Rank Minimization via Alternating Optimization

Oscar F. Lopez, Rajiv Kumar

SLIM
University of British Columbia

Motivation

Acquisition challenges

- highly subsampled data

Data exhibits *low-rank* structure

- SVD-free matrix completion

Benefits of Alternating Optimization procedures

- problems become tractable

Outline

Alternating Optimization

Nuclear Norm Minimization
- factorized formulation
- can we benefit from alternating optimization?

Motivation: Variable Decomposition

$$x = x_1 \cdot x_2 \cdot \ldots \cdot x_s$$

Benefits:

- memory efficiency
- computational efficiency

Motivation: Variable Decomposition

$$\boxed{x} = \boxed{x_1} \cdot \boxed{x_2} \cdot \ldots \cdot \boxed{x_s}$$

Benefits:

- memory efficiency
- computational efficiency
- flexibility

SLIM

Xu, Yin. "A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Appications to Nonnegative Tensor Factorization and Completion". SIAM 2013

Alternating Optimization

Want to solve:

$$\min_{x \in \mathcal{X}} f(x_1, x_2, \ldots, x_s) + \sum_{i=1}^{s} r_i(x_i)$$

f is multi-convex

each r_i is convex

Xu, Yin. "A Block Coordinate Descent Method for Regularized Multiconvex Optimization with Appications to Nonnegative Tensor Factorization and Completion". SIAM 2013

Alternating Optimization

Algorithm :

1.Initialization: choose initial point$(x_1^0, x_2^0, \ldots, x_s^0)$

2. for $k = 1, 2, \ldots, T$ do

3. for $i = 1, 2, \ldots, s$ do

4. $x_i^k \leftarrow \arg\min_{x_i \in \mathcal{X}_i^k} f_i^k(x_i) + r_i(x_i)$

5. end for

6. end for

Output: $(x_1^T, x_2^T, \ldots, x_s^T)$

Alternate optimization between factors

Conclusion of Alternating Minimization

Problem becomes tractable

- solve main problem via simpler subproblems

Computationally efficient

- "shown to be superior than other procedures in both speed and quality" **- Xu, Yin**

Outline

Alternating Optimization

Nuclear Norm Minimization

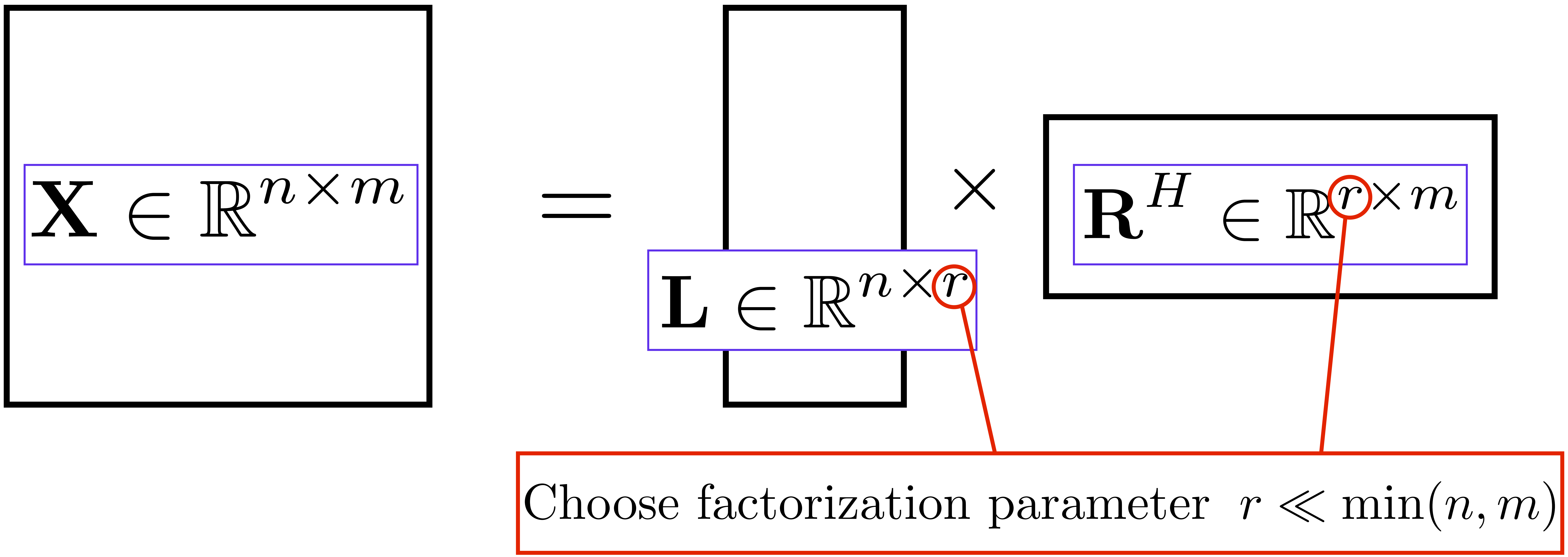
- factorized formulation
- can we benefit from alternating optimization?

Matrix Factorization

$$\mathbf{X} \in \mathbb{R}^{n \times m} \quad = \quad \mathbf{L} \in \mathbb{R}^{n \times r} \quad \times \quad \mathbf{R}^H \in \mathbb{R}^{r \times m}$$

$$\mathbf{X} = \mathbf{L}\mathbf{R}^{\mathbf{H}}$$

Matrix Factorization



$$\mathbf{X} = \mathbf{L}\mathbf{R}^{\mathbf{H}}$$

[Rennie and Srebro 2005]

Nuclear Norm Minimization- Factorized Form

$$\mathbf{X} = \mathbf{L}\mathbf{R}^H$$

- Nuclear norm is bounded by

$$||\mathbf{X}||_* \leq \frac{\mathbf{1}}{\mathbf{2}}(||\mathbf{L}||^{\mathbf{2}}_{\mathbf{F}} + ||\mathbf{R}||^{\mathbf{2}}_{\mathbf{F}})$$

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

- choose r *explicitly* & avoid costly *SVD's*

Nuclear Norm Minimization- Factorized Form

We want to solve

$$\min_{\mathbf{L}\in\mathbb{R}^{n\times r},\mathbf{R}\in\mathbb{R}^{m\times r}} \frac{1}{2}(||\mathbf{L}||_F^2 + ||\mathbf{R}||_F^2) \;\; \text{s.t.} \;\; ||\mathcal{A}(\mathbf{L}\mathbf{R}^H) - \mathbf{b}||_{\mathbf{F}}^{\mathbf{2}} \leq \sigma$$

Nuclear Norm Minimization- Factorized Form

We want to solve

$$\min_{\mathbf{L}\in\mathbb{R}^{n\times r},\mathbf{R}\in\mathbb{R}^{m\times r}} \frac{1}{2}(\underbrace{||\mathbf{L}||_F^2}_{r_1} + \underbrace{||\mathbf{R}||_F^2}_{r_2}) \quad \text{s.t.} \quad \underbrace{||\mathcal{A}(\mathbf{L}\mathbf{R}^H) - \mathbf{b}||_\mathbf{F}^\mathbf{2}}_{f} \leq \sigma$$

convex functions

multi-convex function

Nuclear Norm Minimization- Factorized Form

We want to solve

$$\min_{\mathbf{L}\in\mathbb{R}^{n\times r},\mathbf{R}\in\mathbb{R}^{m\times r}} \frac{1}{2}(\underbrace{||\mathbf{L}||_F^2}_{r_1} + \underbrace{||\mathbf{R}||_F^2}_{r_2}) \;\text{ s.t. }\; \underbrace{||\mathcal{A}(\mathbf{LR}^H) - \mathbf{b}||_\mathbf{F}^\mathbf{2}}_{f} \leq \sigma$$

convex functions

multi-convex function

Let's Alternate!

Alternating Nuclear Norm Minimization

1. Input: $\mathcal{A}, \mathbf{b}$
2. Initialize: $\mathbf{L}^0$ to be the top-r left singular vectors of $\mathbf{b}$
3. for $t = 0, ..., T-1$ do
4.
$$\mathbf{R}^{t+1} = \min_{\mathbf{R} \in \mathbb{R}^{m \times r}} \frac{1}{2}||\mathbf{R}||_F^2 \quad \text{s.t.} \quad ||\mathcal{A}(\mathbf{L}^t \mathbf{R}^H) - \mathbf{b}||_F^2 \leq \sigma_t$$
5.
$$\mathbf{L}^{t+1} = \min_{\mathbf{L} \in \mathbb{R}^{n \times r}} \frac{1}{2}||\mathbf{L}||_F^2 \quad \text{s.t.} \quad ||\mathcal{A}(\mathbf{L}(\mathbf{R}^{t+1})^H) - \mathbf{b}||_F^2 \leq \sigma_t$$
6. end for
7. Return $\tilde{\mathbf{X}} = \mathbf{L}^T (\mathbf{R}^T)^H$

Implementation

$$\min_{\mathbf{R}\in\mathbb{R}^{m\times r}} \frac{1}{2}||\mathbf{R}||_F^2 \quad \text{s.t. } ||\mathcal{A}(\mathbf{L}^t\mathbf{R}^H) - \mathbf{b}||_F^2 \leq \sigma_t$$

- approximately solve a series of $LASSO_\tau$ formulation

$$v(\tau) = \min_{\mathbf{R}} ||\mathcal{A}(\mathbf{L}^t\mathbf{R}^H) - \mathbf{b}||_F^2 \quad \text{s.t. } \frac{1}{2}||\mathbf{R}||_F^2 \leq \tau$$

