



**Utrecht University**

Master in Earth Structure and Dynamics

# **Compressive Acquisition and Wavefield Reconstruction for Ocean Turbulence Monitoring**

*Master Thesis*

Jingming Ruan

Supervisors:  
Dr. Ivan Vasconcelos

Final version, 30/07/2019

Utrecht, July 2019

# Abstract

The acquisition of highly sparse wavefield data, while maintaining information content, can yield significant gains in acquisition productivity, and enable cost-effective autonomous platforms. To achieve these goals, the use of wavefield data jointly with its first- and second-order spatial derivatives allows for relaxation of Nyquist sampling criteria up to a factor of three under regular sampling. This sampling rate can be potentially further decimated by means of compressive sensing and preconditioning approaches. We introduce a multicomponent reconstruction scheme which uses sparse, irregularly-sampled data, containing pressure fields and two orders of its spatial derivatives, to retrieve the desired regularly-sampled, dense data. Here, we employ a jittered sampling scheme, together with a consistent preconditioning for the solver, that aims for higher or lower sampling according to the expected data complexity. The method is tested on a seismic shot-gather decimated in space domain with different sampling patterns. The results verify that by regular sampling with one-third of the original Nyquist rate, the resulting decimated wavefield from the seismic data set is successfully reconstructed by the multicomponent reconstruction. When using irregular sampling, our approach provides observably better reconstructions for the cases that are over-sampled, with sampling rates lower than one-third of the original Nyquist rate, and considerably lower for the lesser complex portions of the data. The multicomponent reconstruction is adapted for 3D wavefield generated from a 2D ocean turbulence model with layer wavespeed configuration for imposing additional derivatives in shot direction and cross-derivatives, along with physics-based preconditioning. The reconstruction has major seismic events well-reconstructed from the over-sampled wavefield.

# Contents

<b>Contents</b>	<b>iii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Gradient-based Reconstruction</b>	<b>4</b>
<b>3 Sampling Pattern</b>	<b>7</b>
3.1 Regular sampling . . . . .	7
3.2 Irregular sampling . . . . .	7
3.3 Data-driven irregular sampling . . . . .	9
<b>4 Preconditioning</b>	<b>11</b>
<b>5 2D Wavefield Reconstruction</b>	<b>13</b>
5.1 Regular sampling . . . . .	13
5.2 Irregular sampling . . . . .	13
<b>6 Turbulence Wavefield Inversion</b>	<b>16</b>
6.1 2D synthetic ocean turbulence wavefield reconstruction . . . . .	18
6.2 3D synthetic ocean turbulence wavefield reconstruction . . . . .	20
<b>7 Conclusions</b>	<b>29</b>
<b>Bibliography</b>	<b>30</b>

# Chapter 1

## Introduction

Since the first publication[10] about seismic observation of thermohaline fine structure in an oceanographic front, various oceanographic structures(e.g. Ocean internal waves, water mass boundaries, etc.) have been imaged from field and synthetic data with applications of seismic oceanography[10, 9, 11, 8, 5, 2, 25, 19, 12, 16]. Seismic oceanography offers a new solution for oceanography studies, meanwhile controllable seismic sources and fast responses of seismic events make seismic oceanography extremely suitable for oceanography monitoring, if provided with stable and continuous wavefield data.

Most seismic oceanography studies, however, are implemented on large-scale seismic acquisition (e.g. towed-marine seismic acquisition), which are extremely expensive to ever scale to on-demand oceanographic studies. Low-running-cost autonomous platforms, with recent examples of such are the SpiceRack[21], or Wave Gliders[14], are designed for long-term oceanographic data acquisition, with endurance up to 12 months, making them suitable devices for monitoring purposes. These platforms allow for swarm surveys [15]: low-footprint, high-productivity surveys of which one notable issue is the high cost-per-unit of the platforms themselves. Thus, the number of such platforms is limited, and thus, they tend to be used to sparsely cover the research areas. Here we choose to design several different sampling schemes based on the well-sampled case, including regular sampling, irregular sampling, and multi-source sampling. Figure 1.2 is a demonstration on the sampling schemes, detailed explanation will be discussed on the following chapters.

Although the topic of sparse sampling and reconstruction have been widely explored in the seismic exploration community over the past two decades, here we revisit the topic with the particular goal of enabling highly-sparse acquisition by autonomous platforms, in order to maximize the wavefield information while minimizing the number of autonomous vehicles in the field.

Large intervals between sparsely distributed streamers put a high demand on wavefield reconstruction in order to mitigate the effects of spatial aliasing, which is caused by a lower-than-Nyquist sampling rate from the distribution of the platforms. Similar problems arise in ocean-bottom or land acquisition geometries. With one alternative to handling spatial gaps, Robertsson, et al.(2008)[20] extends the multichannel sampling theorem from Linden(1959)[13] with the use of both the first- and second order derivatives of the pressure data, successfully lowering the Nyquist sampling rate by means of gradient-based reconstruction. Building on this approach, gradient-based wavefield reconstruction methods relying on basis-pursuit and the use of priors have been proposed [24, 17], relying mostly on first-order gradients.

Figure 1.1 is an example of Multichannel Interpolation by Matching Pursuit (MIMAP)[24], which uses multicomponent seismic measurements for crossline wavefield reconstruction and doesn't require assumptions such as linearity of seismic events to interpolate severely aliased data. Figure 1.1 shows different results from a simple synthetic data, in which both sinc interpolator and multichannel sinc interpolator fail to address the spatial aliasing, while MIMAP has successful reconstruction in t-x domain and f-k domains.

Based on the multichannel sampling theorem, here we introduce a new wavefield reconstruction scheme, called multicomponent reconstruction, which combines multichannel sampling theorem



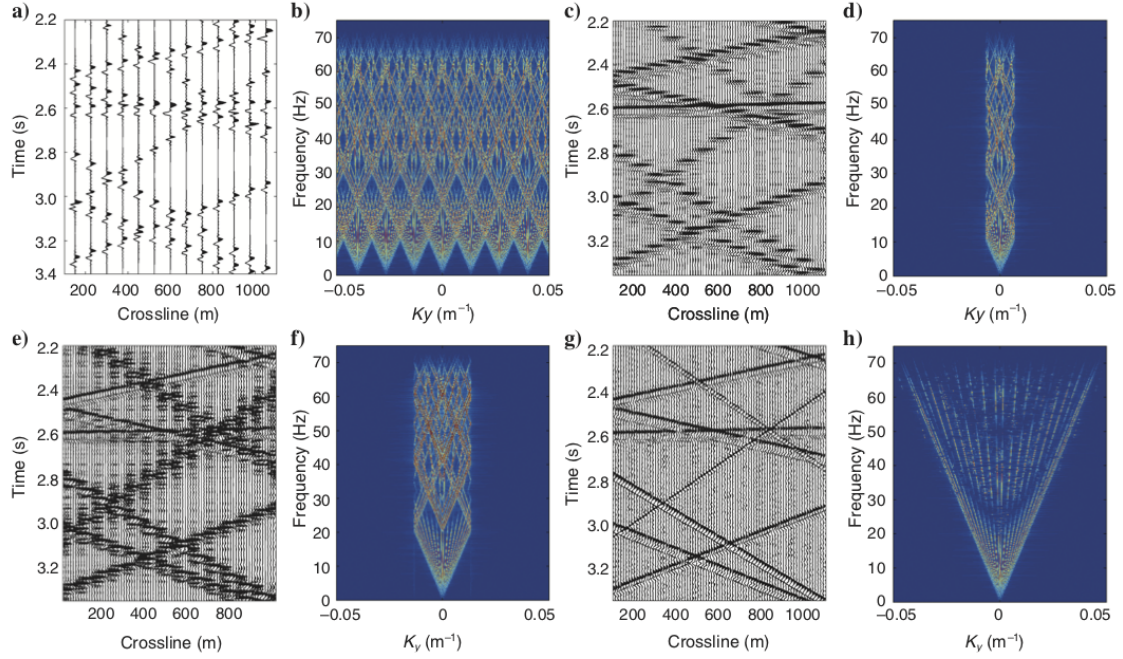


Figure 1.1: Example with simple synthetics: close-up of a region of the t-y domain and f-k transforms of the whole dataset. (a, b) Input pressure, sampled at 75 m; (c,d) pressure reconstructed by using a sinc interpolator; (e,f) pressure reconstructed by using a multichannel sinc interpolator, also having as input the crossline gradients at the samples positions; (g,h) pressure reconstructed with MIMAP, also having as input the crossline gradients at the samples positions[24].

with compressed sensing, allowing for relaxation of Nyquist sampling criteria up to a factor of three under regular sampling. This sampling rate can be potentially further decimated by means of compressive sensing and preconditioning approaches. The retrieval of the wavefield data is achieved by creating a linear sampling operator which the desired model to the observed data and its derivatives, followed by solving the linear-equation with specific types of solvers based on preconditioning and priors.

Starting from testing the multicomponent reconstruction on a benchmark single seismic shot-gather data set, we explore the effects of multichannel sampling theorem, together with sampling patterns and preconditioning. Then a single synthetic seismic shot-gather data set generated from a ocean turbulence model (figure 6.1) is tested with the addition of the preconditioning based on its physics-based priors. In the end of this research, 3D ocean turbulence seismic wavefield reconstruction with the multicomponent reconstruction is achieved.

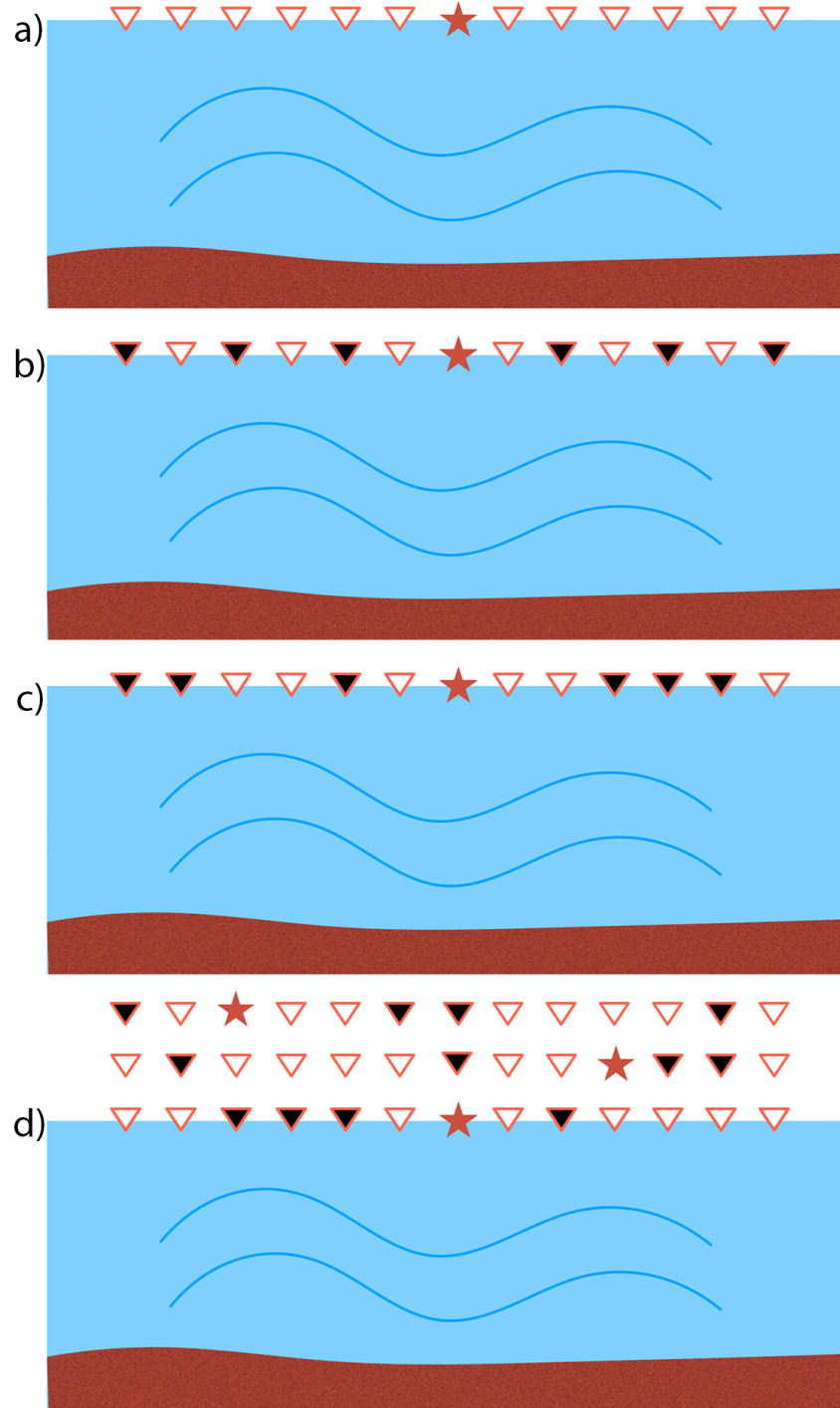


Figure 1.2: Sampling scheme: a) an example of well-sampled case; b) regular sampling(subsampling); c) irregular sampling(subsampling); d) multi-source irregular sampling(subsampling). Stars and triangle represent sources and receivers, respectively.

