Structured tensor missing-trace interpolation in the Hierarchical Tucker format

Curt Da Silva and Felix J. Herrmann

Dec. 3, 2013

"Optimization on the Hierarchical Tucker manifold - applications to tensor completion". Submitted.

SLIM
University of British Columbia

Motivation

3D seismic experiments - 5D data

- expensive to acquire, store
- sample at *sub-Nyquist* rates

Data exhibits *low-rank* structure

- exploit structure for interpolation

Fully sampled data

- simultaneous sources in wave-equation based inversion
- mitigating multiples

[1] Kreimer, N, and Sacchi, M. D. "A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation." (2012)
[2] Gao, Jianjun, Vicente Oropeza, and Mauricio D. Sacchi. "Evaluation of a fast algorithm for the eigen-decomposition of large block Toeplitz matrices with application to 5D seismic data interpolation." *(2011)*

SLIM

Context

Low-rank matrix/tensor completion via *nuclear norm* projection [1]

- Require SVDs on huge data matrices
- Not scalable to large problem sizes

Data completion via Toeplitz embedding [2]

- Problem size - (# data points)^2

Goals

Generalization of Compressive Sensing to multiple dimensions

- what can we learn from 1D/2D recovery?

Randomized source/receiver acquisition

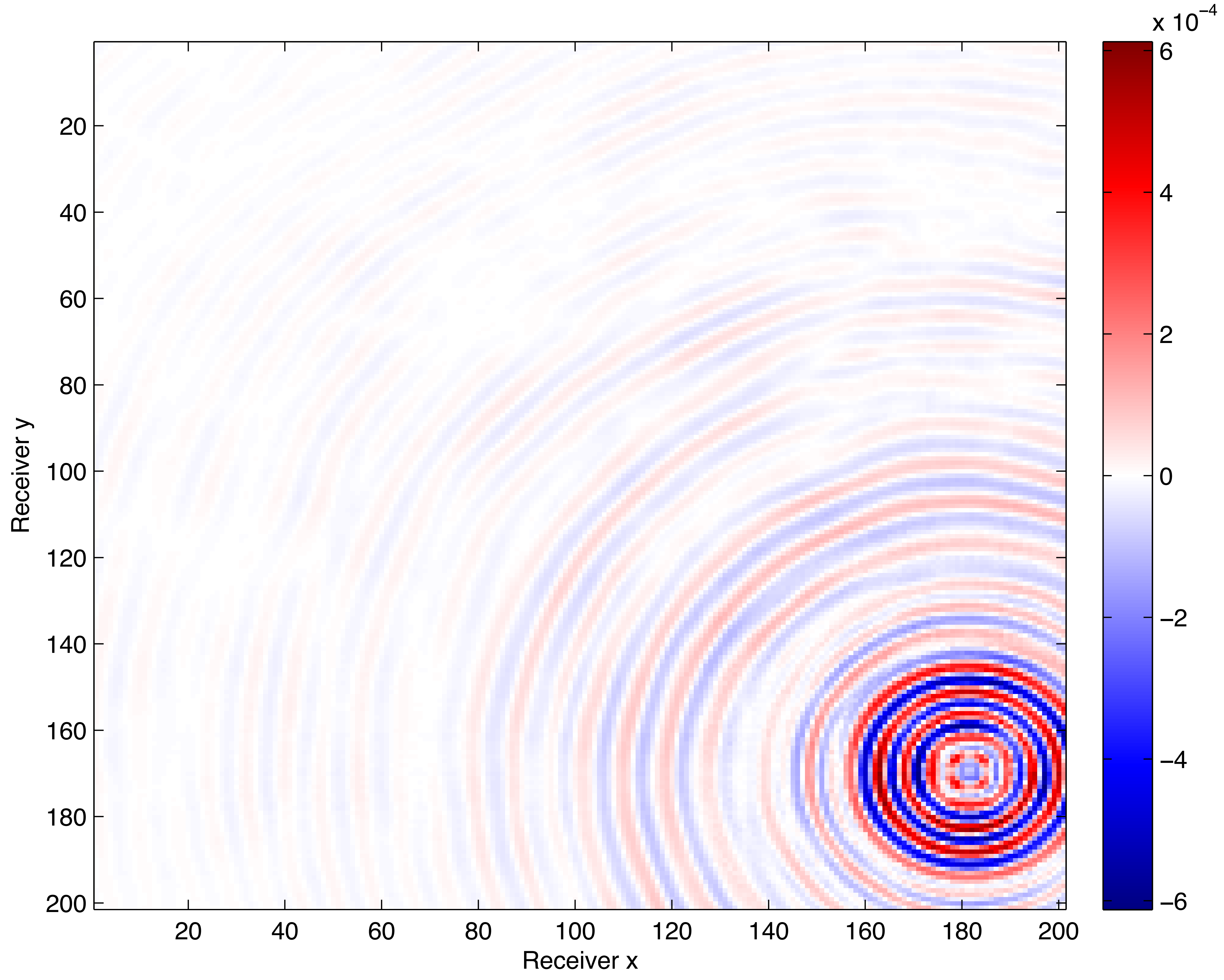
- reduce acquisition financial/time costs

Efficient solver

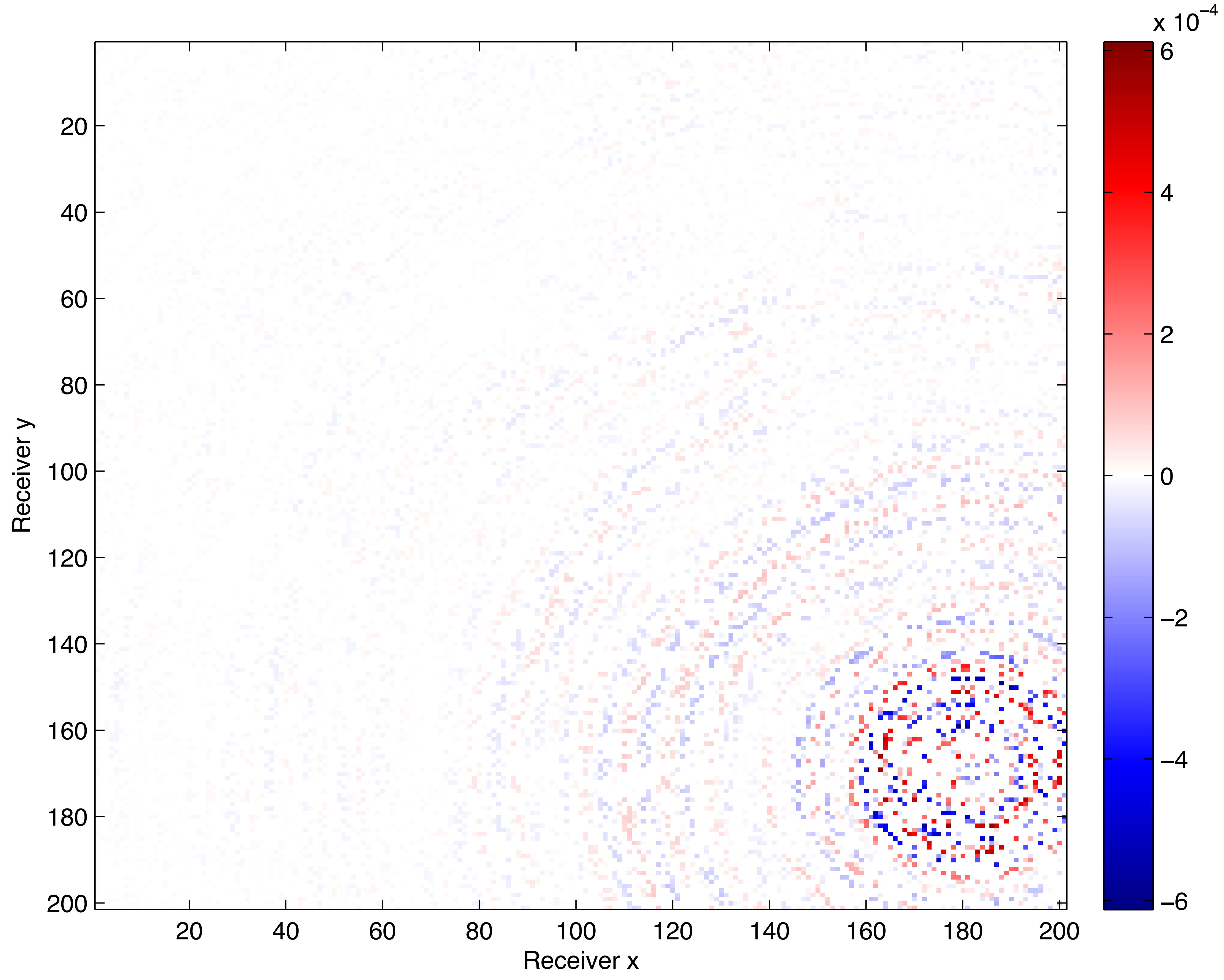
- SVD-free, parallelizable
- # parameters << # data points

7.34 Hz - 75% missing receivers

Common source gather



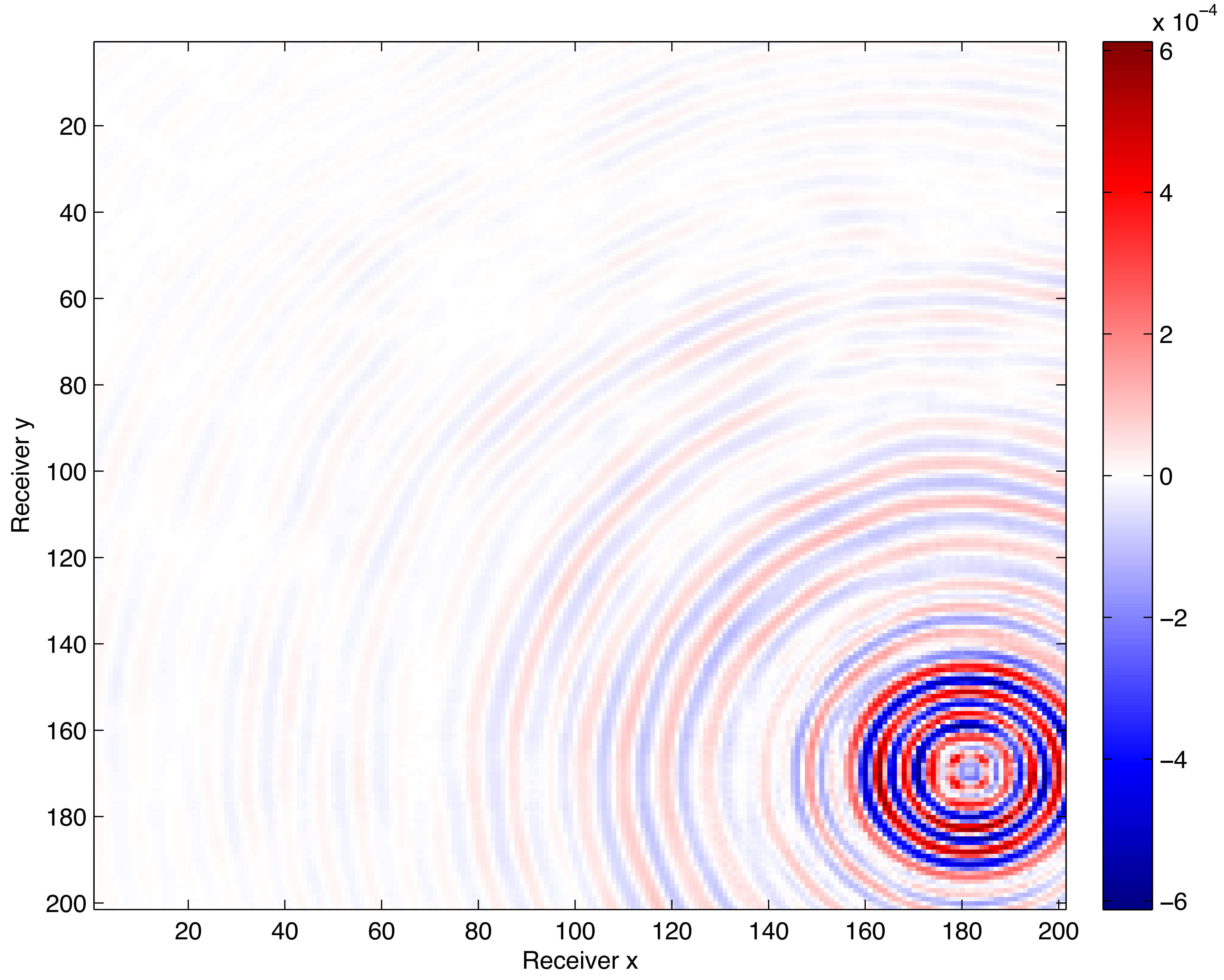
True data



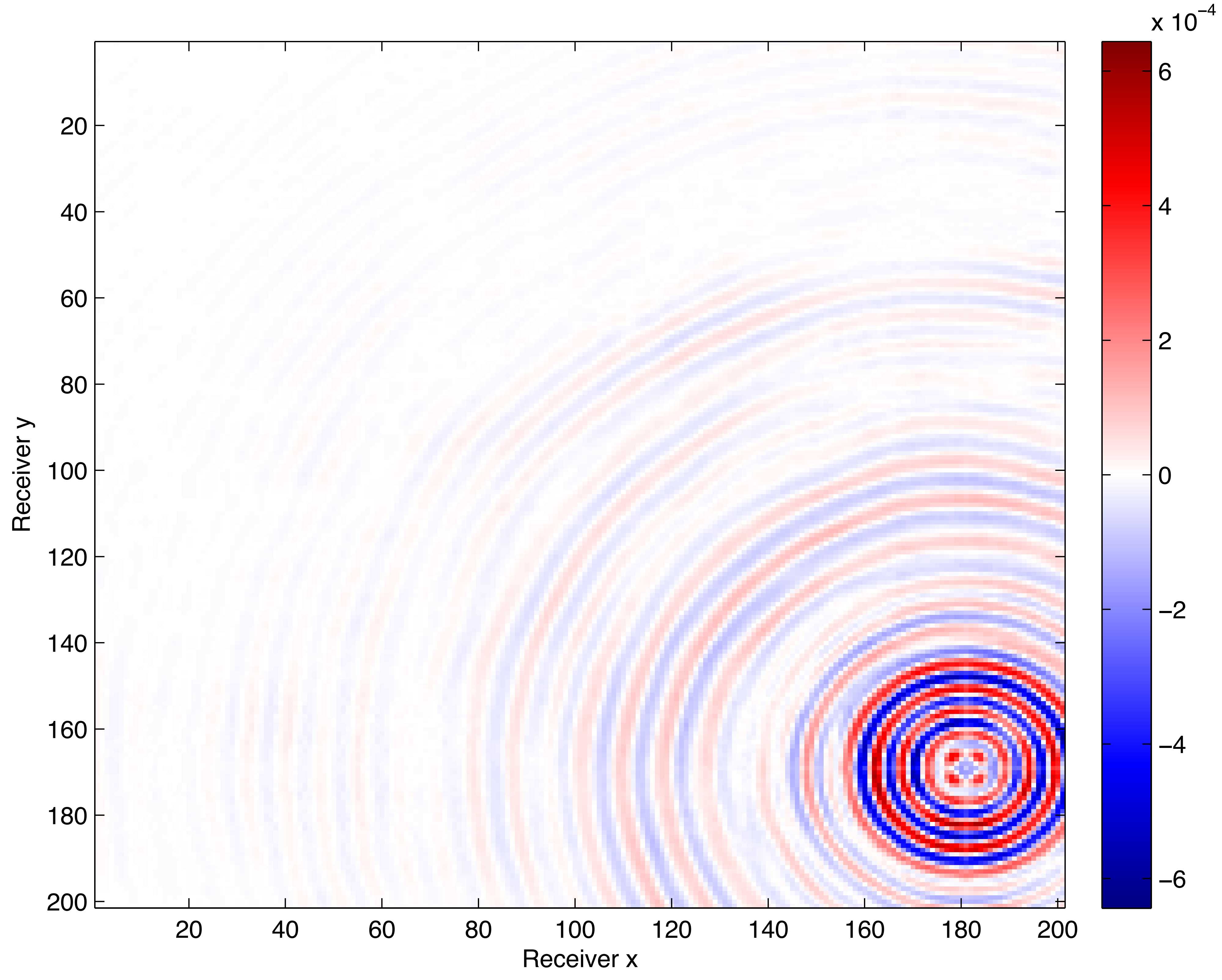
Subsampled data

7.34 Hz - 75% missing receivers

Common source gather



True data



Recovered data - SNR 17.4 dB

Compressive sensing
with sparsity promotion

Successful reconstruction scheme

Signal structure

- sparsity

Sampling

- subsampling decreases sparsity

Optimization

- look for sparsest solution

Multidimensional interpolation

with Hierarchical Tucker

Successful reconstruction scheme

Signal structure

- ***Hierarchical Tucker***

Sampling

- subsampling increases hierarchical rank

Optimization

- fit data in the Hierarchical Tucker format

Matricization

The matricization of a tensor X with dimensions $1, \ldots, d$ along the dimensions $t = (t_1, \ldots, t_r)$ is the matrix formed by placing the dimensions t along the rows and dimensions t^c along the columns

Denoted $X^{(t)}$

SLIM

Example in Matlab

```
n1 = 20; n2 = 20; n3 = 20; n4 = 20;
% Tensor
x = randn(n1,n2,n3,n4);

% Matricization along dimensions 1 and 2
x12 = reshape(x,n1 * n2, n3 * n4);

% Matricization along dimensions 3 and 4
y34 = permute(x,[3 4 1 2]);
x34 = reshape(x, n3 * n4, n1 * n2);

% Matricization along dimensions 1 and 3
y13 = permute(x,[1 3 2 4]);
x13 = reshape(x,n1 * n3, n2 * n4);
```

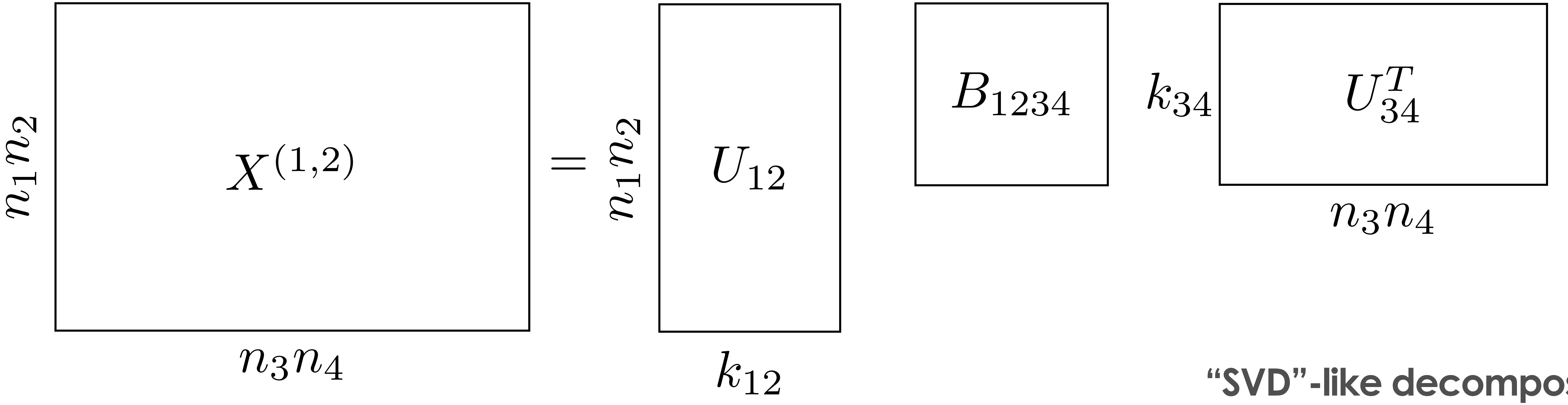
$X^{(1,2)}$: x12

$X^{(3,4)}$: y34, x34

$X^{(1,3)}$: y13, x13

Hierarchical Tucker format

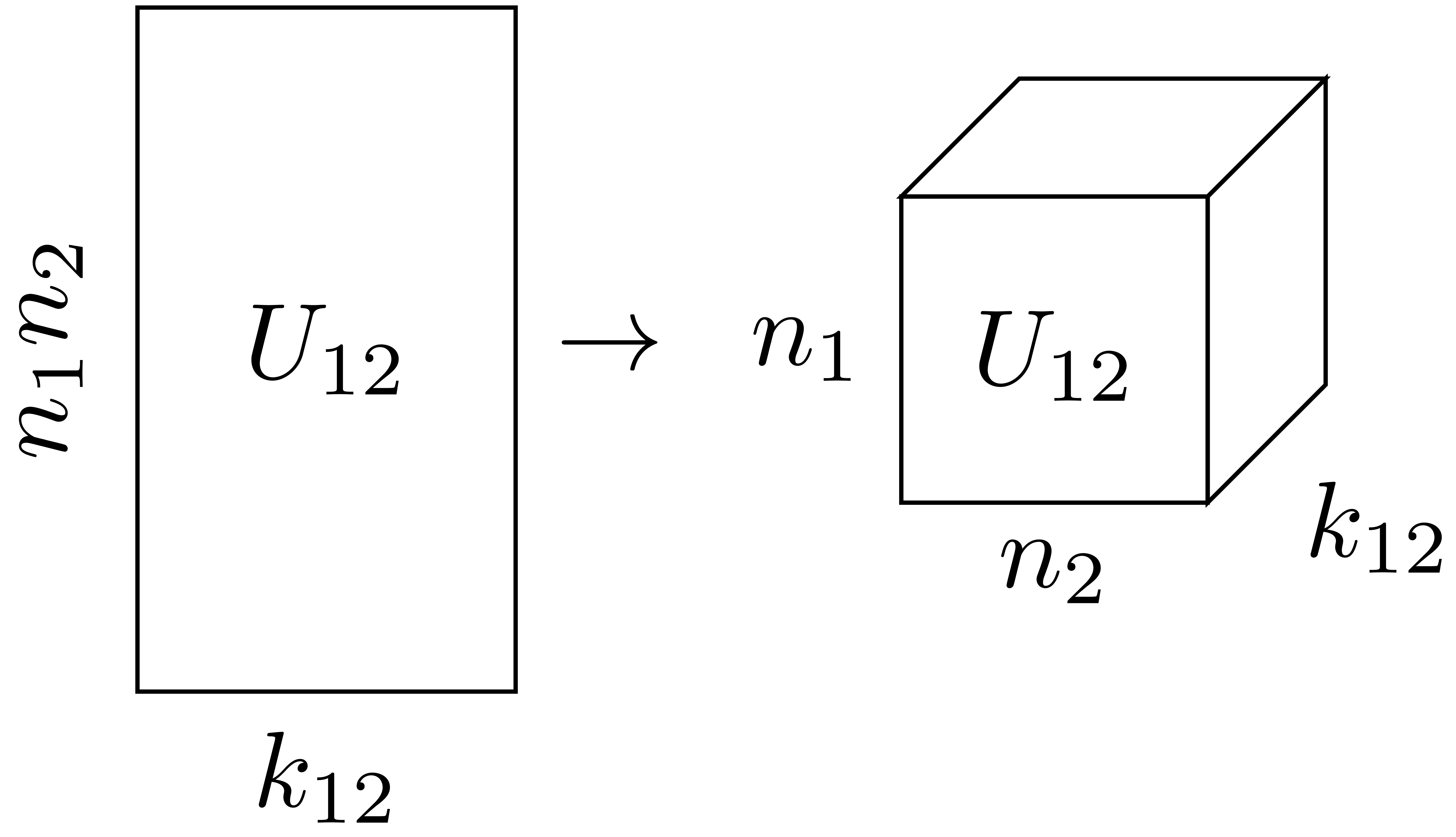
$X - n_1 \times n_2 \times n_3 \times n_4$ tensor



"SVD"-like decomposition

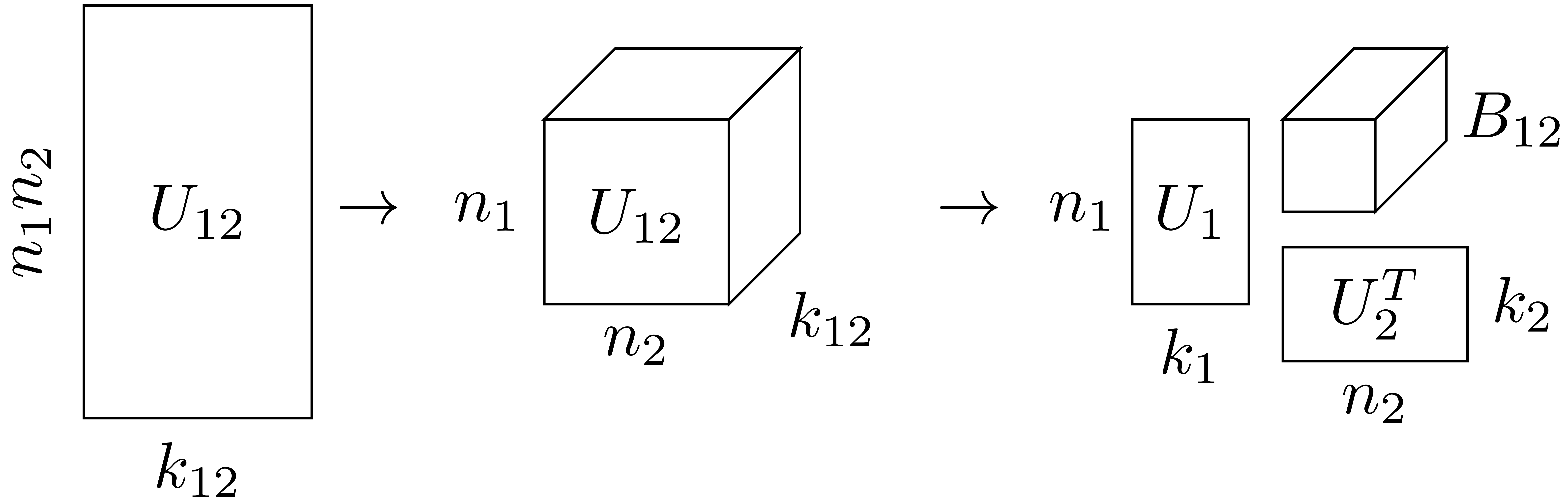
Hierarchical Tucker format

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



Hierarchical Tucker format

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



Hierarchical Tucker format

Intermediate matrices don't need to be stored

U_t, B_t - small parameter matrices

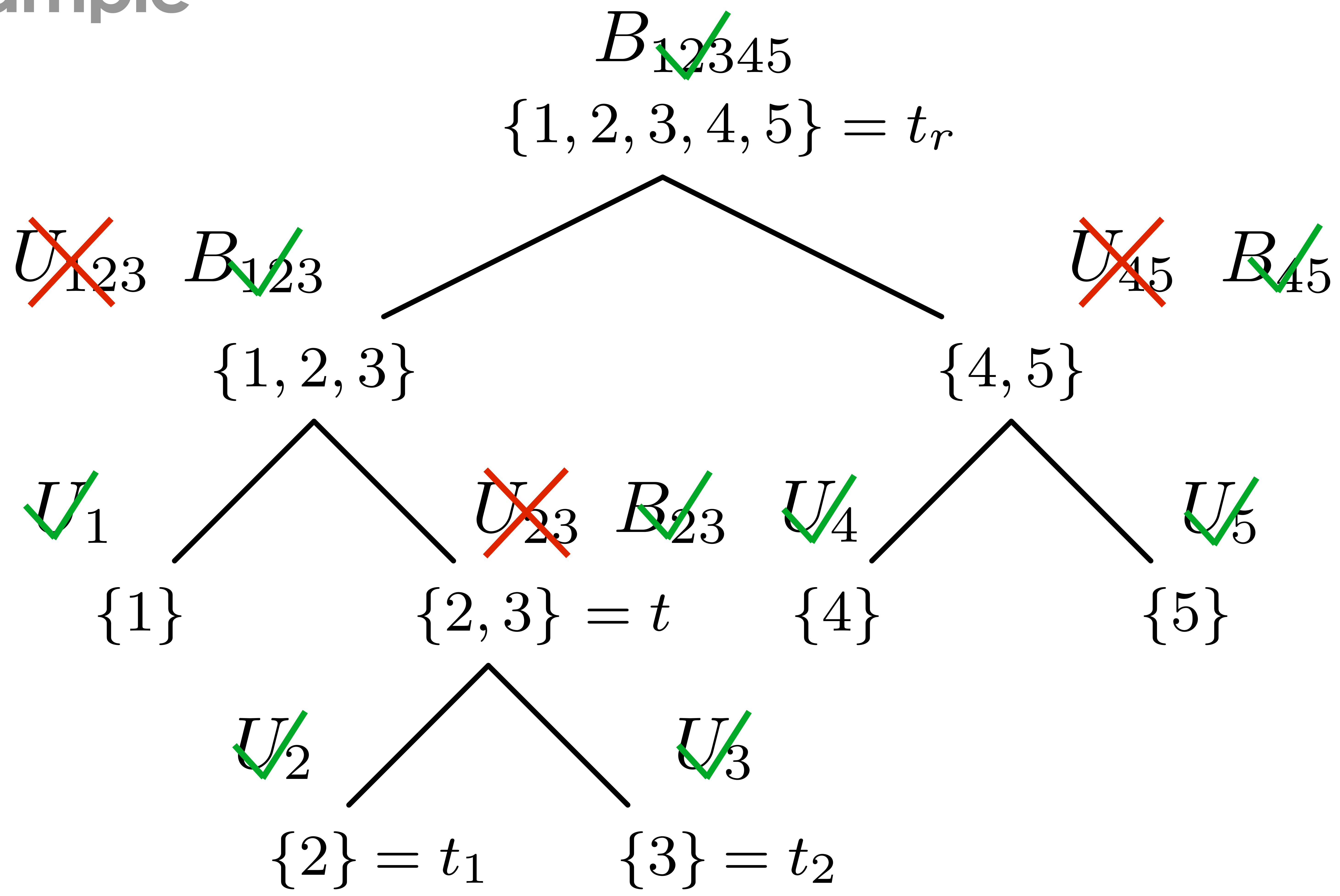
- specify the tensor completely

Separating groups of dimensions from each other

- dimension tree

A. Uschmajew, B. Vandereycken. The geometry of algorithms using hierarchical tensors. Linear algebra and its applications, 2013.

