Low-rank methods for on-the-fly slicing & dicing of seismic data and image volumes

Rajiv Kumar

SLIM
University of British Columbia

Motivation

Massive seismic data volumes of the orders of petabytes

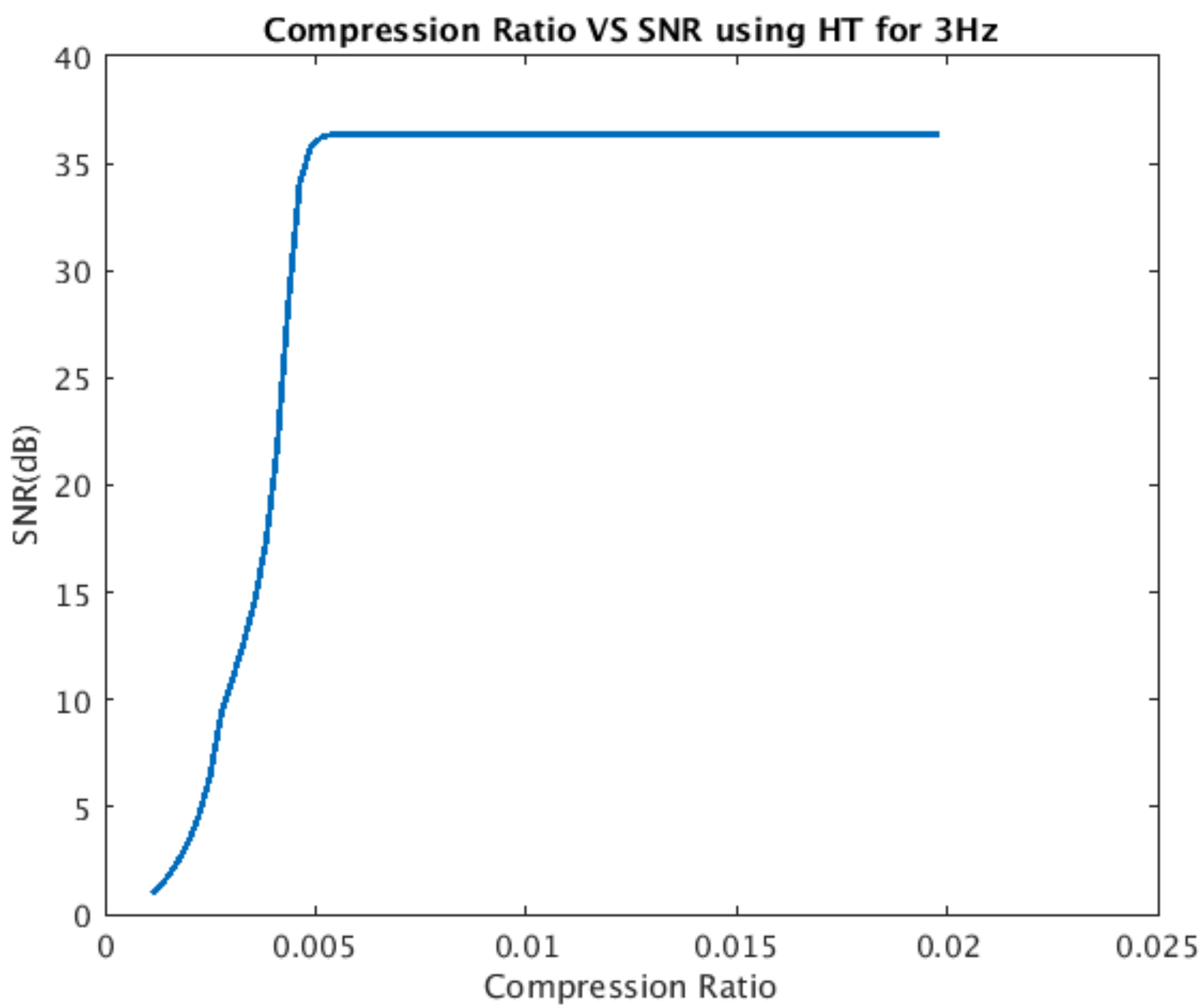
Expensive to store and use in processing

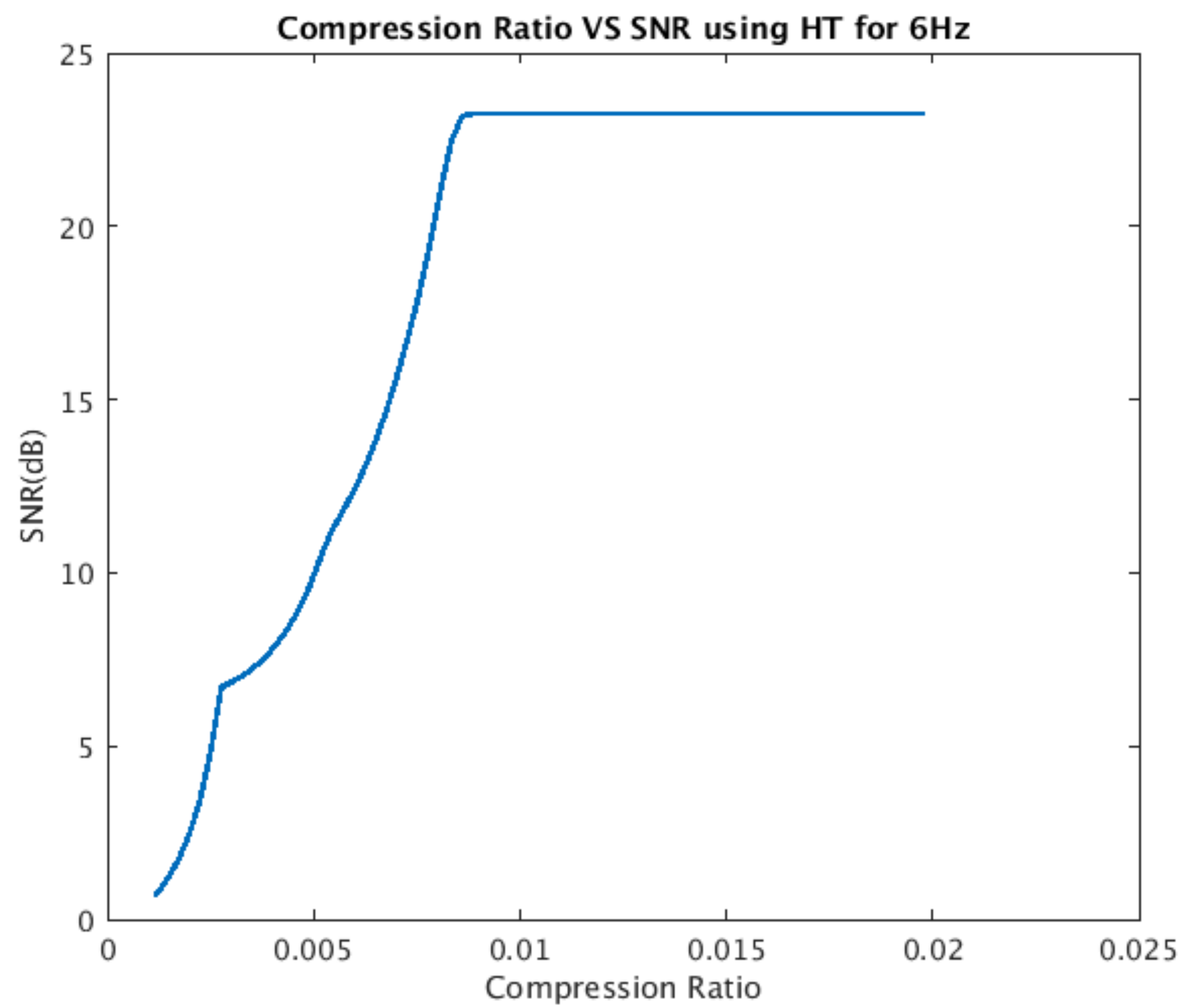
Devise compression techniques to mitigate data handling

Work with compressed data in inversion without forming the full seismic volume

Compression rate (4D seismic volumes)

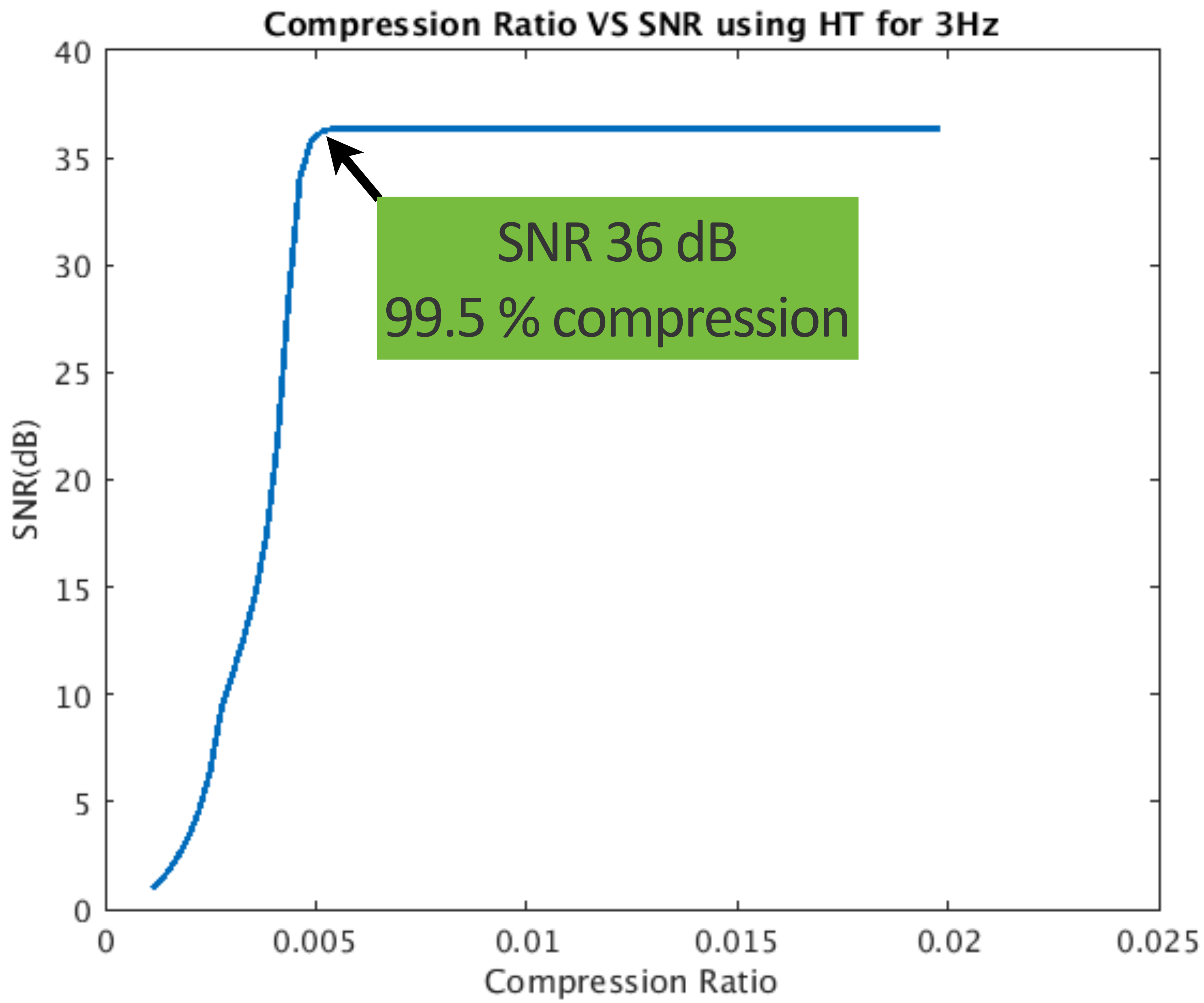
Monochromatic slice - 396 x 396 x 50 x 50

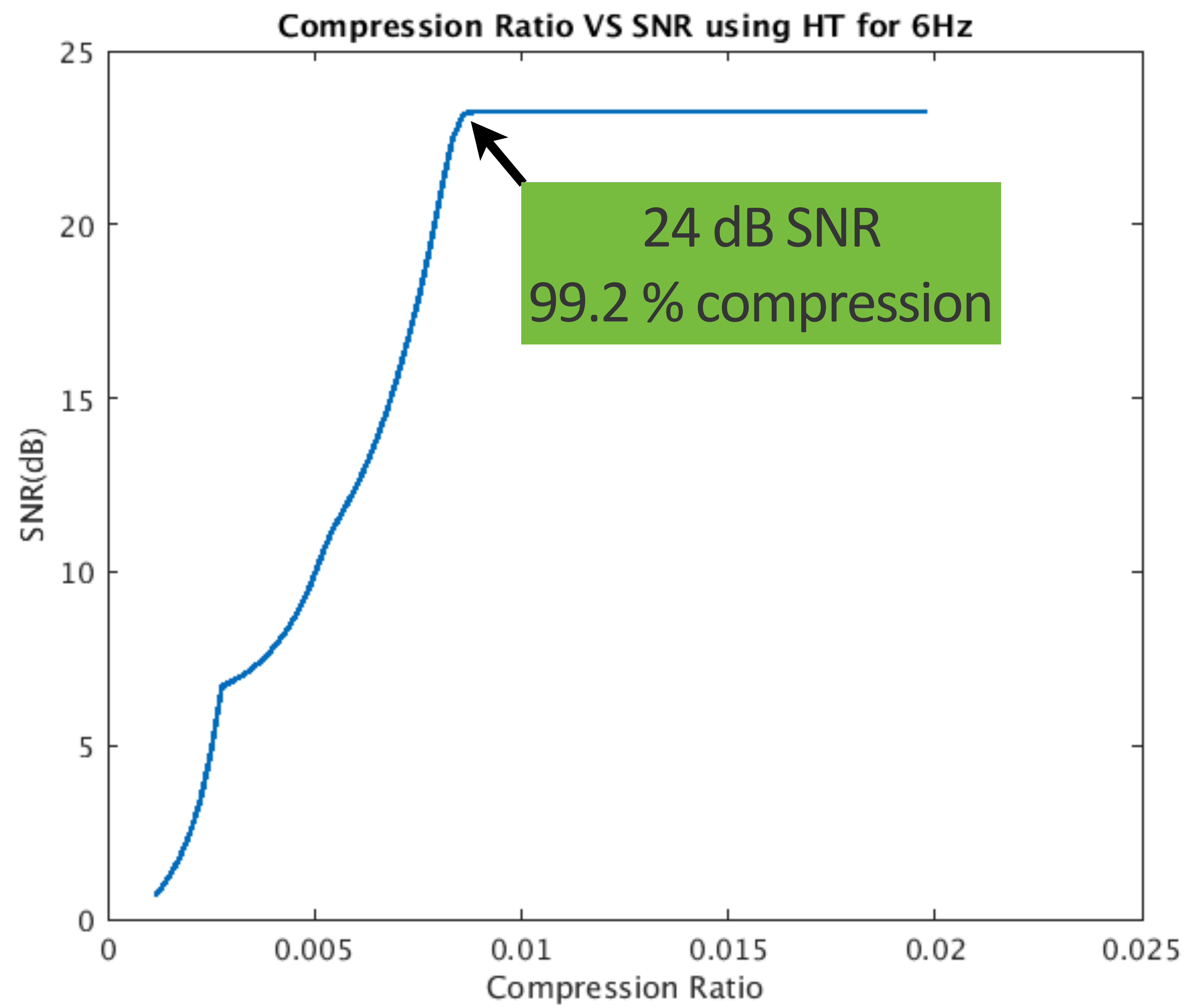




Compression rate (4D seismic volumes)

Monochromatic slice - 396 x 396 x 50 x 50





Compressed seismic volume for 3D FWI

Rajiv Kumar, Curt Da Silva, Yiming Zhang and Felix J. Herrmann



SLIM
University of British Columbia

Problem formulation

Our problem:

$$\min_{m} \sum_{k,l} \|P_k H_{k,l}^{-1}(m) q_{k,l} - d_{k,l}\|_2^2$$

Where:

$H_{k,l}(m) \in \mathbb{C}^{N \times N}$	Helmoltz operator at the kth shot of lth frequency
$m \in \mathbb{R}^N$	Medium parameters
$P_k \in \mathbb{R}^{n \times N}$	Receiver projection operator at the kth shot
$q_{k,l} \in \mathbb{R}^N$	Source at the kth shot of lth frequency
$d_{k,l} \in \mathbb{C}^n$	Observed data at the kth shot of lth frequency

Challenges for Full Waveform Inversion

Less than ideal acquisition

- missing data

Computationally intensive

- # of source experiments

High storage costs for data

- curse of dimensionality

Low rank seismic data

Seismic data is *redundant*

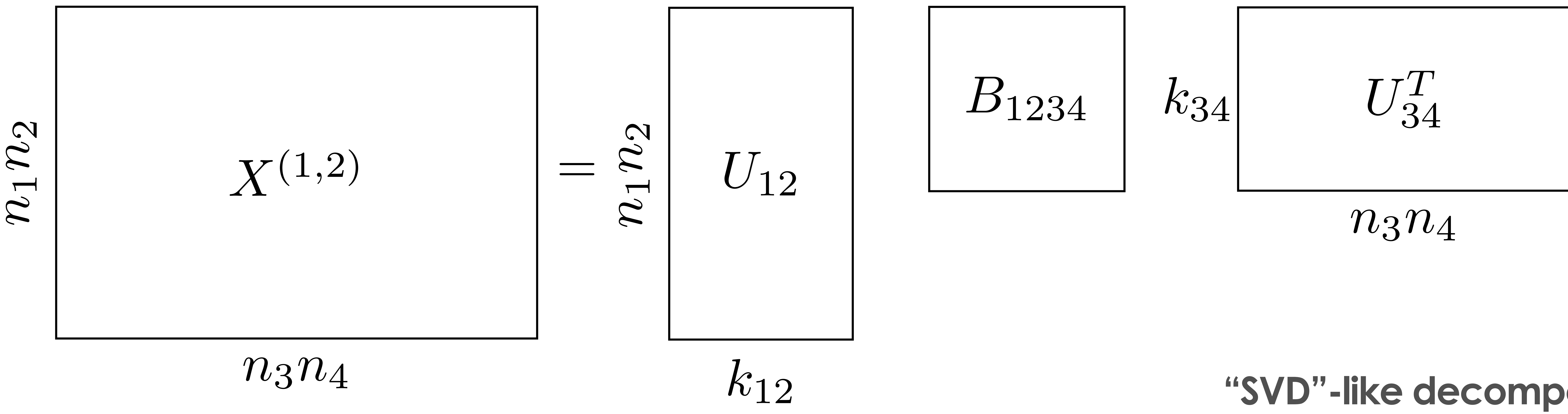
- measuring the same Earth, with slightly different measurements

We can exploit *low rank tensor* structure for data compression

- Hierarchical Tucker format

Hierarchical Tucker format

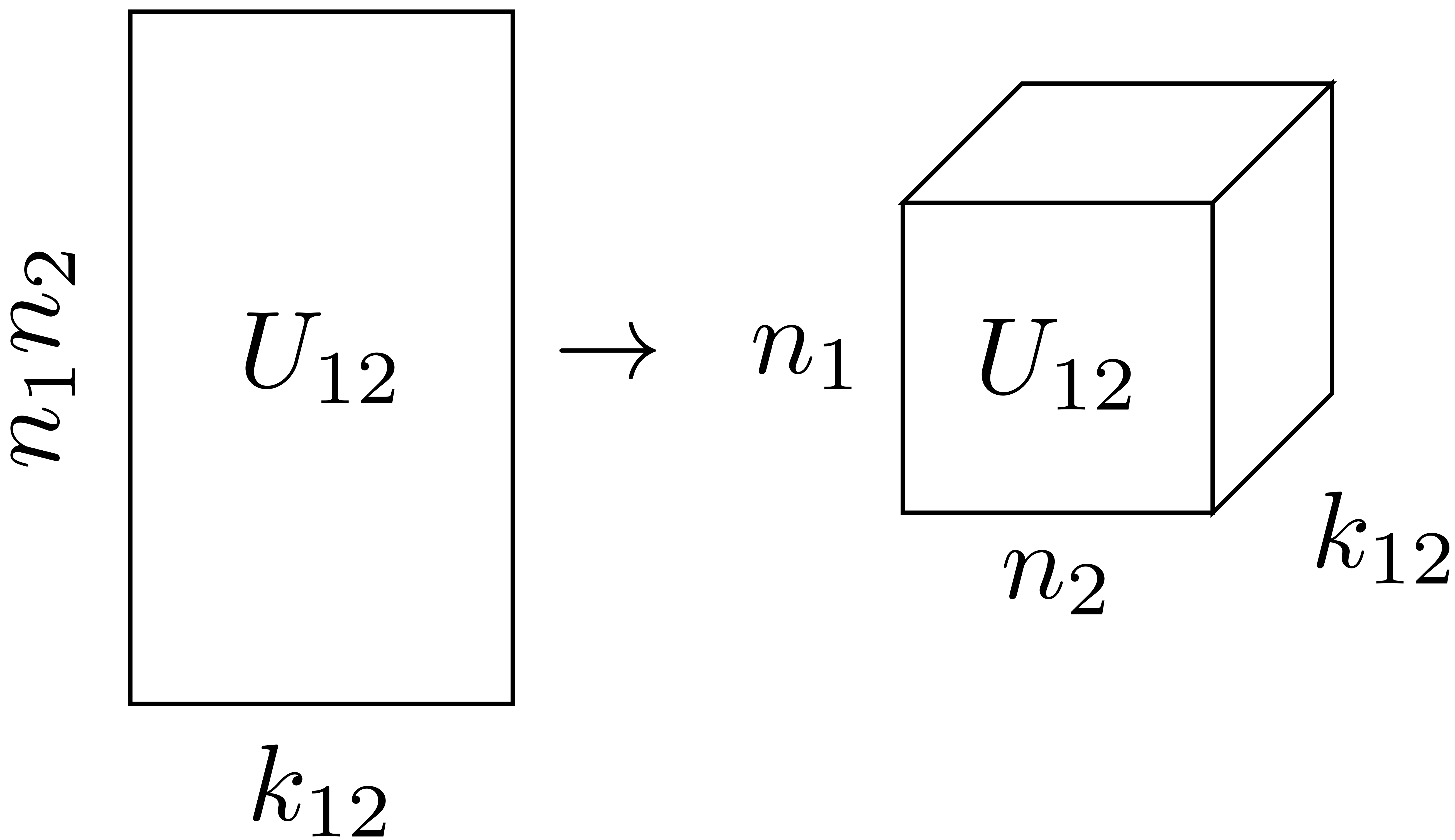
$X - n_1 \times n_2 \times n_3 \times n_4$ tensor



"SVD"-like decomposition

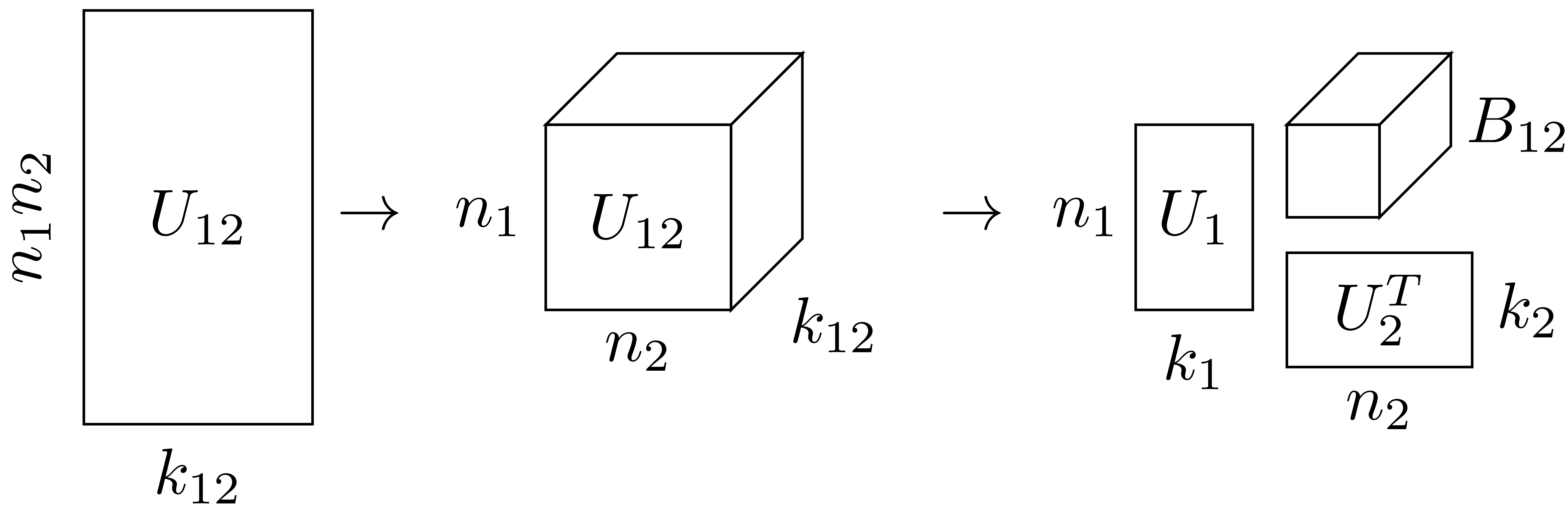
Hierarchical Tucker format

$X - n_1 \times n_2 \times n_3 \times n_4$ tensor



Hierarchical Tucker format

$X - n_1 \times n_2 \times n_3 \times n_4$ tensor



HT Compression

A $100 \times 100 \times 100 \times 100$ cube with max rank 20

Full storage: $100^4 = 10^8$ parameters

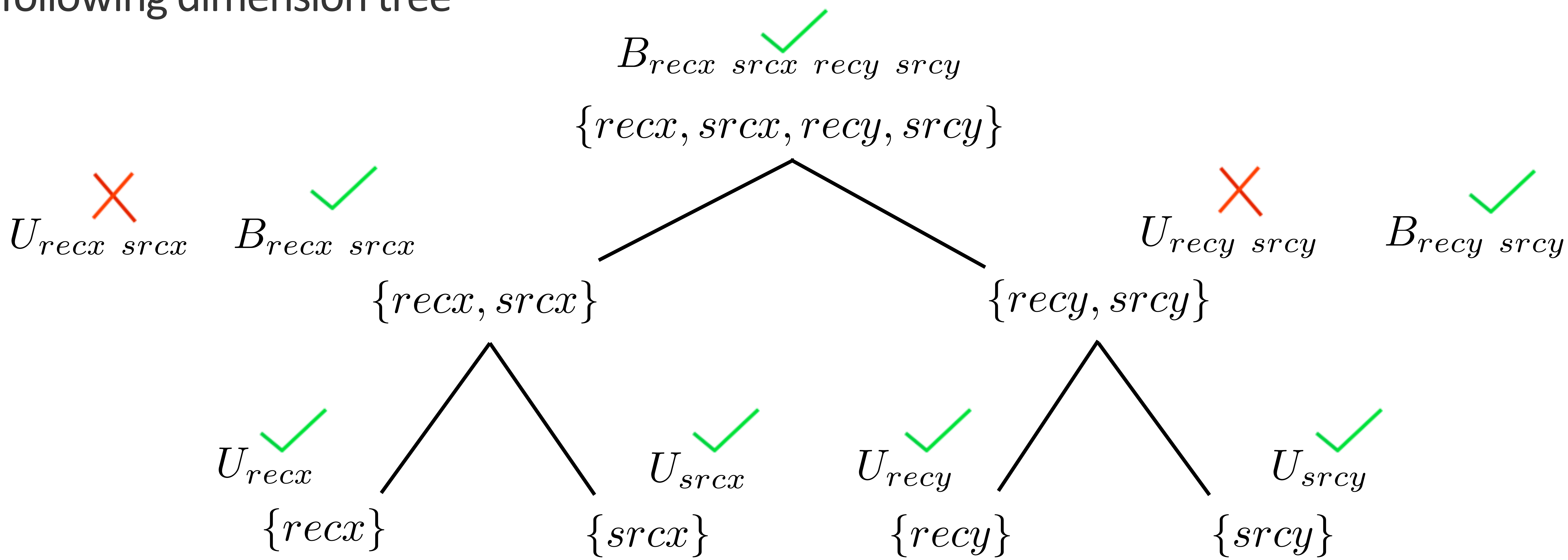
HT storage: 24400 values

Compression of a factor of 99.97%

C. Da Silva, F. Herrmann “Optimization on the Hierarchical Tucker manifold - applications to tensor completion”, 2015

Seismic HT Tensor

For a frequency slice with coordinates (rec x, rec y src x, src y), we can introduce the following dimension tree



