A hybrid stochastic-deterministic method for waveform inversion

T. van Leeuwen, M. Schmidt, M. Friedlander and F. Herrmann

SLIM
University of British Columbia

SLIM

Q: Do we need all the data all the time for FWI?

Q: Do we need all the data all the time?

A: No, we can rely on techniques from stochastic optimization and use random source encoding

[Krebs et. al ’09; Haber et. al ’10; van Leeuwen et. al ’10]

Q: Do we need all the data all the time?
A: No, we can rely on techniques from stochastic optimization and use random source encoding

Q: Do we need randomized source encoding to reap the benefits of stochastic optimization?

SLIM

Overview

- Full waveform inversion
- Conventional optimization
- Stochastic optimization
- Hybrid method
- Results
- Conclusions

SLIM

Full waveform inversion

Least-squares fitting of multi-experiment data

$$\min_{\mathbf{m}} \Phi[\mathbf{m}] = \frac{1}{K}\sum_{i=0}^{K-1} \phi_i[\mathbf{m}]$$

$$\phi_i[\mathbf{m}] = ||\mathbf{d}_i - F[\mathbf{m}]\mathbf{q}_i||_2^2$$

Typically, costs are proportional to K

Source encoding

Replace sequential sources by one simultaneous source: $\widetilde{\mathbf{q}} = \sum_j w_j \mathbf{q}_j$

$$\widetilde{\phi}[\mathbf{m}] = ||\widetilde{\mathbf{d}} - F[\mathbf{m}]\widetilde{\mathbf{q}}||_2^2$$

if $\mathsf{E}\{w_i w_j\} = \delta_{ij}$ **we get** $\mathsf{E}\left\{\widetilde{\phi}[\mathbf{m}]\right\} = \Phi[\mathbf{m}]$

$$\widetilde{\Phi}[\mathbf{m}] = \frac{1}{K}\sum_{i=0}^{K-1} ||\widetilde{\mathbf{d}}_i - F[\mathbf{m}]\widetilde{\mathbf{q}}_i||_2^2$$

SLIM

Source encoding

Expand the sums:

$$\widetilde{\phi}[\mathbf{m}] = \sum_{i,j} w_i w_j \langle \mathbf{d}_i - F[\mathbf{m}]\mathbf{q}_i, \mathbf{d}_j - F[\mathbf{m}]\mathbf{q}_j \rangle$$

if $\mathsf{E}\{w_i w_j\} = \delta_{ij}$ **we get**

$$\mathsf{E}\left\{\widetilde{\phi}[\mathbf{m}]\right\} = \sum_i ||\mathbf{d}_i - F[\mathbf{m}]\mathbf{q}_i||_2^2 = \Phi[\mathbf{m}]$$

Source encoding

Choice of random weights:

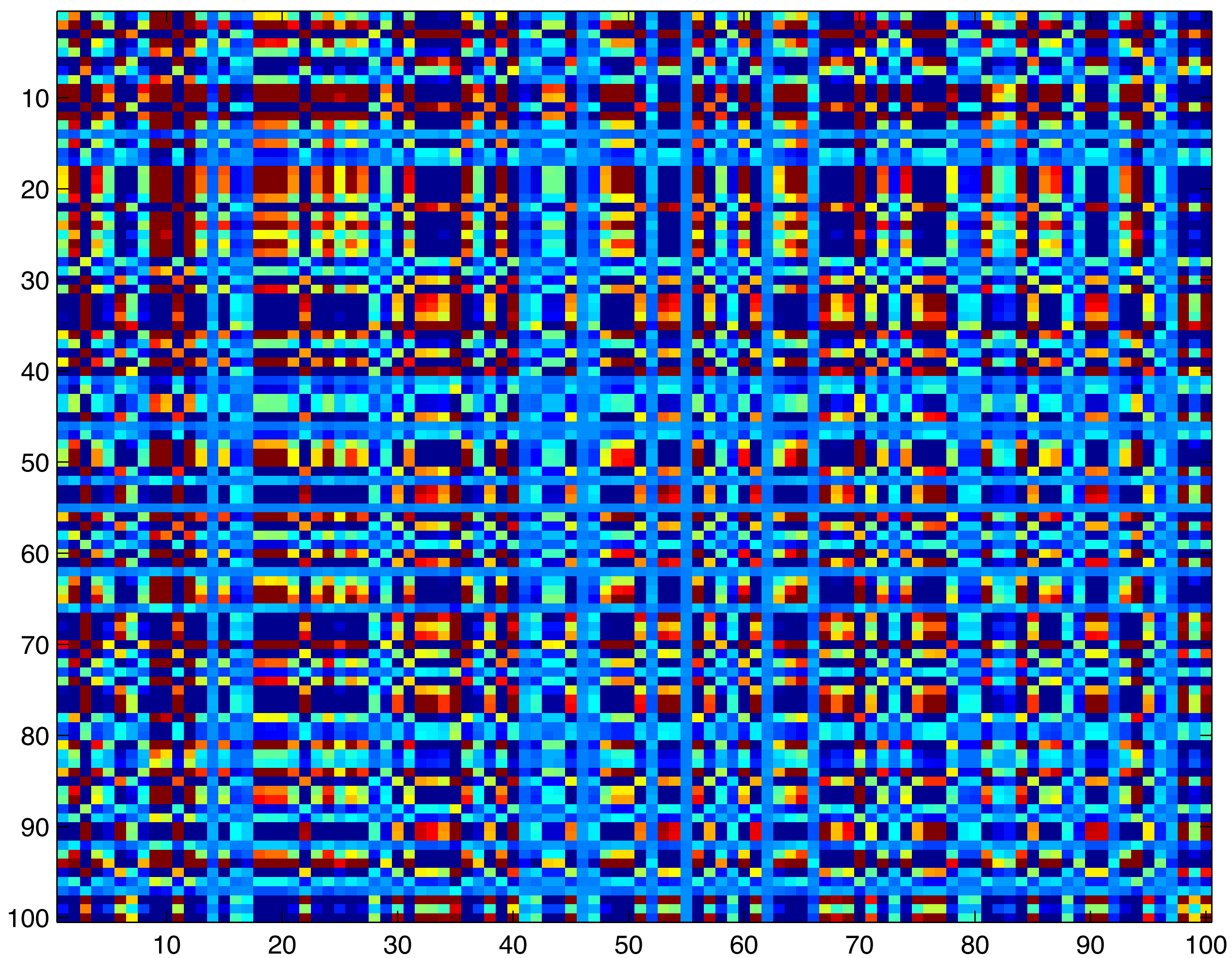
- Gaussian, ± 1, random phases: efficient in sampling the whole matrix, but problematic for marine data

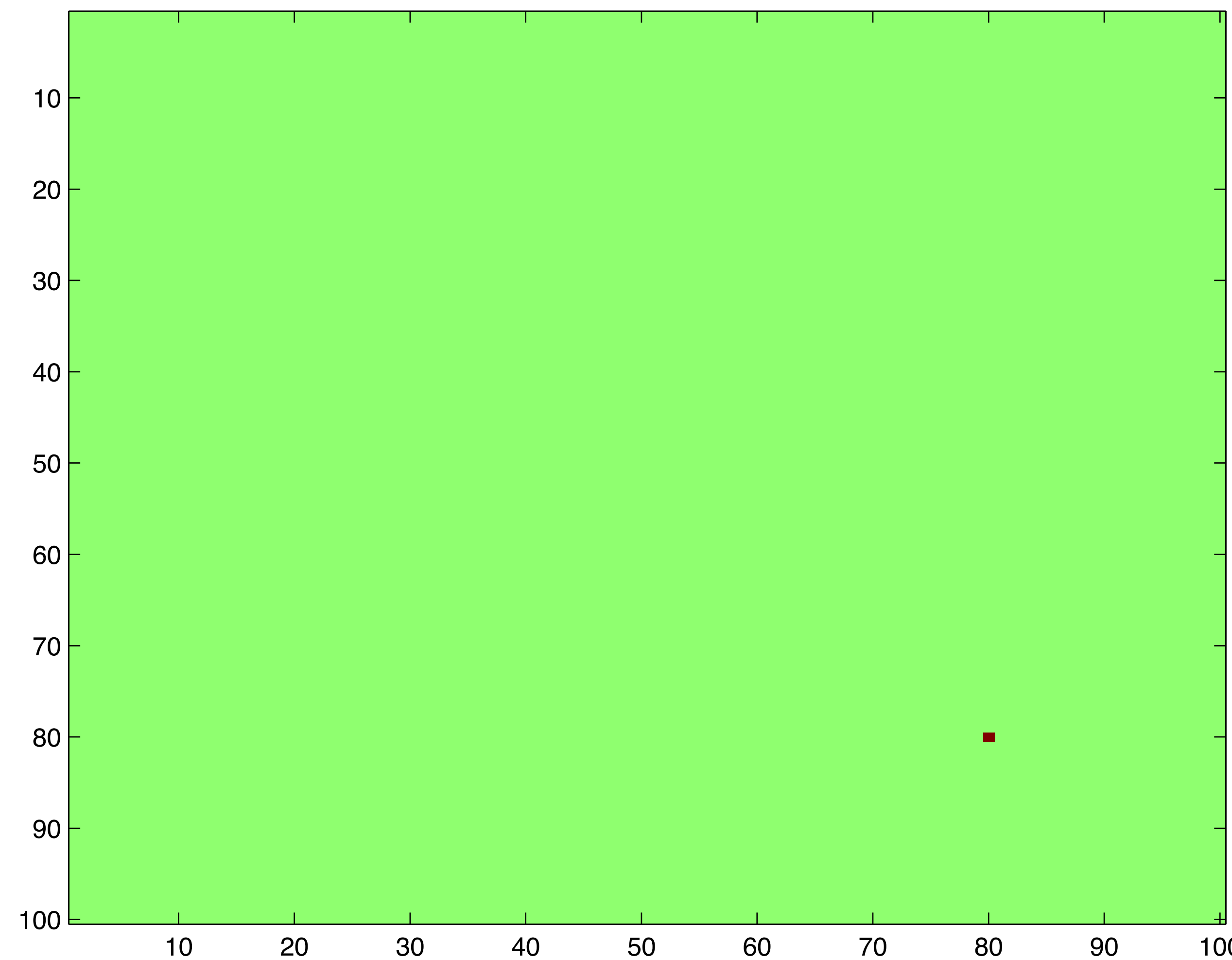
- Random unit vector: less efficient for trace estimation, but applicable for marine data!

Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

K=1



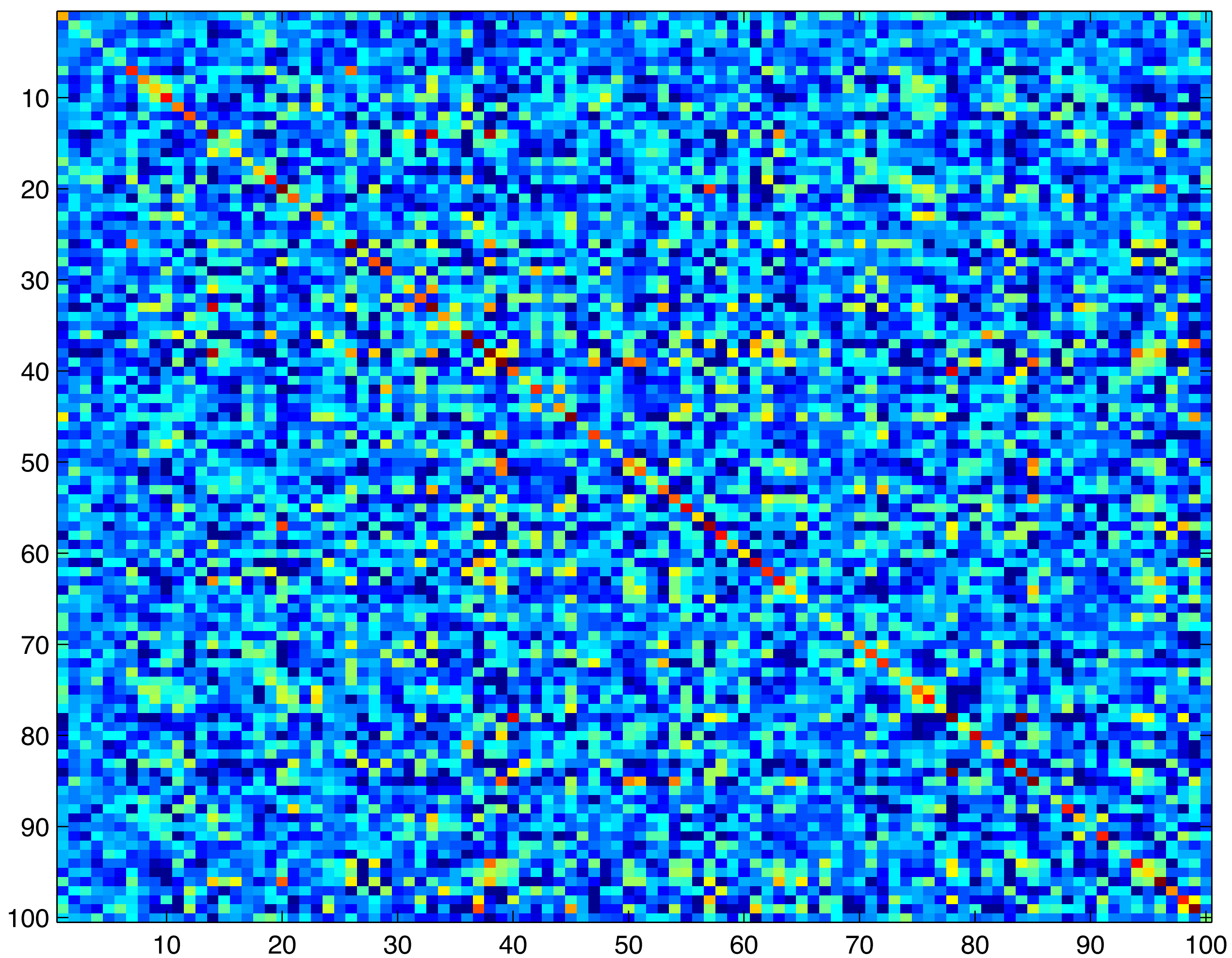


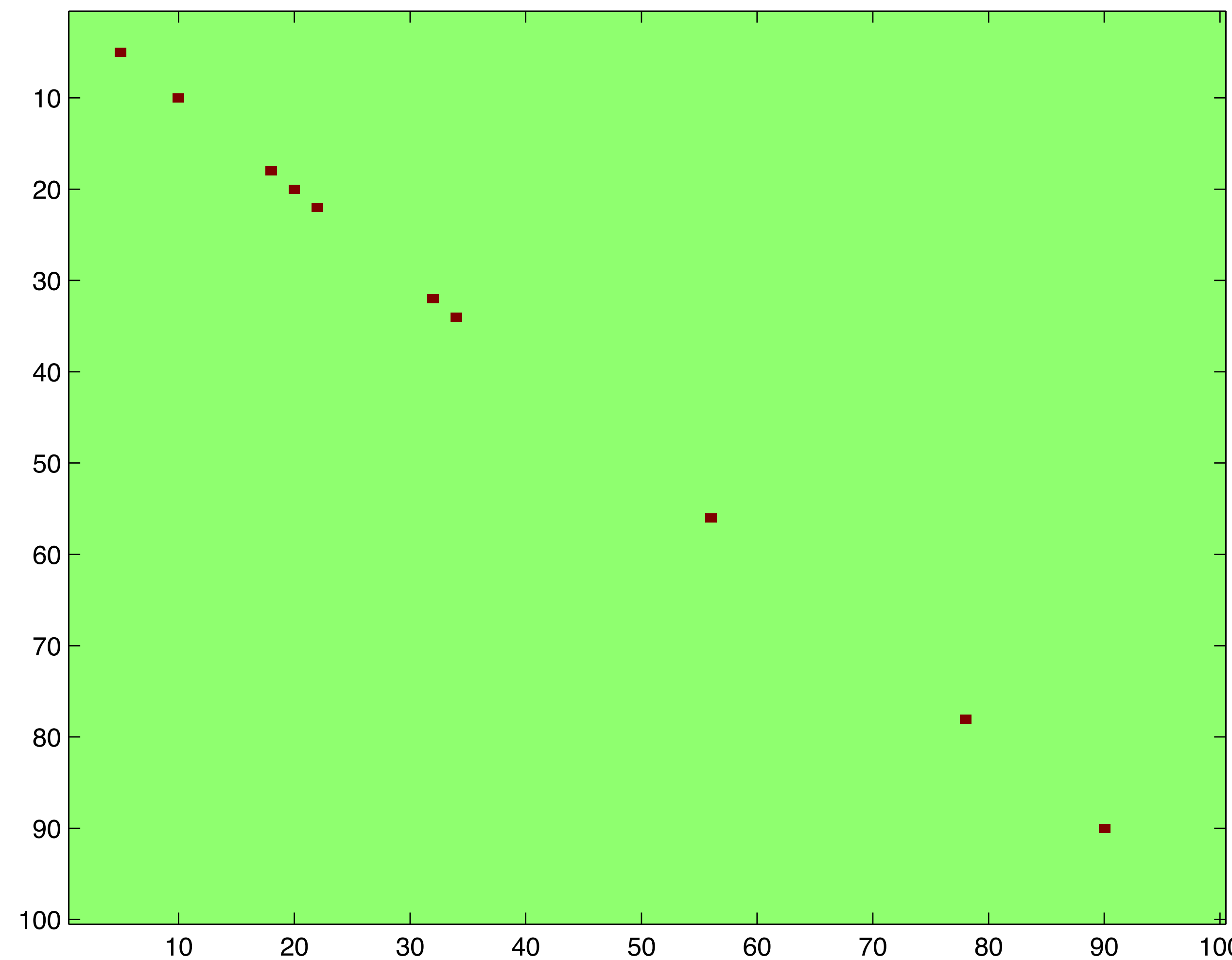
SLIM

Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

K=10

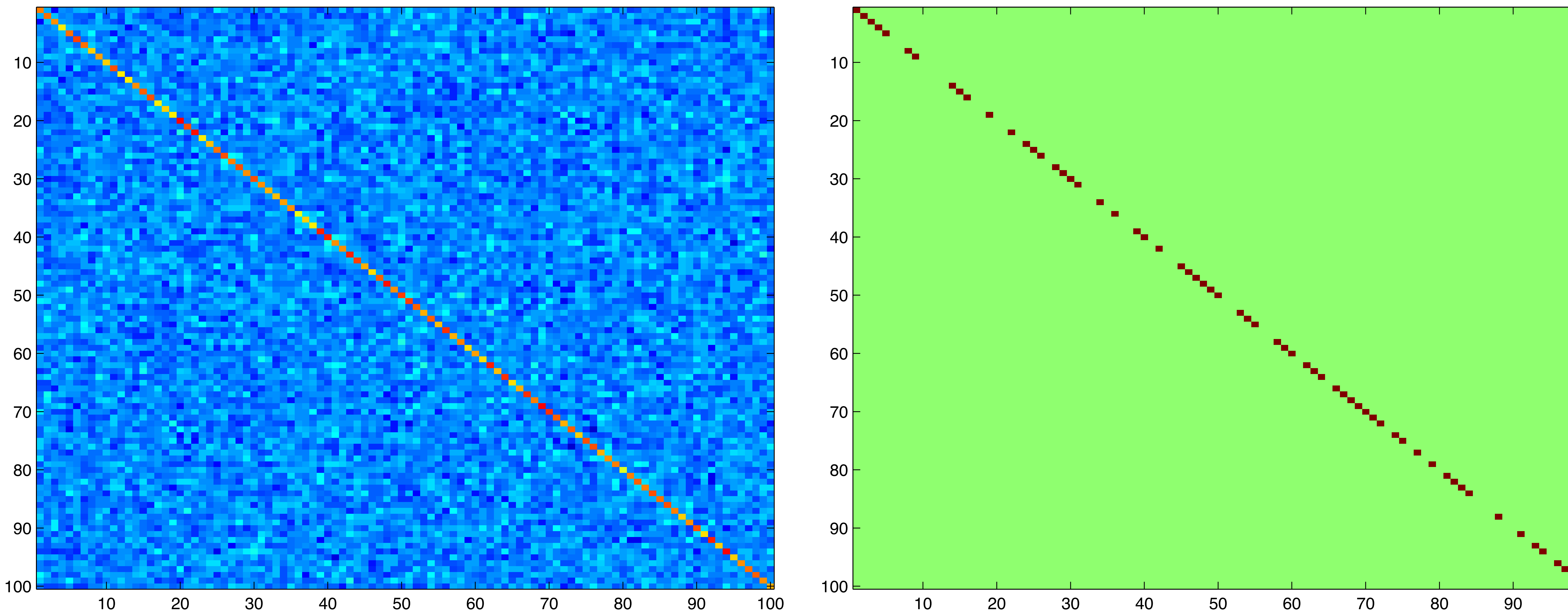




Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

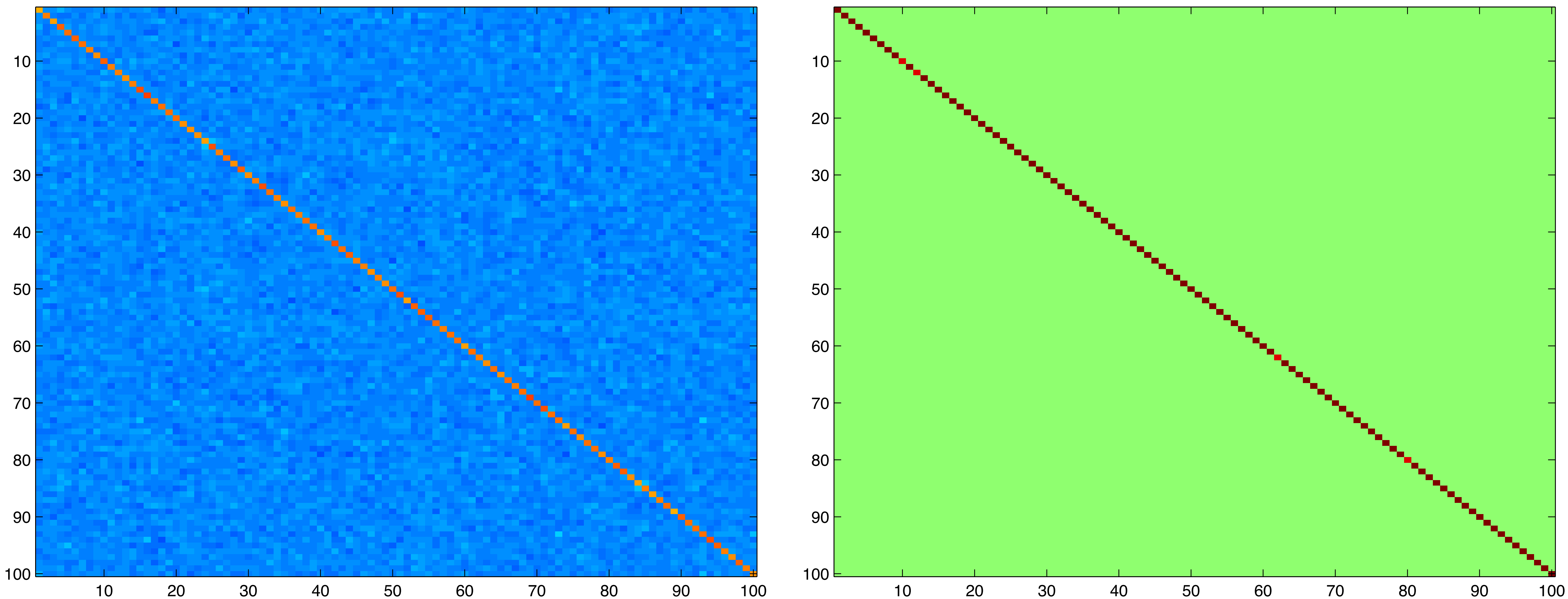
K=100



Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

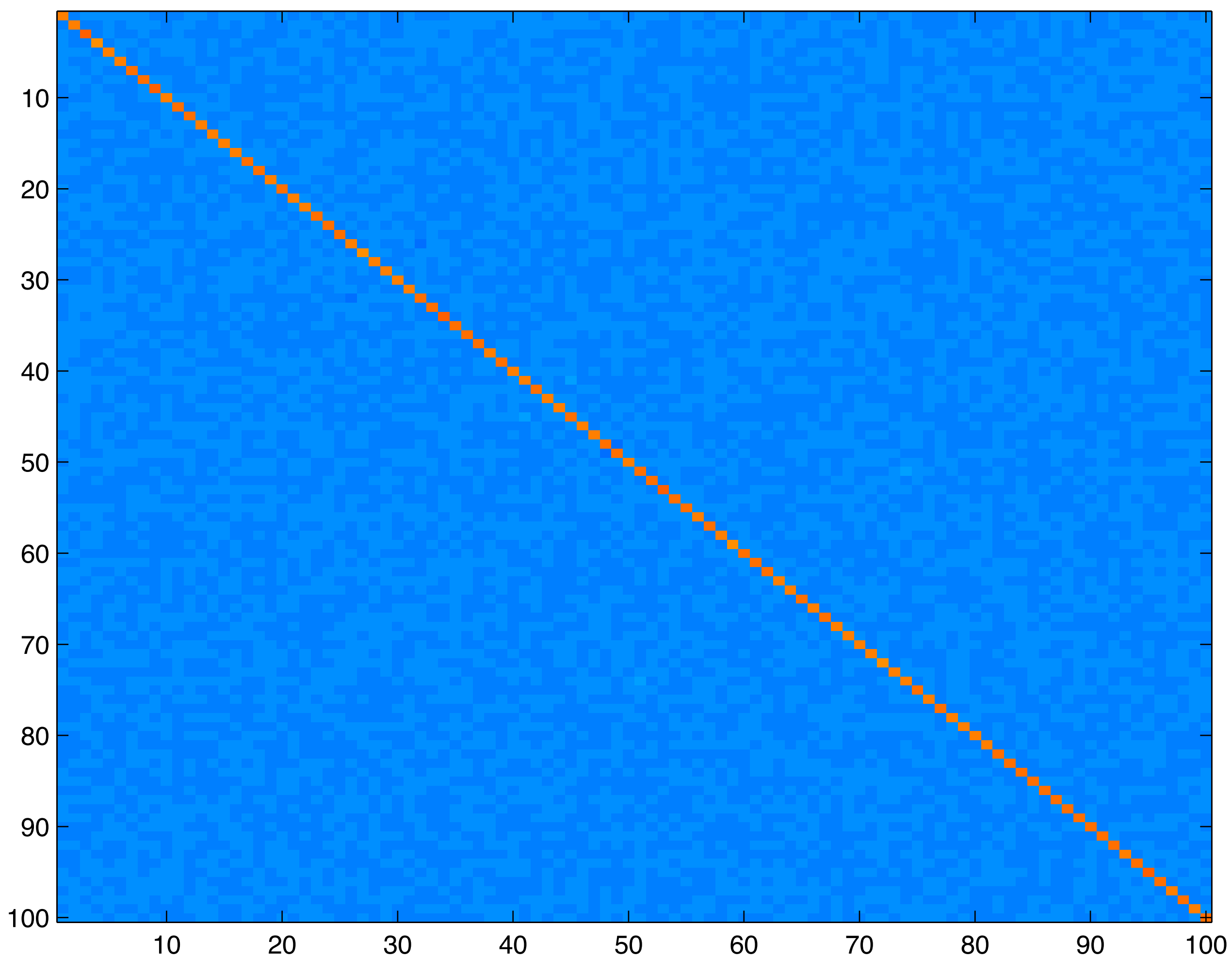
K=1000

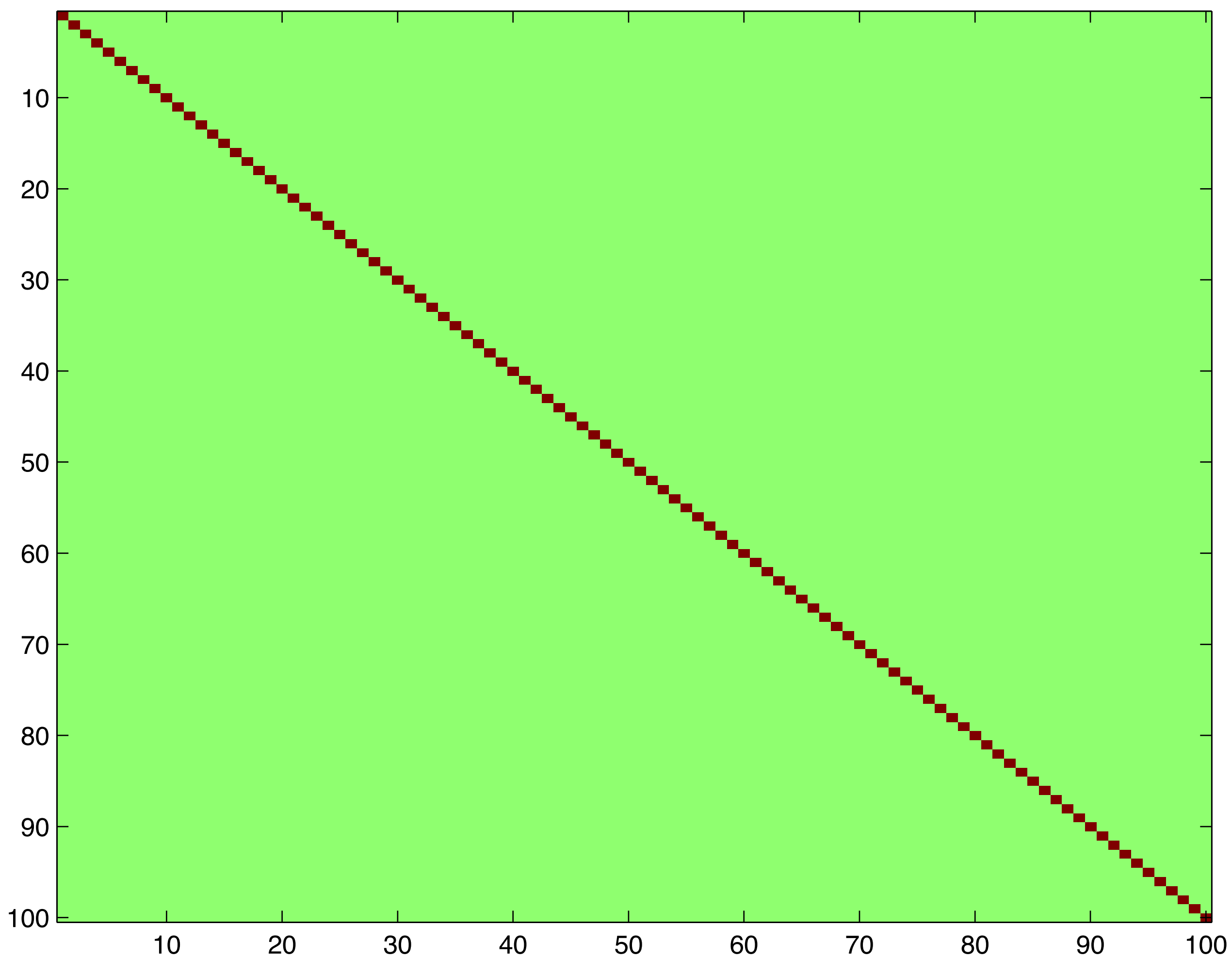


Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

K=10000

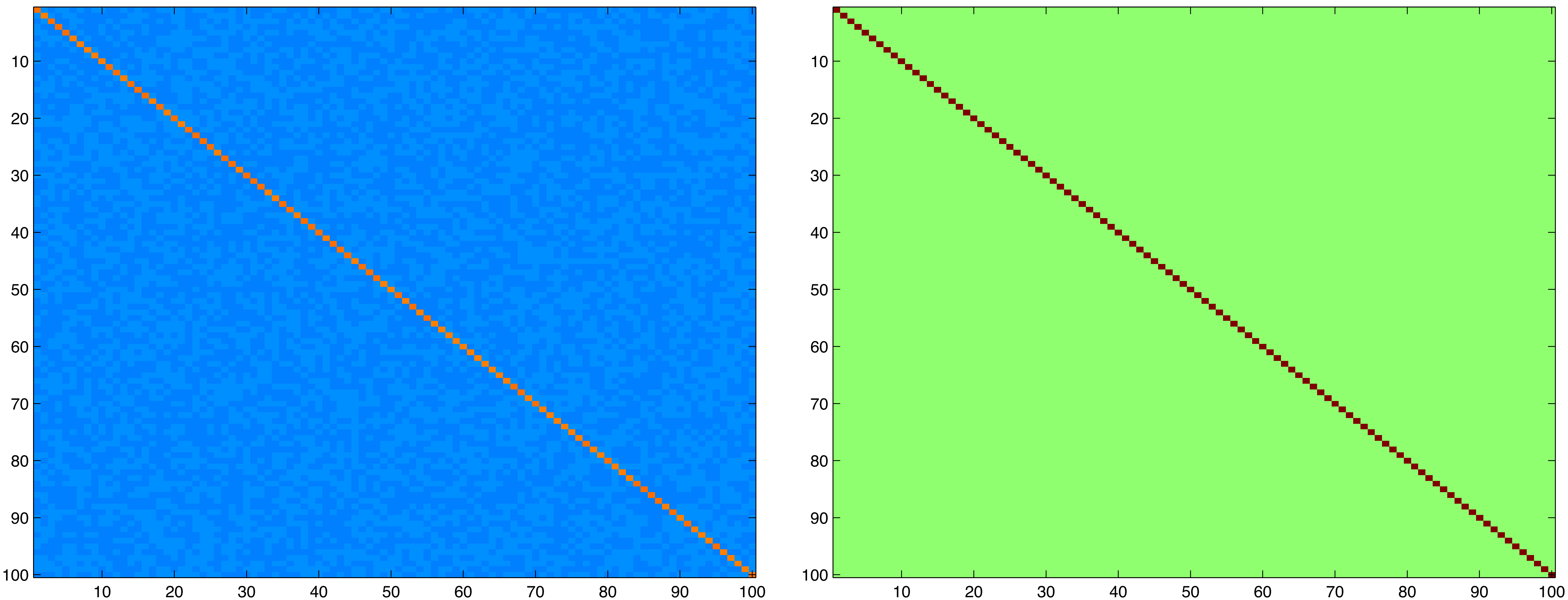




Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

K=100000



Conventional optimization

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \gamma_k \mathbf{s}_k$$
$$\mathbf{s}_k = -H_k^{-1}\left(\frac{1}{K}\sum_{i=0}^{K-1}\nabla\phi_i[\mathbf{m}_k]\right)$$

- **cost per iteration:** $\mathcal{O}(K)$
- **convergence rate:**

$$|\Phi[\mathbf{m}_*] - \Phi[\mathbf{m}_k]| = \mathcal{O}(c^k), \quad 0 < c \leq 1$$

