Large Scale Randomized Learning Guided by Physical Laws with Applications in Full Waveform Inversion

Rui Xie, ***Fangyu Li****, Zengyan Wang, and WenZhan Song

University of Georgia

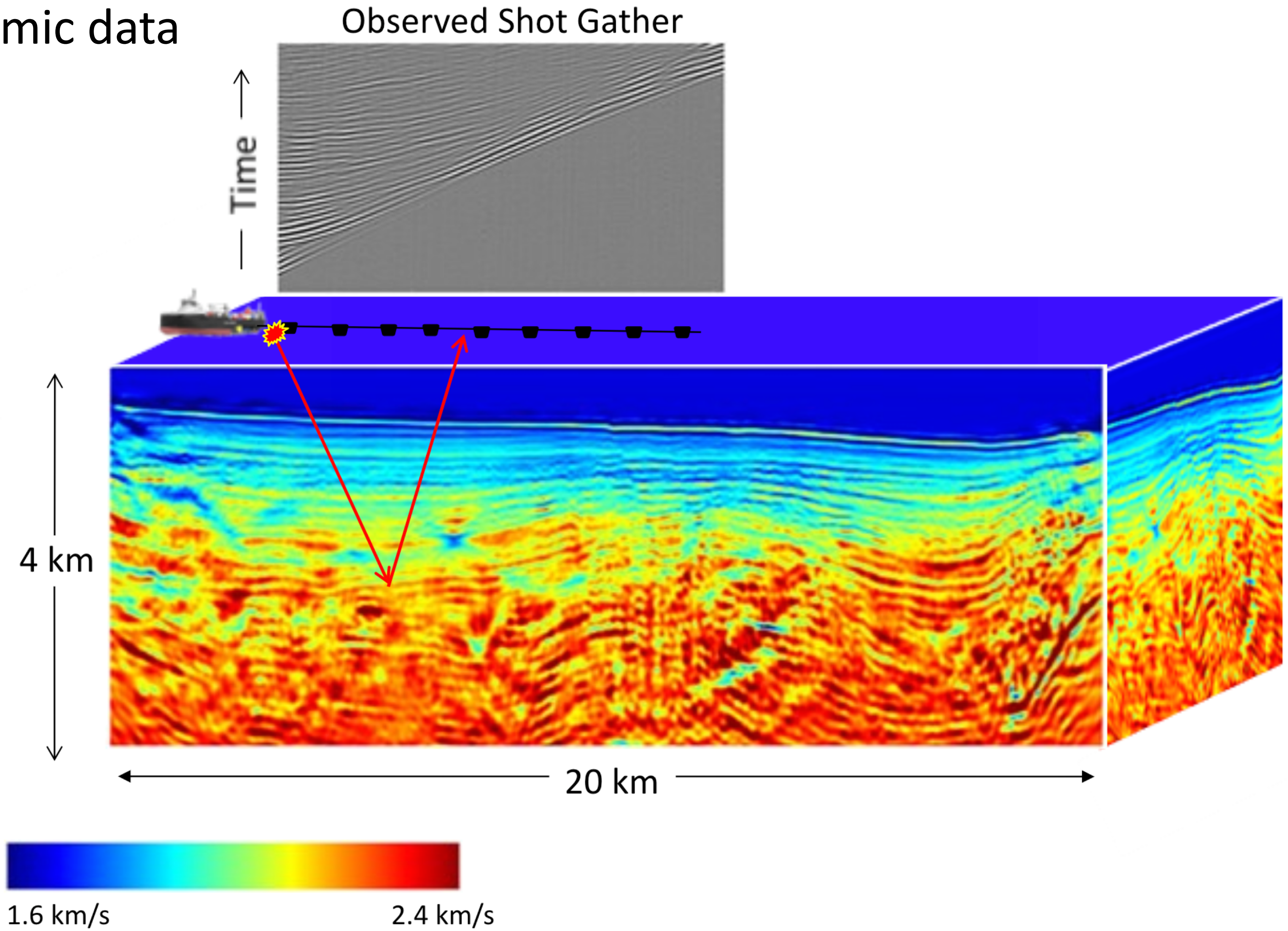
2018/11/27

FWI: Big data challenge

