Delft University of Technology

Machine-learning-based data recovery and its contribution to seismic acquisition
Simultaneous application of deblending, trace reconstruction, and low-frequency
extrapolation

Nakayama, Shotaro; Blacquière, Gerrit

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Machine-learning based data recovery and its contribution to seismic acquisition: simultaneous application of deblending, trace reconstruction and low-frequency extrapolation

Shotaro Nakayama[*][†] **and Gerrit Blacquière**[†]

[*]*INPEX Corporation,*

Akasaka Biz Tower 5-3-1, Akasaka, Minato-ku, Tokyo 107-6332, Japan

[†]*Delft University of Technology,*

Building 23, Stevinweg 1, 2628 CN, Delft, the Netherlands

Email addresses of the corresponding author :

shotaro.nakayama@inpex.co.jp

(September 11, 2020)

Prepared for *Geophysics*

Running head: **ML based data recovery**

ABSTRACT

Acquisition of incomplete data, i.e., blended, sparsely-sampled and narrowband data, allows for cost-effective and efficient seismic operations in the field. This strategy becomes technically acceptable, provided that a satisfactory recovery of the complete data, i.e., deblended, well-sampled and broadband data, is attainable. We, hence, explore a machine-learning approach that simultaneously performs suppression of blending noise, reconstruction of missing traces and extrapolation of low frequencies. We apply a deep convolutional neural network in the framework of supervised learning where we train a network using pairs of incomplete-complete datasets. Incomplete data, which are never used for training and employ different subsurface properties and acquisition scenarios, are subsequently fed into the trained network to predict complete data. We describe matrix representations indicating the contributions of different acquisition strategies to reducing the operational effort in the field. We also illustrate that the simultaneous implementation of source blending, sparse geometry and band limitation leads to a significant data compression where the size of the incomplete data in the frequency-space domain is much smaller than the size of the complete data. This reduction is indicative of survey cost and duration that our acquisition strategy can save. Both synthetic and field data examples demonstrate the applicability of the proposed approach. Despite the reduced amount of information available in the incomplete data, the results obtained from both numerical and field data cases clearly show that the machine-learning scheme effectively performs deblending, trace reconstruction and low-frequency extrapolation in a simultaneous fashion. It is noteworthy that no discernible difference in prediction errors between extrapolated frequencies and preexisting frequencies is observed. The approach potentially allows seismic data to be acquired in a significantly compressed manner, while subsequently recovering data of satisfactory quality.

2

INTRODUCTION

Acquisition of seismic data is almost always considered as a trade-off between data quality and cost. One way to deal with these conflicting objectives is to reduce the acquisition effort in the field and then to address deficiencies in the recorded data through subsequent processing steps. Due to the large imbalance between acquisition and processing cost, the need for the extra effort in processing is still acceptable. Hence, this strategy is capable of lowering the overall project cost while making the desired data quality realizable. For instance, over the last several years, compressive sensing has attained considerable attention in the industry (Herrmann, 2010; Mosher et al., 2012). The technique aims at recovering satisfactory data quality from data recorded with efficient sampling schemes which do not necessarily follow the Nyquist sampling theorem. Blended acquisition, also referred to as simultaneous source acquisition, is a widely accepted way to enhance the data quality while maintaining or even reducing the project cost and time (Beasley et al., 1998; Berkhout, 2008; Bouska, 2010; Abma et al., 2012; Nakayama et al., 2015). These acquisition strategies normally coexist with a subsequent data recovery step, such as deblending and data reconstruction. These processes are generally posed as an inverse problem. The widely used approach is to exploit either the low-rank structure or sparse representation of seismic data in some transform domain(s) (Hennenfent and Herrmann, 2008; Oropeza and Sacchi, 2011; Kutscha and Verschuur, 2012; Kontakis and Verschuur, 2014). Despite its applicability, the need for an iterative process inherently incurs the computational burden. In addition, underlying assumptions used in these approaches inevitably impose constraints on the recovered data, which may lead to the limitation of capturing subsurface complexities in the real world.