Compressed Data

We can compress low frequency seismic data in HT format

- full data -> truncation
- missing data -> interpolation

In either case, we don't have to store the full volume during FWI

Memory usage - 396 x 396 x 50 x 50 volume ~ 5.8 GB

Frequency	HT Parameter Size	SNR	Compression Ratio
3 Hz	48 MB	40.95	99.2%
6 Hz	95 MB	25.76	98.4%

Compressed Data FWI

For 3D FWI with stochastic optimization

- we only need *query-based* access to the data volume

Each iteration of the stochastic algorithm only requires a subset of the full sources

- compress data volume in HT
- extract shots as requested by the algorithm

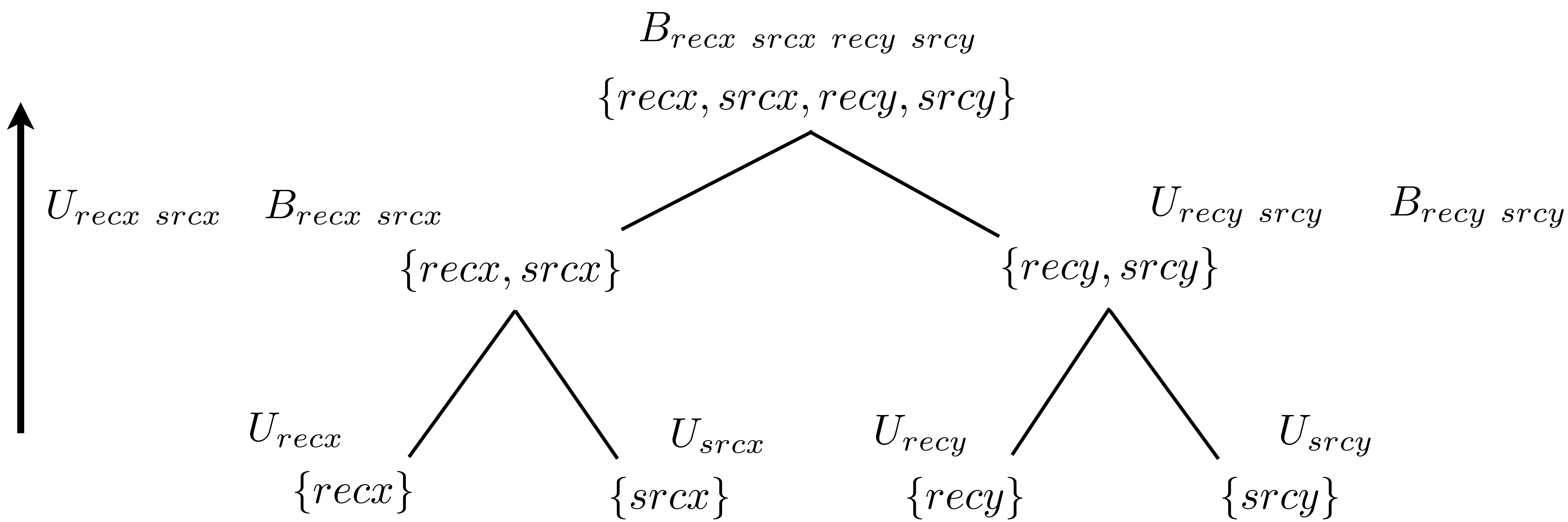
Ideal scenario

Follow the tree structure

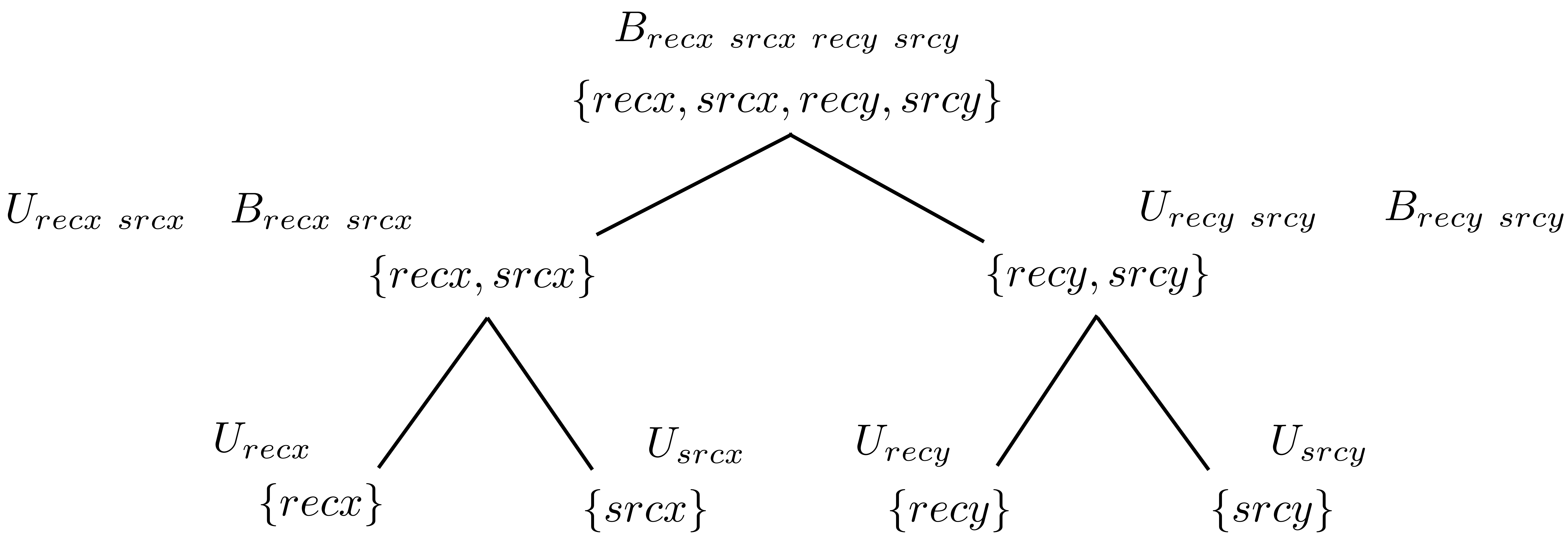
form full seismic data volumes

Extract shot / receiver gather

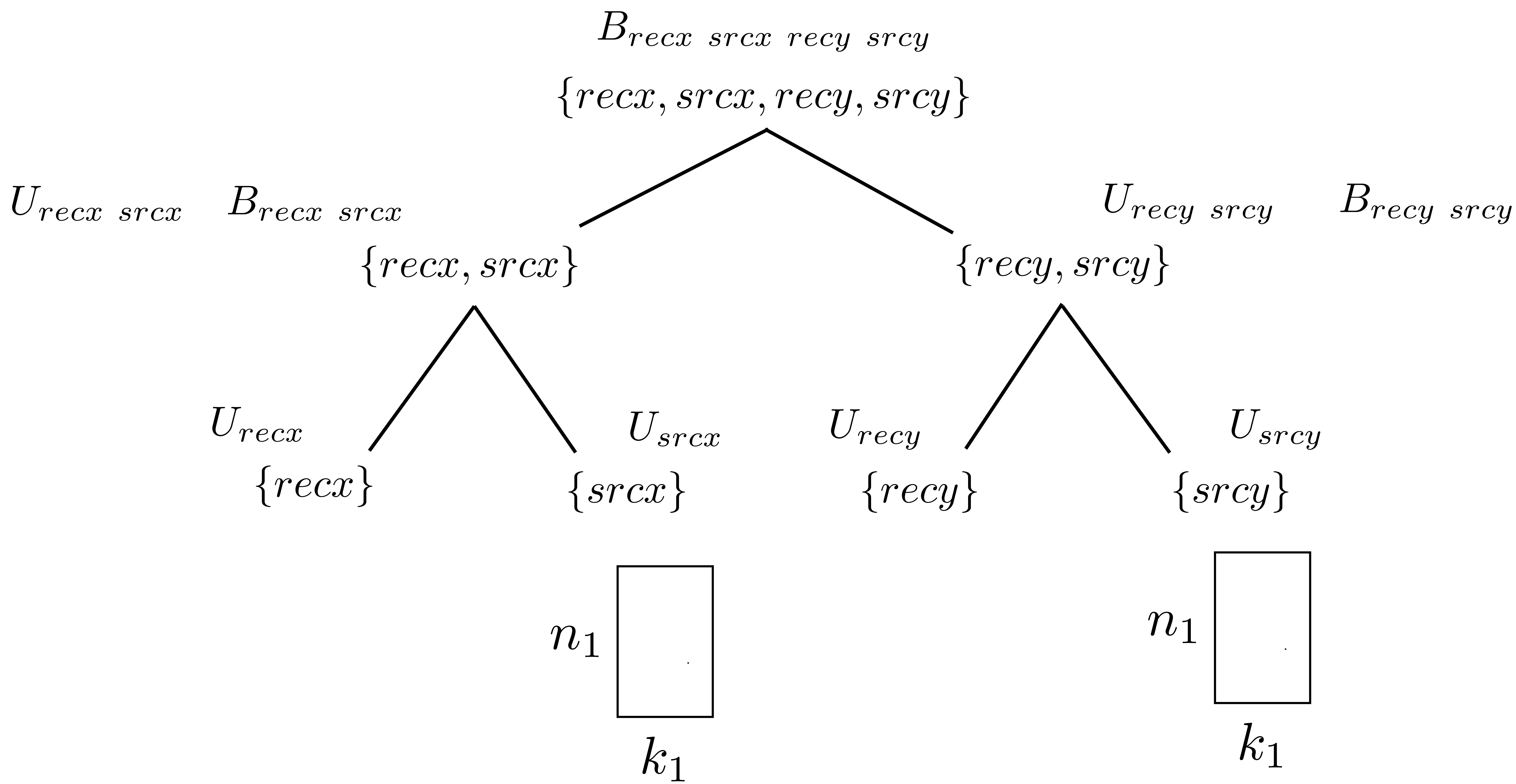
very expensive



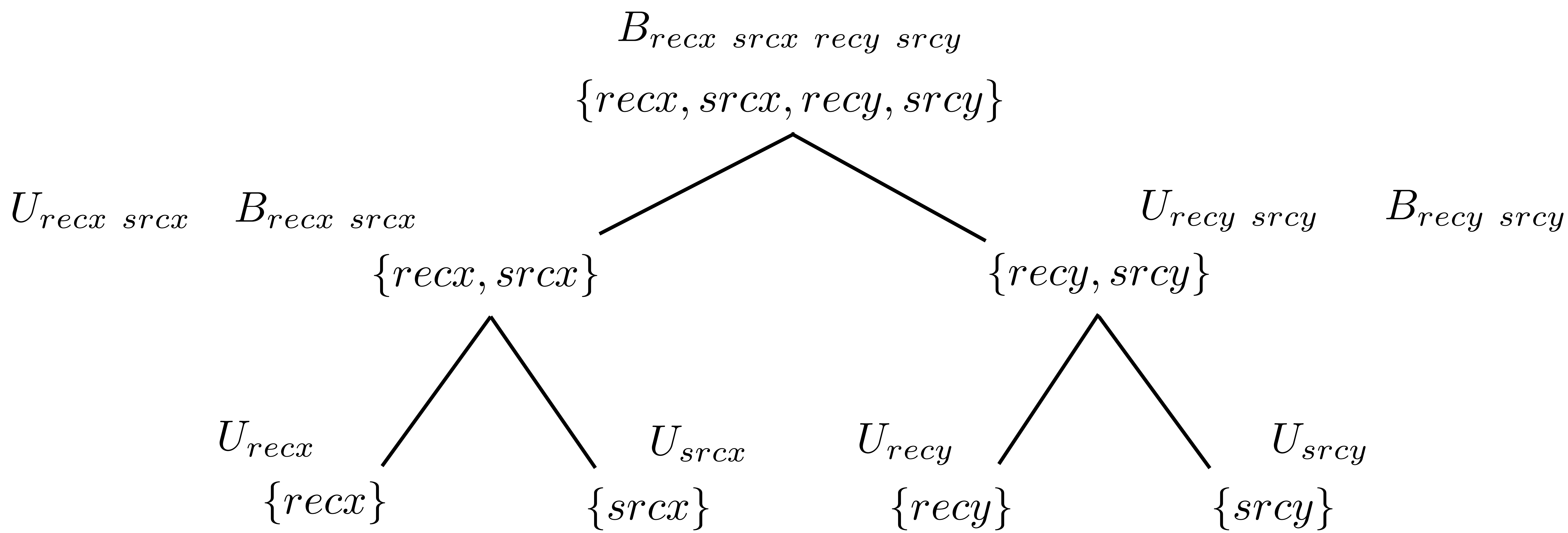
Efficient trick to extract gathers



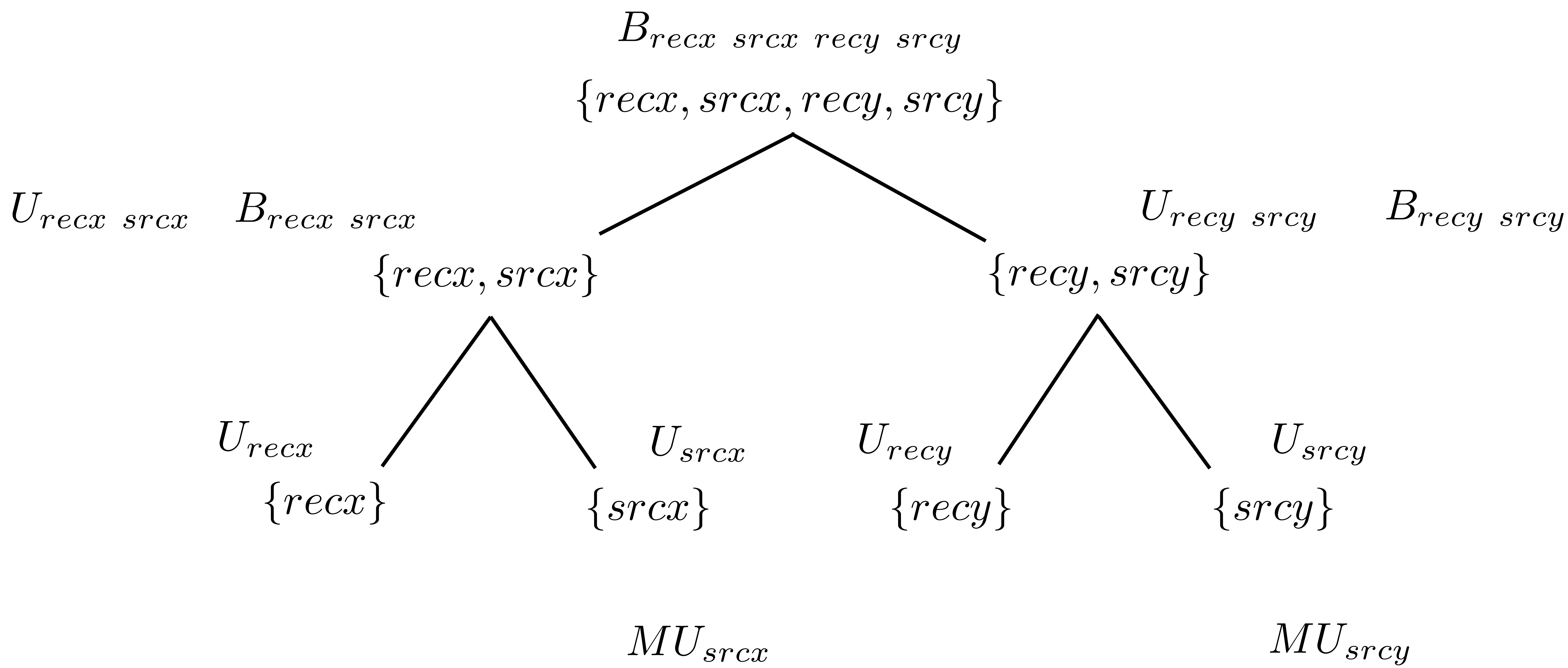
Efficient trick to extract gathers



Efficient trick to extract gathers

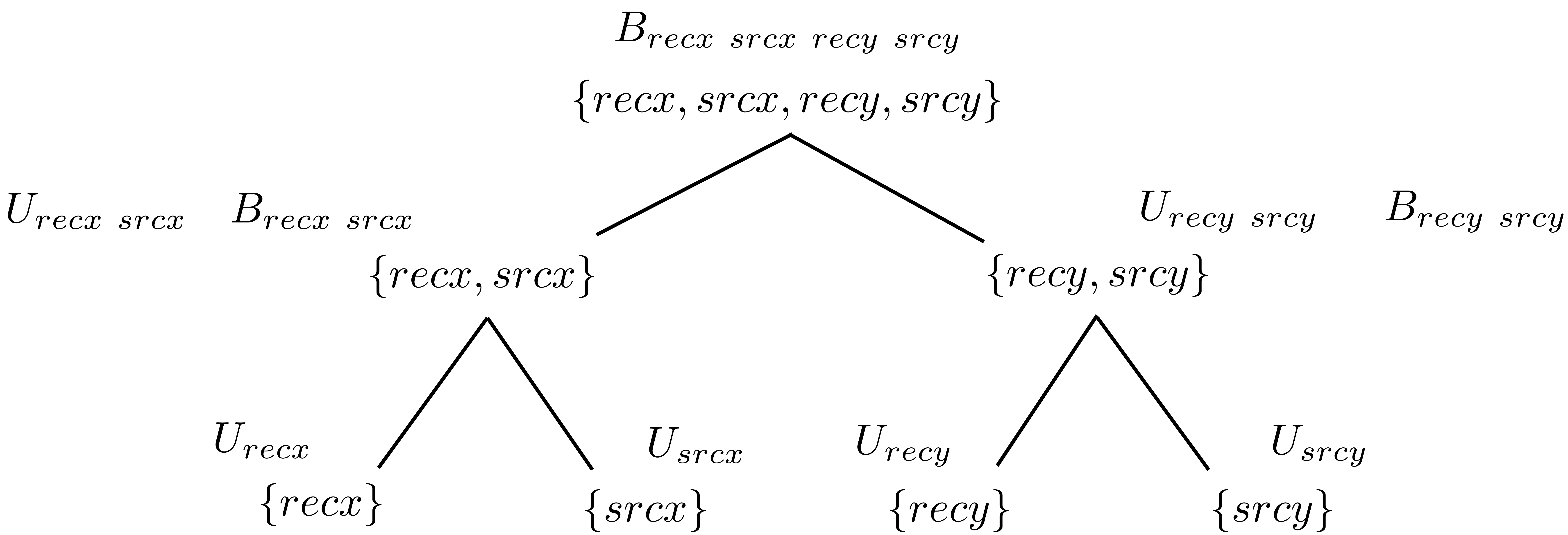


Efficient trick to extract gathers

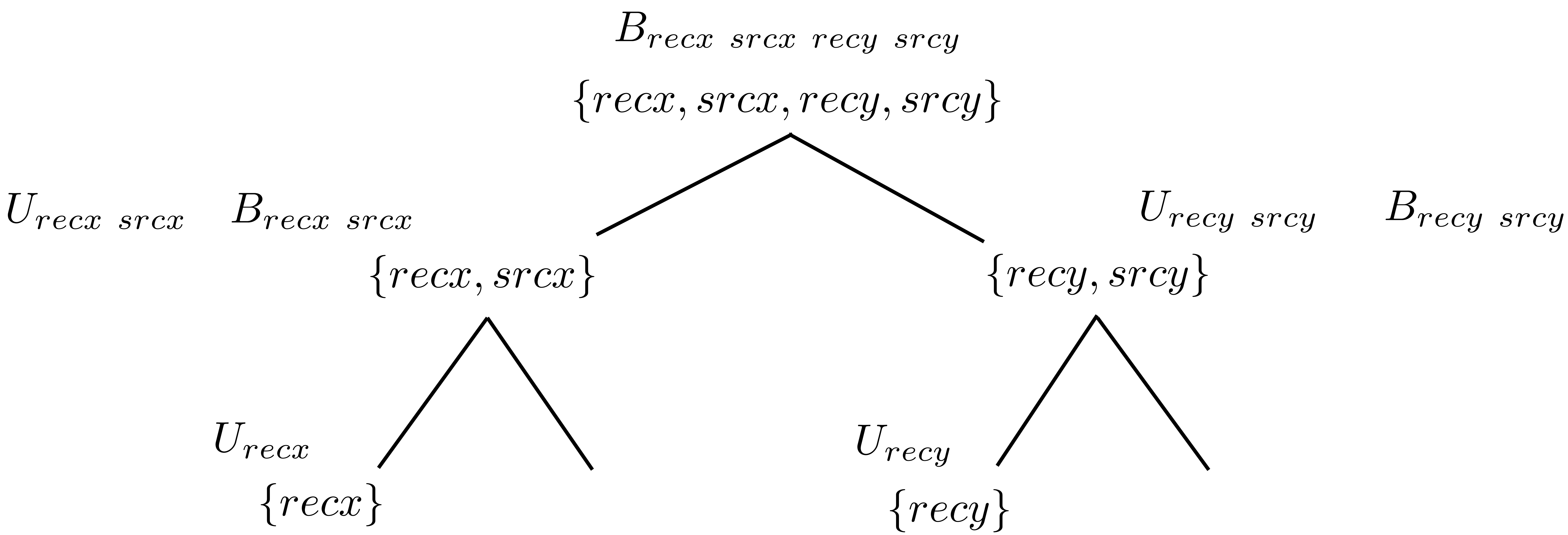


$$M \in \mathrm{R}^{1\times n_1}$$
$$[0, 0, ...1, 0, ...]$$

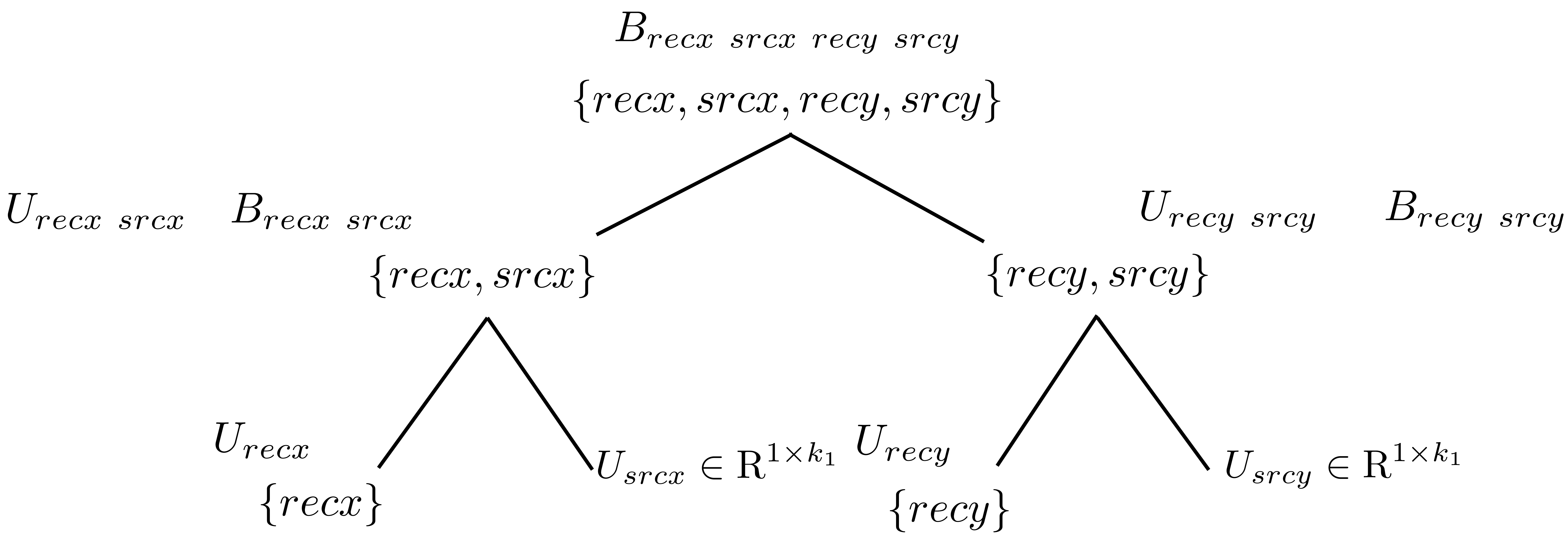
Efficient trick to extract gathers



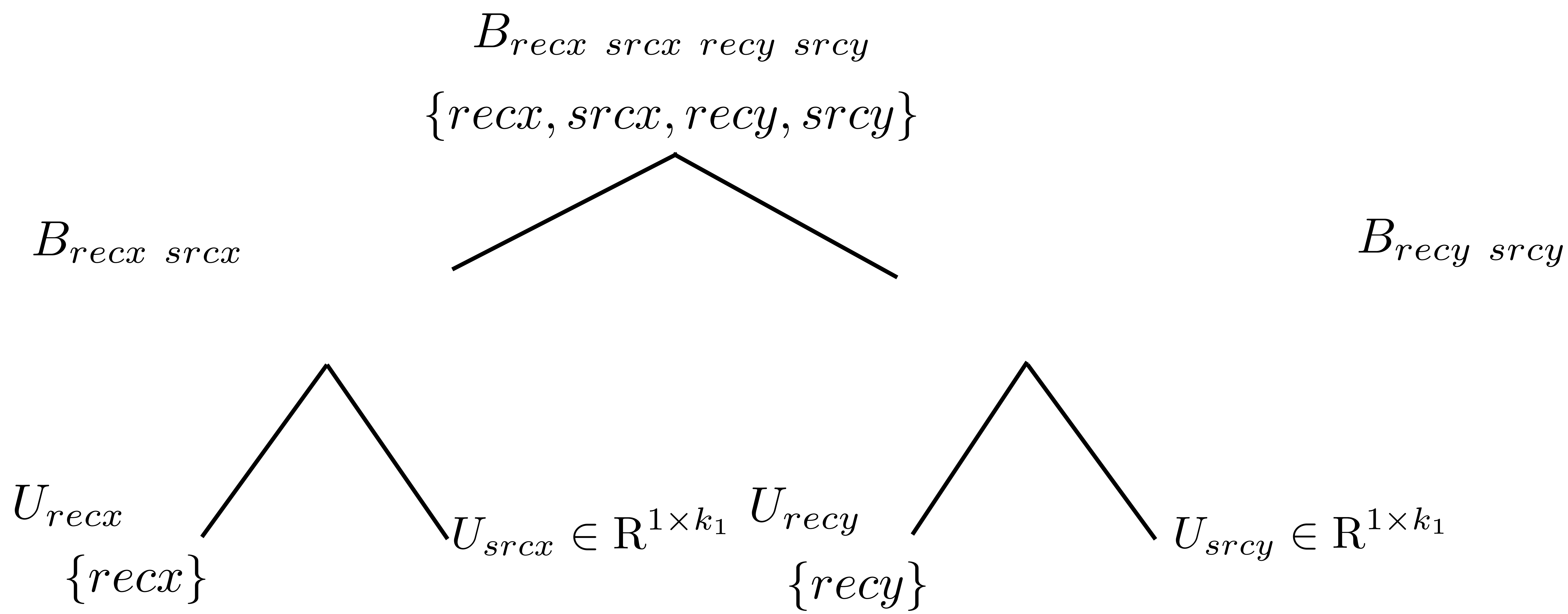