SLIM

Stochastic optimization

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \gamma_k \mathbf{s}_k$$

$$\mathbf{s}_k = -\nabla\phi_i[\mathbf{m}_k], \quad i \sim U[0, K-1]$$

- **assumption:** $\mathsf{E}\{\mathbf{s}_k\} = -\nabla\Phi[\mathbf{m}_k]$
- **cost per iteration:** $\mathcal{O}(1)$
- **convergence rate:**

$$|\Phi[\mathbf{m}_*] - \Phi[\mathbf{m}_k]| = \mathcal{O}(1/k)$$

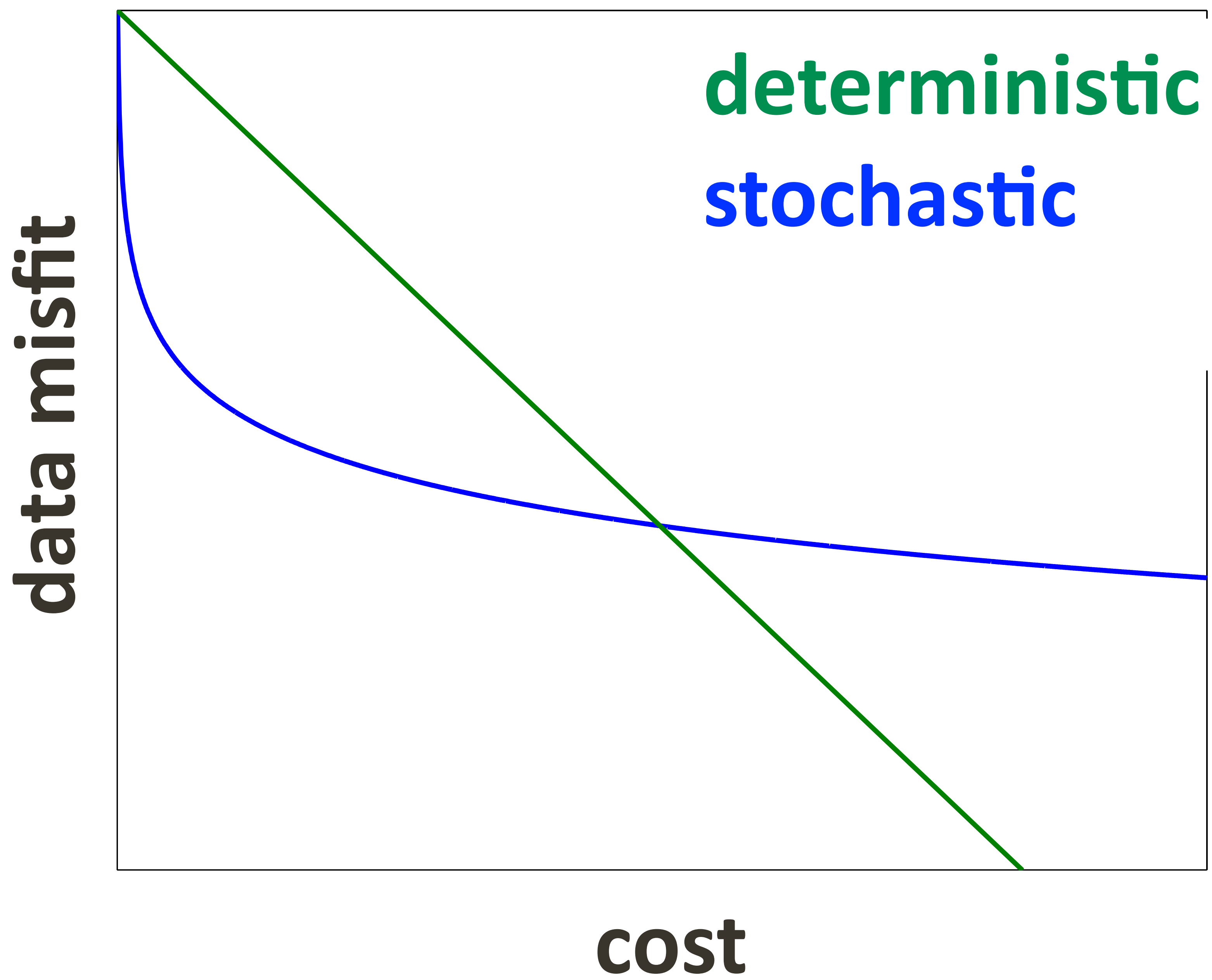
[Robbins et al. '50; Bertsekas et. al '96]

SLIM

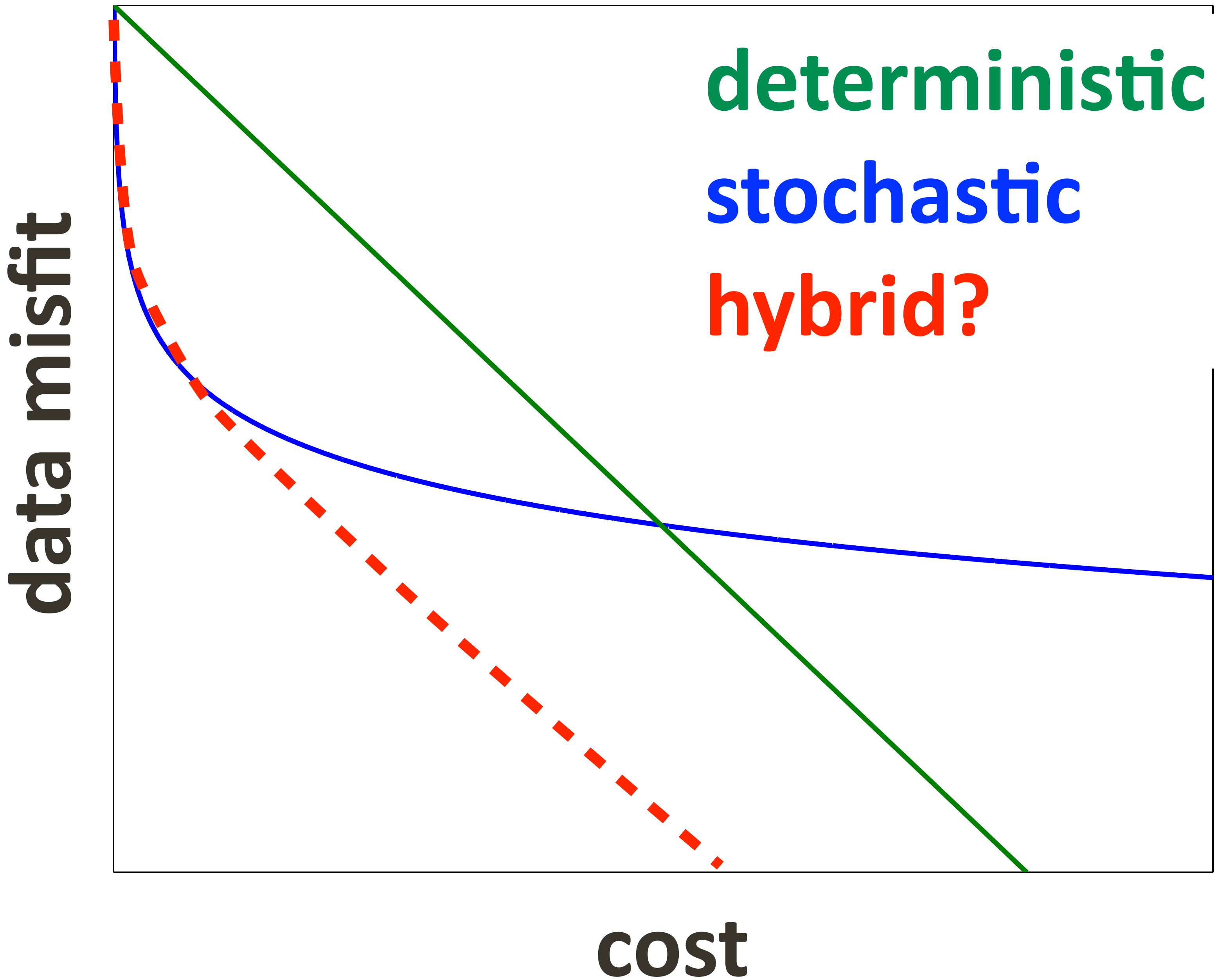
Stochastic optimization

- cheap iterations
- only for encoded data
- sample data with replacement
- slow convergence (relies on law of large numbers)

Stochastic vs. Deterministic



Stochastic vs. Deterministic



SLIM

Hybrid method

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \gamma_k \mathbf{s}_k$$
$$\mathbf{s}_k = -H_k^{-1}\left(\frac{1}{|B_k|}\sum_{i\in B_k}\nabla\phi_i[\mathbf{m}_k]\right)$$

- **assumption:** $|||B_k|^{-1}\sum_{i\in B_k}\nabla\phi_i[\mathbf{m}_k] - \nabla\Phi[\mathbf{m}_k]||_2^2 < C_k$
- **cost per iteration:** $\mathcal{O}(|B_k|)$
- **convergence rate, assuming** $C_k \downarrow 0$

$$|\Phi[\mathbf{m}_*] - \Phi[\mathbf{m}_k]| = \mathcal{O}(c^k), \quad 0 < c \leq 1$$

[Friedlander et. al ’11]

SLIM

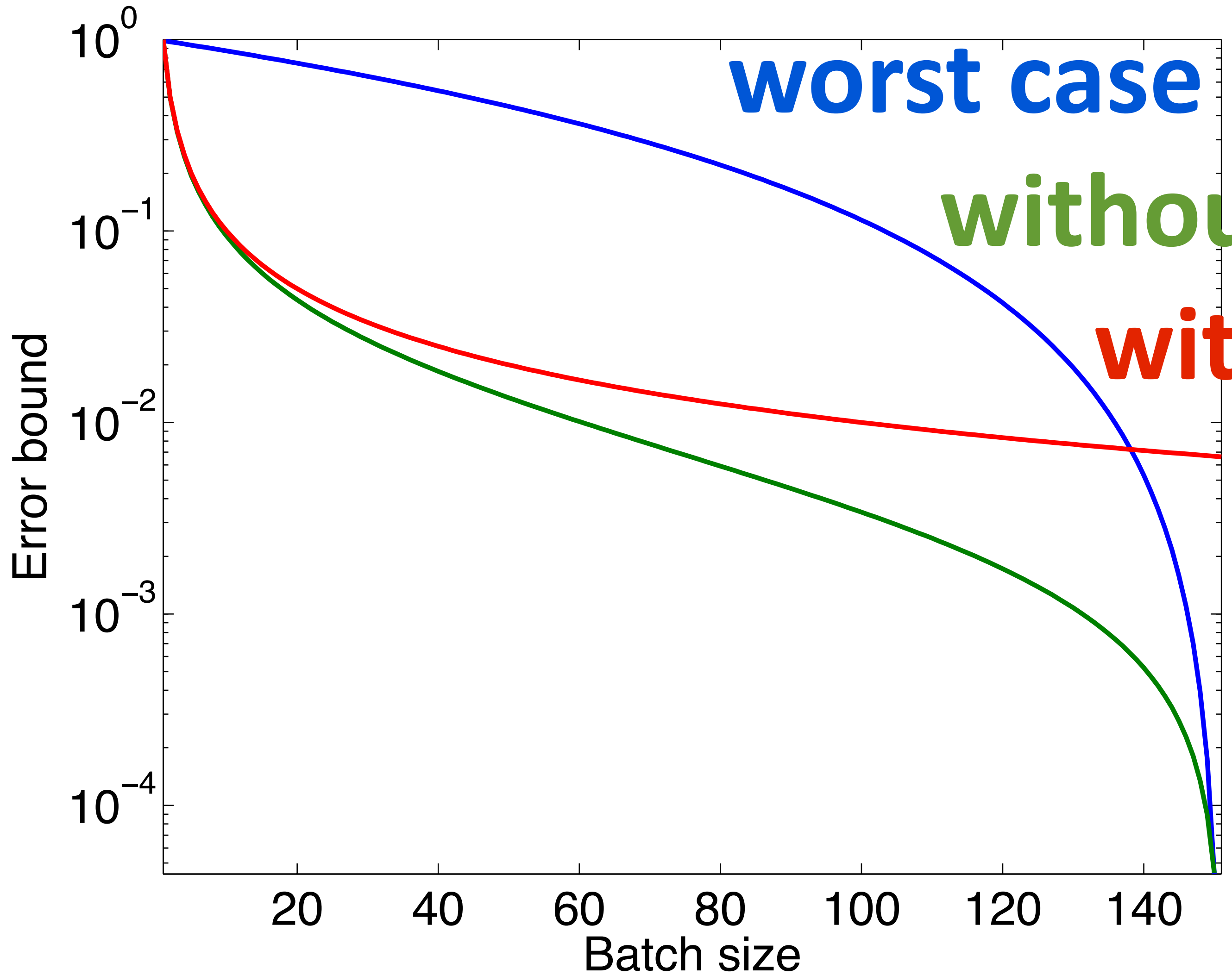
Batching strategy

We can grow the batch according to the natural order, or by random selection without replacement.

The SO strategy relies on drawing with replacement.

Batching strategy

Batching strategy controls theoretical decay of the error.



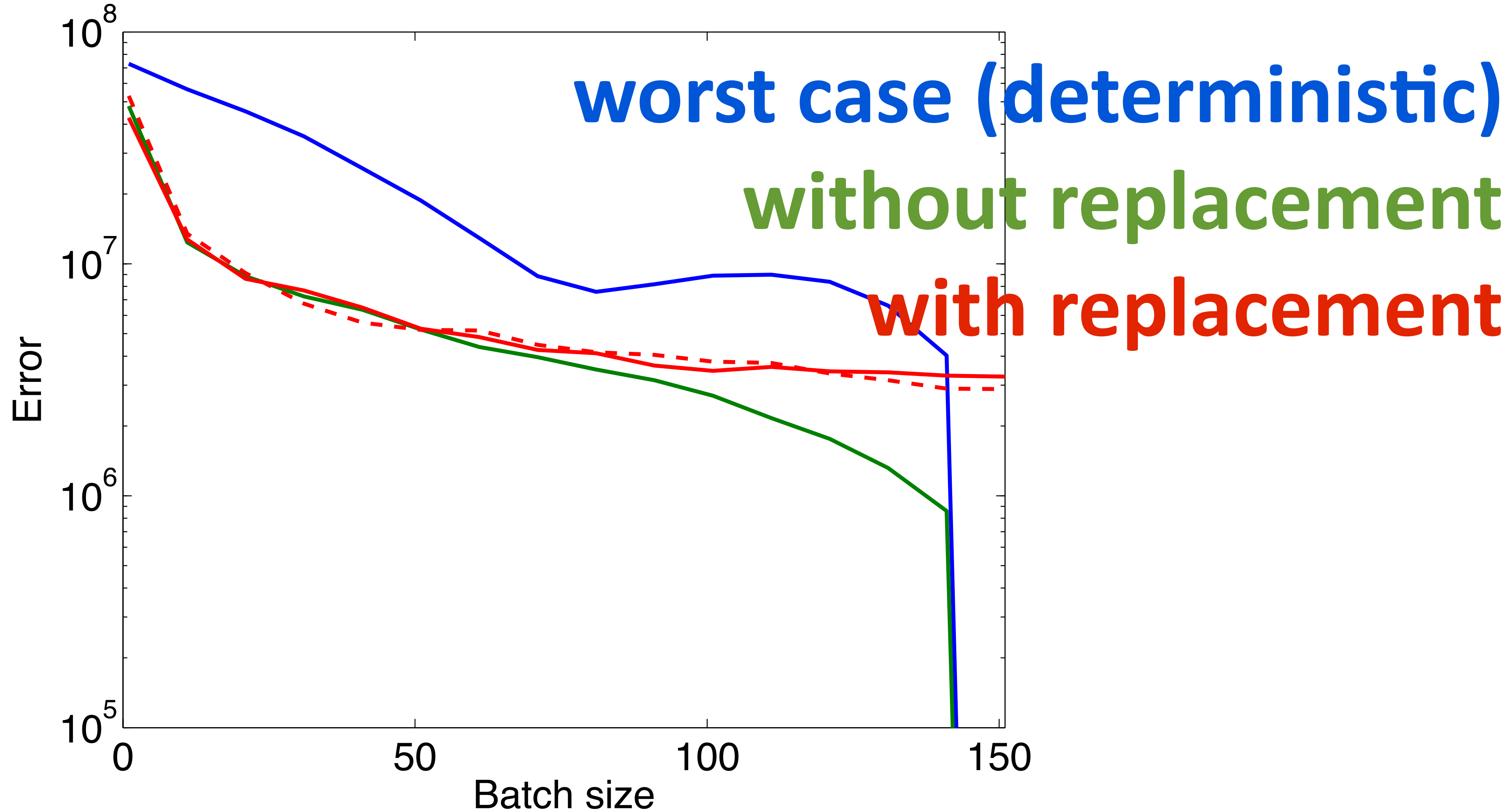
worst case (deterministic)

without replacement

with replacement

Batching strategy

Actual decay of the error



Hybrid method

while not converged **do**

$\mathbf{g}_k = \sum_{i \in B_k} \nabla \phi_i[\mathbf{m}_k]$ gradient

$\mathbf{s}_k = -H_k^{-1} \mathbf{g}_k$ search dir.

find γ s.t. $\sum_{i \in B_k} \phi_i[\mathbf{m}_k + \gamma \mathbf{s}_k] < \sum_{i \in B_k} \phi_i[\mathbf{m}_k]$ linesearch

$\mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{s}_k$ update model

$B_{k+1} = B_k \bigcup \{i\}$ increase batchsize

end while

Hybrid method

- **Will work with any source encoding strategy (none, randomized, plane wave, eigenvectors, ...)**
- **Draw samples without replacement**
- **Random batching strategy is important to efficiently probe the data**

