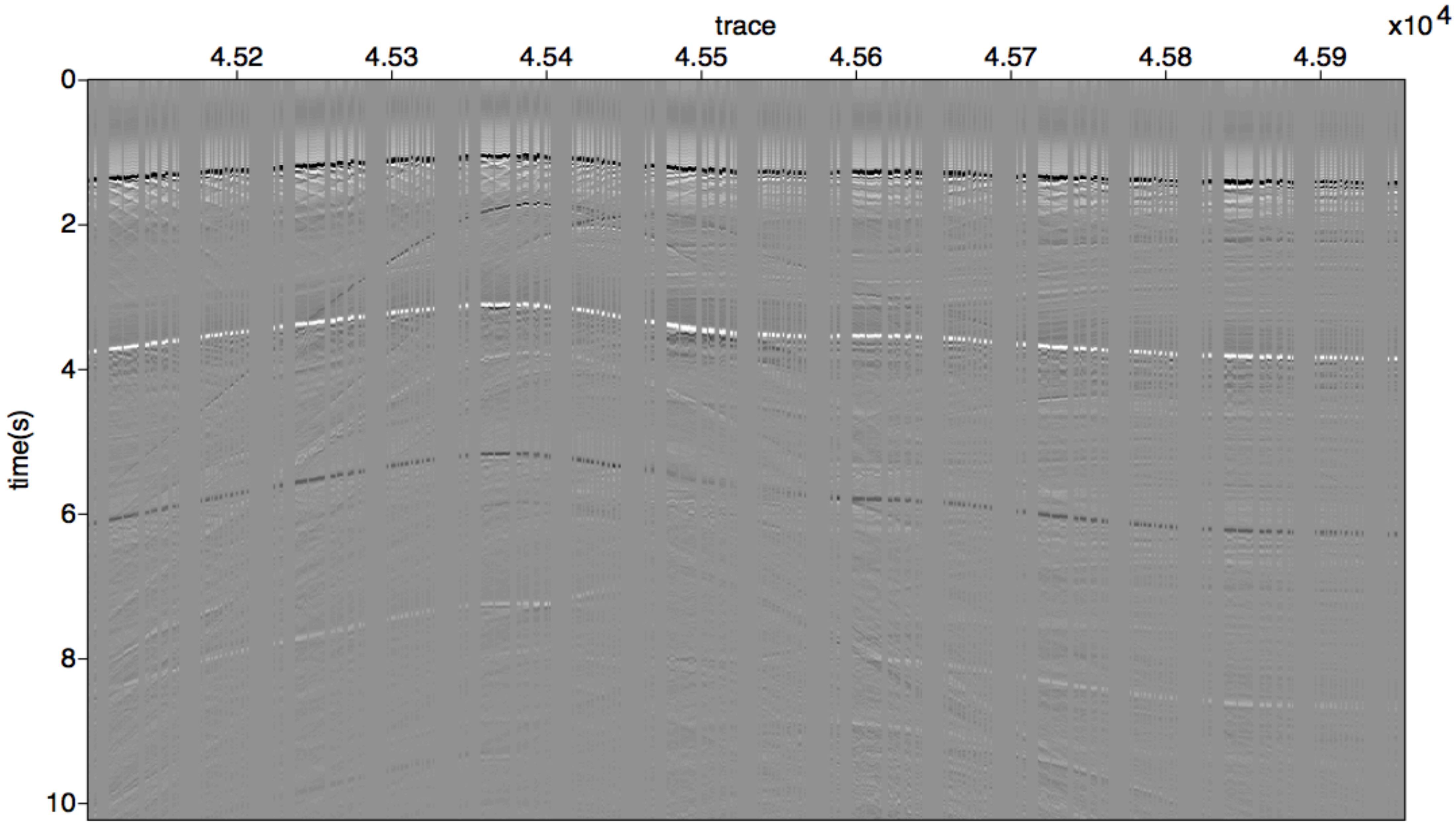
Seismic Data Interpolation using SVD-free Low Rank Optimization

Rajiv Kumar, Sasha Aravkin, Hassan Mansour, and Felix J. Herrmann

SLIM
University of British Columbia

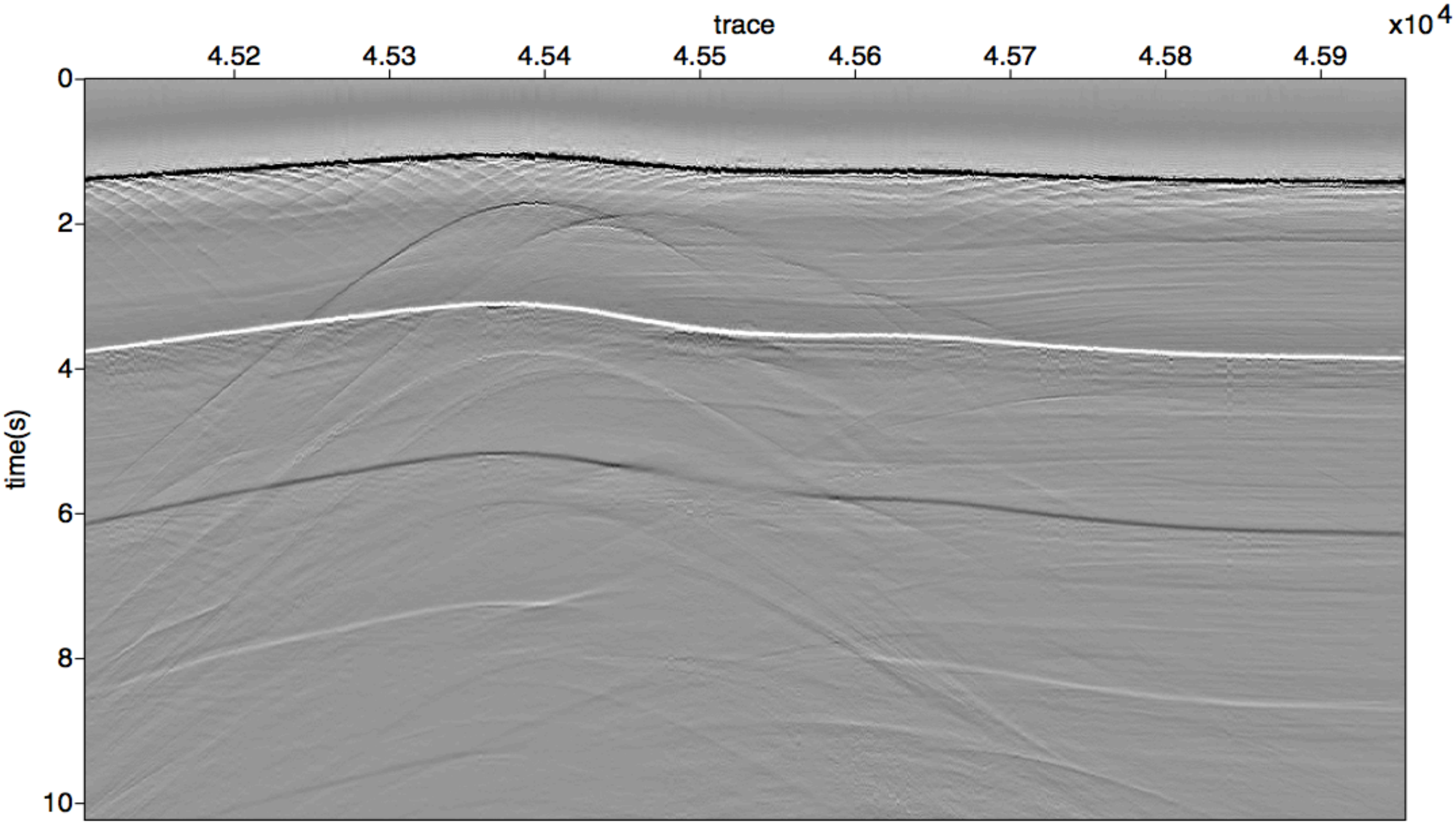
Motivation



Input data

SLIM

Motivation



Interpolation with 2D Curvelet

- ***is sparsity the only inherent structure in seismic??***

Outline

- rank structure in seismic
- matrix factorization
- regularized matrix factorization

Rank Structure in Seismic

SLIM

Setup

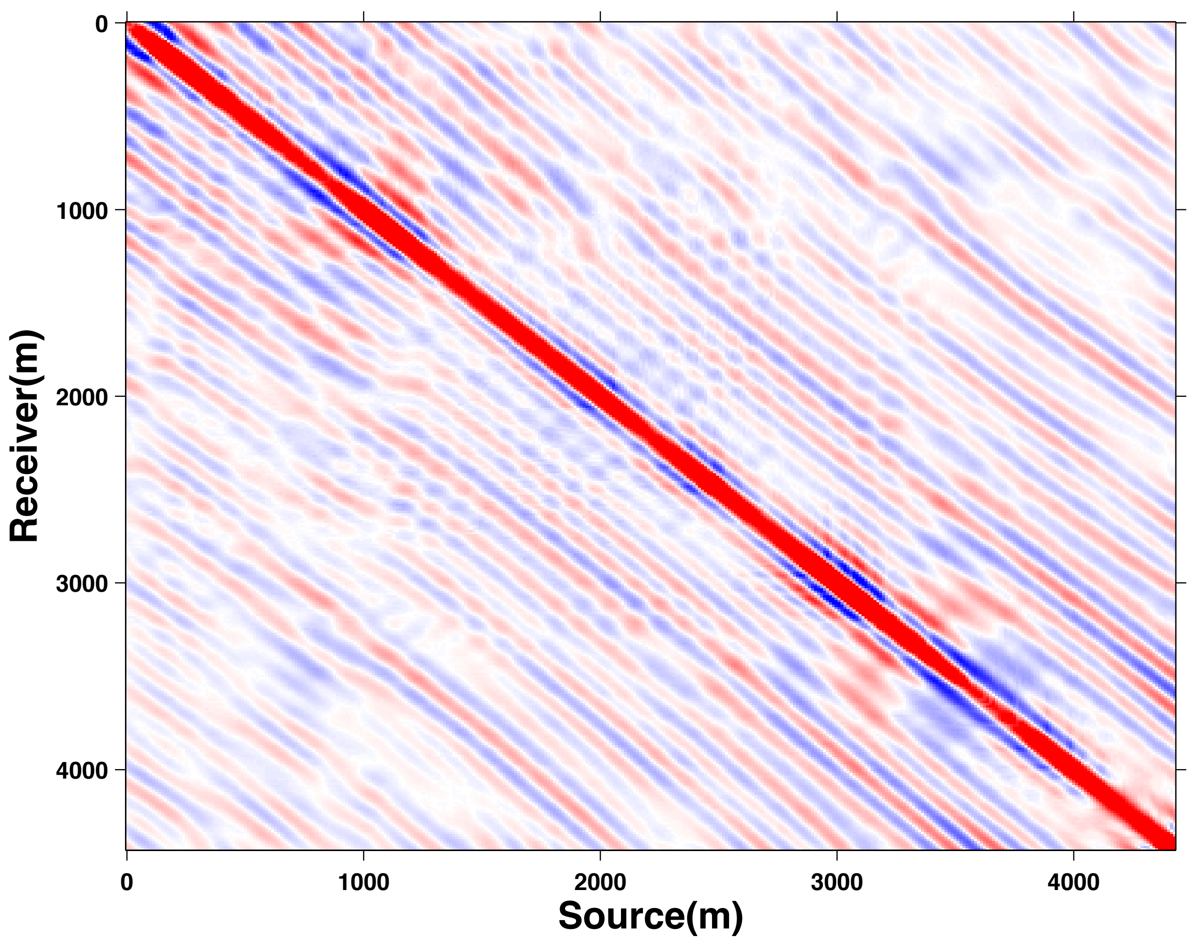
- let $\mathbf{D} \in \mathbb{R}^{n_r \times n_s}$ is spatial acquired data on regular grid
- monochromatic frequency

$$\mathbf{D}_\omega = \begin{pmatrix} D_{1,1} & D_{1,2} & \cdots & D_{1,n_s} \\ D_{2,1} & D_{2,2} & \cdots & D_{2,n_s} \\ \vdots & \vdots & \ddots & \vdots \\ D_{n_r,1} & D_{n_r,2} & \cdots & D_{n_r,n_s} \end{pmatrix}$$

- n_r, n_s represent receiver # and source #

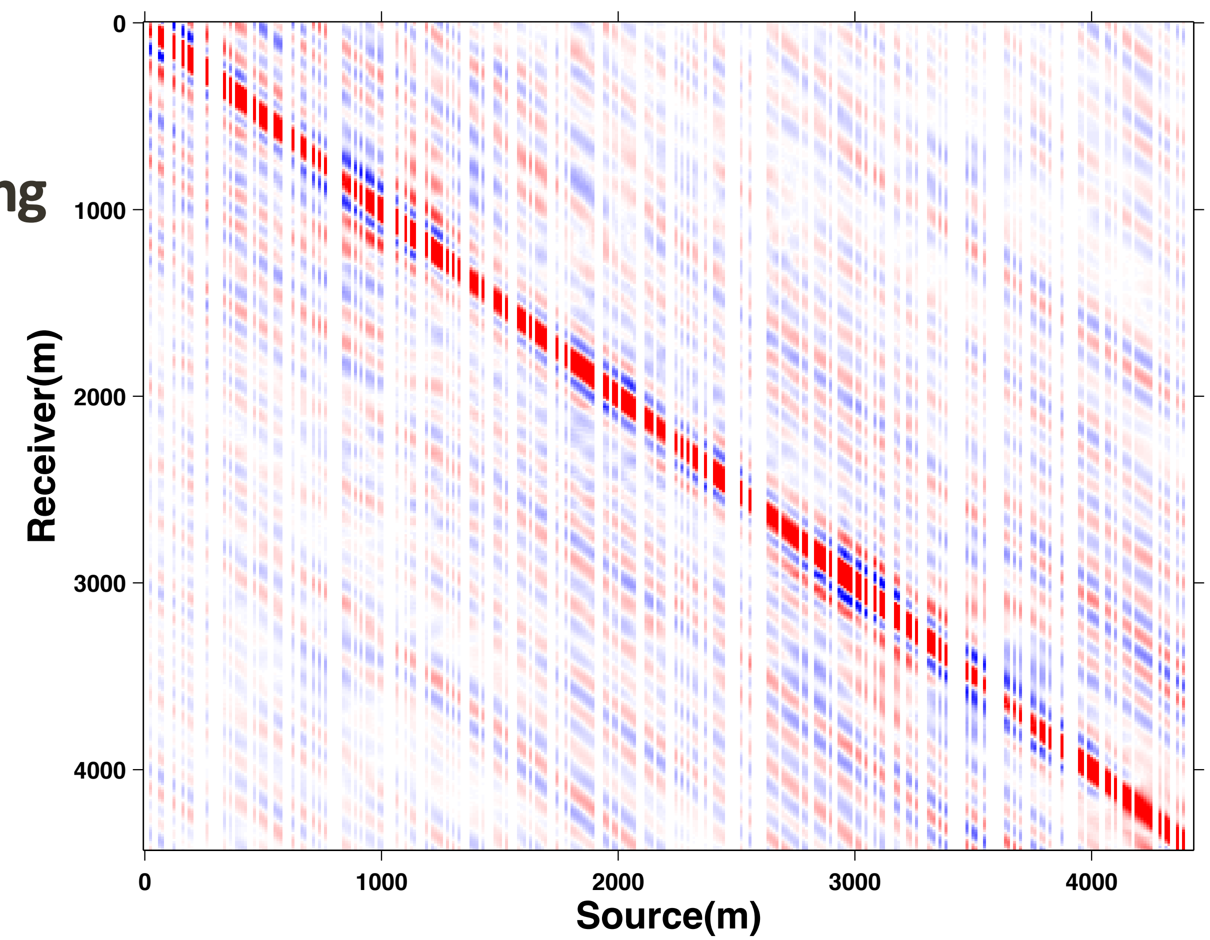
2D Acquisition

Acquisition impediments
[Frequency -12 Hz]



Regular Sampled data

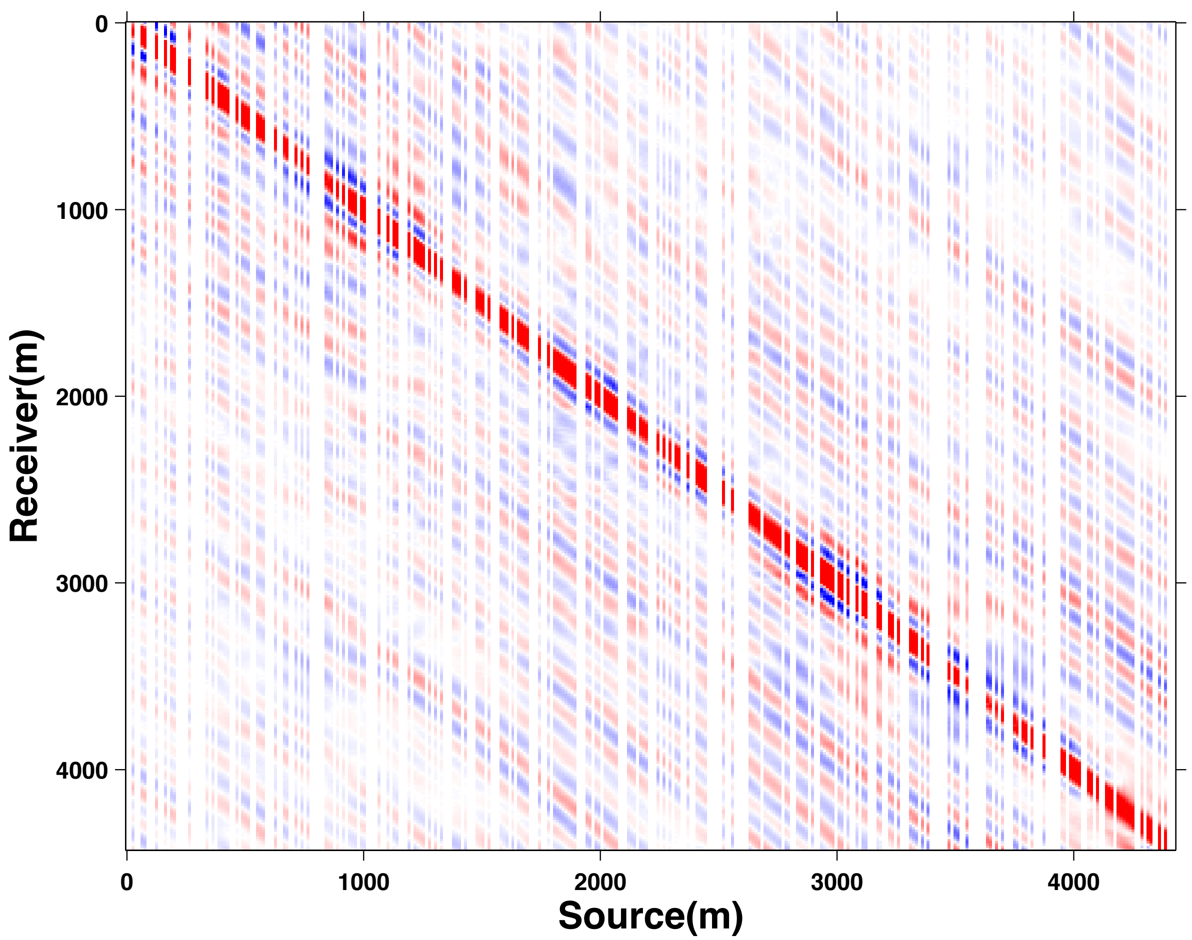
50% missing shots



Irregular Sampled data
[Missing columns do not increase rank]

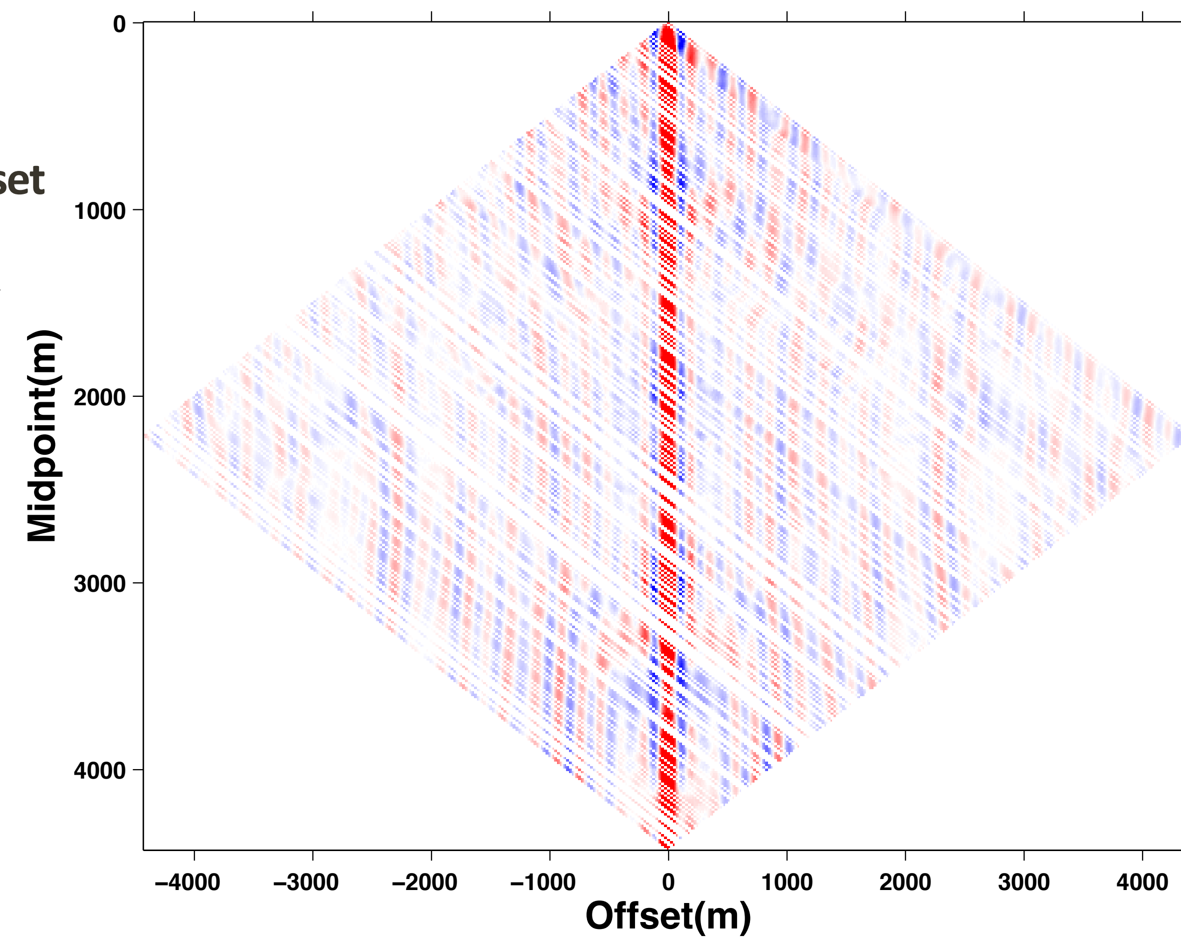
Missing data scenario
[Frequency -12 Hz]

