Fast Approximations of High-Rank Hessians: Applications to Seismic Inversion and Uncertainty Quantification

Mathew Hu, Nick Alger, Omar Ghattas, Rami Nammour

IMAGE24 - W-7: Exposing our Errors

Houston, TX

August 30, 2024

The University of Texas at Austin
Oden Institute for Computational Engineering and Sciences

OPTIMUS
BEYOND FORWARD SIMULATION

TotalEnergies

Introduction

Seismic Inversion $\longrightarrow$ Uncertainty Quantification

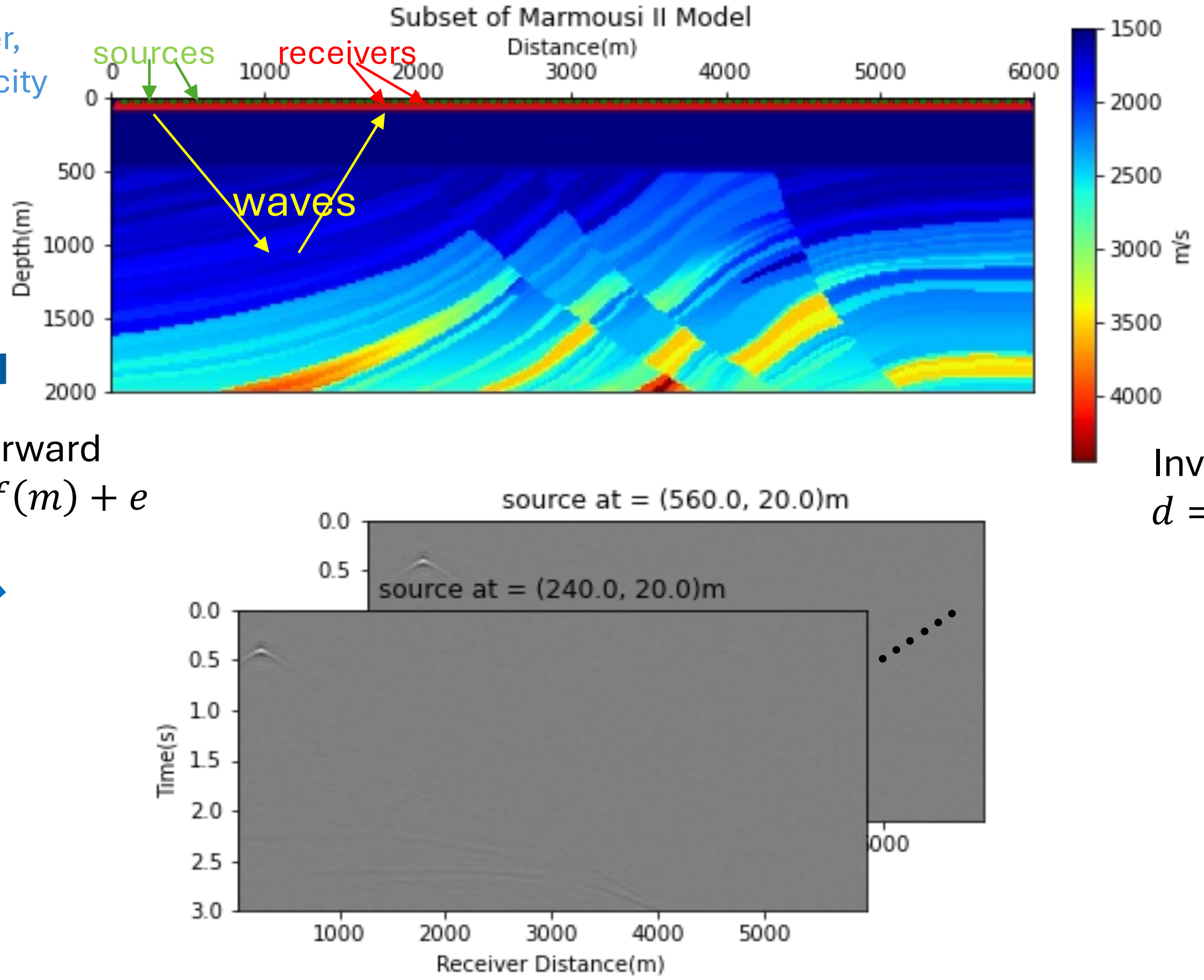
Hessian

Seismic Inversion

Model Parameter,
e.g. P wave velocity

m

Forward
$d = f(m) + e$



Inverse
$d \Longrightarrow m$

d

Data, e.g.
seismogram

Seismic Inversion

- Full-waveform inversion:
$$\min_m \Phi(m) := \| f(m) - d \|_2^2 + \mathcal{R}(m)$$
- Quasi-Newton methods (Newton-CG, L-BFGS) require many iterations (wave simulations).
- Need a preconditioner for the Hessian of Φ:
$$\mathcal{H} = \frac{d^2}{dm^2} \Phi(m)$$

Uncertainty Quantification

- Bayesian inference framework:
$$d\mu_{post}(m|d) \propto \pi_{like}(d|m) d\mu_{prior}(m)$$
- Information extraction is challenging.
- The Hessian contains crucial local curvature information to accelerate the Monte Carlo sampling.
$$\mathcal{H} = -\frac{d^2}{dm^2} \log d\mu_{post}(m|d)$$

Hessian Operator

- The whole Hessian matrix is intractable for large-scale problems.
- Hessian matrix-vector multiplication is available via the Adjoint Method ($\approx$4 Forward computations).
- Conventional low-rank methods are not suitable for seismic Hessian which is high-rank (with informative data from multiple sources and receivers).

Research Questions

1. How to approximate the seismic Hessian which is high-rank?

2. How to utilize the approximation to accelerate the seismic inversion?

3. How to utilize the approximation to accelerate seismic uncertainty quantification and hence make it tractable?

Hessian Approximation Methods

ΨDO Probing Method + PSF Method $\Longrightarrow$ Low-rank Symbol Method

Pseudo-differential Operator (ΨDO) Probing

$$(Hv)(x) = \int e^{ix\cdot\xi} s(x,\xi)\hat{v}(\xi)d\xi, \qquad v \in L^2(\mathbb{R}^d)$$

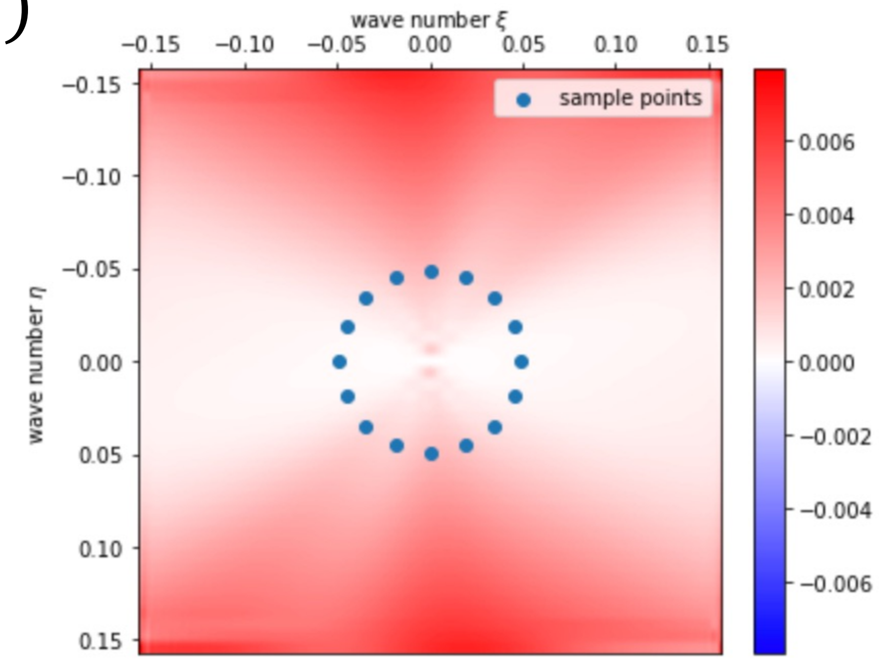
- Symbol $s(x,\xi)$: $\mathbb{R}^d \times \mathbb{R}^d \backslash \{0\} \to \mathbb{R}$ (smooth in x, ξ, asymptotically polynomial in ξ)

$$s(x,\xi) = e^{-ix\cdot\xi}\left(He^{ix\cdot\xi}\right)(x)$$

Bao and Symes (1996) use the symbol's asymptotic behavior to apply it in $\mathcal{O}\left(N^d \log N\right)$ vs $\mathcal{O}\left(N^{d+1}\right)$ for simulation:

$$s(x,\xi) \approx |\xi| s_1(x, \arg\xi) \approx |\xi| \sum_{k,l} c_k(x) e^{il \arg\xi}$$

⇒ Only need to compute symbol s on angles.



