Blended-acquisition design of irregular geometries towards faster, cheaper, safer and better seismic surveying

Nakayama, Shotaro; Blacquière, Gerrit; Ishiyama, Tomohide; Ishikawa, Satoshi

DOI
10.1111/1365-2478.12701

Publication date
2019

Document Version
Final published version

Published in
Geophysical Prospecting

Geophysical Prospecting

EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

Geophysical Prospecting, 2019, **67**, 1498–1521 doi: 10.1111/1365-2478.12701

Blended-acquisition design of irregular geometries towards faster, cheaper, safer and better seismic surveying

Shotaro Nakayama[1,2]*, Gerrit Blacquière[1], Tomohide Ishiyama[2,3] and Satoshi Ishikawa[2]

[1]*Delft University of Technology, Mekelweg 5, 2628 CD Delft, the Netherlands,* [2]*INPEX Corporation, Akasaka Biz Tower, 5-3-1 Akasaka, Minato-ku, Tokyo, 107-6332, Japan, and* [3]*ADNOC Research and Innovation Center, Abu Dhabi, United Arab Emirates*

Received February 2018, revision accepted September 2018

ABSTRACT

The application of blended acquisition has drawn considerable attention owing to its ability to improve the operational efficiency as well as the data quality and health, safety and environment performance. Furthermore, the acquisition of less data contributes to the business aspect, while the desired data density is still realizable via subsequent data reconstruction. The use of fewer detectors and sources also minimizes operational risks in the field. Therefore, a combined implementation of these technologies potentially enhances the value of a seismic survey further. One way to encourage this is to minimize any imperfection in deblending and data reconstruction during processing. In addition, one may derive survey parameters that enable a further improvement in these processes as introduced in this study. The proposed survey design workflow iteratively performs the following steps to derive the survey parameters responsible for source blending as well as the spatial sampling of detectors and sources. The first step is the application of blending and sampling operators to unblended and well-sampled data. We then apply closed-loop deblending and data reconstruction. The residue for a given design from this step is evaluated and subsequently used by genetic algorithms to simultaneously update the survey parameters related to both blending and spatial sampling. The updated parameters are fed into the next iteration until they satisfy the given termination criteria. We also propose a repeated encoding sequence to form a parameter sequence in genetic algorithms, making the size of problem space manageable. The results of the proposed workflow are outlined using blended dispersed source array data incorporating different scenarios that represent acquisition in marine, transition zone and land environments. Clear differences attributed solely to the parameter design are easily recognizable. Additionally, a comparison among different optimization schemes illustrates the ability of genetic algorithms along with a repeated encoding sequence to find better solutions within a computationally affordable time. The optimized parameters yield a notable enhancement in the deblending and data reconstruction quality and consequently provide optimal acquisition scenarios.

Key words: Seismic acquisition, Sampling, Iterative scheme.

INTRODUCTION

In a conventional seismic survey, detectors and sources are deployed at regular spatial intervals. Each source then emits a

*E-mail: S.nakayama@tudelft.nl; shotaro.nakayama@inpex.co.jp

uniform signature with sufficiently large temporal and spatial separations with respect to other sources. This allows the energy from the former shot to decay to an acceptable level or to propagate outside the area of interest before the data associated with the following shot arrive at the detectors. However, operations become costly and time consuming along with an increase in the health, safety and environment (HSE) exposure in the field. These factors potentially curtail our opportunities to acquire seismic data, particularly when desiring denser sampling along with longer apertures for all azimuths. One hence needs to design a seismic survey in a more efficient way while satisfying geophysical requirements.

Over the last decade, blended acquisition, or sometimes referred to as simultaneous source acquisition, has received considerable attention in the industry owing to its ability to drastically change the business and technical aspects of seismic data acquisition. Furthermore, the enhancement in the survey efficiency significantly minimizes the HSE risk during the operation. This has consequently resulted in the widespread application of the technique. Beasley, Ronald and Zerong (1998) proposed the simultaneous deployment of two sources at both ends of a streamer cable. They demonstrated the possibility of separating interference noise using conventional processes such as multichannel filtering, dip moveout and prestack migration. Berkhout (2008) proposed the concept of blending in acquisition and processing. Blended acquisition is a way in which seismic data are continuously recorded along with a significant overlap among consecutive shots in time and space as well as in temporal and spatial frequencies to produce 'blended shot records'. A common practice for processing this type of data is to deblend the blended records first rather than processing them directly (although it is expected that the emphasis will shift to the latter in the future). In most cases, deblending is posed as an inversion problem that iteratively estimates unblended data in some transform domain such as the Fourier (Mahdad, Doulgeris and Blacquière 2011), Radon (Moore *et al.* 2008), Curvelet (Lin and Herrmann 2009) and Focal (Kontakis and Verschuur 2015) domains. The deblended data then enable us to make use of standard processing algorithms and to treat data as if they were acquired in a conventional, unblended manner.

The acquisition of less data also contributes to the business aspect as well as a reduction in the environmental footprint. Subsequent data reconstruction then allows the processed data to attain the desired data density. A number of studies have demonstrated both practical and theoretical aspects of this technique. For instance, the well-known and widely used approaches are transformation-based ones such as the use of the Fourier (Xu *et al.* 2005), Curvelet (Hennenfent and Herrmann 2008), Radon (Verschuur, Vrolijk and Tsingas 2012) and Focal (Kutscha and Verschuur 2012) domains. As the forward and inverse operators for these domains are not orthogonal, data reconstruction is generally posed as an inversion problem to minimize the loss of information in the input data. The methods utilize the redundancy of the data that assumes the desired data can be efficiently described in the transform domain. Along with given constraints such as sparseness of the solution and/or a guide from non-aliased low frequency components, the transform coefficients are iteratively estimated, meaning that the desired signals are separated from undesired aliasing noise in the transform domain. This is possible when the aliasing noise has different properties in the model space from the signal. Once model parameters are estimated, optimum data recovery is attainable via the inverse transformation even when input data do not meet Nyquist sampling theorem. This potentially allows us to ease survey requirements related to the spatial sampling criteria.

As described previously, deblending resembles data reconstruction in the sense that both methods are generally posed as an inversion problem that iteratively estimates signals in the transform domain where they are known to be sparse and well separated from undesired events such as the blending noise and aliasing noise. For example, Kutscha and Verschuur (2012) and Kontakis and Verschuur (2015) dealt with data reconstruction and deblending respectively using similar inversion frameworks. Some recent studies have jointly handled both deblending and data reconstruction (Cheng and Sacchi 2015; Ishiyama *et al.* 2017; Li *et al.* 2013). In principle, deblended and reconstructed data are iteratively estimated in the transform domain together with prior knowledge and/or constraints to data. By making use of the combination of operators responsible for both blending and spatial sampling schemes derived from the measured data, the forward process to generate an estimate of blended data on the measured grid is possible. This subsequently allows for evaluation of the residual between the measured data and the estimated data at each iteration. Data misfit is then used for the next iteration to minimize the residual further. In existing blended acquisition schemes, irregularity or randomness is often embedded into the survey parameters such as a random time delay for each source, allowing for effective source separation (Baardman and van Borselen 2013). The source wavefield is made incoherent in at least one of the sorting domains by the use of a random time delay, a randomized distance between concurrent sources for each blended shot, a unique encoding for each

source, or their combination. Similarly, irregularity is also of importance when sampling detectors and sources. Herrmann (2010) showed that the spatial sampling of data in an irregular fashion is a key element for implementing compressive sensing (CS). Irregularity permits the energy of spectral leakage to spread over the entire spectrum and to behave as if it were random noise. In CS, these random-like artefacts are well-separated in a certain transform domain. Sparsity-based inversion subsequently recovers the desired signals. The relevant and conventional survey parameters for both land and marine environments are the four spatial sampling intervals and four apertures of the template geometries (Vermeer 2012). In a conventional way, as long as detectors and sources are regularly deployed, the number of parameters is still manageable (Ishiyama, Mercado and Belaid 2012; Nakayama, Belaid and Ishiyama 2013). However, this is not the case with blending and irregular geometries. Despite their potential benefits, designing a survey incorporating these techniques is rather intricate as irregularity inherently requires numerous selections of survey parameters, unlike acquisition in a regular manner. Therefore, it is obviously worthwhile to seek an efficient and practical way to design a survey in order to incorporate blending and irregular geometries. Additionally, different types of survey parameters interactively influence the quality of the subsequent processing steps and eventually that of the final subsurface image. To deal with this, we have to optimize the parameters simultaneously rather than sequentially.

In this paper, we introduce a survey design workflow that iteratively optimizes the survey parameters responsible for both the blending and spatial sampling of detectors and sources, leading to satisfactory deblending and data reconstruction results. The workflow includes a closed-loop approach allowing for a robust deblending and data reconstruction. The residue for a given survey design is evaluated and subsequently input into another system based on genetic algorithms (GAs) to update the blending and sampling operators. A repeated encoding sequence (RES) is proposed and implemented to form a parameter sequence for GAs, making the size of the problem space affordable. Several numerical examples incorporating the dispersed source array (DSA) concept (Berkhout 2012) outline the results of the proposed workflow.

FRAMEWORK OF THE SURVEY DESIGN WORKFLOW

Using the WRW (W stands for wave propagation and R for reflection) model proposed by Berkhout (1982), we describe the seismic data for a given frequency and recorded by detectors at a depth z_d for sources at a depth z_s as:

$$\mathbf{P}(z_d; z_s) = \mathbf{D}(z_d)\mathbf{X}(z_d, z_s)\mathbf{S}(z_s). \tag{1}$$

Each column and row of the data matrix $\mathbf{P}(z_d, z_s)$ represent a shot and detector gather respectively. For example, the vectors corresponding to the jth shot and ith detector gathers are written as $\mathbf{p}_j$ and $\mathbf{p}_i^\dagger$, where the dagger symbol (†) represents a row vector. A collection of $\mathbf{P}(z_d, z_s)$ for each frequency enables seismic data to be stored as a three-dimensional matrix. $\mathbf{D}(z_d)$ and $\mathbf{S}(z_s)$ are the detector and source matrices respectively. The columns and rows in $\mathbf{D}(z_d)$ represent the spatial coordinates and detector arrays, whereas the columns and rows in $\mathbf{S}(z_s)$ are the source arrays and spatial coordinates respectively. $\mathbf{X}(z_d, z_s)$ is the Earth's transfer function containing the subsurface impulse responses. It can be regarded as unblended data with perfect spatial sampling. The amplitude and phase information for a given frequency is embedded into each element of each matrix. In the following, we assume the sources and detectors to be located at the same depth, i.e. $z_s = z_d = z_0$. For notational simplicity, the detector and source depth indices as well as the superscript of + in the source matrix indicating the direction of wave propagation are hereinafter omitted unless necessary.

With the use of a point detector (rather than a detector array), the distribution of the zero and non-zero elements in $\mathbf{D}$ dictates the spatial sampling of the detectors. The information embedded into the non-zero elements is mainly attributable to the responses of the sensor and recording system used in a given acquisition system. Similarly, with a point source, the distribution of the zero and non-zero elements in $\mathbf{S}$ dictates the spatial sampling of the sources. The information embedded into non-zero elements is mainly attributable to the source signature. Since our primary focus is on the effect of spatial sampling, we make the assumption of delta-functioned detector and source responses, where the elements of $\mathbf{D}$ and $\mathbf{S}$ become either zero or one. Equation (1) therefore depicts the impact of the spatial detector and source sampling on the ability of $\mathbf{P}$ to be a representation of $\mathbf{X}$. In an ideal situation, i.e. carpet detectors and sources, $\mathbf{D}$ and $\mathbf{S}$ are identity matrices. In this case, $\mathbf{P}$ equates to $\mathbf{X}$. However, in practice, imperfections (related to the spatial sampling) in $\mathbf{D}$ and $\mathbf{S}$ hinder $\mathbf{P}$ from resembling $\mathbf{X}$. Therefore, we need to pay proper attention to the distributions of detectors and sources in order for the subsequent processes to be able to retrieve $\mathbf{X}$ from $\mathbf{P}$, which is our goal.

Berkhout (2008) proposed the theoretical framework of source blending by introducing a blending operator, $\mathbf{\Gamma}$:

$$\mathbf{S}' = \mathbf{S}\mathbf{\Gamma}, \tag{2}$$

where $\mathbf{S}'$ represents the blended source matrix. Each column and row of $\mathbf{\Gamma}$ are a shot experiment and location of a source-to-be-blended respectively. Each element of $\mathbf{\Gamma}$ contains the blending code(s) such as the amplitude and phase terms applied to the corresponding source. With *N*-fold blended sources, i.e. *N* sources are blended in one blended experiment leading to one blended shot record, the blending codes of *N* source units are stored in one column of $\mathbf{\Gamma}$. Linearly adding the wavefields of those individual sources forms one blended source wavefield, after detection resulting in one blended shot record. Using the data matrix expression in equation (1) , the results of blended experiments, $\mathbf{P}'$, can be formulated as:

$$\mathbf{P}' = \mathbf{D}\mathbf{X}\mathbf{S}' = \mathbf{D}\mathbf{X}\mathbf{S}\mathbf{\Gamma} = \mathbf{P}\mathbf{\Gamma}. \tag{3}$$

Equation (3) implies that starting with unblended and regularly well-sampled data $\mathbf{X}$, any blended and spatial sampling schemes can be modelled by multiplication with $\mathbf{D}$, $\mathbf{S}$ and $\mathbf{\Gamma}$. As discussed previously, when designing $\mathbf{D}$ and $\mathbf{S}$, our choices are making their elements either zero or one. On the other hand, many choices are available to design $\mathbf{\Gamma}$. In a linear system, one element of the blending operator $\mathbf{\Gamma}$ for the frequency ω is given by:

$$\gamma_{i,j} = a_{i,j}(\omega)\exp\left[-j\phi_{i,j}(\omega)\right], \tag{4}$$

where $a_{i,j}$ is the amplitude term corresponding to the ith source of the jth blended experiment, and $\phi_{i,j}$ is its phase. Any type of source code can be accommodated in equation (4). In the case where the blending codes are time delays $\tau_{i,j}$, the corresponding phase can be written as $\phi_{i,j} = \omega\tau_{i,j}$. Note that the dispersed source array concept, where the total blended wavefield is generated by a plurality of narrowband sources with different centre frequencies, is also described by equations (3) and (4) (albeit implicitly, as the equations are formulated in the frequency domain).