## Chapter 2

# Gradient-based Reconstruction

The multichannel sampling theorem by Linden (1959)[13] extends the Nyquist-Shannon sampling theorem with wavefield data and its first derivative, and is able to lower the minimum sampling rate, relative to the conventional Nyquist rate, by a factor of two (eq. 2.1). Bringing this concept to exploration geophysics, Robertsson et al.(2008)[20] show that the Nyquist wavenumber can be further increased up to a factor of three when both first and second-order derivatives are available.

$$\begin{aligned} p(y_0) &= \sum_{m=-\infty}^{\infty} p\left(\frac{m}{\sigma}\right) \text{sinc}(\sigma y_0 - m) \\ p(y_0) &= \sum_{m=-\infty}^{\infty} \left\{ p\left(\frac{2m}{\sigma}\right) + \left(y_0 - \frac{2m}{\sigma}\right) \times \partial_y p\left(\frac{2m}{\sigma}\right) \right\} \text{sinc}^2\left(\sigma \frac{y_0}{2} - m\right) \end{aligned} \quad (2.1)$$

For seismic wavefield acquisition, each receiver (e.g. wave glider) will be equipped with multicomponent sensors, which are able to record scalar data, its first-order spatial derivatives (from particle velocity measurements) in desired directions. Second-order derivatives could, in principle, be obtained from closely-spaced arrays of multicomponent sensors, as shown in figure 2.1. While on the source side, closely-spaced arrays of phase-controlled sources can generate data as if excited dipole sources, in order to measure derivatives in the source direction by the 3D-sensor array of receivers. Here, we assume the traces sampled at the source location can still have high signal-to-noise ratio which contain turbulence reflection or diffraction signals, rather than dominated by direct arrivals.

Based on the multichannel sampling theorem, to build a multicomponent reconstruction method for wavefield reconstruction, the problem is simply represented as:

$$\mathbf{d}_{obs} = \mathbf{S} \mathbf{d}_{mod} \quad (2.2)$$

where  $\mathbf{d}_{obs}$  is a vector of the observed scalar (pressure) data and its first- and second-order derivatives in the sub-sampled direction(s);  $\mathbf{d}_{mod}$  is the desired wavefield model. In this research, we assume that data are well sampled in time domain but sub-sampled in space domain, having Nyquist wavenumber(s) lower than the highest wavenumber component of the signal. The sampling operator  $\mathbf{S}$  is composed of a restriction operator  $\mathbf{R}_r$ , keeping only the traces that can be recorded by the receivers, together with operators for first- and second-order derivatives in the sub-sampled direction  $\mathbf{D}_1$  and  $\mathbf{D}_2$ . For reconstruction using pressure data  $\mathbf{d}_t$  and its first spatial derivative  $\mathbf{d}_1$  recorded by the receivers after decimation:

$$\mathbf{d}_{obs} = \begin{pmatrix} \mathbf{d}_t \\ \mathbf{d}_1 \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} \mathbf{R}_r \\ \mathbf{D}_1 \end{pmatrix} \quad (2.3)$$

For reconstruction with pressure data and its first- and second-order spatial derivatives  $\mathbf{d}_1$  &  $\mathbf{d}_2$  recorded by the receivers after decimation:

$$\mathbf{d}_{obs} = \begin{pmatrix} \mathbf{d}_t \\ \mathbf{d}_1 \\ \mathbf{d}_2 \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} \mathbf{R}_r \\ \mathbf{D}_1 \\ \mathbf{D}_2 \end{pmatrix} \quad (2.4)$$

Here, we use the Fourier-based models for spatial derivatives:

$$\begin{aligned} \mathbf{D}_1 &= \mathbf{R}_r * \mathbf{F}^{-1} * \mathbf{diag}(ik_n) * \mathbf{F} \\ \mathbf{D}_2 &= \mathbf{R}_r * \mathbf{F}^{-1} * \mathbf{diag}[(ik_n)^2] * \mathbf{F} \end{aligned} \quad (2.5)$$

Where  $\mathbf{F}$  is the Fourier transform and  $\mathbf{F}^{-1}$  is the inverse Fourier transform;  $\mathbf{diag}(ik_n)$  and  $\mathbf{diag}[(ik_n)^2]$  are the diagonal matrices with wavenumber. Figure 2.2 is the derivatives data calculated from a 2D wavefield data (figure 3.1). Note that Restriction operator  $\mathbf{R}_r$  is also applied in each derivative operator, for gradients data can only be attained through receivers.

In this research, we numerically obtain the gradients and decimated wavefield by applying the forward operators to the original relatively well-sampled wavefield scalar data, for it is a convenient method to generate derivative data without noise and error. While this approach may need revisiting when processing the field scalar data and field gradient data, which is beyond the scope of this research.

The reconstruction of wavefield is carried by inversion of the linear system. The choice of solvers depends on data structure, thus it can vary according to the implemented priori information. LSQR[18] is the mostly used solver for inversion in this study, which minimizes the l2-misfit between observed and predicted data (eq.2.6). LSQR is believed to be good at solving over-determined problems, however, as will be shown in our results, if given appropriate preconditioning, LSQR can still give satisfying inversion from under-determined problems, for example in our case, the wavefield reconstruction from over-sampled data with sampling rate lower than threefold subsampling.

$$LSQR: \quad \text{minimize} \quad \|\mathbf{d}_{obs} - \mathbf{S}\mathbf{d}_{mod}\|_2 \quad (2.6)$$

The a priori information we rely on in this research refer to data structure in different basis and they are implemented by data-driven sampling and preconditioning, which will be explained in the next chapters.

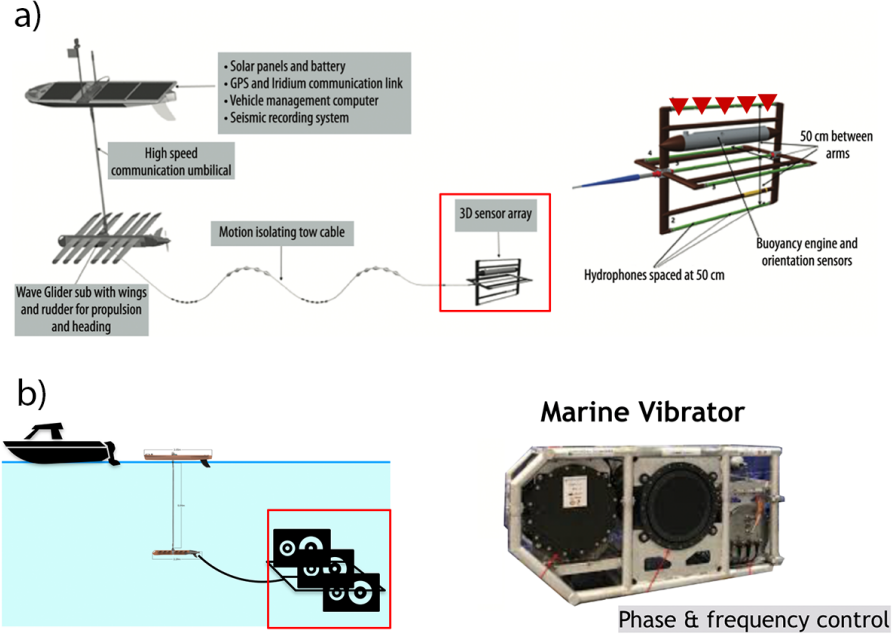


Figure 2.1: Autonomous platform: a) Wave glider[14] and the 3D sensor array used for multichannel sampling, capable of gradients measurement. b) Source composition, including an array of marine vibrators, guarantee gradients in source direction.

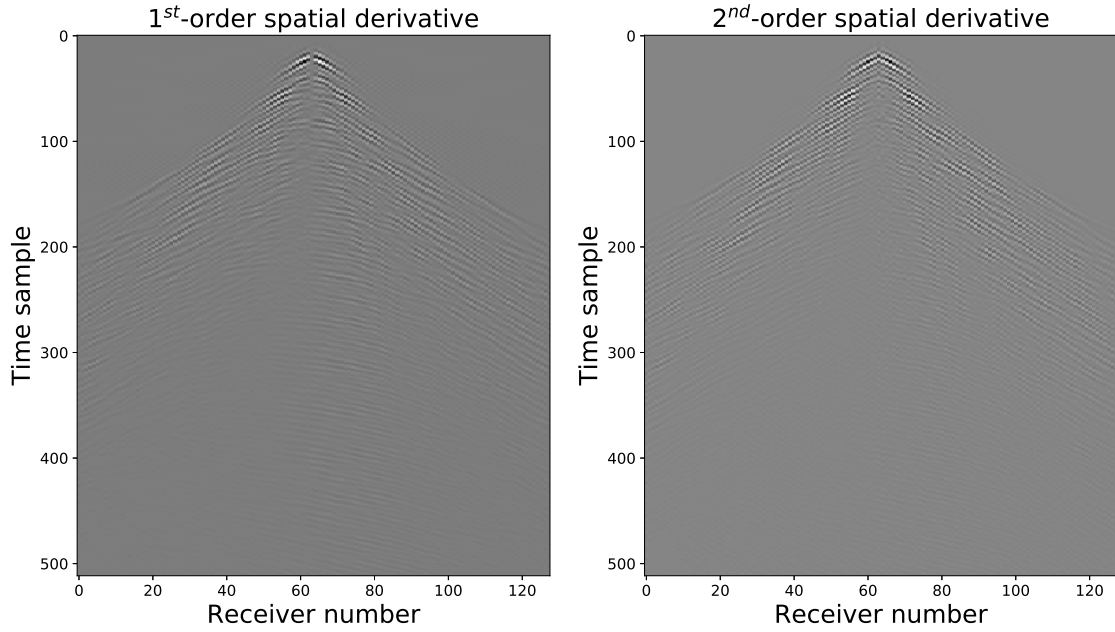


Figure 2.2: First- and second-order spatial derivatives calculated from the 2D benchmark shot-gather using the Fourier representation.

## Chapter 3

# Sampling Pattern

### 3.1 Regular sampling

Different sampling schemes are discussed in this research, including regular sampling, irregular sampling and data-driven irregular sampling. Regular sampling is used to verify the prediction by the multichannel sampling theorem. Decimation up to a factor of three is made to a benchmark single seismic shot-gather from van den Berg and Friedlander (2013)[23], with the source location in the middle of the 128 evenly distributed receivers, as is shown in figure 3.1 along with corresponding frequency-wavenumber spectrum. Seismic acquisitions normally are known for their relatively large sampling intervals in at least one spatial direction (e.g. crossline), and for being well-sampled in the time domain. Thus, only situations with spatial subsampling are covered in this project. Compared with the original f-k domain, the spatial aliasing is pronounced after decimation, and the order of alias increases with larger sampling interval.

### 3.2 Irregular sampling

Inspired by compressed sensing, for over-sampled cases with sampling rate lower than one-third of the original Nyquist rate, irregular sampling is applied to maximize the wavefield information while minimizing the number of samples in the field. Irregular sampling is implemented with jittered sub-sampling[6], as shown in figure 3.2, by first regularly choose nodes with specific interval according to the decimation rate, with actual sampling locations jittering between the nodes and the last trace before next node, thus every trace is equally likely to be sampled.

Herrmann (2010)[7] gives an example about the different sampling schemes and their imprint in the Fourier domain, in figure 3.3, showing that given same amount of samples, regular sampling will create sparse and strong alias in Fourier domain, while irregular sampling tends to generate weak sub-sampling artifacts, compared with original wavenumber components. The artifacts created by the regular sampling can dominate the  $l_2$ -misfit of the reconstruction, hindering the retrieval of the original Fourier spectrum. Therefore, irregular sampling can better minimize the number of samples, while maintaining the major signal components of the original wavefield.

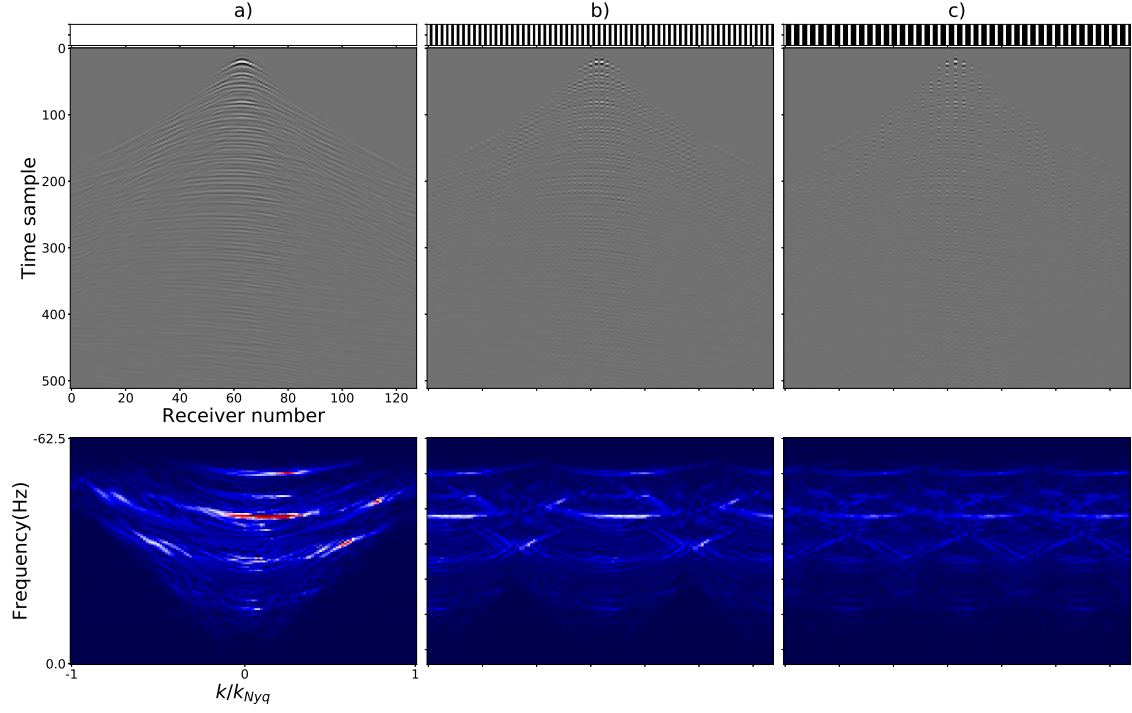


Figure 3.1: 2D benchmark shot-gather dataset[23] with regular sampling and corresponding Fourier spectrum: a) Original wavefield; b) regular sampling scheme with half of the original Nyquist sampling rate; c) regular sampling scheme with one-third of the original Nyquist sampling rate.

