A deep-learning based Bayesian approach to seismic imaging and uncertainty quantification

Ali Siahkoohi, Gabrio Rizzuti, and Felix J. Herrmann

School of Computational Science and Engineering
Georgia Institute of Technology

February 2020

SLIM
Georgia Institute of Technology

Inverse problems

Estimate unknown parameters of a system via indirect measurements

- seismology: estimate the speed of sound in subsurface of the Earth
- medical imaging: infer visual representations of the interior of a body (X-ray radiography, MRI)

Inverse problems

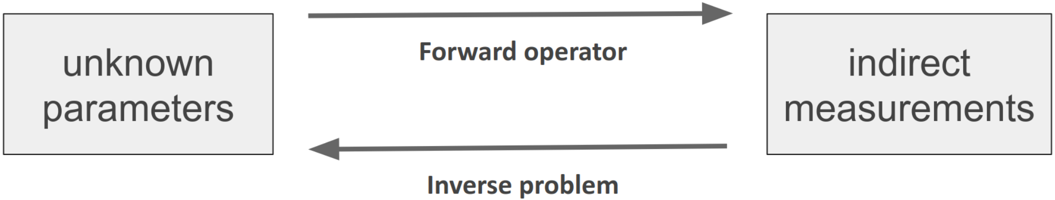


Figure 1: A generic inverse problem

Seismic data acquisition

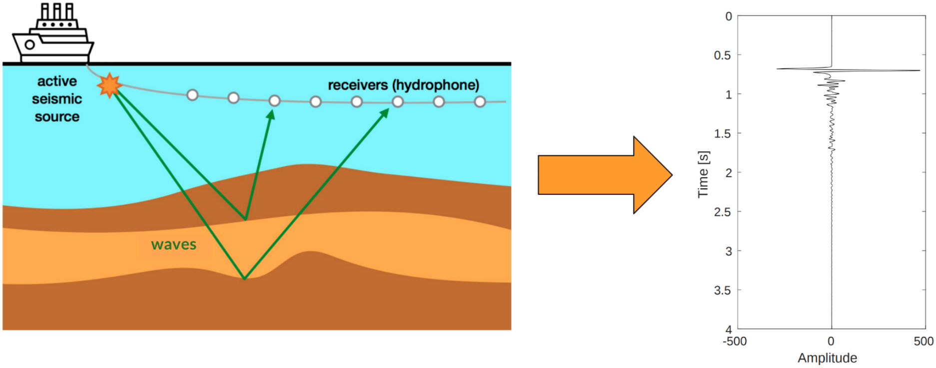


Figure 2: A schematic representation of a seismic survey.[1]

[1] Zhilong Fang. "Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion". PhD thesis. University of British Columbia, 2018.

Shot record - one source location

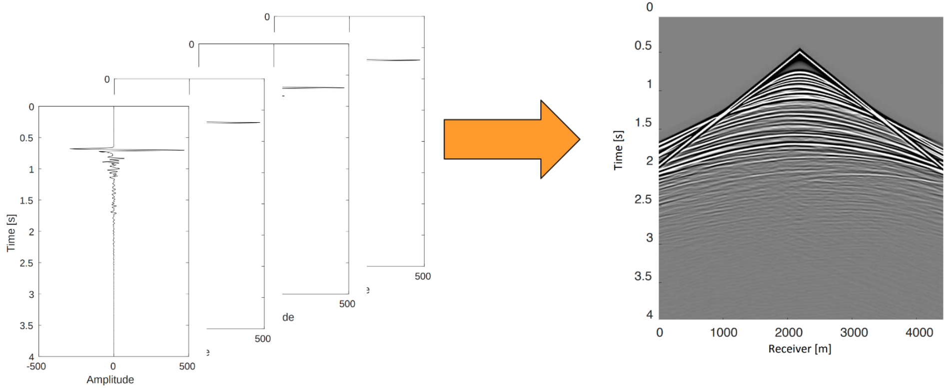


Figure 3: Shot record— a collection of traces associated with one source.[2]

[2] Fang, "Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion".

Seismic data volume

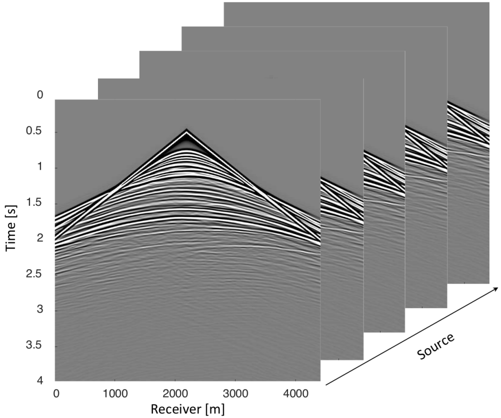


Figure 4: The entire seismic data obtained from one survey.[3]

[3]Fang, "Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion".

Objective of exploration seismology

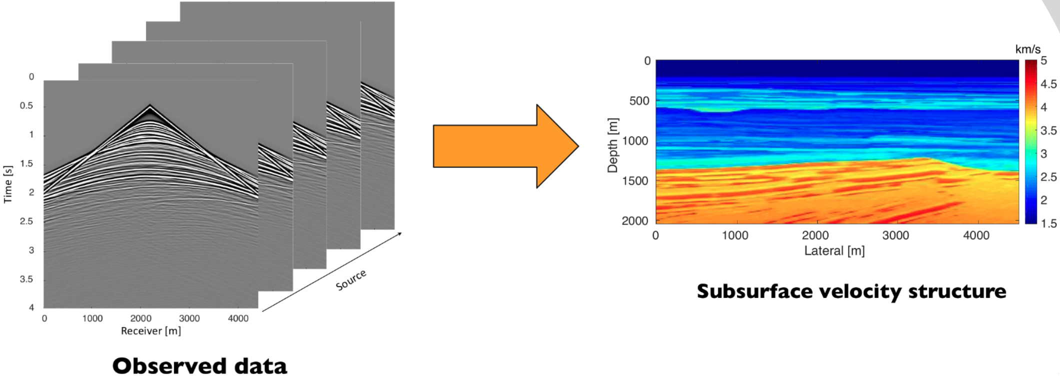


Figure 5: Ultimate goal.[4]

[4] Fang, "Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion".