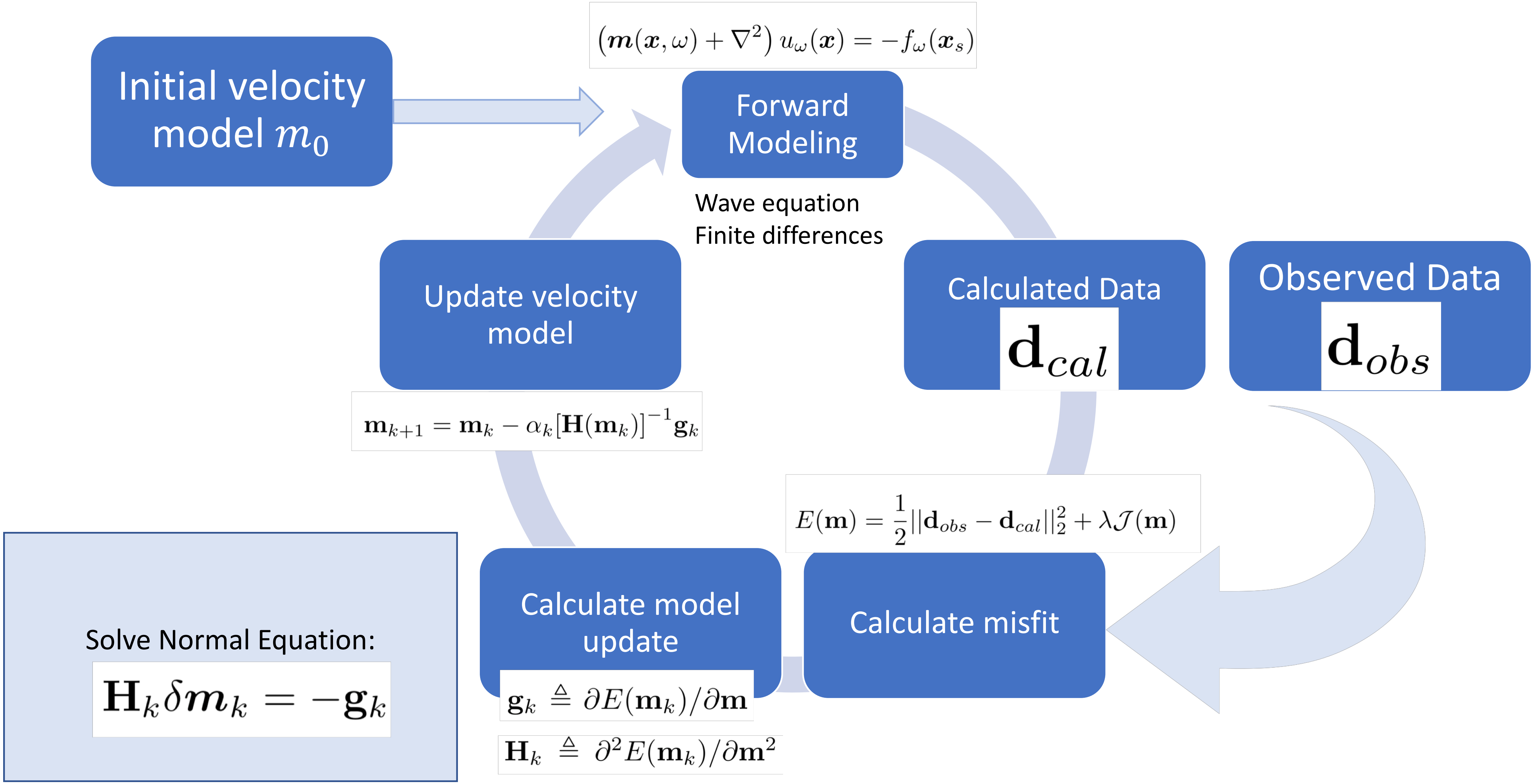
- Extract the most from seismograms (full data) with the best physics (wave equations) using the most efficient approach (efficient optimization).

- FWI is a data-driven imaging technique as billions of data points are involved in the computation.