Efficient trick to extract gathers



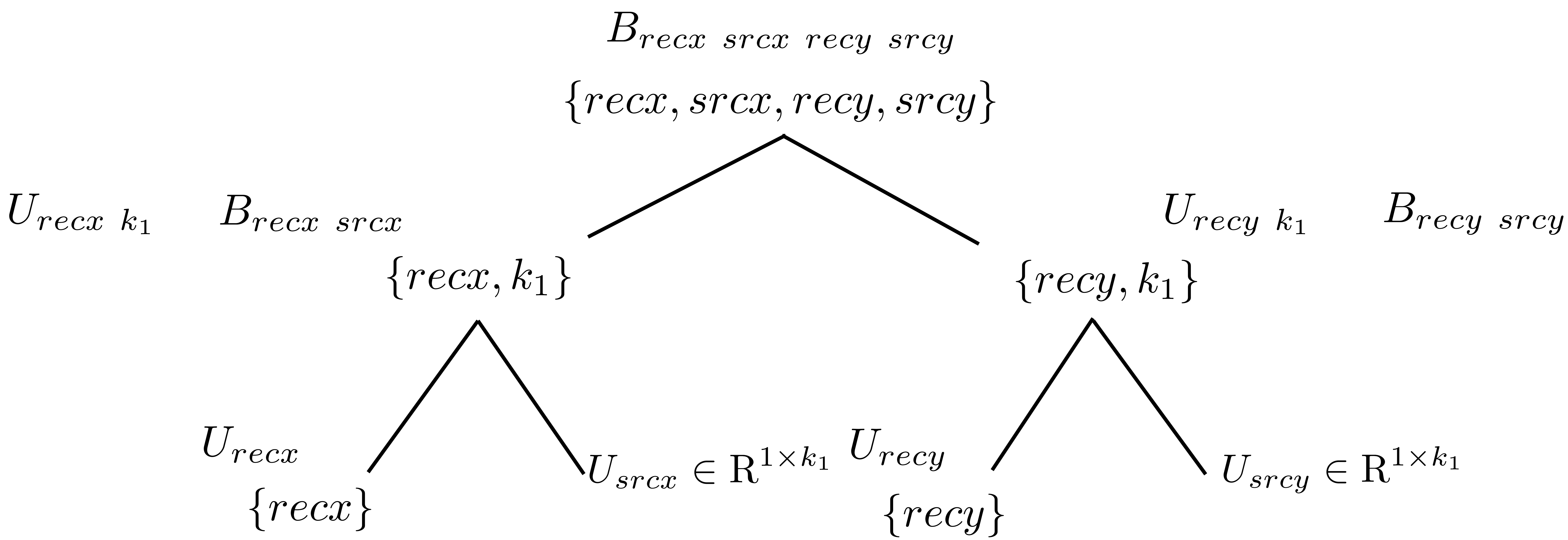
Efficient trick to extract gathers



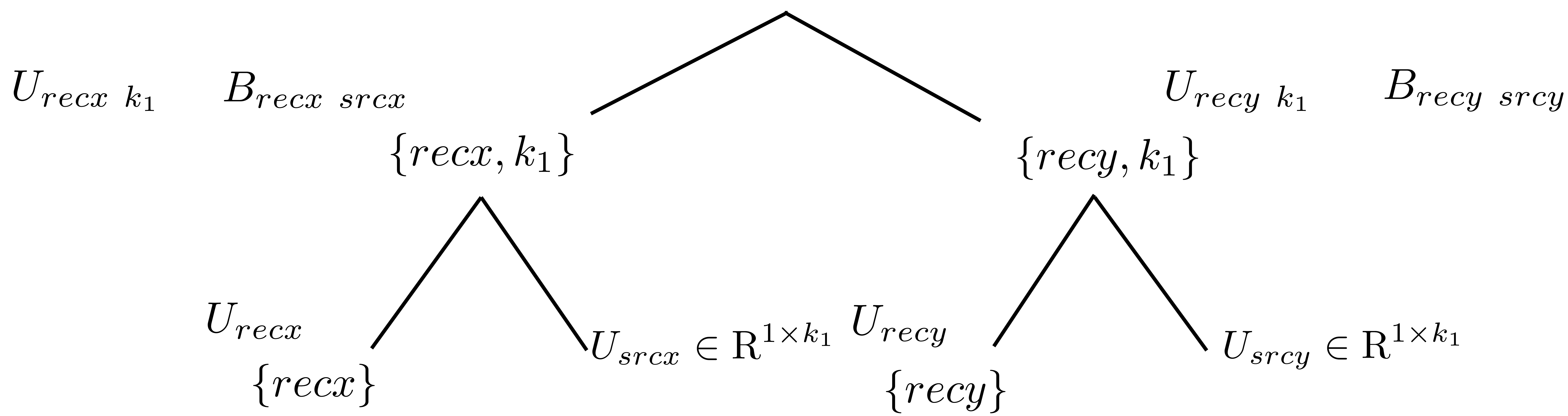
Efficient trick to extract gathers



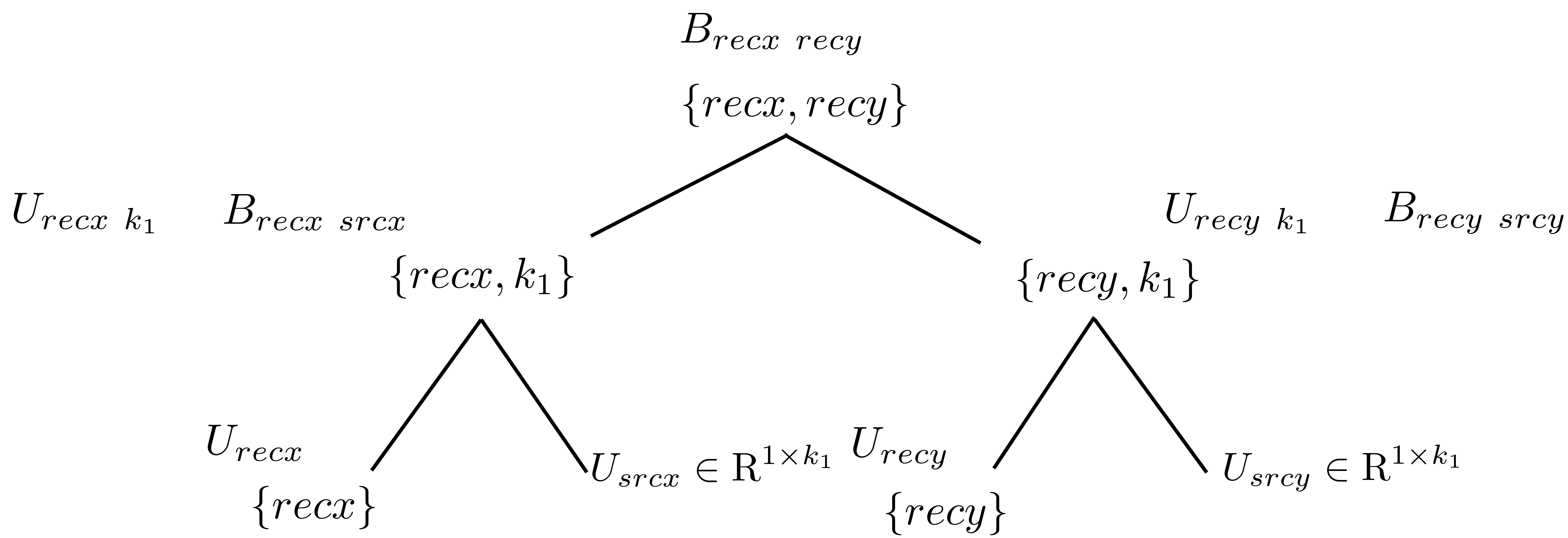
Efficient trick to extract gathers



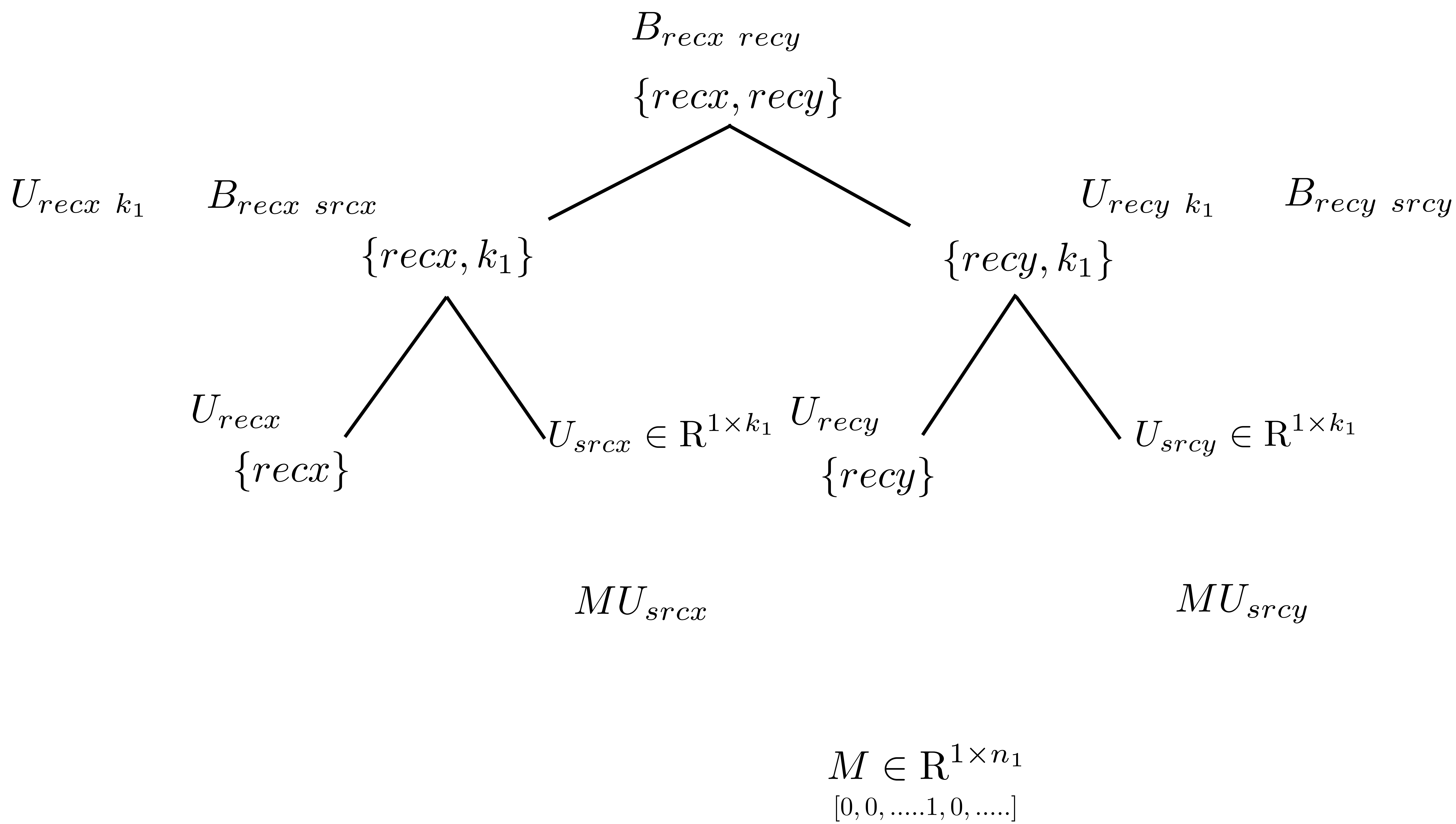
Efficient trick to extract gathers



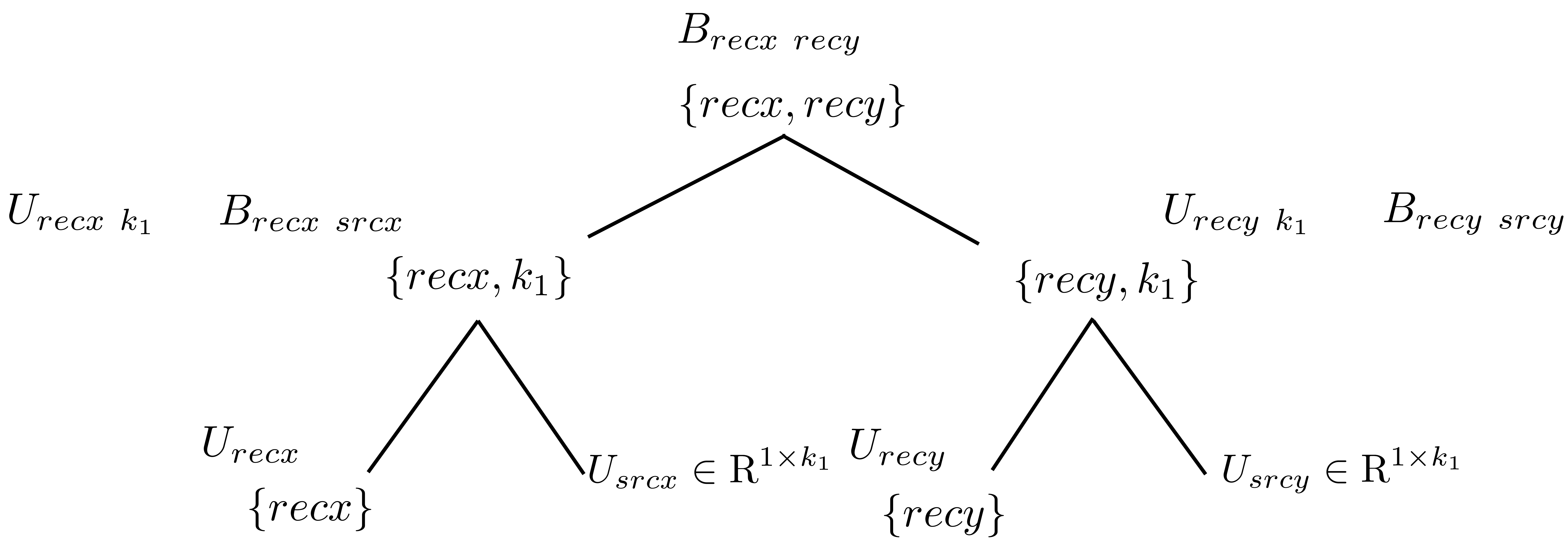
Efficient trick to extract gathers



Efficient trick to extract gathers



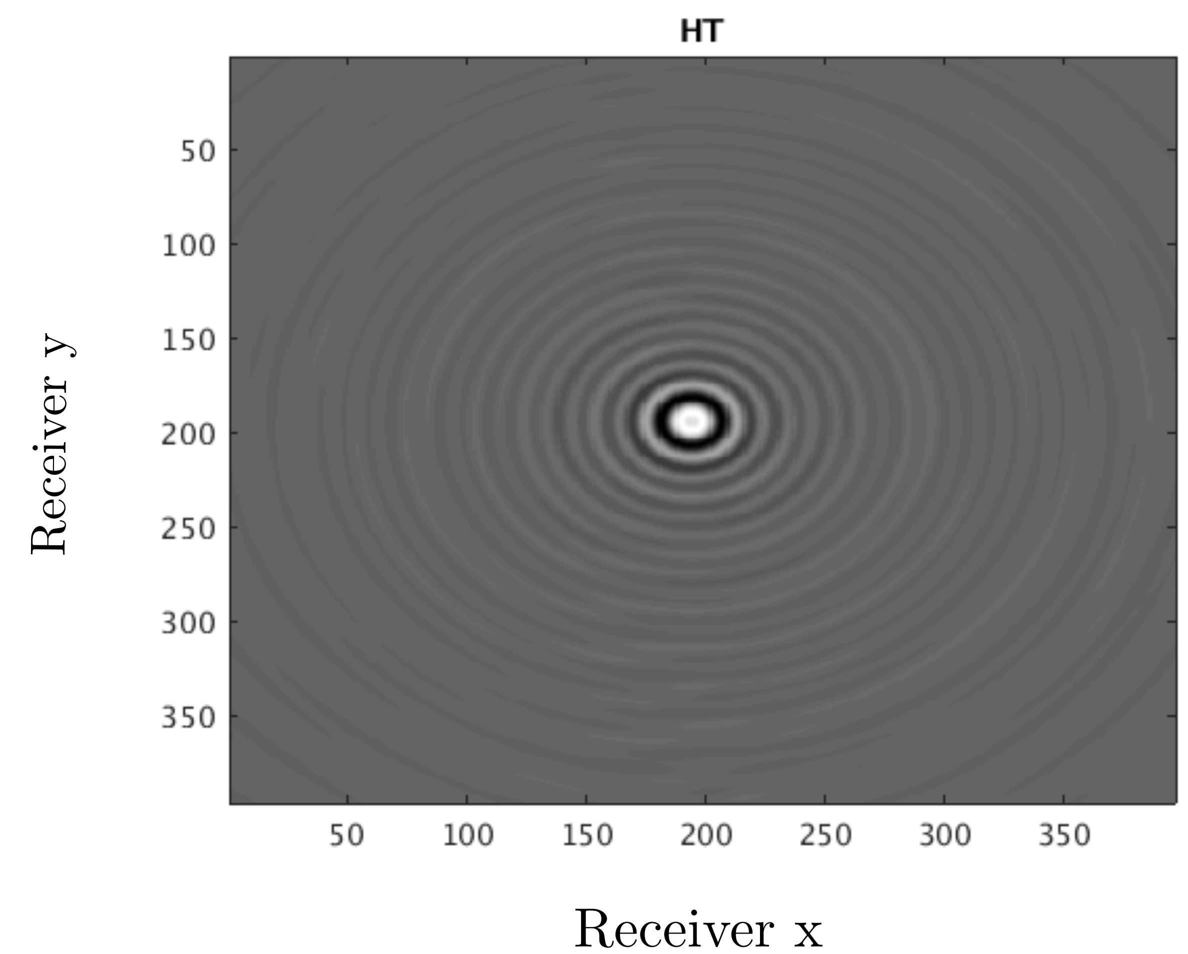
Efficient trick to extract gathers



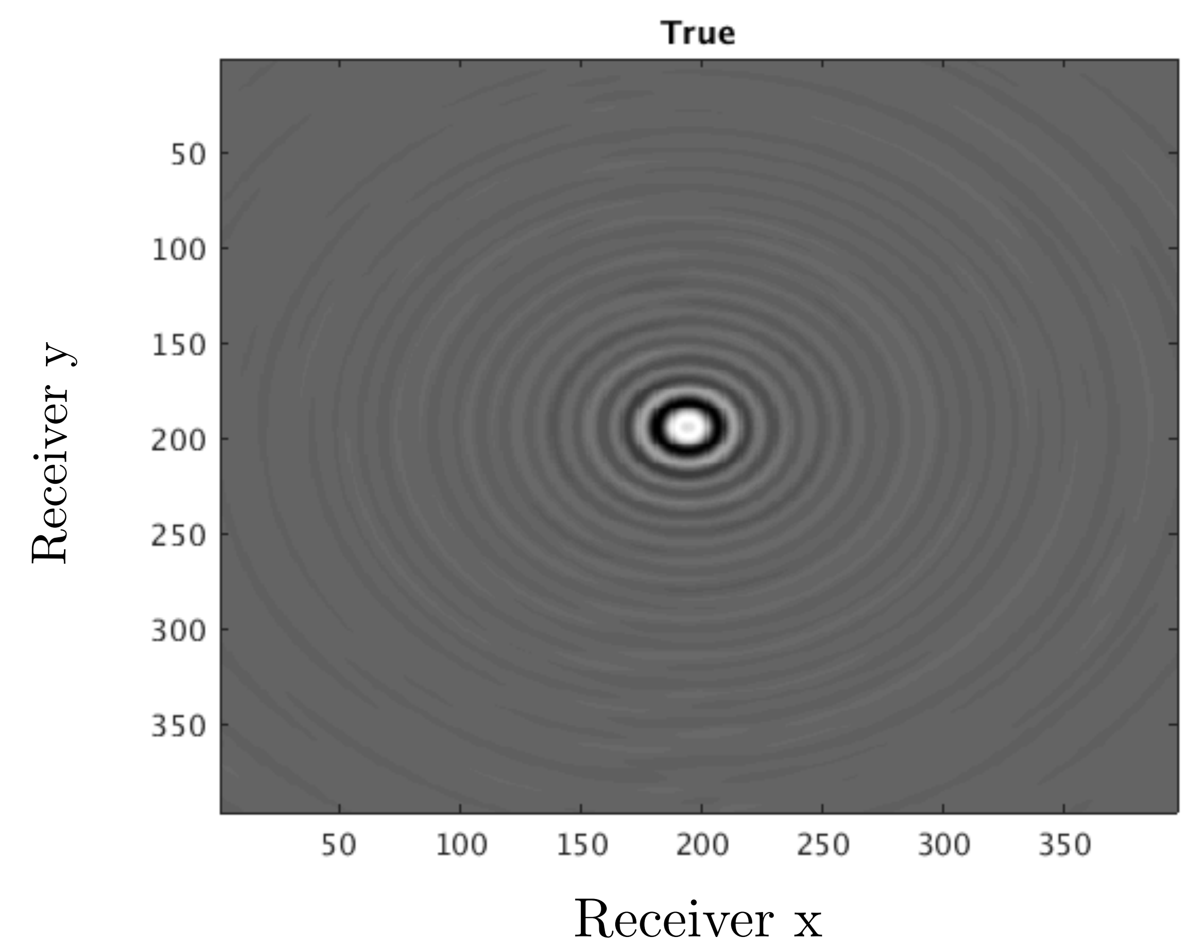
Slicing & dicing for data domain

Only need HT parameters and src index

```
function gather = gathers_extr(dimTree,x,index,mode)
% Extracts any common shot gathers or receiver gathers from HT parameters
%
% Usage:
%   gather=gather_extr(dimTree,x,index,mode);
%
% Input:
%   x        -    vectorized HT parameters
%   dimTree  -    generate U ( HT leaf bases) and B (HT interior nodes)
%
%   index    -    the index you want to extract from src or rec location
%   mode     -    1   :common shot gathers
%                 2   :common receiver gathers
%
% Output:
%   gather   -    the common shot gather or receiver gather
```

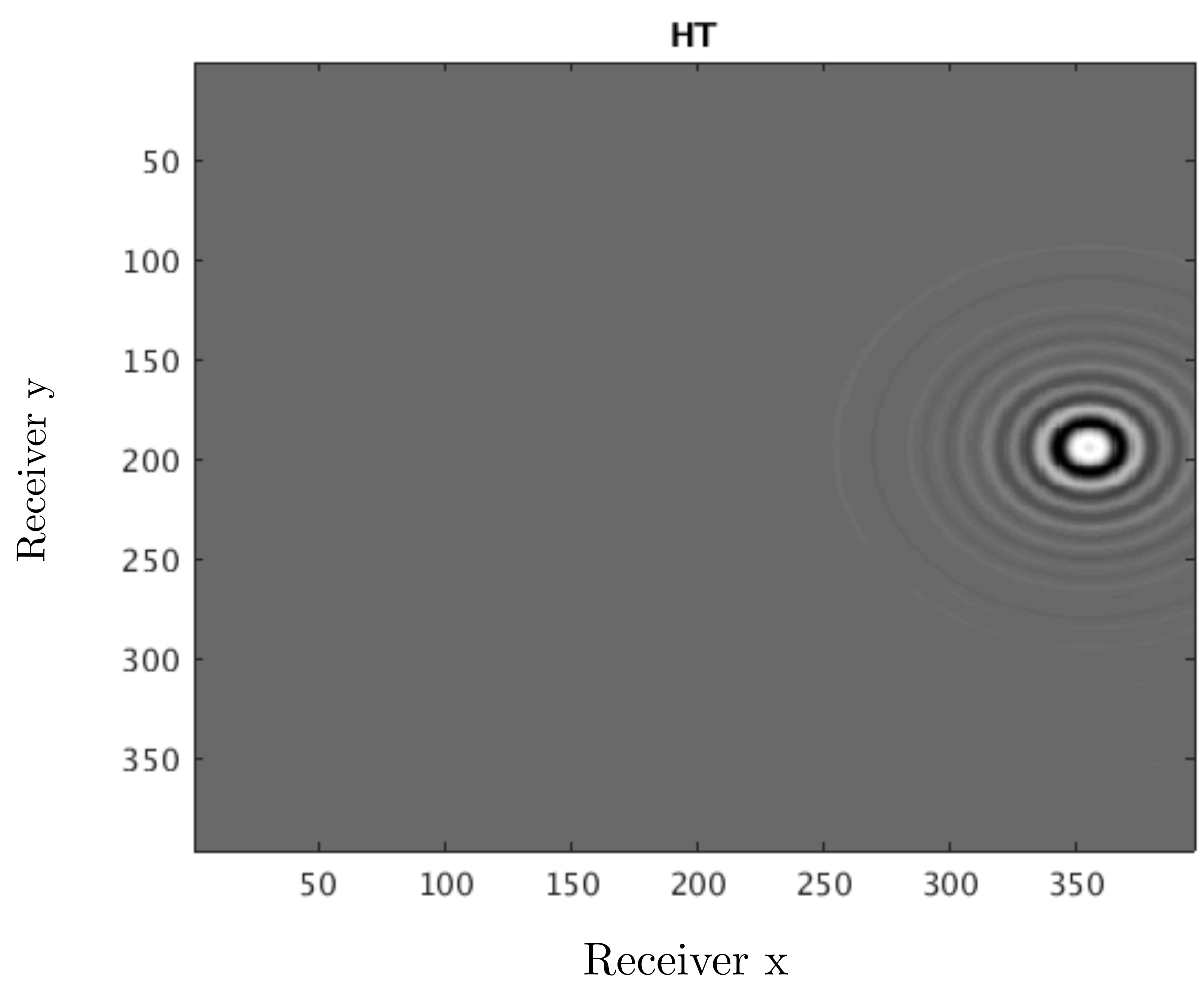


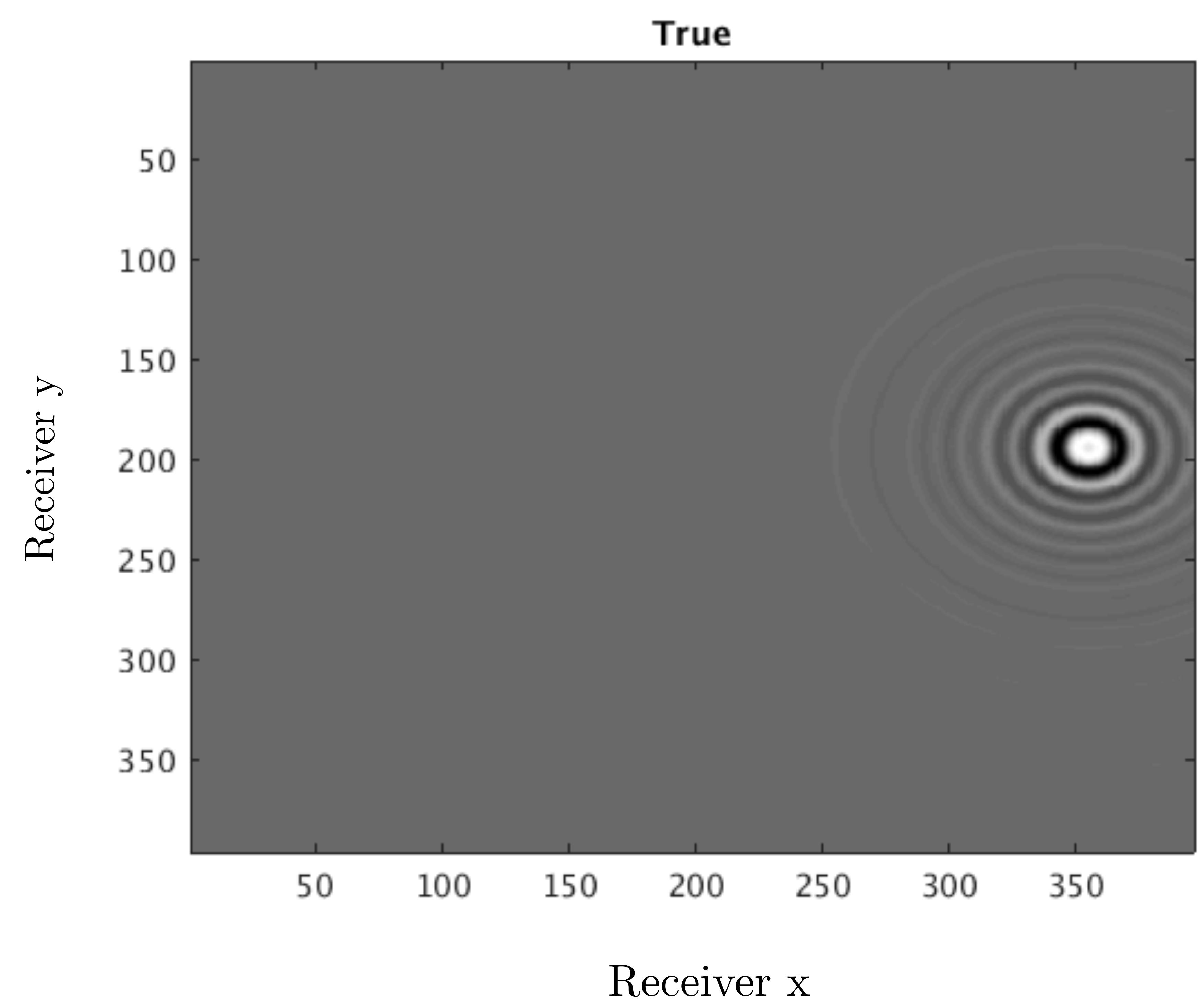

```
gather1=gathers_extr(dimTree,x,[25 25],'1');
```