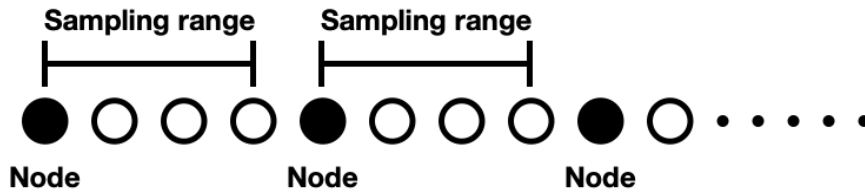


Figure 3.2: Jittered subsampling: nodes are regularly distributed based on the number of samples, the actual sampling location will vary between the nodes and the last trace before next node.

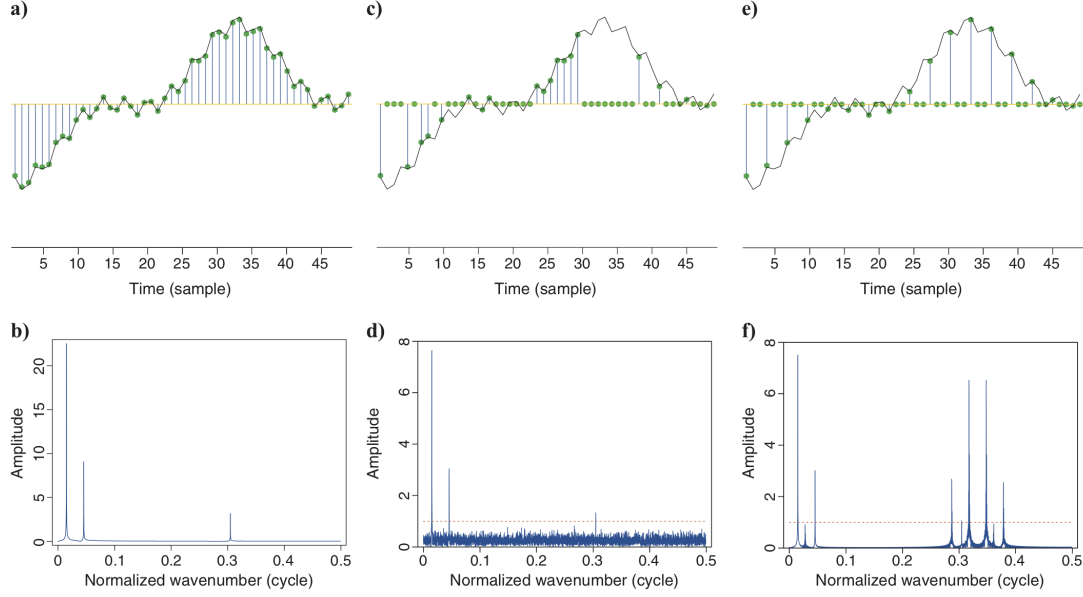


Figure 3.3: Different sampling schemes and their imprint in the Fourier domain for a simple signal: (a,b) Regularly sampled above Nyquist rate and its wavenumber spectrum; (c,d) irregularly threefold subsampled with a discrete uniform distribution and its wavenumber spectrum; (e,f) regularly threefold subsampled and its wavenumber spectrum[7]

### 3.3 Data-driven irregular sampling

Based on the data structure of the 2D benchmark wavefield, a data-driven irregular sampling is also introduced for reconstruction from over-sampled data. The purpose-built irregular sampling is developed based on the priori information of the model, hence the sampling scheme will vary with different data sets. Figure 3.4 is a partitioned sampling scheme (fourfold decimation) based on the 2D example data, with threefold regular subsampling at the center and sixfold irregular subsampling on both flanks, in order to better sample the more complex part of the data (near offsets), while sparsifying the data where the signal is less complex (far offsets). Prior information is required for applying data-driven irregular sampling, thus in this study partitioned sampling is working as an example of implementing priori information in sampling.



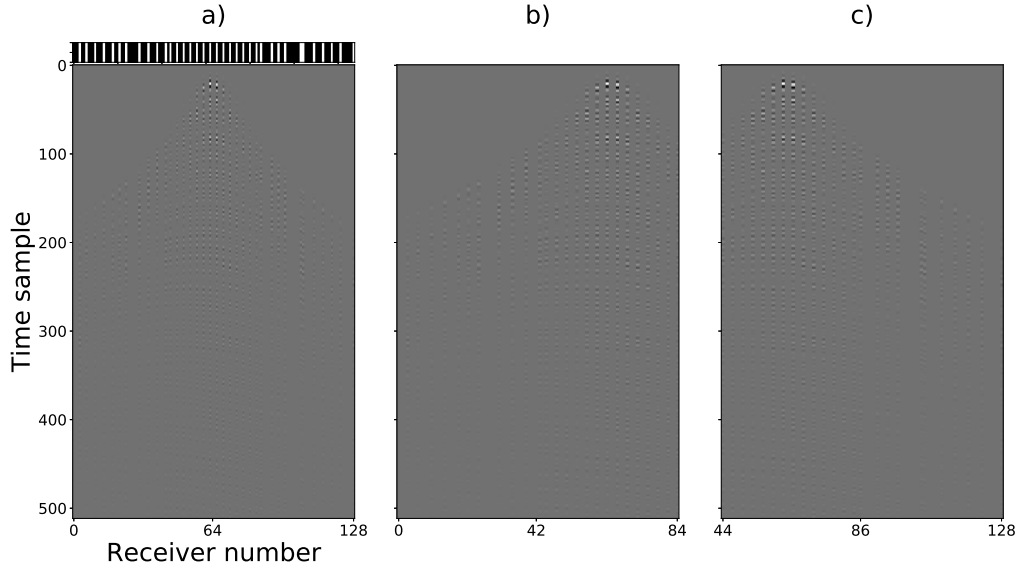


Figure 3.4: Data-driven irregular sampling and corresponding partition of the wavefield based on data structure: a) the priori-based sampling scheme with regular threefold subsampling in near offsets(trace No. 44-84), while irregularly sixfold subsampled on far offsets(trace No. 0-43, 85-128); (b,c) The partition of priori-based sampled wavefield, consisting of two part(trace No.0-84 & trace No.44-128), based on different wavenumber components of both flanks.

## Chapter 4

# Preconditioning

The role of preconditioning is significant when solving an under-determined problem, for which the space of acceptable solutions is narrowed by imposing priori information, in a cost of additional computation (i.e., the cost of preconditioning operators themselves). In this study, we discuss the effect of physics-based priors through preconditioning operators, such as imposing causality in time-space, and enforcing the dispersion relation in f-k domain.

The solving of the problem in eq.2.6 gives the result in space-time domain. However, as is shown in figure 3.1, the seismic shot-gather data set is well-confined and sparser in the Fourier domain compared with space-time domain. Hence, it is more favorable to solve the linear system in the Fourier domain. The preconditioning for solver is implemented by transforming model into the Fourier domain, and the linear equation then becomes:

$$\begin{aligned} \mathbf{d}_{mod} &= \mathbf{F}^{-1} \mathbf{d}_f \\ \mathbf{d}_{obs} &= \mathbf{S} \mathbf{F}^{-1} \mathbf{d}_f \end{aligned} \quad (4.1)$$

The LSQR solver is now solving the following new problem, which gives the model in the Fourier domain  $\mathbf{d}_f$ :

$$\text{minimize} \quad \|\mathbf{d}_{obs} - \mathbf{S} \mathbf{F}^{-1} \mathbf{d}_f\|_2 \quad (4.2)$$

Similarly, physics-based priors can be imposed in the linear equation. For example, the 2D wavefield data in figure 3.1 is acquired in water(i.e., near-constant wavespeed) over a well-behaved geology. Space-time preconditioner can be created with the assumed constant wavespeed and offset, and f-k preconditioner can be made with wavespeed and maximum frequency component. For the Fourier spectrum, we approximate the dispersion relation with an hourglass-shaped preconditioner centered at zero, composed of two opposing triangles in the f-k domain. The preconditioners are implemented as masks and inserted in the linear system:

$$\begin{aligned} \mathbf{d}_{obs} &= \mathbf{S} \mathbf{P}_{xt} \mathbf{F}^{-1} \mathbf{P}_{k\omega} \hat{\mathbf{d}}_f \\ \mathbf{d}_{mod} &= \mathbf{P}_{xt} \mathbf{F}^{-1} \mathbf{P}_{k\omega} \hat{\mathbf{d}}_f \end{aligned} \quad (4.3)$$

Eq.4.3 is the linear equation including the preconditioner ( $\mathbf{P}_{xt}$ ) in space-time domain and the preconditioner( $\mathbf{P}_{k\omega}$ ) in Fourier domain. The inversion of the new linear equation will give the result of  $\hat{\mathbf{d}}_f$ , instead of the model in the Fourier domain. The model is given by the re-application of the preconditioner.

The LSQR solver is used in this study for its stability, though it is designed for solving over-determined problems. Our reconstruction results show that given appropriate preconditioners, LSQR is still able to perform well in inverting our linear equation, in the case that the sampling rate is lower than threefold subsampling Nyquist rate.

Implementing a priori information that enforces sparsity is another preconditioner for wavefield reconstruction. Wavefield data are generally sparser in the Fourier domain than in the space-time domain, and it can be even sparser in other basis. For example, the Curvelet transform[3, 4] is a multi-dimensional and multi-directional extension of the Fourier transform, which will project wavefield data into a even sparser domain. Starting from the sparsity of the data, sparsity-promoting solvers (e.g. SPGL1[22] and FISTA[1], ) tend to have better inversion in the areas with large amount of traces missing, compared with LSQR under the same preconditioner.

$$\begin{aligned}
 SPGL1(BPDN): \quad & \text{minimize} \quad \left\| \hat{\mathbf{d}}_f \right\|_1 \quad \text{subject to} \quad \left\| \mathbf{d}_{obs} - \mathbf{SP} \mathbf{d}_{mod} \right\|_2 \leq \sigma \\
 FISTA: \quad & \text{minimize} \quad \left\| \mathbf{d}_{obs} - \mathbf{SP} \mathbf{d}_{mod} \right\|_2 + \epsilon \left\| \hat{\mathbf{d}}_f \right\| \quad \text{subject to} \quad \mathbf{d}_{obs} = \mathbf{SP} \hat{\mathbf{d}}_f
 \end{aligned} \tag{4.4}$$

The 2D wavefield example has the feature of well-aligned hyperbolic-like events in the time-space domain, which translate into coherent features with interference notches in the Fourier domain, thus being a good example for implementing L1-solvers. In contrast, scattering in the ocean turbulence wavefield appears as more random diffractions in the time-space domain, leading to a far less sparse, incoherent f-k domain spectrum, confined by a simple dispersion relation, thus favouring the use of the LSQR solver.

## Chapter 5

# 2D Wavefield Reconstruction

Before proceeding to ocean turbulence wavefield reconstruction, a benchmark 2D shot-gather seismic wavefield data set[23] is used for exploring and examining the multicomponent reconstruction method with different sampling patterns.

### 5.1 Regular sampling

The multichannel sampling theorem is built on regular sampling condition, thus we test our reconstruction procedure first on regular sampling with half-decimation and threefold-decimation, respectively.

Figure 5.1 is the reconstruction result from regular sampling with sampling rate as half of the original Nyquist rate. The case in which only pressure data is used in the multicomponent reconstruction fails to retrieve the missing traces and the spatial aliasing is still pronounced in the Fourier domain. While the case in which both pressure data and its first-order spatial derivative are used successfully recovers the missing traces and nullifies the spatial aliasing in the Fourier domain. Both of the reconstructions are solved with LSQR solver (400 iterations for each) in space-time domain without preconditioner. The result verifies the multichannel sampling theorem by Linden(1959)[13], that the Nyquist sampling rate can be lower to a half of the original Nyquist rate when the scalar data and its first derivative are available.

The reconstruction results from threefold decimation are shown in figure 5.2. Similar to the previous example, the decimated data set is reconstructed in space-time domain with LSQR solver (400 iterations) and without preconditioner. The case with pressure data only again fails to recover the missing traces, while the use of first-order spatial derivative is able to reduce the spatial aliasing but also fails to reconstruct the original wavefield. The last case verifies the prediction by Robertsson, et al.(2008)[20], when the pressure data and both of its spatial derivatives are available, the effective Nyquist wavenumber is increased by a factor of three, which enables successful wavefield reconstruction from threefold decimation data set.

