Released to public domain under Creative Commons license type BY (https://creativecommons.org/licenses/by/4.0).
Copyright (c) 2019 SLIM group @ Georgia Institute of Technology.

Deep-learning based ocean bottom seismic wavefield recovery

Ali Siahkoohi, Rajiv Kumar, and Felix J. Herrmann

SLIM

Georgia Institute of Technology

SEG19
SAN ANTONIO, TX
15-20 SEP 2019
INTERNATIONAL EXPOSITION
89TH ANNUAL MEETING

Problem setup

Ocean bottom node (OBN) geometry:

- assume a desirable source sampling, via (simultaneous-source) randomized marine acquisition
- very sparse receivers scattered throughout the ocean bottom, **but** on a grid

Objective:

Reconstruct the information in the missing receivers

Why a neural net?

Most of previous methods rely on linear mathematical models:

- superposition of prototype waveforms from a fixed or learned dictionary or in terms of a matrix factorizations

- Particularly, matrix completion can be considered as a two-layer *linear* neural net

Using a *nonlinear* neural net, we find an *implicit* deep factorization

Main contribution

A supervised learning technique for wavefield reconstruction that **does not need any external training data**

i.e., training data is extracted from the acquired data

Seismic data in a 3D survey

Seismic data is 5D: $(t, \mathrm{Src}\ x, \mathrm{Src}\ y, \mathrm{Rec}\ x, \mathrm{Rec}\ y)$

$$\Downarrow$$

Taking Fourier transfer w.r.t. time: $(\omega, \mathrm{Src}\ x, \mathrm{Src}\ y, \mathrm{Rec}\ x, \mathrm{Rec}\ y)$

$$\Downarrow$$

Monochromatic seismic data is 4D: $(\mathrm{Src}\ x, \mathrm{Src}\ y, \mathrm{Rec}\ x, \mathrm{Rec}\ y)$

Demanet, L., 2006, Curvelets, wave atoms, and wave equations: PhD thesis, California Institute of Technology.
Silva, C. D., and F. J. Herrmann, 2013, Hierarchical Tucker tensor optimization - applications to tensor completion: Presented at the , Sampling Theory and Applications conference.

Matricization of monochromatic seismic data

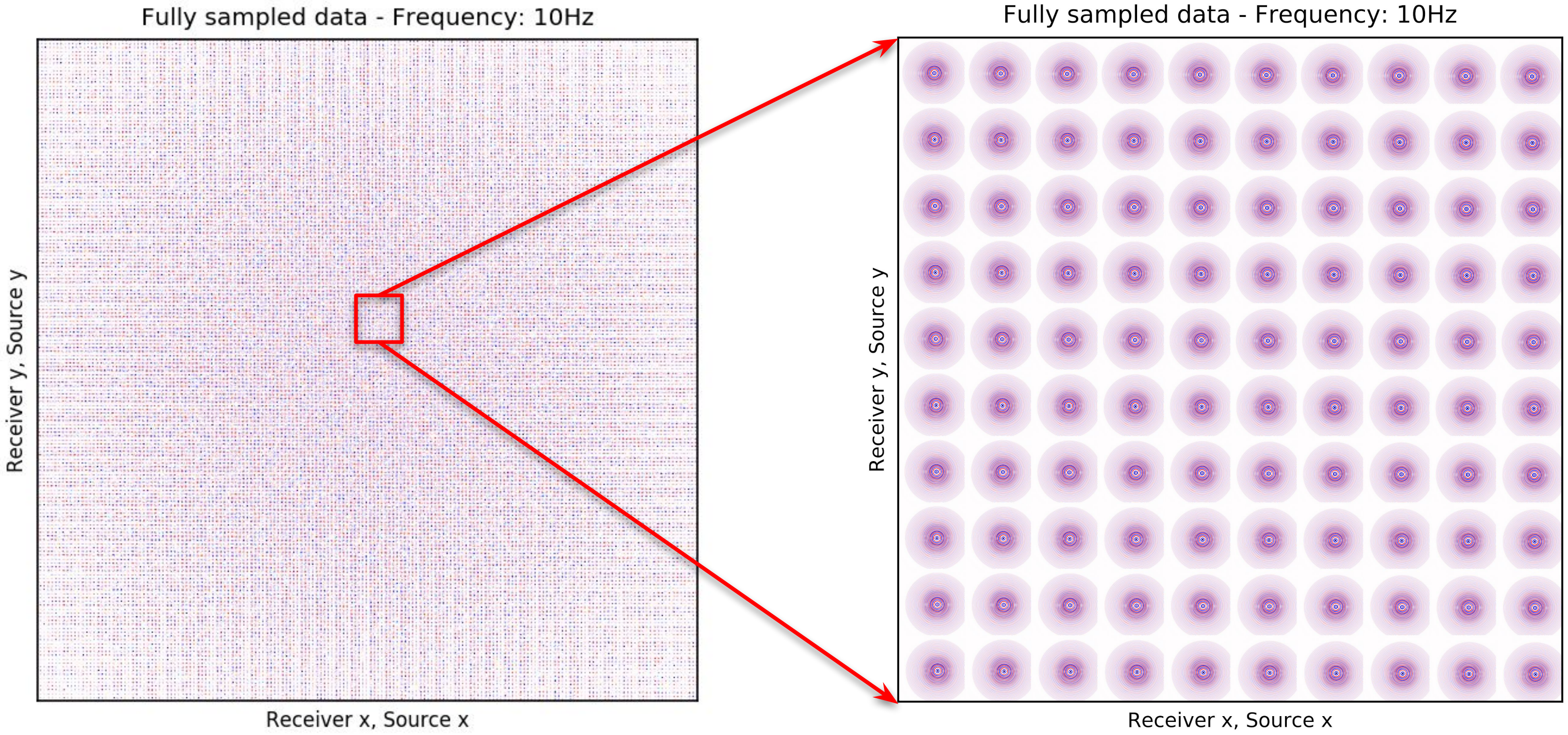
Our framework operators on *monochromatic frequency slices*

Two choices for matricization of monochromatic seismic data

- $(\text{Rec } x,\ \text{Rec } y) \times (\text{Src } x,\ \text{Src } y)$
- $(\text{Rec } y,\ \text{Src } y) \times (\text{Rec } x,\ \text{Src } x)$

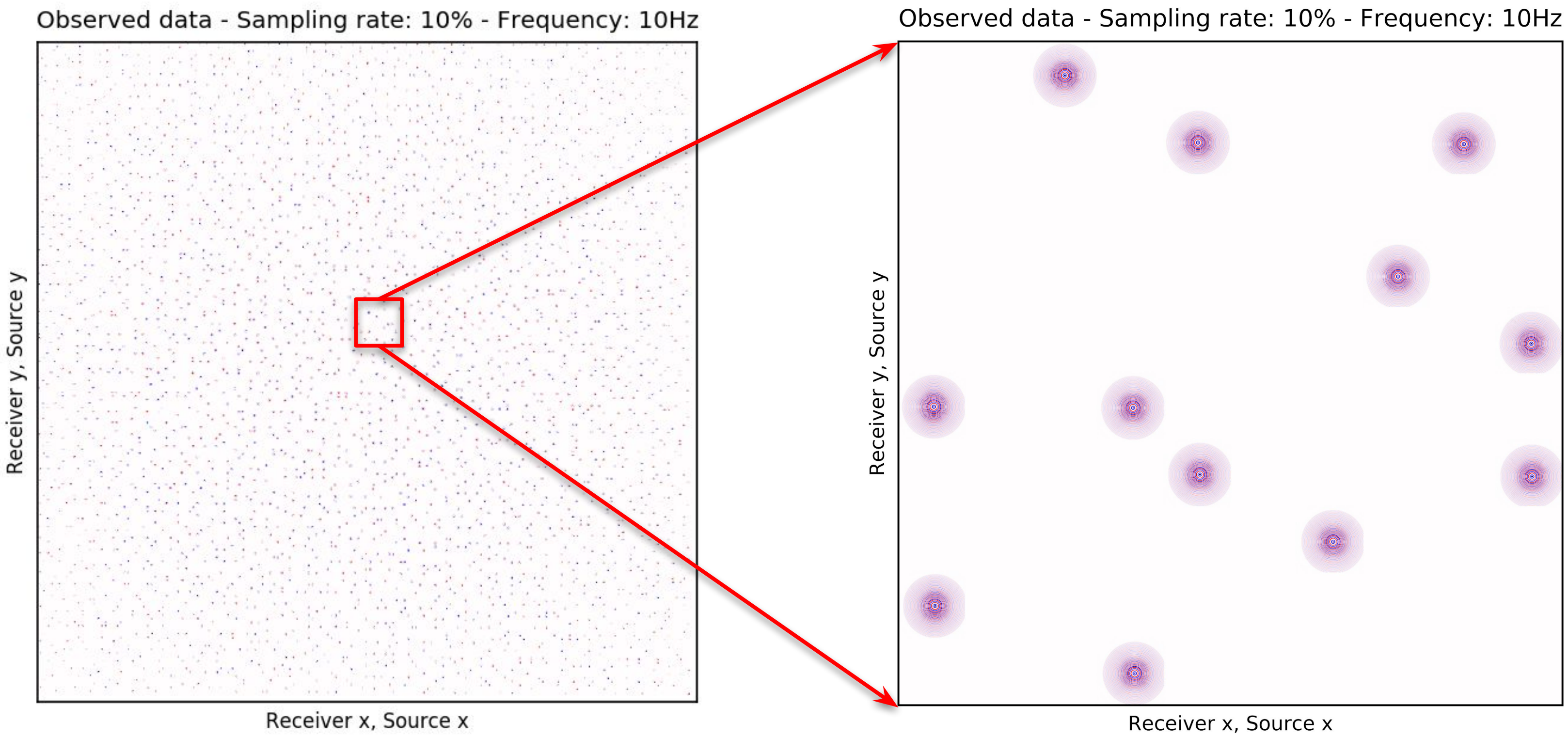
Fully-sampled data

$(\mathbf{Rec\ y},\ \mathbf{Src\ y}) \times (\mathbf{Rec\ x},\ \mathbf{Src\ x})$ domain



