Application of randomized sampling schemes to curvelet-based sparsity-promoting seismic data recovery

Reza Shahidi[1] (rshahidi@gmail.com), Gang Tang[12] (gtangthu@gmail.com), Jianwei Ma[3] (jma@hit.edu.cn) and Felix J. Herrmann[1] (fherrmann@eos.ubc.ca)

[1]Seismic Laboratory for Imaging and Modeling, Department of Earth and Ocean Sciences, The University of British Columbia, 6339 Stores Road, Vancouver, BC, Canada, V6T 1Z4

[2]Research Institute of Petroleum Exploration and Development, PetroChina, Beijing, China

[3]Institute of Applied Mathematics, Harbin Institute of Technology, Harbin, China

ABSTRACT

Reconstruction of seismic data is routinely used to improve the quality and resolution of seismic data from incomplete acquired seismic recordings. Curvelet-based Recovery by Sparsity-promoting Inversion, adapted from the recently-developed theory of compressive sensing, is one such kind of reconstruction, especially good for recovery of undersampled seismic data. Like traditional Fourier-based methods, it performs best when used in conjunction with randomized subsampling, which converts aliases from the usual regular periodic subsampling into easy-to-eliminate noise. By virtue of its ability to control gap size, along with the random and irregular nature of its sampling pattern, jittered (sub)sampling is one proven method that has been used successfully for the determination of geophone positions along a seismic line. In this paper, we extend jittered sampling to two-dimensional acquisition design, a more difficult problem, with both underlying Cartesian, and hexagonal grids. We also study what we term separable and non-separable two-dimensional jittered samplings. We find hexagonal jittered sampling performs better than Cartesian jittered sampling, while fully non-separable jittered sampling performs better than separable jittered sampling. Two other 2D randomized sampling methods, Poisson Disk sampling and Farthest Point sampling, both known to possess blue-noise spectra, are also shown to perform well.

Keywords: curvelets, irregular subsampling, acquisition design, jittered sampling, blue noise

INTRODUCTION

Motivation

Seismic data, especially higher-dimensional data, is often extremely large, requiring a large number of shots for data acquisition and thus in turn high costs. Due to economical reasons it is desirable to reduce the number of geophones used to acquire seismic data. Seismic data also often has missing trace information due to complex acquisition conditions, so there is less information than required by the usual Nyquist criterion for perfect reconstruction of the complete data. Thus, reducing the amount of acquired data as well as effective reconstruction of complete seismic data from acquired incomplete seismic data are very important issues (Moldoveanu, 2010; Mosher et al., 2012). We are led to design sampling methodologies that minimize the necessary number of traces to be acquired while maintaining good quality reconstructions, by the two competing requirements of more information and less shots.

The Shannon/Nyquist sampling theorem states that for regular subsampling of data (the term subsampling is used since the full data is assumed to already be regularly sampled on a fine grid to make it discrete), it is necessary to sample at a frequency at or above the Nyquist frequency, twice the maximum frequency present in the data. Some methods to interpolate regularly-sampled data with an underlying plane-wave assumption, and without aliasing, exist in the literature, such as Spitz's f-x method ((Spitz, 1991) and Gulunay's f-k method (Gulunay and Chambers, 1997). However, these methods require geophones to be regularly spaced (Abma and Kabir, 2005), which is not always possible in practice. On the other hand, randomized (or nondeterministic) irregular subsampling is an alternative to regular subsampling. The irregularity is designed to allow for seismic data of a higher bandwidth to be acquired. A relatively new theory called "compressed sensing" (CS) (Candès et al.,

2006b; Candès, 2006; Donoho, 2006) provides a new paradigm for the reconstruction of signals of geophysical interest accurately from a number of samples far fewer than that dictated by the Nyquist rate.

Unfortunately, the sizes of unsampled gaps in the data are not controlled when sampling according to distributions such as the discrete uniform random distribution. This in turn can affect the quality of reconstruction with Curvelet-based Recovery by Sparsity-promoting Inversion (CRSI) greatly (Hennenfent and Herrmann, 2008) because there may be too much unsampled contiguous data, leading to excessively large gaps (larger than the support of the localized curvelet basis functions used for reconstruction) for which the curvelet coefficients cannot be determined or even constrained. Jittered sampling was introduced in the one-dimensional case to help resolve the issue of large unsampled sections of the seismic data in (Cook, 1986; Hennenfent and Herrmann, 2007, 2008), where gap size is limited, and also randomness maintained, as a result of which aliasing artifacts may be more easily removed. It is natural to extend it to two dimensions (László, 1995) to hope for similar improvements over uniform random sampling in the higher-dimensional case.

We find that jittered sampling is an example of a wider class of subsampling methods called "blue-noise" subsampling methods, all of which control maximum gap size in their subsampling configurations. We therefore not only use jittered sampling, but other blue-noise subsampling methods in order to design seismic acquisition surveys for improved reconstruction quality from fewer shots.

Blue-noise subsampling

A blue-noise signal refers to a signal whose energy is concentrated at high wavenumbers with little energy concentrated at lower non-zero wavenumbers. Sampling patterns with blue-noise spectra, common in the field of image processing, scatter aliasing artifacts

out of the signal band into high wavenumbers (Dippé and Wold, 1985; Ignjatovic and Bocko, 2005), to avoid interference with the baseband signal.

The benefit of irregular sampling methods may be more clearly expressed from a signal processing viewpoint by the duality of multiplication and convolution in the spatial and wavenumber domains. Given a two-dimensional array $\mathbf{F}$ representing the fully-sampled data on a regular grid, a subset of the data is selected with a subsampling function $\mathbf{S}$, which when multiplied elementwise with $\mathbf{F}$, gives an incomplete subset $\mathbf{\Sigma}$ of the full data. More succinctly, we may use the Hadamard product $\circ$ to express this subsampling as:

$$\mathbf{\Sigma} = \mathbf{S} \circ \mathbf{F} \tag{1}$$

In the wavenumber domain, this relation becomes

$$\hat{\mathbf{\Sigma}}(\mathbf{k}) = (\hat{\mathbf{S}} * \hat{\mathbf{F}})(\mathbf{k}), \tag{2}$$

where $*$ is the convolution operator, and the $\hat{}$ superscript indicates a function in the wavenumber domain. When the spatial subsampling function $\mathbf{S}$ regularly subsamples the complete fully-sampled data $\mathbf{F}$, as is the case in traditional seismic with regular spacing of geophones, full replicas of the complete data are obtained in the wavenumber domain. These replicas are themselves evenly spaced with spacing inversely proportional to the regular spacing in the spatial domain (Oppenheim et al., 1999). Thus, if the spatial sampling interval is too large (i.e. the signal is sampled below the Nyquist rate), then the support of these replicas in the wavenumber domain will overlap and there will be aliasing, leading to distortion in the reconstructed signal.

On the other hand, when the subsampling function $\mathbf{S}$ irregularly subsamples the fully-sampled data, the aliasing artifacts are no longer full replicas. With proper design of the subsampling function these replicas are replaced by incoherently distributed

artifacts in the spectrum (Xu et al., 2004), which are easier to remove using traditional noise removal techniques such as iterative soft-thresholding (Hennenfent and Herrmann, 2008). If we subsample with what is called a blue-noise based method, see e.g. (Dippé and Wold, 1985), these artifacts will be moved to a higher band, where they may be more easily muted out as they interfere less with the baseband wavenumber-domain signal $\hat{\mathbf{F}}(\mathbf{k})$ than with a white-noise based subsampling method, namely uniform random subsampling. More importantly, the gap size between neighboring samples is also controlled, avoiding major problems with amplitude recovery which occur when the spatial support of a curvelet is entirely unsampled.

Mathematically, it is possible to demonstrate that blue-noise subsampling methods do indeed lead to less aliasing than other (e.g. regular) subsampling techniques. We consider the one-dimensional case for simplicity - the proof for higher-dimensional cases is analogous. For a regular subsampling of the (discrete) data, the subsampling function $\mathbf{S}$ can be represented as:

$$\mathbf{S} = \sum_{l=-\infty}^{\infty} \delta(n - lT), \tag{3}$$

where δ is a Dirac delta function, and T is the (integer) sampling period. This function is designed to be periodic since the discrete Fourier transform is only defined for periodic functions, and the function to be samples, $\mathbf{F}$, is also extended to be periodic.

In the discrete-time frequency domain, $\hat{\mathbf{S}}(\omega)$ will be of a similar form, except that the spacing between impulses will be $\omega_s = \frac{2\pi}{T}$, and the impulses will be scaled by a factor of $\frac{2\pi}{T}$ (Oppenheim et al., 1999).

Now, if we choose an irregular subsampling of the data, the subsampling function

$\mathbf{S}$ can instead be represented as:

$$\mathbf{S}(n) = \sum_{l=-\infty}^{\infty} \sum_{m=0}^{M} \delta(n - l\bar{T} + t_m), \tag{4}$$

where the t_m's are integers in $[0, \bar{T})$, and M is the total number of (sub)samples. By the time-shifting property of the Fourier transform, the transform of $\mathbf{F}(n-p)$ for an arbitrary integer p is $\hat{\mathbf{F}}(\omega)e^{j\omega p}$, where $\hat{\mathbf{F}}(\omega)$ is the discrete Fourier transform of $\mathbf{F}(n)$. We may thus derive the discrete Fourier transform of $\mathbf{S}(n)$ to be:

$$\hat{\mathbf{S}}(\omega) = \sum_{m=0}^{M} e^{-j\omega t_m} \frac{2\pi}{\bar{T}} \sum_{l=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi l}{\bar{T}}\right) =$$

$$\frac{2\pi}{\bar{T}}\hat{\mathbf{B}}(\omega) \sum_{l=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi l}{\bar{T}}\right) \approx \frac{2\pi}{\bar{T}}\hat{\mathbf{B}}(\omega) \left(\delta(\omega) + \sum_{|l|\geq L>0} \delta\left(\omega - \frac{2\pi l}{\bar{T}}\right)\right)$$

Here $\hat{\mathbf{B}}$ is a blue-noise spectrum because it is the Fourier transform of the binary irregular subsampling function $\mathbf{S}$. It thus by definition has very little non-zero low-frequency energy. Additionally, because this blue-noise spetrum has little non-zero low-frequency energy, we can assume that the impulse functions for non-zero low frequencies may be removed from the infinite periodic sum of impulses, which is done above since L is an integer greater than 0.

So, similar to Equation 2,

$$\hat{\mathbf{\Sigma}}(\omega) = (\hat{\mathbf{S}} * \hat{\mathbf{F}})(\omega) \approx \frac{2\pi}{\bar{T}}\hat{\mathbf{B}}(\omega) \left(\delta(\omega) + \sum_{|l|>L>0} \delta\left(\omega - \frac{2\pi l}{\bar{T}}\right)\right) * \hat{\mathbf{F}}(\omega) =$$

$$\frac{2\pi}{\bar{T}}\hat{\mathbf{B}}(\omega) \left(\hat{\mathbf{F}}(\omega) + \sum_{|l|>L>0} \hat{\mathbf{F}}\left(\omega - \frac{2\pi l}{\bar{T}}\right)\right)$$

This means that even though the replicas now have a smaller spacing between them (i.e. $\frac{2\pi}{\bar{T}}$), where $[0, \bar{T})$ is the domain of the signal in discrete time, these replicas

will be muted out for low frequencies due to modulation with the blue-noise spectrum $\hat{\mathbf{B}}(\omega)$ of the underlying sampling impulse pattern. The higher that L is in the above equation, the more impulses are muted out. This means that in fact there will be fewer aliasing artifacts in the subsampled signal, and thus also fewer such artifacts in the reconstruction from this signal, highlighting the advantages of using blue-noise subsampling methods as in this paper.

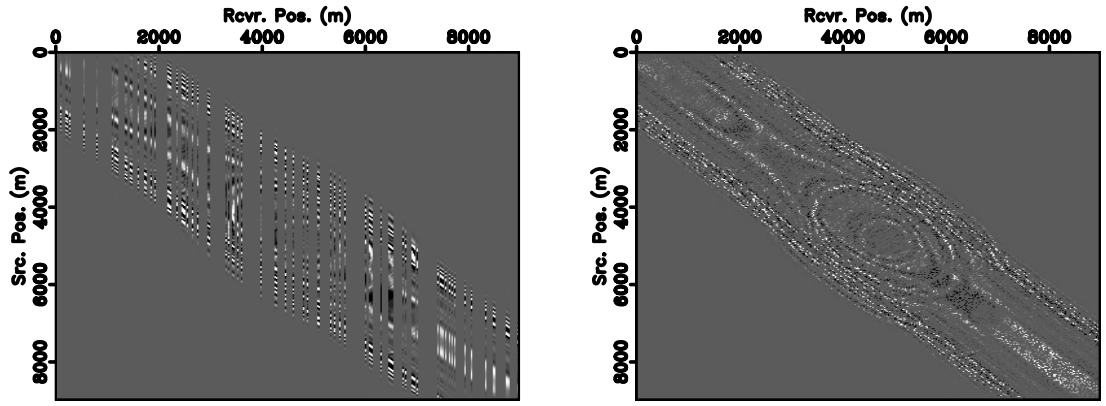
Additionally, in the context of time slices from three or higher-dimensional seismic data, higher frequencies correspond to events with steeper dips. These steeper dips come from slower returns, and thus may be muted out if they are known to come from returns slower than the minimum possible velocity for the medium.

CRSI and 2D irregular subsampling schemes

Curvelet-based Recovery by Sparsity-promoting Inversion (CRSI) (Herrmann and Hennenfent, 2008) is an example of the application of compressed sensing to the reconstruction of seismic data. It is derived from the discrete curvelet transform (Candès et al., 2006a) which sparsifies typical seismic data, in conjunction with a subsampling method that favours recovery. Curvelets have a wide variety of applications in many fields (Ma and Plonka, 2010), including seismic data processing, with their localized plane-wave-like frame elements and well-documented sparsity for seismic data wavefronts (Candès et al., 2006a; Herrmann et al., 2007; Chauris and Nguyen, 2008). We choose irregular and randomized (non-deterministic) subsampling for the subsampling method used for acquisition of seismic data reconstructed with CRSI.

In the context of seismic acquisition, subsampling design corresponds to the determination of geophone locations. In this paper, we focus on the placement of receivers, however it is easy to extend such methods to source placement by source-receiver reciprocity. Several one-dimensional subsampling methods have already been used in

conjunction with CRSI, e.g. 1D jittered sampling. In these one-dimensional subsampling techniques, one determines the possibly nonuniform placement of geophones along a line, as in Figure 1(a). However, there are only a few practical higher-dimensional sampling strategies for CRSI in the literature which can be used for seismic acquisition design. The focus of this paper is on 2D subsampling methods, as shown in Figure 1(b) which corresponds to a common shot gather in two dimensions and with reconstruction done across time slices.



(a) (b)

Figure 1: 25% random samples of time slice via: (a) 1D sampling and (b) 2D hexagonal jittered sampling

Regular subsampling has been traditionally used in practice, as it has been deemed easier to perform, for example in the field, where regular placement of sources and receivers has been considered easier and in the context of marine acquisition, where geophones are towed behind a boat along a straight line and shots fired at regular intervals as the boat travels along the inline direction(Vermeer, 1990). But performance is poor when regular subsampling is used to form the input to CRSI, because of the existence of periodic aliases (Zwartjes and Sacchi, 2007) both in the Fourier domain and equivalently in the sparsifying curvelet domain.

