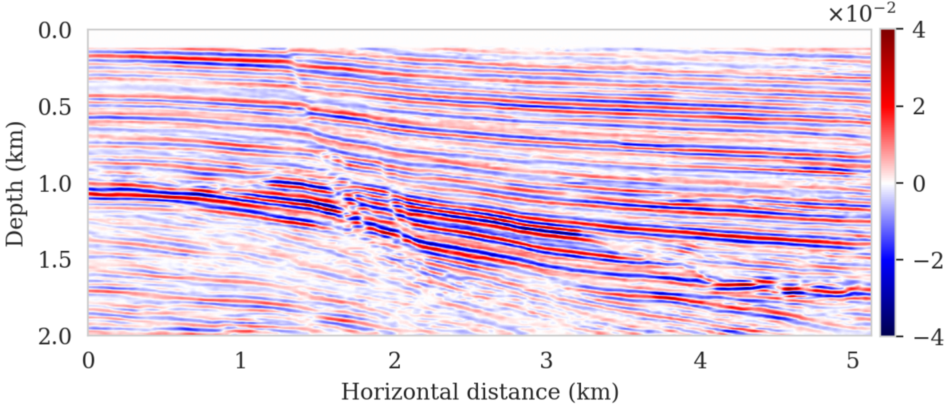
$$\boldsymbol{\xi}_k \sim \mathrm{N}(\mathbf{0}, \alpha_k \mathbf{M}_k)$$

sampling the posterior distribution

requires at least 10k iterations

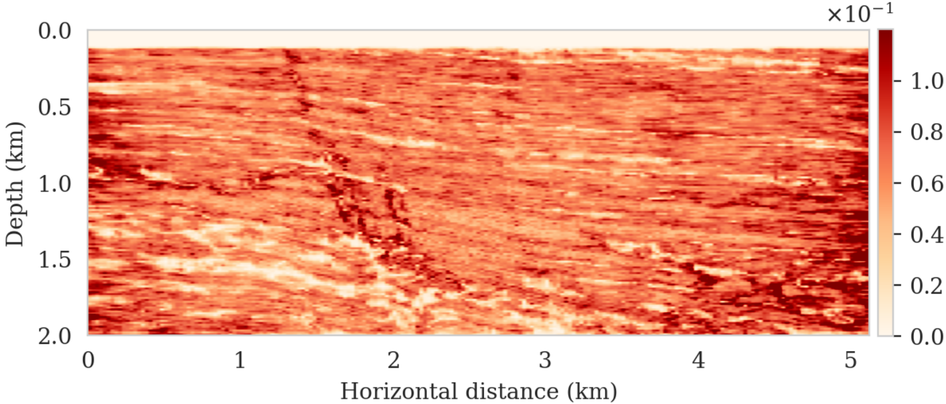
Max Welling and Yee Whye Teh. "Bayesian Learning via Stochastic Gradient Langevin Dynamics". In: *Proceedings of the 28th International Conference on International Conference on Machine Learning*. ICML'11. 2011, pp. 681–688. DOI: 10.5555/3104482.3104568.

Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann. "Uncertainty quantification in imaging and automatic horizon tracking—a Bayesian deep-prior based approach". In: *SEG Technical Program Expanded Abstracts 2020*. Sept. 2020. URL: https://arxiv.org/pdf/2004.00227.pdf.



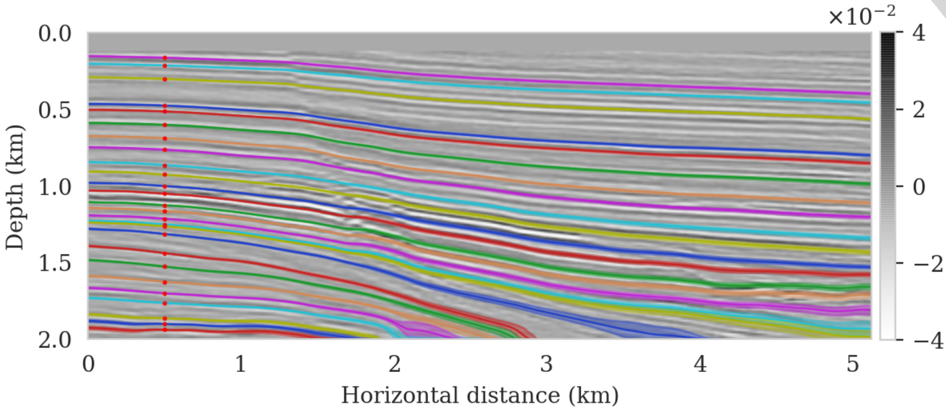
Conditional mean estimate

Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann. "Uncertainty quantification in imaging and automatic horizon tracking—a Bayesian deep-prior based approach". In: *SEG Technical Program Expanded Abstracts 2020*. Sept. 2020. URL: `https://arxiv.org/pdf/2004.00227.pdf`.



Pointwise standard deviation, normalized by the envelope of conditional mean

Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann. "Uncertainty quantification in imaging and automatic horizon tracking—a Bayesian deep-prior based approach". In: *SEG Technical Program Expanded Abstracts 2020*. Sept. 2020. URL: https://arxiv.org/pdf/2004.00227.pdf.



Uncertainty in horizon tracking due to uncertainties in imaging

Xinming Wu and Sergey Fomel. "Least-squares horizons with local slopes and multi-grid correlations". In: *GEOPHYSICS* 83 (4 2018), pp. IM29–IM40. DOI: 10.1190/geo2017-0830.1. URL: https://doi.org/10.1190/geo2017-0830.1.

Xinming Wu and Sergey Fomel. *xinwucwp/mhe*. 2018. URL: https://github.com/xinwucwp/mhe.

Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann. "Uncertainty quantification in imaging and automatic horizon tracking—a Bayesian deep-prior based approach". In: *SEG Technical Program Expanded Abstracts 2020*. Sept. 2020. URL: https://arxiv.org/pdf/2004.00227.pdf.

Inverse problems and Bayesian inference

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.

Bayes' rule

$$p\left(\mathbf{x} \mid \mathbf{y}\right) \propto p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{x})$$

posterior distribution $p\left(\mathbf{x} \mid \mathbf{y}\right)$

likelihood function $p(\mathbf{y} \mid \mathbf{x})$

prior distribution $p(\mathbf{x})$

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

Sampling the posterior

Markov chain Monte Carlo (MCMC) sampling

guarantees of producing (asymptotically) exact samples

high-dimensional integration/sampling

costs associated with the forward operator

requires choosing a prior distribution

Christian Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Science & Business Media, 2004. DOI: 10.1007/978-1-4757-4145-2.

Prior art: MCMC

Alberto Malinverno and Victoria A Briggs. "Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes". In: *GEOPHYSICS* 69.4 (2004), pp. 1005–1016. DOI: 10.1190/1.1778243.

Alberto Malinverno and Robert L Parker. "Two ways to quantify uncertainty in geophysical inverse problems". In: *GEOPHYSICS* 71.3 (2006), W15–W27. DOI: 10.1190/1.2194516.

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". In: *SIAM Journal on Scientific Computing* 34.3 (2012), A1460–A1487. eprint: http://epubs.siam.org/doi/pdf/10.1137/110845598. URL: http://epubs.siam.org/doi/abs/10.1137/110845598.

Anandaroop Ray, Sam Kaplan, John Washbourne, and Uwe Albertin. "Low frequency full waveform seismic inversion within a tree based Bayesian framework". In: *Geophysical Journal International* 212.1 (Oct. 2017), pp. 522–542. ISSN: 0956-540X. DOI: 10.1093/gji/ggx428. eprint: https://academic.oup.com/gji/article-pdf/212/1/522/21782947/ggx428.pdf. URL: https://doi.org/10.1093/gji/ggx428.

Zhilong Fang, Curt Da Silva, Rachel Kuske, and Felix J. Herrmann. "Uncertainty quantification for inverse problems with weak partial-differential-equation constraints". In: *GEOPHYSICS* 83.6 (2018), R629–R647. DOI: 10.1190/geo2017-0824.1.

Georgia K Stuart, Susan E Minkoff, and Felipe Pereira. "A two-stage Markov chain Monte Carlo method for seismic inversion and uncertainty quantification". In: *GEOPHYSICS* 84.6 (Nov. 2019), R1003–R1020. DOI: 10.1190/geo2018-0893.1.

Zeyu Zhao and Mrinal K Sen. "A gradient based MCMC method for FWI and uncertainty analysis". In: *SEG Technical Program Expanded Abstracts 2019*. Aug. 2019, pp. 1465–1469. DOI: 10.1190/segam2019-3216560.1.

M Kotsi, A Malcolm, and G Ely. "Uncertainty quantification in time-lapse seismic imaging: a full-waveform approach". In: *Geophysical Journal International* 222.2 (May 2020), pp. 1245–1263. ISSN: 0956-540X. DOI: 10.1093/gji/ggaa245.

Andrew Curtis and Anthony Lomax. "Prior information, sampling distributions, and the curse of dimensionality". In: *Geophysics* 66.2 (2001), pp. 372–378.

Felix J. Herrmann, Ali Siahkoohi, and Gabrio Rizzuti. "Learned imaging with constraints and uncertainty quantification". In: *Neural Information Processing Systems (NeurIPS) 2019 Deep Inverse Workshop*. Dec. 2019. URL: https://arxiv.org/pdf/1909.06473.pdf.