ΨDO in One Hessian Action

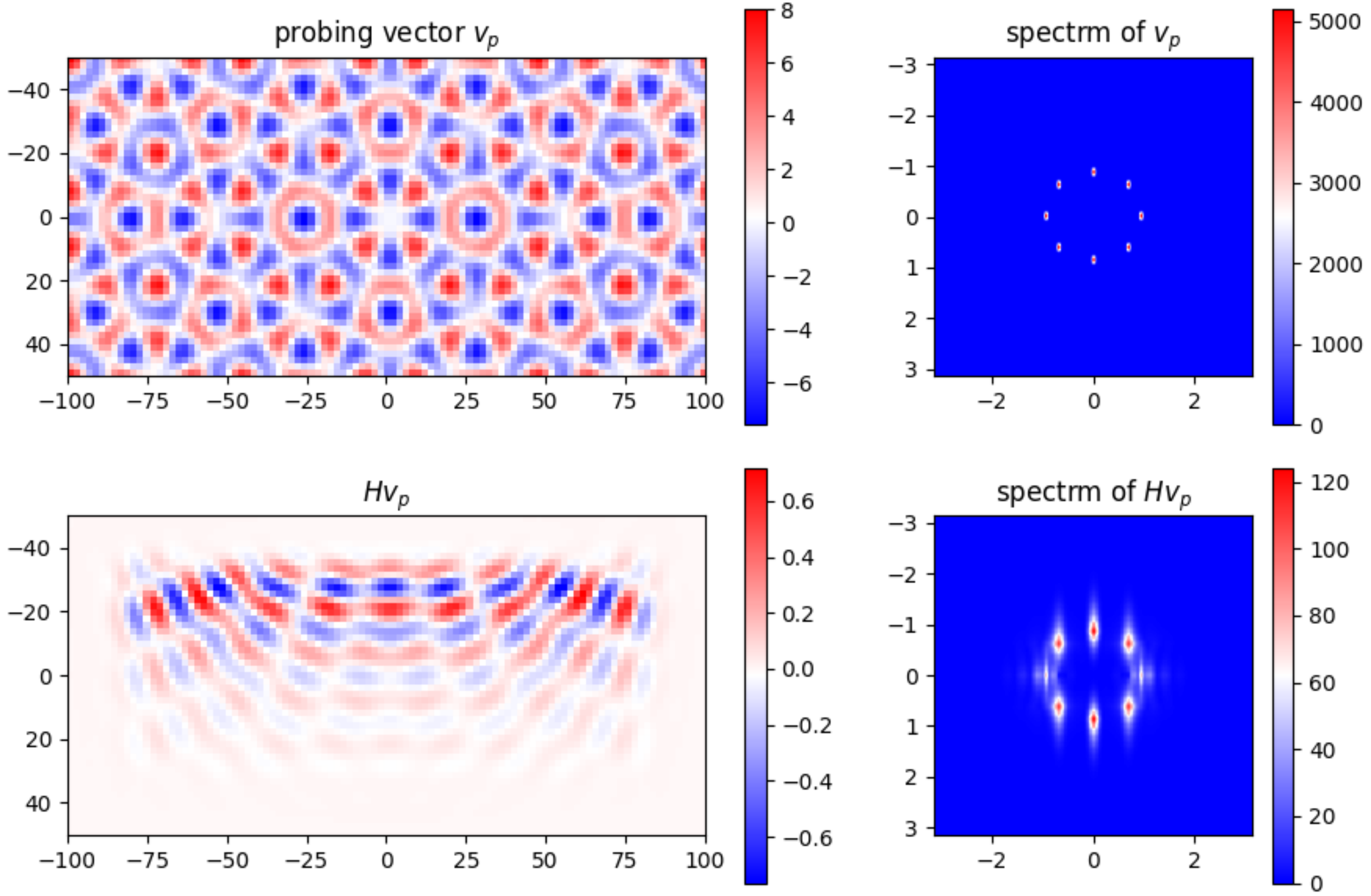
Pick sample points ξ_k in different directions.

- Naive scheme: Compute symbol columns $s(x,\xi_1)$, $s(x,\xi_2)$, …, $s(x,\xi_n)$ individually (n Hessian actions).
- Advanced scheme: The angles are preserved in frequency space $\Rightarrow$ Compute on all angles together (one Hessian action).

$$v(x) = \sum_{k=1}^{n} e^{ix\cdot\xi_k} \Rightarrow (Hu)(x) = \sum_{k=1}^{n} \left(He^{ix\cdot\xi_k}\right)(x) = \sum_{k=1}^{n} s(x,\xi_k)e^{ix\cdot\xi_k}$$

Then extract each symbol column $s(x,\xi_k)$ from the product Hv.

ΨDO in One Hessian Action (Example)

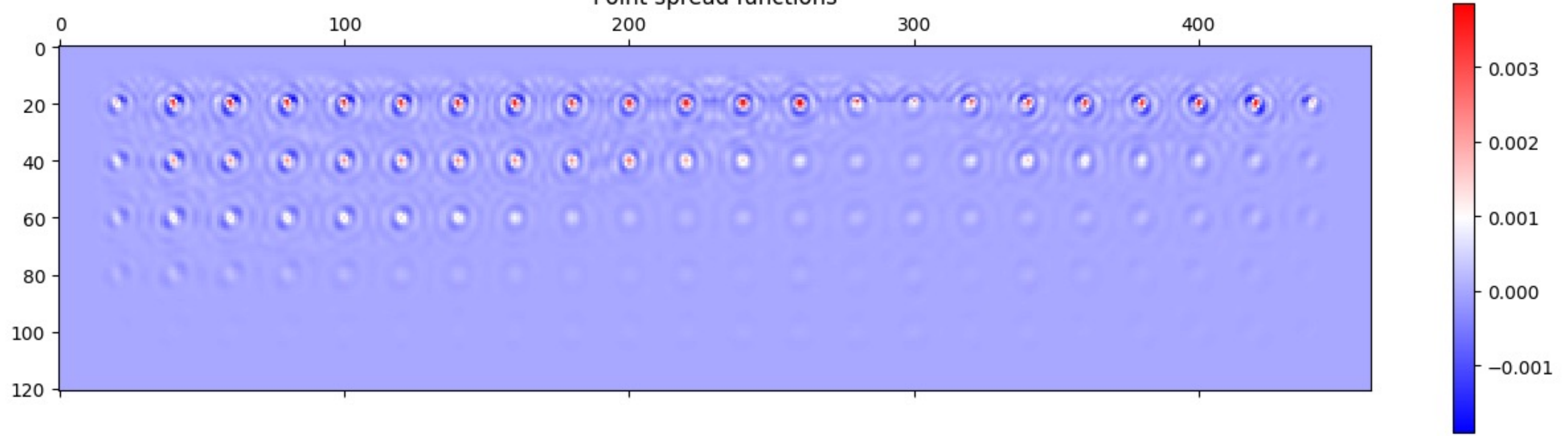


Point Spread Function Method (PSF)

