SIGNAL
ANALYSIS &
IMAGING GROUP

UNIVERSITY OF
ALBERTA

Tensor factorization and its application to multidimensional seismic data recovery

Recent Advances and the Road Ahead
SEG 2015, New Orleans

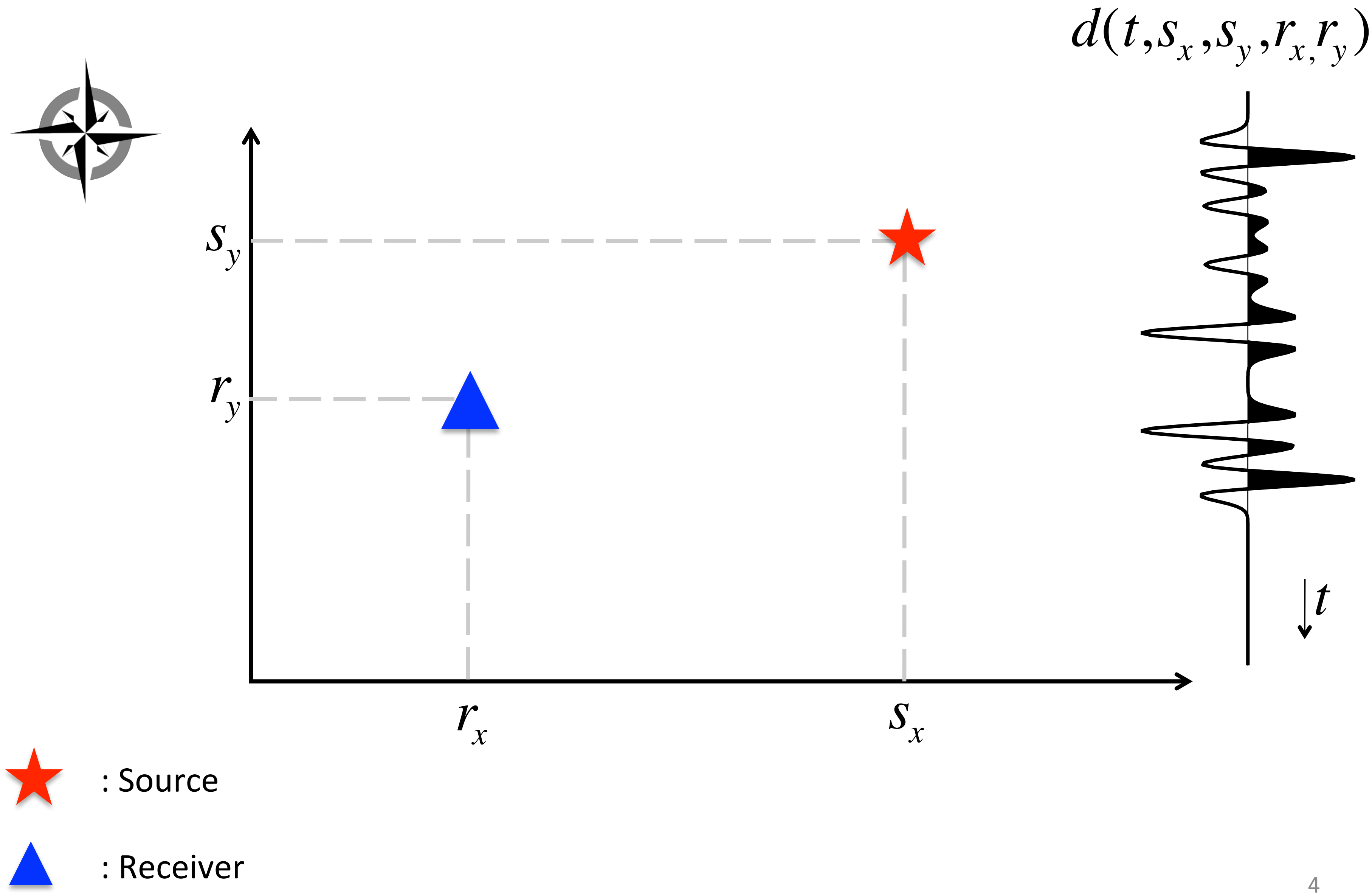
Mauricio D Sacchi
University of Alberta

Acknowledgments

- Contributors to the project:
  - Vicente Oropeza
  - Jianjun Gao
  - Nadia Kreimer
  - Aaron Stanton
  - Kevin Cheng
  - Ke Chen
- Industrial Sponsors of SAIG & NSERC

Why 5D cubes ?

Source receiver coordinates



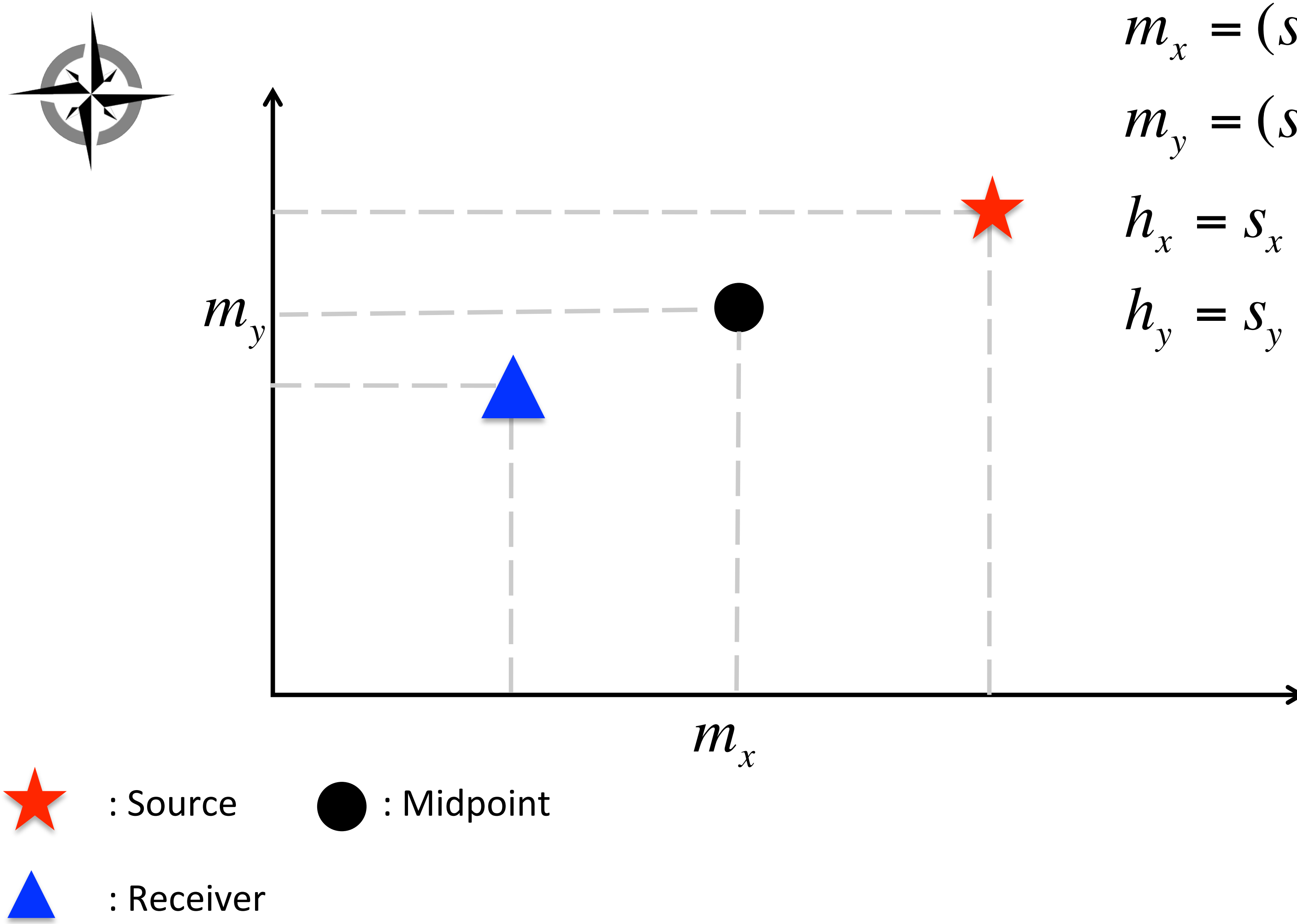
Midpoint-offset coordinates

$$m_x = (s_x + r_x)/2$$

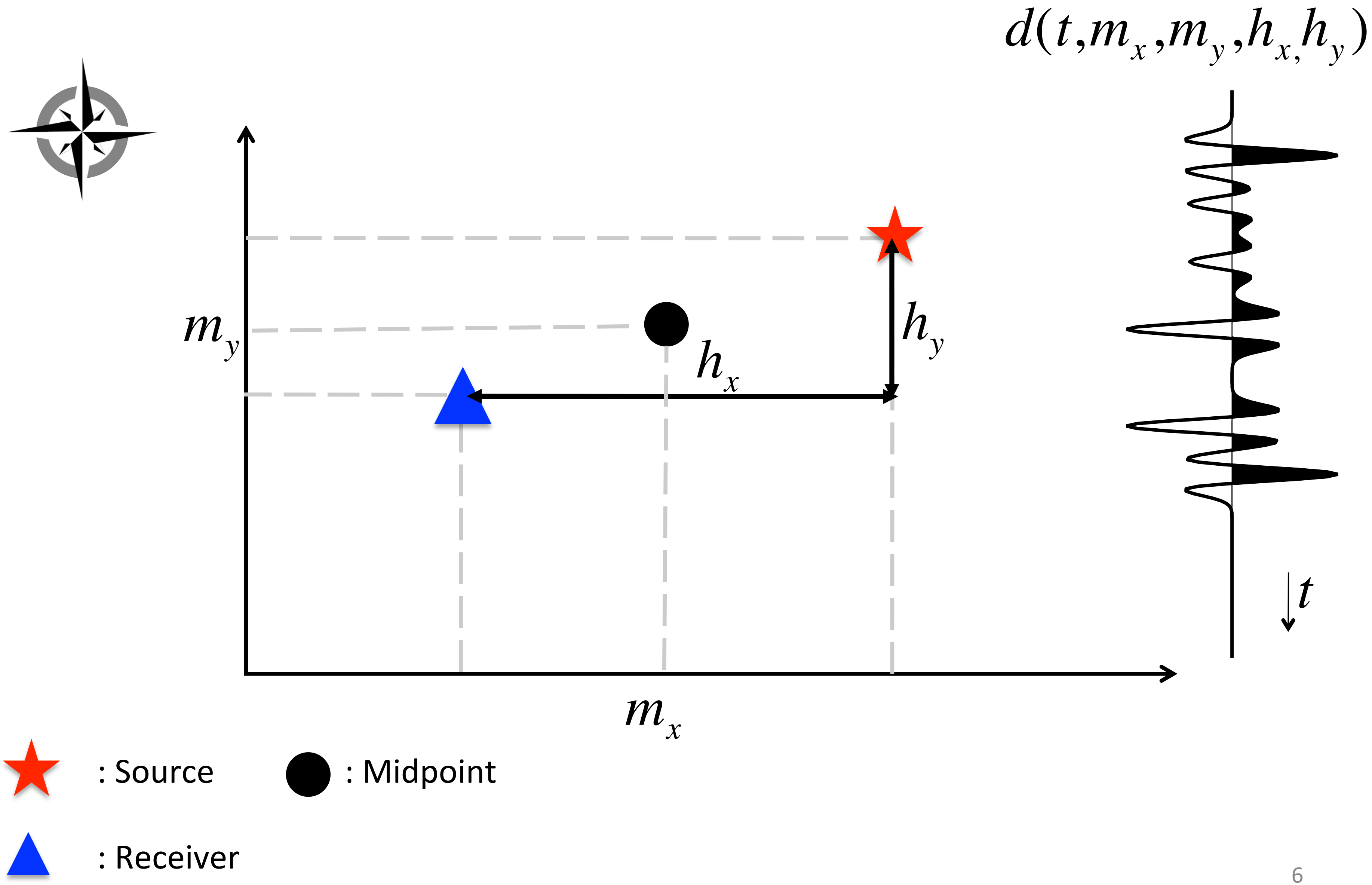
$$m_y = (s_y + r_y)/2$$

$$h_x = s_x - r_x$$

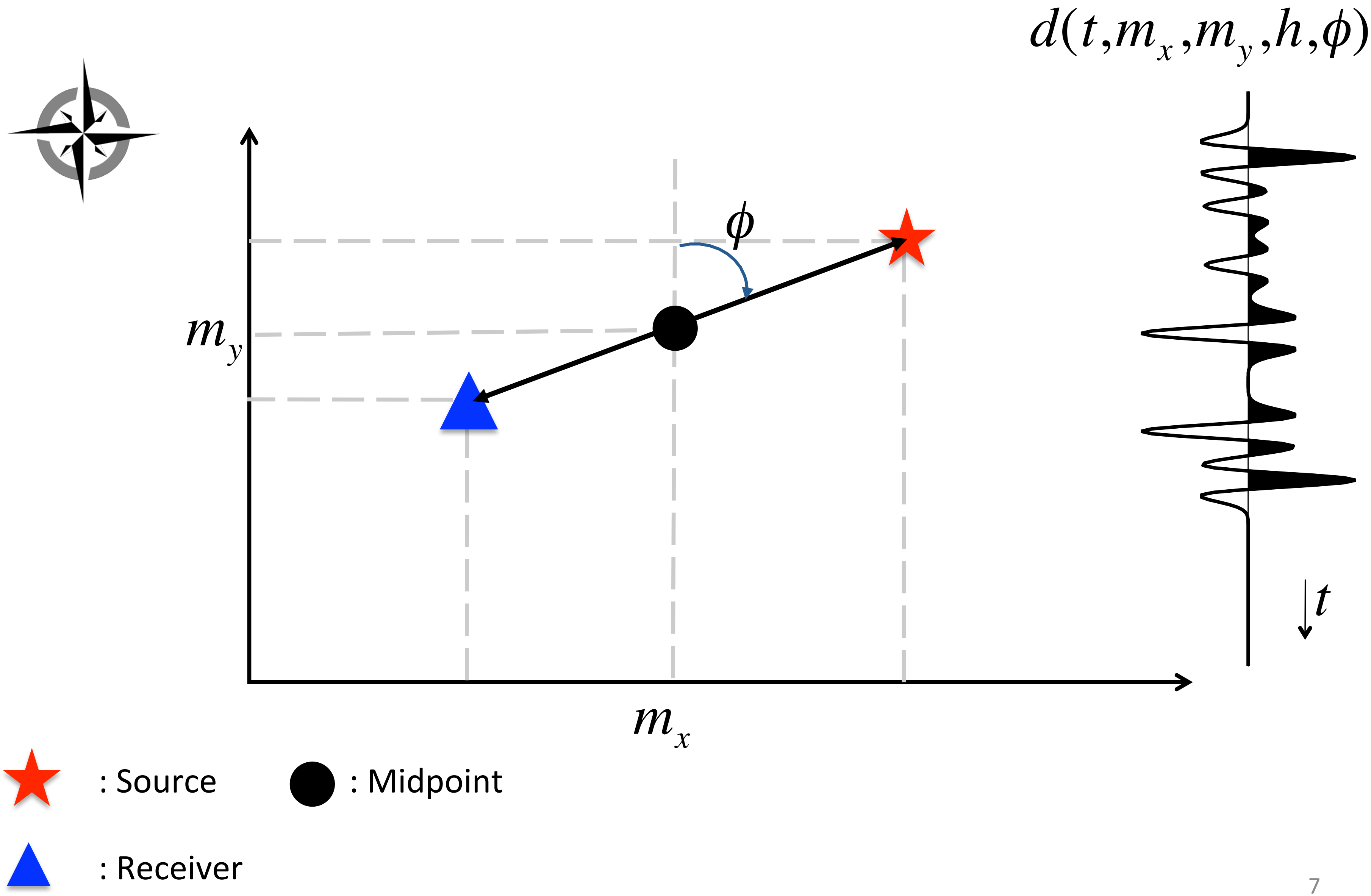
$$h_y = s_y - r_y$$



Midpoint-offset coordinates



Midpoint-offset coordinates



5D data (4 spatial coordinates + time)