Wave speed in the subsurface

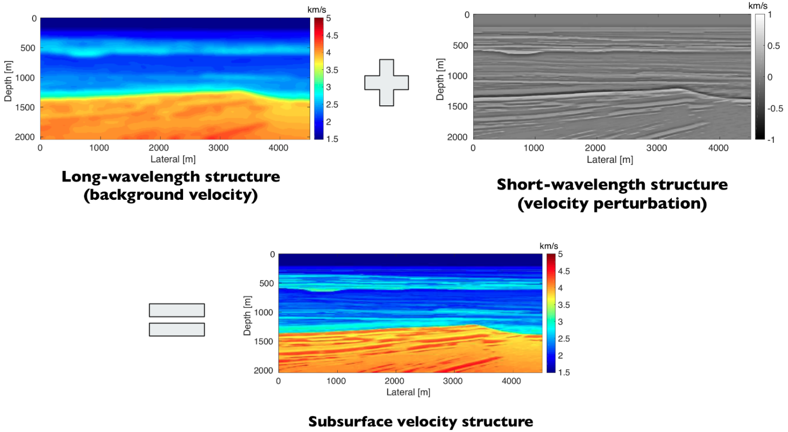


Figure 6: Long- and short-wavelength components of velocty model.[5]

[5] Fang, "Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion".

Seismic imaging

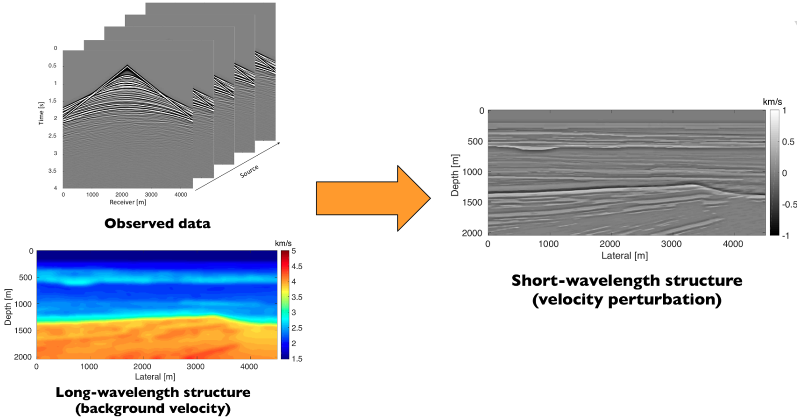


Figure 7: Goal of seismic imaging.[6]

[6] Fang, "Source estimation and uncertainty quantification for wave-equation based seismic imaging and inversion".

Nonlinear forward operator

$$F(\mathbf{m}, \mathbf{q}_i) = \mathbf{PA}(\mathbf{m})^{-1}\mathbf{q}_i, \quad i = 1, 2, \ldots, N \tag{1}$$

where

- $F(\cdot, \cdot)$: nonlinear forward operator
- $\mathbf{m}$: squared-slowness model
- $\mathbf{q}_i$: source signature of i^{th} source experiment
- $\mathbf{P}$: restriction operator, restricting wavefields to the receiver locations
- $\mathbf{A}(\cdot)$: discretized wave equation
- N: number of source experiments

Taylor's series expansion

$$
\begin{aligned}
F(\mathbf{m}_0 + \delta\mathbf{m}, \mathbf{q}_i) &= F(\mathbf{m}_0, \mathbf{q}_i) + \nabla F(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|_2^2) \\
\mathbf{d}_i^{(\text{obs})} &= \mathbf{d}_i^{(\text{pred})} + \underbrace{\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)}_{\text{linear forward op.}}\ \delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|_2^2)
\end{aligned} \tag{2}
$$

where

- $\mathbf{m}_0$: background (long-wavelength) squared-slowness model
- $\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$: unknown squared-slowness *perturbation model*
- $\mathbf{d}_i^{(\text{obs})} = F(\mathbf{m}_0 + \delta\mathbf{m}, \mathbf{q}_i)$: observed data
- $\mathbf{d}_i^{(\text{pred})} = F(\mathbf{m}_0, \mathbf{q}_i)$: predicted (simulated) data
- $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i) = \nabla F(\mathbf{m}_0, \mathbf{q}_i)$: linear *forward* operator
- $\mathcal{O}(\|\delta\mathbf{m}\|_2^2)$: *linearization* error

Seismic imaging

$$\delta \mathbf{d}_i \;= \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta \mathbf{m} \quad + \quad \underbrace{\mathcal{O}(\|\delta \mathbf{m}\|_2^2)}_{\text{unknown linearization error}} \tag{3}$$

where

- $\delta \mathbf{d}_i = \mathbf{d}_i^{(\text{obs})} - \mathbf{d}_i^{(\text{pred})}$: data *residual*

Seismic imaging— challenges

Involves an inconsistent, mildly ill-conditioned linear inverse problem due to:

- presence of shadow zones and complex structures in the subsurface
- coherent linearization error—i.e., $\delta\mathbf{d}_i = \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{m} + \mathcal{O}(\|\delta\mathbf{m}\|_2^2)$
- noise in observed data—i.e., $\mathbf{d}_i^{(\text{obs})} = F(\mathbf{m}, \mathbf{q}_i) + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim p_{\text{noise}}(\boldsymbol{\epsilon})$

Requires prior/regularization. Due to Earth's heterogeneity:

- not possible to precisely encode our prior knowledge
- do not have access to samples from *ground-truth* prior to utilize data-driven priors

Seismic imaging— challenges

Computational challenges

- applying $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)$ or $\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)^T$ involves two *expensive* PDE solves:

$$\mathbf{J}(\mathbf{m}_0, \mathbf{q}_i) = -\mathbf{P}\underbrace{\mathbf{A}(\mathbf{m}_0)^{-1}[\nabla\mathbf{A}(\mathbf{m}_0)(\overbrace{\mathbf{A}(\mathbf{m}_0)^{-1}\mathbf{q}_i}^{\text{PDE solve}})]}_{\text{PDE solve}} \tag{4}$$

- many source experiments—i.e., N is large

Bayesian inversion

Why Bayesian?

- uncertainty quantification
- incorporating uncertainty into the inversion—e.g., conditional mean estimate

Ultimate goal:

- sampling the posterior distribution

Challenges:

- need for a prior distribution
- expensive to sample the posterior

A Bayesian approach to seismic imaging

Bayes' rule:

$$p_{\text{post}}\left(\delta\mathbf{m} \mid \{\delta\mathbf{d}_i, \mathbf{q}_i\}_{i=1}^{N}\right) \propto p_{\text{noise}}\left(\{\delta\mathbf{d}_i, \mathbf{q}_i\}_{i=1}^{N} \mid \delta\mathbf{m}\right) p_{\text{prior}}\left(\delta\mathbf{m}\right) \tag{5}$$

where

- p_{post}: posterior distribution density
- p_{noise}: density of the noise distribution
- p_{prior}: prior distribution density

Prior distribution

Conventional methods— handcrafted and unrealistic priors [7],

- Gaussian or Laplace distribution prior in the physical/transform domain
- tend to bias the outcome of inversion

Pretrained generative models as an implicit prior [8,9],

- i.e., requires samples from the *ground-truth* prior distribution
- allows for MCMC sampling in the low-dimensional latent space

---

[7] Zhilong Fang et al. "Uncertainty quantification for inverse problems with weak partial-differential-equation constraints". In: *GEOPHYSICS* 83.6 (2018), R629–R647. DOI: 10.1190/geo2017-0824.1.