Example



Hierarchical Tucker format

Storage $\leq dNK + (d-2)K^3 + K^2$

Compare to N^d storage for the full tensor

Effectively breaking the curse of dimensionality when $K \ll N \quad d \geq 4$

Low frequency data compresses in HT

Hierarchical Tucker example

For a $100 \times 100 \times 100 \times 100$ cube with max rank 20

$N = 100, d = 4, K = 20$

Full storage: $N^d = 10^8$ values

HTucker storage: 24400 values

Compression of a factor of *99.97%*

Seismic Hierarchical Tucker

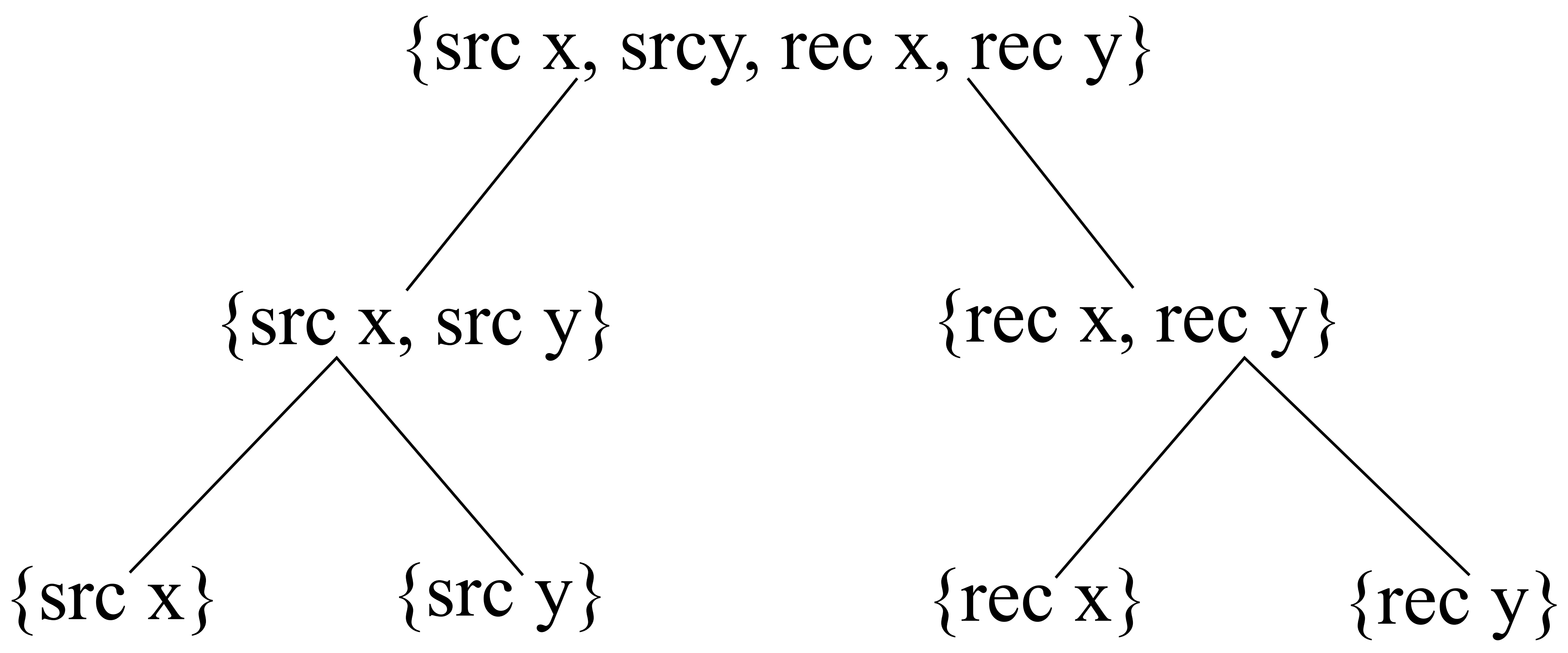
We consider a 3D seismic survey with coordinates
(src x, src y, rec x, rec y, time)

We take a Fourier transform in time and restrict ourselves to a single frequency slice

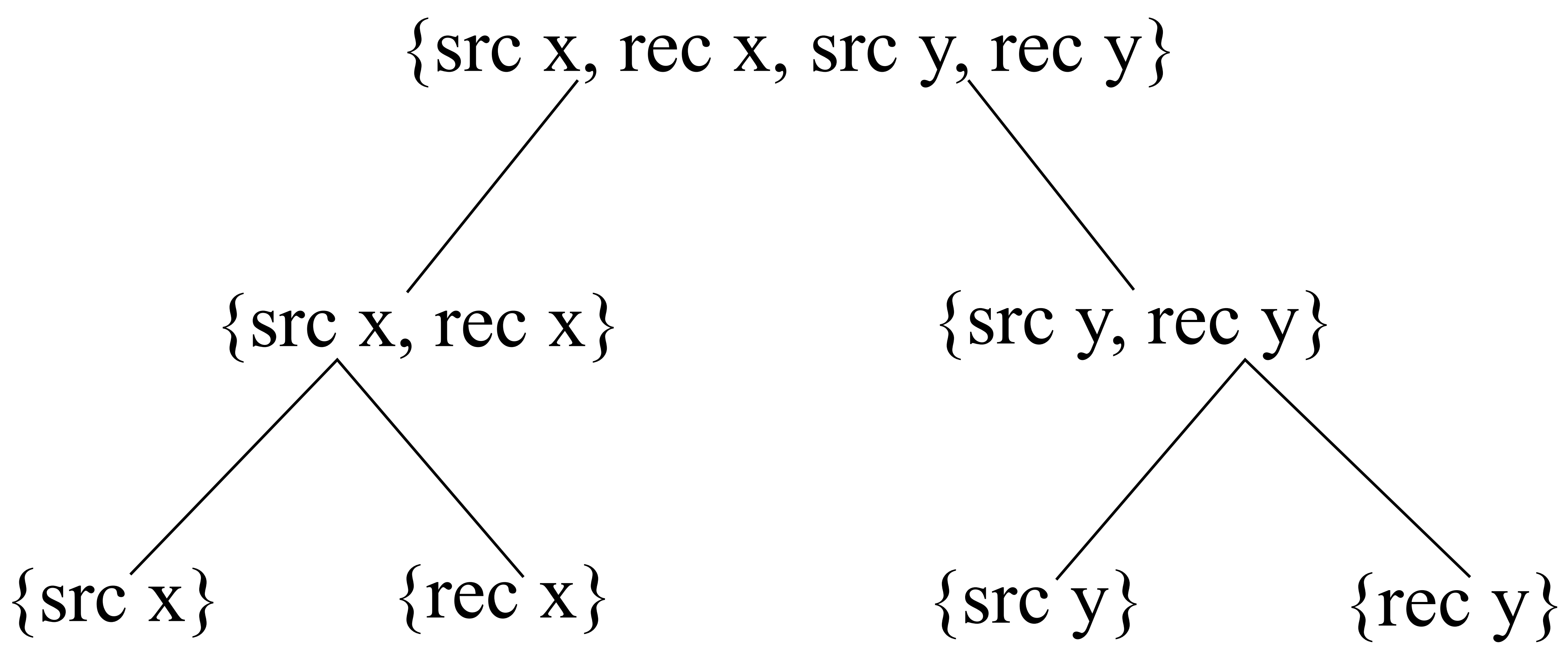
Seismic Hierarchical Tucker

For a frequency slice with coordinates (src x, src y, rec x, rec y),

there are essentially two choices of dimension splitting (by reciprocity)

