SINBAD Consortium 2010

Randomized Full Waveform Inversion

Peyman P. Moghaddam

SLIM
University of British Columbia

Motivation

- Cost of the FWI is proportional to the number of shots and it requires hundreds of RTM (Reverse Time Migration).
- Dimensionality reduction with compressive sensing aims at compressing the data volume in inversion.
- Stochastic optimization provides better solution to randomized inversion problem than conventional optimization.

Overview

- Full Waveform Inversion (FWI)
- Simultaneous Source Experiment
- Randomized FWI
- Stochastic Optimization Methods
- Examples
- Conclusion
- Future Plans

Full Waveform Inversion

- Mathematically FWI can be formed as

$$\min_{\sigma} J(\sigma) = \frac{1}{2}||\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)||_2 \text{ subject to } \mathbf{F}(\mathbf{u}, \sigma) = \mathbf{0}$$

$$\mathbf{F}(\mathbf{u}, \sigma) = (\omega^2\sigma^2 + \nabla^2)\mathbf{u} + q = 0$$

$\mathbf{d} : data$

$\mathbf{D} : \text{ detection operator}$

$\sigma : sloweness$

$\mathbf{u}(\sigma) : wavefield$

$\omega : \text{ angular frequency}$

$q : \text{ source}$

$\nabla^2 : Laplacian$

(Ben Hadj Ali 08)

Full Waveform Inversion

- Gradient of the cost function with respect to slowness is defined as,

$$\partial J(\sigma)/\partial\sigma = -\Re\{(\partial \mathbf{u}/\partial\sigma)^H \mathbf{D}^T[\mathbf{d} - \mathbf{D}\mathbf{u}(\sigma)]\}$$

$$\Re : \ \text{real part}$$

$$(.)^H : Hermitian$$

  - with $\partial \mathbf{u}/\partial\sigma$ defined as,

$$\partial \mathbf{u}/\partial\sigma \quad = \quad -2\omega^2(\omega^2\sigma^2 + \nabla^2)^{-1}\sigma\mathbf{u} = \mathbf{K}$$

$$\mathbf{K} : \ \text{de-migration operator, linearized Born}$$

(Plessix 2009)

Conventional Optimization

- Limited-memory BFGS (Plessix 2009)

$$\sigma_{k+1} = \sigma_k - \tau \mathbf{H}_k \nabla J(\sigma_k)$$

$$\mathbf{H}_k: \text{ inverse Hessian}$$

$$\tau: \text{ line search value}$$

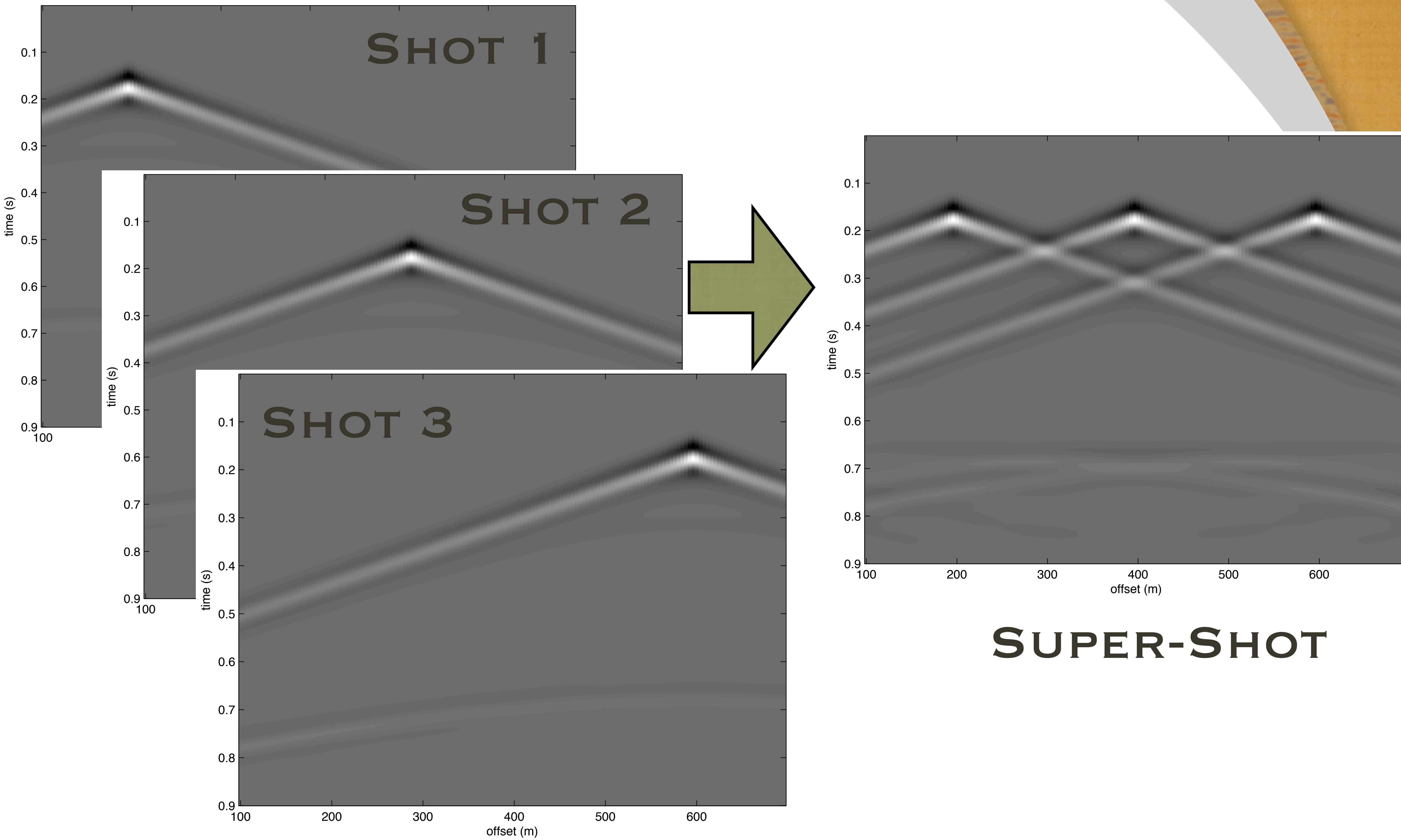
- Preconditioned Gradient method (Ravaut 2004)

$$\sigma_{k+1} = \sigma_k - \tau diag(\mathbf{K}^T\mathbf{K} + \epsilon\mathbf{I})^{-1} \nabla J(\sigma_k)$$

- Conjugate Gradient method (Virieux 2009)

SLIM

Simultaneous Source Experiment



(Krebs 2009)

Simultaneous Source Experiment

updates read

$$\sigma_{k+1} = \sigma_k - \tau \nabla J(\sigma_k, \mathbf{Q})$$

$\mathbf{Q}$: all sources

with $$\nabla J(\sigma_k, \mathbf{Q}) \approx \frac{1}{N_s} \sum_{i=1}^{i=N_s} \nabla J(\sigma_k, \mathbf{Q}_i)$$

$\mathbf{Q_i}$: a simultaneous source

$$\mathbf{E}\left(\frac{1}{N_s} \sum_{i=1}^{i=N_s} \nabla J(\sigma_k, \mathbf{Q}_i) \right) \rightarrow \nabla J(\sigma_k, \mathbf{Q})$$

$\mathbf{E}(.)$: *expectation*

Randomized FWI

$\sigma \longleftarrow \sigma_0$ initial model

$\{J(\sigma, \mathbf{Q_i}), \nabla_\sigma J(\sigma, \mathbf{Q}_i)\} \longleftarrow$ new randomized super-shot

$\textbf{While} \|\nabla_\sigma J(\sigma)\| \geq \epsilon$

$\sigma \longleftarrow$ update model with $J(\sigma, \mathbf{Q}_i), \nabla_\sigma J(\sigma, \mathbf{Q}_i)$