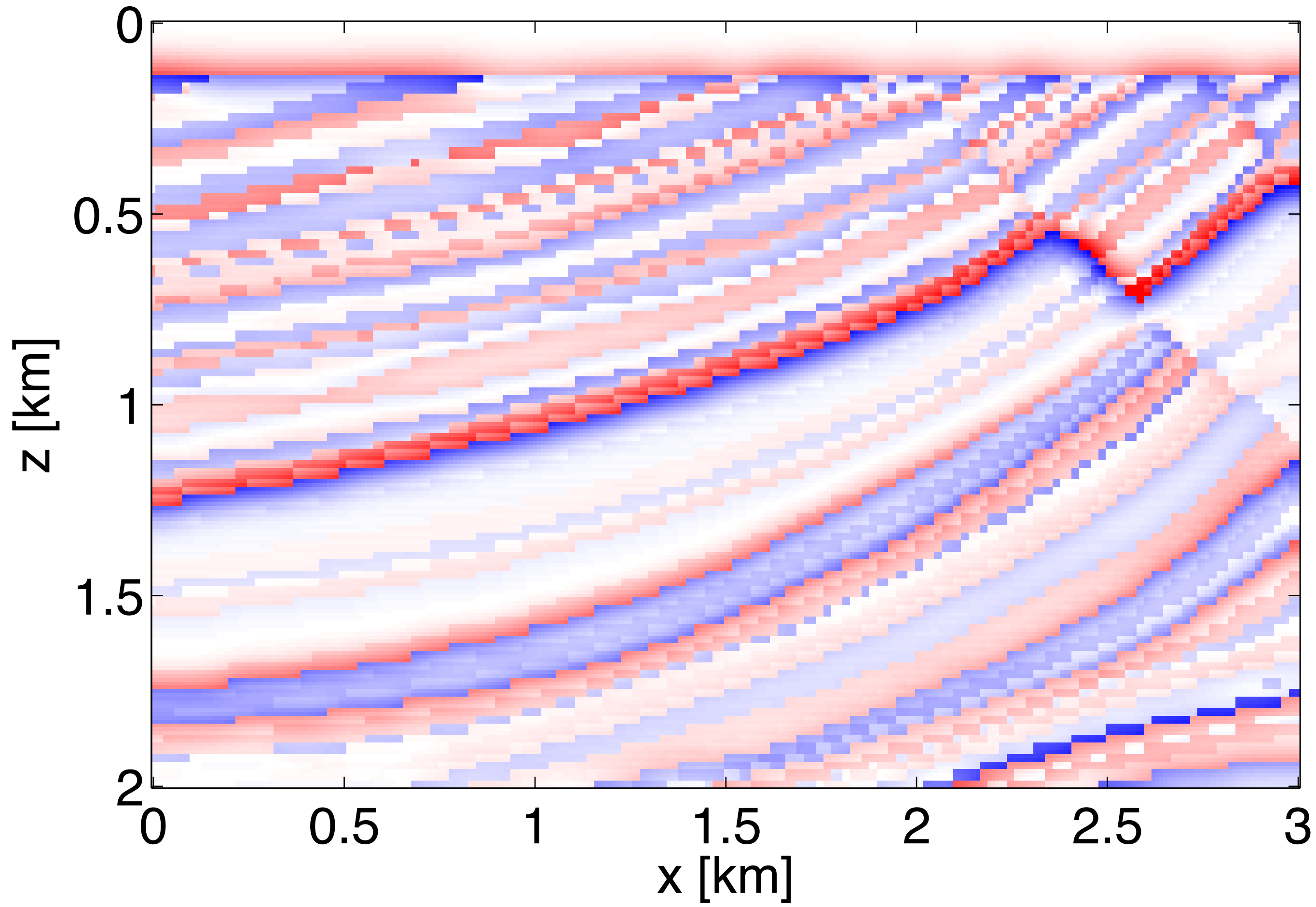
Results

Two scenario's :