Missing column do not increase rank



Midpoint-Offset transform

Missing column DOES increase rank

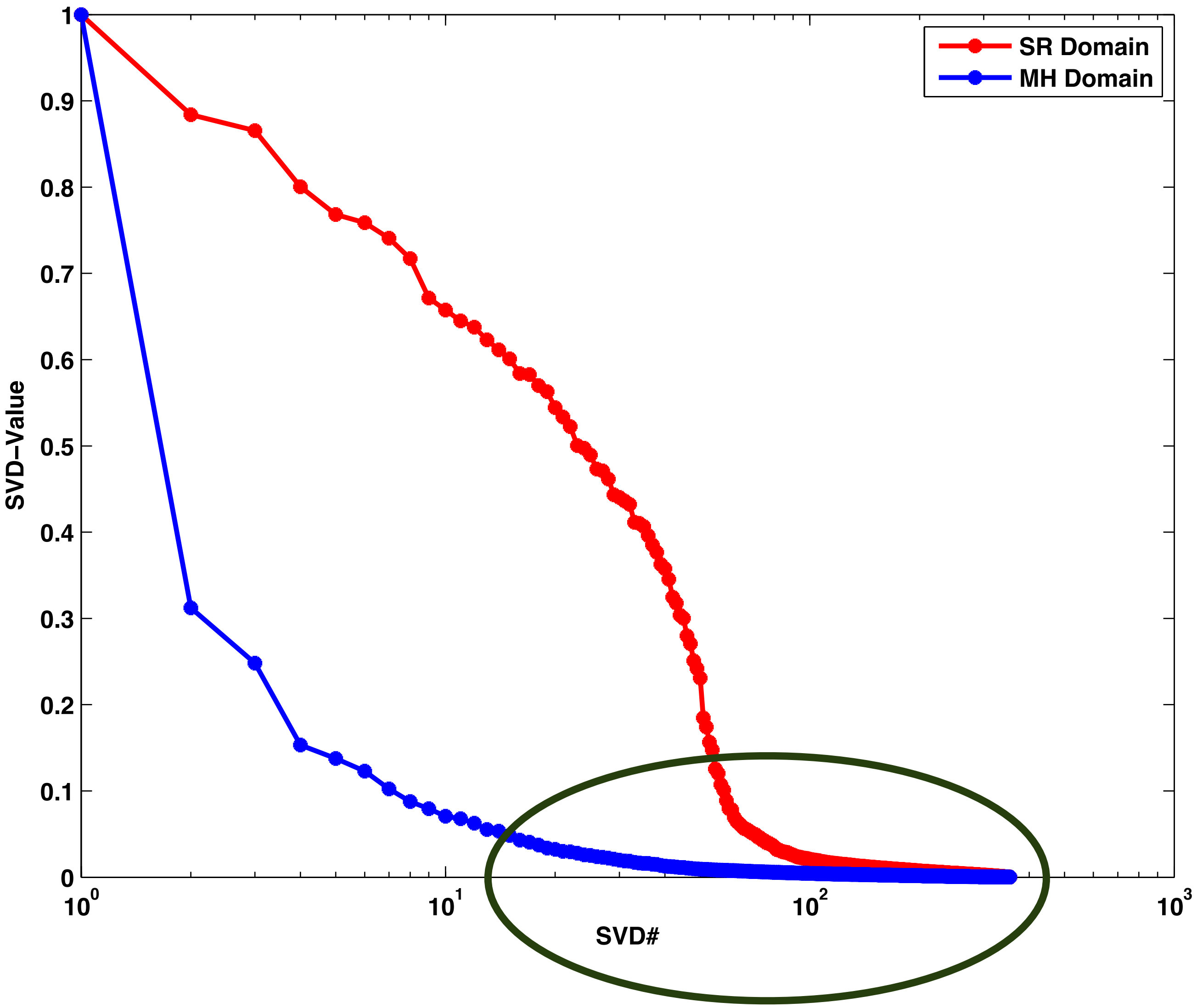


- **Midpoint -** $0.5(s+r)$
- **Offset -** $(s-r)$

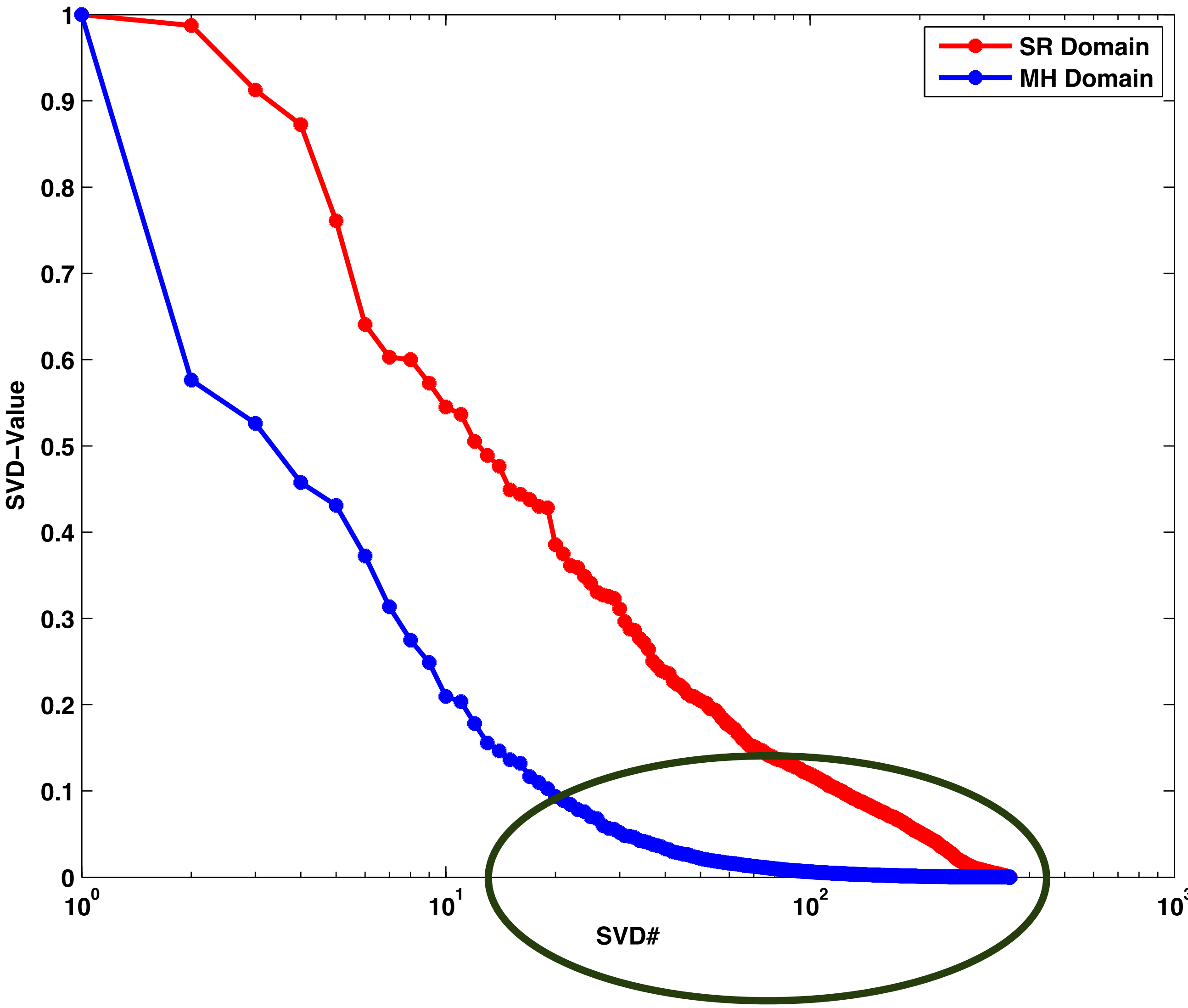
where s **is source position and** r **is receiver position**

Regular sampled data
[Singular value decay]

Freq : 12 Hz

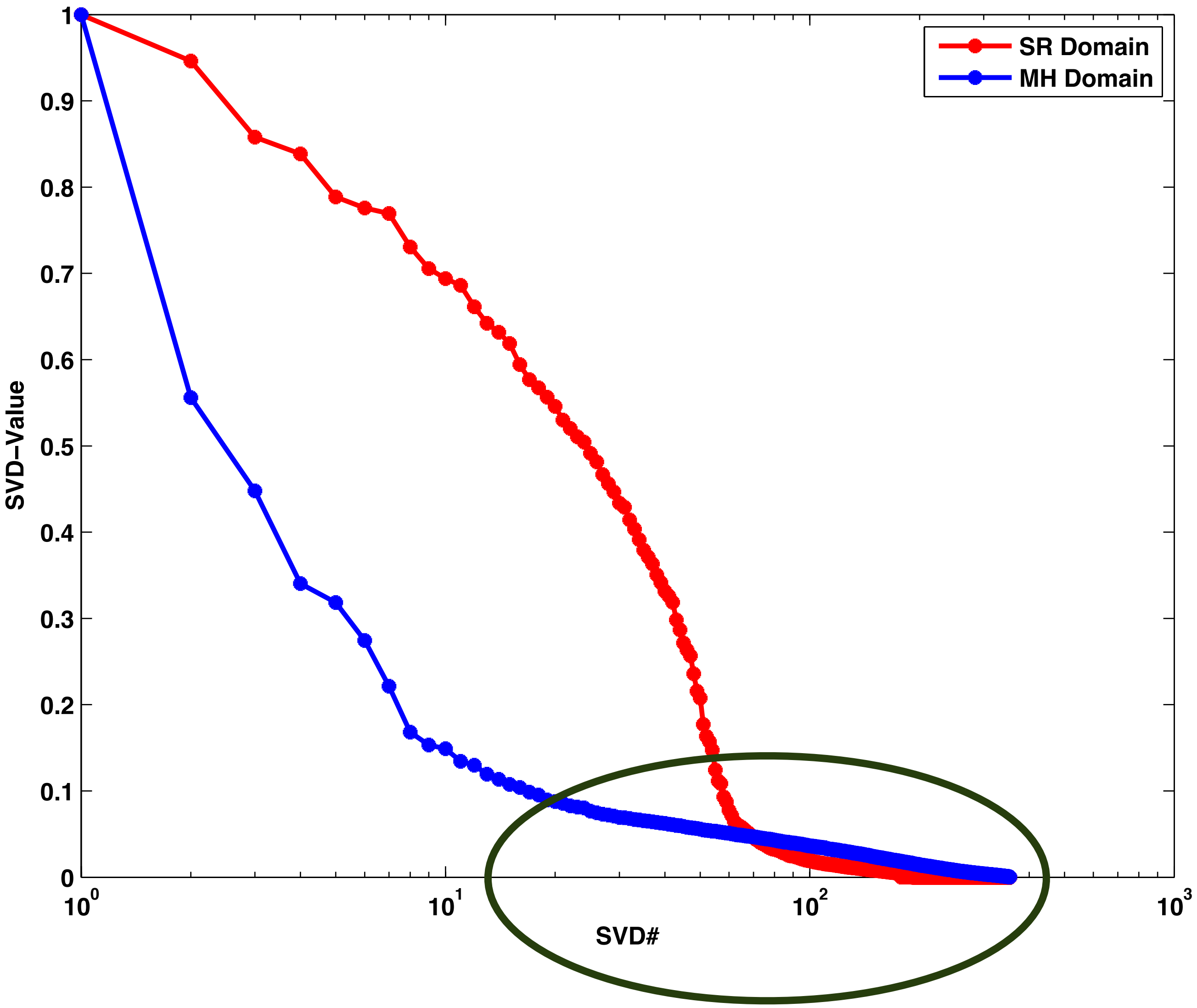


Freq : 60 Hz

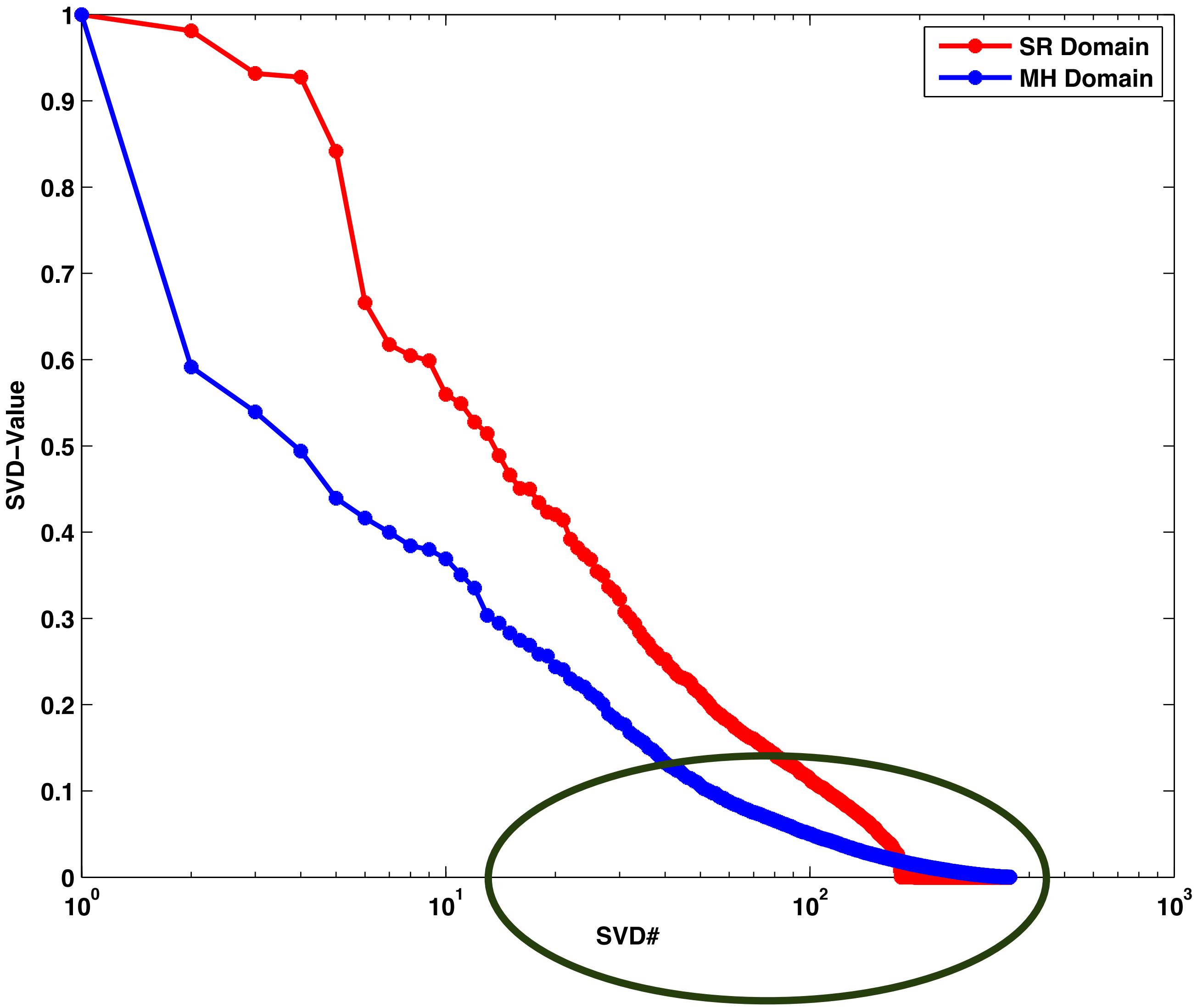


50 % Missing data
[Singular value decay]

Freq : 12 Hz

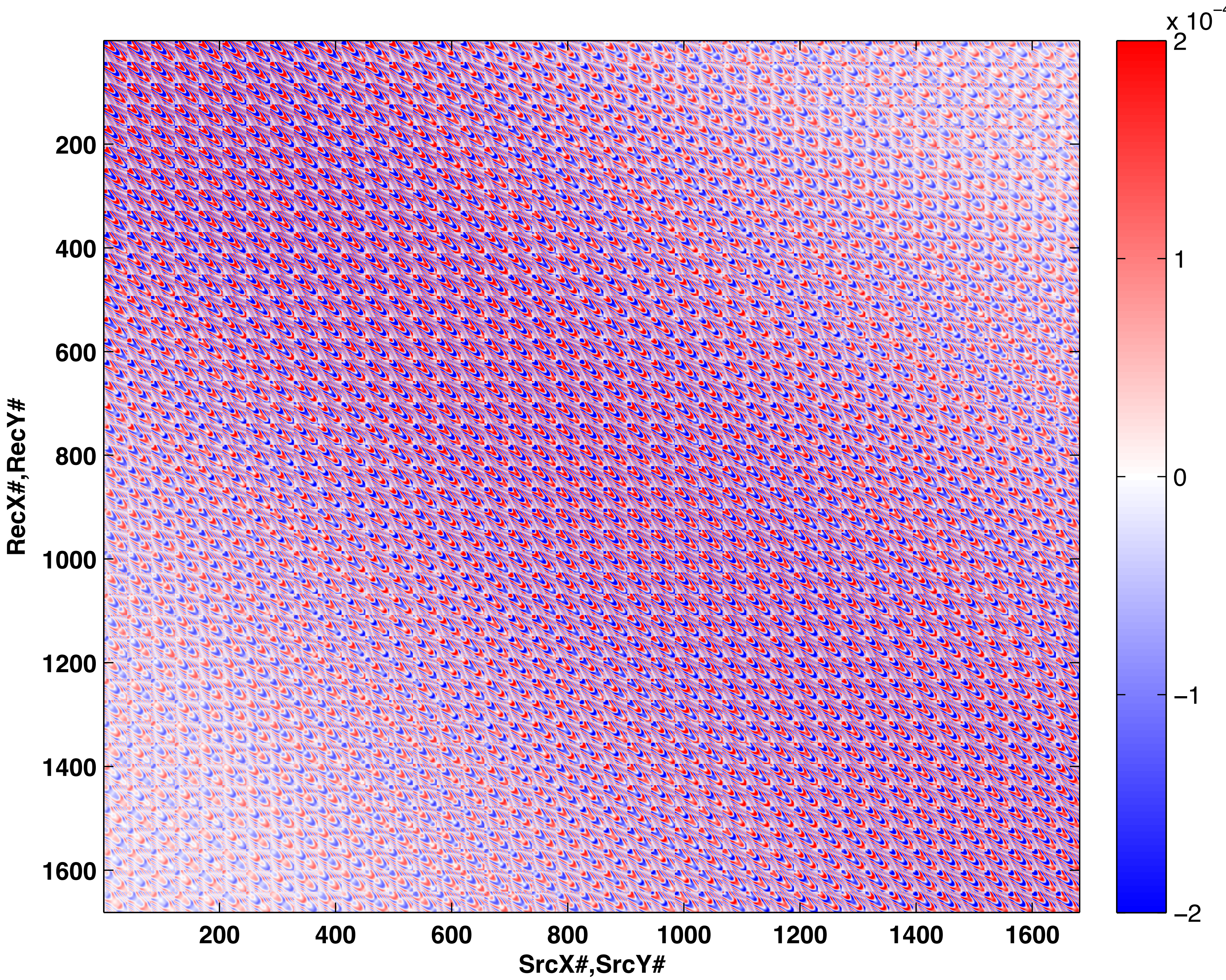


Freq : 60 Hz

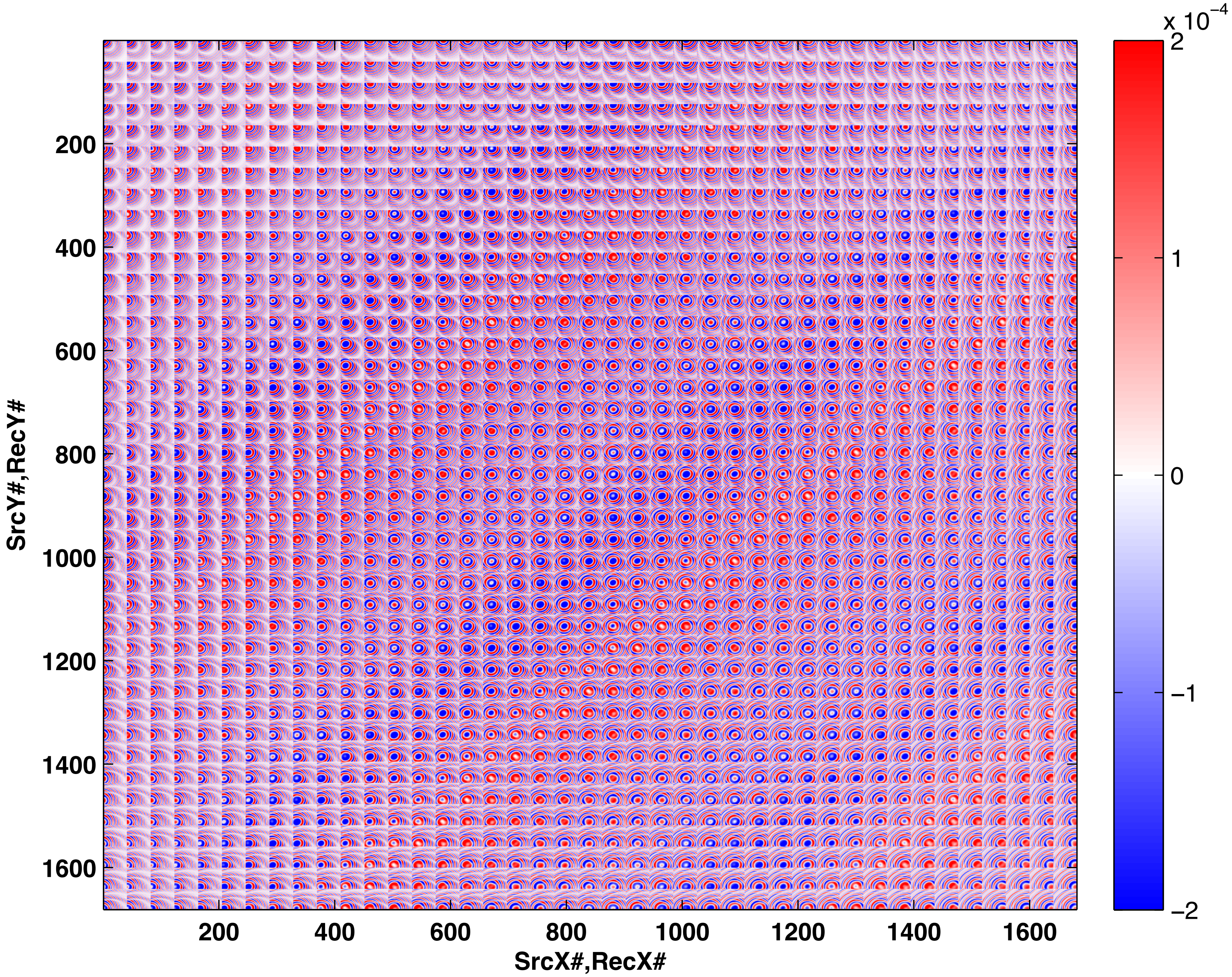


3D Acquisition

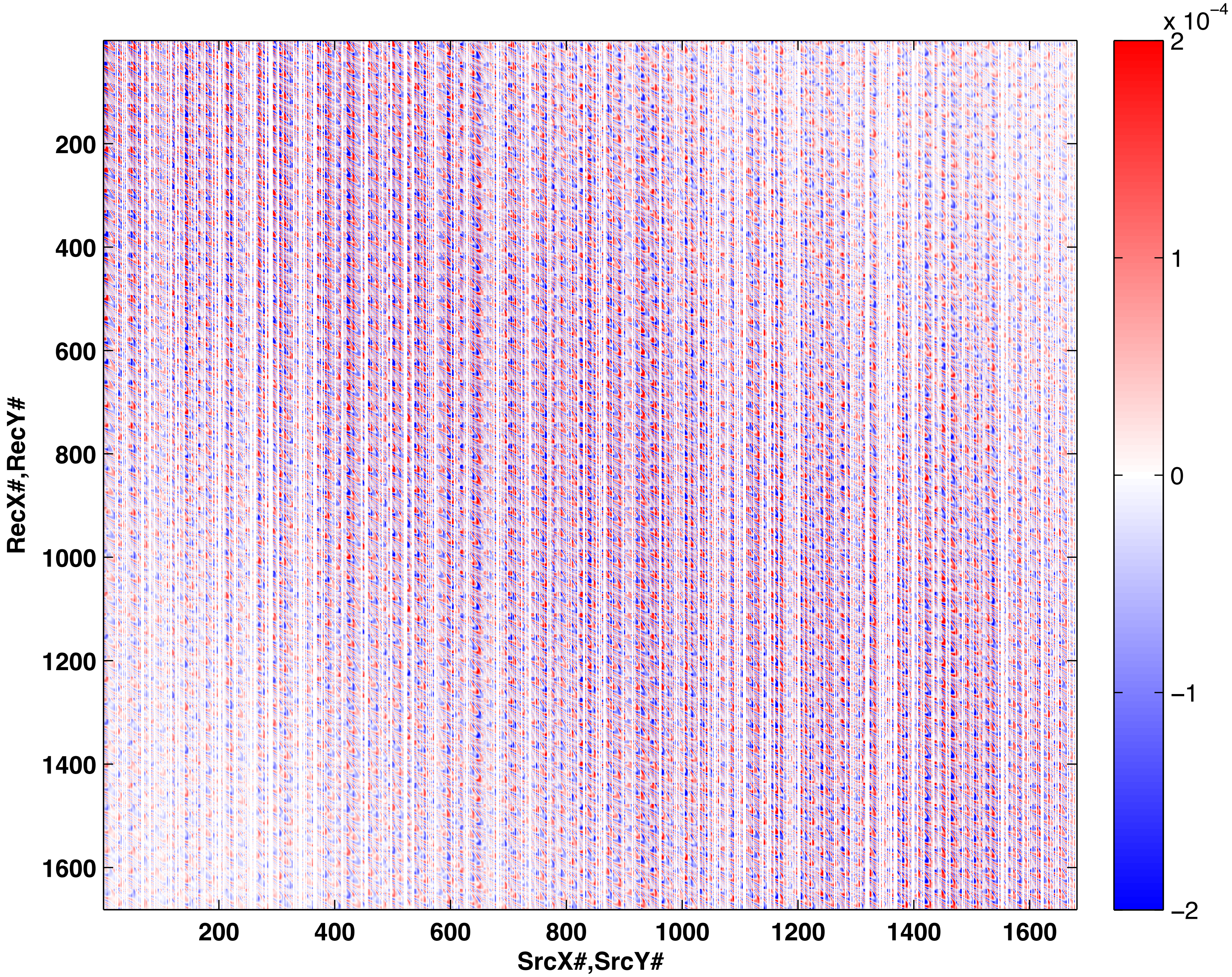
3D Acquisition
[Regular sampled data]



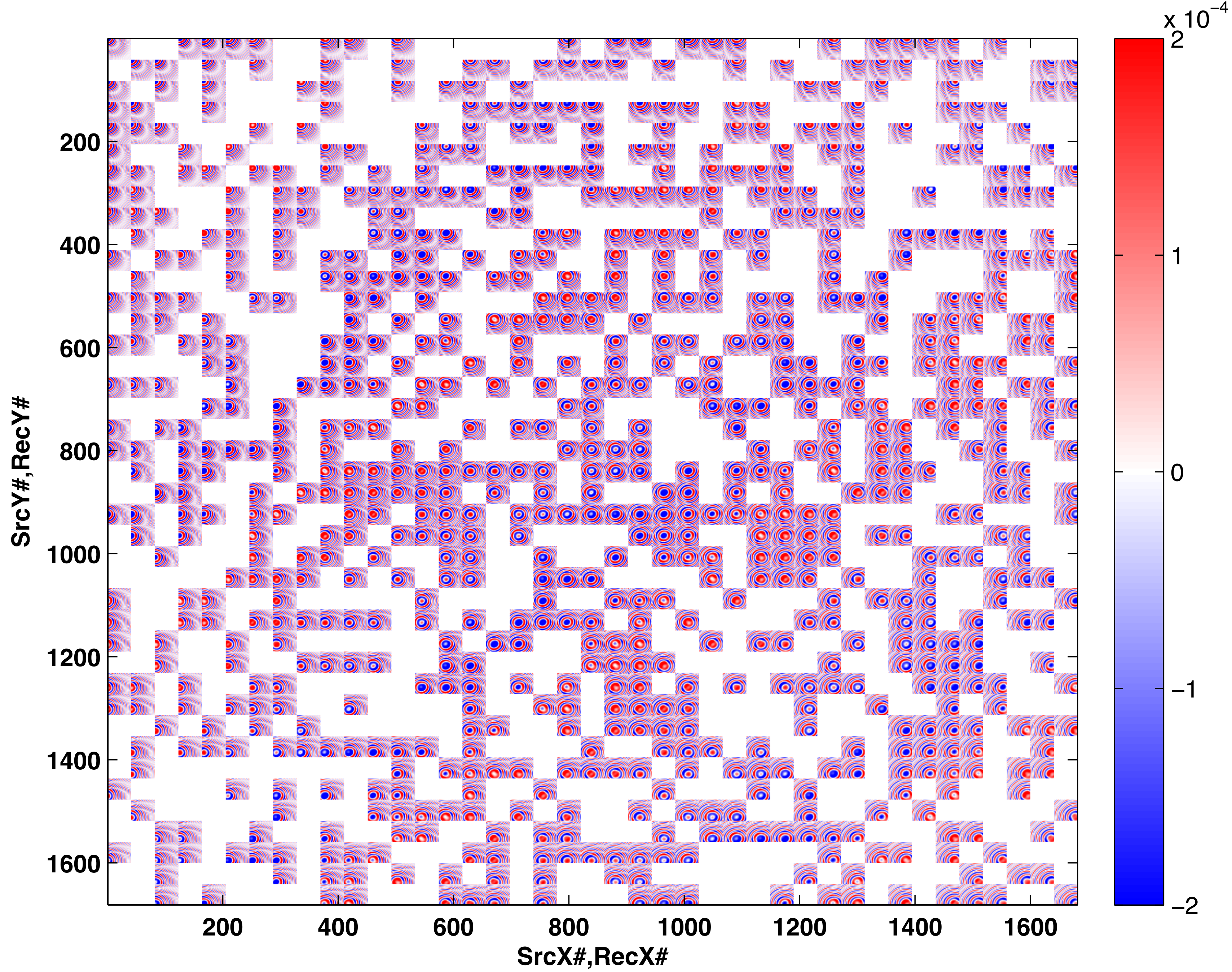
3D Acquisition
[Regular sampled data - “Transform” domain]