$\{J(\sigma, \mathbf{Q_i}), \nabla_\sigma J(\sigma, \mathbf{Q}_i)\} \longleftarrow$ new randomized super-shot

end

(Moghaddam 2010, Krebs 2009)

Stochastic Optimization Approaches

- Stochastic Gradient Descent

$$\sigma_{k+1} = \sigma_k - \tau \nabla J(\sigma_k, \mathbf{d}_k)$$

- Integrated Stochastic Gradient Descent (iSGD)

$$\sigma_{k+1} = \sigma_k - \eta_k \overline{\nabla J(\sigma_k)}$$

with $\overline{\nabla J(\sigma_k)}$ is averaging on the past numbers of gradients with weights,

$$\overline{\nabla J(\sigma_k)} = \frac{\sum_{i=k-m}^{k} e^{\alpha[i-(k-m)]} \nabla J(\sigma_i, \mathbf{d}_i)}{\sum_{i=k-m}^{k} e^{\alpha[i-(k-m)]}}$$

(Moghaddam 2010)

Stochastic Optimization

- Limited Memory BFGS (quasi-Newton methods),

$$\sigma_{k+1} = \sigma_k - \eta_k \mathbf{H}_k \nabla J(\sigma_k, \mathbf{d}_k)$$

- with updates:

$$\mathbf{H}_{k+1} = \mathbf{V}_k^T \mathbf{H}_k \mathbf{V}_k + \rho_k \mathbf{s}_k \mathbf{s}_k^T$$

$$\mathbf{s}_k = \sigma_{k+1} - \sigma_k \qquad \mathbf{y}_k = \nabla J(\sigma_{k+1}) - \nabla J(\sigma_k)$$

$$\mathbf{V}_k = \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^T$$

- On-line Limited Memory BFGS

$$\mathbf{H}_0 = \sum_{i=1}^{m} \frac{\mathbf{s}_{k-i}^T \mathbf{y}_{k-i}}{\mathbf{y}_{k-i}^T \mathbf{y}_{k-i}}$$

(Schraudolph 2007)

Stochastic Optimization

- Regular Limited Memory BFSG (quasi-Newton methods),

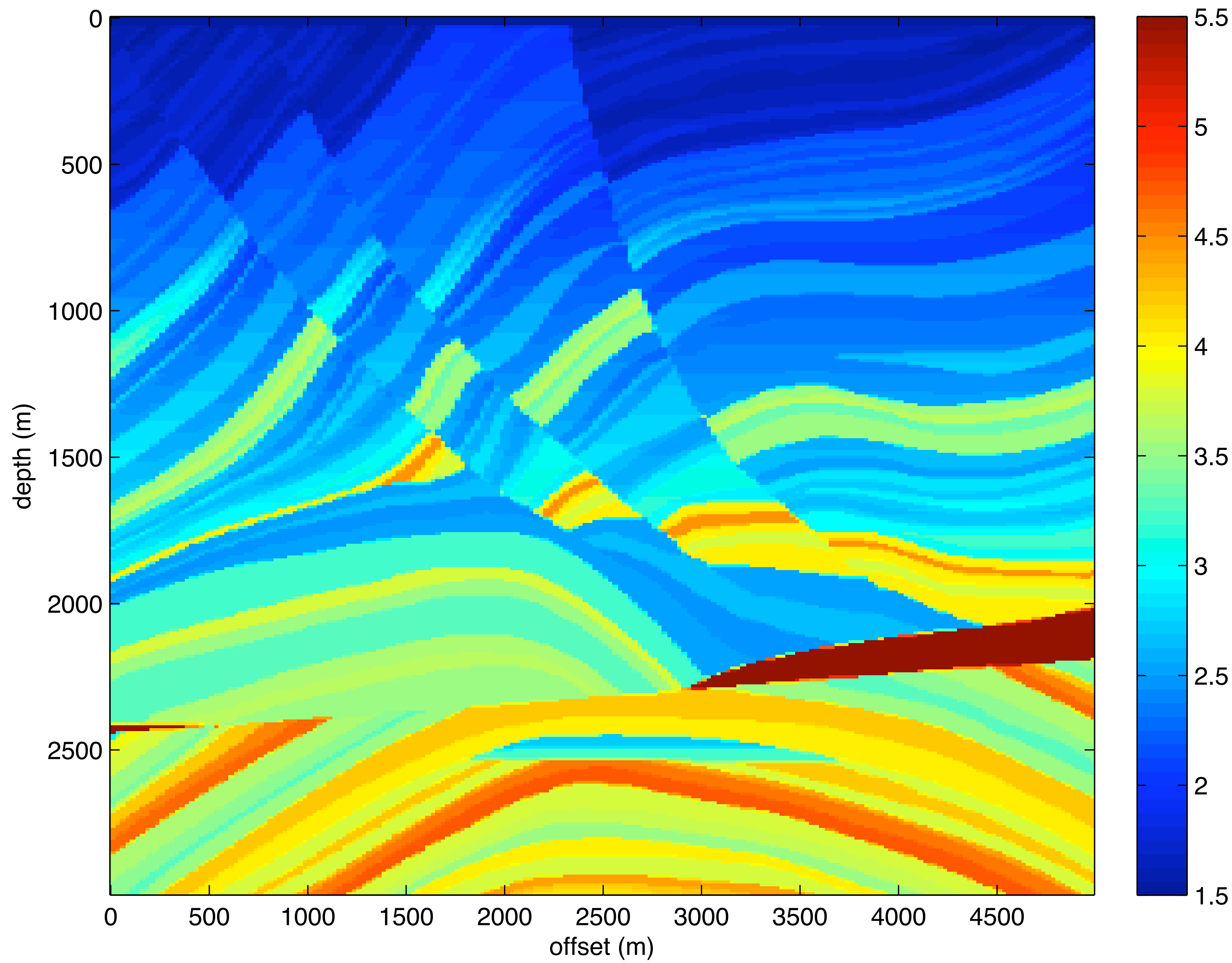
$$\begin{array}{ll} \min_H & ||\mathbf{H}_{k+1} - \mathbf{H}_k||_F \\ \text{subject to} & \mathbf{H}_{k+1}^T = \mathbf{H}_{k+1}, \mathbf{H}_{k+1}^T \mathbf{y}_k = \mathbf{s}_k \end{array}$$

$$\mathbf{s}_k = \sigma_{k+1} - \sigma_k \qquad \mathbf{y}_k = \nabla J(\sigma_{k+1}) - \nabla J(\sigma_k)$$

- Integrated Limited Memory BFSG (iLBFGS) ,

$$\begin{array}{ll} \min_H & ||\mathbf{H}_{k+1} - \sum_k \mathbf{H}_k||_F \\ \text{subject to} & \mathbf{H}_{k+1}^T = \mathbf{H}_{k+1}, \mathbf{H}_{k+1}^T \mathbf{y}_k = \mathbf{s}_k \end{array}$$