- Non-linear migration
- multiscale FWI

Frequency domain FD with `full' acquisition

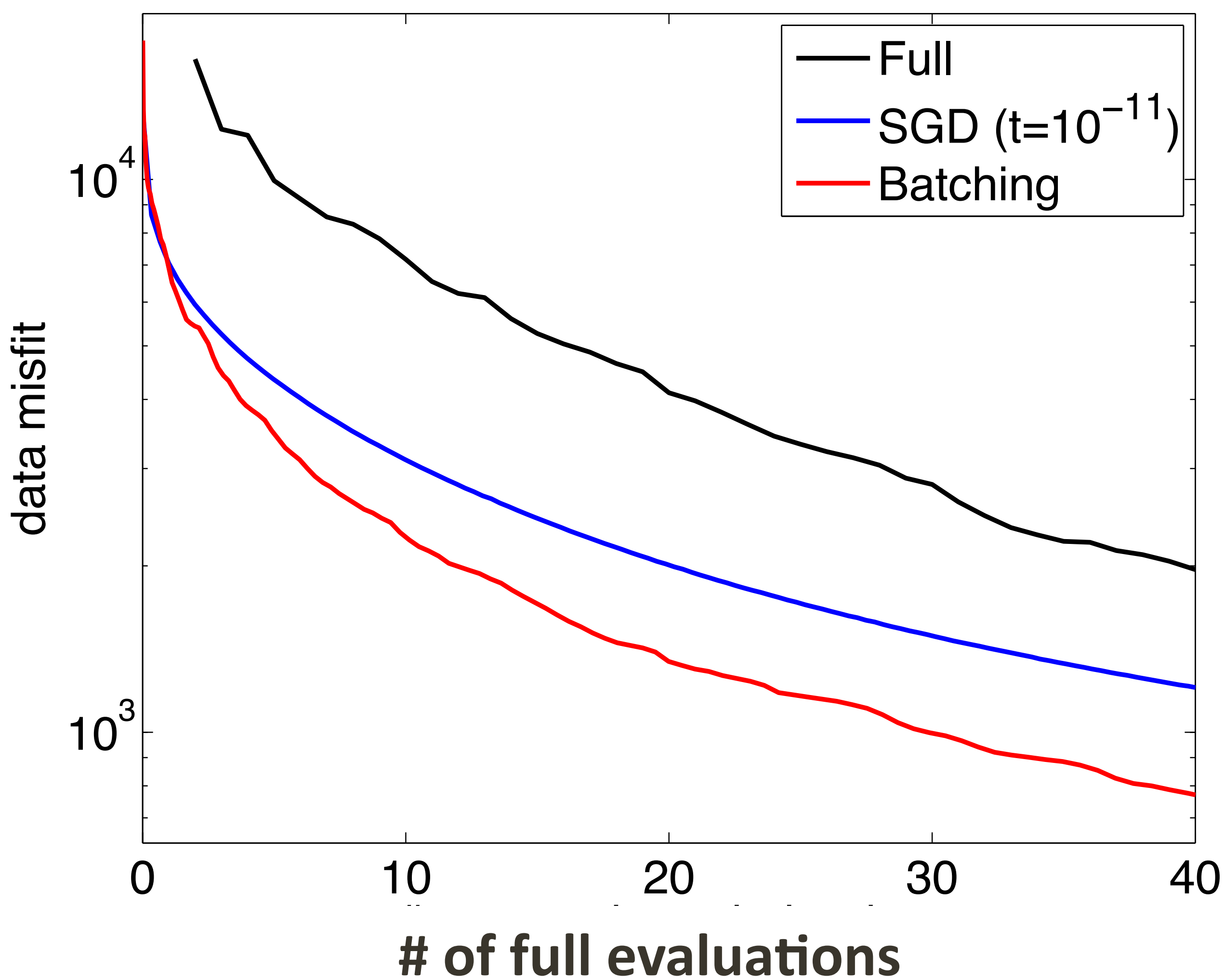
Non-linear migration

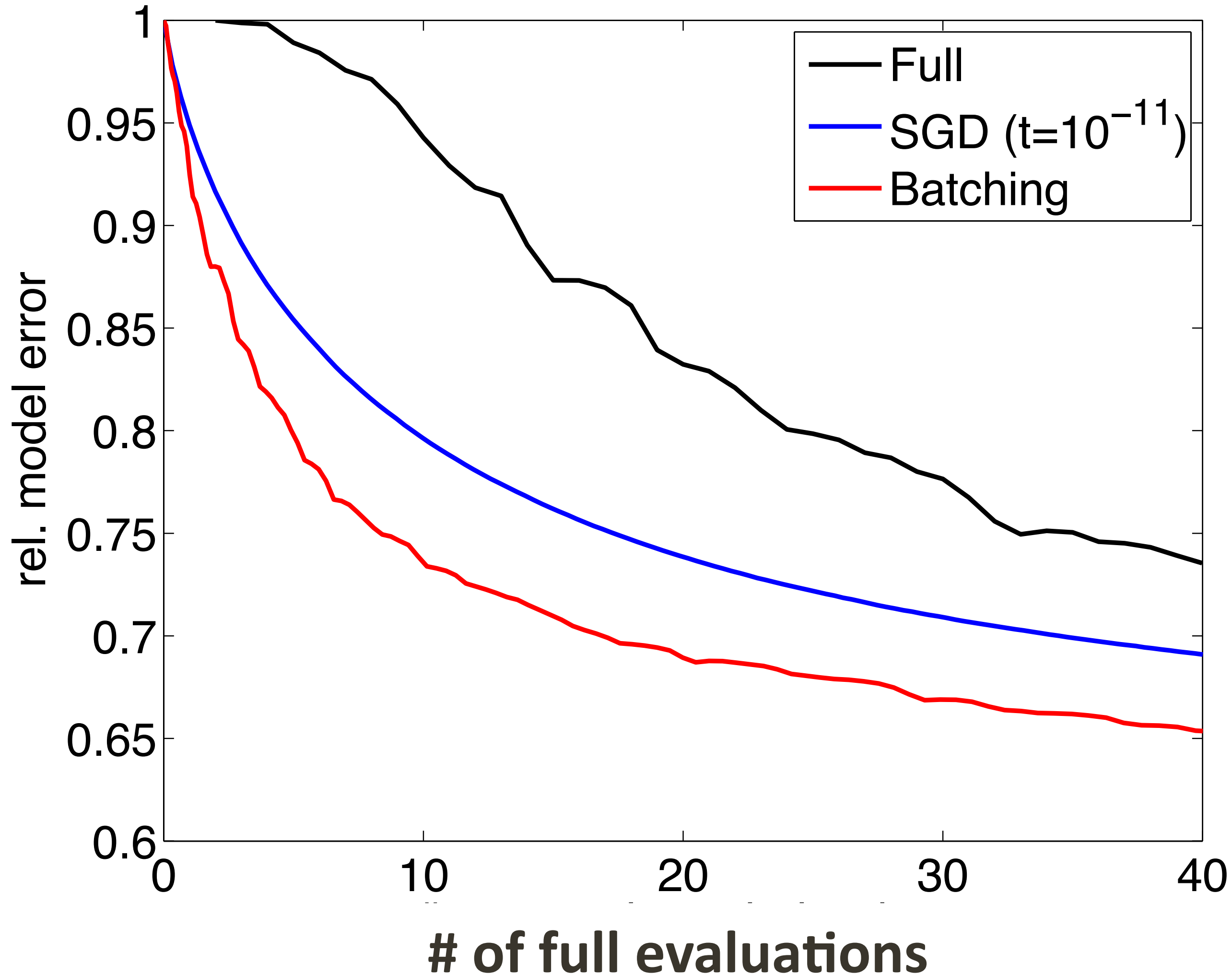


151 sources/receivers, frequencies [5-25] Hz

15 Hz Ricker

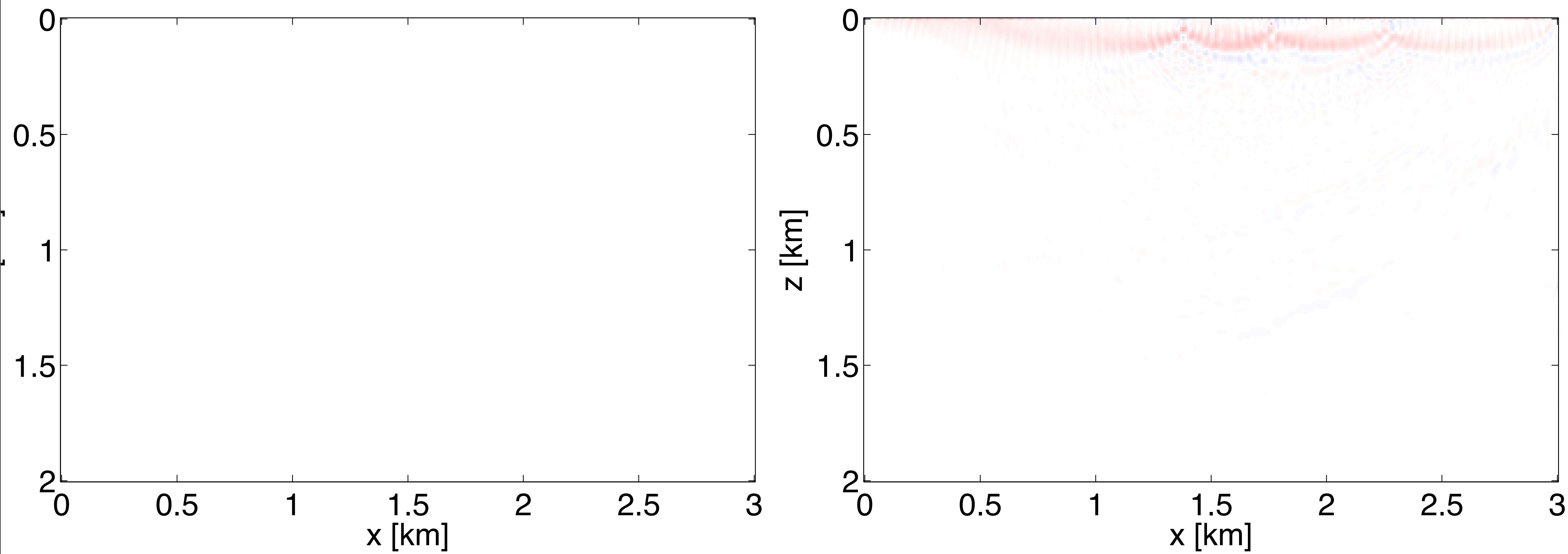
Non-linear migration





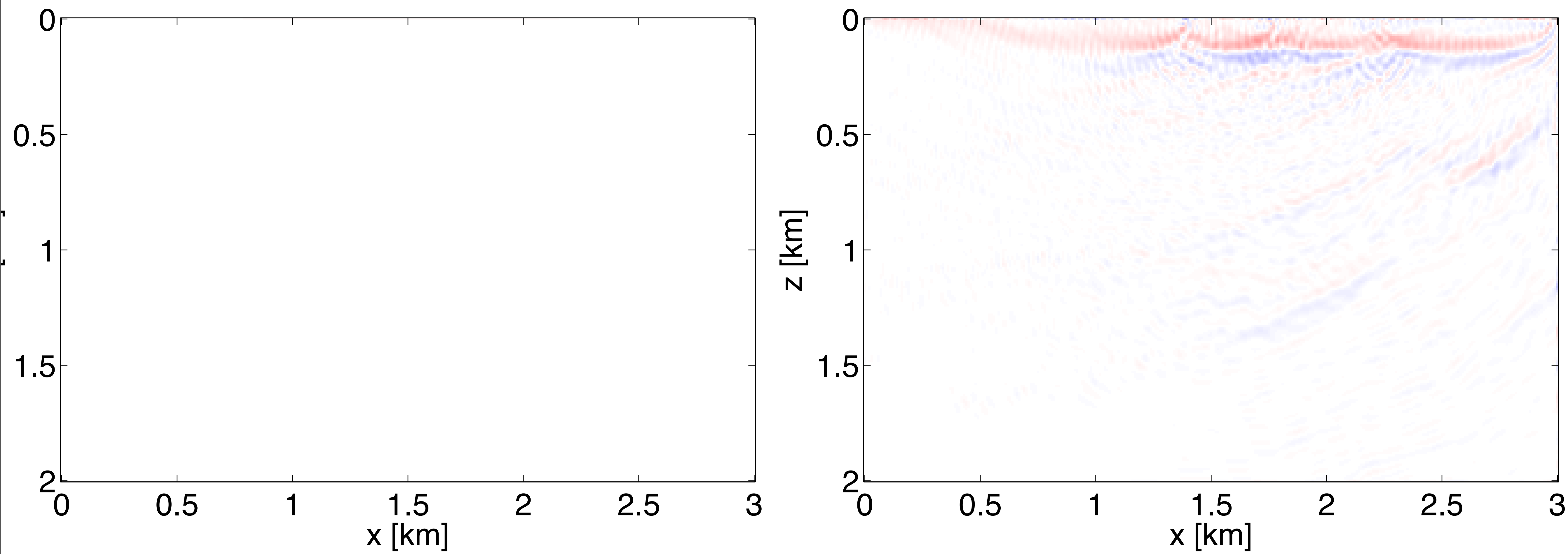
Non-linear migration

0.18 full evaluations



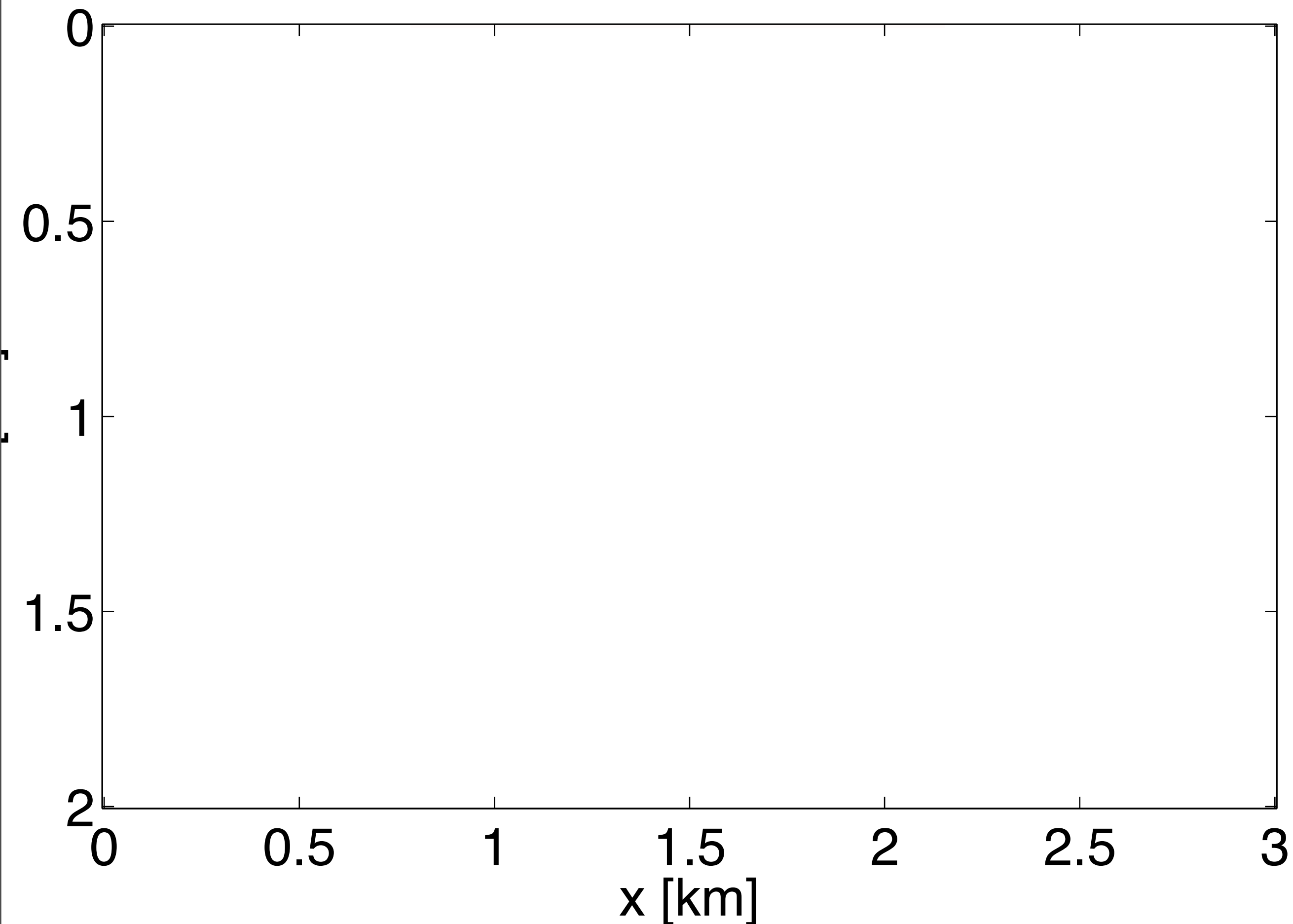
Non-linear migration

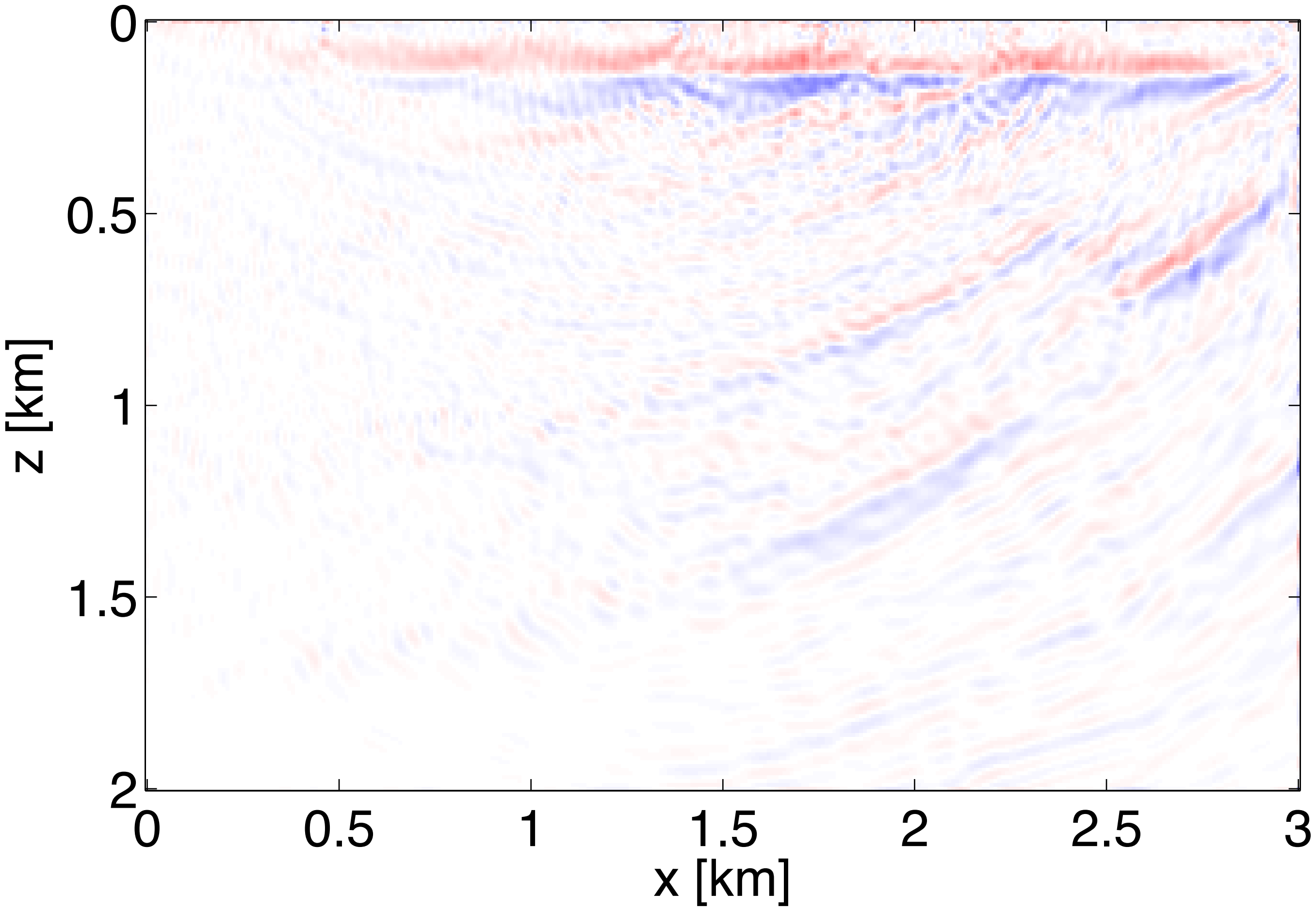
0.43 full evaluations



Non-linear migration

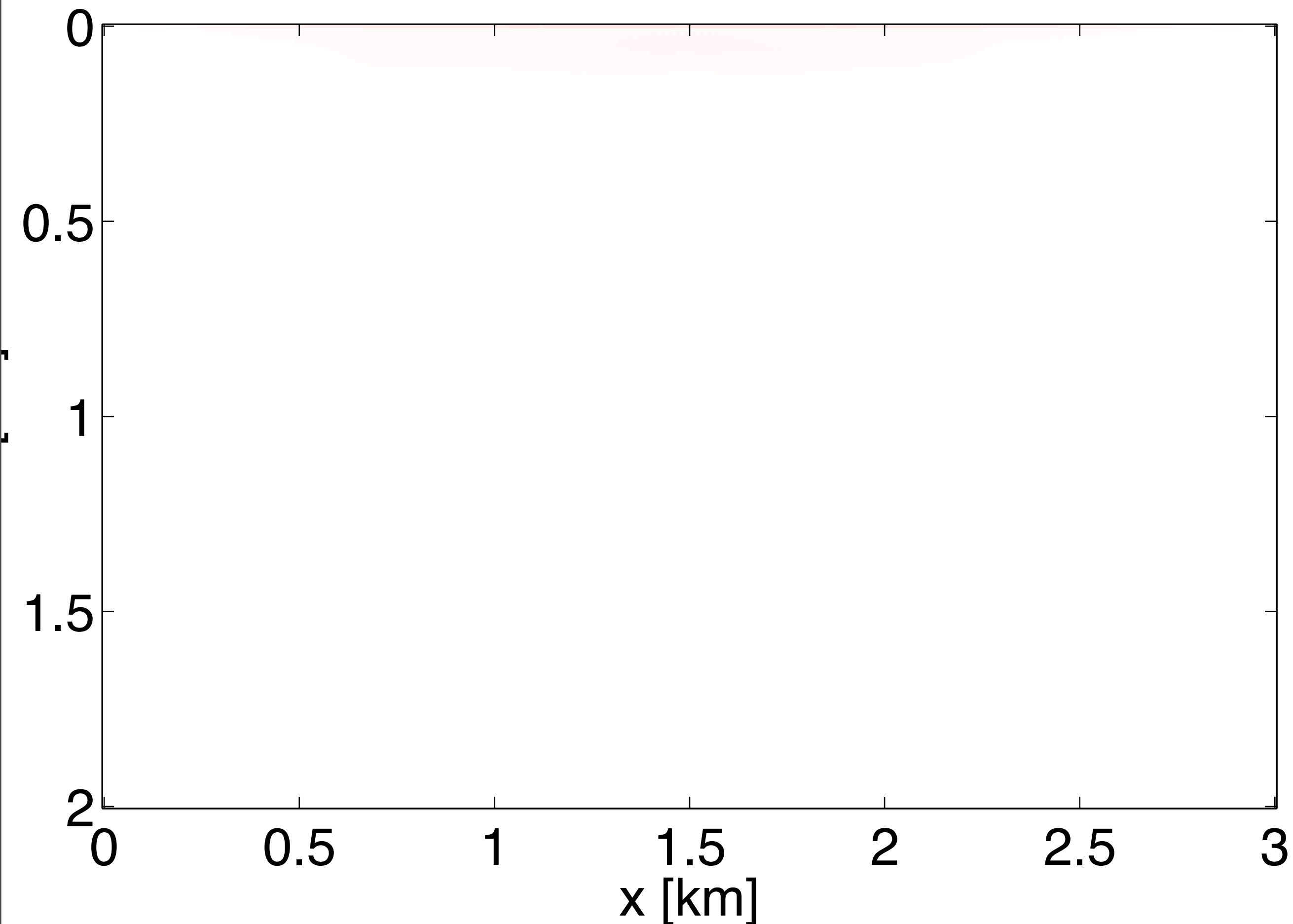
0.89 full evaluations

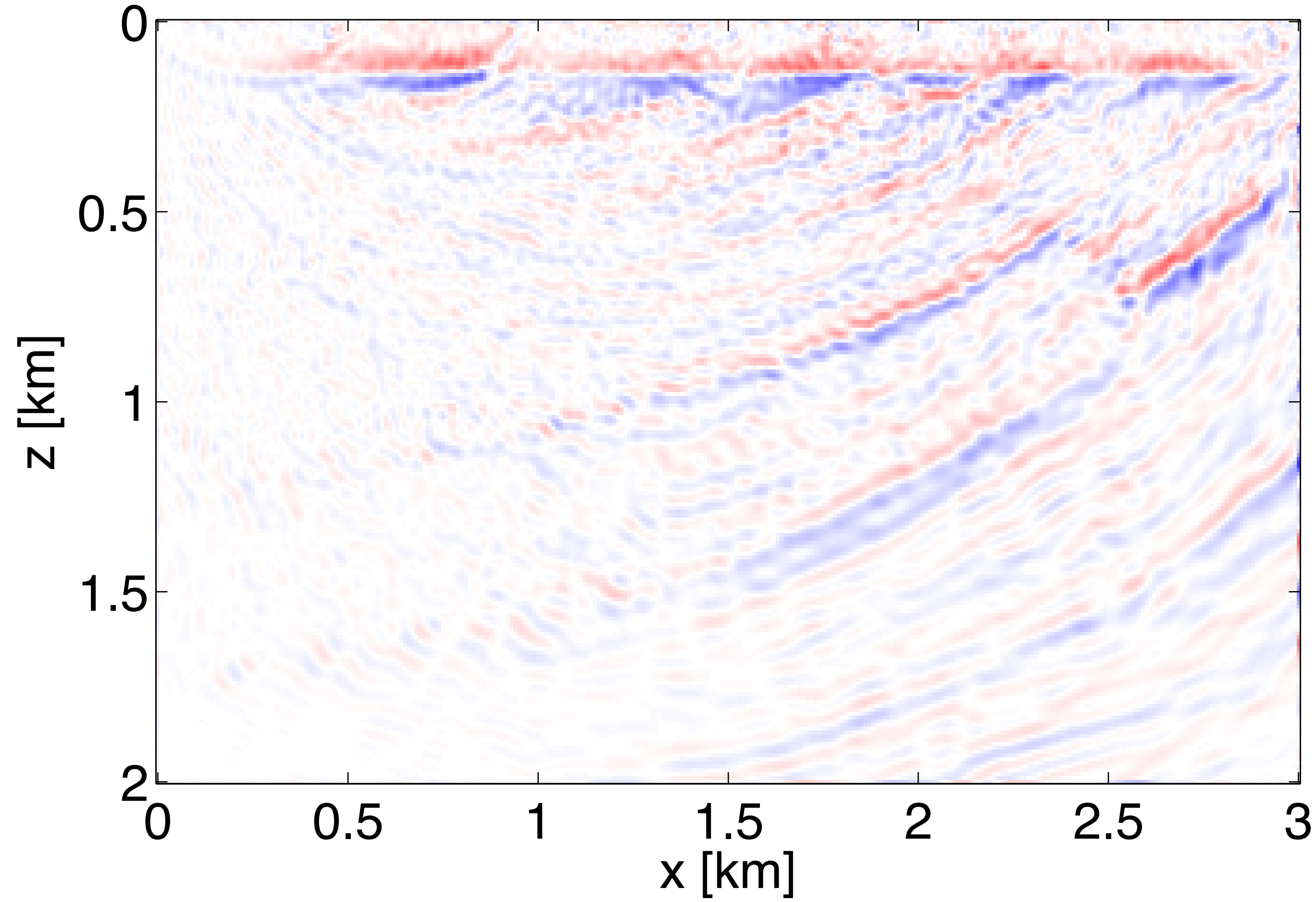




Non-linear migration

1.66 full evaluations