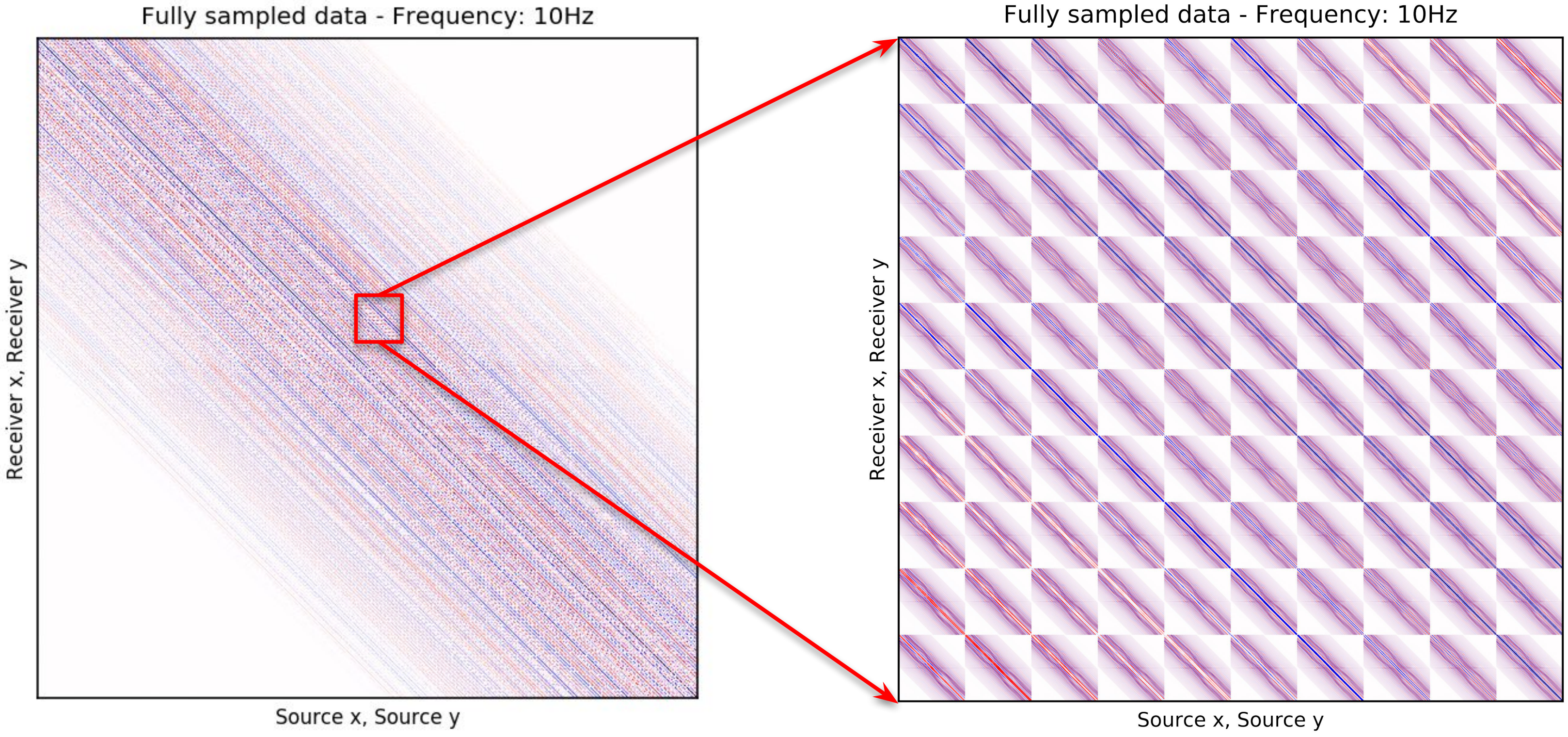
Observed data – Sampling rate 10%

$(\mathbf{Rec\ y,\ Src\ y}) \times (\mathbf{Rec\ x,\ Src\ x})$ domain



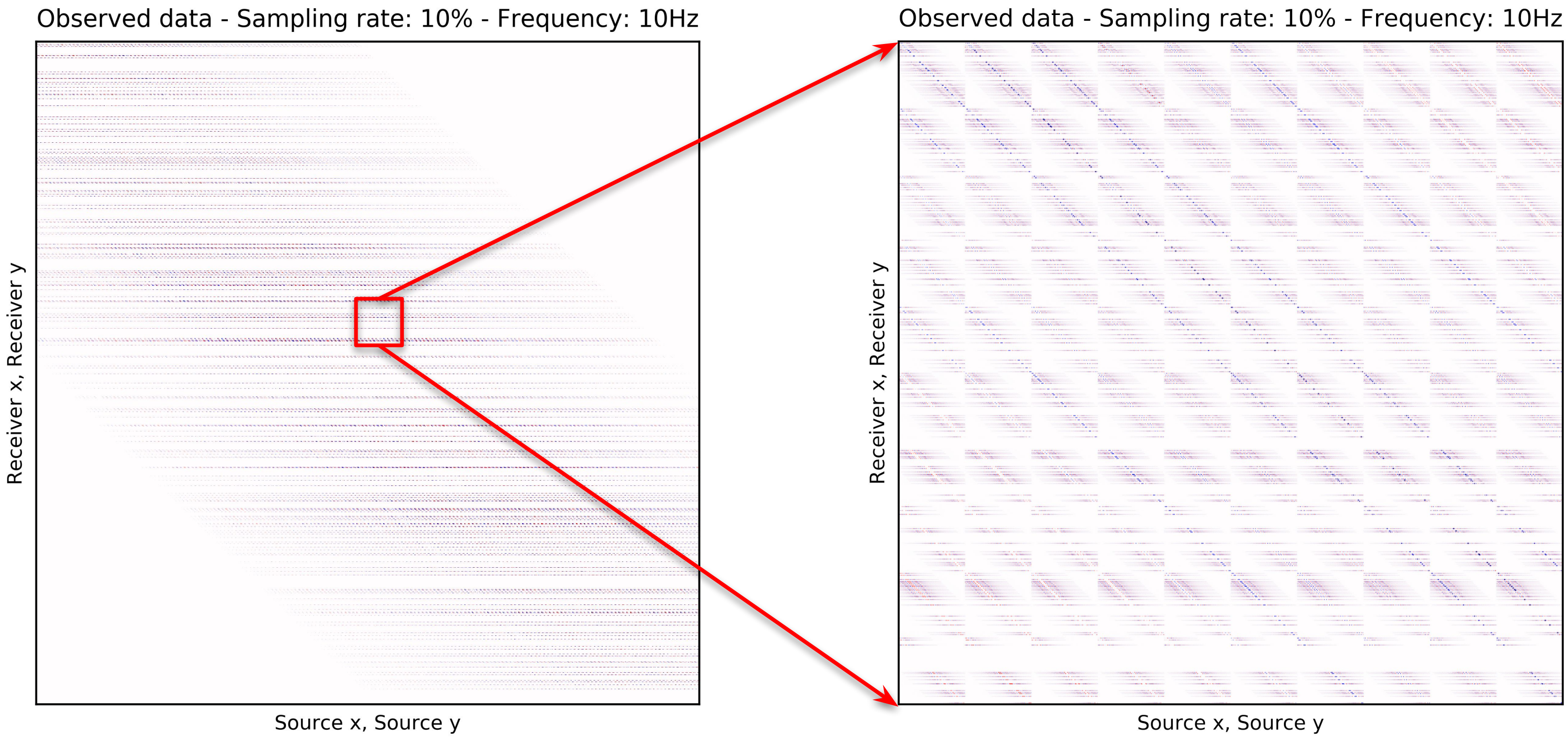
Fully-sampled data

$(\text{Rec x}, \ \text{Rec y}) \times (\text{Src x}, \ \text{Src y})$ domain

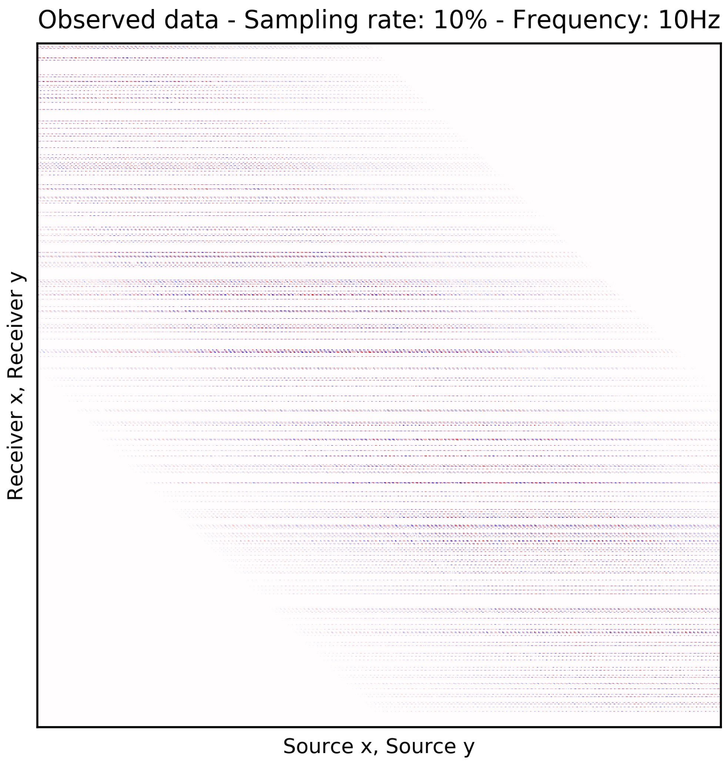


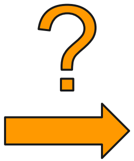
Observed data – Sampling rate 10%

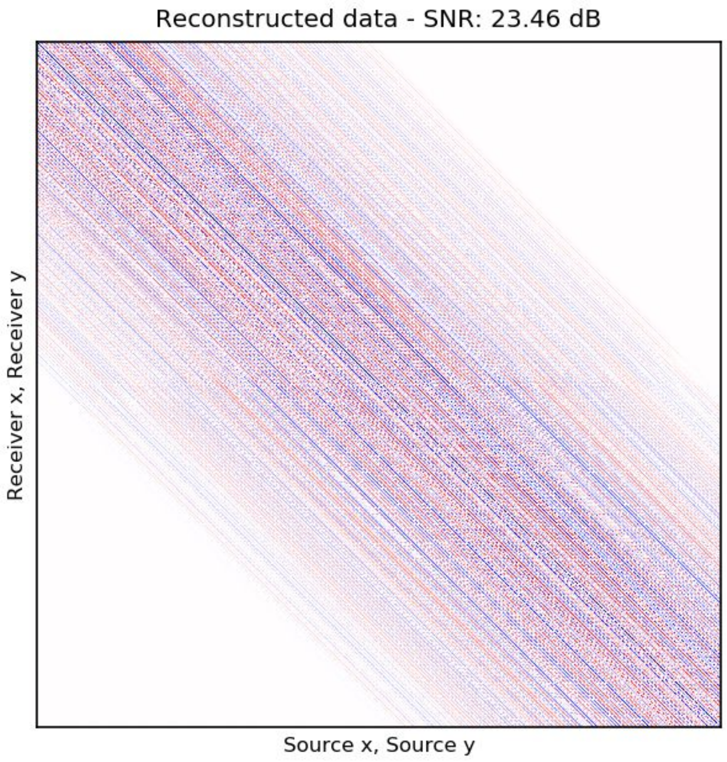
$(\mathrm{Rec\ x},\ \mathrm{Rec\ y}) \times (\mathrm{Src\ x},\ \mathrm{Src\ y})$ domain