Sampling the posterior

Variational inference

approximate the posterior with a parametric and easy-to-sample distribution

sampling is turned into an optimization problem

known to scale better than MCMC in high-dimensional problems

requires choosing a prior distribution

David M Blei, Alp Kucukelbir, and Jon D McAuliffe. "Variational inference: A review for statisticians". In: *Journal of the American statistical Association* 112.518 (2017), pp. 859–877.

Michael I Jordan, Zoubin Ghahramani, Tommi S Jaakkola, and Lawrence K Saul. "An Introduction to Variational Methods for Graphical Models". In: *Machine Learning* 37.2 (1999), pp. 183–233. DOI: 10.1023/A:1007665907178.

Posterior inference with Normalizing Flows (NFs)

sampling the posterior distribution directly via NFs

Purely data-driven approach

$$\min_{\theta} \mathbb{E}_{\mathbf{y},\mathbf{x}\sim p(\mathbf{y},\mathbf{x})} \left[\frac{1}{2} \left\| G_\theta(\mathbf{y},\mathbf{x}) \right\|^2 - \log \left| \det \nabla_{y,x} G_\theta(\mathbf{y},\mathbf{x}) \right| \right]$$

$$G_\theta(\mathbf{y},\mathbf{x}) = \begin{bmatrix} G_{\theta_y}(\mathbf{y}) \\ G_{\theta_x}(\mathbf{y},\mathbf{x}) \end{bmatrix}, \ \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_y \\ \boldsymbol{\theta}_x \end{bmatrix}$$

conditional NF, $G_\theta : \mathcal{Y} \times \mathcal{X} \to \mathcal{Z}_y \times \mathcal{Z}_x$

expectation estimated via training pairs, $\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^n \sim p_{y,x}(\mathbf{y},\mathbf{x})$

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, 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.

Inference with the data-driven approach

Sampling from $p(\mathbf{x} \mid \mathbf{y})$

$$G_{\theta_x}^{-1}(G_{\theta_y}(\mathbf{y}), \mathbf{z}) \sim p\left(\mathbf{x} \mid \mathbf{y}\right), \quad \mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$$

Posterior density estimation

$$p_\theta(\mathbf{x} \mid \mathbf{y}) = p_{\mathbf{z}}\big(G_{\theta_x}(\mathbf{y}, \mathbf{x})\big) \left| \det \nabla_x G_{\theta_x}(\mathbf{y}, \mathbf{x}) \right|$$

Seismic imaging example

Seismic imaging forward model

$$\delta\mathbf{d} = \mathbf{J}[\mathbf{m}_0, \mathbf{q}]\delta\mathbf{m} + \boldsymbol{\eta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$$

linearized observed data $\delta\mathbf{d}$

linearized Born modeling operator $\mathbf{J}$

source signature $\mathbf{q}$

smooth background model $\mathbf{m}_0$

unknown perturbation model $\delta\mathbf{m}$

linearization error and noise $\boldsymbol{\eta}, \boldsymbol{\epsilon}$

Training dataset for the data-driven approach

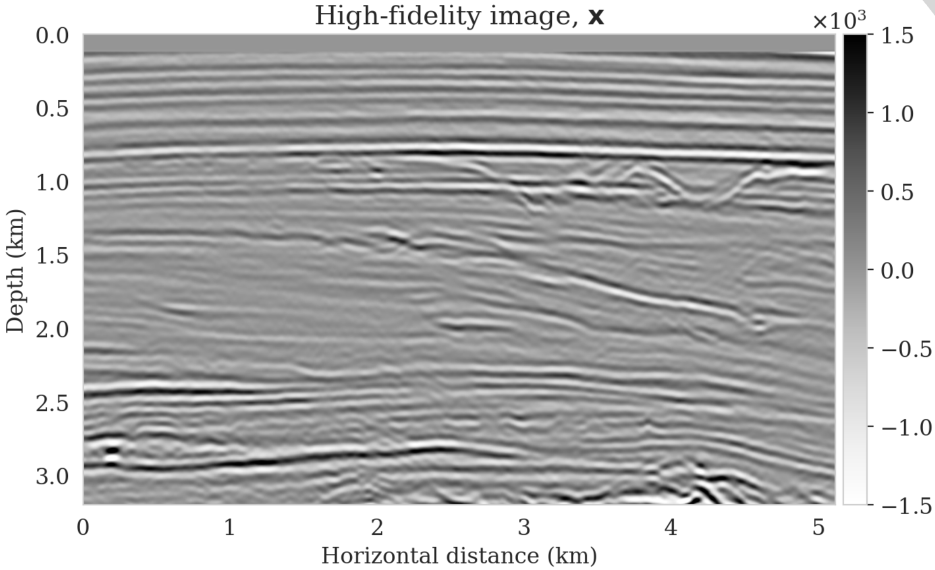
training pairs

$$\{\mathbf{y}_i, \mathbf{x}_i\}_{i=1}^{n} \sim p(\mathbf{y}, \mathbf{x})$$

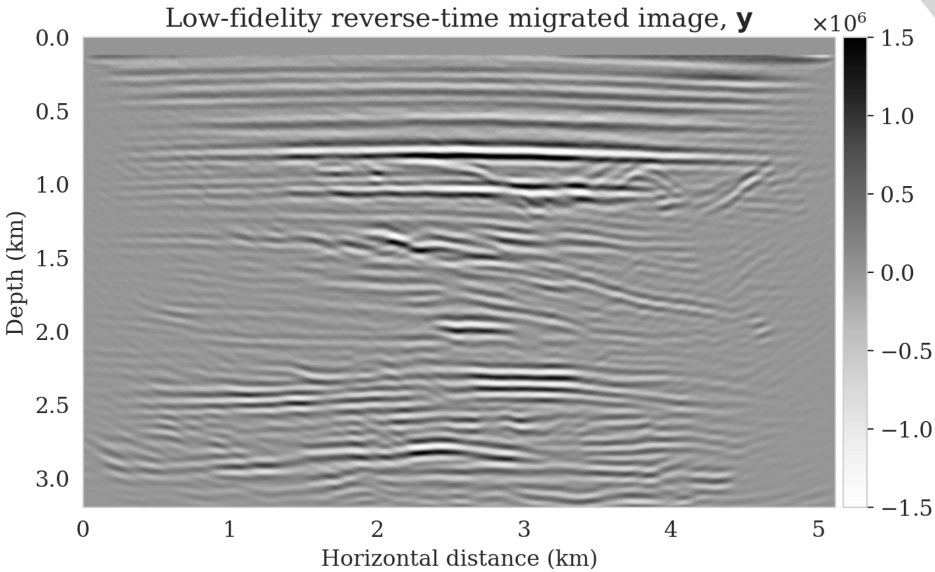
where

$$(\mathbf{y}_i, \mathbf{x}_i) = \left(\mathbf{J}^\top (\mathbf{J}\delta\mathbf{m}_i + \boldsymbol{\epsilon}_i), \delta\mathbf{m}_i\right)$$

WesternGeco. *Parihaka 3D PSTM Final Processing Report*. Tech. rep. New Zealand Petroleum Report 4582. 2012. URL: https://wiki.seg.org/wiki/Parihaka-3D.

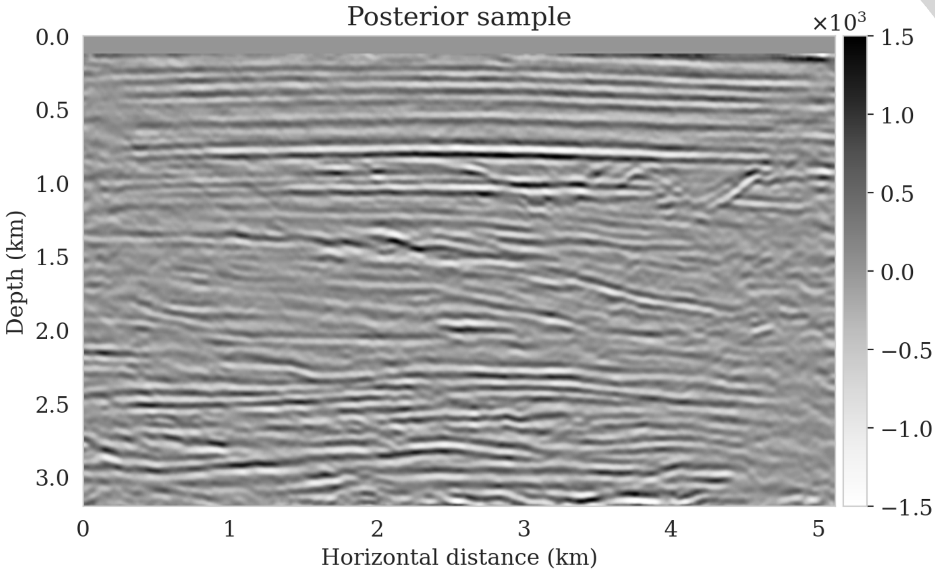


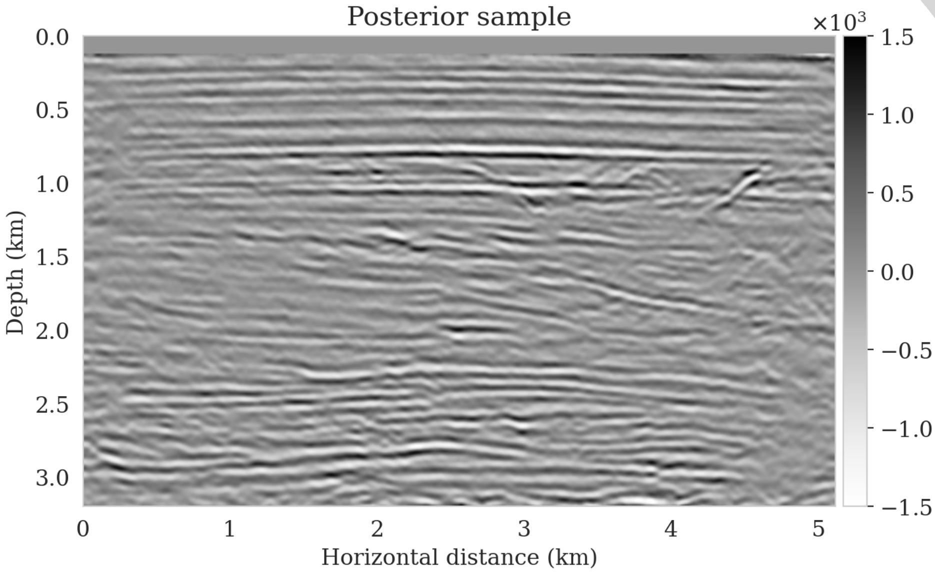
High-fidelity image, **x**
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
$\times 10^3$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