On the other hand, randomized subsampling can render high-amplitude coherent aliases into easy-to-remove incoherent noise in the wavenumber domain, so that CRSI

reconstruction may be achieved by a simple thresholding operation (Hennenfent and Herrmann, 2008). Typically, the inline direction is sampled more densely than the crossline direction (Long, 2004), so that sampling intervals along each axis may be different. There is also some freedom in positioning the cables behind the boat (in the case of marine acquisition), and this positioning can be done deliberately in an irregularly-spaced manner. As will be seen, introducing randomization leads to better recovery with CRSI from the incomplete acquired seismic data. The choice of coordinates is also important, since for example in two dimensions, common-midpoint/offset coordinates render regular sampling patterns in the source-receiver plane more irregular and thus better-suited for regularization using sparsity-promoting methods such as CRSI. However this choice of coordinate system is not a focus of the current research.

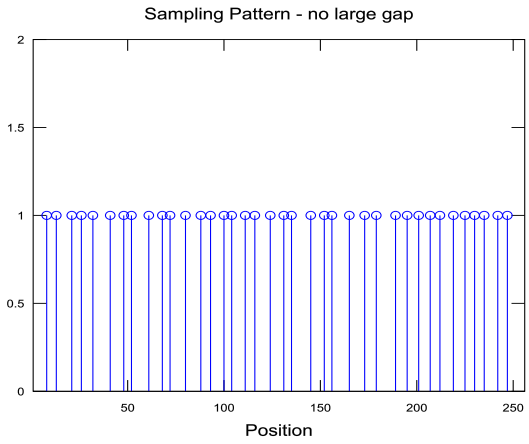
CRSI and blue-noise subsampling of seismic images

The underlying spectrum of the binary jittered sampling pattern, which consists of ones where samples are taken, and zeros elsewhere, has not been looked at in much depth in the context of jittered subsampling of seismic data. As stated in the work by (Grundland et al., 2009), jittered sampling has good blue-noise spectrum characteristics, meaning that spectral energy is concentrated at high frequencies with little energy concentrated at lower non-zero frequencies. Sampling patterns with blue-noise spectra, common in the fields of computer graphics and image processing, scatter aliasing artifacts out of the signal band into high frequencies (Dippé and Wold, 1985; Ignjatovic and Bocko, 2005).

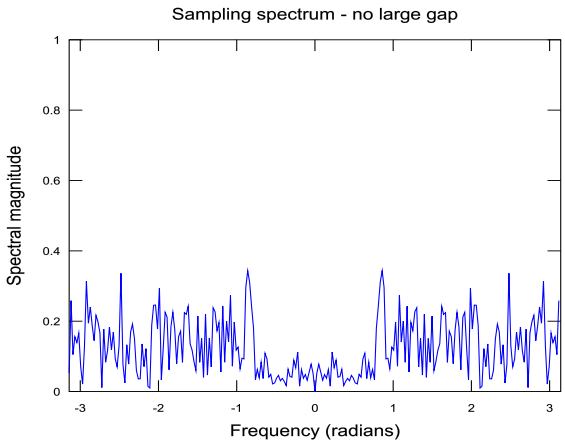
A significant property of blue-noise sampling patterns in the context of acquisition design is the limitation they place on the spacing between geophones. Because of the way that the sampling pattern for a certain placement of geophones is defined, with

ones indicating geophone positions, geophones cannot be placed too far apart with a blue-noise sampling pattern. Otherwise, there would be a large region in the sampling pattern consisting exclusively of zeros, leading to significant low-wavenumber energy, in violation of the definition of a blue-noise spectrum. Therefore, there is a natural limitation on any geophone not being placed too far away from the other geophones in a blue-noise sampling pattern.

This is illustrated in more depth in Figures 2 and 3. In Figure 2(a), we see a one-dimensional binary sampling pattern where sampling gap sizes are neither too small nor too large , and in Figure 2(b) we see its normalized spectrum (with DC component zeroed out), where there is not much low-wavenumber spectral energy. On the other hand, in Figure 3(a), the same sampling pattern as in Figure 2(a) is depicted, except now several samples near the centre of the sampling domain have been removed. Figure 3(b) shows the spectrum of this sampling pattern, where there is more low-wavenumber energy than in the spectrum for the original sampling pattern, e.g. the two peaks circled in red in that figure. The spectrum no longer possesses the characteristic of blue noise. We find from this simple example imposing a blue-noise condition on the sampling spectrum will necessarily impose limits on the sizes of unsampled regions, i.e. distances between neighboring geophones in a seismic survey.

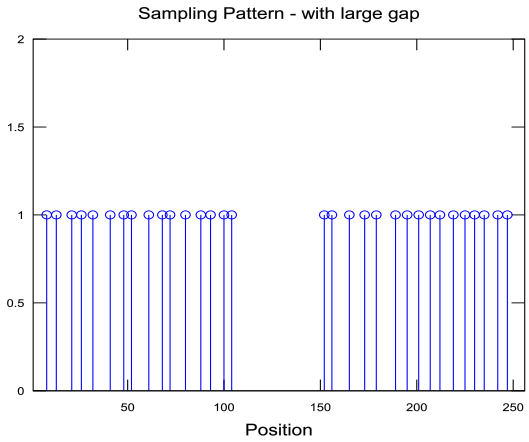


(a)

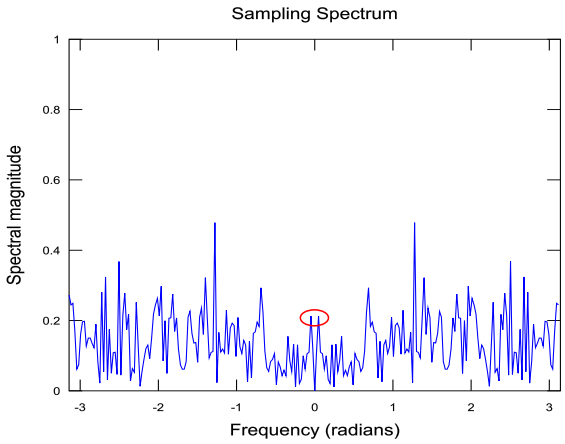


(b)

Figure 2: (a) a sampling pattern and (b) its blue-noise spectrum



(a)



(b)

Figure 3: (a) the same sampling pattern as in Figure 2(a) with a large number of consecutive samples removed and (b) its spectrum which now has substantial low-wavenumber energy (peaks encircled in red).

Therefore, sampling patterns other than jittered sampling, but which also possess blue-noise spectra, are worth exploring in the context of seismic acquisition design, and also in conjunction with CRSI reconstruction, because of their ability to leave no large sections of the survey unsampled. Another important consideration is the capability for progressive sampling, where shots can be taken one at a time until the reconstructed result is of acceptable quality. Jittered sampling does not offer this capability up-front since it is not clear in which order the jittered samples should be taken. On the other hand, two other blue-noise subsampling methods, Poisson Disk sampling (Cook, 1986) and Farthest Point sampling (Eldar et al., 1997), are both progressive subsampling methods, and are easily described algorithmically. Those two methods are widely used in the image processing and computer graphics literature and are known to possess blue-noise spectra (Mitchell, 1987).

In the field of magnetic resonance imaging (Lustig et al., 2009) (MRI), Poisson Disk sampling provides a locally uniform distribution for compressive sensing. However in that context, k-space or frequency space is sampled, and not the spatial domain. In this paper, we test all of these irregular and randomized sampling schemes to generate subsampled seismic data as inputs to seismic reconstruction by CRSI. Though not tested for this paper, the recent approach of coil-sampling of Moldoveanu (Moldoveanu, 2010) is another method based on jittered sampling where multiple vessels perform a seismic survey along coils with randomized centers.

Finally, the use of Ocean Bottom Seismic (OBS) nodes is an exciting recent development in marine seismic acquisition where acoustic seismic sensors are placed directly on the ocean floor. The use of such nodes leads to improved seismic imaging, with better reservoir characterisation and anisotropic velocity-model building (Olofsson et al., 2012). There is no reason why the acquisition design methods used in this paper cannot be used in conjunction with these nodes to improve seismic image quality even further.

Main Contributions

The main contributions of this paper are the application and analysis of blue-noise subsampling methods to the problem of two-dimensional seismic acquisition design. By virtue of their blue-noise characteristics, such methods are able to control the sizes of unsampled regions in a seismic survey. They are thus especially suited to seismic data reconstruction with localized basis functions, such as curvelets, the support of which must include some sampled data. Jittered, Poisson Disk and Farthest-Point subsampling methods all have blue-noise characteristics and are found to give similar results in terms of the quality of the reconstructed seismic data volumes. All of these methods improve upon regular subsampling when used in conjunction with CRSI reconstruction, by virtue of their aperiodicity and randomness, Although these conclusions are obtained with CRSI, other transform-based recovery methods based on Fourier or wavelet bases may also be used in conjunction with these sampling methods, though this is not the focus of this paper.

CURVELET-BASED RECONSTRUCTION BY SPARSITY-PROMOTING INVERSION

Restriction operator

Two-dimensional Curvelet-based Recovery by Sparsity-Promoting Inversion (CRSI) reconstruction is a sparse optimization problem. Although the discussion here is limited to CRSI in two dimensions, it is relatively straightforward to generalize the method to higher dimensions. Defining vec to be a vectorization operator on a 2D data matrix, the reconstruction from an incomplete 2D seismic dataset follows the forward model

$$\mathbf{b} = \mathbf{R}\text{vec}(\mathbf{F}), \tag{5}$$

where $\mathbf{b} \in \mathbb{R}^N$ represents the vector of acquired incomplete 2D data with missing traces. Here $\mathbf{F} \in \mathbb{R}^{N_s \times N_r}$ is the to-be-recovered 2D signal, on a regular unaliased output grid, i.e. the adequately-sampled data. Here $N_s \times N_r$ is the number of samples in the full seismic data, so for example on a 2D time slice, N_s is the number of samples on the source axis and N_r the number of samples along the receiver axis. Furthermore, $\mathbf{R} \in \mathbb{B}^{N \times N_s N_r}$ with $\mathbb{B} = \{0, 1\}$, is a restriction operator that collects N acquired samples from the full seismic data, $\mathbf{F}$, $N_s N_r \gg N$, Thus $\mathbf{R}$ is a binary sampling matrix, on which both acquired data $\mathbf{b}$, and recovery of the model $\mathbf{F}$ depend, and each row of which consists entirely of zeros, except for a single 1 corresponding to the sample being taken.

If the restriction operator $\mathbf{R}$ can be expressed as the Kronecker product of two matrices $\mathbf{R}_s \in \mathbb{R}^{N_1 \times N_s}$ and $\mathbf{R}_r \in \mathbb{R}^{N_2 \times N_r}$, with $N_1 \cdot N_2 = N$, i.e., $\mathbf{R} = \mathbf{R}_s \otimes \mathbf{R}_r$, then we call the sampling associated with $\mathbf{R}$ *separable*. Otherwise, we call the sampling method *non-separable*. If only certain traces are sampled, but each trace consists of complete data, then $\mathbf{R} = \mathbf{R}_t \otimes \mathbf{I}_m$, with $\mathbf{R}_t$ the restriction to the subsampled traces, and $\mathbf{I}_m$ the mxm identity matrix. The matrix $\mathbf{R}_t$ is a restriction operator corresponding to a 1D subsampling of traces. This matrix can thus use any 1D subsampling scheme, e.g. $\mathbf{R}_r$ for 1D regular subsampling, $\mathbf{R}_{j1}$ for 1D jittered subsampling or $\mathbf{R}_u$ for 1D uniform random subsampling. It is also possible to do complete 2D subsampling in which case $\mathbf{R}$ will not be expressible as a Kronecker product in a non-trivial way, e.g. $\mathbf{R}_{j2}$ for fully two-dimensional jittered sampling. Several choices of two-dimensional binary restriction matrices $\mathbf{R}$ in the source-receiver plane are shown in Figure 4, with blue dots representing the ones in the matrix. Subsequently, Figure 5 shows the wavenumber spectra of all of these sampling patterns. Horizontal and/or vertical gridding artifacts are evident in the spectra when Kronecker or outer products are taken. In these cases there is a concentration of horizontal and/or vertical wavenumber energy in the sampling pattern due to its lower-rank nature.

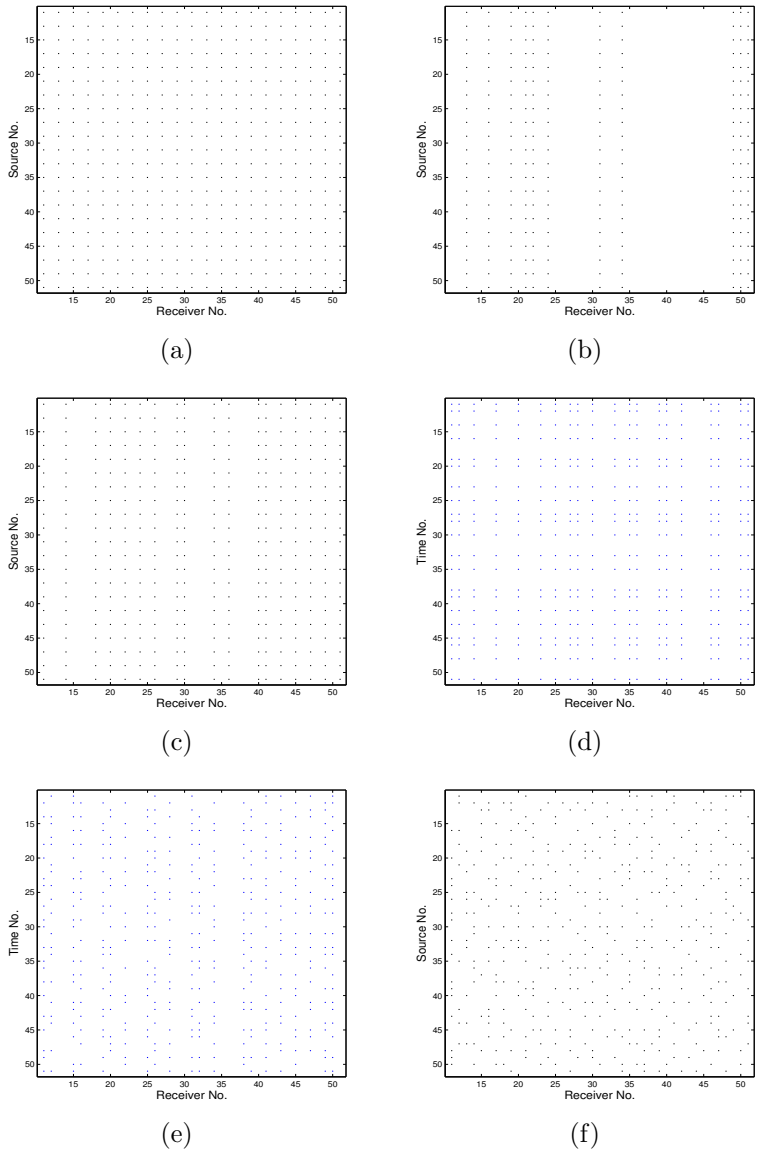


Figure 4: Different types of sampling: (a) regular, (b) uniform random along one axis and regularly sampled along other, (c) jittered sampling along one axis and regularly sampled along other, (d) 2D jittered sampling, with jittered sample positions along one axis, and positions determined by another jittered sampling pattern on other axis, (e) 2D jittered sampling, jittered positions along receiver axis, and different jittered positions along each cable inline (non-separable), (f) Fully 2D jittered sampling.