Objective: Recovering missing receivers







Proposed method

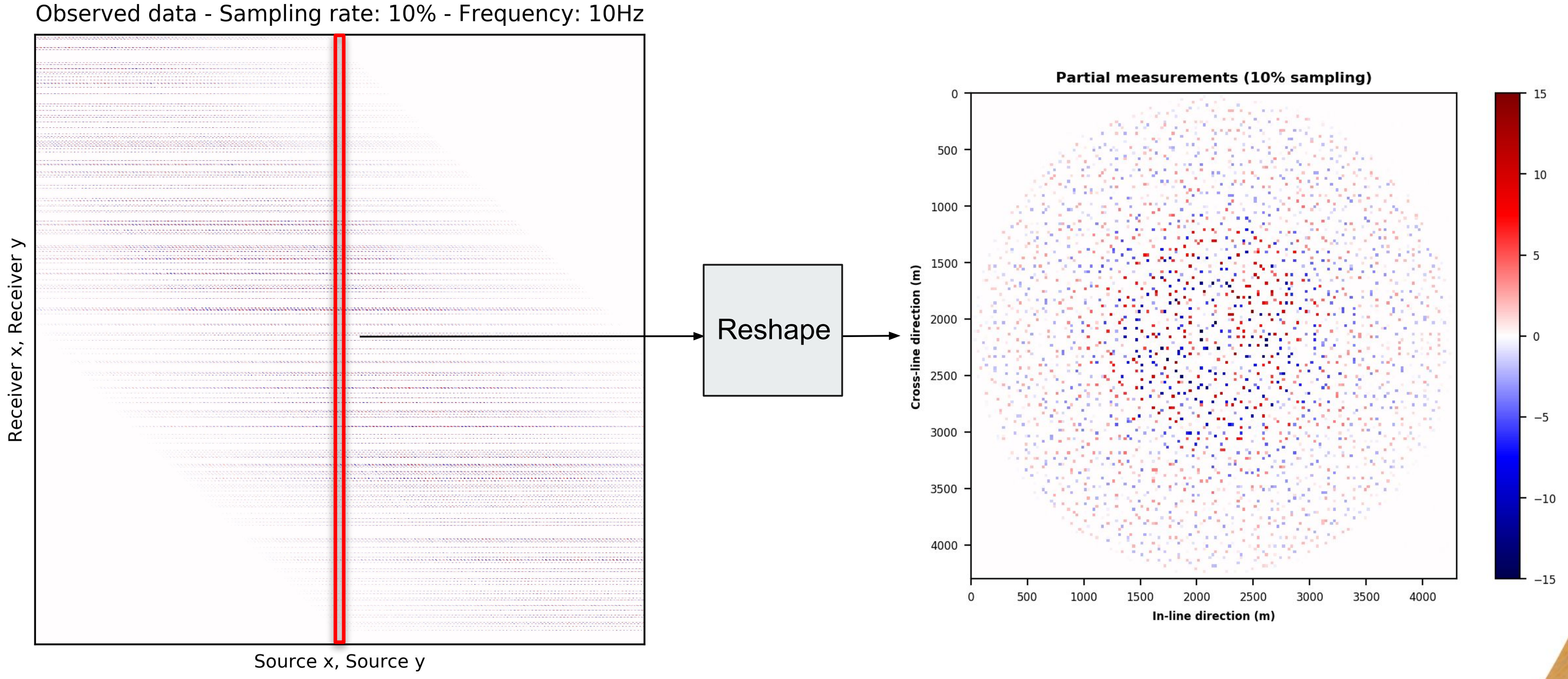
0. Pre-train a neural network (more on this soon. For now, assume we have this)

1. Extract single-source frequency slices

   i.e., columns of $(\text{Rec } x,\ \text{Rec } y) \times (\text{Src } x,\ \text{Src } y)$

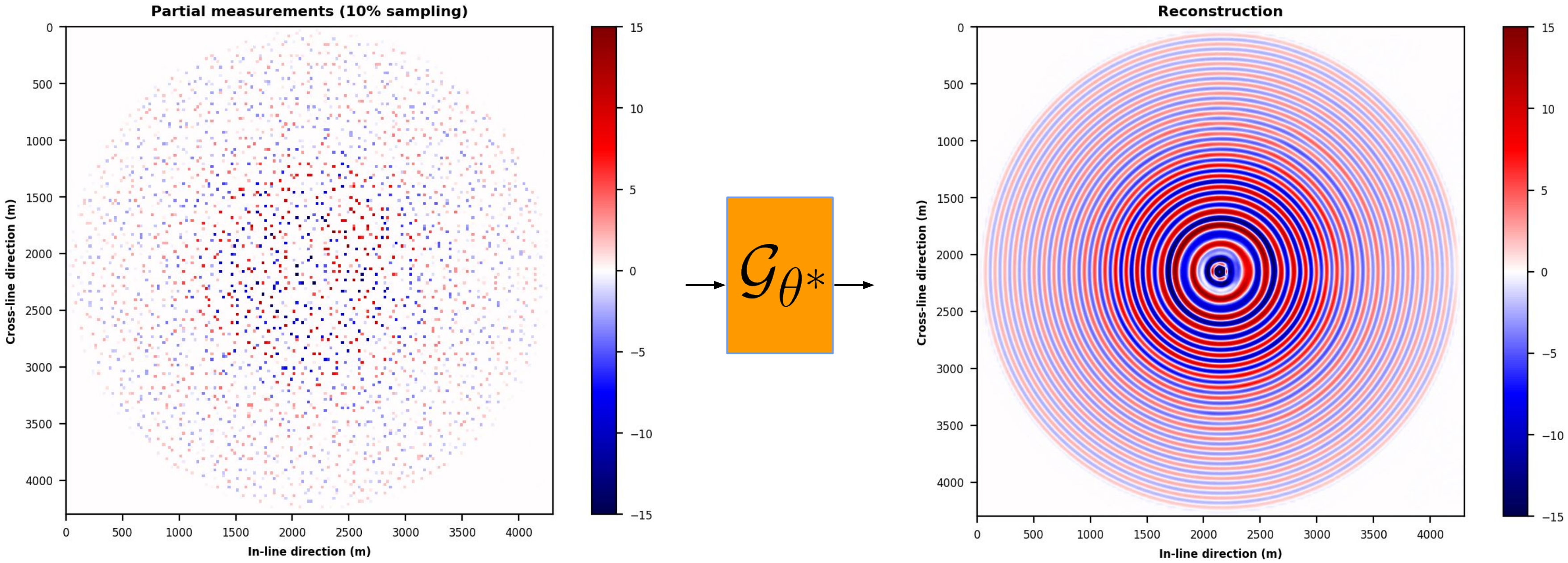
2. Reconstruct the missing values by feeding the extracted slices to the pre-trained neural network

Proposed method:
Step 1: Extract and reshape



Proposed method:

Step 2: Reconstruction via the pre-trained neural net



$\mathcal{G}_{\theta^*}$ is the pre-trained neural net.

Pre-training a neural net: Training data

Problem:

► we need training data pairs, i.e., subsampled and fully-sampled frequency slices

Solution:

► extract fully-sampled single-receiver frequency slices and subsample them with an arbitrary *training mask*

Underlying assumption:

► source-receiver reciprocity holds + dense source sampling

Steps to Extract training pairs

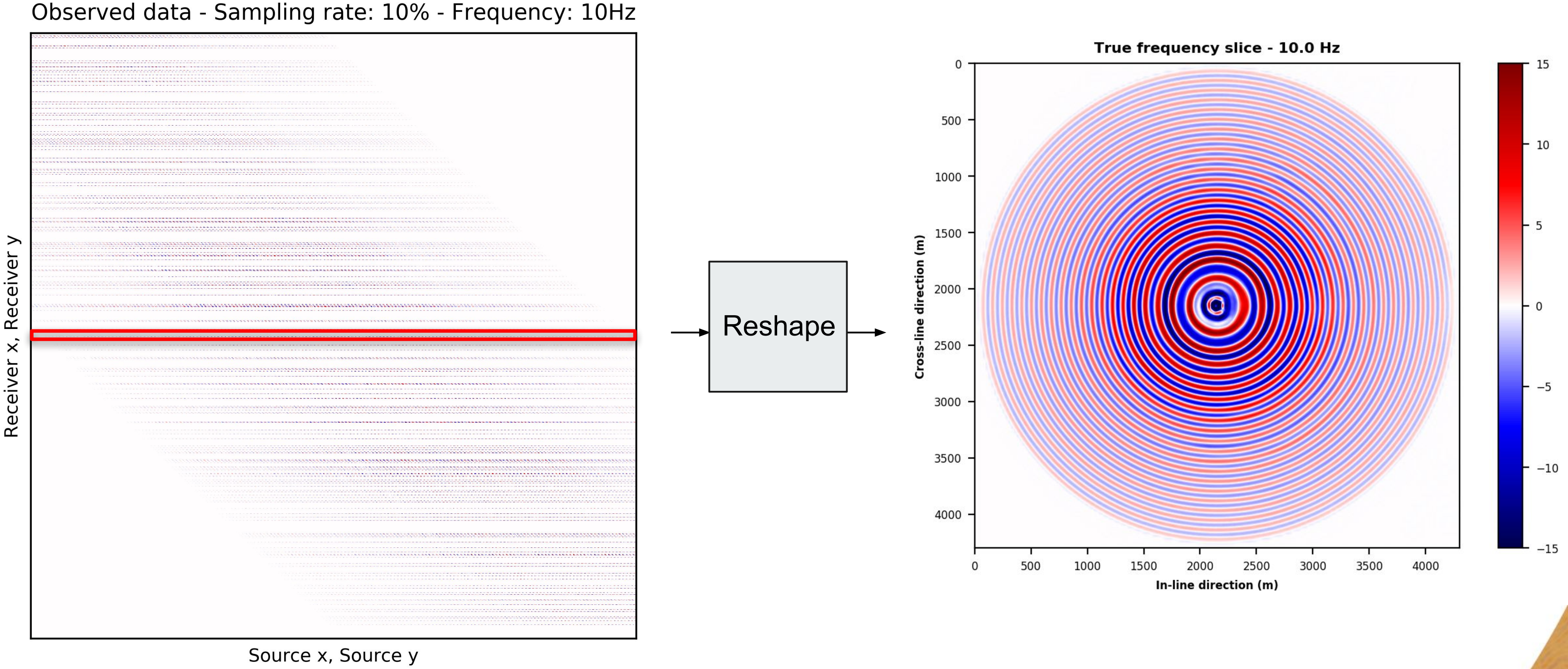
1. Extract and reshape single-receiver slices for existing receivers (fully sampled rows in $(\text{Rec } x,\ \text{Rec } y) \times (\text{Src } x,\ \text{Src } y)$ domain)
   - as many slices as recording receivers we have in the field
   - desired output of the network during training

2. Choose a training mask

3. Apply the training mask to artificially subsampled extracted single-receiver slices
   - input of the network during training