When the decimation rate of regular sampling is higher than threefold decimation, the multicomponent reconstruction fails to recover the missing traces. Irregular sampling and preconditioner are then imposed to push the limit of subsampling.

### 5.2 Irregular sampling

With our over-sampled data reconstruction examples, we explore the effect of compressive sensing by means of purpose-built irregular sampling and solver preconditioning.

The benchmark 2D wavefield data is characterized with distinctive partitions, with more complex central traces (near offsets) and relatively simple flanks (far offsets). The near-offset traces have higher amplitude and their frequency-wavenumber components are more complex, while the far-offset traces are of lower amplitude and near-constant slopes (simpler Fourier spectrum).

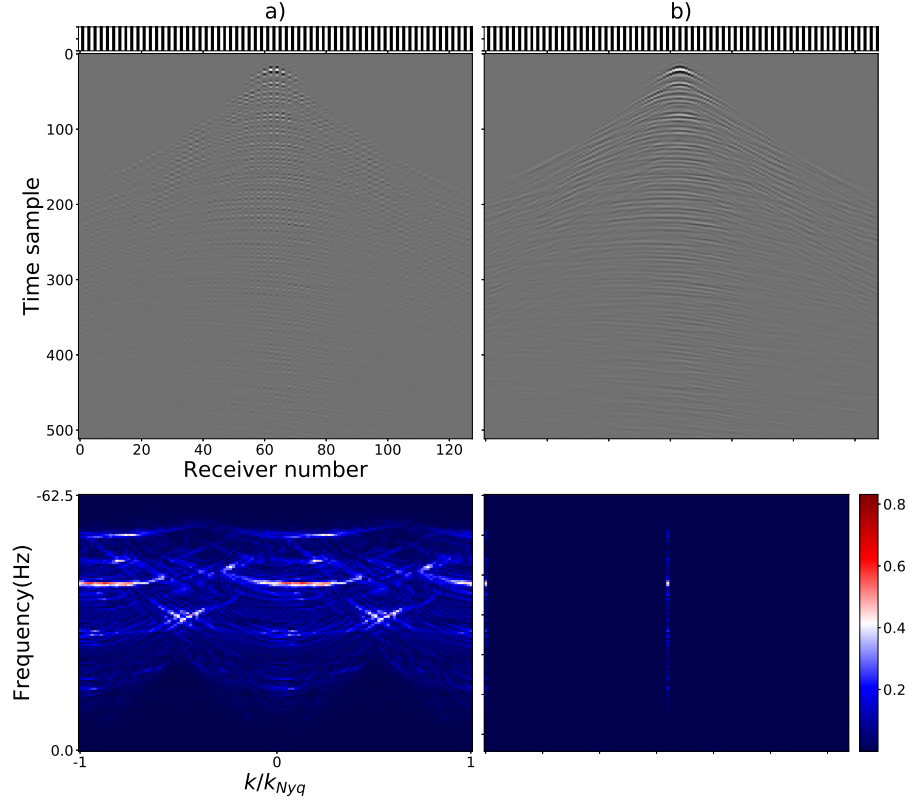


Figure 5.1: Wavefield reconstruction from regularly half subsampled using the multicomponent reconstruction method. Barcode on the top represent sample distribution (white bars for sampled traces, dark bars for missing traces).

Based on this priori information, a data-driven partitioned sampling scheme (figure 3.4) is created by regularly sample the near offsets part with one-third of the original Nyquist rate, and irregularly sample the far offsets part with jittered subsampling. The interval of the jittered sampling nodes is 3, making the possible maximum interval between two traces 6 times the original sampling rate. The overall sampling rate of the scheme is one-fourth of the original Nyquist rate. Our goal here is to set irregular sampling that takes advantage of data structure that is known a priori. In this example, the priori information is that the medium structure is expected to be well-behaved, e.g., quasi-layered.

In this case, preconditioning is implemented in sampling pattern and the linear system. The linear system is solved in the Fourier domain, according to the eq.4.1. Due to different signs of wavenumber components in both flanks, in order to reduce spatial aliasing from reconstruction, the original linear system based on the decimated data set is divided into two linear systems, with one recovering the central traces and left flank, another recovers the central traces and right flanks, shown in figure 3.4. Both share the near-offset part for its extended components in the Fourier domain are the constraint for solving the under-determined problem.

Figure 5.3 is the reconstruction result from the partitioned sampling scheme. The near-offset part is successfully reconstructed and the far-offset part is also retrieved considerably well given the fact that both flanks are over-sampled, and the errors are mainly cluster around the areas where a large number of traces are missing. Our cases show that by combining application-specific sampling schemes and preconditioning, together with multichannel sampling, the overall residual is notably reduced for over-sampled data sets.

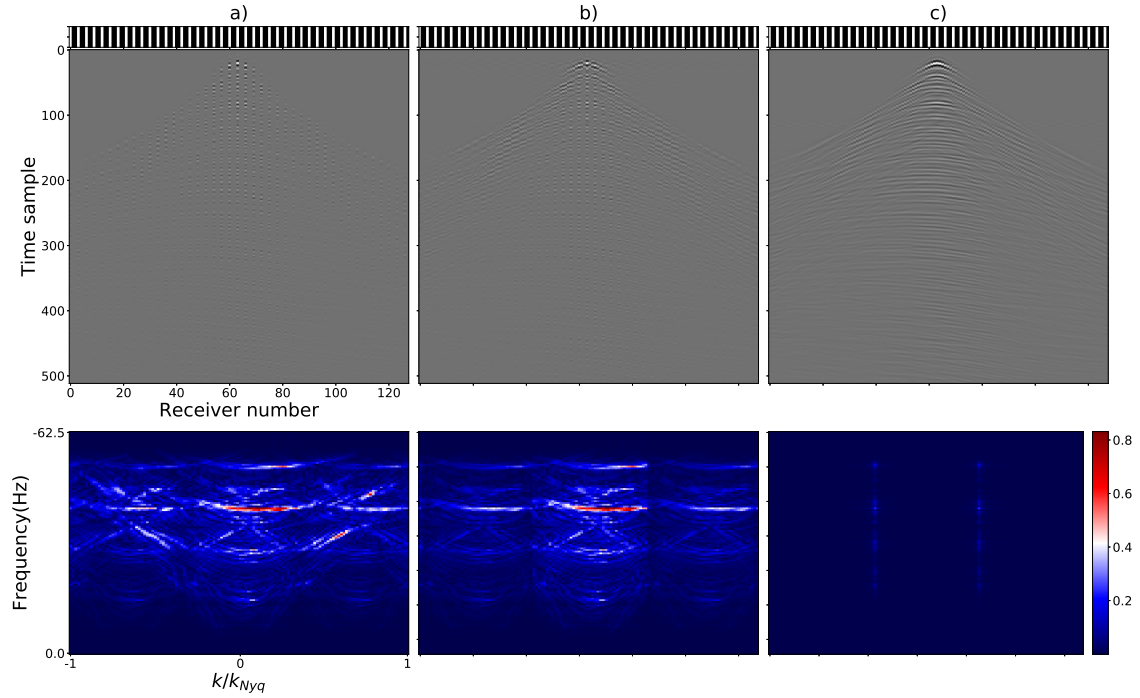


Figure 5.2: Wavefield reconstruction from regularly threefold subsampled using the multicomponent reconstruction method. Barcode on the top represent sample distribution (white bars for sampled traces, dark bars for missing traces)

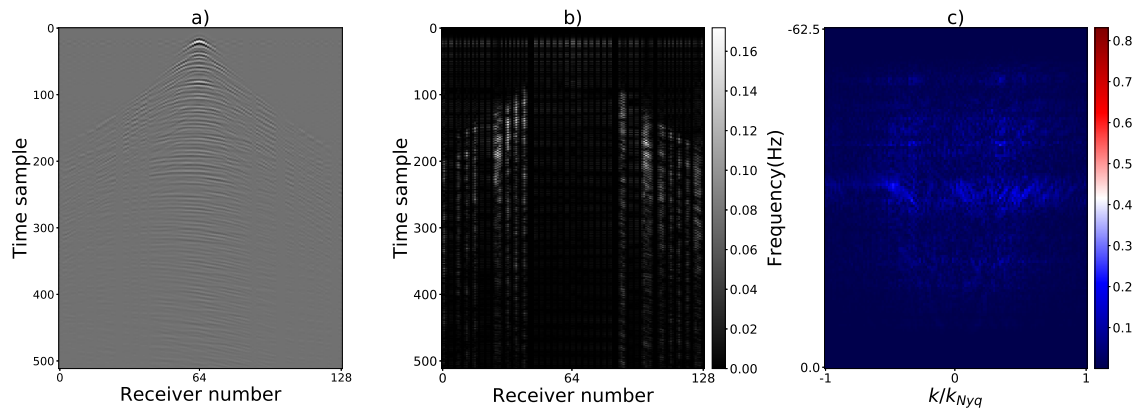


Figure 5.3: Wavefield reconstruction from the data-driven irregular sampling scheme using the multicomponent reconstruction method.

## Chapter 6

# Turbulence Wavefield Inversion

Moving on to ocean turbulence wavefield reconstruction, we choose to test our multicomponent reconstruction method on synthetic data sets, which are generated from the 2D model with its wavespeed configuration shown in figure 6.1. Temperature and salinity are affected by ocean turbulence, which interacts with different water layers and induces density variation in the boundaries of the turbulence. Therefore, despite the wavespeed in water is relatively homogeneous, the presence of ocean turbulence will still be able to generate medium contrasts in the ocean.

Notice that turbulence will vary with time as well as seismic acquisition requires time to make measurements, hence the medium is supposed to change during each shot measurement. This property makes strong constraint to small scale seismic acquisition, yet the model used in this research has the size of  $4000m \times 2000m$ , with the coverage of the shots and receivers up to  $3000m$  and trace interval as  $10m$  and can be several times larger for spatial over-subsampling, which dominates the scale of turbulence variations in the model, leading to an assumption that the medium is considered to be static during seismic acquisition.

The original seismic wavefield is generated by running shots in the model and each seismic shots are made separately and only one shot is measured for each acquisition. For the full set of data, the shots and receivers are located on the top of the model, and both of them share the same locations. We here assume that the receivers that are closed to the source locations still have high signal-to-noise ratio over the full wavefield, instead of being saturated by direct arrivals.

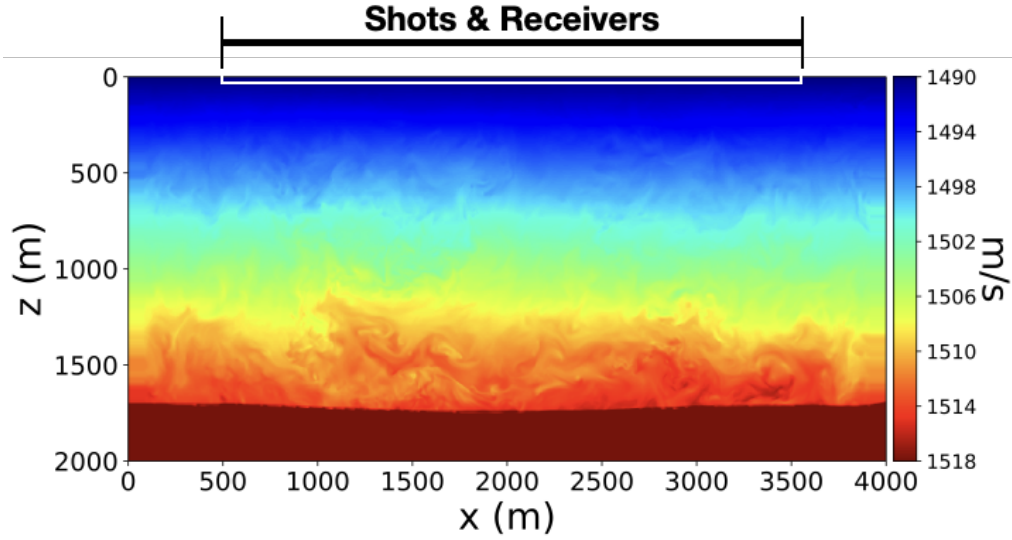


Figure 6.1: 2D ocean turbulence model, with layered wavespeed configuration while overall wavespeed is relatively homogeneous.