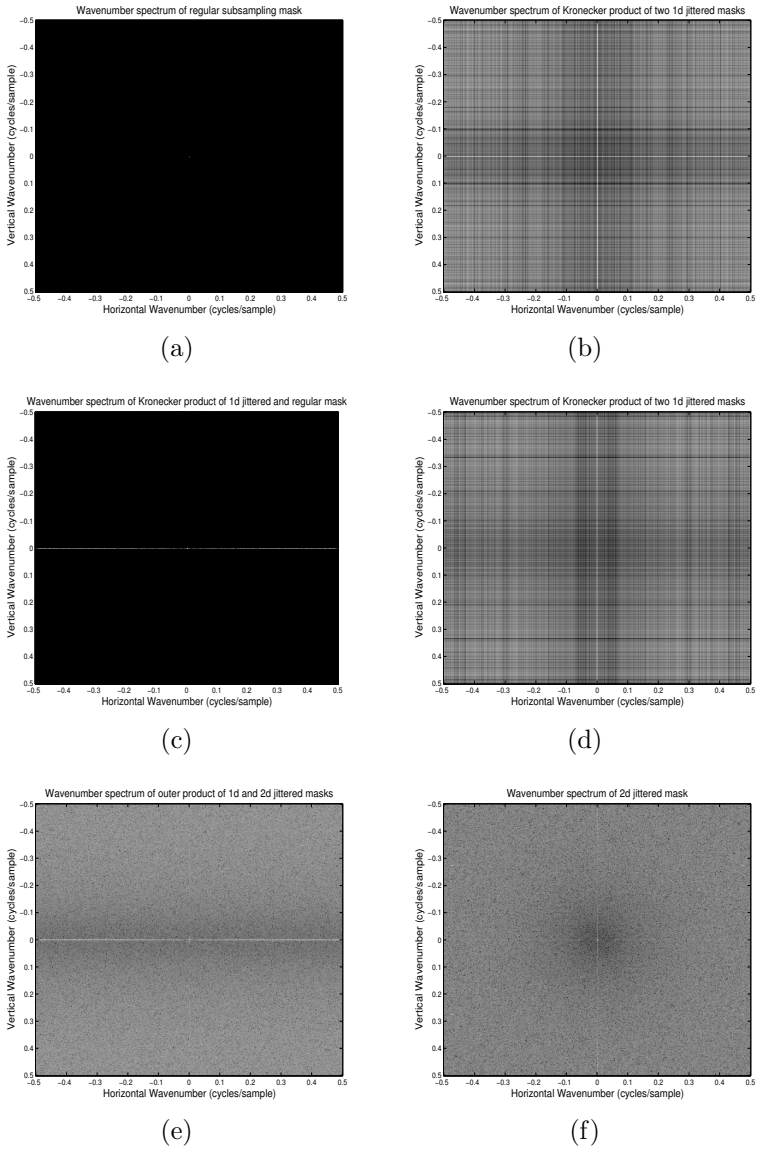


Figure 5: Wavenumber spectra from different sampling patterns: (a) regular, (b) uniform random along one axis and regularly sampled along other, (c) jittered sampling along one axis and regularly sampled along other, (d) 2D jittered sampling, with jittered sample positions along one axis, and positions determined by another jittered sampling pattern on other axis, (e) 2D jittered sampling, jittered positions along receiver axis, and different jittered positions along each cable inline (non-separable), (f) Fully 2D jittered sampling.

In Figure 4(b), the positions of the cables on the boat are chosen according to a discrete uniform random distribution. In Figure 4(c), an example is shown of where geophone positions (in the context of land surveys) or cable positions (for marine surveys) follow a 1-D jittered distribution.

In Figure 4(d), both horizontal and vertical coordinates of geophone positions (for land surveys) or cable positions and positions of shots (for marine surveys) are jittered with the same distribution. Both jittered distributions can also be independent of each other, as illustrated in Figure 4(e). Figure 4(f) shows fully two-dimensional jittered sampling for positioning of geophones, where positions are not determined by the Kronecker product of two one-dimensional jittered distributions, but by a jittered distribution from a two-dimensional rectangular tiling of the acquisition plane.

CRSI method

Equation 5 is an underdetermined inverse problem, because any model that agrees with incomplete data $\mathbf{b}$ after being restricted by matrix $\mathbf{R}$ will be a solution. According to compressive sampling theory, it has been suggested to reformulate the problem as follows:

$$\mathbf{b} = \mathbf{A}\mathbf{x} \quad \text{with} \quad \mathbf{A} \stackrel{\text{def}}{=} \mathbf{R}\mathbf{C}^H, \tag{6}$$

where $\mathbf{C}$ is a sparsifying transform (e.g. the discrete curvelet transform for seismic data), $\mathbf{C}^H$ is its adjoint, and $\mathbf{x} \in \mathbb{R}^P$ with $P \gg N$ is the (redundant) representation of $\mathbf{s}$ in the sparsifying domain, which by virtue of being the representation of the model $\mathbf{b}$ in the sparsifying domain will contain many zero elements. In (Herrmann and Hennenfent, 2008), the discrete curvelet transform is adopted as the sparsifying transform for seismic data.

As stated in (Herrmann and Hennenfent, 2008), as opposed to regular subsampling,

randomized subsampling will change Equation 6 to

$$\mathbf{A}^H(b) = \mathbf{A}^H\mathbf{A}\mathbf{x} \approx \mathbf{x} + \mathbf{n}, \tag{7}$$

where the spectral leakage is approximated by additive white Gaussian noise $\mathbf{n}$. Thus the recovery problem turns into a much simpler "denoising" problem. So a randomized subsampling would be preferred if it can make the off-diagonal energy in $\mathbf{A}^H\mathbf{A}$ spread out as evenly as possible across each row's off-diagonal elements.

The curvelet transform gives a compressible representation of $\mathbf{s}$. In other words, the vector $\mathbf{x}$ has few large and many small non-zero coefficients. These properties make it possible to successfully recover $\mathbf{s}$ according to the theory of *compressive sampling* (Candès et al., 2006b; Candès, 2006; Donoho, 2006). The CRSI method promotes sparsity in the curvelet domain as a regularization term and gives a solution to 6 by solving the optimization problem:

$$\mathbf{P}: \quad \begin{cases} \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{R}\mathbf{C}^H\mathbf{x} = \mathbf{b}, \\ \widetilde{\mathbf{f}} = \mathbf{C}^H\widetilde{\mathbf{x}}, \end{cases}$$

where $\|\mathbf{x}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^{N} |x_i|$ is the ℓ_1 norm. The recovered vector that solves $\mathbf{P}$ is $\widetilde{\mathbf{x}}$ and $\widetilde{\mathbf{f}}$ is the estimate of the recovered data obtained by applying $\mathbf{C}^H$.

RANDOMIZED SUBSAMPLING METHODS

According to sampling theory, aliasing occurs at frequencies higher than the Nyquist limit for a signal when it is regularly subsampled. If we subsample in an irregular manner to make the sizes of unsampled regions unequal, it is possible for these aliases to be reduced in amplitude so that they can be more easily removed. Randomized subsampling is a way of achieving this, by distributing the subsample positions

randomly, so that every discrete position has a finite probability of being sampled (Dippé and Wold, 1985). But it is not enough to simply employ discrete uniform random subsampling, where each location on a grid has exactly the same probability of being chosen, since this leaves the possibility of the size of gaps between sample positions being too large so that some curvelet coefficients become unconstrained and can take on any value.

As already explained, jittered random (sub)sampling is a better choice than uniform random subsampling. Because each region is sampled, the size of gaps can be controlled, and a blue-noise spectrum is obtained. (Hennenfent and Herrmann, 2008) introduced 1D jittered sampling in conjunction with CRSI and proved its effectiveness. In the current paper jittered sampling is extended to two dimensions.

Generalization of jittered (sub)sampling to 2D

Randomised sampling in dimensions higher than one is a challenging problem, in large part due to the "curse of dimensionality" (Bengtsson et al., 2008), where the number of sampling configurations is exponential in the number of dimensions being sampled. For one-dimensional jittered sampling, all positions are located along one line, and so there is no more than one way to subdivide the corresponding axis into cells.

But in two dimensions, there are more choices for the underlying tiling of the field, two examples being Cartesian and hexagonal tilings. A Cartesian (unjittered) tiling simply consists of the usual regularly spaced rectangular tiling, while a hexagonal (unjittered) tiling simply consists of the usual regularly spaced hexagonal tiling. As well, the sampling method can result from the Kronecker product of two one-dimensional samplings, in which case we call the sampling method *separable*. Both the Cartesian and hexagonal jittered samplings are non-separable subsampling methods since they cannot be expressed as the Kronecker product of two one-dimensional subsamplings.

Later in the text, we sometimes refer to such subsamplings as fully two-dimensional jittered samplings since they cannot be reduced down to simpler 1D subsamplings. Jittered sampling with hexagonal tiling is described in more detail since it was found to be a particularly good randomized sampling method.

Comparisons between Cartesian and hexagonal jittered sampling

Regular hexagonal grids require 13.4% fewer samples than Cartesian grids to reconstruct the same image information from the sampled data in two dimensions (Mersereau, 1979). In 3D, this decrease in sampling density becomes 29.3% Theußl et al. (2001) In general, this implies that larger bandwidth signals can be subsampled with hexagonal grids than with Cartesian grids using the same number of samples.

An example of a regular hexagonal sampling grid is shown in Figure 6. In that figure, k and l are integers, and T_x and T_y are the horizontal and vertical sampling periods respectively. The hexagonal sampling pattern can be thought of as the union of two Cartesian samplings staggered with respect to one another by half a sample.

For 2D jittered sampling, the field to be sampled is tiled with Cartesian or hexagonal grids, then a random perturbation is made around the center of each tile while making sure the sample is kept inside its tile, after which a quantization operation is performed to round the sample position to the nearest integer grid point defining the full data. This quantization operation becomes more important when the subsampling ratio is small, i.e. when many subsamples are taken. Cartesian (rectangular) and hexagonal jittered strategies are shown in Figures 7(a) and 7(b) respectively. The subsampling depends on the tiling itself because of the restriction that random perturbations keep a sample within a tile or a subset thereof. For example, if the underlying tiling consists of very small tiles, then by necessity, the spacing between neighboring subsamples will also be small. The amounts of theoretical aliasing from both Cartesian and hexagonal jittered sampling are compared, as shown in Figure 8, as calculated by the formulae

in the appendix of this paper. This theoretical aliasing is defined as the ratio of the magnitude of the largest non-zero spike in the frequency spectrum to the spectral magnitude at zero.

This aliasing in the expected power spectrum is plotted as a function of the perturbation ratio, defined as the ratio of the maximum allowed jitter ζ to the cell radius γ. In this case, since the inherent packing densities of Cartesian jittered and hexagonal jittered sampling are different, the cell sizes must be chosen so that the same number of samples is taken for each method.

Figure 9 illustrates examples of the quantities ζ and γ as they relate to both Cartesian and hexagonal jittered samplings.

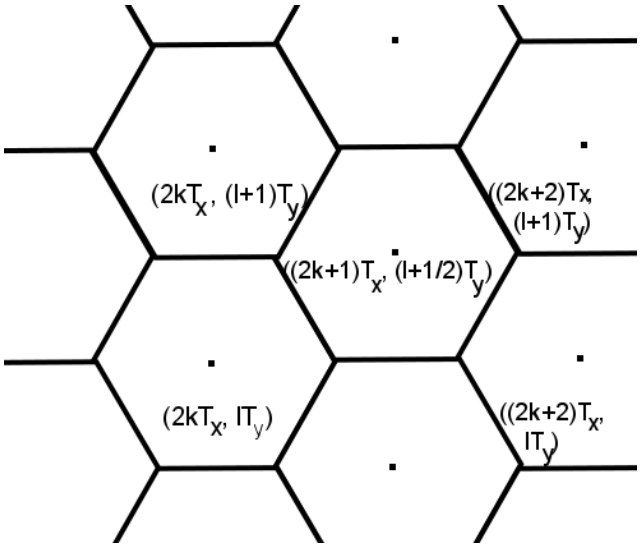
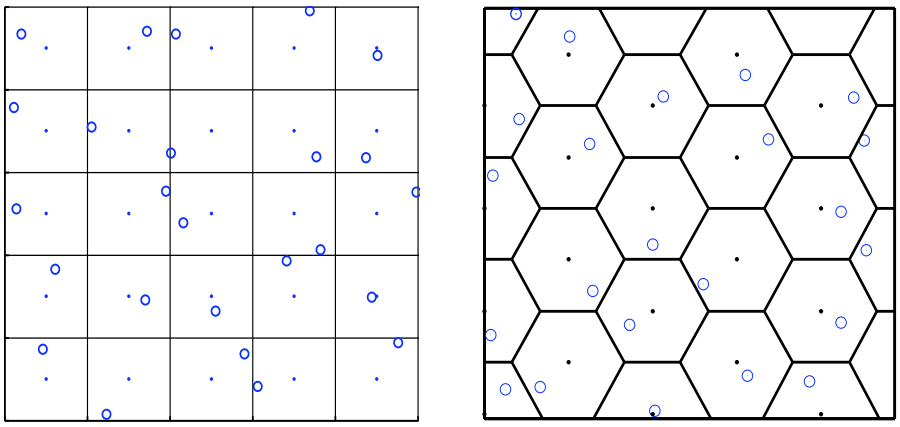


Figure 6: Regular hexagonal sampling grid

Jittered hexagonal undersampling ($0 < \zeta < \gamma$)

When $\zeta = 0$, there is no jitter, corresponding to regular hexagonal undersampling, with full-amplitude aliasing replicas. As the value of ζ increases, the peak aliasing magnitudes decrease. In Figure 8, the expressions derived in the appendix are used to plot the amount of aliasing for the expected spectra of hexagonal and Cartesian jittered sampling vs. the perturbation ratio $\beta = \frac{\zeta}{\gamma}$ as β ranges from 0 to 1. Note that



(a) (b)

Figure 7: Diagrams of sampling pattern: (a) Cartesian jittered sampling and (b) hexagonal jittered sampling. Solid dots are regular grid centres and circles are sample positions.

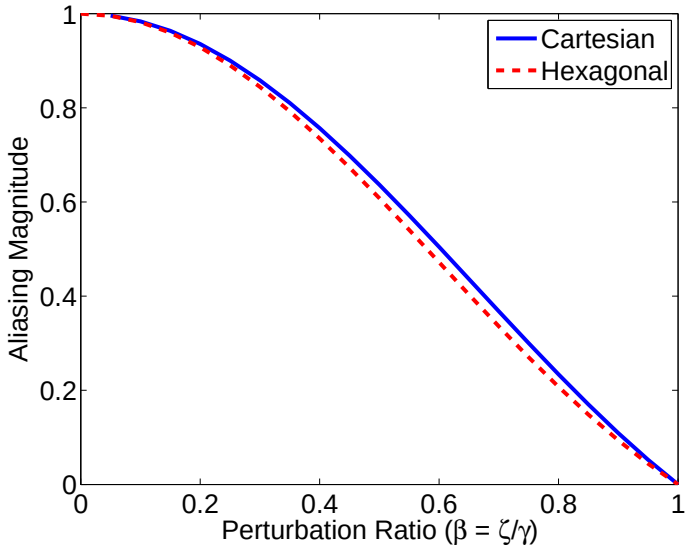


Figure 8: Comparison of the amount of theoretical aliasing for hexagonal versus Cartesian jittered sampling.