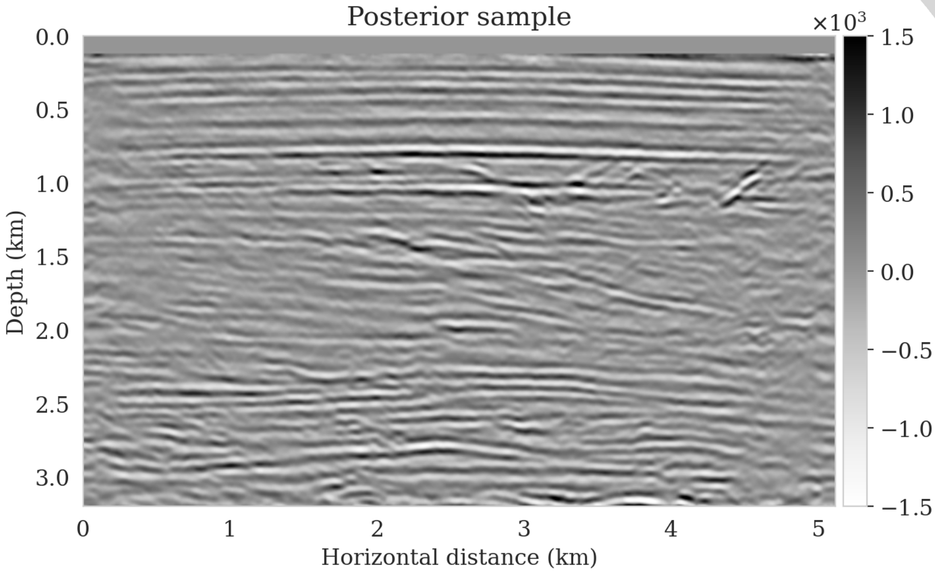

Low-fidelity reverse-time migrated image, **y**
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
$\times 10^6$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

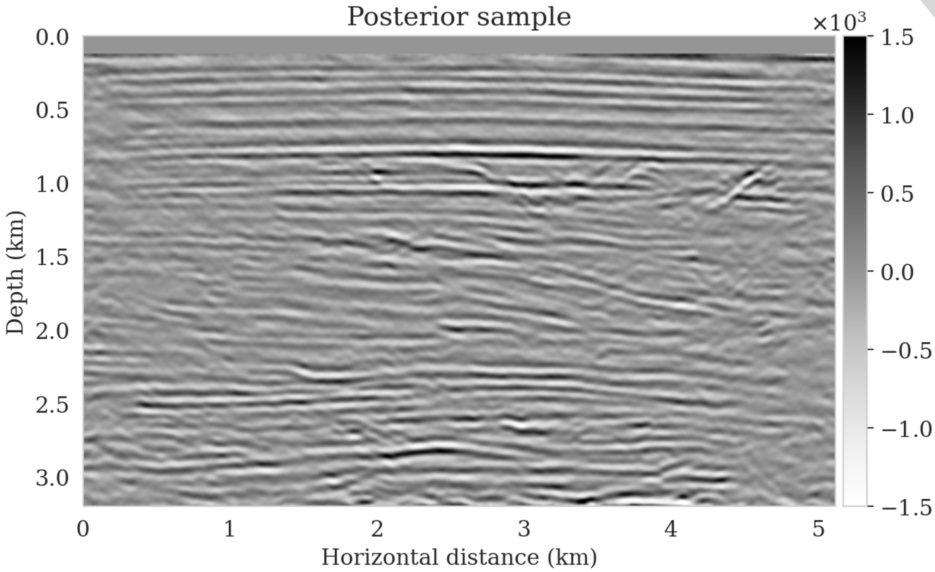
Samples from the posterior

$$G_{\theta_x}^{-1}(G_{\theta_y}(\mathbf{y}), \mathbf{z}) \sim p\left(\mathbf{x} \mid \mathbf{y}\right), \quad \mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$$


Posterior sample
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

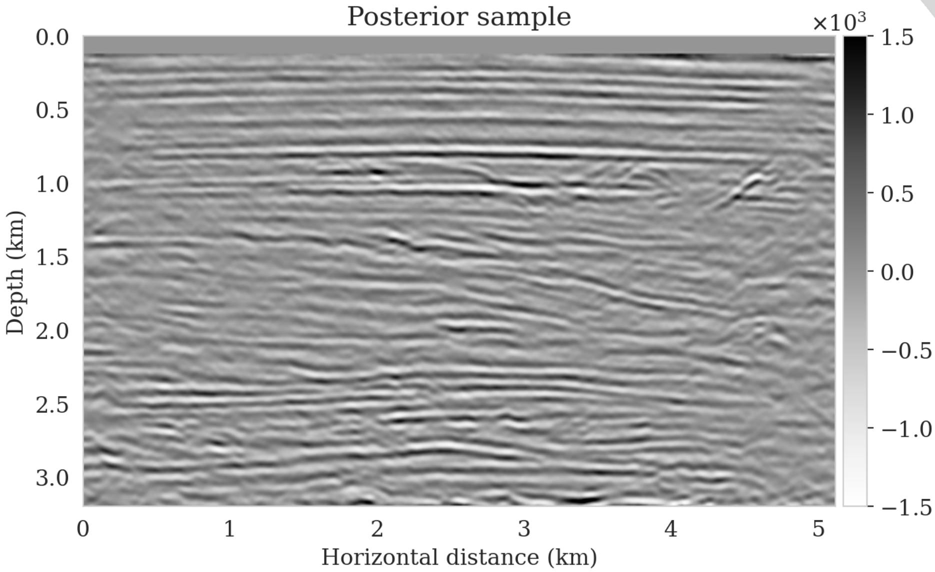

Posterior sample
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5


Posterior sample
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



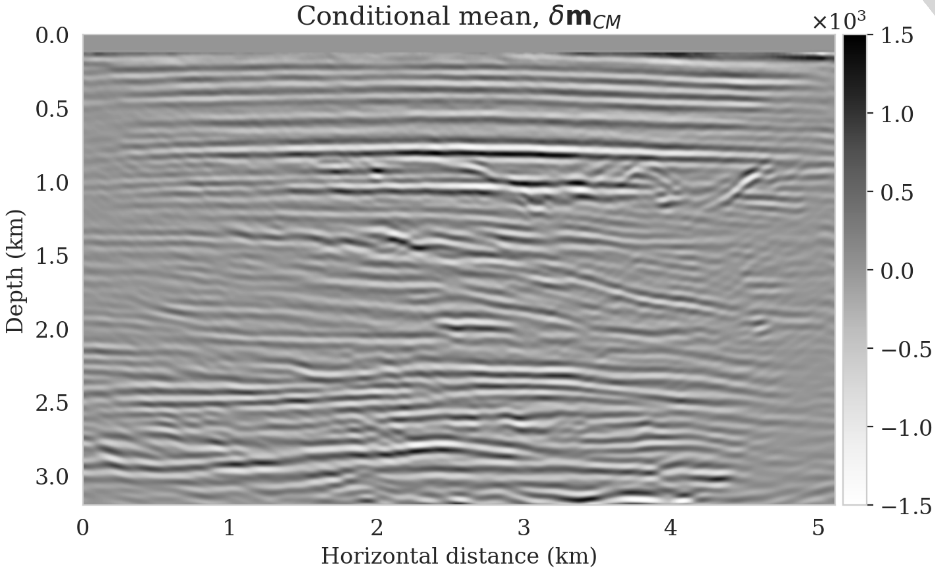
Posterior sample
$\times 10^3$
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SLIM

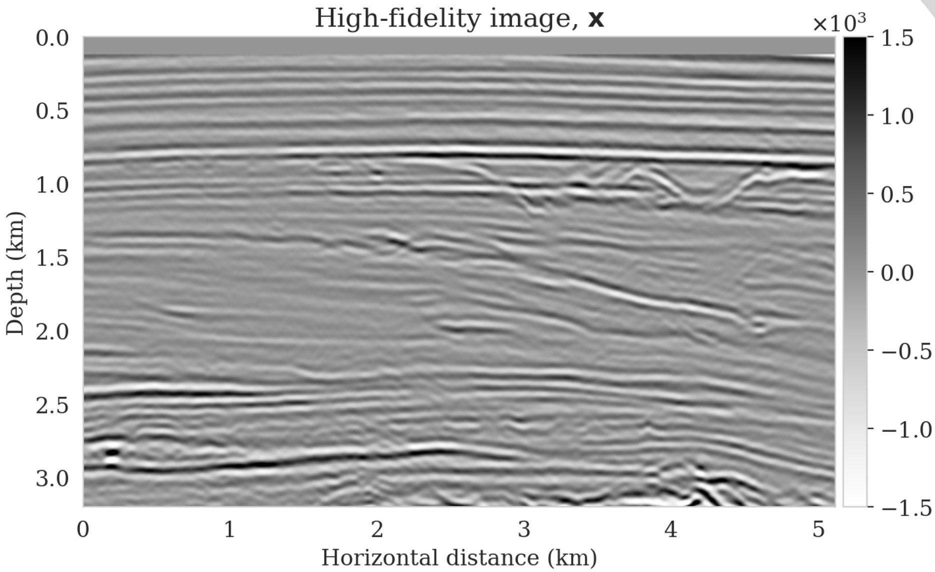

Posterior sample
$\times 10^3$
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