where τ is the regularization parameter

Experiments: Nelson 2-D seismic line

1024 x 401 x 401 matrix

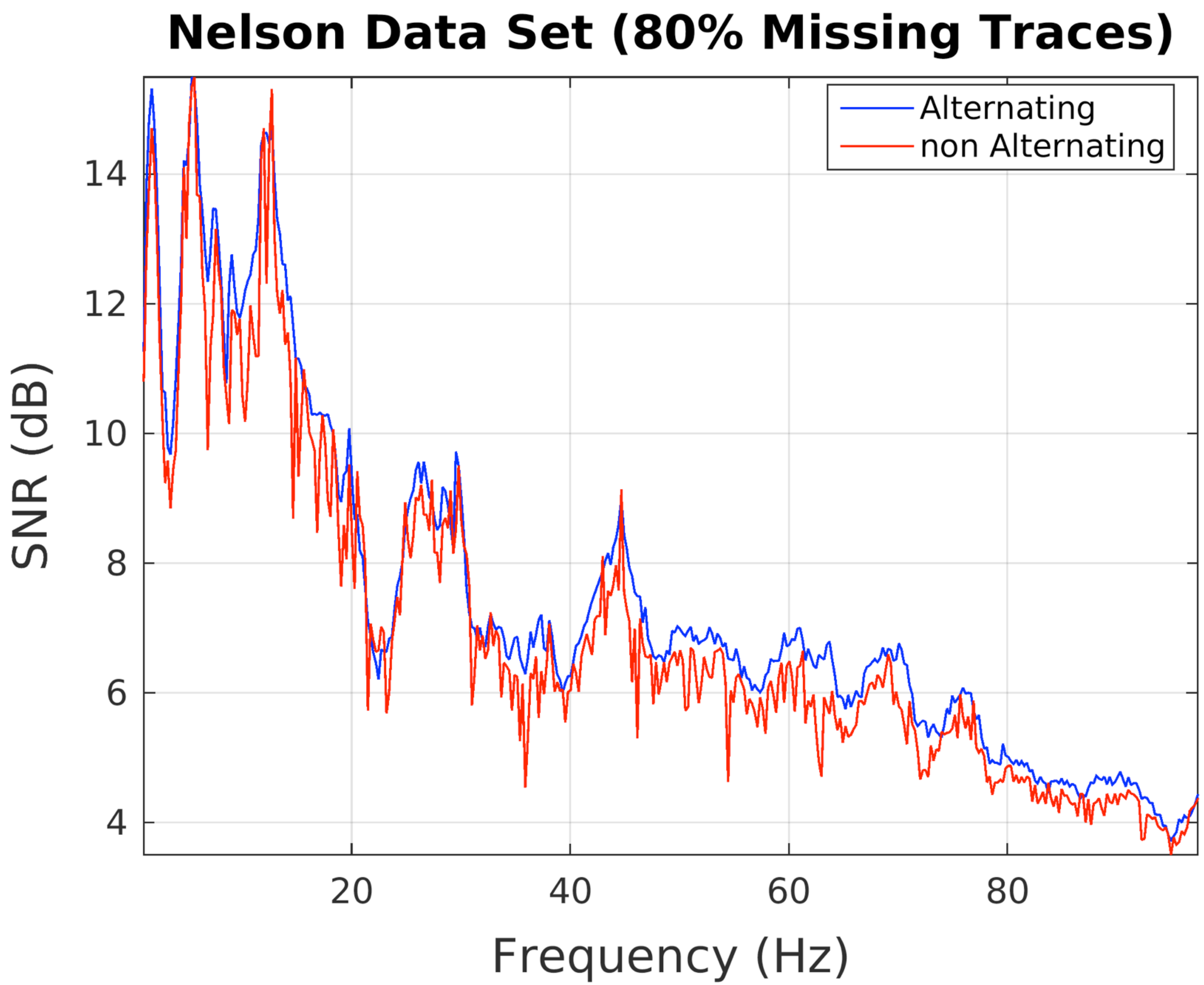
- 80% missing traces
- Factorization parameter adjusted from low to high frequency
- Jittered subsampling
- solve with SPGL1