The importance of low frequencies in seismic data is well recognized in several aspects such as deep penetration of source energy, lessened side lobes of the wavelet, absolute impedance estimation and improved convergence of full-waveform inversion (Ten Kroode et al., 2013; Berkhout et al., 2017). However, the emission of low-frequency components is often troublesome in the field. For example, it requires large and/or dedicated devices as well as extra shooting effort (Dellinger et al., 2016; Wei et al., 2018). In addition to the operational difficulty, this possibly incurs a financial burden as well. Hence, the creation of missing low frequencies in a processing step, hereinafter referred to as low-frequency extrapolation, rather than actually collecting them in the field, is of value. Several studies have been carried out to extrapolate low frequencies, e.g., by using the envelope of the recorded signal or convolving a broadband wavelet to an estimated reflectivity model (Wu et al., 2014; Zhang et al., 2017). Nevertheless, finding an intrinsic relationship between recorded and missing frequencies is still an extremely difficult task.

In recent years, machine learning (ML) techniques have become increasingly popular in various domains including seismic applications. Several studies have demonstrated their capability to handle a deblending or data reconstruction problem with a performance that is comparable to that of existing geophysical approaches (Siahkoohi et al., 2019; Wang et al., 2019; Sun et al., 2020). Furthermore, once the networks are trained, the prediction can be done at a significantly reduced computational cost. Similarly, some recent studies on synthetic data using ML to extend the bandwidth showed encouraging outcomes, where the particular emphasis was given to low-frequency extrapolation (Ovcharenko et al., 2019; Sun and Demanet, 2020). However, the aforementioned processes have been, so far, treated individually as separate tasks. This study, therefore, explores an ML scheme that aims at optimal data recovery from seismic data acquired in a blended, sparsely-sampled and nar-

rowband manner. We describe matrix representations that indicate the effects of different acquisition scenarios on the survey efficiency, which helps to intuitively indicate the value of the proposed acquisition strategy. Synthetic and field data examples demonstrate the capability of the proposed method to simultaneously handle suppression of seismic interference, reconstruction of missing traces and extrapolation of low frequencies. Additionally, we discuss the potential benefits of the proposed scheme making a cost-effective and efficient seismic survey realizable without compromising data quality.

MATRIX REPRESENTATIONS OF DIFFERENT ACQUISITION STRATEGIES

We first describe matrix representations that indicate the contributions of sparse acquisition geometries, source blending and band limitation to the survey efficiency. We can arrange (2D or 3D) seismic data into a 3D data matrix in the frequency-space domain. With this arrangement, the two horizontal axes correspond to the spatial coordinates of detectors and sources, respectively. The vertical axis corresponds to frequency components. Based on the notation in Blacquière and Nakayama (2019), we describe the seismic response acquired with the perfect spatial sampling and broad bandwidth, called complete date in this study, as

$$\mathbf{X} \in \mathbb{C}^{n_\mathrm{d} \times n_\mathrm{s} \times n_\omega}, \tag{1}$$

where n_d and n_s are equal to the numbers of the detector and source grids along the acquisition surface where the detector and source intervals satisfy the Nyquist sampling criterion, and n_ω corresponds to the number of frequency components. A vertical planar section of $\mathbf{X}$, parallel to the detector coordinate axis, corresponds to a common shot gather.

A vertical planar section of $\mathbf{X}$, parallel to the source coordinate axis, then corresponds to a common detector gather. A horizontal planar section of $\mathbf{X}$ corresponds to seismic data for a given frequency. Therefore, each element contains a complex number that represents a monochromatic seismic response acquired by a given detector-source pair.