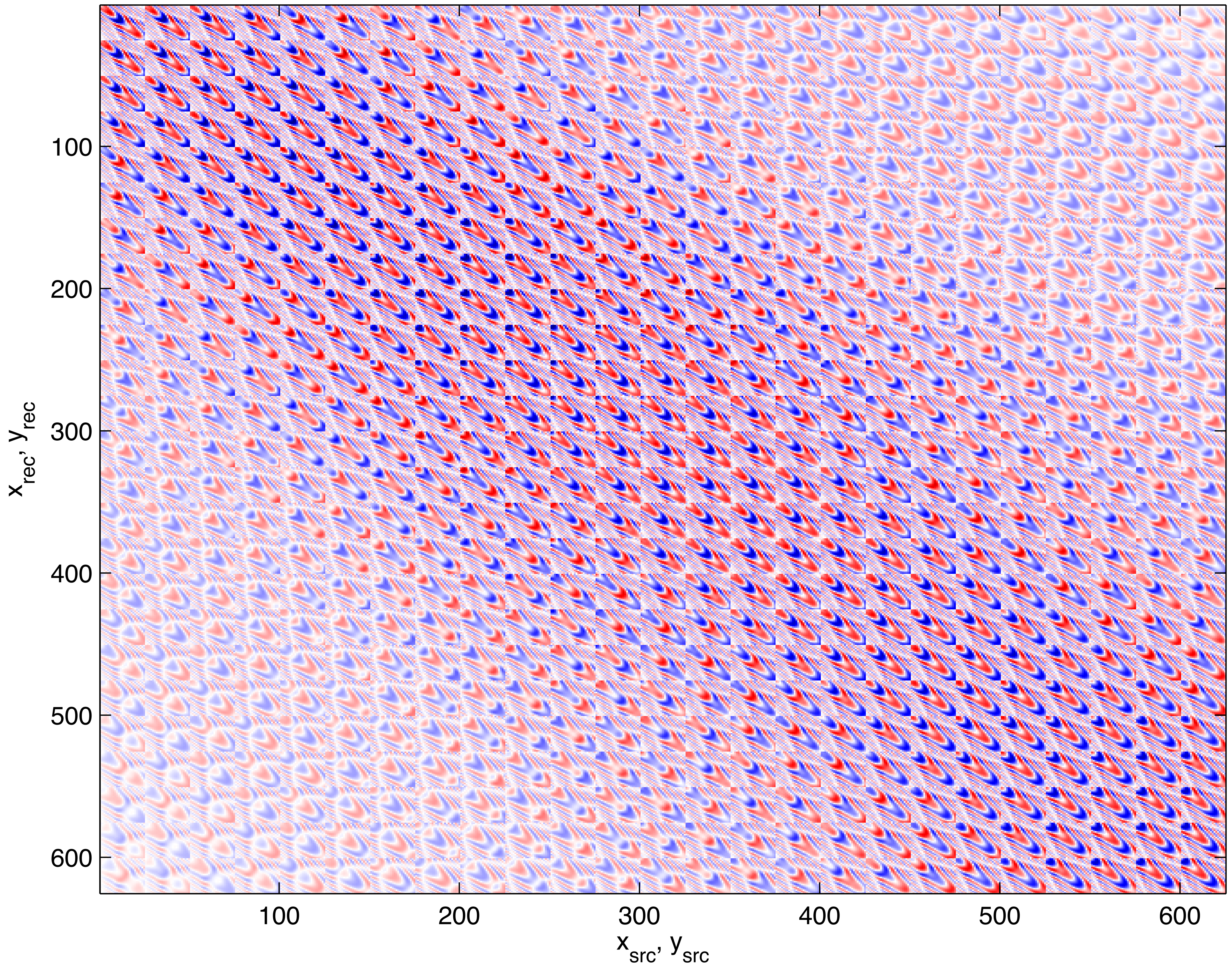


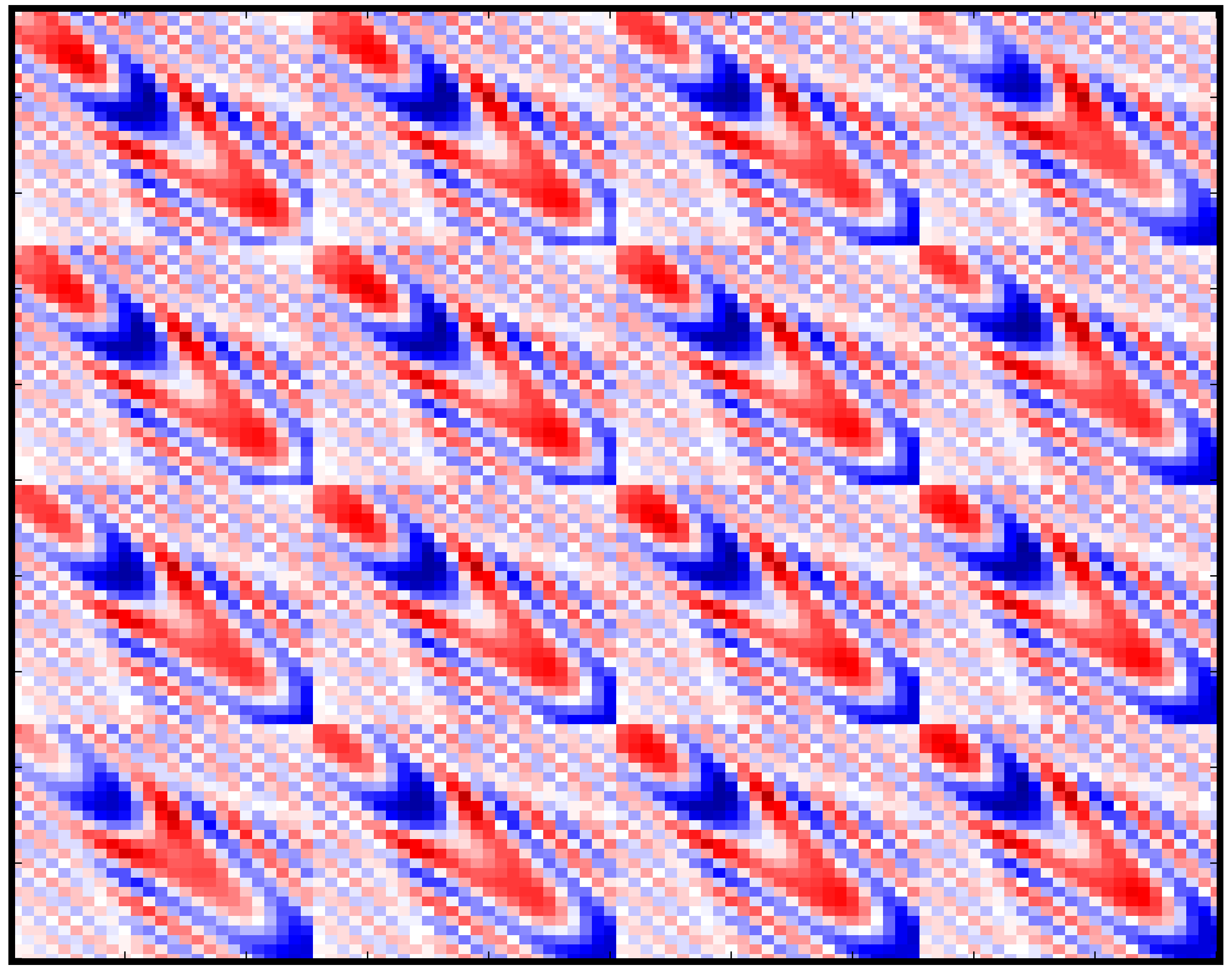
Canonical Decomposition

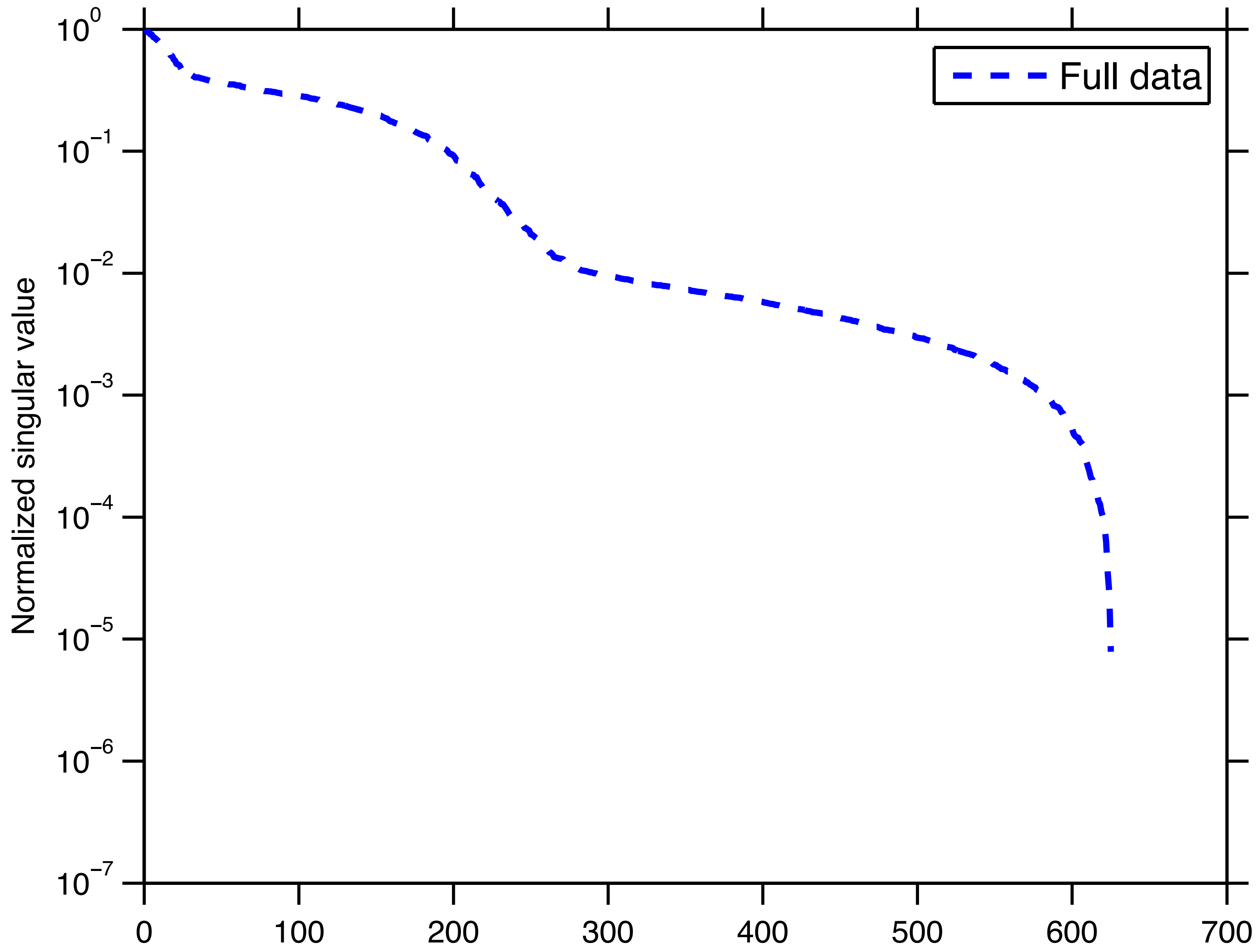


Non-canonical Decomposition

Matricizations

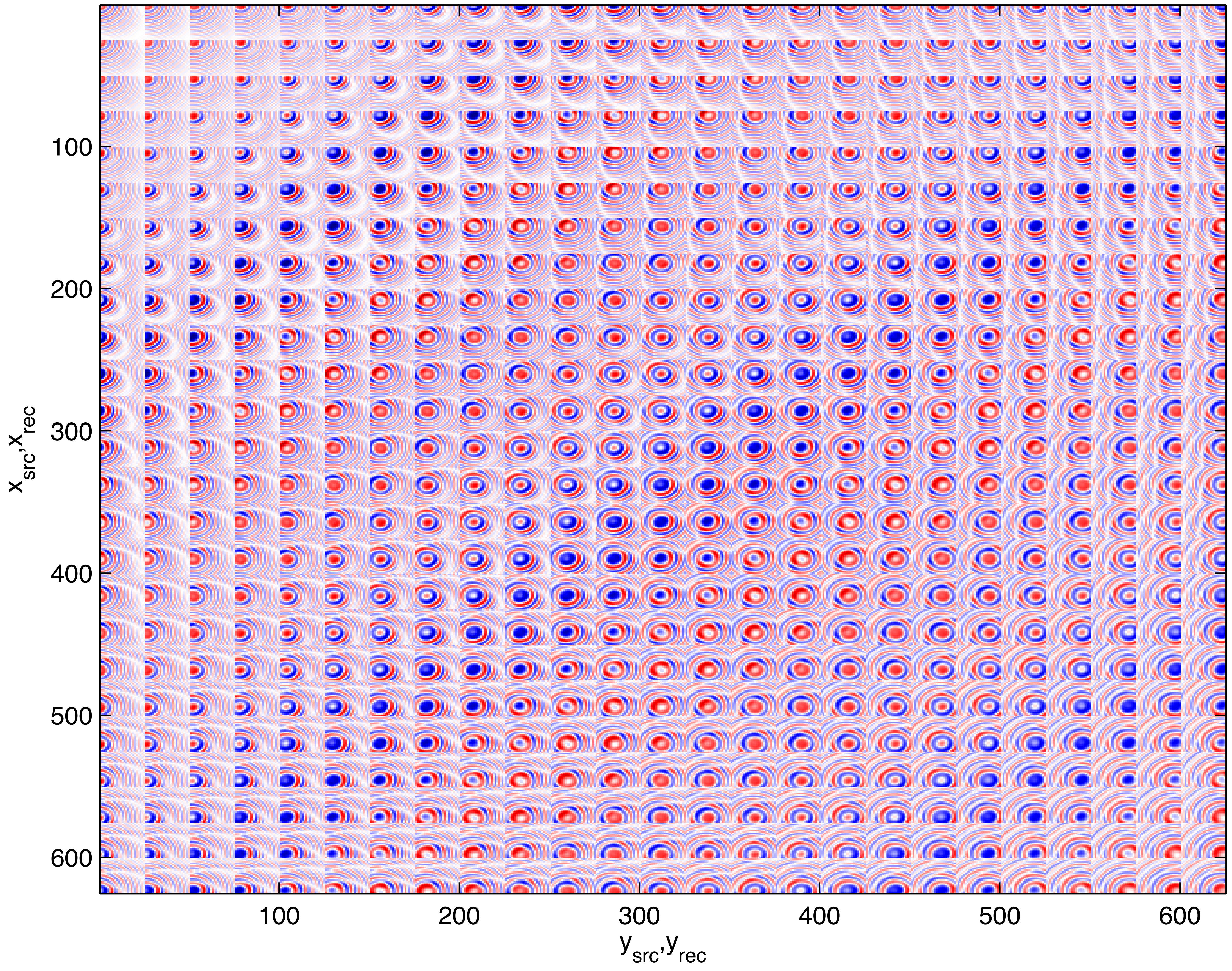


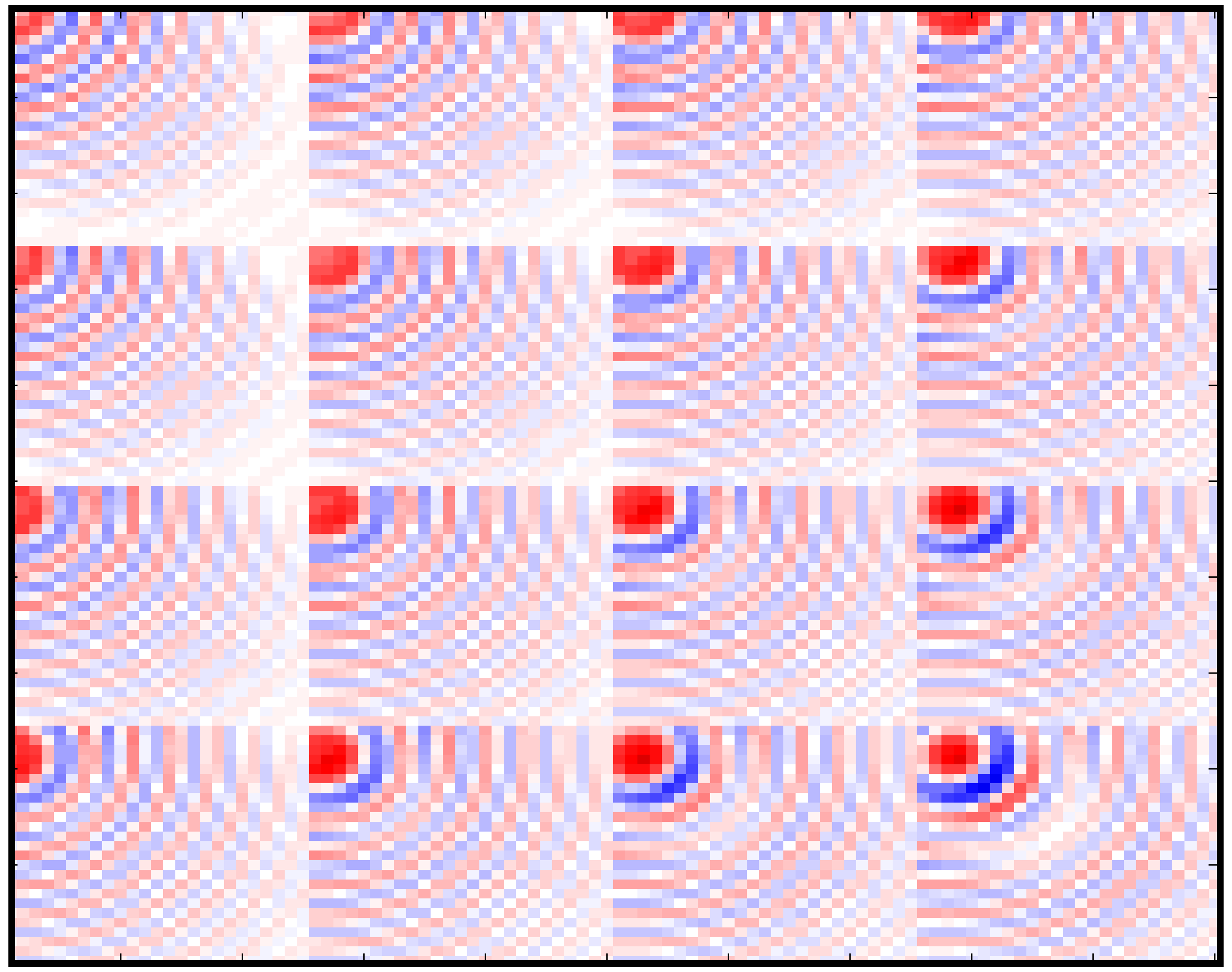


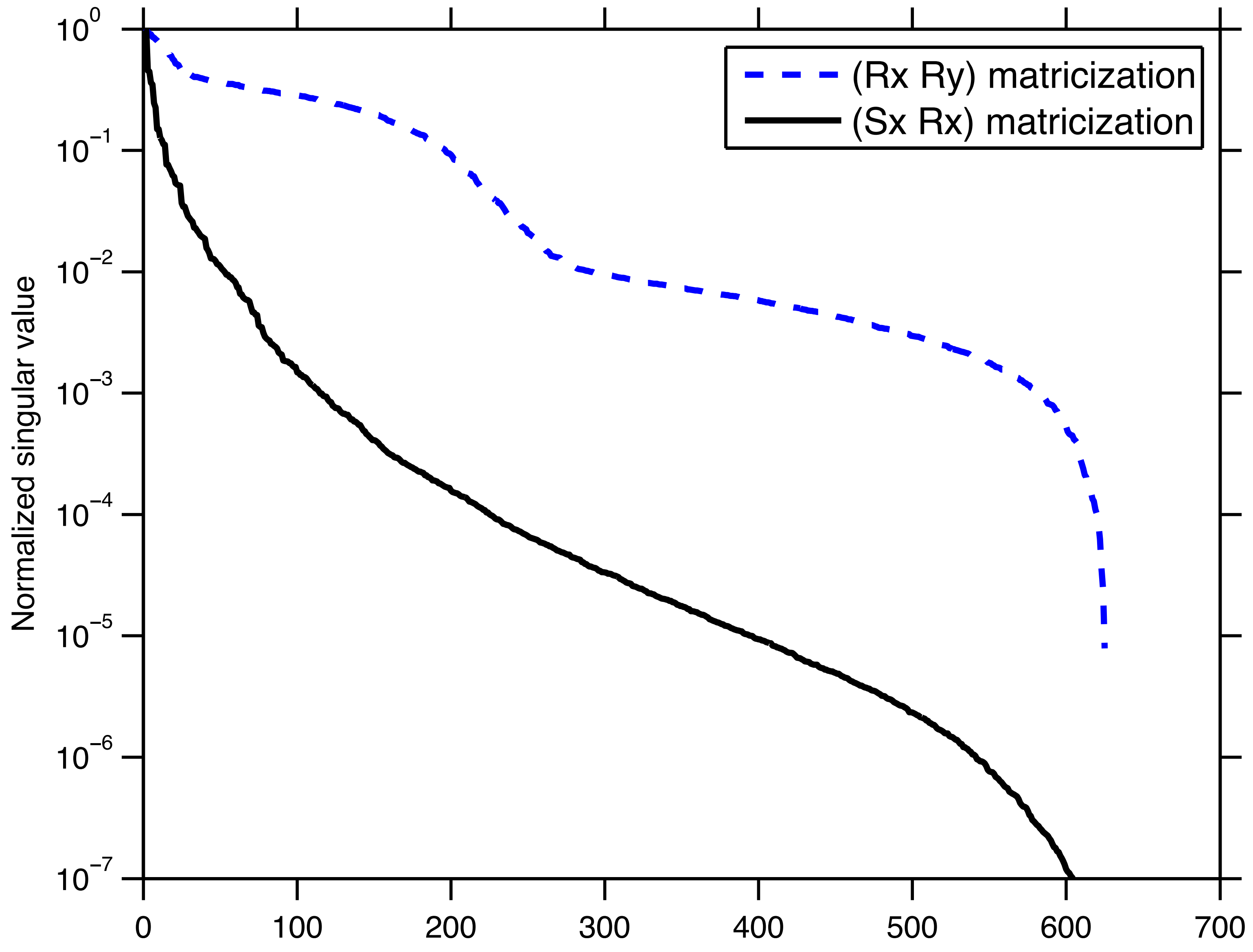


(Rec x, Rec y) matricization - Canonical ordering

Matricizations







(Src x, Rec x) matricization - Noncanonical ordering

Multidimensional interpolation

with Hierarchical Tucker

Successful reconstruction scheme

Signal structure

- Hierarchical Tucker

Sampling

- ***subsampling increases hierarchical rank***

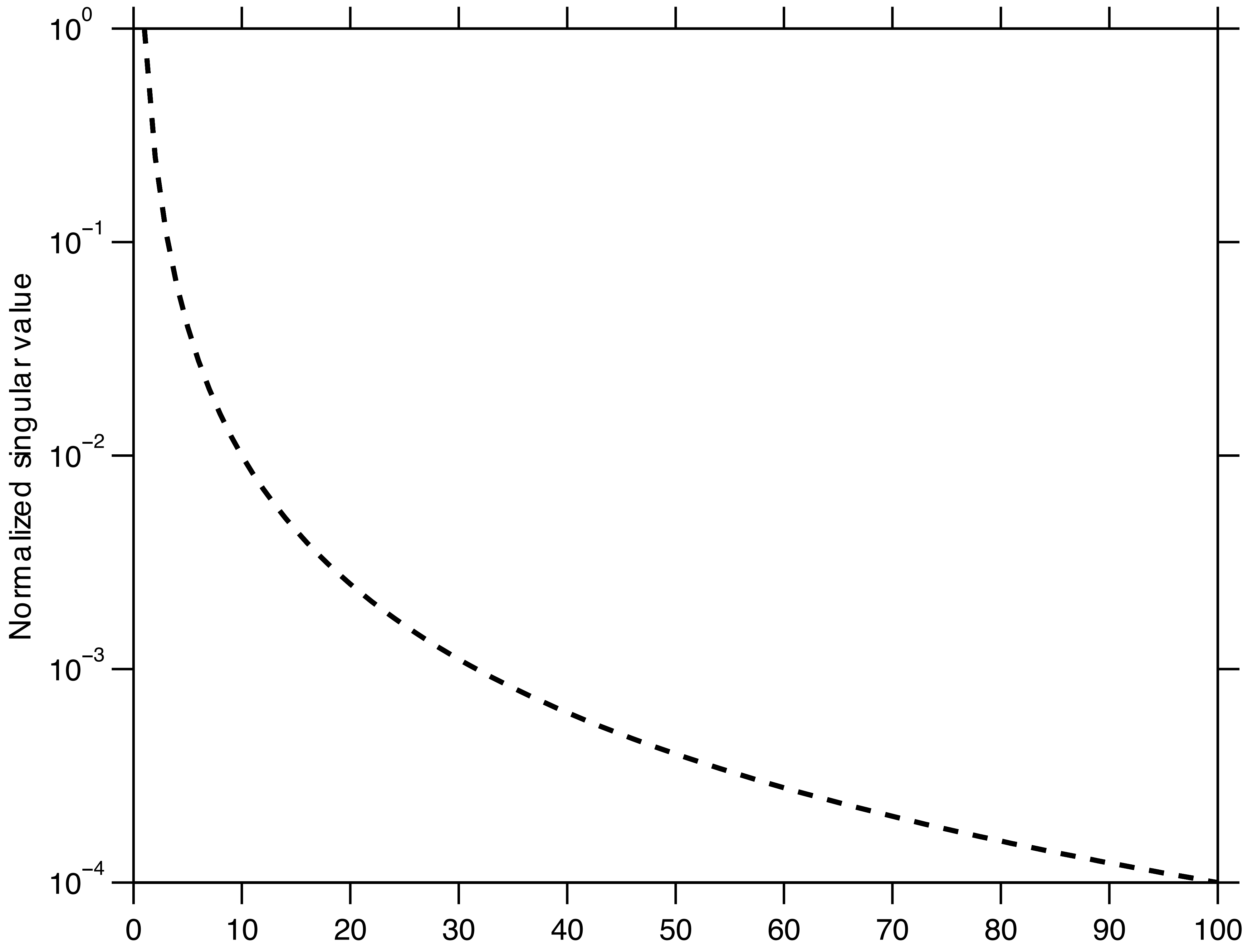
Optimization

- fit data in the Hierarchical Tucker format

Matrix Completion

$$\mathbf{X}$$

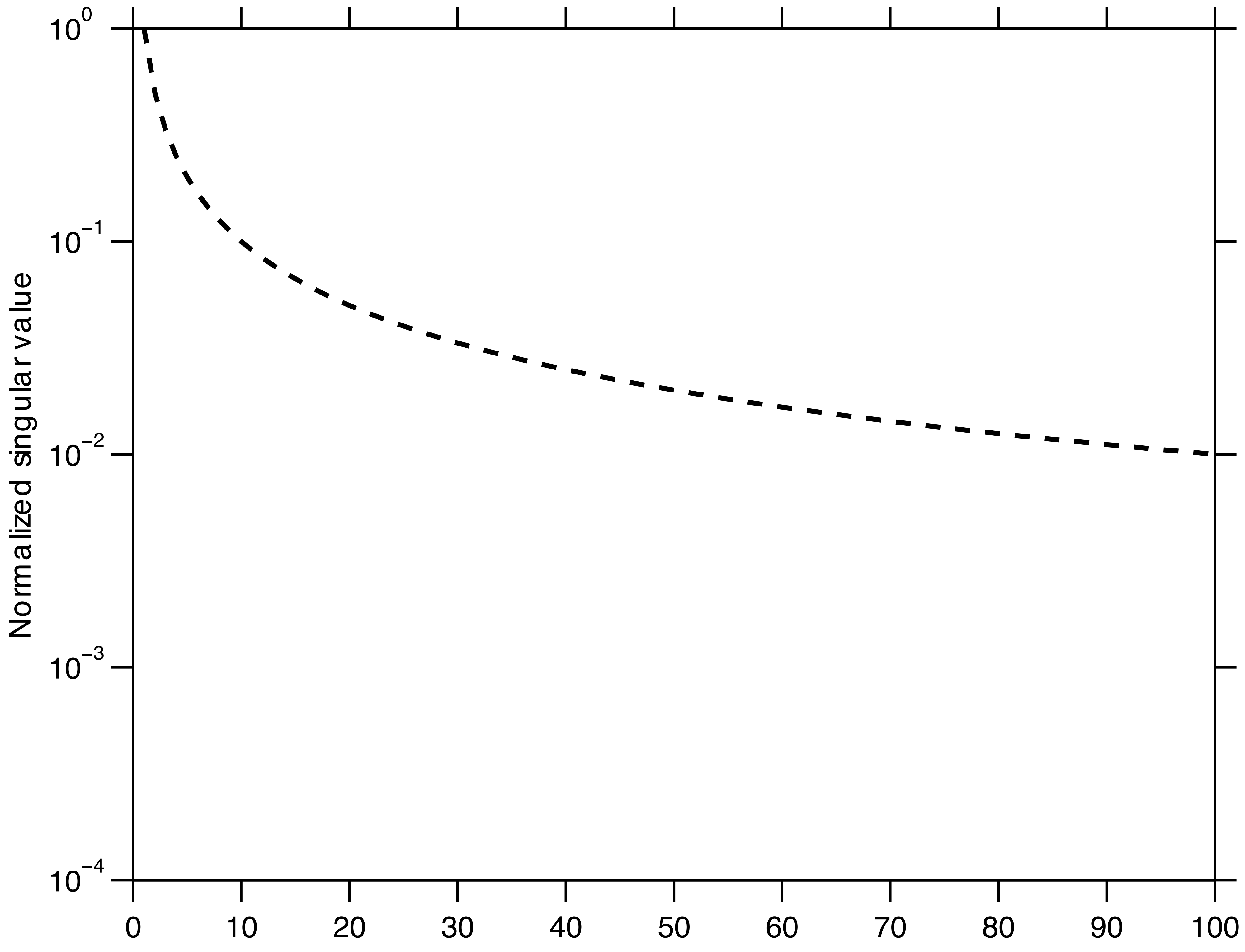
$$\begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}$$



Matrix Completion

$$\mathcal{A}(\mathbf{X})$$

$$\begin{bmatrix} * & * & * & 0 & * \\ * & 0 & 0 & * & 0 \\ * & * & * & * & * \\ * & * & 0 & * & * \\ 0 & * & * & * & 0 \end{bmatrix}$$



Tensor Completion

Structure - recover a tensor X which has low hierarchical rank

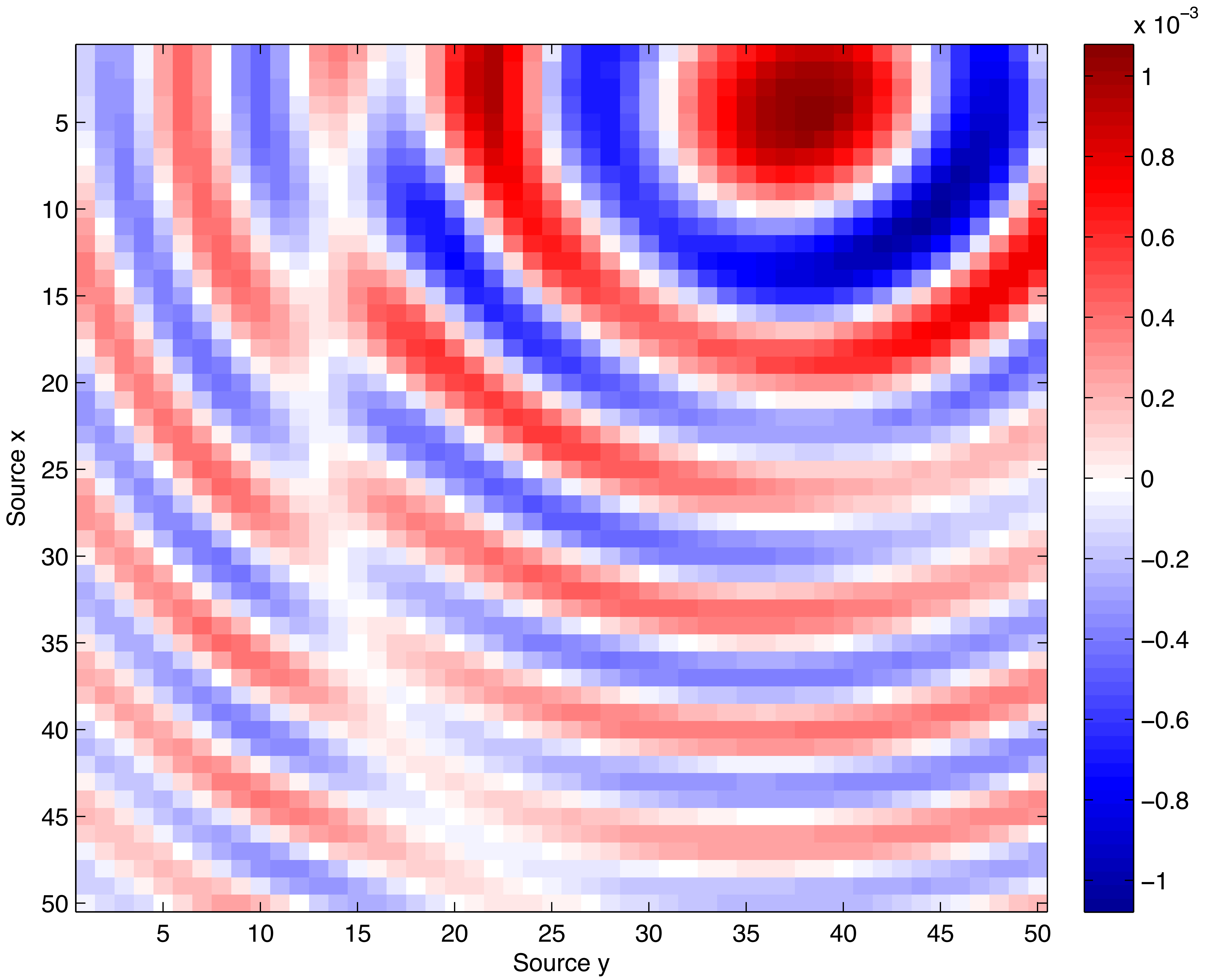
- Well represented in HT

Sampling - random removal of points increases rank

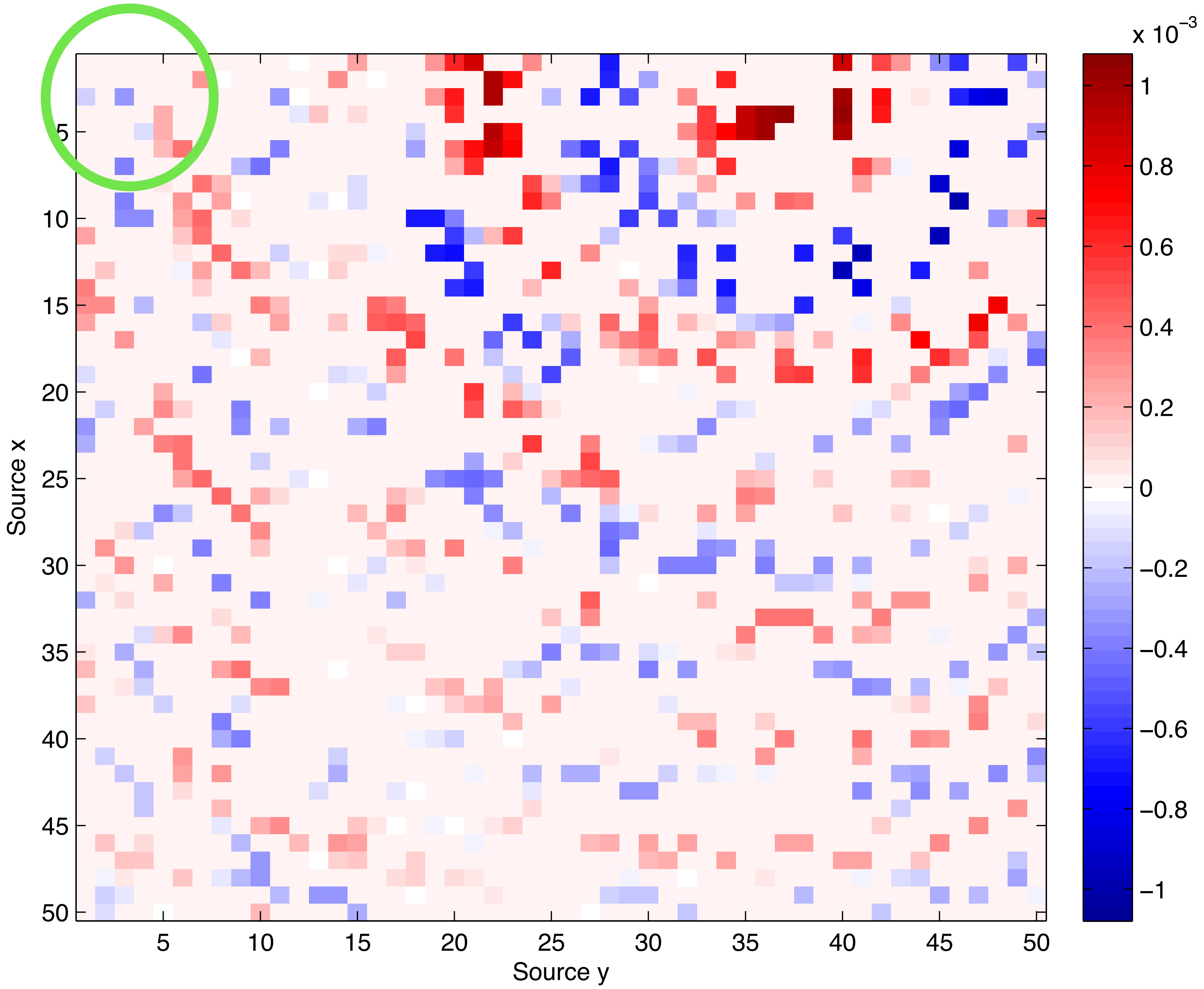
- Poorly represented in HT
- Idealized sampling