Compare: Alternating vs Non Alternating

- 6 alternations, 15 iterations per alternation
- 180 iterations total for both

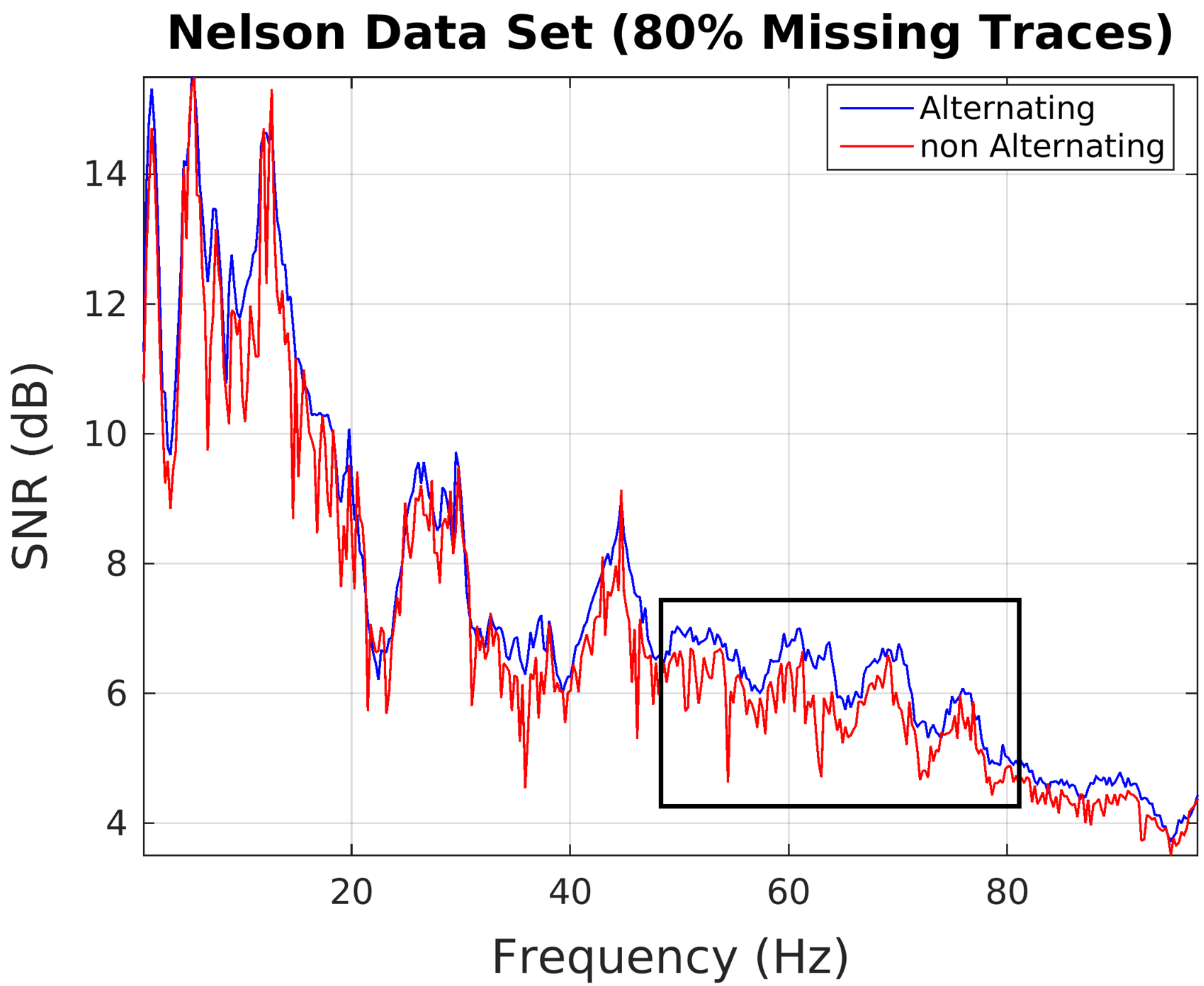
Experiments

Average Improvement
.5 dB



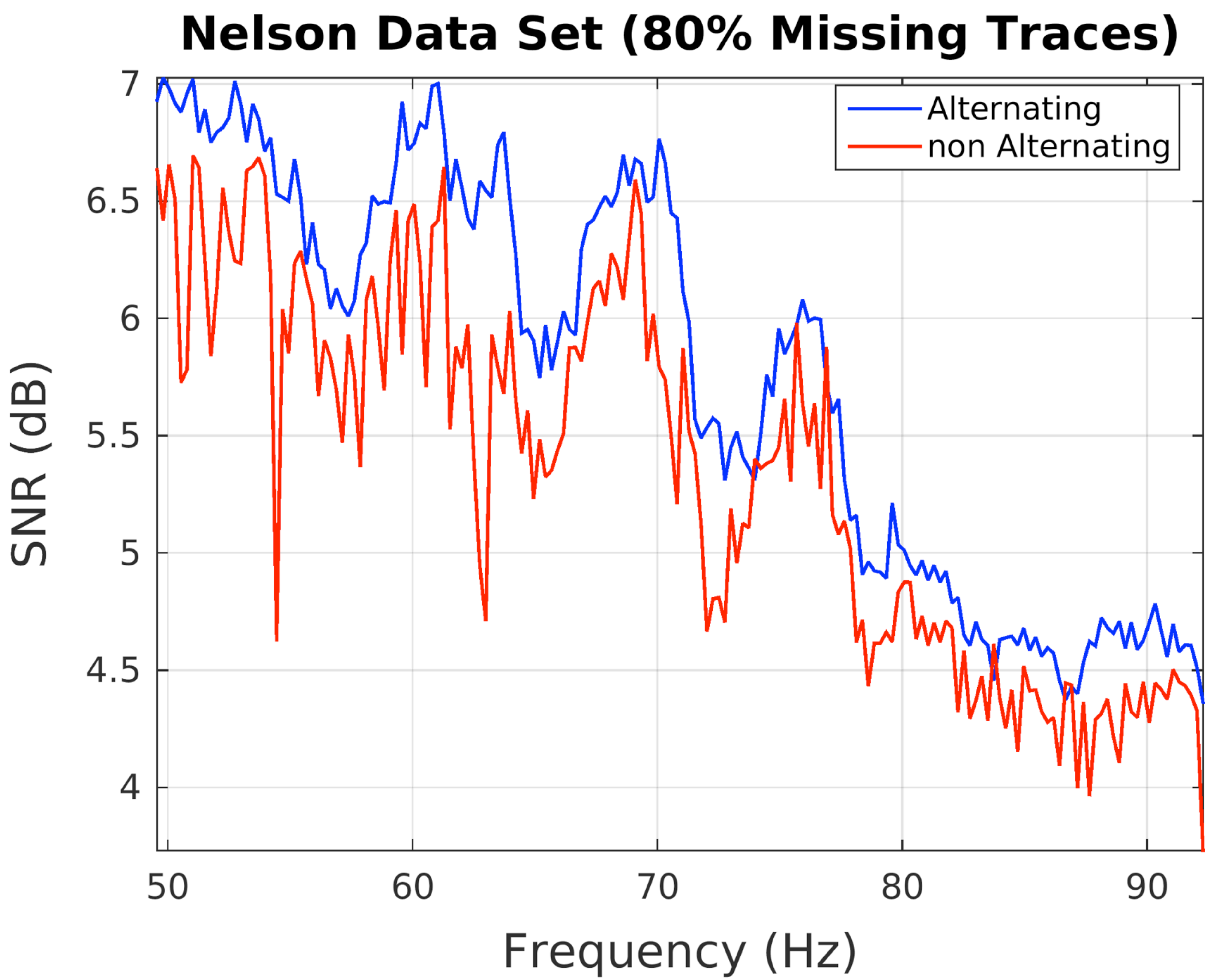
Experiments

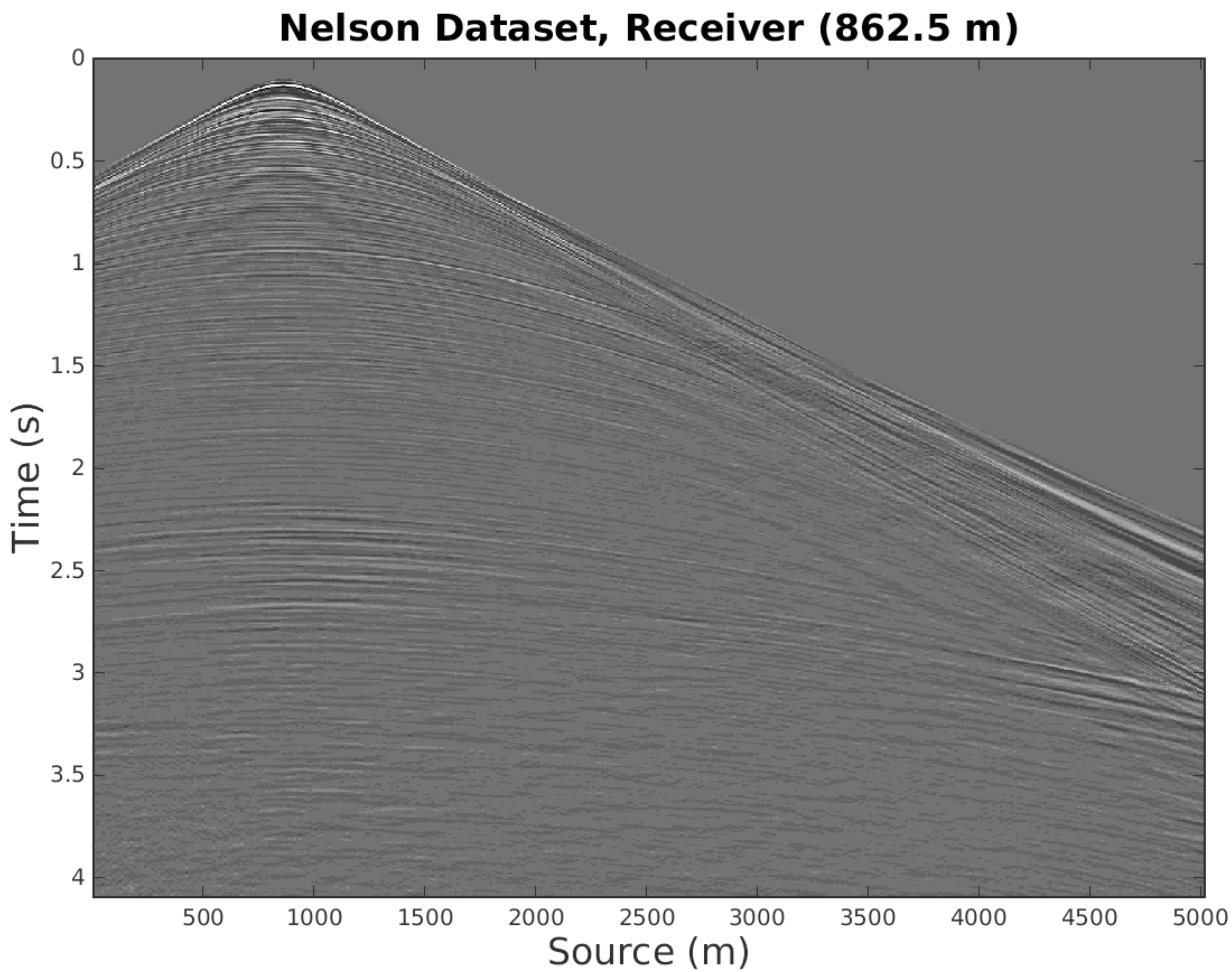
Average Improvement
.5 dB

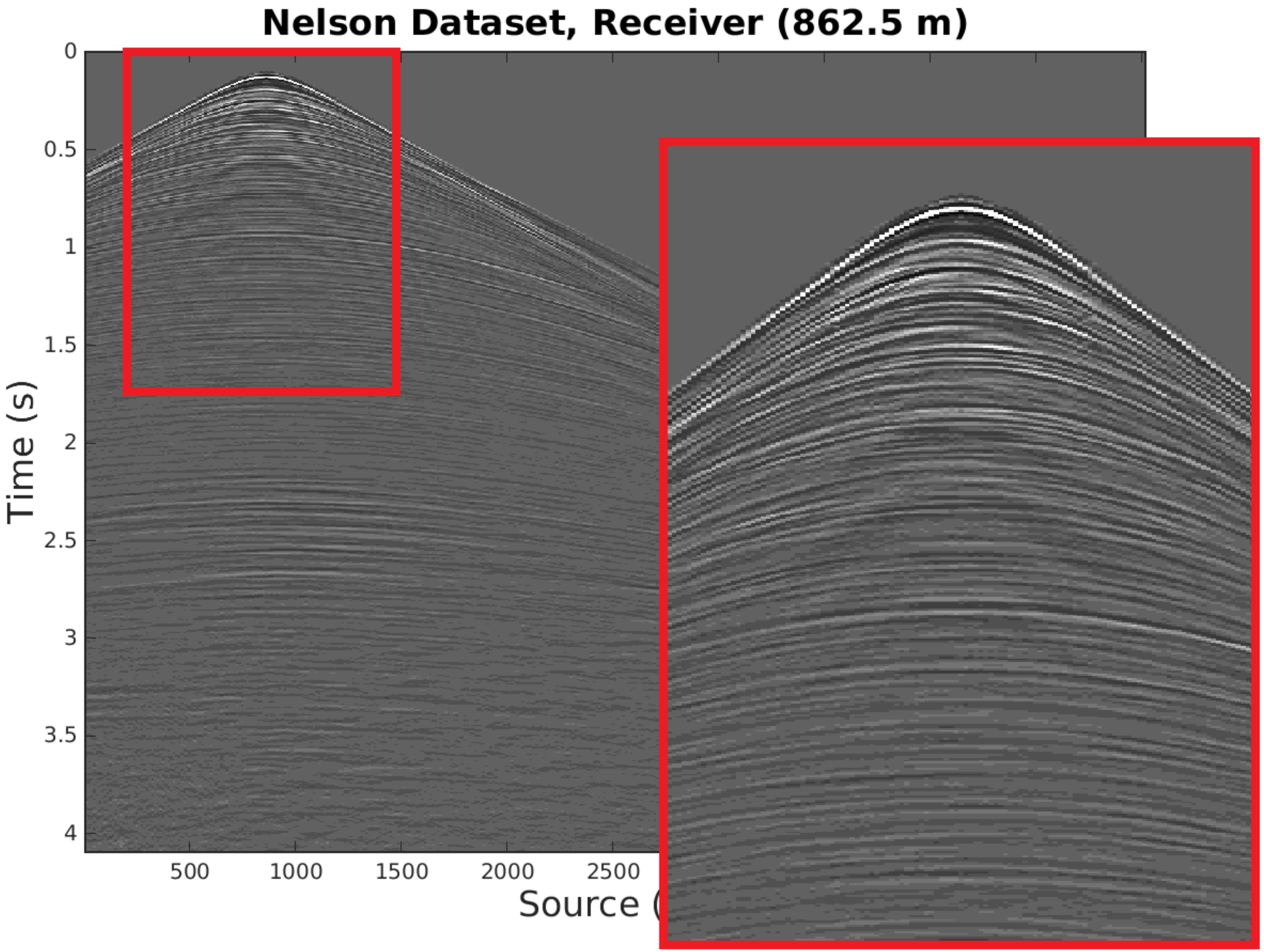


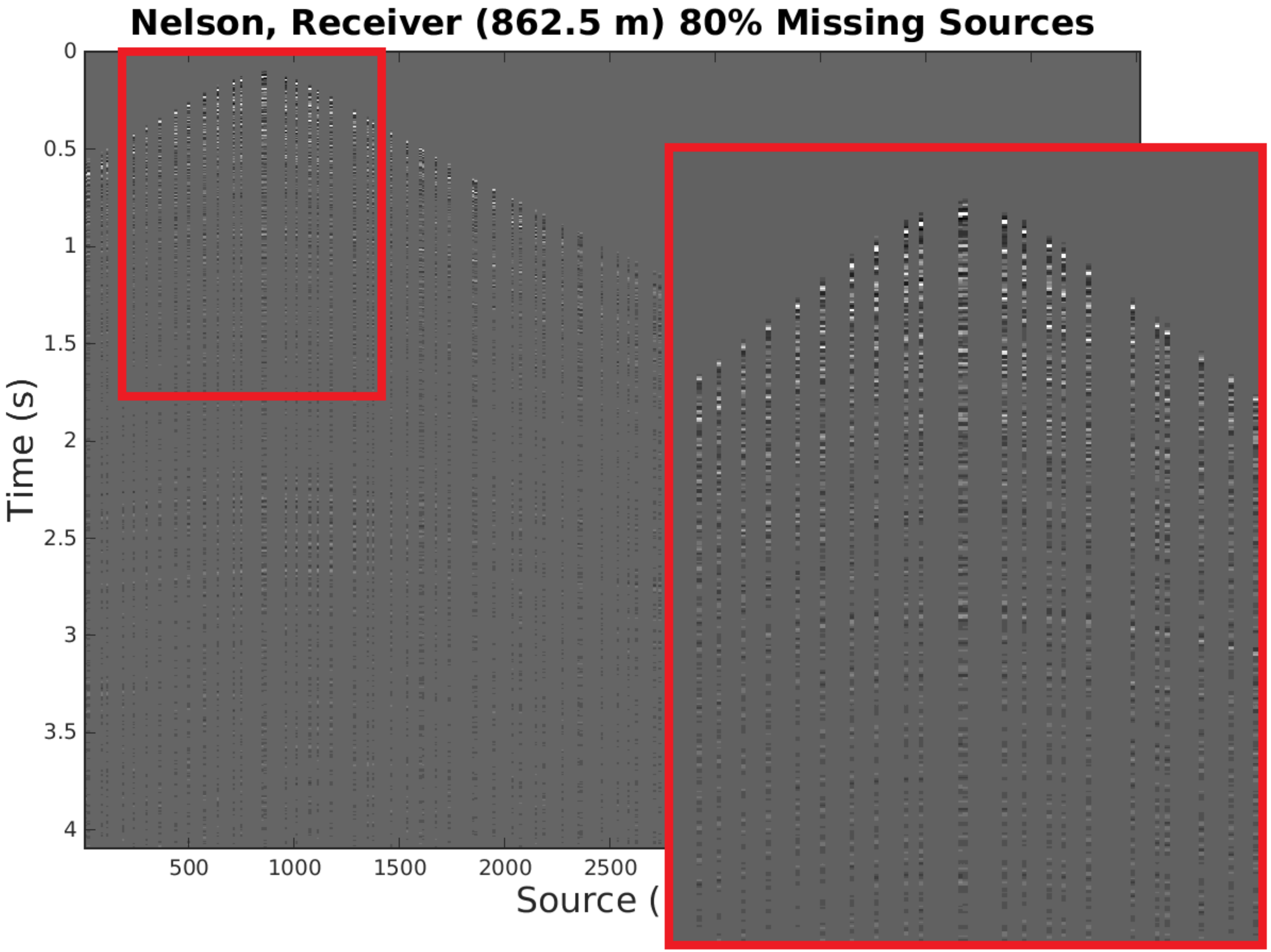