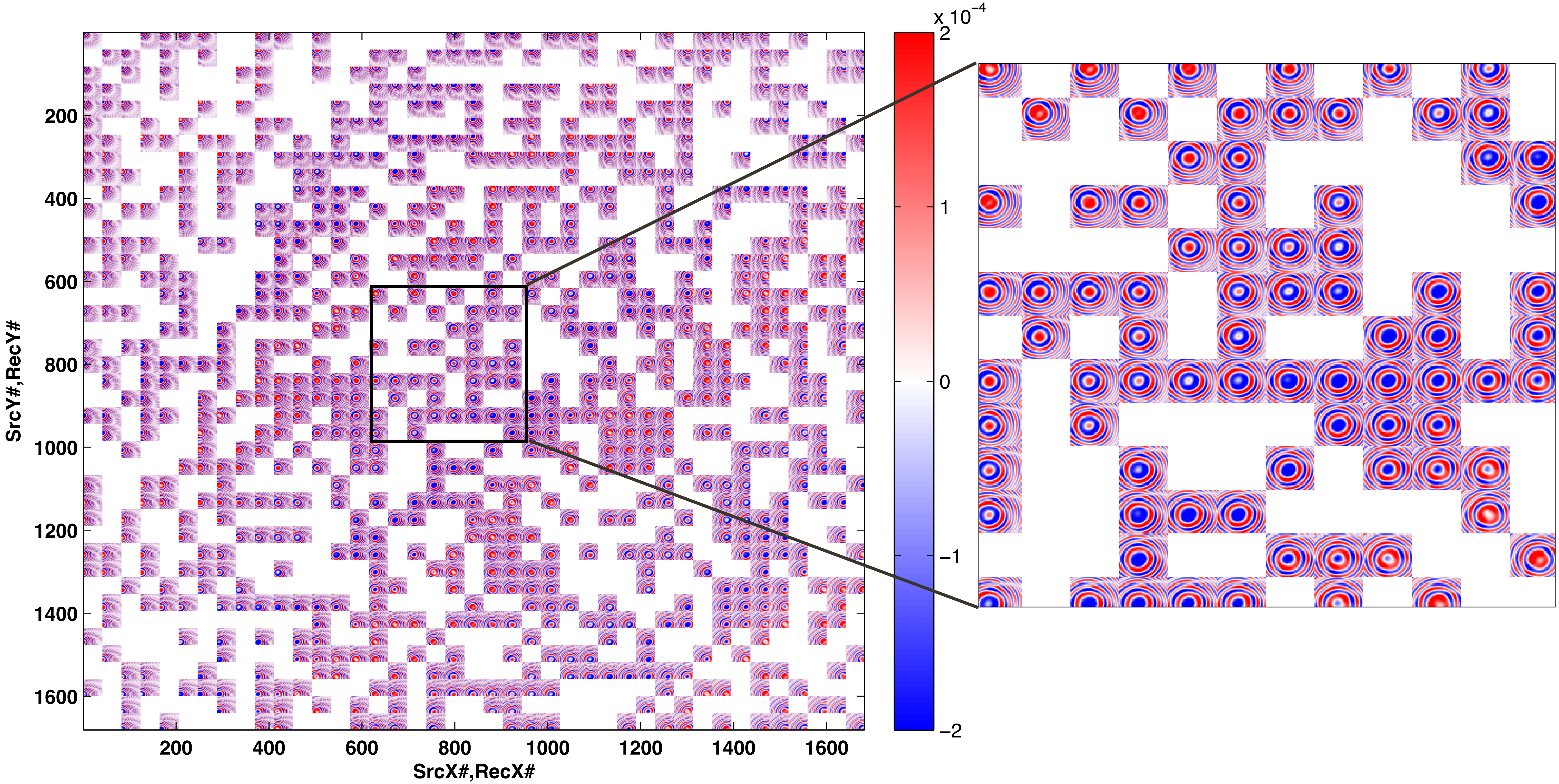
3D Acquisition
[Irregular sampled data]



3D Acquisition
[Irregular sampled data - "Transform" domain]

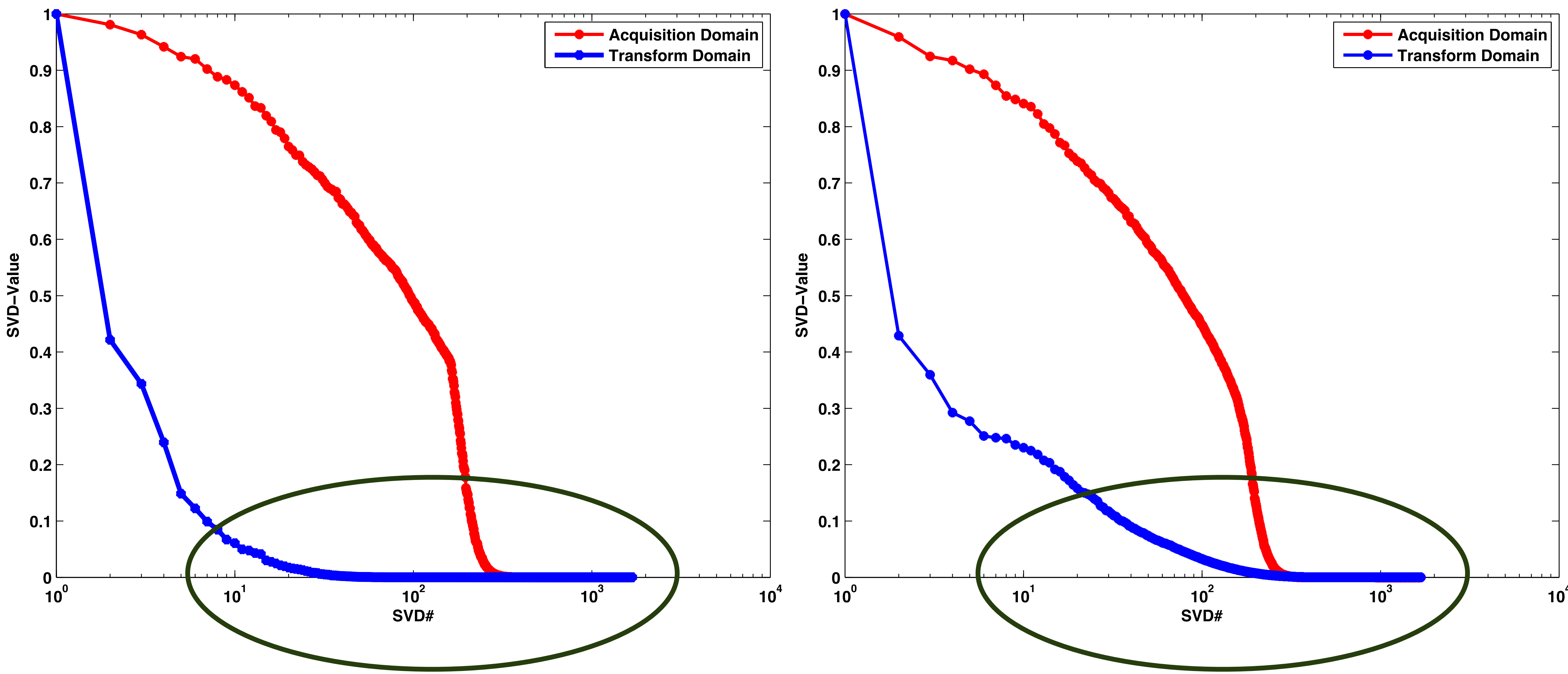


3D Acquisition
[Irregular sampled data - “Transform” domain]



3D Acquisition
[Singular value decay]

- **Acquisition domain - (SourceX, SourceY) & (ReceiverX, ReceiverY)**
- **Transform domain - (SourceX, ReceiverX) & (SourceY, ReceiverY)**



Original Data

Missing Data

Low Rank domain - Seismic

- 2D acquisition :
  - Midpoint-Offset
- 3D acquisition :
  - (SourceX,ReceiverX) and (SourceY,ReceiverY)
  - Midpoint-Offset-Azimuth

Sparsity vs Rank

Sparsity

- ***sparse/compressible***
- missing traces make signal ***less sparse*** in transform domain
- recovery using ***sparsity promoting*** scheme

Rank

- ***low rank/fast decay*** of singular values
- missing traces ***increase rank*** in "transform domain"
- recovery using ***rank penalization*** scheme

Observations

- Sampling become *incoherent* in “transform” domain
- *Slow decay* of singular values in “transform” domain

Low Rank Optimization

Methodology

- given data matrix $\mathbf{b}$ (in a "*transform* domain")

$$(BPDN_\sigma) \qquad \min_X ||X||_* \quad \text{s.t. } ||\mathcal{A}(X) - \mathbf{b}||_2^2 \leq \sigma$$

where $*$ is the *nuclear* norm given by :

$$\|X\|_* = \sum_{i=1}^{m} \lambda_i = \|\lambda\|_1$$

where λ_i are the *singular* values

Methodology

- $\mathcal{A}$ is the sampling operator in the "*transform*" domain defined by :

$$\mathcal{A} = \mathbf{RM}$$

where

$\mathbf{R}$: restriction operator

$\mathbf{M}$: measurement operator

Impediments

- requires repeated application of *SVD* for projection
- *expensive* to compute for large system
- can we exploit *rank structure* *“**SVD free**”?*

Matrix Factorization

(Recht et al.)

SLIM

Methodology

- given data matrix $\mathbf{b}$ (in a "*transform* domain")
- parameterize $X \in \mathbb{R}^{n \times m}$ using factors:

$$X = LR'$$

where $L \in \mathbb{R}^{n \times k}$ and $R \in \mathbb{R}^{m \times k}$

- note that rank $k << max(m, n)$ is enforced through parametrization

Methodology

- solve the least-squares problem :

$$\min_{L,R} \|\mathcal{A}(LR') - \mathbf{b}\|_2^2$$

- implemented using *L-BFGS* with 300 iterations

Experiment & Results

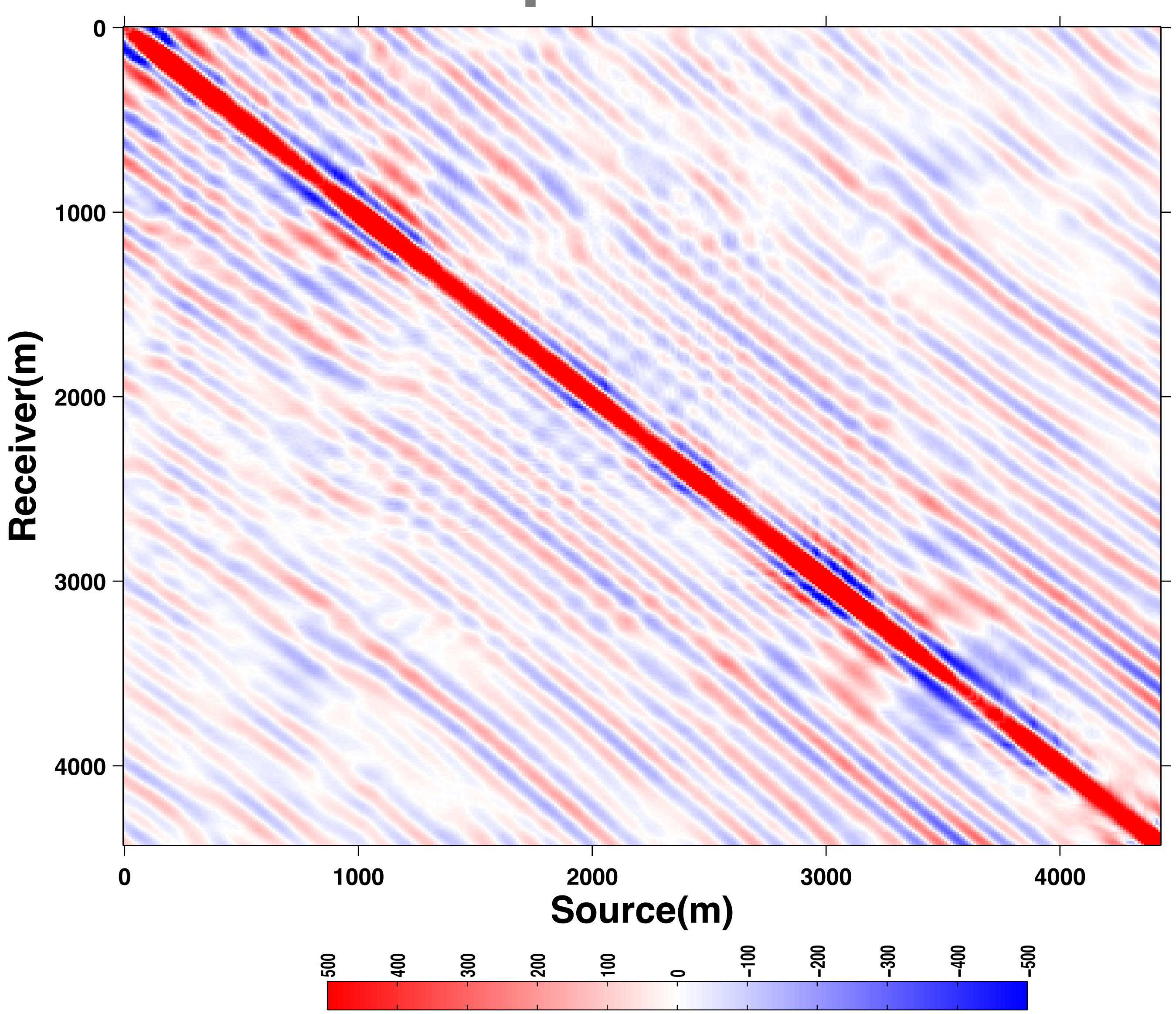
2D Interpolation

2D Acquisition

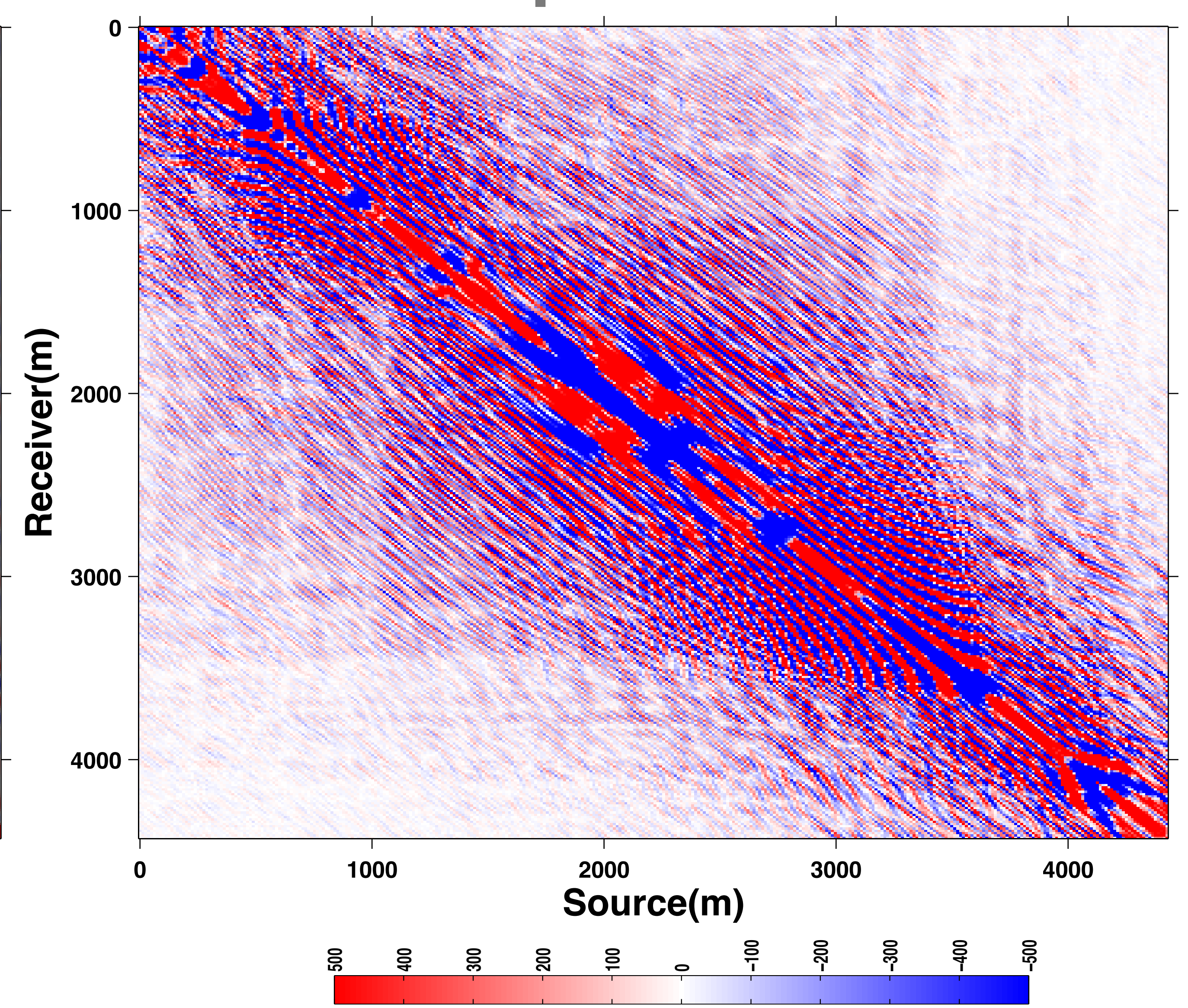
- Gulf of Suez data set
- 354 receivers, 354 sources, 1024 time samples
- monochromatic frequency slice
- rank adjusted according to frequencies