Idealized recovery

75% random entries removed
Common receiver gather



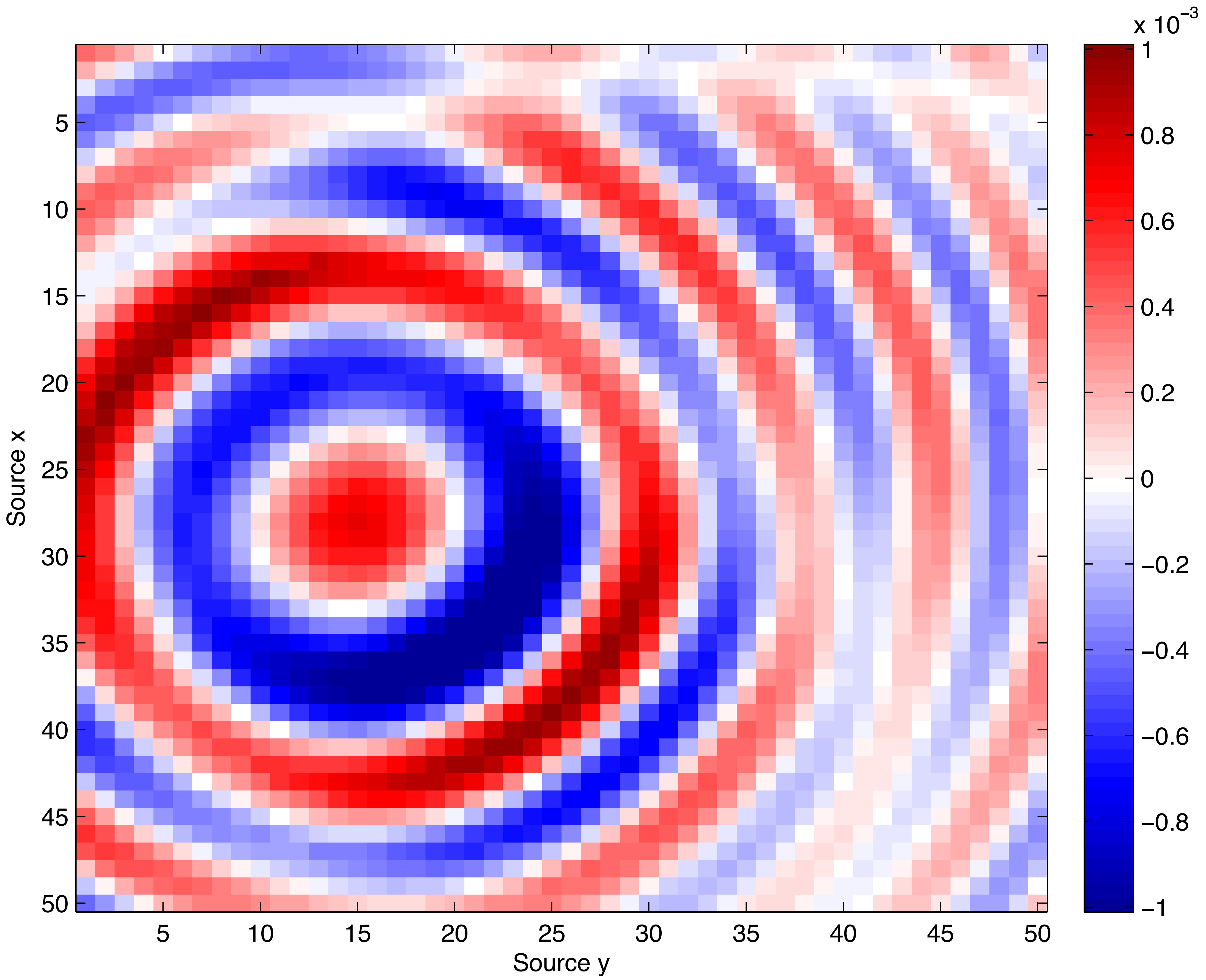
True data



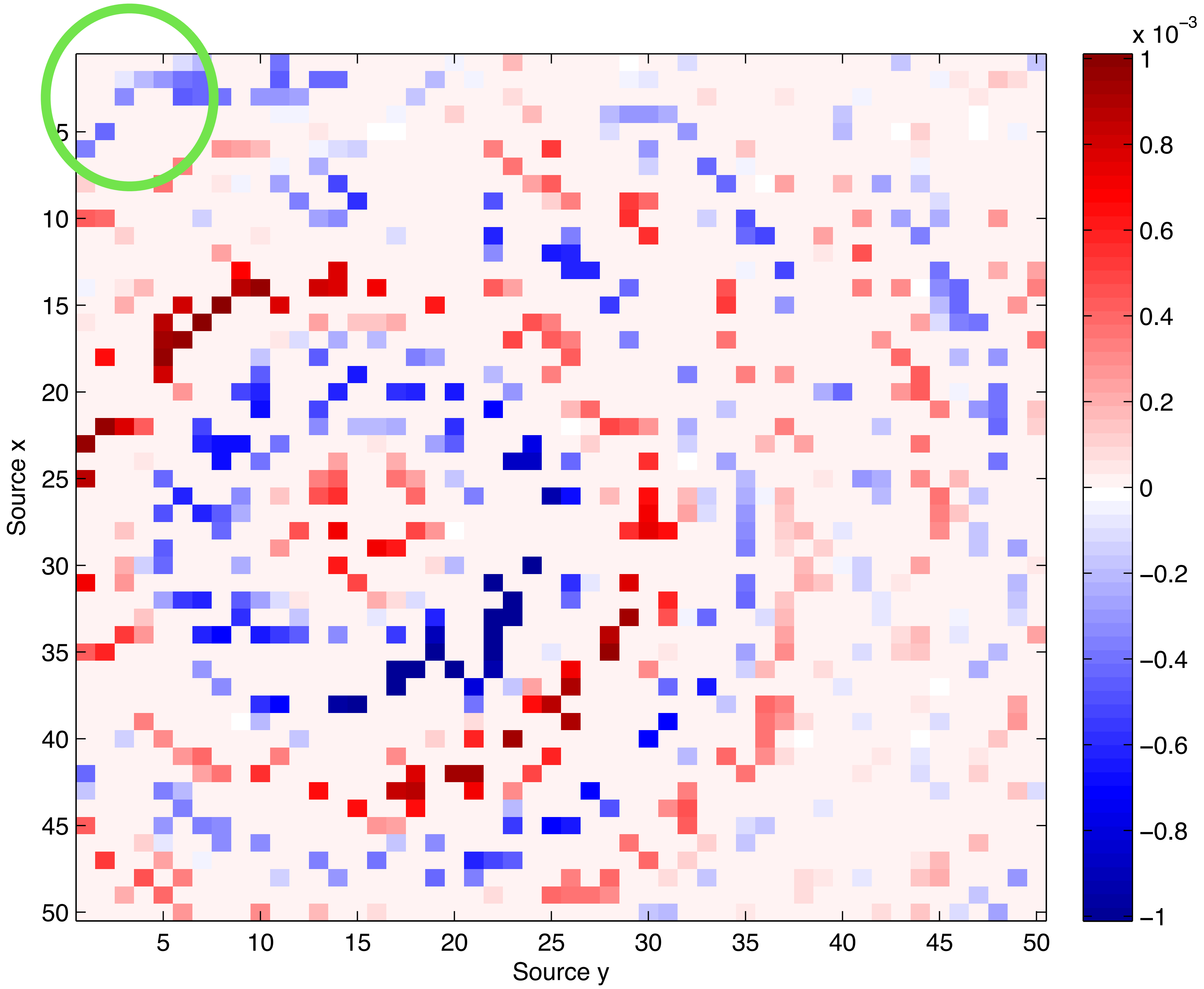
Subsampled data

Idealized recovery

75% random entries removed
Common receiver gather



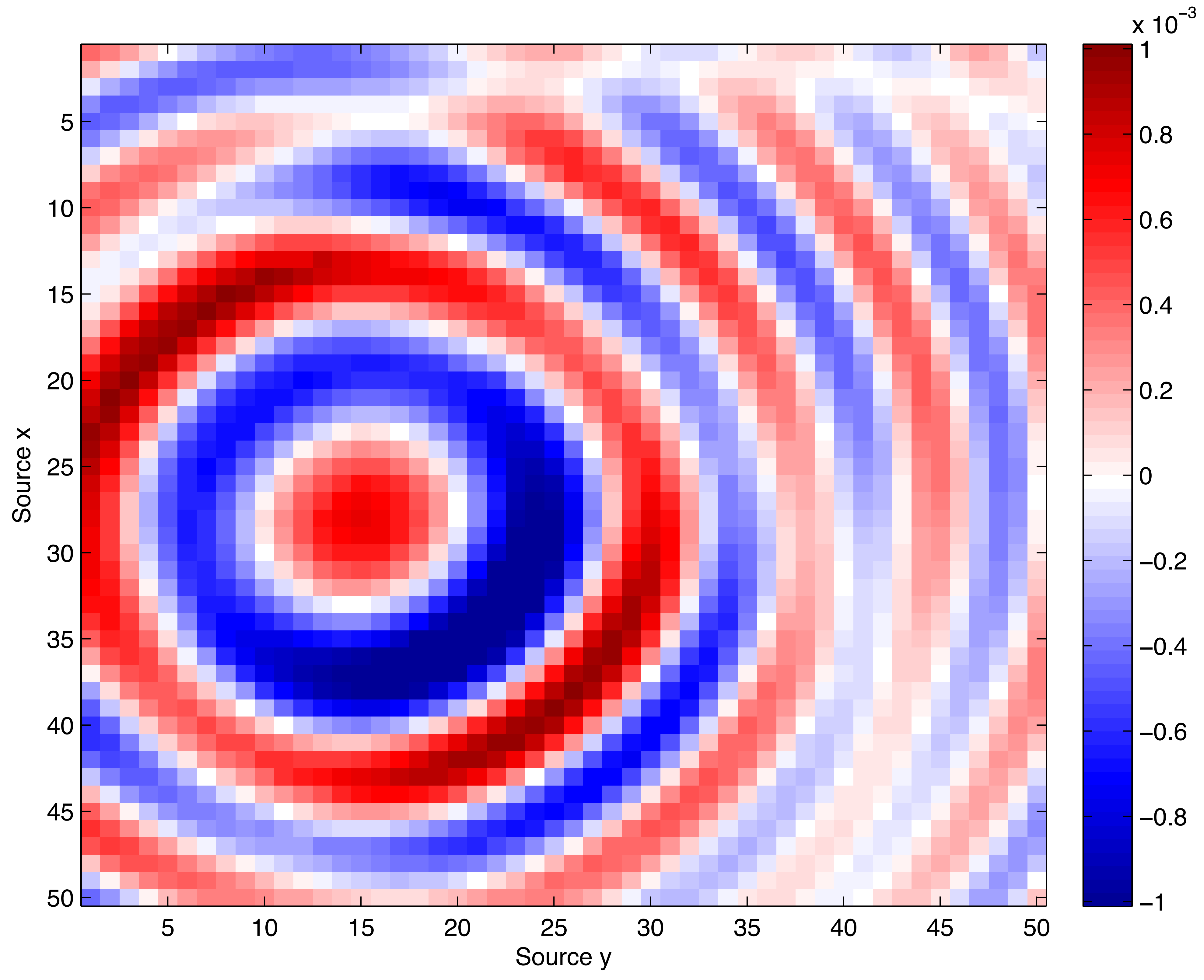
True data



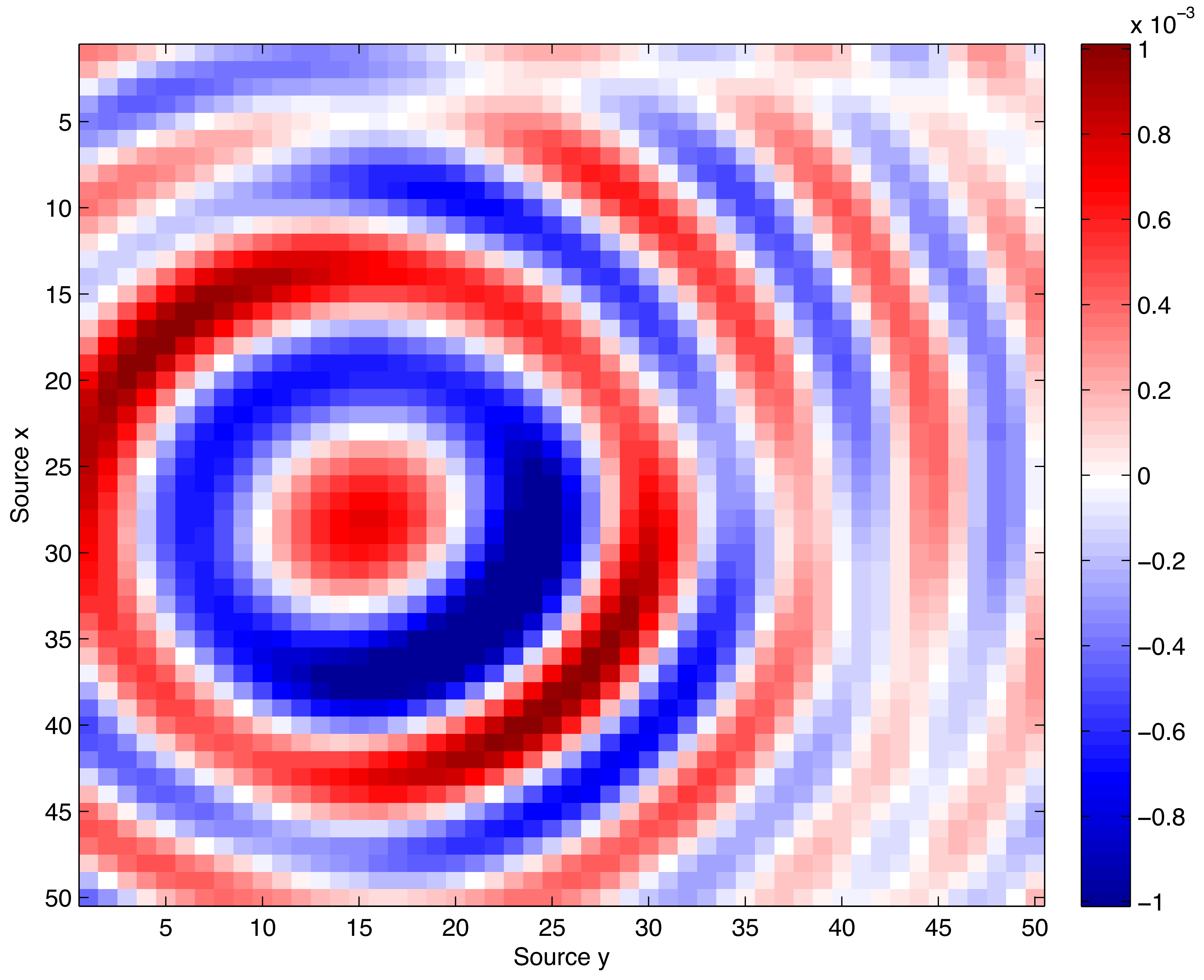
Subsampled data

Idealized recovery

75% random entries removed
Common receiver gather



True data



Recovered data - SNR 19.3 dB

SLIM

Sampling

Sampling $(x_{\mathrm{src}}, y_{\mathrm{src}}, x_{\mathrm{rec}}, y_{\mathrm{rec}})$ points from the data

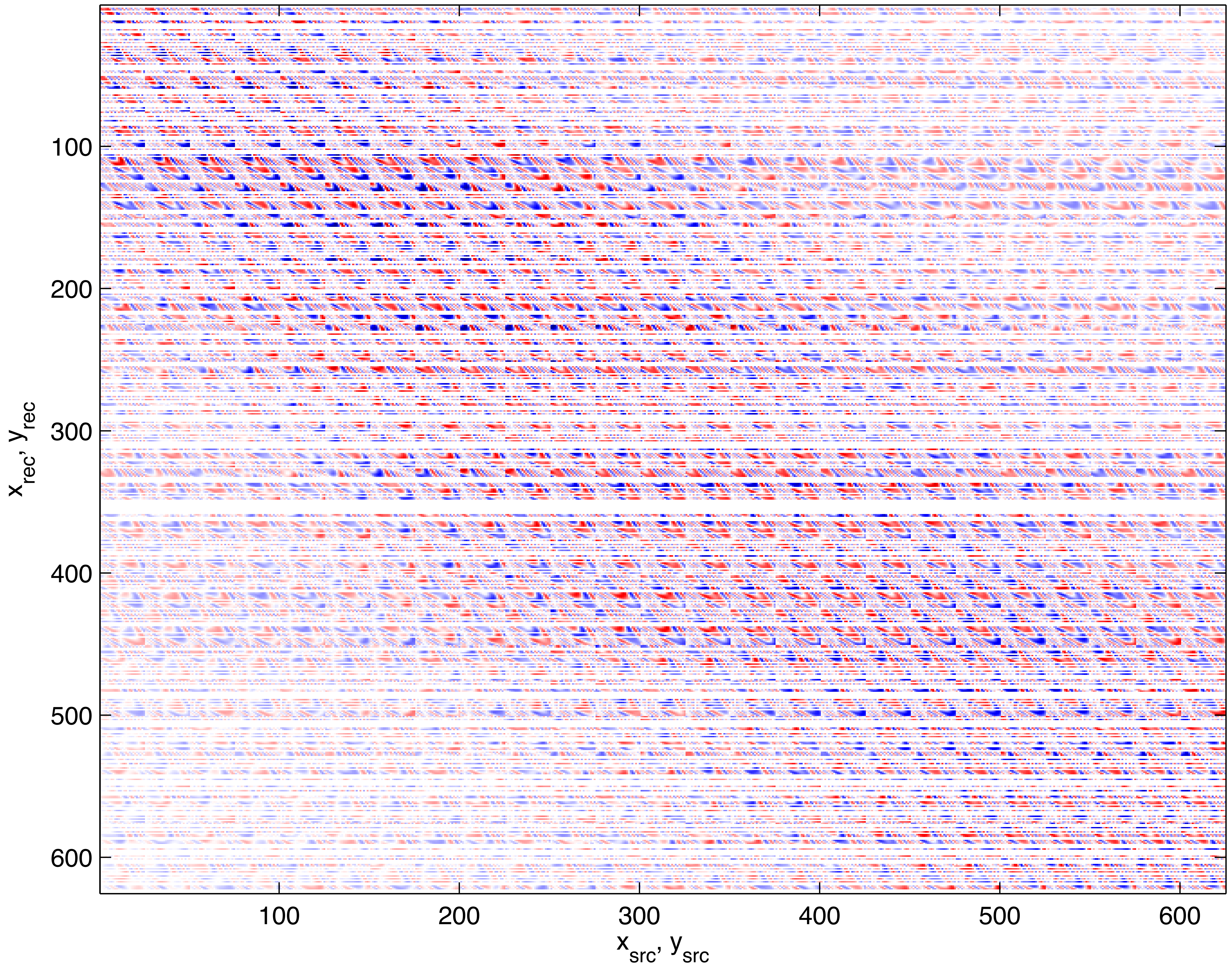
- idealized recovery
- impossible to physically implement

Sampling $(x_{\mathrm{rec}}, y_{\mathrm{rec}})$ points from the data

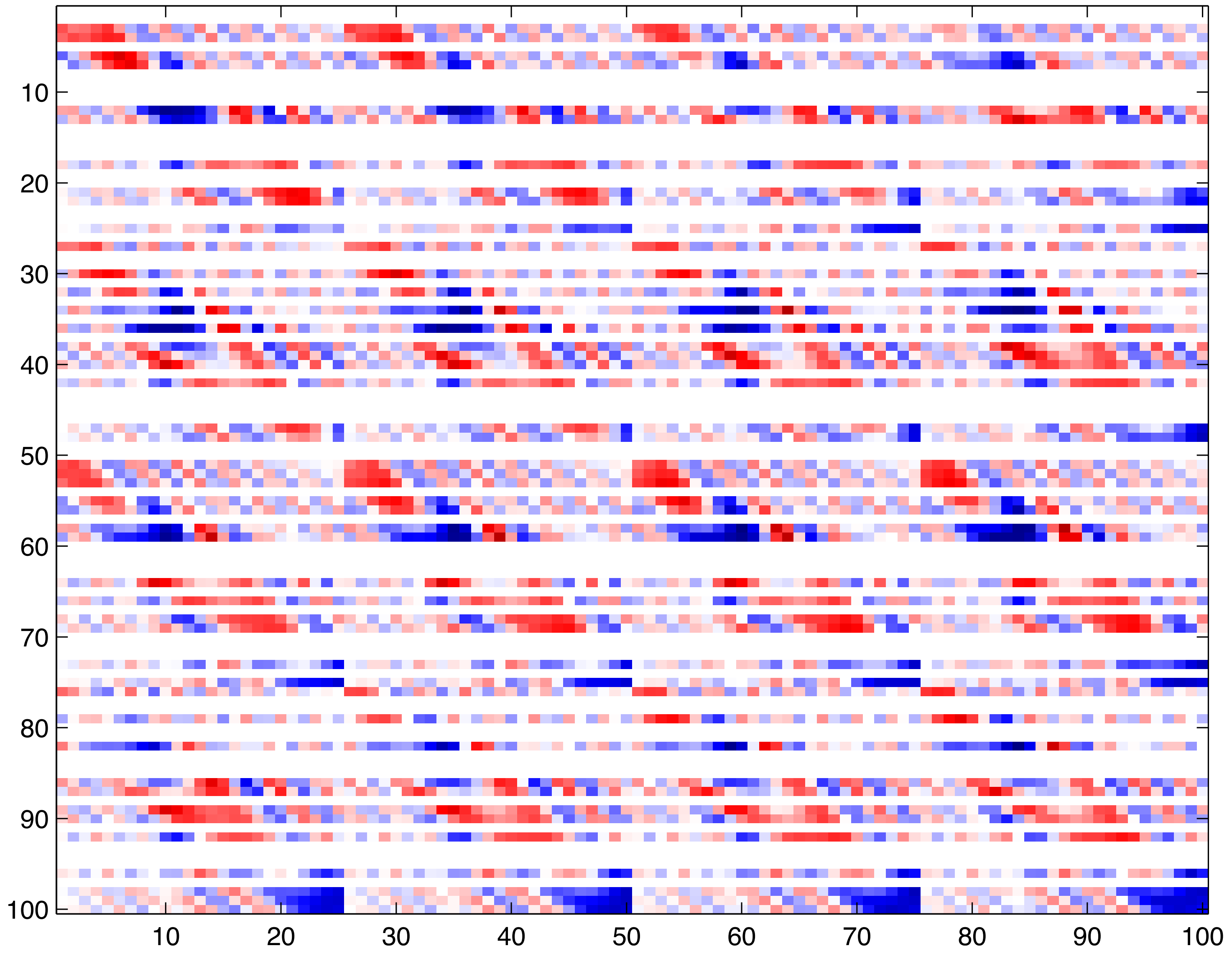
- less idealized
- possible to acquire data - e.g. ocean bottom nodes

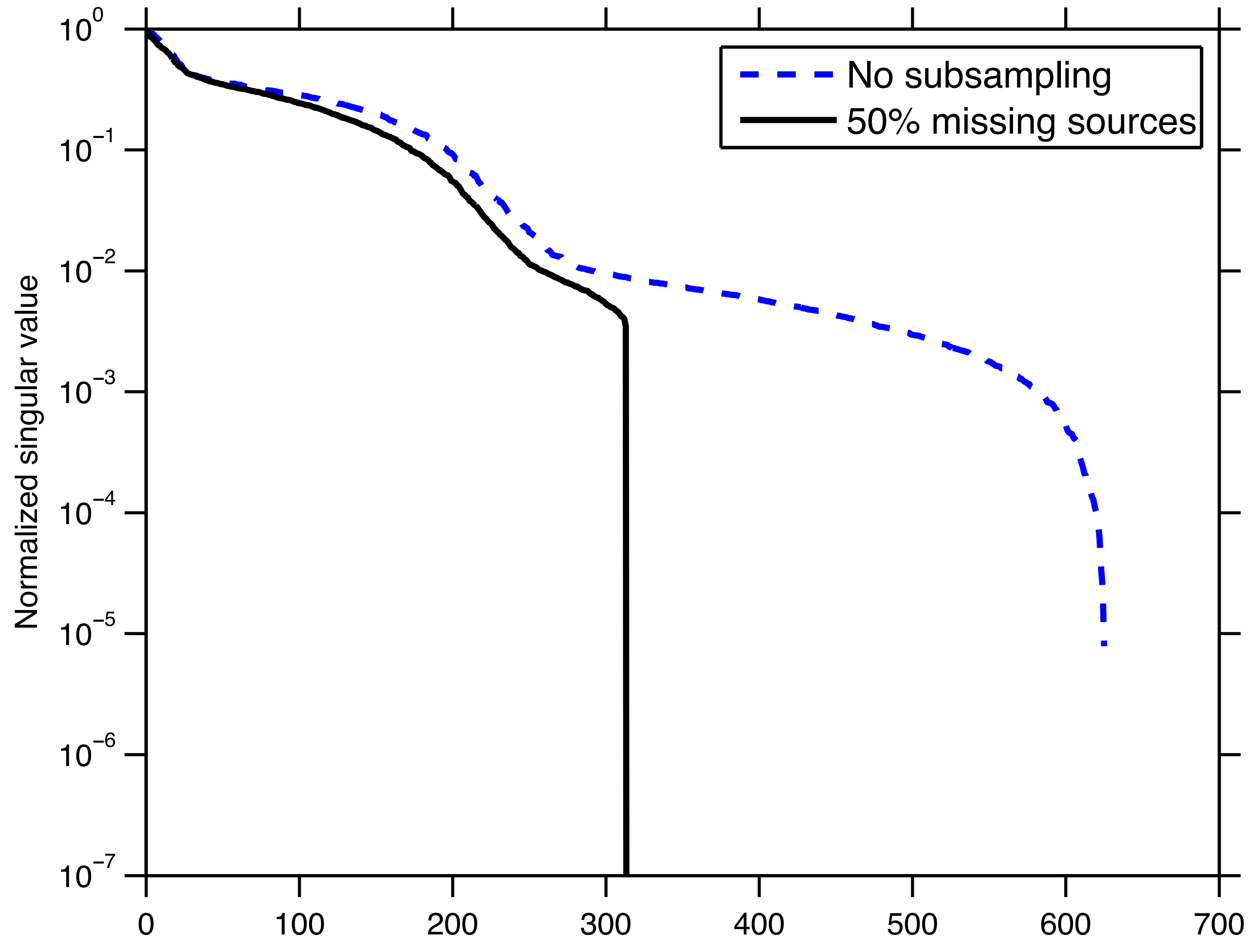
Realistic recovery

50% random receivers removed



(Rec x, Rec y) matricization - Canonical ordering

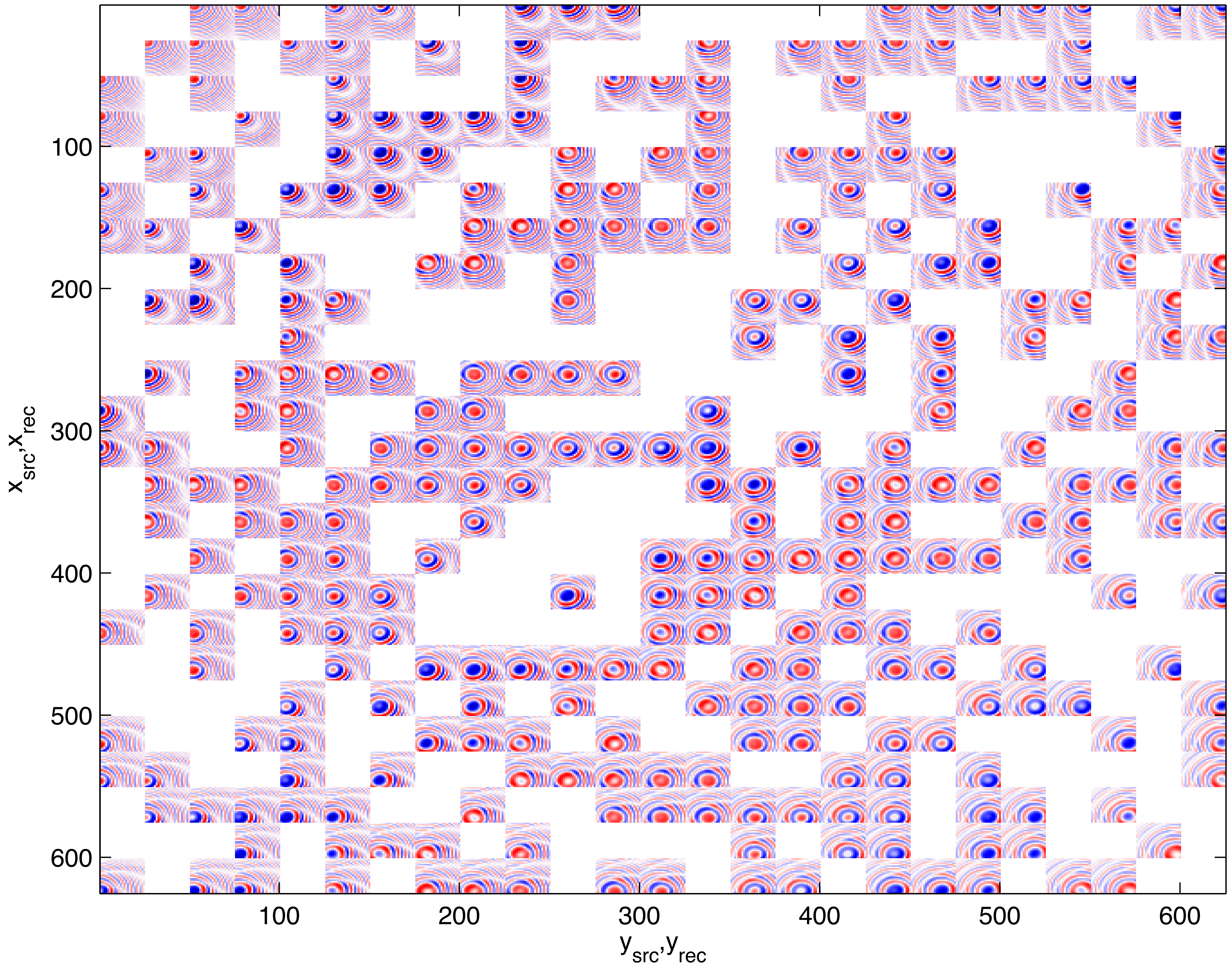


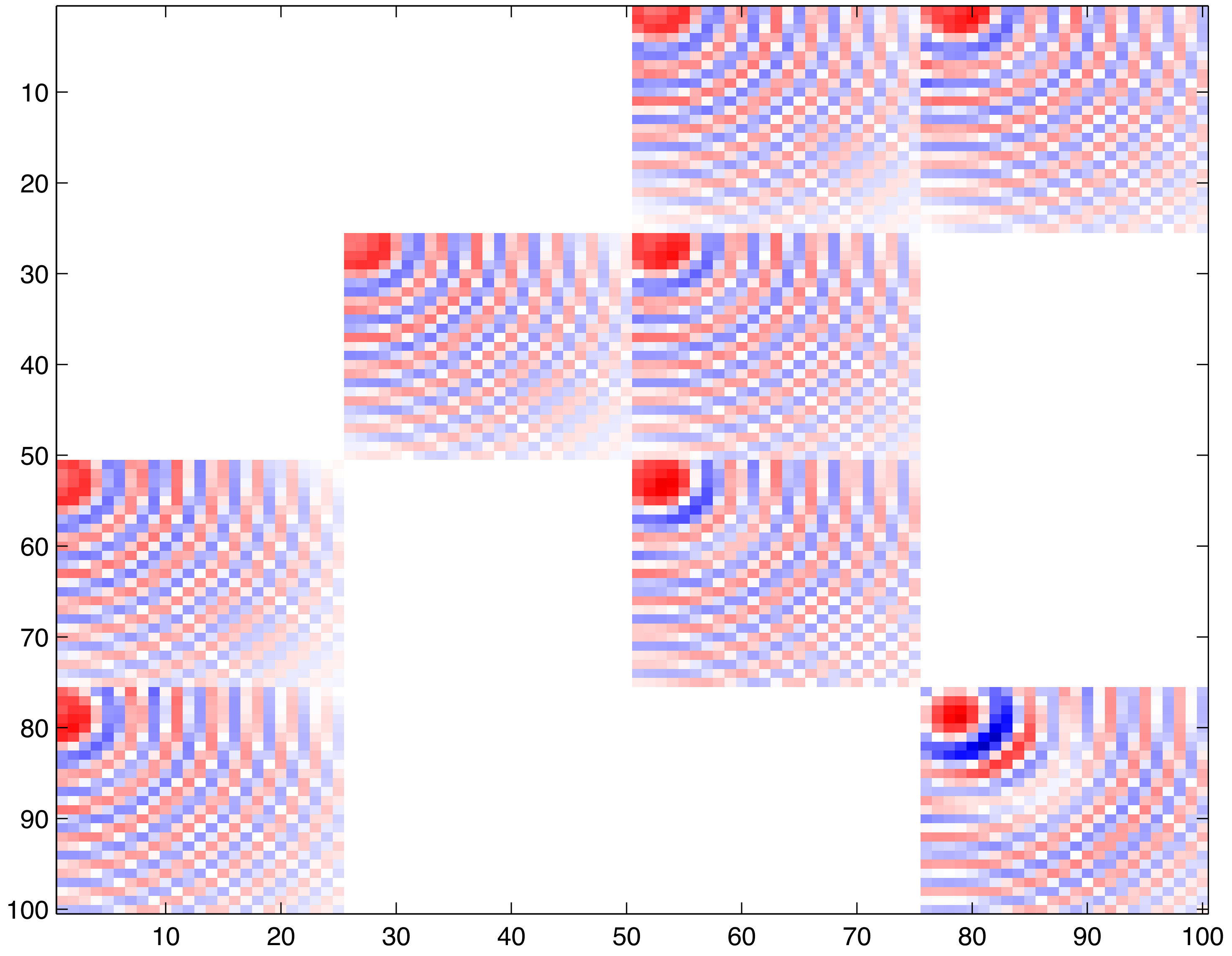


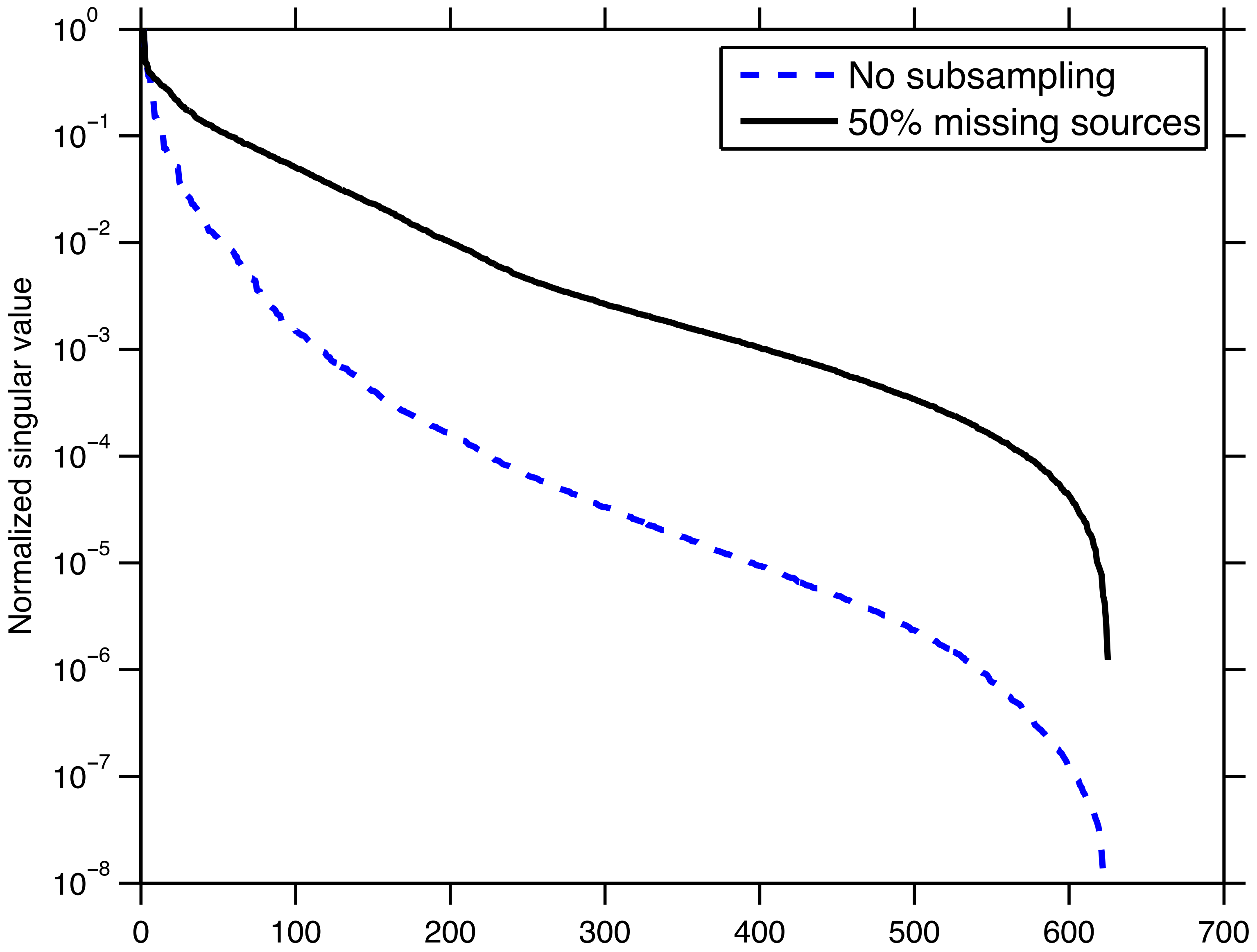
SLIM

Realistic recovery

50% random receivers removed







(Src x, Rec x) matricization - Noncanonical ordering

Data organization

In summary:

(rec x, rec y) organization

- High rank
- Missing sources operator - removes columns
- Poor recovery scenario

(src x, rec x) organization

- Low rank
- Missing sources operator - removes blocks
- Closer to ideal recovery scenario

Multidimensional interpolation
with Hierarchical Tucker

Successful reconstruction scheme

Signal structure

- Hierarchical Tucker

Sampling

- subsampling increases hierarchical rank

Optimization

- ***fit data in the Hierarchical Tucker format***

Optimization

Given data b with missing sources and/or receivers, subsampling operator A, full tensor expansion operator

$$\phi : (U_t, B_t) \rightarrow \mathbb{C}^{n_1 \times \ldots n_d}$$

solve

$$\min_{x=(U_t, B_t)} \frac{1}{2} \|A\phi(x) - b\|_2^2$$

A. Uschmajew, B. Vandereycken. *The geometry of algorithms using hierarchical tensors*. Linear algebra and its applications, 2013
C. Da Silva and F. J. Herrmann, *Optimization on the Hierarchical Tucker manifold - applications to tensor completion*, 2013

Differential geometry

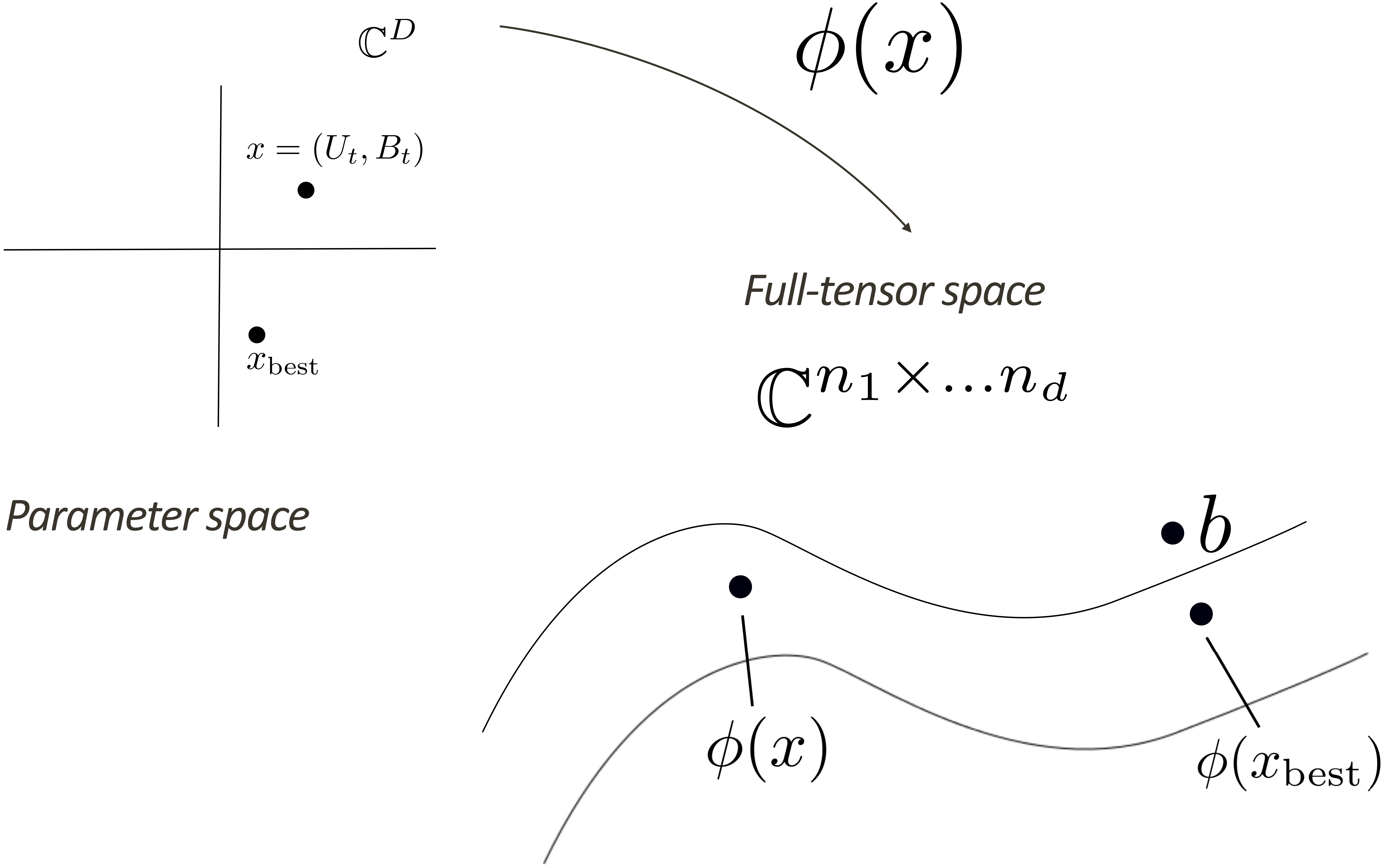
HT tensors parametrize a submanifold of full tensor space $\mathbb{C}^{n_1 \times \dots n_d}$

- Nonlinear, nonconvex space
- Generalization of curved surfaces

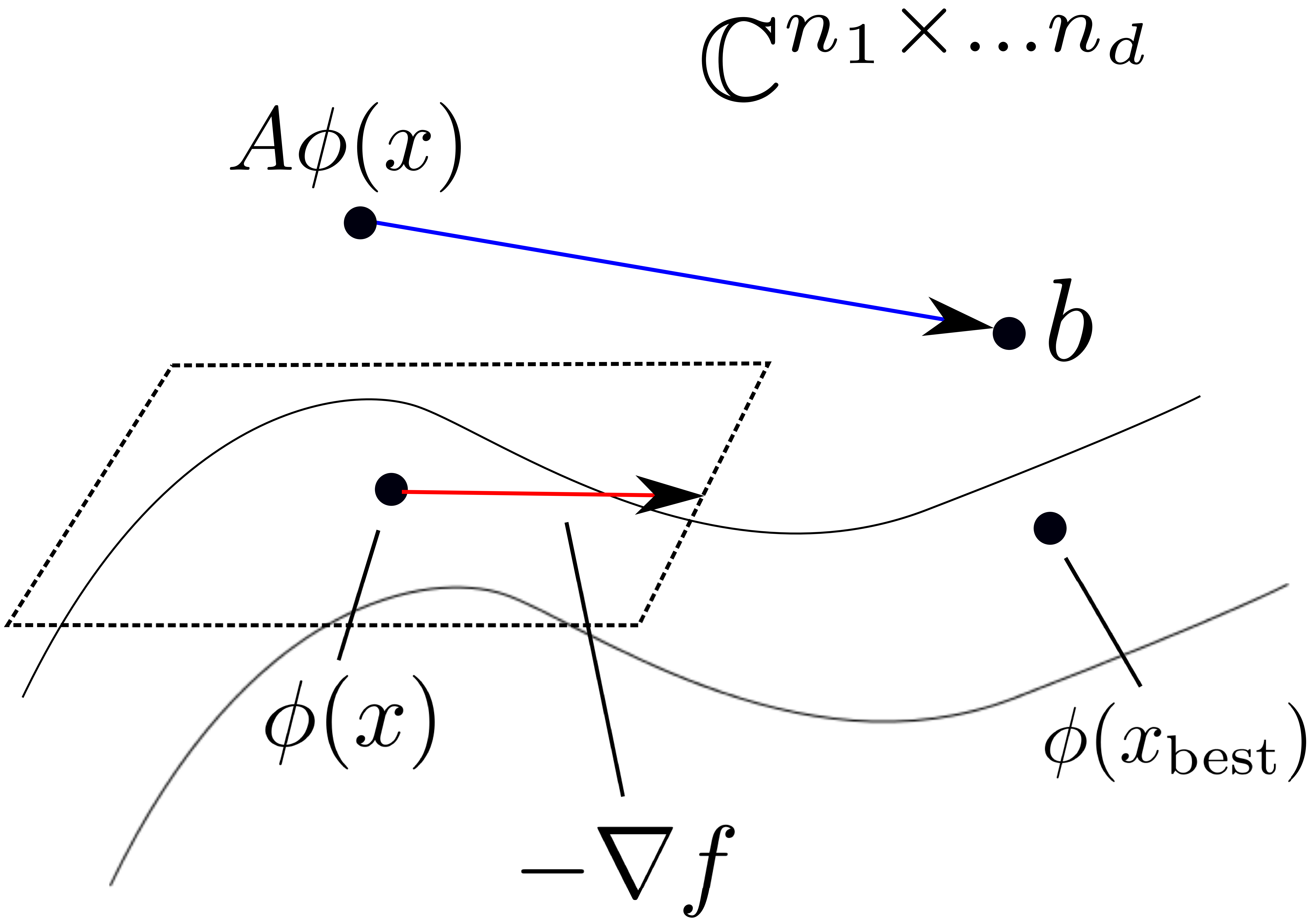
Steepest Descent, Conjugate gradient, Gauss-Newton

- ***without*** SVDs in the full tensor space

Optimization program



Optimization program



Derivatives

Derivatives of a particular node with respect to its children can be computed efficiently, i.e. via

$$(I - U_{t_l} U_{t_l}^H) \langle U_{t_r}^T \circ_2 Z, B_t \rangle_{(2,3),(2,3)}$$
$$(I - U_{t_r} U_{t_r}^H) \langle U_{t_l}^T \circ_1 Z, B_t \rangle_{(1,3),(1,3)}$$

$P \circ_i Q$ multiplies Q by P in dimension i, $\langle X, Y \rangle_{(1,3),(1,3)} = \sum_{i_1, i_3} \overline{X}_{i_1,:,i_3} Y_{i_1,:,i_3}$

The chain rule gives the gradient of the function ϕ

Derivatives

Only involves matrix-matrix multiplications of small matrices compared to the full-tensor space

Parallelizable - multilinear product can be done in parallel

SVD-free - no large-scale SVDs, unlike nuclear norm-based methods

Results

Synthetic BG Group data

Unknown model

- 68 x 68 sources with 401 x 401 receivers, data at 7.34 Hz

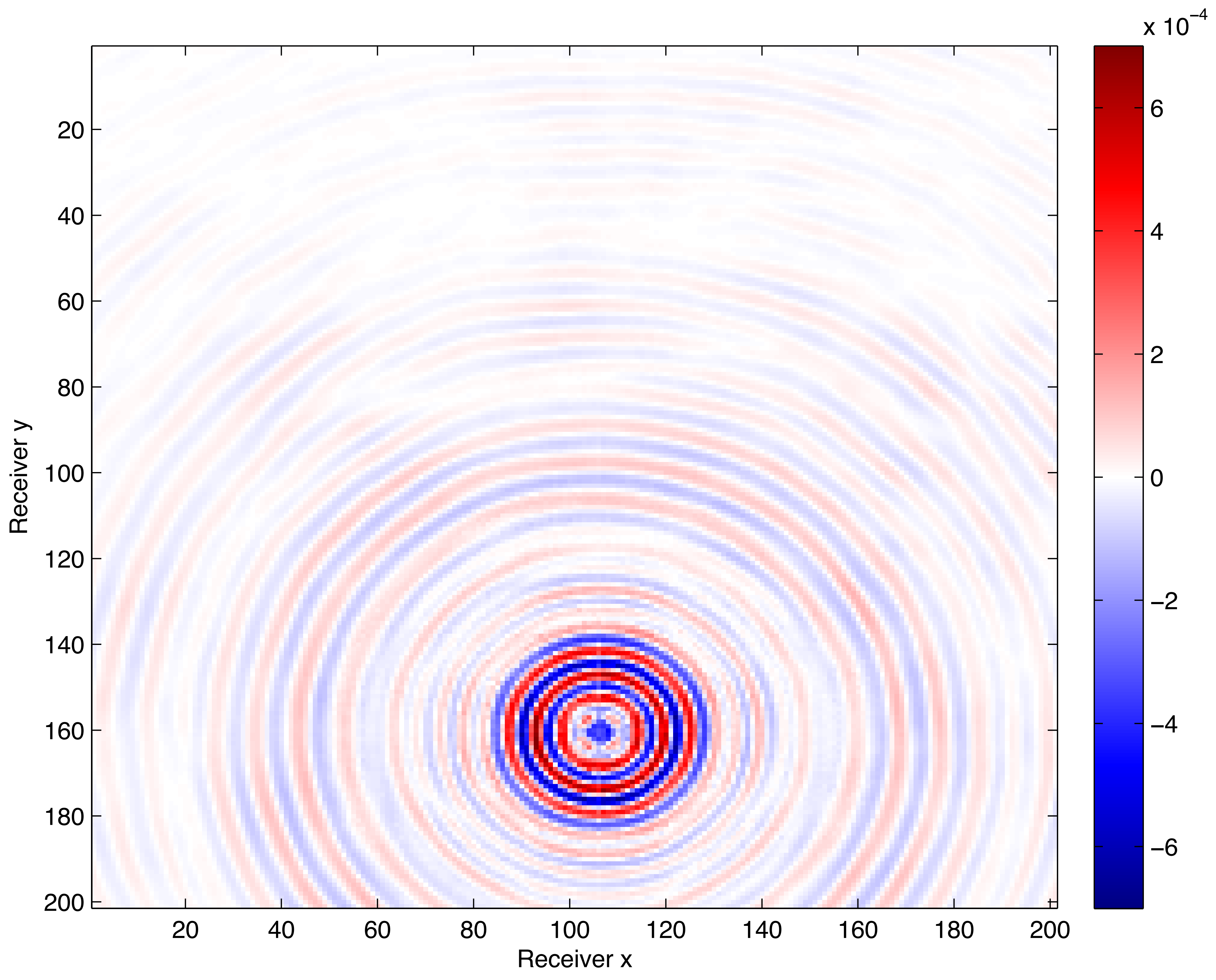
Receivers subsampled to 201 x 201

Recovered with Gauss-Newton

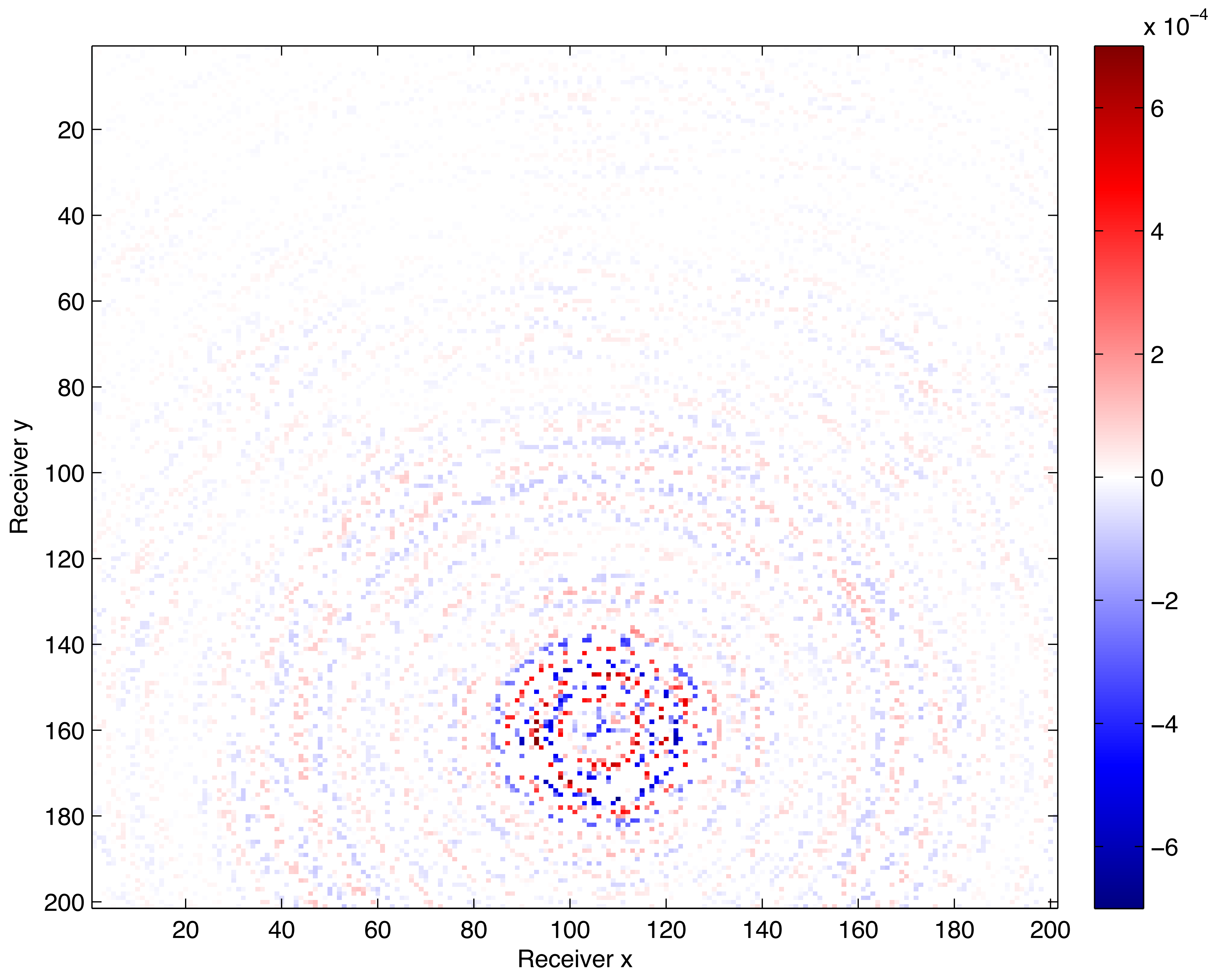
- code available at https://www.slim.eos.ubc.ca - HTOpt toolbox
- ~ 1 hour serial running time