Figure 1 is a schematic illustrating the proposed survey design workflow. The blue- and red-filled steps in Figure 1 correspond, respectively, to the forward process to generate $\mathbf{P}'$ from $\mathbf{X}$ and the inverse process for deblending and data reconstruction to obtain $\langle\mathbf{X}\rangle$ from $\mathbf{P}'$. The angled brackets $\langle$ and $\rangle$ indicate estimated data. Since we deal with survey design, we consider $\mathbf{X}$ to be known in this study. The objective function based on the residue between $\mathbf{X}$ and $\langle\mathbf{X}\rangle$ for a given survey design is defined as:

$$J = \sum_{\omega} \|\mathbf{X} - \langle\mathbf{X}\rangle\|_2^2. \tag{5}$$

If the objective function is sufficiently small, the procedure stops. If not, genetic algorithms (see the green-filled step in Fig. 1) update the blending and sampling operators that are subsequently fed into the next iteration. Therefore, our approach iteratively optimizes $\mathbf{D}$, $\mathbf{S}$ and $\mathbf{\Gamma}$ to minimize the objective function in equation (5), meaning that optimum deblending and reconstruction quality is attainable using the resultant survey design.

DEBLENDING AND DATA RECONSTRUCTION

In the deblending and data reconstruction step (see the red box in Fig. 1), we go from $\mathbf{P}'$ to $\langle\mathbf{X}\rangle$. Obviously, in this step, we do not make use of the fact that $\mathbf{X}$ is known. We use the given $\mathbf{D}$, $\mathbf{S}$ and $\mathbf{\Gamma}$ derived from the overall scheme. According to equation (3), a trivial solution to obtain $\langle\mathbf{X}\rangle$ is the use of the generalized pseudo-inverse matrix as:

$$\langle\mathbf{X}\rangle = \mathbf{D}^H(\mathbf{D}\mathbf{D}^H)^{-1}\mathbf{P}'(\mathbf{S}'^H\mathbf{S}')^{-1}\mathbf{S}'^H, \tag{6}$$

where the superscript H denotes the Hermitian conjugate. However, this particular solution of the underdetermined system often leads to an undesired result as blending noise and missing traces still remain in the data. Mahdad *et al.* (2011) introduced sparseness and coherency to an iterative deblending scheme to constrain the solution. Ishiyama *et al.* (2017) then extended it to a generalized deblending and data reconstruction workflow as follows. A least-squares criterion in combination with sparseness and coherency constraints is used to minimize the following objective function:

$$J = \sum_{\omega} \left\| \Delta\mathbf{P}' \right\|_2^2 = \sum_{\omega} \left\| \mathbf{P}' - \mathbf{D}\langle\mathbf{X}\rangle\mathbf{S}' \right\|_2^2. \tag{7}$$

Figure 2 shows a closed-loop system to solve the inverse problem. The inversion is designed to find the optimum parameters in the model space, in our case, the Fourier domain. With the transform operator, $\mathbf{L}$, we can write the forward and inverse transformations as:

$$\mathbf{L}\mathbf{X} = \mathbf{M}, \tag{8}$$

and

$$\mathbf{X} = \mathbf{L}^H\mathbf{M}. \tag{9}$$

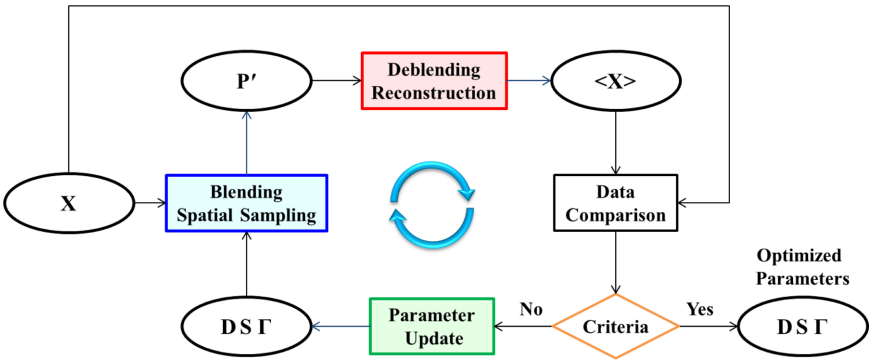

Figure 1 Schematic illustrating the proposed survey design workflow. The blue-filled step is forward modelling to produce blended and irregularly sampled data (going $\mathbf{P}'$ from $\mathbf{X}$ by applying $\mathbf{D}, \mathbf{S}$ and $\boldsymbol{\Gamma}$). The red-filled step indicates deblending and data reconstruction (going $\langle\mathbf{X}\rangle$ from $\mathbf{P}'$). A termination criterion is based on the residue between $\mathbf{X}$ and $\langle\mathbf{X}\rangle$. If it is sufficiently small, the procedure stops and produces output from the last iteration. If not, the green-filled step indicating GAs updates the blending and sampling operators.

A successful estimate of the model parameters, $\langle\mathbf{M}\rangle$, provides an estimate of the deblended and reconstructed data according to:

$$\langle\mathbf{X}\rangle = \mathbf{L}^H \langle\mathbf{M}\rangle . \tag{10}$$

The inversion scheme also contains the forward modelling step, the multiplication of $\langle\mathbf{X}\rangle$ with the known operators, $\mathbf{D}$ and $\mathbf{S}'$. This allows for a direct comparison between the observed data, $\mathbf{P}'$, and the estimated data, $\mathbf{D}\langle\mathbf{X}\rangle\mathbf{S}'$. As a consequence, an estimate of $\langle\mathbf{X}\rangle$ from $\mathbf{P}'$ can be obtained through minimizing the objective function in equation (7).

The estimated data from the $(i+1)$th update is obtained via a gradient descent scheme as:

$$\langle\mathbf{X}\rangle_{i+1} = \left\langle \underline{\mathbf{X}} \right\rangle_i + \alpha_i \Delta \left\langle \underline{\mathbf{X}} \right\rangle_i , \tag{11}$$

where

$$\Delta \left\langle \underline{\mathbf{X}} \right\rangle_i = \mathbf{D}^H (\mathbf{D}\mathbf{D}^H)^{-1} \Delta \underline{\mathbf{P}}'_i (\mathbf{S}'^H \mathbf{S}')^{-1} \mathbf{S}'^H , \tag{12}$$

with

$$\Delta \underline{\mathbf{P}}'_i = \mathbf{P}' - \mathbf{D} \left\langle \underline{\mathbf{X}} \right\rangle_i \mathbf{S}' . \tag{13}$$

Here, $\left\langle \underline{\mathbf{X}} \right\rangle_i$ represents a deblended and reconstructed estimate with sparseness and coherency constraints on the ith update, $\langle\mathbf{X}\rangle_i$. α_i is a scaling factor for the update that acts as a step size to minimize the residual by satisfying:

$$\sum_{\omega} \left\| \Delta \underline{\mathbf{P}}'_i - \alpha_i \mathbf{A}_i \right\|_2^2 \rightarrow \min, \tag{14}$$

with

$$\mathbf{A}_i = \mathbf{D} \Delta \left\langle \underline{\mathbf{X}} \right\rangle_i \mathbf{S}' . \tag{15}$$

It is computed as follows. We write the objective function to be solved, $J(\Delta \underline{\mathbf{P}}'_i - \alpha_i \mathbf{A}_i)$, as:

$$\begin{aligned} J\left(\Delta \underline{\mathbf{P}}'_i - \alpha_i \mathbf{A}_i\right) &= \sum_{\omega} \mathrm{tr}\left[\left(\Delta \underline{\mathbf{P}}'_i - \alpha_i \mathbf{A}_i\right)^H \left(\Delta \underline{\mathbf{P}}'_i - \alpha_i \mathbf{A}_i\right) \right] \\ &= \sum_{\omega} \mathrm{tr}\left[\Delta \underline{\mathbf{P}}'^H_i \Delta \underline{\mathbf{P}}'_i - \alpha_i \Delta \underline{\mathbf{P}}'^H_i \mathbf{A}_i - \alpha_i \mathbf{A}_i^H \Delta \underline{\mathbf{P}}'_i + \alpha_i^2 \mathbf{A}_i^H \mathbf{A}_i \right], \end{aligned} \tag{16}$$

where tr[ ] indicates the sum of the diagonal elements of a matrix. The partial derivative of $J(\Delta \underline{\mathbf{P}}'_i - \alpha_i \mathbf{A}_i)$ with respect to α_i is:

$$\frac{\partial}{\partial \alpha_i} J\left(\Delta \underline{\mathbf{P}}'_i - \alpha_i \mathbf{A}_i\right) = \sum_{\omega} \mathrm{tr}\left[-\Delta \underline{\mathbf{P}}'^H_i \mathbf{A}_i - \mathbf{A}_i^H \Delta \underline{\mathbf{P}}'_i + 2\alpha_i \mathbf{A}_i^H \mathbf{A}_i \right]. \tag{17}$$

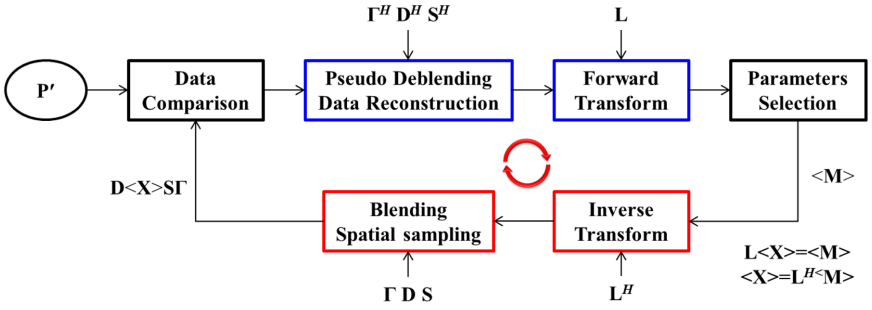

Figure 2 Step-by-step illustration of closed-loop deblending and data reconstruction (after Ishiyama *et al.* 2017).

Therefore, the scalar factor that makes equation (17) zero is given by:

$$\alpha_i = \frac{\sum\limits_{\omega} \mathrm{tr}\left[\Delta\underline{\mathbf{P}'}_i{}^H \mathbf{A}_i + \mathbf{A}_i^H \Delta\underline{\mathbf{P}'}_i\right]}{2\sum\limits_{\omega} \mathrm{tr}\left[\mathbf{A}_i^H \mathbf{A}_i\right]} = \frac{\sum\limits_{\omega} \mathrm{tr}\left[\Delta\underline{\mathbf{P}'}_i{}^H \mathbf{A}_i\right]}{\sum\limits_{\omega} \mathrm{tr}\left[\mathbf{A}_i^H \mathbf{A}_i\right]}. \tag{18}$$

An iterative estimate of the deblended and reconstructed data consists of: muting based on causality, combined median filtering and filtering in the wavenumber–frequency domain to pass coherent events while suppressing undesired incoherent ones and amplitude thresholding in time–space domain. Since our data are stored as a three-dimensional (3D) matrix, as described previously, these processes are performed with a 3D implementation. The parameters for the filters and threshold level are updated through the course of the iterations, starting from harsh and then moving to mild ones.

SURVEY DESIGN WITH GENETIC ALGORITHMS

When designing a survey or assessing a given acquisition geometry, common midpoint-based attributes such as fold and sampling at different offset as well as azimuth ranges are widely used, as illustrated in Cordsen, Galbraith and Peirce (2000) and Vermeer (2012). Although they can quickly provide beneficial information on the anticipated data quality from a given acquisition geometry, these attributes inherently disregard the effect of processing. With the application of blending along with irregular geometries, the quality of deblended and reconstructed data is inevitably a major concern; thus, one eventually needs another means. One solution would be to build a field-wide earth model with finer grids followed by the forward modelling of any anticipated acquisition scenarios (e.g. Regone 2007). This certainly requires enormous amounts of computer resources, funds and time, which may not be always allocated to every occasion.

In this study, we implement genetic algorithms (GAs), which allow us to optimize the blending and sampling operators simultaneously rather than sequentially. GAs are classified as a metaheuristic and are generally capable of handling optimization problems with non-convexity, the existence of many local minima, non-differentiability and a large problem space. GAs are inspired by biological evolution through the process of natural selection. It was first introduced in 'On The Origin of Species' by Darwin (1859), which describes the biological development of species and survival of minor advantageous mutations. Holland (1975) originated the concept of GAs and demonstrated how the theory of evolution can be exploited for optimization problems based on binary string representations. Such strings are considered as biological chromosomes, and evolution processes are described in the natural selection such as mutation, selection and crossover. Over several decades, the original definition of GAs has gradually evolved, and the technology has been widely adapted to a variety of optimization problems. Numerous successful applications of GAs are easily recognizable in different domains such as biomedicine (Monteagudo and Reyes 2015), arts (Davies *et al.* 2016), architecture (Bak, Rask and Risi 2016), music (Scirea *et al.* 2016), games (Liebana *et al.* 2015) and recently machine learning (Kramer 2016).