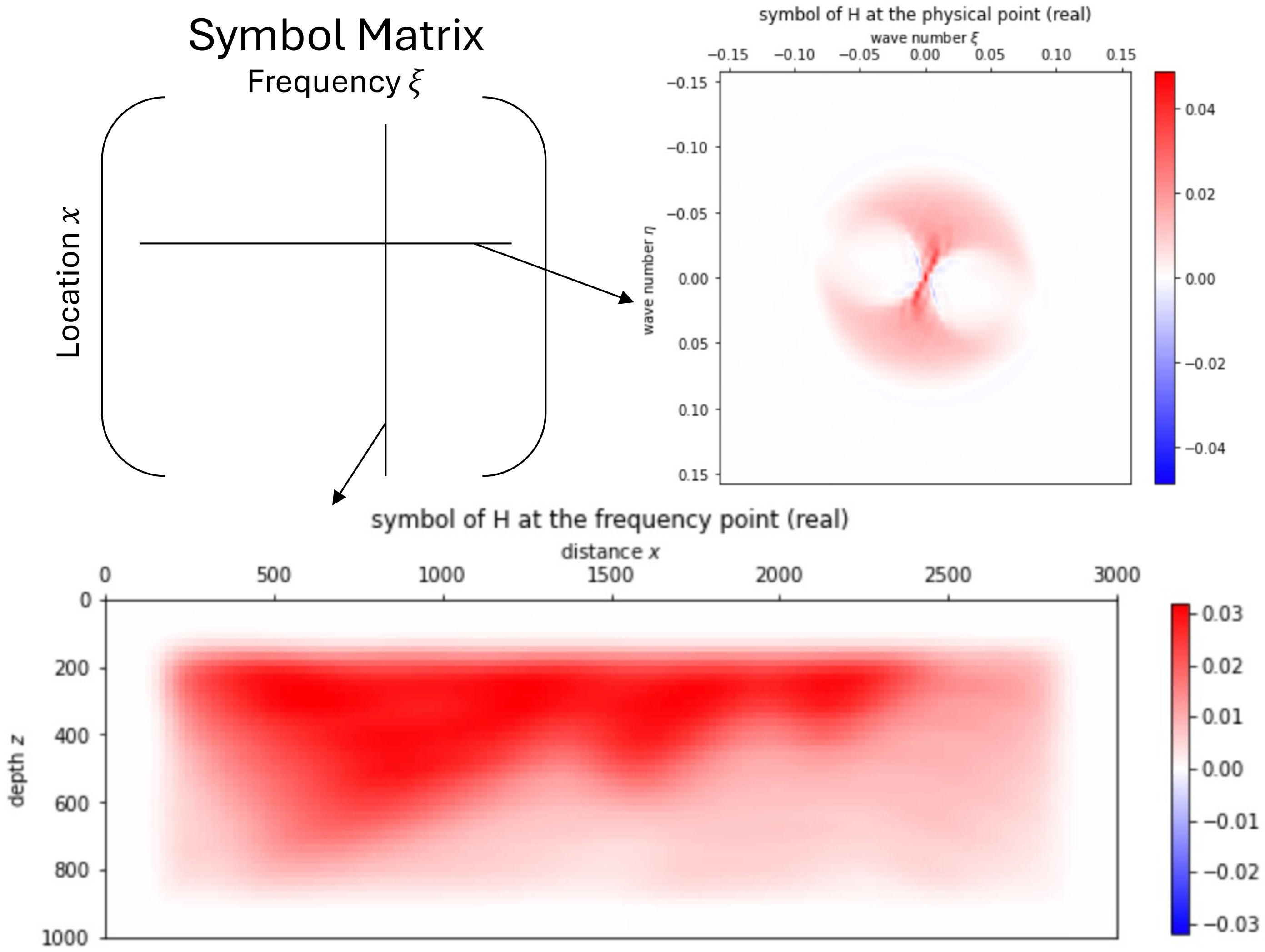
- PSFs are local ⇒ Compute multiple PSFs in one Hessian action.
- PSFs are translation invariant ⇒ Approximate PSFs with nearby PSFs.

⇒ Approximate H as a sum of Product-Convolutions.

$$Hu = \sum_{i=1}^{r} \underbrace{\phi_{x_i}}_{\text{sampled PSFs}} * (\underbrace{w_i}_{\text{RBF-weighting}} \cdot u)$$



Observation: Duality between PSF and ΨDO



PSF Method: Row of symbol

$$s(x_k,\cdot) = \mathcal{F}\big[\underbrace{H^T(\delta_x)(x_k - x)}_{\text{PSFs}}\big]$$

ΨDO Method: Cols of symbol

$$s(\cdot,\xi_k) = \frac{H\left(\mathcal{F}^{-1}\left[\delta_{\xi_k}\right]\right)(x)}{\mathcal{F}^{-1}\left[\delta_{\xi_k}\right](x)}$$

ΨDO + PSF = Low-Rank Symbol Method

- Both PSF and ΨDO method utilize the fact that the symbol of the seismic Hessian (at a smooth background) is low-rank.

XR Algorithm:

1. Convert the PSF method's result to low-rank symbol: $S_{psf} = X_0 Q$
2. Refine X_0 to fit the symbol columns $S[:, I_C]$ from the ΨDO method:
$$X = \underset{X}{\operatorname{argmin}} \ |XQ[:, I_c] - S[:, I_C]|_F^2 + \alpha |X - X_0|_F^2$$
3. Create Hessian approximation from the low-rank symbol: $S_{new} = XQ$

Accelerate Seismic Inversion

$$m^* = \underset{m}{\operatorname{argmin}}\, \Phi(m) := \| f(m) - d \|_2^2 + \mathcal{R}(m)$$

Preconditioner $P = \left(\widetilde{H} + R\right)^{-1}$ applies to L-FBGS algorithm:

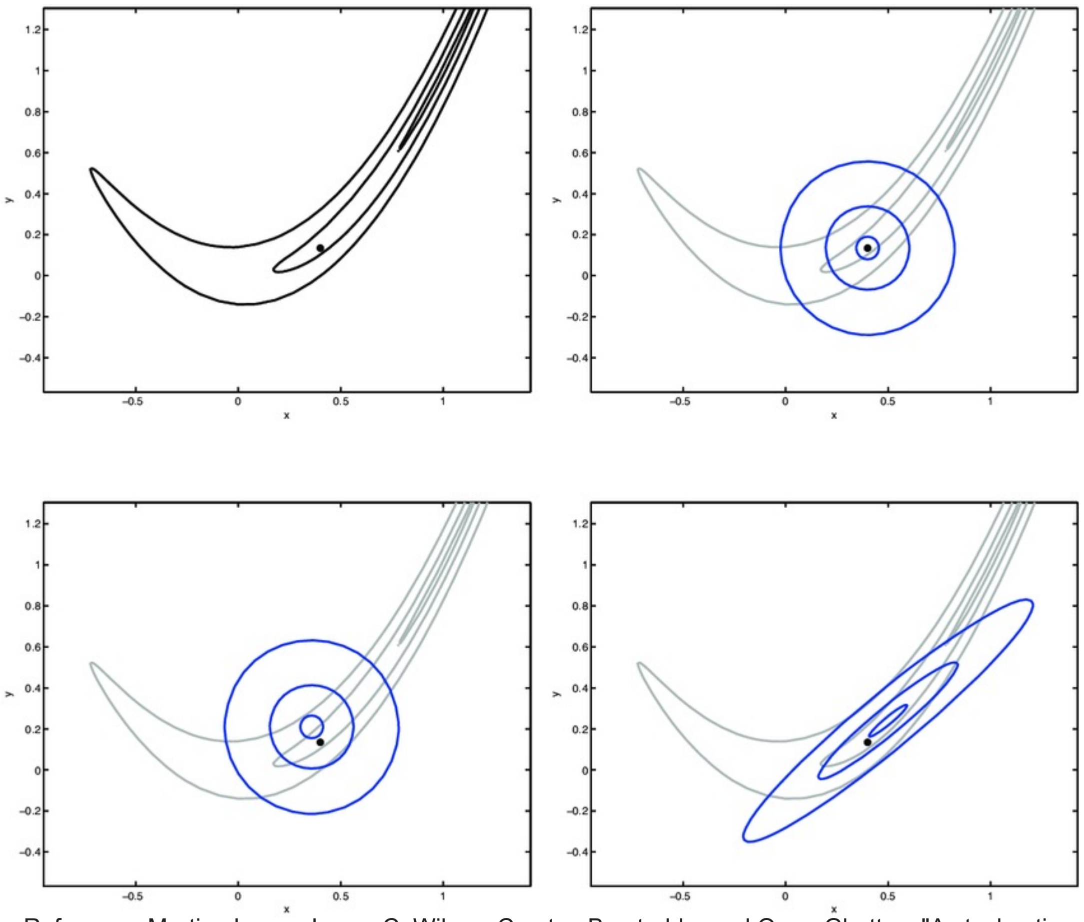
Algorithm 1 Preconditioned L-BFGS

Choose starting point x_0, and integer $m > 0$
$k \leftarrow 0$ ▷ Initialization
repeat ▷ Newton Loop
  Choose $\boxed{H_k^0 = \gamma_k' P}$
  Compute $p_k = -H_k \nabla f_k$ via two-loop recursion.
  Compute α_k which satisfies Wolf conditions ▷ Line search
  $x_{k+1} \leftarrow x_k + \alpha_k p_k$
  if $k > m$ **then**
    Discard vector pair $\{s_{k-m}, y_{k-m}\}$ from storage
  end if
  Compute and save $s_k = x_{k+1} - x_k$ and $y_k = g_{k+1} - g_k$
  $k \leftarrow k + 1$
until convergence

Accelerate Uncertainty Quantification

- Build a Laplace Approximation to the posterior:
$$\mathcal{N}\left(m^*, (\tilde{H} + R)^{-1}\right)$$
- Draw samples from this distribution.
- These samples are more likely to be accepted.
- Implemented by the generalized preconditioned Crank-Nicolson algorithm (gpCN).

Visualizations of different types of proposal distributions for MCMC



Reference: Martin, James, Lucas C. Wilcox, Carsten Burstedde, and Omar Ghattas. "A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion." *SIAM Journal on Scientific Computing* 34, no. 3 (2012): A1460-A1487.