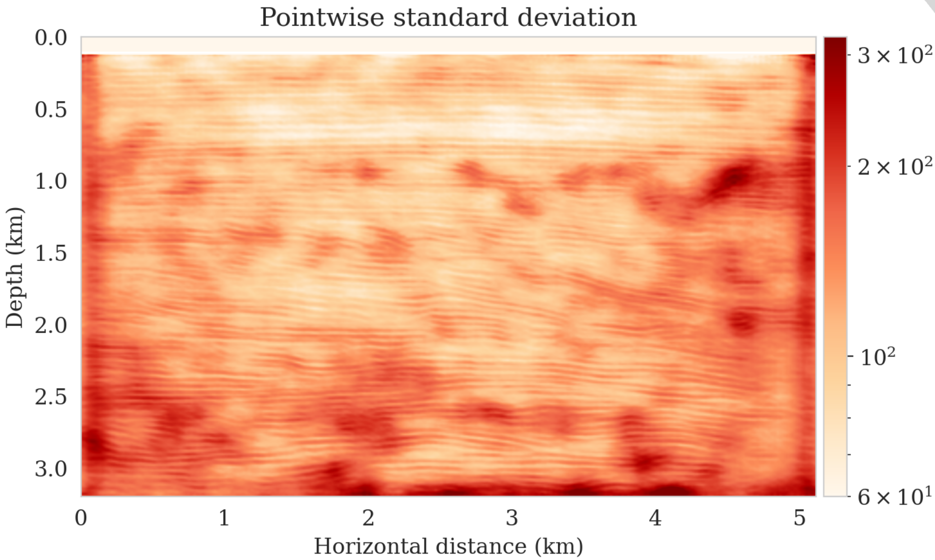
Conditional mean estimate

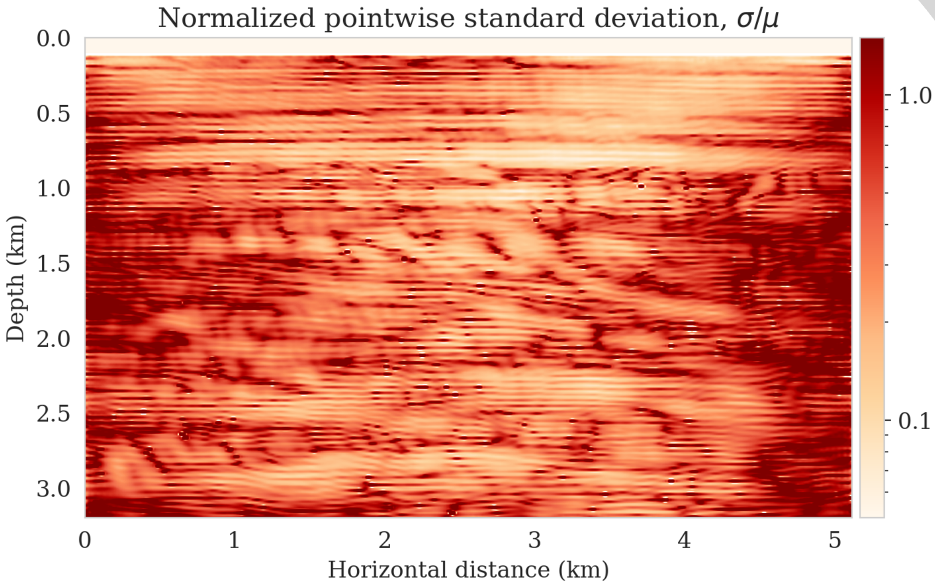
$$\mathbb{E}\left[\mathbf{x} \mid \mathbf{y}\right] = \int \mathbf{x}\, p\left(\mathbf{x} \mid \mathbf{y}\right) \mathrm{d}\mathbf{x}$$



Conditional mean, $\delta \mathbf{m}_{CM}$
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)
$\times 10^3$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5


High-fidelity image, **x**
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
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5


Pointwise standard deviation
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)
3×10^2
2×10^2
10^2
6×10^1


Normalized pointwise standard deviation, σ/μ
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)
1.0
0.1

Purely data-driven approach

learns the prior distribution from training data

samples from the posterior virtually for free in test time

not specific to one observation $\mathbf{y}$

needs supervised pairs of model and data

heavily relies on the training data to generalize

not directly tied to physics/data

Purely physics-based approach

$$\min_{\phi} \mathbb{E}_{\mathbf{z}\sim \mathrm{N}(\mathbf{0},\mathbf{I})} \left[\frac{1}{2\sigma^2} \left\| F\big(T_{\phi}(\mathbf{z})\big) - \mathbf{y} \right\|_2^2 - \log p\big(T_{\phi}(\mathbf{z})\big) - \log \left| \det \nabla_z T_{\phi}(\mathbf{z}) \right| \right]$$

NF, $T_{\phi} : \mathcal{Z} \to \mathcal{X}$

prior distribution, $p(\mathbf{x})$

specific to one observation, $\mathbf{y}$

Gabrio Rizzuti, Ali Siahkoohi, Philipp A. Witte, and Felix J. Herrmann. "Parameterizing uncertainty by deep invertible networks, an application to reservoir characterization". In: *SEG Technical Program Expanded Abstracts*. Sept. 2020, pp. 1541–1545. DOI: 10.1190/segam2020-3428150.1.

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.

Xuebin Zhao, Andrew Curtis, and Xin Zhang. "Bayesian Seismic Tomography using Normalizing Flows". In: (2020). DOI: 10.31223/X53K6G. URL: https://eartharxiv.org/repository/view/1940/.

Inference with the physics-based approach

sampling from $p(\mathbf{x} \mid \mathbf{y})$

$$T_\phi(\mathbf{z}) \sim p\left(\mathbf{x} \mid \mathbf{y}\right), \quad \mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I}),$$

posterior density estimation,

$$p_\phi(\mathbf{x} \mid \mathbf{y}) = p_{\mathbf{z}}\left(T_\phi^{-1}(\mathbf{x})\right) \left| \det \nabla_x T_\phi^{-1}(\mathbf{x}) \right|$$

Purely physics-based approach

tied to the physics and data

no training data needed

requires choosing a prior distribution $p(\mathbf{x})$

repeated evaluations of F and $\nabla F^{\top}$

specific to one observation, $\mathbf{y}$

Multi-fidelity preconditioned scheme

exploit the information in the pretrained conditional NF (data-driven approach)

tie the results to physics and data

Multi-fidelity preconditioned scheme

Use the density encoded by the pretrained conditional NF as a prior

$$p_{\text{prior}}(\mathbf{x}) := p_{\mathbf{z}}\big(G_{\theta_x}(\mathbf{y}, \mathbf{x})\big) \left| \det \nabla_x G_{\theta_x}(\mathbf{y}, \mathbf{x}) \right|$$

allows for using a data-driven prior

removes the bias introduced by hand-crafted priors

can be used as a prior density in inverse problems

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.

Multi-fidelity preconditioned scheme

Use transfer learning and initialize $T : \mathcal{Z} \to \mathcal{X}$ by

$$T_{\theta_x}(\mathbf{z}) := G_{\theta_x}^{-1}(G_{\theta_y}(\mathbf{y}), \mathbf{z})$$

can significantly reduce $(5\times)$ the cost of the purely physics-based approach

can be used as an implicit prior by reparameterizing the unknown with T

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*. Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 399–409. URL: http://proceedings.mlr.press/v119/asim20a.html.

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.