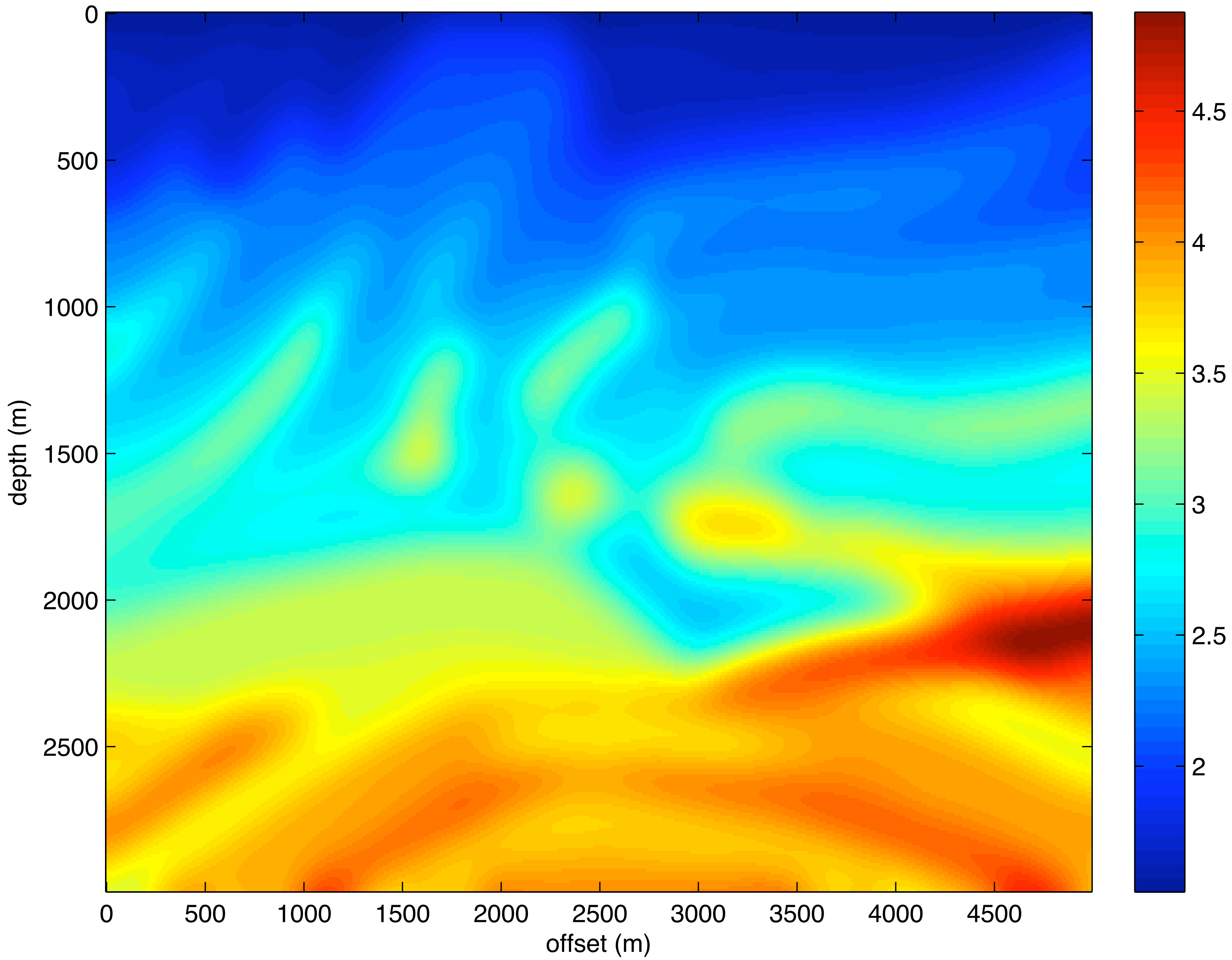
Examples (Marmoussi Model)



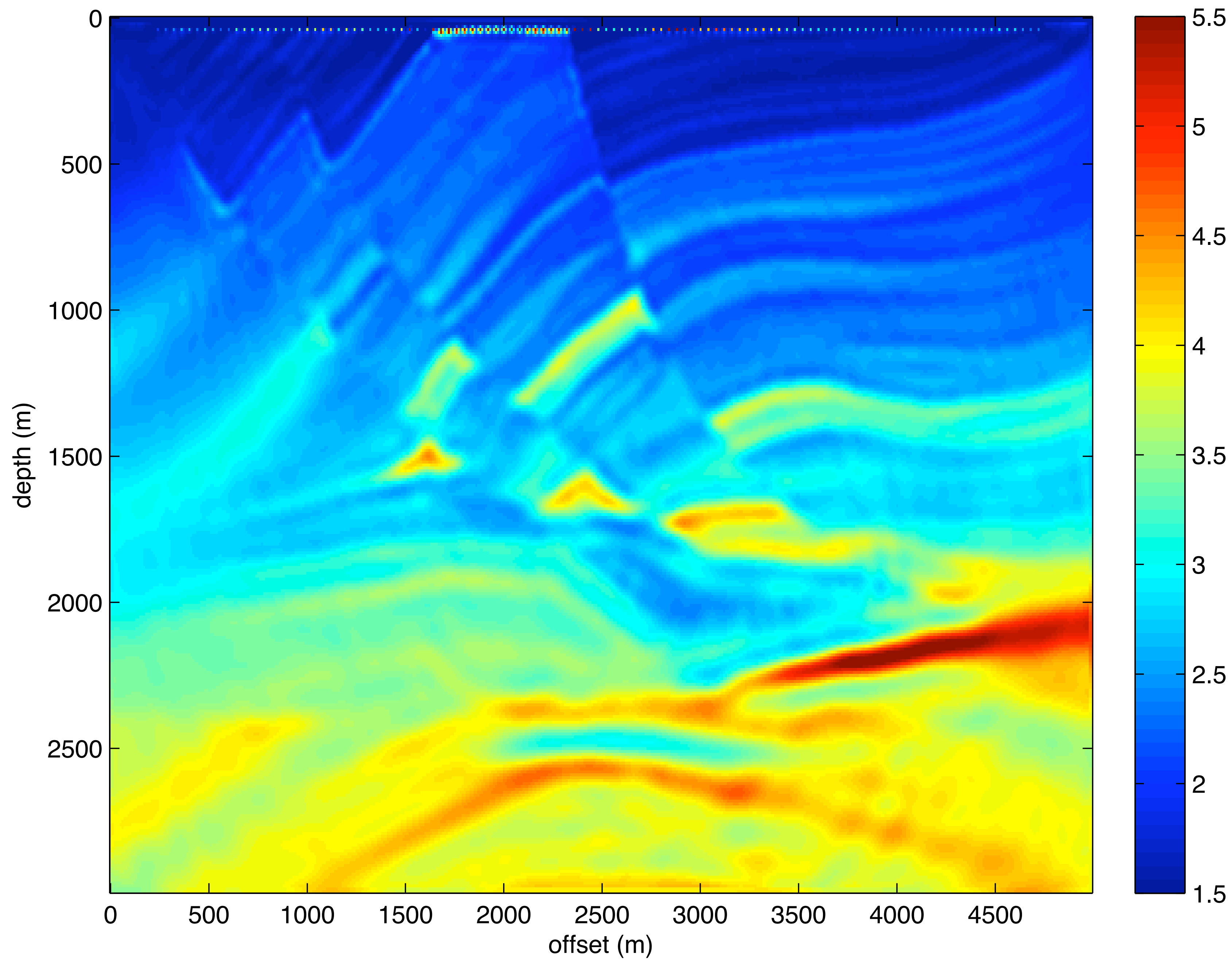
- 113 shots with 40 (m) spacing, 249 receivers with 20 (m) spacing, WAZ survey with 5 (km) max. aperture, Ricker source with 10 Hz central frequency, 3.6 second recording time with .9 (ms) time sampling.