7.34 Hz - 75% missing receivers

Common source gather



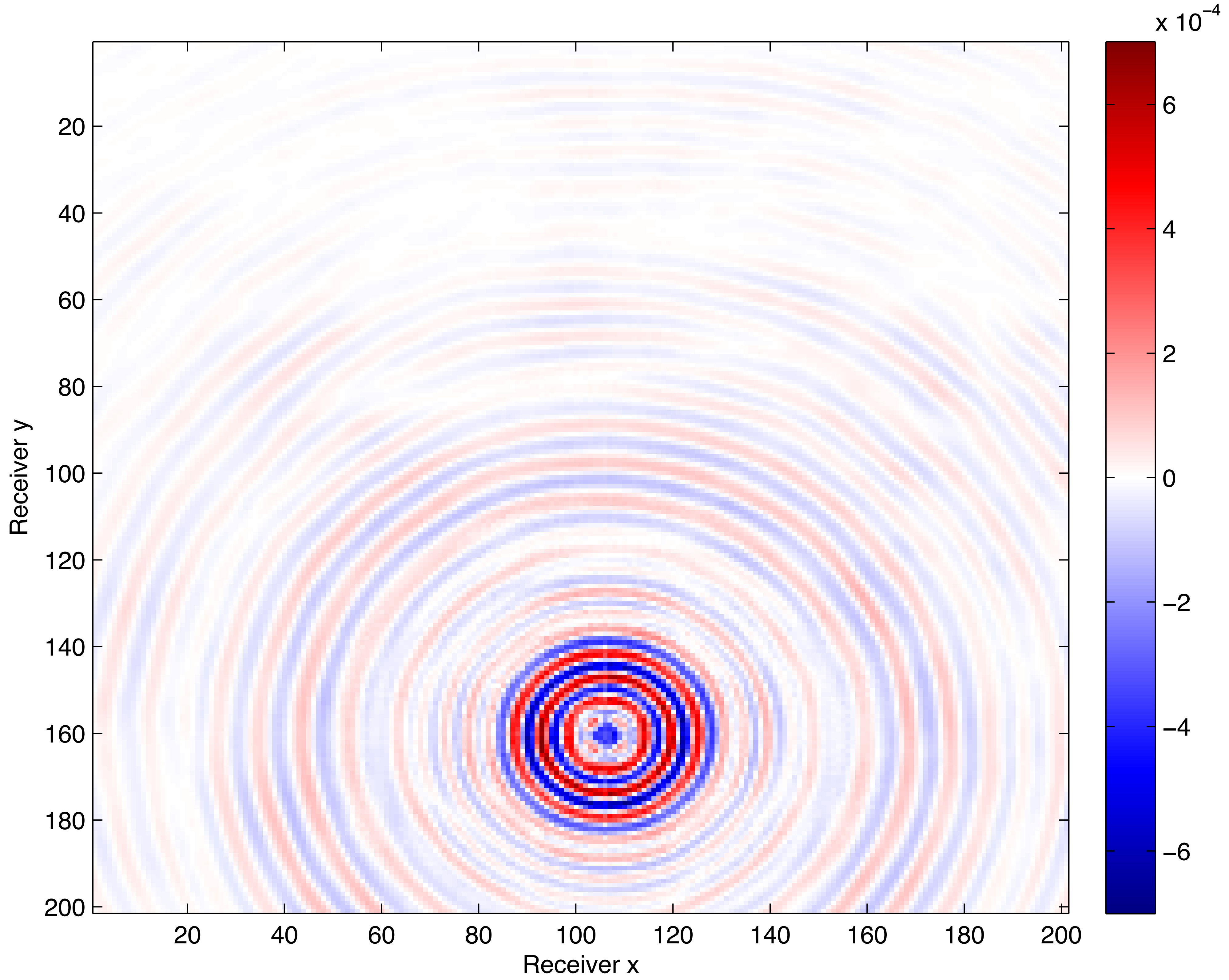
True data



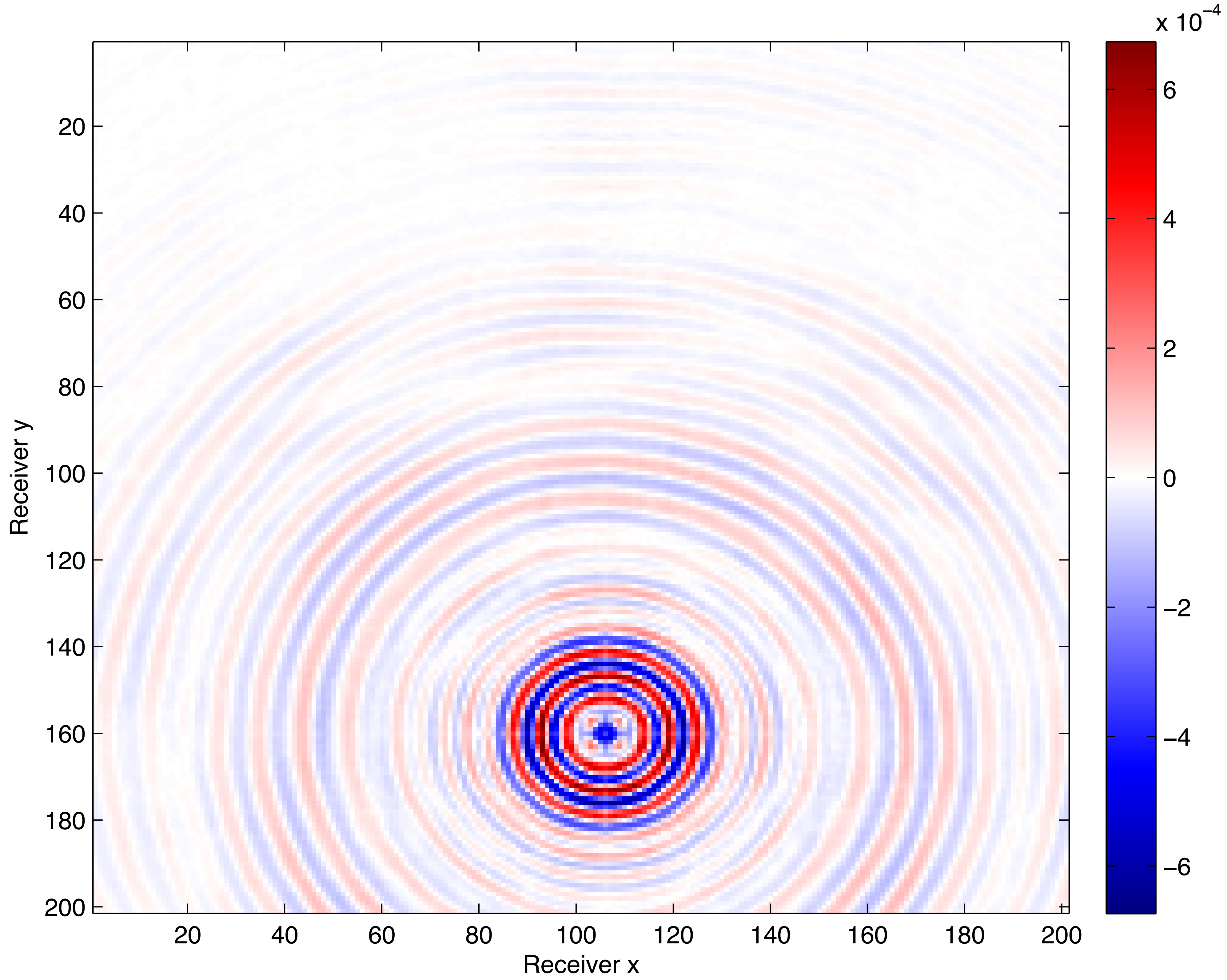
Subsampled data

7.34 Hz - 75% missing receivers

Common source gather



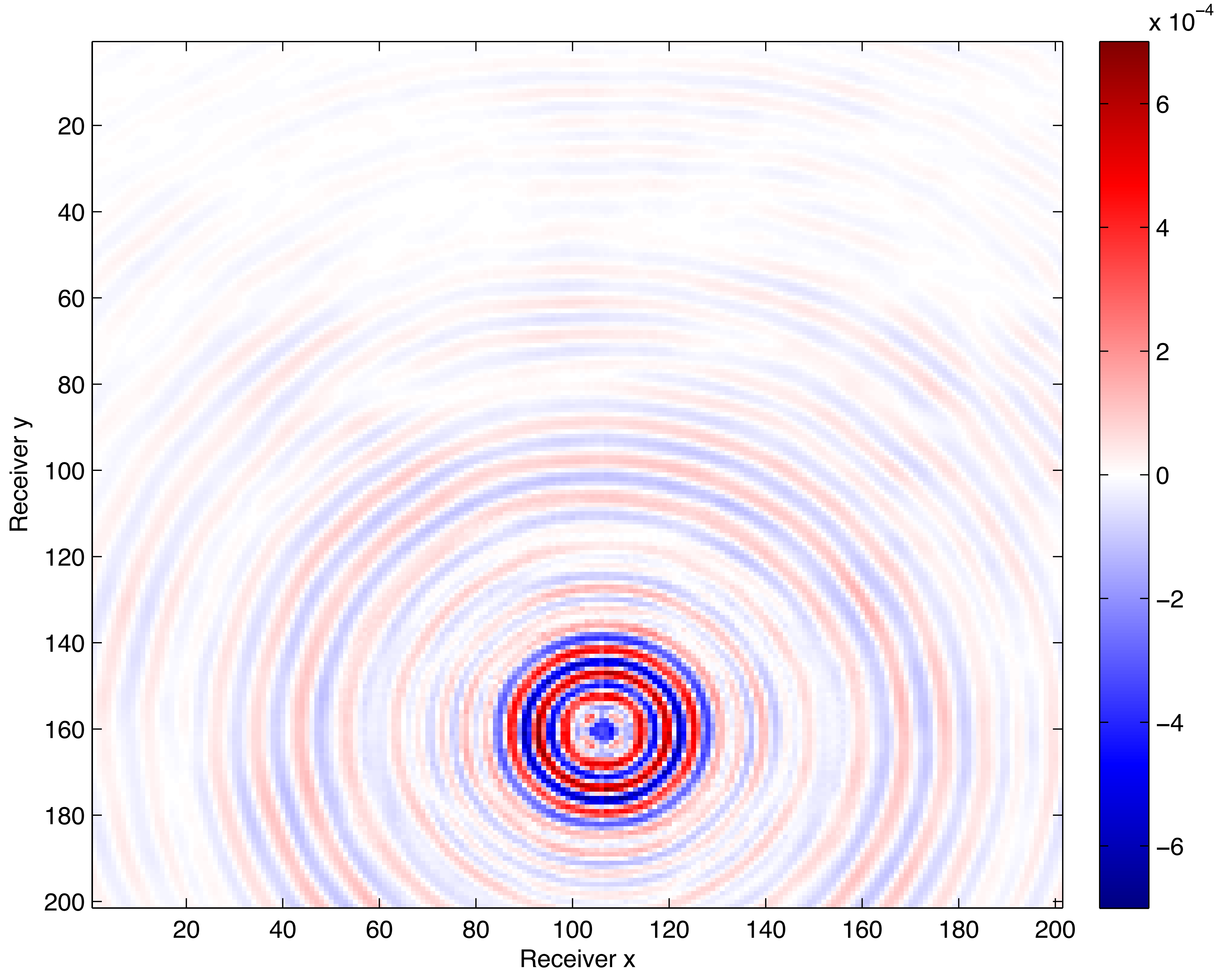
True data



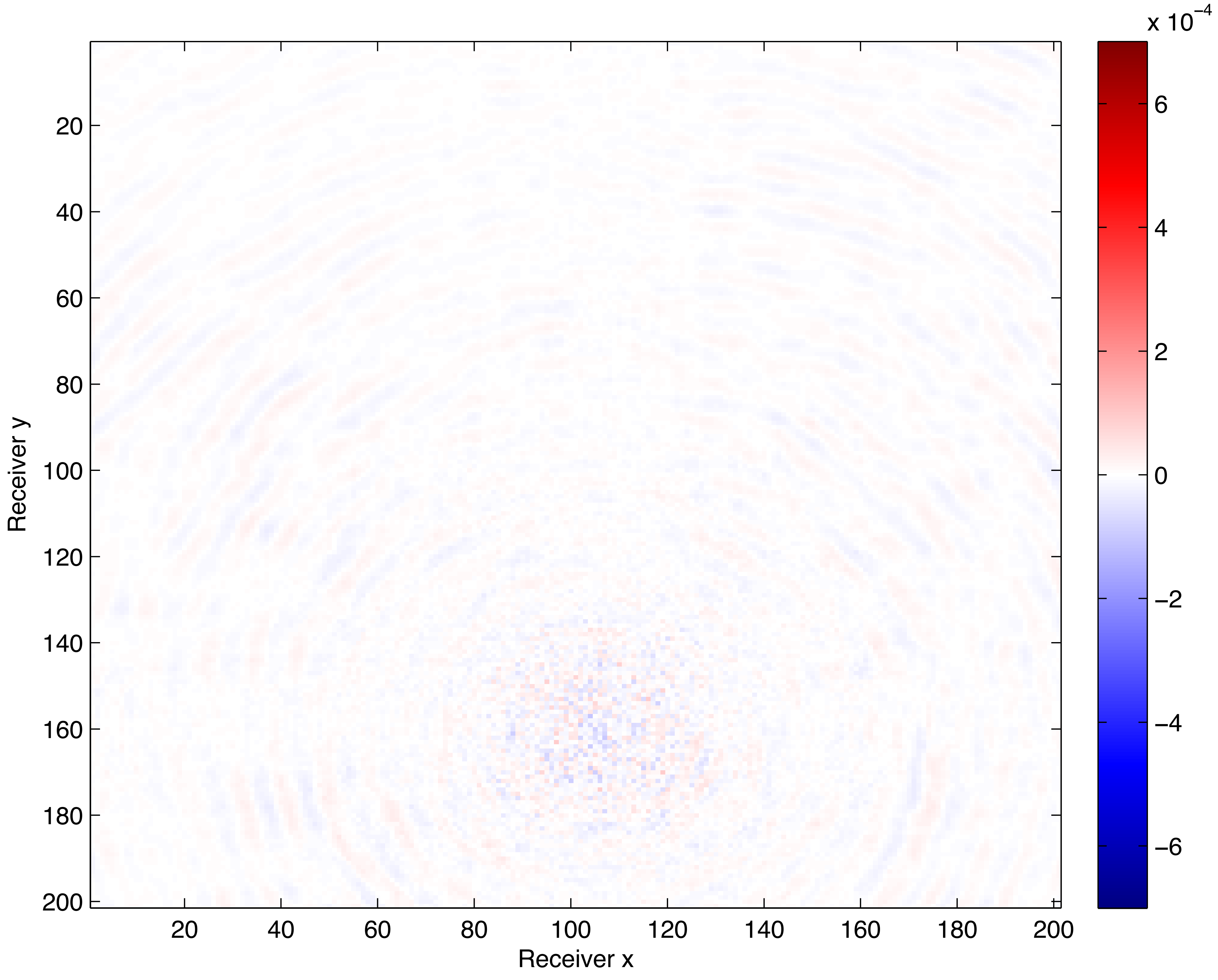
Recovered data - SNR 17.6 dB

7.34 Hz - 75% missing receivers

Common source gather



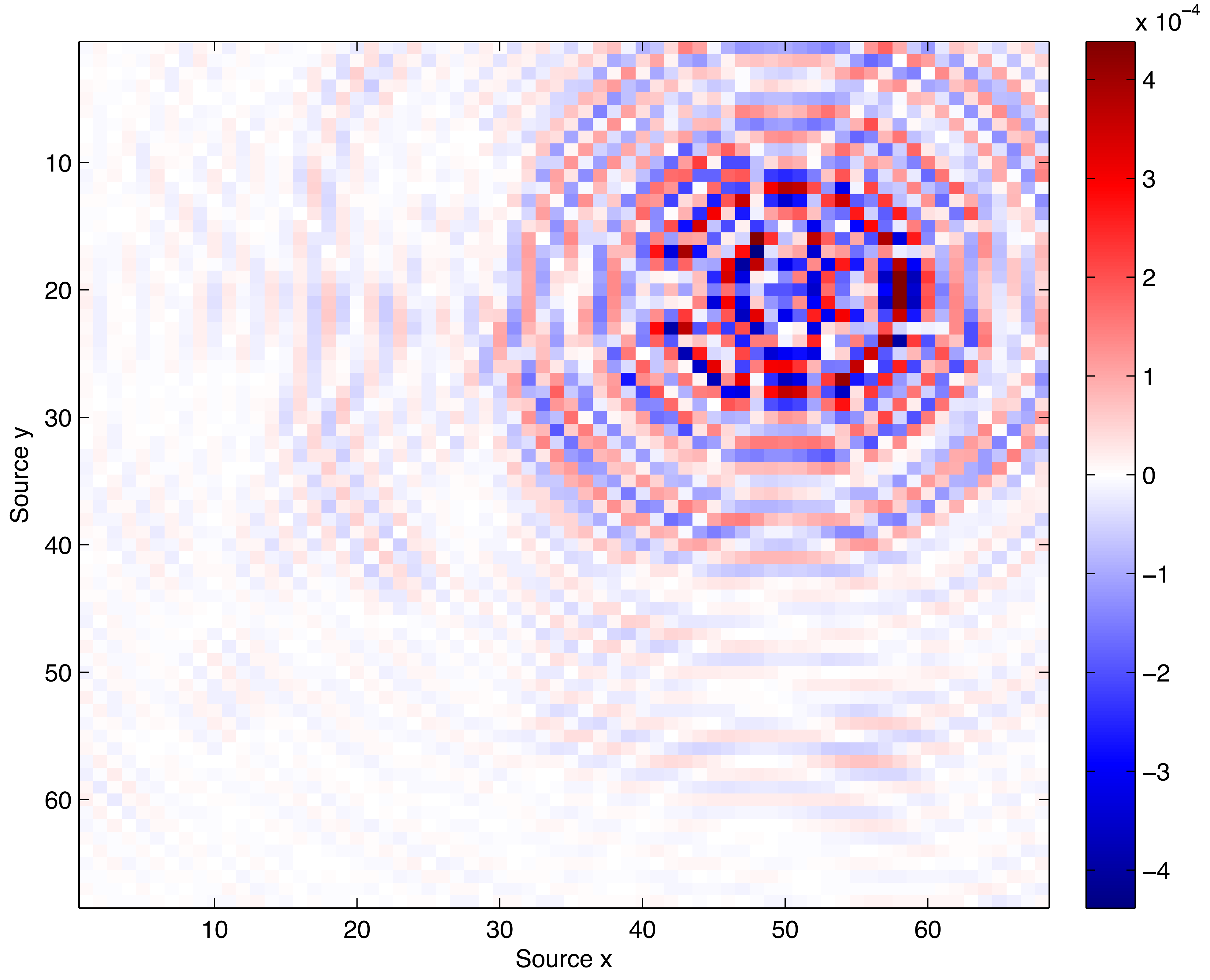
True data



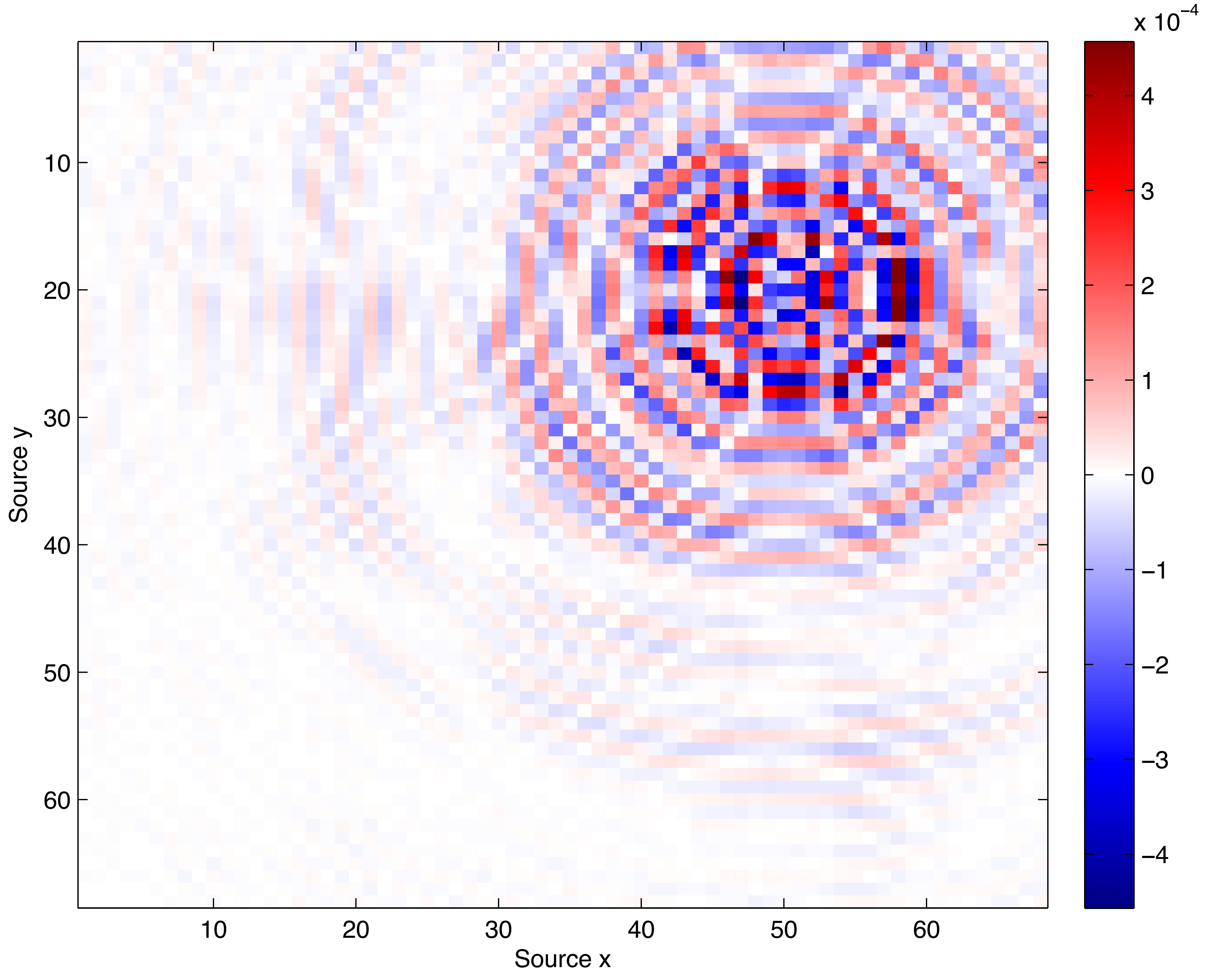
Difference

7.34 Hz - 75% missing receivers

Common receiver gather - no data initially



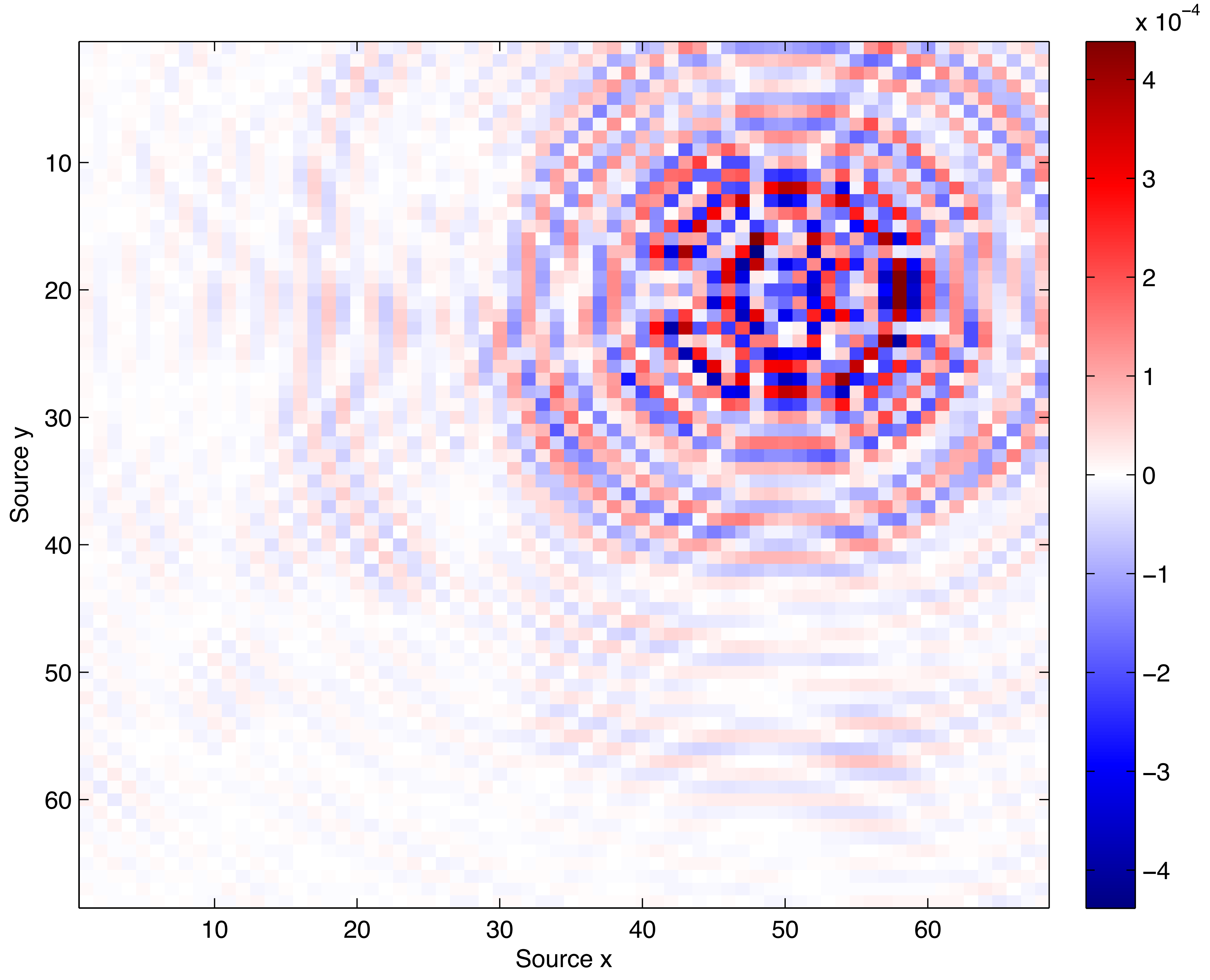
True data



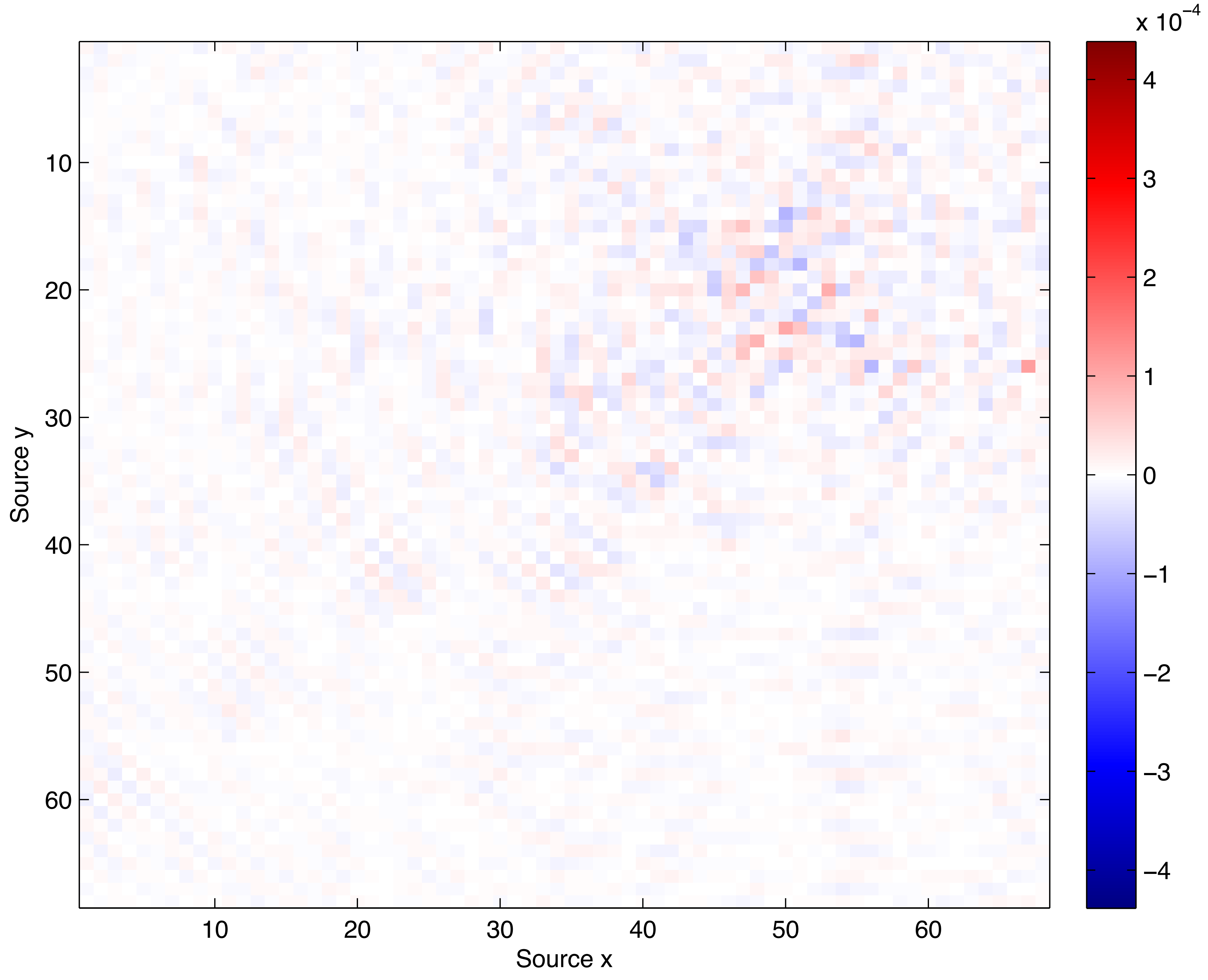
Recovered data - SNR 16.2 dB

7.34 Hz - 75% missing receivers

Common receiver gather - no data initially



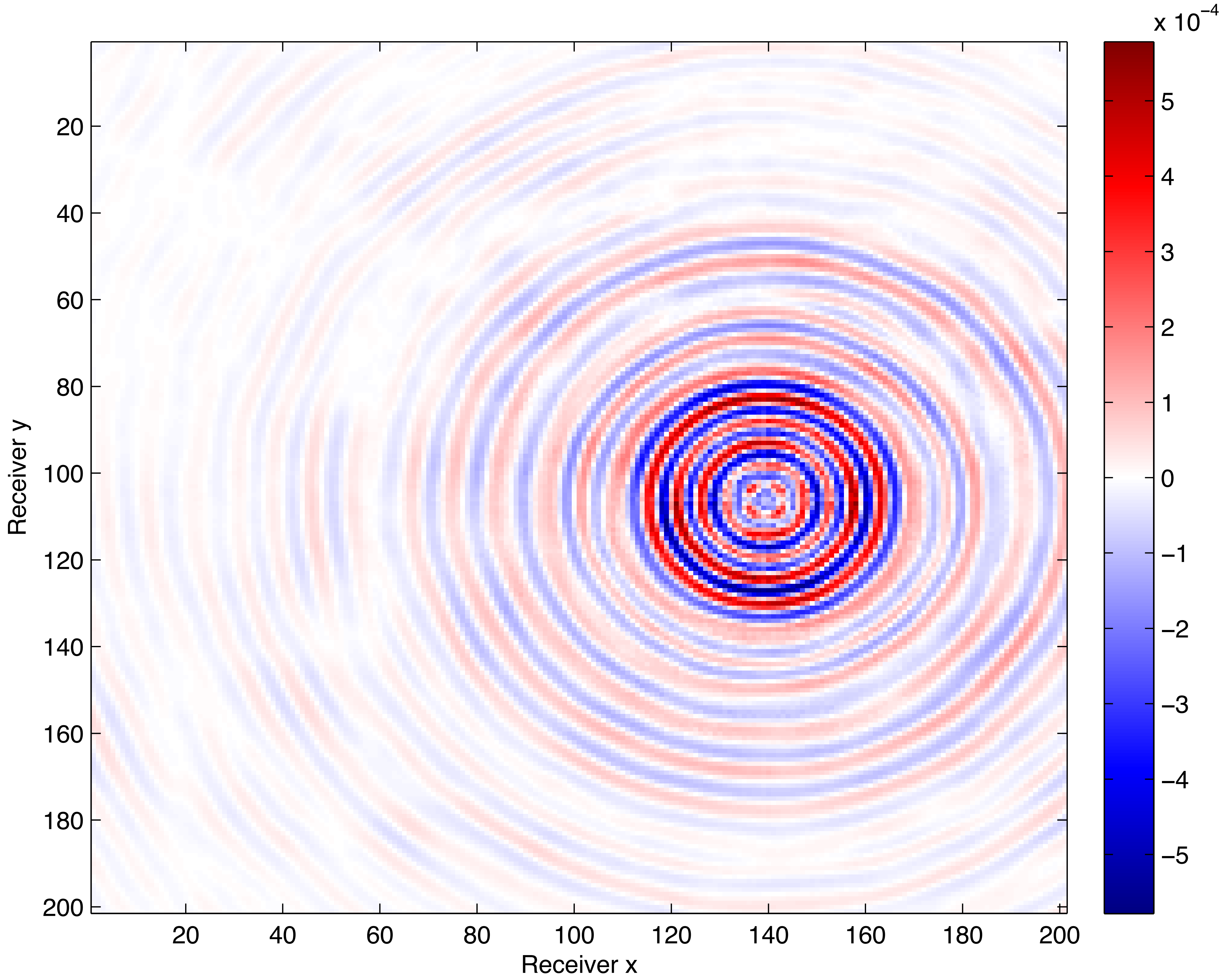
True data



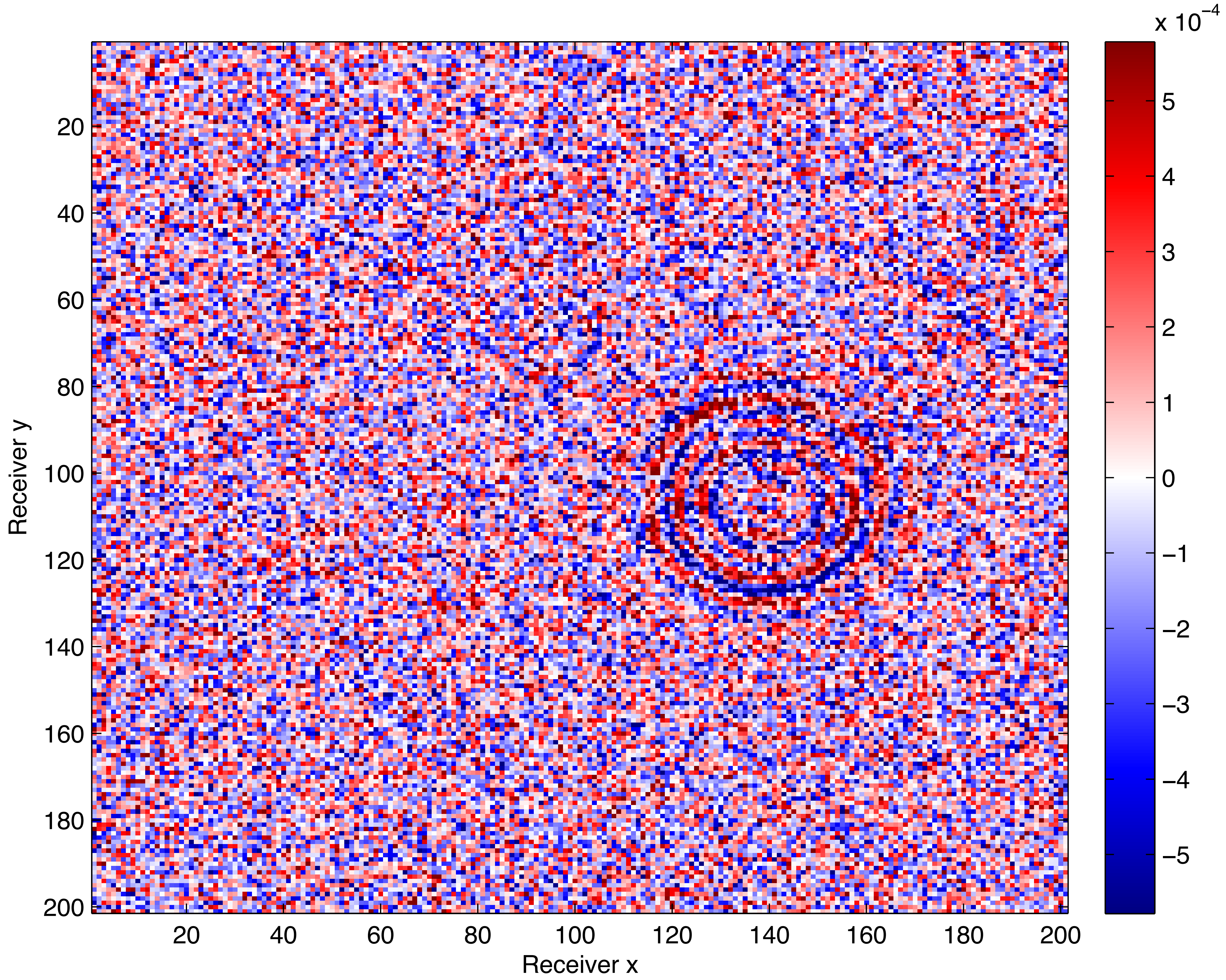
Difference

7.34 Hz - simultaneous receivers - 90% data reduction

Common source gather



True data

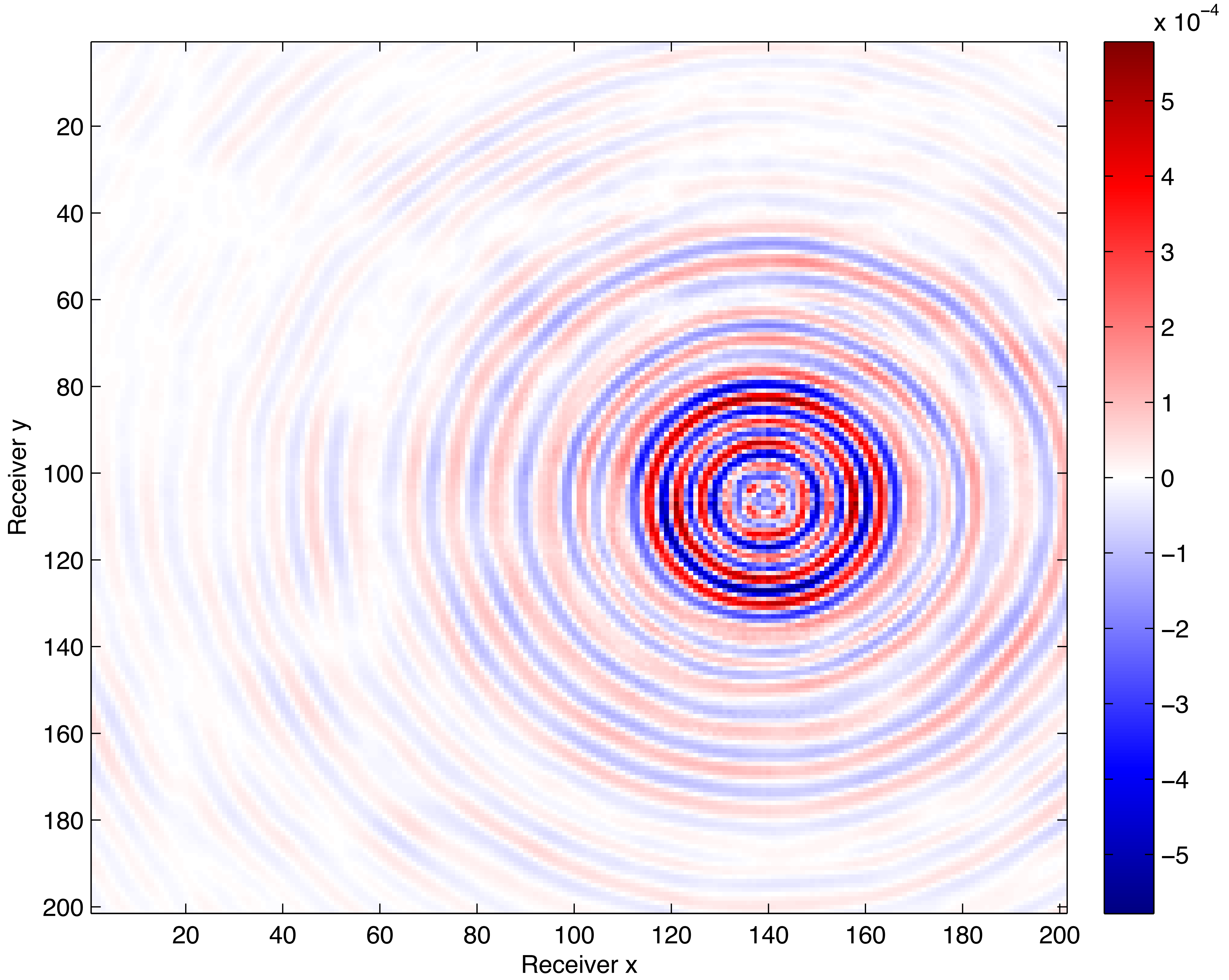


Input data - $A^T Ab$

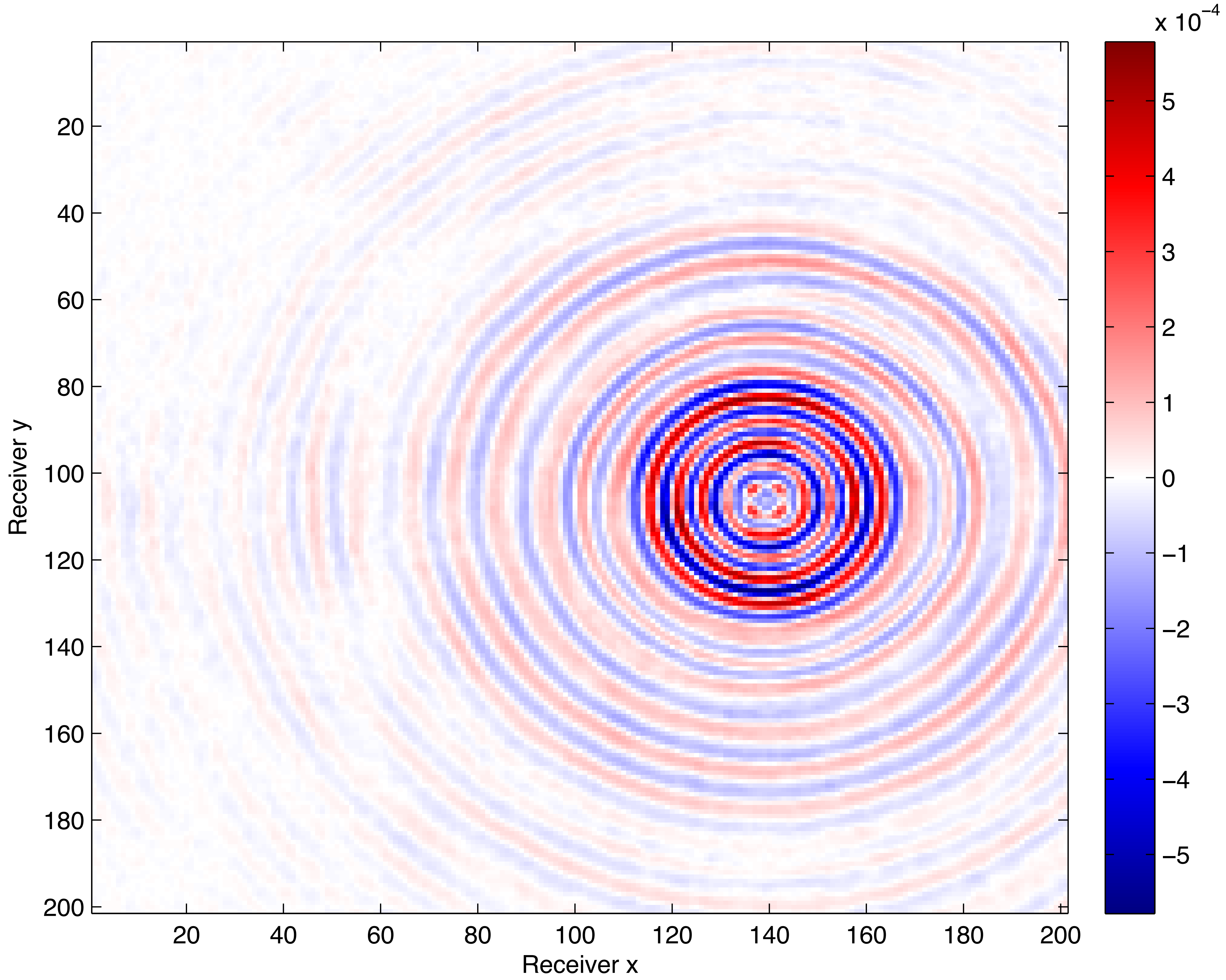
A - subsampling operator b - full data

7.34 Hz - simultaneous receivers - 90% data reduction

Common source gather



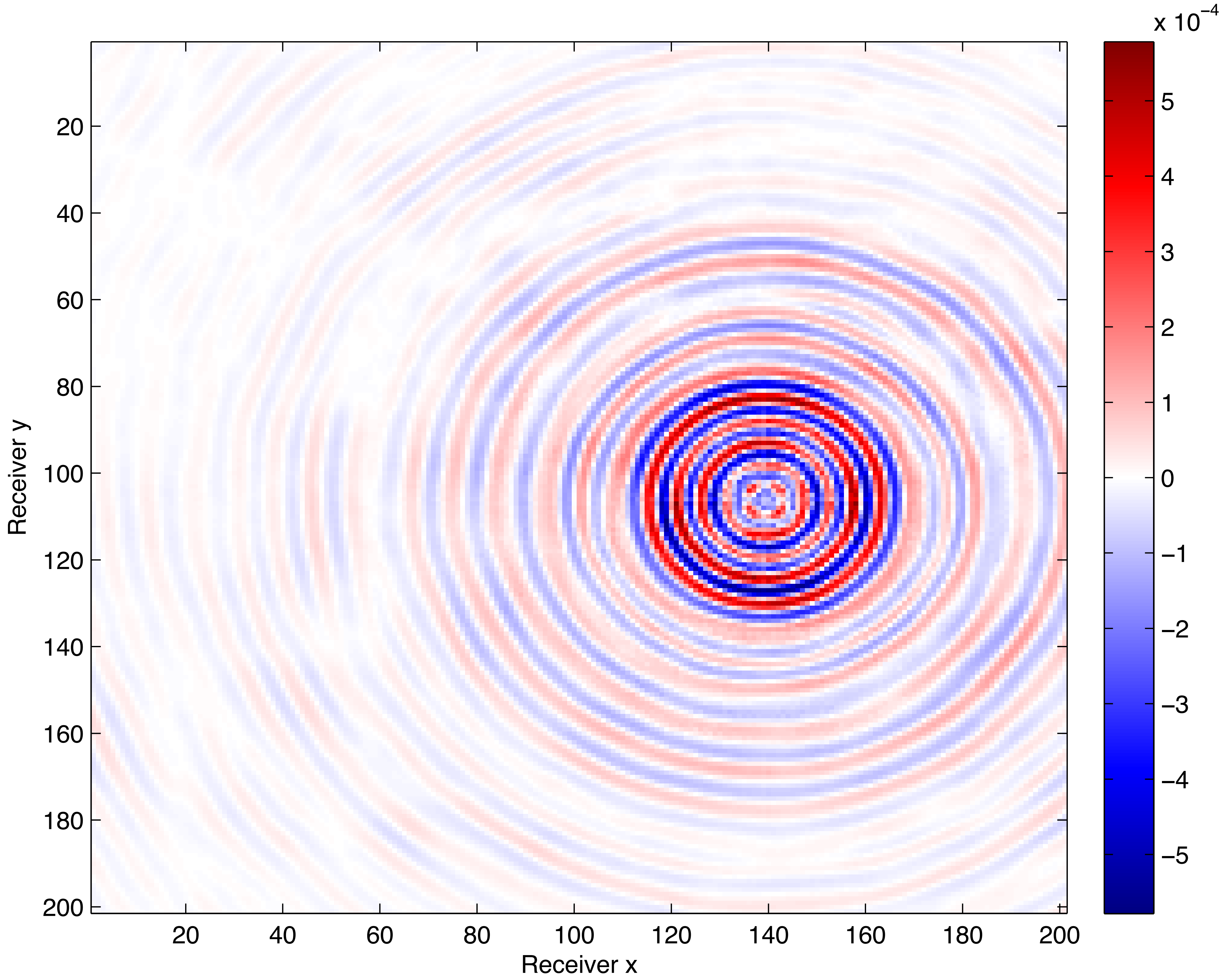
True data



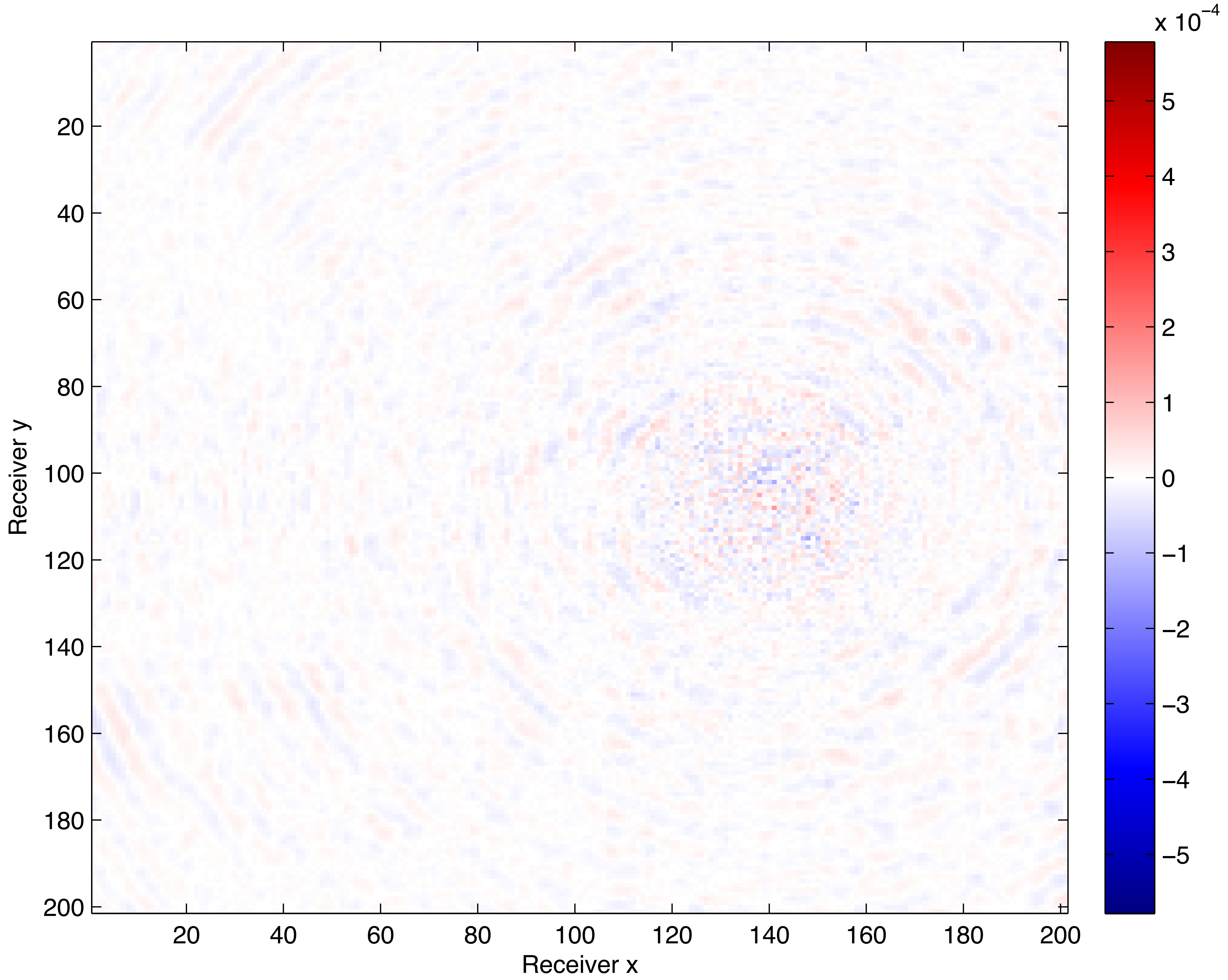
Recovered data - SNR 16.4 dB

7.34 Hz - simultaneous receivers - 90% data reduction

Common source gather



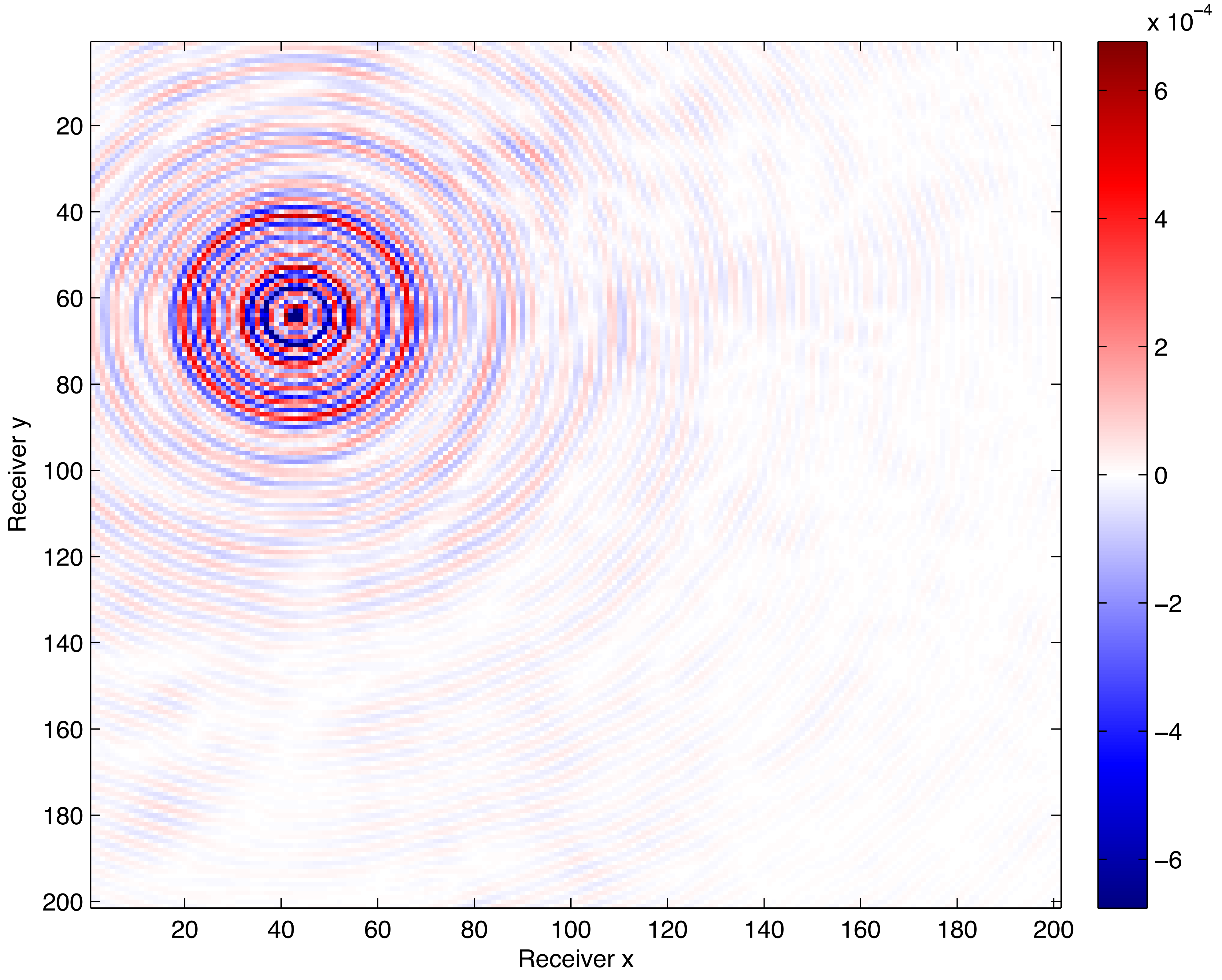
True data