As discussed previously, Fig. 1 illustrates the proposed survey design workflow. The step in the green box corresponds to GAs that iteratively update the blending and sampling operators that constitute of a set of parameter vectors:

$$\mathbf{x}_{i,j} = \left[\mathbf{d}_{i,j}, \mathbf{s}_{i,j}, \boldsymbol{\gamma}_{i,j}\right]^T, \tag{19}$$

with

$$\mathbf{d}_{i,j} \in \{0,1\}^k, \quad \mathbf{s}_{i,j} \in \{0,1\}^l, \quad \boldsymbol{\gamma}_{i,j} \in \{0,1\}^{l\times m}. \tag{20}$$

Here, i and j represent the numbers of populations and generations, and $\mathbf{d}_{i,j}$, $\mathbf{s}_{i,j}$ and $\boldsymbol{\gamma}_{i,j}$ are binary vectors indicating detector sampling, source sampling and blending operators for the ith individual solution in the jth generation that we aim to update through stochastic operators in GAs. k, l and m dictate the dimensions of the parameter vectors. k and l are attributed to the numbers of detectors and sources to be designed for the geometry under consideration. m then equates to the required bit length to parametrize a given blending code per single source. The following steps describe the implementation of the technique to our survey design workflow.

1 Initialization: A set of parameter vectors, $\mathbf{x}_{i,0}$, called the initial population having n individuals is randomly generated across a given problem space as:

$$\begin{pmatrix} \mathbf{x}_{1,0} \\ \mathbf{x}_{2,0} \\ \vdots \\ \mathbf{x}_{n,0} \end{pmatrix} = \begin{pmatrix} \left[\mathbf{d}_{1,0}, \mathbf{s}_{1,0}, \boldsymbol{\gamma}_{1,0}\right]^T \\ \left[\mathbf{d}_{2,0}, \mathbf{s}_{2,0}, \boldsymbol{\gamma}_{2,0}\right]^T \\ \vdots \\ \left[\mathbf{d}_{n,0}, \mathbf{s}_{n,0}, \boldsymbol{\gamma}_{n,0}\right]^T \end{pmatrix}. \tag{21}$$

2 Data comparison: Once a population in generation j is created, it goes to the forward (going from $\mathbf{X}$ to $\mathbf{P}'_{i,j}$) and inverse (going from $\mathbf{P}'_{i,j}$ to $\langle\mathbf{X}_{i,j}\rangle$) processes, as described in the previous sections. Once $\langle\mathbf{X}_{i,j}\rangle$ is obtained using a parameter vector $\mathbf{x}_{i,j}$, the objective function in equation (5) for the ith

solution of the jth generation, equivalent to the so-called fitness function, can be written as:

$$J_{i,j}(\mathbf{x}_{i,j}) = \sum_{\omega} \left\| \mathbf{X} - \langle \mathbf{X}_{i,j} \rangle \right\|_2^2. \tag{22}$$

On the basis of the residual in equation (22), **D**, **S** and **Γ** are updated by the following stochastic operators.

3 Selection: Selection allocates more copies of solutions with smaller misfits than solutions with larger misfits, and thus imposes the principle of natural selection on the candidate solutions. Although a number of selection procedures have been proposed, the main purpose of selection is to give more weight to better solutions than to worse ones. Our workflow utilizes roulette-wheel selection. For the jth generation, each individual solution possesses an expected selection probability according to its objective function given by:

$$G(\mathbf{x}_i) = j_i / \sum_{i=1}^{n} j_i, \tag{23}$$

with

$$j_i = \exp(\beta J_i / \min_{i \in n} J_i), \tag{24}$$

where β is a dimensionless parameter controlling the diversity in the selection. A smaller value of β results in more diversity in selection, whereas a larger value of β provides a chance that solutions having a small misfit are selected.

4 Crossover: Crossover combines the information of two parental solutions to create new, possibly better solutions. There are various ways to accomplish this. However, the principle of this process is that a new solution has to combine parental information in a predefined manner. Our parameter sequence contains different parameter vectors attributable to the blending and sampling operators. Crossover among different vectors is not viable as each employs different lengths and constraints. Thus, we apply a single crossover per parameter vector.

5 Mutation: While crossover operates on two parental solutions, mutation, locally yet randomly, modifies a single solution. Although there are many variations of mutation, it commonly involves one or more changes that are made to an individual solution. In the case where mutation involves a subtle change, refinement in the vicinity of a candidate solution is performed. On the other hand, a process with a larger change potentially has the ability to escape from local minima. We control this process using two parameters. One is the probability of mutants to appear in a given population. The other one is the mutation rate that describes the number of values to be altered within a single solution.

Step 1	Main code: 01001011011100101010
Step 2	Base codes 1&2: 0100101101&1100101010
Step 3	Base codes 3&4: 1011010010&0101010011
Step 4	Primary source: 1342432112314 -----
	Secondary source: 4213112343241 -----

Figure 3 Steps to generate an RES. Twenty binary numbers exemplify a way to form long parameter vectors from a main code. Four different colours distinguish four base codes.

6 New population: The population created by selection, crossover and mutation replaces the original parental population on the basis of elitist replacement. Since the new individuals do not necessarily provide better results, the technique allows us to preserve some better ones from the parental population in a new generation.

7 By repeating steps 2–6, **D**, **S** and **Γ** are iteratively updated to minimize the objective function in equation (22) such that effective deblending and data reconstruction can be realized. Within the stochastic operators, the constraints on the blending and sampling operators can be imposed in order to avoid the generation of undesired solutions.

IMPLEMENTATION OF A REPEATED ENCODING SEQUENCE

Although genetic algorithms (GAs) generally have the ability to handle the large problem size, survey design inherently provides a significant number of solutions. This is more obvious when dealing with the parameters involved in irregularity, which is unfortunately true in our case. In addition, k and l in equation (20) increase with the size of a survey, resulting in parameter vectors that are extremely long. Hence, a reduction of parameter space makes the proposed survey design workflow practical and computationally affordable. In this respect, we propose a repeated encoding sequence (RES) inspired by a nucleic acid sequence of deoxyribonucleic acid (DNA) consisting of a chain of four nucleobases: adenine (A), cytosine (C), guanine (G) and thymine (T). We make use of this as an analogy to form a parameter sequence in GAs to reduce the problem space. Figure 3 exemplifies a way to generate an RES using 20 binary numbers.

We first create a main code having a random-like feature (step 1). The prior constraints are embedded into this main code. It is then divided into two halves to make two base codes (step 2). These codes are flipped to create two more base codes (step 3). The four base codes obtained in this way are finally

combined in a predetermined order so that even a long parameter sequence can be formed, like a chain of four nucleobases in DNA (step 4). The order has to be predetermined to ensure the reproducibility of the solution.

Additionally, it is well known that DNA has a double helix structure in which one nucleobase bonds only with one other specific nucleobase. This is referred to as base pairs where A bonds only with T, while C bonds only with G. We also use this analogy for blended acquisition. We predefine base pairs. For example, as illustrated in step 4 in Fig. 3, base code 1 bonds with base code 4, while base code 2 bonds only with base code 3. When blending, we first define a parameter sequence of a primary source. Then, the parameters of a secondary source that blends with the primary one are automatically defined according to the concept of the base pairs. This ensures that sources that overlap each other hold different properties, leading to effective deblending, e.g. by applying different spectral properties to overlapped sources.

With an RES, the optimization deals with a single main code rather than with a whole parameter sequence. Therefore, k and l in equation (20) become sufficiently manageable, which enables a significant reduction in the parameter space. Although the four base codes are repeated, each one possesses a random-like feature and also appears in an irregular manner based on a predetermined order. Therefore, the resulting survey parameters still employ the property of irregularity, which enhances the effectiveness of deblending and data reconstruction. Figure 4 shows a comparison of shot records having different spatial sampling in the time–space and frequency–wavenumber domains. Figure 4(a,f) depict well-sampled data that contain no spatial aliasing. The other three cases employ 50% detector decimation, keeping one out of two detectors. Because of this, they show undersampling-related artefacts. Different spatial sampling schemes result in different imprints in the frequency–wavenumber domain. Regularly sampled data show coherent artefacts due to periodically missing traces (Fig. 4b,f). On the contrary, randomly sampled data show Gaussian-noise-like aliasing artefacts (Fig. 4c,g). Irregularly sampled data with an RES notably exhibit a similar characteristic that consequently allows deblending and reconstruction to effectively estimate the desired signals while clearly separating them from undesired events (Fig. 4d,h).

NUMERICAL EXAMPLES

A subset of synthesized two-dimensional Marmousi data is used to numerically simulate several acquisition scenarios incorporating the DSA concept. Berkhout (2012) introduced a dispersed source array (DSA) that utilizes a set of source units, each having a dedicated narrow frequency range. Caporal and Blacquière (2015) discussed its benefits from different perspectives, some of which are reviewed as follows. A DSA permits each narrowband source to be independently deployed to satisfy its own spatial sampling criteria determined by its frequency range. This subsequently allows for proper sampling of entire frequency ranges and addresses both the oversampling of lower frequencies and undersampling of higher frequencies. In a marine survey, the method minimizes destructive interference from a source ghost in the spectrum using so-called ghost matching. By towing narrowband sources at different and proper depth levels, source notches insignificantly overlap with the frequencies emitted from these sources.

Table 1 Properties of four DSA sources

	Frequency Range (Hz)	Spatial Distribution (%)
Source type 1	2–4–6–10	6.7
Source type 2	4–8–12–20	13.3
Source type 3	8–16–24–40	26.7
Source type 4	16–32–48–80	53.3

Notes: Four corner frequencies, (1) low-cut, (2) low-pass, (3) high-pass and (4) high-cut, describe the frequency range of each source type. The spatial distribution of each source type is expressed by percentage, e.g. 25% means keeping one out of four shots.

Table 2 Acquisition configurations for unblended and well-sampled data (**X**) and blended and irregularly sampled data (**P′**)

	X	**P'**
Detector interval	10 m at regular	irregular
The number of detectors	120	96 (20% decimation)
Source interval	10 m at regular	irregular (see Table 1)
The number of sources	120	(see Table 1)
The number of activated sources	1	2 with 600 m separation

Note: Parameters related to source sampling vary with four DSA source types as shown in Table 1.

Table 1 summarizes the spectral properties and spatial distribution of four DSA source types used in this study. Four corner frequencies: (1) low-cut, (2) low-pass, (3) high-pass and (4) high-cut, describe the frequency range of each source type. The spatial distribution of a source type is expressed by percentage, e.g. 25% means keeping one out of four shots. Table 2 shows a comparison of acquisition configurations between unblended and well-sampled data and

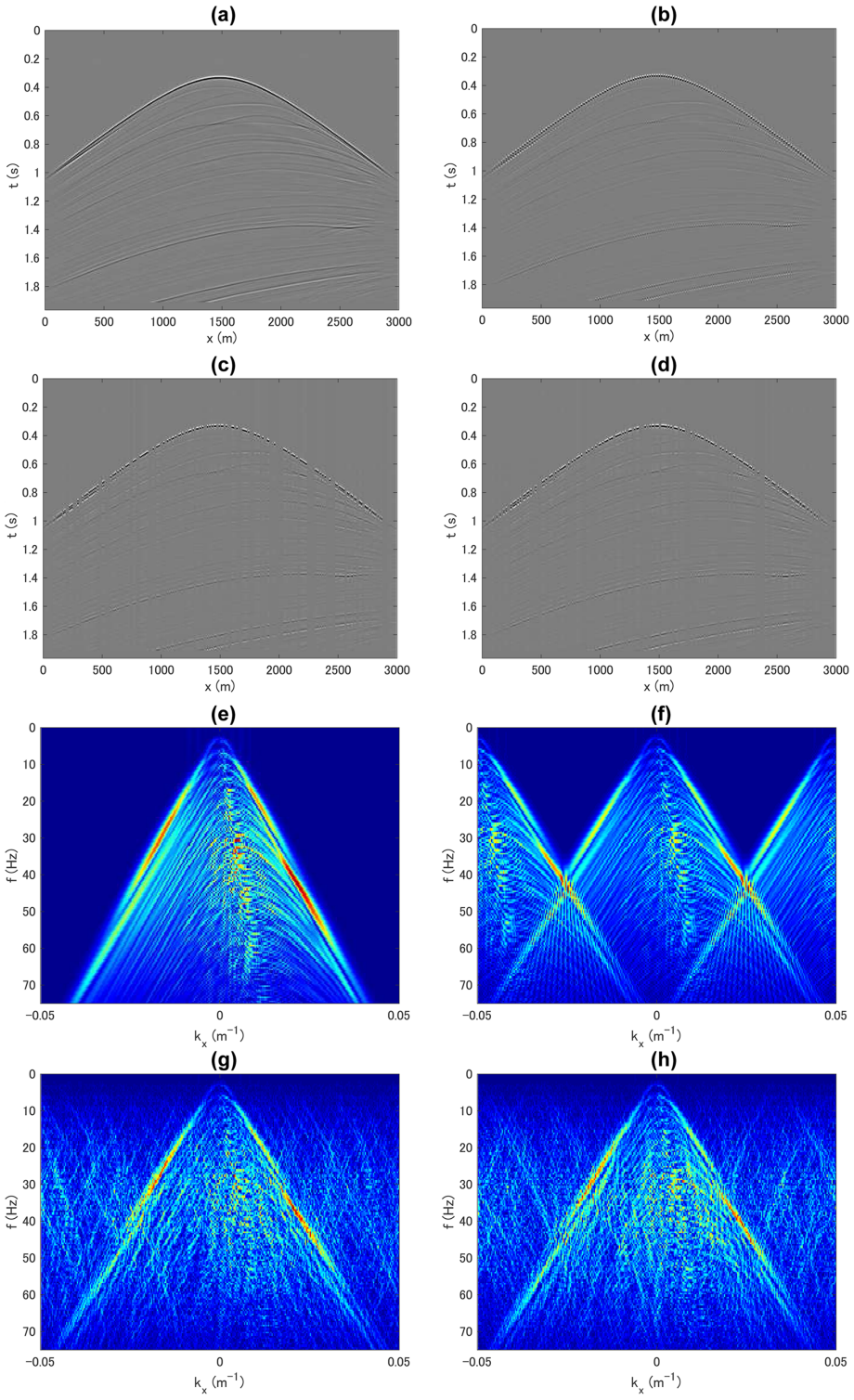

Figure 4 Shot gathers in time–space and frequency–wavenumber domains with different spatial sampling: well-sampled (a and e), regularly sampled (b and f); randomly sampled (c and g); irregularly sampled with an RES (d and h). Each undersampled case employs 50% detector decimation, keeping one out of two detectors.