Experiments

Average Improvement
.5 dB

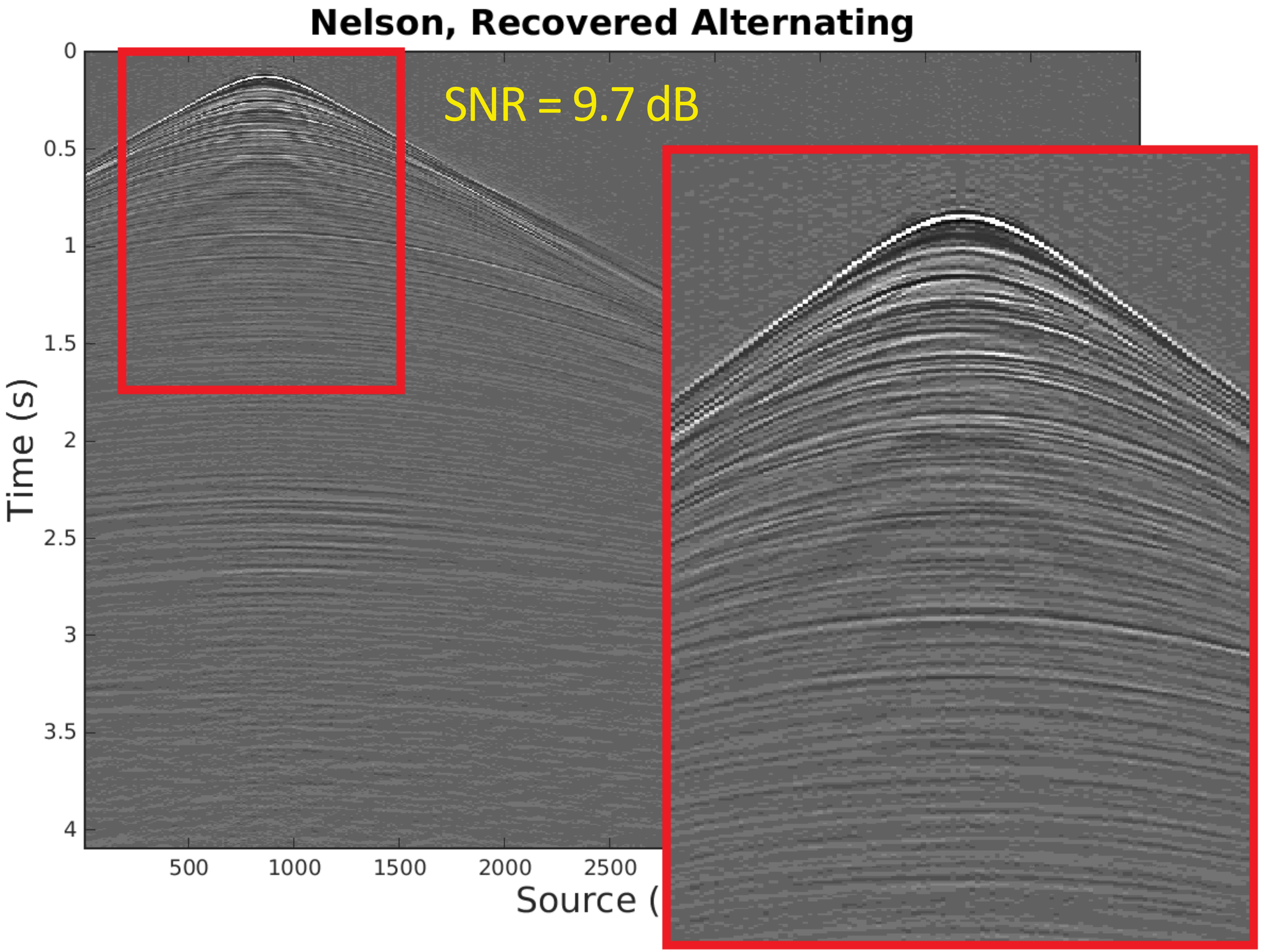


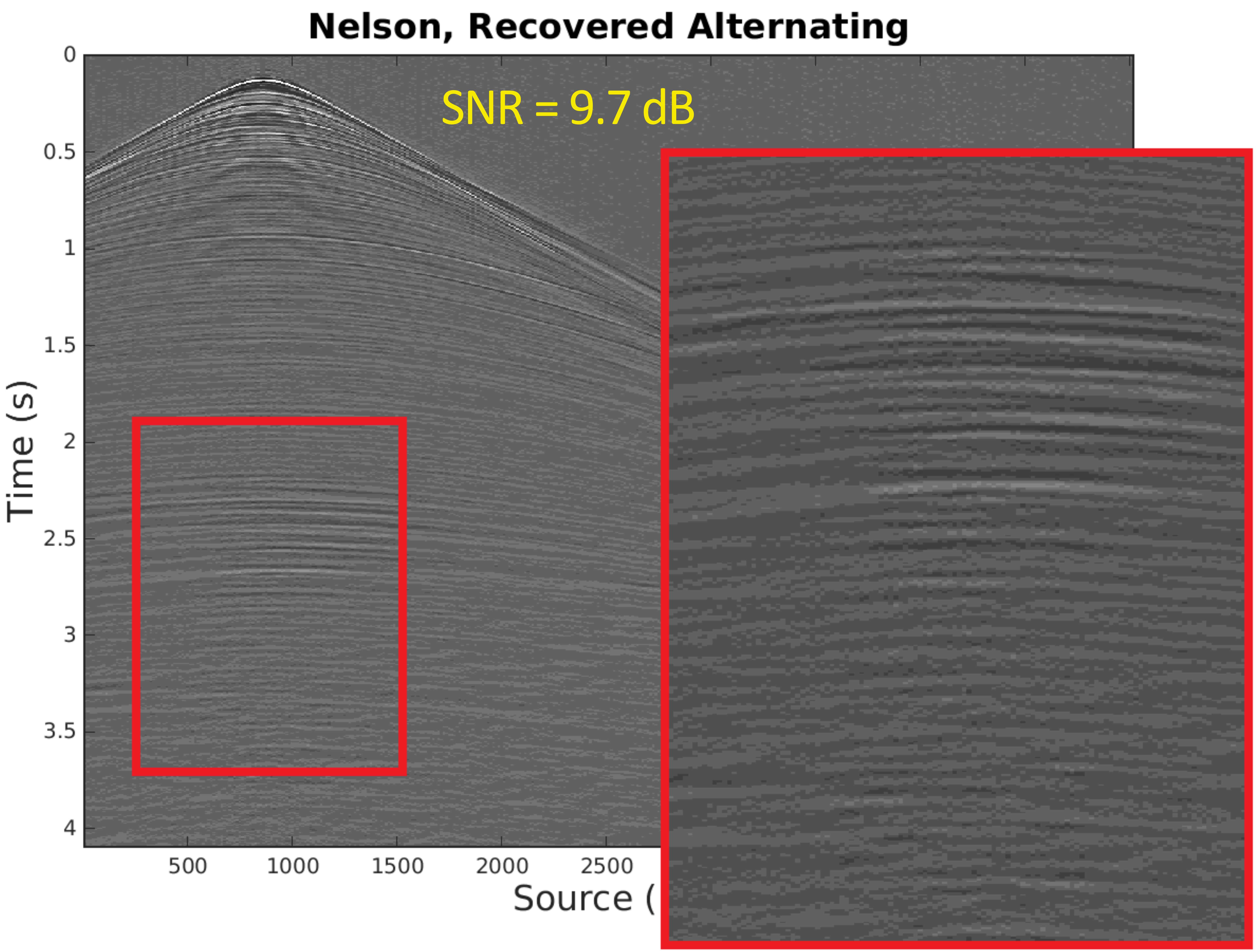

Nelson Dataset, Receiver (862.5 m)
Time (s)
0
0.5
1
1.5
2
2.5
3
3.5
4
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
Source (m)

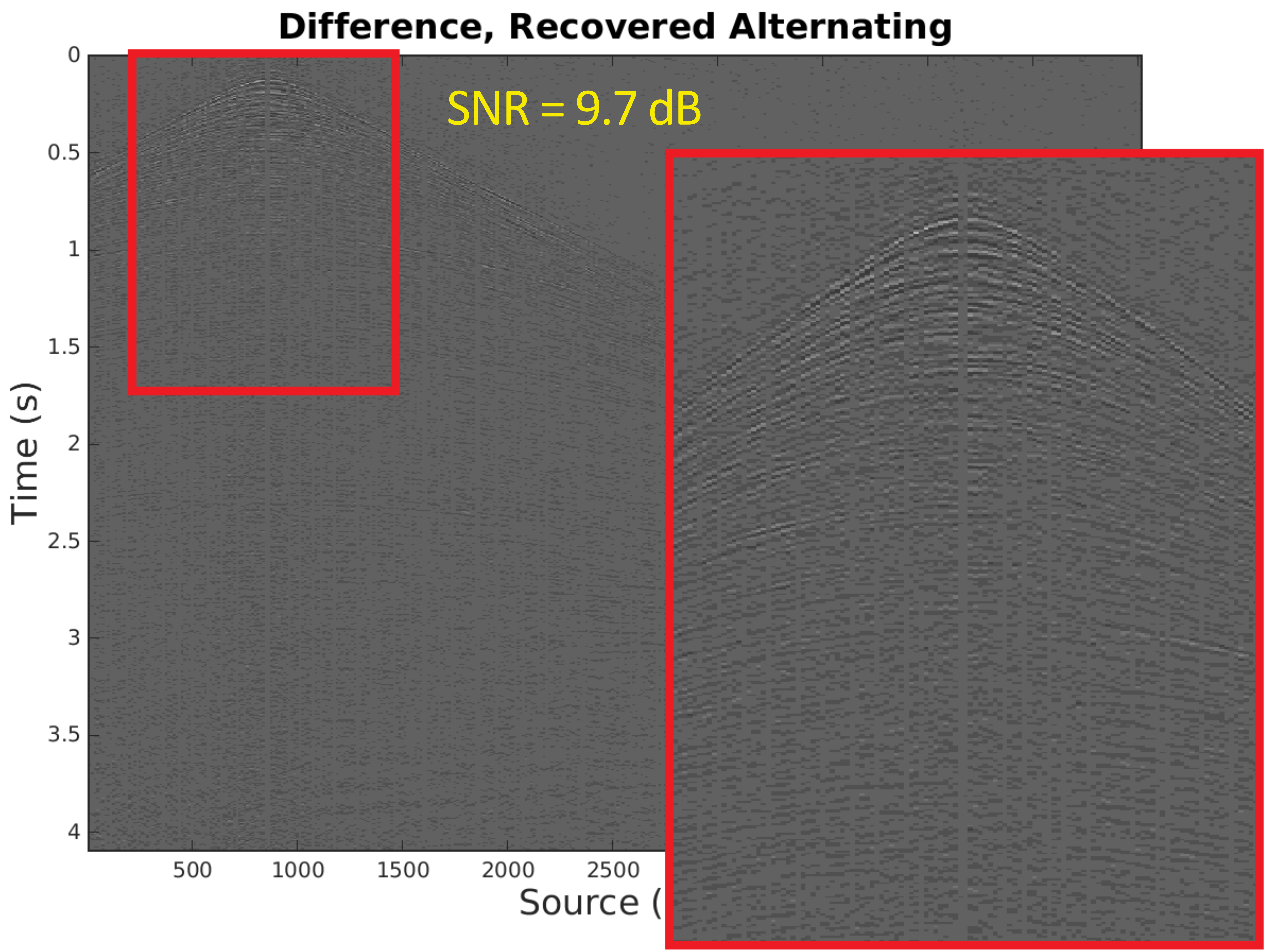

Nelson Dataset, Receiver (862.5 m)
Time (s)
0
0.5
1
1.5
2
2.5
3
3.5
4
500
1000
1500
2000
2500
Source (