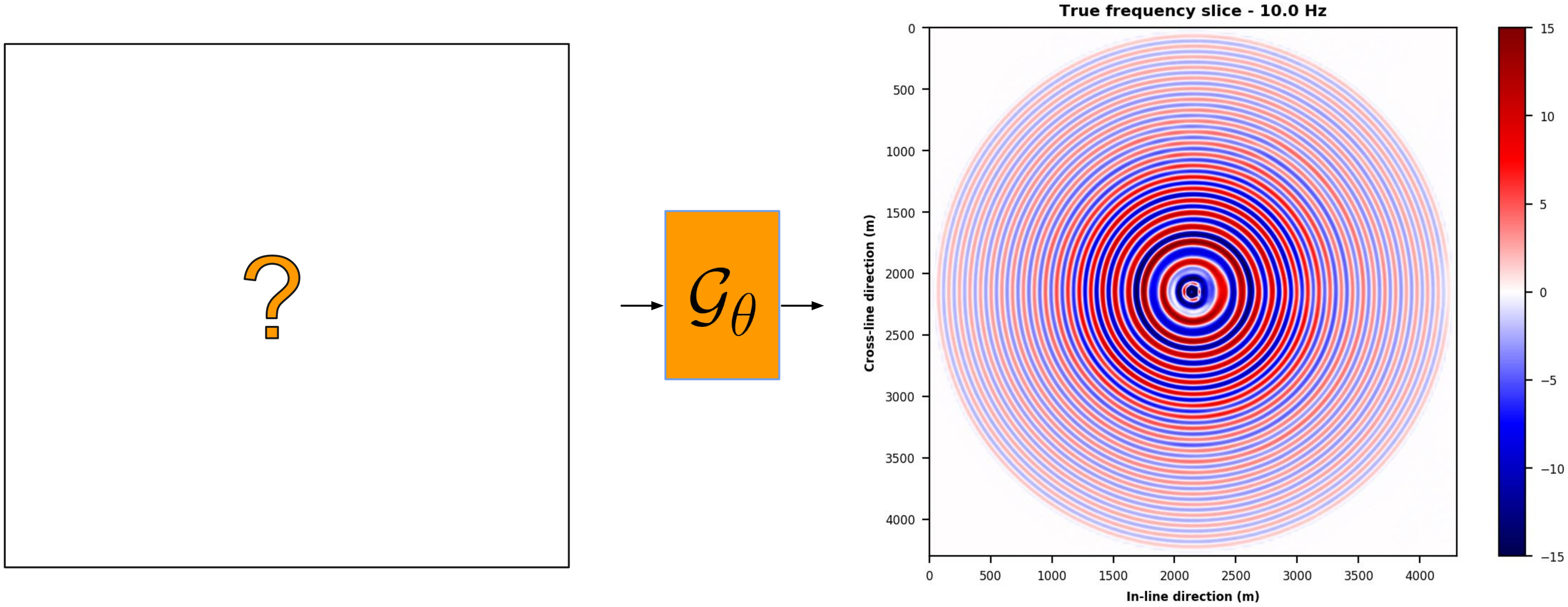
Training data: fully-sampled slices

Extract single-receiver slices for existing receivers, i.e., from acquired data



Training data: subsampled slices

What about the the input for supervised learning?



Steps to Extract training pairs

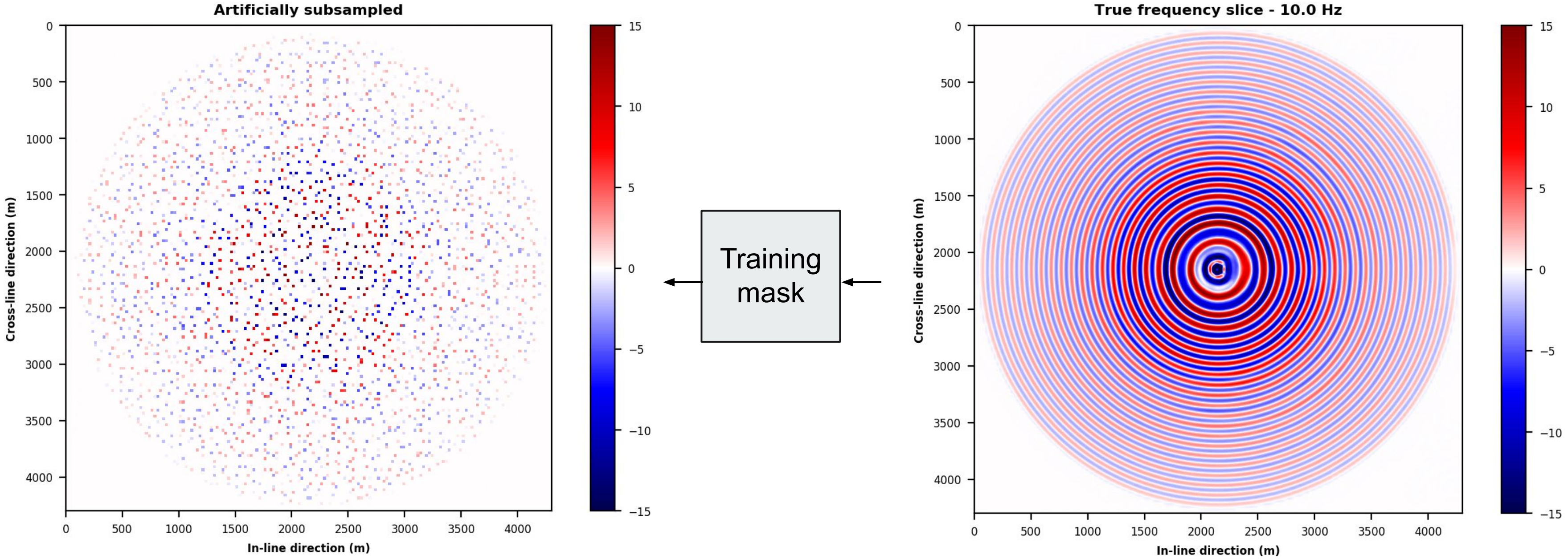
1. Extract and reshape single-receiver slices for existing receivers (fully sampled rows in $(\mathrm{Rec}\ x,\ \mathrm{Rec}\ y) \times (\mathrm{Src}\ x,\ \mathrm{Src}\ y)$ domain)
   - as many slices as recording receivers we have in the field
   - desired output of the network during training

2. Choose a training mask

3. Apply the training mask to artificially subsampled extracted single-receiver slices
   - input of the network during training

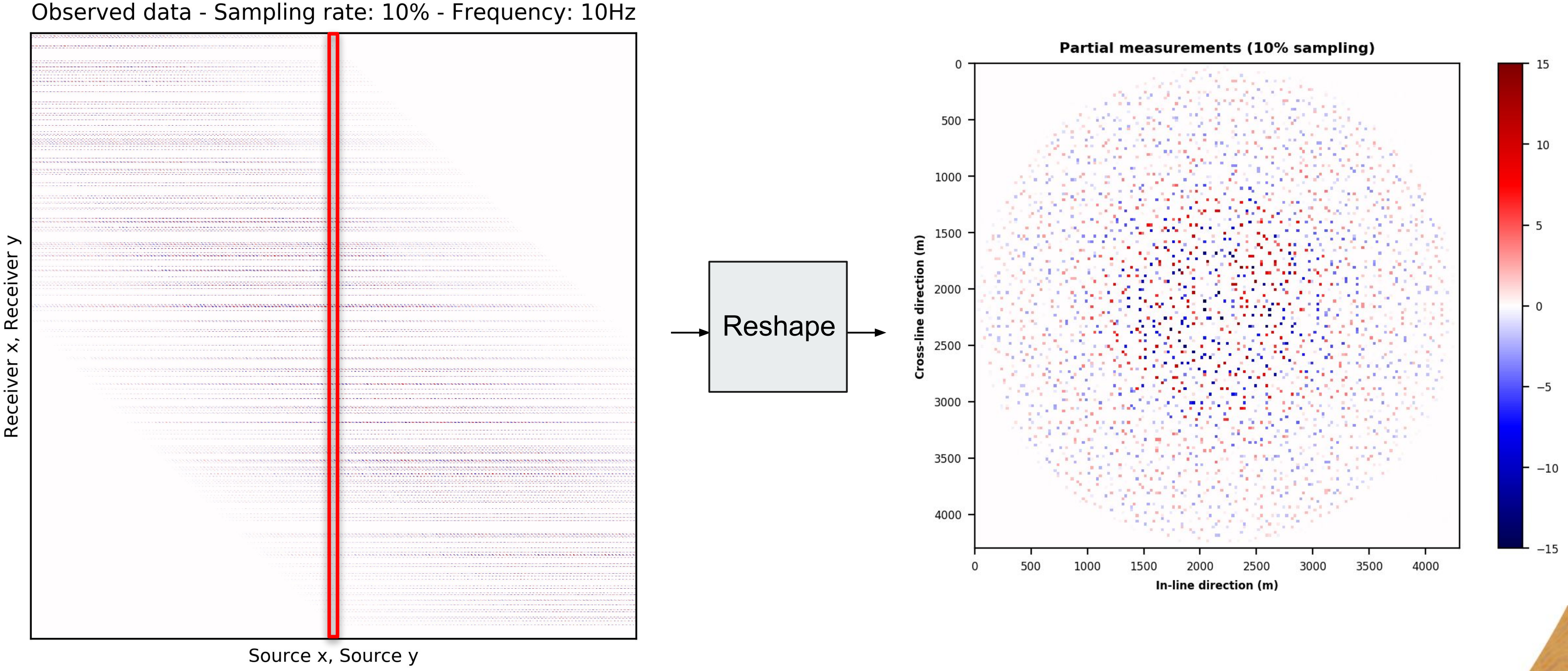
Training data: subsampled slices

Arbitrarily subsample the extracted fully-sampled slice with a training mask



Choosing training mask?

Reminder: The objective is to fill-in the columns



Training mask

We are free in choosing the training mask for artificial subsampling

We choose a random training mask equal to the randomly missing receiver sampling mask

- ► we know the missing pattern of receivers

Our experiments show that a random training mask is essential for successful wavefield recovery, even when receivers are on a periodic grid.

Training objective and CNN architecture

We use Generative Adversarial Networks (GANs) (Goodfellow et al., 2014)

GANs are based on an adversarial training procedure, i.e. involves two networks:

- Generator: is trained to reconstruct the artificially subsampled single-receiver slices

- Discriminator: is trained to distinguish between true single-receiver slices and reconstructed slices

We use a ResNet (He et al., 2016) based architecture for Generator and a fully-convolutional CNN with down-sampling for Discriminator.

Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S. Least squares generative adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision 2017, pages 2794-2802.
Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 5967–5976, July 2017.

Training framework: GANs

$$\min_{\theta} \mathop{\mathbb{E}}_{\mathbf{x}\sim p_X(\mathbf{x}),\, \mathbf{y}\sim p_Y(\mathbf{y})} \left[\left(1 - \mathcal{D}_\phi \left(\mathcal{G}_\theta(\mathbf{x})\right)\right)^2 + \lambda \left\| \mathcal{G}_\theta(\mathbf{x}) - \mathbf{y} \right\|_1 \right],$$

$$\min_{\phi} \mathop{\mathbb{E}}_{\mathbf{x}\sim p_X(\mathbf{x}),\, \mathbf{y}\sim p_Y(\mathbf{y})} \left[\left(\mathcal{D}_\phi \left(\mathcal{G}_\theta(\mathbf{x})\right)\right)^2 + \left(1 - \mathcal{D}_\phi(\mathbf{y})\right)^2 \right].$$