Preconditioned MAP estimation

SLIM

$$\min_{\mathbf{z}} \frac{1}{2\sigma^2} \left\| F\left(T_{\theta_x}(\mathbf{z})\right) - \mathbf{y} \right\|_2^2 + \frac{1}{2} \left\| \mathbf{z} \right\|_2^2$$

initializing $\mathbf{z} = \mathbf{0}$ acts as an implicit prior

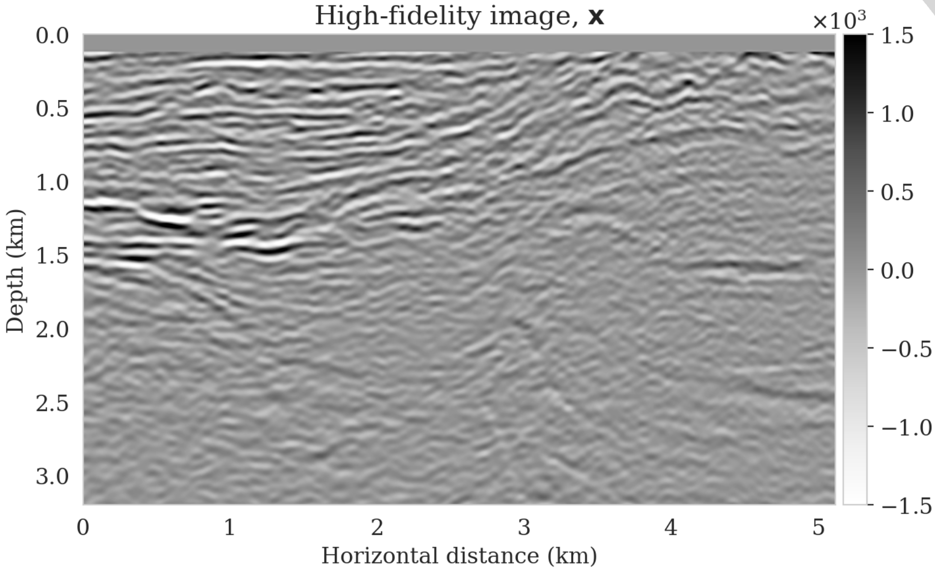
since T is invertible, in principle it can represent any unknown $\mathbf{x} \in \mathcal{X}$

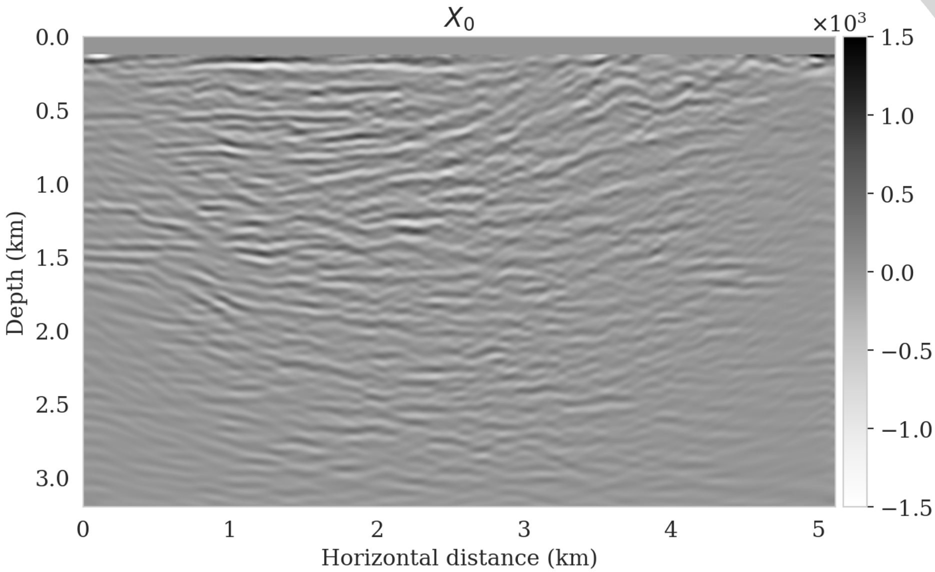
can be used if $\mathbf{y}$ is out-of-distribution

final estimate is tied to the physics/data

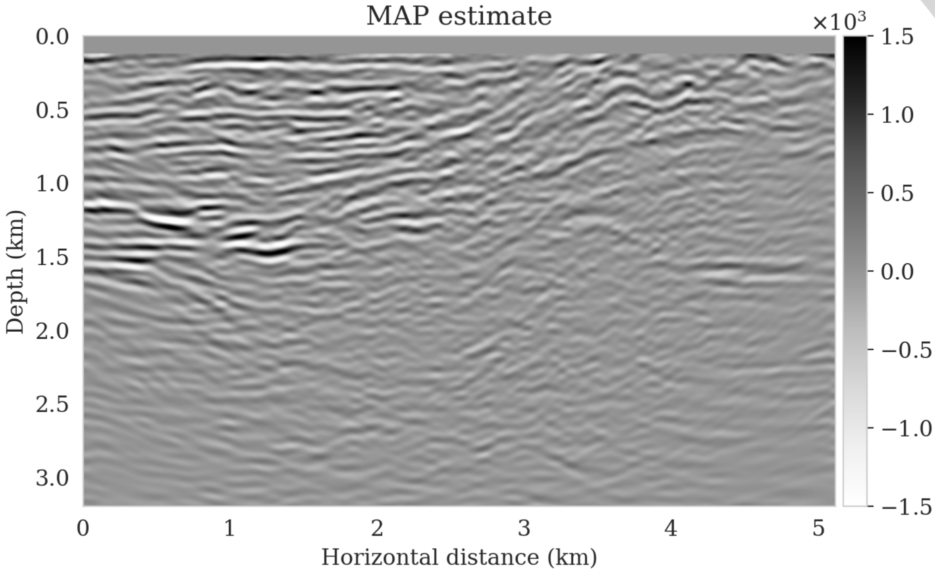
Muhammad Asim, Max Daniels, Oscar Leong, Ali Ahmed, and Paul Hand. "Invertible generative models for inverse problems: mitigating representation error and dataset bias". In: *Proceedings of the 37th International Conference on Machine Learning*. Vol. 119. Proceedings of Machine Learning Research. PMLR, 2020, pp. 399–409. URL: http://proceedings.mlr.press/v119/asim20a.html.

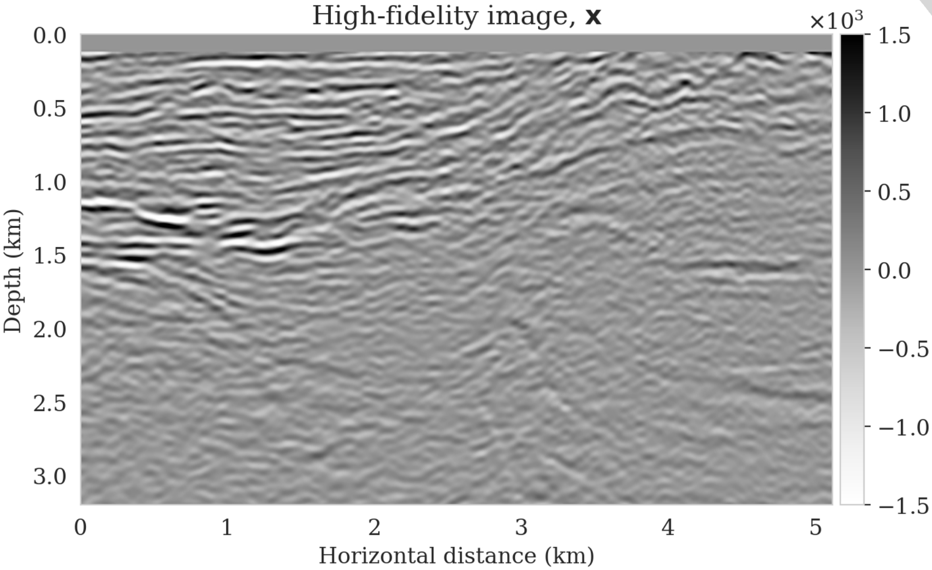
Seismic imaging example

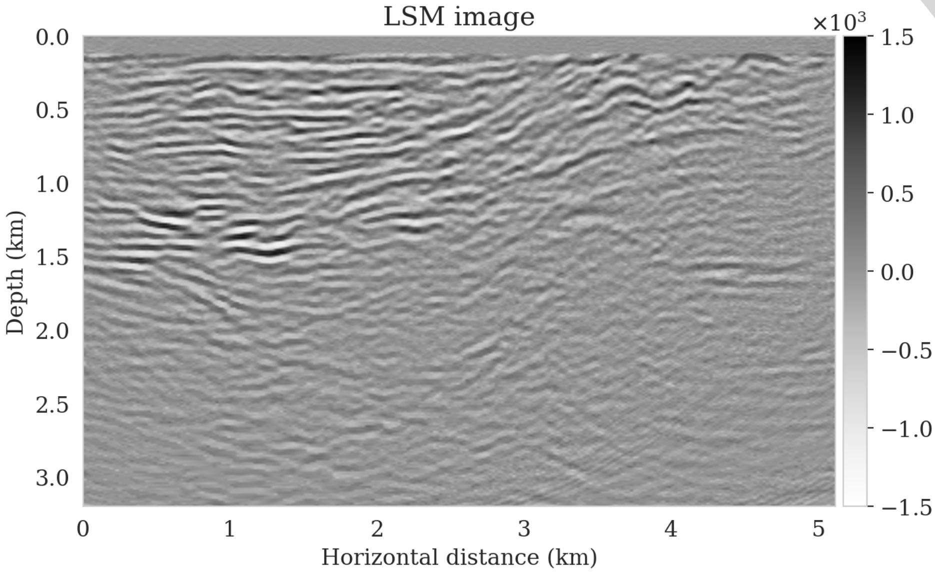

High-fidelity image, x
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)


X_0
$\times 10^3$
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)

SLIM


MAP estimate
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


High-fidelity image, **x**
×10³
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
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5


LSM image
×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)

Preconditioned physics-based approach

$$\min_{\theta_x} \mathbb{E}_{\mathbf{z}\sim \mathrm{N}(\mathbf{0},\mathbf{I})} \left[\frac{1}{2\sigma^2} \left\| F\big(T_{\theta_x}(\mathbf{z})\big) - \mathbf{y} \right\|_2^2 - \log p_{\text{prior}}\big(T_{\theta_x}(\mathbf{z})\big) - \log \left| \det \nabla_z T_{\theta_x}(\mathbf{z}) \right| \right]$$

transfer learning $T_{\theta_x}(\mathbf{z}) := G_{\theta_x}^{-1}(G_{\theta_y}(\mathbf{y}), \mathbf{z})$