Numerical Results

Quadratic Toy Model

The Marmousi Model

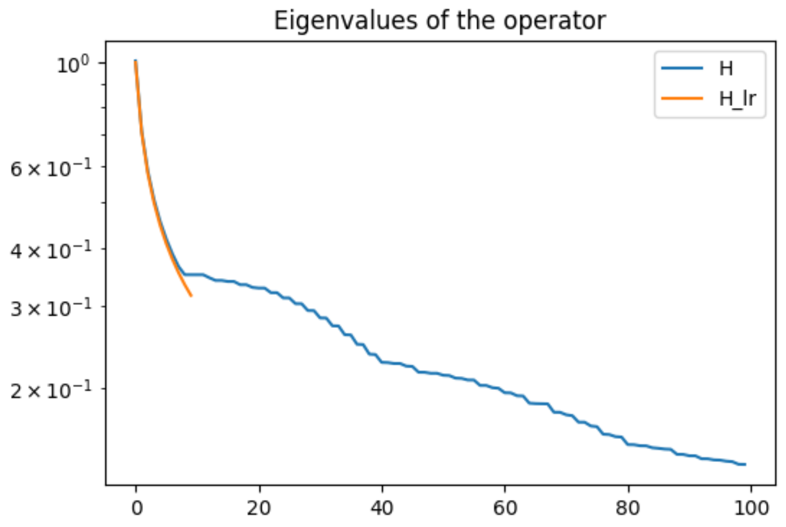
Quadratic Toy Model

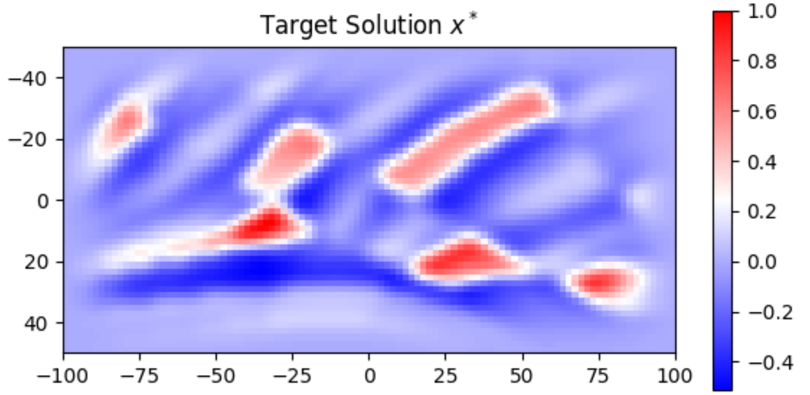
- Hessian = Low-rank matrix + ΨDO
- True model x^*. Right hand side $b = Hx^*$.
- Suppose $x = Wy$. (W suppresses boundaries).

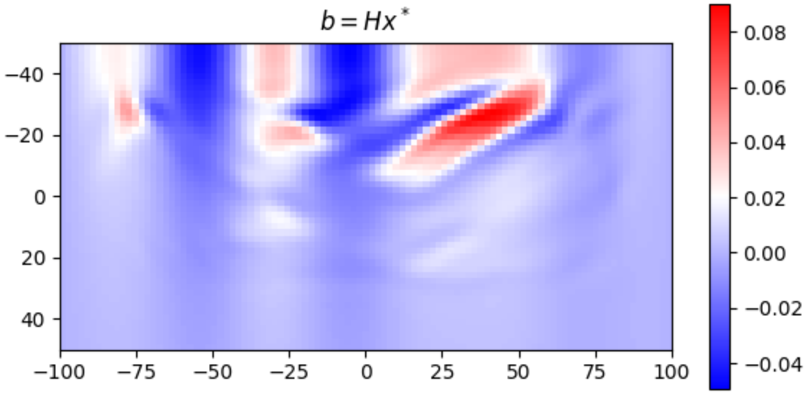
$$\begin{cases} cost(x) = \frac{1}{2}x^T Hx - b^T x \\ grad(x) = Hx - b \\ hess(x) = H \end{cases} \qquad \begin{cases} cost(y) = \frac{1}{2}y^T WHWy - b^T Wy \\ grad(y) = WHWy - Wb \\ hess(y) = WHW \end{cases}$$

Model	Quadratic Model
Dimensions	100m × 50m
Grids	101 × 51 = 5151

- The toy model is linear
- The posterior is Gaussian.



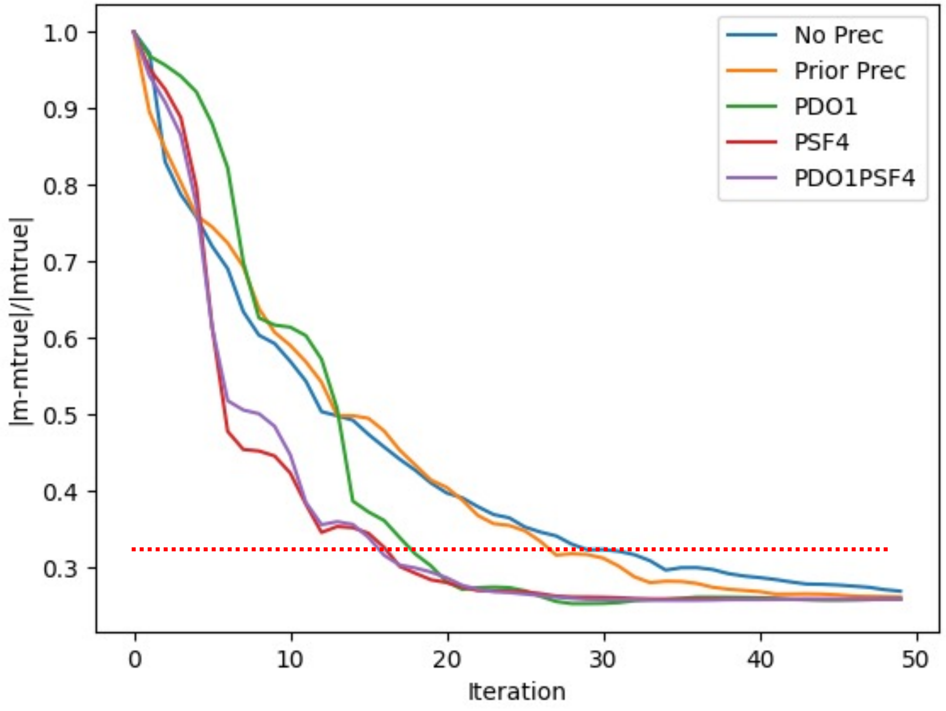




Inversion Solution: Preconditioned L-BFGS

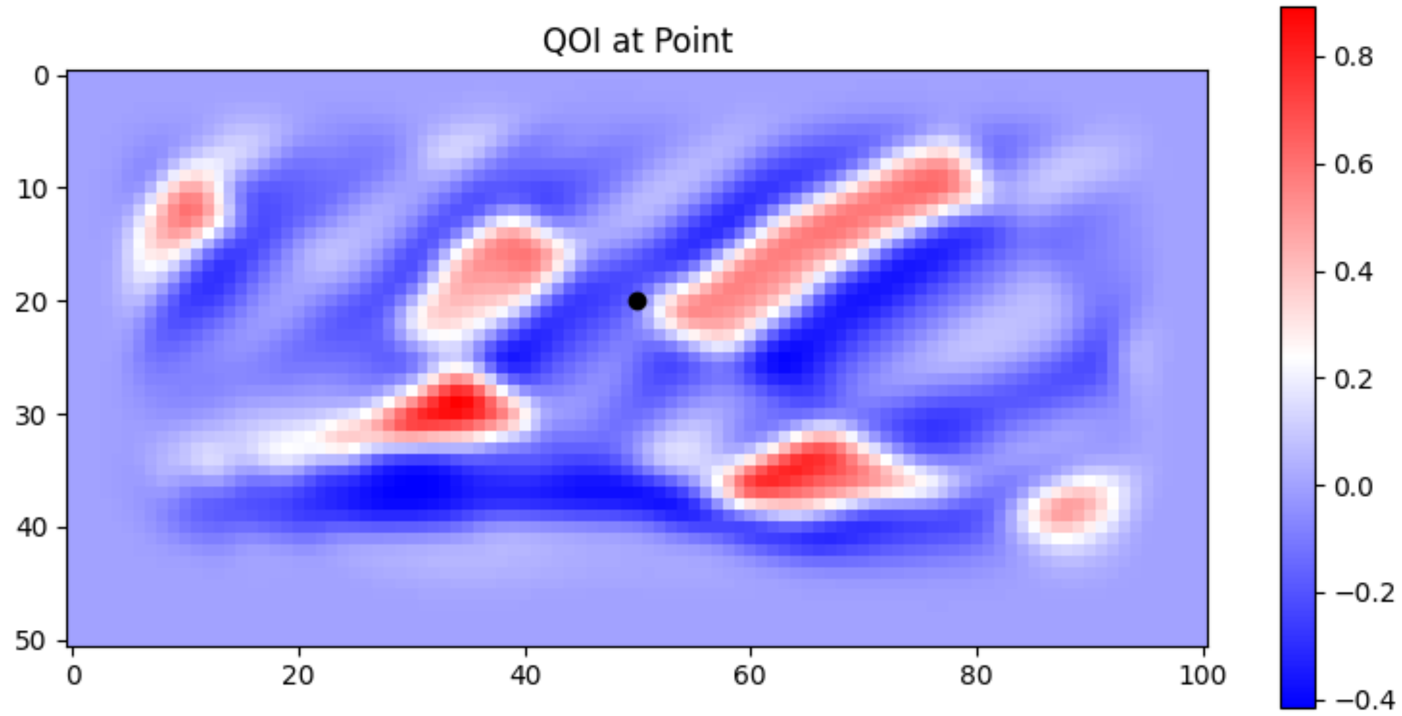
- Decay slower at beginning.
  - Our methods do not address the low-rank part.
- Decay faster later.