Reflection seismic data



FWI: a data-driven inversion



FWI Consider the optimization problem

$$E(\mathbf{m}) = \frac{1}{2}||\mathbf{d}_{obs} - \mathbf{d}_{cal}||_2^2 + \lambda\mathcal{J}(\mathbf{m})$$

- Ill-conditioned problem

It is often the case that first-order methods return a solution far from the minimizer.

- Second-order methods

Most second-order algorithms prove to be more robust to ill conditioning. By using the curvature information, second-order methods properly rescale the gradient and enjoy fast local convergence.

Newton’s method

- Newton's method is an extremely powerful technique—in general the convergence is quadratic.
- The Newton's method is used to get the optimal model perturbation through the n $\delta \mathbf{m}_k$ equation:

$$\mathbf{H}_k \delta \boldsymbol{m}_k = -\mathbf{g}_k$$

and then update the velocity model according to

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k [\mathbf{H}(\mathbf{m}_k)]^{-1} \mathbf{g}_k$$

where α_k is the learning rate.

$$\mathbf{g}_k = \frac{\partial E(\mathbf{m}_k)}{\partial \mathbf{m}} = -\Re\left\{\left[\frac{\partial \mathcal{F}(\mathbf{m}_k)}{\partial \mathbf{m}}\right]^\dagger (\mathbf{d}_{obs} - \mathcal{F}(\mathbf{m}_k))\right\}$$

$$= \Re\left\{\mathbf{J}_k^\dagger \delta\mathbf{d}_k\right\},$$

where $\delta\mathbf{d}_k = \mathbf{d}_{obs} - \mathcal{F}(\mathbf{m}_k)$

Gradient

Hessian

$$\mathbf{H}_k = \frac{\partial^2 E(\mathbf{m}_k)}{\partial \mathbf{m}^2} = \Re\left\{\mathbf{J}_k^\dagger \mathbf{J}_k + \frac{\partial \mathbf{J}_k^\dagger}{\partial \boldsymbol{m}^T}[\delta \mathbf{d}_1^* \ \cdots \ \delta \mathbf{d}_k^*]\right\}$$

Number of parameters involved:
10^5 *to* 10^7 unknowns in 2D FWI

Realization of the Hessian at certain receivers and frequencies

$$\mathbf{H}(\mathbf{m}_k) = \sum_{s=1}^{N} \mathbf{A}_s^T(\mathbf{m}_k)\mathbf{A}_s(\mathbf{m}_k) + \mathbf{Q}(\mathbf{m}_k)$$

Forming the Hessian is expensive, which costs $O(N(N_z N_x)^2)$ and solving the normal equation costs $O\left((N_z N_x)^3\right)$.

Remedy

- Idea: Sub-sample only a few terms, say s, from $\sum_{s=1}^{N} \mathbf{A}_s^T(\mathbf{m}_k)\mathbf{A}_s(\mathbf{m}_k)$, without forming them, to form $\widetilde{\mathbf{H}}(\mathbf{m}_k)$ so that the cost can be reduced to $O(s(N_z N_x)^2)$.

- When $N_z N_x$ is large, use the iterative solver such as Conjugate Gradient to solve the normal equation.

Non-uniform sampling strategies

- leverage scores sampling
  - For $A \in \mathbb{R}^{n \times d}$, the i-th leverage scores of A is

$$\tau_i(\mathbf{A}) = \mathbf{a}_i^T (\mathbf{A}^T \mathbf{A})^\dagger \mathbf{a}_i$$