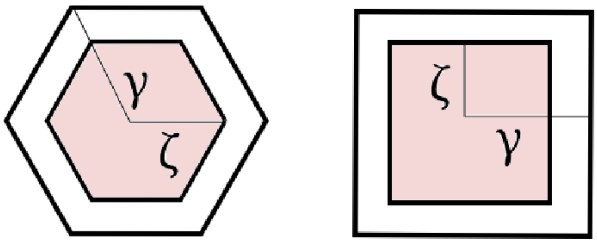


Figure 9: Illustration of the maximum allowed jitter ζ and cell size γ for Cartesian (left) and hexagonal (right) jittered samplings.

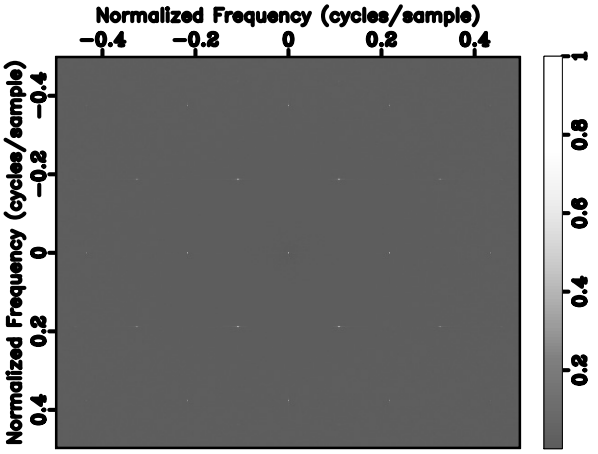
for optimal (i.e. with $\zeta = \gamma$) jittered sampling the amount of aliasing for the expected spectrum is zero, however for any typical instance of optimal jittered sampling, the actual amount of aliasing will be non-zero due to the variance of the random jittered sampling spectrum.

Optimally-jittered hexagonal undersampling ($\zeta = \gamma$)

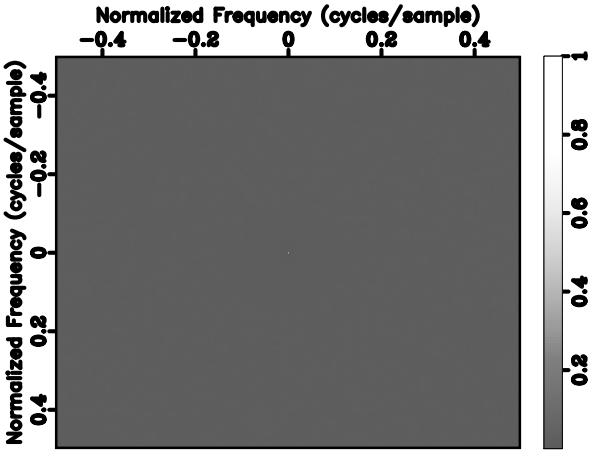
As stated in the previous section, for optimal jittered sampling, when the amount of jitter for each grid cell is maximal, it can be easily shown that the coefficients of all the impulses in Equation 18 as derived in the Appendix, are zero except the impulse corresponding to the DC component at the origin, with coefficient 1.

In other words, ignoring gridding effects, the expected value of the Gram matrix $\mathbf{G} \stackrel{\text{def}}{=} \mathbf{A}^H\mathbf{A}$ is the identity matrix. However there will still be some aliasing since the variances of the entries of this matrix are non-zero. These variances (ignoring gridding effects) are given by the expression $\frac{8}{3\sqrt{3}\gamma^2}(1 - |\psi_Z(\mathbf{f})|^2)$, as derived in the appendix of this paper. The variances are not dependent on the jitter ζ, but are dependent on the wavenumber vector $\mathbf{f} = (f_1, f_2)$. For the zero-wavenumber, the variance of the spectrum is 0, meaning that the base signal replica will always be reproduced exactly without being scaled.

The spectra of two jittered sampling patterns are shown in Figure 10. Figure 10(a)



(a)



(b)

Figure 10: Example spectrum for hexagonal jittered sampling with: (a) $\zeta = 0.3\gamma$ and (b) $\zeta = \gamma$

shows an example spectrum for suboptimal jittered sampling. The sampling spectrum for an instance of optimal hexagonal jittered sampling is shown in Figure 10(b), where there are no spiky aliases, but just wideband noise of low magnitude. Observe that both spectra exhibit blue noise properties with little low-wavenumber energy.

Alternative randomized subsampling methods

As stated in the recent work by (Grundland et al., 2009), jittered sampling attempts to replicate a blue-noise spectrum, yielding good reconstruction results. There are other sampling methods that have spectra which are more "blue", in the sense that there is more of a clear threshold wavenumber below which the spectrum magnitude is insignificant (except for the zero wavenumber), and above which there is more substantial spectral energy. Two such methods explored in this paper are Poisson Disk sampling and Farthest Point sampling due to their progressive nature allowing for the subsampling to be stopped when the reconstructions are of sufficiently good quality.

Poisson Disk sampling selects n points sequentially at random, and only keeps a sample if it is at a sufficient distance away from all previously selected samples. This distance can be slowly shrunk as the number of samples, and hence the sampling density, increases, so that the number of samples taken is not limited to be too small. This also leads to a blue noise spectrum, as Figures 11(a) and 11(c) show.

Farthest Point sampling is based on the computational geometry concept of the Voronoi diagram (Eldar et al., 1997). Farthest Point sampling repeatedly places the next sample point to be the farthest point from all previously selected samples. This ensures that the image is adequately sampled, while maintaining some randomness by initially selecting a small number of randomly placed seed samples. It can be proven that the farthest point from the previous samples is a vertex of the Voronoi diagram of those samples. So points are added one at a time to the sample set, chosen from among the Voronoi diagram vertices, and the Voronoi diagram updated incrementally. Please see Figures 11(b) and 11(d).

Both Poisson Disk and Farthest Point sampling methods lead to sampling patterns which do not leave arbitrarily large unsampled image regions. Therefore, these methods

are especially suitable for a reconstruction method like CRSI, where the basis consists of curvelets localized in space.

Instead of pure Farthest Point sampling, a discrete approximation is used to simplify the generation of sample points. A series of two-dimensional array $\mathbf{Q}_i$ with dimensions equal to that of the sampling grid is used, with elements equal to the squared distance of each grid point to the closest previously selected sample. The i here is the iteration number and these arrays are maintained to help in the generation of subsequent sample points. At each iteration, the coordinates of the array element with maximum $\mathbf{Q}$ value are chosen as the coordinates of the next sample, i.e. (x_{n+1}, y_{n+1}). Then the distance map is updated:

$$\mathbf{Q}_{n+1}(i,j) = \min(\mathbf{Q}_n(i,j), (i - x_{n+1})^2 + (j - y_{n+1})^2),$$

for all array positions (i, j) in the image. If there is more than one maximal array element, then one is chosen at random. If the discrete grid is fine enough, this results in a good approximation to the continuous version of Farthest Point sampling, without dealing with the intricacies of computational geometry constructs such as the Voronoi diagram.

RESULTS AND DISCUSSION

In this section, experiments are conducted on a time slice from a 2D prestack dataset in source/receiver coordinates in Figure 12. For the curvelet parameters used for CRSI, 6 scales and 16 angles at the coarsest scale were used and 64 angles were used at the finest scale. The seismic data was subsampled at a rate of 25% of the full discrete data. The signal-to-noise ratio (SNR) is defined as $20 \log_{10} \frac{\|\mathbf{s}_0\|_2}{\|\mathbf{s}-\mathbf{s}_0\|_2}$, where $\mathbf{s}_0$ is the original data, and $\mathbf{s}$ is the interpolated data.

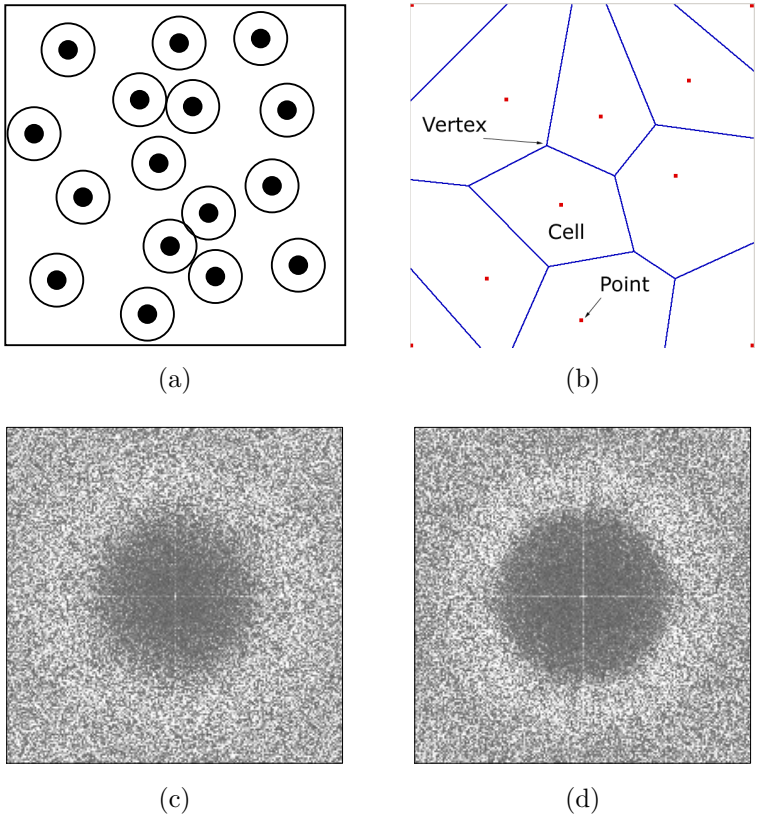


Figure 11: Sampling schemes of (a) Poisson Disk (b) Farthest Point, (c) is a typical Poisson Disk sampling spectrum, and (d) is a typical Farthest Point sampling spectrum

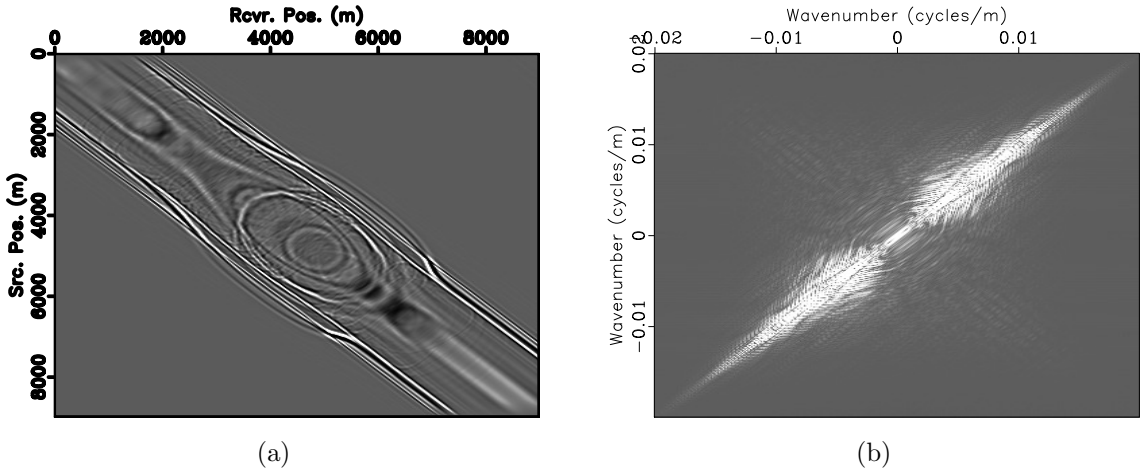
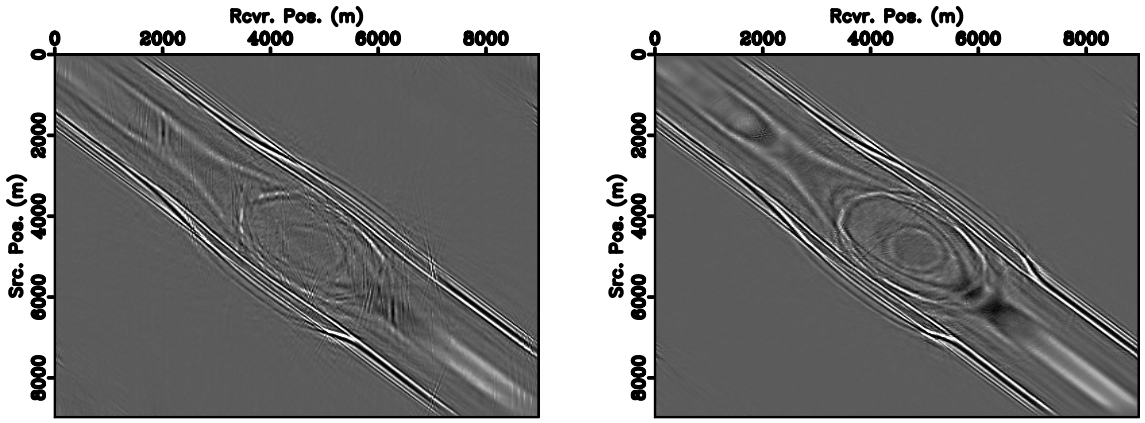


Figure 12: (a) Reference synthetic seismic data and (b) Fourier spectrum of (a)

First the reconstruction results from 1D uniform random subsampling versus fully two-dimensional random subsampling are compared. It is expected that taking samples with a fully 2D uniform random distribution would lead to better results because the two dimensional nature of the data is taken into account. In Figure 13 it is seen that this is indeed the case. The reconstruction in Figure 13(b) is free of the many artifacts which are clearly visible in Figure 13(a). The signal-to-noise ratio of the result in Figure 13(b) is also $\sim$ 3 dB higher than that of the Figure 13(a) result.



(a) (b)

Figure 13: (a) Reconstruction from 1D samples in Figure 1(a) (SNR=6.77 dB) and (b) Reconstruction from 2D samples in Figure 1(b) (SNR=9.75 dB)

To compare 2D hexagonal jittered undersampling with discrete uniform random undersampling, a synthetic time slice is chosen for the experiments. As the sampling rate varies from 10% to 50%, the SNR values are recorded, as shown in Figure 14 for different undersampling rates. It is found that 2D jittered sampling gives higher SNRs than discrete uniform random sampling for all sampling rates. For example, for a subsampling rate of 50%, the uplift of jittered sampling over discrete uniform random sampling is approximatly 2 dB. As expected, from Figure 14, it is seen that as the subsampling rate increases from 10% to 50%, the SNR of the reconstruction increases. What should be also noted is that because the CRSI reconstruction scheme is based on compressed sensing, the sampling rate necessary for perfect reconstruction

is not related to the maximum wavenumber of the underlying wavefield, as is the case with Nyquist sampling, but instead on the wavefield's complexity. This leads to a further reduction in the number of samples necessary to achieve a given SNR for less complicated wavefields. More information on this observation may be found in (Herrmann et al., 2009), and also later in this paper.