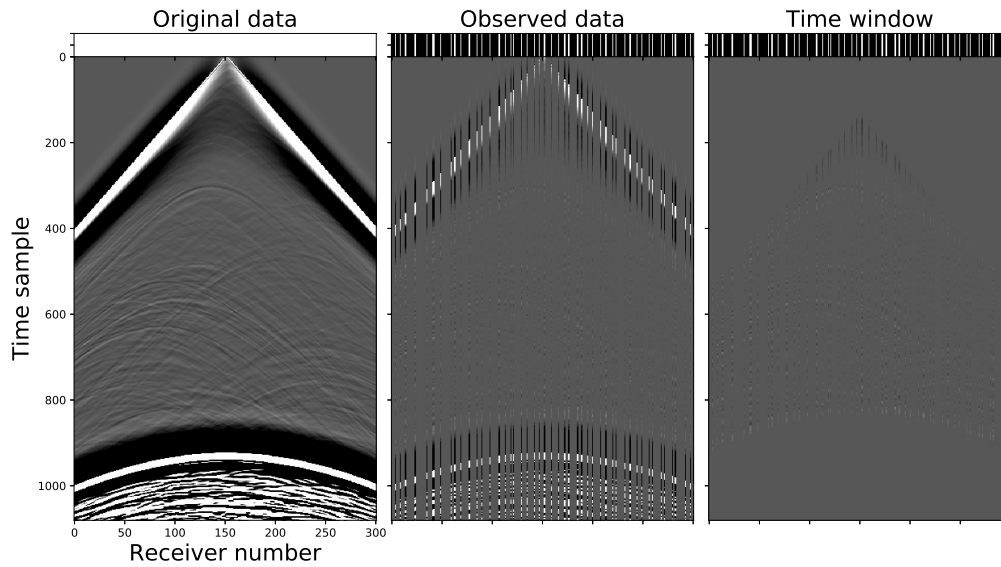


Figure 6.2: Irregularly sixfold subsampled turbulencen wavefield with jittered subsampling and the filtering of direct waves and sediment reflections based on offset and wavespeed. Barcode on the top represent sample distribution (white bars for sampled traces, dark bars for missing traces).



## 6.1 2D synthetic ocean turbulence wavefield reconstruction

Based on the example 2D wavefield reconstruction, a shot-gather generated from the model with the source location in the middle of the 301 evenly distributed receivers is explored in order to make the method adapt to the ocean turbulence wavefield. To this end, we choose the central shot-gather data to include as many seismic events as possible. Figure 6.2 shows the data structure of the wavefield, including strong direct arrivals, sediment reflections, turbulence diffraction and reflection. Among which the turbulence diffraction and reflection between the direct arrivals and sediment reflections are of most value, for they are necessary for turbulence inversion. Though the later arrivals also contain information from turbulence, the strong sediment signals overshadow turbulence seismic signals which may strongly influence event reconstruction.

In the context of reconstruction, we suppress direct arrivals and sediment reflections from the data. Because of the low amplitude of the turbulence diffractions and reflections, solvers tend to reach convergence too early because they will fit the louder direct and sediment-related events first, leaving the desired turbulence-related events poorly reconstructed. Accordingly, the data set after filtering is normalized with the maximum value of the whole space-time domain to help numerical convergence of the iterative solvers. Figure 3.3) is the normalized turbulence seismic wavefield and its corresponding f-k domain.

In our cases, the gradient-based reconstruction successfully reconstructs the turbulence signals under regular sampling with interval three times as the original sampling interval. Following to the procedure of the multicomponent reconstruction, irregular sampling and preconditioners are imposed to further decimate the data set and keep reconstruction results comparable to the original wavefield.

We choose to implement physics-based priori information based on data structure in space-time domain and f-k domain. The preconditioner in space-time domain is designed in the same way with the filter which filters the direct waves and sediment signals, by means of the calculation from offsets. As for preconditioner in f-k domain, it starts with the data structure, which is determined by the relatively homogeneous wavespeed of the model and has the shape of "hourglass" (shown in figure 6.3). The f-k preconditioner is imposed by creating a mask in f-k domain built according to the hourglass-like structure of the Fourier spectrum. Furthermore, due to the dispersed apexes of the hyperbolas in the turbulence wavefield in time-space, its Fourier spectrum is relatively homogeneous within the mask, hence LSQR solver is selected for turbulence wavefield reconstruction.

Figure 6.2b is a irregular sampling case using jittered subsampling with sampling rate as one-sixth of the original sampling rate. The partitioned sampling used in 2D example data is not suitable for the dispersed hyperbola distribution in the turbulence wavefield. The multicomponent reconstruction based on eq.4.3 is applied to retrieve the original wavefield from the subsampled data set, solving with LSQR solver in 600 iterations. After the inversion of the linear equation, preconditioning and scaling are applied back to retrieve the original wavefield. The result and its relative error in space-time domain and f-k domain are shown in figure 6.4. The reconstruction is analogous to the original shot-gather data, despite the original data is subsampled with six times fewer traces. The errors mainly focus on low wavenumber and high frequency components, subject to high-order spatial aliasing.

The 2D turbulence wavefield reconstruction provides guides for 3D wavefield reconstruction, since the filtering is based on offsets and the mask in f-k domain is designed according to the wavespeed in the ocean.

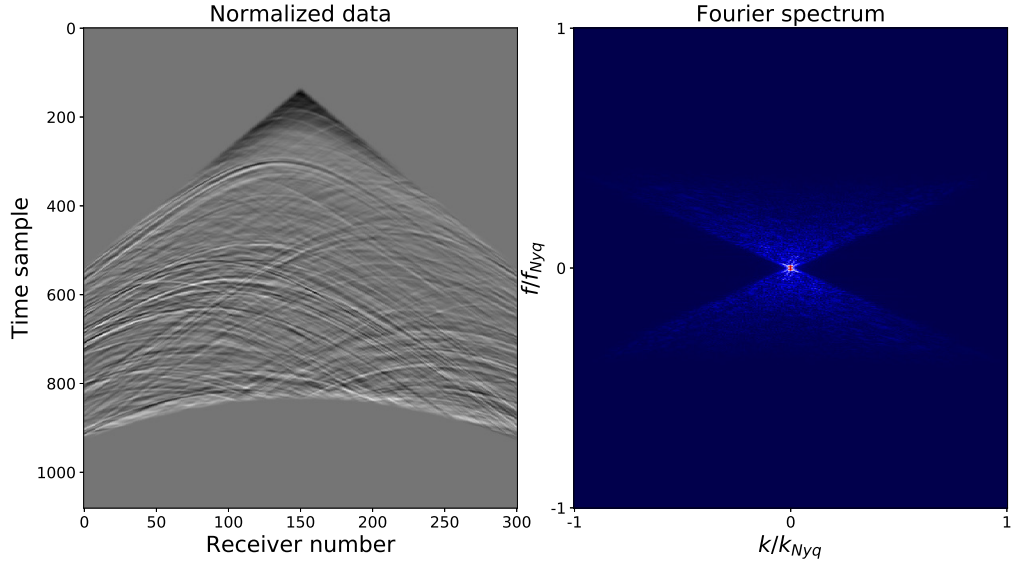


Figure 6.3: Normalized wavefield and its Fourier spectrum after the filtering of direct arrivals and sediment reflections.

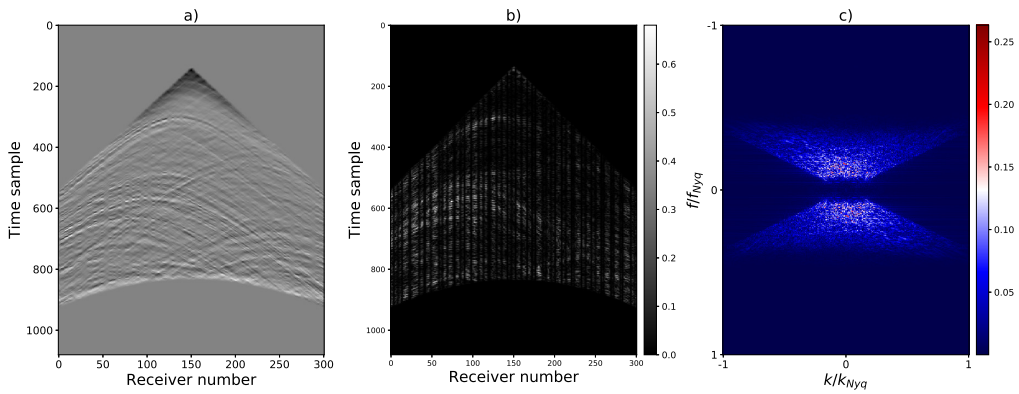


Figure 6.4: Wavefield reconstruction from irregularly sixfold subsampled filtered turbulence wavefield with compressive sensing and physics-based preconditioning: a) reconstruction result in space-time domain; b) corresponding residual in space-time domain; c) its relative errors in the Fourier domain, maximum normalization.

## 6.2 3D synthetic ocean turbulence wavefield reconstruction

The 3D synthetic ocean turbulence wavefield discussed in this research is still generated from the same 2D space of the model, but with the same number of shots as the number of receivers. More specifically, the shots and receivers share the same location, and the assumption that close-to-shot receivers can measure the shots with high signal-to-noise ratio remains valid.

Decimation rate discussed in 3D sampling pattern refer to the number of shots and receivers, instead of the number of traces. For example, threefold decimation is equal to keeping nine times fewer traces from the original data set. More importantly, the decimation is acting on the same location from shot side and receivers side, meaning that each location will either have both shot and receiver or none of them, depending on the sampling pattern.

In 3D cases, because of the increase of dimensionality, derivatives used in the multicomponent reconstruction will not only be limited to the first- and second-order derivatives in the receiver direction, but also with derivatives in shot direction and cross-derivatives, with the maximum order of two. The l.h.s and the sampling operator now become:

$$\mathbf{d}_{obs} = \begin{pmatrix} \mathbf{d}_t \\ \mathbf{d}_{r1} \\ \mathbf{d}_{r2} \\ \mathbf{d}_{s1} \\ \mathbf{d}_{s2} \\ \mathbf{d}_{r1s1} \\ \mathbf{d}_{r2s1} \\ \mathbf{d}_{r1s2} \\ \mathbf{d}_{r2s2} \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} \mathbf{R}_r \\ \mathbf{D}_{r1} \\ \mathbf{D}_{r1} \\ \mathbf{D}_{s1} \\ \mathbf{D}_{s2} \\ \mathbf{D}_{r1s1} \\ \mathbf{D}_{r2s1} \\ \mathbf{D}_{r1s2} \\ \mathbf{D}_{r2s2} \end{pmatrix} \quad (6.1)$$

Where  $\mathbf{d}_{rx}$  and  $\mathbf{d}_{sx}$  represent derivatives in receiver direction and shot direction, respectively.  $\mathbf{d}_{rxsx}$  are the cross-derivatives measured by the receivers. On the sampling operator side,  $\mathbf{D}_{rx}$  and  $\mathbf{D}_{sx}$  are corresponding operators for derivatives calculation in receiver direction and shot direction, in which we use the similar Fourier representation as eq. 2.5.  $k_r$  and  $k_s$  share the same axis since shot intervals and receiver intervals and their numbers are identical.  $\mathbf{D}_{rxsx}$  is the corresponding cross-derivatives operators:

$$\begin{aligned} \mathbf{D}_{r1s1} &= \mathbf{R}_{rs} * \mathbf{F}^{-1} * \mathbf{diag}(ik_{rn}) * \mathbf{diag}(ik_{sn}) * \mathbf{F} \\ \mathbf{D}_{r2s1} &= \mathbf{R}_{rs} * \mathbf{F}^{-1} * \mathbf{diag}[(ik_{rn})^2] * \mathbf{diag}(ik_{sn}) * \mathbf{F} \\ \mathbf{D}_{r1s2} &= \mathbf{R}_{rs} * \mathbf{F}^{-1} * \mathbf{diag}[(ik_{rn})] * \mathbf{diag}(ik_{sn}^2) * \mathbf{F} \\ \mathbf{D}_{r2s2} &= \mathbf{R}_{rs} * \mathbf{F}^{-1} * \mathbf{diag}[(ik_{rn})^2] * \mathbf{diag}(ik_{sn}^2) * \mathbf{F} \end{aligned} \quad (6.2)$$

$\mathbf{R}_{rs}$  is the restriction operator for decimating the receivers and shots,  $\mathbf{diag}[(ik_{rn})]$  and  $\mathbf{diag}[(ik_{sn}^2)]$  are diagonal matrices with wavenumber in shot direction and receiver direction. Cross-derivatives are necessary for recovering the missing shots from the decimated data set. Figure 6.7 is part of the 3D reconstruction result from a threefold regular decimation case that no cross-derivatives are included in the multicomponent reconstruction. The reconstruction fails to recover the missing shots, but has the sampled shot well reconstructed, which is in line with the result in the 2D reconstruction case.

Filtering of the direct waves and sediment reflections is applied based on offsets. Preconditioners in 3D reconstruction are derived from the 2D turbulence wavefield reconstruction procedure. The mask in space-time domain is in the same design with the filter, while the mask in the Fourier domain is more complicated due to the increased dimensionality and the filtering in space-time domain. Figure 6.5 is the wavenumber spectrum at four specific frequencies scaled by the highest frequency component of the original wavefield, indicating similar f-k relation with the 2D single shot data set.

As a result of the filtering in space-time domain, artifacts are created in a form of constant wavenumber components in the f-k domain, and their amplitude and range vary with frequency.