Preconditioner	L-BFGS Steps
No Preconditioner	169
Prior-Preconditioner	108
PDO1	**92**
PSF4	**55**
PDO1+PSF4	**69**



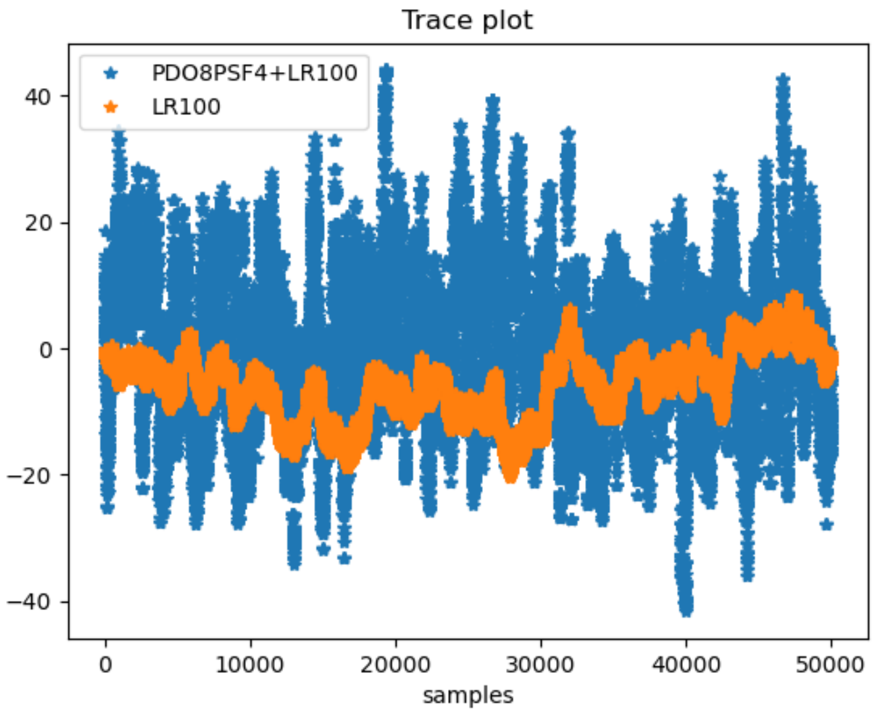
Uncertainty Quantification: MCMC results

- We draw 50k samples from the posterior via MCMC-gpCN.
- Quantity of Interest (QoI) = The value at a certain position

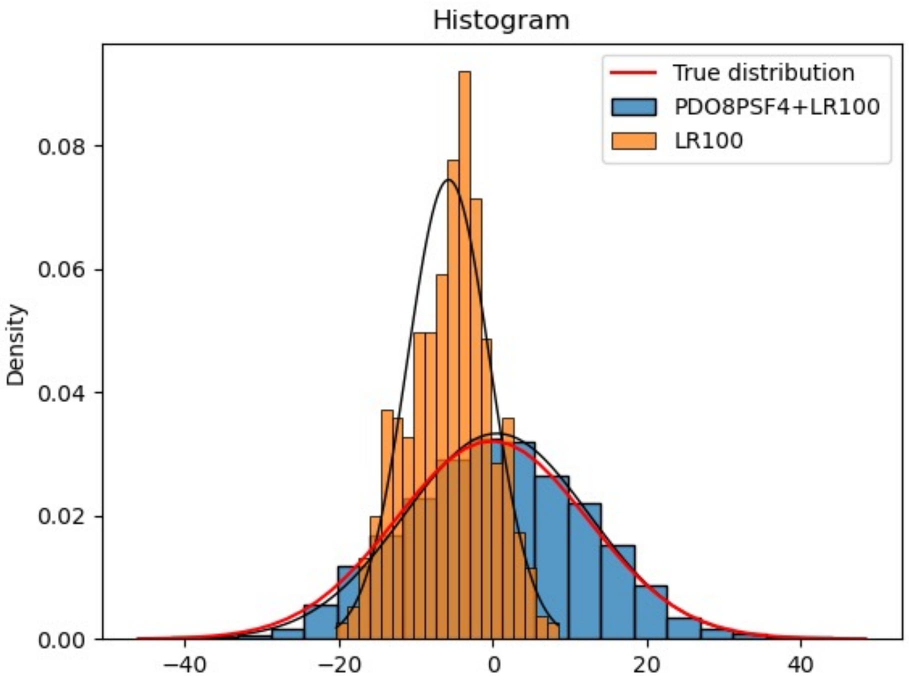


Uncertainty Quantification: MCMC results

Our methods draw samples with more mixture.



Our methods derive nearly the true distribution

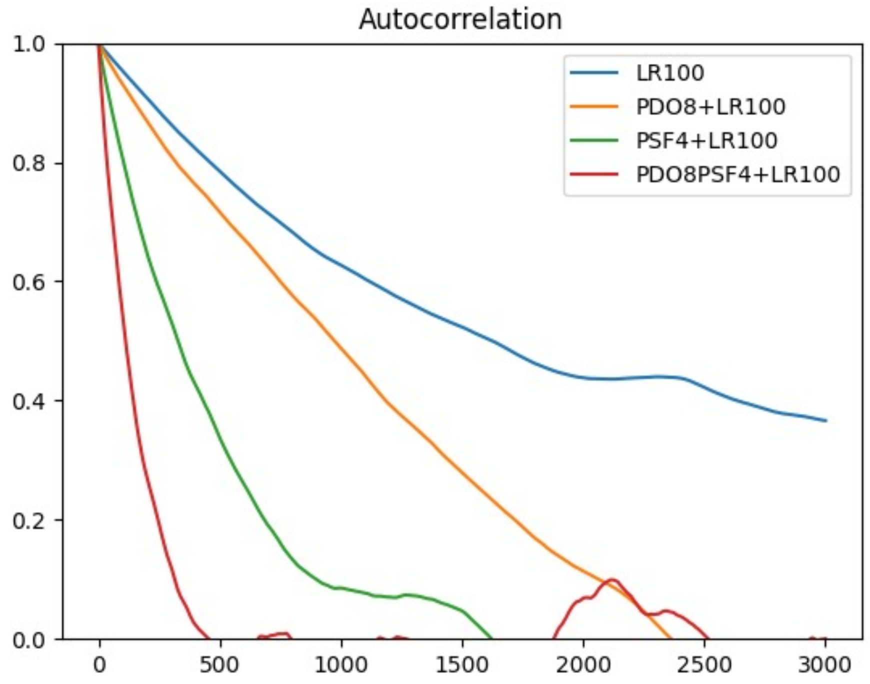


Uncertainty Quantification: MCMC results

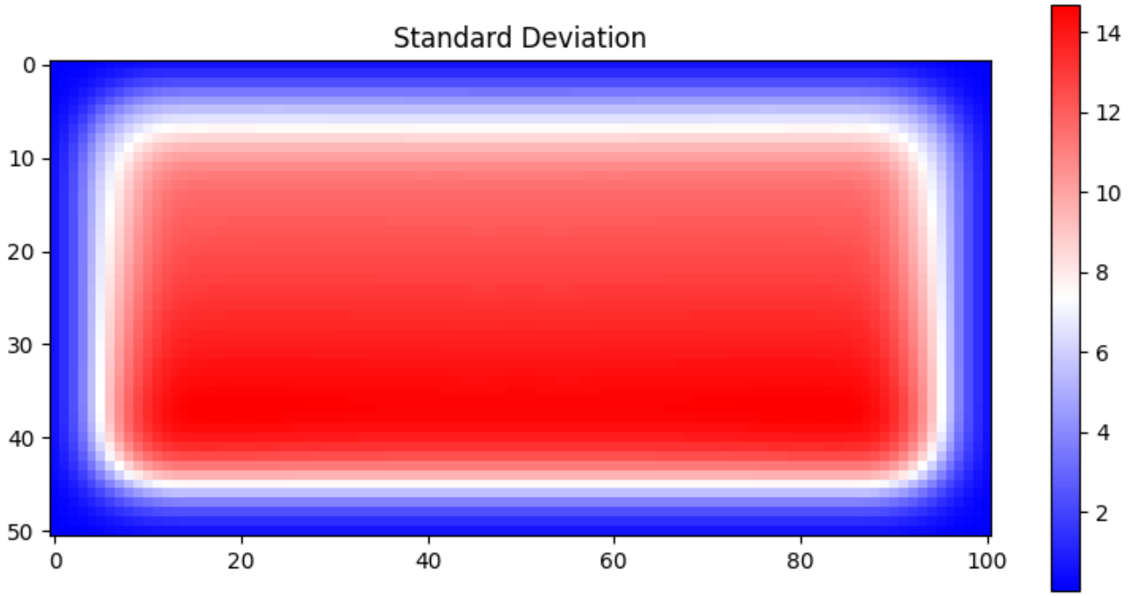
Our methods achieve larger effective samples size (ESS)