SLIM

Examples (Initial Model)

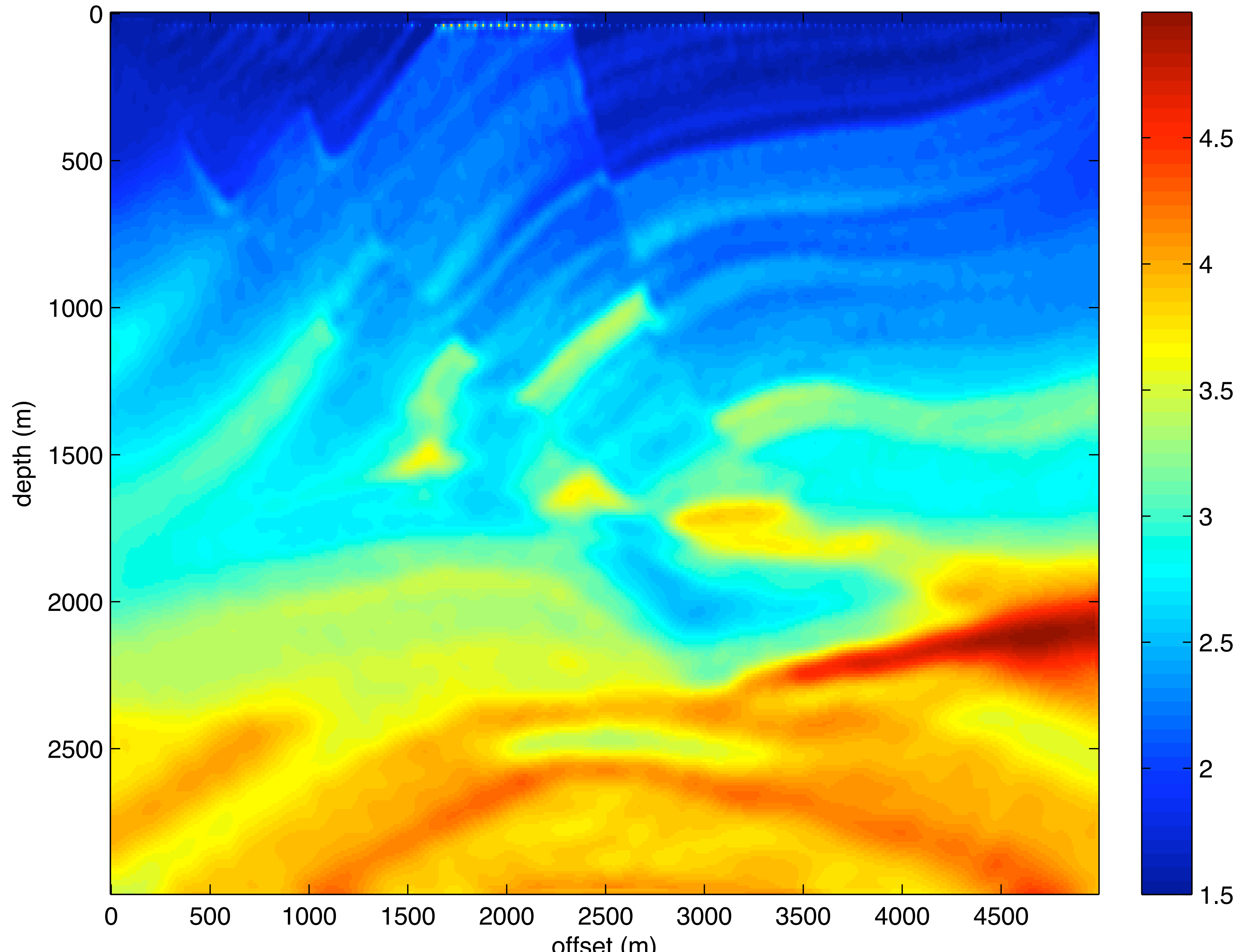


Examples (Inverted Model)



- inverted model after 18 iterations of LBFGS, 113 sequential shots, 50 frequency components has been used from 5 to 33 Hz with .55 Hz resolution

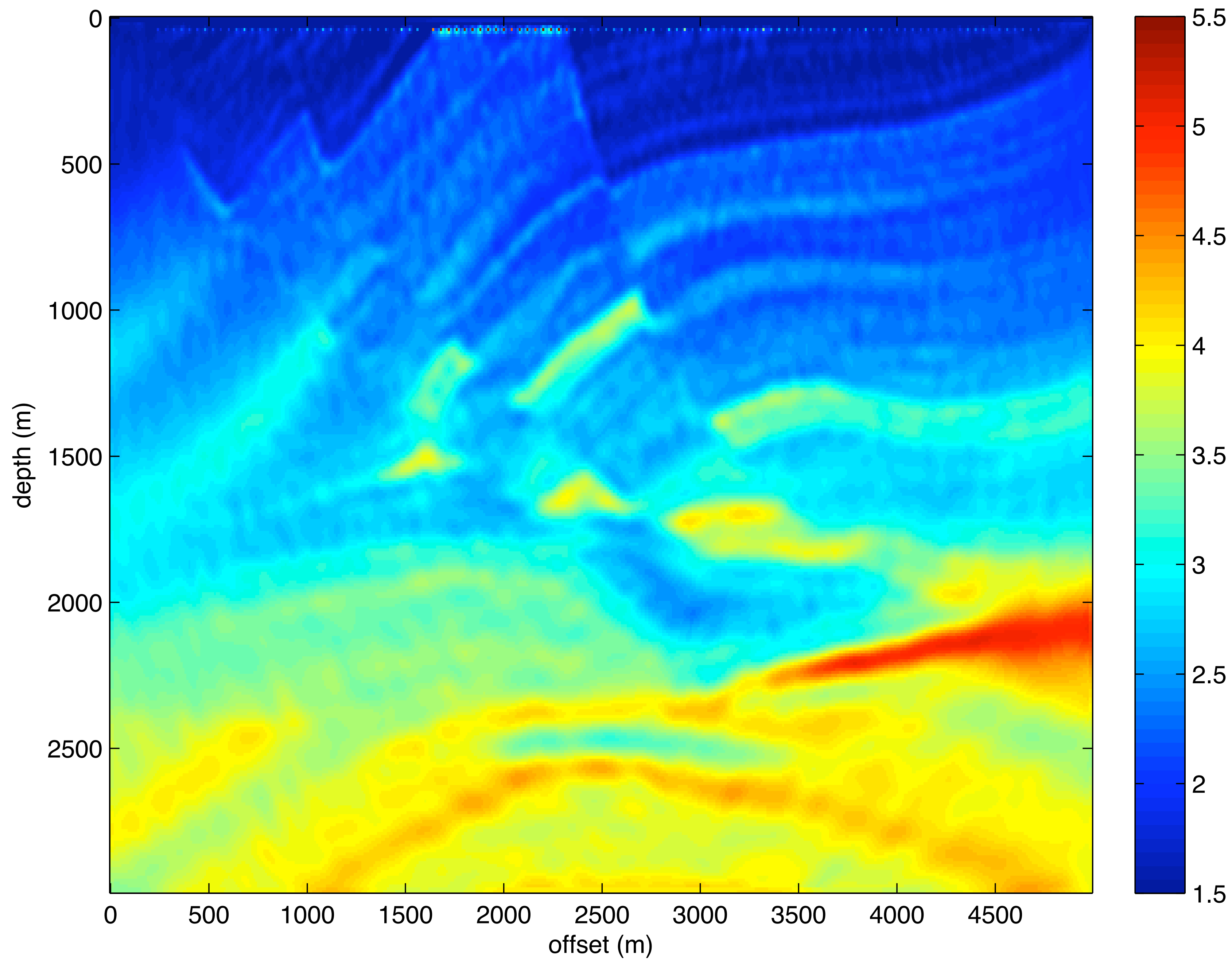
Randomized FWI (Inverted Model)



- Stochastic Gradient Descent, SNR= 4.65 dB, $\mathbf{SNR} = 20\log_{10}(\frac{\|\delta\mathbf{m} - \delta\tilde{\mathbf{m}}\|_2}{\|\delta\mathbf{m}\|_2})$