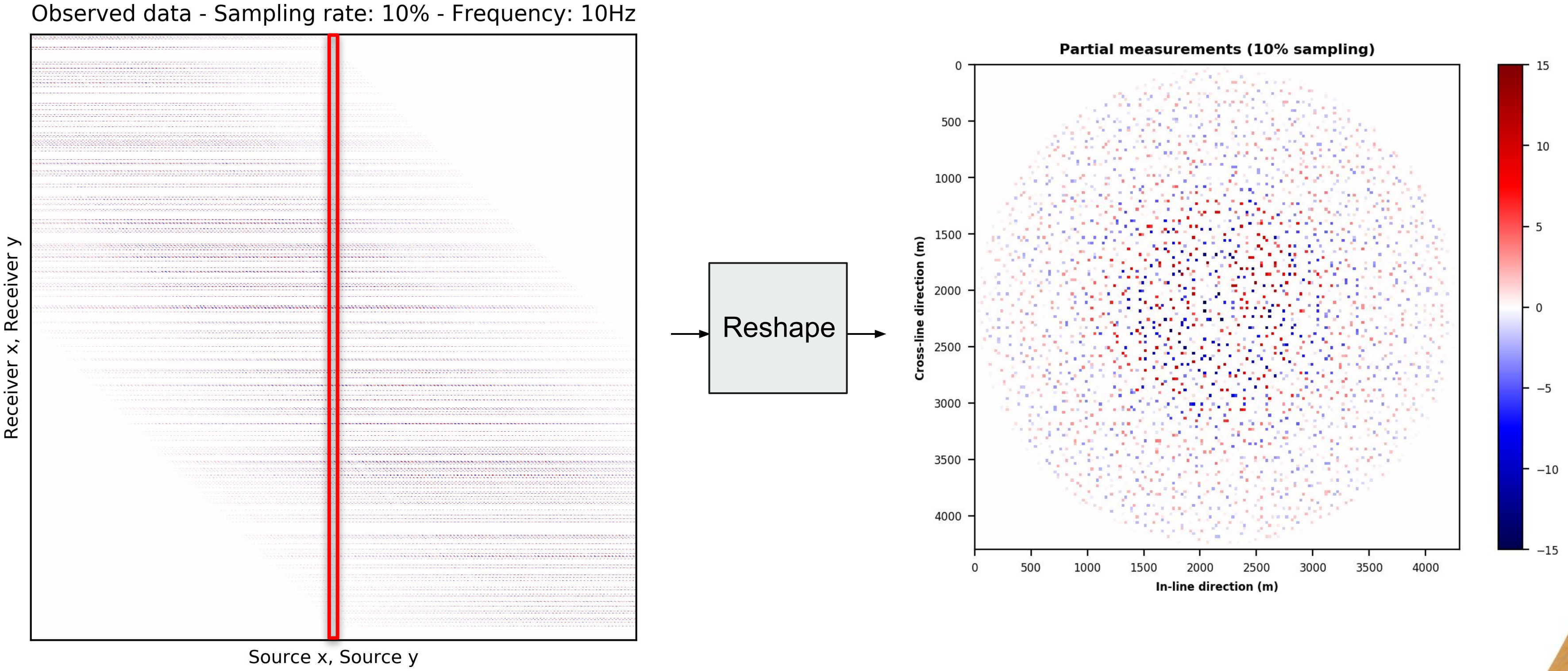
$\{\mathbf{x},\ \mathbf{y}\}$ Input/output pairs, drawn from the probability distributions $p_X(\mathbf{x})$ and $p_Y(\mathbf{y})$

$\mathcal{G}_\theta(\mathbf{x})$ Generator

$\mathcal{D}_\phi$ Discriminator

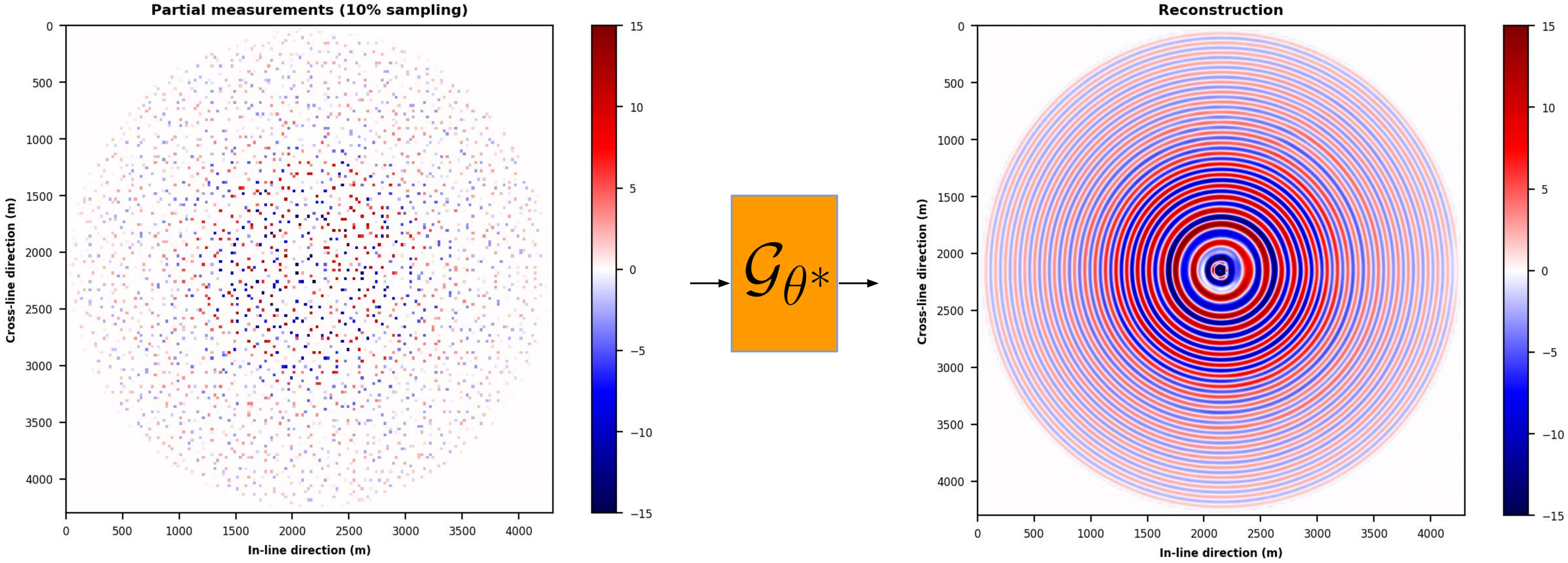
ℓ_1-norm misfit term weighted by λ ensures that each realization of $\mathcal{G}_\theta(\mathbf{x})$ maps to a particular $\mathbf{y}$, i.e., $\mathbf{x} \mapsto \mathbf{y}$ rather than solely fooling the discriminator.

Testing Stage: reconstruction
Extract all the single-source slices (columns)



Testing Stage: reconstruction

Apply the trained neural network to all columns



It works because....

Desirable source sampling, i.e., finely sampled sources

Source-receiver reciprocity holds under certain conditions

We hope Convolutional Neural Networks to perform well on testing data, i.e., reciprocal frequency slices

does not need any external training data

Dataset

Numerically simulated data on 3D BG Compass model

- 172×172 2D periodic grid of sources
- 172×172 2D periodic grid of receivers
- $25\ \mathrm{m}$ spatial sampling in both horizontal directions
- strong vertical and lateral variations

We processed the data for imaging by muting direct/turning waves

Numerical experiments

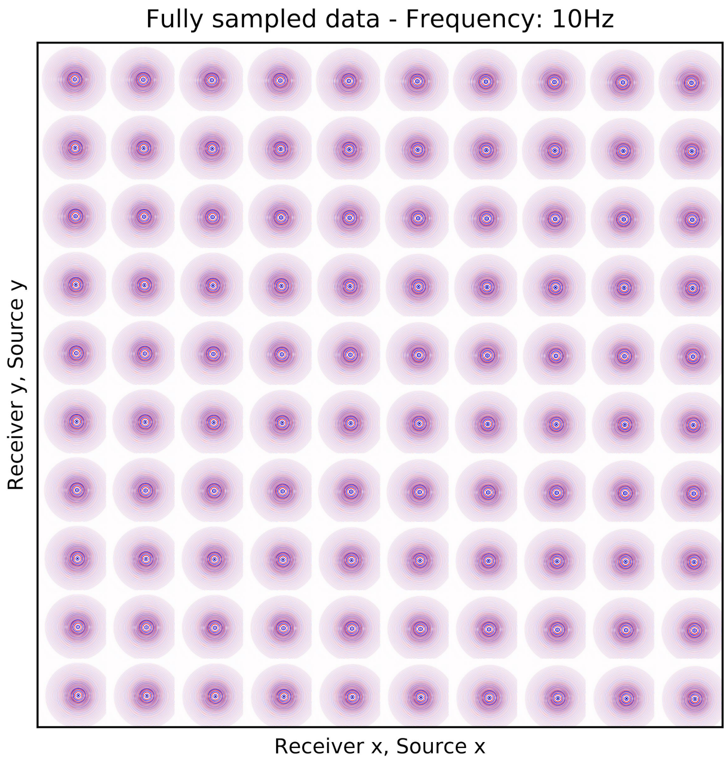
Applied to 3, 5, 10, and 15 Hz monochromatic data:

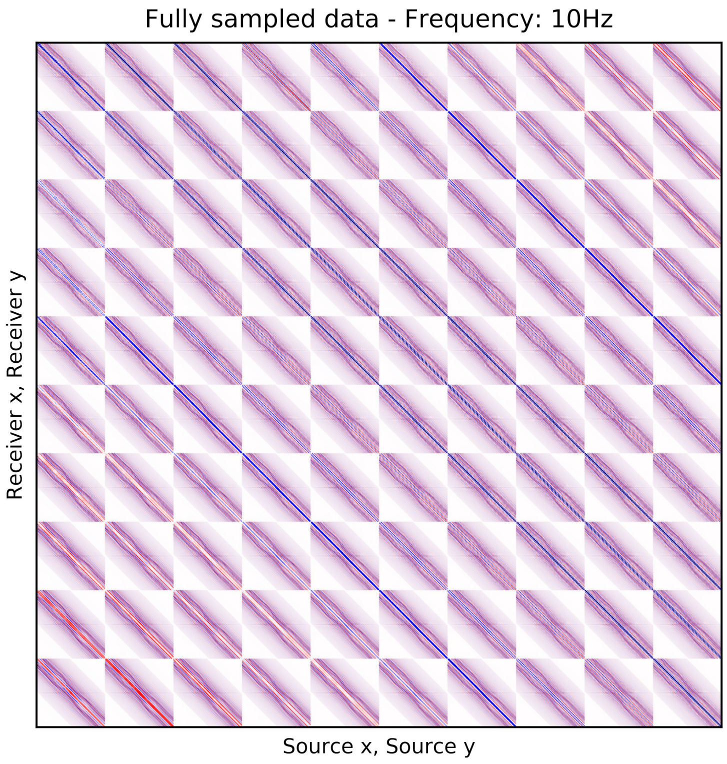
- ► missing 90% of receivers, randomly
- ► missing 90% of receivers, periodically

Training mask:

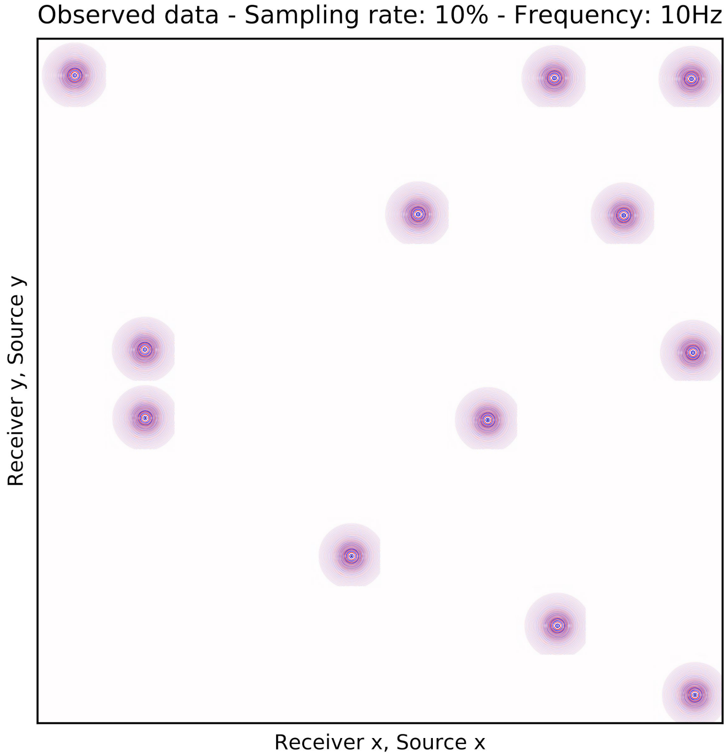
- ► experiments show that using a periodic training mask degrades the results
- ► for both cases (random and periodic), we train a single neural net using a random training mask