Regular sampled data

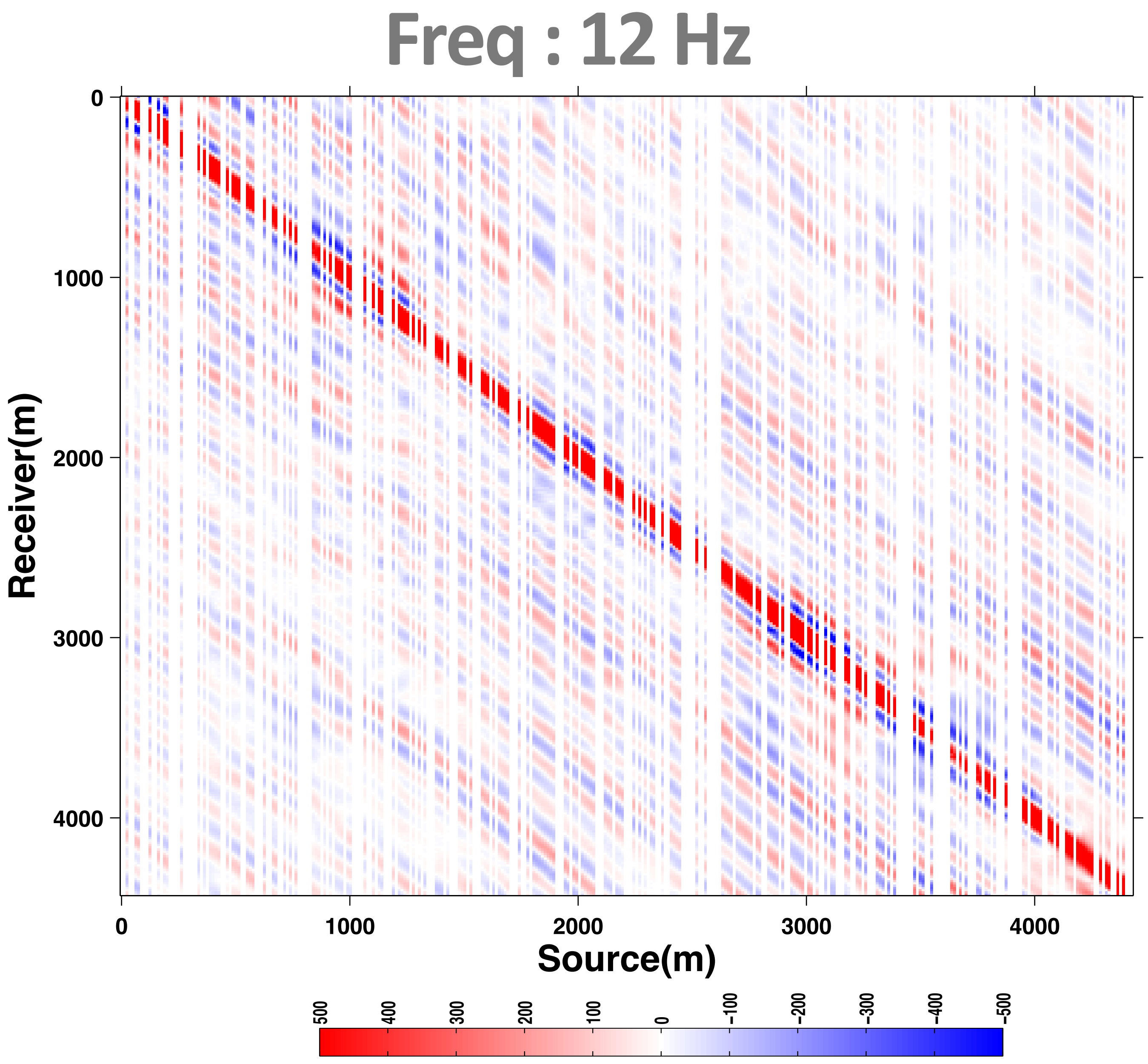
Freq : 12 Hz

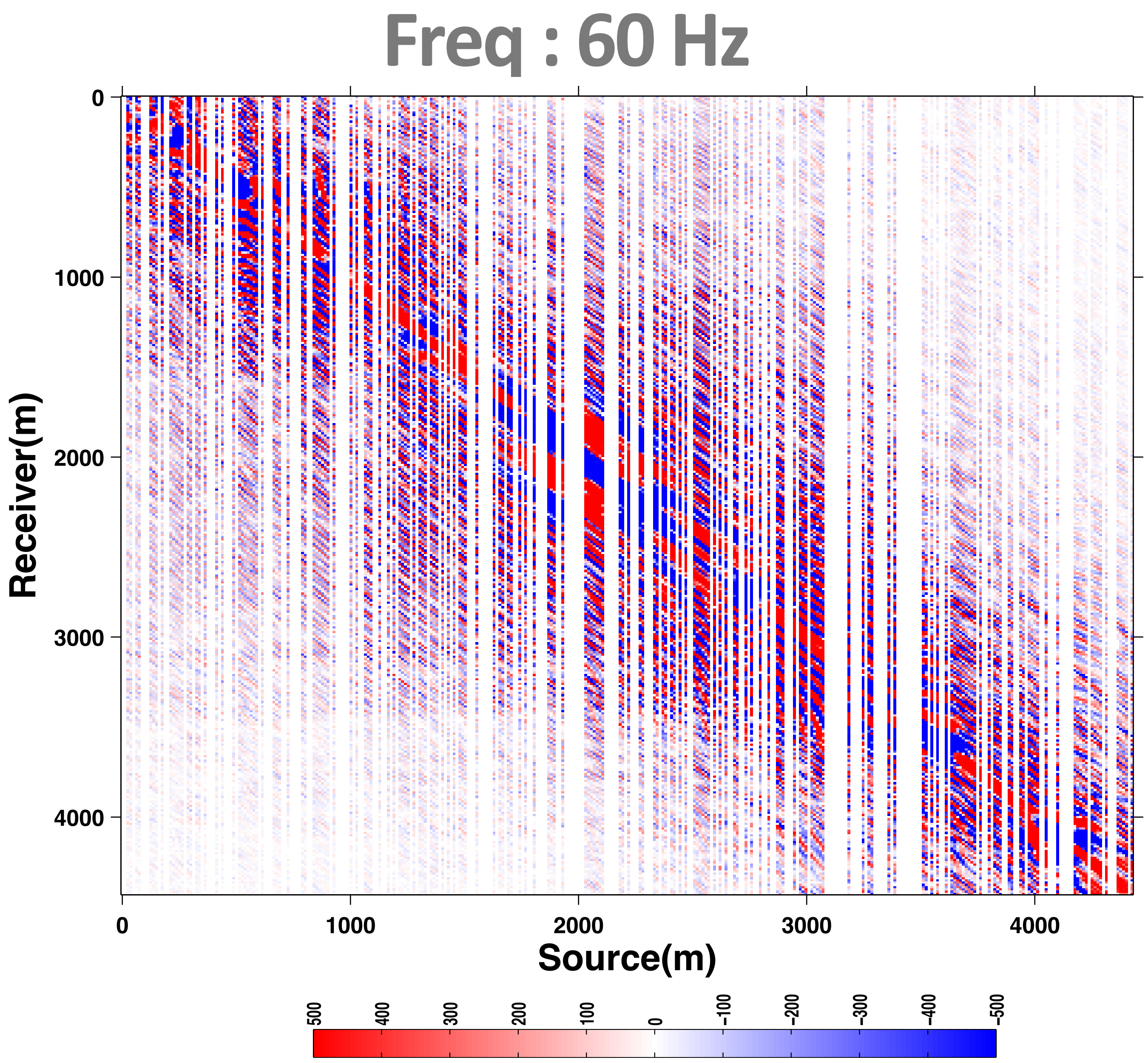


Freq : 60 Hz

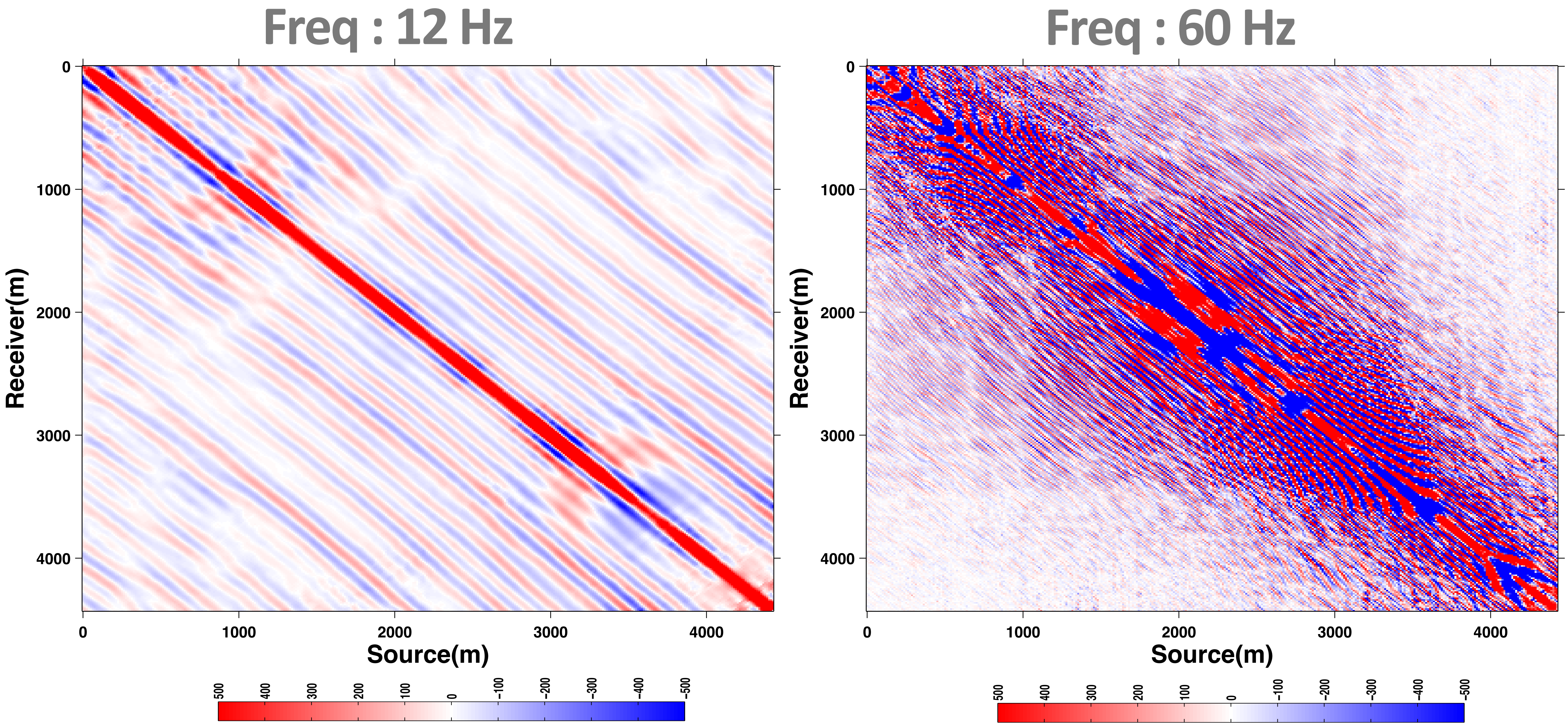


50% Missing Shots





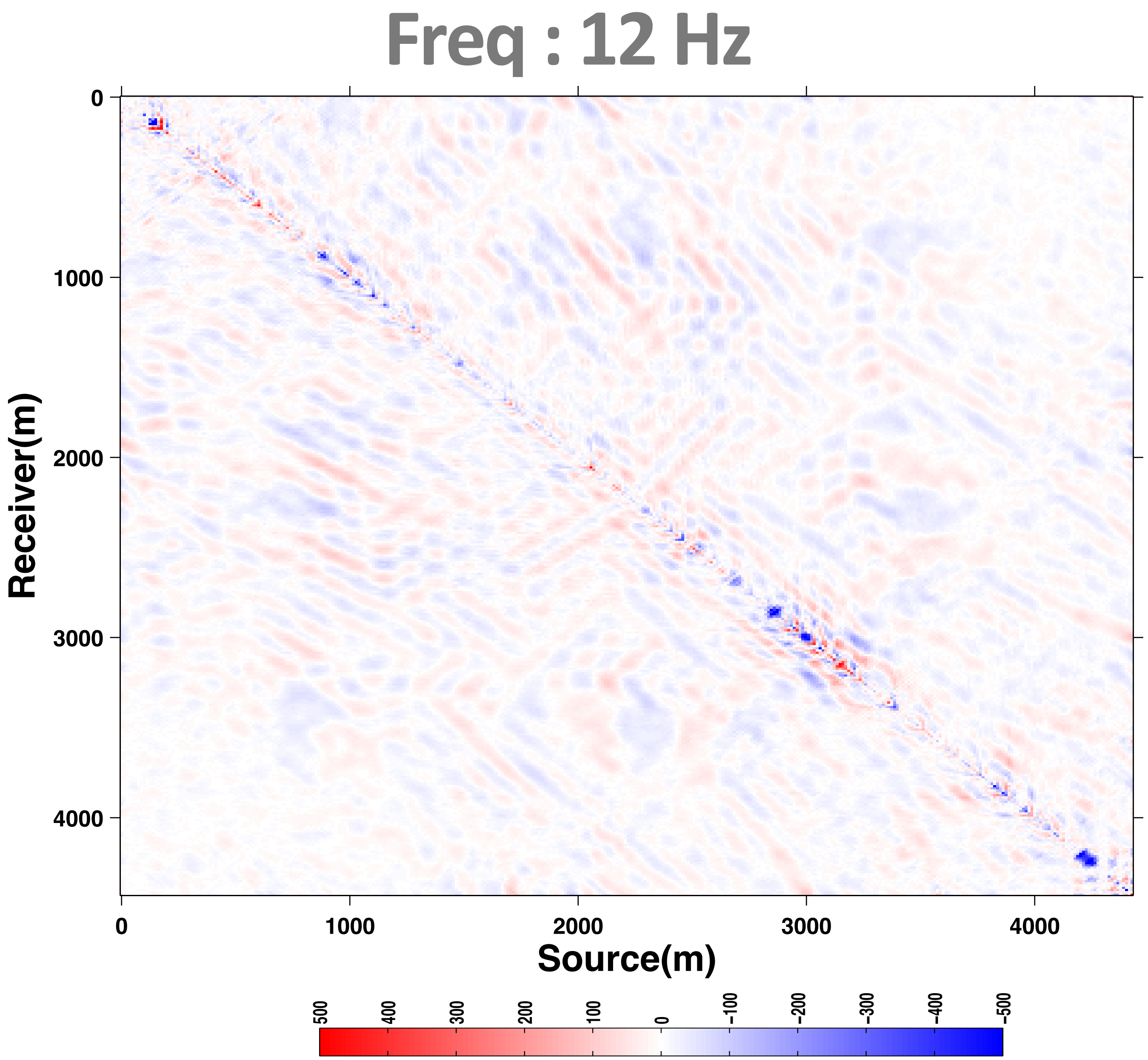
Recovery using MH formulation

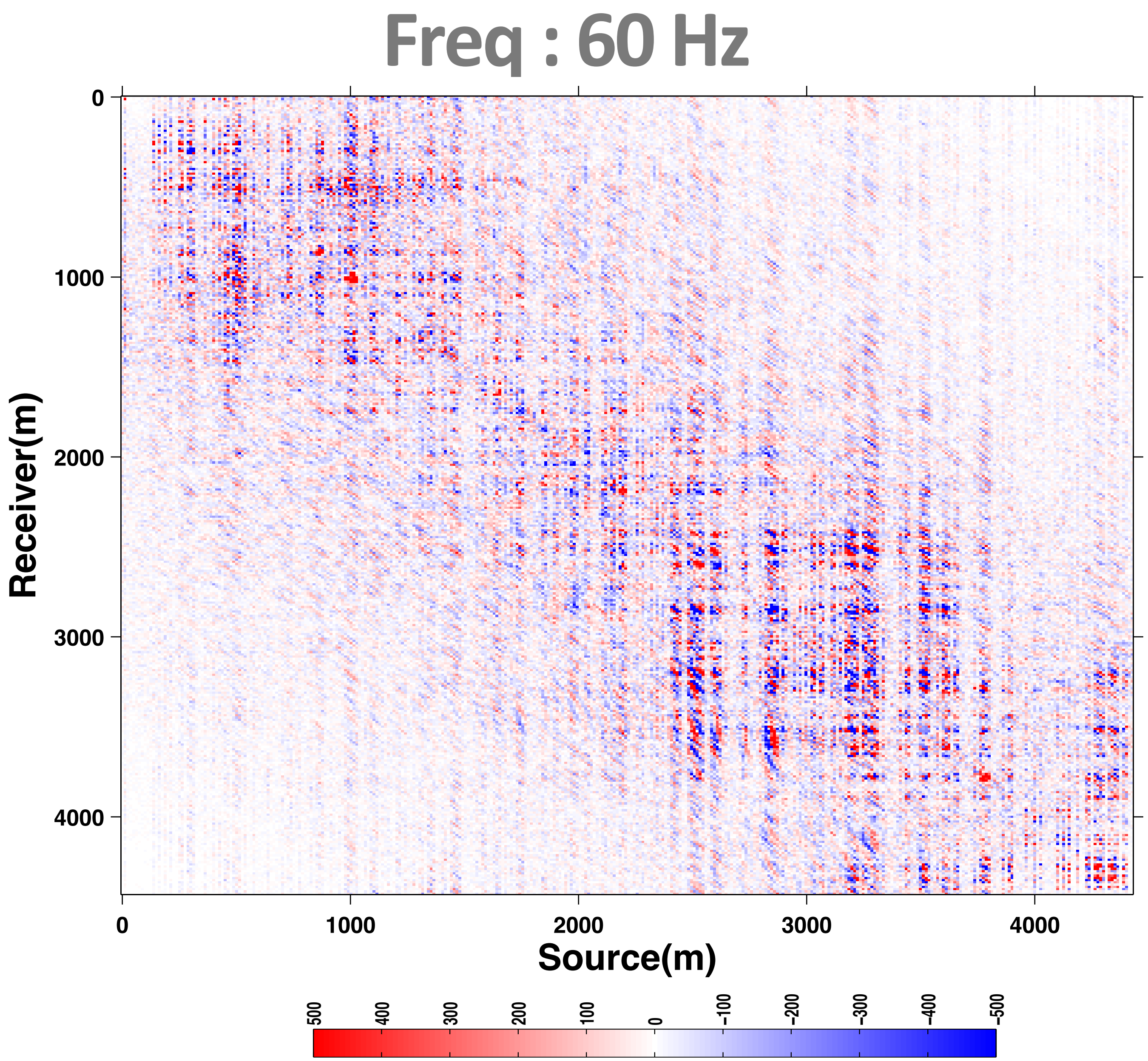


SNR = 13.8 db, Rank=5

SNR = 16.1 db, Rank=30

Residual


Freq : 12 Hz
Receiver(m)
Source(m)
0
1000
2000
3000
4000
500
400
300
200
100
0
-100
-200
-300
-400
-500


Freq : 60 Hz
Receiver(m)
Source(m)
0
1000
2000
3000
4000
500
400
300
200
100
0
-100
-200
-300
-400
-500

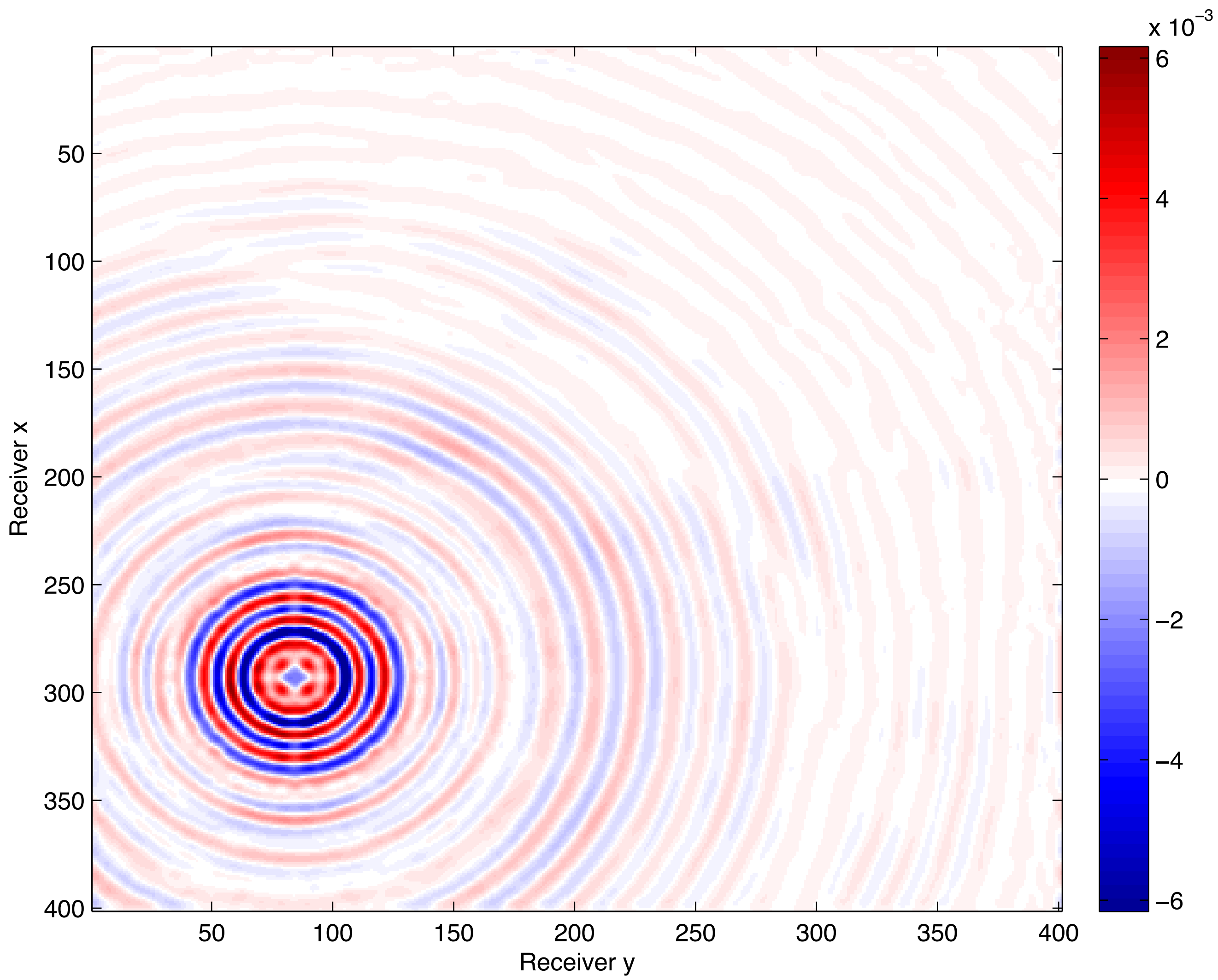
3D Interpolation

3D Acquisition

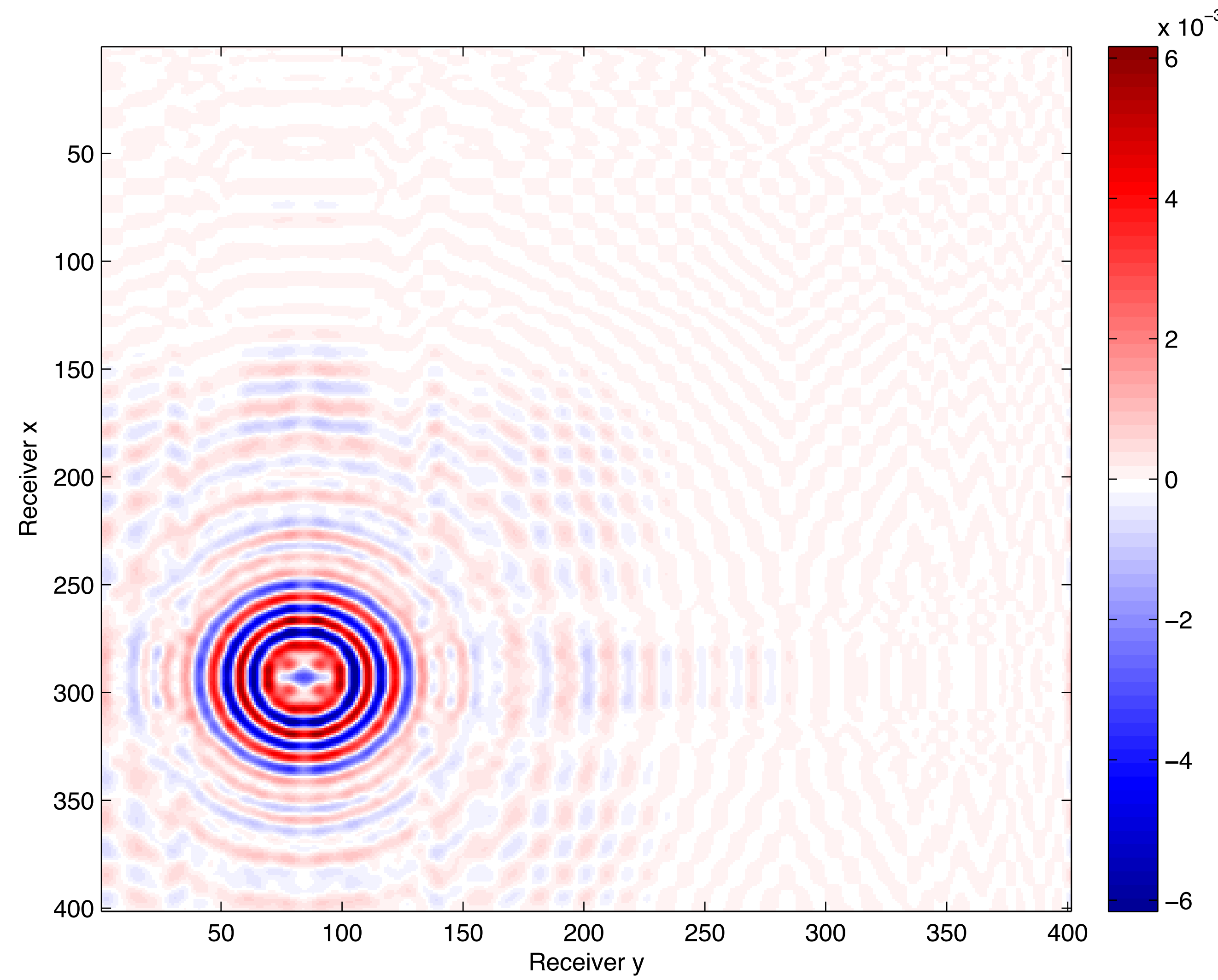
- Synthetic BG data set
- 200 *randomly subsampled* shots out of 4624 i.e 4% of shots
- 1201 x 1201 receivers
- Implementation of formulation on smaller grid with 401 x 401 receivers
- monochromatic frequency slice at 7.34 Hz
- recovery for rank 10

Recovery - Shot Gather
[Rank=10, 500 Iterations]

True data
(SrcX,SrcY = 50,15)

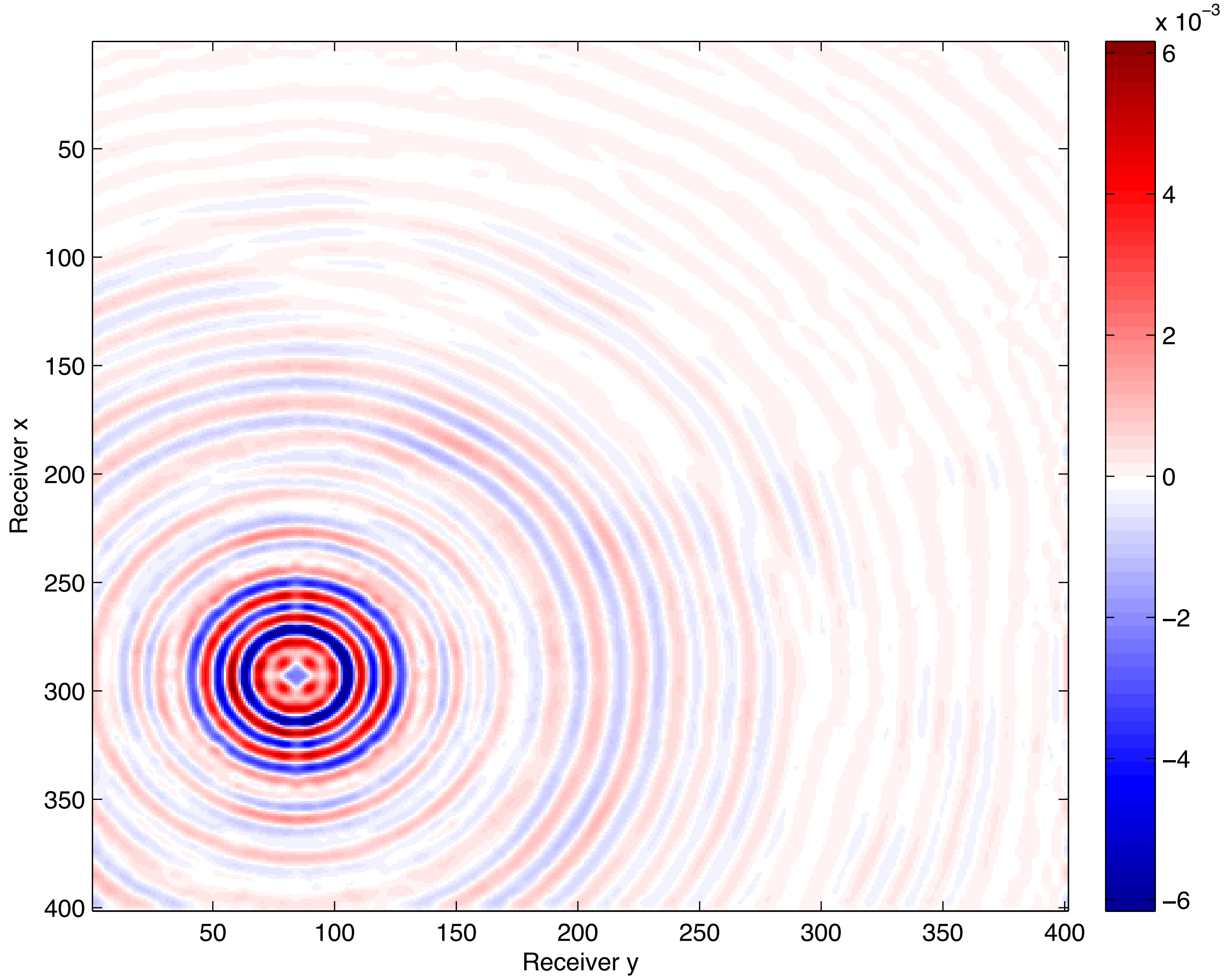


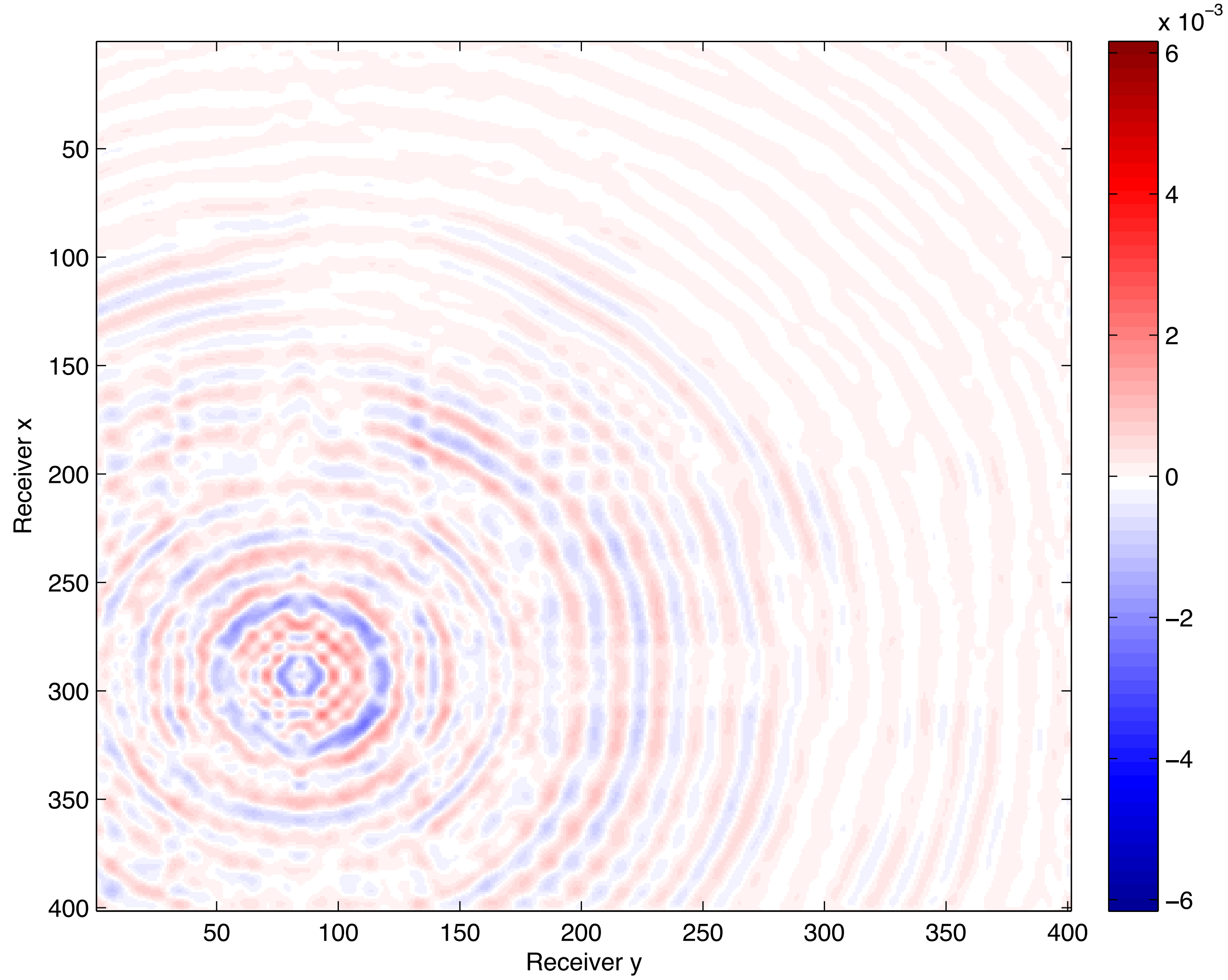
Approximated data
SNR=8.14 db



Residual - Shot Gather
[Rank=10, 500 Iterations]