- Source Receiver coordinates $d(t, s_x, s_y, r_{x,} r_y)$
- Midpoint, inline and cross-line offsets $d(t, m_x, m_y, h_{x,} h_y)$
- Midpoint, offset and azimuth $d(t, m_x, m_y, h, \phi)$

5D data (4 spatial coordinates + frequency)

$$f(t) \leftrightarrow F(\omega)$$

- Source Receiver coordinates $D(\omega, s_x, s_y, r_x, r_y)$
- Midpoint, inline and cross-line offsets $D(\omega, m_x, m_y, h_x, h_y)$
- Midpoint, offset and azimuth $D(\omega, m_x, m_y, h, \phi)$

From 5D data in the frequency domain to 4th order tensors

$$D(\omega, m_x, m_y, h_x, h_y) \rightarrow D_{ijkl}$$

$$m_x \rightarrow i$$

$$m_y \rightarrow j$$

$$h_x \rightarrow k$$

$$h_y \rightarrow l$$

Sampling

- In general, 5D volumes are irregularly sampled in space due to
    - Logistic constraints
    - Insufficient equipment
    - Acquisition costs
    - Provincial/Municipal regulations
    - Environmental constraints

Problem

- The Problem: algorithms for
  - Seismic Imaging and
  - Inversion for QI

  require regular *and dense* data.

Solution

- **Constrained inversion is adopted to solve the seismic reconstruction problem**

- **Assumption: Simplicity in the data representation**

    - **Predictability**

      Spitz 91

    - **Sparsity**

      Sacchi et al. 98, Liu and Sacchi 2004

    - **Rank**

      Trickett et al. 2010, Kreimer and Sacchi 2011

Stanton & Sacchi, All Roads Lead to Rome: Predictability, Sparsity, Rank and pre-stack data reconstruction. Recorder, December 2013

Simplicity in k-space ("Sparsity")

- MWNI (Minimum Weighted Norm Interpolation)
- ALFT (Anti leakage Fourier Transform)
- POCS (Projection onto Convex Sets)
- Sparse Fourier Reconstruction
- Matching Pursuit Reconstruction
- and 10^{10} versions of the aforementioned algorithms

Reconstruction techniques

- Established technology:
  - Industry has mainly adopted methods based on PEFs and/or Fourier synthesis with simplicity in k-space
- Recent developments:
  - Methods that assume that seismic data can be embedded into a low rank matrix/tensor.
    - Interesting area of research with connections to Data Analytics, Big Data, Collaborative Filtering, Personalized Medicine etc etc etc…

Rank-based Reconstruction Techniques

- Methods that assume that seismic data can be embedded into a low rank matrix or tensor:
  - **Rank reduction of Block Hankel forms (Cadzow / MSSA)**
    - Trickett et al 2010
    - Oropeza & Sacchi 2011
    - Gao et. al, 2013
  - **Rank reduction of Multi-liner arrays or tensors**
    - Kreimer and Sacchi, 2011 (HO SVD)
    - Kreimer et al 2013 (Minimum Nuclear Norm Tensor Completion)
    - Gao et al 2015 (Tensor Completion via Parallel Matrix Factorization)

Recommender System

- A recommendation system (or recommender system) is an algorithm that attempts to predict the rating that a user will give to an item. Recommendation systems have become quite popular in the field of e-commerce for predicting ratings of movies, books, news, research articles etc.

Netflix Prize

- From http://www.netflixprize.com/
- *"The Netflix Prize sought to substantially improve the accuracy of predictions about how much someone is going to enjoy a movie based on their movie preferences"*
- Netflix provided a training data set of 100,480,507 ratings that 480,189 users gave to 17,770 movies (only 1.17% of the elements of the data table/matrix are known)
- On September 21, 2009 Netflix awarded the $1M Grand Prize to team BellKor's Pragmatic Chaos.

Matrix/Tensor Completion and the famous NETFLIX problem

Movie →

User ↓

	Taxi Driver	Sense and Sensibility	Battleship Potemkin	Raging Bull	Titanic	Alexander Nevsky
John	5	1	5	4	1	3
Mary	1	4	?	1	4	?
Pepe	4	2	2	3	4	?
Adrian	3	1	?	3	3	?
Tony	?	?	?	?	4	?
Kevin	3	3	?	3	2	?
Jianjung	2	1	?	2	4	?
Natasha	?	?	3	?	5	3

Hypothetical portion of the Netflix matrix

Matrix completion with minimal math

$$\text{Find } M_{ij}, \quad \text{such that } S_{ij} M_{ij} = M_{ij}^{obs} \text{ and } \operatorname{rank}(M) = K$$

S_{ij} : Sampling operator

Tensor completion with minimal math

Find M_{ijkl}, such that $S_{ijkl}M_{ijkl} = M_{ijkl}^{obs}$ and multi-rank$(M) = K$

S_{ijkl} : Sampling operator

Algorithm

Simple matrix completion algorithm

$$M^{obs} = SM$$

$$M^{k} = \alpha M^{obs} + (1 - \alpha S) R[M^{k-1}]$$

$$R[\ ] = \text{rank reduction}$$

Algorithm $\alpha = 1$

Simple matrix completion algorithm

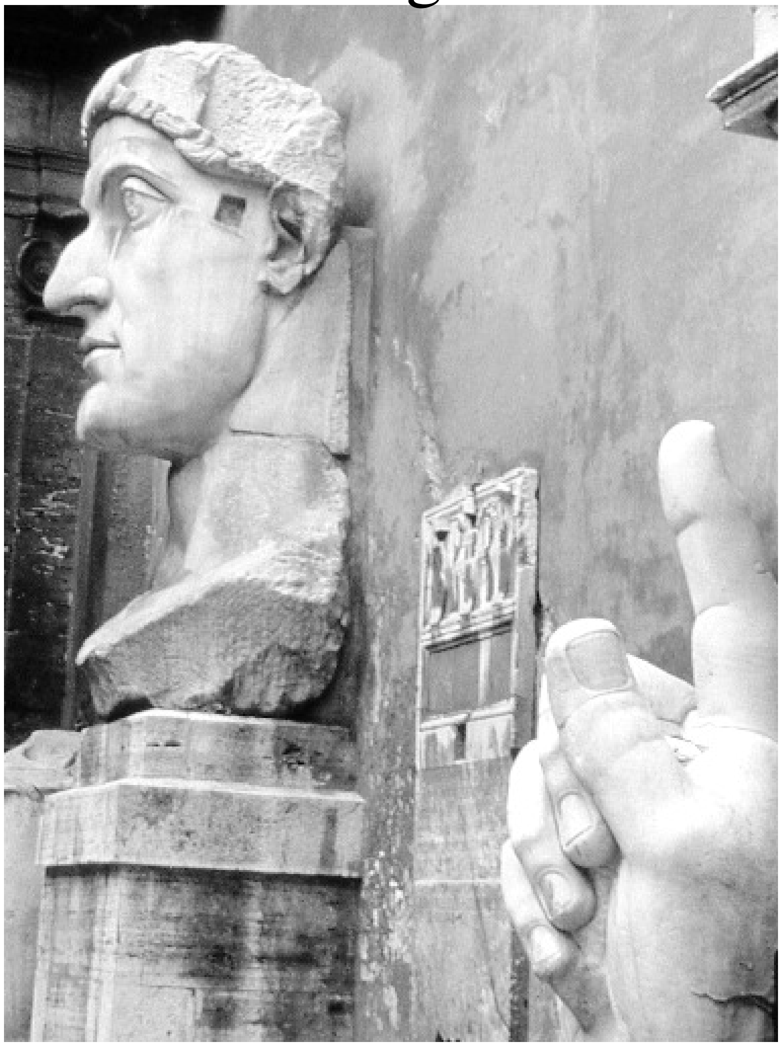
Insert existing data

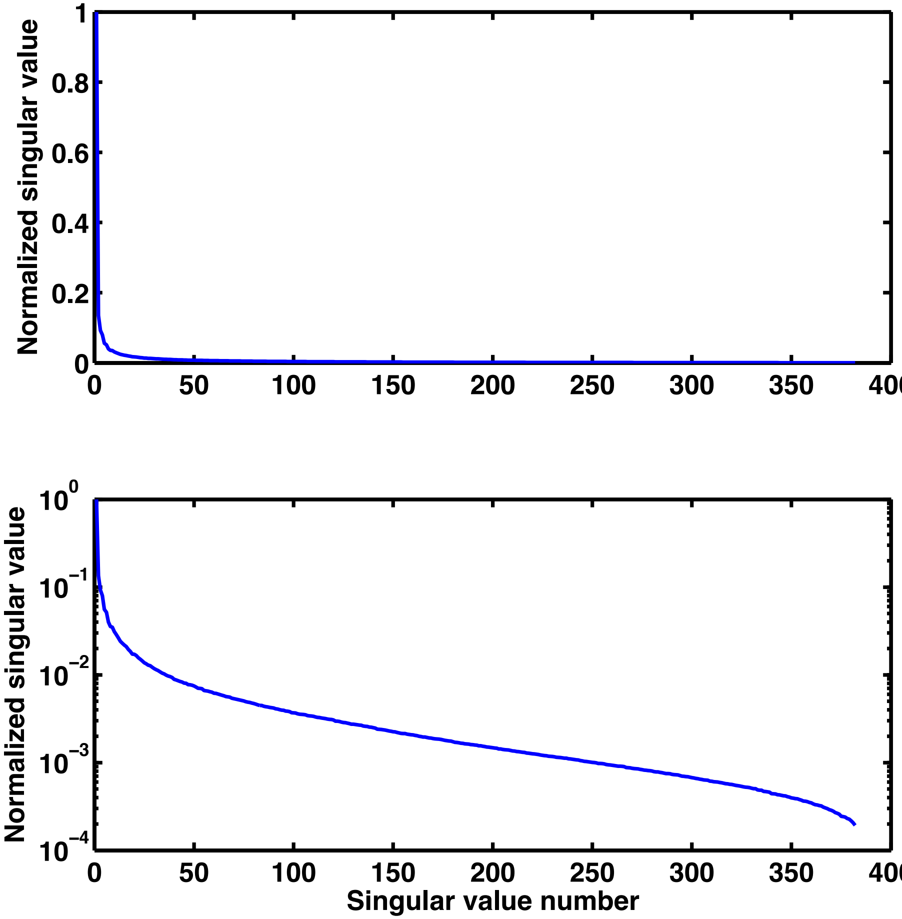
$$M^k = M^{obs} + (1 - S)R[M^{k-1}]$$

Replace low rank approximation in pixels with missing data

Constantine the Great (c. 280-337)