[8] Lukas Mosser, Olivier Dubrule, and M Blunt. "Stochastic Seismic Waveform Inversion Using Generative Adversarial Networks as a Geological Prior". In: *Mathematical Geosciences* 84.1 (2019), pp. 53–79. DOI: 10.1007/s11004-019-09832-6.

[9] Dhruv Patel and Assad A Oberai. "Bayesian Inference with Generative Adversarial Network Priors". In: *Neural Information Processing Systems (NeurIPS) 2019 Deep Inverse Workshop*. Dec. 2019. URL: https://arxiv.org/pdf/1907.09987.pdf.

Prior distribution

Joint inversion and training a generative model [10],

- does *not* require a pretrained generative model
- fast posterior sampling— feed-forward evaluation of the generative model

Proposed approach— an implicit structured *deep prior* [11,12,13],

- i.e., reparameterize $\delta\mathbf{m}$ w/ a randomly initialized deep CNN
- promotes *natural* images, but not unnatural noise

[10] 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.

[11] V. Lempitsky, A. Vedaldi, and D. Ulyanov. "Deep Image Prior". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 2018, pp. 9446–9454. DOI: 10.1109/CVPR.2018.00984.

[12] Zezhou Cheng et al. "A Bayesian Perspective on the Deep Image Prior". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019, pp. 5443–5451.

[13] Yulang Wu and George A McMechan. "Parametric convolutional neural network-domain full-waveform inversion". In: *GEOPHYSICS* 84.6 (2019), R881–R896. DOI: 10.1190/geo2018-0224.1.

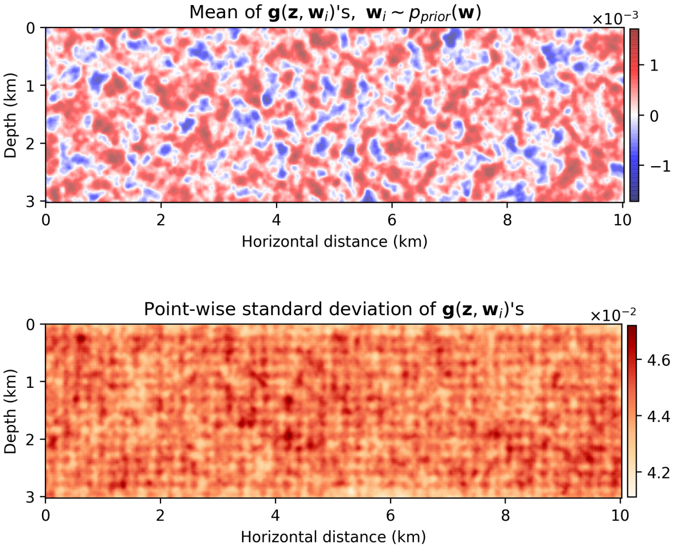
Prior distribution— implicit deep prior

$$\delta\mathbf{m} = \mathbf{g}(\mathbf{z}, \mathbf{w}), \quad \mathbf{w} \sim p_{\text{prior}}(\mathbf{w}) := \mathrm{N}(\mathbf{w} \mid \mathbf{0}, \frac{1}{\lambda^2}\mathbf{I}) \tag{6}$$

where

- $\mathbf{g}(\cdot, \cdot)$: a randomly initialized deep CNN
- $\mathbf{z} \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$: fixed input to the CNN
- $\mathbf{w}$: unknown CNN weights—e.g., convolutional kernels and biases
- $p_{\text{prior}}(\mathbf{w})$: Gaussian prior on $\mathbf{w}$
- λ: a hyperparameter

First and second order statistics of the implicit deep prior



Negative log-likelihood

$$
\begin{aligned}
-\log p_{\text{noise}}\left(\{\delta\mathbf{d}_i, \mathbf{q}_i\}_{i=1}^{N} \mid \mathbf{w}\right) &= -\sum_{i=1}^{N} \log p_{\text{noise}}\left(\delta\mathbf{d}_i, \mathbf{q}_i \mid \mathbf{w}\right) \\
&= \frac{1}{2\sigma^2} \sum_{i=1}^{N} \|\delta\mathbf{d}_i - \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\mathbf{g}(\mathbf{z}, \mathbf{w})\|_2^2 \quad + \quad \underbrace{\text{const},}_{\text{independent of } \mathbf{w}}
\end{aligned}
\tag{7}
$$

where

- p_{noise}: Gaussian distribution on the noise
- σ^2: estimated noise variance

Negative log-posterior

$$
\begin{aligned}
& -\log p_{\text{post}}\left(\mathbf{w} \mid \{\delta\mathbf{d}_i, \mathbf{q}_i\}_{i=1}^{N}\right) \\
& = \frac{1}{2\sigma^2}\sum_{i=1}^{N} \|\delta\mathbf{d}_i - \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\mathbf{g}(\mathbf{z}, \mathbf{w})\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{w}\|_2^2 \quad + \underbrace{\text{const},}_{\text{independent of } \mathbf{w}}
\end{aligned}
\tag{8}
$$

where

- p_{post}: posterior distribution density on $\mathbf{w}$

Point estimators [14]

Maximum likelihood estimator:

$$\begin{aligned}\widehat{\delta \mathbf{m}_{MLE}} &= \underset{\delta \mathbf{m}}{\operatorname{argmin}} - \log p_{\text{noise}} \left(\{\delta \mathbf{d}_i, \mathbf{q}_i\}_{i=1}^N \mid \delta \mathbf{m} \right) \\ &= \underset{\delta \mathbf{m}}{\operatorname{argmin}} \frac{1}{2\sigma^2} \sum_{i=1}^{N} \|\delta \mathbf{d}_i - \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\delta \mathbf{m}\|_2^2\end{aligned} \tag{9}$$

- computationally feasible for large-scale problems
- does not regularize, not suitable for ill-posed problems
- does not incorporate uncertainty

[14] Ozan Oktem. *Bayesian inversion for tomography through machine learning*. https://imaging-in-paris.github.io/semester2019/slides/w3/Oktem.pdf. Accessed: 2020-02-04.