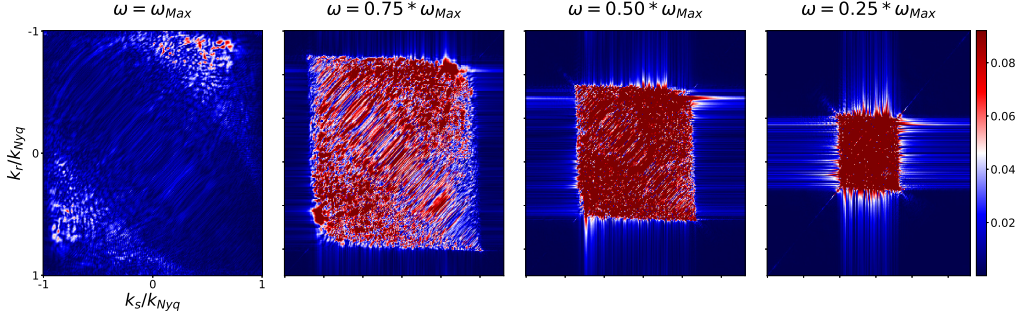


Figure 6.5: Normalized 3D turbulence wavefield wavenumber spectrum for specific frequencies with first arrivals and sediment reflections filtered; constant wavenumber components outside the high-amplitude rectangle on the center are considered to be the artifacts created by filtering of the direct waves and sediment reflections.

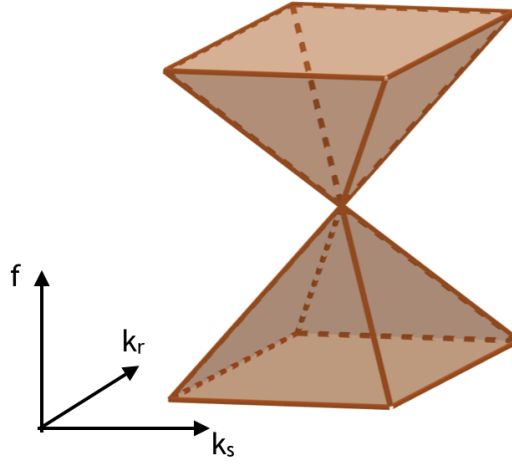


Figure 6.6: "opposing-pyramids" mask in Fourier domain.

The artifacts are of high amplitude in low frequency component in a form of a "cross" shown in figure 6.5. We then build the mask in  $f$ - $k$  domain based on the wavespeed as an extension of the hourglass-like 2D mask design, with "opposing-pyramids" with the shared tip at zero-frequency in  $\omega - k_s - k_r$  domain, as shown in figure 6.6. The mask keeps only the high-amplitude rectangular part for each frequency below the highest frequency component, and leaves the artifacts out.

This 3D pyramidal-hourglass structure of the mask will deteriorate the reconstruction over low frequency components since it set most the low frequency components as zero. On the other hand, the mask can improve the recovery of middle-to-high frequency components, the major components for the reflections and diffractions, because the amplitudes of low frequency components are considerably larger than other frequency components. By filtering the lower frequency components, solvers tend to lower the misfit on middle-to-high frequency components, resulting in well-reconstructed reflections and diffractions.

Figure 6.8 is part of the reconstruction result from fourfold decimation with irregular sampling, having sixteen times fewer traces than the original data set. The inversion is based on eq.2.2 and 6.1, using LSQR solver with 1000 iterations. Within these eight shot-gather panels, only the first and the last shot are recorded. Traces in the six missing shots are recovered by the multicomponent reconstruction, as well as the missing receivers in the two sampled shots. The residual figure 6.9 indicates that the sampled shots are well-reconstructed, and the residual is positively related to the interval from the nearest sampled shot. The two middle shots are of relatively high residual,

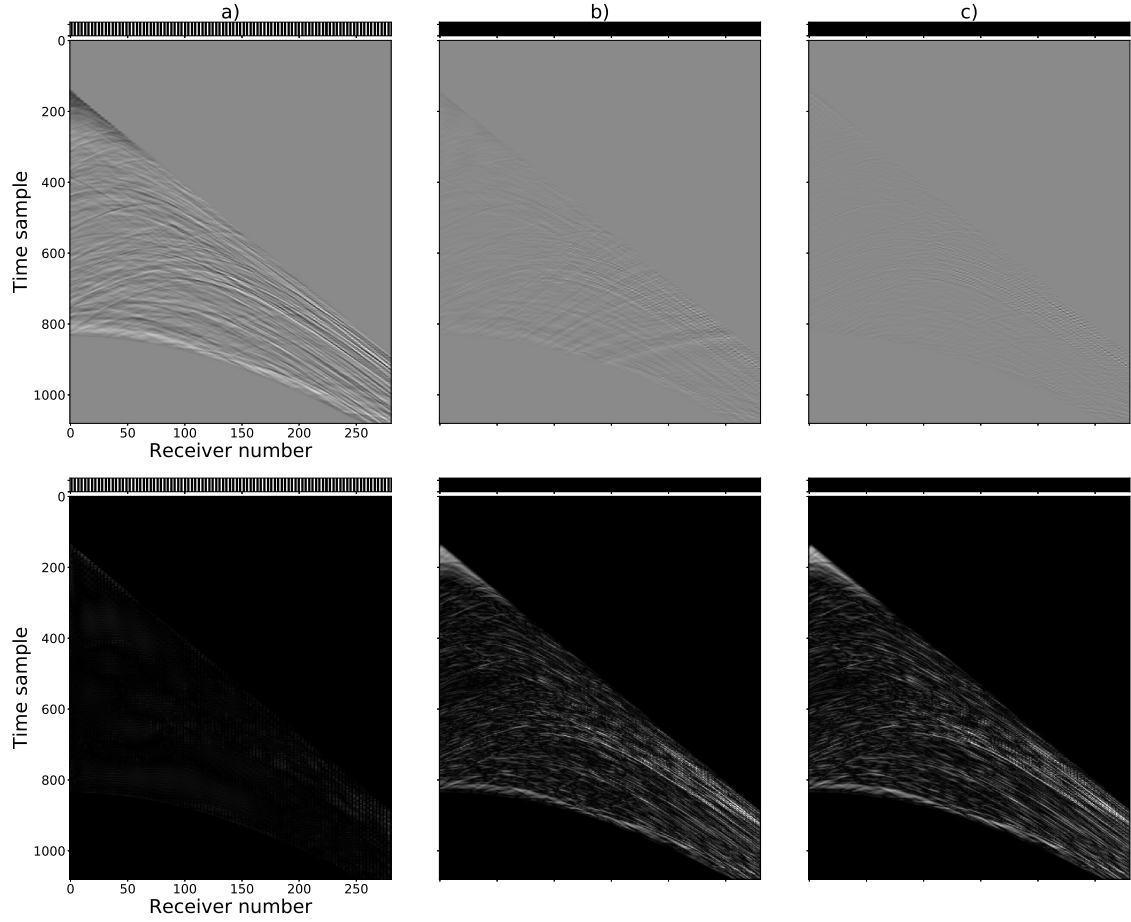


Figure 6.7: 3 shot-gathers (shot 0 - shot 2) from 3D turbulence wavefield reconstruction without cross-derivatives, barcode on the top represent sample distribution (white bars for sampled traces, dark bars for missing traces): a) reconstruction on shot 0 (sampled shot) and its residual in space-time domain; b) reconstruction on shot 1 (unsampled shot) and its residual in space-time domain; c) reconstruction on shot 2 (unsampled shot) and its residual in space-time domain.

nevertheless major seismic events are still appropriately recovered. Densely sampled locations have of course more accurate reconstruction than the eight panels in figure 6.8, and the receiver-gather panels has the same pattern, as shown in figure 6.10 and 6.11.

For shots and receivers that are located at the left boundary of the sampling range, both shot-gather and receiver-gather panels show aliasing that is only limited to the first two shots and receivers, as shown in figure 6.12. We suggest that the aliasing is result from the design of the mask in  $f$ - $k$  domain, which might also filter some of the high-frequency, high-wavenumber components. However, as this aliasing is only limited to the two shot-gather and receiver-gather panels on the left boundary, the preconditioner in  $f$ - $k$  domain is considered to be effective thus at this point we make no further alteration on the mask.

The 3D turbulence wavefield reconstruction is developed based on the the procedure of 2D shot-gather turbulence wavefield reconstruction, with the similar sampling pattern preconditioners in space-time domain and  $f$ - $k$  domain. With the addition of cross-derivatives, the multicomponent reconstruction successfully recovers the traces from missing shots and receivers from fourfold decimation. The retrieved wavefields are generally comparable to the original wavefield with acceptable errors, clustering around the area where a large number of shots and receivers are missing, however, the major turbulence seismic events are distinguishable.

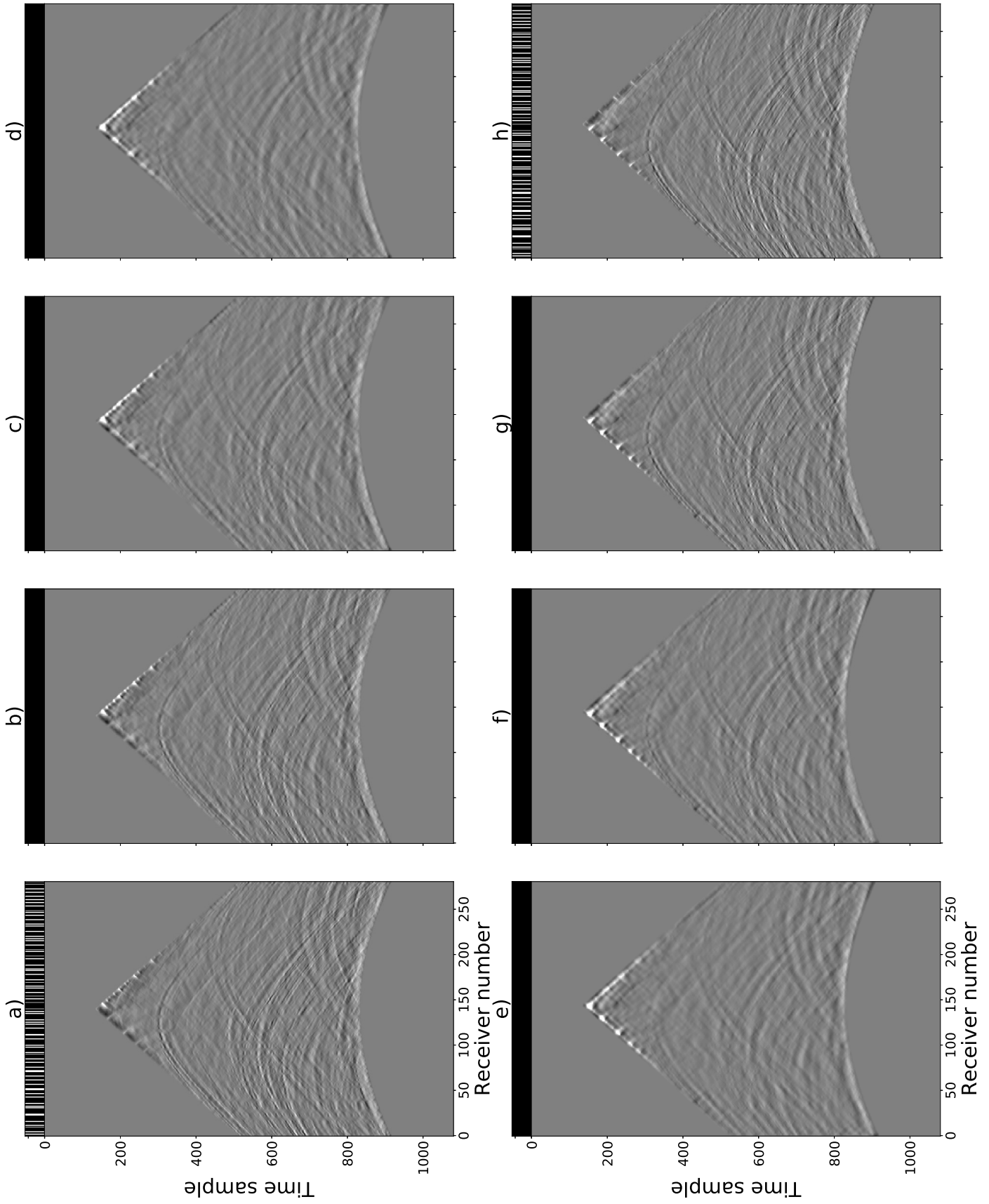


Figure 6.8: 8 shot-gathers (shot 140 - shot 147) from the 3D turbulence wavefield reconstruction, barcode on the top represent sample distribution (white bars for sampled traces, dark bars for missing traces)