M : True Image

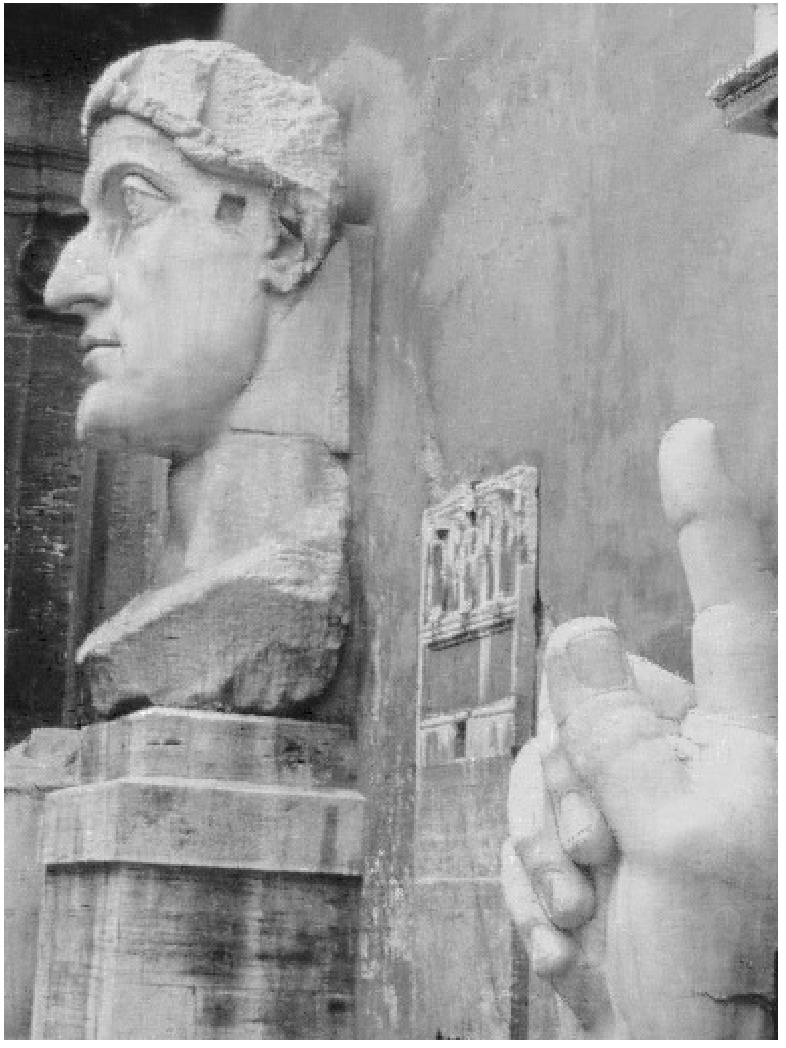




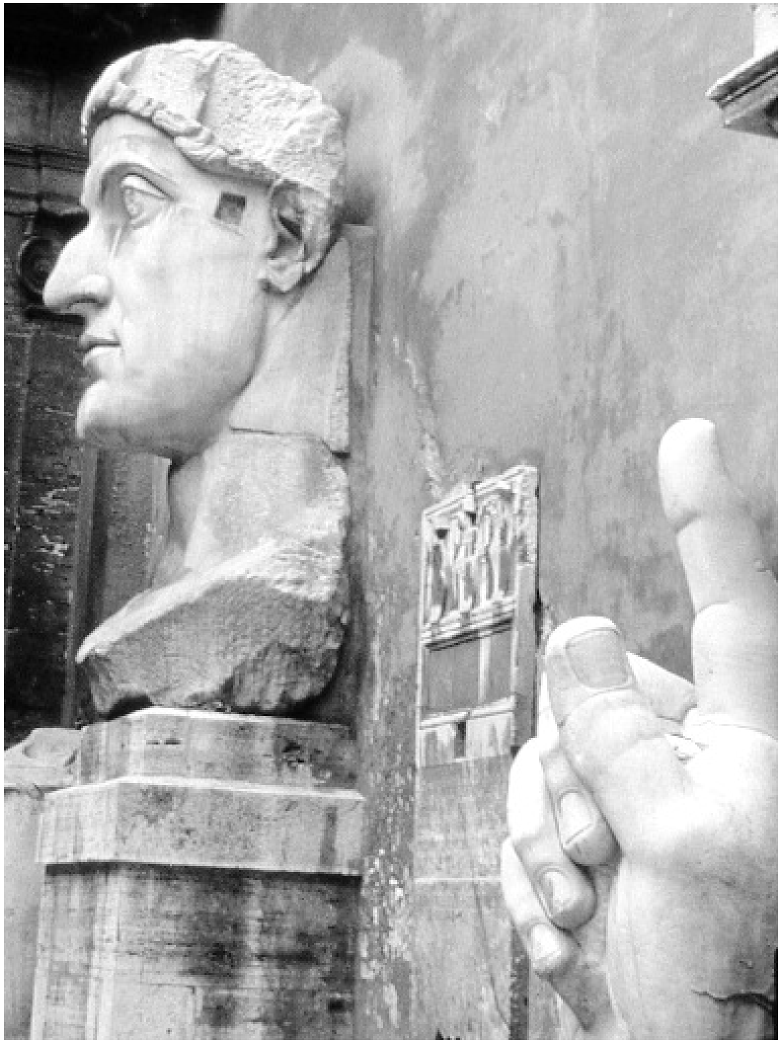
Constantine the Great after decimation

$$M^{obs} = SM$$

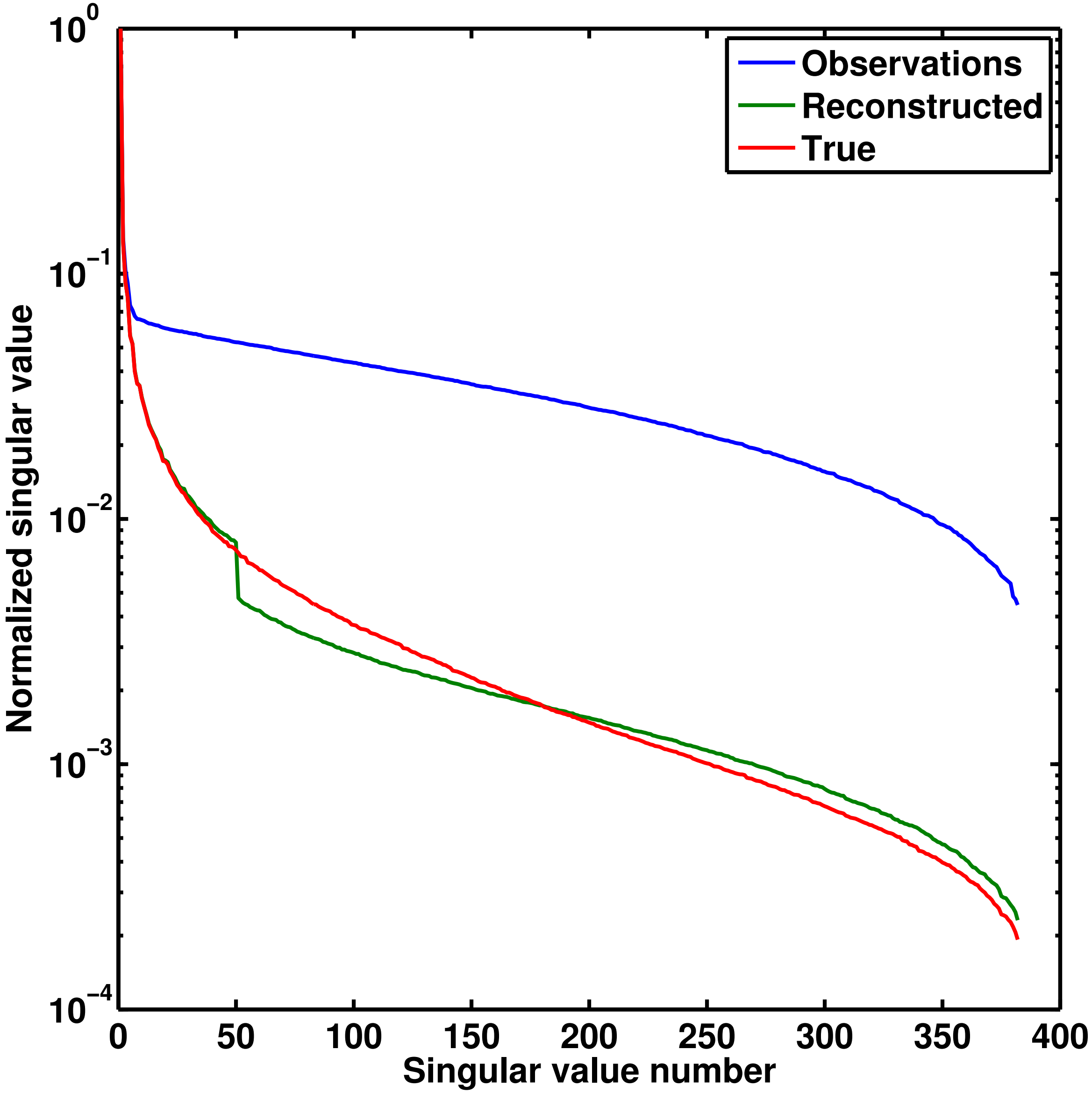
Constantine the Great after reconstruction



Constantine the Great – original image



Constantine the Great – Singular values



Real Data Example (WCSB)

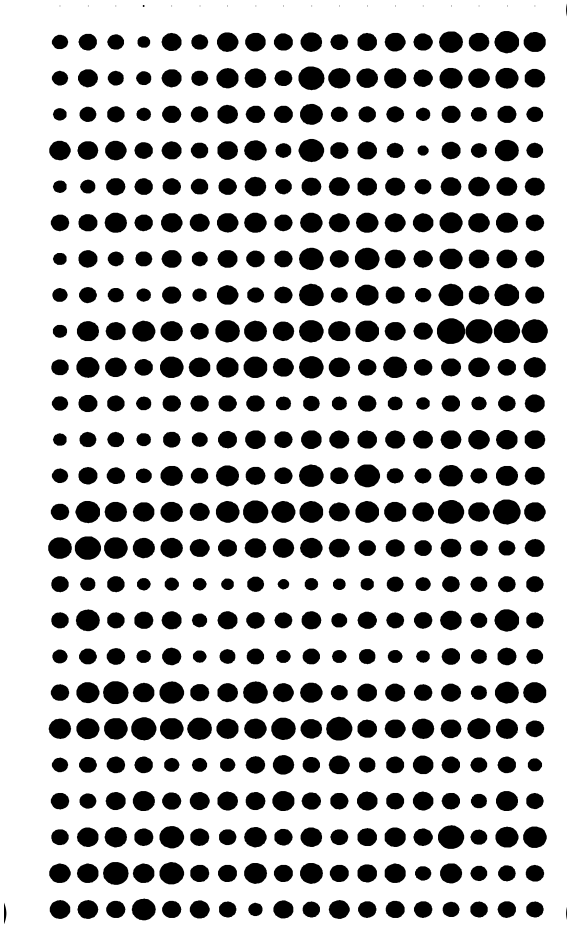
$$D(\omega, m_x, m_y, h, \phi) \rightarrow D(\omega, i, j, k, l) \rightarrow \mathbf{D}$$

$\mathbf{D}$: 4th order tensor

Real Data Example (WCSB)

- Regularization of Fold

$$\mathbf{D}^{obs} = \mathbf{SD}$$



Fold Map

PMF (Xu et al., 2013; Gao, Sacchi, Stanton 2015)

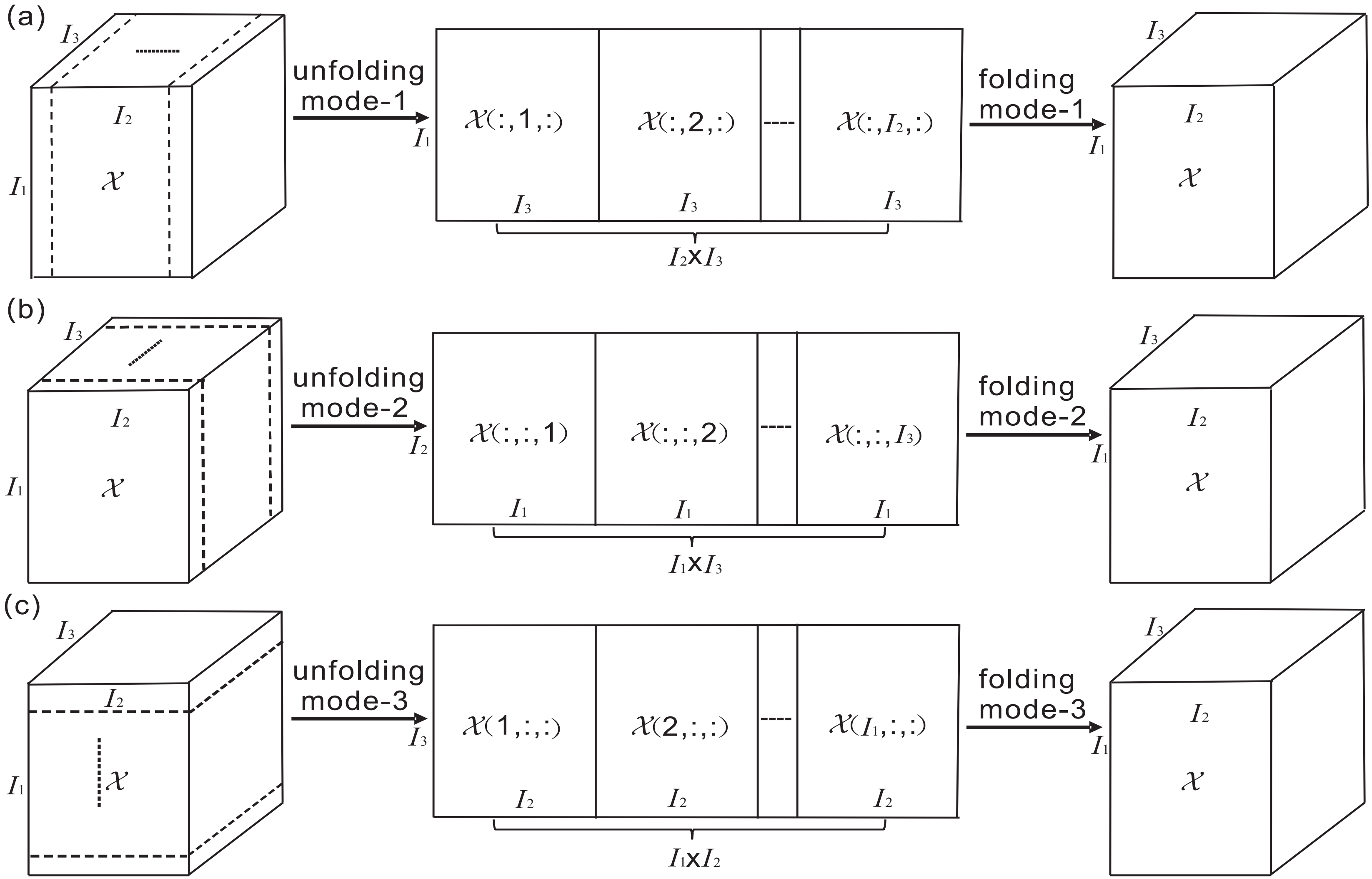
- In the Parallel Matrix Factorization method we minimize

$$\| \mathbf{D}^{obs} - \mathbf{S}\mathbf{D} \|_F$$

- Subject to

$$\mathrm{Rank}\,[\mathrm{Unfold}_i(\mathbf{D})] = k_i \quad i = 1,2,3,4$$

Unfolding and folding



Reconstruction algorithm

- The math leads to a simple algorithm

$$\mathbf{D}^{n} = \alpha \mathbf{D}^{obs} + (1 - \alpha \mathbf{S}) R[\mathbf{D}^{k-1}]$$