Slicing & dicing for data domain

```
gather2=gathers_extr(dimTree,x,[25 45],'1');
```





3D FWI Example

3D FWI Example

Overthrust model

- 20 km x 20 km x 4.6 km - 50 m spacing, 500m water layer
- 50 x 50 sources, 200m spacing - 2500 shots
- 401 x 401 receivers, 50m spacing
- 3Hz - 5Hz frequency range, single freq. inverted at a time
- compression rate of 99.7%

3D FWI Example

Stochastic algorithm with three full passes through the data

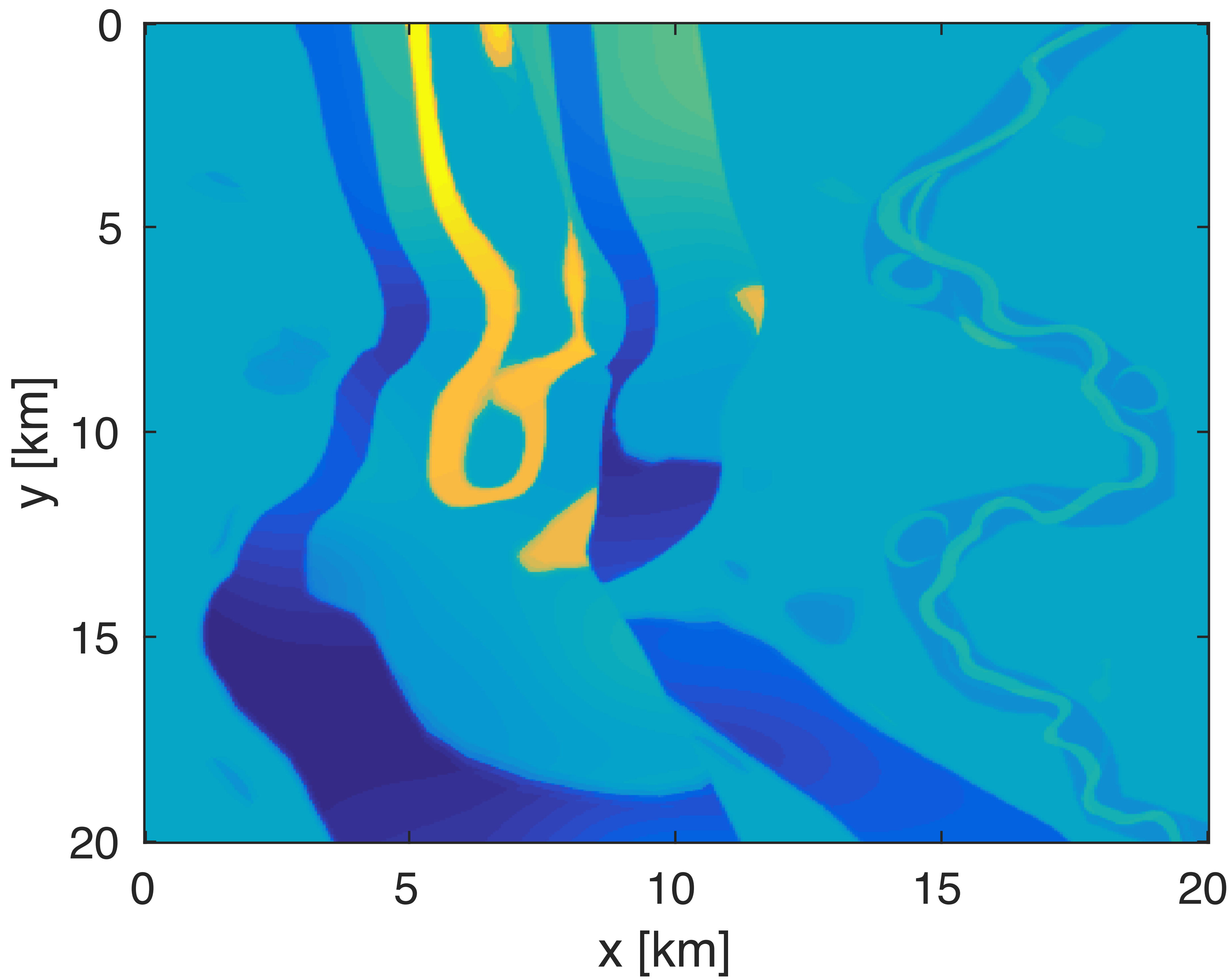
3D FWI Example

We compare stochastic FWI with

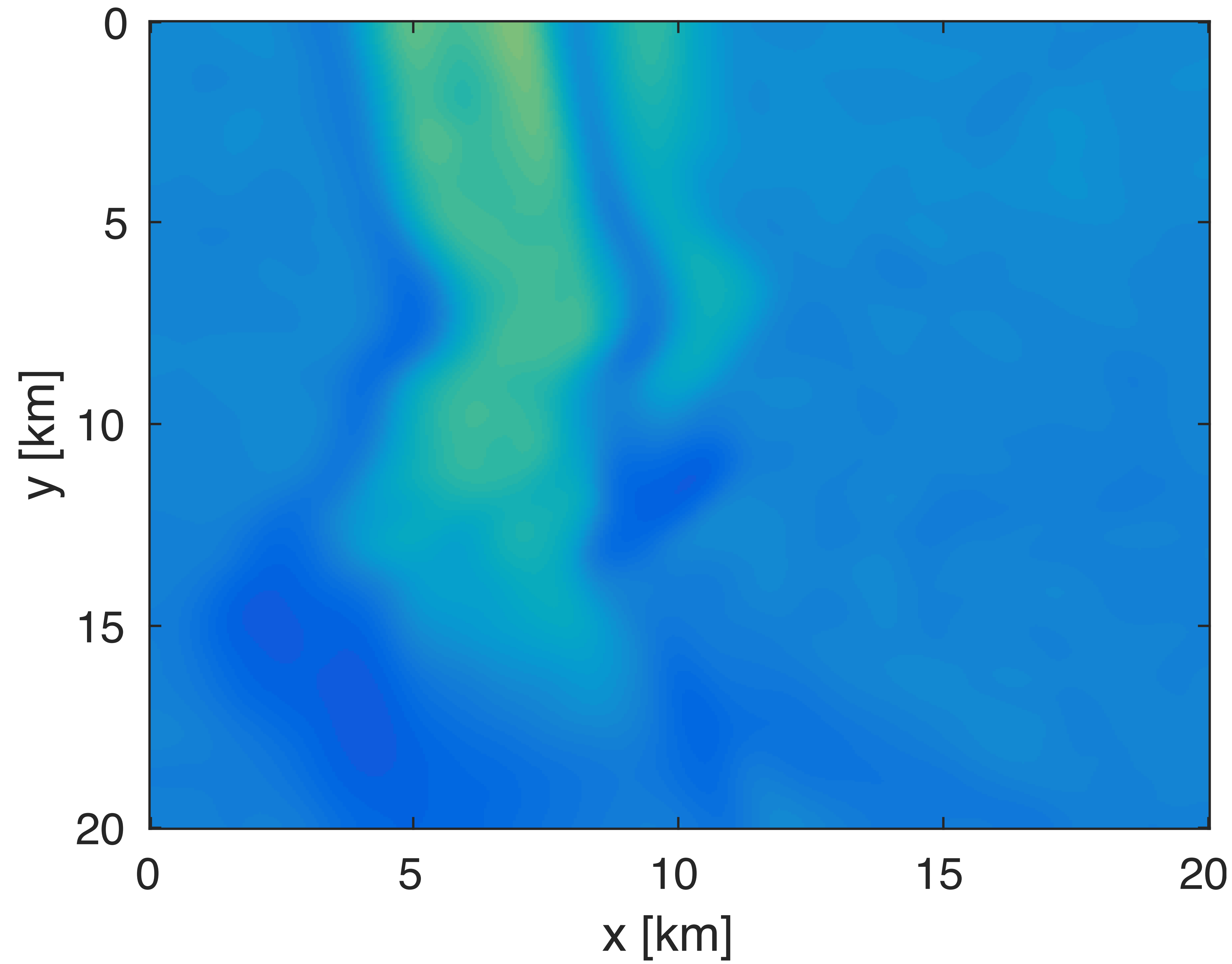
- full data
- compressed data

Exactly the same source indices chosen by the algorithm in all three trials

z=1000m slice

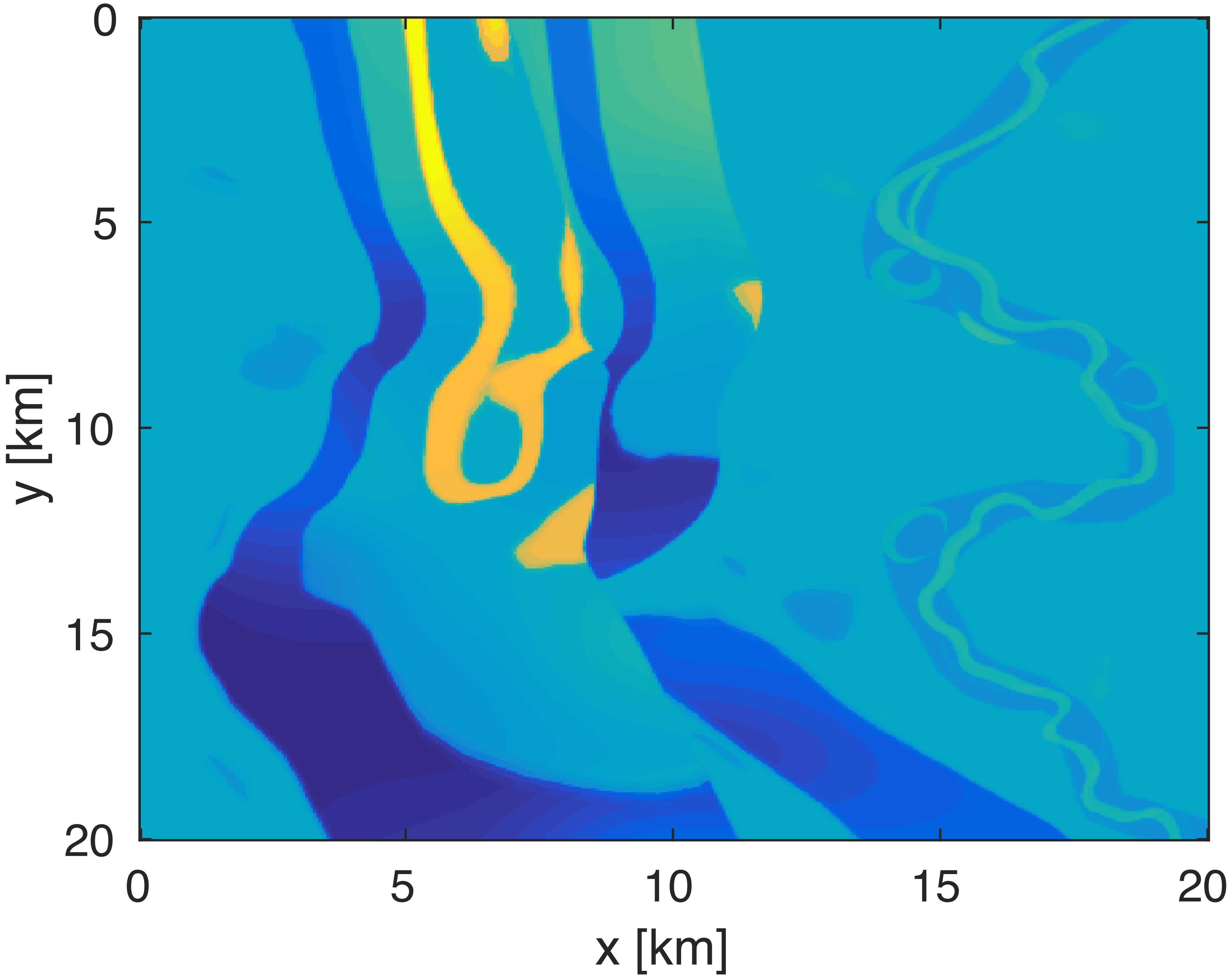


True model

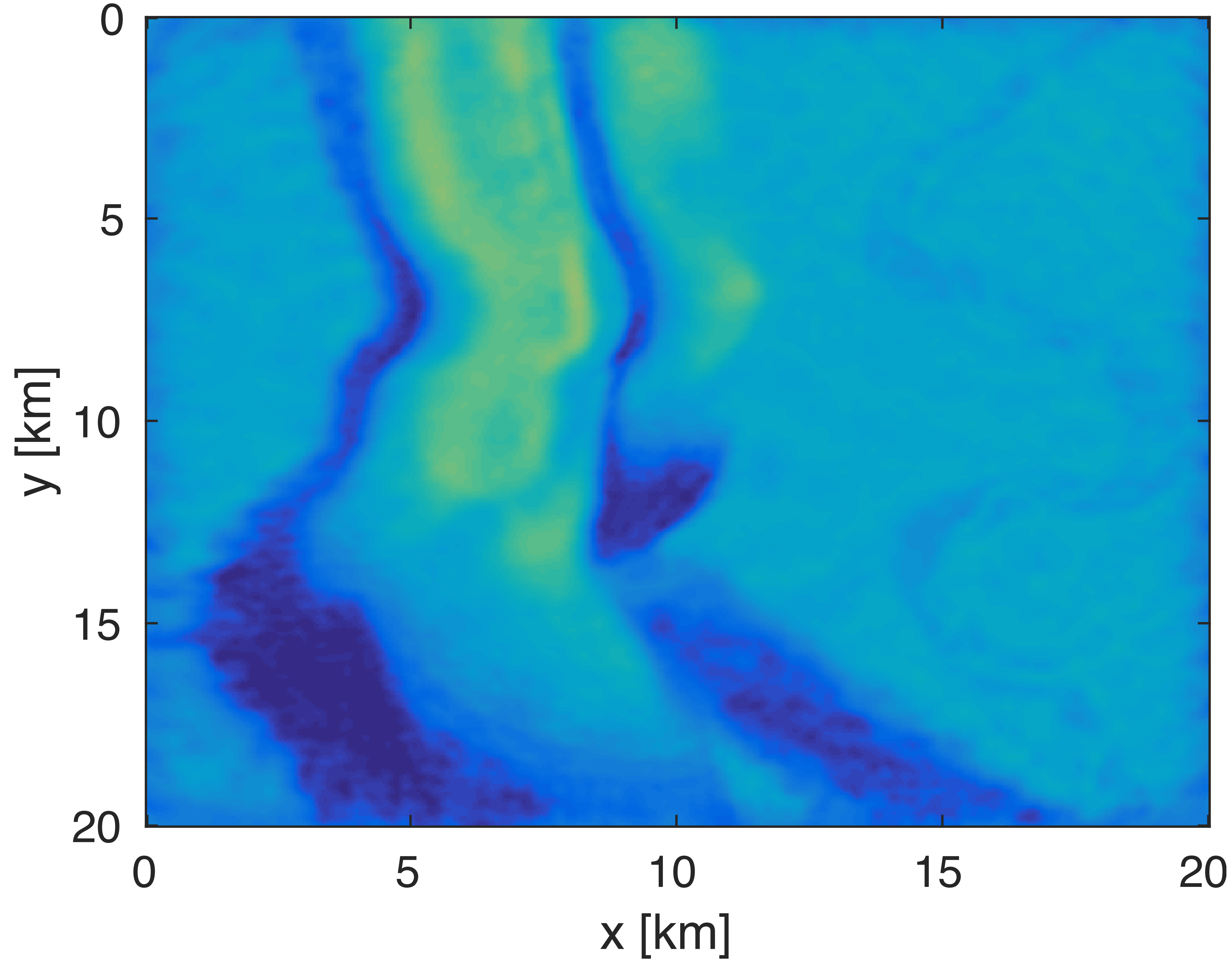


Initial model

z=1000m slice

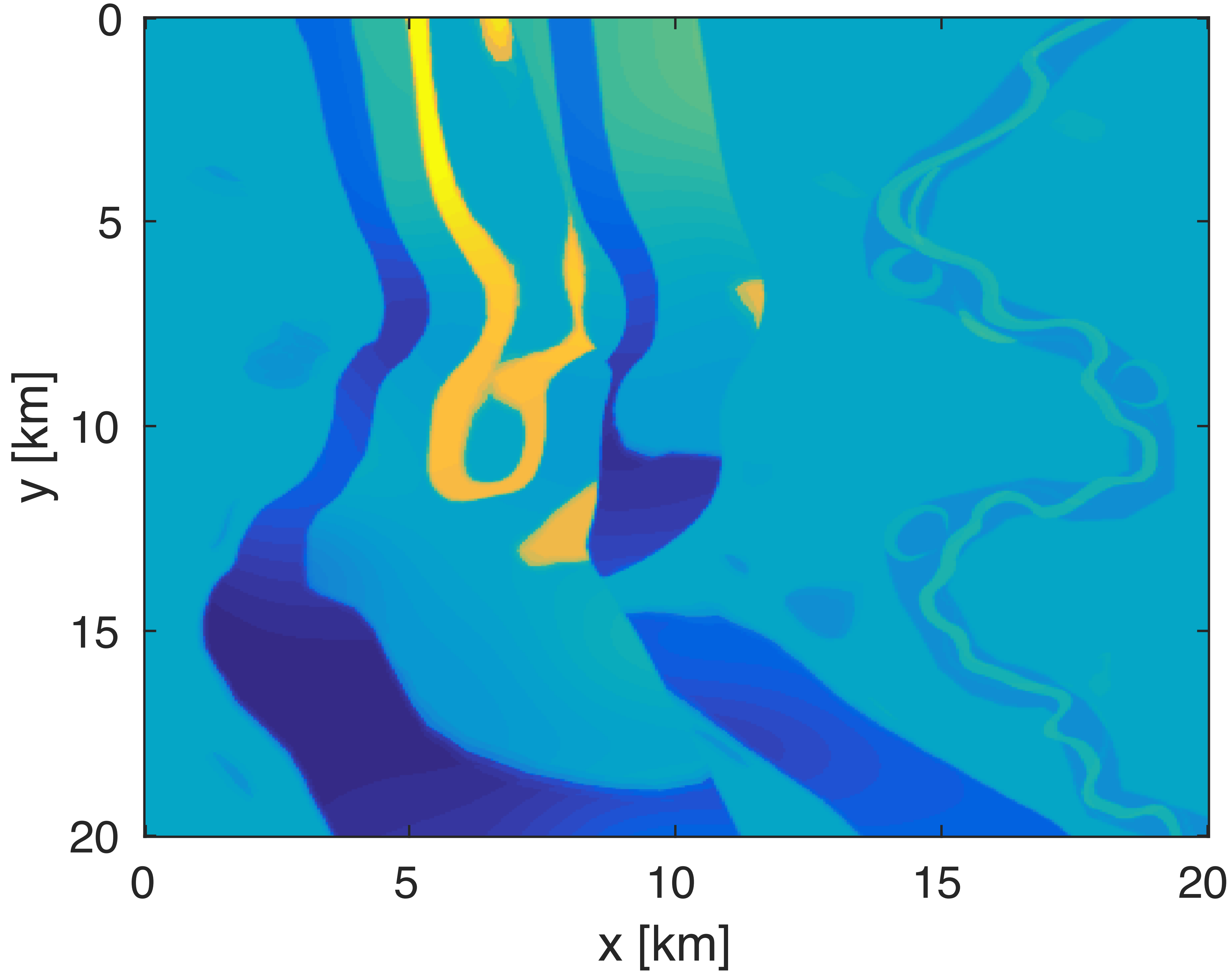


True model

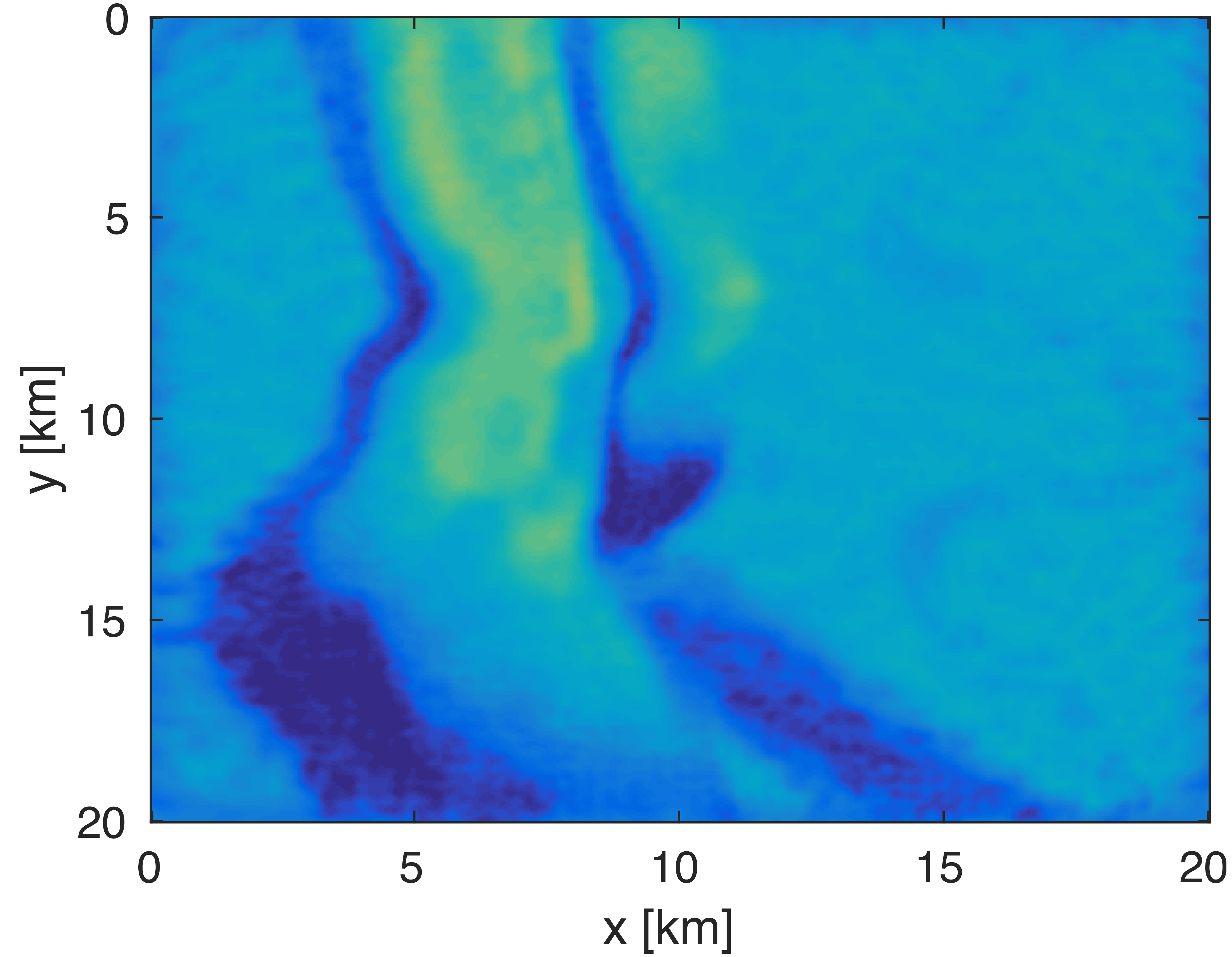


Full data

z=1000m slice

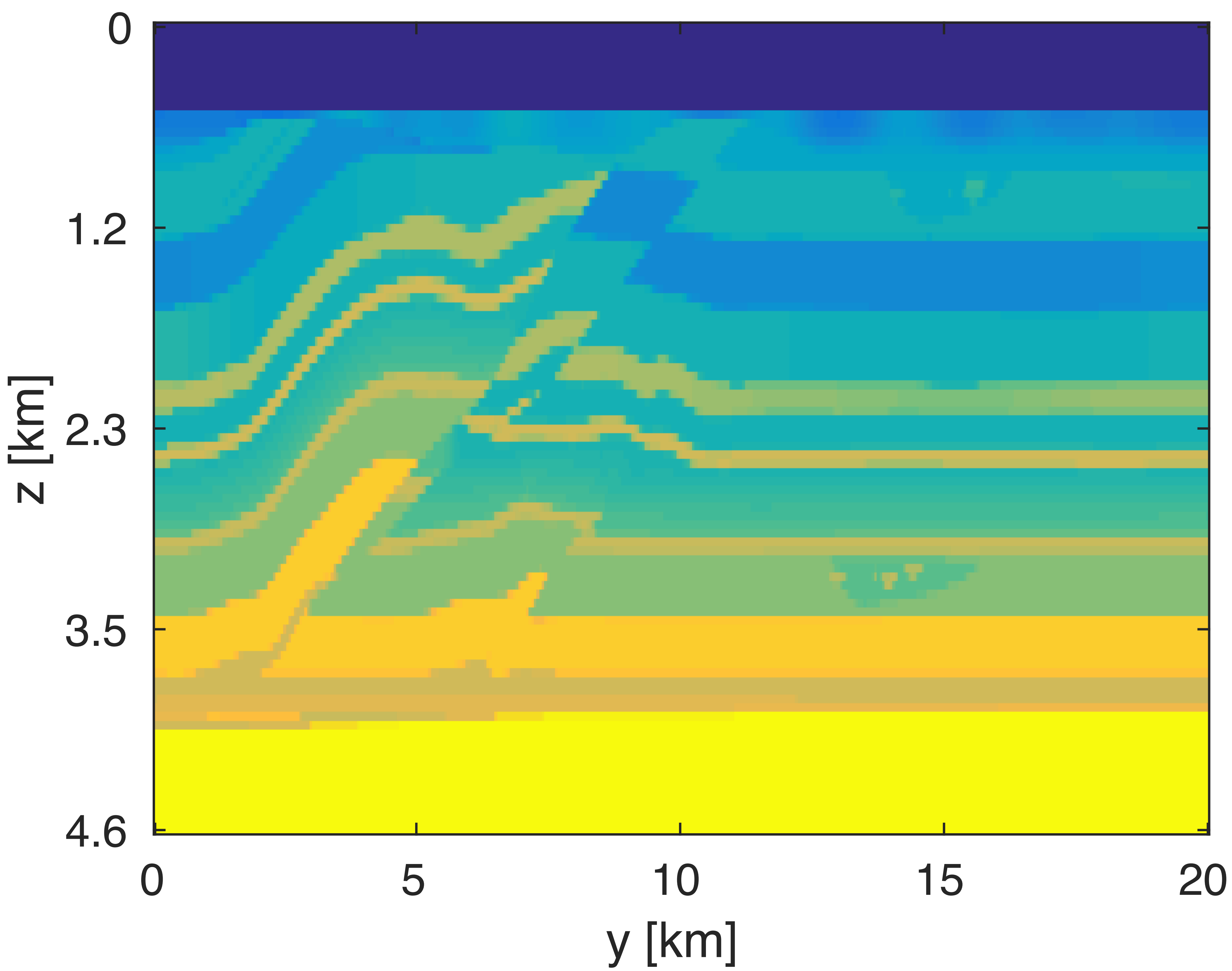


True model

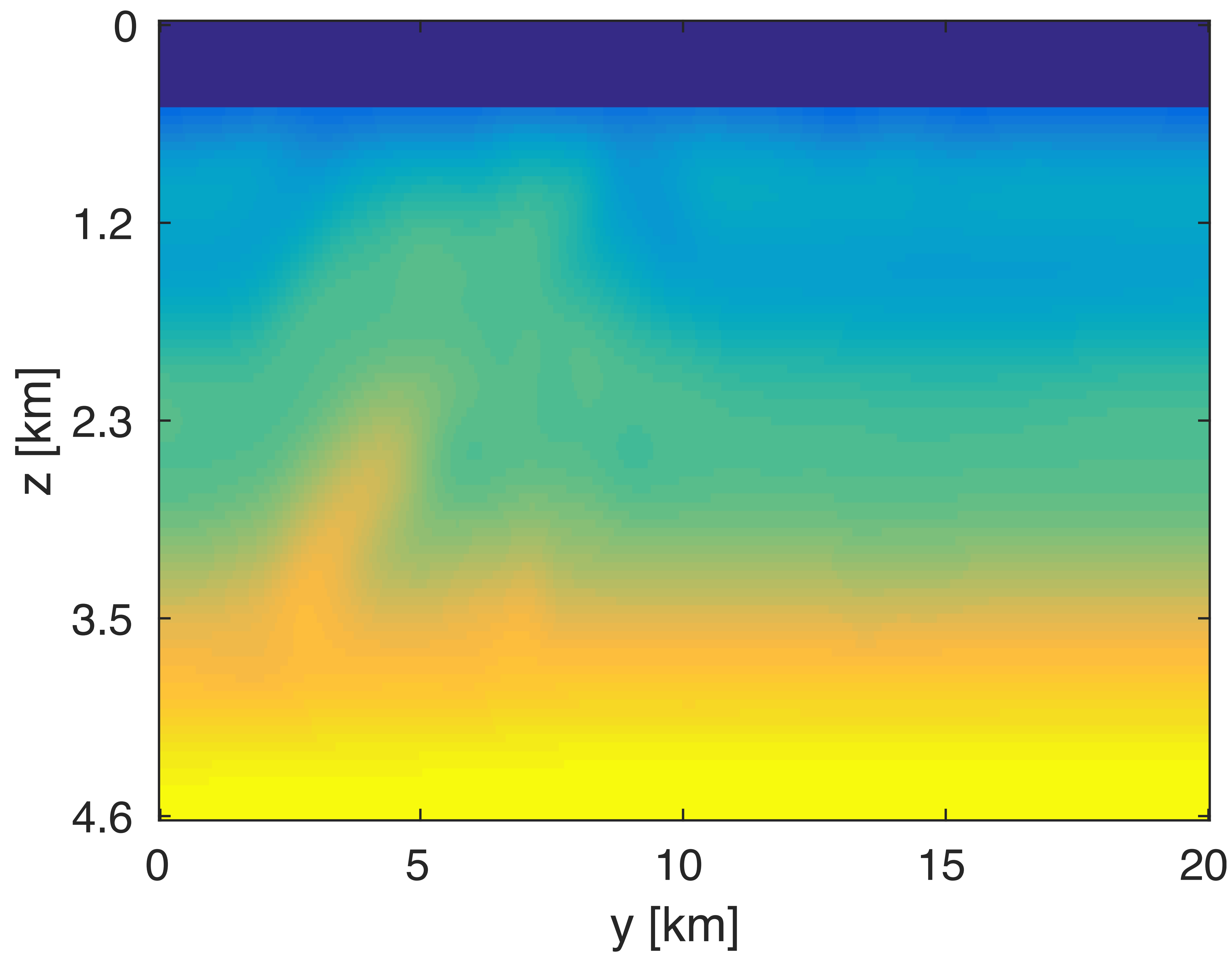


Compressed data

x=12.5km slice

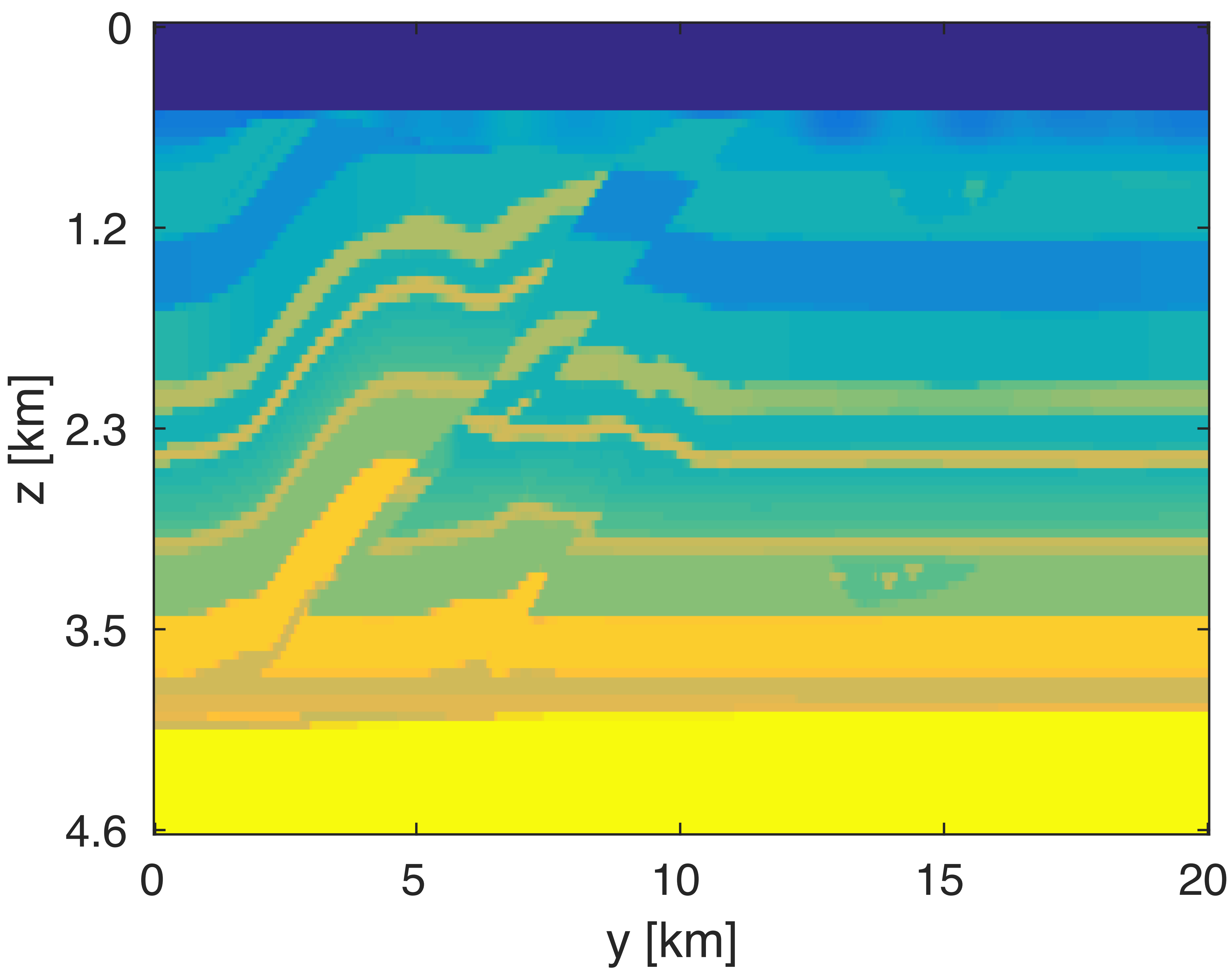


True model

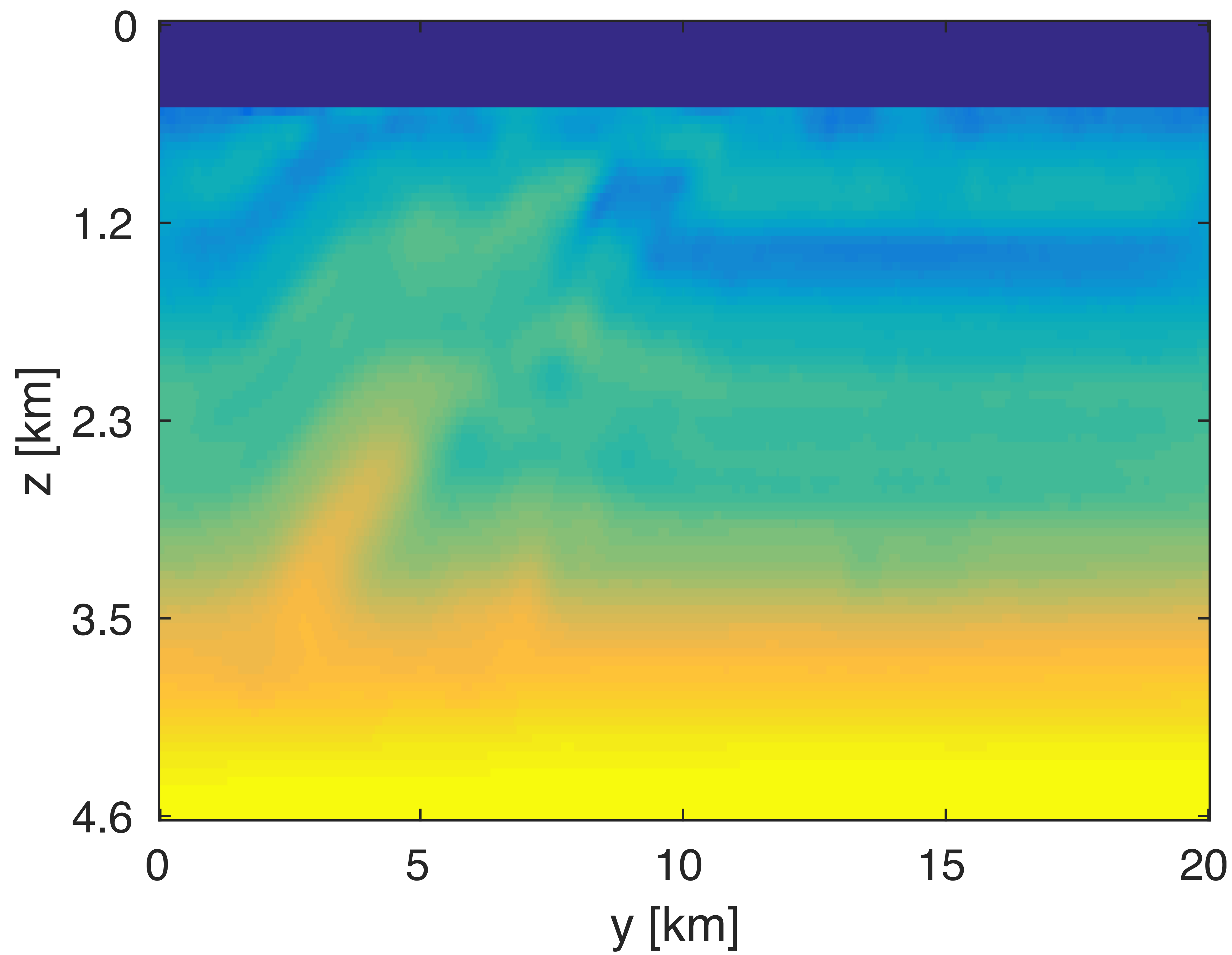


Initial model

x=12.5km slice

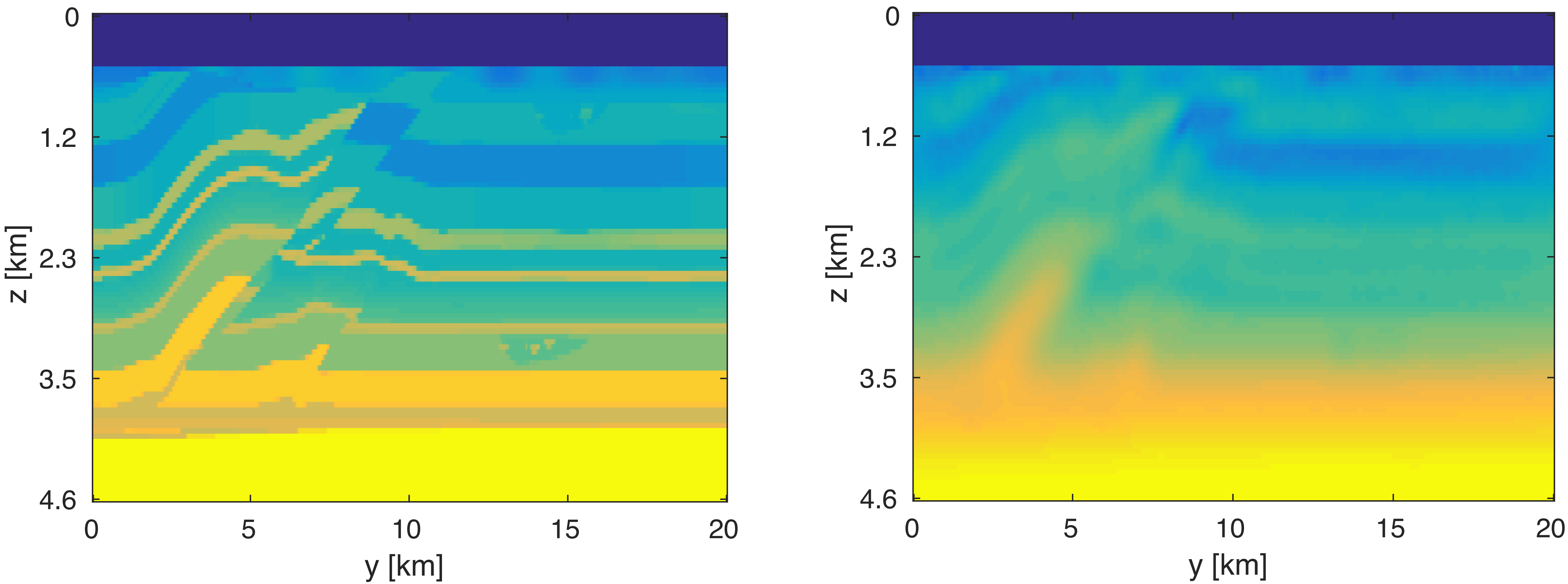


True model



Full data

x=12.5km slice



True model

Compressed data

Data Compressed FWI

We can still work with compressed data in a 3D FWI workflow

- reduced memory costs
- comparable results

Trickier for higher frequencies, but still workable

2D & 3D Image volumes

Rajiv Kumar, Curt Da Silva, Yiming Zhang and Felix J. Herrmann



SLIM
University of British Columbia

Motivation

Form subsurface offset image volumes

Velocity analysis

Targeted imaging

SLIM

Motivation