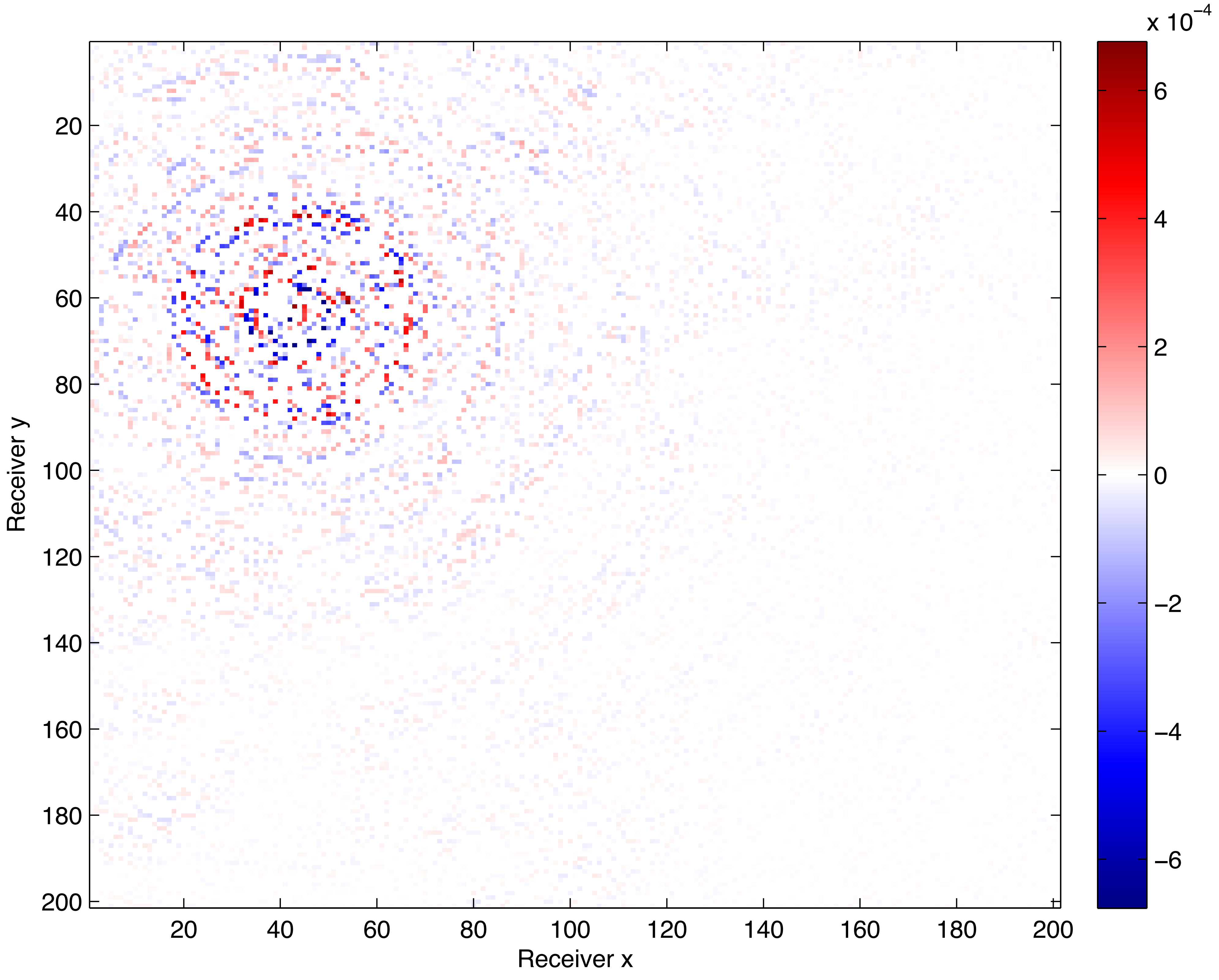
Difference

12.3 Hz - 75% missing receivers

Common source gather



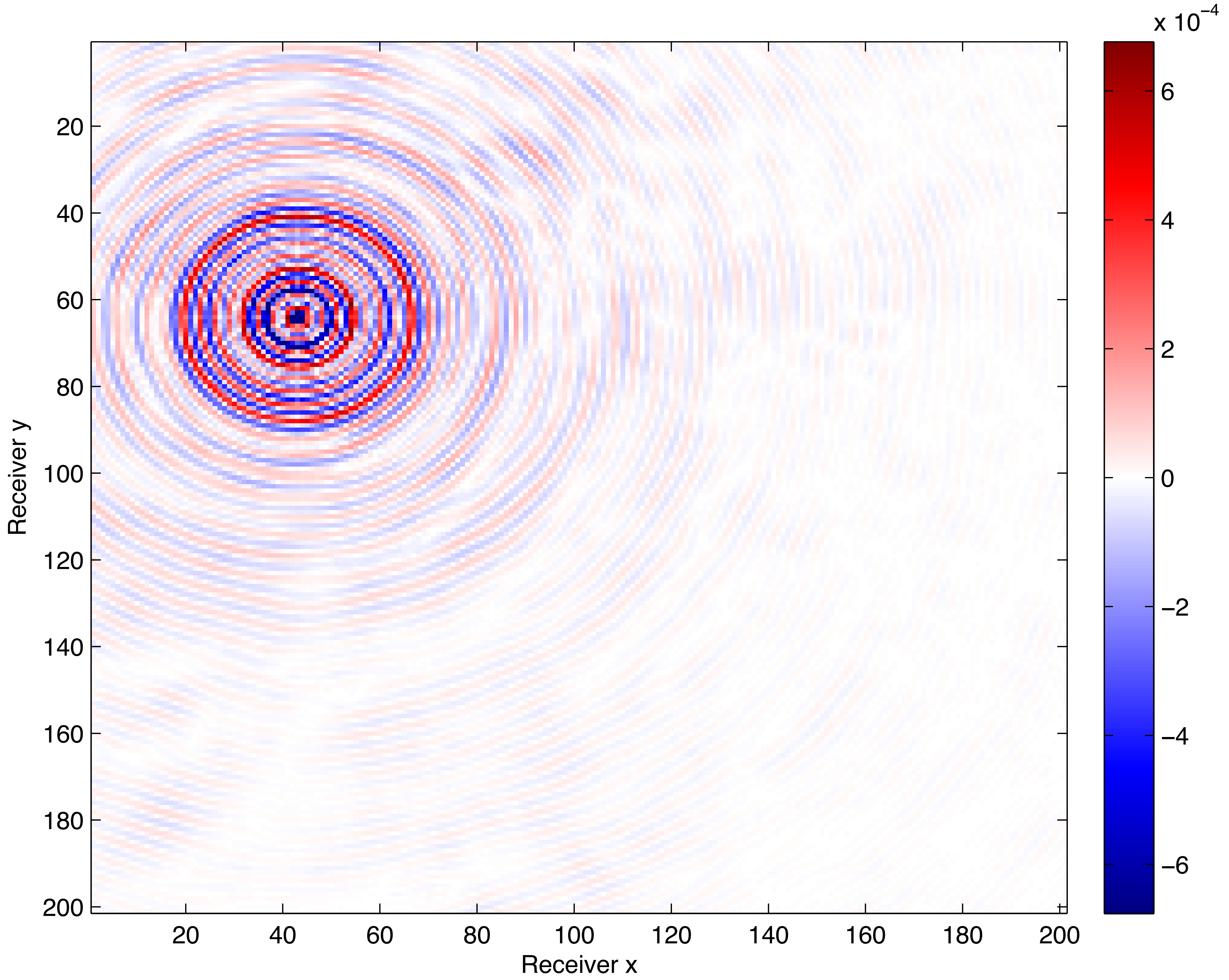
True data



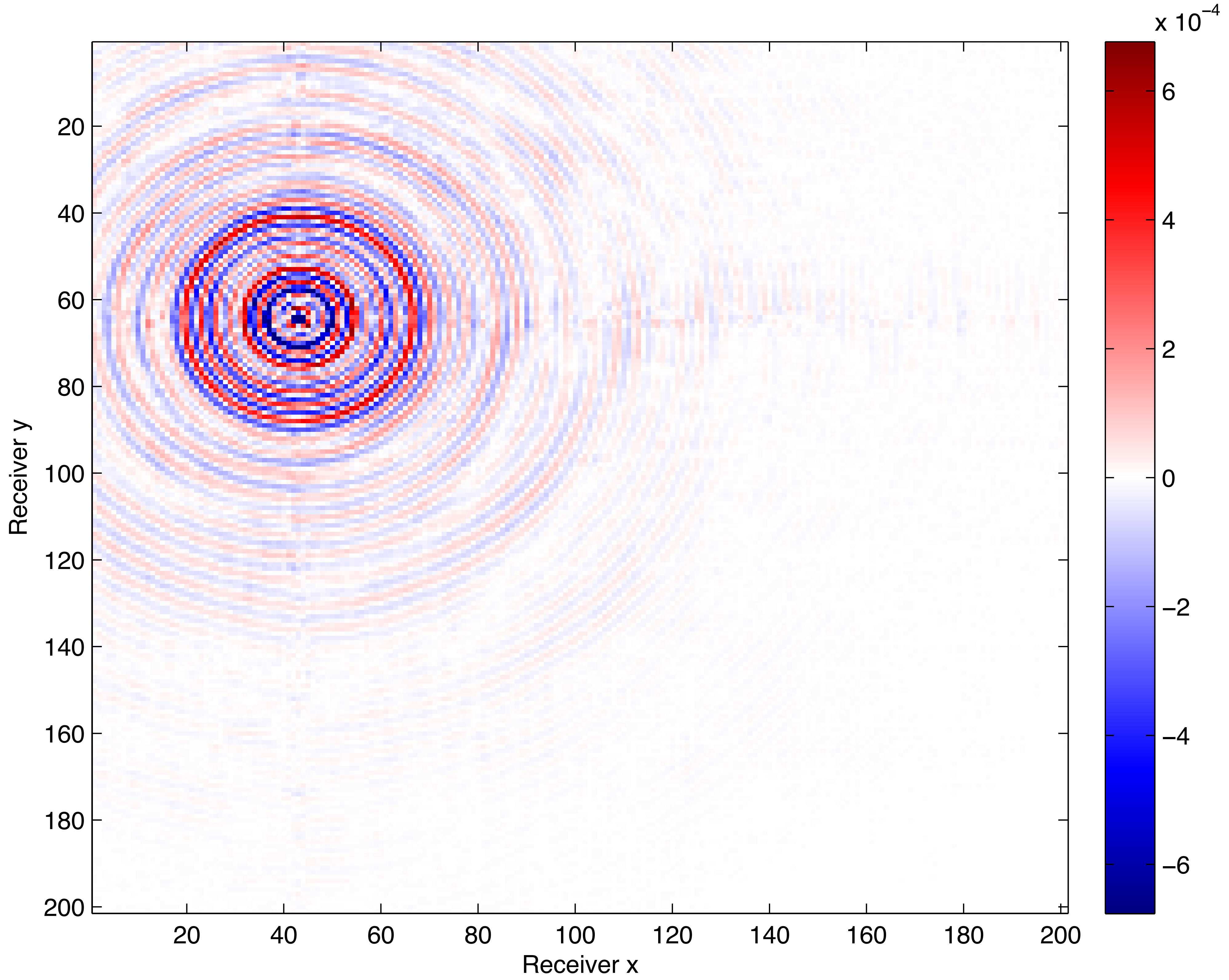
Subsampled data

12.3 Hz - 75% missing receivers

Common source gather



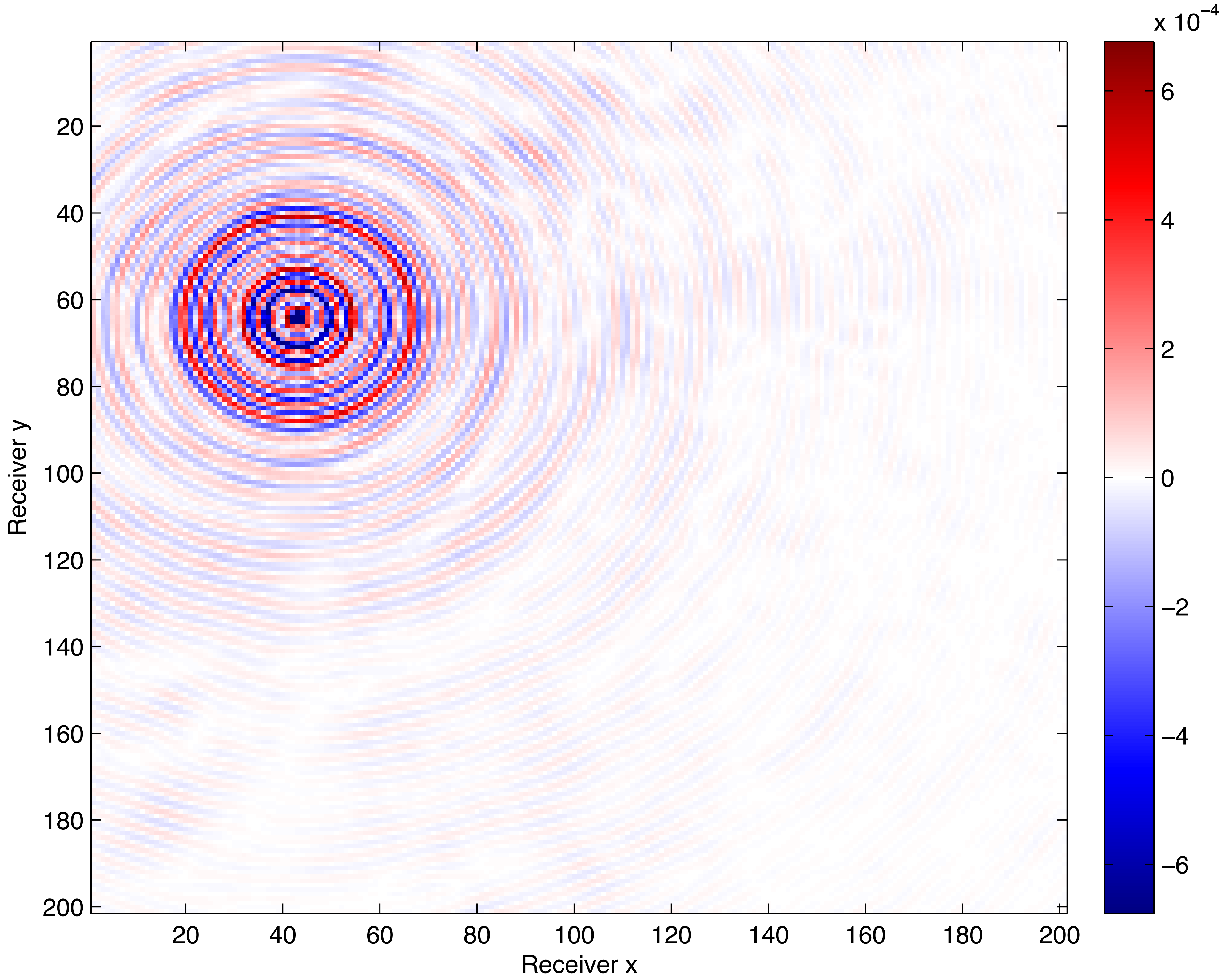
True data



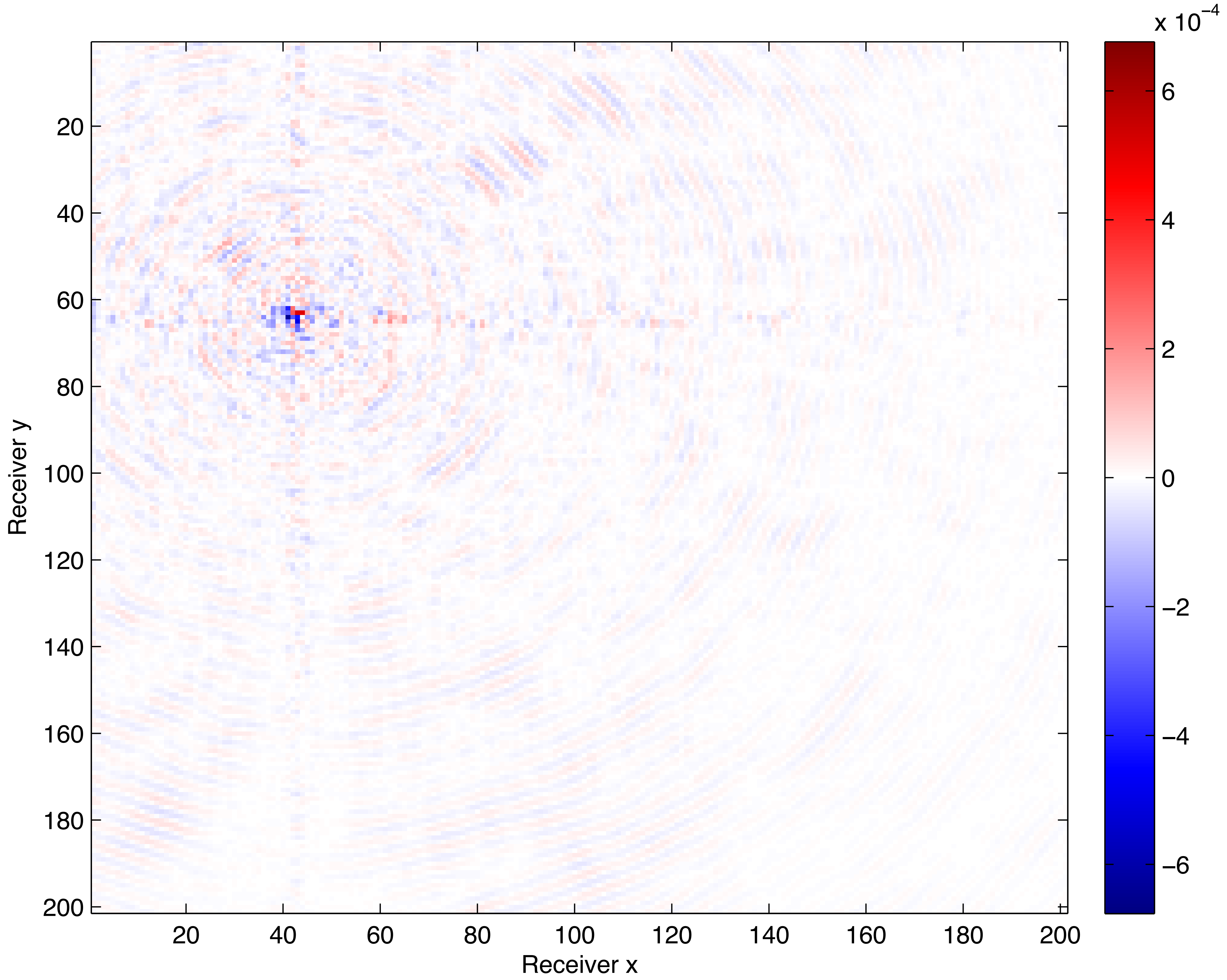
Recovered data - SNR 11.9 dB

12.3 Hz - 75% missing receivers

Common source gather



True data



Difference

Conclusion

3D seismic data has an underlying structure that we can exploit for interpolation (Hierarchical Tucker format)

Different schemes for organizing data - important for recovery

Conclusion

We can interpolate HT tensors with missing entries using the Riemannian manifold structure of the HT format

Achieve good results from largely subsampled data (75% missing receivers, 90% reduction in simultaneous receivers)

Can use this method to create full volumes from subsampled data

- Migration, multiple removal, etc.

What next?

Simulating waves in random media

$m(x,y)$ - earth medium as a function of space, x, random variable, y

$u(x,y)$ - wavefield depending on x, y

$$\left(\frac{1}{m(x,y)^2}\frac{\partial^2}{\partial t^2} + \Delta\right) u(x,y) = f(x)$$

SLIM

Why are we interested?

Full Waveform Inversion with *uncertainty quantification*

- we are interested in computing

$$\tilde{m}(x) = \mathbb{E}_y(m(x,y)|d)$$