- 16 Randomized simultaneous shots, 4 frequencies, 40 times speed up

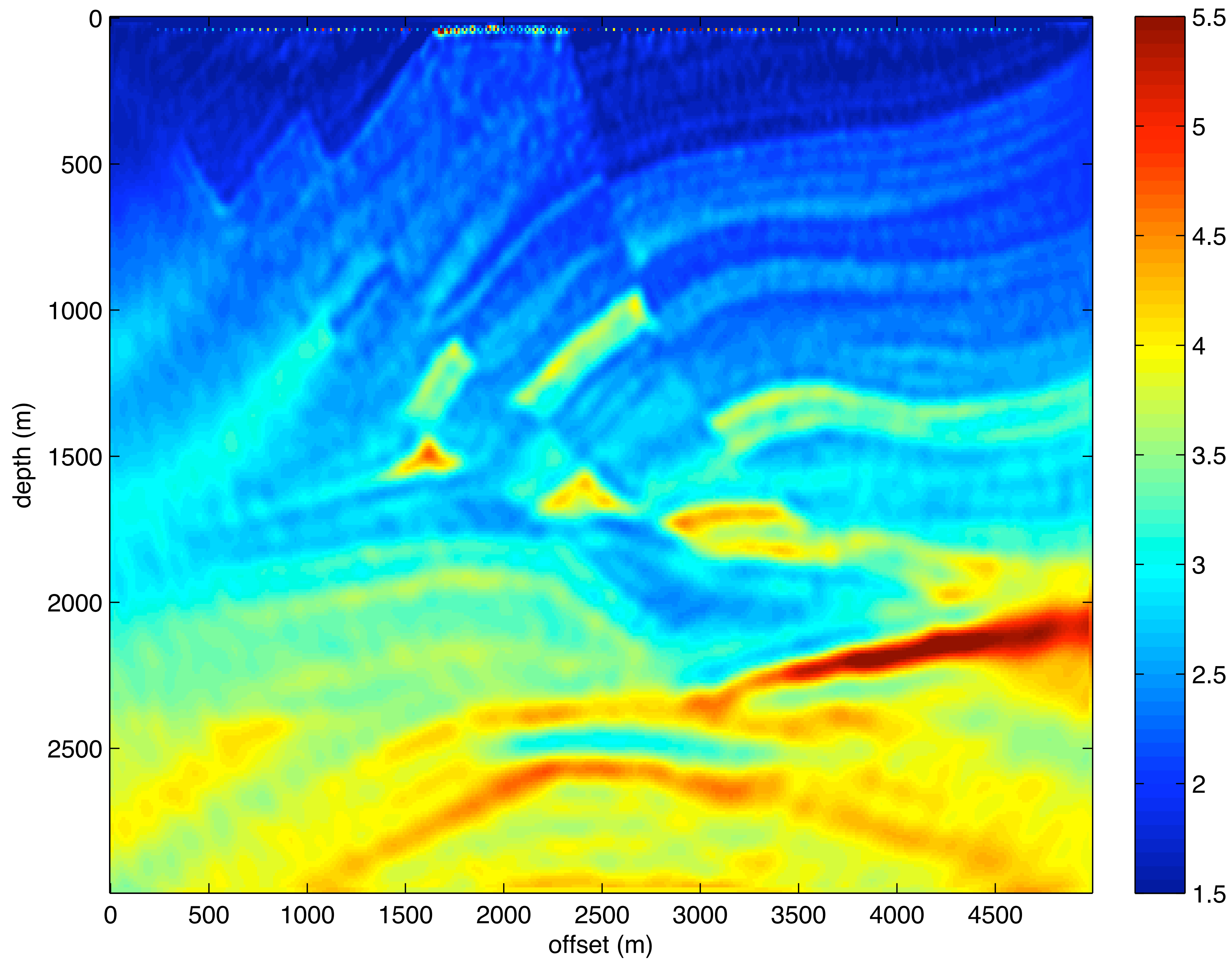
Randomized FWI (Inverted Model)



- online-LBFGS, SNR= 7.17 dB

- 16 Randomized simultaneous shots, 4 frequencies, 40 times speed up

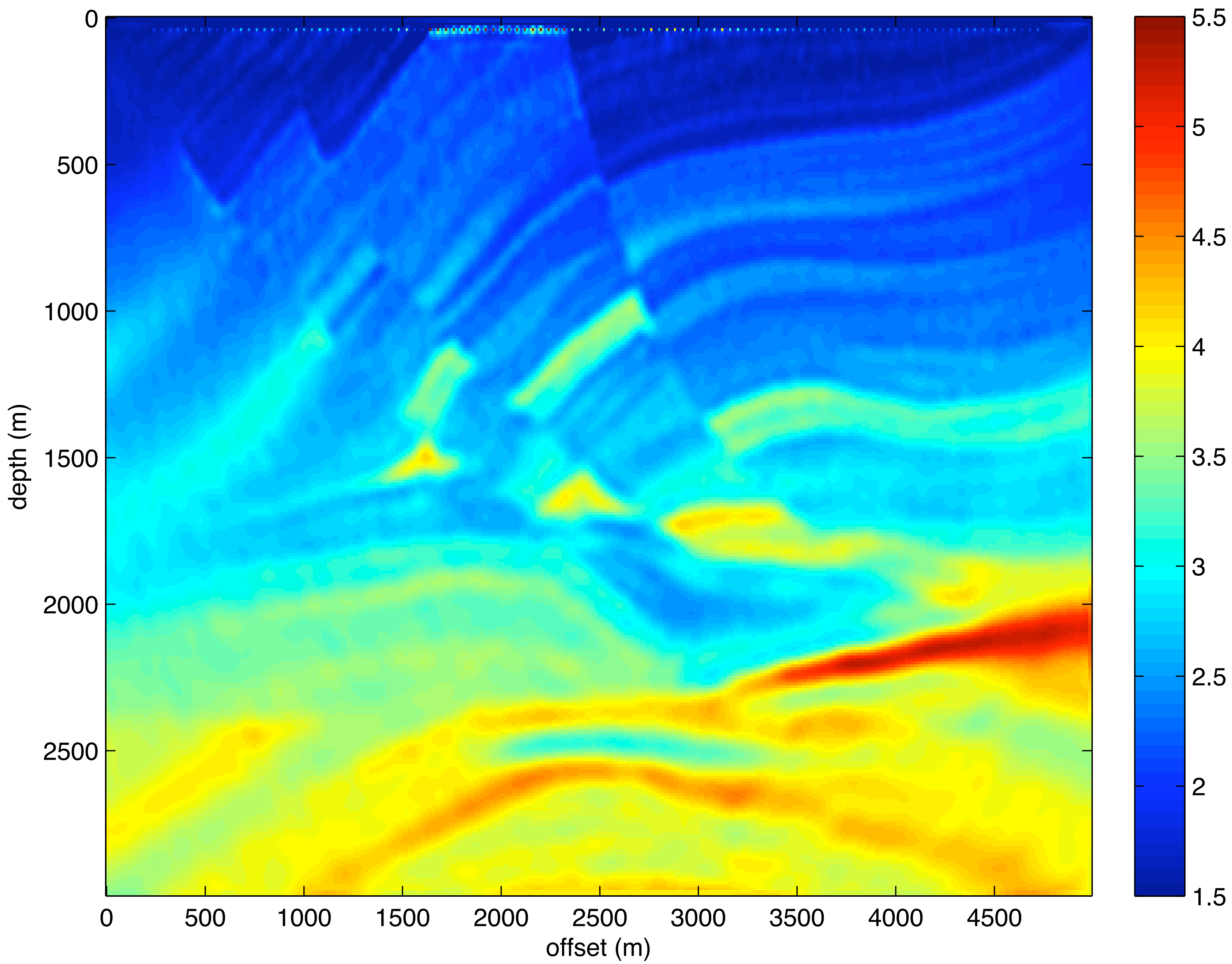
Randomized FWI (Inverted Model)