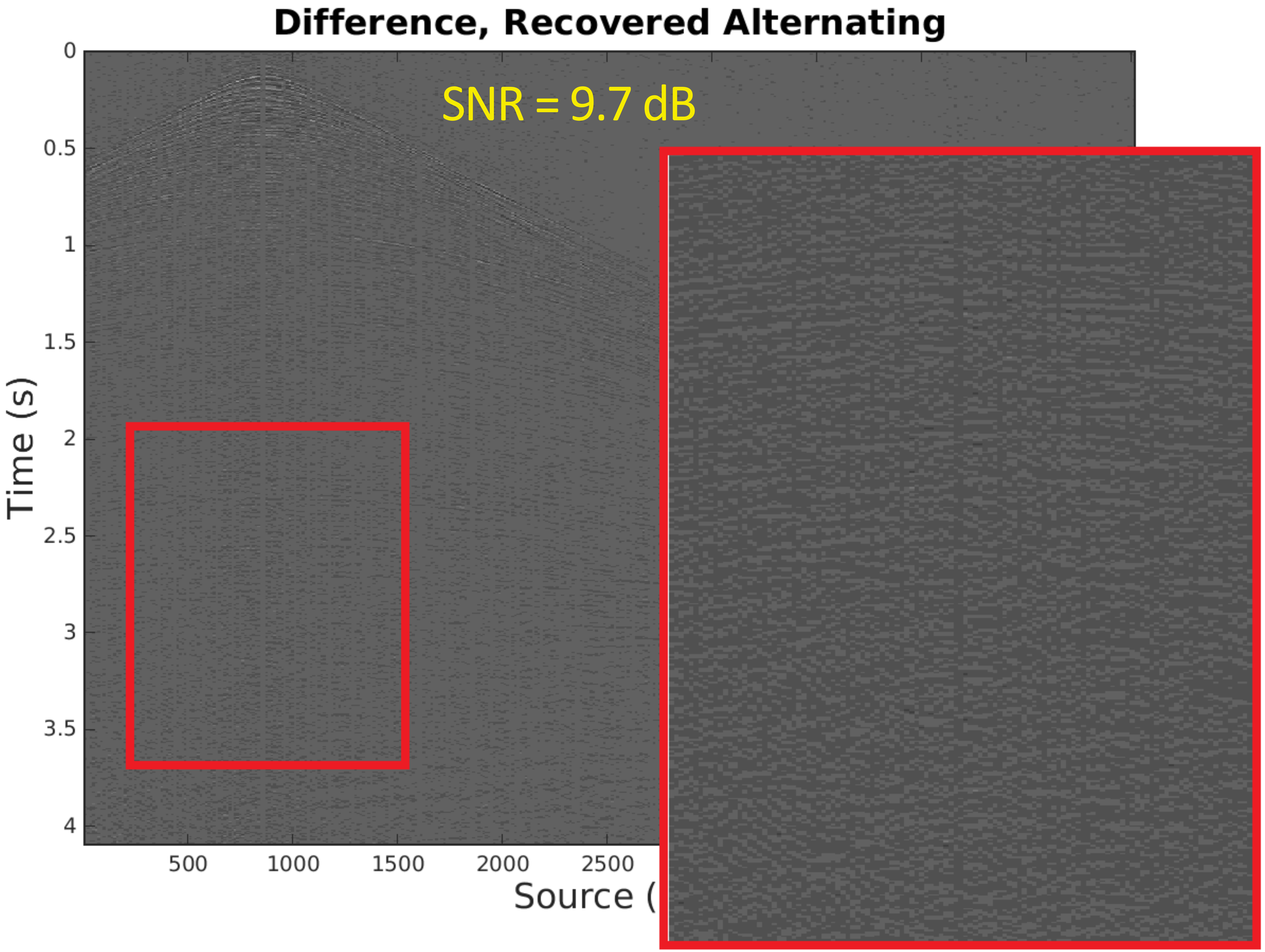


Nelson, Receiver (862.5 m) 80% Missing Sources
Time (s)
Source (
0
0.5
1
1.5
2
2.5
3
3.5
4
500
1000
1500
2000
2500


Nelson, Recovered Alternating
SNR = 9.7 dB
Time (s)
0
0.5
1
1.5
2
2.5
3
3.5
4
500
1000
1500
2000
2500
Source (


Nelson, Recovered Alternating
SNR = 9.7 dB
Time (s)
Source (
0
0.5
1
1.5
2
2.5
3
3.5
4
500
1000
1500
2000
2500


Difference, Recovered Alternating
SNR = 9.7 dB
0
0.5
1
1.5
2
2.5
3
3.5
4
Time (s)
500
1000
1500
2000
2500
Source (


Difference, Recovered Alternating
SNR = 9.7 dB
Time (s)
0
0.5
1
1.5
2
2.5
3
3.5
4
500
1000
1500
2000
2500
Source (

Experiments: Gulf Of Mexico 7Hz Slice

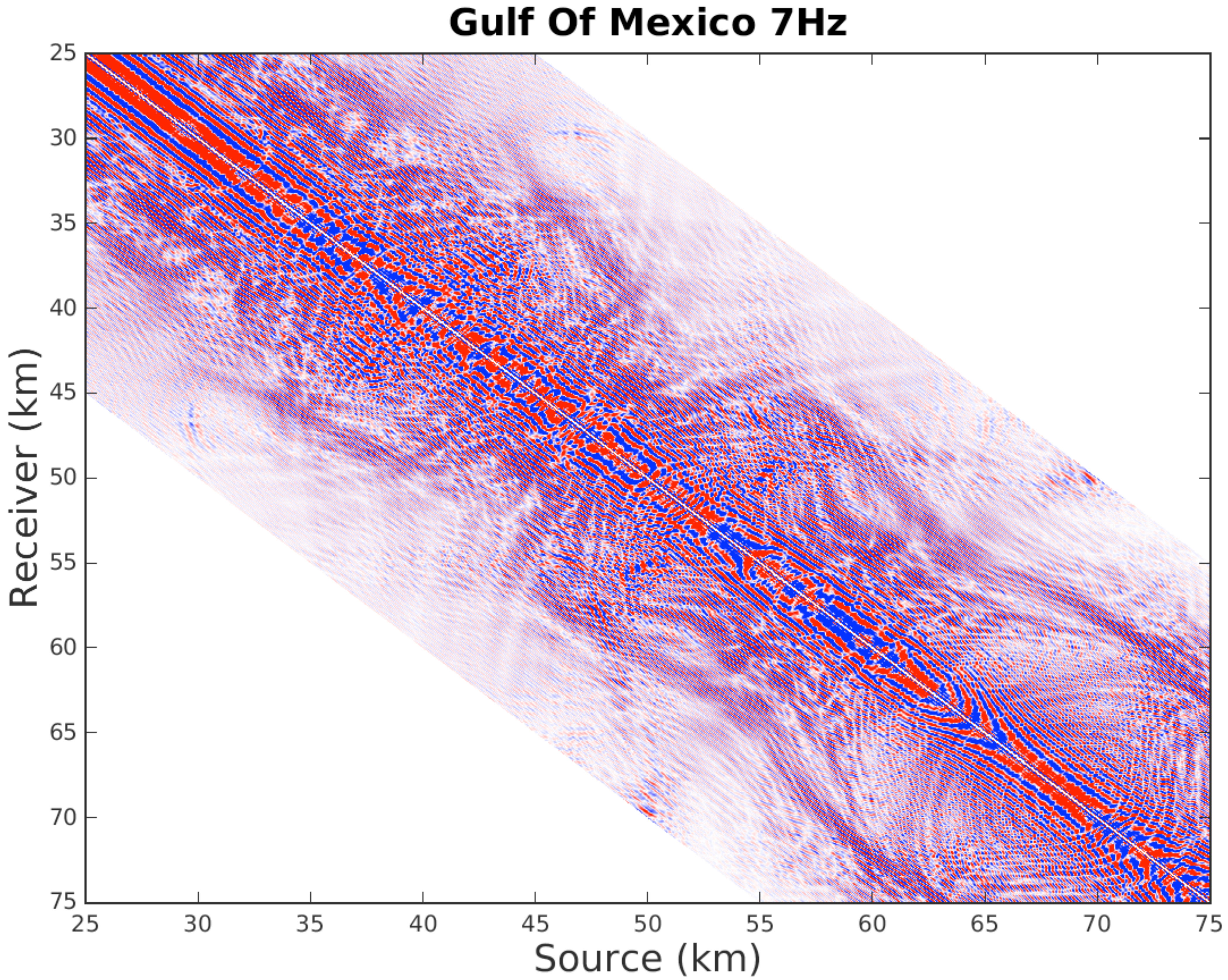
4001 x 4001 matrix (factorization parameter = 80)

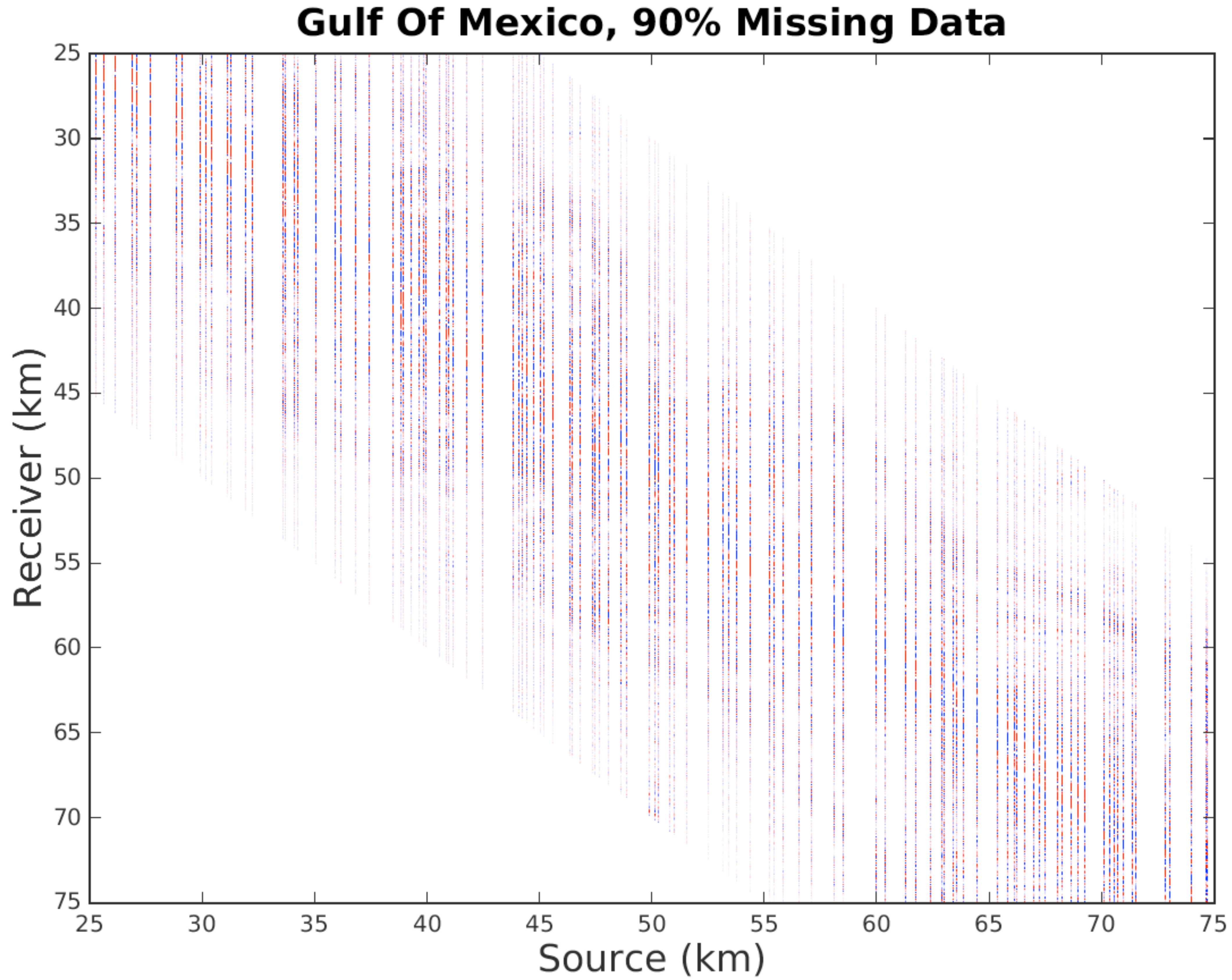
- 90% missing traces
- JIttered subsampling
- solve with SPGL1