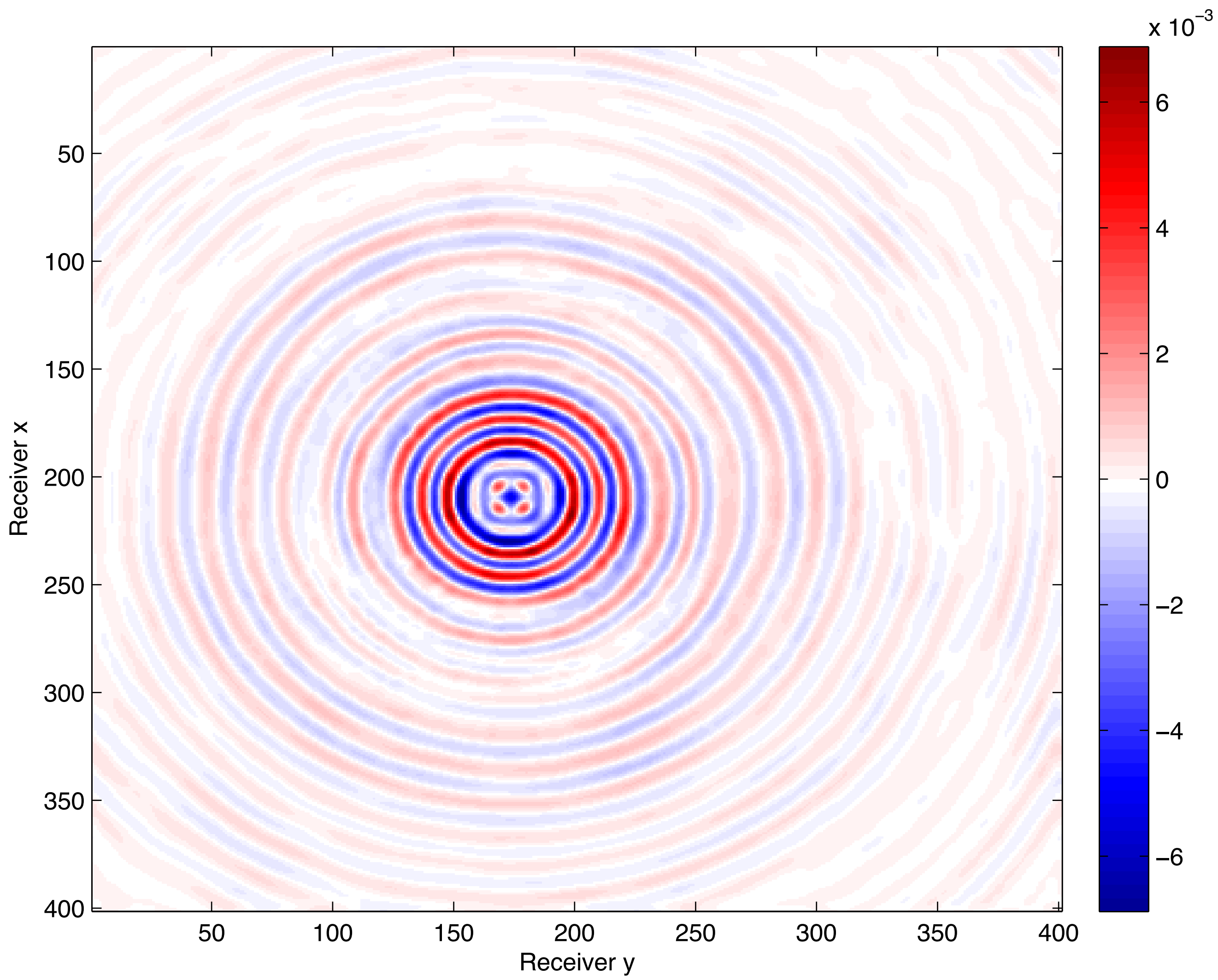
True data
(SrcX,SrcY = 50,15)



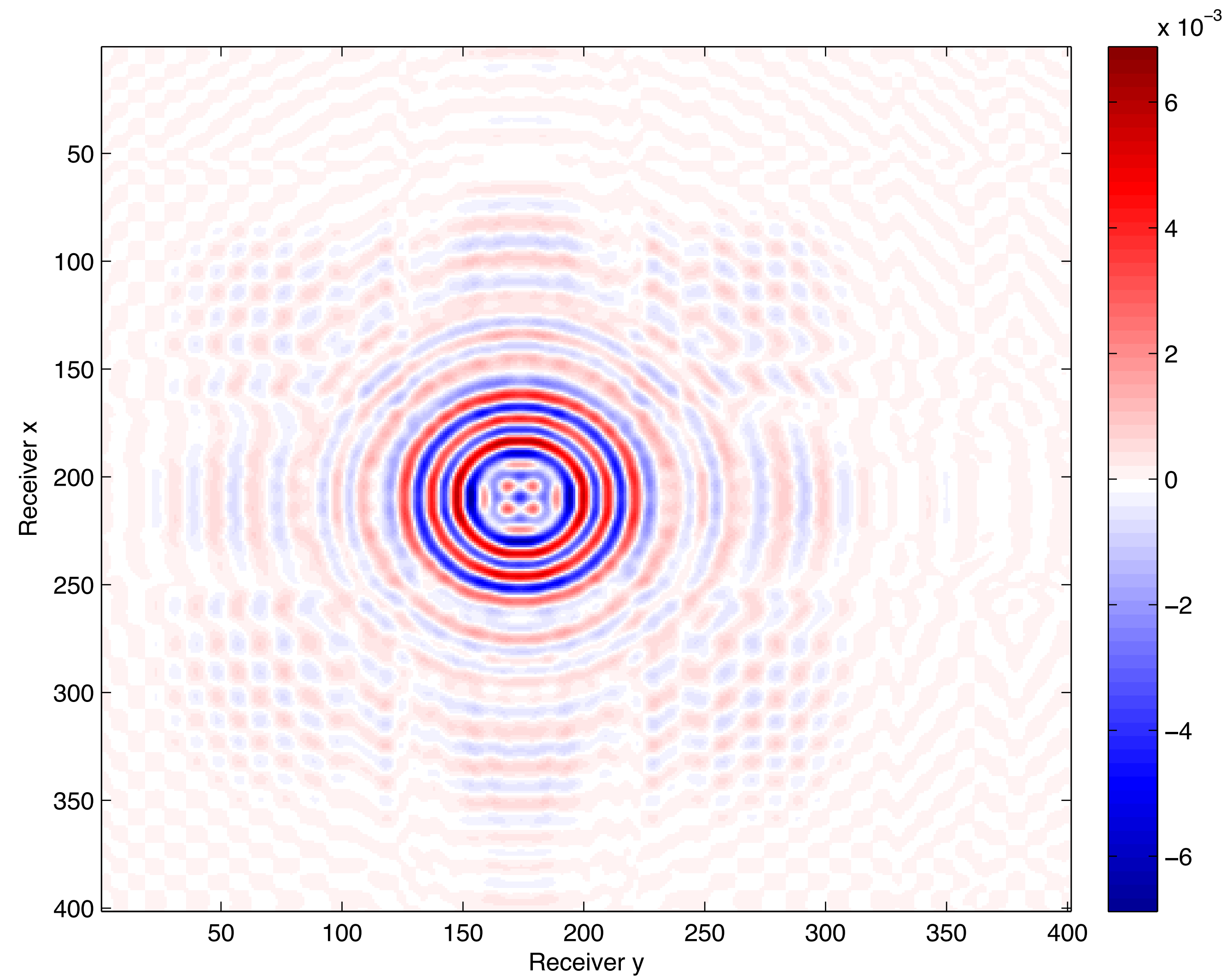


Recovery - Shot Gather
[Rank=10, 500 Iterations]

True data (SrcX,SrcY = 36,30)



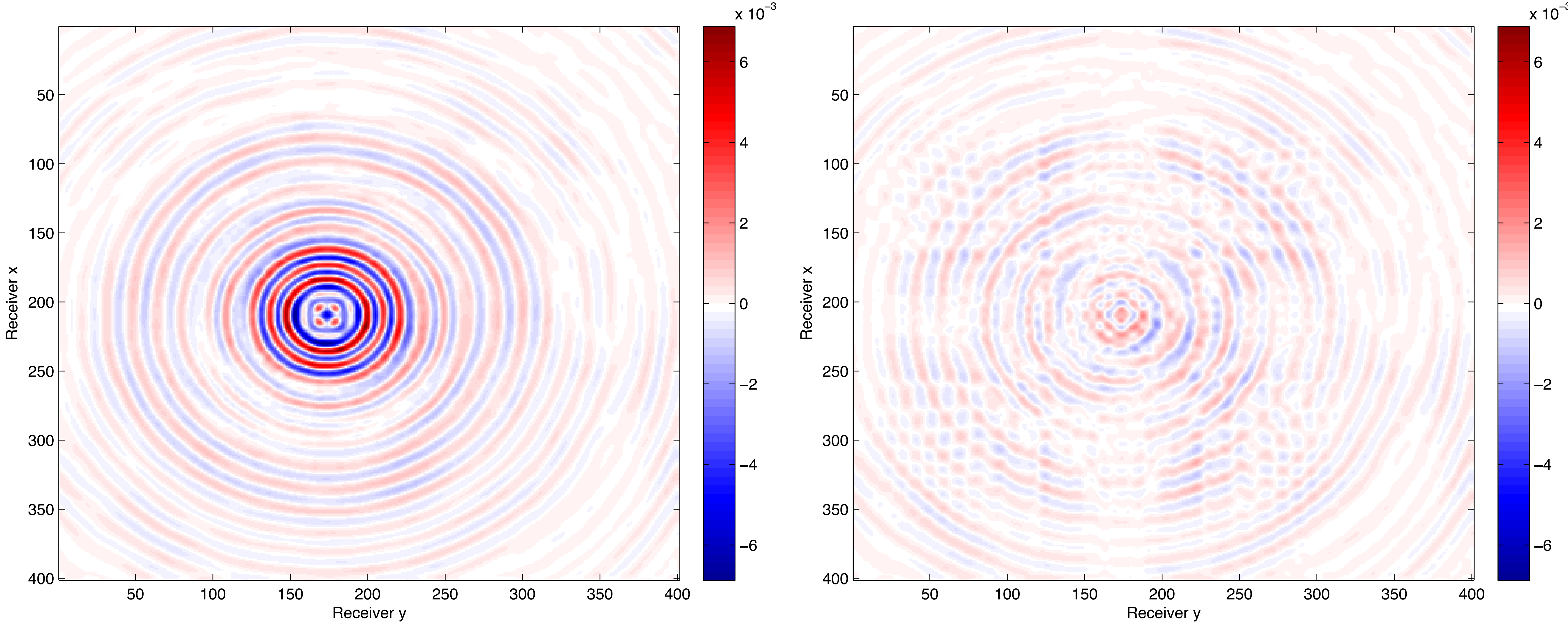
Approximated data SNR=8.49 db



Residual - Shot Gather
[Rank=10, 500 Iterations]

True data

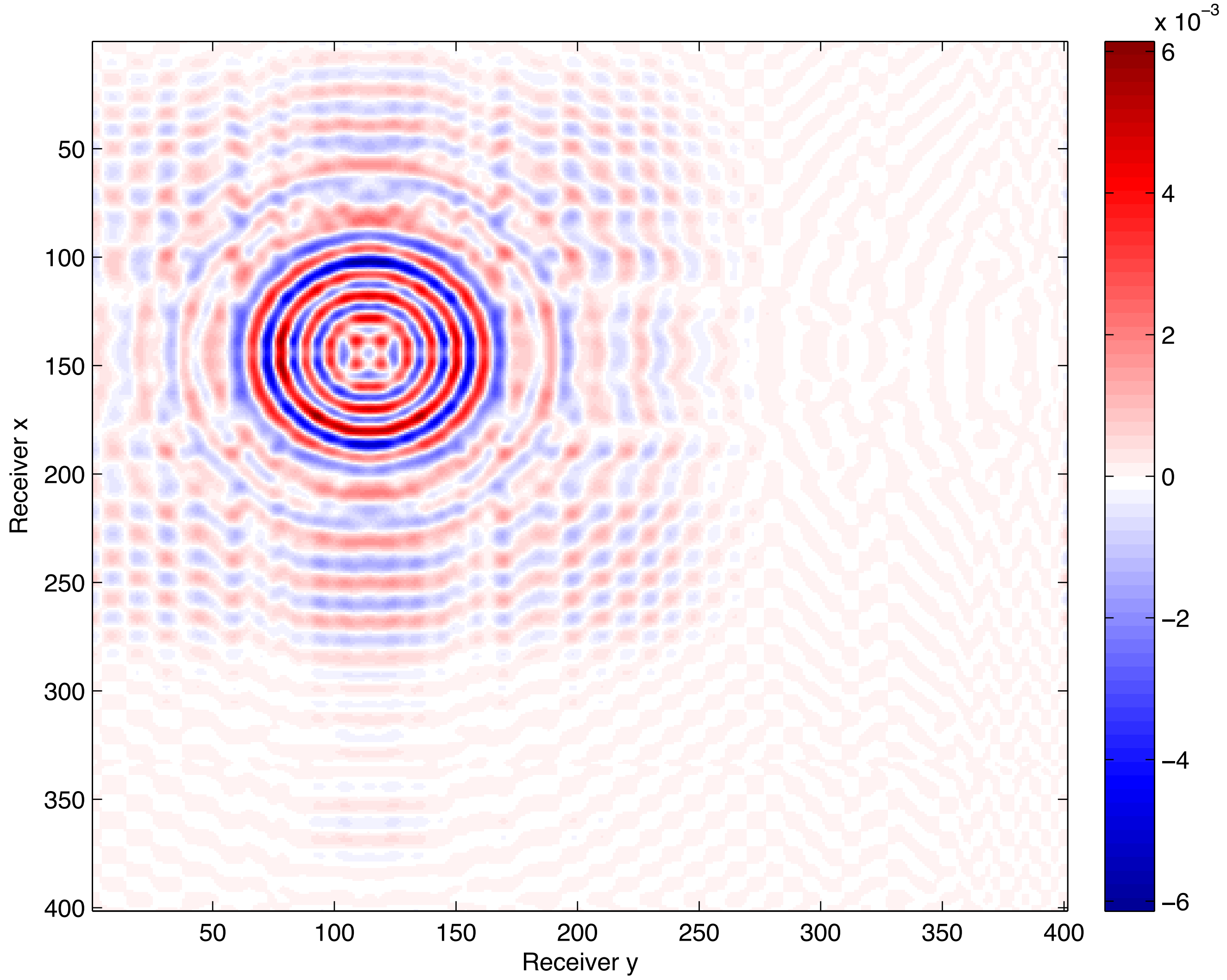
(SrcX,SrcY = 36,30)



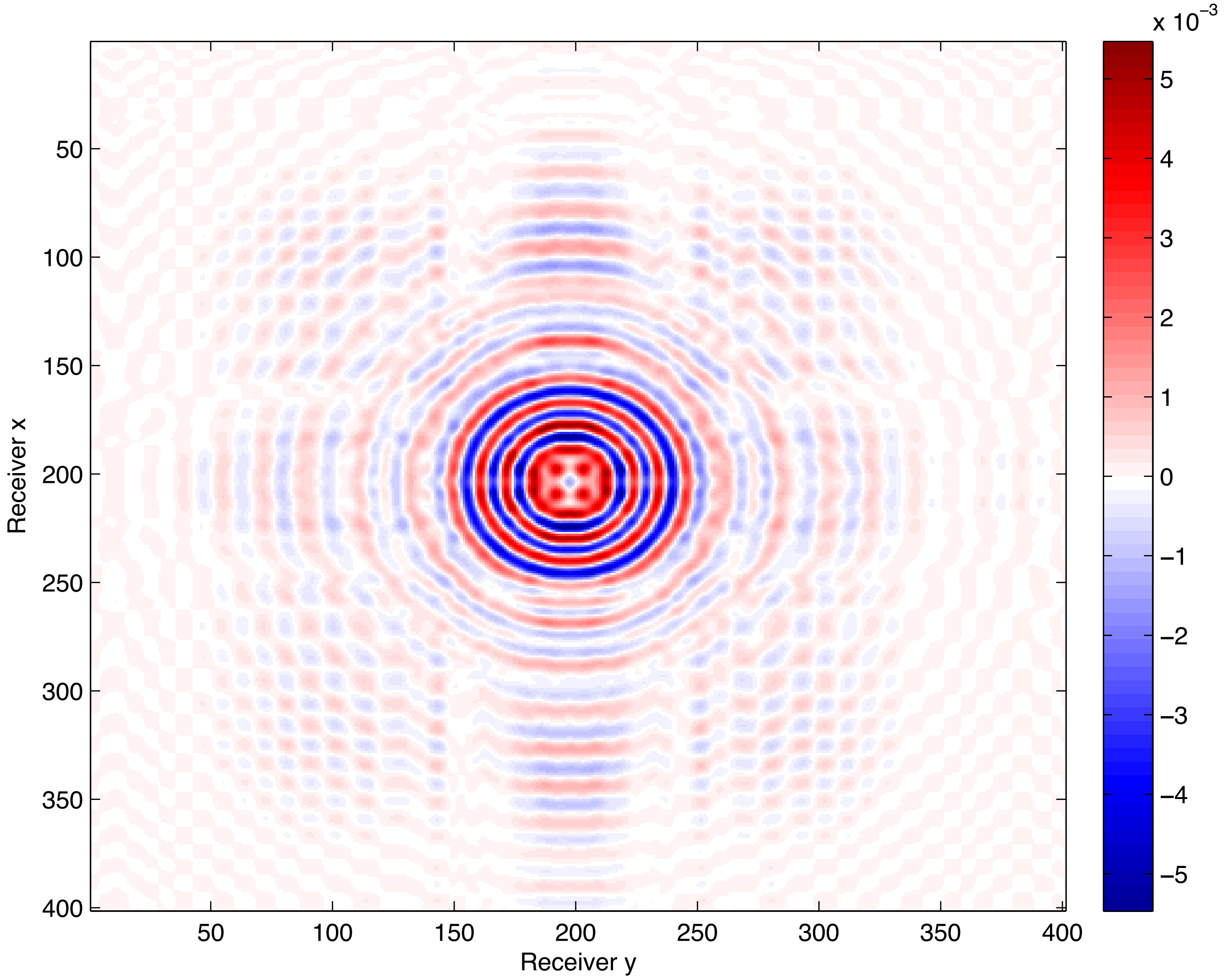
Monday, 3 December, 12

Recovery - Shot Gather
[Rank=10, 500 Iterations]

Interpolated data (SrcX,SrcY = 25,20)



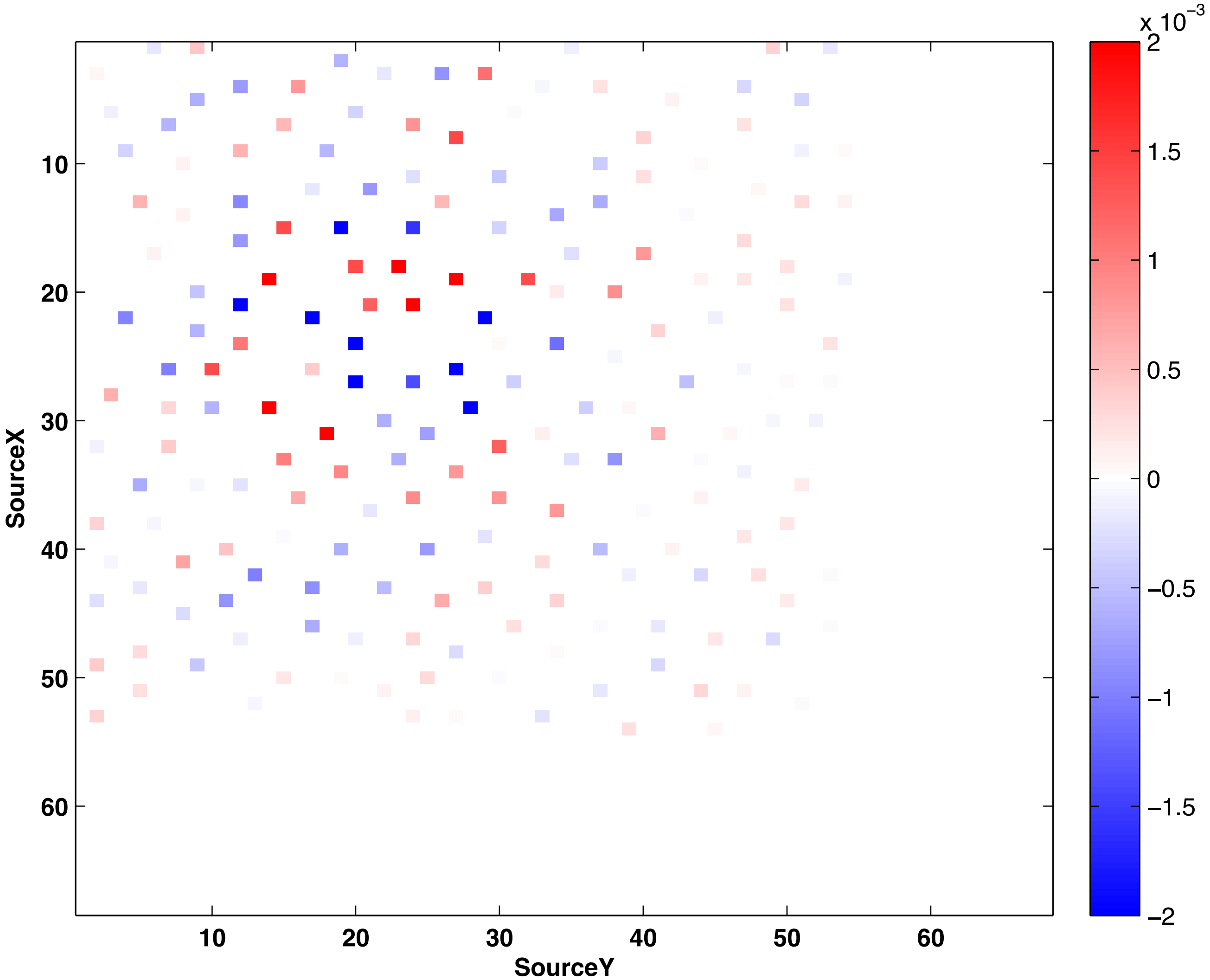
Interpolated data (SrcX,SrcY = 35,34)



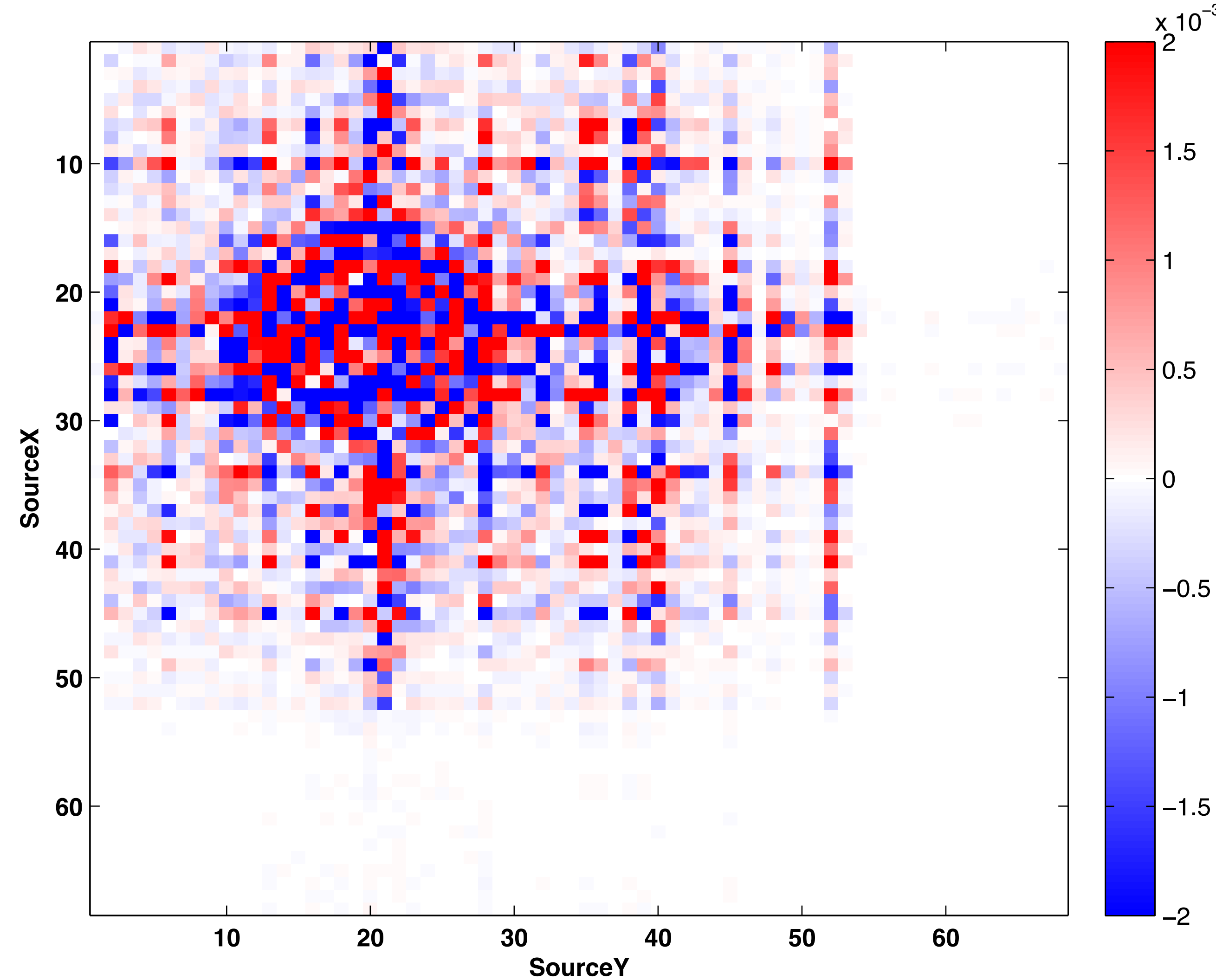
Recovery - Receiver Gather

[Rank=10, 500 Iterations]

Acquired data (RecX,RecY = 35,30)



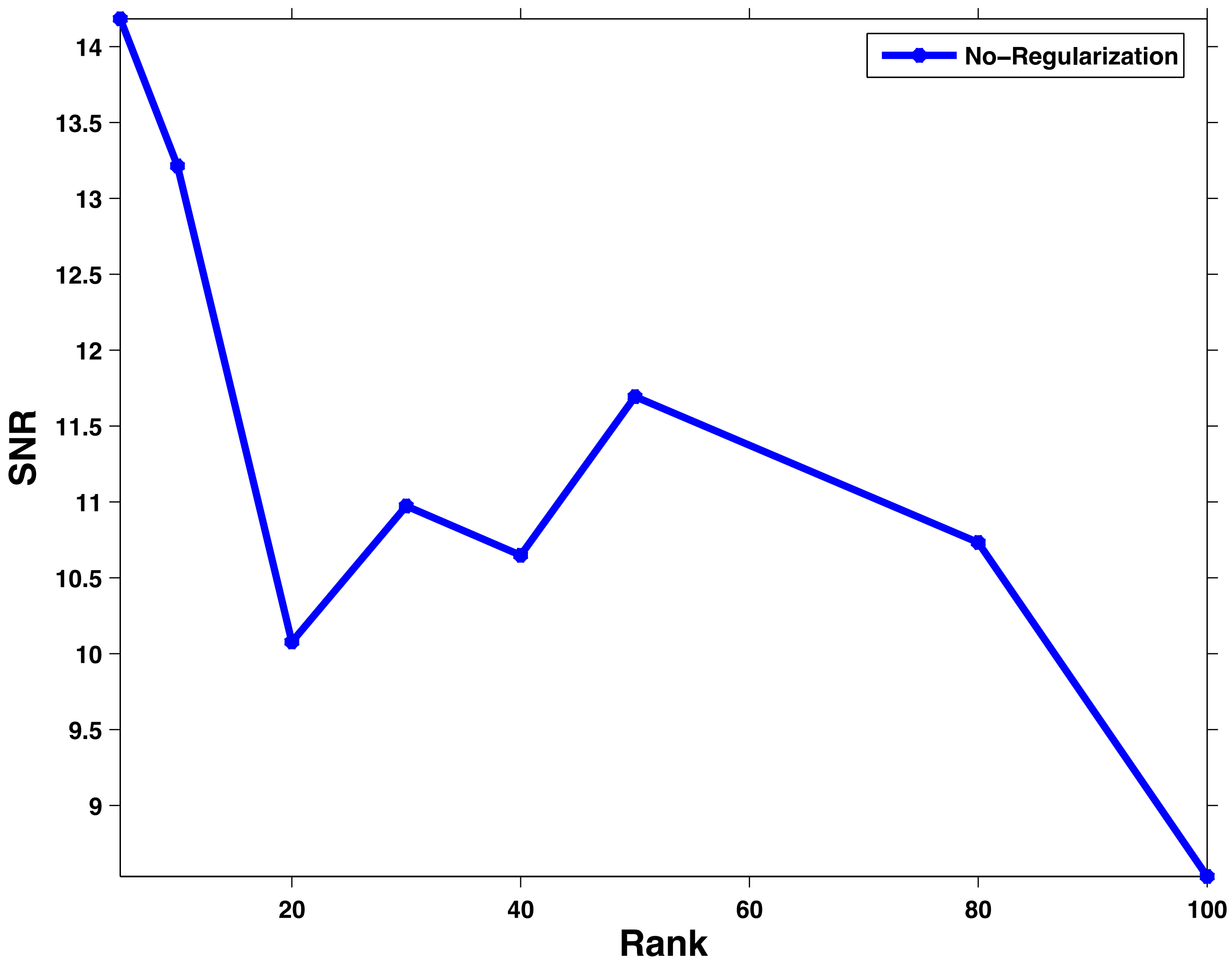
Interpolated data (RecX,RecY = 35,30)