In practice, acquisition of $\mathbf{X}$ is too expensive. As mentioned previously, the deployment of sparse detector and source geometries is one way to enhance the survey efficiency. The data matrix acquired with a practical acquisition geometry is expressed as

$$\mathbf{P} \in \mathbb{C}^{f_\mathrm{d} n_\mathrm{d} \times f_\mathrm{s} n_\mathrm{s} \times n_\omega} \quad (0 < f_\mathrm{d} < 1 \quad \text{and} \quad 0 < f_\mathrm{s} < 1), \tag{2}$$

where f_d and f_s indicate decimation factors for detector and source side, respectively. A smaller value of f_d or f_s indicates fewer active detector points or source points along the acquisition surface. The matrix representations in Figure 1a illustrate the effect of sparse geometries. With missing detectors, corresponding sections of $\mathbf{X}$, parallel to the source coordinate axis, become empty. Similarly, with missing sources, corresponding sections of $\mathbf{X}$, parallel to the detector coordinate axis, become empty. By removing zero elements and keeping non-zero elements, it is possible to further rearrange the data matrix in a smaller form. This change in the size of the data matrix is indicative of acquisition cost and time that we can potentially save. Nevertheless, a detailed discussion on the particular relationship between matrix size and acquisition effort is beyond the scope of this paper.

Blended acquisition is another means of contributing to improving the survey efficiency, thereby reducing the size of the data matrix. We describe a blended data matrix as

$$\mathbf{P}' \in \mathbb{C}^{n_\mathrm{d} \times n_\mathrm{s}' \times n_\omega} \quad (n_\mathrm{s} > n_\mathrm{s}'), \tag{3}$$

where n'_s corresponds to the number of blended shots. Figure 1b exemplifies matrix representations in the case of a blending fold of two. Elements with two different colors indicate data acquired by two different sources. In blended acquisition, these two sources are activated in an overlapping fashion and blended in one blended experiment, leading to one blended shot record. Therefore, each element with mixed colors in Figure 1b indicates blended data comprising of a contribution from two shot records. With this blending scenario, the size of the data matrix along the source axis is reduced by a factor of two.

Acquisition of limited frequencies is also of help in reducing the size of the data matrix (Figure 1c), defined as

$$\dot{\mathbf{P}} \in \mathbb{C}^{n_\mathrm{d} \times n_\mathrm{s} \times f_\omega n_\omega} \quad (0 < f_\omega \leq 1), \tag{4}$$

where f_ω indicate a decimation factor along the frequency axis. As mentioned previously, our primary focus is on a situation where low-frequency components are not recorded in the field. In this study, we therefore assume $\dot{\mathbf{P}}$ to be a data matrix without low frequencies. As shown in Figures 1a-c, $\mathbf{P}$, $\mathbf{P}'$ and $\dot{\mathbf{P}}$ are far smaller than $\mathbf{X}$, making these techniques justifiable instead of acquiring $\mathbf{X}$, provided that the subsequent data recovery leads to reasonable $\langle \mathbf{X} \rangle$ where angle brackets indicate estimation.

Additionally, the use of all mentioned strategies in a combined manner further contributes to lowering the acquisition effort, which is the objective of this study. We describe blended, sparsely-sampled and narrowband data, called incomplete data in this study, as

$$\dot{\mathbf{P}}' \in \mathbb{C}^{f_\mathrm{d} n_\mathrm{d} \times f_\mathrm{s} n'_\mathrm{s} \times f_\omega n_\omega}. \tag{5}$$

This acquisition scenario significantly reduces the size of the data matrix as depicted in Fig-

ure 1d. As mentioned, $\dot{\mathbf{P}}'$ is assumed to be a situation where low frequencies are missing. The data recovery scheme introduced in this study aims to simultaneously handle suppression of blending noise, reconstruction of missing traces and extrapolation of missing low frequencies such that prediction of the complete data from the incomplete data is possible.