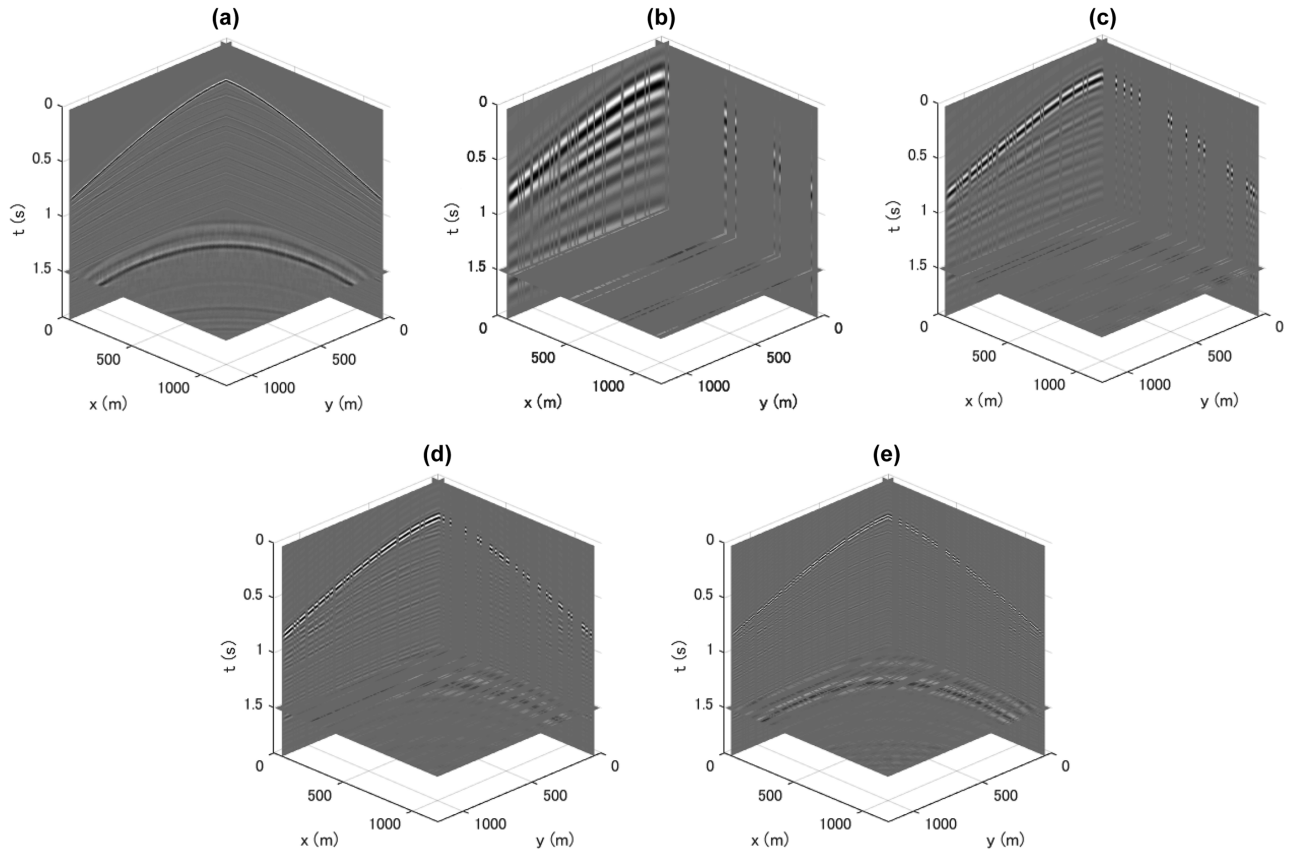

Figure 5 3D view of a subset of synthesized Marmousi data: (a) well-sampled and broadband data; (b) randomly sampled and narrowband data from source type 1; (c) source type 2; (d) source type 3; and (e) source type 4. A section parallel to x direction is a common detector gather while one parallel to y direction is a common shot gather. Spectral properties and spatial distribution of four DSA source types are shown in Table 1.

blended and irregularly-sampled data. In addition to four DSA sources, 20% detector decimation, keeping four out of five detectors, is applied to each scenario. Figure 5(a) shows a three-dimensional (3D) view of data matrix from unblended and well-sampled data containing a broad frequency range. It consists of 120 detectors and 120 sources with a sampling interval of 10 m, located at the surface, i.e. $z_s = z_d = 0$. We use these data as an input for our numerical examples and as a reference to analyse the deblending and reconstruction results. Figure 5(b– e) are 3D views of data matrices from unblended and irregularly sampled data. They are acquired by four source units, each having its own bandwidth and spatial sampling criteria, as previously defined. The sampling operators of the detectors and sources are derived from the use of a random realization according to spatial sampling schemes specified in Tables 1 and 2. In this study, the blending scheme has the same blending performance indicator (BPI) (Berkhout and Blacquière 2014) having a value of two. Each blended shot record employs 600 m spatial separation between two sources.

The proposed approach tries to find survey designs allowing for the satisfactory retrieval of broadband, deblended and reconstructed data from a set of narrowband records acquired in a blended manner along with irregular geometries. For comparison purposes, we also show the results with blending and sampling operators created by a realization from random variables according to a discrete uniform distribution. Instead of a single realization, we use parameters that provide a median objective function value among 500 random realizations for each case. We assume that the median value can represent an anticipated situation where we rely on a random realization to embed irregularity into blending and sampling operators. We use 'random' to describe this type of data. In terms of computation time, the optimized design and random design are not comparable. However, this comparison is still worthwhile to differentiate between the proposed approach and one of common practices to implement irregularity and randomness to survey parameters. We apply 100 iterations for the previously described deblending and reconstruction. Additionally, each

blended data set employs a time shift ranging from 0 ms to 256 ms.

To quantify the performance of deblending, Mahdad *et al.* (2011) utilized the signal-to-noise ratio (SNR) expressed as:

$$\mathrm{SNR} = 10\log_{10}\left(\frac{\sum\limits_{\omega}\|\mathbf{X}\|_2^2}{\sum\limits_{\omega}\|\mathbf{X}-\langle\mathbf{X}\rangle\|_2^2}\right). \tag{25}$$

This SNR definition allows for a direct comparison with the desired output, which may not be suitable for some situations. However, the principle of acquisition design permits $\mathbf{X}$ to be known; thus, this quantitative measurement is applicable to our case.

Example 1: applicability of the approach

Three blended dispersed source array (DSA) scenarios are numerically simulated: (1) DSA with encoded time delay, (2) DSA with encoded time delay and sweep and (3) DSA with encoded time delay and partial sweep (left column in Figs 6–8). Although data with encoded sweep signatures employ a longer length in time, they are clipped for display purposes. The first and second scenarios are purposely designed to resemble acquisition in marine and land environments. The third scenario corresponds to a situation with a mixture of explosive and vibrator sources that can possibly occur in transition zone acquisition. Alternatively, these scenarios can be considered as situations that utilize different types of seismic sources for a given environment. All the scenarios employ the same blending performance indicator (BPI) of two along with the same numbers of detectors and sources. Each blended shot record employs 600 m spatial separation between two sources (Table 2). The optimized design using the proposed method is compared to one with randomly designed blending and sampling operators for each scenario. Our approach optimizes the spatial distribution of detectors and four DSA source units as well as the encoded signatures applied to each source.

Although both of the optimized and random designs have the same BPI and the same numbers of detectors and sources, the optimized ones achieve better continuity and less blending noise regardless of the acquisition scenario (middle column in Figs 6–8). They present a notable difference attributable solely to the way that the blending and sampling operators are designed. The resultant SNR values are as follows: scenario 1) 8.36 dB with random and 10.92 dB with optimized, 2) 8.74 dB with random and 11.07 dB with optimized and 3) 8.83 dB with random and 11.22 dB with optimized. The difference plots evidently reflect these differences, where smaller errors are easily recognizable in the optimized designs (right column in Figs 6–8). Mahdad *et al.* (2011) described a deblended record with an SNR of approximately 12 dB as the result close to the desired output. Although the experimental settings are different, the level of this value can still be used as a reference to quantitatively assess the resultant $\langle\mathbf{X}\rangle$ with respect to $\mathbf{X}$. Since our cases involve both the blending of source wavefields and the use of fewer detectors and sources unlike their study, the deblended and reconstructed data with randomly designed operators obtain a suboptimal quality. On the other hand, the optimized designs are almost comparable to the level of desired quality. The results clearly indicate that the choice of blending and sampling operators is one of the factors determining the quality of deblending and reconstruction. They also confirm the viability of our approach for deriving the blending and sampling operators, leading to the improvement of these processes for different blending and spatial sampling schemes. This then implies the wide applicability of the proposed method to acquisition in various situations.

Example 2: extensibility of the approach

In this example, blended dispersed source array (DSA) scenarios are numerically simulated using eight different subsets within the Marmosi model. We first optimize the blending and sampling operators of the detectors and four DSA source units in a given subset called model 1. The optimized operators are then applied to other subsets within the Marmosi model, called models 2–8 as:

$$\begin{pmatrix}\mathbf{P}_1'\\ \mathbf{P}_2{}'\\ \vdots\\ \mathbf{P}_8{}'\end{pmatrix} = \begin{pmatrix}\mathbf{D}_1\mathbf{X}_1\mathbf{S}_1\boldsymbol{\Gamma}_1\\ \mathbf{D}_1\mathbf{X}_2\mathbf{S}_1\boldsymbol{\Gamma}_1\\ \vdots\\ \mathbf{D}_1\mathbf{X}_8\mathbf{S}_1\boldsymbol{\Gamma}_1\end{pmatrix}. \tag{26}$$

These data are compared to those with randomly designed operators. Each data set has the same blending performance indicator and the same numbers of detectors and sources. Figure 9 shows a comparison of the signal-to-noise ratio (SNR) for each model. Although the operators are established using model 1, they reasonably achieve a notable enhancement in the SNR in all cases. Figures 10–12 show comparisons between data with randomly designed operators and those with optimized operators from models 1, 3 and 6. They clearly illustrate that the optimized blending and sampling operators from model 1 provide a significant uplift on the deblending and reconstruction quality in other models. The