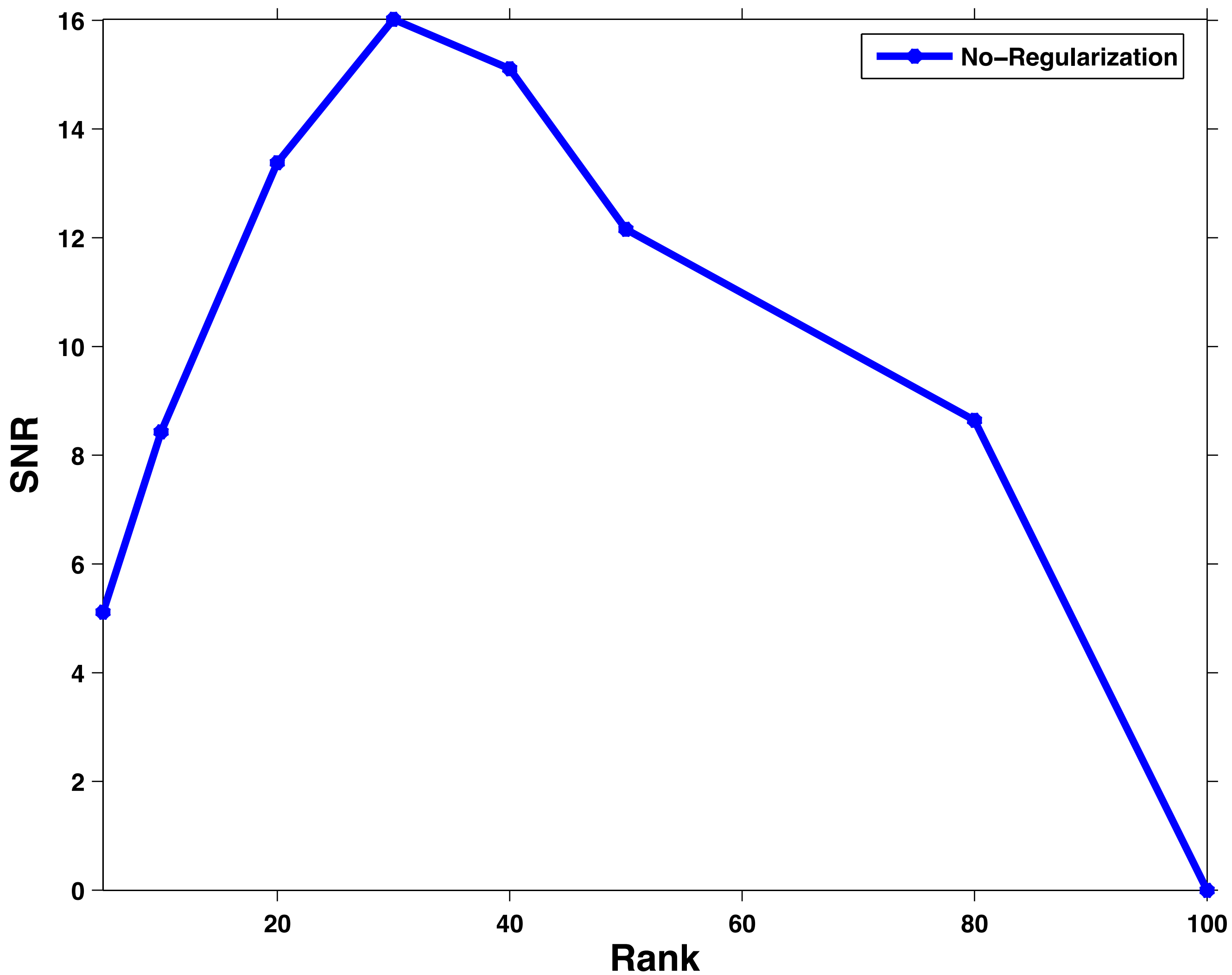
Recovery from 96% missing data

Limitation
[SNR vs Rank - 2D Interpolation]

Better SNR can achieve by increasing k, but unconstrained formulation starts over-fitting the data



Freq : 12 Hz



Freq : 60 Hz

Regularized Matrix Factorization

SLIM

Methodology

- given data matrix $\mathbf{b}$ (in a "*transform* domain")

- solve the following formulation :

$$(LASSO_\tau) \qquad \min_X ||\mathcal{A}(X) - \mathbf{b}||_2^2 \quad \text{s.t.} \quad ||X||_* \leq \tau$$

where τ is a rank regularization parameter.

(Recht et al.)

SLIM

Methodology

- Parametrize $X \in \mathbb{R}^{n\times m}$ using factors:

$$X = LR'$$

where $L \in \mathbb{R}^{n\times k}$ and $R \in \mathbb{R}^{m\times k}$

- Solve the following rank penalized problem:

$$\min_{L,R} \|\mathcal{A}(LR') - \mathbf{b}\|_2^2 \quad \text{s.t.} \; \|LR'\|_* \leq \tau$$

(Recht et al.)

Methodology

- nuclear norm is given by :

$$\|X\|_* = \inf_{X=LR'} \frac{1}{2} \left\| \begin{bmatrix} L \\ R \end{bmatrix} \right\|_F^2$$

where $\|.\|_F^2$ is sum of squares of all entries

- Therefore, $\|LR'\|_* \leq \frac{1}{2} \left\| \begin{bmatrix} L \\ R \end{bmatrix} \right\|_F^2$

  and we can avoid SVDs for projection.

Methodology

- We solve the BPDN problem

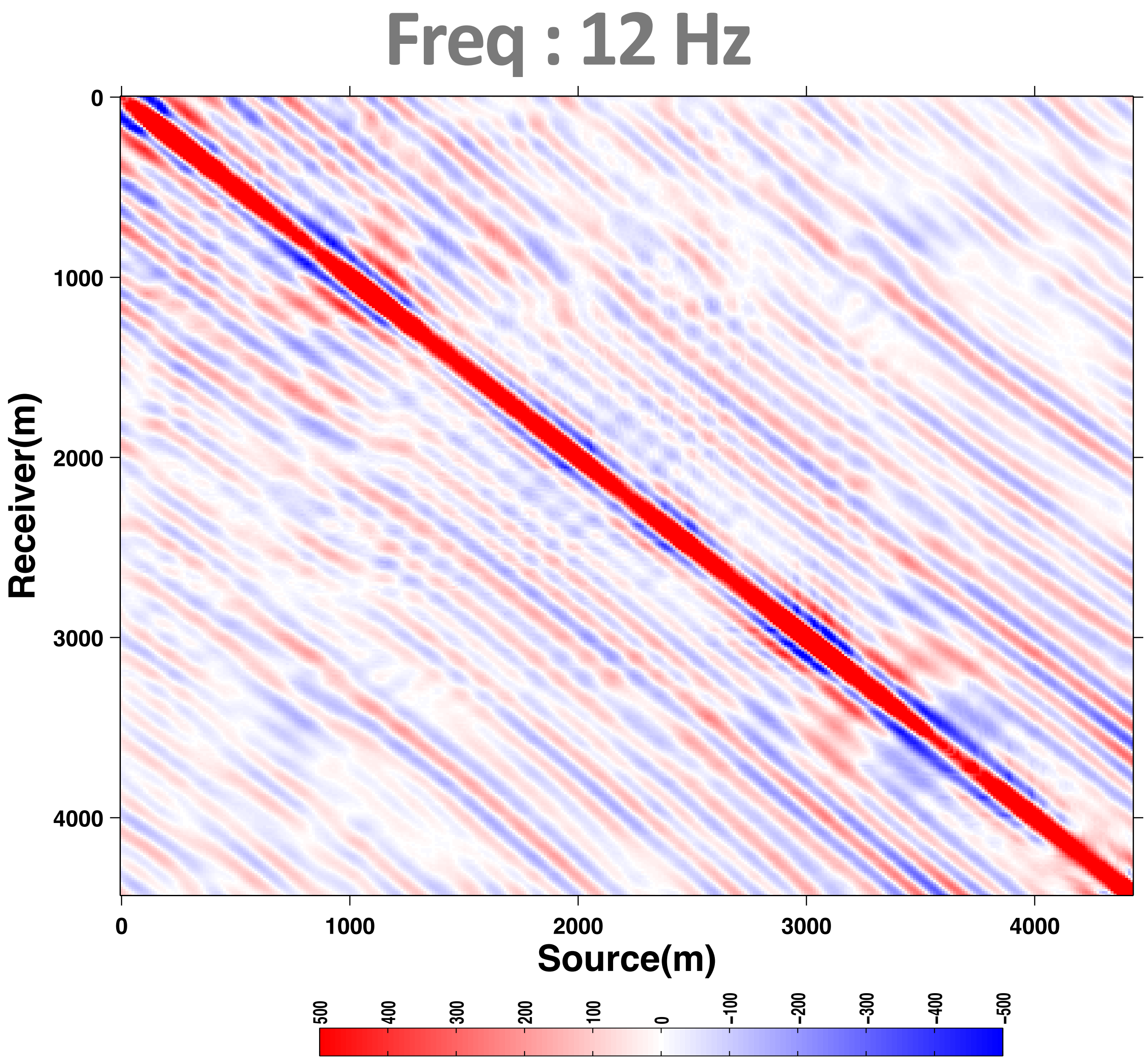
$$\min_X ||X||_* \quad \text{s.t. } ||\mathcal{A}(X) - \mathbf{b}||_2^2 \leq \sigma$$

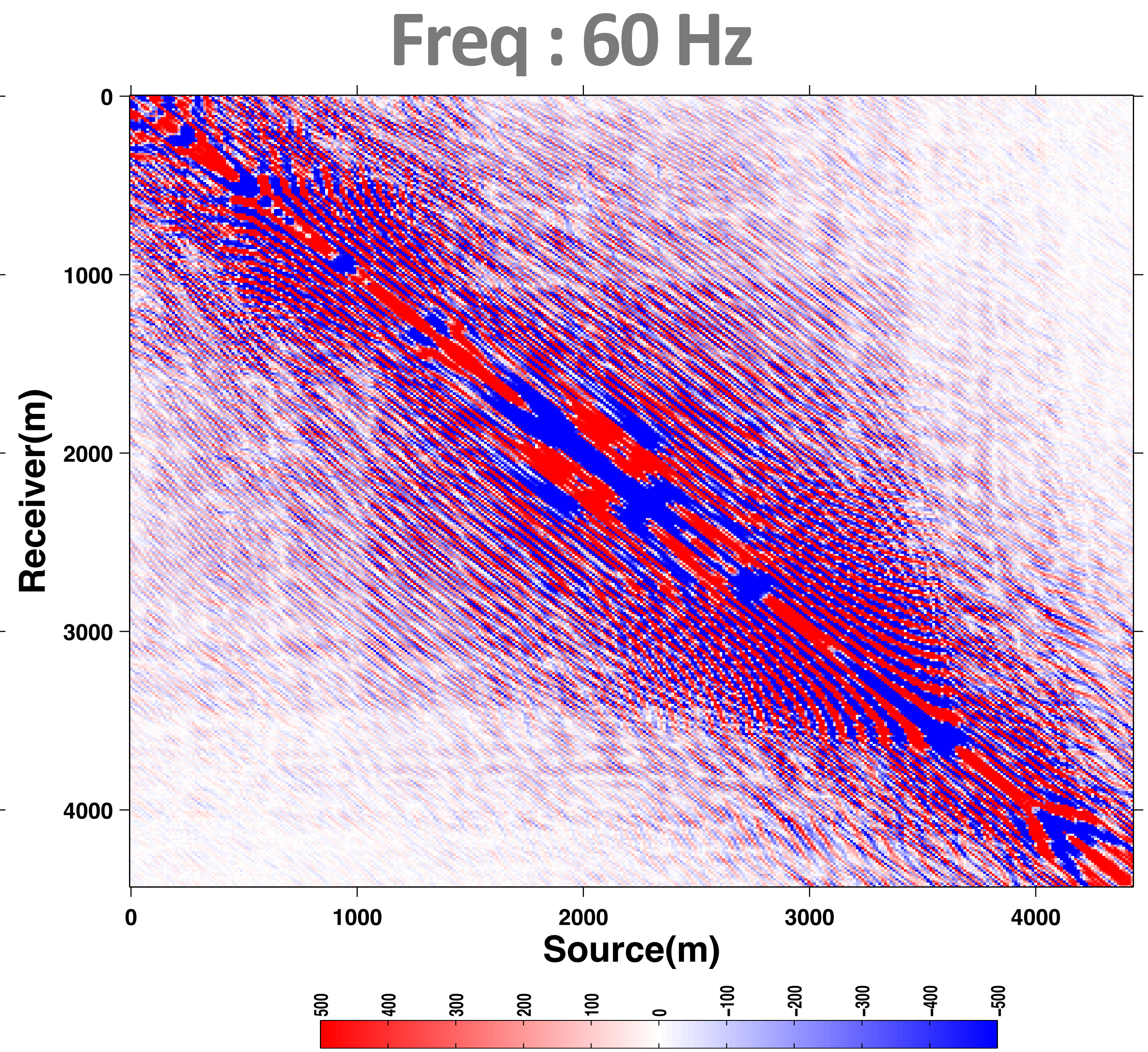
- We use 300 iterations of modified $\mathrm{SPG}\ell 1$ algorithm for the factorized formulation.

- Pareto curve uses the true dual of the $\|\cdot\|_*$ norm, but projection is done using factors.

Experiment & Results

Regular sampled data

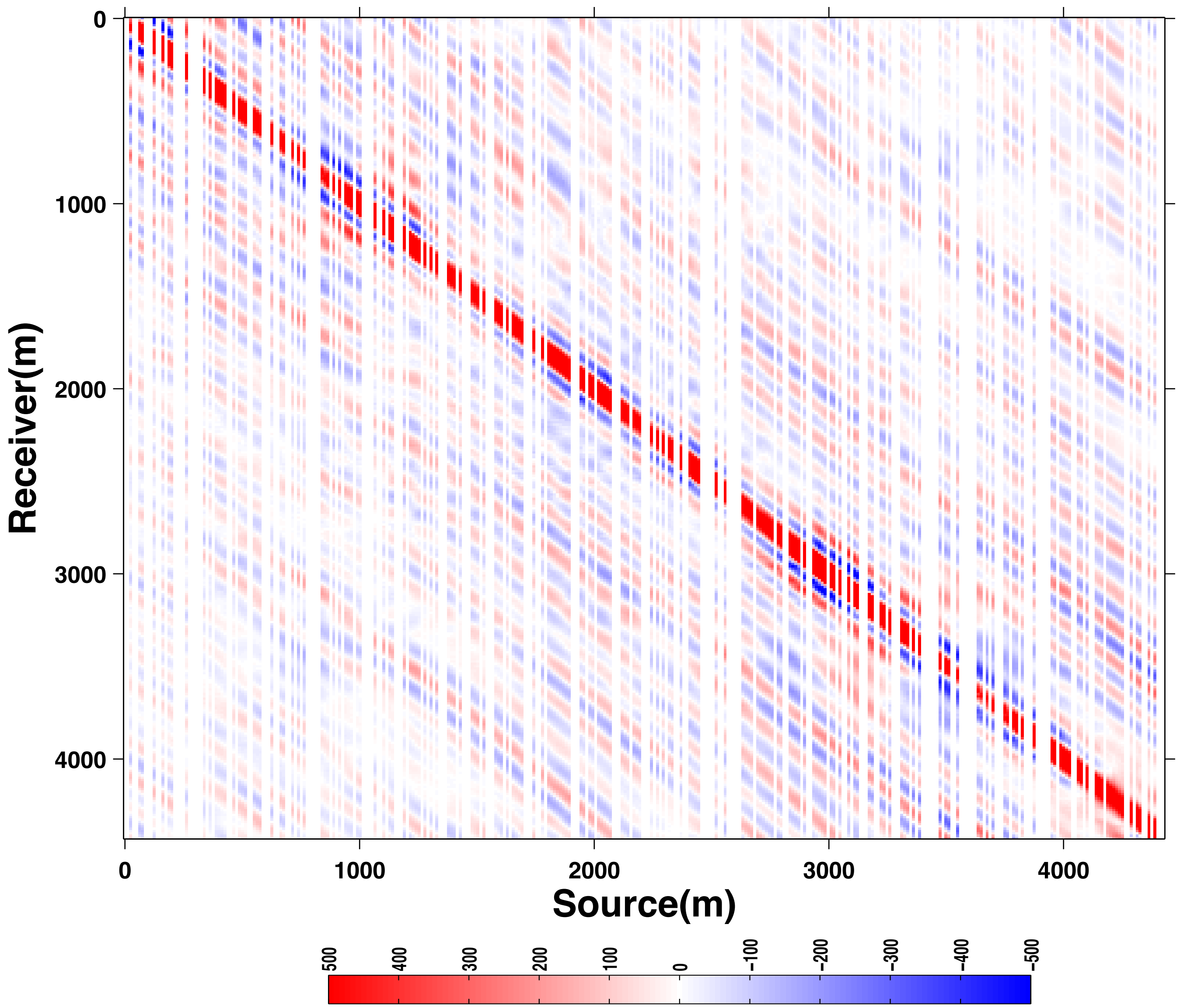




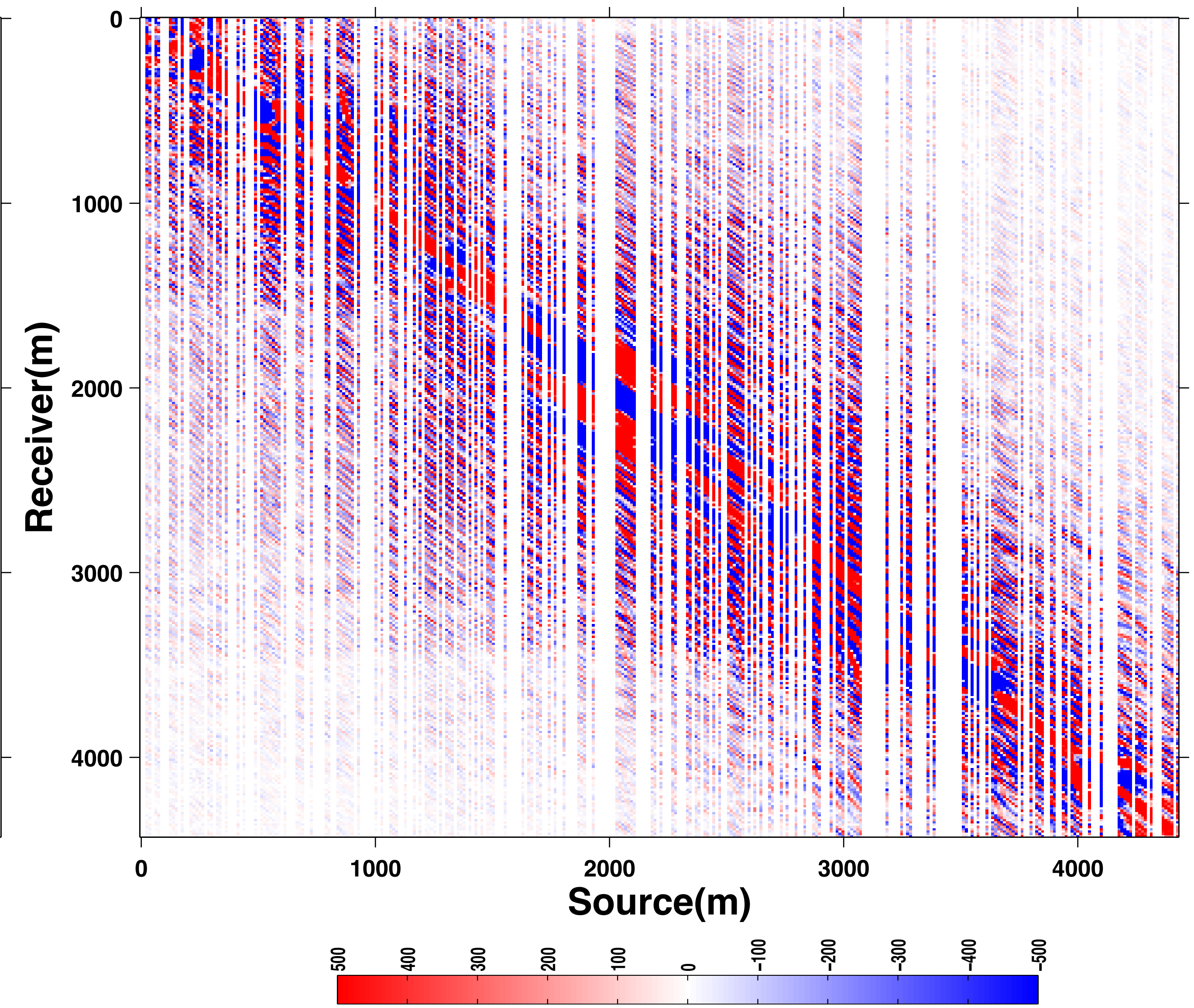
50% Missing Shots

SLIM

Freq : 12 Hz



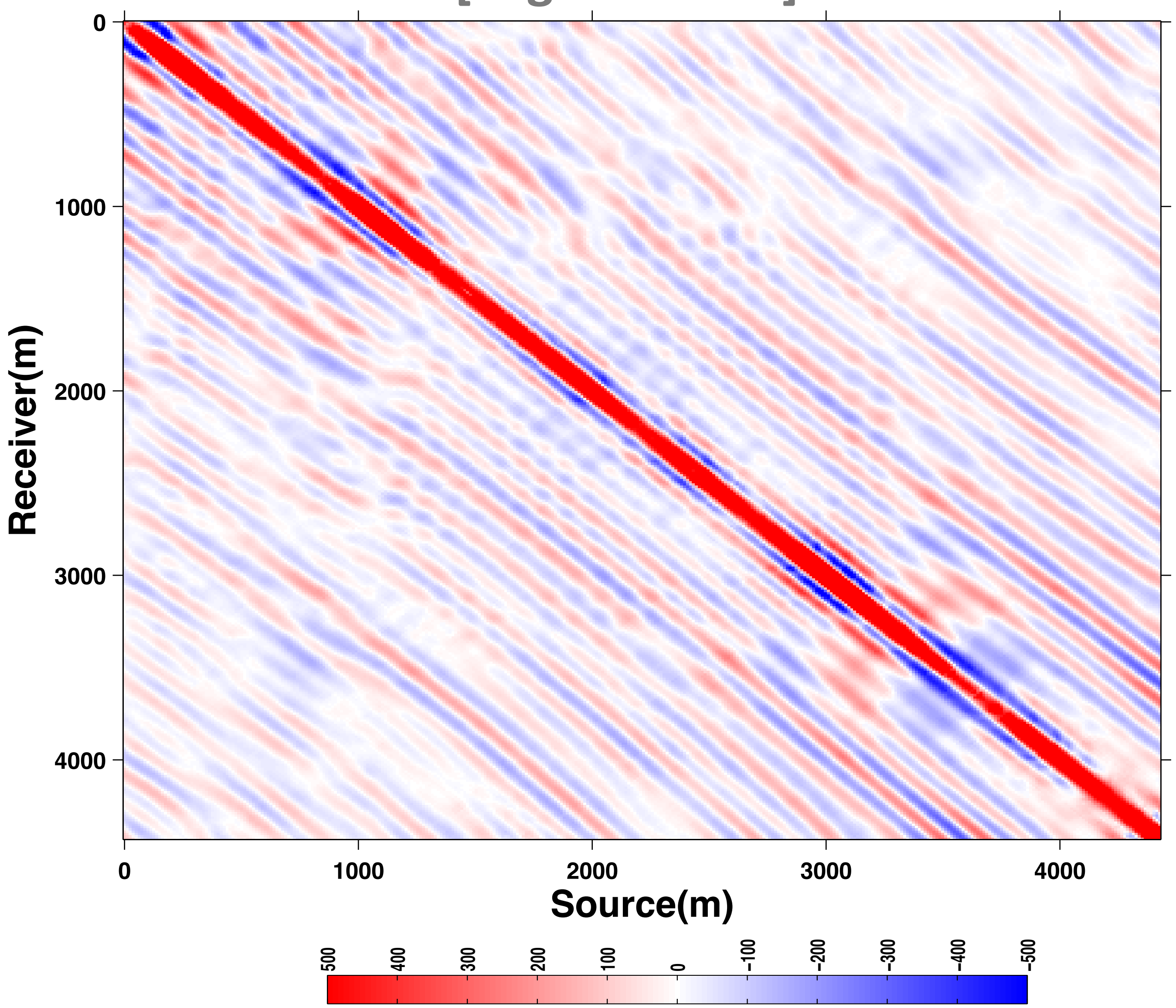
Freq : 60 Hz



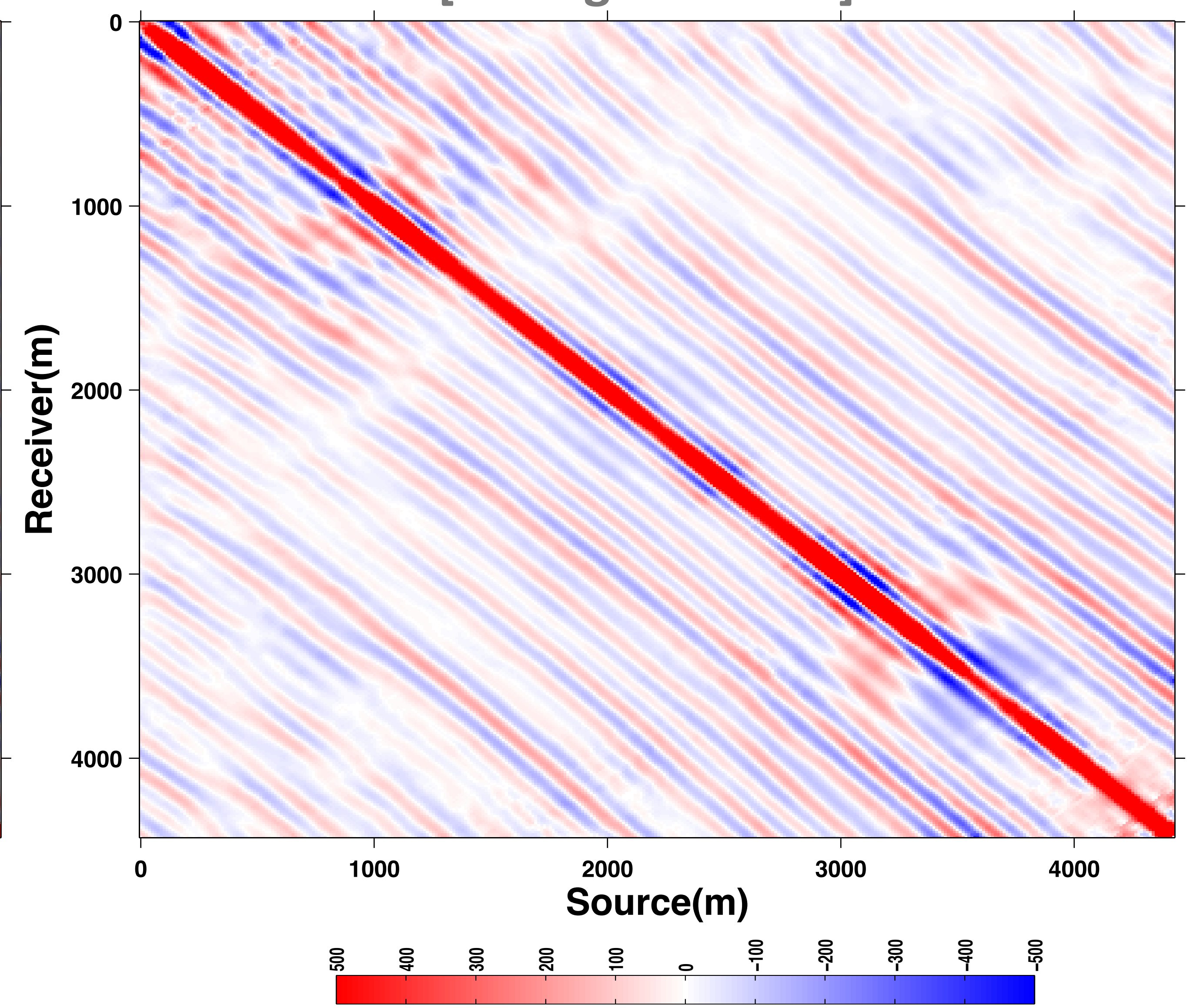
Recovery using MH formulation
[12 Hz]