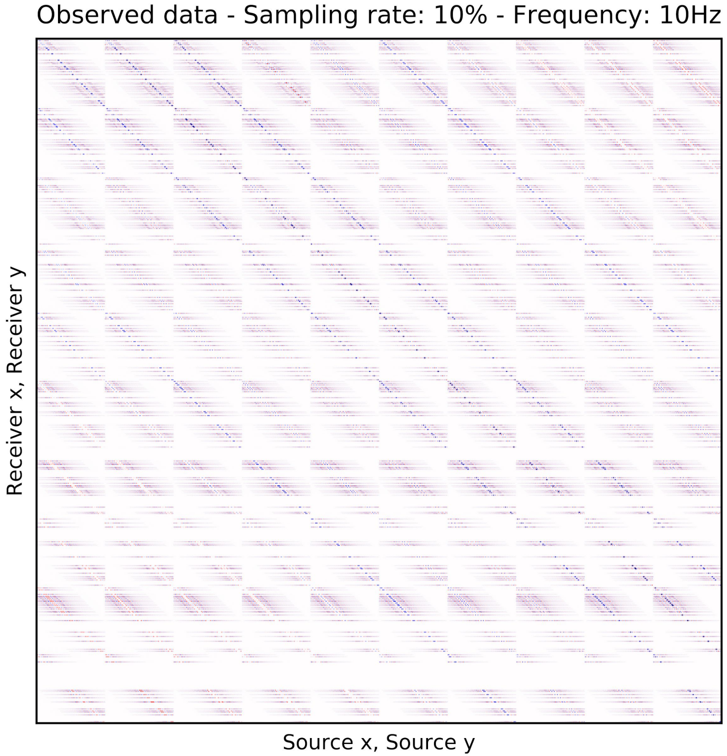
Fully-sampled data - 10 Hz



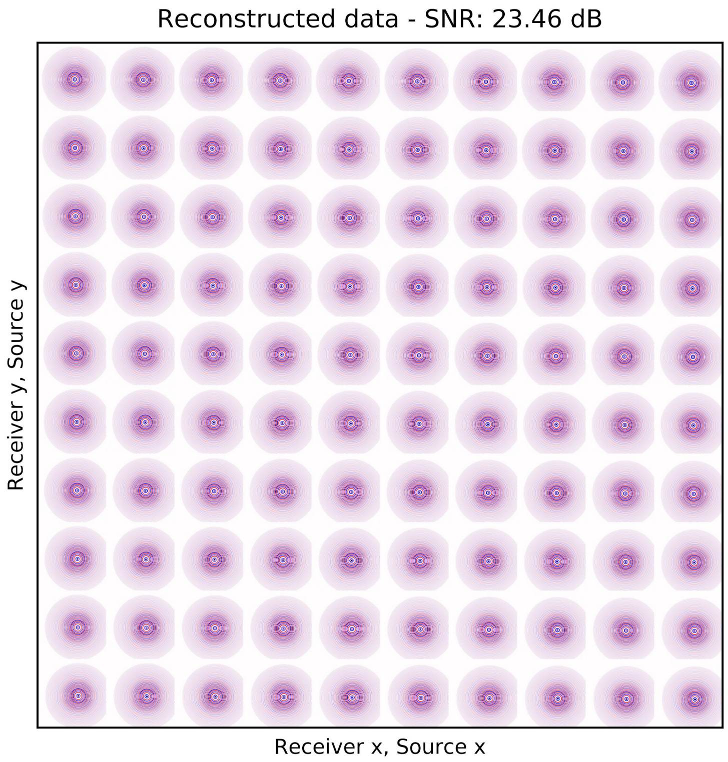


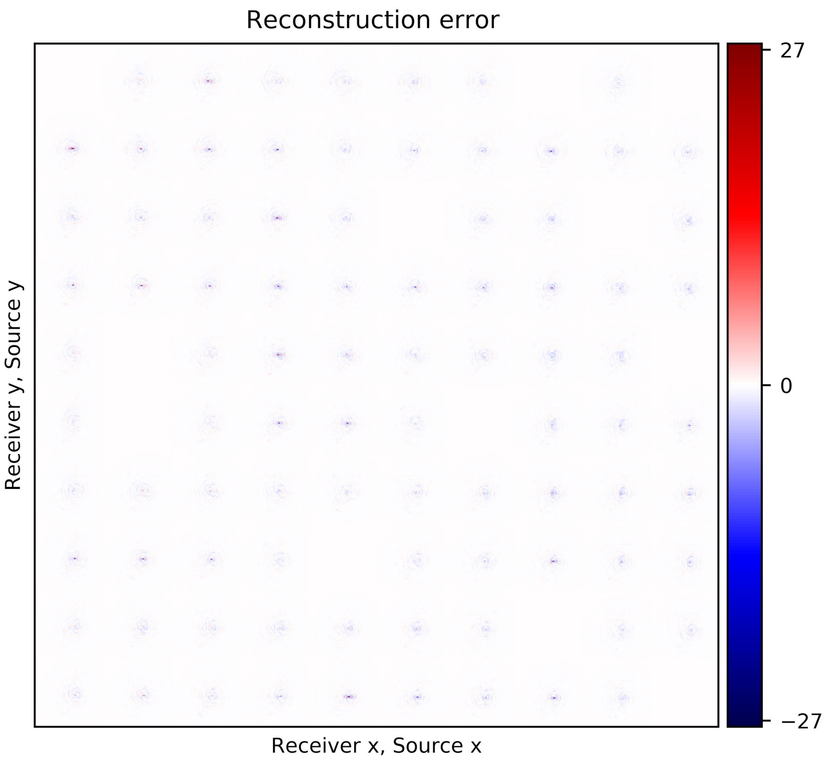
Observed data – Sampling rate 10%, randomly



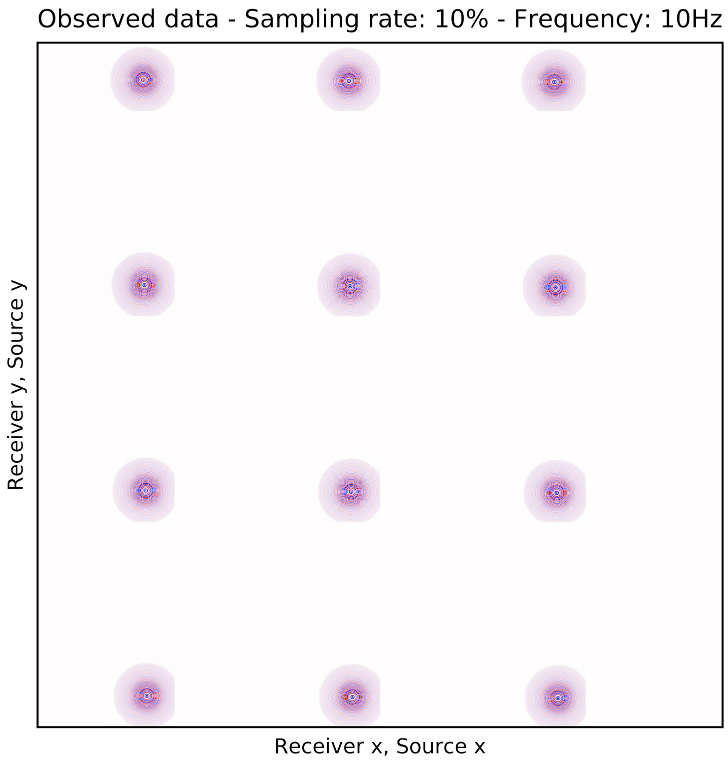


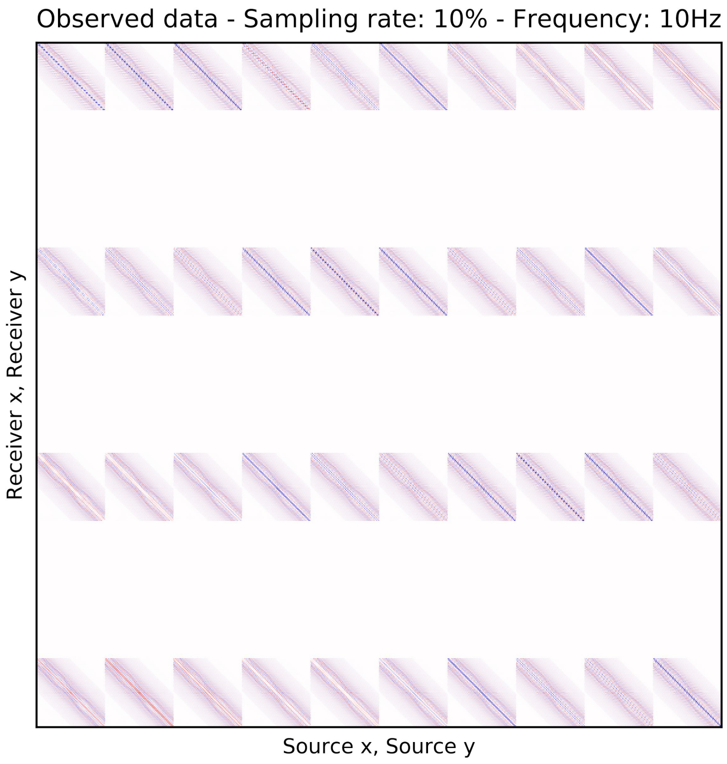
Recovered data - 10 Hz - random case
$(\text{Rec y, Src y}) \times (\text{Rec x, Src x})$ domain





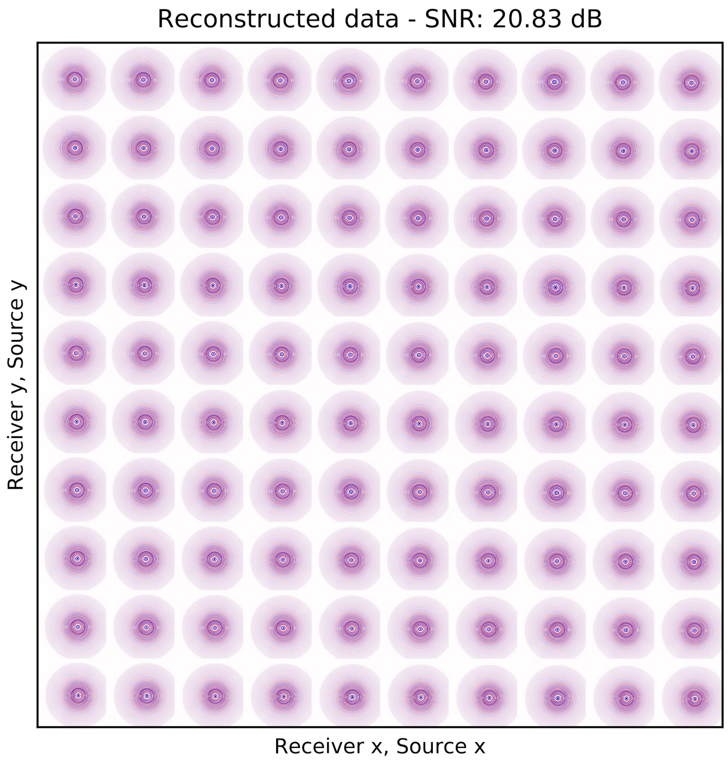
Observed data – Sampling rate 10%, periodically