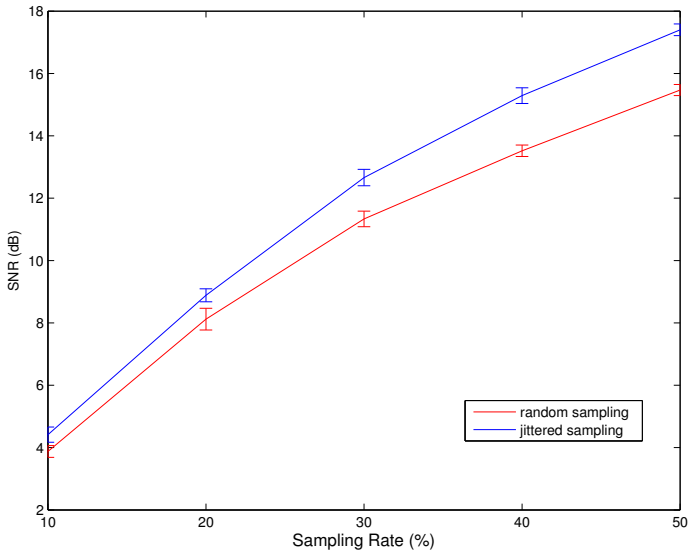


Figure 14: SNR of reconstruction from 2D hexagonal jittered undersampling and discrete uniform random undersampling, for different sampling rates

The spectra of the various sampling methods discussed in this paper are shown in Figure 15. In Figure 15(a), there are full copies of the original spectrum in Figure 12 of the same magnitude as the original spectrum. In Figure 15(b), where discrete random sampling is introduced along one axis, these aliases are less pronounced, and also it can be seen along the zero source wavenumber that there is dimming in the spectrum, due to the jittered sampling taken in the crossline direction. The spectrum in Figure 15(c) is for a sampling similar to that in (b), except that instead of discrete uniform random sampling, jittered sampling is used in the crossline direction. This leads to a slight improvement in the sampling spectrum, with the incoherent noise less obvious. In Figure 15(d), each direction is sampled with a 1D jittered distribution,

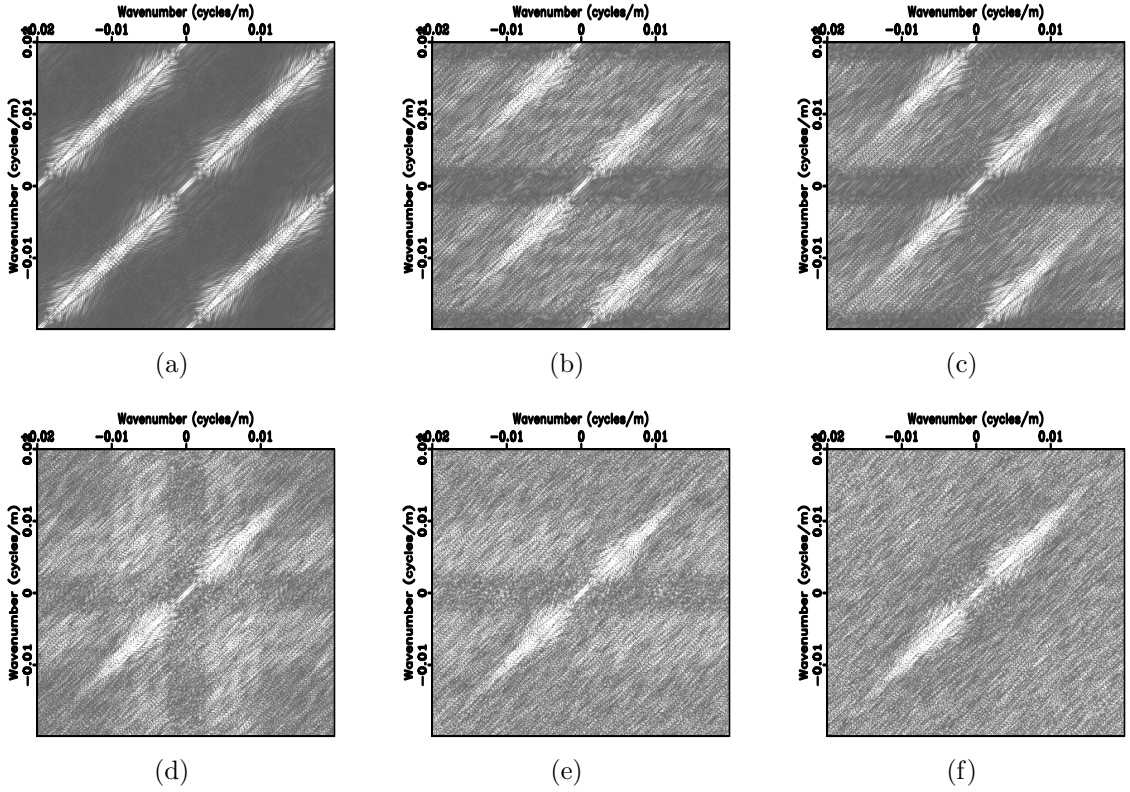


Figure 15: Spectra of data after sampled with different 2D sampling methods: (a) regular sampling, (b) regular along source axis, discrete uniform random along receiver axis, (c) regular along source axis, jittered along receiver axis, (d) 2D jittered sampling, jittered along receiver and source axes, same jittered positions for each receiver (separable), (e) 2D jittered sampling, jittered along receiver and source axes, different jittered positions for each receiver (non-separable), (f) fully 2D jittered sampling (non-separable)

each one possibly being a different instance of such a distribution. The overall 2-D sampling pattern is expressible as the Kronecker product of two one-dimensional jittered sampling distributions. The aliases are much less visible than in Figures (b) and (c), due to the fact that regular samples are not taken along any direction. Also, there is dimming of the spectrum, however this time along both the zero source and receiver frequencies. The spectrum in Figure 15(e) corresponds to a jittered sampling pattern as in Figure **??**, where now different jittered sampling positions are taken for each receiver position. This is again better than the previous spectra, with very little aliasing or noise present. Finally, in Figure 15(f), it is seen the sampling spectrum

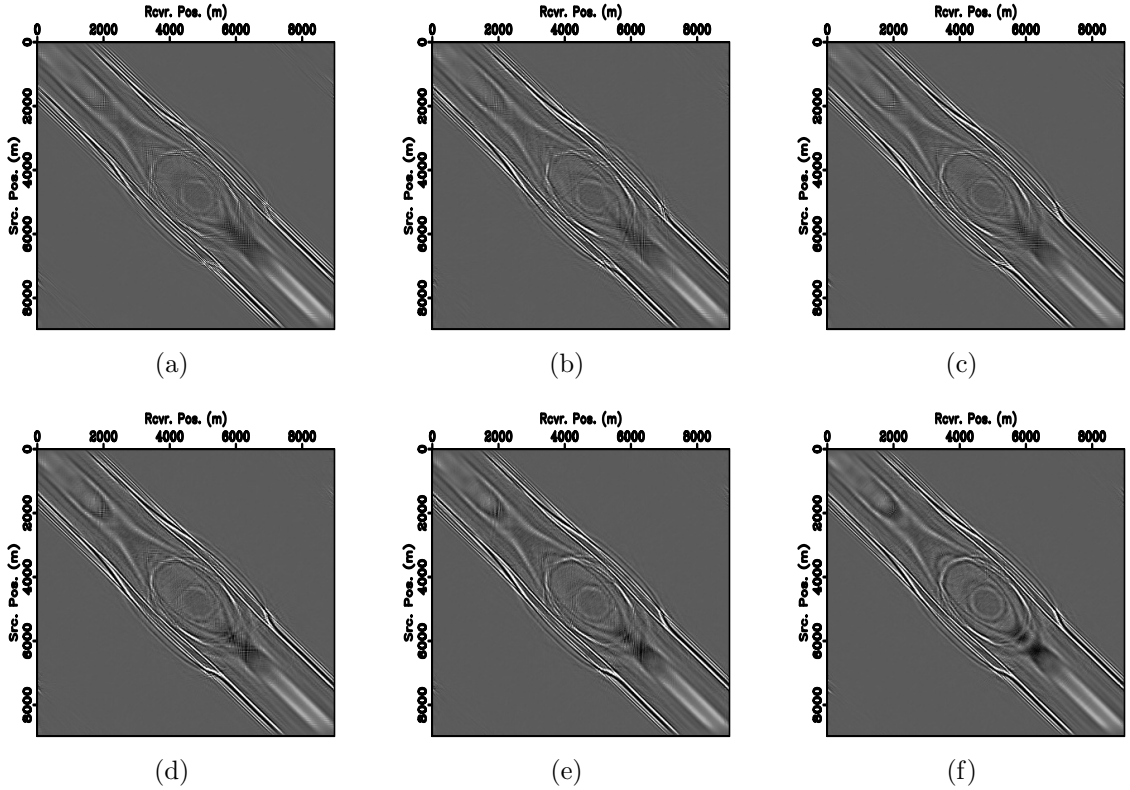


Figure 16: Reconstructions from 75% missing traces: (a) 2D regular sampling, SNR=3.91 dB, (b) regular along source axis, discrete uniform random along receiver axis, SNR=7.30 dB, (c) regular along source axis, jittered along receiver axis, SNR=8.94 dB, (d) 2D jittered sampling, jittered sampling along receiver and source axes, same source pos. for all receivers, SNR=9.65 dB, (e) 2D jittered sampling along receiver and source axes, different source pos. for each receiver, SNR=10.03 dB, (f) Fully 2D jittered sampling, SNR=10.86 dB

for fully 2D (optimal) jittered sampling, and now there is no aliasing, with only some incoherent noise that can be easily filtered out from the spectrum.

The reconstructions from the incomplete data determined by the various sampling methods discussed in this paper are shown in Figure 16. As expected, the results progressively improve from Figure 16(a) to 16(f), as the sampling positions become less constrained and less regular. There is a notable ($\sim$ 1.64 dB) improvement even when samples are only taken irregularly in one dimension and regularly in the other, over completely 2D regular subsampling.

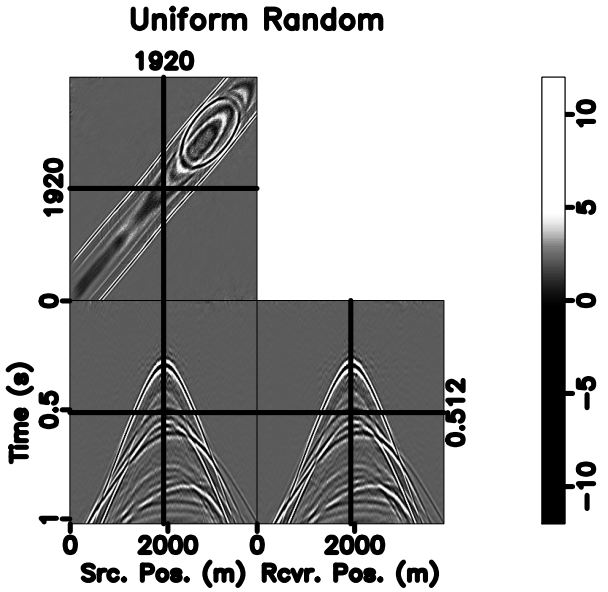
Next, 3D seismic acquisition patterns are employed, including 2D Poisson disk sampling and Farthest Point sampling. Here, only 25% of the total number of traces are sampled with different sampling methods, and then the recovery is performed with 3D CRSI. The interpolated results for the different schemes are shown in Figs. 17 and 18, and the residuals in Figs. 19 and 20. Again, jittered sampling is better than discrete uniform random sampling, while Poisson Disk and Farthest Point sampling give reconstruction qualities similar to those from jittered sampling.

The residuals (the original model minus the reconstructions) are plotted in Figures 19 and 20. From this latter figure, it is seen that Poisson Disk and Farthest Point sampling have residuals that look quite similar to each other, and both have slightly more energy than the residual for jittered sampling. The difference between the Farthest Point/Poisson Disk and the jittered sampling residuals is evident at the apex of the residual in the time slice at $t = 0.4$. The jittered sampling residual contains less energy than the residual from discrete uniform random sampling, most notably at the elliptical structure in the time slice.

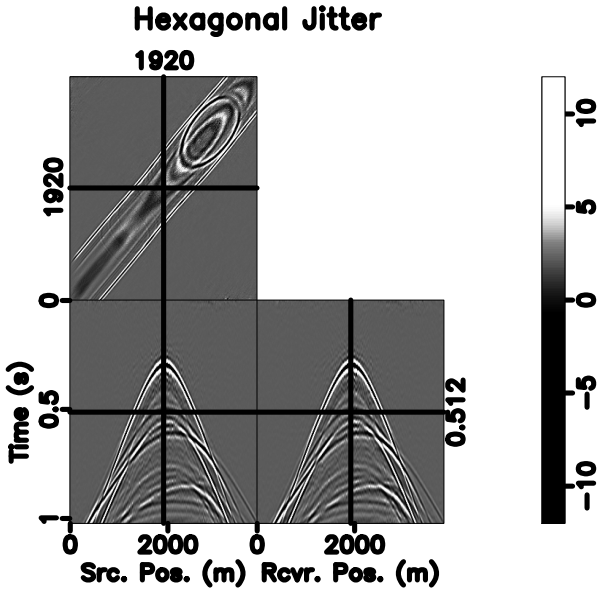
In Figures 21 and 22 are plotted the stacks of the CRSI results from Figures 17 and 18 respectively. The stacks of the residuals in Figures 19 and 20 are plotted in Figures 23 and 24 respectively. Clearly the residual stacks are dimmer than the stacks of the actual CRSI results, which is to be expected since the CRSI results are close to the full data and there is not much error between the two.

Field data

The SAGA dataset (Verschuur et al., 1992) and the reconstruction from CRSI with hexagonal jittered sampling are shown in Figure 25 and Figure 26 respectively.