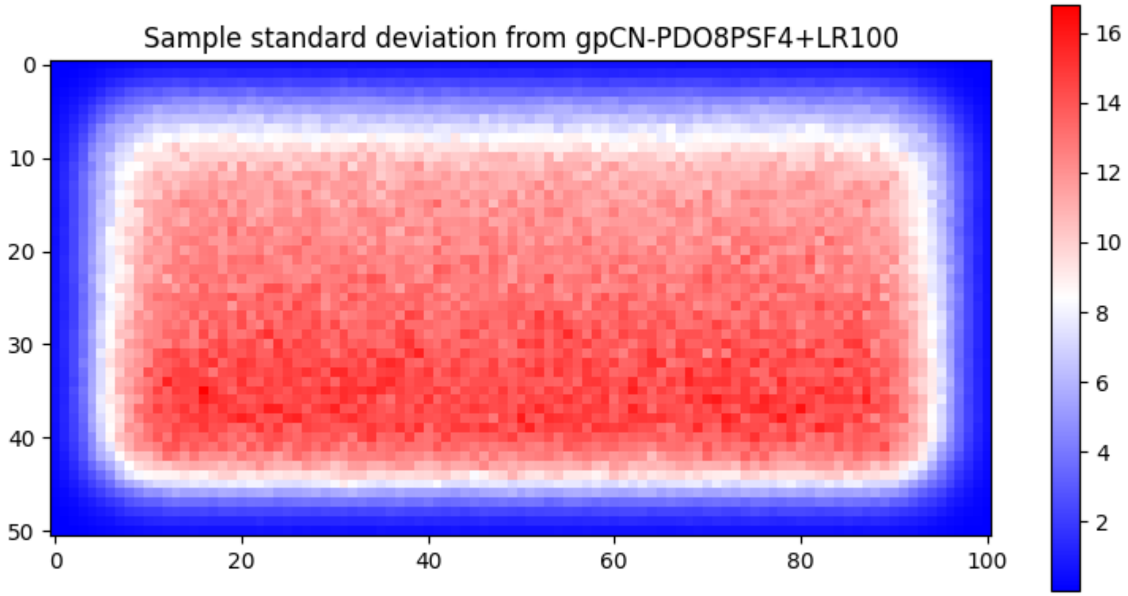
Sampling Method	Computational Cost	ESS
gpCN-LR100	200Hess+50kFwd	7.89
gpCN-PDO8+LR100	208Hess+50kFwd	24.20
gpCN-PSF4+LR100	204Hess+50kFwd	57.48
gpCN-PDO8PSF4+LR100	212Hess+50kFwd	**132.66**

Our methods draw samples with faster decaying autocorrelation.



Uncertainty Quantification: MCMC results

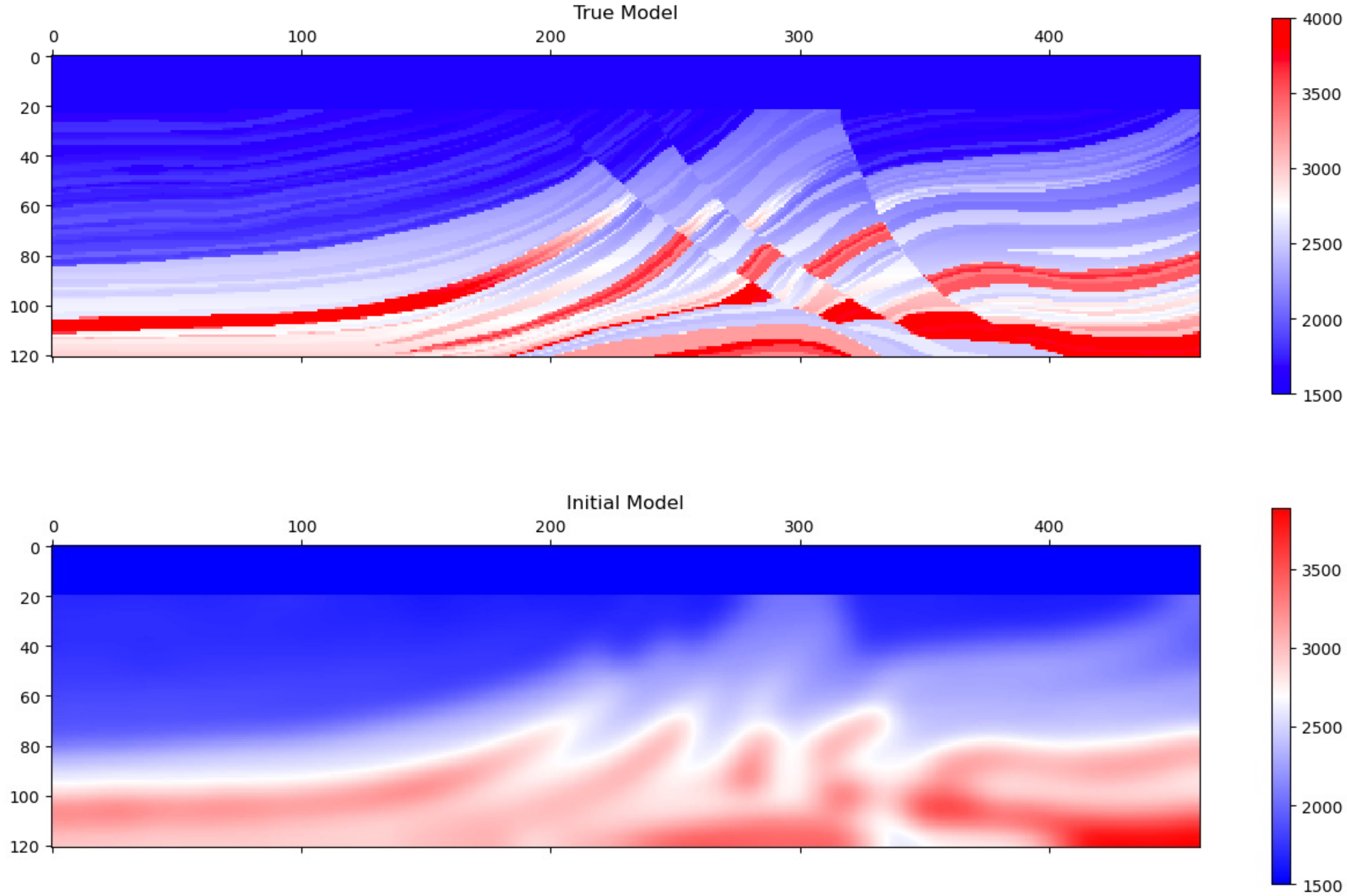




Marmousi Model

$$\Phi(x) := \| f(m_0 + Wx) - d \|_2^2 + \mathcal{R}(m)$$

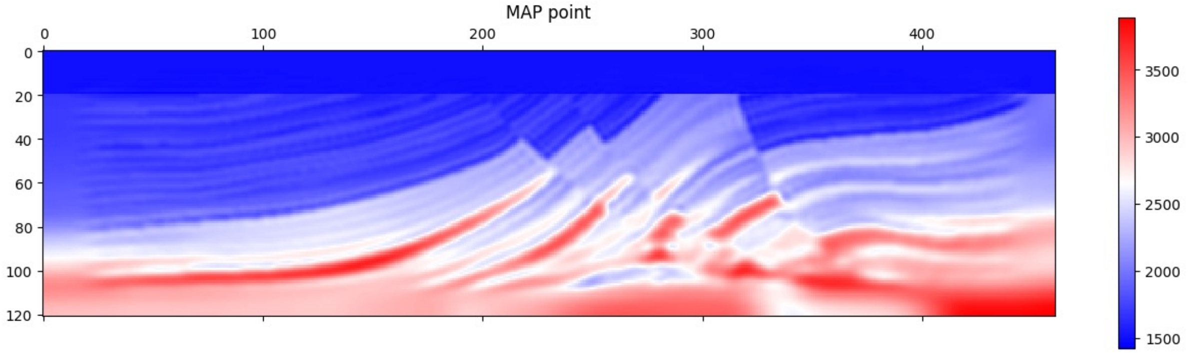
Model	Marmousi + water layer
Dimensions	9200m × 2400m
Grids	461 × 121 = 55781
Grid Size	20m × 20m
Simulation Time T	8s
Time Interval dt	0.0025s
Number of Sources N_s	92
Number of Receivers N_r	153
Source wave frequency	4-12Hz