MACHINE LEARNING FRAMEWORK

Following our previous study (Nakayama and Blacquière, 2020), we utilize a deep convolutional neural network (CNN) in the framework of supervised learning. CNNs are implemented with linear convolutions optionally followed by non-linear operations. As compared to standard fully-connected neural networks with similarly-sized layers, CNNs have much fewer connections and parameters. Although CNNs already exist for long, notable experimental results have been achieved more recently (LeCun et al., 2015). One of the most well-known cases was the ImageNet Large Scale Visual Recognition Challenge in 2012, where a CNN achieved an error rate of 15.3% in an image classification task, compared to 26.2% by the second-best entry (Krizhevsky et al., 2012; Russakovsky et al., 2015). CNNs have also provided remarkable performances in a variety of areas such as speech recognition (Hinton et al., 2012), biomedicine (Leung et al., 2014) and natural language understanding (Sutskever et al., 2014). Additionally, the implementation of CNNs in the seismic industry has recently emerged, as mentioned previously.

In this study, we apply a network architecture based on the U-Net (Ronneberger et al., 2015; Enokiya et al., 2018). Figure 2 depicts our network architecture applied for the field data example. It consists of four encoding and four decoding blocks with skip pathways that enable the encoder and decoder feature maps to be directly connected. This helps us to transfer some detailed information which may be smeared through down- and up-sampling

processes. Figure 3a shows layers applied to encoders. Each of them utilizes the repeated application of a convolutional layer (LeCun et al., 1998), a rectified linear unit (Hahnloser et al., 2000) and a batch normalization layer (Ioffe and Szegedy, 2015), along with a residual learning framework to make accuracy gains with increasing network depth realizable (He et al., 2016). In each encoding block, there is a stack of three layers having 1×1, 3×3 and 1×1 convolutions based on a bottleneck design (He et al., 2016). Here, the 1×1 convolutional layers account for reducing and restoring the depth dimension, meaning that the 3×3 convolutional layer between the two employs smaller input and output dimensions. Each encoding block halves the spatial resolution via a 2×2 max pooling layer with a stride of 2, while doubling the depth dimension (Nagi et al., 2011). In between encoding and decoding steps, there is a center block (Figure 3b). Each decoding block doubles the spatial resolution via a 2×2 up-convolution layer, while halving the depth dimension (Dumoulin and Visin, 2016). The two feature maps from an up-convolution layer and from the corresponding encoding block through the skip pathway are concatenated (Figure 3c). The following 3×3 convolutional layer further halves the depth dimension. A set of layers utilizing a residual framework along with a bottleneck design are subsequently applied. The last decoding block then accounts for mapping each sampling point of the output data, regarded as $\langle \mathbf{X} \rangle$ (Figure 3d). For the numerical example, a similar yet simpler architecture (three pairs of encoders and decoders) is used. The applied ML framework is empirically derived, which does not necessarily guarantee the best performance for a data recovery problem. However, an investigation into the choice of an optimum network architecture along with its hyperparameters is outside of the scope of this study.

For the application of supervised learning to data recovery, we first create a set of suitably chosen input-output training pairs, i.e., pairs of incomplete-complete data. Prior to

feeding these datasets into the training process, we apply pseudo-deblending (Mahdad et al., 2011) to the incomplete data and insert zero elements that correspond to data associated with missing detectors and sources. Additionally, we transform the data matrix to the time-space domain where the two horizontal axes correspond to detector and source coordinates, and the vertical axis corresponds to recording time. These operations make the size of the input data identical to that of output data. This means that our input and output data are 3D matrices in the time-space domain with the size of $n_\mathrm{d} \times n_\mathrm{s} \times n_\mathrm{t}$ where n_t is the number of time samples. In our case, we apply a 2D CNN with multiple channels where the time axis is treated as the channel axis. Hence, the spatial information along the detector and source coordinate axes is compressed and then de-compressed through the encoding-decoding process, while vice versa for the dimensional information along the time axis.

In general, seismic data exhibit a rapid amplitude decay due to wavefront divergence and attenuation losses. In a marine environment, the seabed is normally responsible for the largest impedance contrast. As pointed out by Sun and Demanet (2020), the trained network may be biased by strong events, e.g., reflectors at shallow levels, while weak events, e.g., reflectors at deep levels, may be neglected. To compensate for the amplitude decay and ensure proper recovery of weak reflectors, we apply a time gain to pairs of incomplete-complete data prior to the training. This means that our trained network is designed to deal with gain-compensated data. The gain can be easily reversed after the prediction.