Point estimators [15]

Maximum a-posteriori (MAP) estimator:

$$
\begin{aligned}
\widehat{\delta \mathbf{m}_{MAP}} &= \mathbf{g}(\mathbf{z}, \widehat{\mathbf{w}_{MAP}}), \\
\text{where,} \quad \widehat{\mathbf{w}_{MAP}} &= \underset{\mathbf{w}}{\operatorname{argmin}} - \log p_{\text{post}} \left(\mathbf{w} \,\middle|\, \{\delta \mathbf{d}_i, \mathbf{q}_i\}_{i=1}^{N} \right) \\
&= \underset{\delta \mathbf{m}}{\operatorname{argmin}} \frac{1}{2\sigma^2} \sum_{i=1}^{N} \|\delta \mathbf{d}_i - \mathbf{J}(\mathbf{m}_0, \mathbf{q}_i)\mathbf{g}(\mathbf{z}, \mathbf{w})\|_2^2 + \frac{\lambda^2}{2} \|\mathbf{w}\|_2^2
\end{aligned}
\tag{10}
$$

- incorporates prior information, suitable for most ill-posed problem
- does not incorporate uncertainty

[15] Oktem, *Bayesian inversion for tomography through machine learning*.

Point estimators [17]

Conditional mean estimator:

$$\widehat{\delta\mathbf{m}} := \mathbb{E}_{\mathbf{w}\sim p_{\text{post}}(\mathbf{w}|\{\delta\mathbf{d}_i,\mathbf{q}_i\}_{i=1}^N)}\left[\mathbf{g}(\mathbf{z},\mathbf{w})\right] = \int p_{\text{post}}(\mathbf{w}|\,\{\delta\mathbf{d}_i,\mathbf{q}_i\}_{i=1}^N)\mathbf{g}(\mathbf{z},\mathbf{w})d\mathbf{w} \tag{11}$$

- incorporates prior information, suitable for most ill-posed problem
- Bayesian learning reduces overfitting [16]
- incorporates uncertainty into inversion
- involves high-dimensional integration

[16] Cheng et al., "A Bayesian Perspective on the Deep Image Prior".
[17] Oktem, *Bayesian inversion for tomography through machine learning*.

Monte Carlo integration

Approximating the integration by a sum with Monte Carlo integration:

1. sample the posterior distribution, $\widehat{\mathbf{w}}_j \sim p_{\text{post}}(\mathbf{w} \mid \{\delta \mathbf{d}_i, \mathbf{q}_i\}_{i=1}^N), \; j = 1, \dots, T$
2. approximate the expectation by the mean using the samples,

$$\mathbb{E}_{\mathbf{w} \sim p_{\text{post}}(\mathbf{w} \mid \{\delta \mathbf{d}_i, \mathbf{q}_i\}_{i=1}^N)} \left[\mathbf{g}(\mathbf{z}, \mathbf{w})\right] \simeq \frac{1}{T} \sum_{j=1}^{T} \mathbf{g}(\mathbf{z}, \widehat{\mathbf{w}}_j) \tag{12}$$

Sampling the posterior distribution

Bayesian inference in deep CNNs,

- generally intractable due to high-dimensional parameters space
- a popular approach to sample the posterior is *stochastic gradient Langevin dynamics* (SGLD) [18]

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \frac{\epsilon}{2}\nabla_{\mathbf{w}} L^{(j)}(\mathbf{w}_k) + \eta_k, \quad \eta_k \sim \mathrm{N}(\mathbf{0}, \epsilon\mathbf{I}), \tag{13}$$

where

- ϵ: stepsize
- $L^{(j)}(\mathbf{w}) = \frac{N}{2\sigma^2}\|\delta\mathbf{d}_j - \mathbf{J}(\mathbf{m}_0, \mathbf{q}_j)\mathbf{g}(\mathbf{z}, \mathbf{w})\|_2^2 + \frac{\lambda^2}{2}\|\mathbf{w}\|_2^2$ approximates the negative-log posterior

[18] 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.

Uncertainty quantification

Pointwise standard deviation as a measure of uncertainty,

- i.e., pointwise standard deviation among $\widehat{\mathbf{w}}_j \sim p_{\text{post}}(\mathbf{w} \mid \{\delta\mathbf{d}_i, \mathbf{q}_i\}_{i=1}^N)$

We expect to see more uncertainty in regions that are more difficult to image—e.g.,

- location of the reflectors
- deeper parts of the model
- close to boundaries and fault zone

Numerical experiment— setup

Synthetic dataset simulated by solving the acoustic wave equation,

- 2D Overthrust velocity model
- finite-difference simulations w/ Devito [19,20]
- 369 shot records w/ 369 receivers
- 27 m source/receiver sampling
- 2 seconds recording time
- Ricker source wavelet w/ 8 Hz central frequency

[19] F. Luporini et al. "Architecture and performance of Devito, a system for automated stencil computation". In: *CoRR* abs/1807.03032 (2018). arXiv: 1807.03032. URL: http://arxiv.org/abs/1807.03032.

[20] M. Louboutin et al. "Devito (v3.1.0): an embedded domain-specific language for finite differences and geophysical exploration". In: *Geoscientific Model Development* 12.3 (2019), pp. 1165–1187. DOI: 10.5194/gmd-12-1165-2019. URL: https://www.geosci-model-dev.net/12/1165/2019/.

Numerical experiment— setup

Simultaneous source experiments [21],

- an economic way to sample seismic data in practice
- wave equation is *linear* in source—i.e., $F(\mathbf{m}, \mathbf{q}) = \mathbf{PA}(\mathbf{m})^{-1}\mathbf{q}$,
- linearly combine observed data w/ normally distributed weights
- generate associated forward operators, $\mathbf{J}(\mathbf{m}_0, \mathbf{q})$, with random superposition of source signatures

[21] Felix J. Herrmann, Yogi A. Erlangga, and Tim T.Y. Lin. "Compressive simultaneous full-waveform simulation". In: *Geophysics* 74.4 (2009), A35–A40. DOI: 10.1190/1.3115122. URL: http://library.seg.org/doi/abs/10.1190/1.3115122.