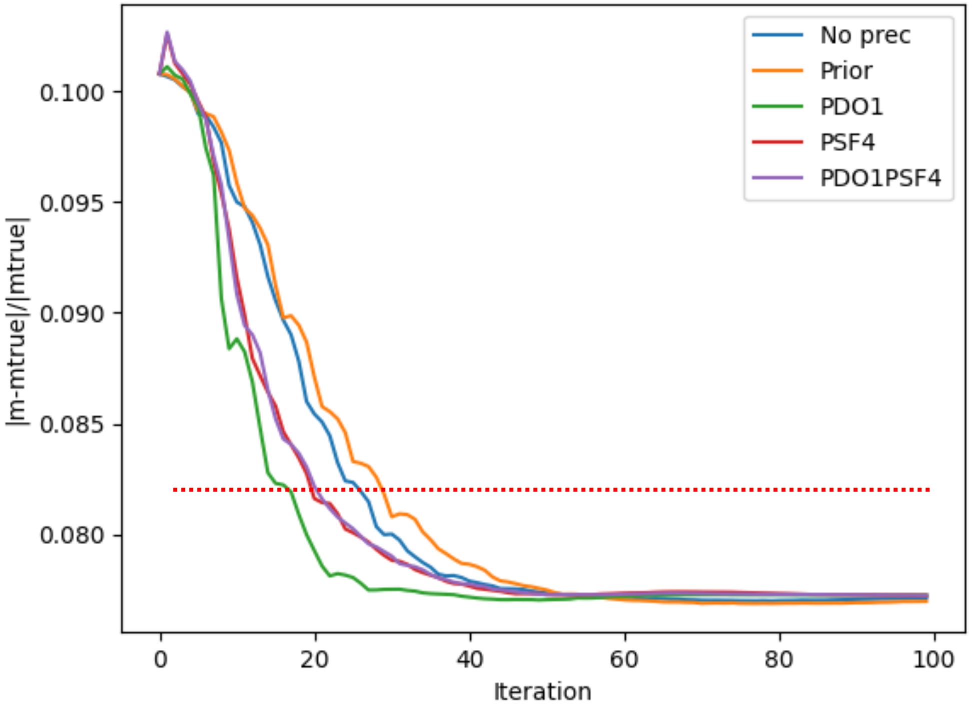


Inversion Solution: Preconditioned L-BFGS

- Our methods achieve faster decaying solution errors.

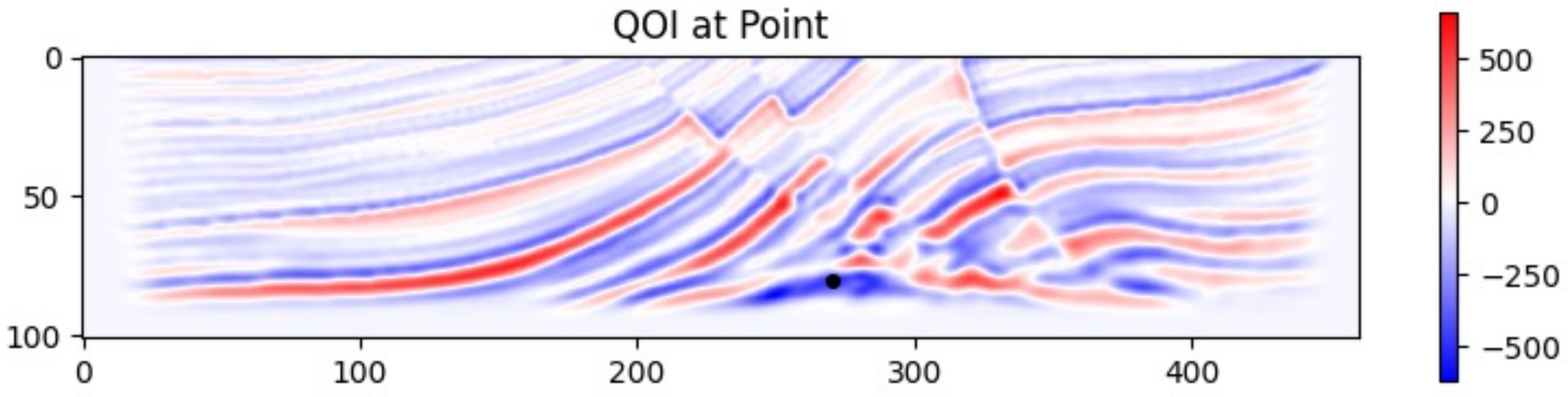
Preconditioner	L-BFGS Steps
No Preconditioner	333
Prior-Preconditioner	465
PDO1	**267**
PSF4	**231**
PDO1+PSF4	**238**





Uncertainty Quantification: MCMC results

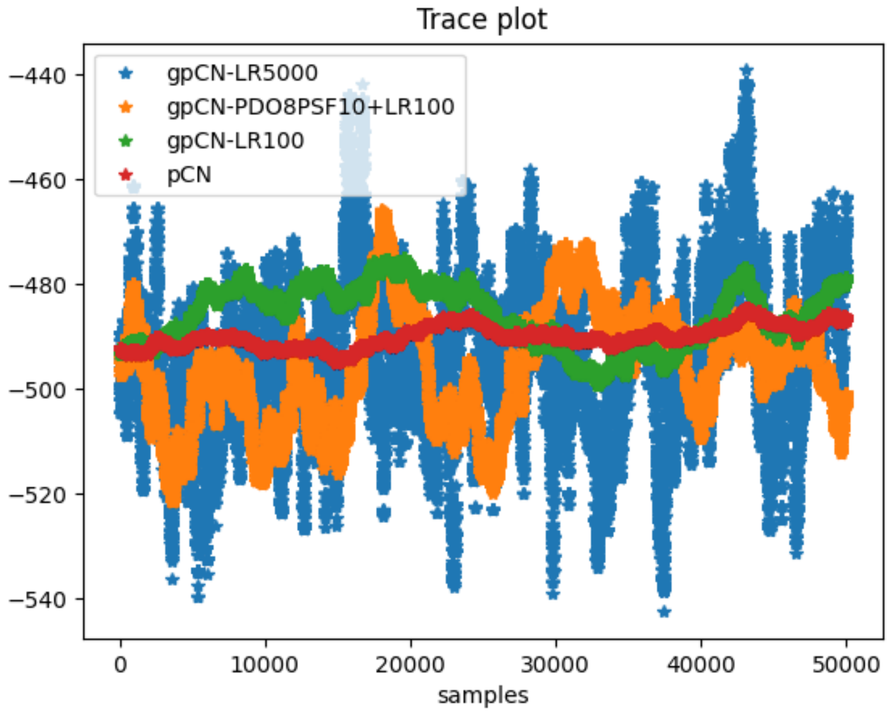
- We draw 50k samples from the posterior via MCMC-pCN/gpCN.
- Quantity of Interest (QoI) = The value $m - m_0$ at a certain position



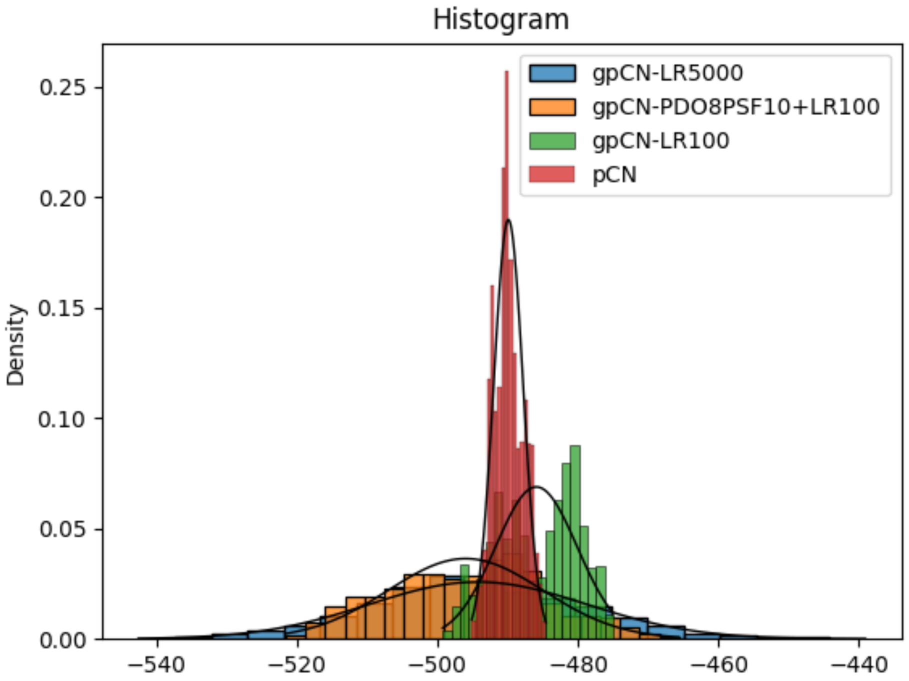
- The Marmousi model is non-linear (FWI). The correct UQ is unavailable.
- We benchmark against the low-rank approximation with 5000 eigenvalues.