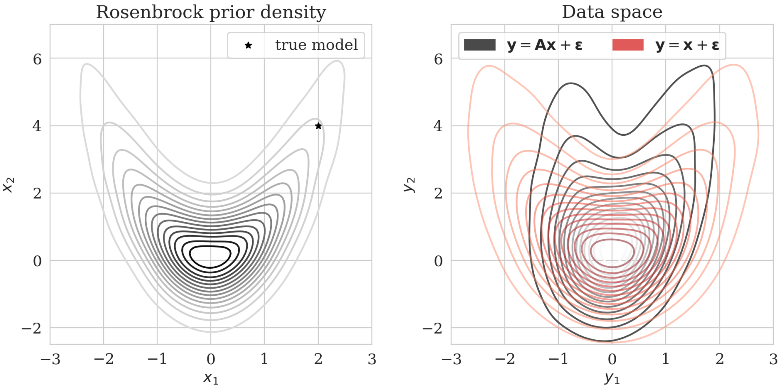
- corrects for errors due to out-of-distribution data
- less evaluations of F and $\nabla F^\top$

learned prior density, $p_{\text{prior}}(\mathbf{x}) := p_{\mathbf{z}}\big(G_{\theta_x}(\mathbf{y}, \mathbf{x})\big) \left| \det \nabla_x G_{\theta_x}(\mathbf{y}, \mathbf{x}) \right|$

fast to adapt to new observation

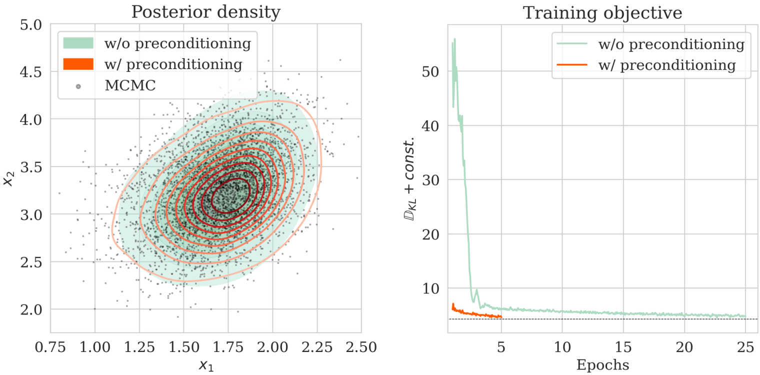
Ali Siahkoohi, Gabrio Rizzuti, Mathias Louboutin, Philipp Witte, and Felix J. Herrmann. "Preconditioned training of normalizing flows for variational inference in inverse problems". In: *3rd Symposium on Advances in Approximate Bayesian Inference*. Jan. 2021. URL: https://openreview.net/pdf?id=P9m1sMaNQ8T.

2D Rosenbrock toy example



(left) prior, (right) low- and high-fidelity data

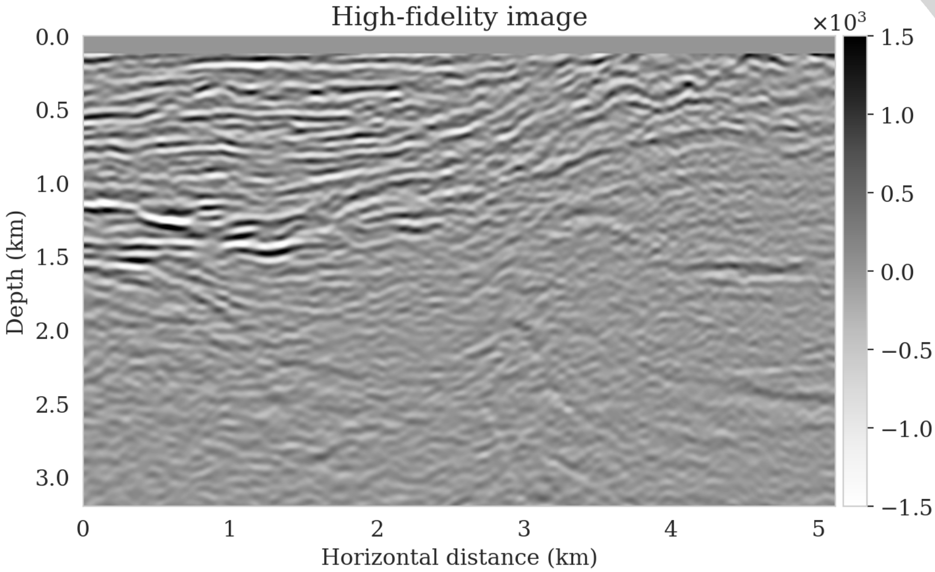
Ali Siahkoohi, Gabrio Rizzuti, Mathias Louboutin, Philipp Witte, and Felix J. Herrmann. "Preconditioned training of normalizing flows for variational inference in inverse problems". In: *3rd Symposium on Advances in Approximate Bayesian Inference*. Jan. 2021. URL: `https://openreview.net/pdf?id=P9m1sMaNQ8T`.

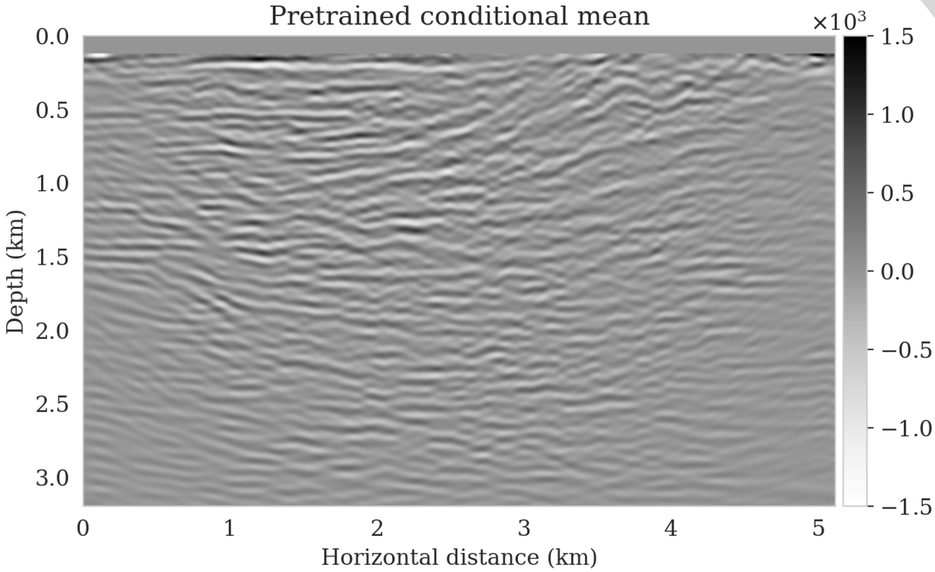


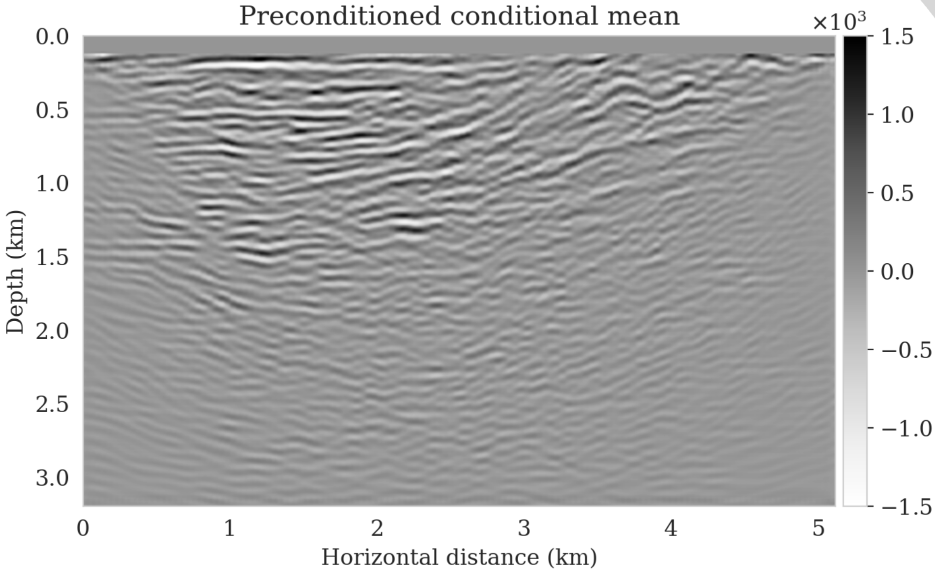
(left) without vs with preconditioning vs MCMC, (right) loss with and without preconditioning

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.