SNR improvement by 6db

SNR = 19.3 db, Rank=20
[Regularization]



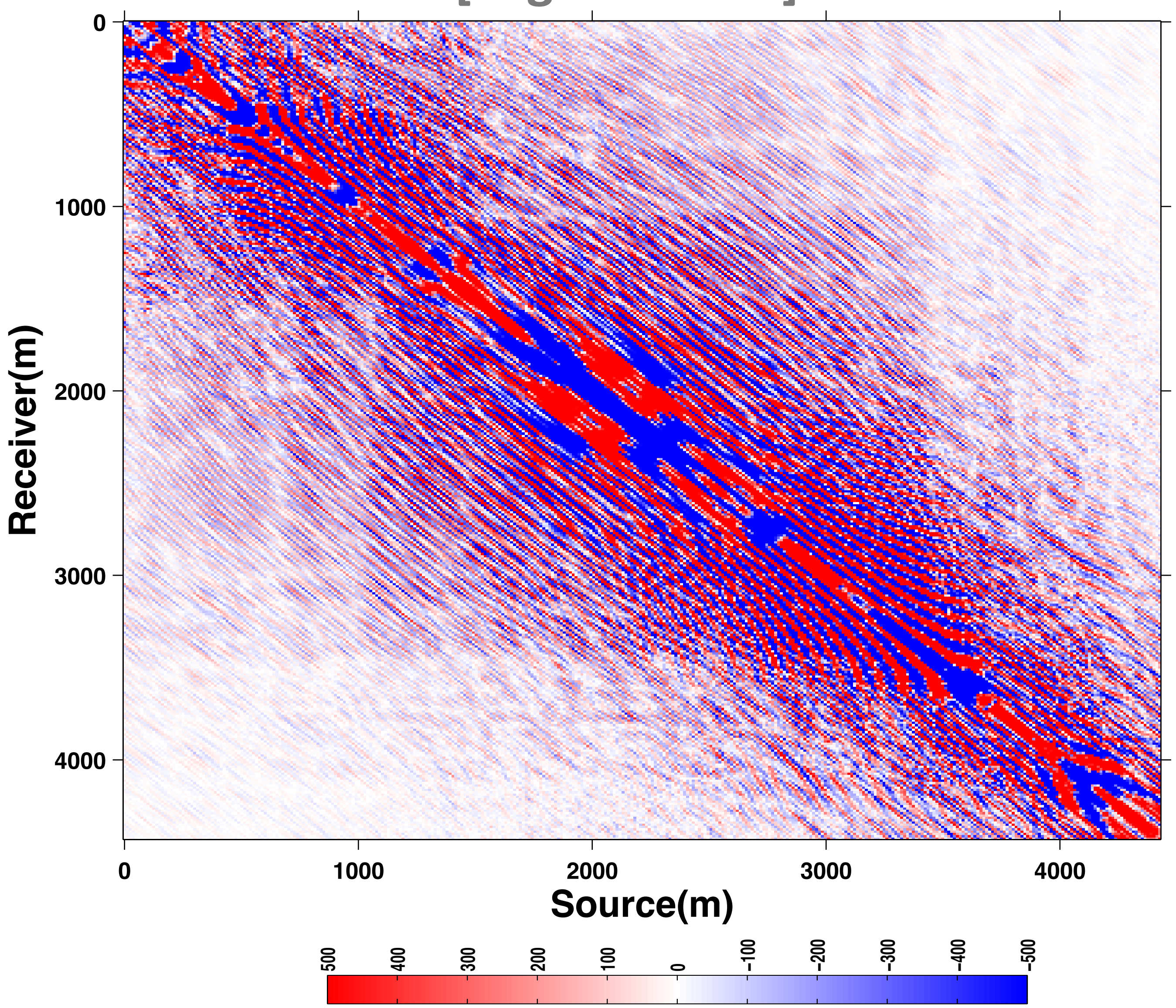
SNR = 13.8 db, Rank=10
[No Regularization]



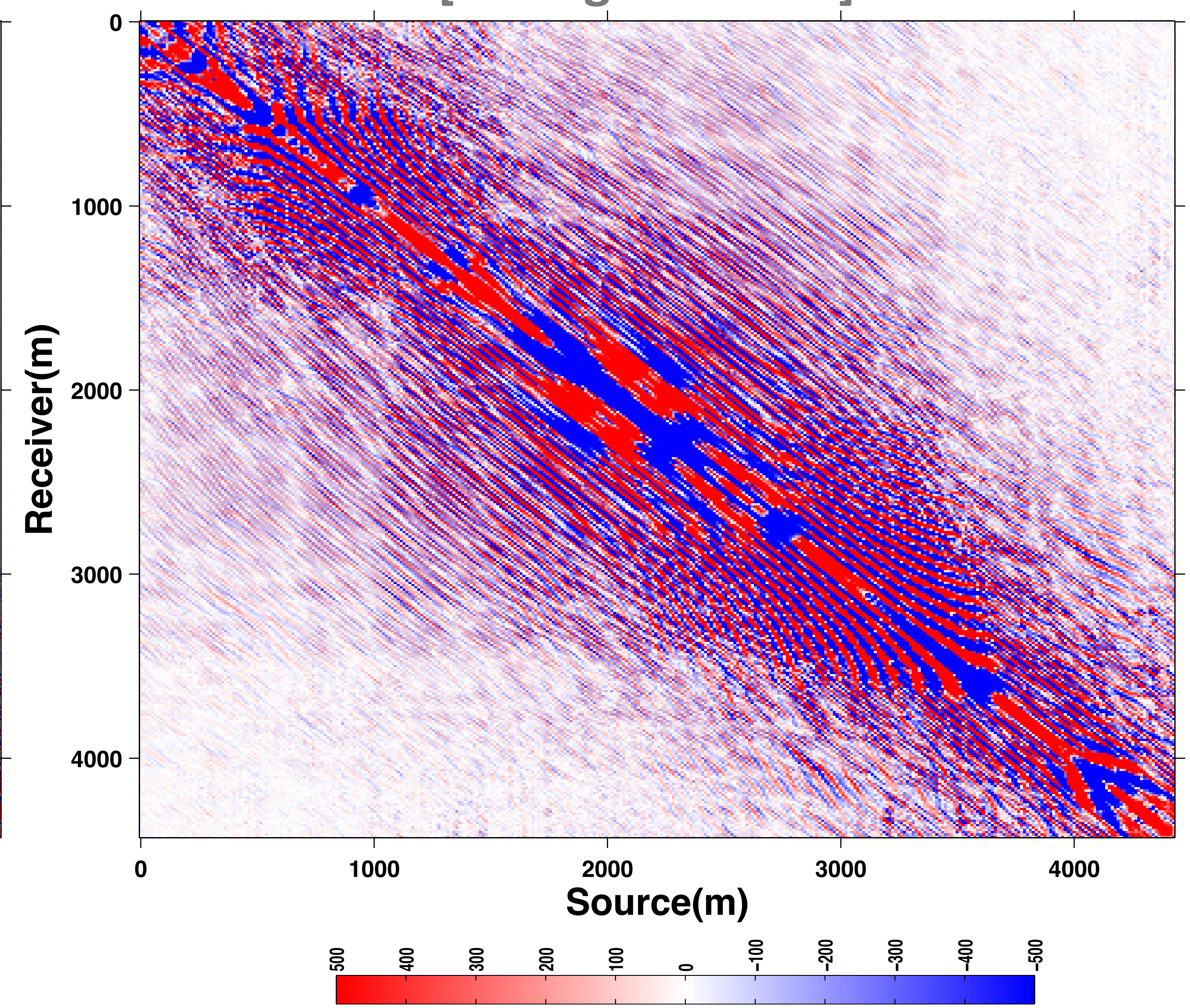
Recovery using MH formulation
[60 Hz]

SNR improvement by 1db

SNR = 17.5 db, Rank=40
[Regularization]

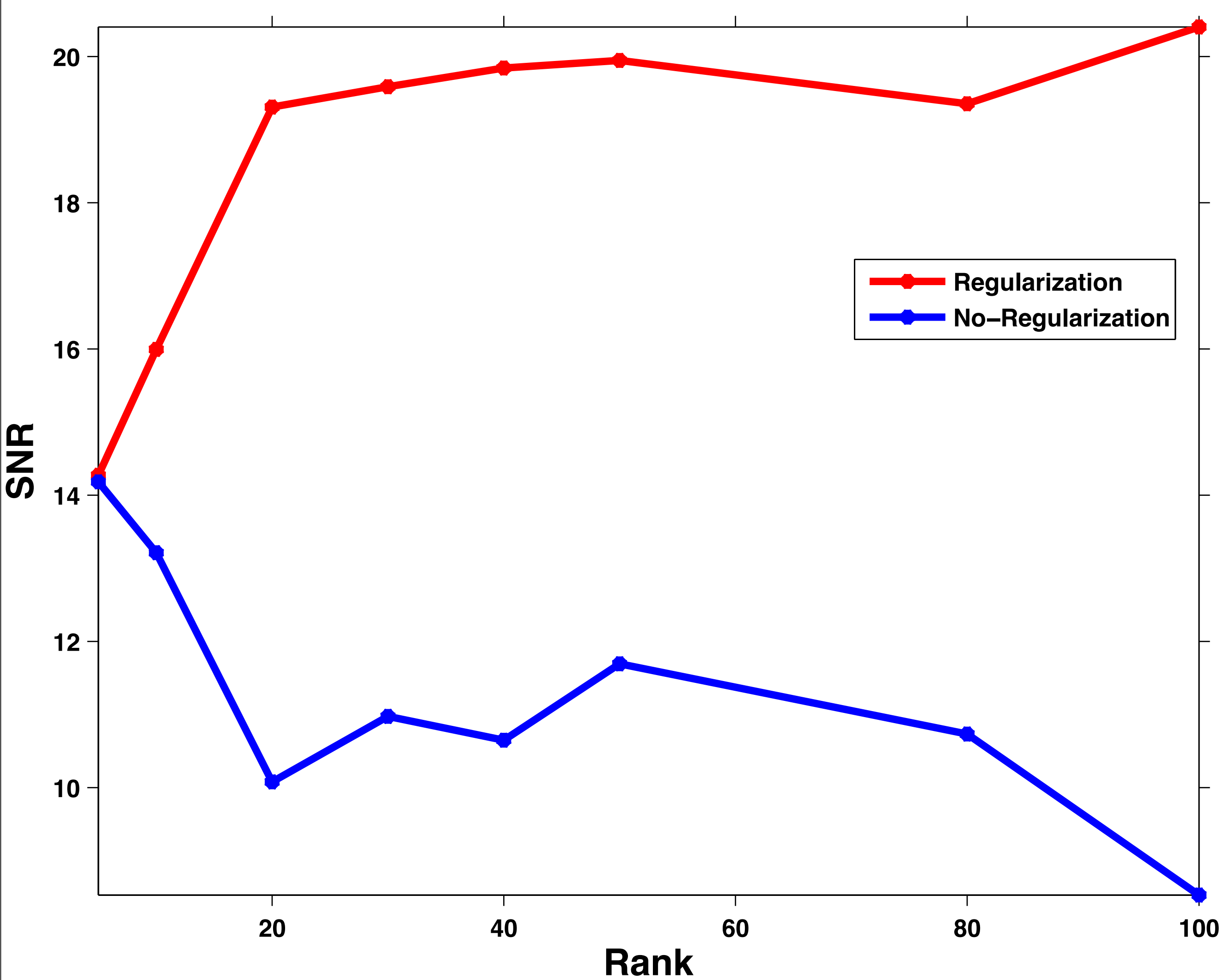


SNR = 16.1 db, Rank=30
[No Regularization]

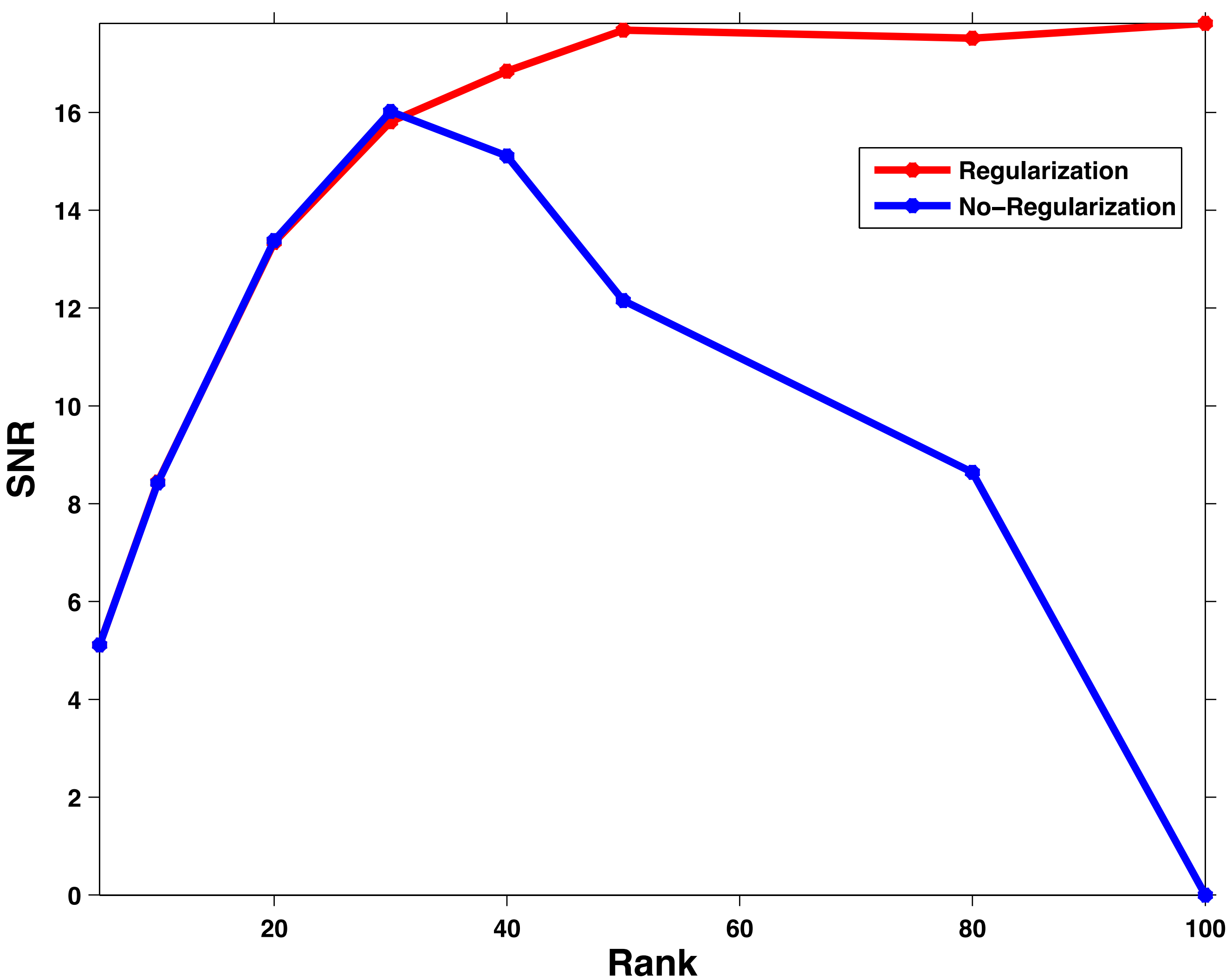


2D Interpolation
[SNR vs Rank]

Higher SNR can be achieved by increasing k



Freq : 12 Hz

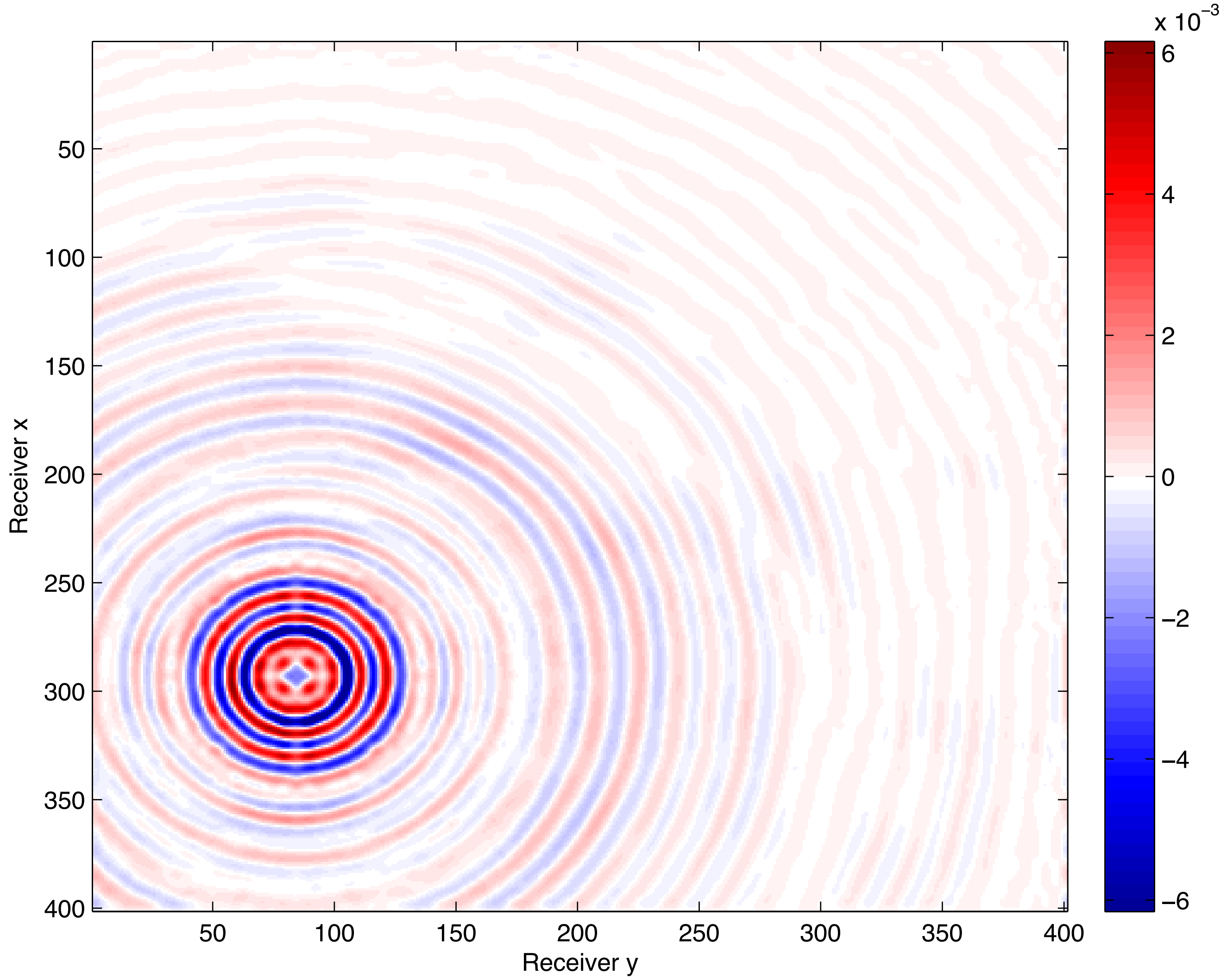


Freq : 60 Hz

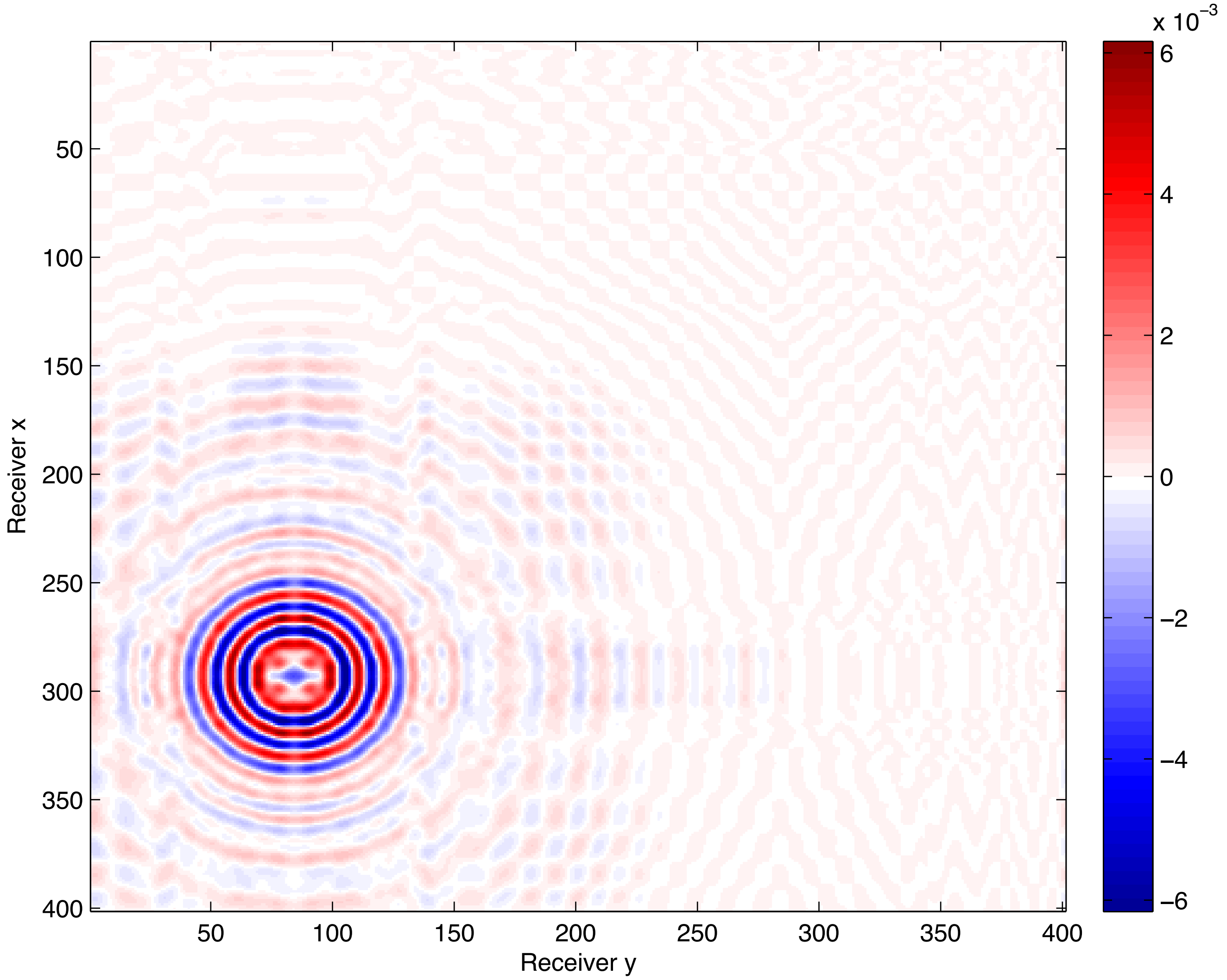
Recovery - Shot Gather

[Rank=10, 500 Iterations]

True data
(SrcX,SrcY = 50,15)