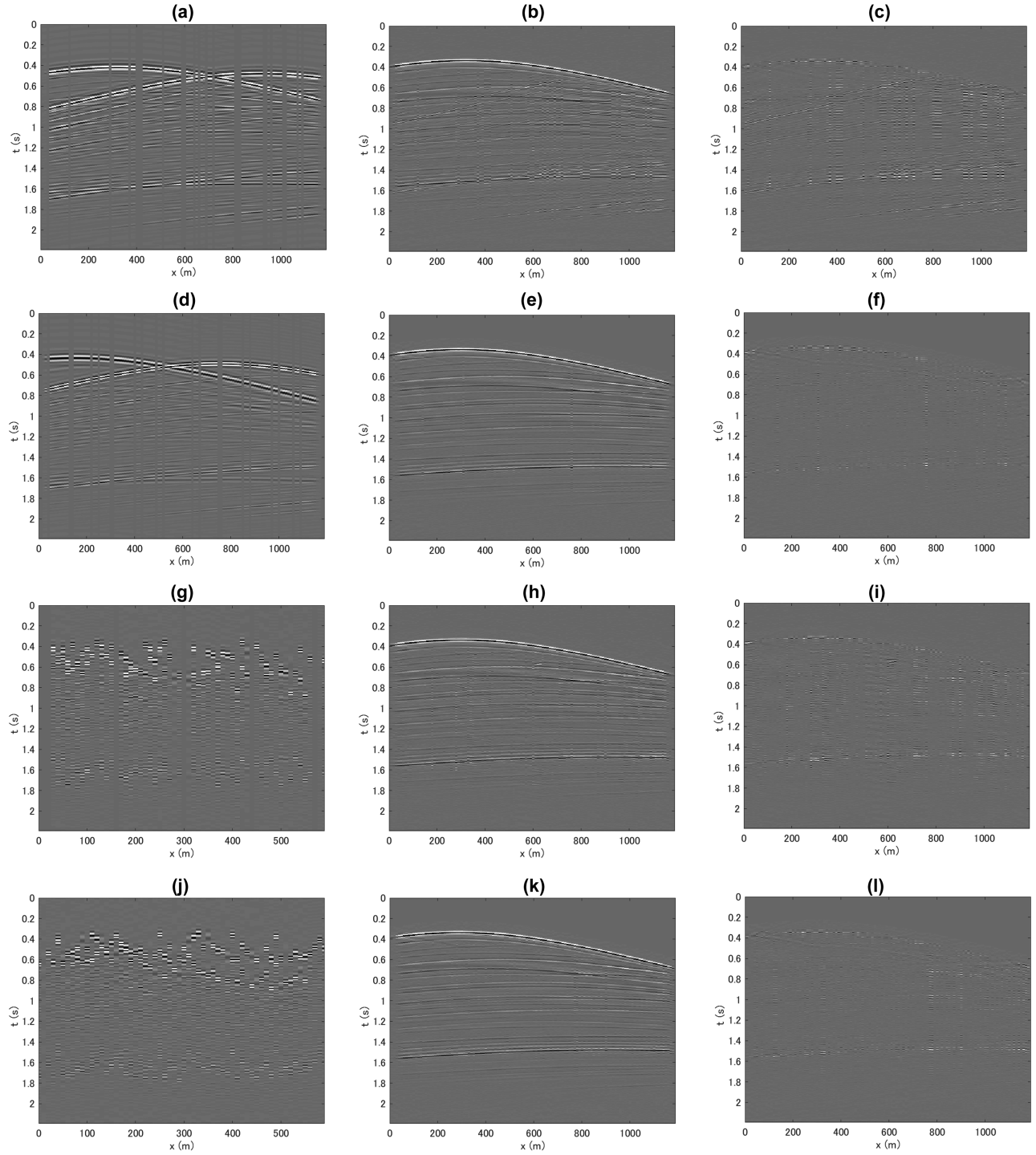

Figure 6 Optimized survey design in scenario 1 (DSA with time delay). Three columns differentiate: blended and irregularly sampled data; deblended and reconstructed data; and difference between input (unblended and well-sampled data) and output (deblended and reconstructed data). Data with randomly designed operators are shown in the odd rows, whereas ones with optimized operators are in the even rows. The top two rows are data in common shot domain, while the bottom two rows are ones in common detector domain. Optimized operators achieve an SNR of 10.92 dB, whereas randomly designed operators obtain an SNR of 8.36 dB.

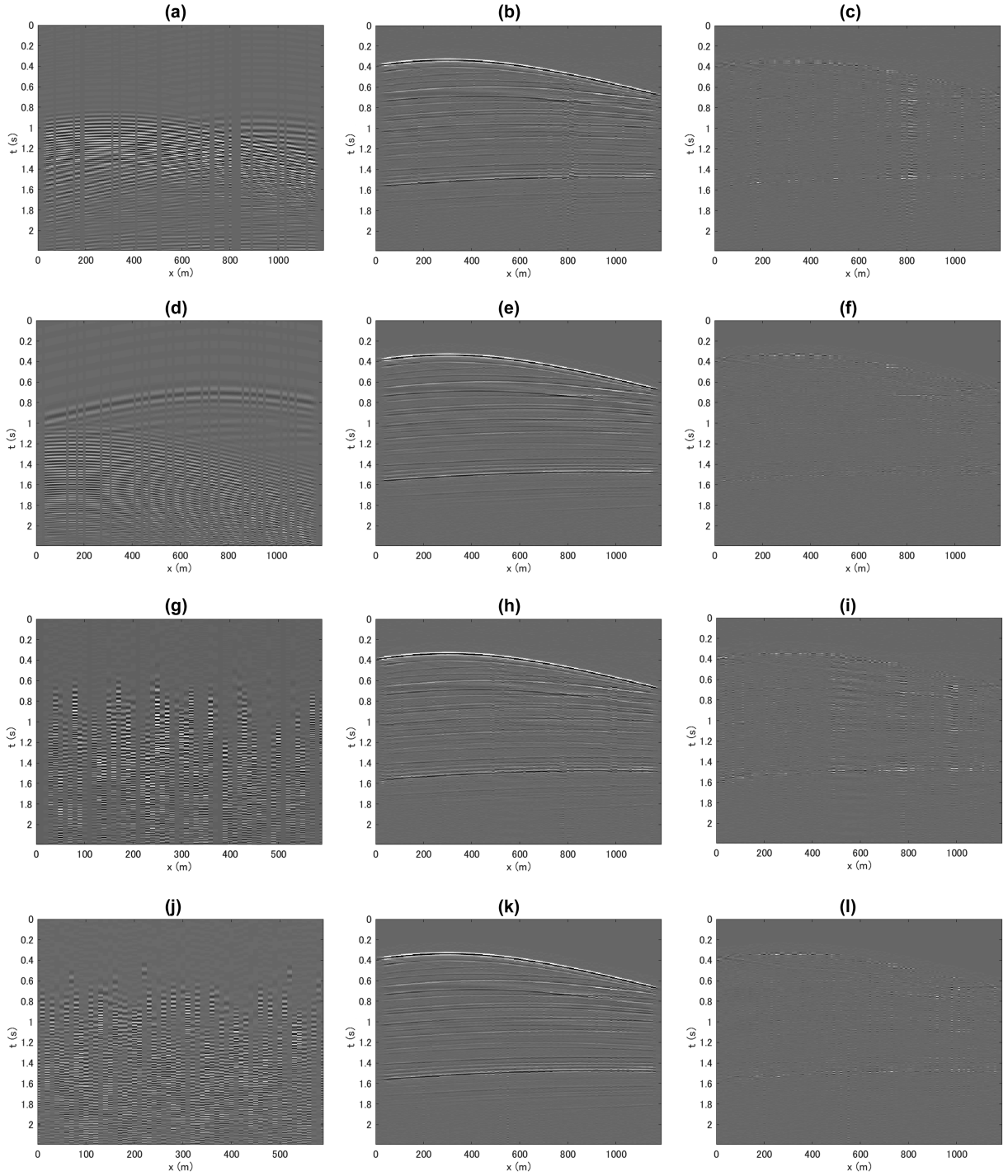

Figure 7 Optimized survey design in scenario 2 (DSA with time delay and sweep). Three columns differentiate: blended and irregularly sampled data; deblended and reconstructed data; and difference between input (unblended and well-sampled data) and output (deblended and reconstructed data). Data with randomly designed operators are shown in the odd rows, whereas ones with optimized operators are in the even rows. The top two rows are data in common shot domain, while the bottom two rows are ones in common detector domain. Optimized operators achieve an SNR of 11.07 dB, whereas randomly designed operators obtain an SNR of 8.74 dB.

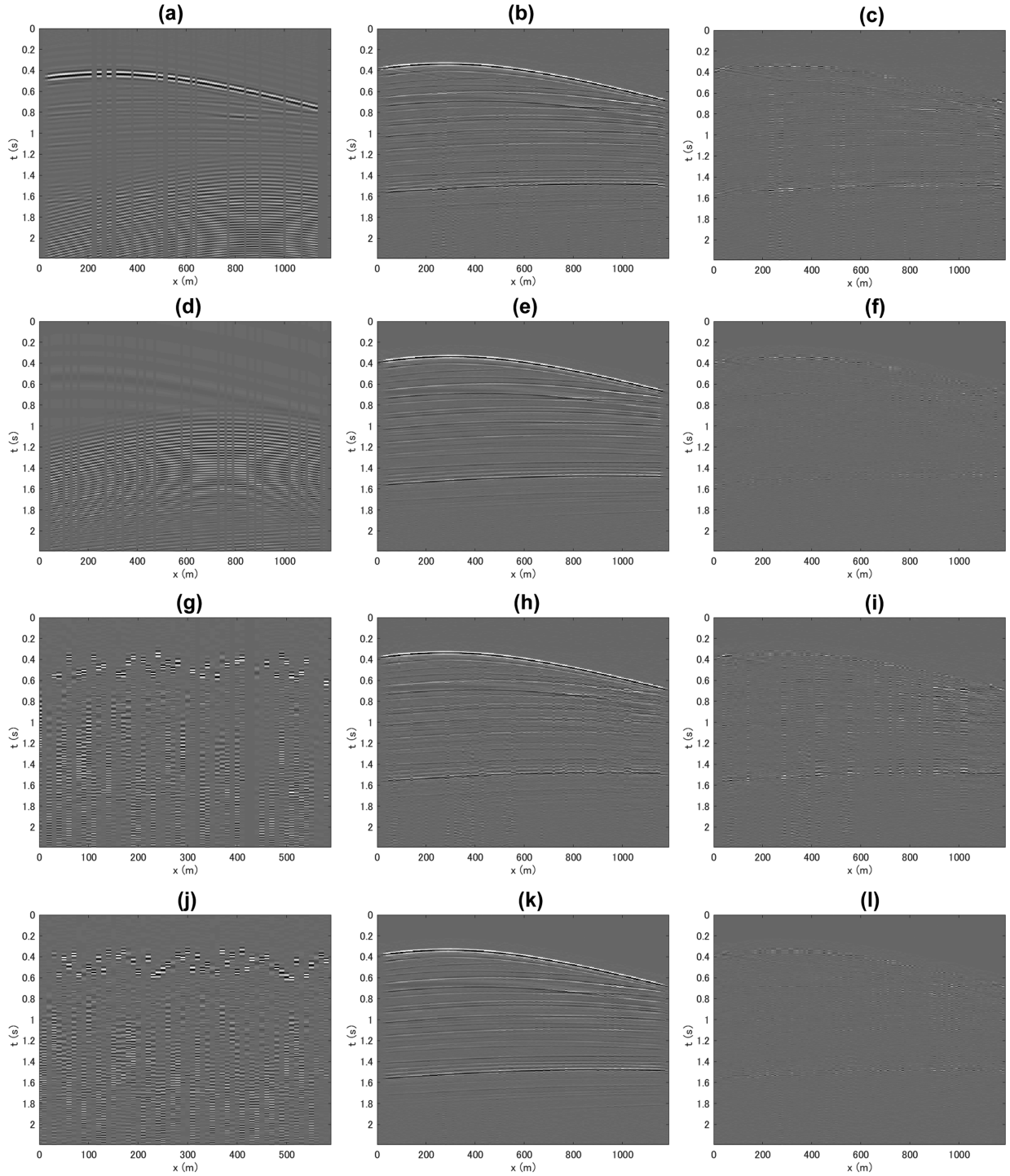

Figure 8 Optimized survey design in scenario 3 (DSA with time delay and partially sweep). Three columns differentiate: blended and irregularly sampled data; deblended and reconstructed data; and difference between input (unblended and well-sampled data) and output (deblended and reconstructed data). Data with randomly designed operators are shown in the odd rows, whereas ones with optimized operators are in the even rows. The top two rows are data in common shot domain, while the bottom two rows are ones in common detector domain. Optimized operators achieve an SNR of 11.22 dB, whereas randomly designed operators obtain an SNR of 8.83 dB.

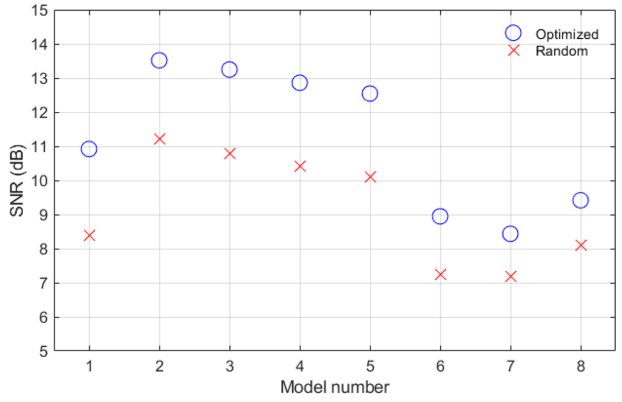

Figure 9 SNR values from different models. Blue circles indicate SNRs from optimized designs, while red crosses indicate SNRs from random designs. For results represented by blue circles, sampling and blending operators are established with model 1 and then applied to other models.

difference in the SNR among models is presumably attributable to the subsurface complexity that varies from one model to another. For instance, models 6–8 tend to have several crossing events that likely make deblending and reconstruction more challenging. On the other hand, few events are recognizable in models 2–5. Nevertheless, designs with optimized operators show better deblending and reconstruction results for all models. This suggests that optimized operators using our approach possibly enhance these processes for an area having similar subsurface responses. This implies that optimization at certain locations representing the area of interest sufficiently provides effective operators applicable to the entire field without the need for the location-by-location update of the survey parameters. This example hence highlights the potential extensibility of the method towards field-wide implementation.

Example 3: viability of the approach

Mosher, Kaplan and Janiszewski (2012) proposed a way to design a survey geometry allowing for the optimum data reconstruction of irregularly sampled data to implement the concept of compressive sensing. Their approach utilizes Monte Carlo simulation and subsequently selects a survey geometry showing the best data recovery among several hundreds of realizations. In this respect, we make a comparison among four different optimization schemes: (1) genetic algorithms (GAs) with a repeated encoding sequence (RES), (2) GAs without an RES, (3) Monte Carlo simulation with an RES and (4) Monte Carlo simulation without an RES. For a fair comparison in terms of computation time, we apply the same number of realizations, 800.