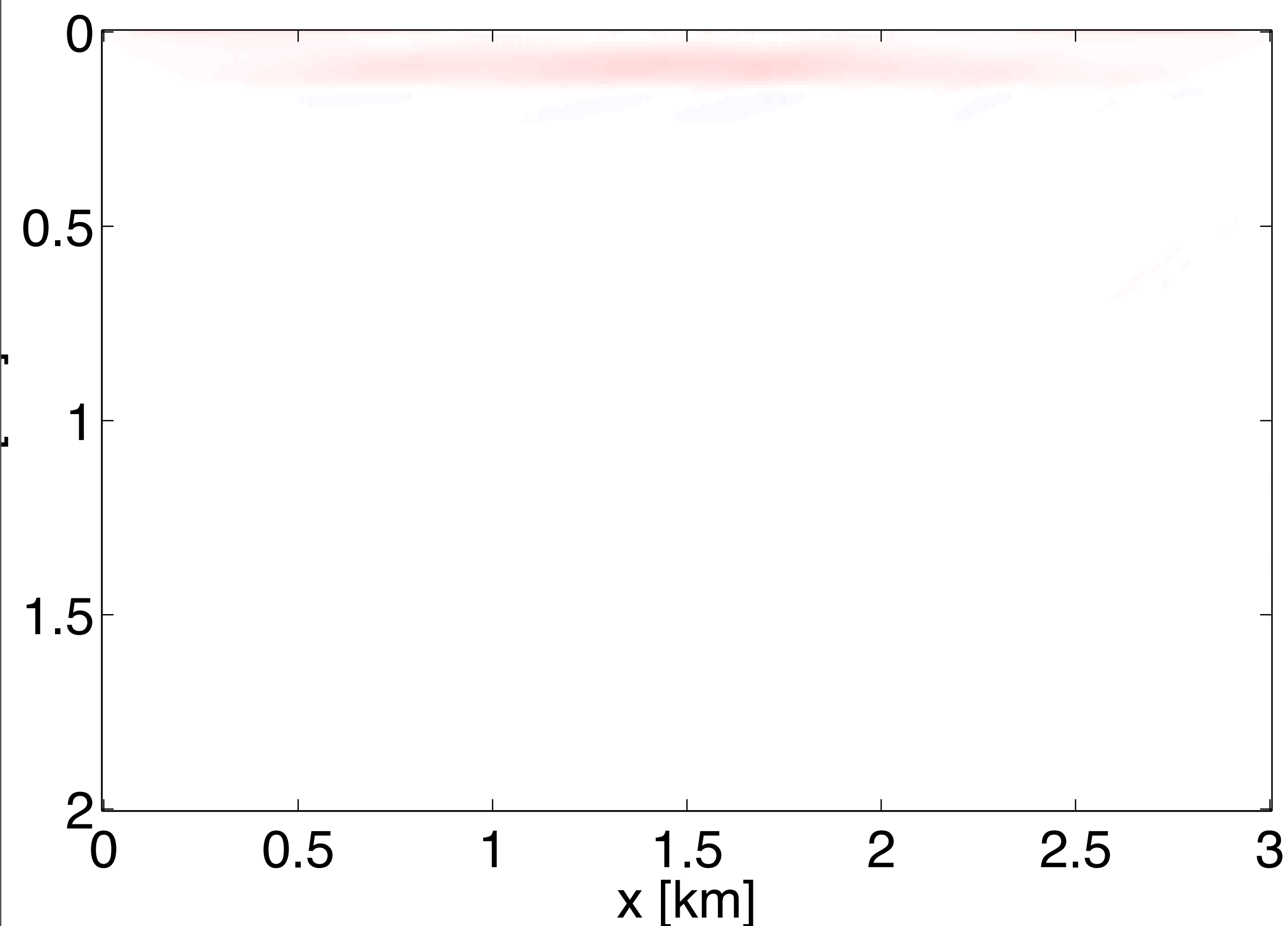


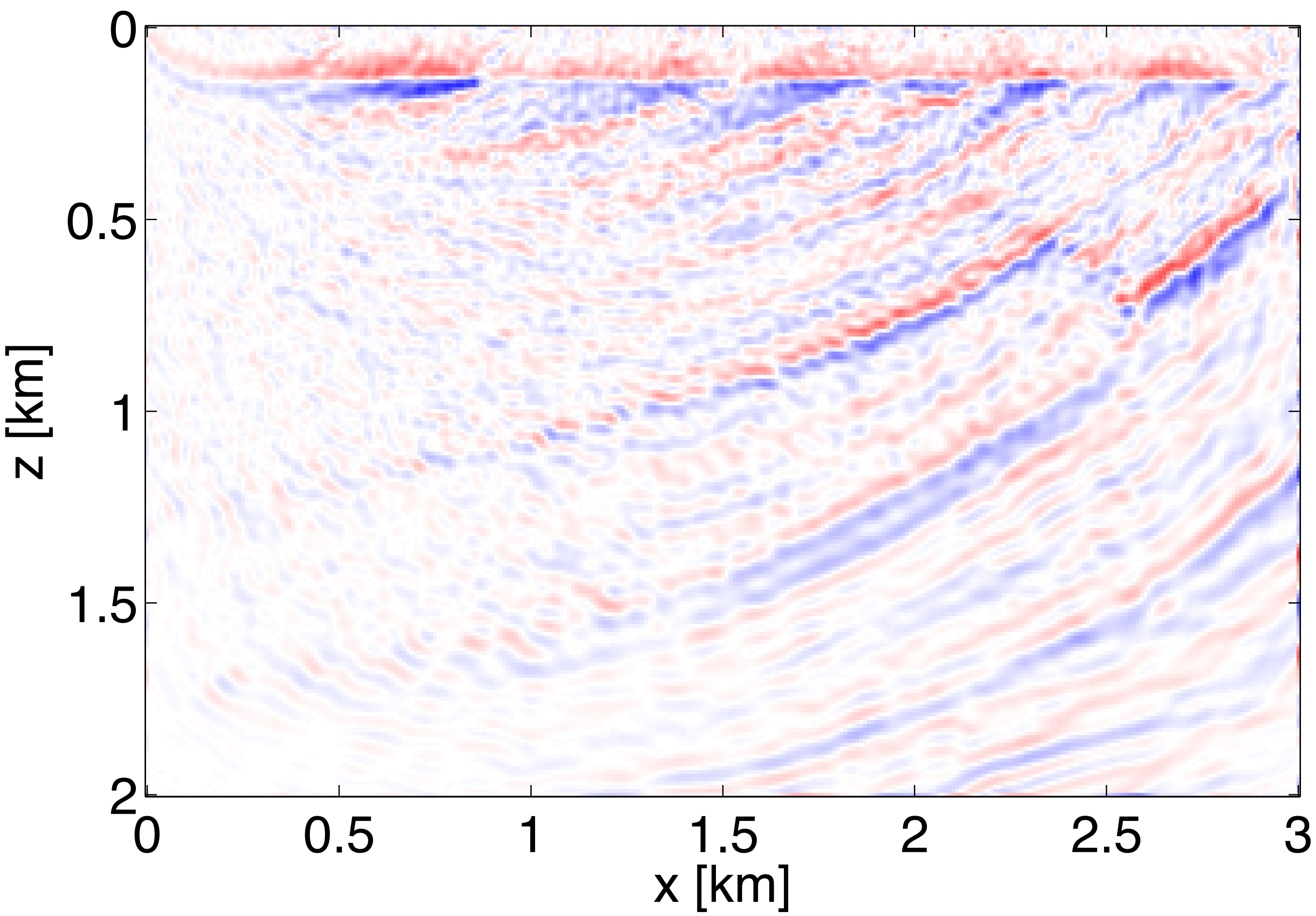


SLIM

Non-linear migration

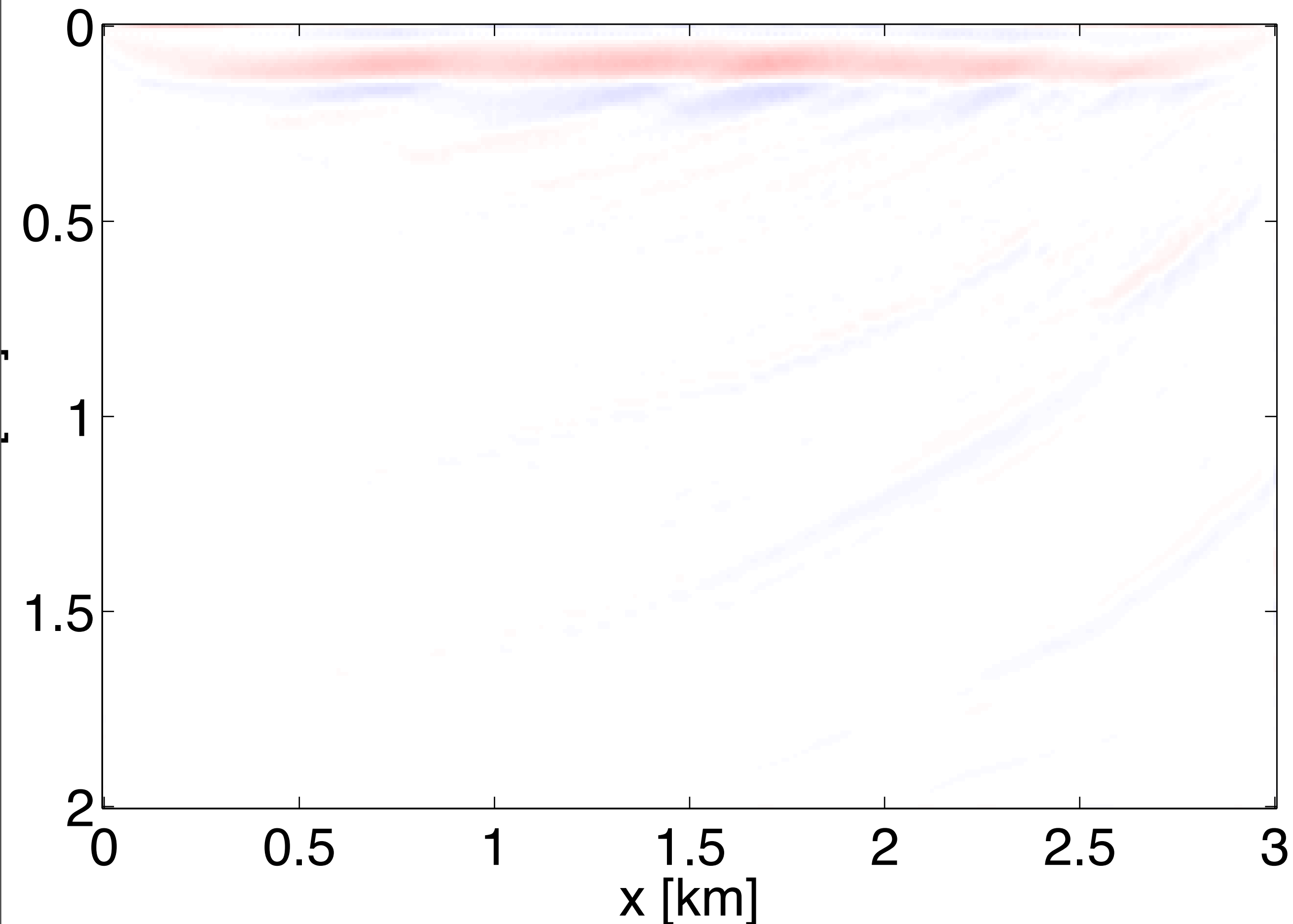
2.87 full evaluations

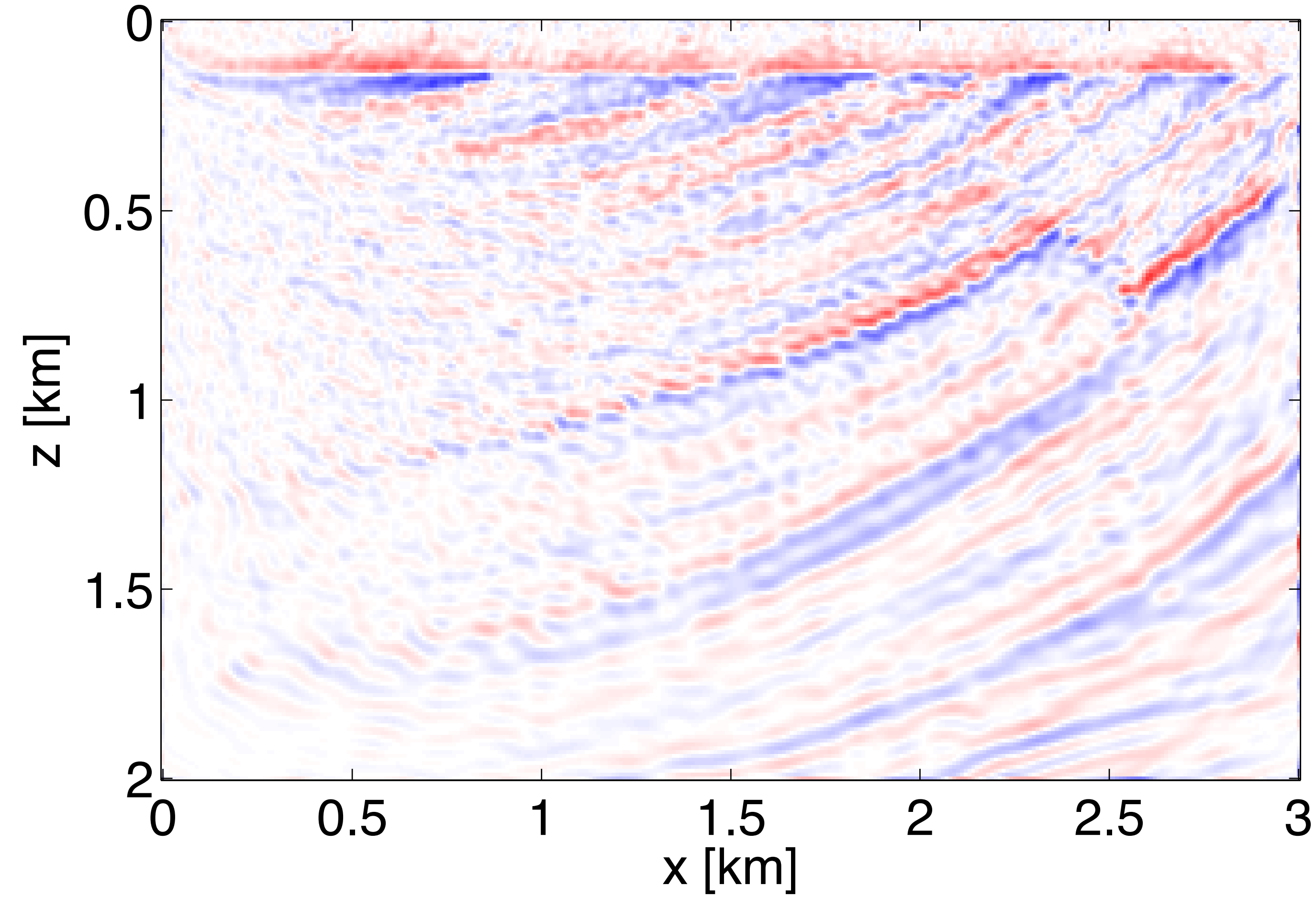




Non-linear migration

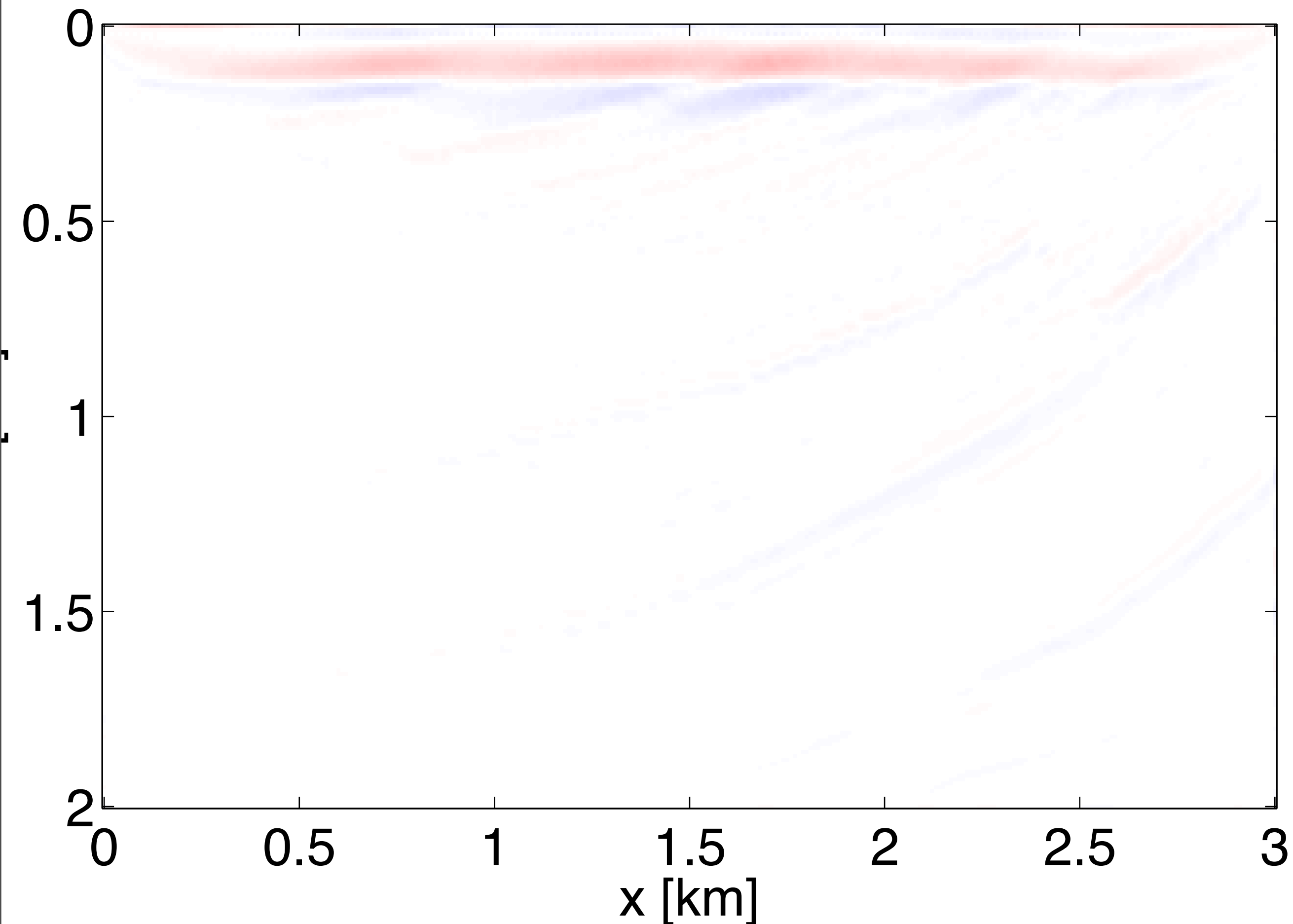
4.65 full evaluations

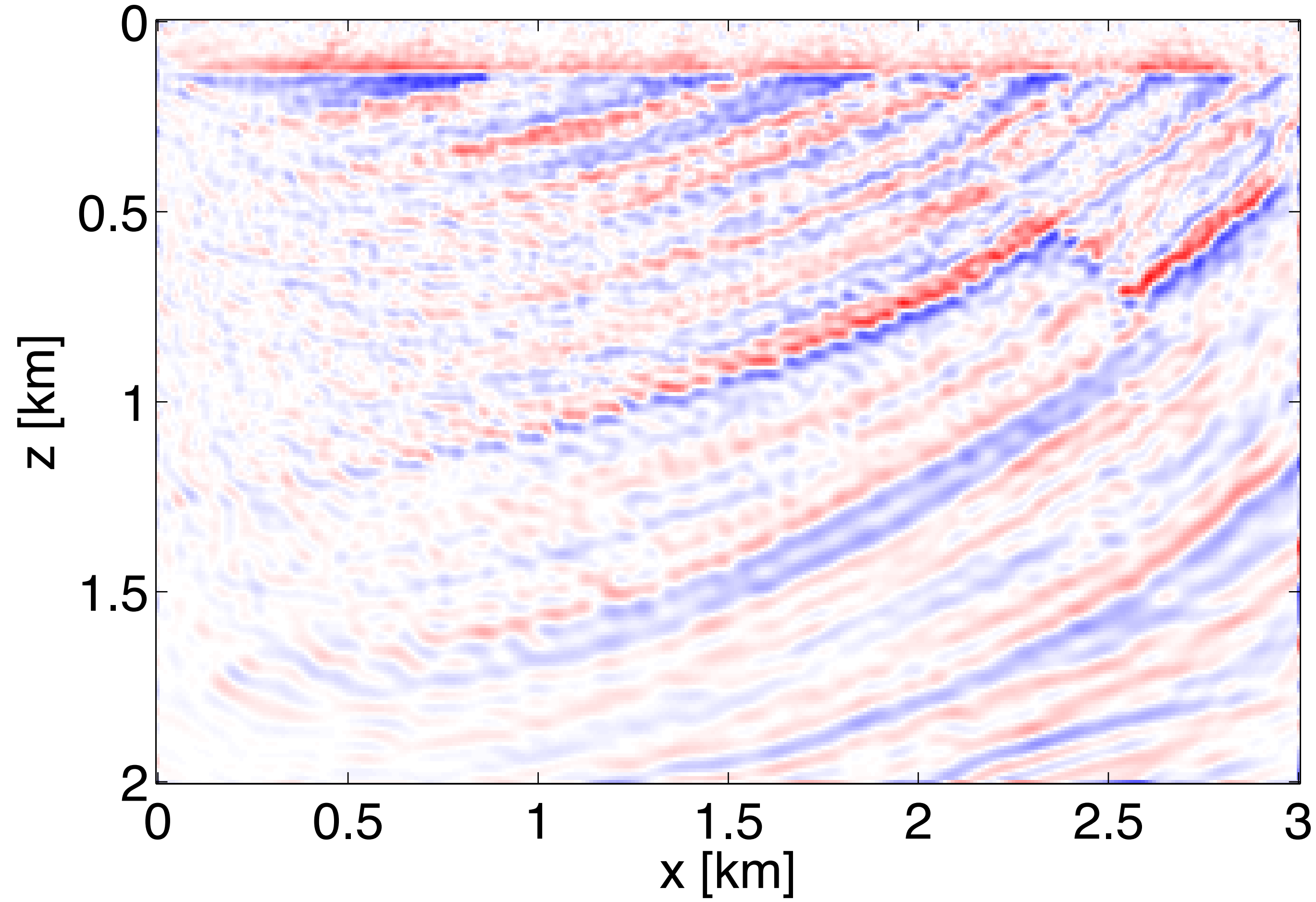




Non-linear migration

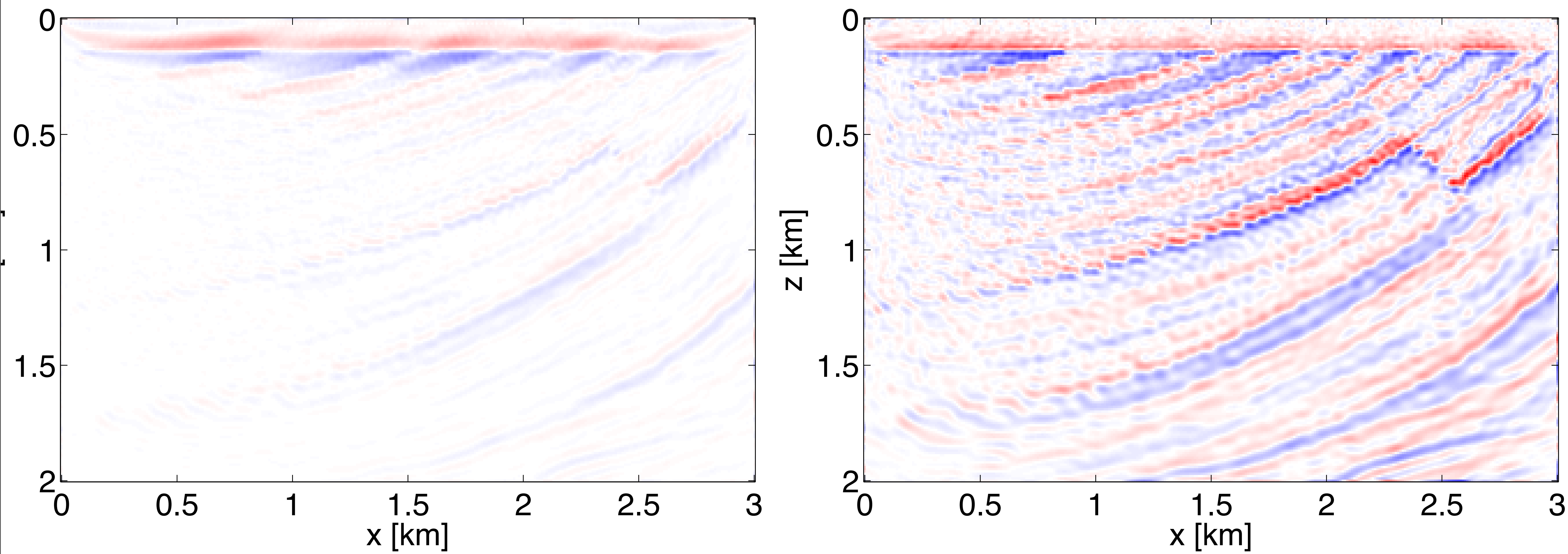
7.15 full evaluations





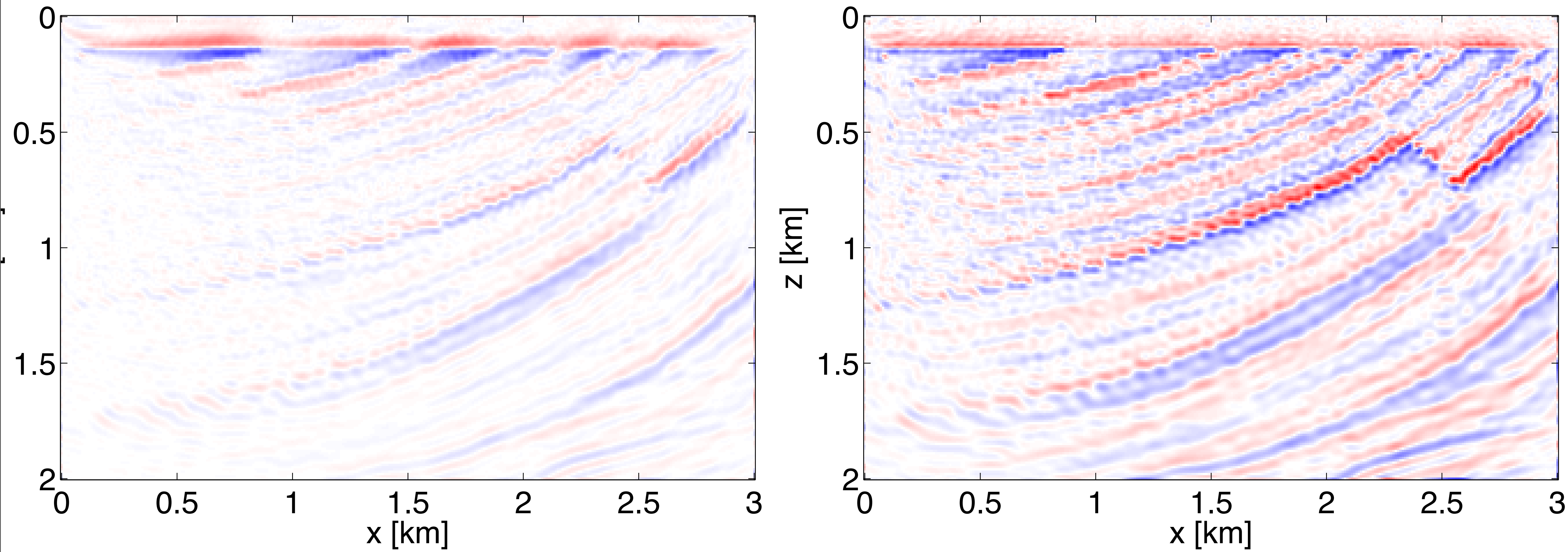
Non-linear migration

10.87 full evaluations



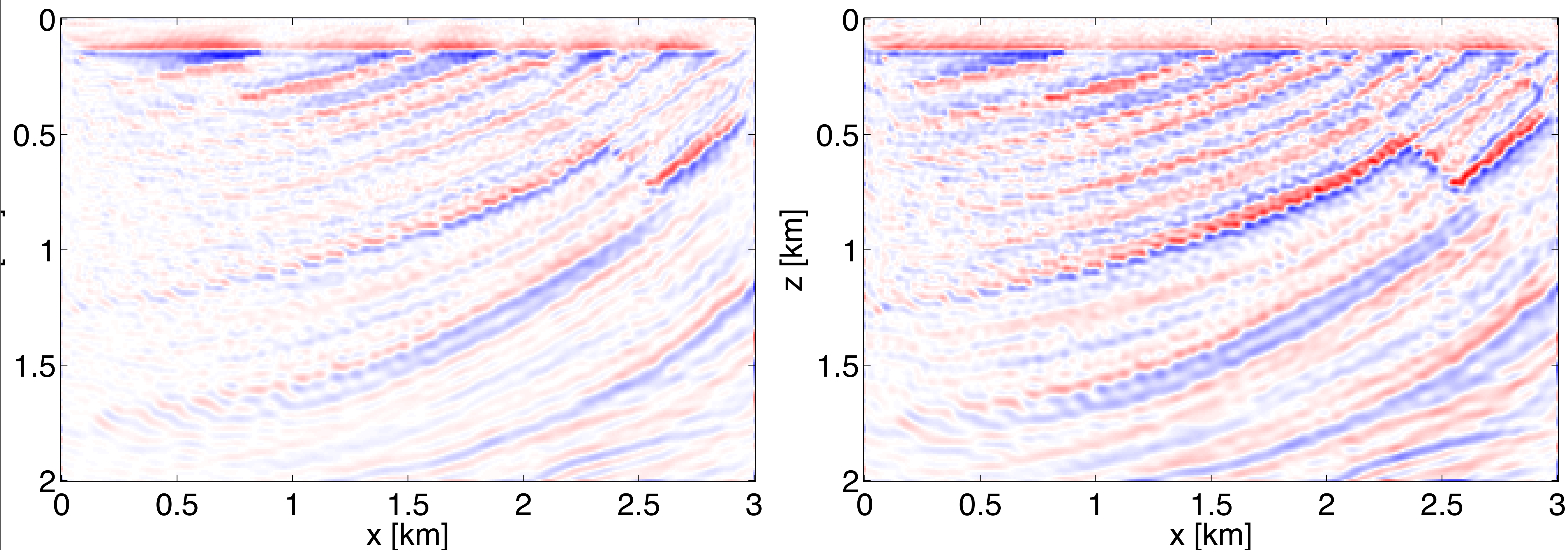
Non-linear migration

16.2 full evaluations



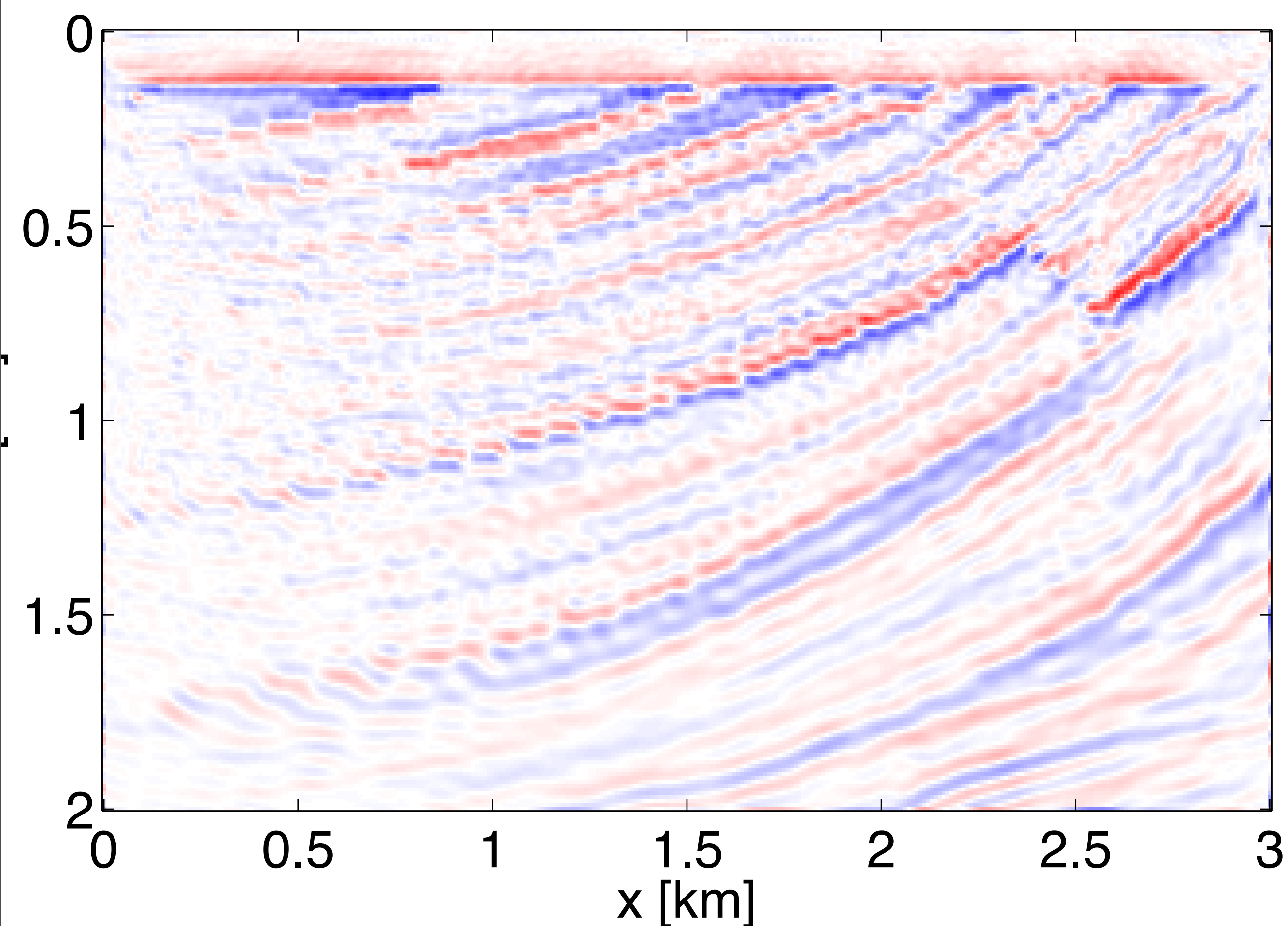
Non-linear migration

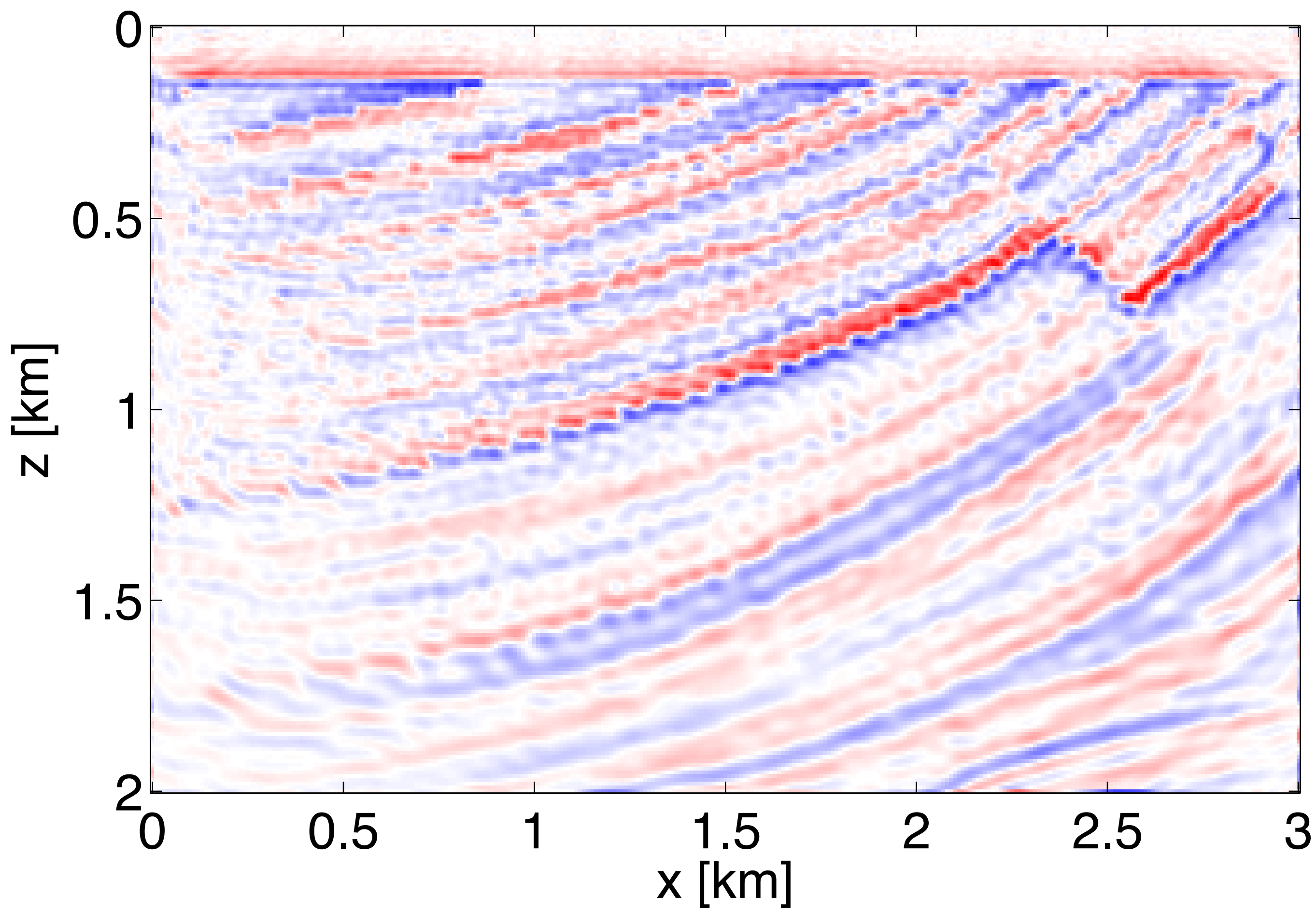
22 full evaluations