R : Rank reduction over all modes (PMF, HO-SVD, R-QR etc)

Note:

R: Amplitude Thresholding (POCS, Abma & Kabir 2006)

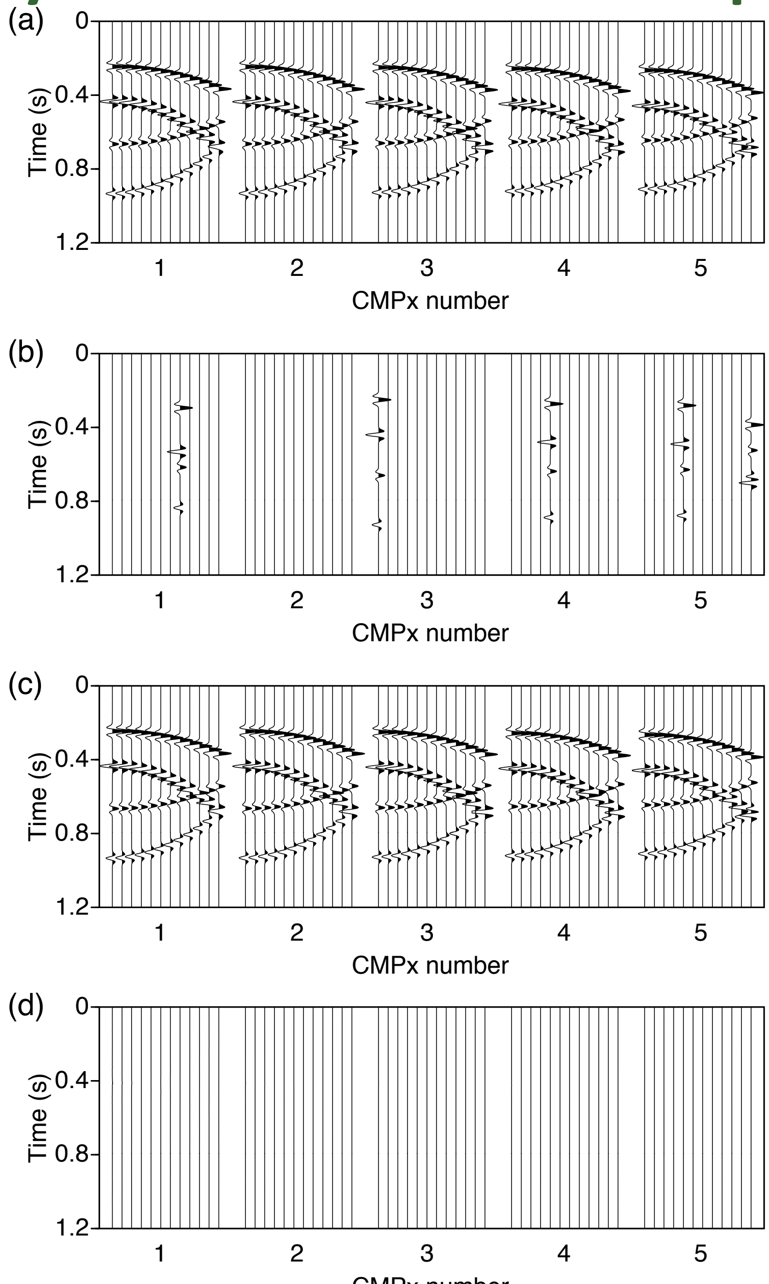
Reconstruction algorithm

- In PMF

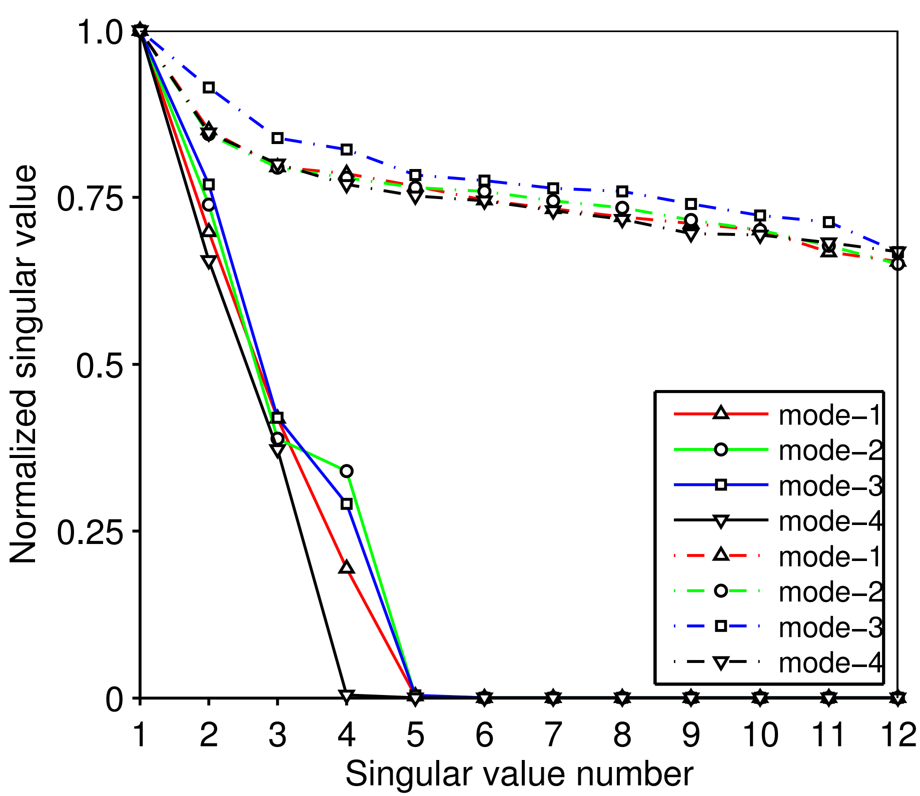
$$\mathbf{D}^{n} = \alpha \mathbf{D}^{obs} + (1 - \alpha \mathbf{S}) R[\mathbf{D}^{k-1}]$$