Blended dispersed source array data are numerically simulated using the same blending performance indicator and the same numbers of detectors and sources. Each blended shot record employs 600 m spatial separation between two sources (Table 2). Figures 13 and 14 show comparisons among four optimization schemes in common shot and detector domain. From top to bottom, the rows correspond to: (1) GAs with an RES, (2) GAs without an RES, (3) Monte Carlo simulation with an RES and (4) Monte Carlo simulation without an RES. Each result obtains a higher signal-to-noise ratio (SNR) compared to the use of a random realization, e.g. an SNR of 8.36 dB from data with randomly designed operators in Fig. 6. However, a certain difference among them is recognizable in the SNR values: 10.92, 10.17, 10.22 and 9.63 dB respectively, although the number of realizations is the same. With an RES, both GAs and Monte Carlo simulation achieve a higher SNR, implying that an RES helps to efficiently find solutions within the limited number of realizations by reducing the problem space. A comparison between GAs and Monte Carlo simulation confirms that the ability of the evolution process derived from stochastic operators in GAs enables us to quickly reach acceptable solutions. With a sufficient number of realizations, Monte Carlo simulation would provide a similar outcome to GAs. Additionally, an RES would become ineffective, as it adds a constraint to the size of the search space. However, achieving 'a sufficient number' is unfortunately quite unrealistic unless an unlimited amount of resources, budget and time are available. As a consequence, both GAs and the RES contribute to make the proposed approach viable to update blending and spatial sampling operators in an effective and efficient manner.

DISCUSSION

Although the use of random realizations is still a common practice to embed irregularity into survey parameters, some recent studies have proposed to manage it particularly for the implementation of the compressive sensing concept. As previously mentioned, Mosher *et al.* (2012) utilized Monte Carlo simulation to find irregular acquisition geometries allowing for optimum data recovery from several hundreds of realizations. Jamali-Rad *et al.* (2016) proposed a way to enhance data reconstruction of sparsely sampled data through optimizing the spatial locations of detectors and sources. The principle of their approach is to find survey geometries that

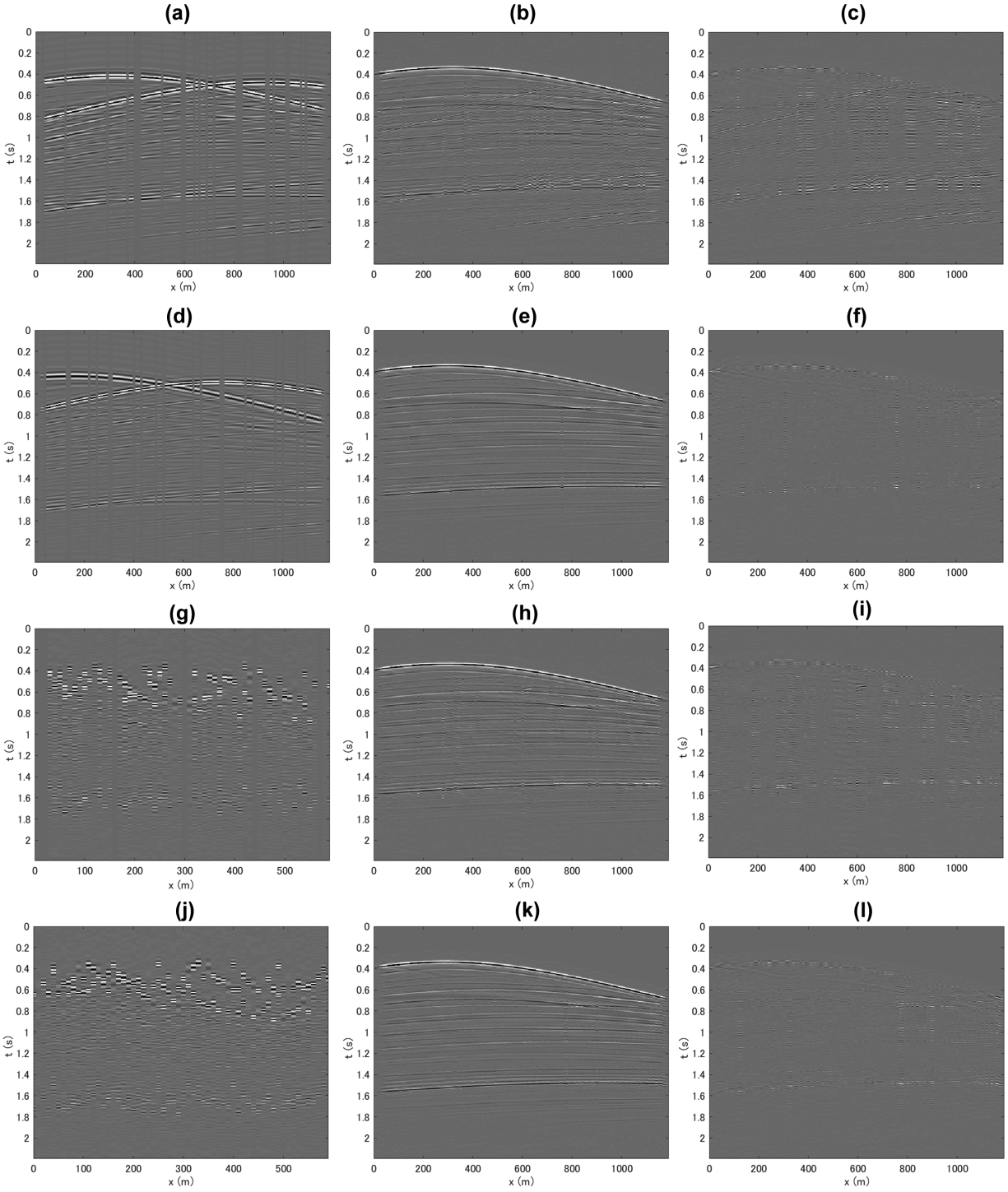

Figure 10 Comparison between optimized and randomly selected survey designs from model 1. Three columns differentiate: blended and irregularly sampled data; deblended and reconstructed data; and difference between input (unblended and well-sampled data) and output (deblended and reconstructed data). Data with randomly designed operators are shown in the odd rows, whereas ones with optimized operators are in the even rows. The top two rows are data in common shot domain, while the bottom two rows are ones in common detector domain. Optimized operators achieve an SNR of 10.92 dB, whereas randomly designed operators obtain an SNR of 8.36 dB.

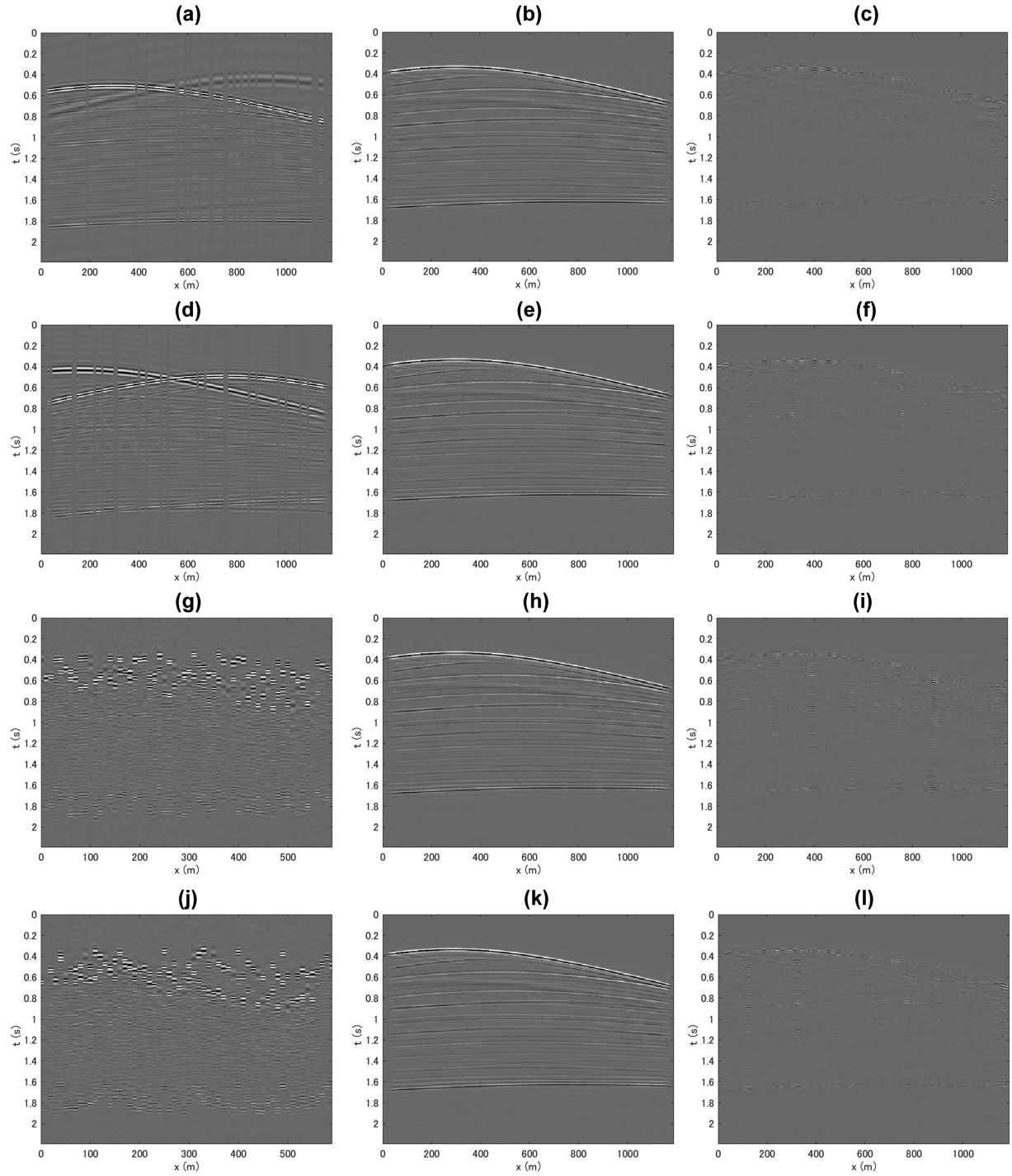

Figure 11 Comparison between optimized and randomly selected survey designs from model 3. Three columns differentiate: blended and irregularly sampled data; deblended and reconstructed data; and difference between input (unblended and well-sampled data) and output (deblended and reconstructed data). Data with randomly designed operators are shown in the odd rows, whereas ones with optimized operators are in the even rows. The top two rows are data in common shot domain, while the bottom two rows are ones in common detector domain. Optimized operators achieve an SNR of 13.25 dB, whereas randomly designed operators obtain an SNR of 10.78 dB.

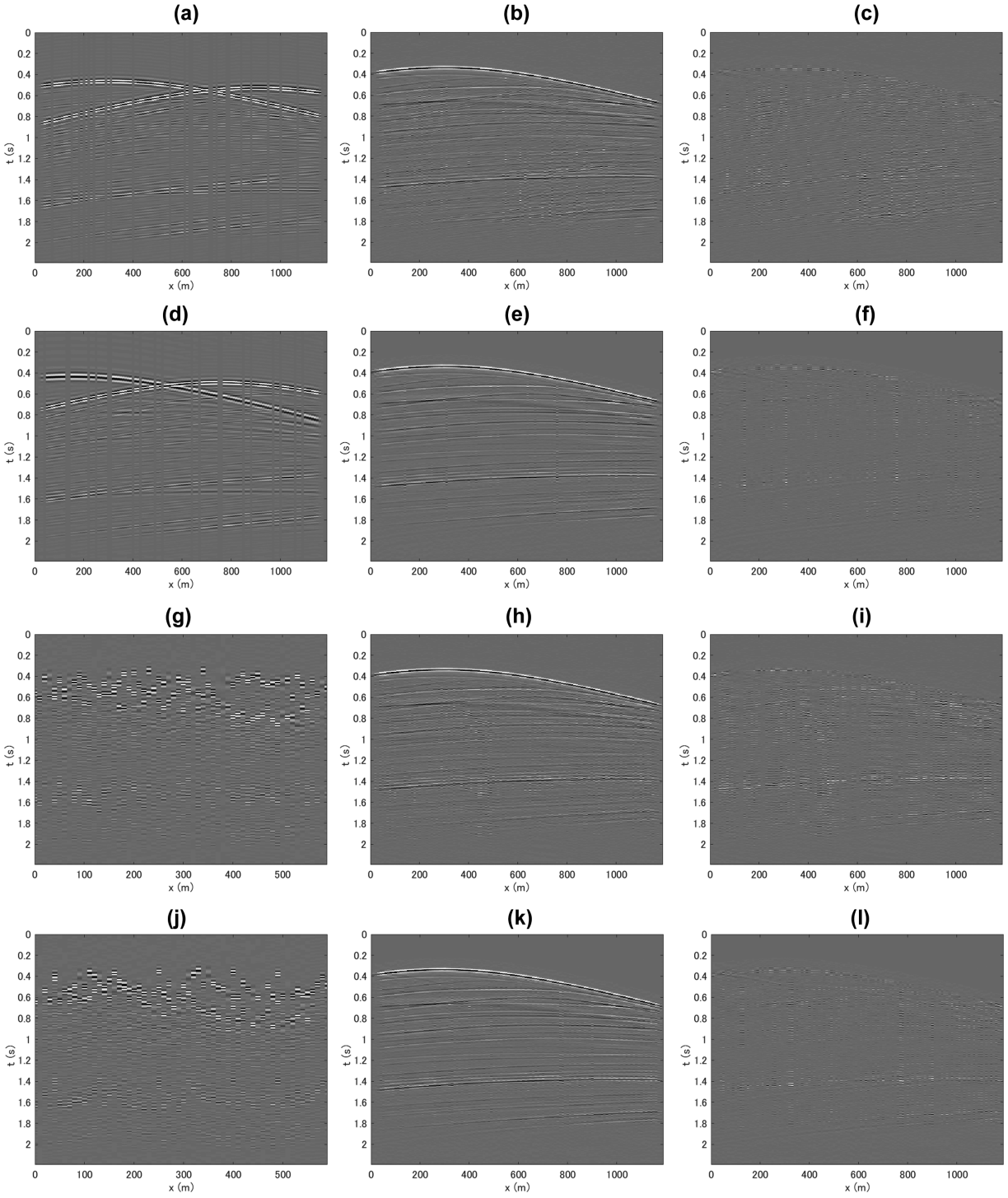

Figure 12 Comparison between optimized and randomly selected survey designs from model 6. Three columns differentiate: blended and irregularly sampled data; deblended and reconstructed data; and difference between input (unblended and well-sampled data) and output (deblended and reconstructed data). Data with randomly designed operators are shown in the odd rows, whereas ones with optimized operators are in the even rows. The top two rows are data in common shot domain, while the bottom two rows are ones in common detector domain. Optimized operators achieve an SNR of 8.94 dB, whereas randomly designed operators obtain an SNR of 7.25 dB.