Non-linear migration

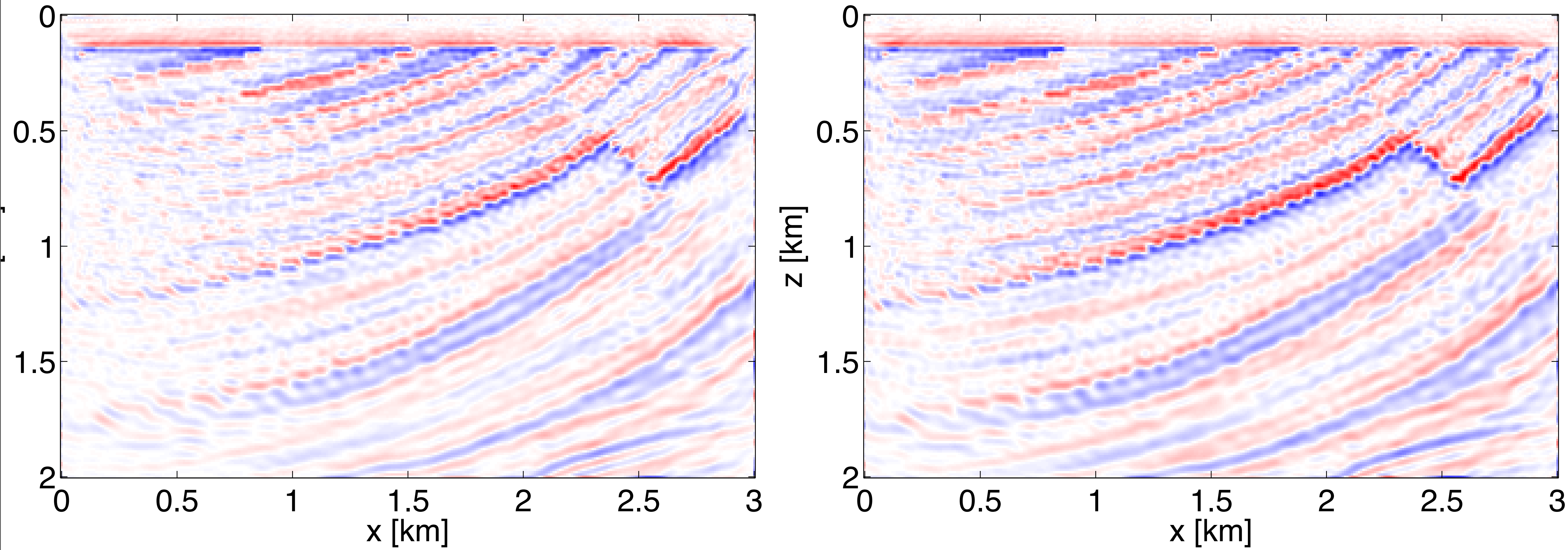
30.5 full evaluations





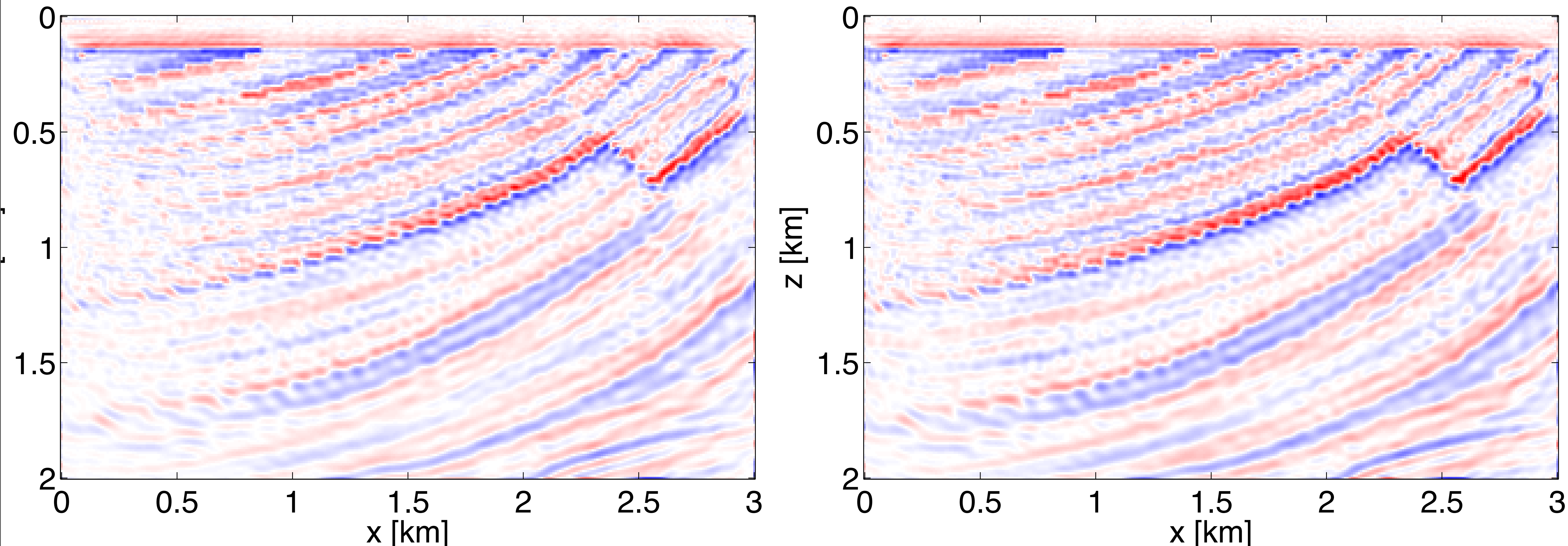
Non-linear migration

39.7 full evaluations



Non-linear migration

50.24 full evaluations

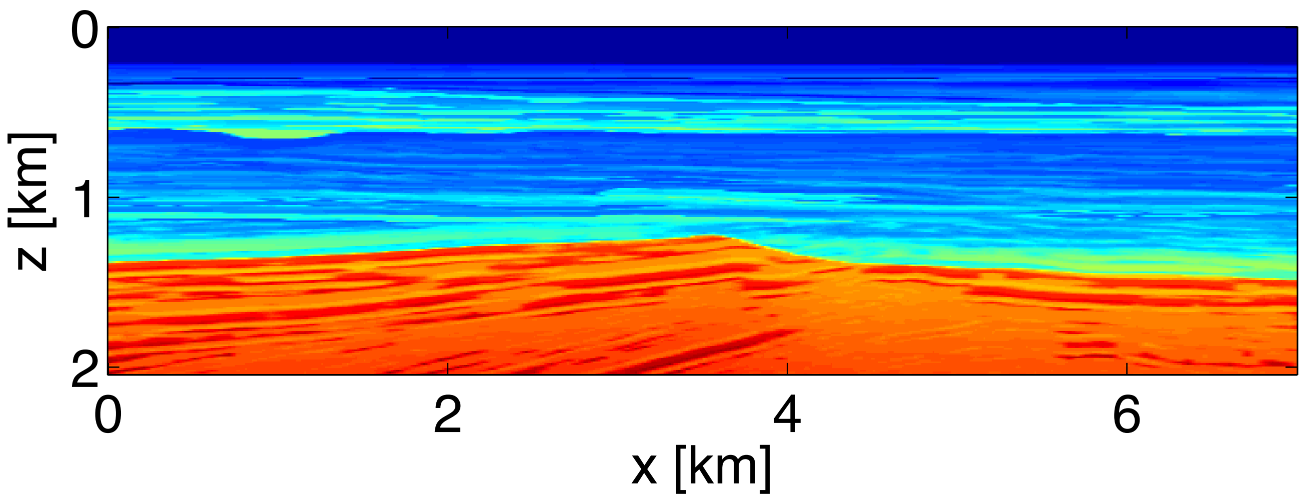


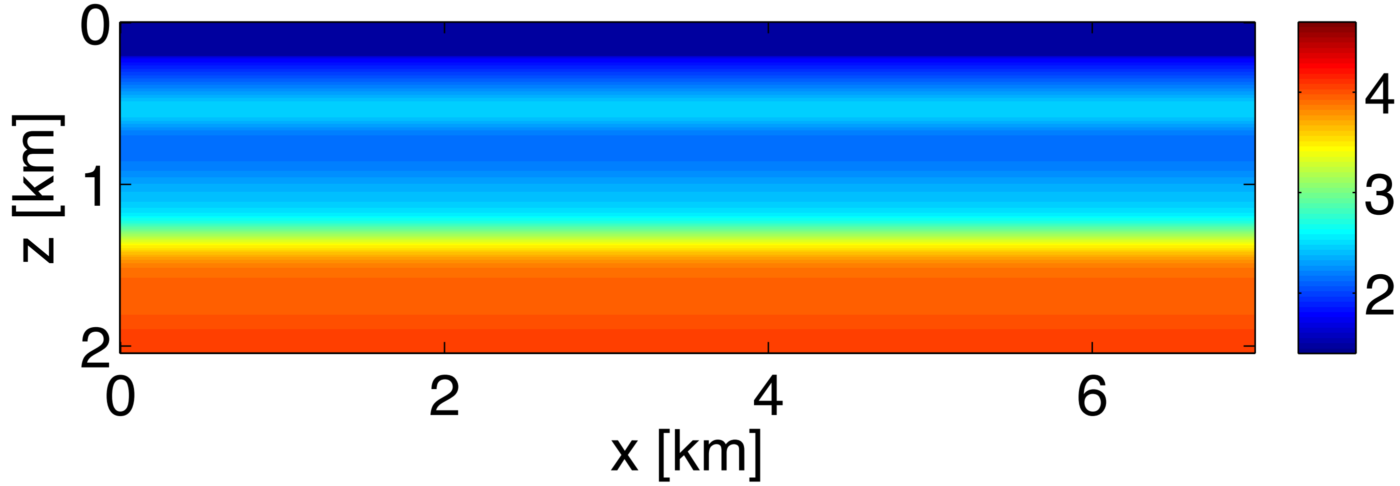
SLIM

Non-linear migration

- Fast progress in the beginning, decent convergence rate in the end
- Same quality with less PDE solves
- Better quality with same # PDE solves

FWI

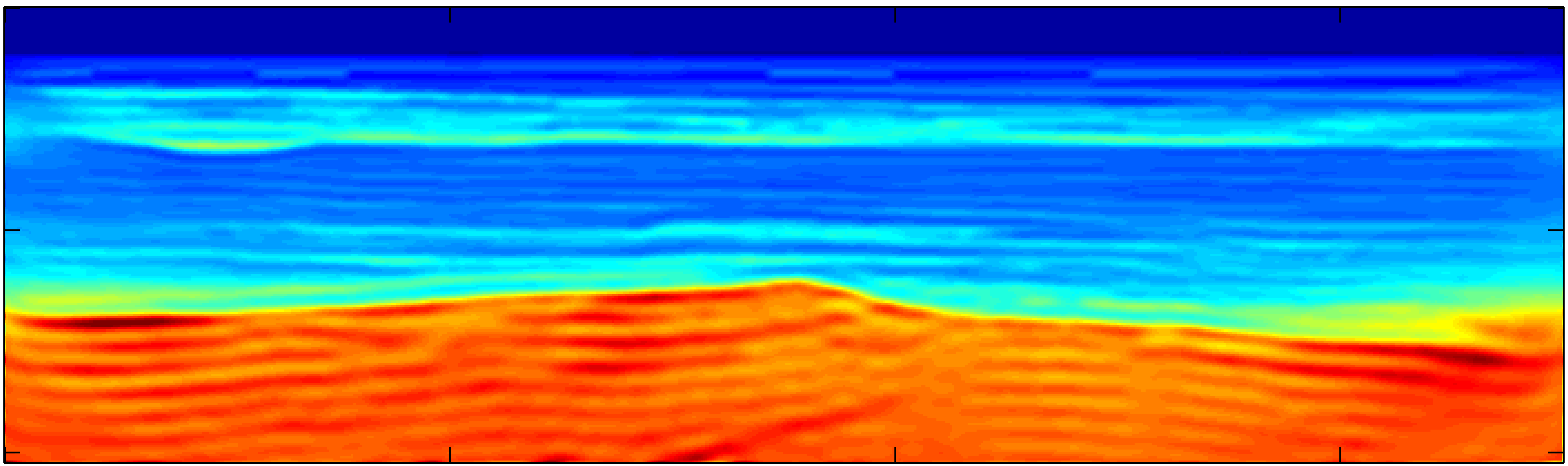




data for
141 sources, 281 receivers, 15 Hz Ricker

multiscale frequency domain inversion:
[2.5-20] Hz in 16 bands

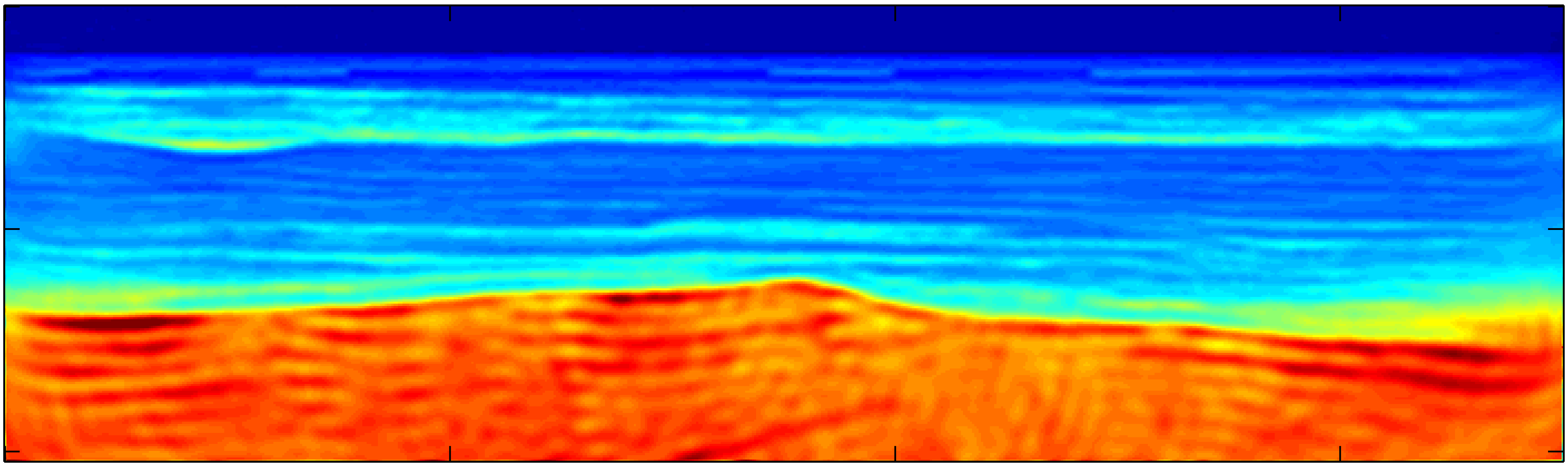
FWI



traditional L-BFGS