Numerical experiment— setup

Noise in the data

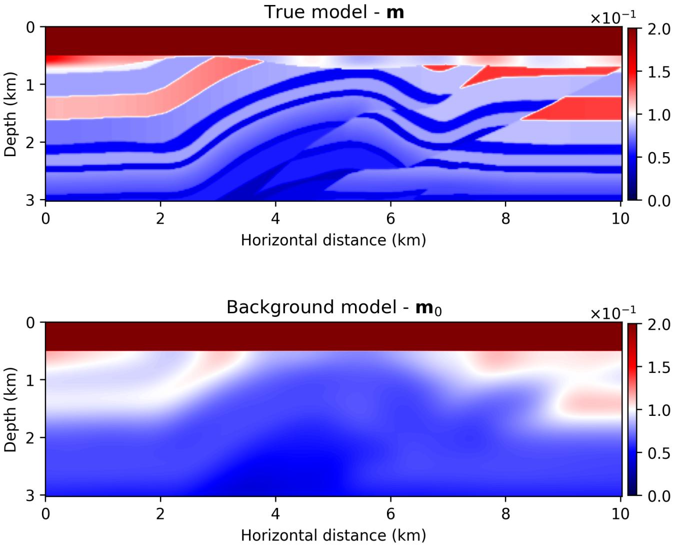
- inherent (non-Gaussian) linearization error— directly estimated variance $= \ 0.490$
- measurement noise drawn from $\mathrm{N}(\mathbf{0}, 2\mathbf{I})$
- overall signal-to-noise ratio of data $= \ -11.37\ \mathrm{dB}$

Deep learning framework

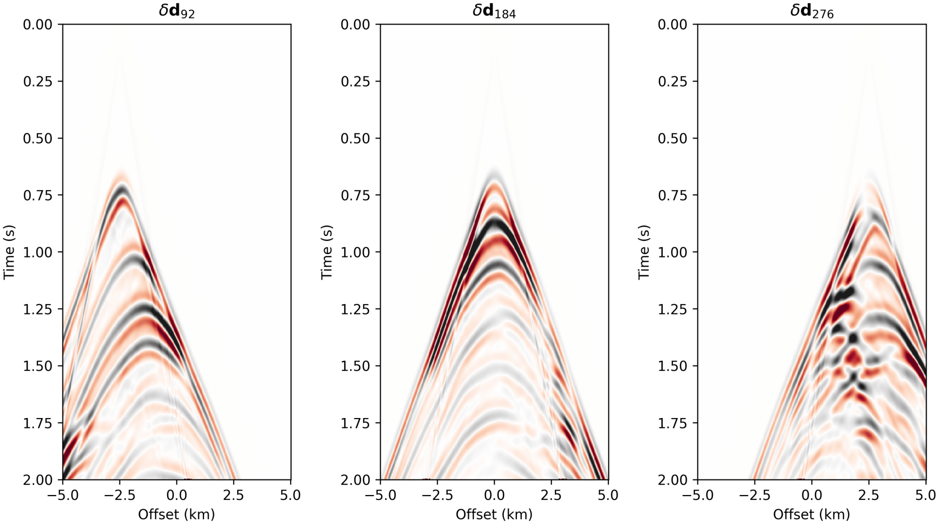
- PyTorch [22] library
- integrate Devito's linear forward operator into PyTorch— allowing us to compute the gradients w/ automatic differentiation.

[22] Adam Paszke et al. "PyTorch: An Imperative Style, High-Performance Deep Learning Library". In: *Advances in Neural Information Processing Systems 32*. 2019, pp. 8024–8035. URL: `http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf`.

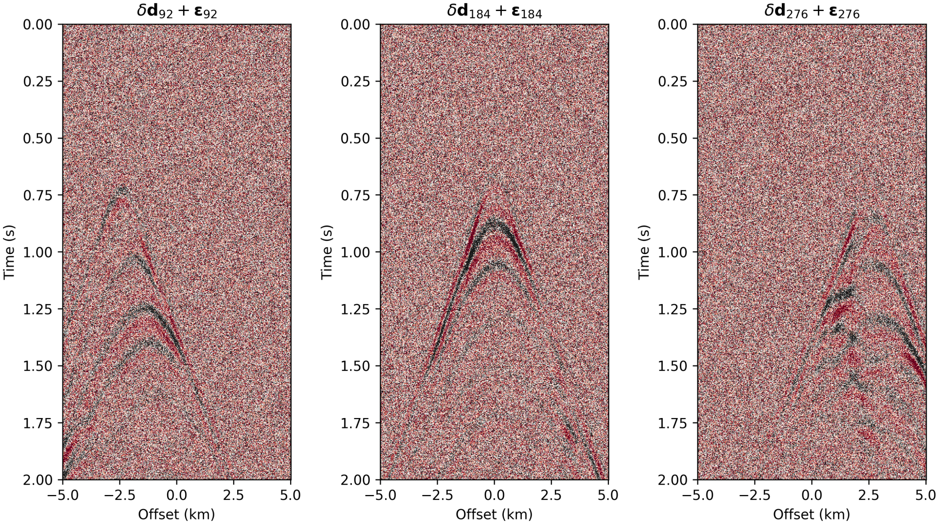
Numerical experiment— squared-slowness model