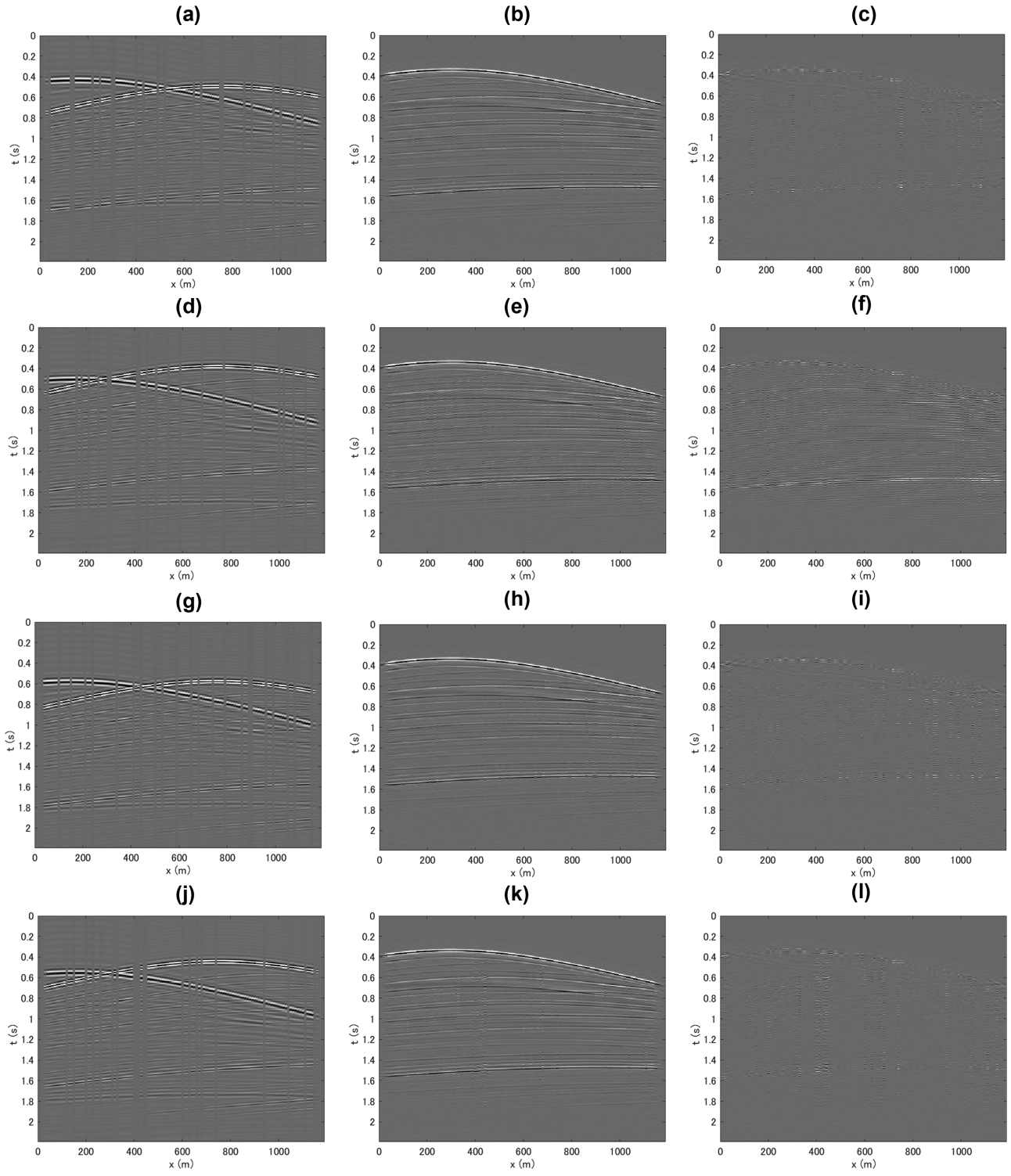

Figure 13 Comparison among different optimization schemes in common shot domain. Each row corresponds to different optimization scheme: GAs with an RES (a–c); GAs without an RES (d–f), Monte Carlo simulation with an RES (g–i); and Monte Carlo simulation without an RES (j–l). Three columns differentiate: blended and irregularly sampled data; deblended and reconstructed data; and difference between input (unblended and well-sampled data) and output (deblended and reconstructed data). SNR values are 10.92 dB for GAs with an RES; 10.17 dB for GAs without an RES, 10.22 dB for Monte Carlo simulation with an RES; and 9.58 dB for Monte Carlo simulation without an RES.

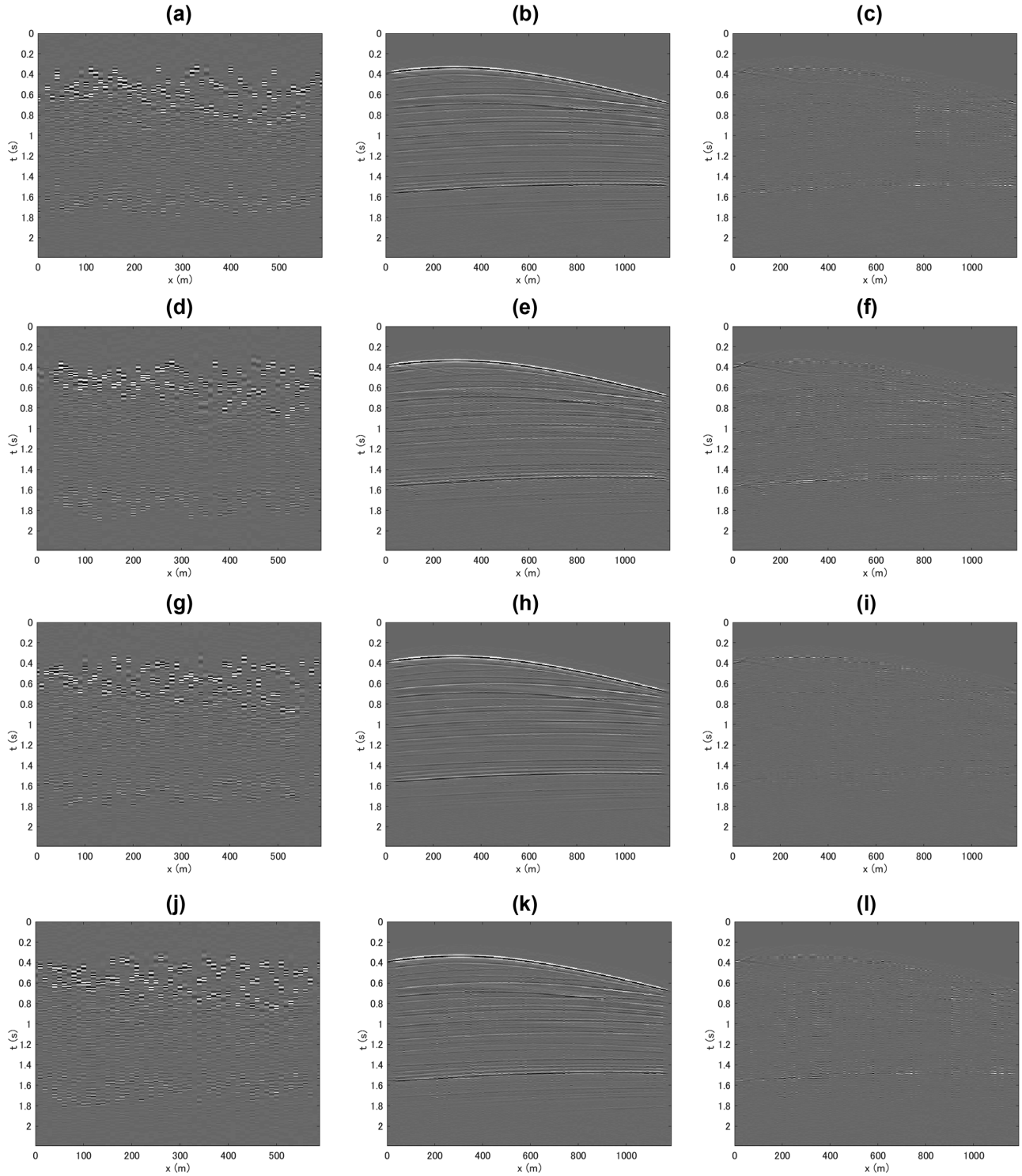

Figure 14 Comparison among different optimization schemes in common detector domain. Each row corresponds to different optimization scheme: GAs with an RES (a–c); GAs without an RES (d–f), Monte Carlo simulation with an RES (g–i); and Monte Carlo simulation without an RES (j–l). Three columns differentiate: blended and irregularly sampled data; deblended and reconstructed data; and difference between input (unblended and well-sampled data) and output (deblended and reconstructed data). SNR values are 10.92 dB for GAs with an RES; 10.17 dB for GAs without an RES, 10.22 dB for Monte Carlo simulation with an RES; and 9.58 dB for Monte Carlo simulation without an RES.

can minimize the maximum mutual coherency of a dictionary matrix such that densely sampled data can be effectively recovered from sparsely sampled data. Campman *et al.* (2017) also showed its applicability to design activation times of overlapped sources instead of the use of time dithering from a random realization. Additionally, Wu, Blacquière and Groenestijn (2015) introduced blended acquisition using shot repetition that activates multiple shots at the same location within a short time interval. They suggested that the design of shot-repetition codes enables effective deblending by making activation times of repeated shots spatially inhomogeneous and uncorrelated.

Nevertheless, few studies have attempted to design both blending and sampling operators. Our study demonstrates that the simultaneous optimization of these operators leading to an improvement in deblending and data reconstruction is attainable. Additionally, numerically simulated examples highlight several aspects of the value of the proposed workflow. First, the method is widely applicable. Our forward modelling can accommodate various types of encoded signatures. This accordingly enables us to simulate several scenarios that resemble acquisition in different environments. The inverse model then makes robust deblending and reconstruction of these data achievable, as described in Ishiyama *et al.* (2017). Genetic algorithms (GAs) along with a repeated encoding sequence (RES) allow for the simultaneous update of blending and sampling operators and subsequently derive an optimum design for scenarios having various blending codes coupled with spatial sampling criteria. Second, the results indicate the potential extensibility of our approach. The optimization for given locations that can represent the area of interest sufficiently provides effective operators applicable to the entire field. By using an RES, the problem space becomes almost irrelevant to the size of a survey as the optimization needs to deal with a single main code only. Then, long parameter sequences can be generated by the combination of four base codes like a chain of four nucleobases in deoxyribonucleic acid. By making use of the base pairs as an analogy, we can simply apply different properties to sources that overlap each other, which certainly makes deblending more effective. These elements surely encourage us to pursue our research further towards field-wide application. Third, the comparison among different optimization schemes illustrates that GAs with an RES contribute to make the proposed approach viable to update blending and sampling operators in a computationally affordable manner. As a consequence, the proposed survey design workflow leads to the optimum acquisition scenario, allowing for an enhancement in the deblending and data reconstruction quality as well as providing significant economical benefits.

The proposed workflow starts with unblended and well-sampled data, **X**, that contain anticipated subsurface responses in the area of interest. This indicates that the approach can derive optimum acquisition scenarios on field-by-field basis. The quality and quantity of the available subsurface information are crucial elements. A limited knowledge of subsurface geology potentially results in uncertain outcomes. On the other hand, the existence of legacy seismic data and/or suitable well data, e.g. density and sonic logs as well as time to depth relationship with a sufficient coverage, is of great help in generating reliable **X**. As discussed previously, the optimization for given locations that can represent the area of interest sufficiently provides effective operators applicable to the entire field. Therefore, our approach unlikely requires a field-wide subsurface model, making the method computationally affordable. It is worth noting that proper attention is required for areas where the deployment of detectors and/or sources is impossible or limited, e.g. due to surface obstructions. To design survey parameters specifically for these areas, separate analysis has to be performed by imposing constraints that can incorporate necessary operational restrictions into parameter vectors.