~10 full evaluations per frequency band

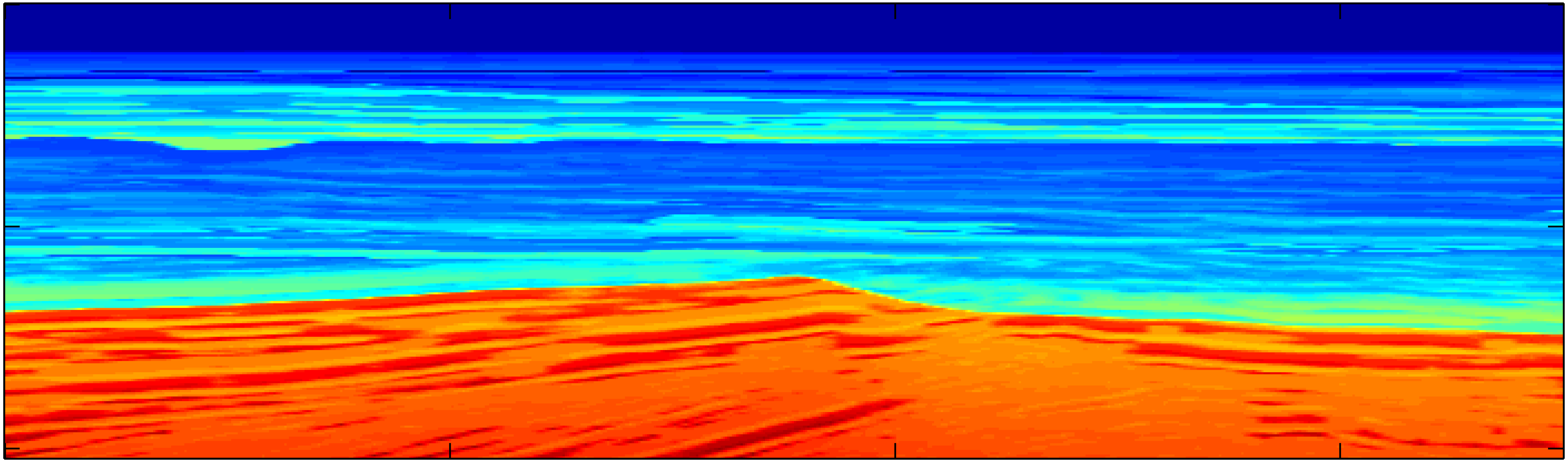
FWI



hybrid method

~2 full evaluations per frequency band

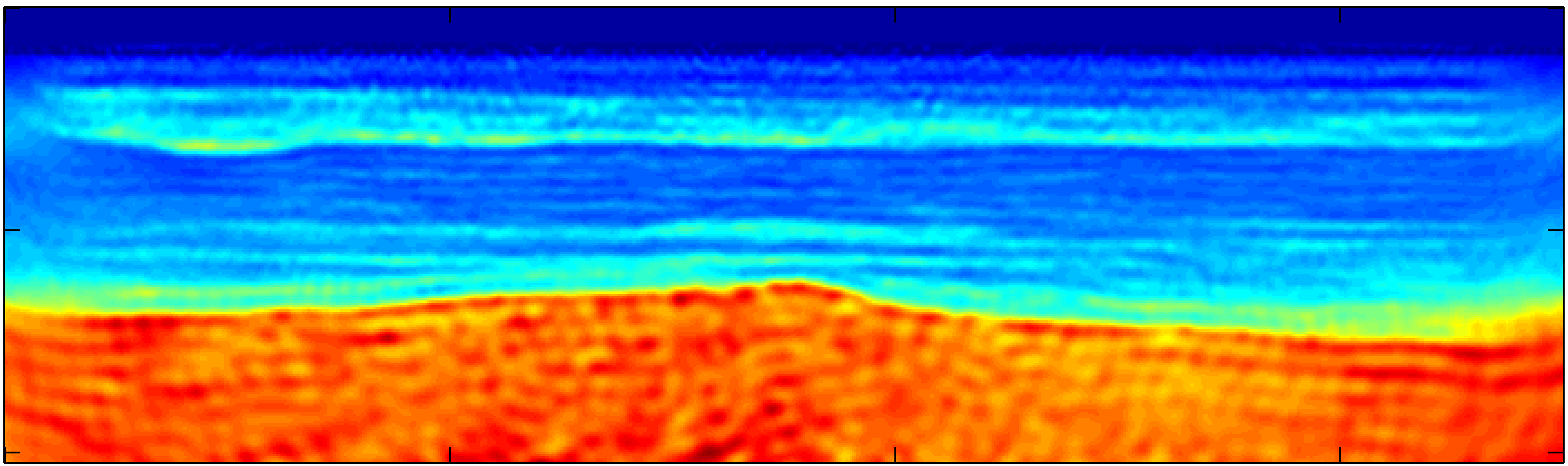
FWI



FWI

- Cost per frequency band equivalent to 2 evaluations of the full misfit.
- Total cost of the inversion equivalent to 17 evaluations of the full misfit.

FWI: time domain data



~2 full evaluations per frequency band

data generated with time domain FD

SLIM

Conclusions

- Hybrid method gives both speed-up of stochastic method and convergence rate of deterministic method
- Not restricted to randomized source encoding: can be applied to marine data!

SLIM

Acknowledgements

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.