$R[\ ]$ = average low rank-approximation over all modes

Synthetic. Size of patch 12X12X12X12

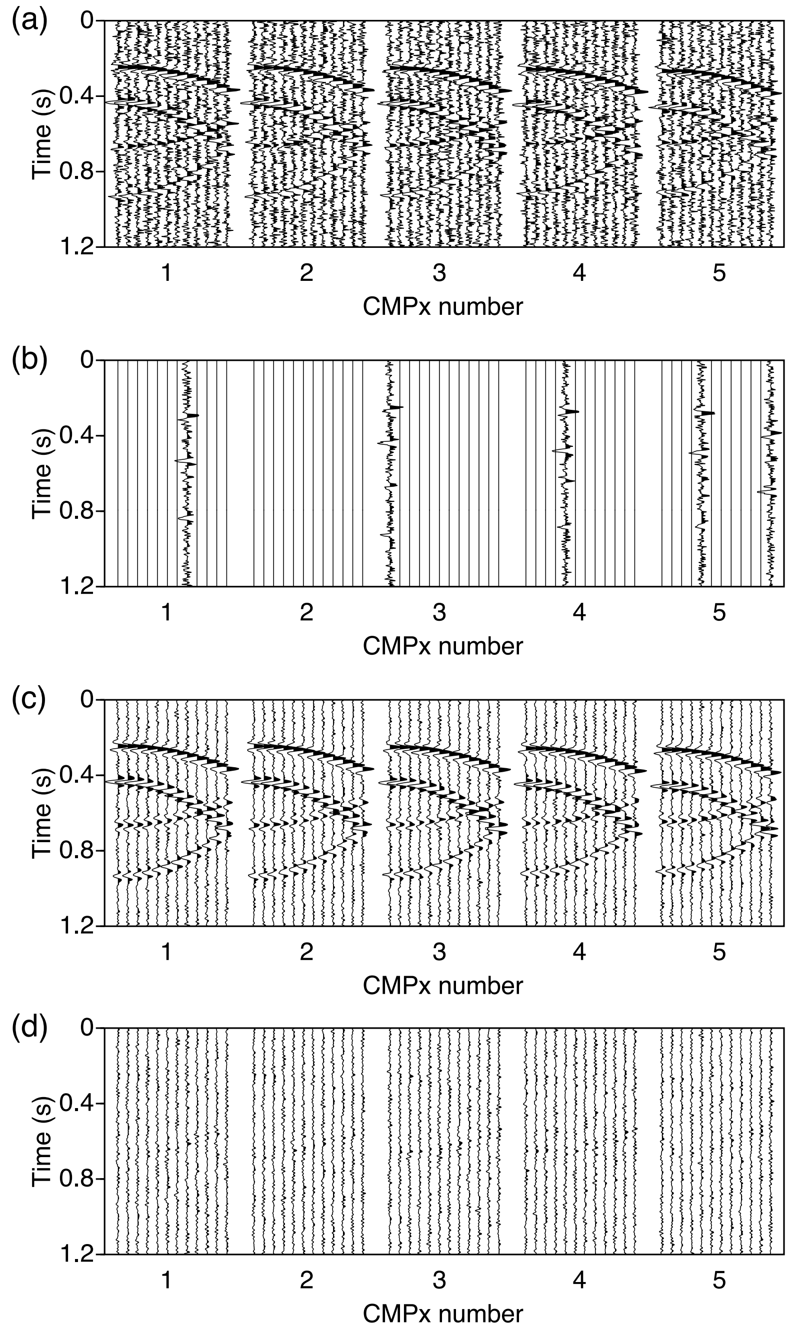


CMPx vs hx for fixed CMPy and hy

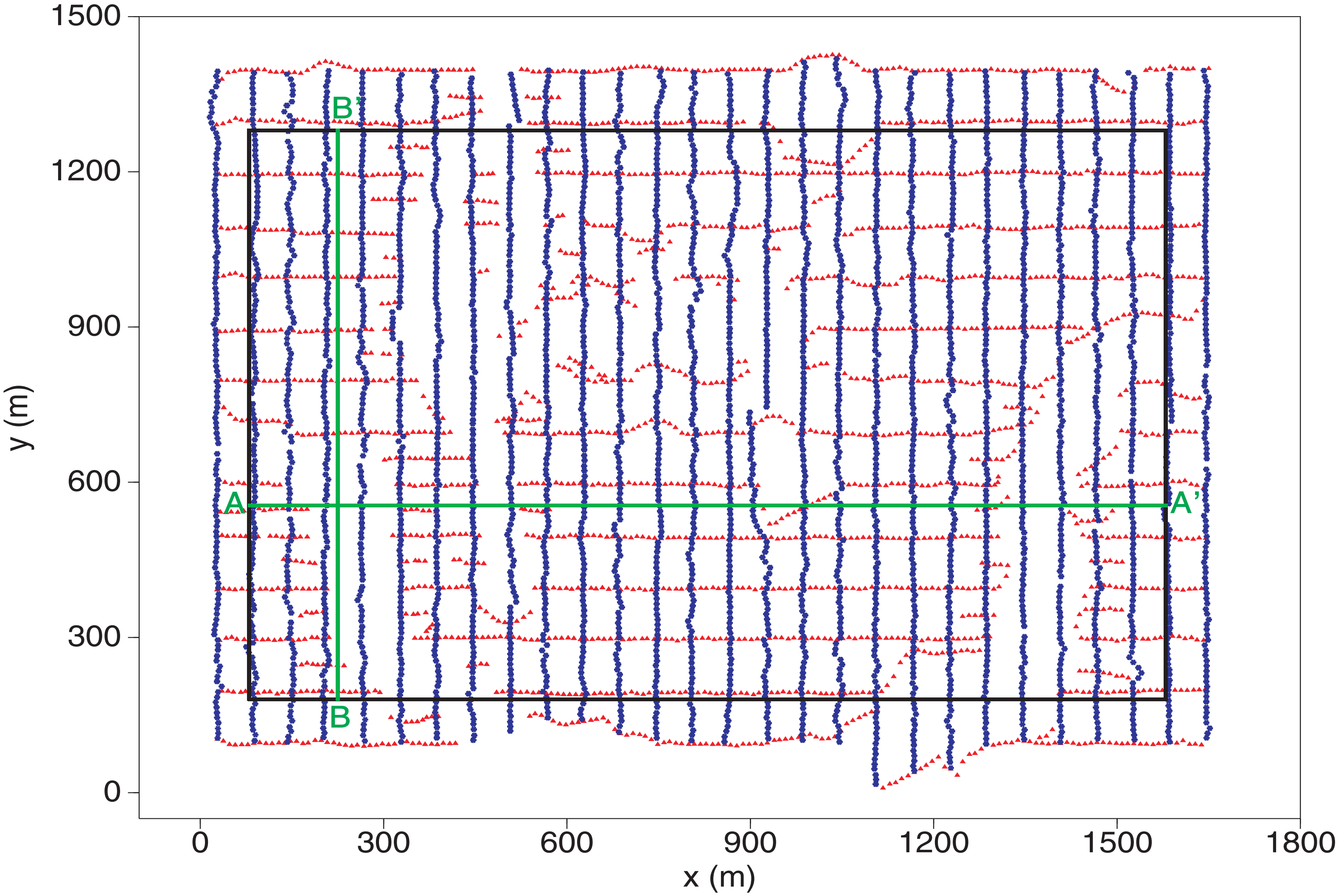


SNR=1

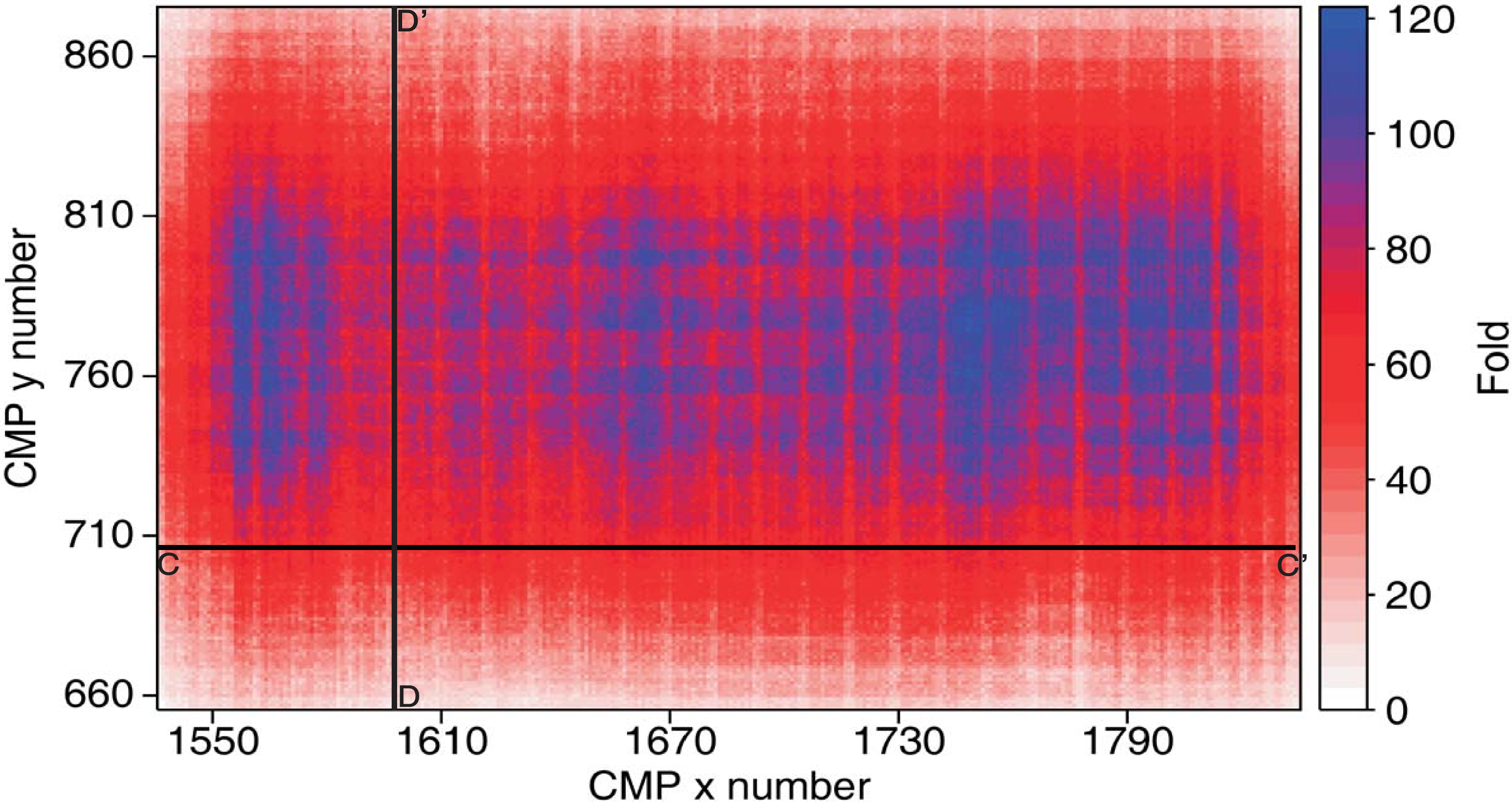
CMPx vs hx for fixed CMPy and hy



Field data example (WCB)



Fold Map



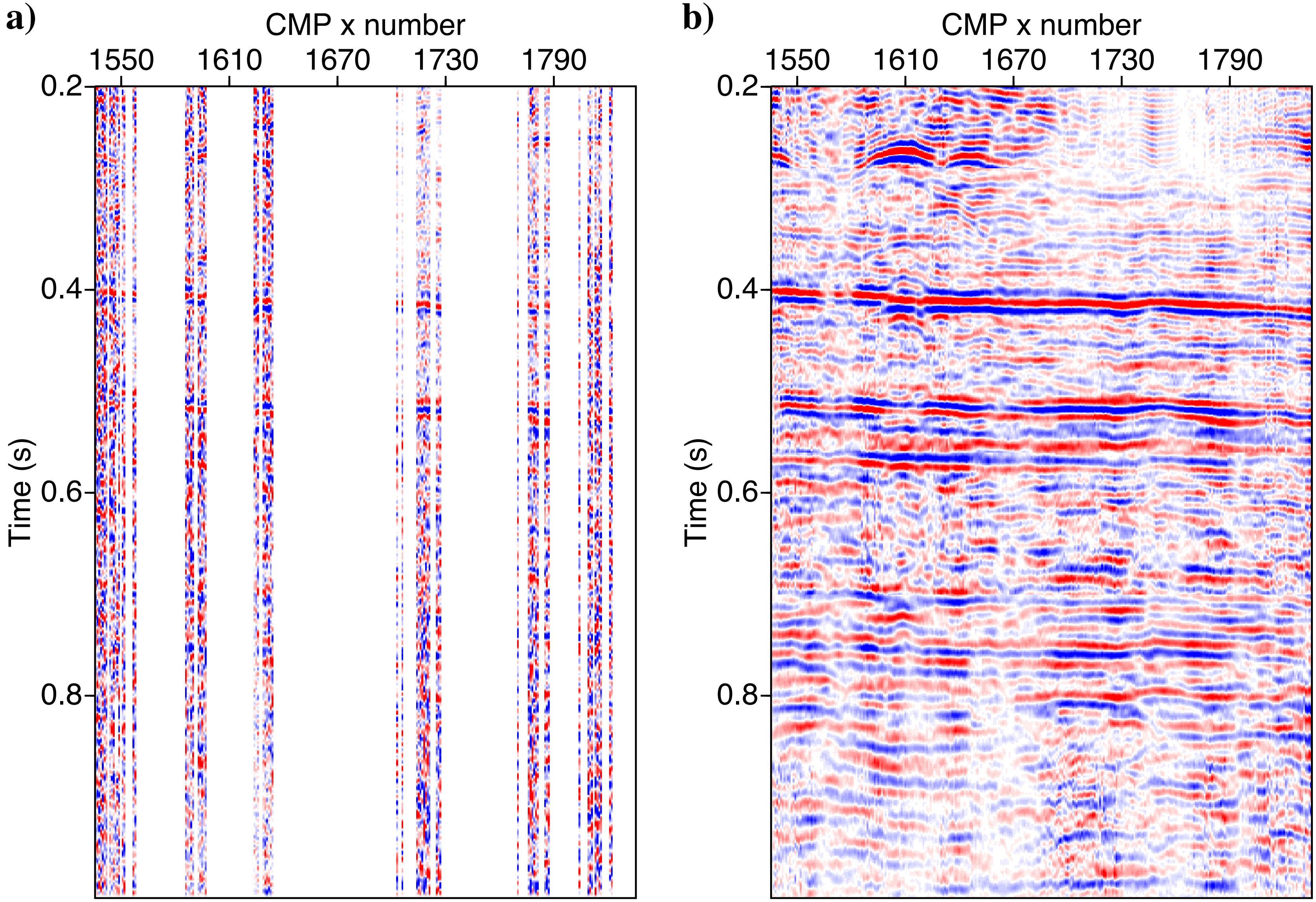
D’
D
C
C’
860
810
760
710
660
CMP y number
1550
1610
1670
1730
1790
CMP x number
120
100
80
60
40
20
0
Fold

Processing Parameters

- 5m X 5m CMP Bins
- 100m offset sectors (x and y)
- 300 CMPx and 220 CMPy bins
- All survey was divided in 2640 overlapping blocks
- Each block has about 85% of missing traces (15% alive)
- First part of analysis is with reconstruction in offset-midpoint