Computation of full-subsurface offset volumes is prohibitively expensive in 3D (storage & computation time)

Past

Can **not** form full *E* **but** *action* on (random) vectors allows us to get information from *all* or *subsets* of *subsurface points*

Motivation

Computation of full-subsurface offset volumes is prohibitively expensive in 3D (storage & computation time)

Present

Can ~~**not**~~ form full *E* **using** *action* on (random) vectors allows us to get information from *all* or *subsets* of *subsurface points*

Efficient ways to extract information from highly compressed image volumes

Game changer for 3D WEMVA

Extended images

Given two-way wave equations, source and receiver wavefields are defined as

$$H(\mathbf{m})U = P_s^T Q$$
$$H(\mathbf{m})^* V = P_r^T D$$

where

$H(\mathbf{m})$: discretization of the Helmoltz operator

Q : source

D : data matrix

P_s, P_r : samples the wavefield at the source and receiver positions

$\mathbf{m}$: slowness

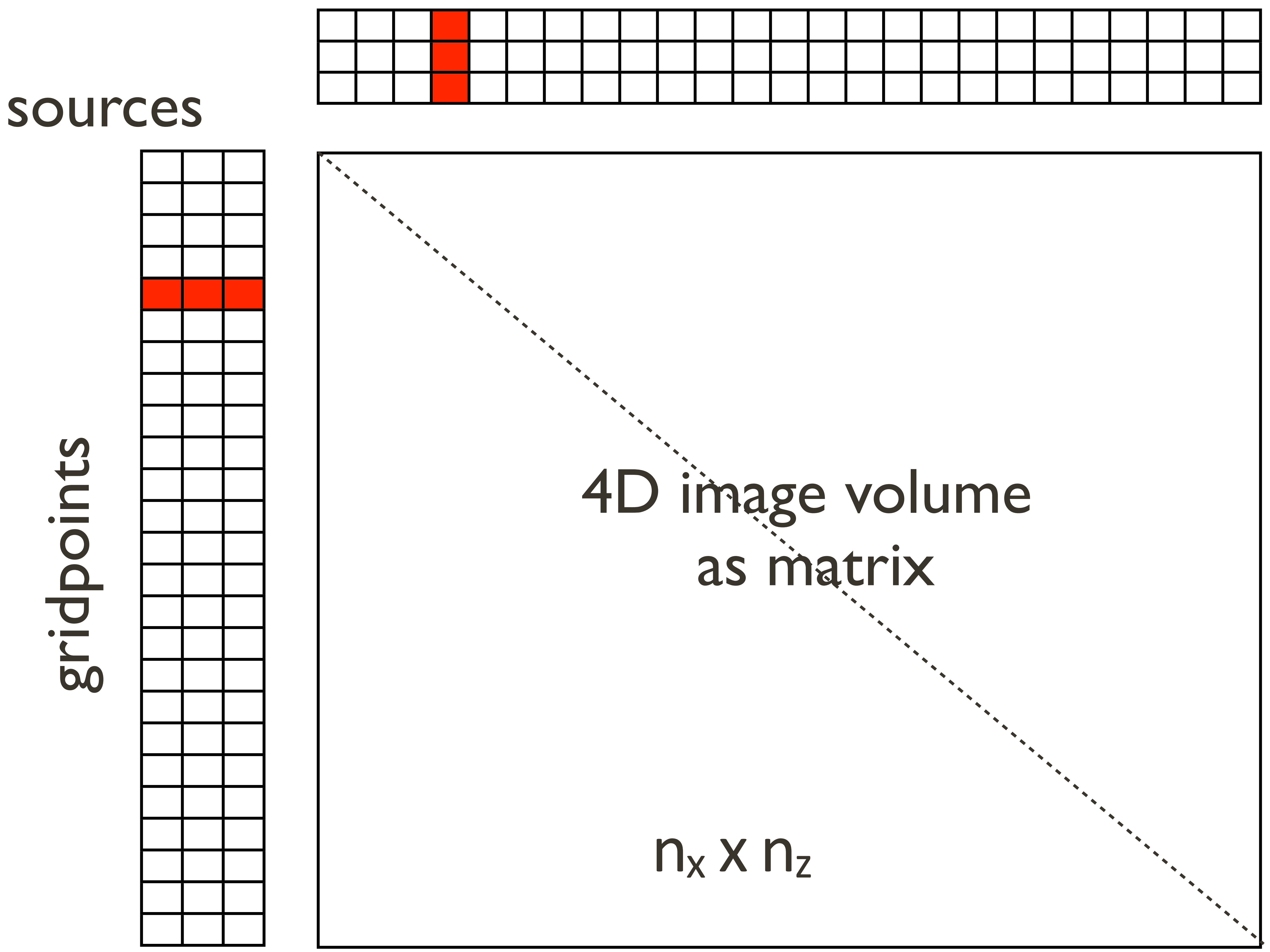
Extended images

Organize wavefields in monochromatic data *matrices* where each *column* represents a *common* shot gather

Express image volume *tensor* for *single* frequency as a *matrix*

$$E = UV^*$$

Extended images



In 3D, E is 6D tensor for each monochromatic slice

Tristan van Leeuwen, Rajiv Kumar, and Felix J. Herrmann, "Enabling affordable omnidirectional subsurface extended image volumes via probing", Geophysical Prospecting, 2016

SLIM

Extended images (Past)

Too expensive to compute *(storage and computational time)*

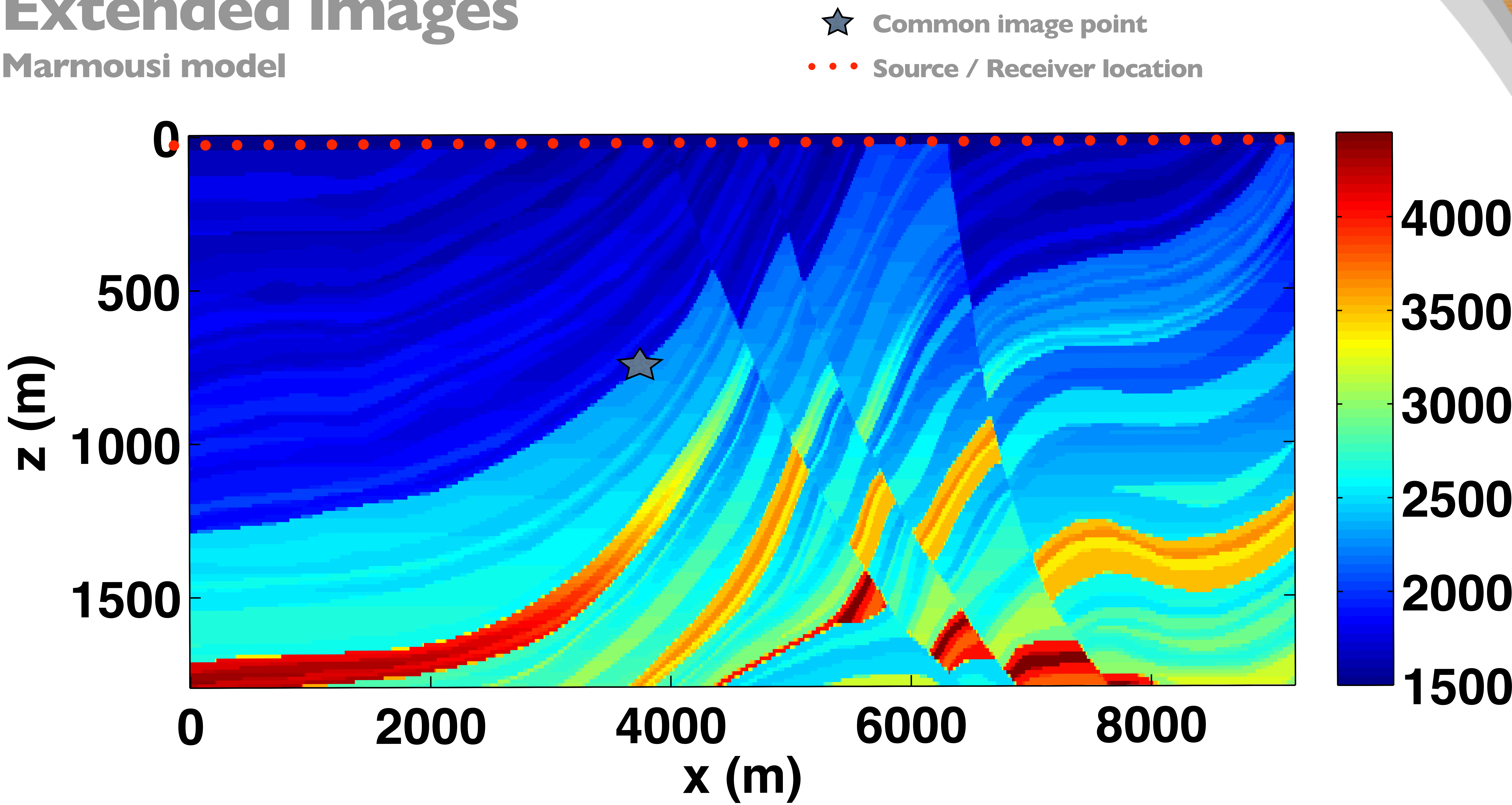
Instead, *probe* volume with *tall* matrix $W = [\mathbf{w}_1, \ldots, \mathbf{w}_l]$

$$\widetilde{E} = EW = H^{-1} P_s^T Q D^* P_r H^{-1} W$$

where $\mathbf{w}_i = [0, \ldots, 0, 1, 0, \ldots, 0]$ represents *single* scattering points