Uncertainty Quantification: MCMC results

Our methods draw samples with more mixture.



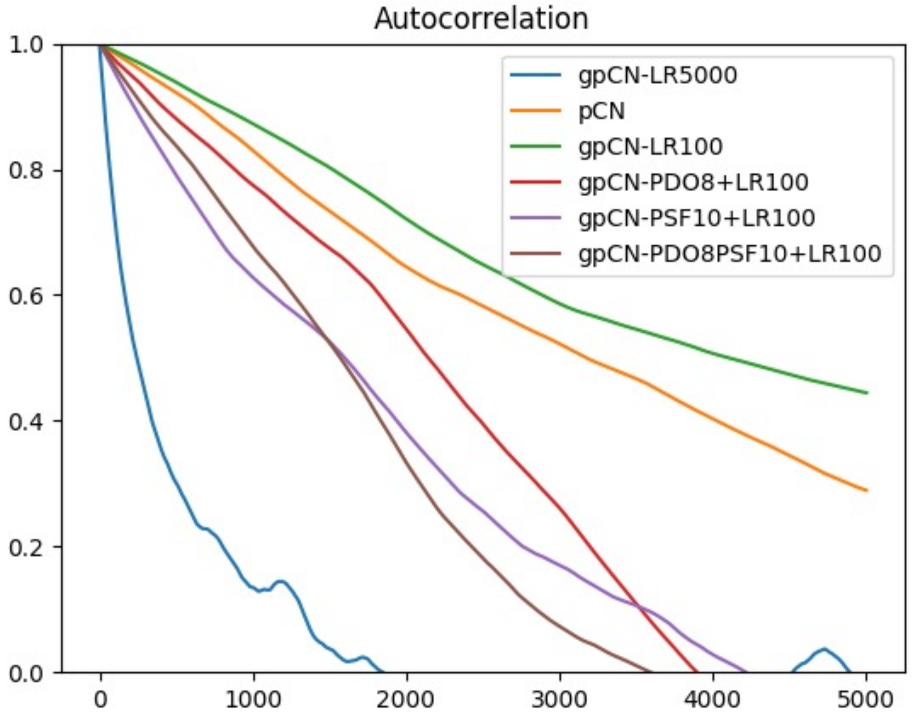
Our methods derive a distribution that is closer to the baseline.



pCN & gpCN-LR100 are trapped in a small region ⇒ Under-estimated variation ⇒ non-interpretable UQ

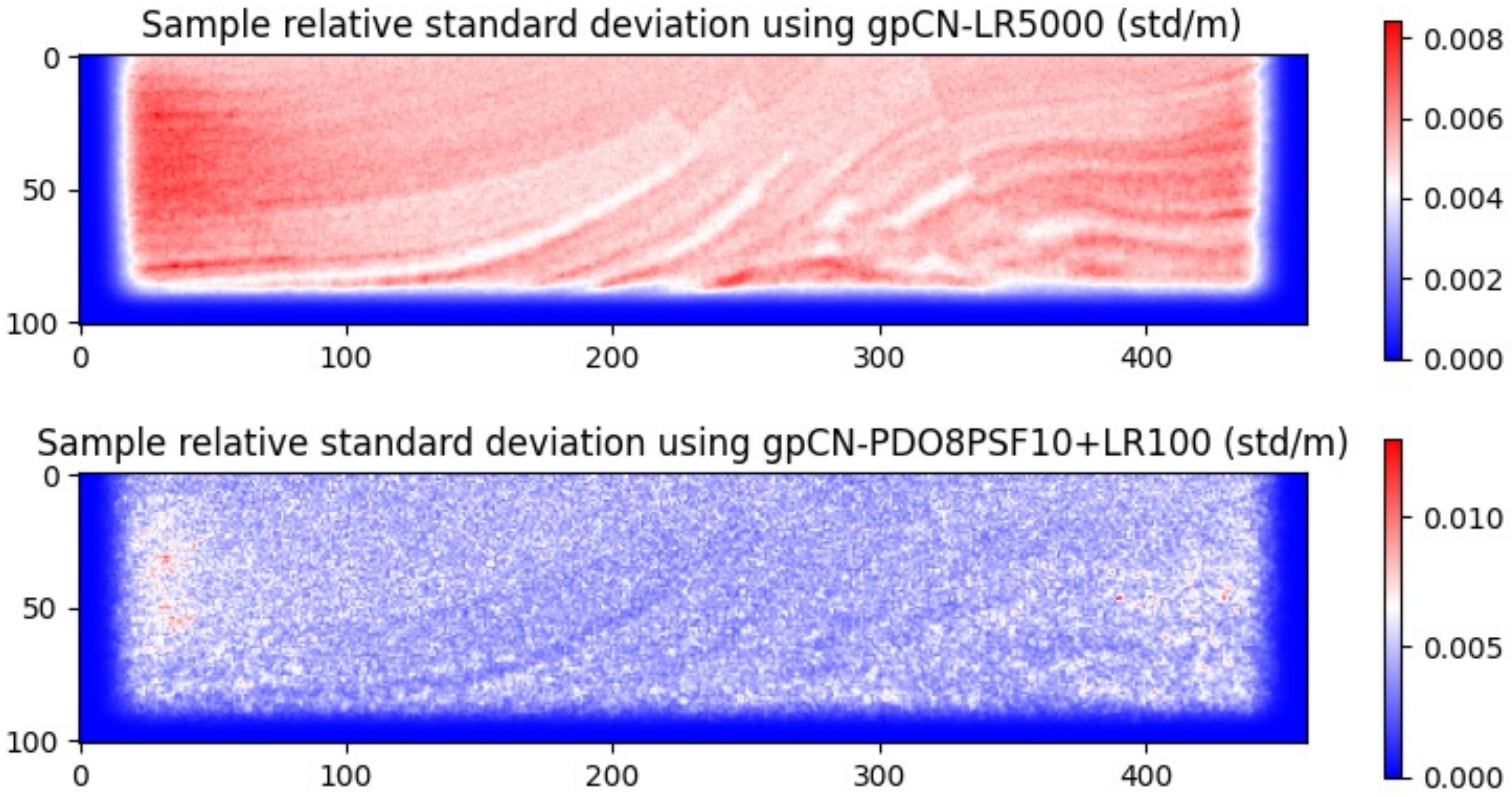
Uncertainty Quantification: MCMC results

- The autocorrelation seems to indicate that pCN and gpCN-LR100 are well, but they under-estimate the uncertainty in fact.



Algorithm	Computational Cost	ESS
gpCN-LR5000 (Baseline)	10kHess+50kFwd	53.12
pCN	0Hess+50kFwd	7.03
gpCN-LR100	200Hess+50kFwd	6.11
gpCN-PDO8+LR100	208Hess+50kFwd	12.10
gpCN-PSF10+LR100	210Hess+50kFwd	12.32
gpCN-PDO8+PSF10+LR100	218Hess+50kFwd	**13.49**

Uncertainty Quantification: MCMC results



Contributions

- Develop novel Hessian approximation algorithms based pseudo-differential operators, which also applies to general hyperbolic problems.
- Accelerate the solution of deterministic seismic inversions via preconditioned quasi-Newton methods.
- Enable large-scale seismic inversion uncertainty quantification via Markov Monte-Carlo with generalized preconditioned Crank-Nicolson kernel.

Future Direction

- Investigate the trade-off between the cost of Hessian approximation and the cost of sampling.
- Test on other benchmark models (e.g. Salt Model).
- Compare with Variational Bayesian Inference.

Acknowledgement

- **TotalEnergies** for sponsoring this research and allowing me to present the results.
- **Dr. Omar Ghattas** and **Dr. Rami Nammour** for their advising and support.
- **Dr. Nick Alger**, **Dr. Stefan Henneking**, and **Dr. Lianghao Cao** for helpful discussions.

The University of Texas at Austin
Oden Institute for Computational Engineering and Sciences

OPTIMUS
BEYOND FORWARD SIMULATION

TotalEnergies

Thank you!

Connect me on LinkedIn!

The University of Texas at Austin
Oden Institute for Computational Engineering and Sciences

OPTIMUS
BEYOND FORWARD SIMULATION

TotalEnergies