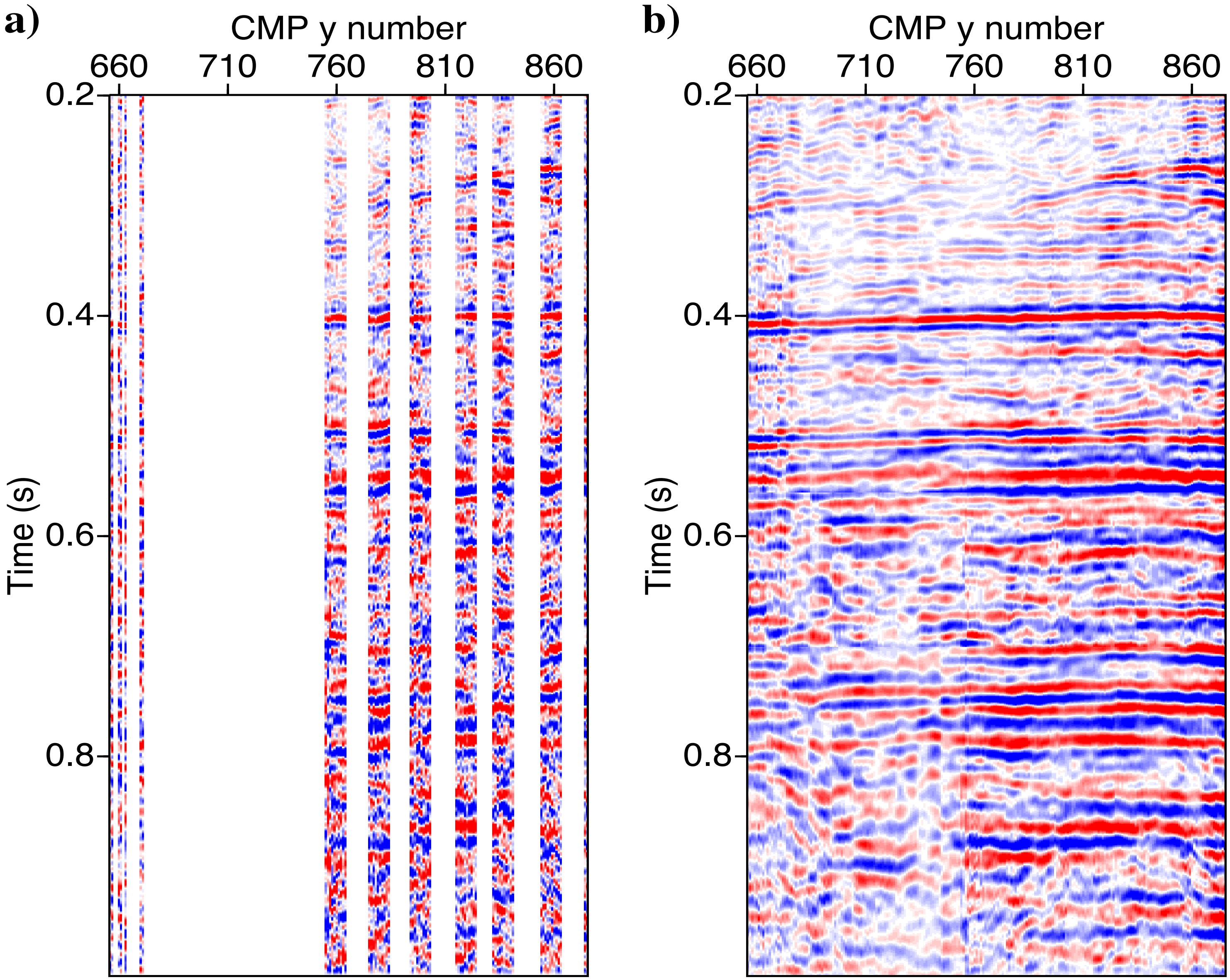
Fix offsets and CMPy

Near offset sector



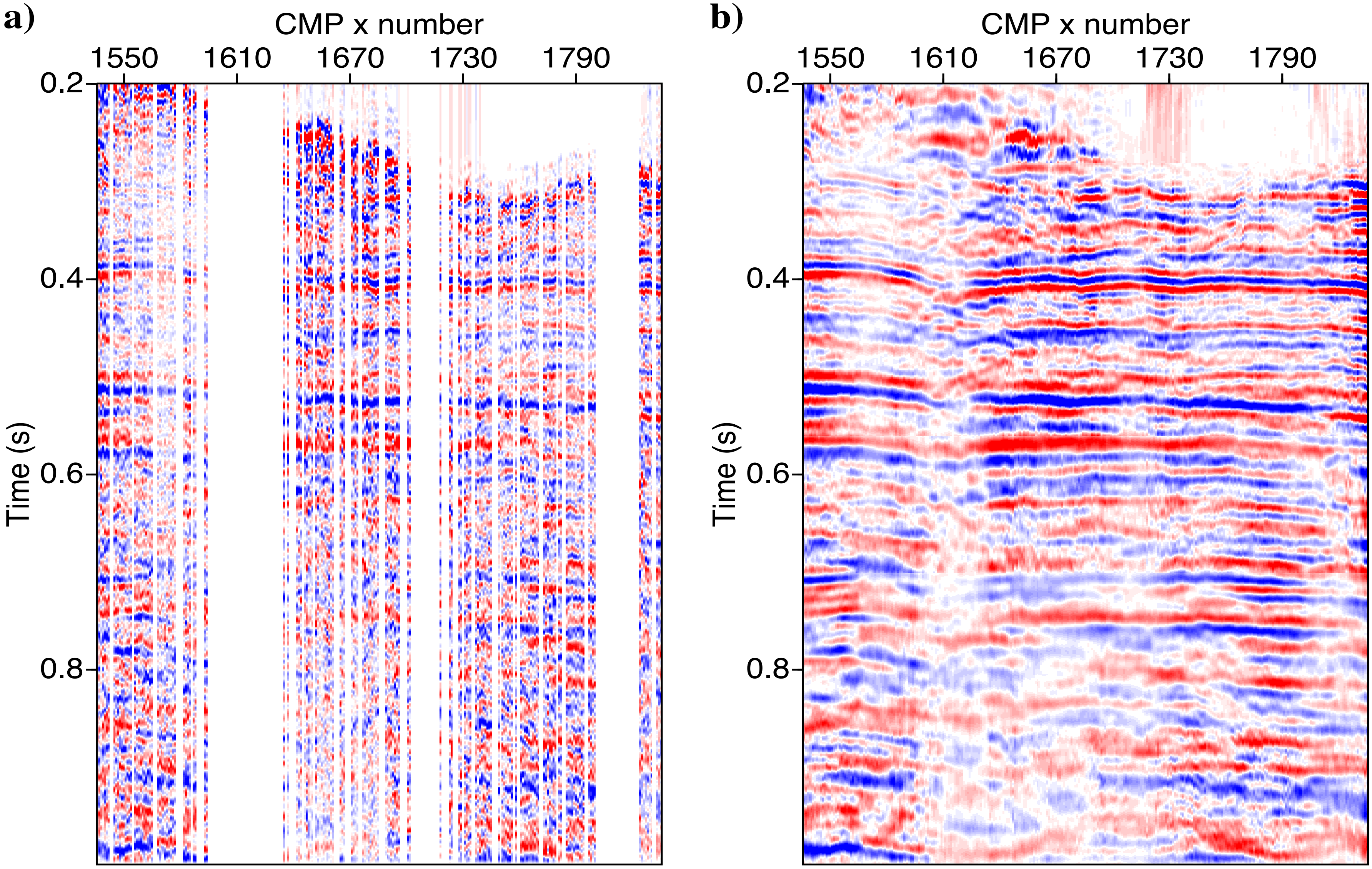
Fix offsets and CMPx

Near offset sector



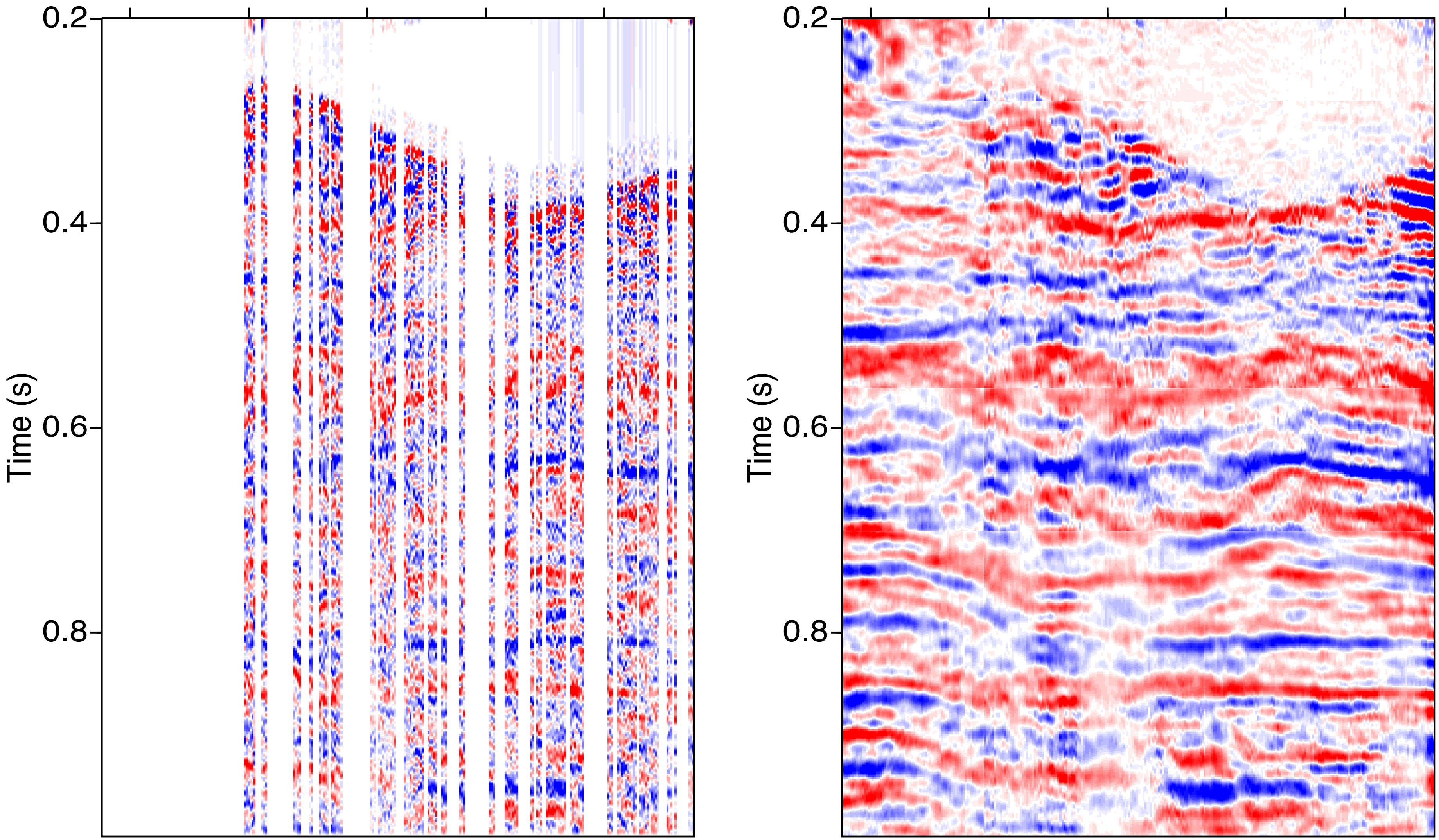
Fix offsets and CMPy

Mid range offset sector

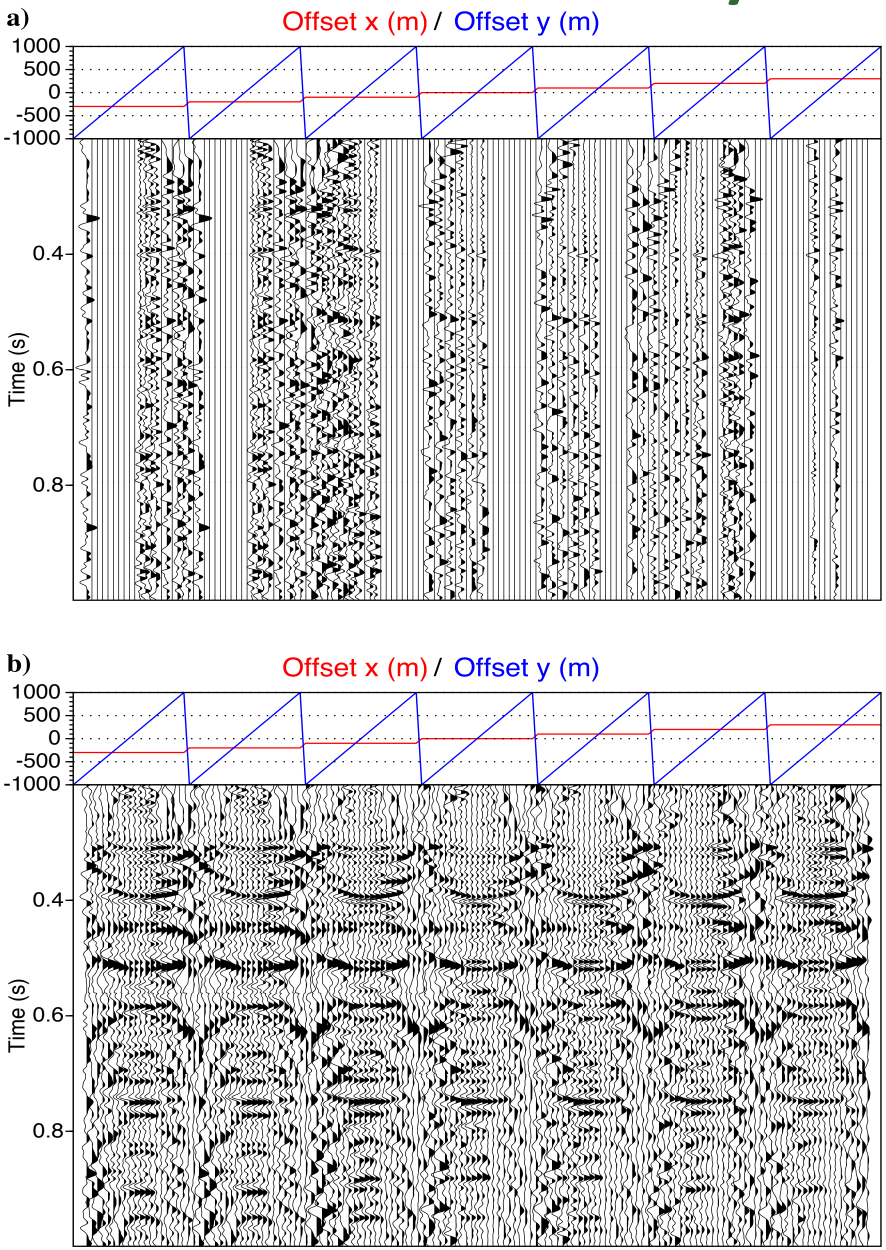


Fix offsets and CMPy

Mid range offset sector

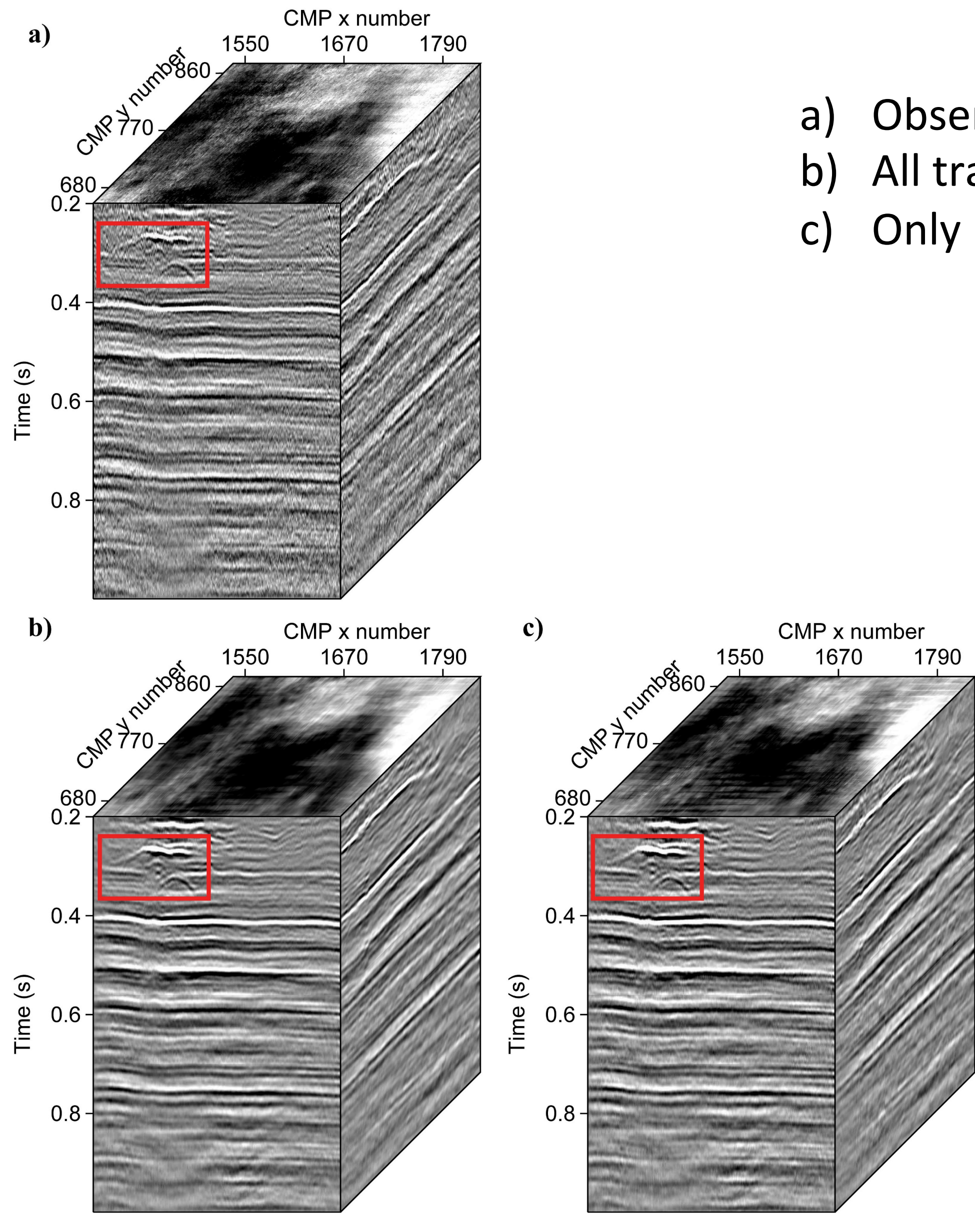


Fix CMPx and CMPy



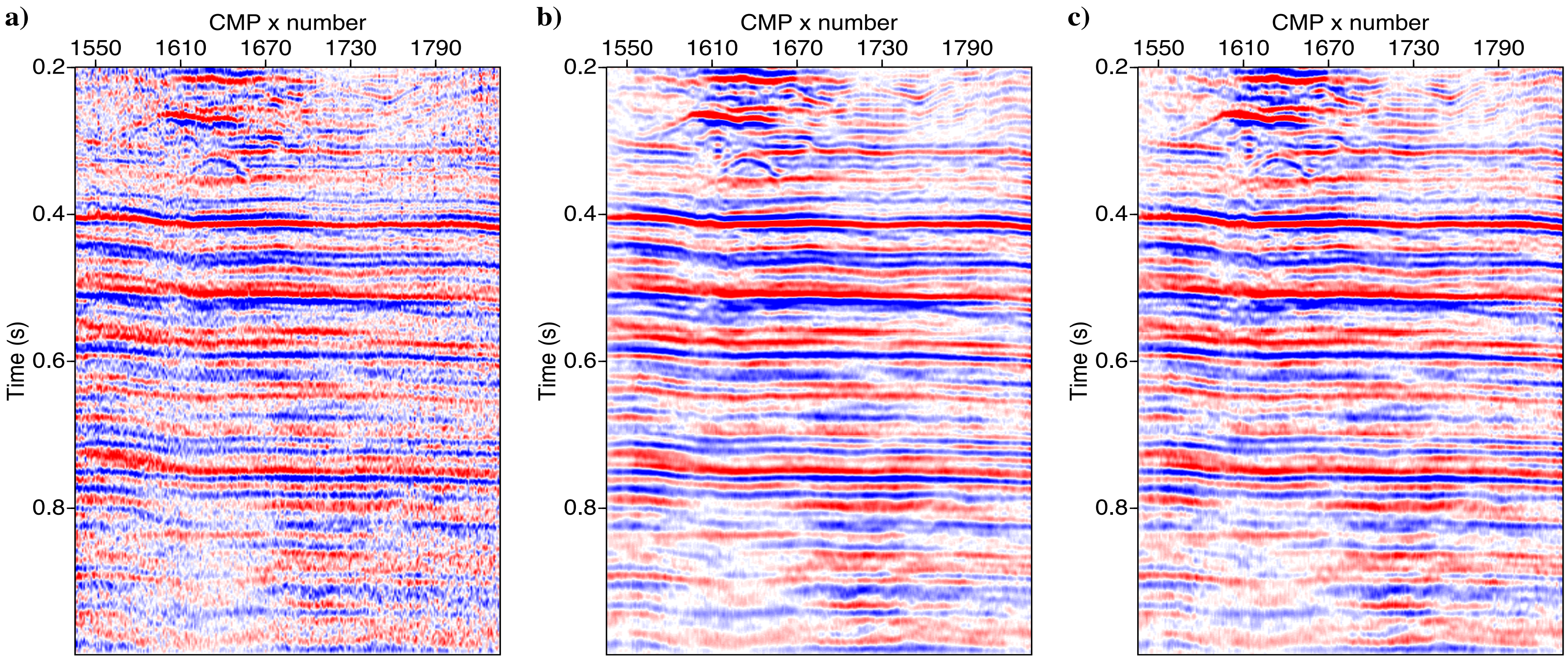
Before and after reconstruction of one CMP