- iLBFGS, SNR= 9.10dB

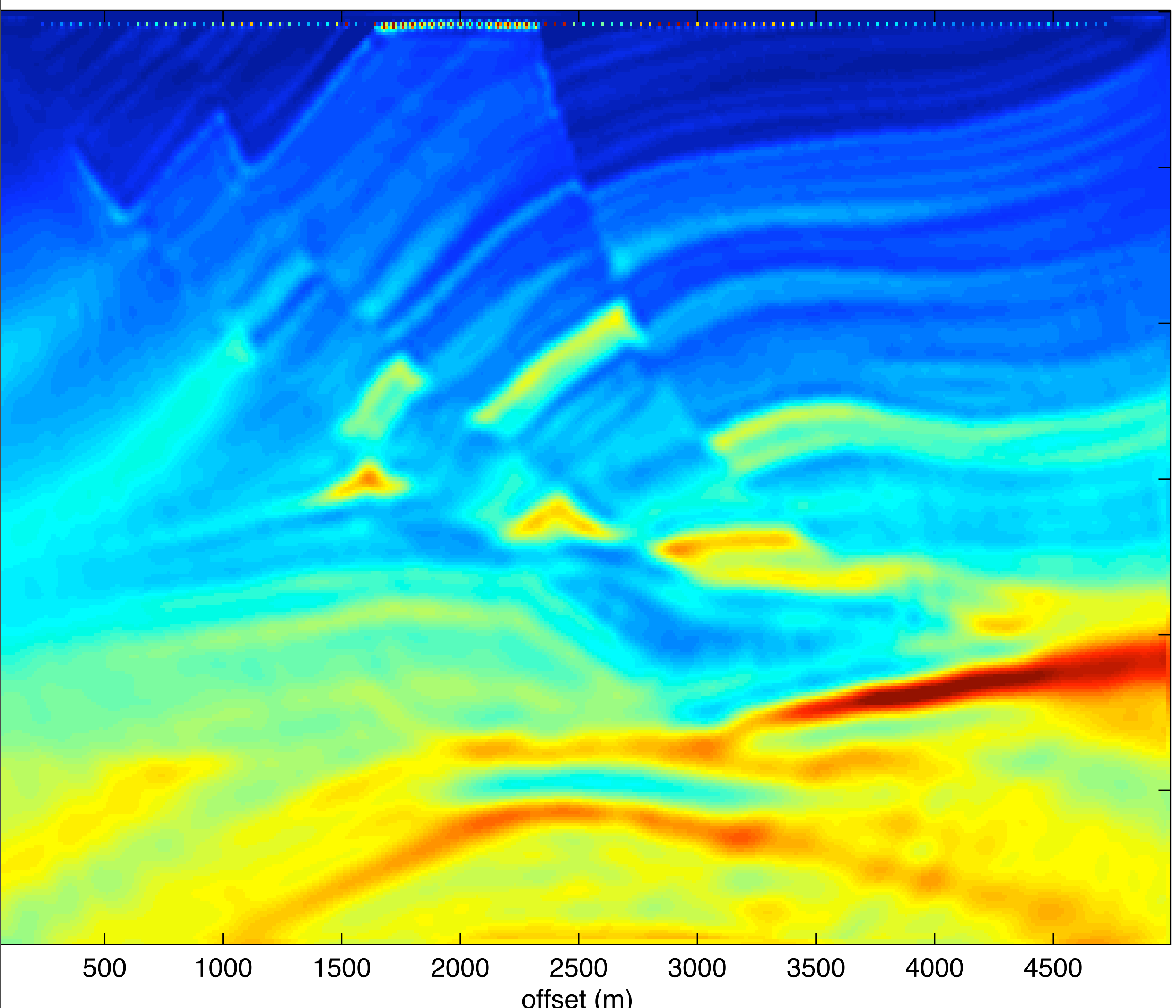
- 16 Randomized simultaneous shots, 4 frequencies, 40 times speed up

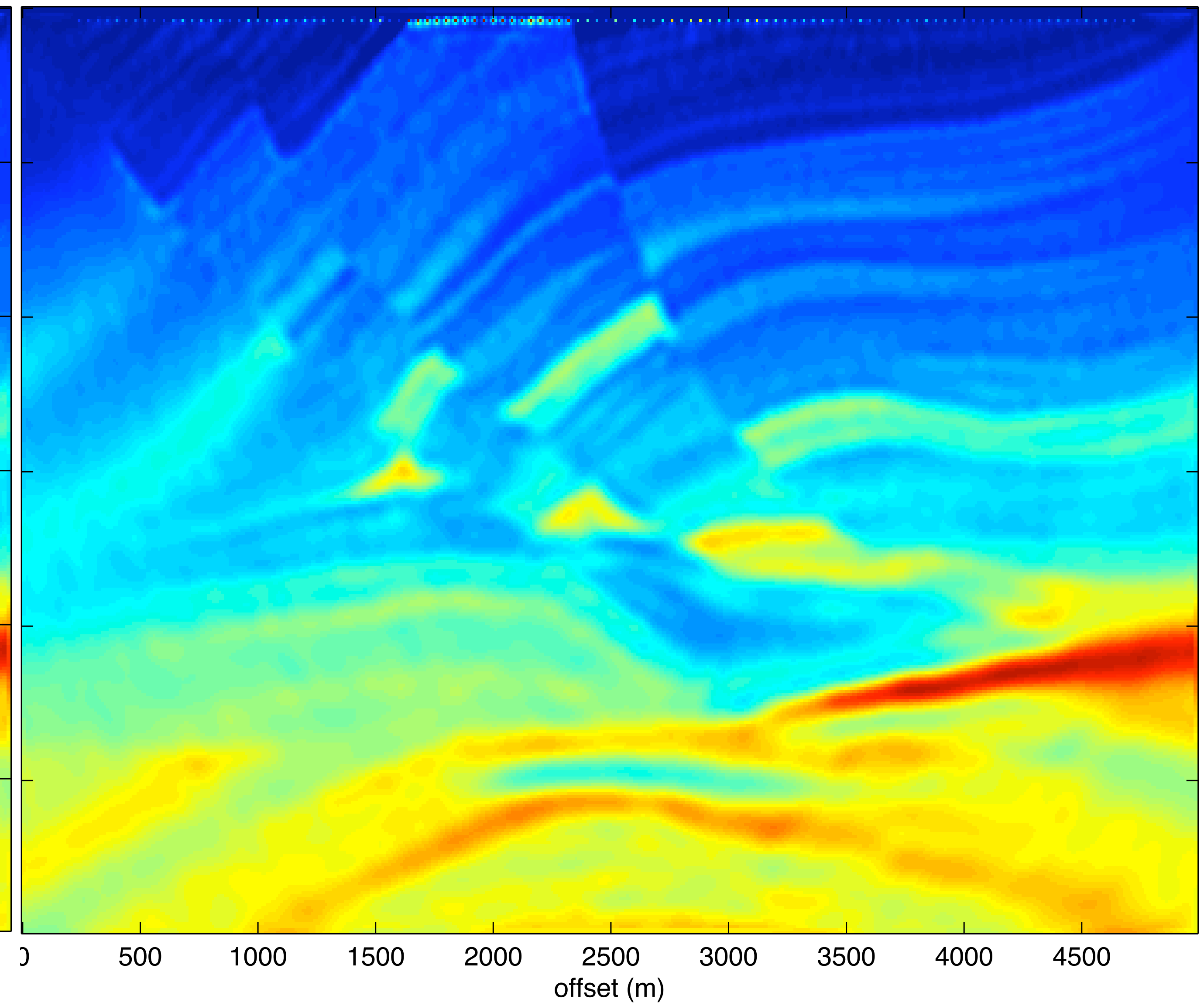
Randomized FWI (Inverted Model)



- iSGD, SNR= 10.85dB
- 16 Randomized simultaneous shots, 4 frequencies, 40 times speed up