Compare: Alternating vs Non Alternating

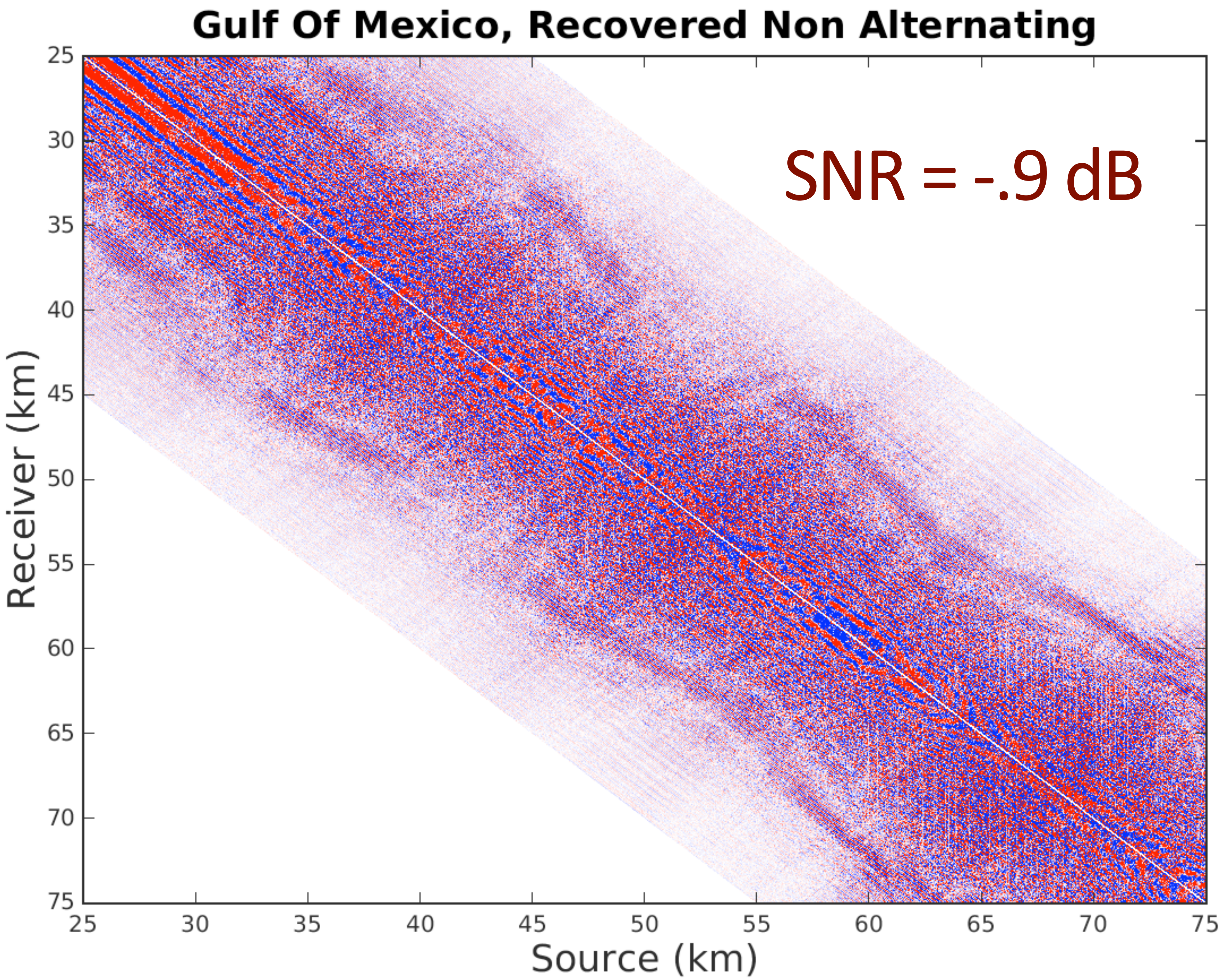
- 10 alternations, 15 iterations per alternation
- 300 iterations total for both

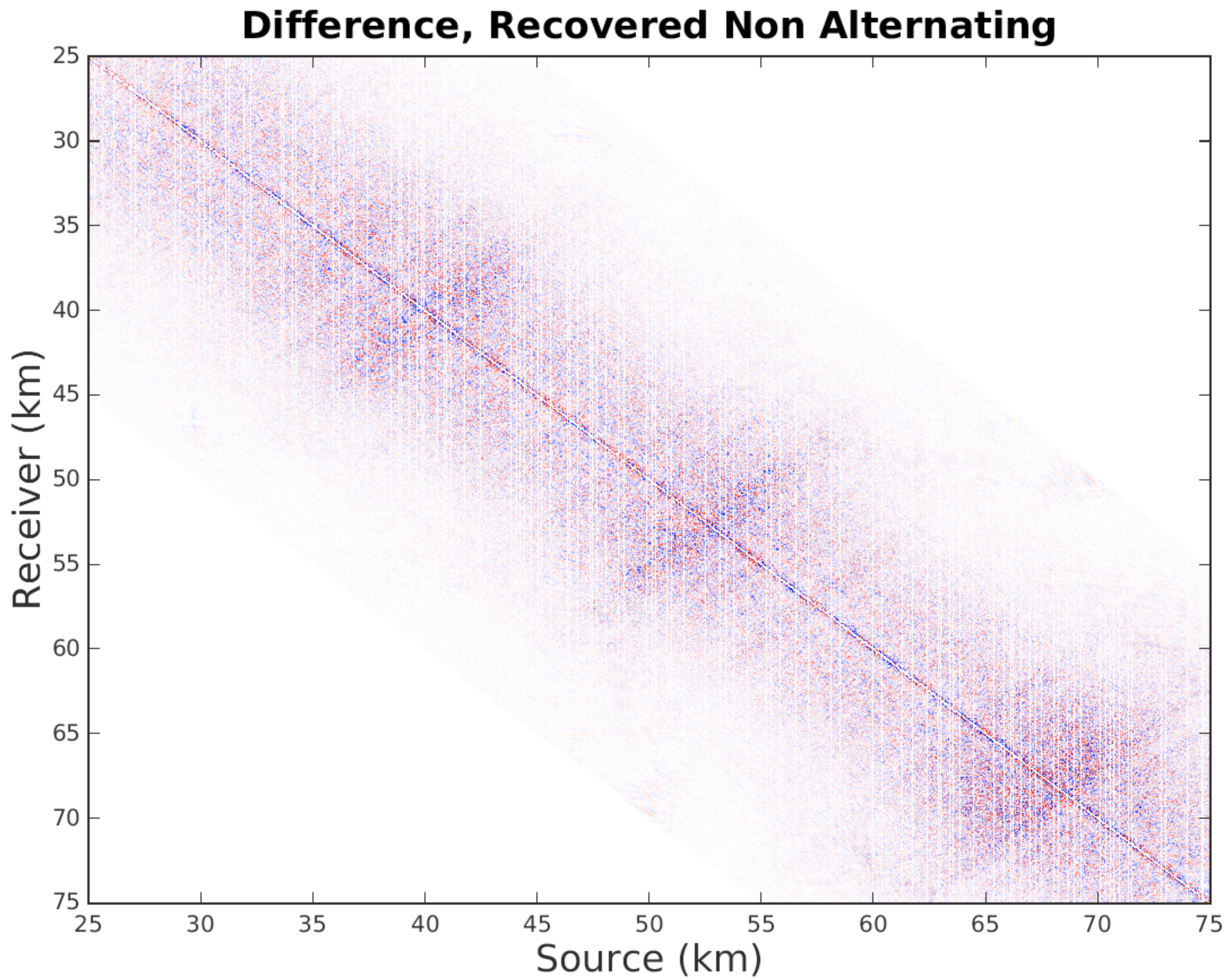
Experiments: Gulf Of Mexico 7Hz (90% missing)