Sampling probability $\pi_i = \tau_i$

- row norm squares sampling

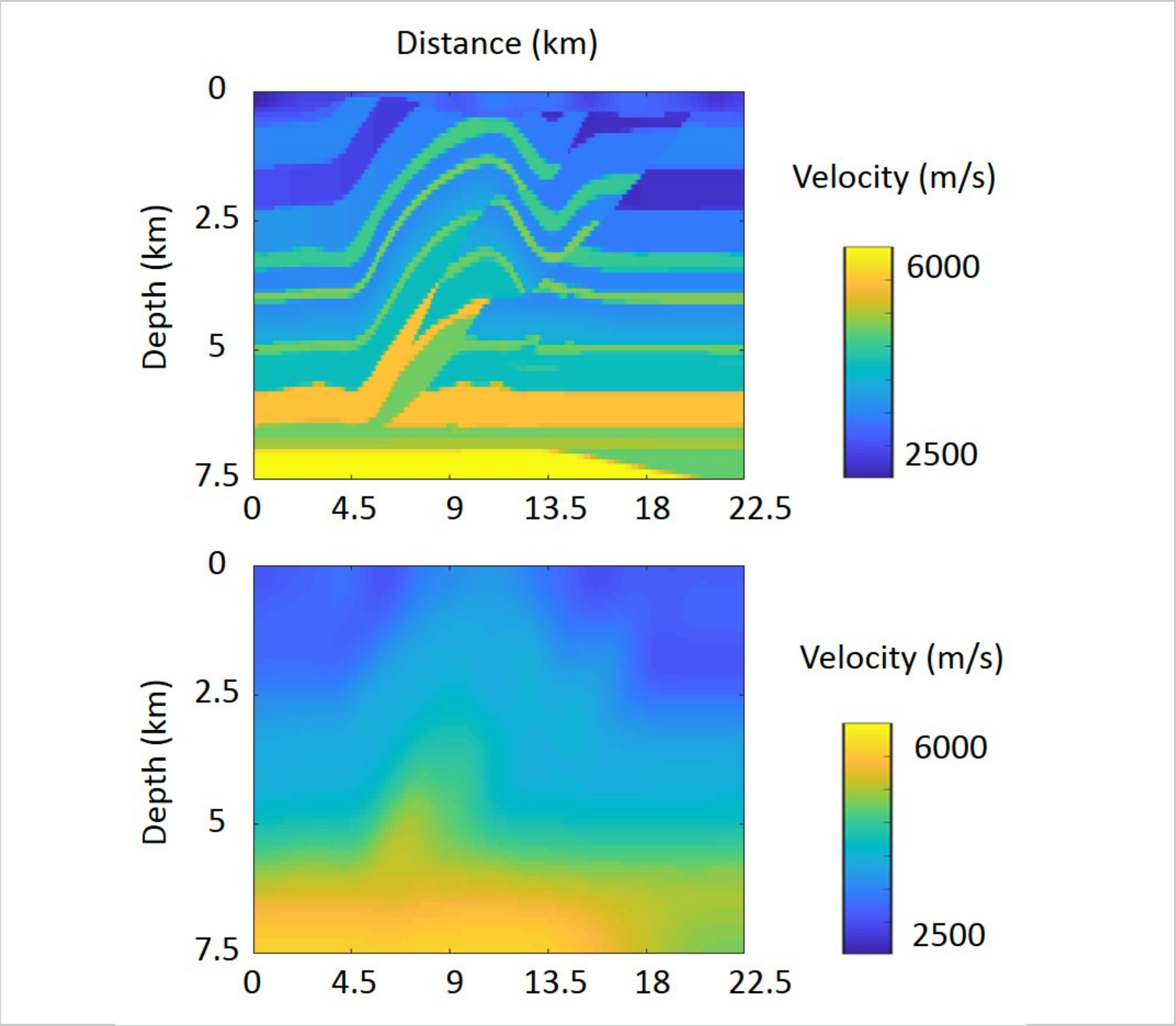
Sampling probability

$$\pi_i = \frac{\|\mathbf{A}_i\|_F^2}{\|\mathbf{A}\|_F^2}$$

Numerical experiments

- 2D SEG/EAGE overthrust model
  - 801×187 grid cells in a 2-D section with 25 m horizontal and vertical grid intervals
  - There are 100 sources and 100 receivers laid on the surface, which are spread out with 25 m spatial interval.
  - A multi-scale inversion approach is adopted in our numerical experiments in frequency bands $0.5 - 4$ Hz in every 0.5 Hz.