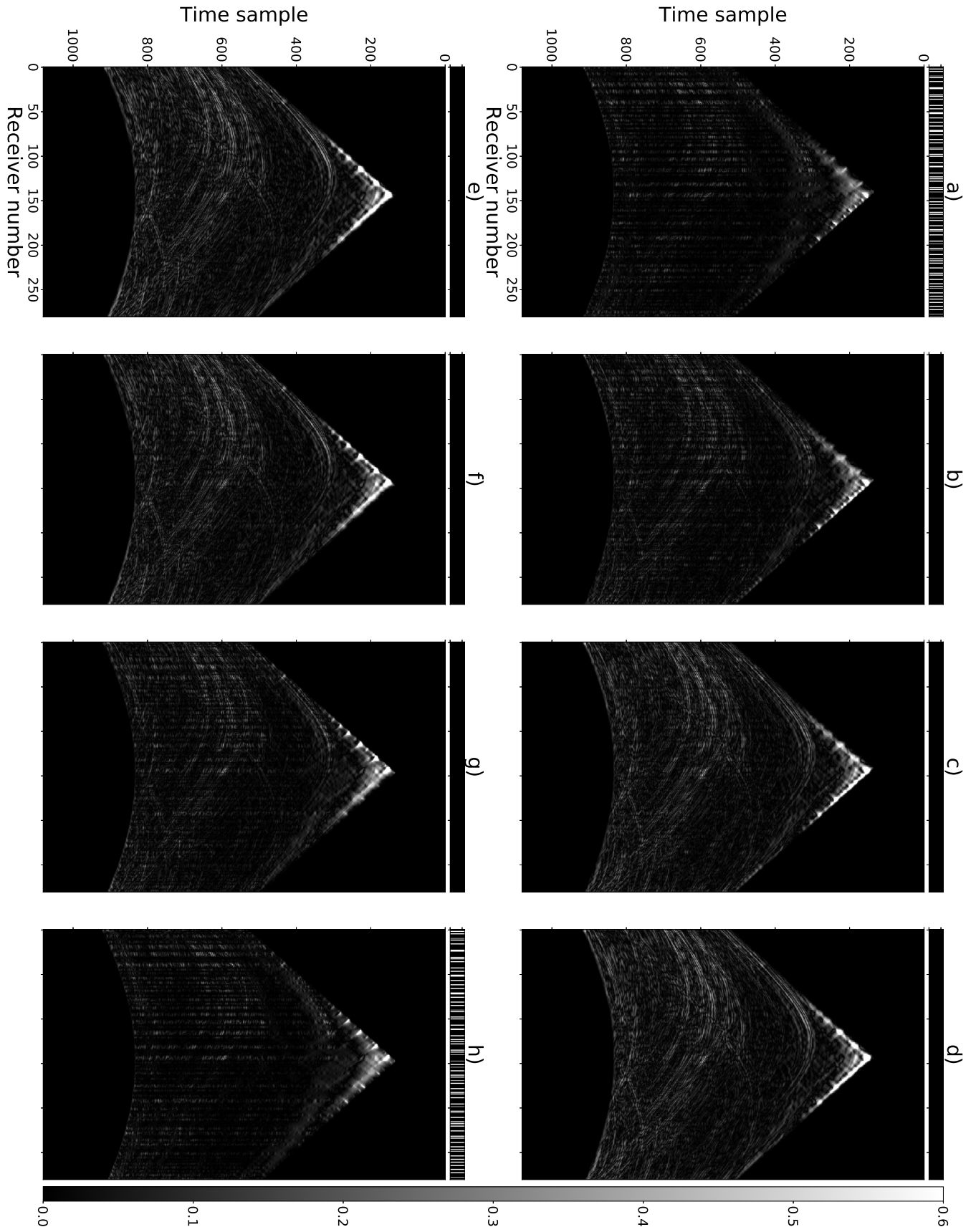


Figure 6.9: Relative residual for the 8 shot-gathers (shot 140 - shot 147) from the 3D turbulence wavefield reconstruction. Overall low residuals at sampled shot, while the residuals increase with sampling interval.

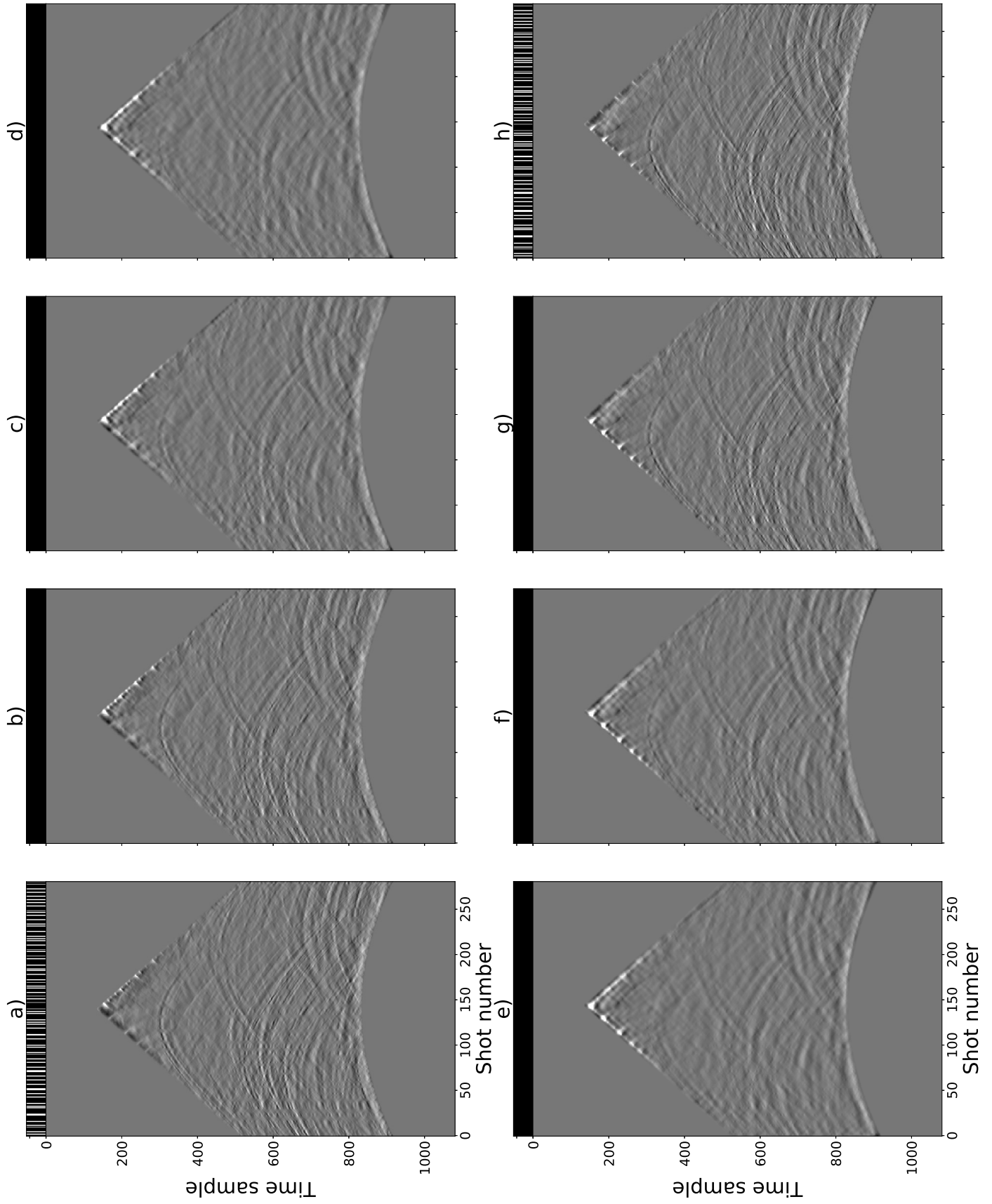


Figure 6.10: 8 receiver-gathers (receiver 140 - shot 147) from the 3D turbulence wavefield reconstruction, barcode on the top represent sample distribution (white bars for sampled traces, dark bars for missing traces)



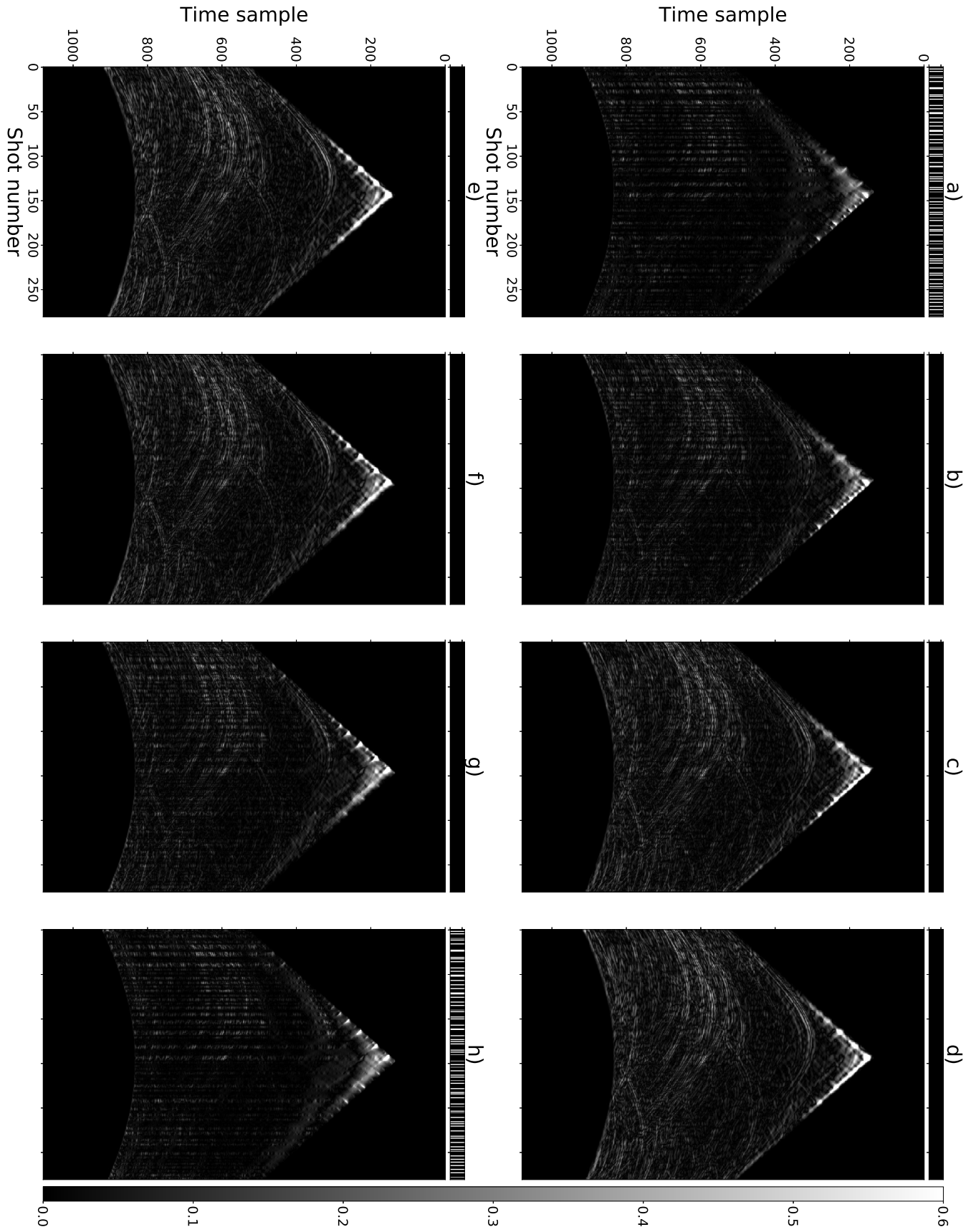


Figure 6.11: Relative residual for the 8 receiver-gathers (receiver 140 - shot 147) from the 3D turbulence wavefield reconstruction. Similar patterns with shot-gather residuals: residuals increase with sampling interval.

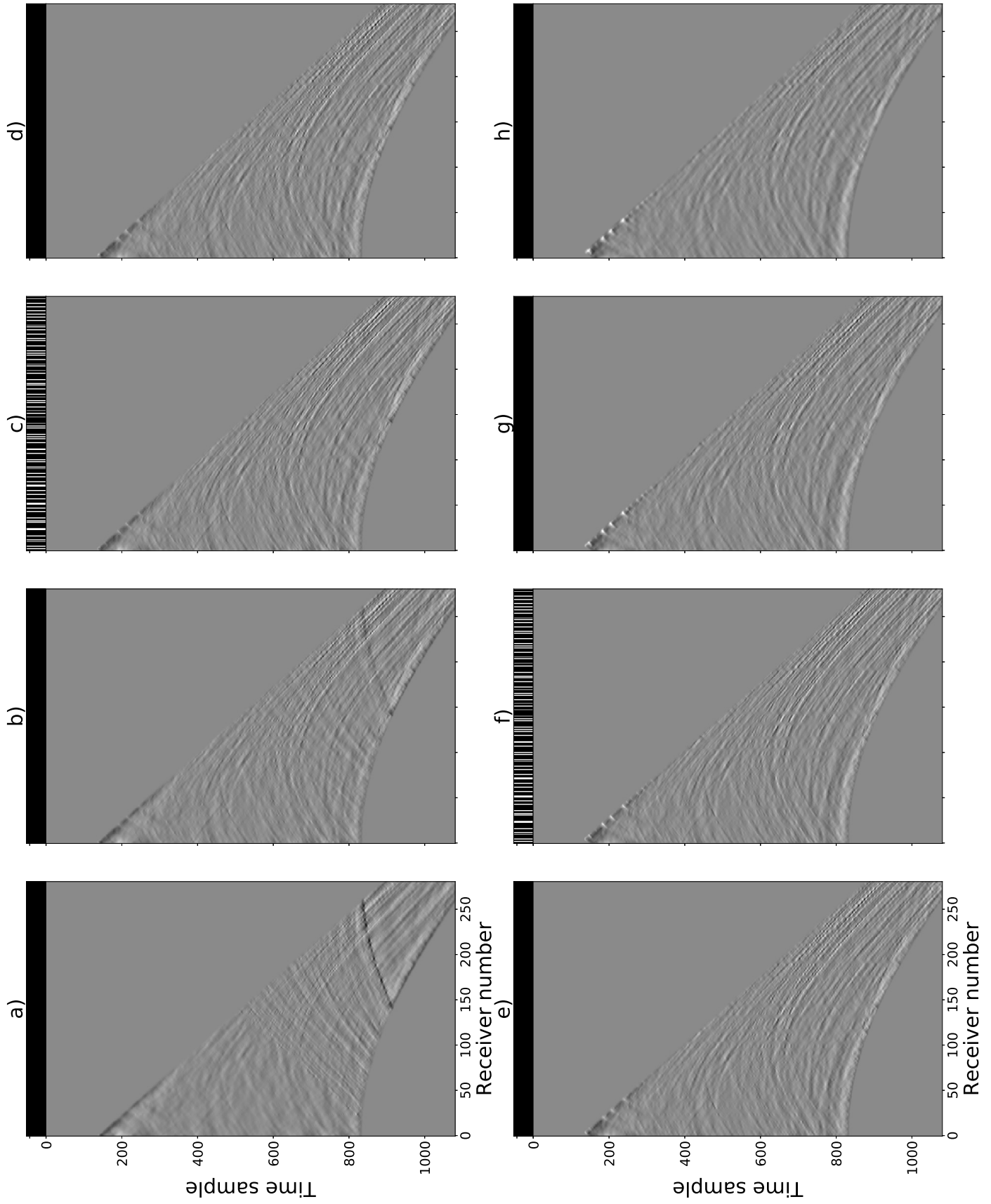


Figure 6.12: Left boundary shots (shot 0 - shot 7) from the 3D turbulence wavefield reconstruction. Alias is pronounced on shot 1 & shot 2, and decay fast with shot location.