SYNTHETIC DATA EXAMPLE

We numerically simulate 20,000 small complete datasets using a full wavefield modelling scheme (Berkhout, 2014). In the complete data, 32 detectors and 32 sources are deployed

10

regularly, both with a 20 m interval. We derive 20,000 subsurface models, all with three anticlinal reflectors. While synthesizing each complete dataset, we arbitrarily alter the subsurface structures, i.e., geometry and depth of each anticline, as well as subsurface properties, i.e., reflectivity of each interface and propagation velocity between each reflector. Additionally, for incomplete datasets, we apply 50% detector and 50% source decimation in an irregular fashion. The blending fold is two, and the blending code is random time dither between 0 s and 0.064 s. Low frequencies are missing in the incomplete data, where the applied low-cut and low-pass frequency are 10 Hz and 16 Hz, respectively. This indicates that each dataset employs different subsurface responses and acquisition scenarios. This experimental setup consequently corresponds to a significant reduction in the size of the incomplete data with respect to the complete data. We arbitrarily select 19,000 complete-incomplete data pairs to train the network, while the remaining 1,000 pairs are used for the purpose of testing. Hence, these testing datasets employ acquisition configurations and subsurface responses that differ from those of the training datasets.

Figures 4 and 5 show the results of our numerical example. Here, the median result is selected among the 1,000 testing datasets in terms of prediction errors, which we assume to be the representative result of the proposed approach. The incomplete data exhibit blending noise and lack a considerable amount of information in frequency and space (Figures 4b and 5b). Despite significant deficiencies in the incomplete data, the applied network reasonably suppresses blending noise, reconstructs missing traces and extrapolates low frequencies in a simultaneous manner (Figures 4c and 5c). It is noteworthy that missing low frequencies in the incomplete data are correctly predicted. Recovery errors are reasonably minimized for the whole frequency range (Figures 4d and 5d). Consequently, the quality of recovered data becomes fairly comparable to that of the complete data.

To quantify the prediction performance, we compute the SNR (signal-to-noise ratio) defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \langle \mathbf{x} \rangle\|_2} \right), \tag{6}$$

where $\mathbf{x}$ is a vectorized form of $\mathbf{X}$. Figure 6 shows the resultant SNR values from the 1,000 testing datasets. The median result in Figures 4 and 5 yields the SNR value of 19.27 dB. A certain amount of variation in the SNR values can be observed among different testing sets. It is noticed that some datasets show relatively large prediction errors. Figures 5a-b show the incomplete data from the 796th testing set which exhibits the lowest SNR value of 11.92 dB among the 1,000 datasets. We also generate different incomplete data and then apply the trained network from our numerical experiment as follows. The two datasets in Figure 7 are simulated with the same subsurface properties and the same frequency content where low frequencies are missing. Additionally, the same numbers of detectors and sources with the same blending fold of two are used. However, we apply different a detector and source distribution as well as different blending code. The recovery result obtained from the incomplete data in Figures 7c-d leads to the improved SNR value of 19.02 dB. This number is fairly comparable to our median result. As the two datasets in Figure 7 are simulated with the same subsurface model, this difference is solely attributable to the choice of survey parameters. Although they use the same number of detectors and sources, the data in Figures 7a-b exhibit relatively large acquisition gaps both in the common shot and detector domains. They are a probable cause of the suboptimal prediction result. Similar observations can be made in other datasets with relatively low SNR values in Figure 6. This is in agreement with (Wang et al., 2020) that showed a larger gap hampers the network from extracting key features to characterize the local data and recovering data from these extracted features. This result indicates that attention should be given towards the choice

of acquisition parameters, which is of potential help in ensuring the performance of ML based data recovery.