Stacks



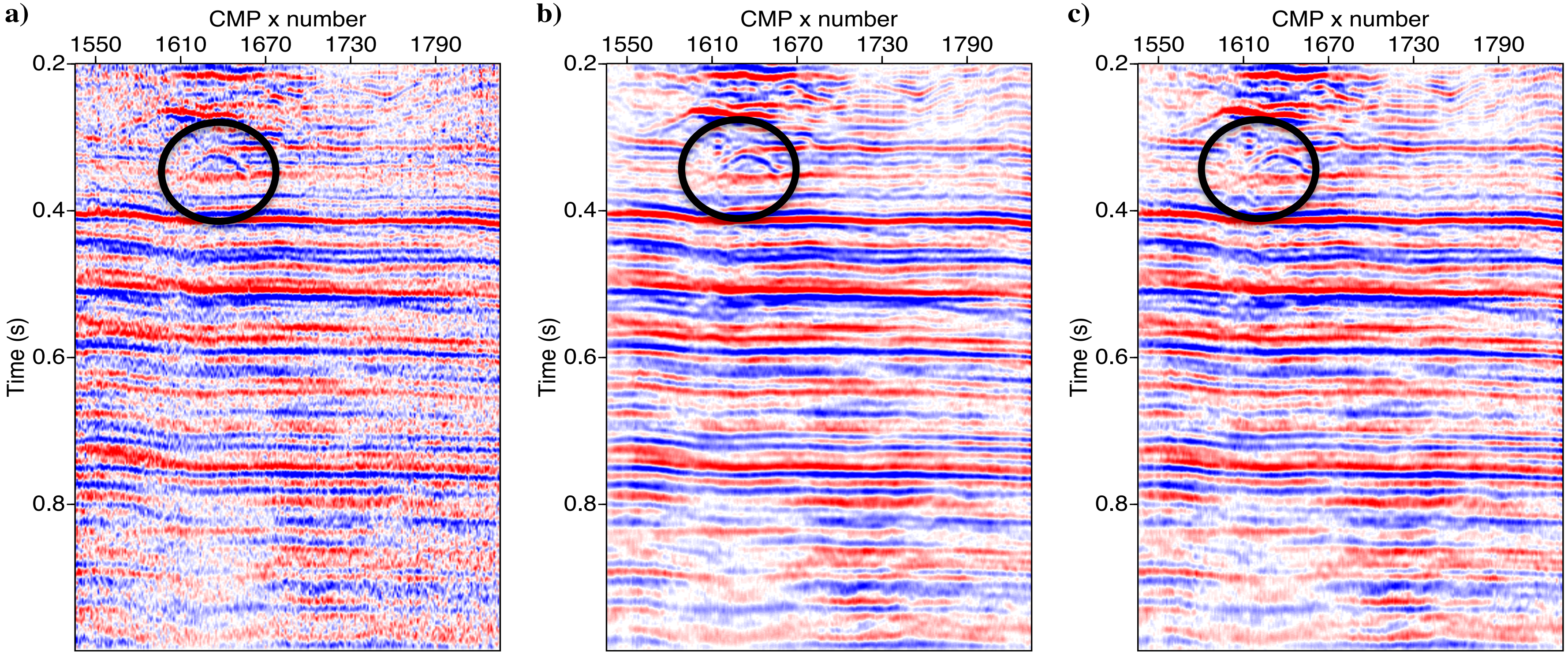
a) Observations
b) All traces (Observed + Reconstructed)
c) Only new traces

Stacks



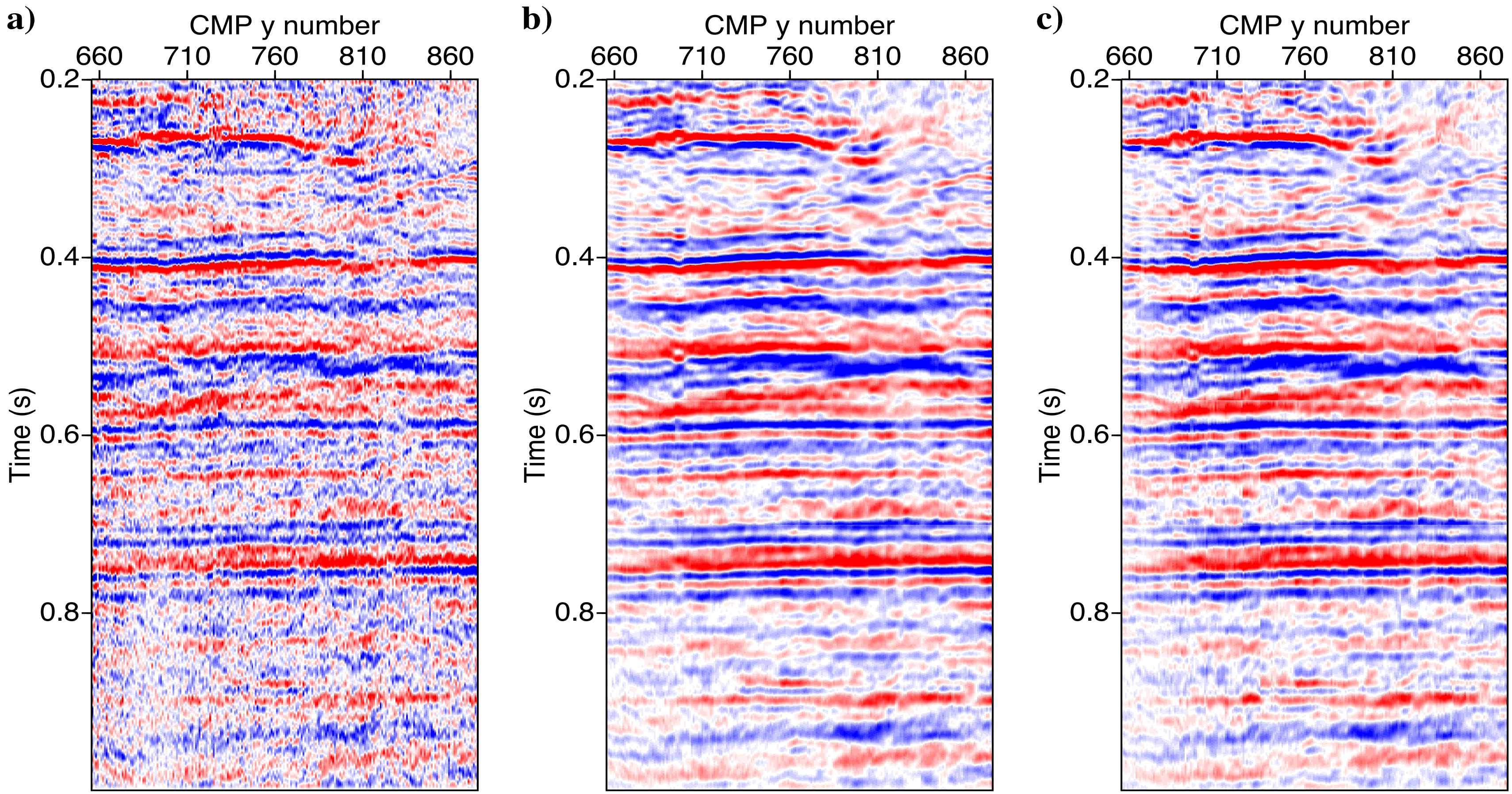
a) Observations b) All traces (Obs +Reconstructed) c) Only new traces

Stacks



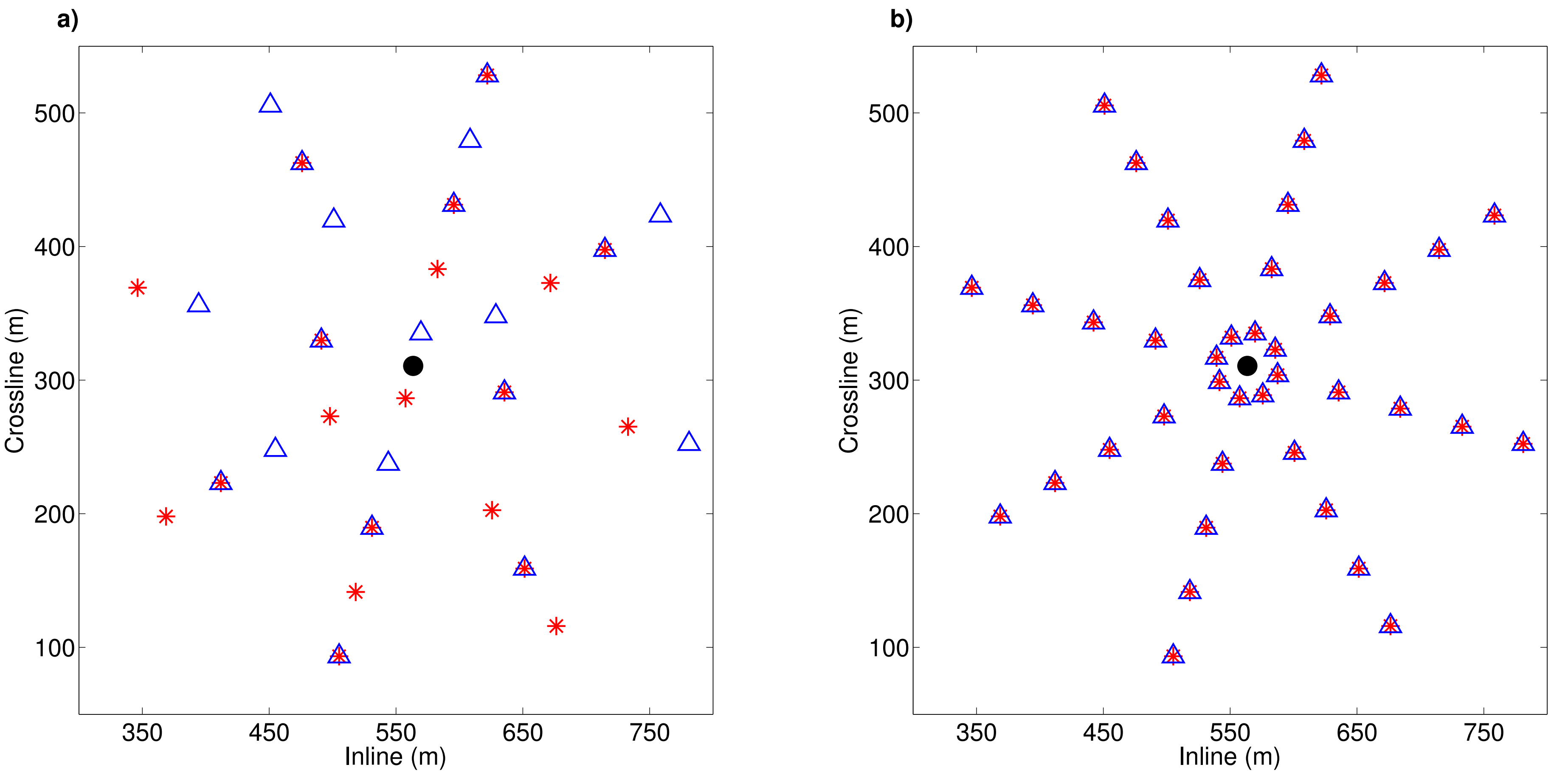
a) Observations b) All traces (Obs +Reconstructed) c) Only new traces