Overthrust model (true) & Initial velocity model m_0

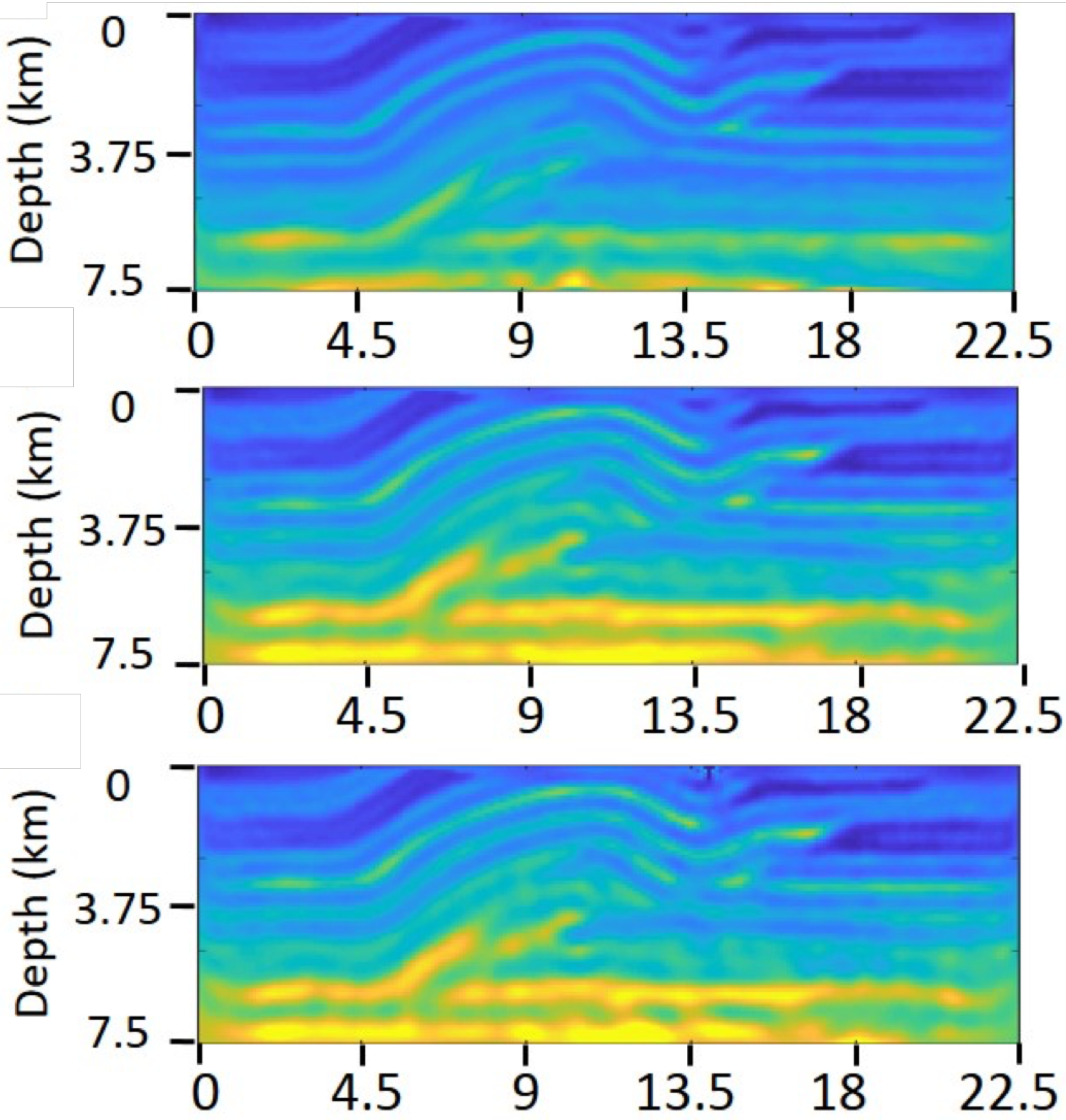


The inverse results using the data frequency band (0.5 – 4Hz).