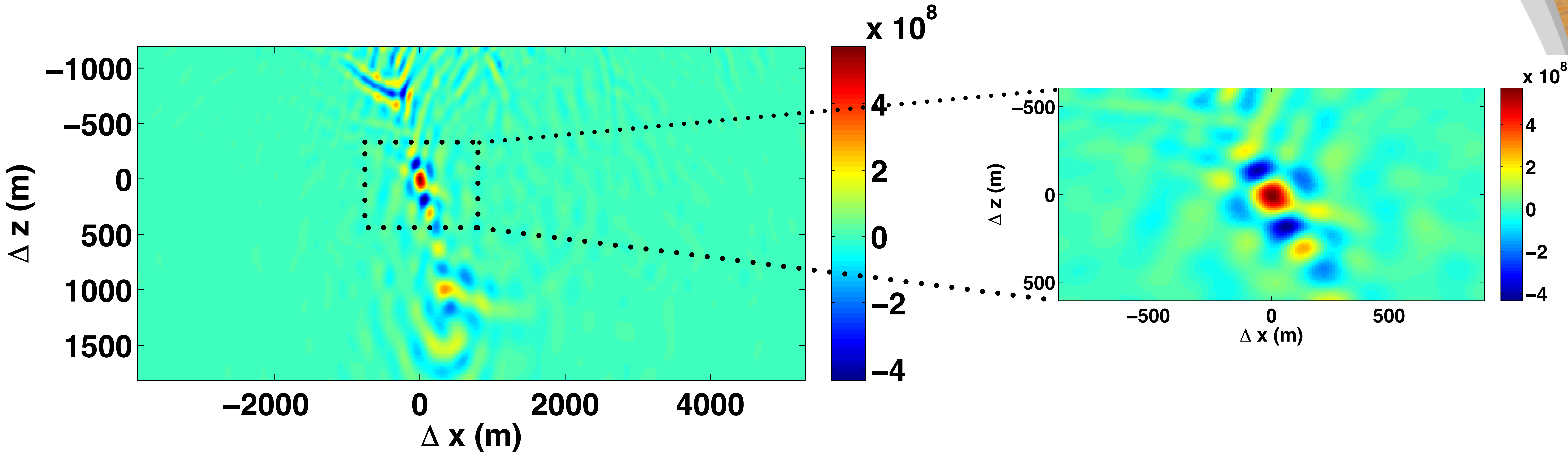
Extended images

Marmousi model



Extended images

common image point gather, 3- 30 Hz



Δ x : Horizontal offset

Δ z : Vertical offset

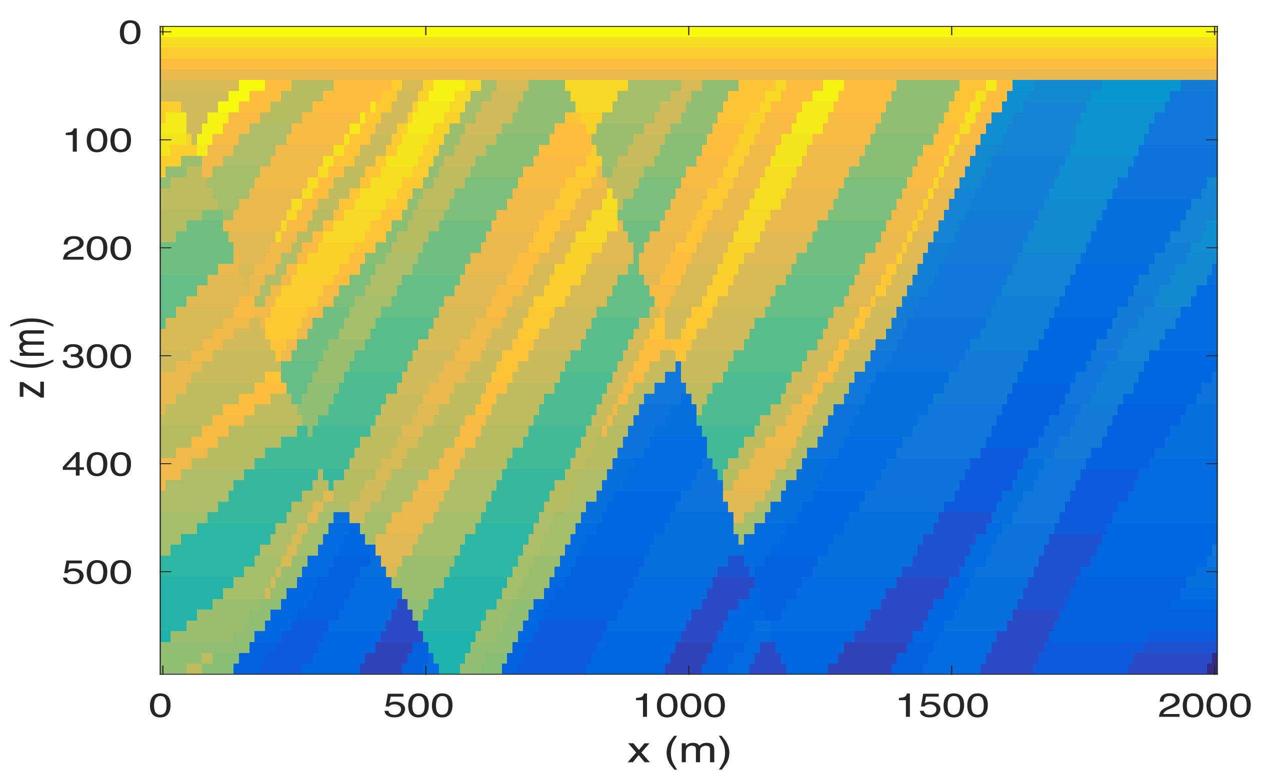
SLIM

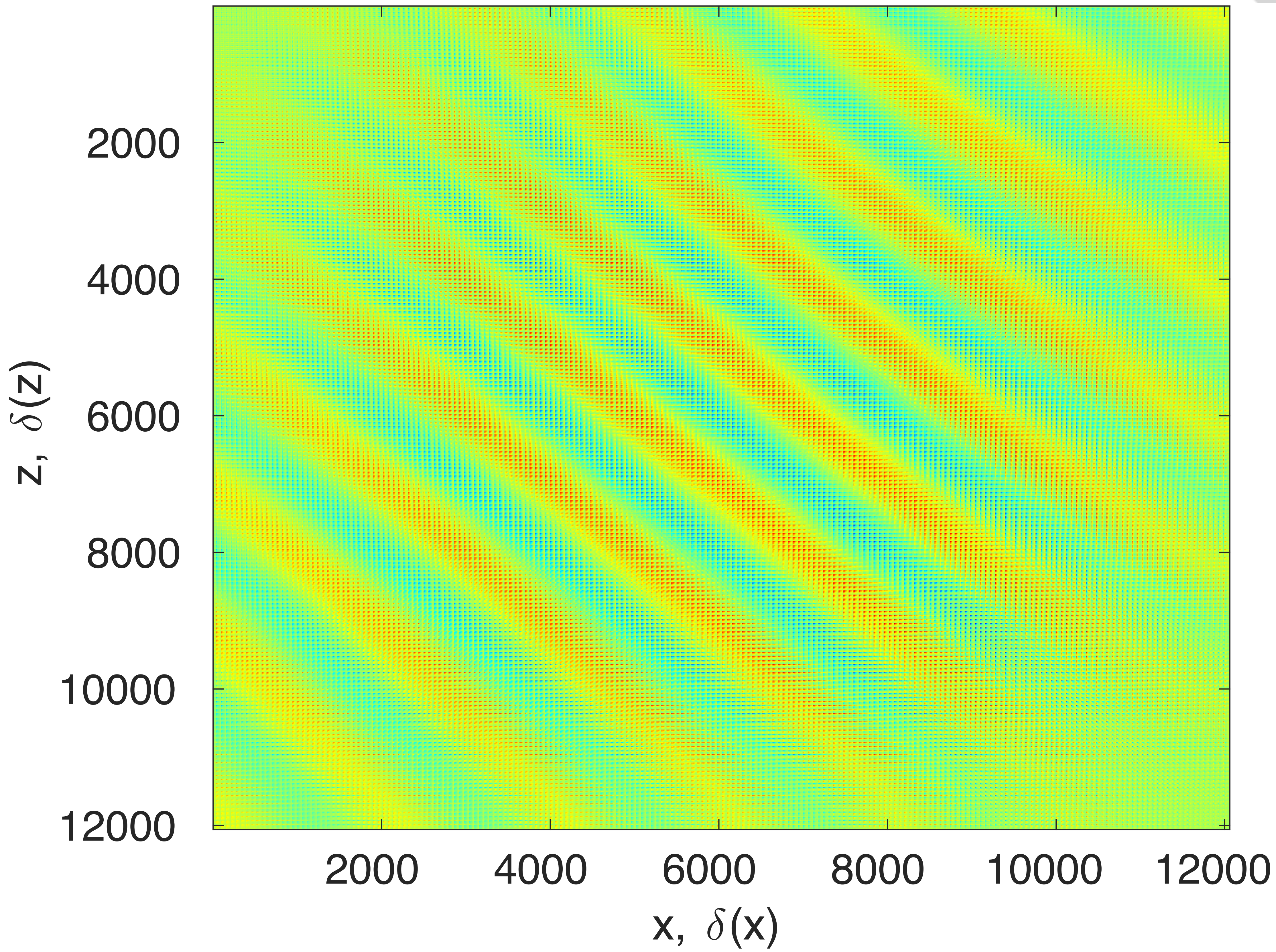
Slicing & Dicing of image volumes

$Y = E(m)W$	Probe full-extended image volume with virtual source
$[Q, R] = \mathrm{qr}(Y)$	QR factorization
$Z = Q^H E(m)$	Probe again with new virtual source
$[U, S, V] = \mathrm{svd}(Z)$	SVD factorization (first few singular values) *
$U \leftarrow QU$	Compute right singular vectors

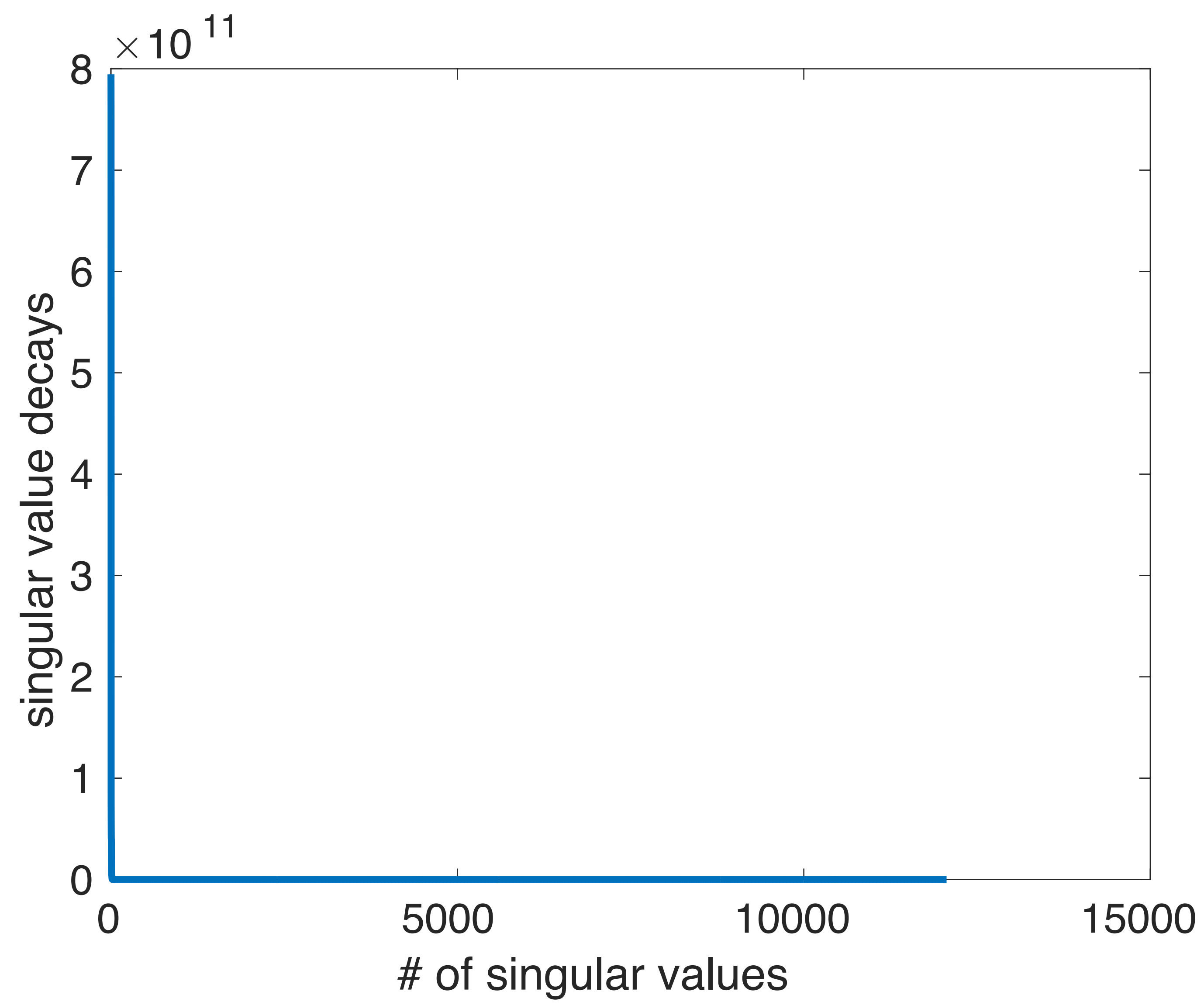
Low-rank representation 5Hz

60 x 201



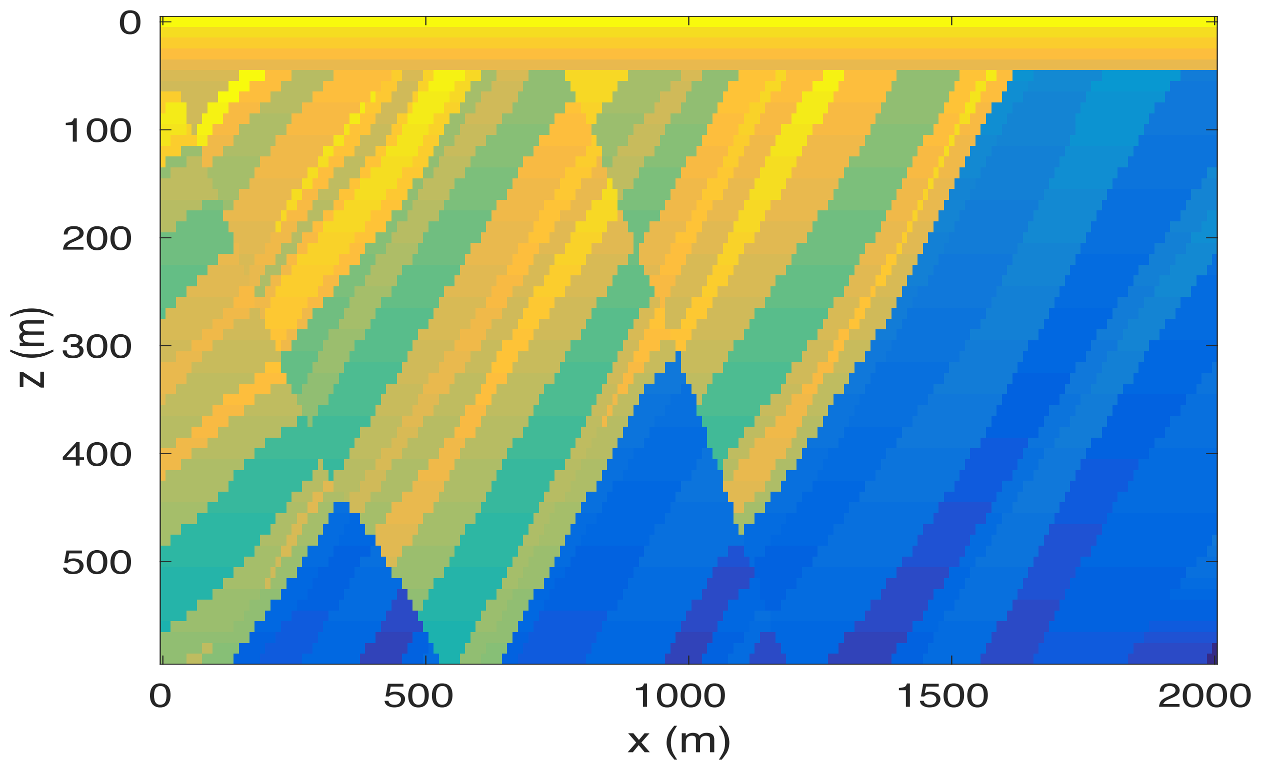


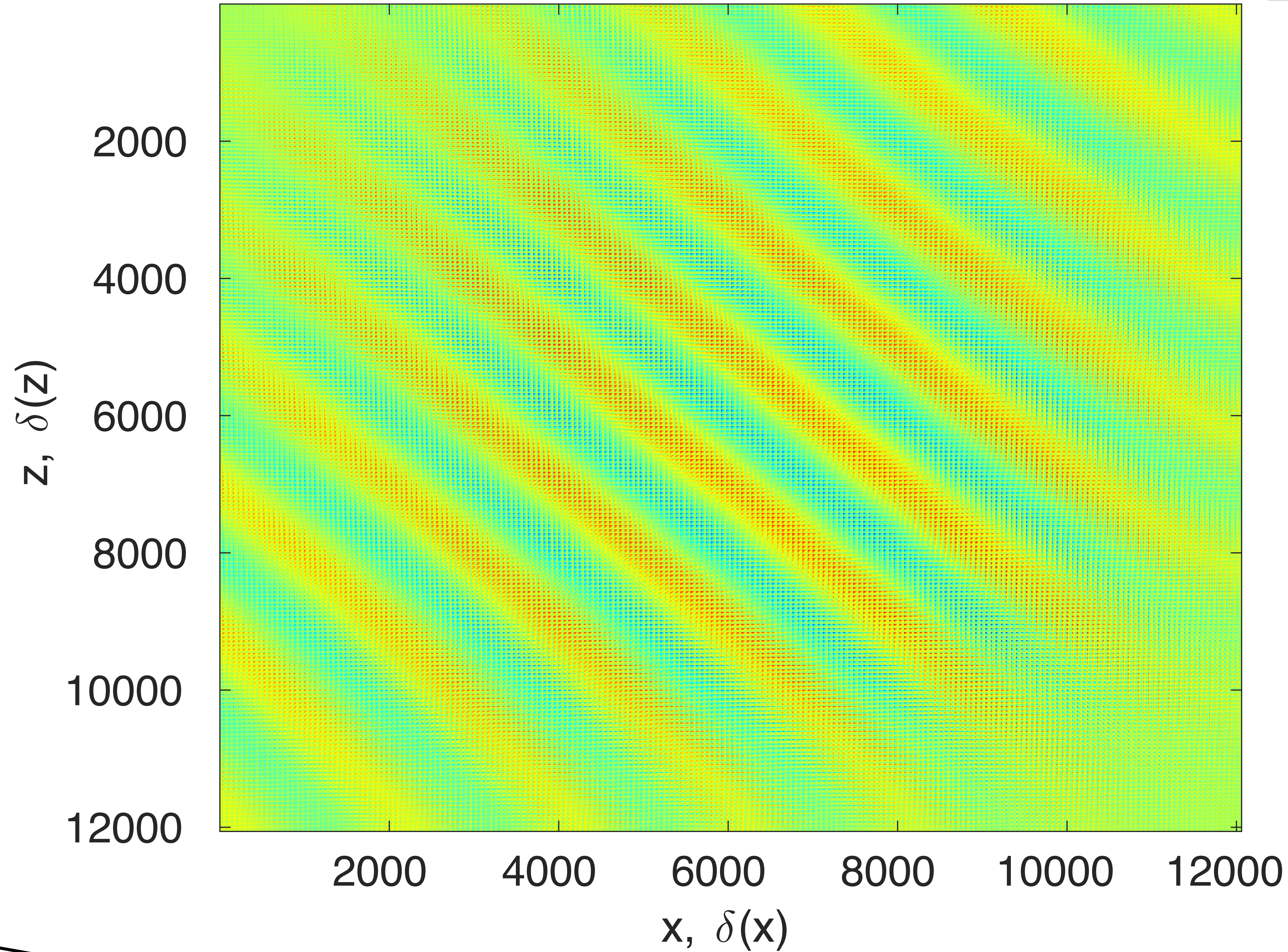
Full E



Low-rank representation 5Hz

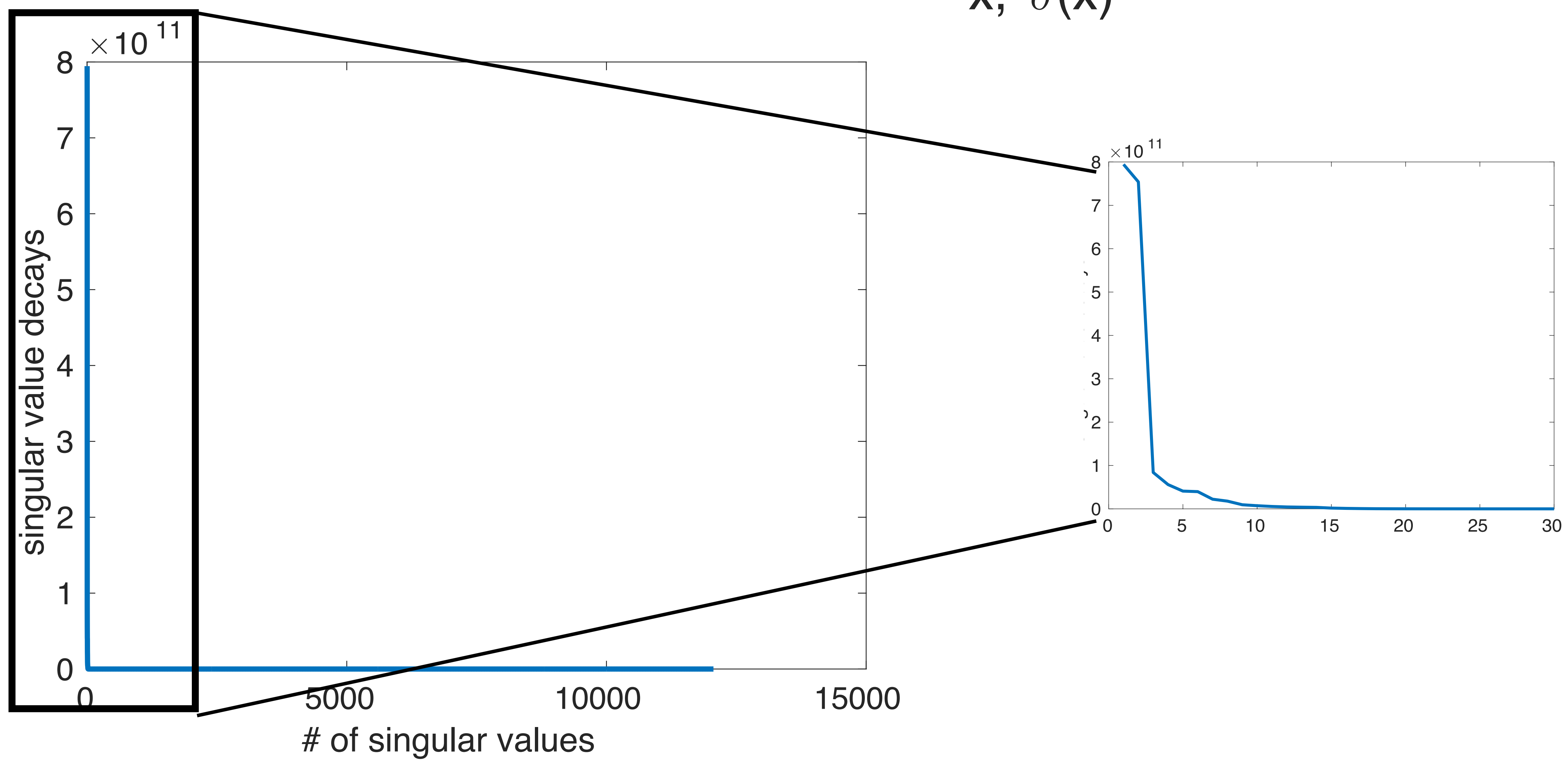
60 x 201





Full E

10 singular vector
error = 1e-04



Take-away message

Computational costs

Full subsurface offset extended images:

	# of PDE solves	"flops for correlations"
conventional	$2N_s$	$N_s \times N_h$
mat-vecs	$4N_x$	$N_s \times N_r$

N_s – # of sources
N_r – # of receivers
N_h – # of subsurface offsets
N_x – # of sample points

Take-away message

Computational costs

Full subsurface offset extended images:

	# of PDE solves	"flops for correlations"
conventional	$2N_s$	$N_s \times N_h$
mat-vecs	$4N_x$	$N_s \times N_r$

N_s – # of sources
N_r – # of receivers
N_h – # of subsurface offsets
N_x – # of sample points

We win when $N_x << N_s$!

Applications

Image gather for QC

Target-imaging

Wave-equation migration velocity analysis

Image-gather

Experimental details

1200 source (75 m spacing) , 2500 receivers (50 m spacing)

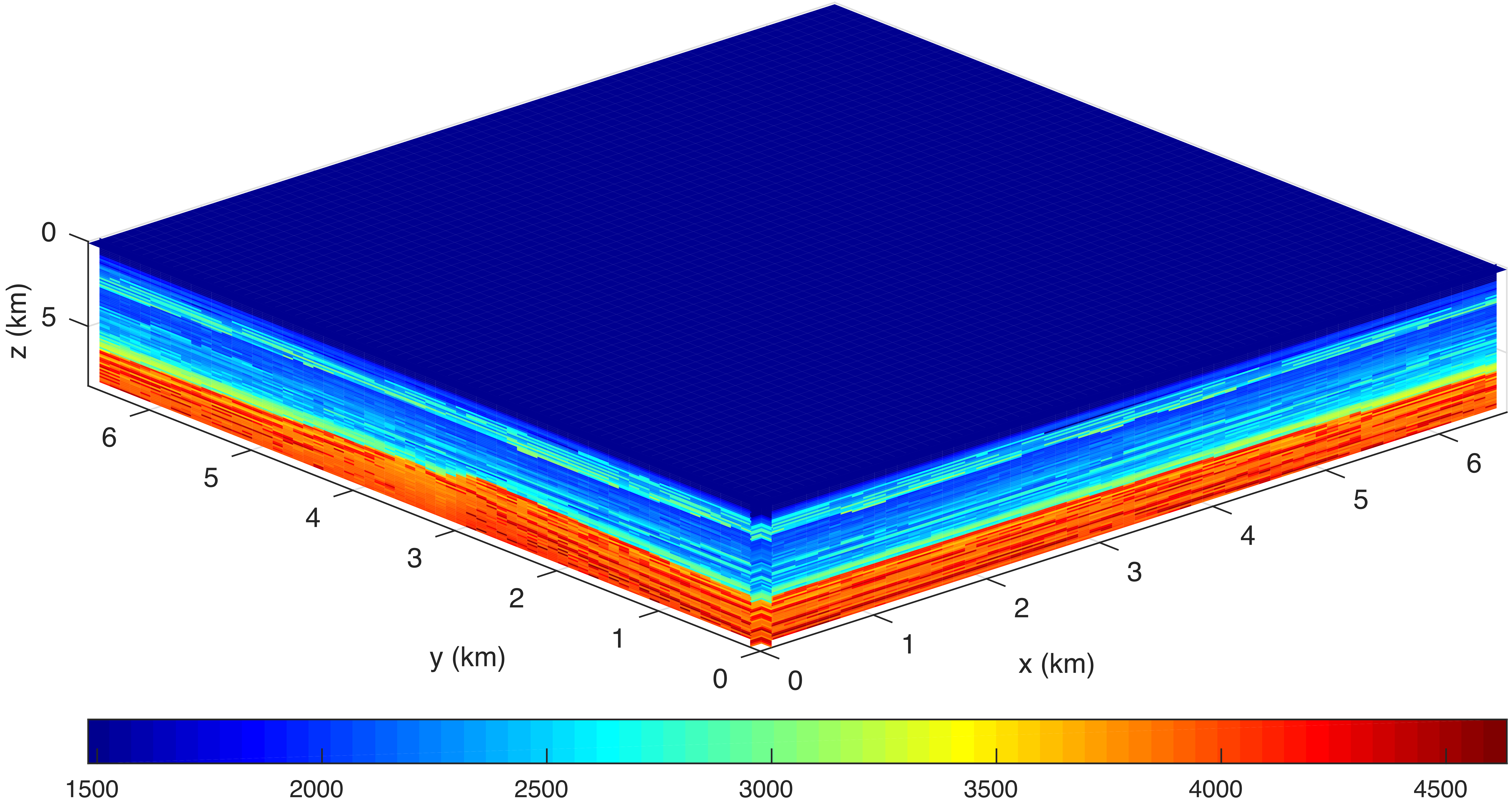
5-12 Hz

OBN acquisition

peak frequency 15 Hz

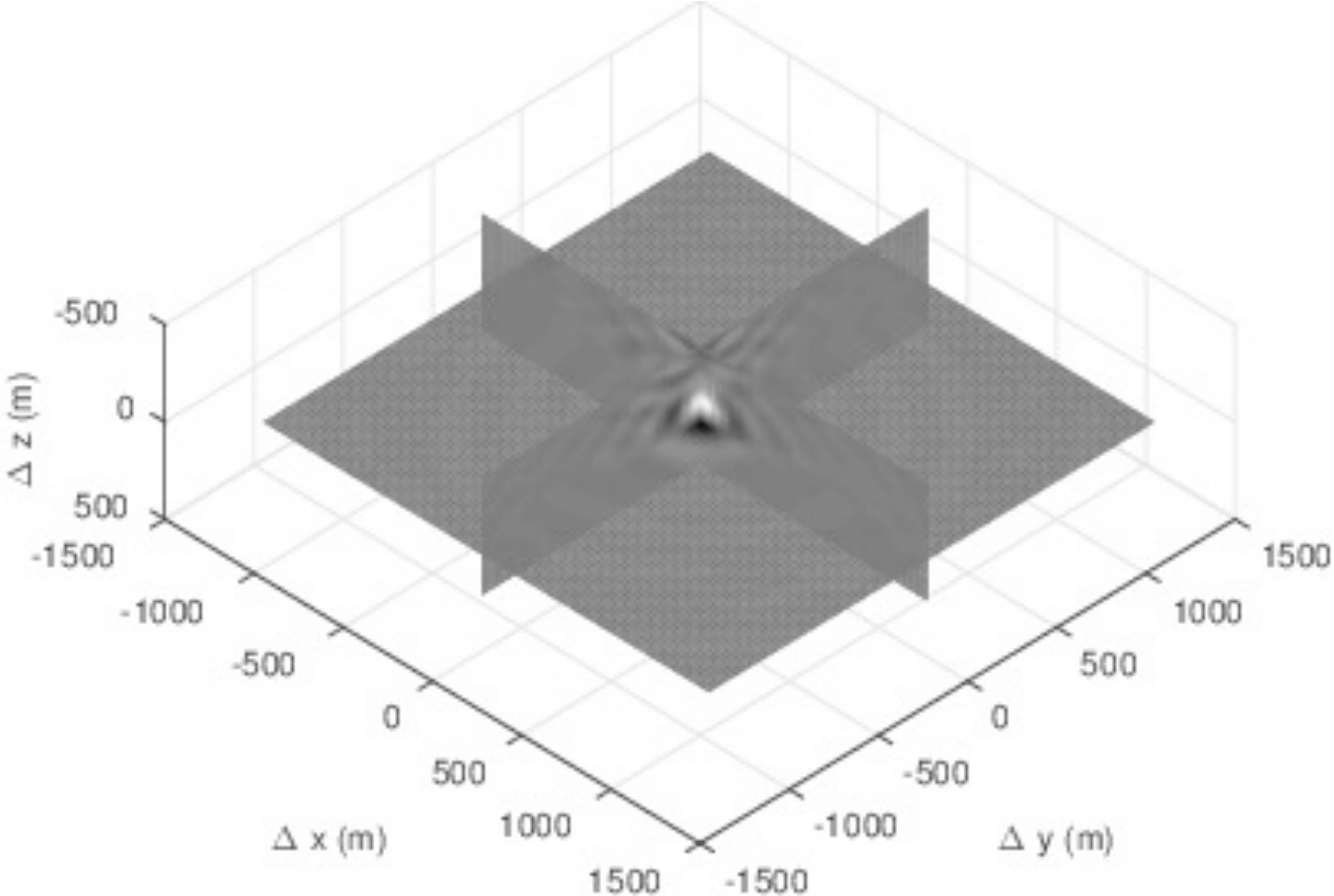
200 probing vectors

3D BG Compass model



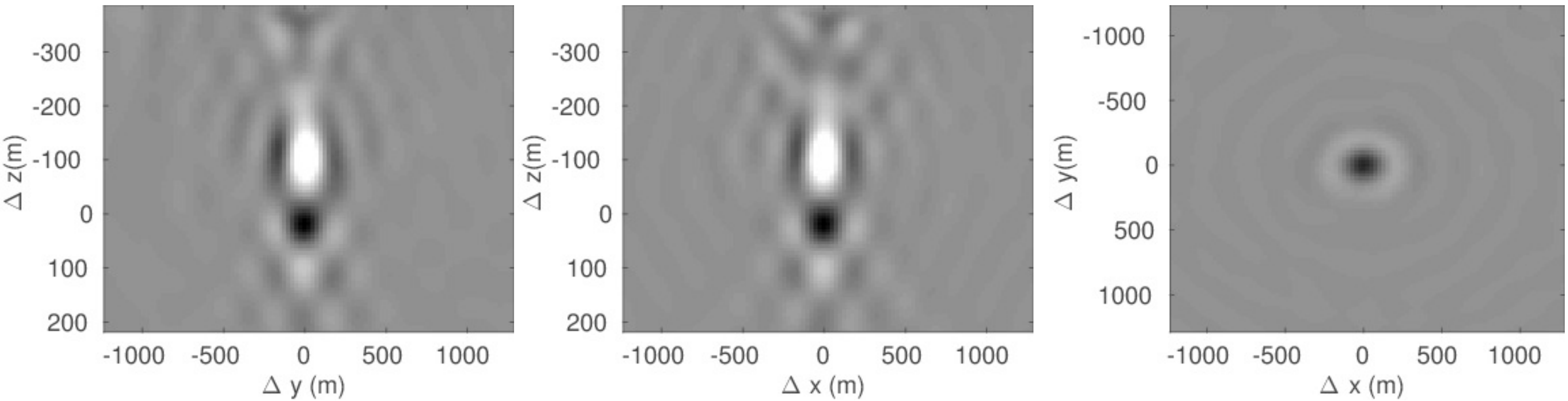
CIG

SLIM



-500
0
500
Δ z (m)
-1500
-1000
-500
0
500
1000
1500
Δ x (m)
1500
1000
500
0
-500
-1000
-1500
Δ y (m)