Seismic imaging example

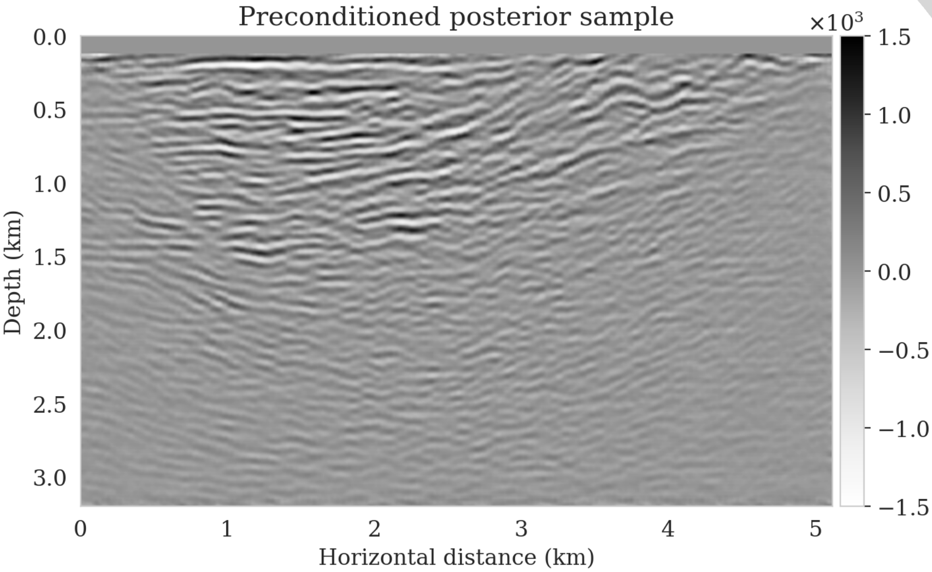

High-fidelity image
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

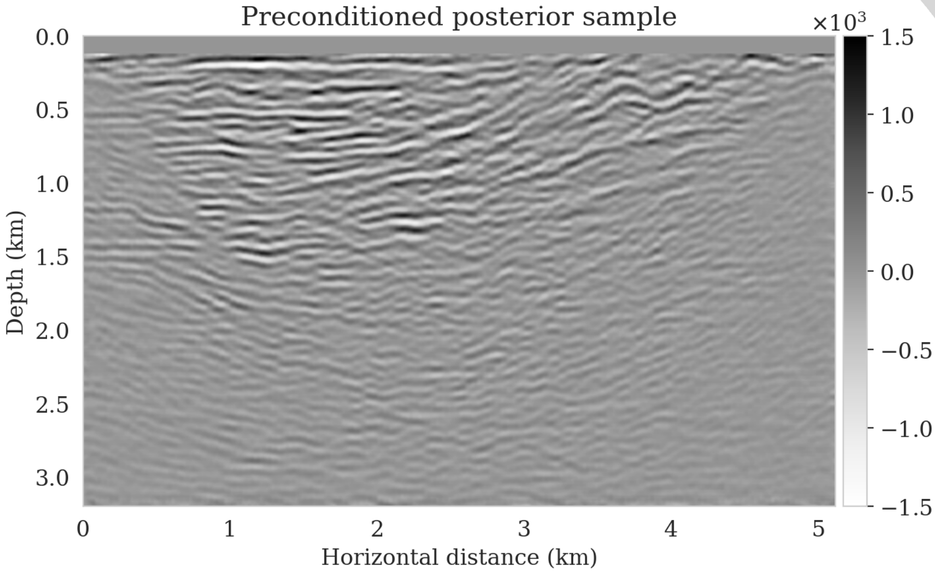

Pretrained conditional mean
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
$\times 10^3$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

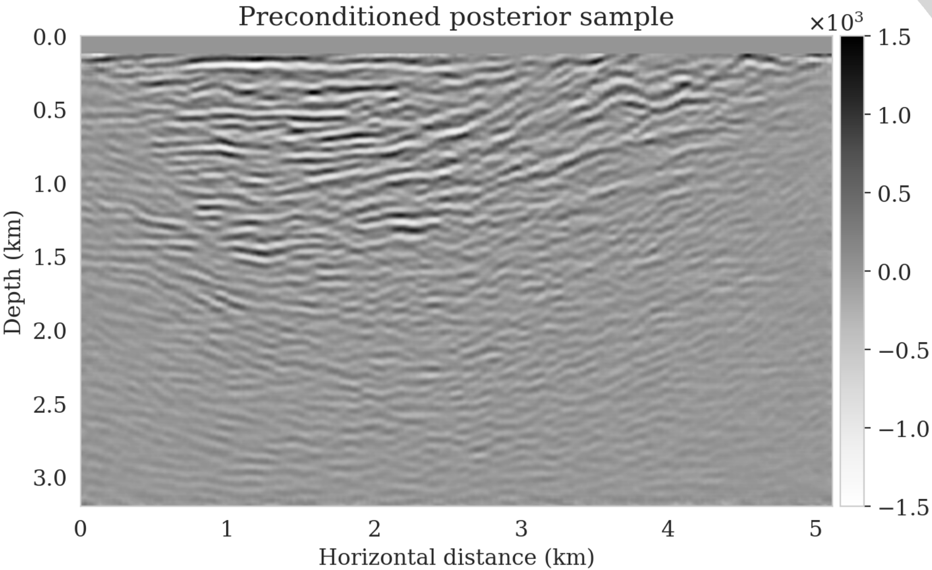


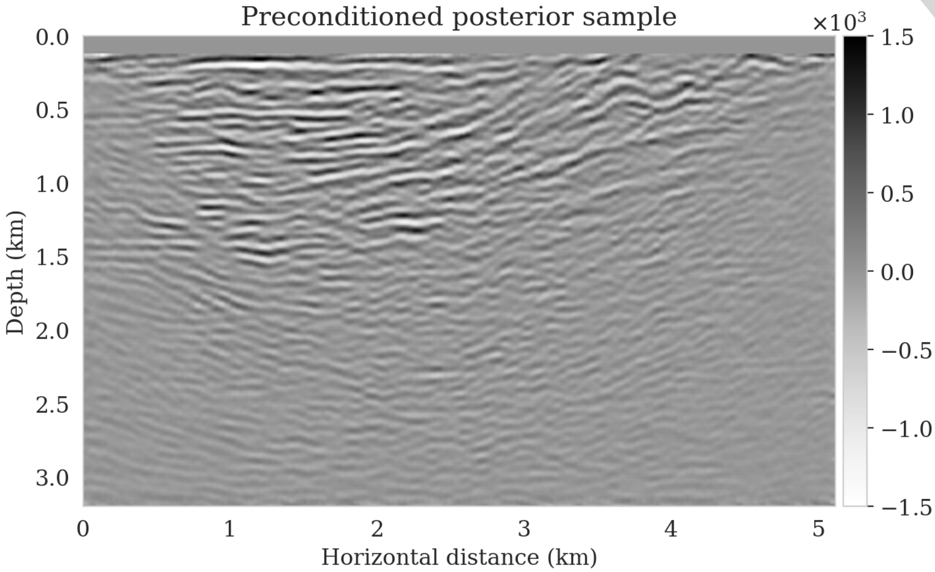
Preconditioned conditional mean
×10^3
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Depth (km)
0
1
2
3
4
5
Horizontal distance (km)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

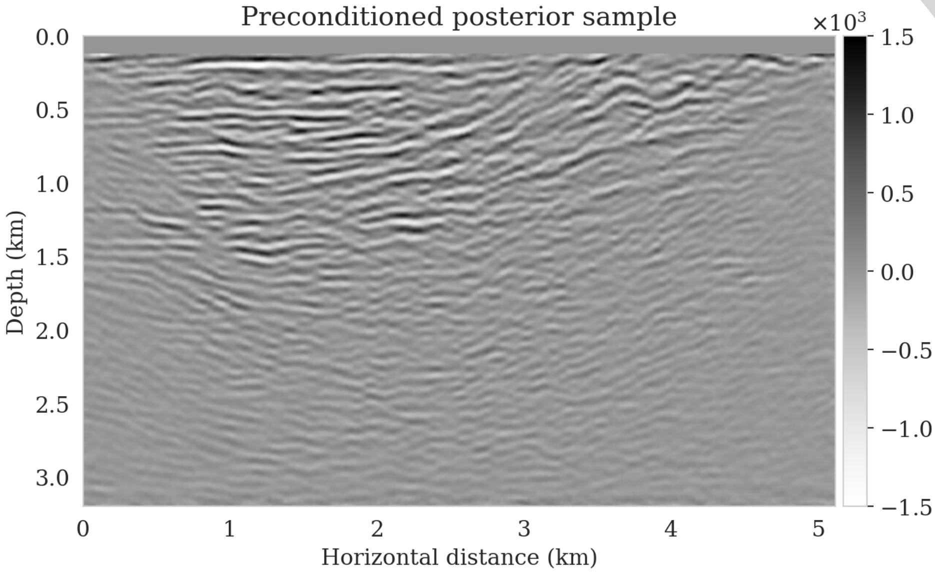
SLIM


Preconditioned posterior sample
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
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5


Preconditioned posterior sample
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

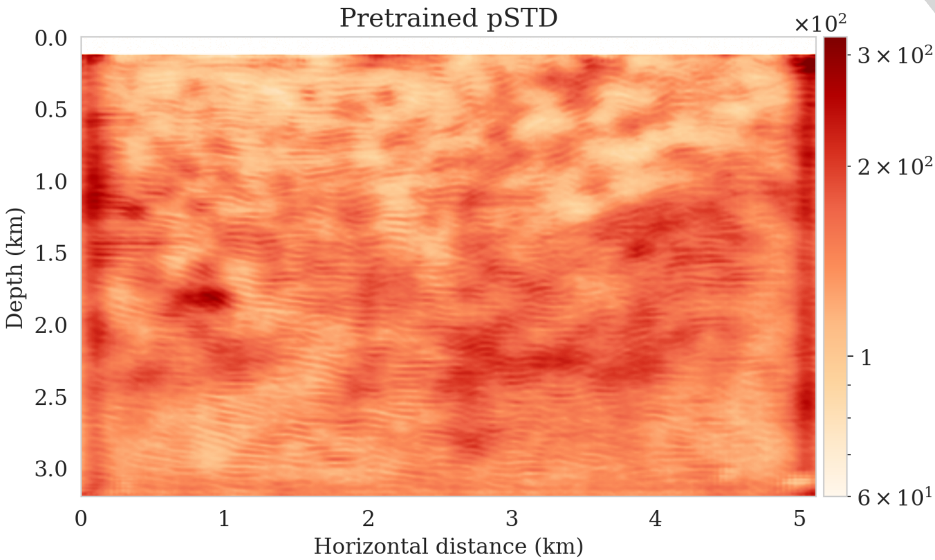

Preconditioned posterior sample
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0
1
2
3
4
5
Horizontal distance (km)
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