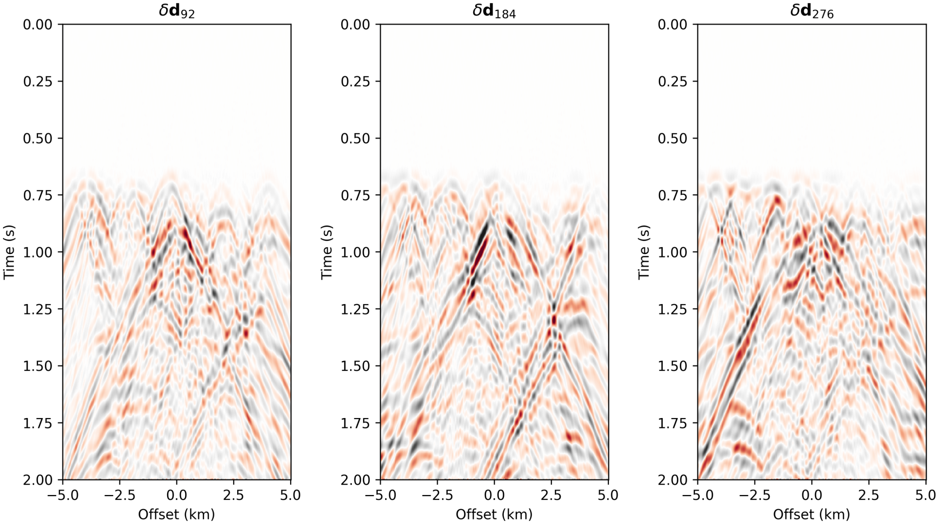
Measurement-noise-free data residual



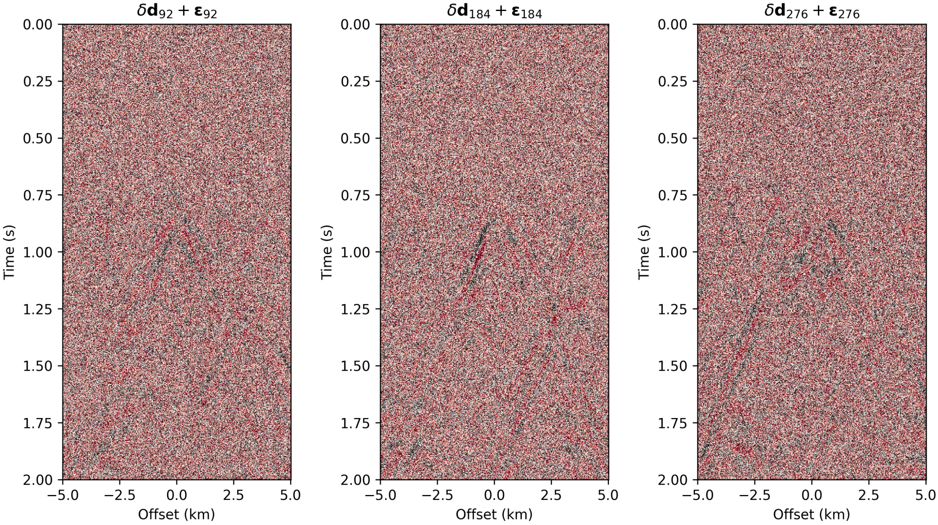
Data residual w/ measurement noise



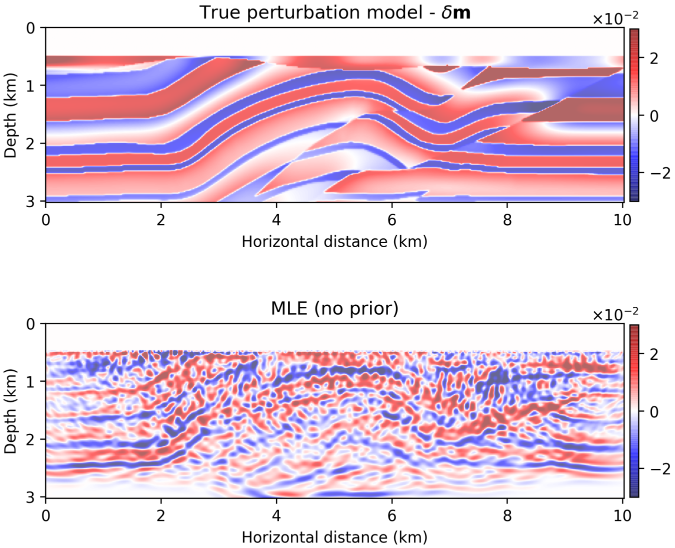
Measurement-noise-free simultaneous source data residual



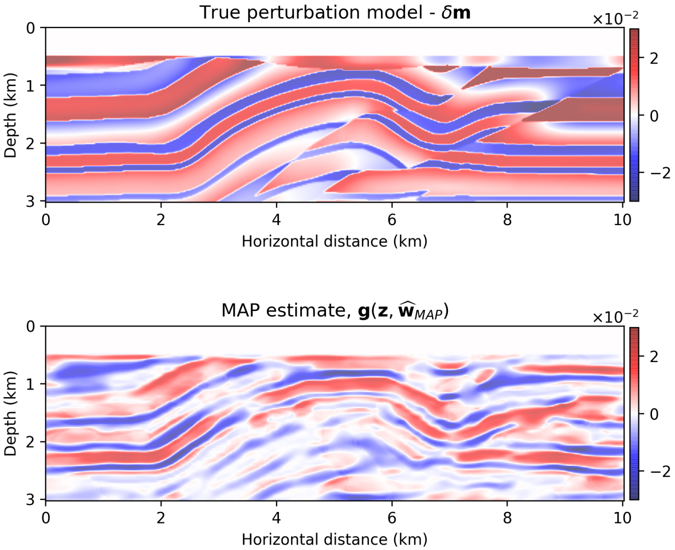
Simultaneous source data residual w/ measurement noise