Comparison



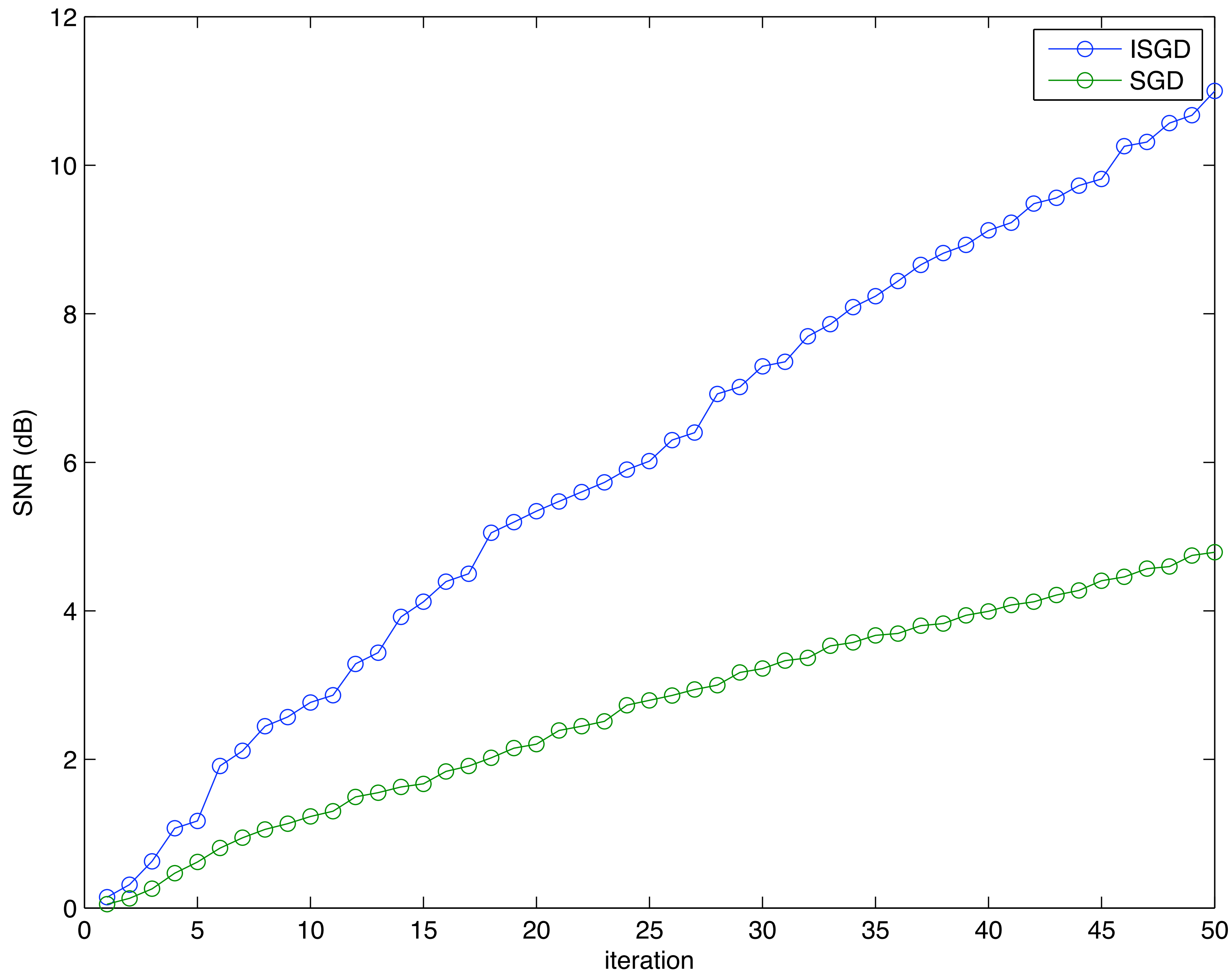


- Inversion for all the shots,
- 1 week on the 32 CPU cluster

- iSGD, SNR= 10.85dB
- 8 hours on the 32 CPU cluster

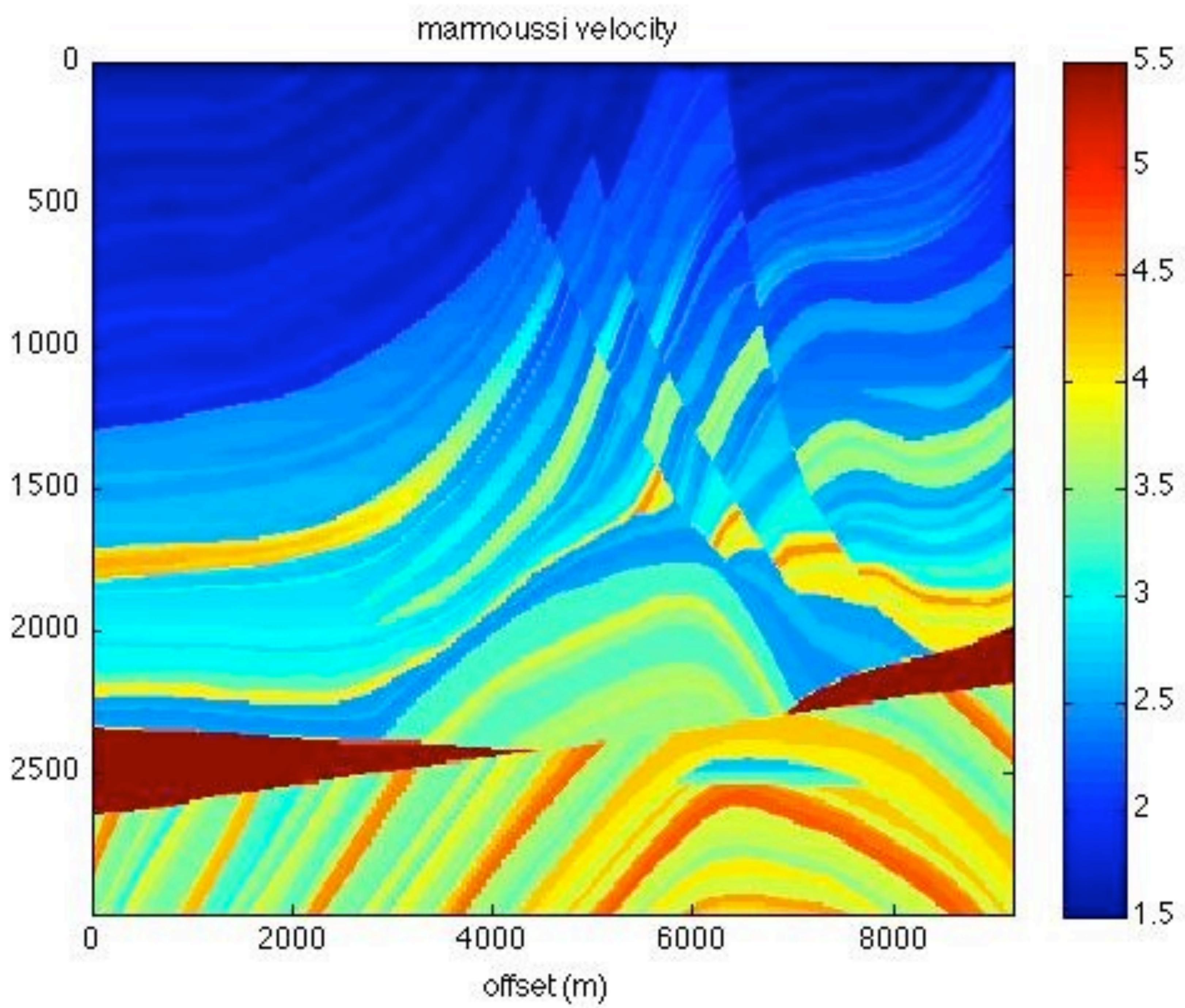
- 16 Randomized simultaneous shots, 4 frequencies, 40 times speed up