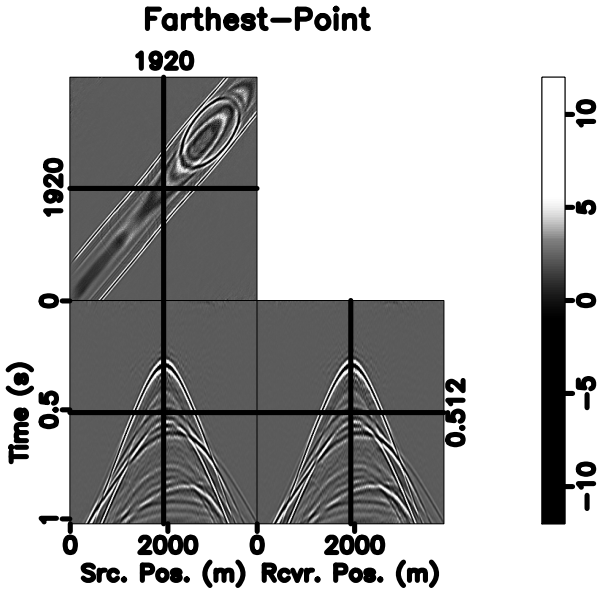


(a)

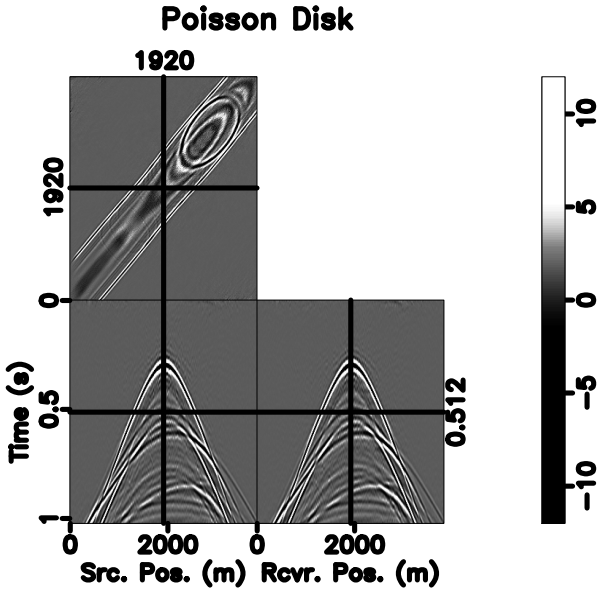


(b)

Figure 17: CRSI reconstruction from 75% missing traces sampled by (a) discrete uniform random sampling, SNR=12.91 dB and (b) jittered sampling, SNR=13.33 dB

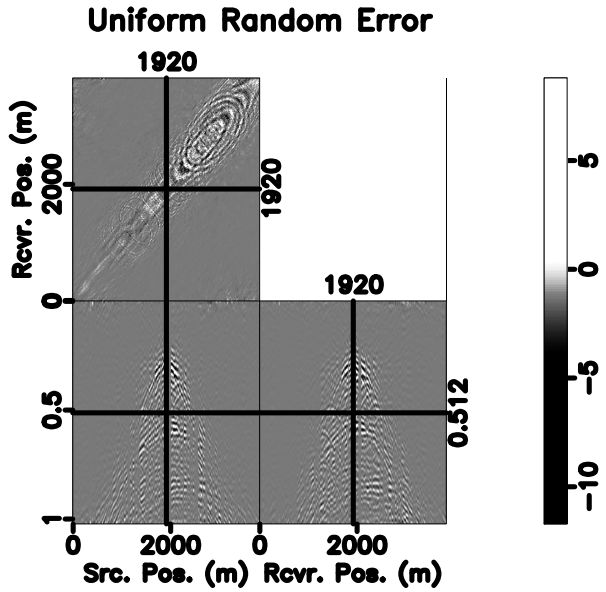


(a)

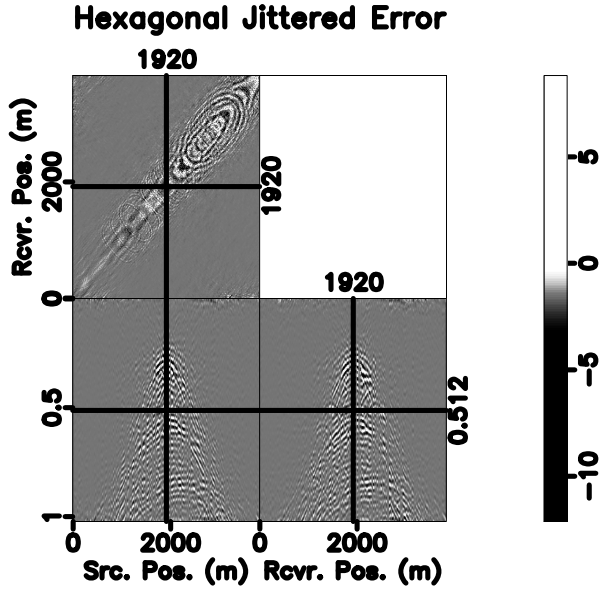


(b)

Figure 18: CRSI reconstruction from 75% missing traces sampled by (a) Farthest Point sampling, SNR=13.10 dB and (b) Poisson Disk sampling, SNR=13.09 dB

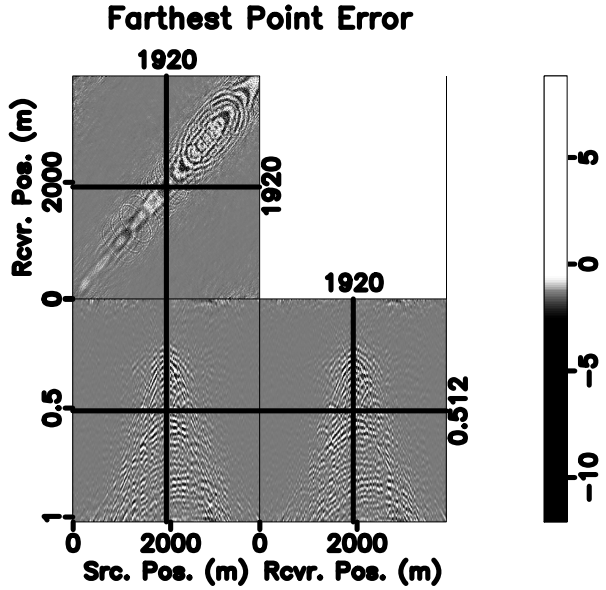


(a)

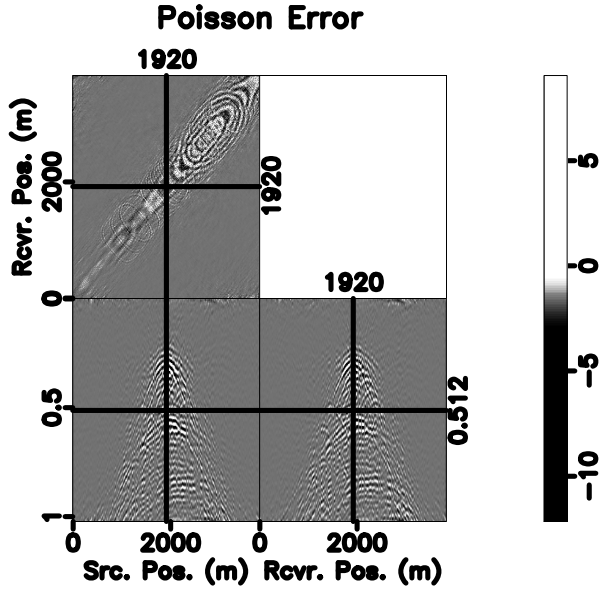


(b)

Figure 19: CRSI reconstruction residuals from (a) random sampling and (b) jittered sampling

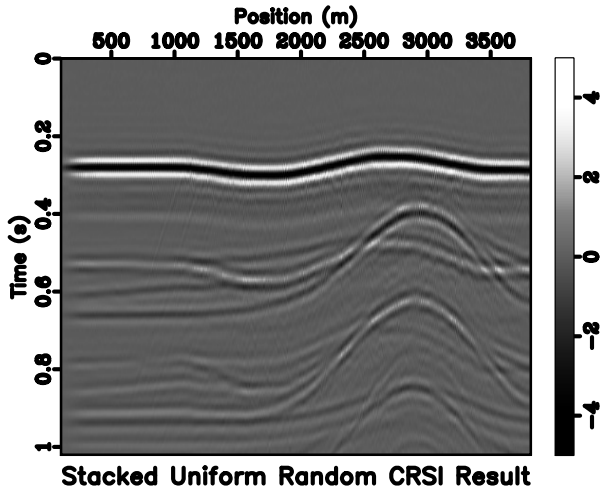


(a)

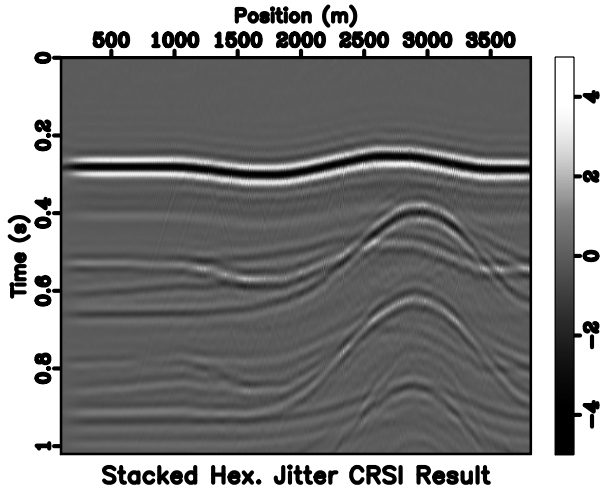


(b)

Figure 20: CRSI reconstruction residuals from (a) Farthest Point sampling and (b) Poisson Disk sampling

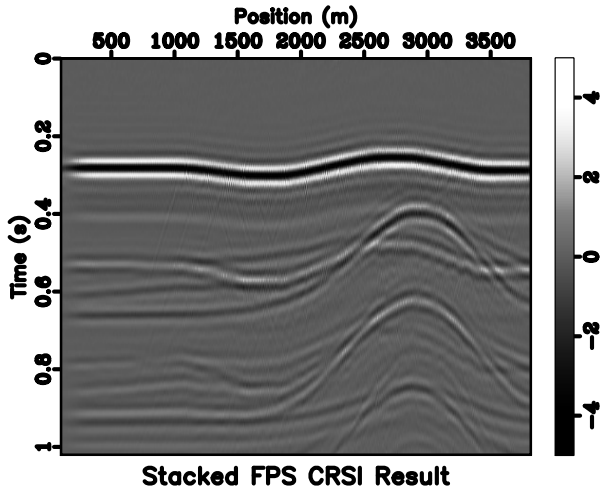


(a)

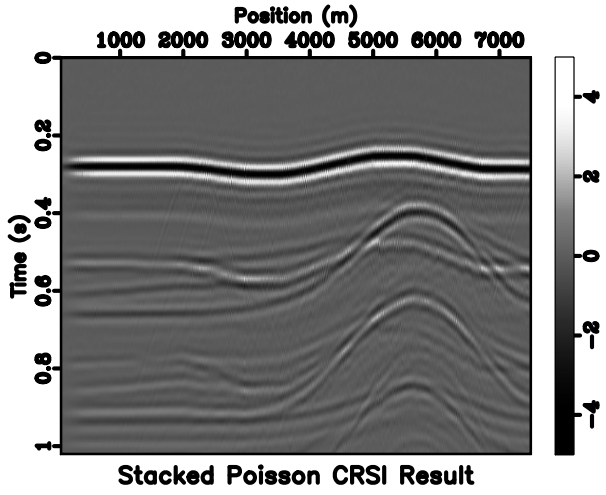


(b)

Figure 21: NMO Stacking of CRSI reconstruction from 75% missing traces sampled by (a) discrete uniform random sampling, and (b) jittered sampling

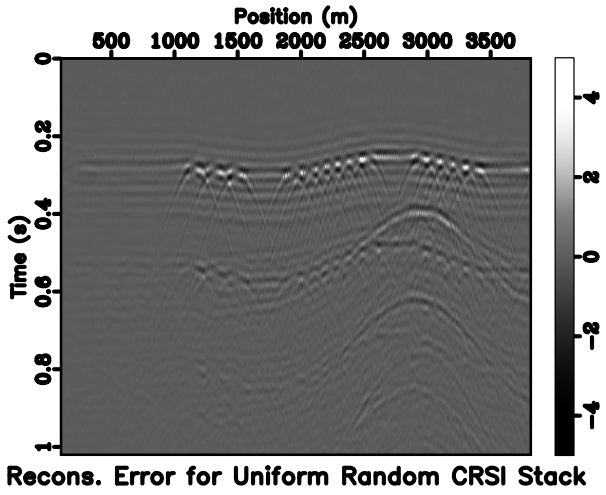


(a)

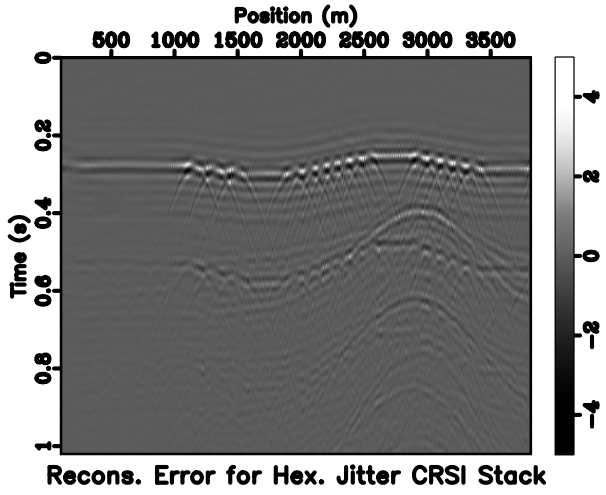


(b)

Figure 22: NMO Stacking of CRSI reconstruction from 75% missing traces sampled by (a) Farthest Point sampling and (b) Poisson Disk sampling

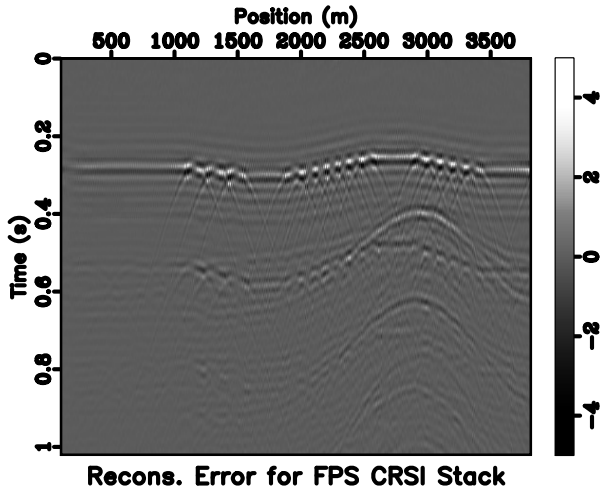


(a)

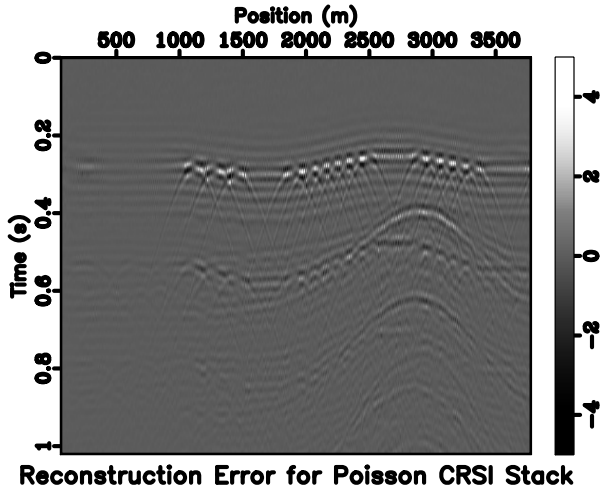


(b)

Figure 23: Error in NMO Stacking of CRSI reconstruction from 75% missing traces sampled by (a) discrete uniform random sampling and (b) jittered sampling



(a)



(b)

Figure 24: Error in NMO Stacking of CRSI reconstruction from 75% missing traces sampled by (a) Farthest Point sampling and (b) Poisson Disk sampling