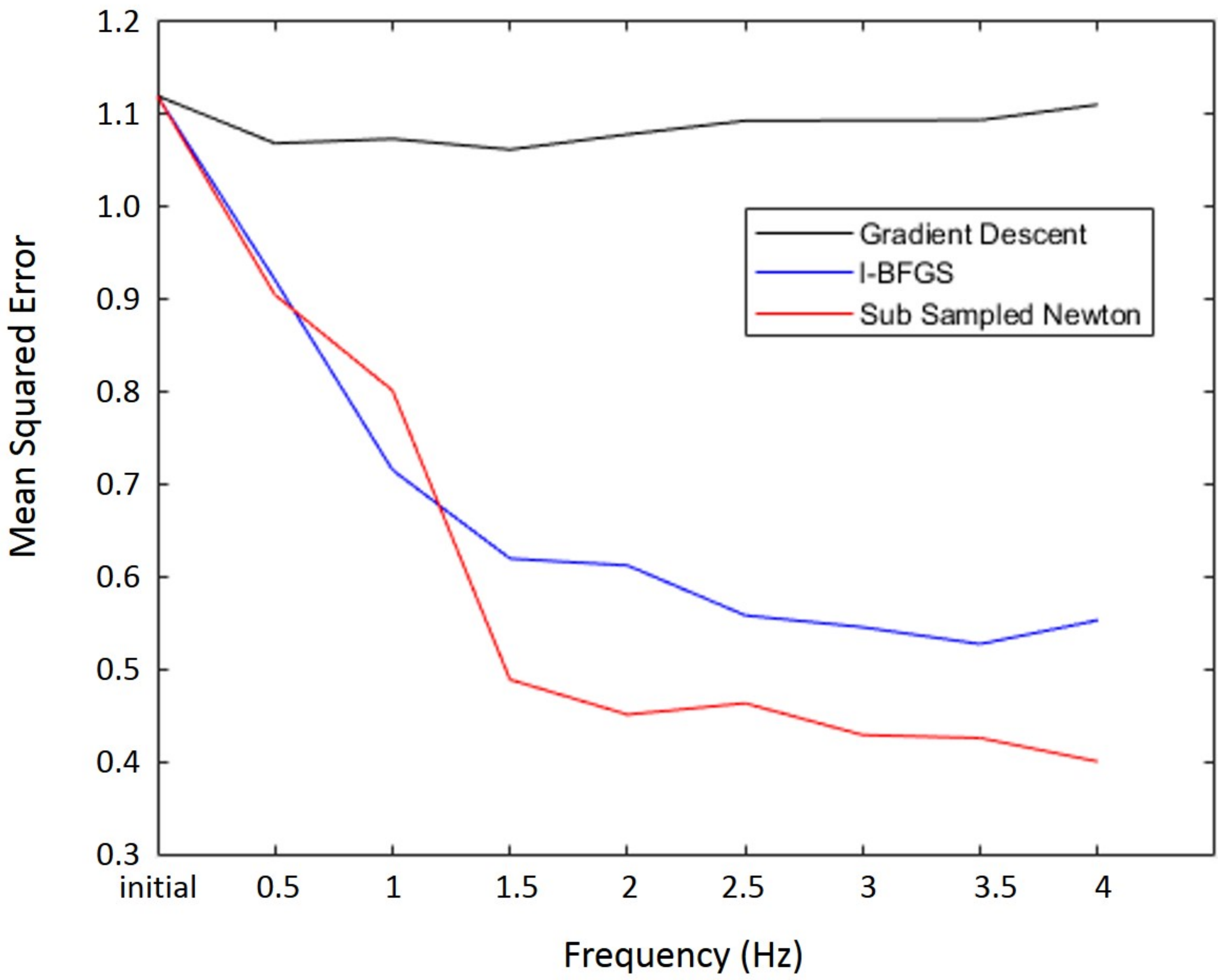
gradient decent

$l - BFGS$

Sub-Sampled Newton



Convergence comparison of different methods.



Conclusion

- An efficient Sub-Sampled Newton (SSN) method to solve complex non-linear system guided by physical laws with application to FWI problem.
- SSN significantly reduces the computational complexity while preserving a fast convergence property, by using the non-uniform subsampling techniques.
- SSN captures the important information in the second order term thus having a rapid rate of convergence.

Future Work

- Efficient sampling algorithm in forward modeling
- Distributed sampling algorithm
- GPU accelerated sub-sampled Newton's method