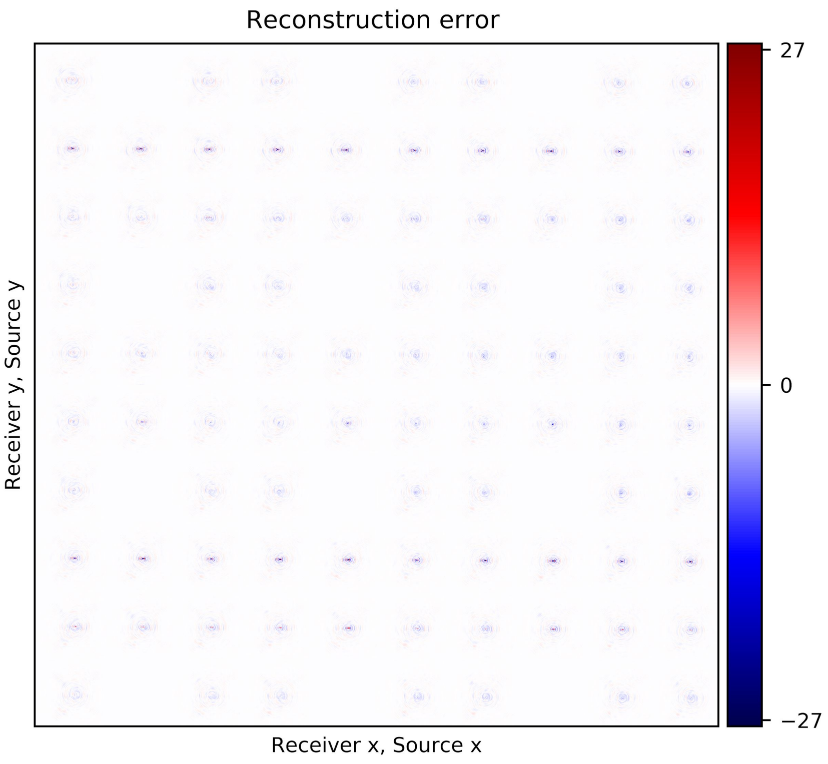




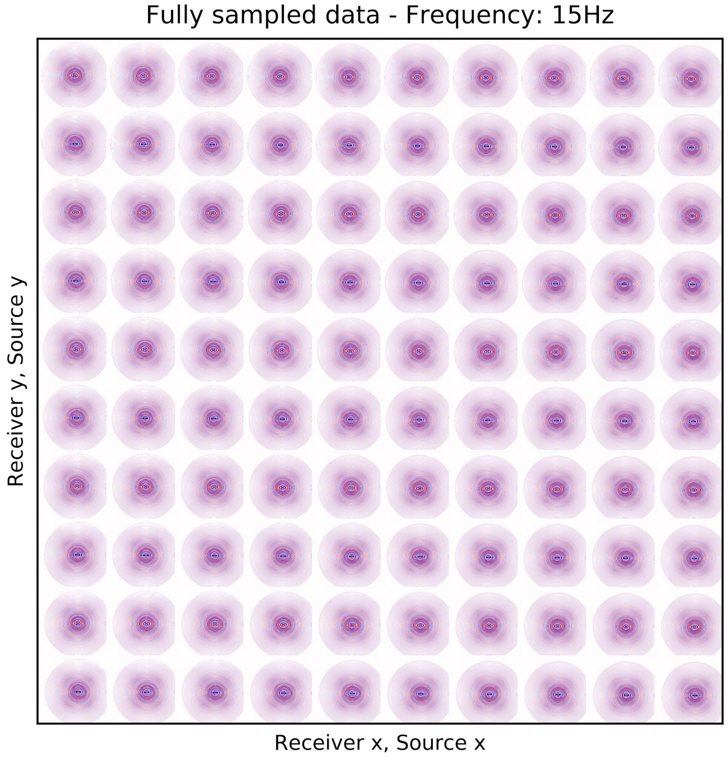
Recovered data - 10 Hz - periodic case

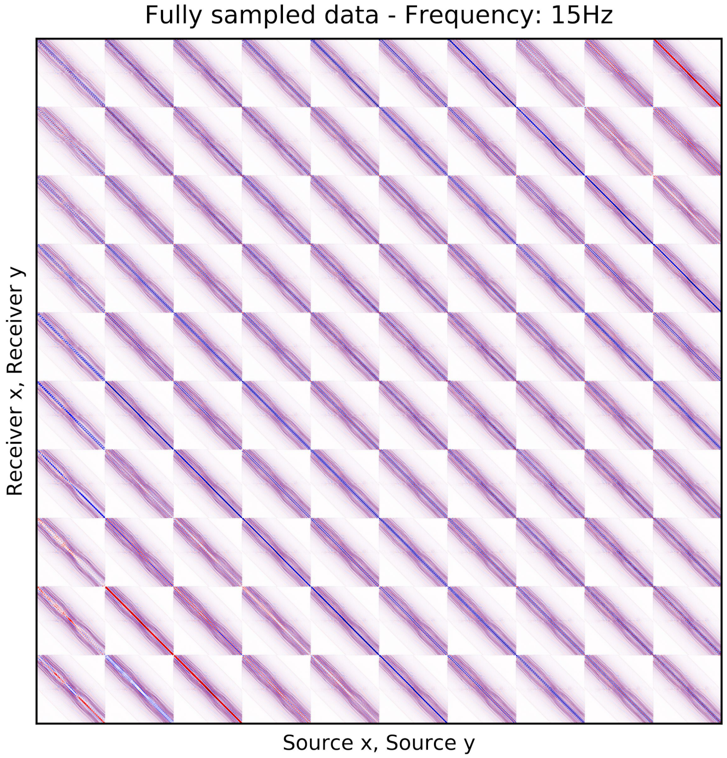
$(\mathbf{Rec\ y},\ \mathbf{Src\ y}) \times (\mathbf{Rec\ x},\ \mathbf{Src\ x})$ domain



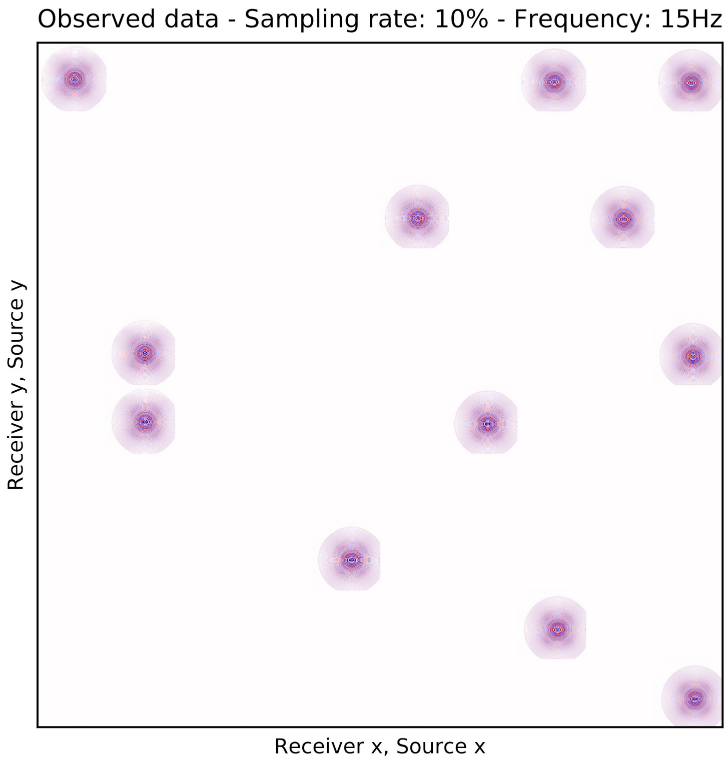


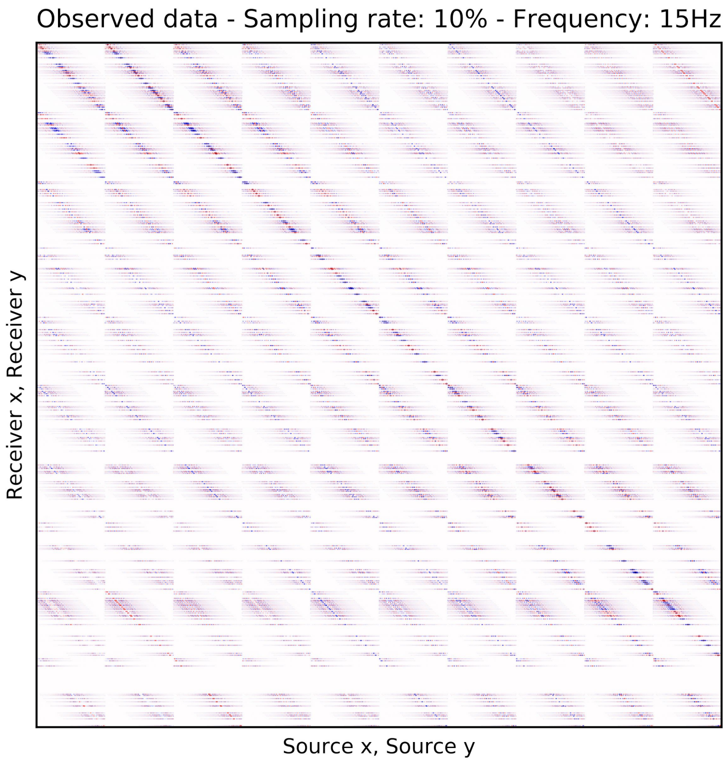
Fully-sampled data - 15 Hz





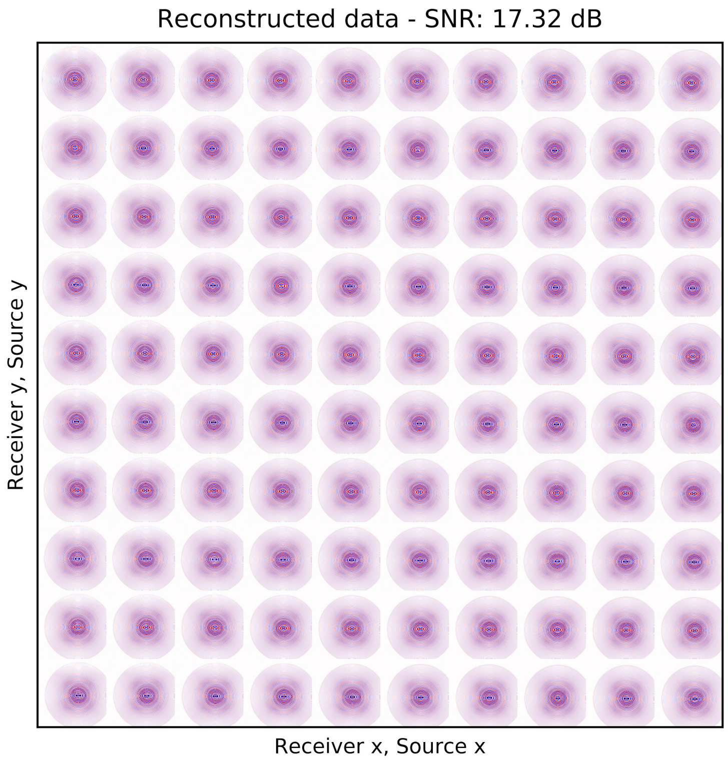
Observed data – Sampling rate 10%, randomly