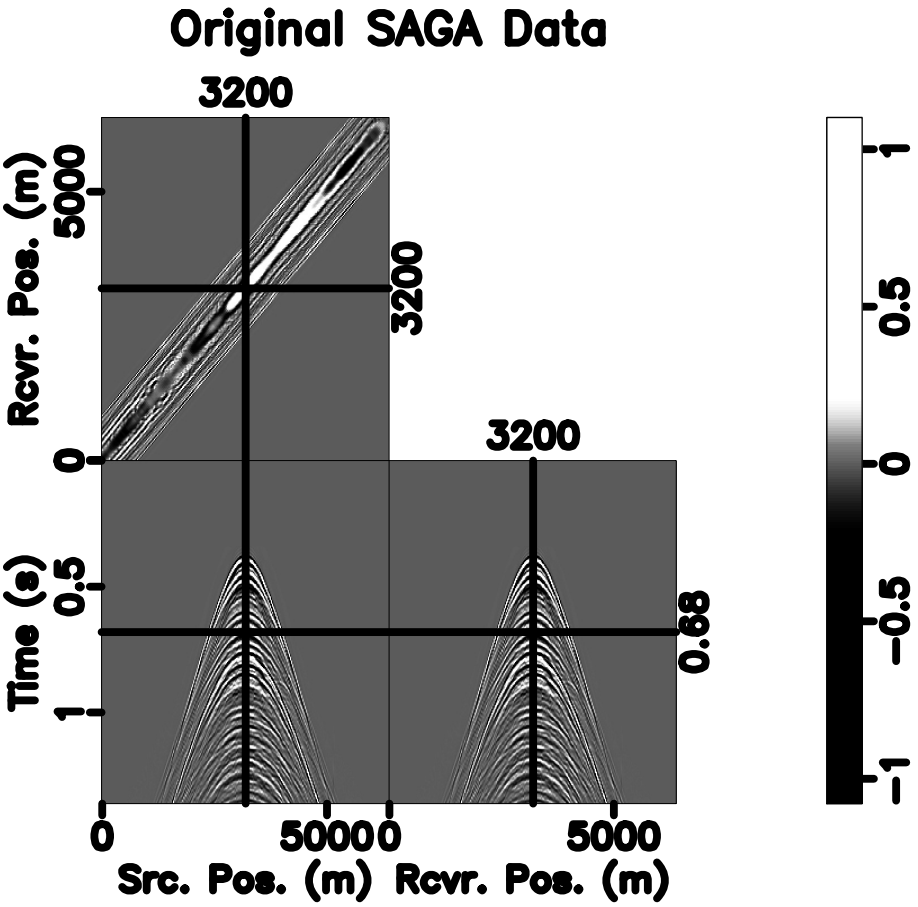


Figure 25: SAGA example: model

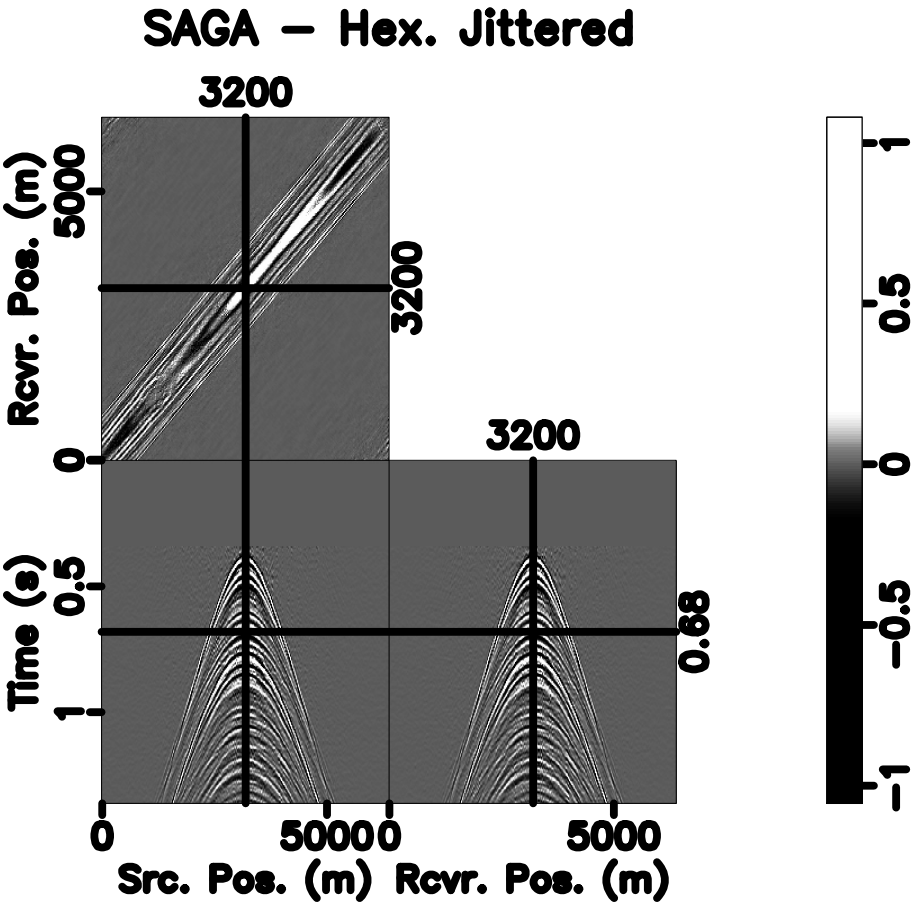
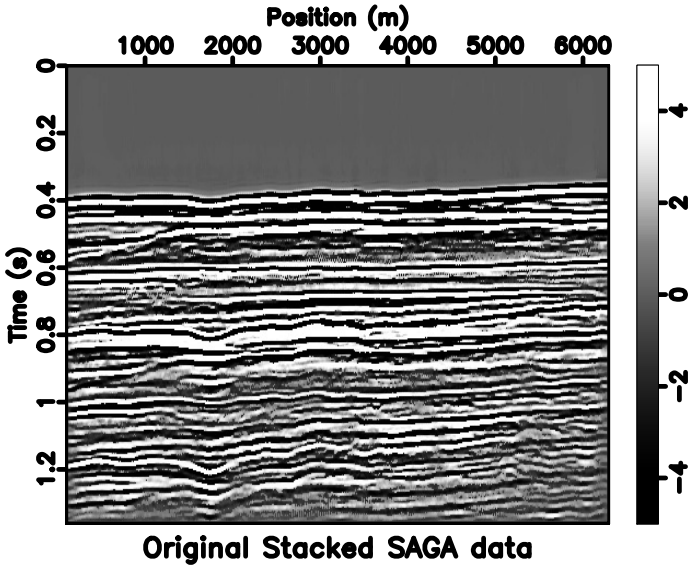
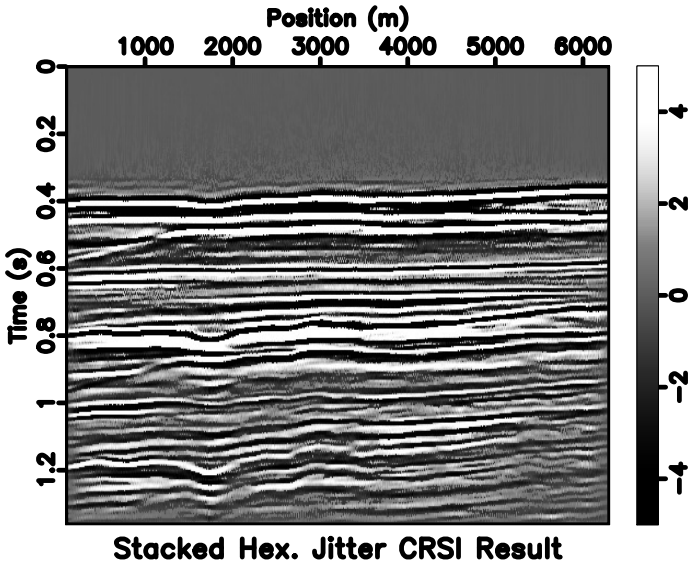


Figure 26: SAGA example: Reconstruction from CRSI after 200 iterations from 2/3 missing traces sampled by jittered hexagonal sampling, SNR=8.61



(a)



(b)

Figure 27: (a) Stack of original SAGA data and (b) Stack of reconstruction with CRSI after 200 iterations from subsampled SAGA data

The stacks of these results are shown in Figure 27(a) and 27(b) respectively. The reconstruction process is stopped relatively early after only 200 iterations, and the SNR of the reconstruction compared to the original full data is approximately 8.61, which is good for such few iterations of CRSI. The reconstruction is faithful to the original with most important structures present.

EXTENSIONS

Non-uniform CRSI

The present work used CRSI, where the curvelet transform is performed on a regular grid. Non-uniform CRSI (NCRSI) (Fénélon, 2008), based on the non-equispaced Fast Discrete Curvelet Transform (NFDCT), can deal with irregular subsampling points or traces better without binning.

Adaptive sampling and reconstruction

If some properties of the model for which data is being acquired are known a priori, then it is possible to sample more densely those areas of the image for which it is known there is more information content. In general, this would correspond to higher-frequency parts of the model since more samples are required to capture the wavefronts that are present in such regions, which are rapidly changing, and for which if a change is not captured by the samples, it will not show up in the reconstruction.

In addition to this, the reconstruction may also be made adaptive. The analytic expressions for the sampling spectra for hexagonal and Cartesian jittered sampling in the appendix, can be used to weight the curvelet coefficient vector so that more curvelet-domain sparsity is enforced in the image at frequencies in the sampling

spectrum where there is less energy. These two extensions are left to future work, along with symmetric sampling, which also allows the sampling density to be reduced.

Insights from compressive sampling theory

The profound implication of compressive sensing, of which this paper is a particular instance, is that the sampling rates are no longer determined by the Nyquist sampling criterion but rather the complexity of the underlying data. Instead, the sampling rate can be chosen in accordance with the recovery quality one desires. This error decays for increasing sampling rates or for more compressive wavefields. In addition, (Herrmann et al., 2009) recently showed that simultaneous acquisition is another instance of compressive sensing, which leads to even better recovery. Combination of simultaneous acquisition with the approach presented in this paper would allow us to obtain an even greater reduction in the number of sources and receivers required for sampling a wavefield. Because both methods are linear, the development of a progressive sampling strategy is envisaged by the authors, where the reconstruction from previously taken samples is updated as new samples are added until a reconstructed wavefield of desired quality is obtained.

CONCLUSIONS

In this paper, sampling schemes with blue-noise patterns for curvelet-based interpolation have been explored. These methods are especially good for seismic acquisition design due to their inherent characteristic of not leaving large regions in a seismic survey unsampled. Jittered sampling, one such blue-noise method which has been used for one-dimensional acquistion design in the past, was extended to two dimensions. It was found that 2D jittered sampling is significantly better than uniform random sampling as it led to reconstructions of significantly higher quality. Hexagonal

jittered sampling is superior to Cartesian jittered sampling due to its less stringent requirements on the sampling density for images of the same bandwidth. Additionally, separable and non-separable jittered sampling were studied, and it was found that completely non-separable jittered sampling is better. Two other blue-noise sampling methods, Farthest Point and Poisson Disk sampling were also tested, and gave good results comparable with 2D hexagonal jittered sampling. In particular, all of these sampling techniques were applied to 2D and 3D seismic data interpolation by CRSI, and excellent reconstruction results were obtained with few shots, as compared to methods previously used in the literature. These acquisition designs are also suitable for other transform-based methods, besides CRSI, e.g. Fourier interpolation.

It is also valuable to investigate the combination of the ideas of spatial subsampling design with the approach of using simultaneous independent sources. This would lead to an even greater savings in the required sampling rate to achieve a desired SNR for the reconstructed wavefield.

ACKNOWLEDGMENTS

This work was in part financially supported by the NSERC Discovery Grant 22R81254 and CRD Grant DNOISE 334810-05 of F.J. Herrmann. The authors would also like to thank the China Scholarship Council, National Basic Research Program of China (973 program, 2007CB209505), and NSFC 40704019. The authors are grateful to SAGA Petroleum A.S. for providing the field data set.

REFERENCES

Abma, R., and N. Kabir, October 2005, Comparisons of interpolation methods: The Leading Edge, 984–989.

Bartlett, M. S., 1963, The spectral analysis of point processes: Journal of the Royal Statistical Society, Series B (Methodological), **25**, 264–296.

Bengtsson, T., P. Bickel, and B. Li, 2008, Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems: Probability and statistics: Essays in honor of David A. Freedman, **2**, 316–334.

Candès, E. J., 2006, Compressive sampling: Presented at the International Congress of Mathematicians, Madrid, Spain.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: SIAM Multiscale Modeling and Simulation., **5**, 861–899.

Candès, E. J., J. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics, **59**, 1207–1223.

Chauris, H., and T. Nguyen, 2008, Seismic demigration/migration in the curvelet domain: Geophysics, **73**, S35–S46.

Cook, R. L., 1986, Stochastic sampling in computer graphics: ACM Transactions on Graphics, **5**, no. 3, 51–72.

Dippé, M. A. Z., and E. H. Wold, 1985, Antialiasing through stochastic sampling: Proceedings of SIGGRAPH 85, 69–78.

Donoho, D. L., 2006, Compressed sensing: IEEE Trans. Inf. Theory, **52**, 1289–1306.

Eldar, Y., M. Lindenbaum, M. Porat, and Y. Y. Zeevi, 1997, The farthest point strategy for progressive image sampling: IEEE Transactions on Image Processing, **6**, no. 9, 1305–1315.

Fénélon, L., 2008, Nonequispaced discrete curvelet transform for seismic data reconstruction: Master's thesis.

Grundland, M., J. Patera, M. Zuzana, and A. Neil, 2009, Image Sampling with Quasicrystals: Arxiv preprint arXiv:0907.3604.

Gulunay, N., and R. E. Chambers, 1997, Generalized fk domain trace interpolation: 67^{th} SEG Annual Meeting, Expanded Abstracts, 1100–1103.

Hennenfent, G., and F. J. Herrmann, 2007, Irregular (sub-)sampling: from aliasing to noise: Presented at the 69^{th} EAGE Conference and Exhibition.

———, 2008, Simply denoise: wavefield reconstruction via jittered undersampling: Geophysics, **73**, V19–V28.

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007, Nonlinear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, 781–799.

Herrmann, F. J., Y. A. Erlangga, and T. Lin, 2009, Compressive simultaneous full-waveform simulation: Geophysics, **74**, A35–A40.

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

Ignjatovic, Z., and M. F. Bocko, 2005, A method for efficient interpolation of discrete-time signals by using a blue-noise mapping method: Proceedings of the IEEE ICASSP 2005, 213– 216.

László, S.-K., 1995, Stochastic sampling of two-dimensional images: COMPUGRAPH-ICS 95, 292–299.

Long, A. S., 2004, The revolution in seismic resolution: High density 3d spatial sampling developments and results: Presented at the ASEG Extended Abstracts.

Lustig, M., M. Alley, S. Vasanawala, D. Donoho, and J. Pauly, 2009, Autocalibrating parallel imaging compressed sensing using l1 spir-it with poisson-disc sampling and joint sparsity constraints: Presented at the ISMRM Workshop on Data Sampling and Image Reconstruction, Sedona '09.

Ma, J., and G. Plonka, 2010, The curvelet transform: IEEE Signal Processing Magazine,

27, 118–133.

Mersereau, R. M., 1979, The processing of hexagonally sampled two dimensional signals: Proceedings of the IEEE, **67**, 930–949.

Mitchell, D., 1987, Generating antialiased images at low sampling densities: Proc. of the 14th annual conf. on computer graphics and interactive techniques, 65–72.

Moldoveanu, N., 2010, Random sampling: A new strategy for marine acquisition: Presented at the 2010 SEG Annual Meeting.