Cross section across CIG



Target-imaging

Experimental details

251 source (30 m spacing) , 751 receivers (10 m spacing)

5-40 Hz

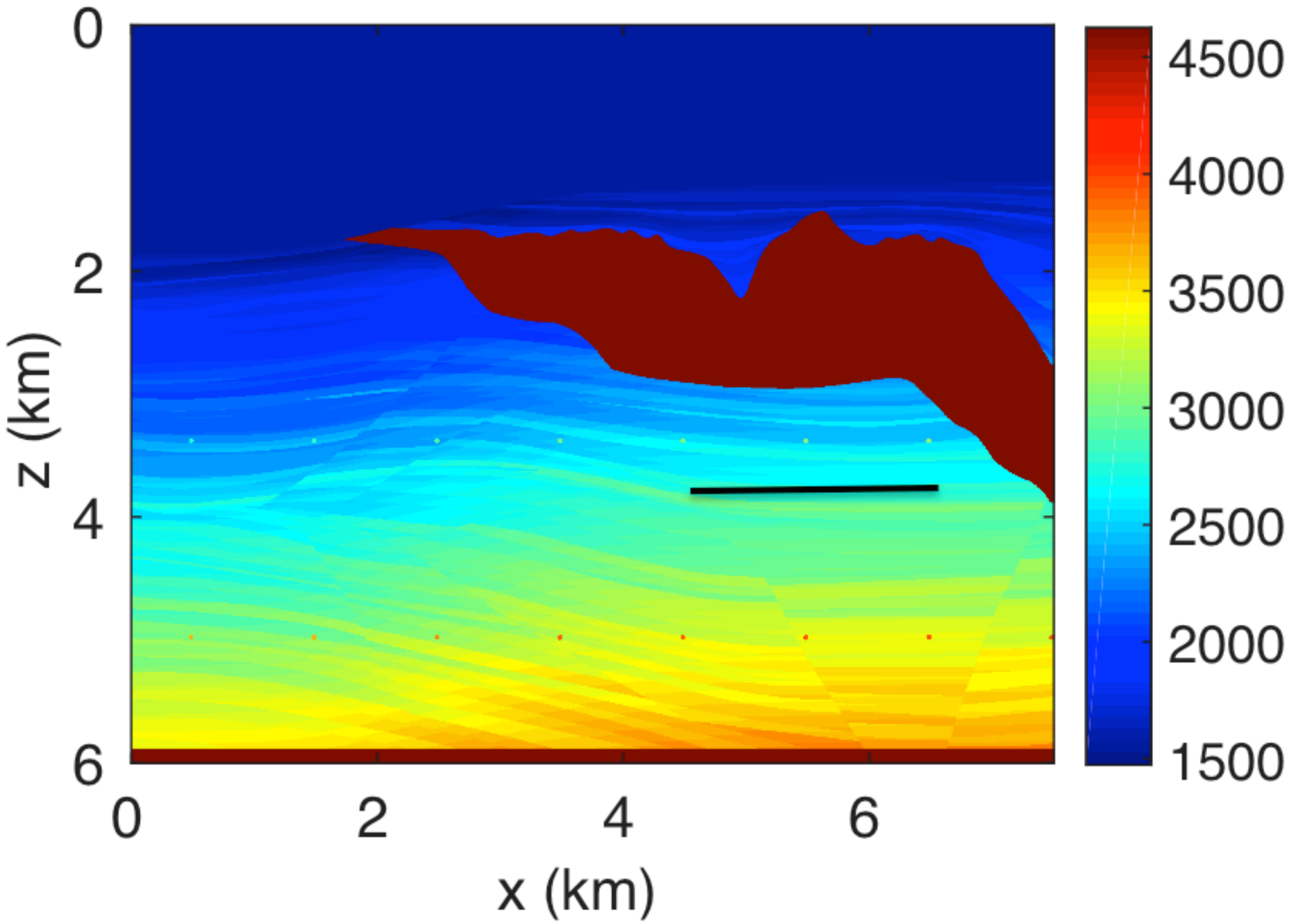
split-spread acquisition

recording length 6s, sampling interval 4ms

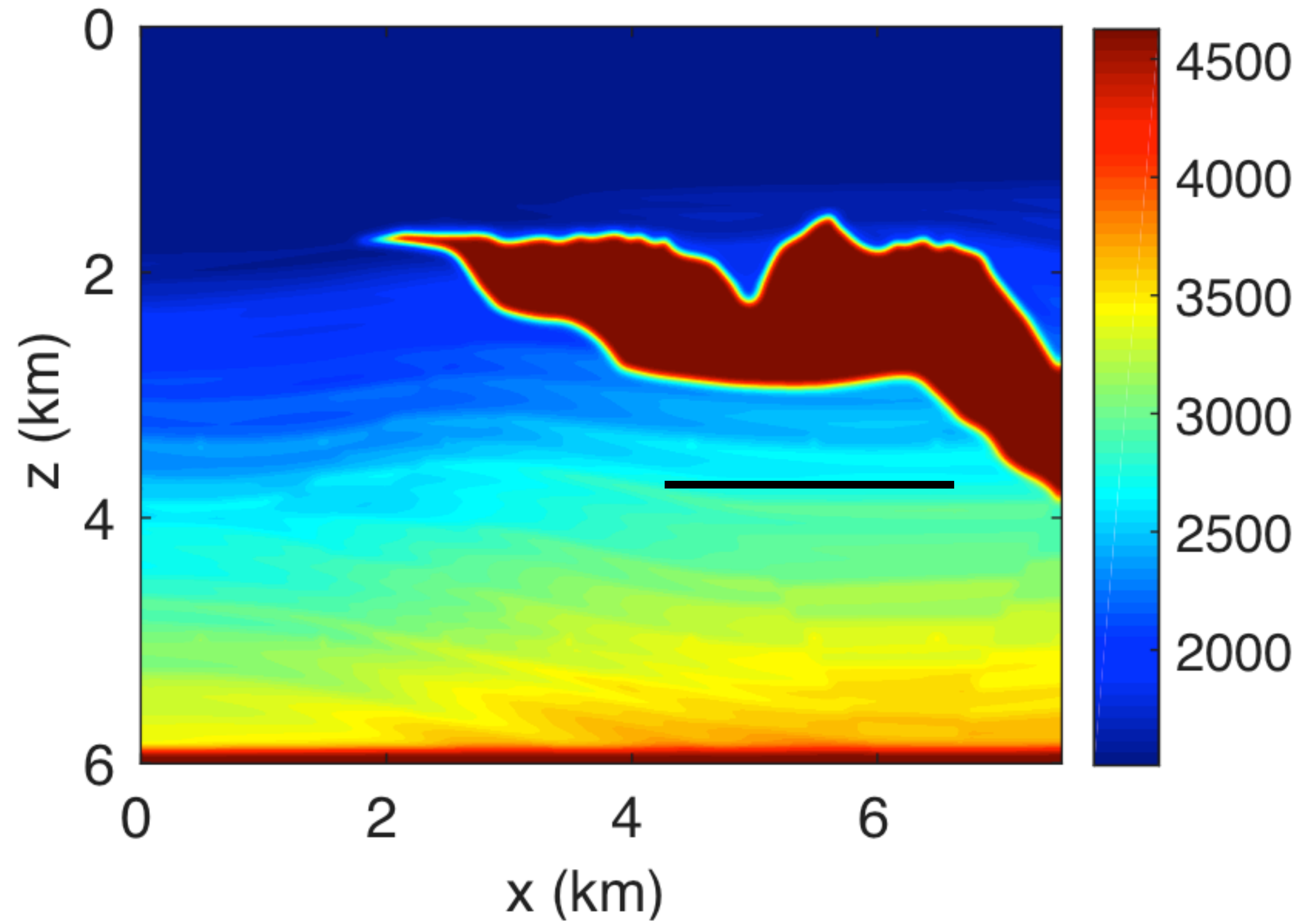
peak frequency 25 Hz

100 probing vectors

Sigsbee model

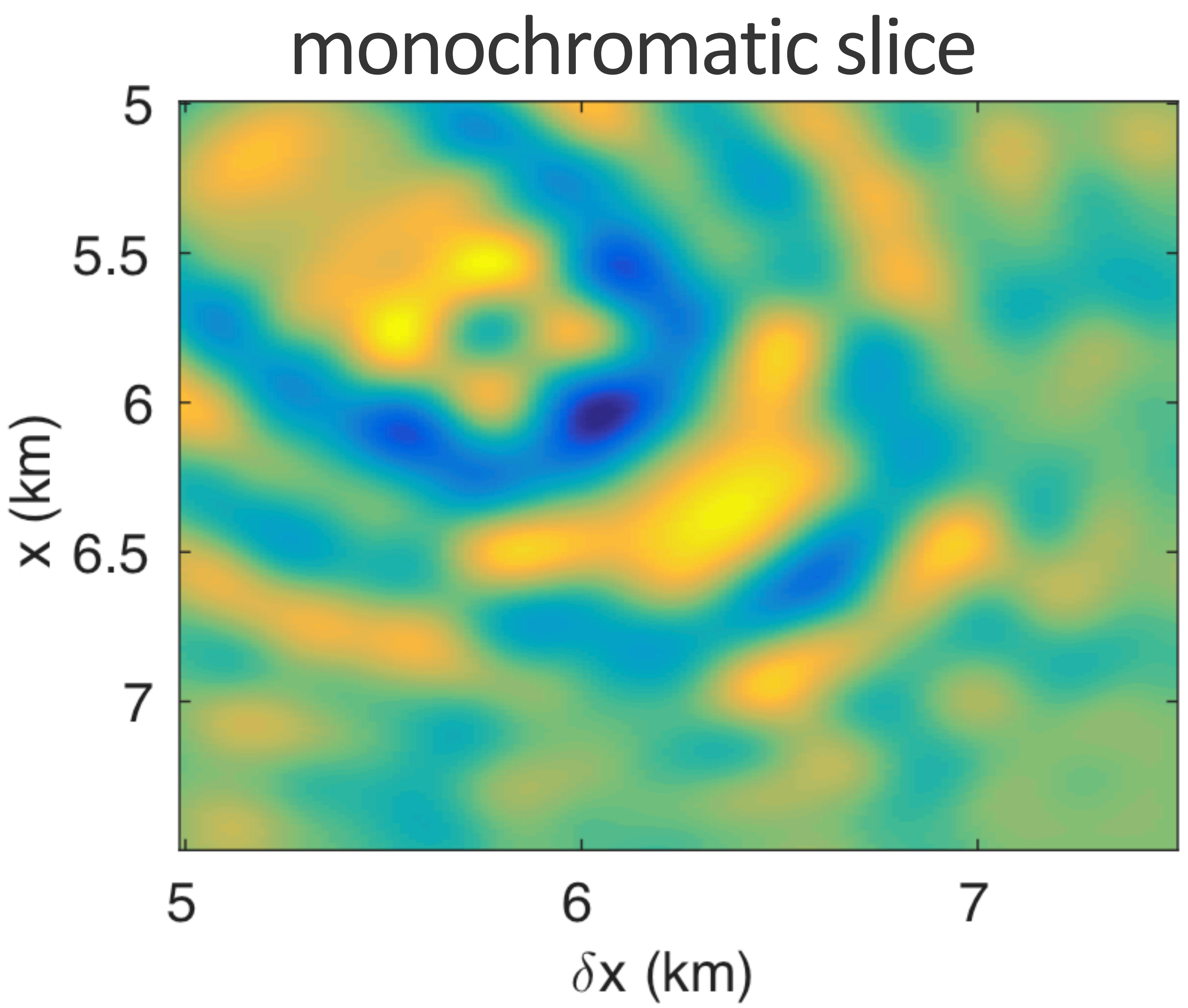


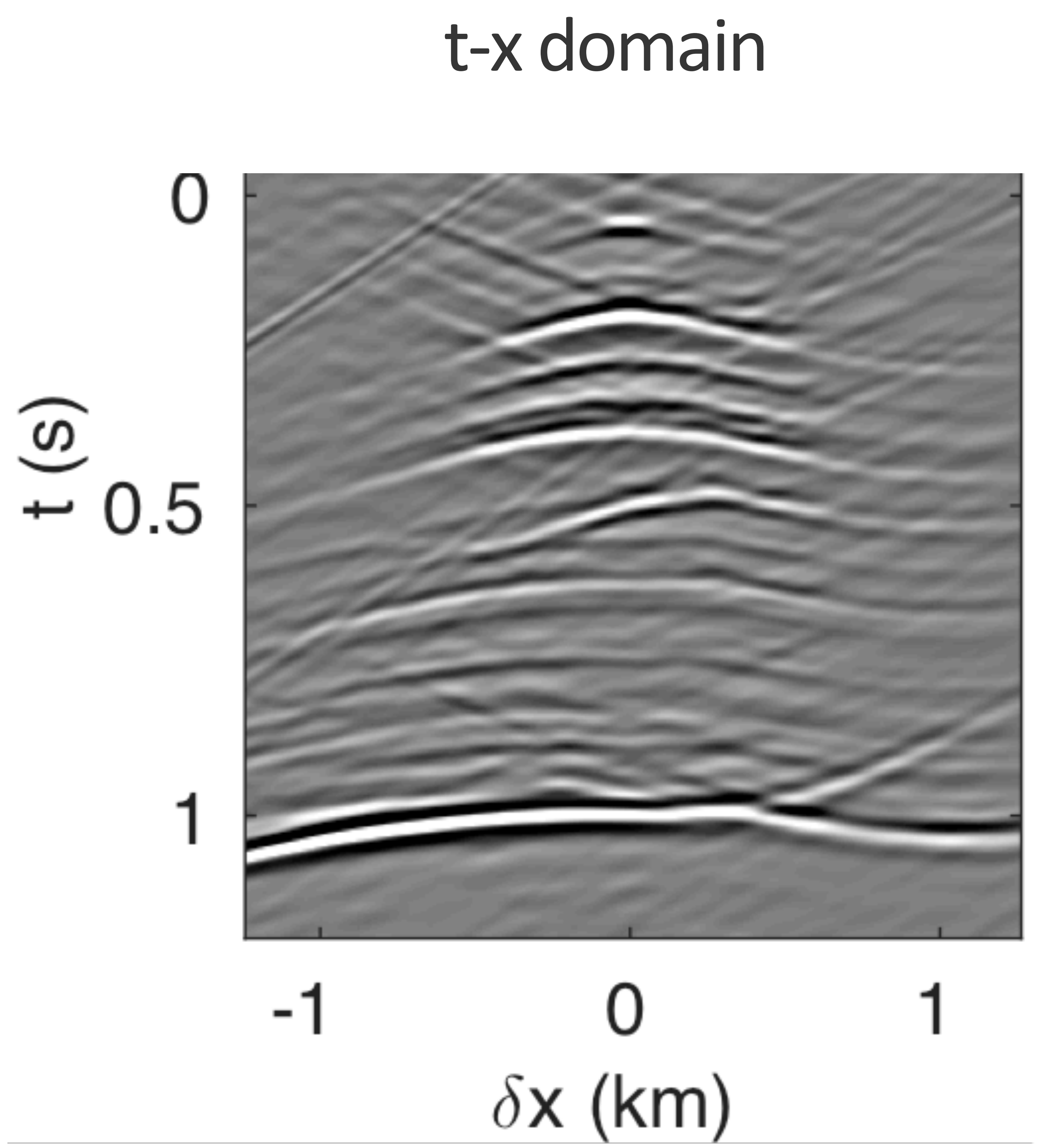
True model

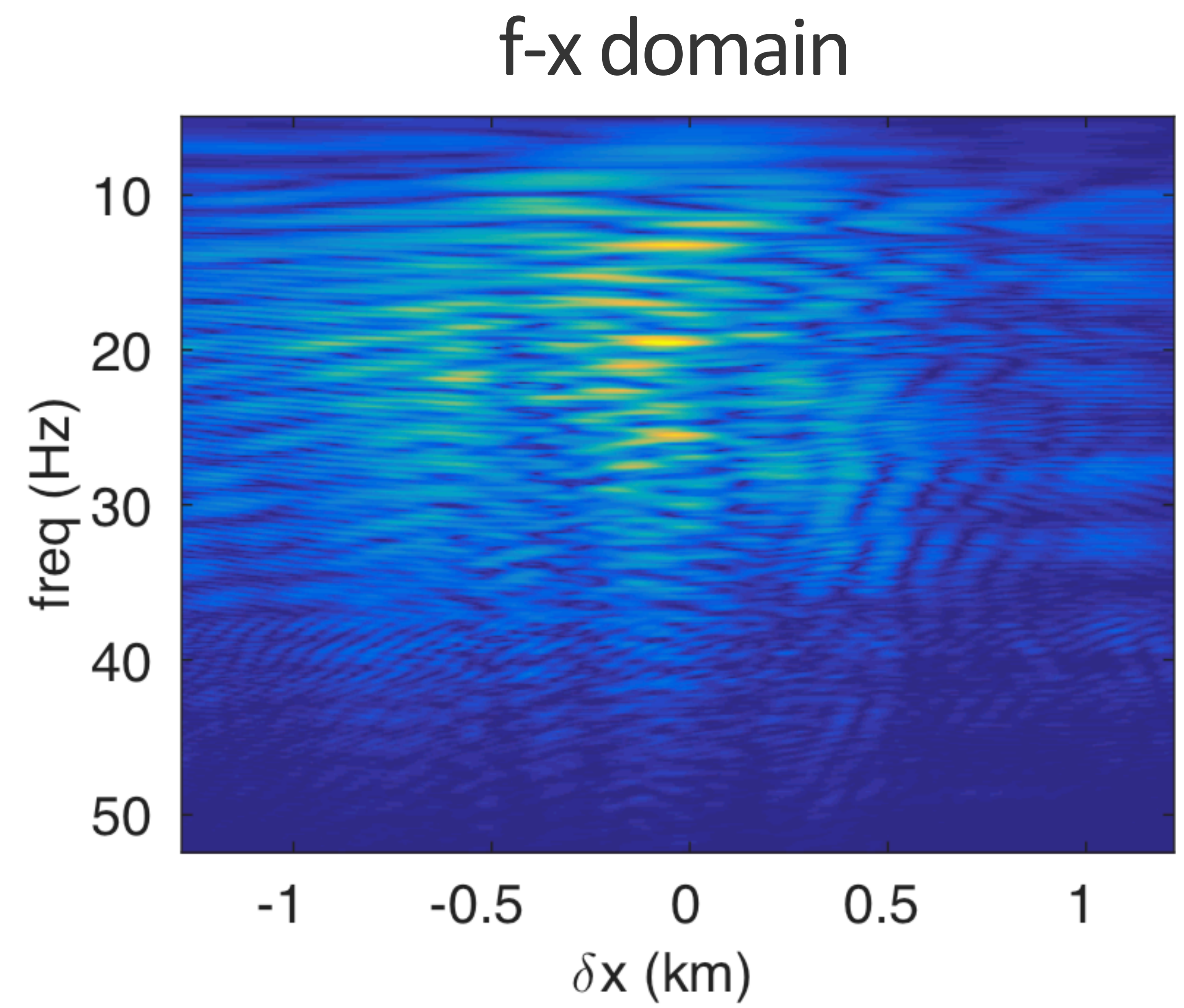


Smooth model

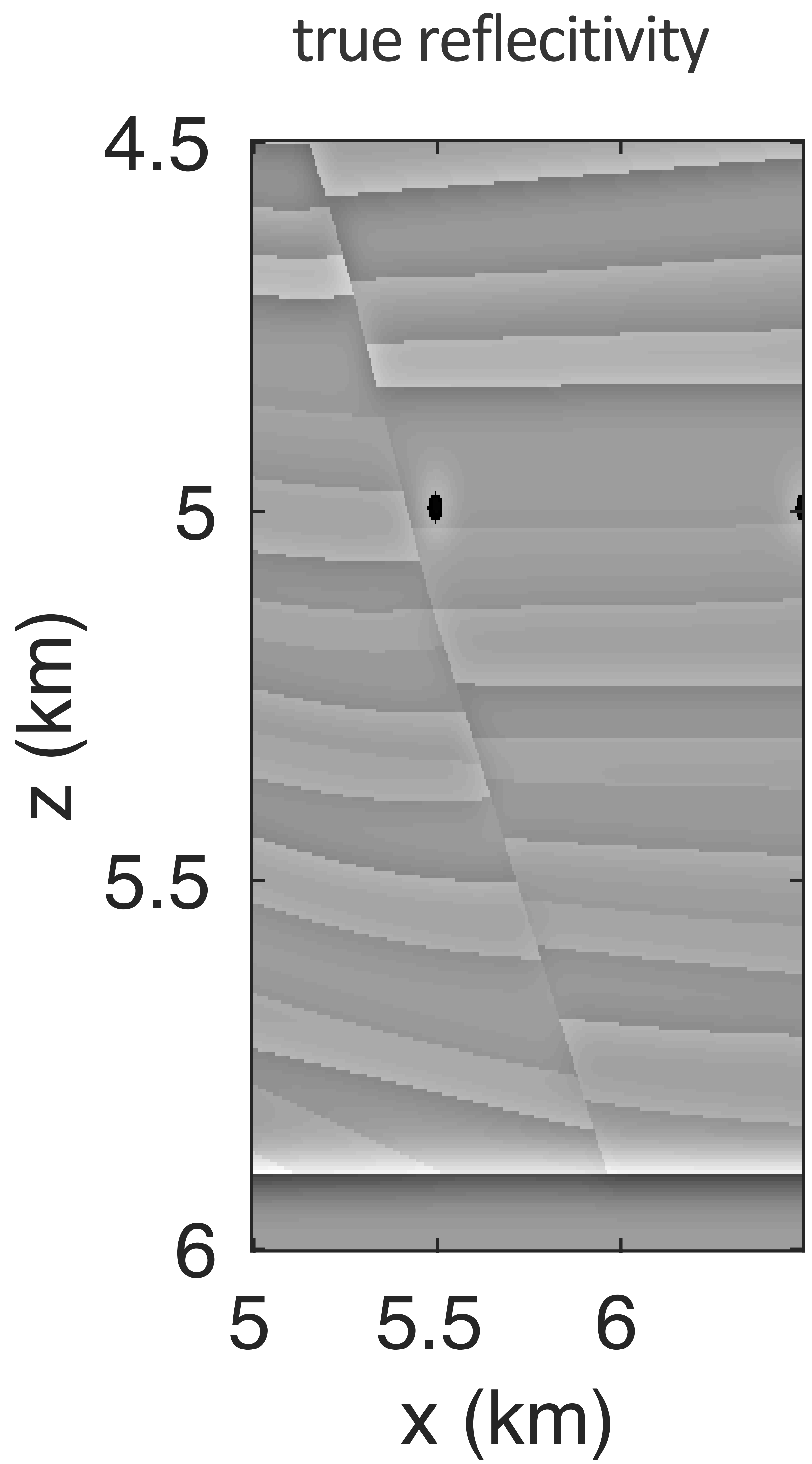
Re-datum data

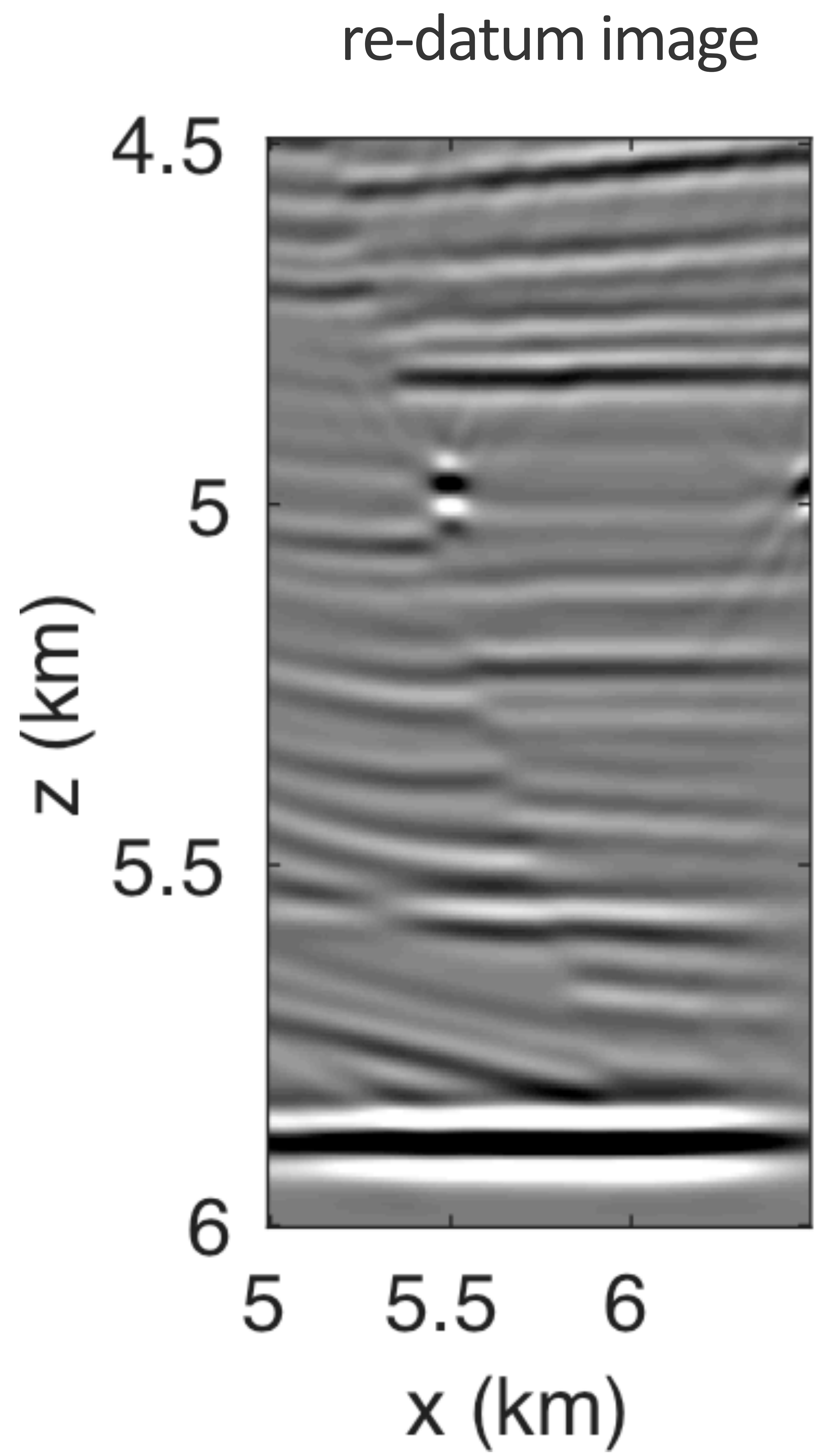






Re-datum image



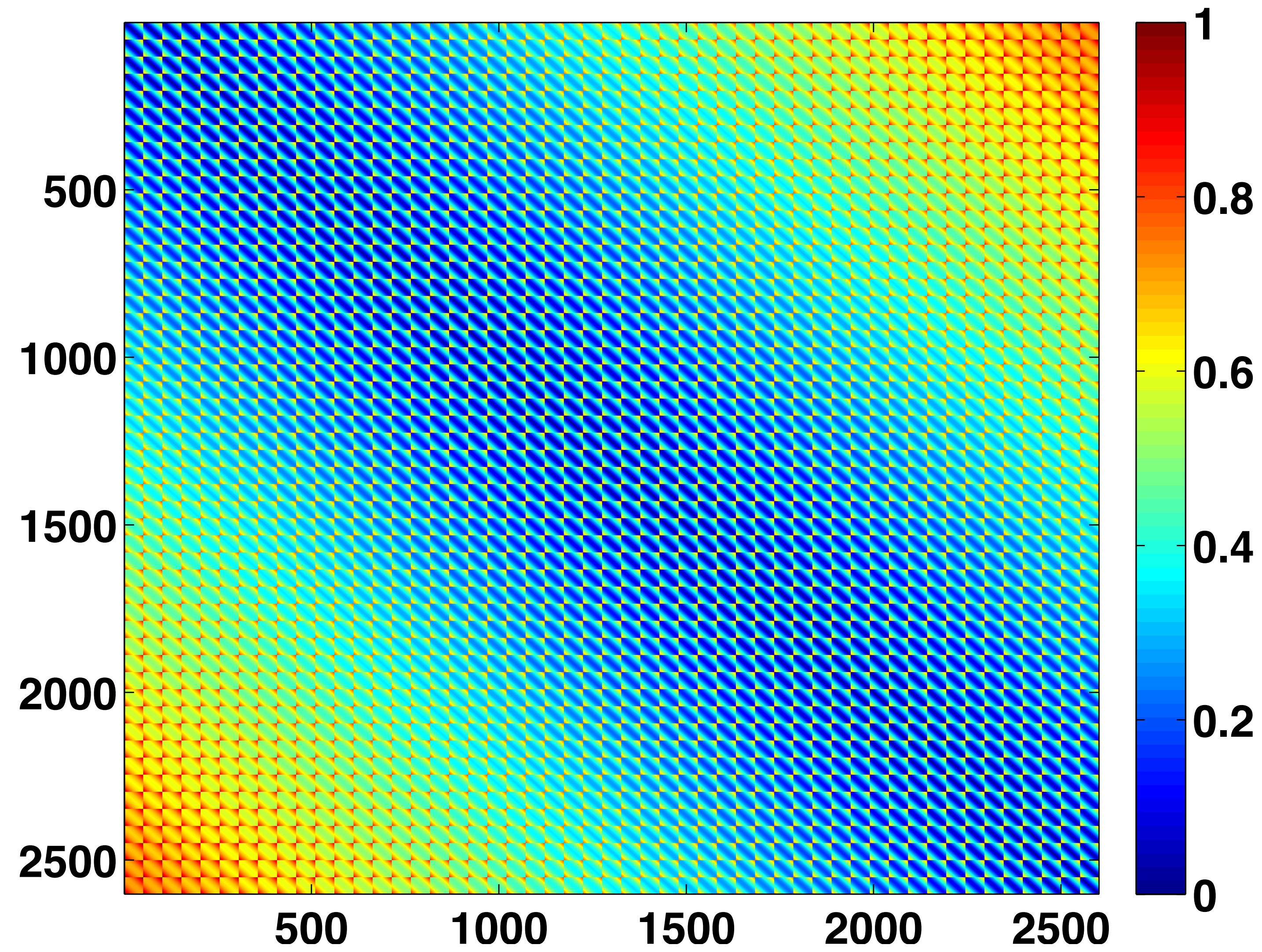


WEMVA

Biondo & Symes, '04 , Symes 2008, Sava & Vasconcelos, '11

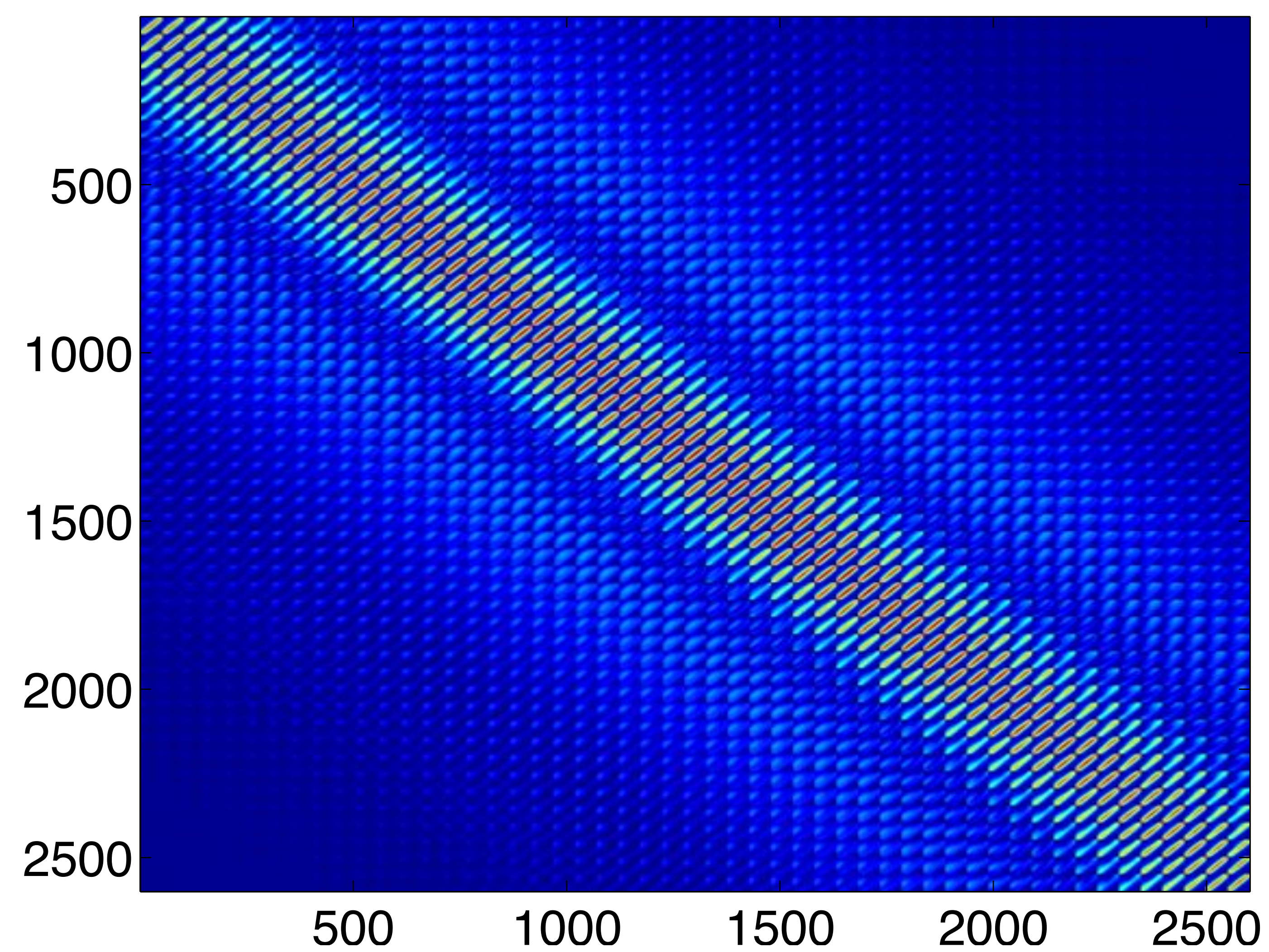
WEMVA

conventional approach



h

$.*$



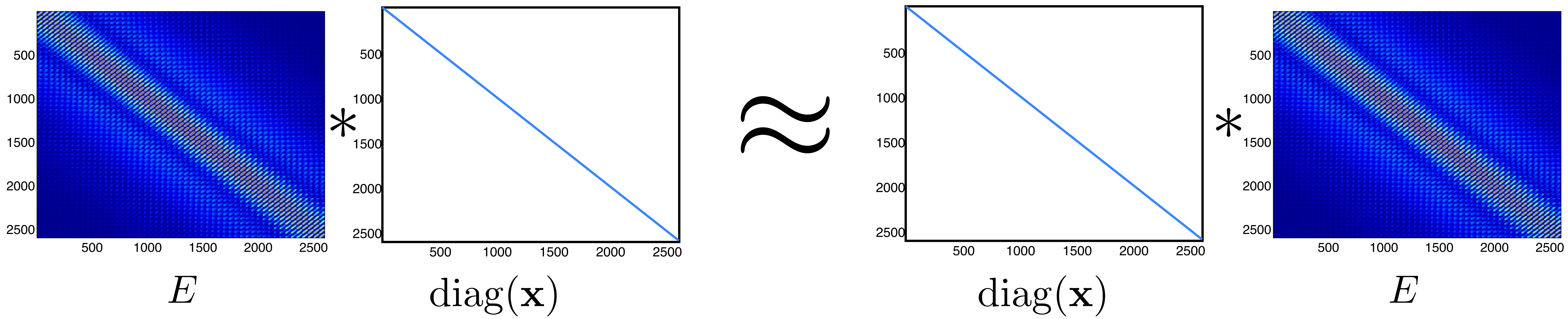
E

$.*$ **stand for element-wise multiplication**

Focusing

propose method approach

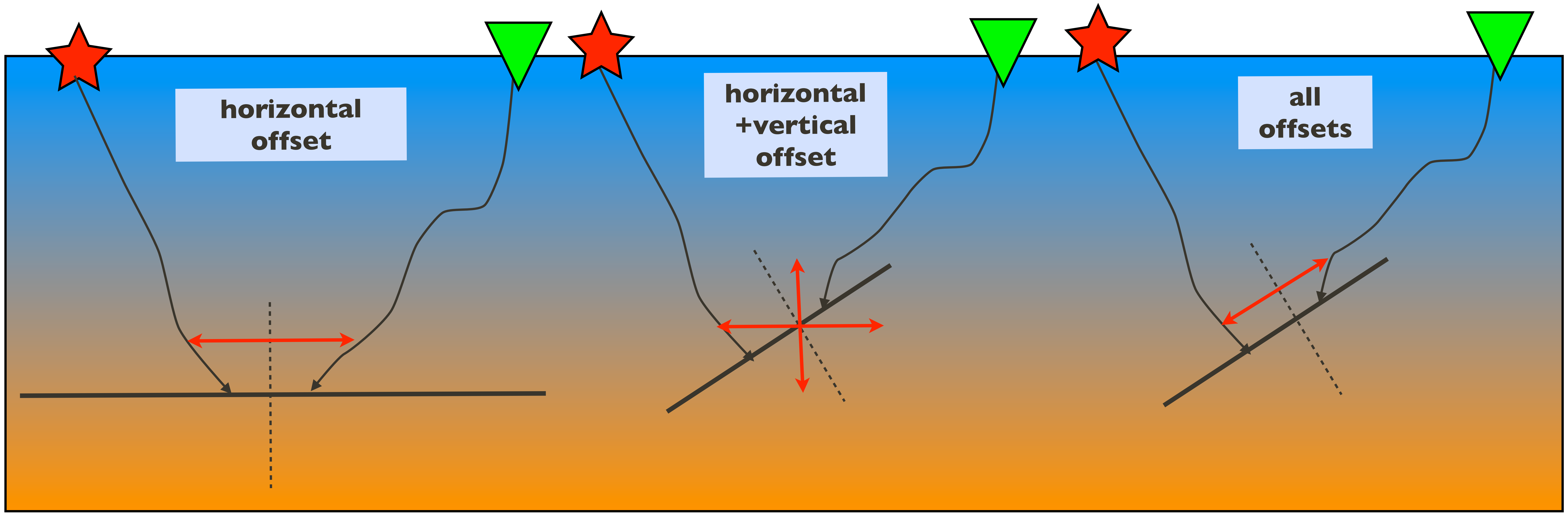
$$E\,\mathrm{diag}(\mathbf{x}) \approx \mathrm{diag}(\mathbf{x})\,E$$



*** matrix-matrix multiplication**

Focusing

where $\mathbf{x}$ represents horizontal, vertical or all offset.



SLIM

Fast WEMVA w/ *randomized* probing

Measure the error in some norm

$$\min_{\mathbf{m}} ||E(\mathbf{m})\mathrm{diag}(\mathbf{x}) - \mathrm{diag}(\mathbf{x})E(\mathbf{m})||_?^2$$

The *Frobenius* norm can be estimated via randomized trace estimation : **Avron and Toledo, 2011**

$$||A||_F^2 = \mathrm{trace}(A^T A)$$

$$\approx \sum_{i=1}^{K} \mathbf{w}_i^T A^T A \mathbf{w}_i = \sum_{i=1}^{K} ||A\mathbf{w}_i||_2^2$$

where $\sum_{i=1}^{K} \mathbf{w}_i \mathbf{w}_i^T \approx I$

Slicing & dicing

Can not store full E for large-scale 2D and 3D

use factorized form

No need to re-estimate E during gradient computations

Gradients

$$DE(m)[\delta m]y = -H(m)^{-1}\frac{\partial H(m)}{\partial m}[\delta m]E(m)y - E(m)\frac{\partial H(m)}{\partial m}[\delta m]H(m)^{-1}y$$