Stacks



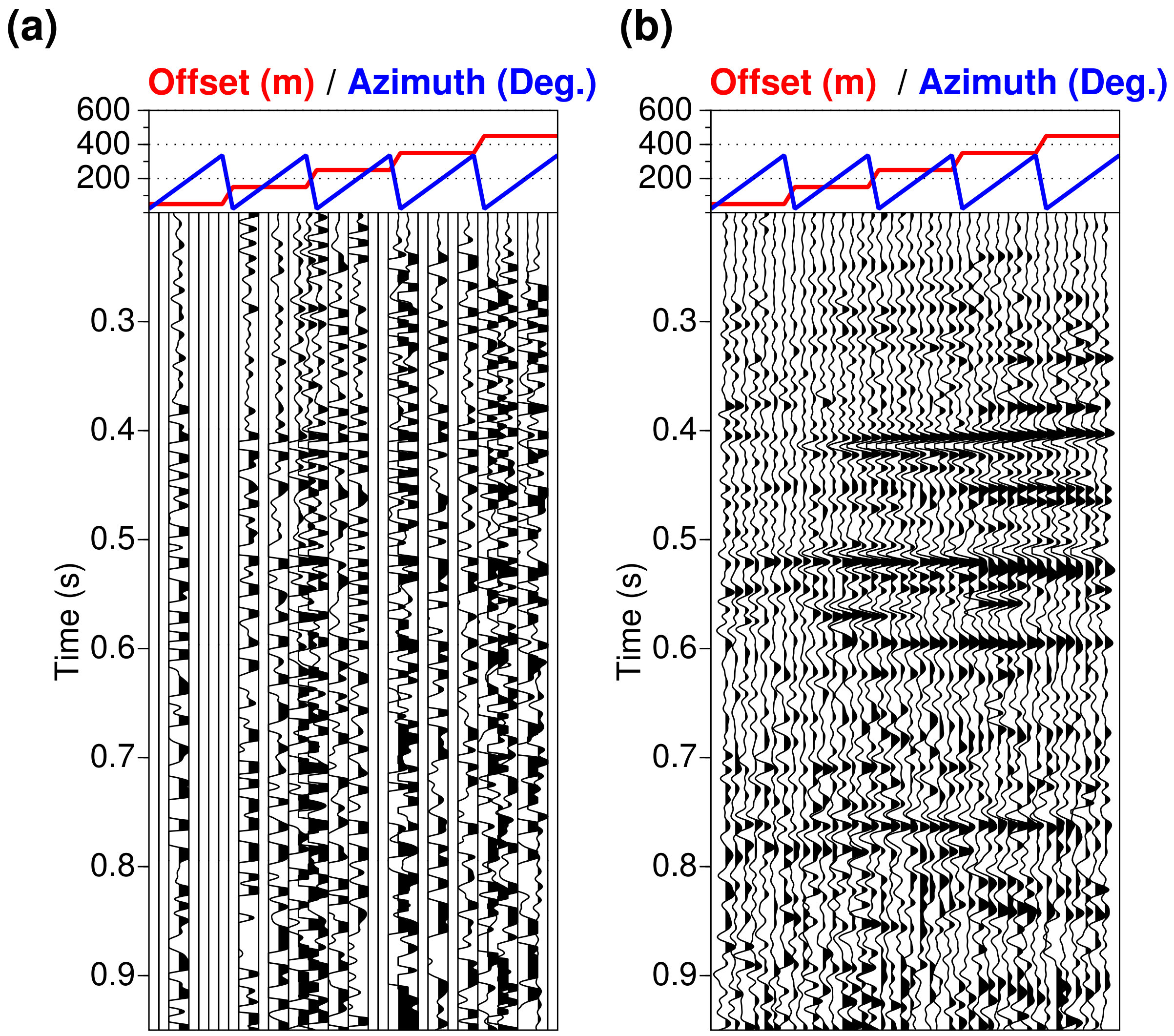
a) Observations b) All traces (Obs +Reconstructed) c) Only new traces

Tensor completion in azimuth offset midpoint

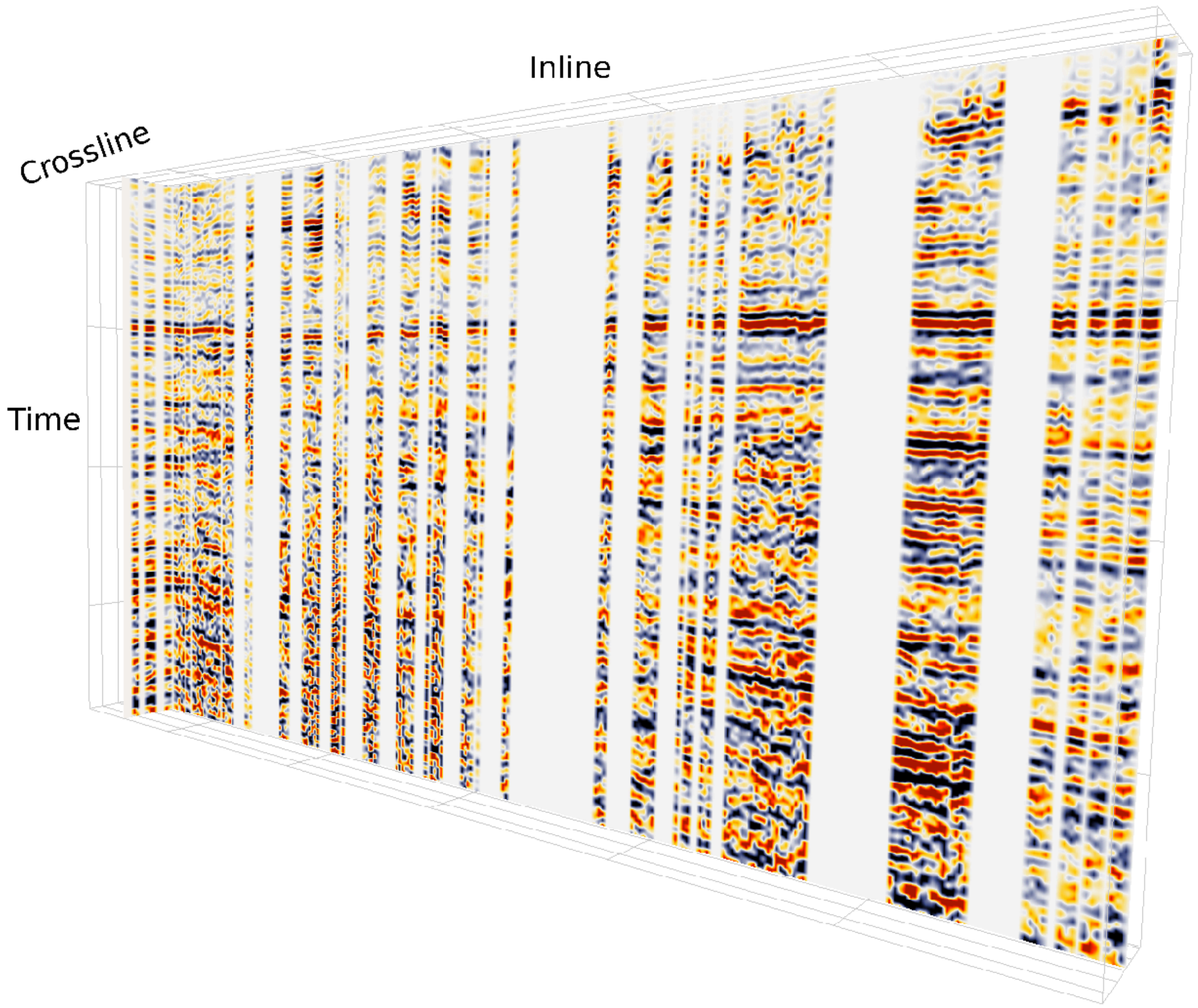


Kreimer et al., 2013

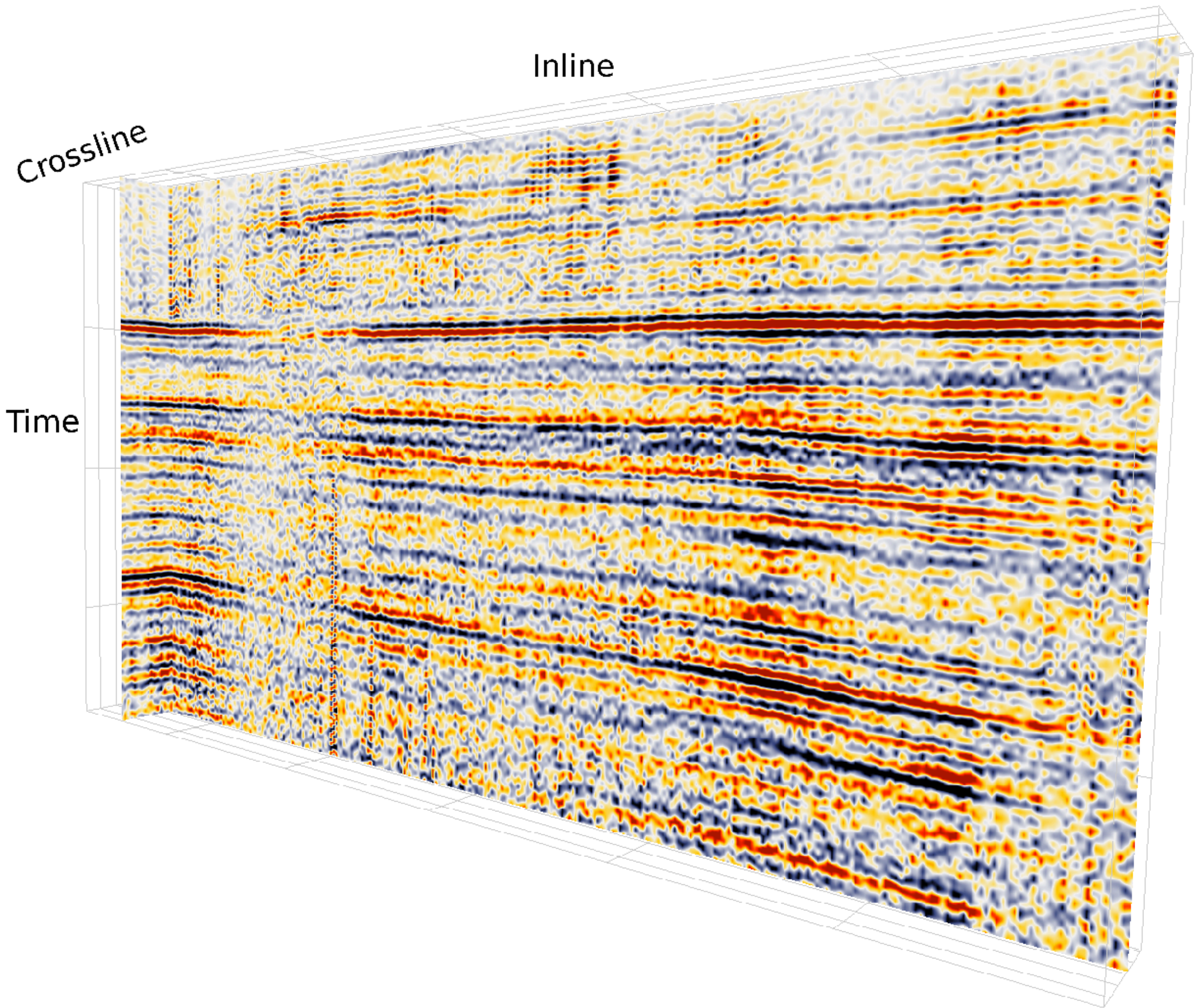
Tensor completion in azimuth offset midpoint



Before

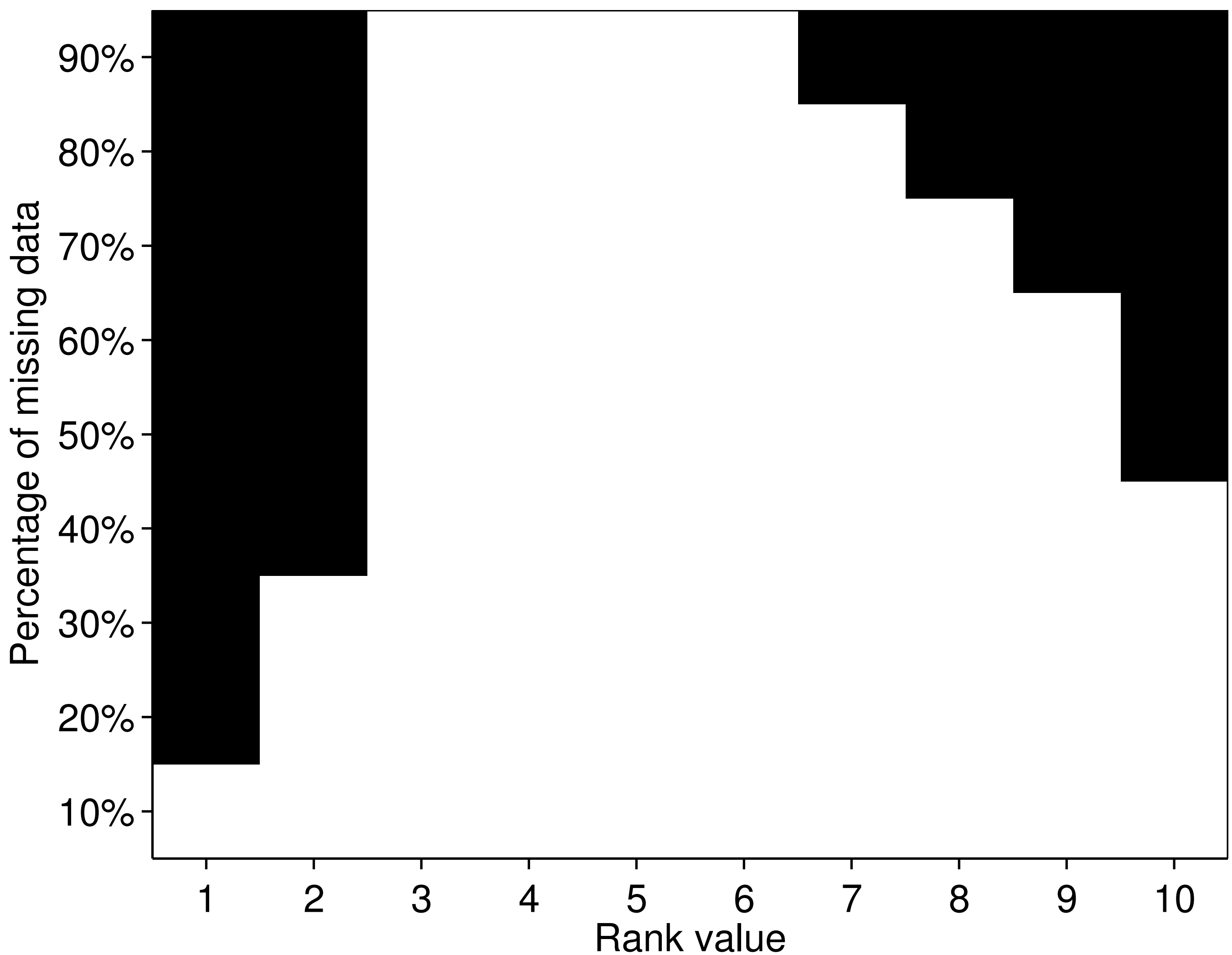


After



Probability of Success (PMF)

Black: recovery was unsuccessful
White: recovery was successful



Conclusions

- Seismic data can be represented in terms of low rank matrices and/or tensors
- In the past, reduced-rank methods for matrices have been used primarily to denoise seismic data
- We are starting to understand how to use multi-linear algebra methods to reconstruct seismic data
- Tensor completion for 5D seismic data reconstruction can cope with spatially varying dips (a problem for reconstruction methods based on Fourier synthesis)

Acknowledgments

- Sergio and Clement for the invitation to speak at this session
- Sponsors of SAIG at the University of Alberta
- NSERC