Mosher, C., S. Kaplan, and F. Janiszewski, 2012, Random sampling: A new strategy for marine acquisition: Presented at the 74th EAGE Conference and Exhibition.

Olofsson, B., P. Mitchell, and R. Doychev, 2012, Decimation test on an ocean-bottom node survey: Feasibility to acquire sparse but full-azimuth data: The Leading Edge, **31**, 457–464.

Oppenheim, A., R. Schafer, and J. Buck, 1999, Discrete-time signal processing: Prentice Hall Signal Processing Series.

P. Brémaud, L. M., and A. Ridolfi, 2002, Power spectra of random spike fields and related processes: Technical report, École Polytechnique Fédérale de Lausanne, School of Computer and Information Sciences. (Technical Report 200280).

Spitz, S., 1991, Seismic trace interpolation in the f-x domain: Geophysics, **56**, 785–794.

Theußl, T., T. Moller, and M. Groller, 2001, Optimal regular volume sampling: Visualization, 2001. VIS'01. Proceedings, 91–546.

Vermeer, G., 1990, Seismic wavefield sampling: Society of Exploration Geophysics.

Verschuur, D., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.

Xu, S., Y. Zhang, D. Pham, and G. Lambaré, 2004, On the orthogonality of anti-leakage fourier transform based seismic trace interpolation: Presented at the 2004 SEG Annual Meeting.

Zwartjes, P. M., and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly

sampled, aliased data: Geophysics, **72**, V21–V32.

APPENDIX - ANALYSIS OF 2D JITTERED SAMPLING

From the technical report of Brémaud et al. (P. Brémaud and Ridolfi, 2002), acquisition designs are if there is a wide sense stationary point process S jittered by a point process Z, so that the jitter point process J is given by $J = S + Z$, then the *Bartlett spectrum* of J, $\mu_J(d\mathbf{f})$ is given by

$$\mu_J(d\mathbf{f}) = |\psi_Z(\mathbf{f})|^2 \mu_S(d\mathbf{f}) + \lambda(1 - |\psi_Z(\mathbf{f})|^2)d\mathbf{f}, \tag{8}$$

where $\mu_S(d\mathbf{f})$ is the Bartlett spectrum of the process S, $\psi_Z(\mathbf{f})$ is the characteristic function of Z, and λ is the *intensity* of the process S. The arguments of functions in Equation 8 are $d\mathbf{f}$ because the power spectral density measures the expected power in small intervals around each point in the wavenumber domain. The Bartlett power spectrum is a generalization of the usual power spectral density for point processes which are not wide-sense stationary (Bartlett, 1963). Since S is a deterministic process, the Bartlett spectrum of S will be the power spectrum and the intensity of S will be the sampling density. So Equation 8 becomes:

$$S_J(\mathbf{f}) = |\psi_Z(\mathbf{f})|^2 S_S(\mathbf{f}) + \lambda(1 - |\psi_Z(\mathbf{f})|^2) = |E\{e^{-2\pi i(\mathbf{f}\cdot\bar{\epsilon})}\}|^2 |\hat{S}(\mathbf{f})|^2 + \lambda(1 - |E\{e^{-2\pi i(\mathbf{f}\cdot\bar{\epsilon})}\}|^2) \tag{9}$$

The above can be shown to be equal to

$$S_J(\mathbf{f}) = (|E\{\hat{J}(\mathbf{f})\}|)^2 + \lambda(1 - |E\{e^{-2\pi i(\mathbf{f}\cdot\bar{\epsilon})}\}|^2), \tag{10}$$

since the jitter point process Z is assumed to be independent of the regular sampling process S.

The expected value and standard deviation of the Fourier spectrum for hexagonal jittered sampling are derived next. This can then be shown to be equivalent to finding

the expectation and standard deviation of the entries of the Gram matrix $\mathbf{A}^H\mathbf{A}$ for the sampling, which can be used to analyze the amount of aliasing present for this type of jittered sampling. Similar expressions for rectangular jittered sampling can also be easily derived.

Analysis of aliasing for 2D regular hexagonal jittered sampling

Following the example in (Hennenfent and Herrmann, 2008) for one-dimensional jittered sampling, expressions are derived to analyze the aliasing for 2D regular hexagonal jittered sampling. In order to do this, the spectrum of this sampling process should be found. Since this process is random, the spectrum will also be a random process. Hence the expectation and variance of this random spectrum are calculated. Let $\frac{\gamma}{2}$ be the radius of any regular hexagon in the hexagonal sampling grid, so that $\gamma = 2W$ as in Figure 6. In this case, the jittered sampling locations $\mathbf{r}_n$ are determined by the formula:

$$\mathbf{r}(k_1, k_2) = \mathbf{r}_1(k_1, k_2) \bigcup \mathbf{r}_2(k_1, k_2), \tag{11}$$

where

$$\mathbf{r}_1(k_1, k_2) = \left(\frac{3}{2}\gamma k_1, \frac{\sqrt{3}}{2}\gamma k_2\right) + \epsilon_{\mathbf{1}k_1,k_2} \tag{12}$$

and

$$\mathbf{r}_2(k_1, k_2) = \left(\frac{3}{2}\gamma k_1 + \frac{3}{4}\gamma, \frac{\sqrt{3}}{2}\gamma k_2 + \frac{\sqrt{3}}{4}\gamma\right) + \epsilon_{\mathbf{2}k_1,k_2}, \tag{13}$$

and the $\epsilon_{\mathbf{i}k_1,k_2}$ are a set of identically uniformly and independently distributed random variables on the hexagon with circumscribing circle of radius $\frac{\zeta}{2}$, centered at the origin, for all indices k_1 and k_2 and $i = 1$ or 2. Call this distribution of the $\epsilon_{\mathbf{i}k_1,k_2}$, $p_{\frac{\zeta}{2}}(\epsilon)$.

Then similar to (Hennenfent and Herrmann, 2008), the sampling operator s is

given by the expression:

$$s(\mathbf{r}) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \delta(\mathbf{r} - \mathbf{r}_1(k_1, k_2)) + \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \delta(\mathbf{r} - \mathbf{r}_2(k_1, k_2)). \tag{14}$$

The Fourier transform of the sampling operator s is

$$\hat{S}(\mathbf{f}) = \frac{4}{3\sqrt{3}\gamma^2} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \delta\left(f_1 - k_1\frac{2}{3\gamma}\right) \delta\left(f_2 - k_2\frac{2}{\sqrt{3}\gamma}\right) \left(1 + (-1)^{k_1+k_2}\right) e^{-2\pi i \mathbf{f}\cdot\epsilon_{\mathbf{1}k_1,k_2}}, \tag{15}$$

where $\mathbf{f} = (f_1, f_2)$ is the wavenumber variable vector. Therefore, the expected value of the Fourier transform of the sampling operator s is:

$$E\{\hat{S}(\mathbf{f})\} = \\ E\{e^{-2\pi i(f_1\bar{\epsilon}_{\mathbf{x}} + f_2\bar{\epsilon}_{\mathbf{y}})}\} \cdot \frac{4}{3\sqrt{3}\gamma^2} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \delta\left(f_1 - k_1\frac{2}{3\gamma}\right) \delta\left(f_2 - k_2\frac{2}{\sqrt{3}\gamma}\right) \left(1 + (-1)^{k_1+k_2}\right). \tag{16}$$

Here, because all the $\epsilon_{\mathbf{1}k_1,k_2}$'s are identically distributed, they can be replaced by a generic $\epsilon = (\bar{\epsilon}_{\mathbf{x}}, \bar{\epsilon}_{\mathbf{y}})$ in the expected value of the sampling operator's Fourier transform. So it remains to compute $E\{e^{-2\pi i(f_1\bar{\epsilon}_{\mathbf{x}} + f_2\bar{\epsilon}_{\mathbf{y}})}\}$, where $(\bar{\epsilon}_{\mathbf{x}}, \bar{\epsilon}_{\mathbf{y}})$ is distributed uniformly on the hexagon $\mathbf{H}_\zeta$ with circumscribing circle of radius $\frac{\zeta}{2}$ centered at the origin. This is in fact the *characteristic* function $\psi_Z(\mathbf{f})$ as in Equation 8, of the distribution $p_{\frac{\zeta}{2}}(\epsilon)$. After some calculations, the following expression is obtained for this characteristic function:

$$\begin{aligned} \psi_Z(\mathbf{f}) = & \frac{4\sqrt{3}}{9\pi^2 f_1 f_2 (f_1^2 - 3f_2^2)\zeta^2} \left(2\sin\left(\frac{\pi\zeta}{2}f_1\right) \sin\left(\frac{\pi\zeta\sqrt{3}}{2}f_2\right) (f_1^2 - 3f_2^2) - \right. \\ & 2f_1 f_2\sqrt{3}\cos(\pi\zeta f_1) + \cos\left(\frac{\pi\zeta}{2}(f_1 + \sqrt{3}f_2)\right) f_1(f_1 + \sqrt{3}f_2) - \\ & \left. \cos\left(\frac{\pi\zeta}{2}(f_1 - \sqrt{3}f_2)\right) f_1(f_1 - \sqrt{3}f_2)\right) \end{aligned} \tag{17}$$

By the sifting property of the δ function, the expected value of $\hat{S}(\mathbf{f})$ may be simplified to:

$$E\{\hat{S}(\mathbf{f})\} = \sum_{\substack{k_1,k_2=-\infty \\ k_1+k_2\ even}}^{\infty} \frac{8\sqrt{3}}{9\pi^2 k_1 k_2 (k_1^2 - 9k_2^2)\zeta^2} \left(\cos\left(\frac{\pi\zeta}{3\gamma}(k_1+3k_2)\right) k_1(k_1+3k_2) - \right.$$
$$\cos\left(\frac{\pi\zeta}{3\gamma}(k_1-3k_2)\right) k_1(k_1-3k_2) + 2\sin\left(\frac{\pi\zeta}{3\gamma}k_1\right)\sin\left(\frac{\pi\zeta}{\gamma}k_1\right)(k_1^2-9k_2^2) \quad (18)$$
$$\left. -6k_1k_2\cos\left(\frac{2\pi\zeta}{3\gamma}k_1\right)\right) \delta\left(f_1 - \frac{k_1}{\frac{3}{2}\gamma}\right)\delta\left(f_2 - \frac{k_2}{\frac{\sqrt{3}}{2}\gamma}\right).$$

The intensity of the regular hexagonal sampling process λ is computed as the sampling density, or the expected number of samples per unit area, $\lambda = \frac{8}{3\sqrt{3}\gamma^2}$.

Let the ratio of the perturbation range to the edge length be $\beta = \frac{\zeta}{\gamma}$. Then the above expression becomes:

$$E\{\hat{S}(\mathbf{f})\} = \sum_{\substack{k_1,k_2=-\infty \\ k_1+k_2\ even}}^{\infty} \frac{8\sqrt{3}k_1}{9\pi^2 k_1 k_2 (k_1^2 - 9k_2^2)\zeta^2} \left(\cos\left(\frac{\pi\beta}{3}(k_1+3k_2)\right) k_1(k_1+3k_2) - \right.$$
$$\cos\left(\frac{\pi\beta}{3}(k_1-3k_2)\right)(k_1-3k_2) + 2\sin\left(\frac{\pi\beta}{3}k_1\right)\sin(\pi\beta k_1)(k_1^2-9k_2^2)$$
$$\left. -6k_1k_2\cos\left(\frac{2\pi\beta}{3}k_1\right)\right) \delta\left(f_1 - \frac{k_1}{\frac{3}{2}\gamma}\right)\delta\left(f_2 - \frac{k_2}{\frac{\sqrt{3}}{2}\gamma}\right). \quad (19)$$

There is still a dependence on the maximum jitter magnitude ζ, but this is expected, since for a given perturbation ratio β, as γ, and thus ζ increases, the hexagonal jittered sampling grid spreads out to become more widely spaced, implying the sampling mask becomes lower in wavenumber. As $(k_1, k_2) \to (0, 0)$, the limit of the above expression is 1, as expected, since this corresponds to the base replica of the sampling spectrum, and should have magnitude of unity. Other values of the spectrum can be calculated by substituting various values of k_1 and k_2, and these values can only be non-zero when the condition that $k_1 + k_2$ is even holds.

Equation 19 gives the expected value of the regular hexagonal jittered sampling spectrum. It is now desired to find the variance of this random spectrum. This can be done by computing the power spectral density for the regular hexagonal jittered sampling process, and then using the fact that

$$\sigma_J^2 = S_J(\mathbf{f}) - [E\{\hat{S}(\mathbf{f})\}]^2 = \lambda(1 - |\psi_Z(\mathbf{f})|^2) \tag{20}$$

Therefore, the variance of the hexagonal jittered sampling process is given by

$$\sigma_J^2 = \frac{8}{3\sqrt{3}\gamma^2}\left(1 - \left(\frac{4\sqrt{3}}{9\pi^2 f_1 f_2 (f_1^2 - 3f_2^2)\zeta^2}\left(2\sin\left(\frac{\pi\zeta}{2}f_1\right)\sin\left(\frac{\pi\zeta}{2}\sqrt{3}f_2\right)(f_1^2 - 3f_2^2) - 2f_1 f_2\sqrt{3}\cos(\pi\zeta f_1) + \cos\left(\frac{\pi\zeta}{2}(f_1 + \sqrt{3}f_2)\right)f_1(f_1 + \sqrt{3}f_2) - \cos\left(\frac{\pi\zeta}{2}(f_1 - \sqrt{3}f_2)\right)f_1(f_1 - \sqrt{3}f_2)\right)\right)^2\right) \tag{21}$$

Ignoring discretization effects, the elements of the Gram matrix $\mathbf{L} = \mathbf{A^H A}$ are each random with mean given by Equation 19 and variance given by Equation 21, but where only the elements of this matrix corresponding to grid points closest to the impulse locations in the frequency response for the regular hexagonal jittered samples are equal to the corresponding magnitudes of the impulse functions and can be non-zero.

Analysis of aliasing for 2D regular Cartesian jittered sampling

The expectation and standard deviation of the sampling spectrum for Cartesian jittered sampling are now found. Since the derivation of these functions for the more complicated case of hexagonal jittered sampling has already been gone through, many of the steps in the present derivation are skipped.

For regular Cartesian jittered sampling, it is found that the characteristic function is:

$$\psi_Z(\mathbf{f}) = \frac{\sin(\pi f_1 \zeta)\sin(\pi f_2 \zeta)}{\pi^2 \zeta^2 f_1 f_2} \tag{22}$$

Hence the expected value of the spectrum for this type of sampling can be found to be:

$$E\{\hat{S}(\mathbf{f})\} = \frac{\sin(\pi k_1 \frac{\zeta}{\gamma})\sin(\pi k_2 \frac{\zeta}{\gamma})}{\pi^2 \zeta^2 k_1 k_2} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \delta\left(f_1 - \frac{k_1}{\gamma}\right) \delta\left(f_2 - \frac{k_2}{\gamma}\right) \tag{23}$$

The variance, once again wideband noise, is computed to be:

$$\sigma_J^2 = \lambda(1 - |\psi_Z(\mathbf{f})|^2) = \frac{1}{\gamma^2}\left(1 - \frac{\sin^2(\pi f_1 \zeta)\sin^2(\pi f_2 \zeta)}{\pi^4 \zeta^4 f_1^2 f_2^2}\right) \tag{24}$$