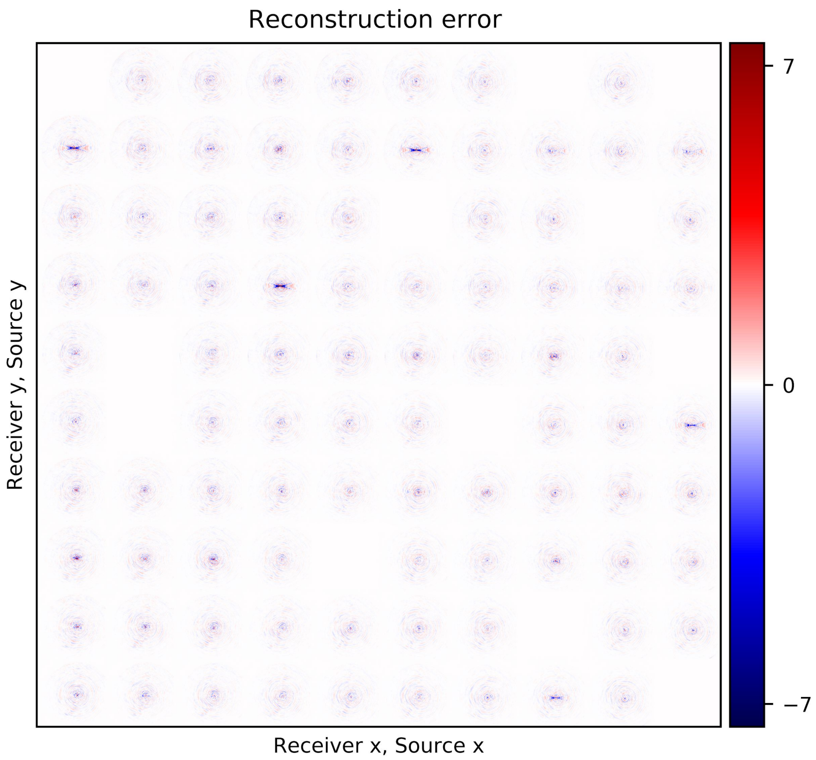




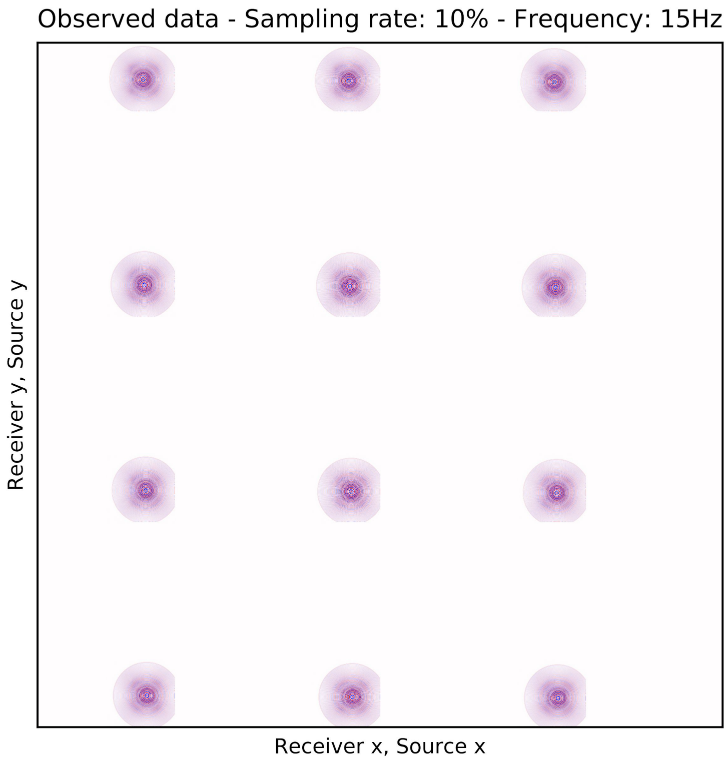
Recovered data - 15 Hz - random case

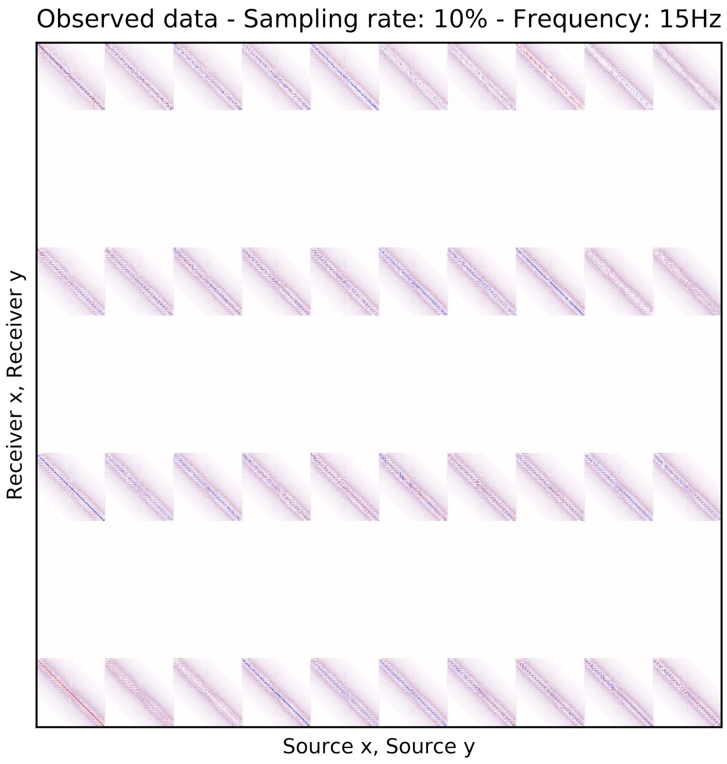
$(\mathbf{Rec\ y},\ \mathbf{Src\ y}) \times (\mathbf{Rec\ x},\ \mathbf{Src\ x})$ domain



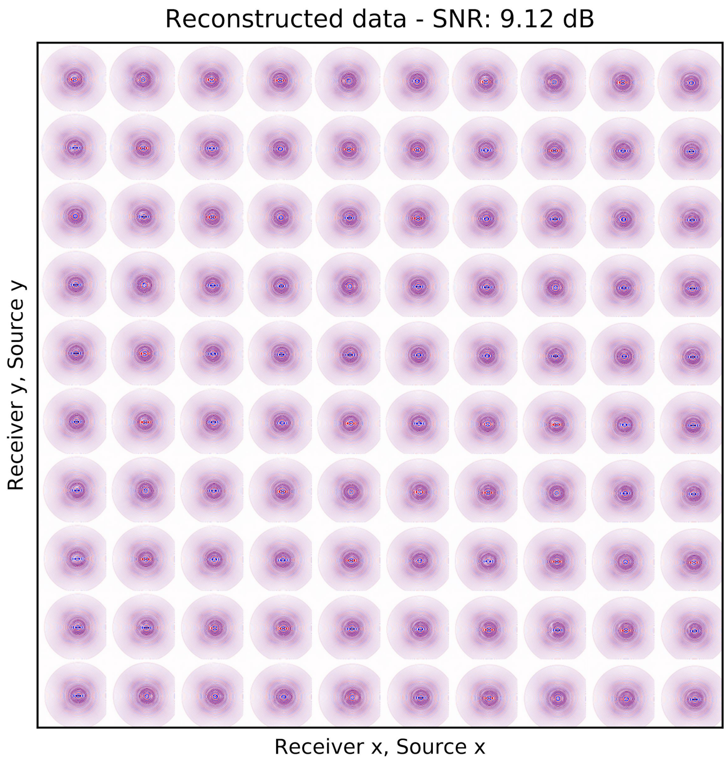


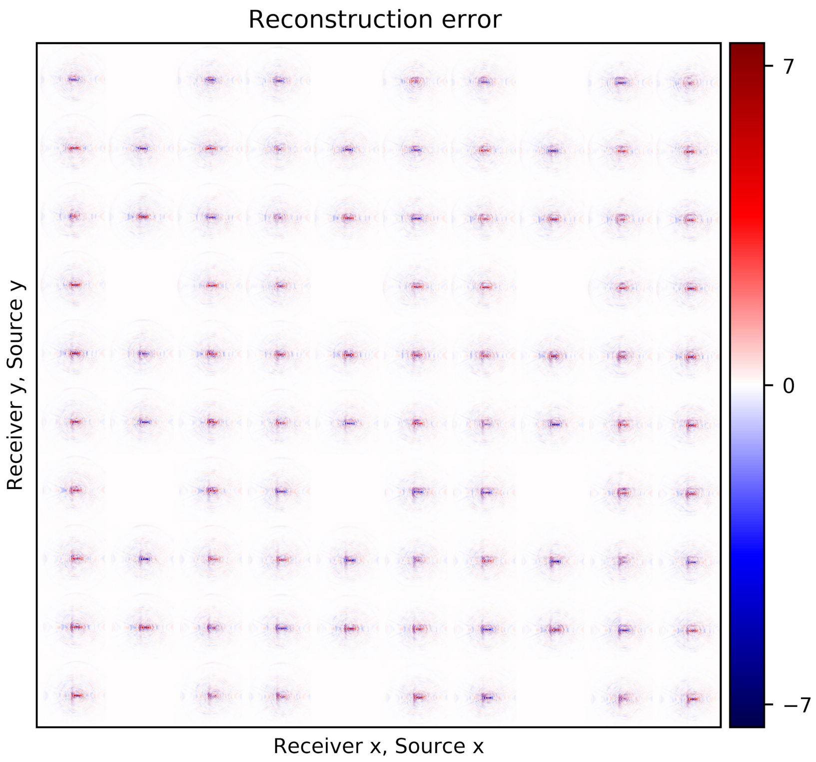
Observed data – Sampling rate 10%, periodically





Recovered data - 15 Hz - periodic case
$(\mathbf{Rec\ y,\ Src\ y}) \times (\mathbf{Rec\ x,\ Src\ x})$ domain





Reconstruction quality

Sampling mask	Frequency	Average recovery SNR
random	3 Hz	32.66 dB
random	5 Hz	29.07 dB
random	10 Hz	23.46 dB
random	15 Hz	17.31 dB
periodic	3 Hz	32.17 dB
periodic	5 Hz	28.32 dB
periodic	10 Hz	20.82 dB
periodic	15 Hz	9.12 dB

Table 1: Average reconstruction SNR for 90% random/periodic missing receivers.

Proposed method

vs

matrix completion method

Conclusions

The method does not need any external training data, assuming:

- source-receiver reciprocity
- desirable source sampling

Experiments show that random training mask is beneficial for recovery

- missing either randomly, or periodically

Future work: perform FWI with data obtained by reconstructing low-frequency spectrum of the observed data.