FIELD DATA EXAMPLE

We selected a subset of 2D towed-streamer data acquired in the Troll field offshore Norway. The survey was performed with 25 m detector and source intervals. Some preprocessing is applied prior to our experiments such as removal of direct arrivals, near-offset interpolation, surface-related multiple elimination and trace equalization. Additionally, the data is arranged into a fixed-spread geometry using source-detector reciprocity. This data is considered as the complete data, $\mathbf{X}$, which we aim to recover.

A further subset of this field data is used to generate 12,000 training datasets, i.e., pairs of complete-incomplete data. To obtain the incomplete datasets, both detectors and sources are arbitrarily and irregularly decimated by 25%. The blending fold is two, and the blending code is random time dither between 0 s and 0.256 s. Low frequencies are also removed by applying the low-cut and low-pass frequency, respectively, of 10 Hz and 16 Hz. These complete-incomplete datasets are used to train the network shown in Figure 2. For testing, we selected a portion of the field data from a different area, such that there is no overlap between training and testing datasets. Using the complete data selected for testing, we generate 500 incomplete datasets, each having different spatial sampling and blending schemes as well as lacking low frequencies. Figures 8 and 9 shows the data recovery result from the field data example. As in the numerical example, we show the median result among 500 testing datasets. Here, we again observe that the ML scheme reasonably suppresses blending noise, reconstructs missing traces and extrapolates low frequencies. The difference plots notably exhibit insignificant prediction errors along with no clear frequency

dependency (Figures 8d and 9d). Despite satisfactory recovery results obtained from the applied network particularly for major reflectors, a close inspection of the prediction residual reveals that some subtle events are still smeared. We expect that further study on the ML framework would enable us to minimize the prediction errors although such a study is beyond the scope of this paper.

Figure 10 shows the SNR values of the 500 testing datasets. Our median result shown in Figures 8 and 8 yields the SNR value of 9.17 dB. As these incomplete datasets are derived from the same complete data, subsurface characteristics are an irrelevant factor of the variation in the resultant SNR values. As described previously, they use the same numbers of detectors and sources along with the same blending fold of two, whereas different geometries and activation times are applied. Hence, the latter parameters are responsible for the differences in the prediction performance among the testing sets. Figure 11 compares the two incomplete datasets that attain the lowest (8.38 dB) and the highest SNR value (9.42 dB), respectively. As observed in the numerical example, the incomplete data with the lowest SNR value shows a relatively large acquisition gap in the common detector domain, which likely deteriorates the quality of data recovery (Figures 11a-b). It is also interesting that a certain variation in the SNR values is still recognizable even among datasets having no significant gap, e.g., our median result (Figures 8b and 9b) and the data in Figures 11c-d. This implies that, besides the size of acquisition gap(s), the choice of spatial sampling and blending schemes is one of the factors determining the quality of the recovered data. Hence, an investigation of the rationale behind this aspect would provide a future research avenue.

DISCUSSION

Although various aspects are still up for discussion, here we focus specifically on the potential benefits of our approach along with future directions related to seismic acquisition.

One of the key features in the applied ML scheme is the capability to deal with multiple tasks in one step. This permits seismic data to be recorded in a highly compressed fashion as illustrated in Figure 1d. With the ML approach, acquisition of the complete data is only needed in a portion of the area for training purposes, while the rest of the area can be acquired in an incomplete manner. Once a network is trained, the computational burden of data recovery is insignificant, unlike in the case of existing iterative algorithms. This way of data acquisition along with the subsequent data recovery therefore potentially leads to a considerable reduction of turnaround time and cost incurred from acquisition to processing.

To make the aforementioned benefits realizable, supplying proper input-output datasets for training in terms of quality and quantity is of primary importance. The collection of more samples for training certainly helps a network to reliably learn subsurface characteristics in the area of interest. However, this inevitably limits the contribution to an efficiency perspective. The collection of fewer training samples is economically beneficial, yet it potentially hampers the network from being properly trained. Hence, a trade-off likely occurs between the collection of more data for training and that of more data to be recovered. Finding a strategy to strike a proper balance for splitting a survey area into two (one for training and the other for testing) would be an important research goal. It is also worth exploring the possibility to train the network using existing data from other fields, e.g., adjacent or neighbouring ones in the same basin, or even synthetic data. This approach obviously makes the prediction challenging as a certain discrepancy in seismic responses

between training and testing datasets is likely anticipated. On the other hand, this would be a highly attractive way from operational and economical perspectives as acquisition of complete data in the area of interest is no longer needed.