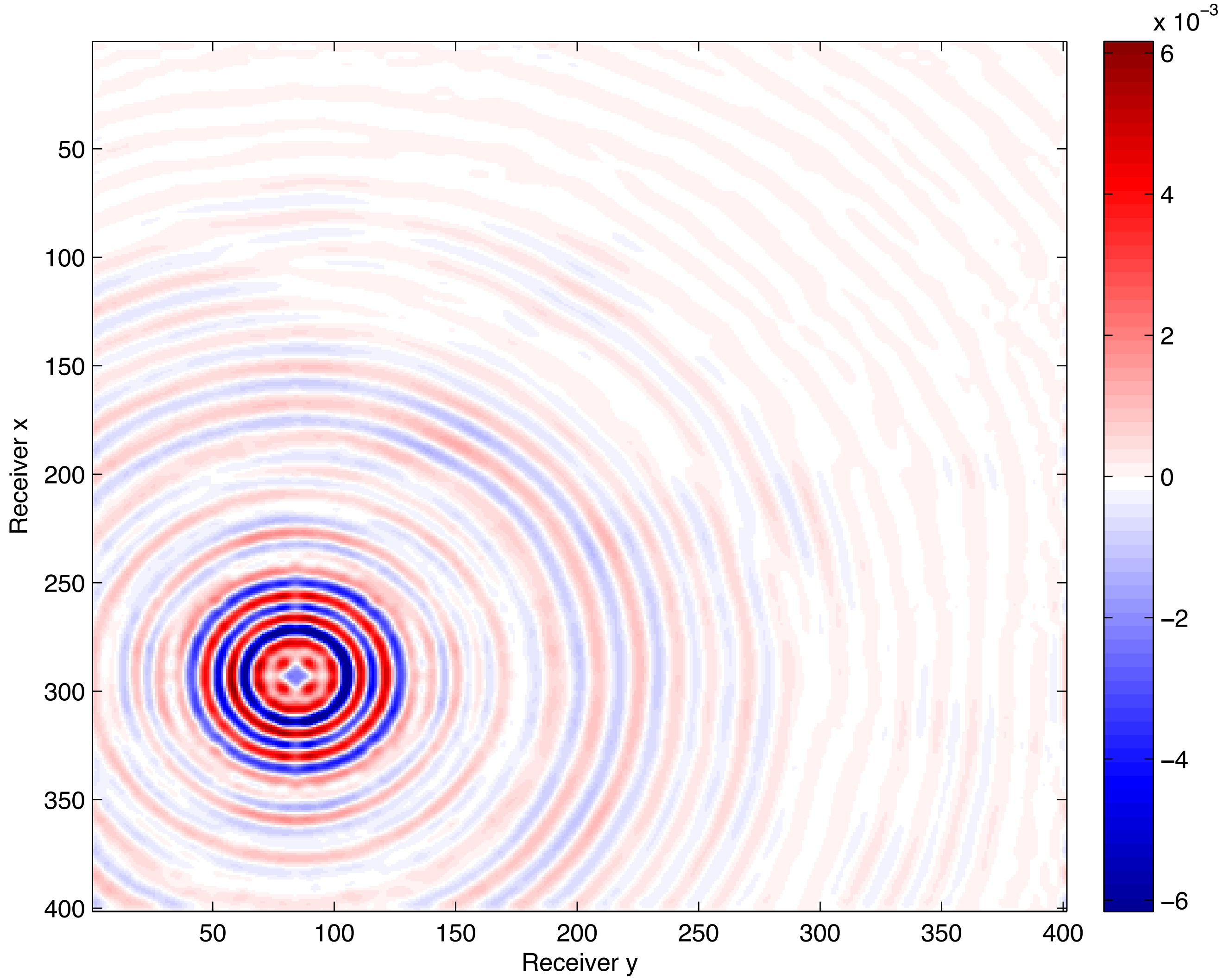


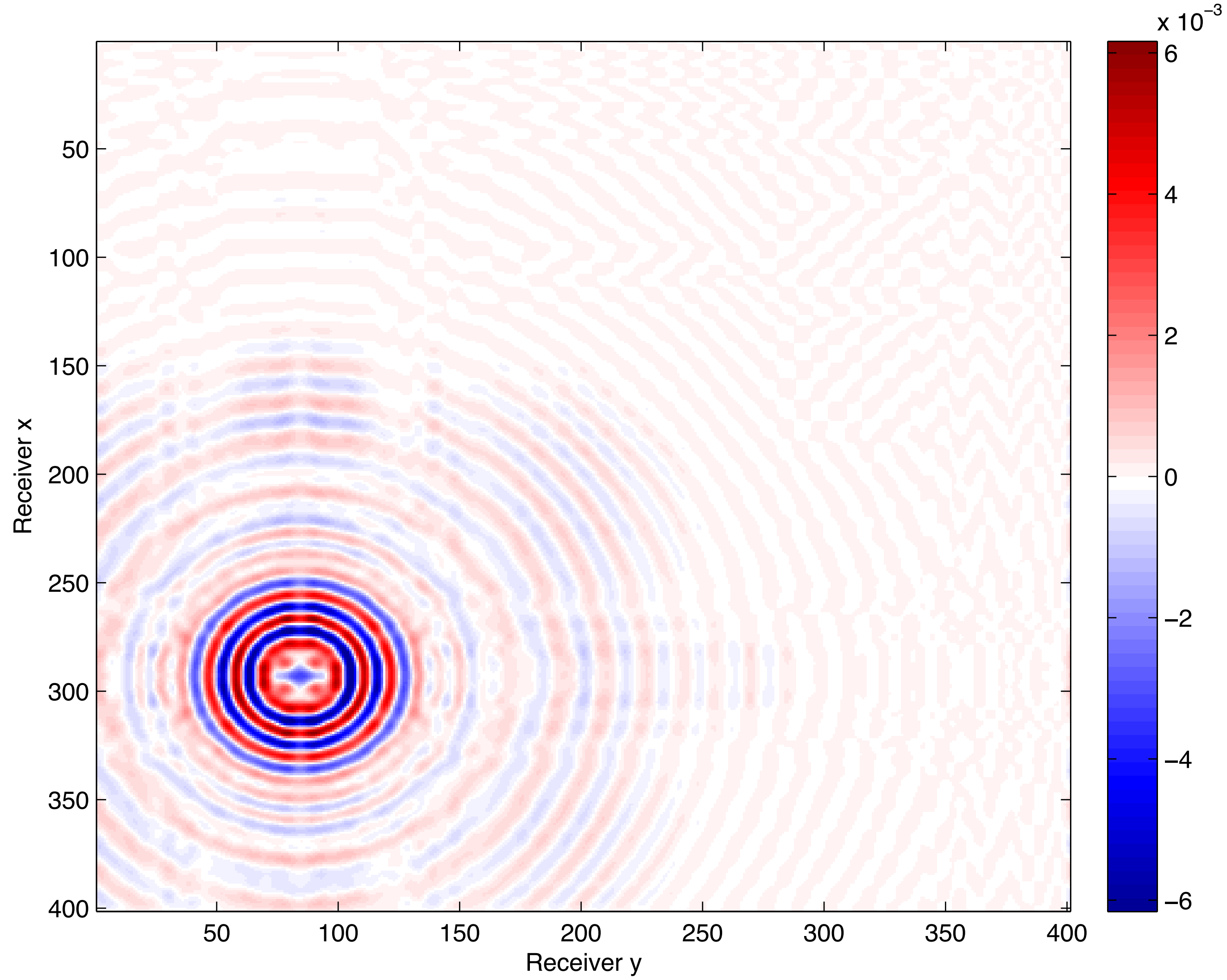
Approximated data
SNR=8.14 db



No-Regularization

Recovery - Shot Gather
[Rank=10, 500 Iterations]

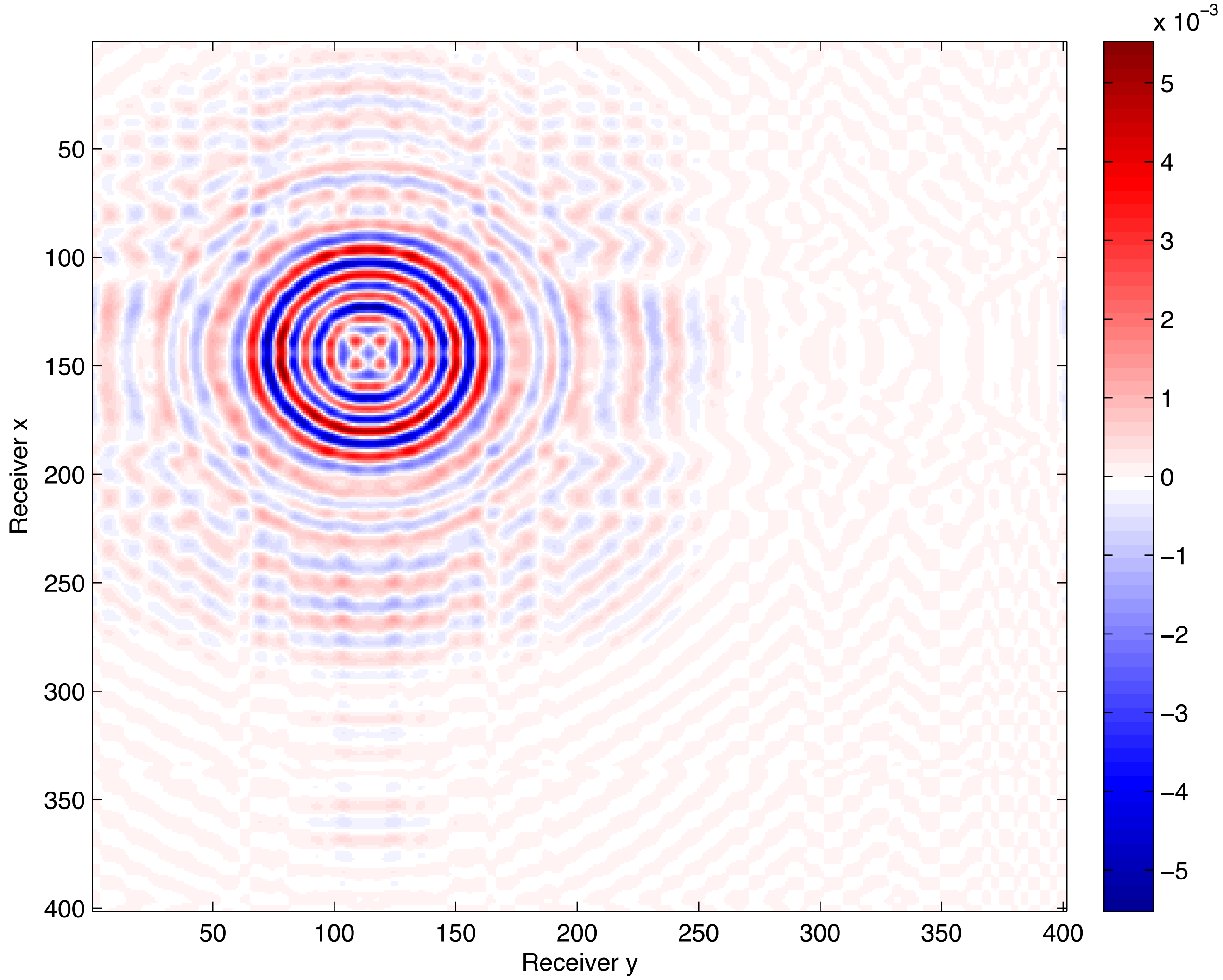




Regularization

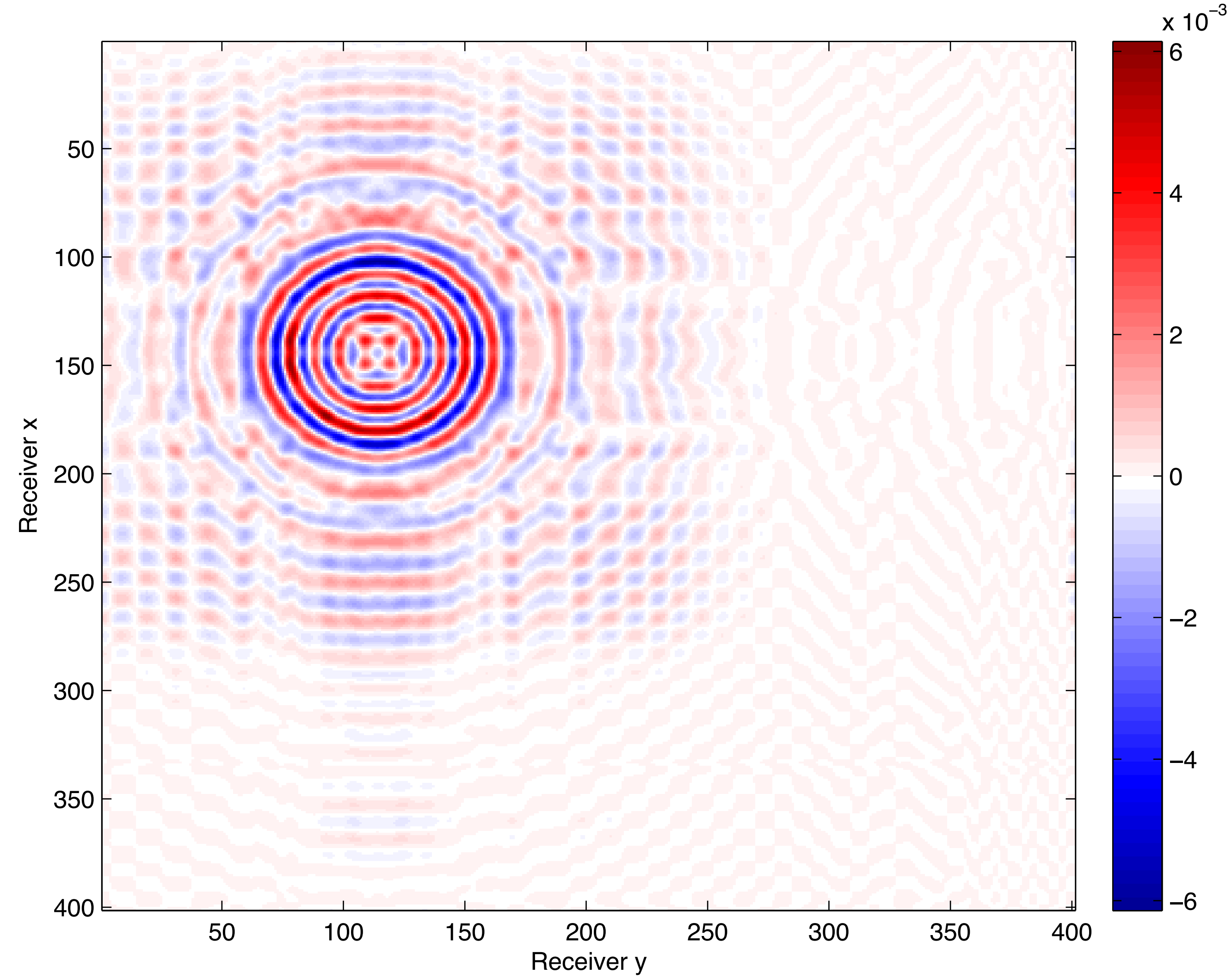
Recovery - Shot Gather
[Rank=10, 500 Iterations]

Interpolated data (SrcX,SrcY = 25,20)



Regularization

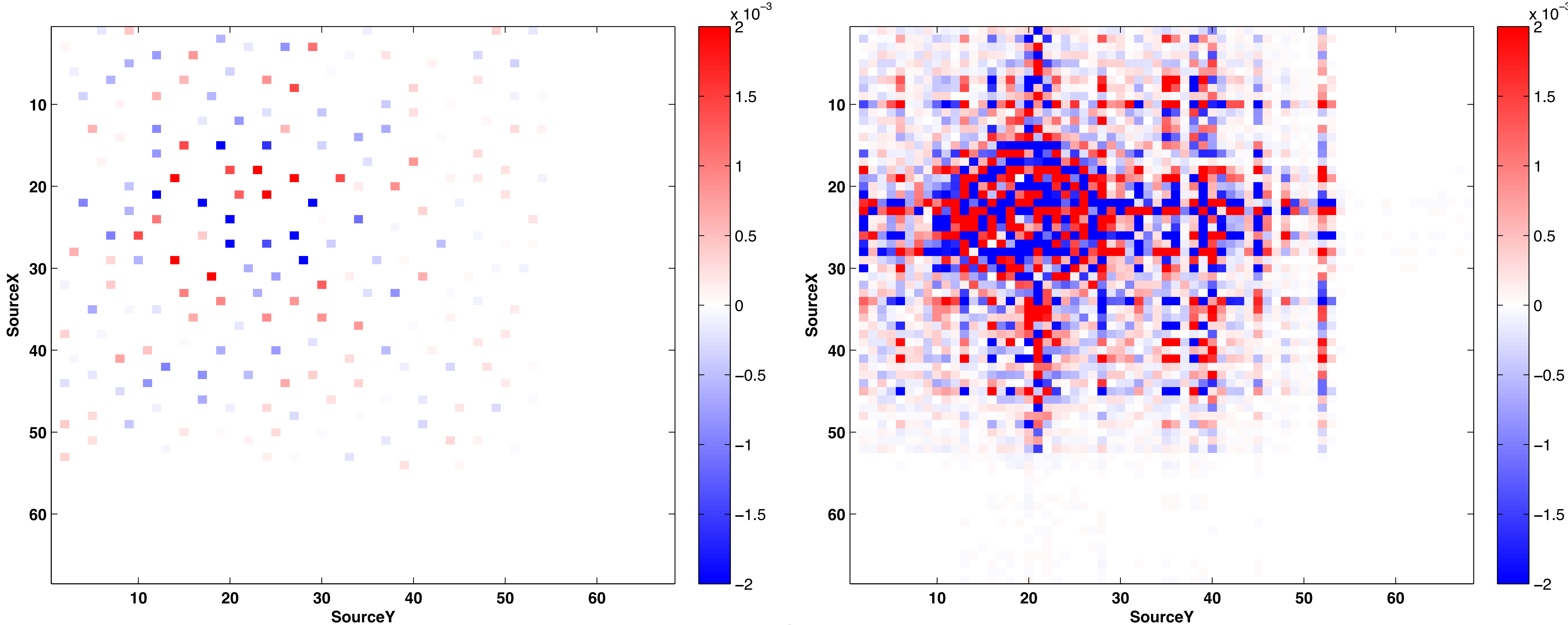
Interpolated data (SrcX,SrcY = 25,20)



No-Regularization

Recovery - Receiver Gather
[Rank=10, 500 Iterations]

Acquired data (RecX,RecY = 35,30) — **Interpolated data (RecX,RecY = 35,30)**



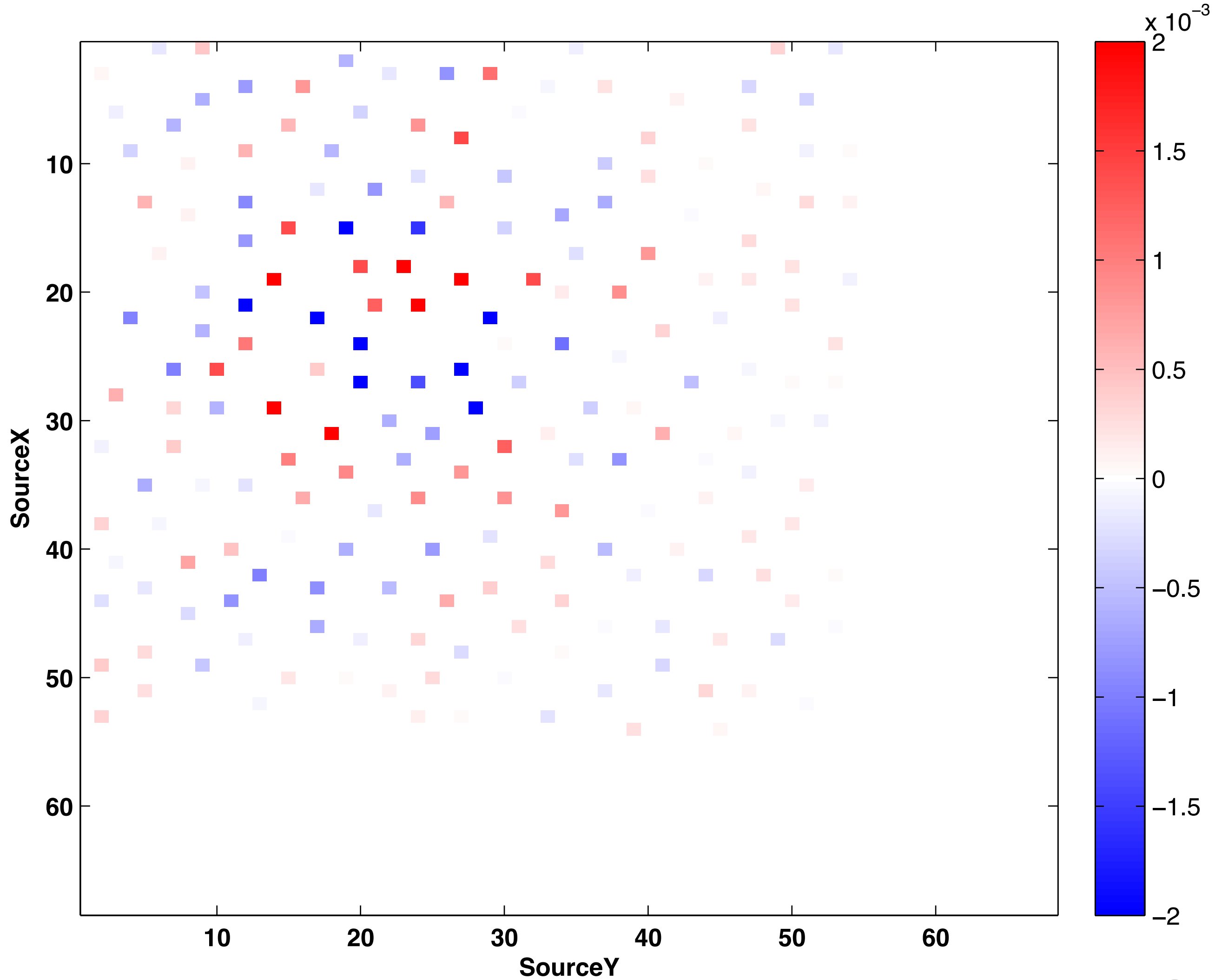
No-Regularization

[98 % missing data]

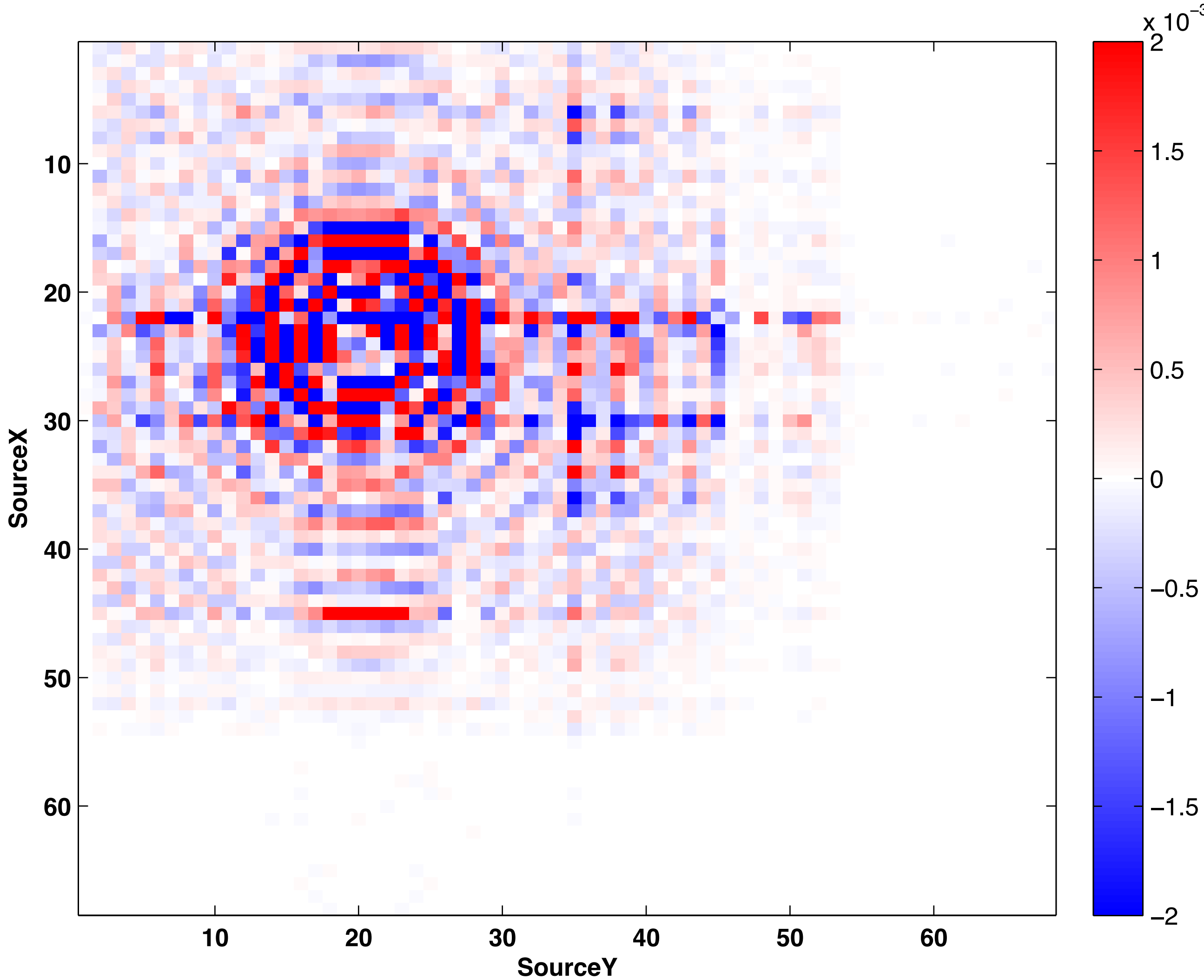
Recovery - Receiver Gather

[Rank=10, 500 Iterations]

Acquired data
(RecX,RecY = 35,30)



Interpolated data
(RecX,RecY = 35,30)



Regularization
[98 % missing data]

Conclusion

- matrix factorization allows *SVD-free* rank penalty methods that work fast on large data
- low-rank structure can be exploited *implicitly* (through factors) and *explicitly* (BPDN)
- low-rank structure holds promise for data recovery and more compact representation.

Future Work

- *weighted* regularized formulation
- Use factorization in *processing*
- implementation of *HSS* structure
- implementation of 3D interpolation in *Midpoint-Offset-Azimuth* domain

References

- **Kumar, R., A. Y. Aravkin, F. J. Herrmann, 2012,** Fast Methods for Rank Minimization with Applications to Seismic-Data Interpolation : TR-2012-04

- **Lee, J., B. Recht, R. Salakhutdinov, N. Srebro, and J. Tropp, 2010**, Practical large- scale optimization for max-norm regularization: Analysis, 23, 1–9.

- **van den Berg, E., and M. P. Friedlander, 2008**, Probing the pareto frontier for basis pursuit solutions: SIAM Journal on Scientific Computing, 31, 890–912.

- **Herrmann, F. J., 2008**, Compressive sampling: a new paradigm for seismic data acquistion and processing?: Presented at the ION.

- **Herrmann, F. J., and G. Hennenfent, 2008**, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, 173, 233–248.

SLIM

Acknowledgements

Thanks everyone for your attention

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.