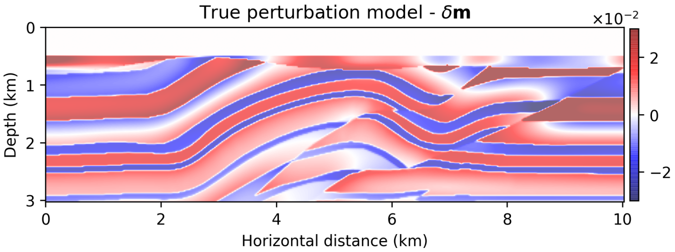
Seismic imaging— MLE

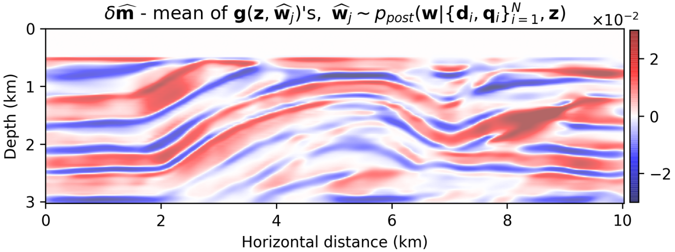


Seismic imaging— MAP estimate

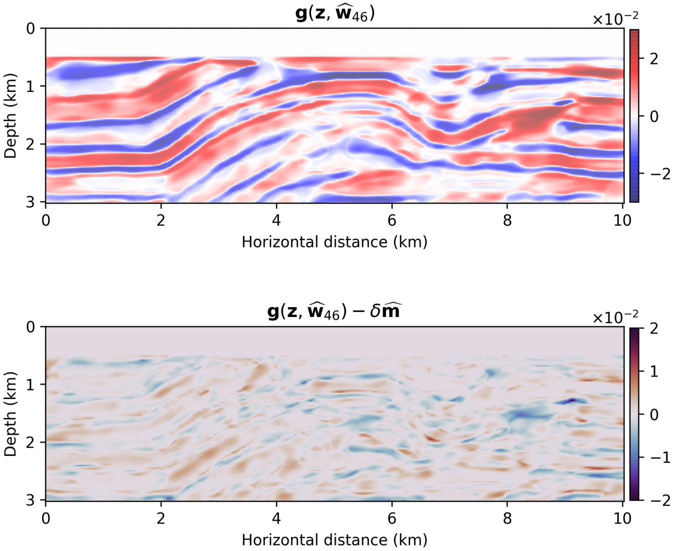


Seismic imaging— conditional mean estimate

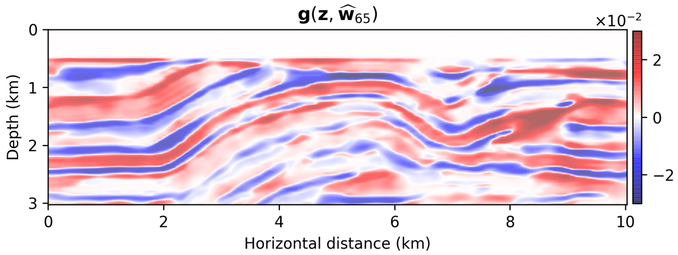


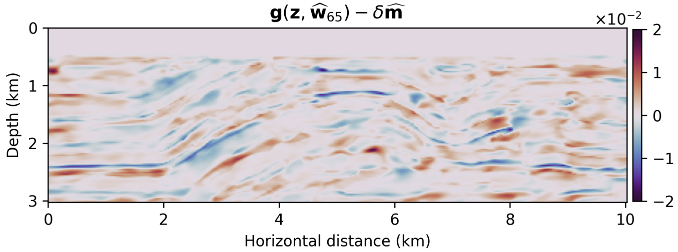


Samples from the posterior distribution

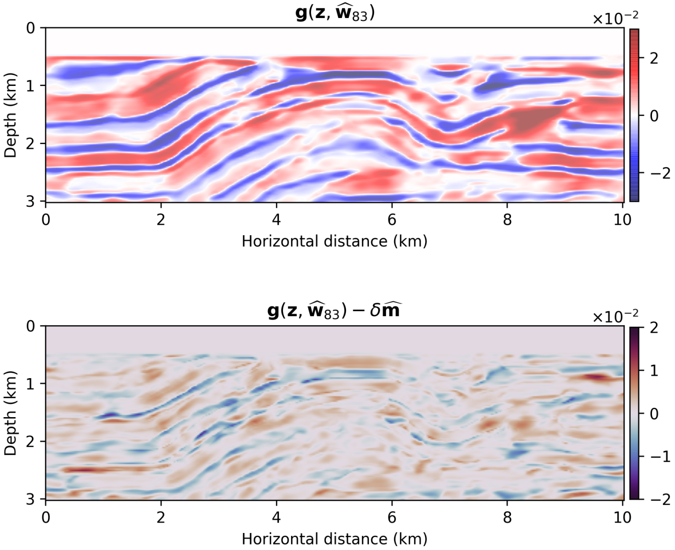


Samples from the posterior distribution

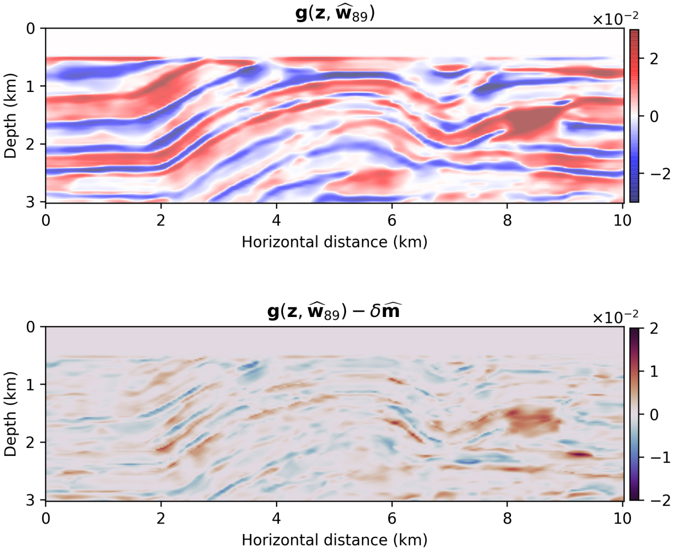




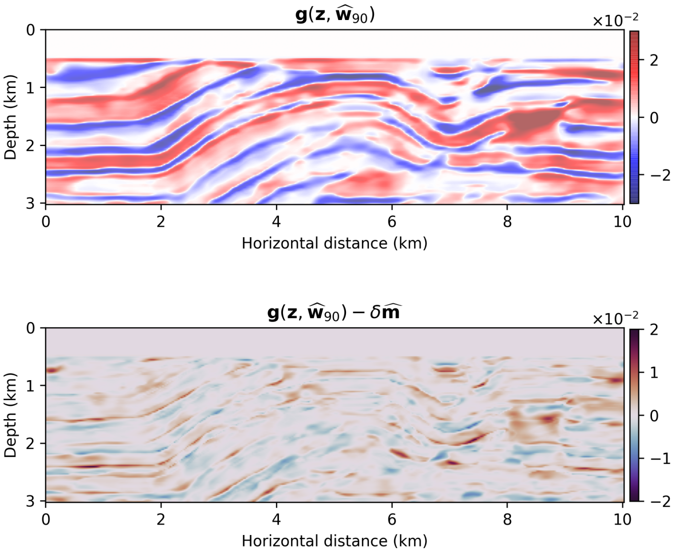
Samples from the posterior distribution



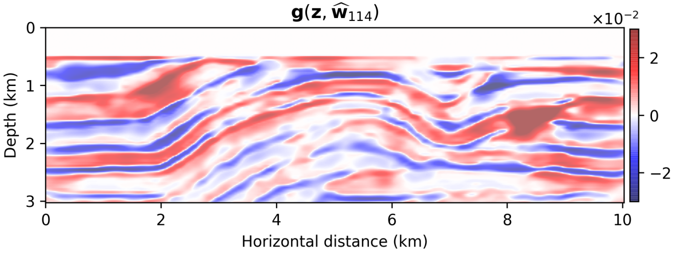
Samples from the posterior distribution

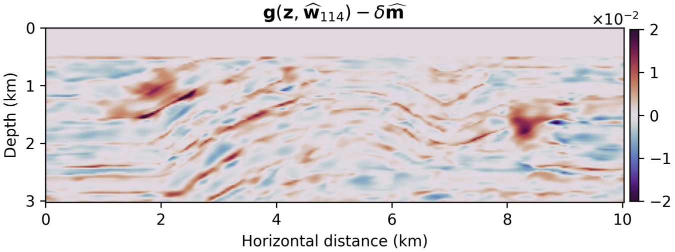


Samples from the posterior distribution

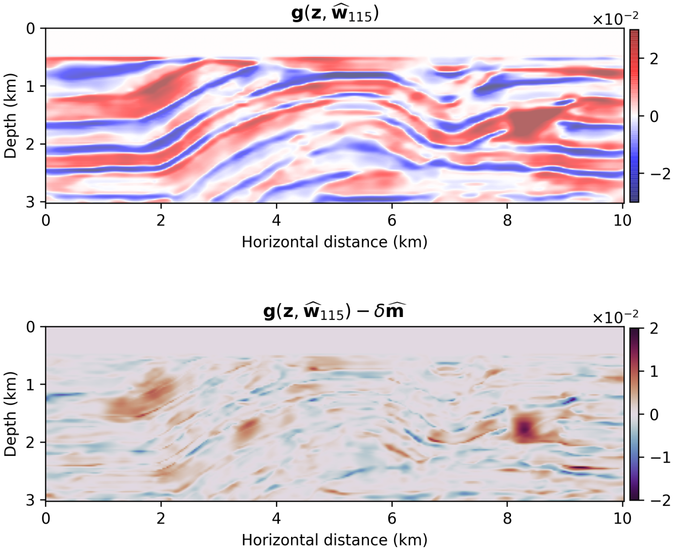


Samples from the posterior distribution

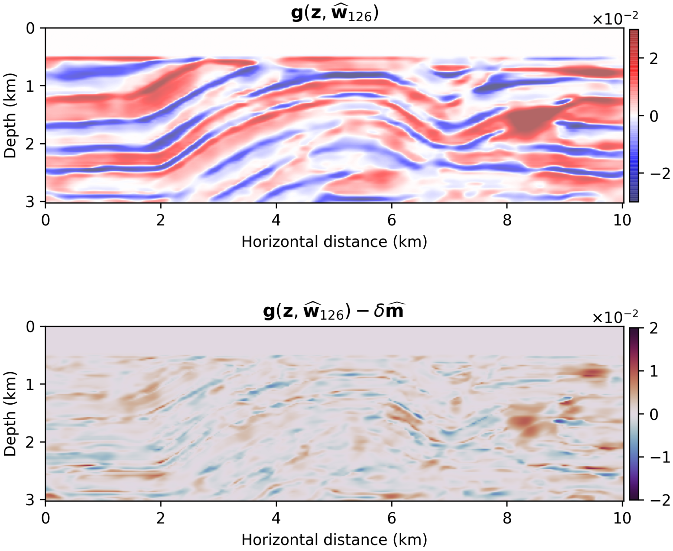




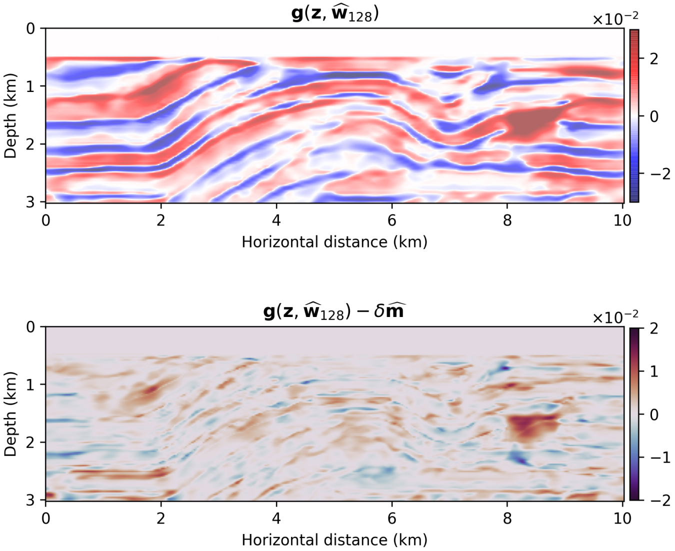
Samples from the posterior distribution



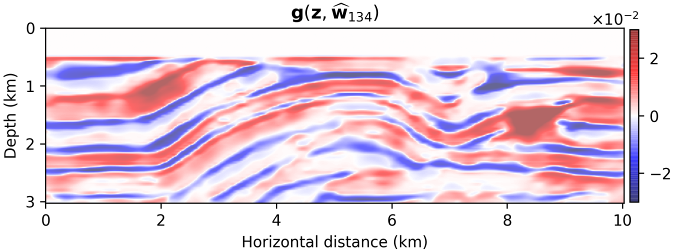
Samples from the posterior distribution