As discussed previously, various studies have proposed different ways of deblending and data reconstruction, such as the use of different inversion schemes and transform domains. Although our forward and inverse models are highly applicable to various acquisition scenarios, different deblending and reconstruction schemes can be accommodated in the proposed survey design workflow. Additionally, several optimization approaches are adaptable for the update of survey parameters. Monte Carlo simulation is a possible option, although it may not be as effective as the proposed method as shown in Figures 13 and 14. Algorithms having similar features to GAs are seemingly more suitable such as ant colony and particle swarm methods; those having global and population-based search. Although the use of these metaheuristics does not necessarily guarantee the best output within a practical computation time, acceptable solutions are expectantly achievable as shown in this study. The application of an RES to various types of optimization methods is fairly simple and straightforward. In this study, we keep the number of blending performance indicator as well as the numbers of detectors and sources the same through the course of the iteration. In the case where little rationale behind the choice of these parameters is available, the process can start without a constraint on them. Once

their ranges that possibly satisfy required data quality are reasonably defined, we can then add the constraint to subsequent iterations for the further update of the survey parameters. Therefore, the contents of our survey design workflow can be easily and freely altered according to the needs of users. The selection can depend on the subsurface characteristics in the area of interest, any technical preferences, operational considerations and even available resources.

The proposed workflow can be utilized for a health, safety and environment (HSE) perspective. For the last few decades, the consideration and awareness towards HSE have been increasingly recognized as the most essential element in a seismic survey. Hence, it is even one of the factors determining a way to acquire a seismic data. Bouska (2010) and Nakayama *et al.* (2015) described that blended acquisition both in land and marine cases does not require operational complexity; thus, standard and common practices used in a conventional, unblended fashion are sufficient for implementation without the need for any specialized equipment or procedure. Therefore, it is highly unlikely that the technique brings any major safety concerns to field operations, while significant minimization of the operational risk in the field is doubtlessly realizable with a shorter survey duration, especially in land, seabed and transition zone acquisition, where adding extra sources introduces a subtle change in the crew size. The deployment of fewer detectors and sources can also improve survey productivity without any complicated procedures. Therefore, the implementation of blending along with efficient acquisition geometries can lead to a potential reduction in HSE incidents in the field. The proposed workflow makes these technologies more technically justifiable by deriving the optimum blending and spatial sampling operators. Furthermore, our approach is capable of embedding any operational constraints including HSE regulations into the optimization scheme. This consequently enables us to design a safer seismic survey without jeopardizing geophysical and business objectives.

The emission of acoustic energy is often of concern, particularly in a marine environment. Seismic sources possibly cause both injuries and behavioural disturbances to marine life such as cetaceans and fish. The effects on marine mammals are often regarded as a particular concern because a certain number of species rely critically on sound for orientation, food finding, and communication (Tyack and Clark 2000). The use of fewer sources surely reduces the areal acoustic footprint leading to less total acoustic energy disposed to the environment. In addition to sound exposure level over the duration of the survey, the level of the peak amplitude in the source wavefield, independent of the survey duration, is another concern that is a potential threat to marine life. Conventional explosive sources instantaneously emit significant amount of energy at broad frequency ranges, e.g. by the synchronization of clustered air guns. In this respect, the use of encoded signatures is also of help in reducing the peak amplitude of the source energy without adversely affecting the subsurface illumination. The proposed workflow helps to find the blending operator including technically and environmentally favourable encoded signatures. In addition, the spatial distribution of the detectors and encoded sources can be optimized to comply with HSE requirements while ensuring the quality of the subsequent deblending and data reconstruction.

There are inherent operational difficulties in carrying out a seismic survey according to a plan precisely. Certain acquisition errors in $\mathbf{D}$, $\mathbf{S}$ and/or $\mathbf{\Gamma}$ are presumably inevitable in the field operation. The iteration process in GAs provides some insights into the sensitivity of the proposed approach with respect to acquisition errors. A significant change in the objective function values from one generation to the next is rarely observable near the stage of termination. Additionally, a certain number of solutions within the last a couple of generations show small variations in parameters as compared to the optimum one, which accounts for slight differences in objective function values among them. This is presumably attributable to the fact that they are generated from the same parental solutions because of the selection process and the elitist replacement utilized in our approach where the better solution employs the higher expected selection probability, and the best solutions in a given generation still survive into the subsequent generation. A comparison among these solutions infers possible consequences caused by minor or partial alteration in the optimized survey design. Assuming that this situation can be considered as the case where the optimum solution becomes one of the similar solutions in the last a couple of generations, a drastic degradation in the deblending and data reconstruction quality would be not incurred as these solutions employ similar objective function values. Therefore, this possibly implies that minor acquisition errors in the optimized design still provide close to the anticipated result. On the other hands, an undesired result may arise from a large discrepancy between actual and planned parameters as it possibly corresponds to an alteration of the optimum solution to one in a far-separated generation. Further analysis is obviously worthwhile to investigate and quantify significances of acquisition errors with taking practical and operational perspectives into consideration. This should deserve one of our future works to further enhance the applicability of the proposed method.

CONCLUSIONS

We present a workflow to design the survey parameters related to blending as well as detector and source sampling. Through the workflow, blending and spatial sampling operators are iteratively and simultaneously optimized to find solutions leading to an improvement in the deblending and reconstruction quality. The method is widely applicable. It can accommodate various blending and spatial sampling schemes that can describe different acquisition scenarios. Our approach is extensible. The updated acquisition parameters for the given location(s) that can represent the area of interest sufficiently provide effective operators applicable to the entire field. Genetic algorithms along with an repeated encoding sequence contribute to making the proposed approach viable. The blending and sampling operators can be updated in a reasonable computation time. The proposed survey-design workflow consequently provides an optimum acquisition scenario that enables us to realize the benefit of blending and efficient spatial sampling; thereby enhancing the survey productivity, managing budgetary constraints, minimizing health, safety and environment exposure in the field, as well as delivering satisfactory data quality.

ACKNOWLEDGEMENTS

The authors would like to express our sincere appreciation and gratitude to all the sponsors of DELPHI consortium, and INPEX Corporation for their continuous supports.

REFERENCES

Baardman R. and van Borselen R. 2013. A simulated simultaneous source experiment in shallow waters and the impact of randomization schemes. SEG Technical Program Expanded Abstracts, 4382–4386.

Bak S.H., Rask N. and Risi S. 2016. Towards adaptive evolutionary architecture. In: *Evolutionary and Biologically Inspired Music, Sound, Art and Design* (eds C. Johnson, V. Ciesielski, J. Correia and P. Machado), pp. 47–62. Springer International Publishing.

Beasley C.J., Ronald E.C. and Zerong J. 1998. A new look at simultaneous sources. SEG Technical Program Expanded Abstracts, 133–135.

Berkhout A.J. 1982. *Seismic Migration, Imaging of Acoustic Energy by Wave Field Extrapolation, Part A: Theoretical Aspects*, 2nd edn. Elsevier.

Berkhout A.J. 2008. Changing the mindset in seismic data acquisition. *The Leading Edge* **27**, 924–938.

Berkhout A.J. 2012. Blended acquisition with dispersed source arrays. *Geophysics* **77**, A19–A23.

Berkhout A.J. and Blacquière G. 2014. Limits of deblending. SEG Technical Program Expanded Abstracts, 110–114.

Bouska J. 2010. Distance separated simultaneous sweeping, for fast, clean, vibroseis acquisition. *Geophysical Prospecting* **58**, 123–153.

Campman X., Tang Z., Jamali-Rad H., Kuvshinov B., Danilouchkine M., Ji Y., *et al.* 2017. Sparse seismic wavefield sampling. *The Leading Edge* **36**, 654–660.

Caporal M. and Blacquière G. 2015. Benefits of blended acquisition with dispersed source arrays (DSA). 77th EAGE Conference and Exhibition, Madrid, Spain, Extended Abstracts, Tu N104 15.

Cheng J. and Sacchi M.D. 2015. Separation and reconstruction of simultaneous source via iterative rank reduction. *Geophysics* **80**, V57–V66.

Cordsen A., Galbraith M. and Peirce J. 2000. *Planning Land 3-D Seismic Surveys*. Society of Exploration Geophysicists.

Darwin C.R. 1859. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, London.

Davies E., Tew P., Glowacki D.R., Smith J. and Mitchell T. 2016. Evolving atomic aesthetics and dynamics. In: *Evolutionary and Biologically Inspired Music, Sound, Art and Design* (eds C. Johnson, V. Ciesielski, J. Correia and P. Machado), pp. 17–30. Springer International Publishing.

Hennenfent G. and Herrmann F.J. 2008. Simply denoise: Wavefield reconstruction via jittered undersampling. *Geophysics* **73**, V19–V28.

Herrmann F.J. 2010. Randomized sampling and sparsity: getting more information from fewer samples. *Geophysics* **75**, WB173–WB187.

Holland J.H. 1975. *Adaptation in Natural and Artificial Systems*. University of Michigan Press.

Ishiyama T., Ali M.Y., Blacquière G. and Nakayama S. 2017. Deblended-data reconstruction for time-lapse seismic monitoring. *Abu Dhabi International Petroleum Exhibition and Conference*, Abu Dhabi, UAE, November 2017, SPE-188720-MS. Society of Petroleum Engineers.

Ishiyama T., Mercado G. and Belaid K. 2012. 3D OBC seismic survey geometry optimization offshore Abu Dhabi. *First Break* **30**, 51–59.

Jamali-Rad H., Kuvshinov B., Tang Z. and Campman X. 2016. Deterministically subsampled acquisition geometries for optimal reconstruction. 78th EAGE Conference and Exhibition, Vienna, Austria, June 2016, Extended Abstracts, We STZ1 04.

Kontakis A. and Verschuur D.J. 2015. Combined focal and coherency-based deblending strategy. SEG Technical Program Expanded Abstracts, 4667–4672.

Kramer O. 2016. *Machine Learning for Evolution Strategies*. Springer International Publishing.

Kutscha H. and Verschuur D.J. 2012. Data reconstruction via sparse double focal transformation: an overview. *IEEE Signal Processing Magazine* **29**, 53–60.

Li C., Mosher C.C., Morley L.C., Ji Y. and Brewer J.D. 2013. Joint source deblending and reconstruction for seismic data. SEG Technical Program Expanded Abstracts, 82–87.

Liebana D.P., Dieskau J., Hunermund M., Mostaghim S. and Lucas S.M. 2015. Open loop search for general video game playing.

Genetic and Evolutionary Computation Conference, Madrid, Spain, July 2015, pp. 337–344. Association for Computing Machinery.

Lin T.T.Y. and Herrmann F.J. 2009. Designing simultaneous acquisitions with compressive sensing. 71st EAGE Conference and Exhibition Extended Abstracts, S006.

Mahdad A., Doulgeris P. and Blacquière G. 2011. Separation of blended data by iterative estimation and subtraction of blending interference noise. *Geophysics* **76**, Q9–Q17.

Monteagudo A. and Reyes J.S. 2015. Evolutionary optimization of cancer treatments in a cancer stem cell context. *Genetic and Evolutionary Computation Conference*, Madrid, Spain, July 2015, pp. 233–240. Association for Computing Machinery.

Moore I., Dragoset B., Ommundsen T., Wilson D., Ward C. and Eke D. 2008. Simultaneous source separation using dithered sources. SEG Technical Program Expanded Abstracts, 2806–2810.

Mosher C.C., Kaplan S.T. and Janiszewski F.D. 2012. Non-uniform optimal sampling for seismic survey design. 74th EAGE Conference and Exhibition Extended Abstracts.

Nakayama S., Belaid K. and Ishiyama T. 2013. Seeking efficient OBC survey designs that still satisfy established geophysical objectives. *First Break* **31**, 65–73.

Nakayama S., Mercado G., Benson M., Belaid K. and Garden M. 2015. Field-wide implementation of time and distance separated source techniques on a 3D OBC survey offshore Abu Dhabi, UAE. *First Break* **33**, 47–53.

Regone C.J. 2007. Using 3D finite-difference modeling to design wide-azimuth surveys for improved subsalt imaging. *Geophysics* **72**, SM231–SM239.

Scirea M., Togelius J., Eklund P. and Risi S. 2016. Metacompose: A compositional evolutionary music composer. In: *Evolutionary and Biologically Inspired Music, Sound, Art and Design* (eds C. Johnson, V. Ciesielski, J. Correia and P. Machado), pp. 202–217. Springer International Publishing.

Tyack P.L. and Clark C.W. 2000. Communication and acoustic behavior of dolphins and whales. In: *Hearing by Whales and Dolphins* (eds W.W.L. Au, R.R. Fay and A.N. Popper), pp. 156–224. Springer.

Vermeer G.J.O. 2012. *3D Seismic Survey Design*, 2nd edn. Society of Exploration Geophysicists.

Verschuur D.J., Vrolijk J.W. and Tsingas C. 2012. 4D reconstruction of wide azimuth (WAZ) data using sparse inversion of hybrid Radon transforms. SEG Technical Program Expanded Abstracts, 1–5.

Wu S., Blacquière G. and Groenestijn G.V. 2015. Shot repetition: An alternative approach to blending in marine seismic. SEG Technical Program Expanded Abstracts, 48–52.

Xu S., Zhang Y., Pham D. and Lambaré G. 2005. Antileakage Fourier transform for seismic data regularization. *Geophysics* **70**, V87–V95.