Comparison



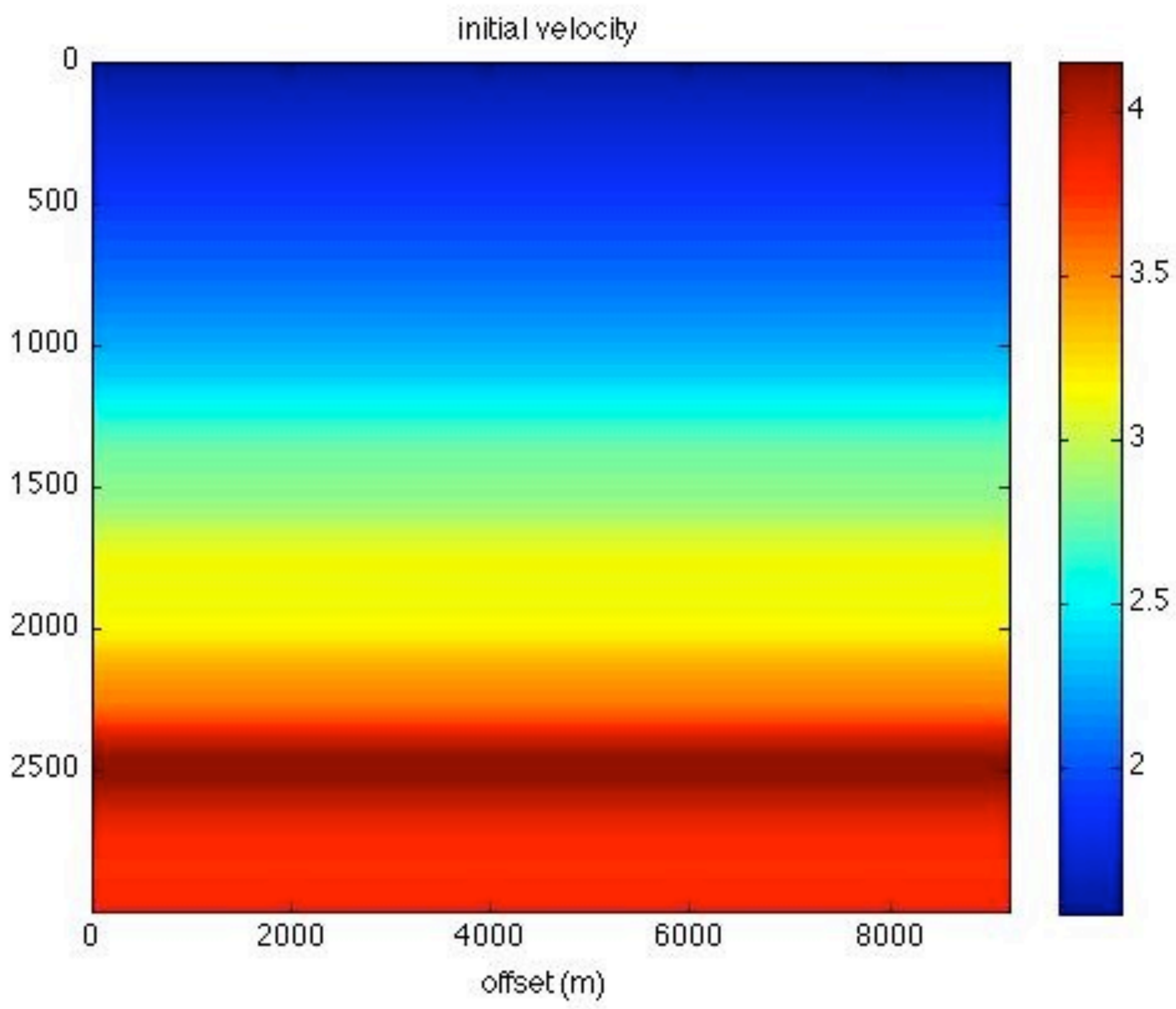
- Comparison between conventional gradient descent and stochastic gradient descent.

Examples (Marmoussi Model)

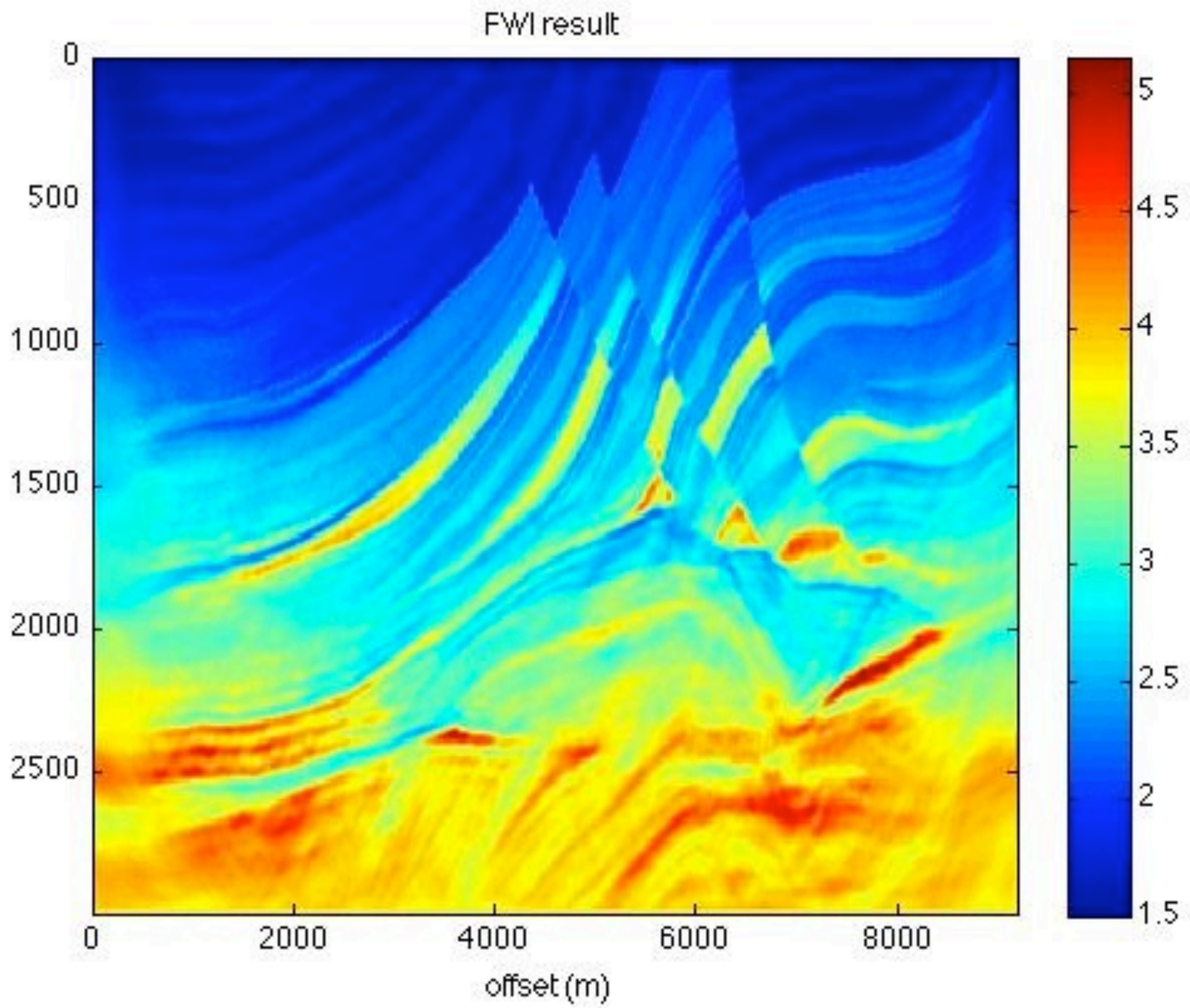


- 900 shots with 10 (m) spacing, 900 receivers with 10 (m) spacing, WAZ survey with 5 (km) max. aperture, Ricker source with 10 Hz central frequency, 3.6 second recording time with .9 (ms) time sampling.

Examples (Initial Model)



Randomized FWI (Inverted Model)



- iSGD method, 1 Randomized simultaneous shots, 900 times speed up!

SLIM

Conclusion

- Super-shot experiment combined with stochastic optimization methods produce promising results for solution for FWI
- Randomized FWI greatly increases the performance of the FWI.
- Dimensionality reduction algorithms, open possibility of replacing migration with FWI with no extra cost.

SLIM

Future Plans

- Further investigation on the choice of random frequency and super-shot
- Stochastic optimization strategies for FWI, improved iLBFGS, Natural gradient
- Regularization for the FWI
- Solving the uniqueness problem, exploiting the multi-scale nature of the FWI

Acknowledgements

- The authors would like to thank Yogi Erlangga and Tim Lin for their Helmoltz operator.

- The authors would like to thank Eldad Haber for discussion on stochastic optimization and regularized inverse problem.

SINBAD

NSERC CRSNG

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE II (375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.