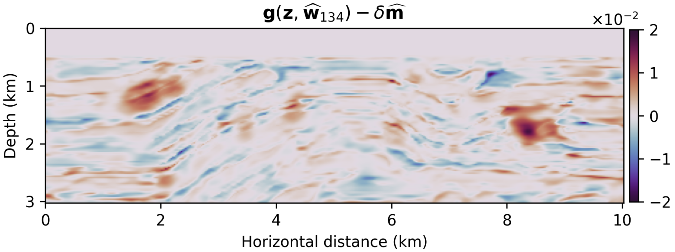


Samples from the posterior distribution

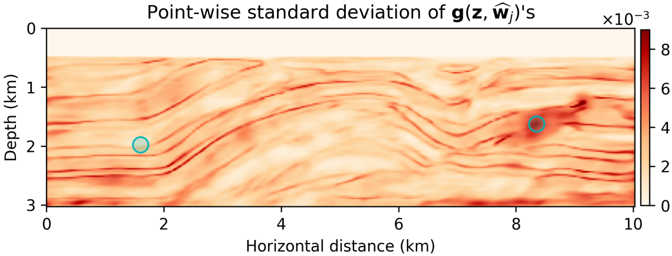


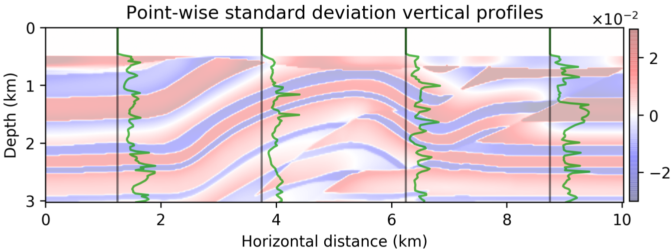
Samples from the posterior distribution



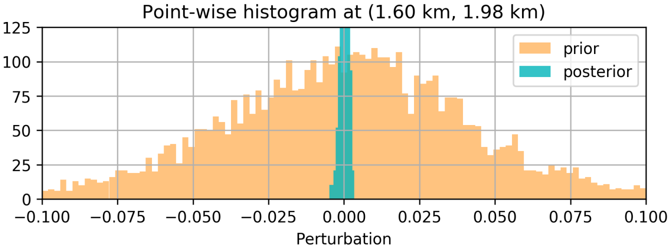


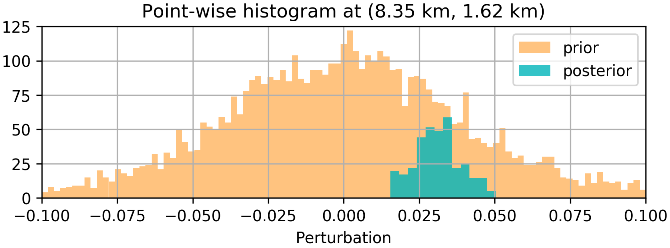
Samples from the posterior distribution





Imaging and uncertainty quantification





Observations and conclusions

- Utilized a structured prior induced by a carefully designed CNN
- Capable of sampling the posterior by running SGLD, albeit being expensive
- Jointly captured uncertainty in the imaging and the reparametrization with a CNN

Observations and conclusions

- Deep prior was partially able to circumvent the imaging artifacts
- The conditional mean demonstrated less artifacts
- Pointwise standard variation coincided with regions that are more difficult to image

Paper and code

Paper: https://arxiv.org/pdf/2001.04567.pdf

Code: https://github.com/alisiahkoohi/seismic-imaging-with-SGLD