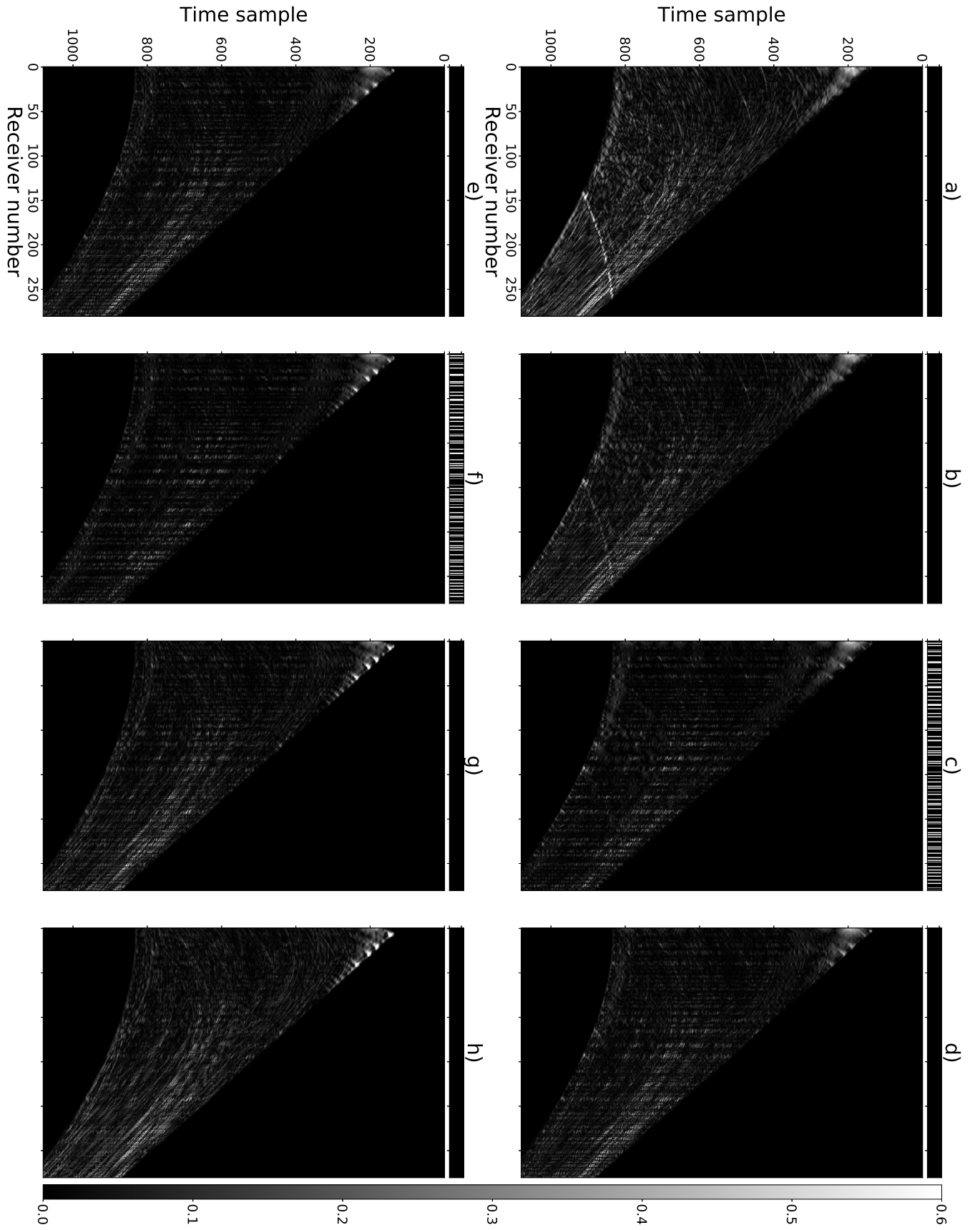


Figure 6.13: Relative residual on the left boundary shots (shot 0 - shot 7) from the 3D turbulence wavefield reconstruction.

## Chapter 7

# Conclusions

In order to develop a seismic wavefield reconstruction method to recover turbulence wavefield from sparse multichannel measurements, we discuss a gradient-based wavefield reconstruction along with compressive sensing, based on multichannel sampling theorem including high-order and multi-dimension spatial derivatives. Given some a priori information about the expected structure of the data, we design bespoke sampling and inverse construction schemes for over-sampled multichannel data.

In the example of 2D wavefield recovery, we test the high-order multichannel reconstruction on a seismic shot-gather, decimated with different sampling schemes. Our results verify the predictions of the multichannel sampling theorem, that for regular sampling, the Nyquist sampling rate can be lower by a factor of three when the pressure data and both of its spatial derivatives are available. Our prior-based irregular sampling scheme samples the wavefield densely and regularly at near offsets, where the data are more complex, while being sparse and irregular at the far offsets, seeking to acquire maximum information with minimum samples. In the accompanying reconstruction approach, preconditioning is designed in the f-k domain, splitting up of the linear system according to prior information on the irregular sampling and wavefield structure. The bespoke compressive sensing and preconditioning approaches increase the effective Nyquist wavenumber and lower the overall reconstruction residual.

The retrieval of turbulence wavefield from sparse multichannel measurements is tested on synthetic wavefields which are generated from an ocean turbulence model, where wavespeed is layered due to presence of turbulence while overall wavespeed is relatively homogeneous. We explore the turbulence wavefield reconstruction by imposing physics-based preconditioners along with the multicomponent reconstruction on a irregular sampled single shot-gather generated from the model. The preconditioners, in space-time domain and f-k domain, further increase the effective Nyquist wavenumber, and enable successful reconstruction from sixfold decimation data set.

Following to 3D wavefield reconstruction, irregular sampling is now affecting both shot and receiver axis. Receivers are able to measure derivatives in shot and receiver directions, as well as cross-derivatives. The multicomponent reconstruction and preconditioners which were used in 2D wavefield reconstruction are adapted for 3D wavefield data set. Our case, which only has one-fourth of the shots and receivers (1/16 of the original traces left), proves a strong reconstruction from the over-sampled data set, with the major seismic events well-reconstructed, while residual cluster around the areas where a large number of shots and receivers are missing.

The multicomponent reconstruction, which uses derivatives data calculated by the Fourier representation from the original wavefield in space-time, will inevitably suffer from noise and errors in field data, hence further research is needed to process a real seismic data set in such scenarios. Meanwhile, the turbulence wavefield reconstruction discussed in the project is limited to large scale acquisition, which enables the assumption of static medium. Understating sampling requirements, acquisition design and signal reconstruction for moving/flowing media is also the topic of future research.

# Bibliography

- [1] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1):183–202, 2009. 12
- [2] Berta Biescas, Valentí Sallarès, Josep Lluís Pelegrí, Francisco Machín, Ramón Carbonell, Grant Buffett, Juan José Dañobeitia, and Alcino Calahorrano. Imaging meddy finestructure using multichannel seismic reflection data. *Geophysical Research Letters*, 35(11), 2008. 1
- [3] E[mmanuel] J[] Candes and D[avid] L[] Donoho. Continuous curvelet transform: I. resolution of the wavefront set. *Applied and Computational Harmonic Analysis*, 19(2):162–197, 2005. 12
- [4] E[mmanuel] J[] Candes and D[avid] L[] Donoho. Continuous curvelet transform: Ii. discretization and frames. *Applied and Computational Harmonic Analysis*, 19(2):198–222, 2005. 12
- [5] Sandro Carniel, Andrea Bergamasco, Jeffrey W Book, Richard W Hobbs, Mauro Sclavo, and Warren T Wood. Tracking bottom waters in the southern adriatic sea applying seismic oceanography techniques. *Continental shelf research*, 44:30–38, 2012. 1
- [6] G[illes] Hennenfent and F[elix] J[] Herrmann. Simply denoise: Wavefield reconstruction via jittered undersampling. *Geophysics*, 73(3):V19–V28, 2008. 7
- [7] Felix J Herrmann. Randomized sampling and sparsity: Getting more information from fewer samples. *Geophysics*, 75(6):WB173–WB187, 2010. 7, 9
- [8] W Steven Holbrook and Ilker Fer. Ocean internal wave spectra inferred from seismic reflection transects. *Geophysical Research Letters*, 32(15), 2005. 1
- [9] W Steven Holbrook, Ilker Fer, Raymond W Schmitt, Daniel Lizarralde, Jody M Klymak, L Cody Helfrich, and Robert Kubichek. Estimating oceanic turbulence dissipation from seismic images. *Journal of Atmospheric and Oceanic Technology*, 30(8):1767–1788, 2013. 1
- [10] W Steven Holbrook, Pedro Páramo, Scott Pearse, and Raymond W Schmitt. Thermo-haline fine structure in an oceanographic front from seismic reflection profiling. *Science*, 301(5634):821–824, 2003. 1
- [11] Dirk Klaeschen, RW Hobbs, Gerd Krahmann, Cord Papenberg, and E Vsemirnova. Estimating movement of reflectors in the water column using seismic oceanography. *Geophysical Research Letters*, 36(24), 2009. 1
- [12] Jean Kormann, Pedro Cobo, and Andres Prieto. Perfectly matched layers for modelling seismic oceanography experiments. *Journal of Sound and Vibration*, 317(1-2):354–365, 2008. 1
- [13] D. A. Linden. A discussion of sampling theorems. *Proceedings of the IRE*, 47(7):1219–1226, 1959. 1, 4, 13

- [14] N[ick] Moldoveanu et al. Marine seismic acquisition with autonomous marine vehicles towing 3d sensor arrays. *Teh Leading Edge*, pages 558–565, 2017. 1, 6
- [15] E[verhard] Muyzert. Design, modelling and imaging of marine seismic swarm surveys. *Geophysical Prospecting*, 66:1535–1547, 2018. 1
- [16] Papia Nandi, W Steven Holbrook, Scott Pearse, Pedro Páramo, and Raymond W Schmitt. Seismic reflection imaging of water mass boundaries in the norwegian sea. *Geophysical Research Letters*, 31(23), 2004. 1
- [17] A[li] Özbek et al. Crossline wavefield reconstruction from multicomponent streamer data: Part 2—joint interpolation and 3d up/down separation by generalized matching pursuit. *Geophysics*, 75(6):WB69–WB85, 2010. 1
- [18] C[hristopher] C[ ] Paige and M[ichael] A[ ] Saunders. Lsq: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)*, 8(1):43–71, 1982. 5
- [19] Cord Papenberg, Dirk Klaeschen, Gerd Krahmann, and RW Hobbs. Ocean temperature and salinity inverted from combined hydrographic and seismic data. *Geophysical Research Letters*, 37(4), 2010. 1
- [20] J[ohan] O[ ] A[ ] Robertsson et al. On the use of multicomponent streamer recordings for reconstruction of pressure wavefields in the crossline direction. *Geophysics*, 70:70–81, 2008. 1, 4, 13
- [21] C[onstantin] Tsingas et al. Spicerack: An autonomous underwater vehicle for efficient seabed seismic acquisition. *SEG Expanded Abstracts*, pages 5953–5957, 2017. 1
- [22] E[ ] van den Berg and M[ ] P[ ] Friedlander. Probing the pareto frontier for basis pursuit solutions. 31(2):890–912, 2008. 12
- [23] E. van den Berg and M. P. Friedlander. Spot: A linear-operator toolbox, August 2013. <https://www.cs.ubc.ca/labs/scl/spot/>. 7, 8, 13
- [24] M[assimiliano] Vassallo et al. Crossline wavefield reconstruction from multicomponent streamer data: Part 1—multichannel interpolation by matching pursuit (mimap) using pressure and its crossline gradient. *Geophysics*, 75(6):WB53–WB67, 2010. 1, 2
- [25] Warren T Wood, W Steven Holbrook, Mrinal K Sen, and Paul L Stoffa. Full waveform inversion of reflection seismic data for ocean temperature profiles. *Geophysical Research Letters*, 35(4), 2008. 1