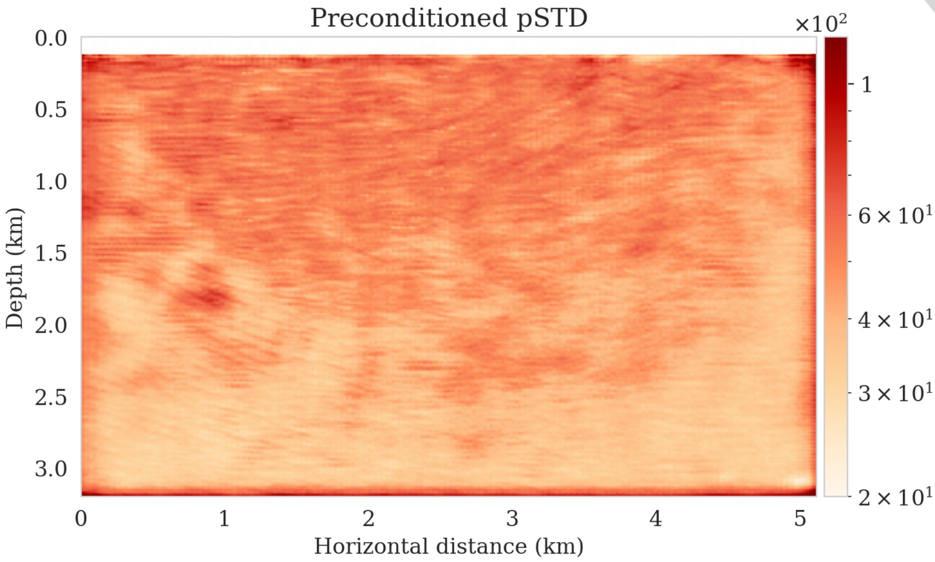

Preconditioned posterior sample
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
×10³
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

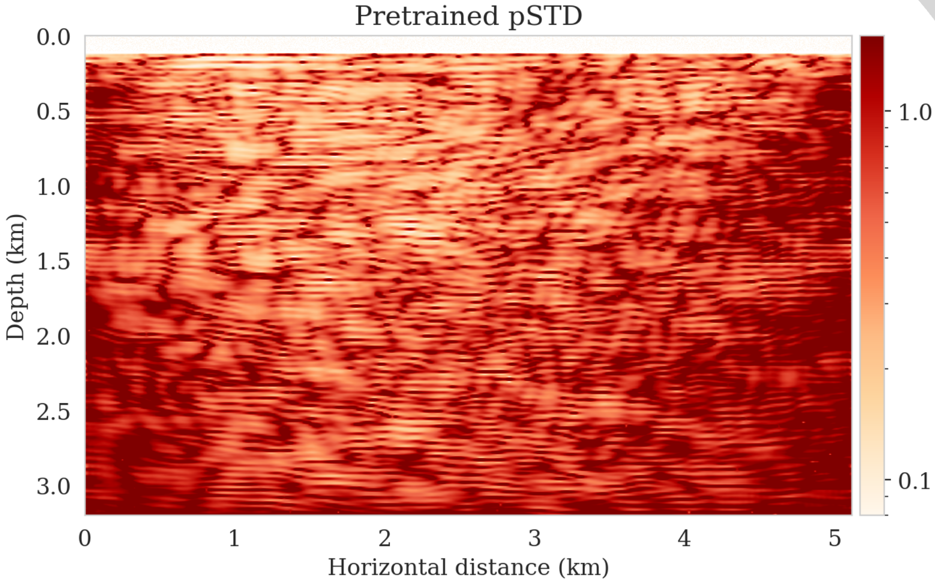


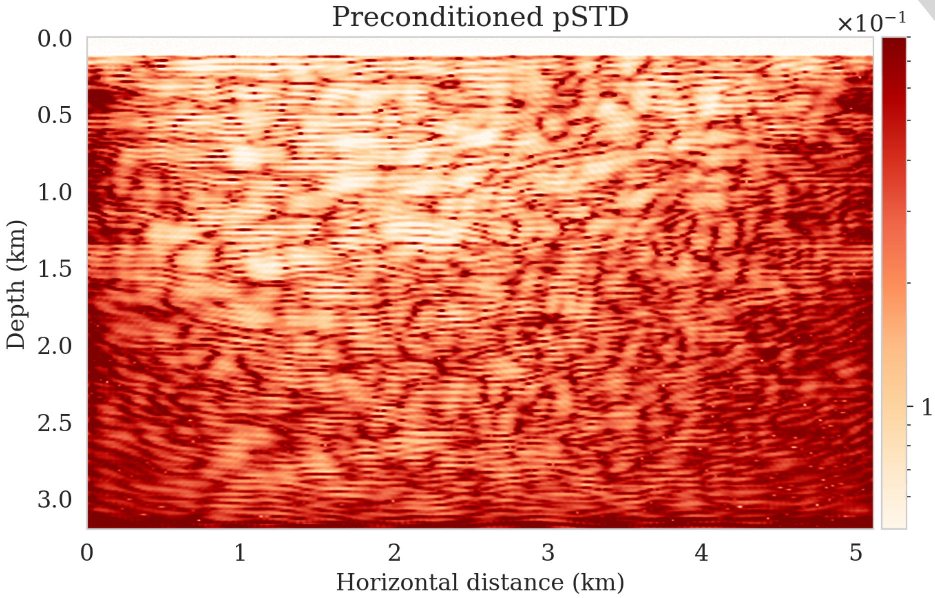
Preconditioned posterior sample
Depth (km)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Horizontal distance (km)
0
1
2
3
4
5
×10^3
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5

SLIM


Pretrained pSTD
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)
$\times 10^2$
3×10^2
2×10^2
1
6×10^1


Preconditioned pSTD
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
×10^2
1
6 × 10^1
4 × 10^1
3 × 10^1
2 × 10^1


Pretrained pSTD
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)
1.0
0.1



Preconditioned pSTD
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)
$\times 10^{-1}$
1

Related work

SLIM

Purely physics-based approach NF-based Bayesian inference

show orders of magnitude speed up compared to traditional MCMC methods

need to train the NF from scratch for a new observation $\mathbf{y}$

use handcrated priors, often negatively bias the inversion

Gabrio Rizzuti, Ali Siahkoohi, Philipp A. Witte, and Felix J. Herrmann. "Parameterizing uncertainty by deep invertible networks, an application to reservoir characterization". In: *SEG Technical Program Expanded Abstracts*. Sept. 2020, pp. 1541–1545. DOI: 10.1190/segam2020-3428150.1.

Xin Zhang and Andrew Curtis. "Seismic tomography using variational inference methods". In: *Journal of Geophysical Research: Solid Earth* 125.4 (2020), e2019JB018589.

Xuebin Zhao, Andrew Curtis, and Xin Zhang. "Bayesian Seismic Tomography using Normalizing Flows". In: (2020). DOI: 10.31223/X53K6G. URL: https://eartharxiv.org/repository/view/1940/.

Related work

Data-driven approaches for directly sampling from the posterior distribution

fast Bayesian inference given new observation $\mathbf{y}$

sample the posterior virtually for free

not directly tied to the physics/data

not reliable when applied to out-of-distribution data

not trivial to scale GAN-based approaches to large-scale problems

---

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`.

Jonas Adler and Ozan Öktem. "Deep Bayesian Inversion". In: *arXiv preprint arXiv:1811.05910* (2018).

Nikola Kovachki, Ricardo Baptista, Bamdad Hosseini, and Youssef Marzouk. *Conditional Sampling With Monotone GANs*. 2021. arXiv: `2006.06755 [stat.ML]`.

Jan-Hinrich Nölke et al. "Invertible neural networks for uncertainty quantification in photoacoustic imaging". In: *arXiv preprint arXiv:2011.05110* (2020).

Related work

Injective network for inverse problems and uncertainty quantification

use an injective map to map data to a low-dimensional space

can be used as a prior (projection operator) in inverse problems

heavily relies on availability of training data to generalize

unclear how the learned projection operator behaves when applied to out-of-distribution data

for new observation $\mathbf{y}$, Bayesian inference requires training an additional NF, involving the forward operator

Konik Kothari, AmirEhsan Khorashadizadeh, Maarten de Hoop, and Ivan Dokmanić. *Trumpets: Injective Flows for Inference and Inverse Problems*. 2021. arXiv: 2102.10461 [cs.LG].

Contributions

Take full advantage of existing training data to provide

- low-fidelity but fast conditional mean estimate
- a first assessment of the image's reliability
- preconditioned physics-based high-fidelity MAP estimate via the learned prior
- preconditioned, scalable, and physics-based Bayesian inference

Conclusions

Obtaining UQ information is rendered impractical when

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

There are strong indications that

- Bayesian inference with normalizing flows can lead to orders of magnitude computational improvements compared to MCMC methods
- preconditioning with a pretrained conditional normalizing flow can lead to another order of magnitude speed up

Future work

- characterize uncertainties due to modeling errors

Code

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

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

https://github.com/slimgroup/Software.SEG2021

Acknowledgment

This research was carried out with the support of

Georgia Research Alliance and partners of the ML4Seismic Center