$$(DE(m)[\cdot]y)^T Z = -\mathrm{diag}(\overline{E(m)y})\frac{\partial H(m)}{\partial m}^H H(m)^{-H} Z - \mathrm{diag}(\overline{H(m)^{-1}y})\frac{\partial H(m)}{\partial m}^H E(m)^H Z$$

$$\nabla f(m) = (DE(m)[\cdot]\mathrm{diag}(s)w - \mathrm{diag}(s)DE(m)[\cdot]w)^T(E(m)\mathrm{diag}(s)w - \mathrm{diag}(s)E(m)w)$$

Experimental details

350 source (40 m spacing) , 700 receivers (20 m spacing)

5-25 Hz

split-spread acquisition

recording length 6s, sampling interval 4ms

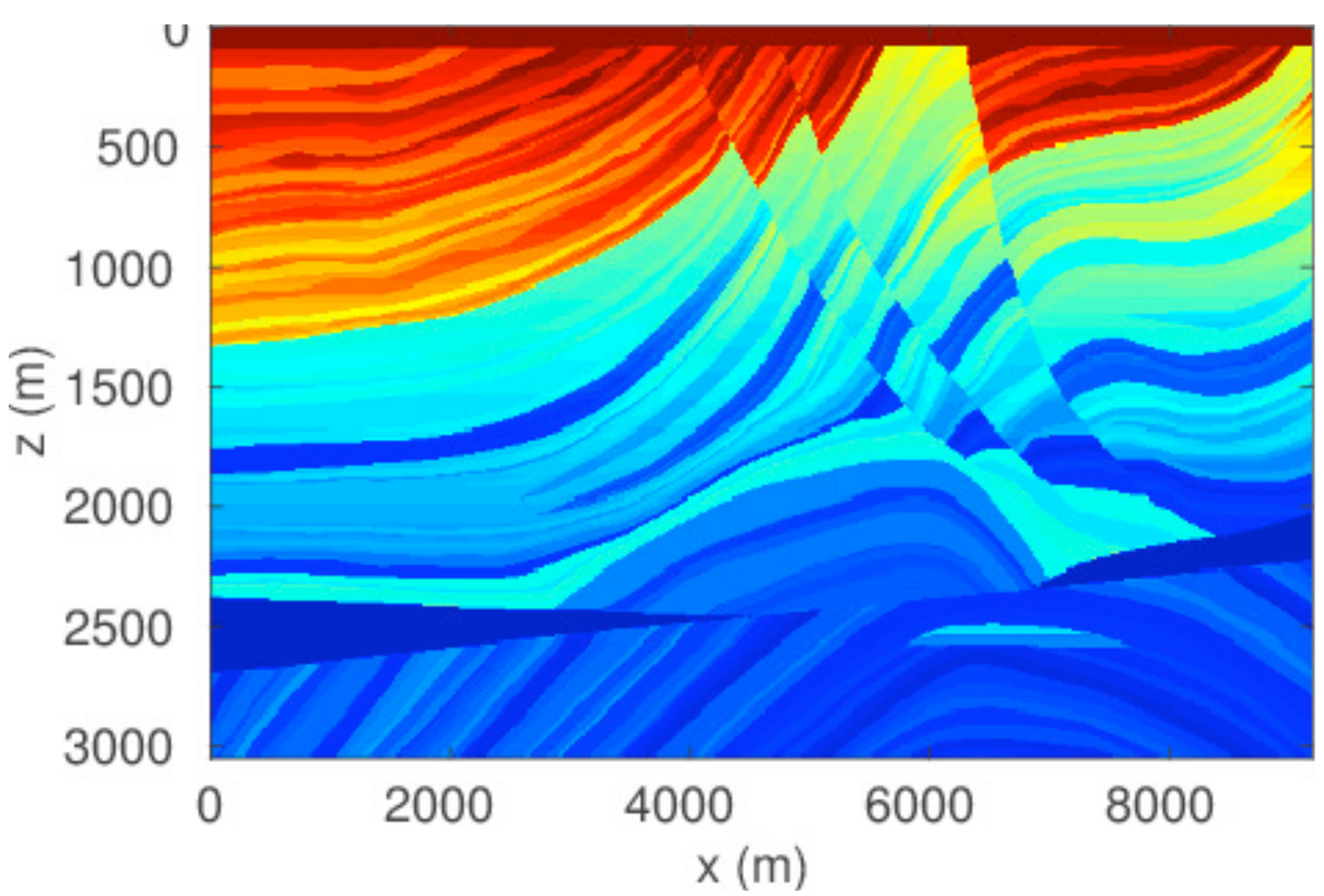
peak frequency 20 Hz

25 LBFGS iterations

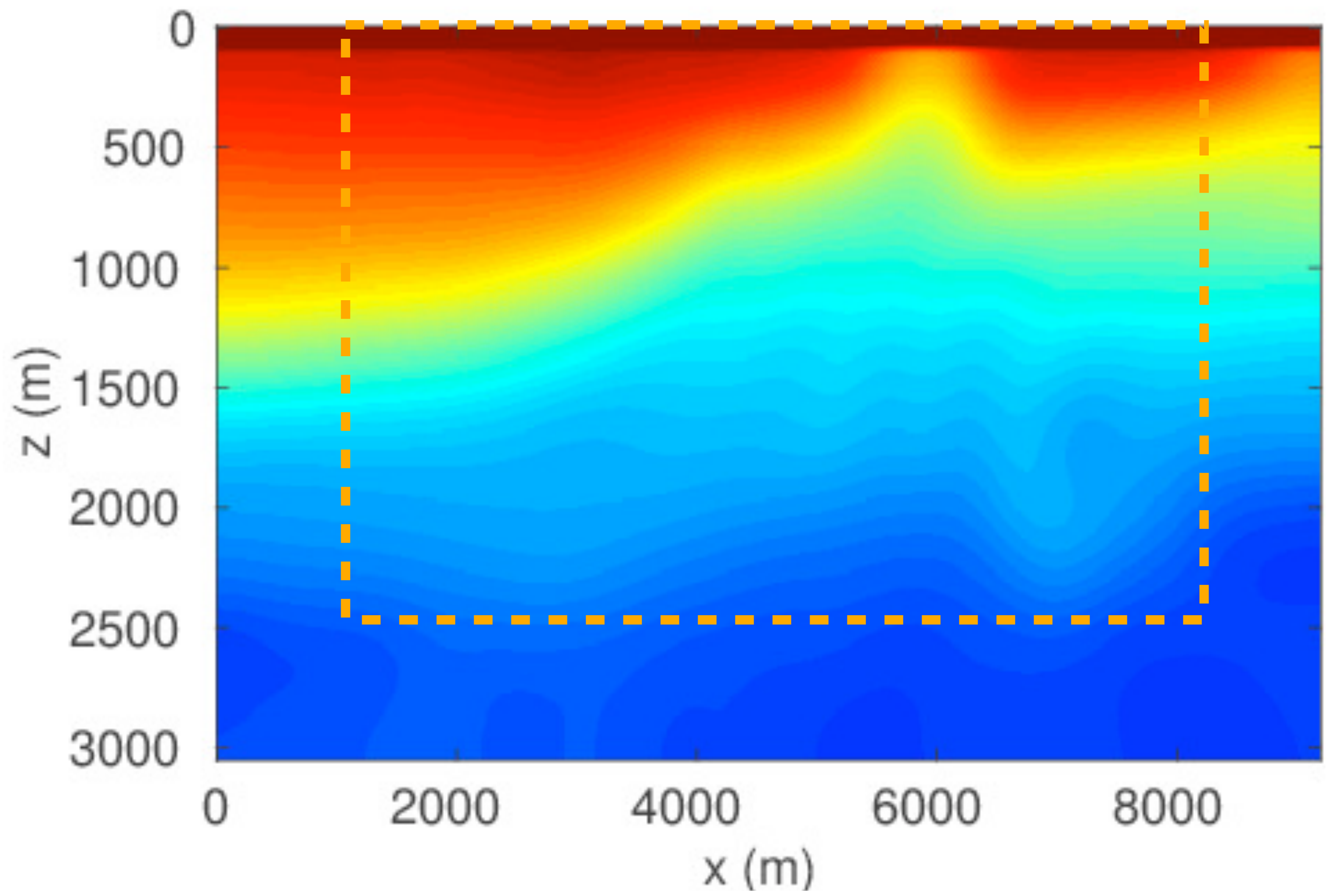
100 probing vectors

Marmousi model

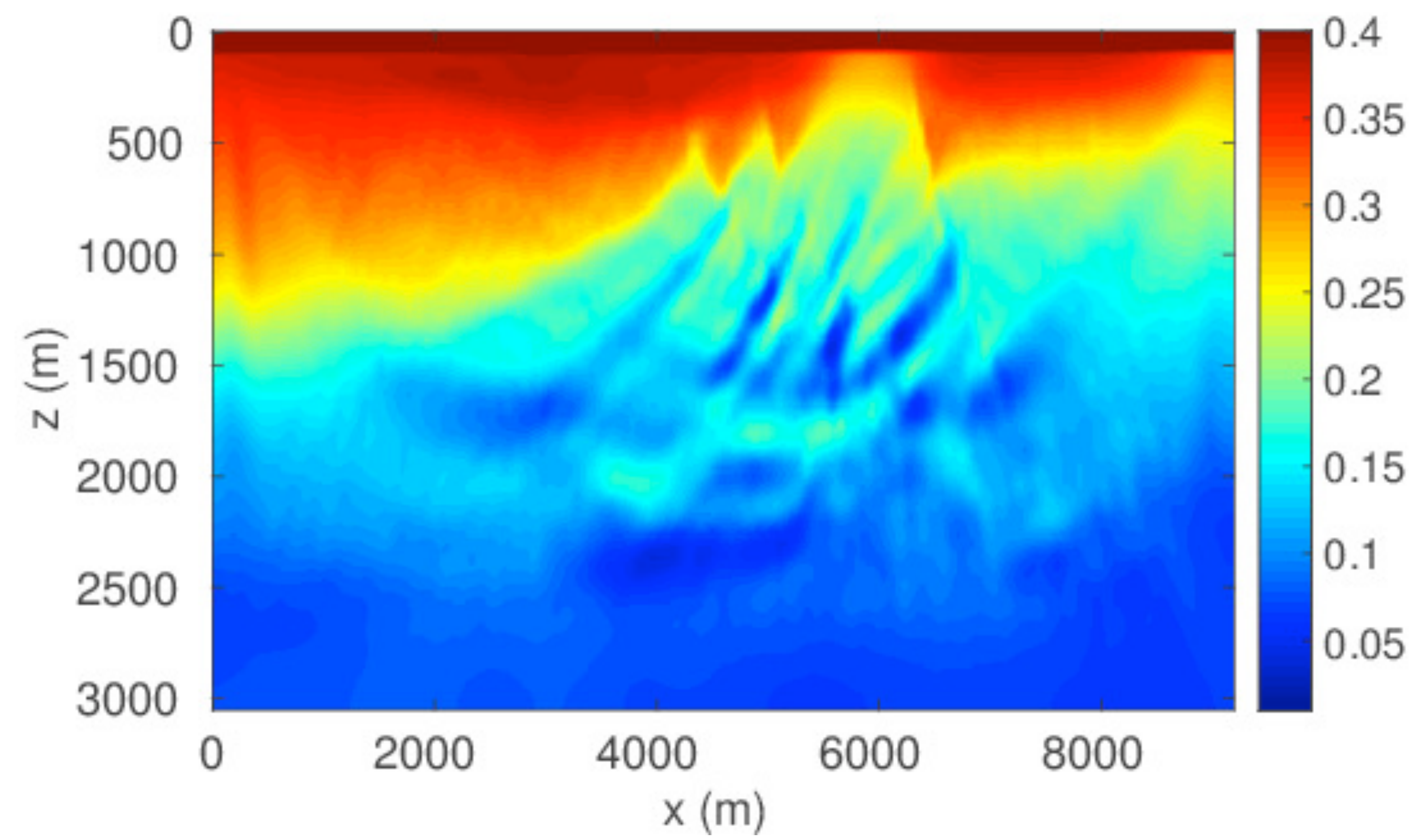
True model



Starting model



Inverted model



Observation

100X computational and memory savings while forming the full-subsurface image volumes in 3D

Efficient way to extract informations from image volumes

Very fast (2D/3D) target-imaging tool

60X reduction in memory and computational cost in 2D WEMVA, a step closer to 3D WEMVA

Conclusion

- Easy way to handle enormous data volumes
  - very high compression ratios (at low-frequencies)
- Efficient data extraction framework
- Can form full-subsurface extended image volumes
- Easy to combine with existing FWI/WEMVA codes

Future work

- Combined w/ frequency-extrapolation
- 3D extension (WEMVA, Target-Imaging etc...)
- Least-square extended image volumes
- Links to time-domain framework

Acknowledgements

This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.

Acknowledgements

The authors wish to acknowledge the SENAI CIMATEC Supercomputing Center for Industrial Innovation, with support from BG Brasil, Shell, and the Brazilian Authority for Oil, Gas and Biofuels (ANP), for the provision and operation of computational facilities and the commitment to invest in Research & Development.

Thank you for your attention