$$\begin{aligned} v(x) &= \mathrm{var}(m|d) \\ &= \mathbb{E}_y((m(x,y) - \tilde{m}(x))^2|d) \end{aligned}$$

Challenges

$m(x, y)$

- what's an appropriate random model?

Typically $u(x, y_1, y_2, \ldots, y_N)$ for some finite random variables y_i

- naive approach - solve wave equation for *each* fixed $y_1, y_2, \ldots, y_N$
  - curse of dimensionality
- exploit low-rank hierarchical tucker structure?

Random models

Write $m(x, y)$ as

$$m(x, y) = \hat{m}(x) + \sum_{i=1}^{N} y_i \psi_i(x)$$

$\hat{m}(x)$ is the mean, $\psi_i(x)$ are basis functions, y_i are random variables

Random layered models

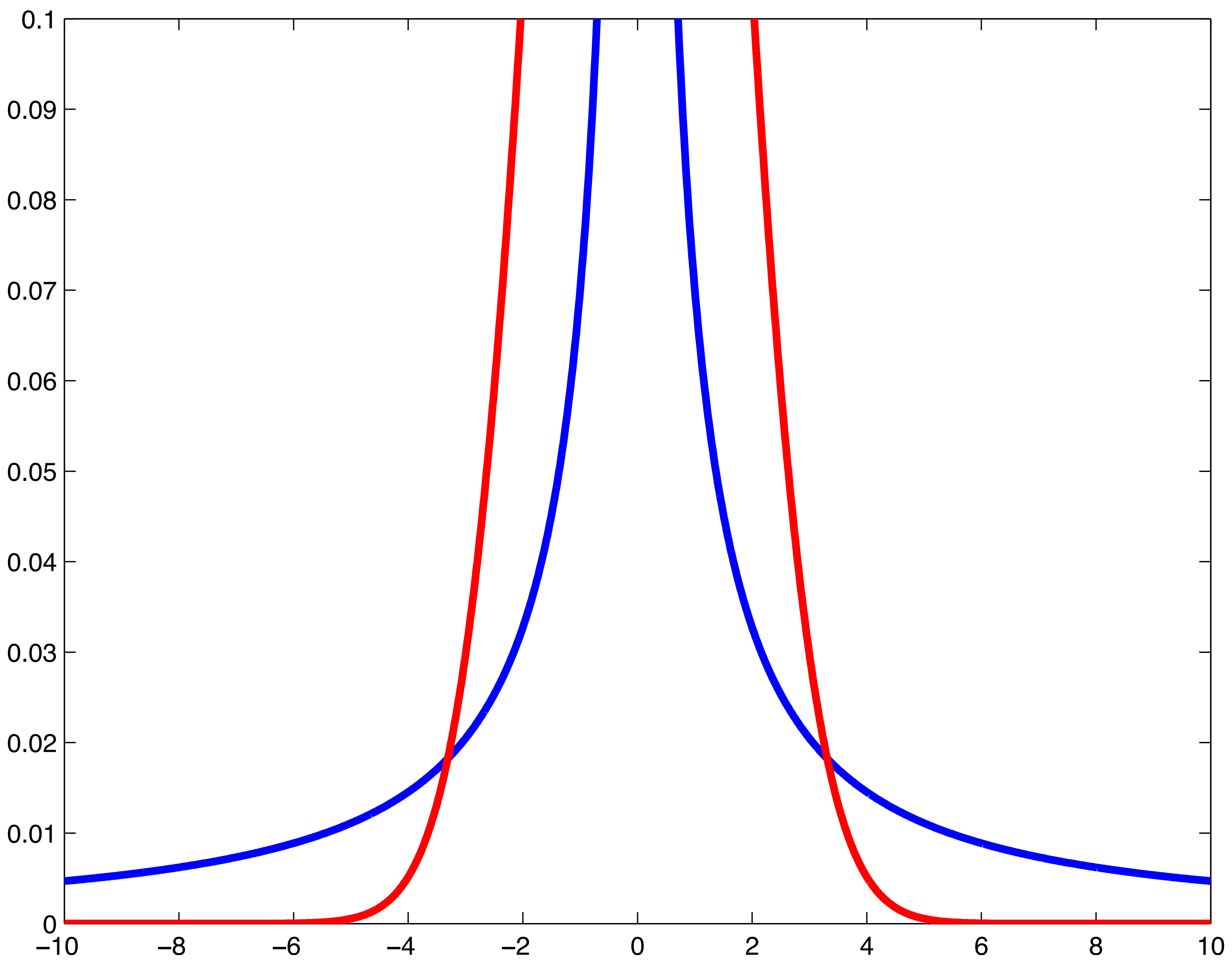
Simple model

- $\psi_i(x)$ - Haar wavelet basis
- y_i - drawn from an α-stable distribution

α-stable distributions

Probability density

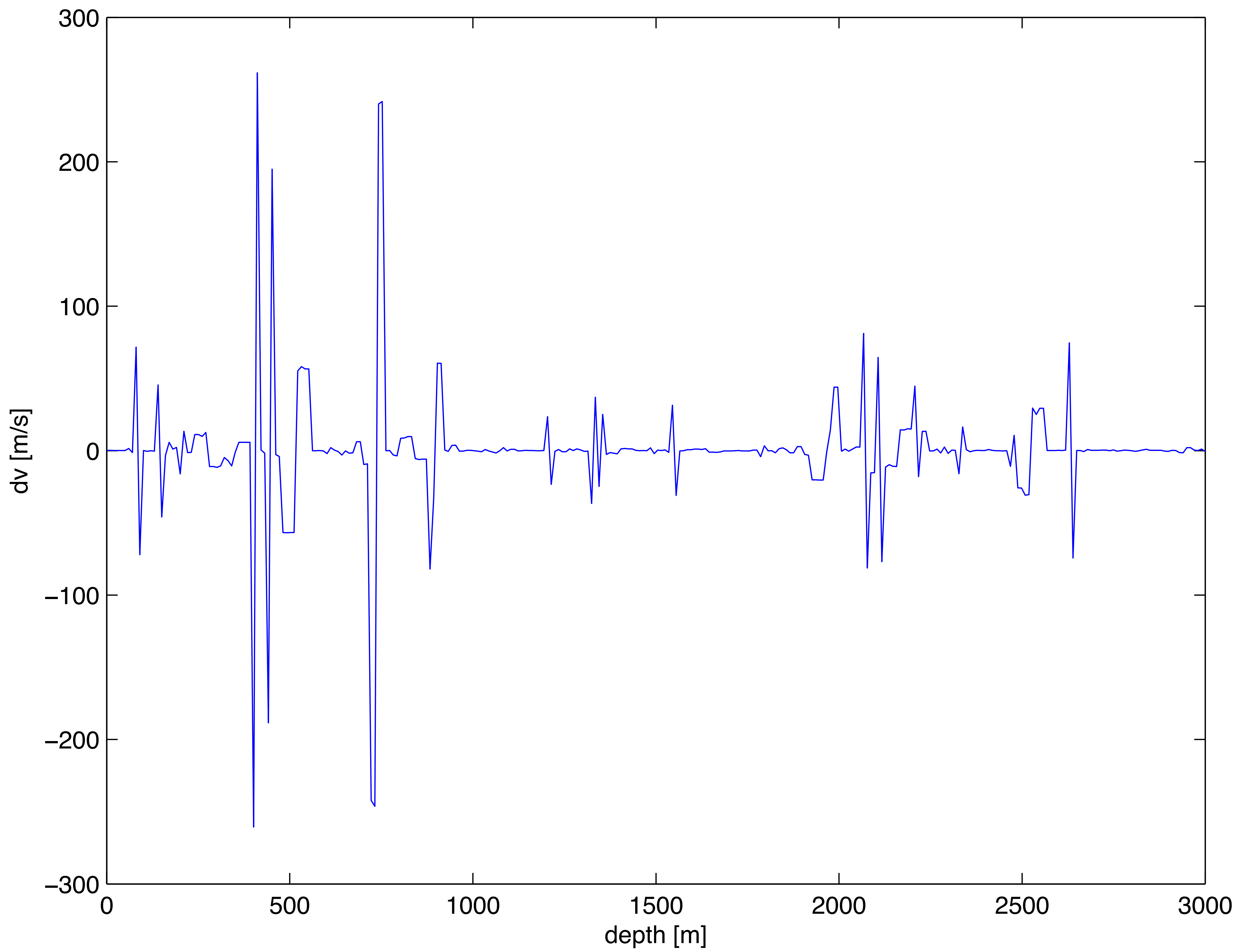
- Blue - α-stable
- Red - Gaussian

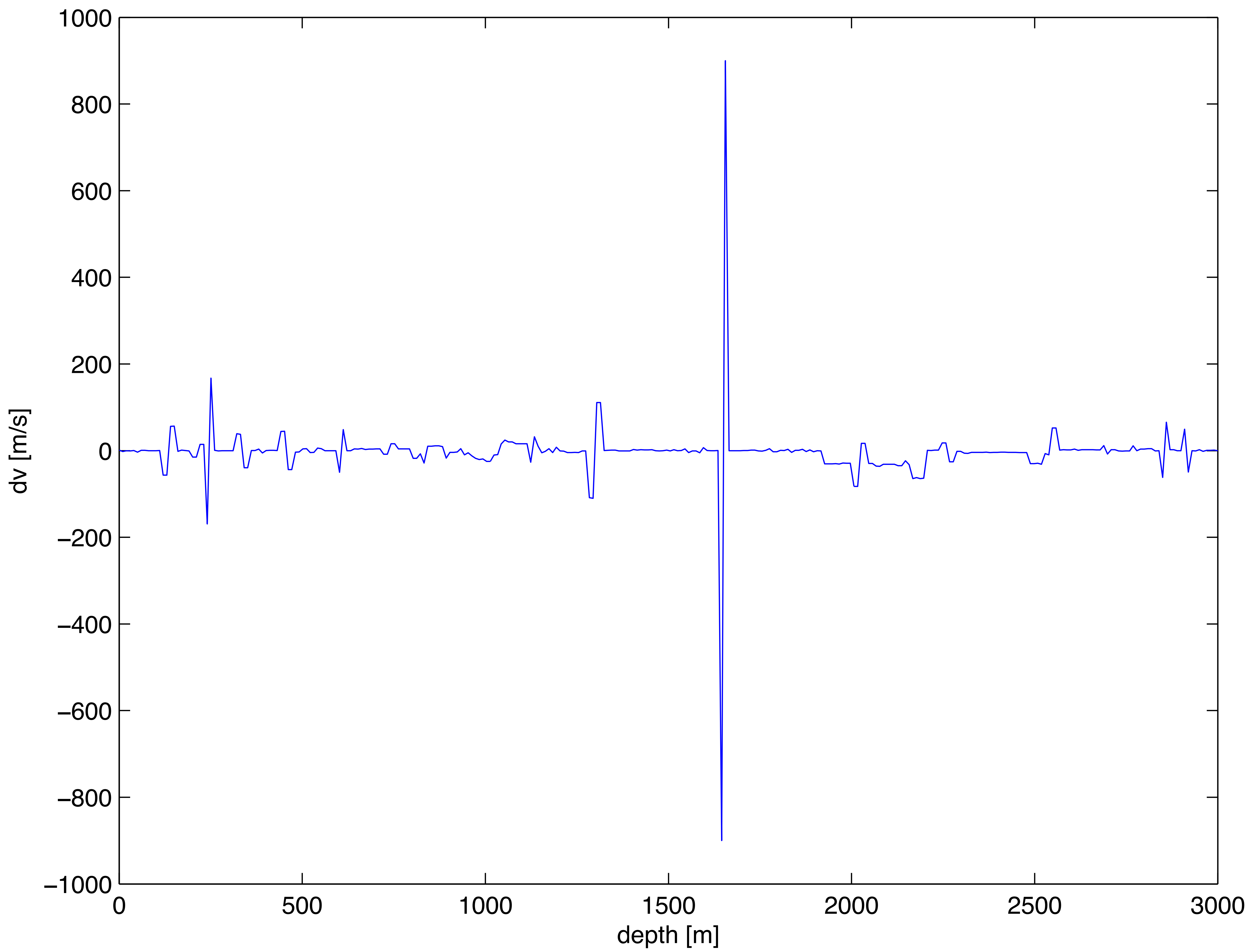


α-stable distributions

- allow *sparse* noise
- allow *large outliers*

Random model perturbations





Future work

Scale-dependent probability - wavelets

- if a coefficient is zero at scale j, it should probably be zero at scale j+1

2D random models

- curvelets + sparse distribution

Efficient implementation for computing random wavefields

Fullwaveform inversion with uncertainty quantification

Acknowledgements

Thank you for your attention

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, BGP, BP, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, WesternGeco, Statoil, and Woodside.