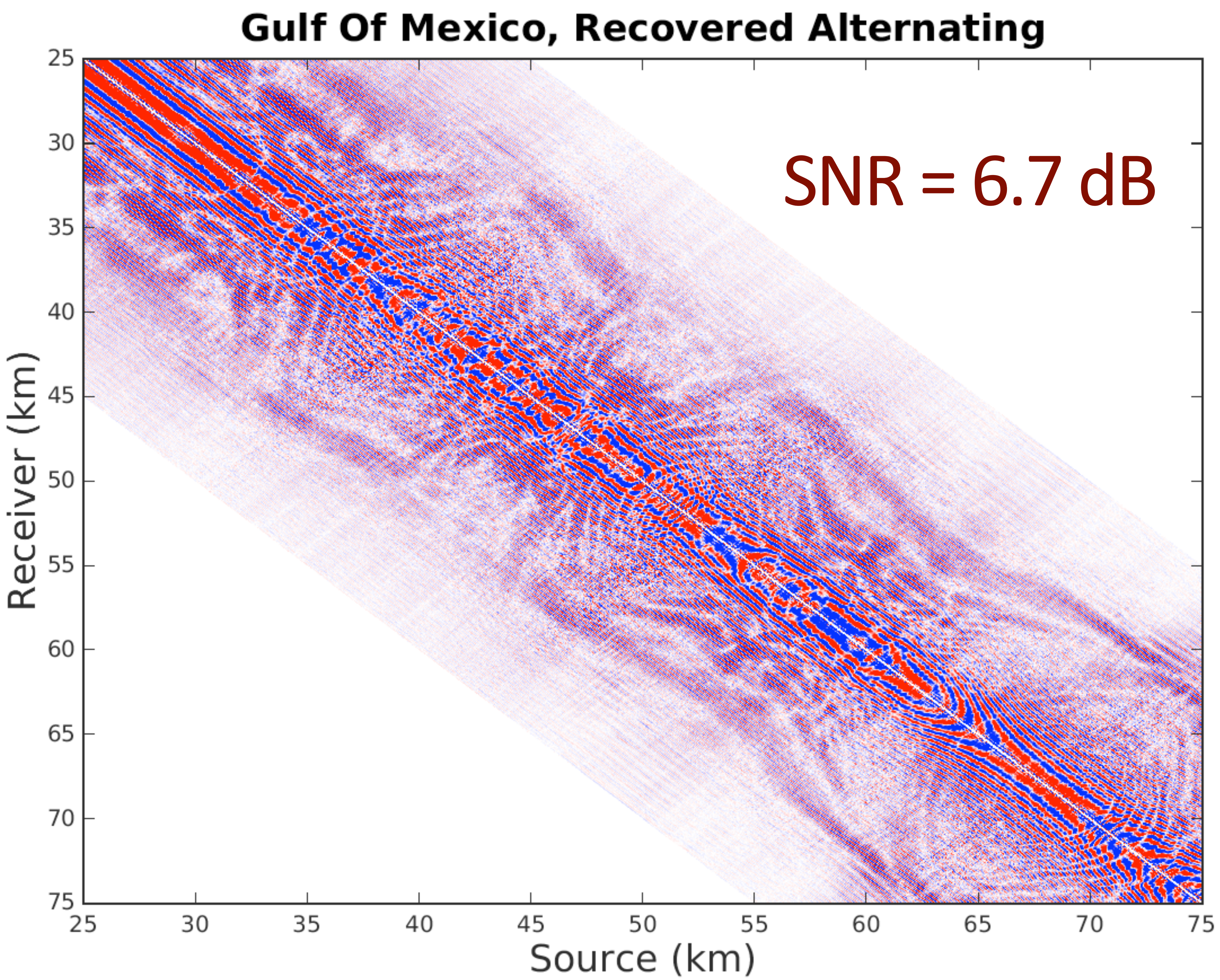


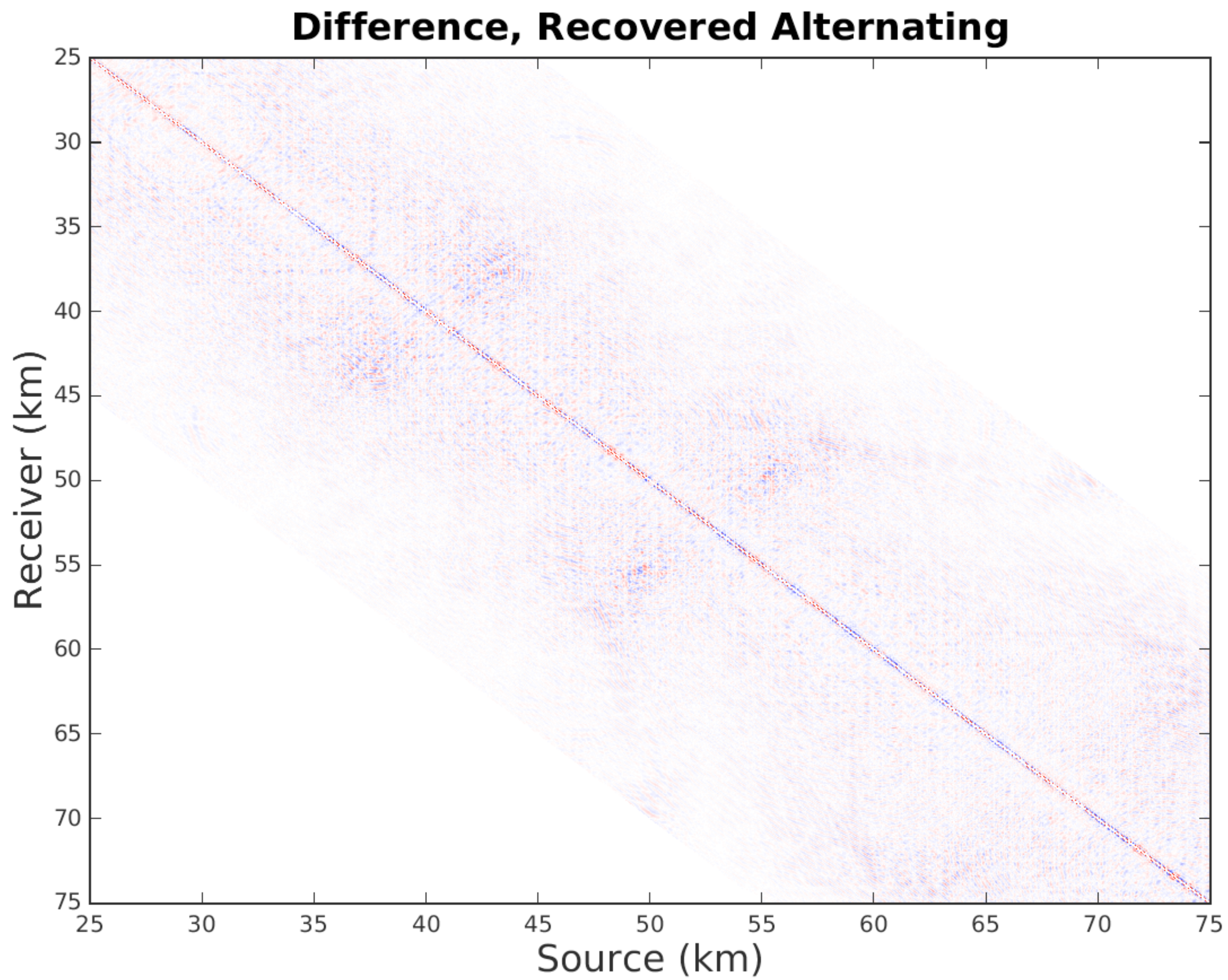
Experiments: Recovered Non Alternating



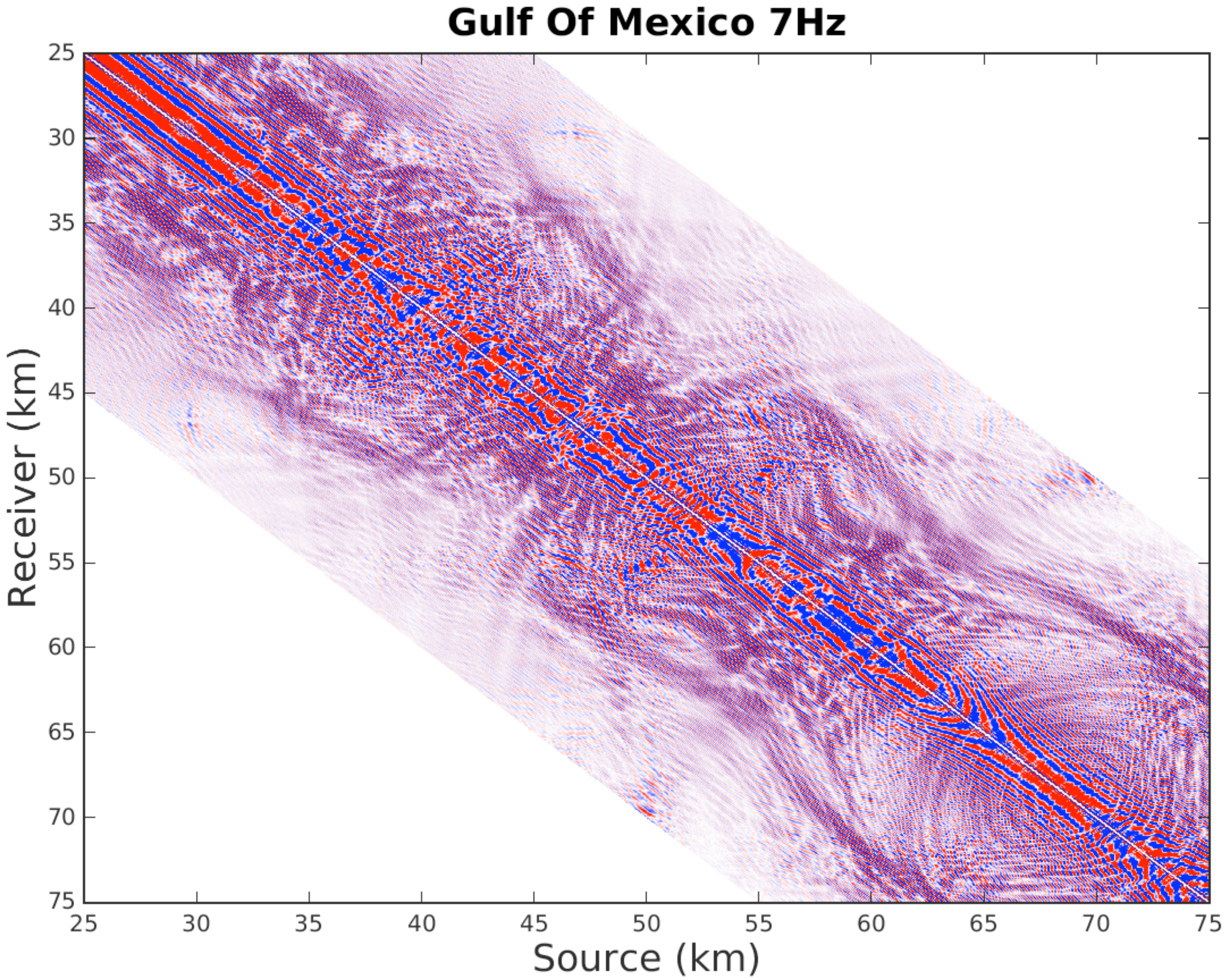


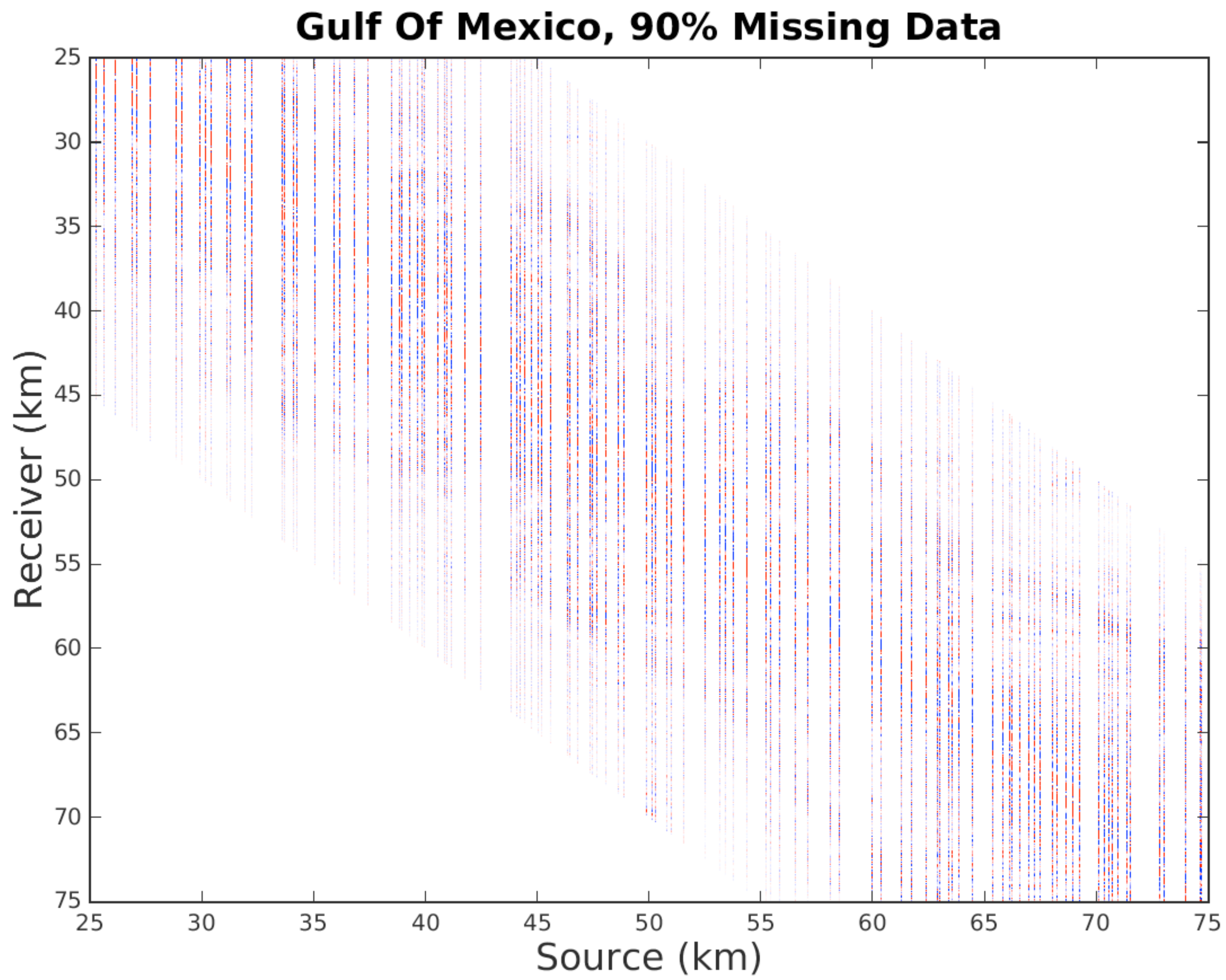
Experiments: Recovered Alternating





Experiments: Gulf Of Mexico 7Hz (90% missing)





Experiments: Regularization + Interpolation

Gulf Of Suez 10 hz slice - 354 x 354 matrix

- irregular data
- varying % missing sources
- Jittered subsampling

Compare: Alternating vs Non Alternating

- 6 alternations, 15 iterations per alternation
- 180 iterations total for both

Experiments: Regularization + Interpolation

solve:

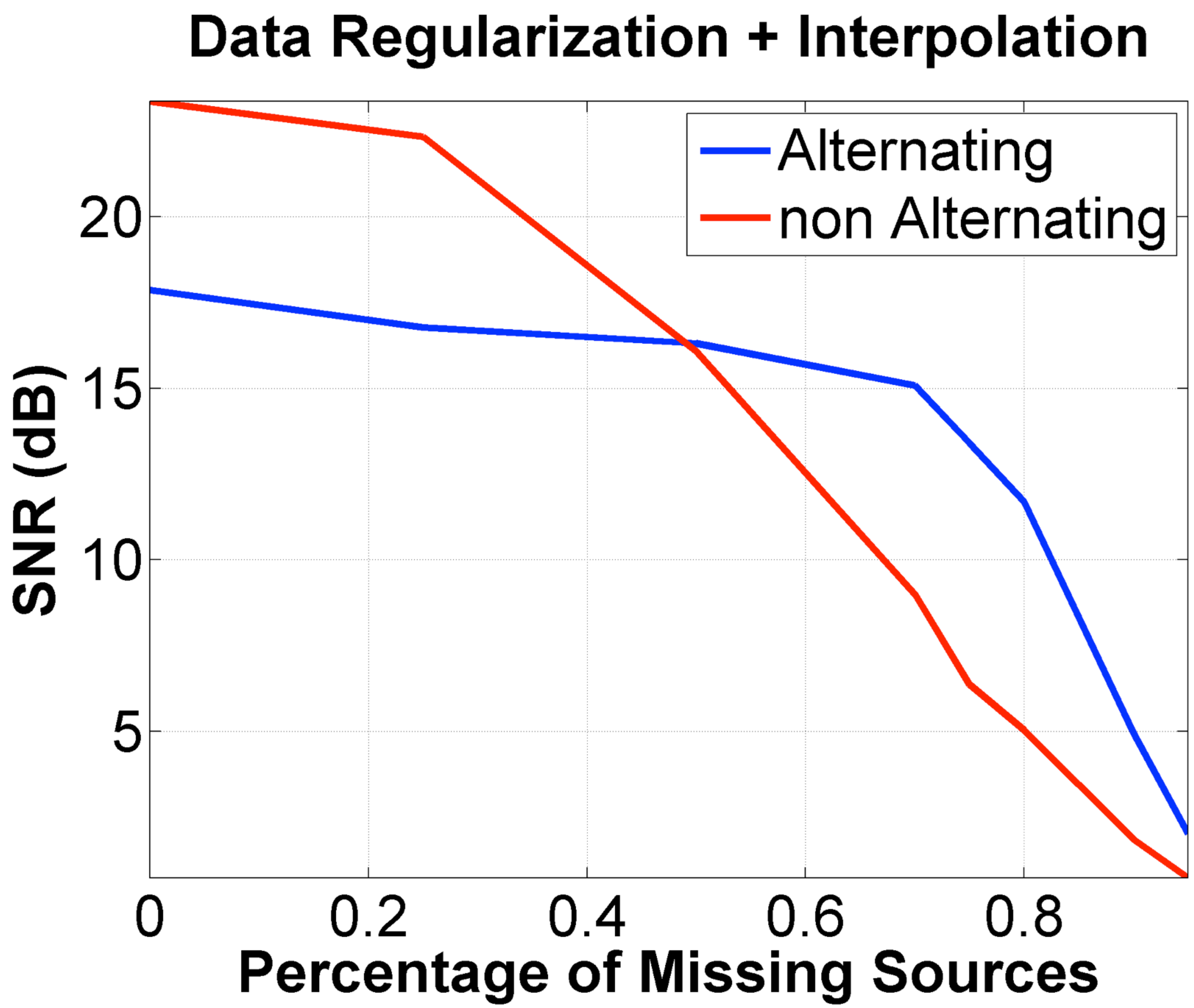
$$\min_{\mathbf{L}\in\mathbb{R}^{n\times r},\mathbf{R}\in\mathbb{R}^{m\times r}} \frac{1}{2}(||\mathbf{L}||_F^2 + ||\mathbf{R}||_F^2) \;\; \text{s.t.} \;\; ||\mathcal{A}(\mathbf{L}\mathbf{R}^H) - b||_F^2 \leq \sigma$$

$$\mathcal{A} = \mathbf{RMN}\mathcal{S}^H$$

where

$\mathbf{R}$: restriction operator
$\mathbf{M}$: measurement operator
$\mathbf{N}$: regularization operator
$\mathcal{S}^H$: transform operator

Experiments: Regularization + Interpolation



Computational Cost
with and without SVD

Percentage missing sources		50.0%		75.0%	
σ		0.1	0.1	0.1	0.1
Matrix completion w/ SVD	SNR (dB)	17.3	18.3	11.6	11.5
	time (sec)	**812**	**937**	**790**	**765**
Matrix completion w/o SVD	SNR (dB)	17.6	18.4	12.6	13.3
	time (sec)	**8**	**10**	**8**	**7**

Computational Cost
Matrix Completion vs Curvelet-based methods

Percentage missing sources		50.0%		75.0%	
σ		0.1	0.1	0.1	0.1
Matrix completion w/ SVD	SNR (dB)	17.3	18.3	11.6	11.5
	time (sec)	812.0	937.0	790.0	765.0
Matrix completion w/o SVD	SNR (dB)	17.6	18.4	12.6	13.1
	time (sec)	**8**	**10**	**8**	**7**
Curvelet-based sparsity promotion	SNR (dB)	17.4	18.6	12.5	12.8
	time (sec)	**879**	**989**	**817**	**1010**

Conclusions

- Alternating Optimization can better handle complicated cases
  - Highly sub sampled data (missing > 80%)
  - Complex data (high frequencies)

- Alternating Optimization does not increase time complexity

Future Work

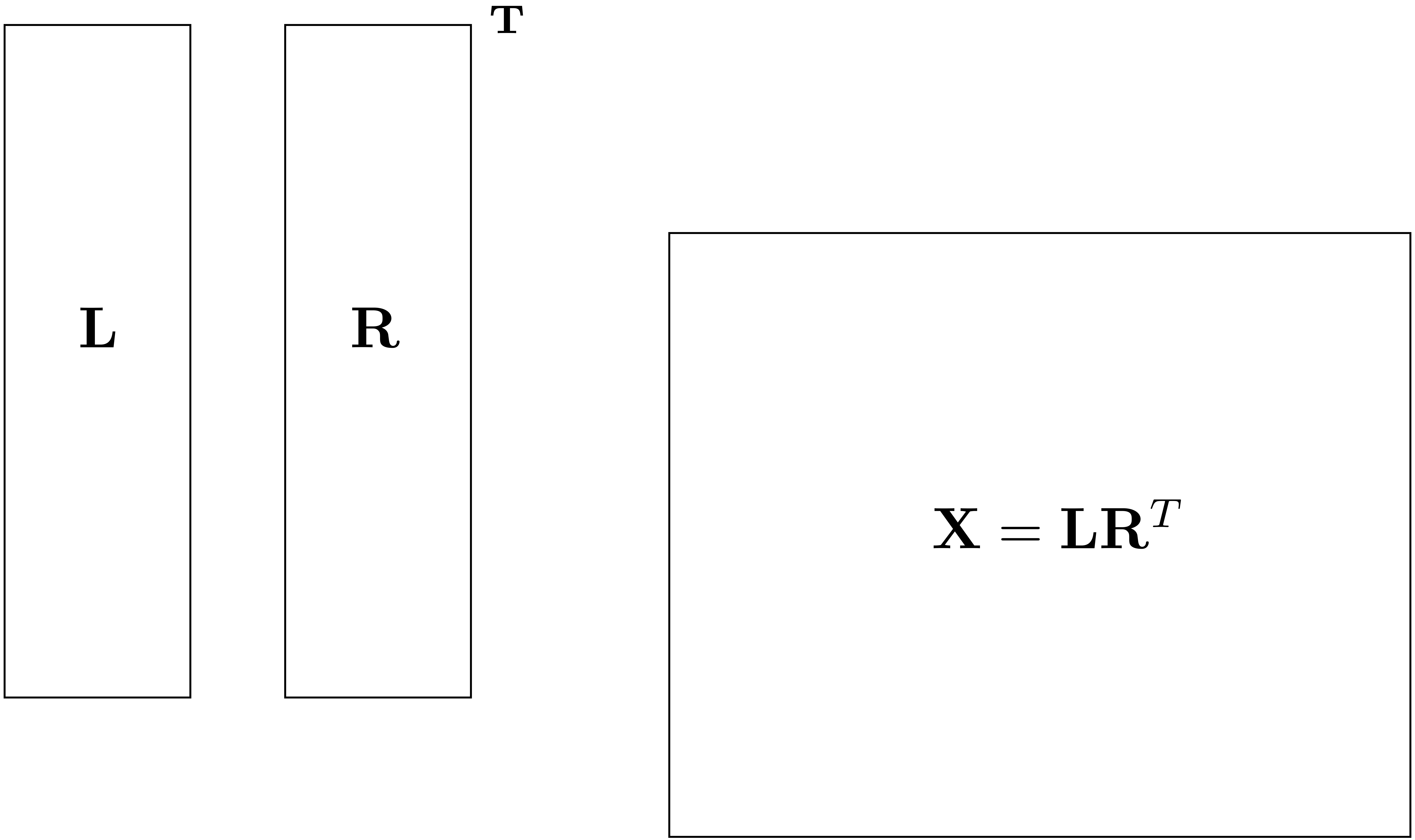
- Analysis of method
  - How many alternations are needed?

- Consider method for other applications
  - Tensor Completion
  - Source Separation
  - Parallel Matrix Completion

Future Work

- Analysis of method
  - How many alternations are needed?
- Consider method for other applications
  - Tensor Completion
  - Source Separation
  - Parallel Matrix Completion

LR parallel matrix multiplication



LR parallel matrix multiplication

	$\mathbf{L}$	$\mathbf{R}^{\mathbf{T}}$
Worker 1	$\mathbf{L_1}$	$\mathbf{R_1}$
Worker 2	$\mathbf{L_2}$	$\mathbf{R_2}$
Worker 3	$\mathbf{L_3}$	$\mathbf{R_3}$
Worker 4	$\mathbf{L_4}$	$\mathbf{R_4}$
Worker 5	$\mathbf{L_5}$	$\mathbf{R_5}$

$\mathbf{L_1R_1^T}$	$\mathbf{L_1R_2^T}$	$\mathbf{L_1R_3^T}$	$\mathbf{L_1R_4^T}$	$\mathbf{L_1R_5^T}$
$\mathbf{L_2R_1^T}$	$\mathbf{L_2R_2^T}$	$\mathbf{L_2R_3^T}$	$\mathbf{L_2R_4^T}$	$\mathbf{L_2R_5^T}$
$\mathbf{L_3R_1^T}$	$\mathbf{L_3R_2^T}$	$\mathbf{L_3R_3^T}$	$\mathbf{L_3R_4^T}$	$\mathbf{L_3R_5^T}$
$\mathbf{L_4R_1^T}$	$\mathbf{L_4R_2^T}$	$\mathbf{L_4R_3^T}$	$\mathbf{L_4R_4^T}$	$\mathbf{L_4R_5^T}$
$\mathbf{L_5R_1^T}$	$\mathbf{L_5R_2^T}$	$\mathbf{L_5R_3^T}$	$\mathbf{L_5R_4^T}$	$\mathbf{L_5R_5^T}$

Acknowledgements

Software Release Coming Soon

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, CGG, Chevron, ConocoPhillips, ION, Petrobras, PGS, Statoil, Total SA, Sub Salt Solutions, WesternGeco, and Woodside.