As for the spatial sampling of detectors and sources, Siahkoohi et al. (2019) applied ML to a data reconstruction problem and demonstrated the superior performance of irregularly-sampled data over regularly-sampled data. They speculated that the principle behind compressive sensing may also be applicable in the case of an ML approach. In existing blended acquisition schemes, the source wavefield is often 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 by their combination (Baardman and van Borselen, 2013). Recent applications of ML to deblending implemented this acquisition strategy (Sun et al., 2020; Baardman and Hegge, 2020). Our numerical and field data examples also incorporate irregularity into the acquisition parameters such as detector and source geometries and activation times. This study shows the variation of data recovery results related to their choices. There have been several studies that aim to design the irregularity in acquisition parameters, such as spatial sampling and blending schemes, to improve the existing, geophysics-based data recovery processes (Mosher et al., 2012; Mueller et al., 2015; Nakayama et al., 2019). Similarly, an investigation into the design of acquisition parameters for the incomplete data would lead to enhancing the performance of ML based data recovery.

As mentioned, the primary focus of this study is to explore the applicability of an ML approach to derive the complete data from the incomplete data and to describe its potential benefits to seismic acquisition. However, further studies are needed to reduce uncertainties in predicted data, particularly in terms of extrapolated low frequencies. Recent

studies have demonstrated to enhance the performance of full waveform inversion by the use of predicted low frequencies (Ovcharenko et al., 2019; Sun and Demanet, 2020; Fang et al., 2020). Additionally, there have been some studies providing theoretical frameworks showing that high and low frequencies are relatable under certain assumption(s), e.g., when small dispersion effects are expected (Li and Demanet, 2015) and the Born approximation reasonably describes wavefields (Ovcharenko et al., 2019). Nevertheless, to fully realize the aforementioned values that low frequency components hold, detailed investigation into the fidelity of predicted data along with a wider range of applications in different geological and geophysical contexts is certainly needed. This should be considered as our future research avenue, allowing us to properly understand the capability as well as limitations of the proposed approach.

CONCLUSIONS

We describe matrix representations that indicate benefits of acquiring incomplete data, i.e., blended, sparsely-sampled and narrowband data, coupled with subsequent data recovery. For this purpose, we implement an ML approach and explore its applicability using synthetic and field data. We utilize the U-Net based network architecture, consisting of an encoding path and a decoding path, in the framework of supervised learning. Although the acquisition scenarios applied in this study significantly compress the size of the data in the frequency-space domain, the results obtained from both numerical and field data examples clearly demonstrate that the trained network effectively performs suppression of blending noise, reconstruction of missing traces and extrapolation of low frequencies in a simultaneous fashion. It is noteworthy that no discernible difference in prediction errors between extrapolated frequencies and preexisting frequencies is recognized, which is hardly realiz-

able with existing geophysical approaches. Additionally, this study shows the variation of recovery results, attributable solely to the choice of acquisition parameters. This infers that properly designed survey parameters are of help in enhancing the quality of recovered data. Although further studies are needed, the ML scheme potentially enables seismic acquisition in a blended, sparsely-sampled and narrowband manner without seriously compromising the data quality.

REFERENCES

Abma, R., Q. Zhang, A. Arogunmati, and G. Beaudoin, 2012, An Overview of BP's Marine Independent Simultaneous Source field trials: SEG Technical Program Expanded Abstracts 2012, 1–5.

Baardman, R., and R. Hegge, 2020, Machine learning approaches for use in deblending: The Leading Edge, **39**, 188–194.

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

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

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

———, 2014, Review Paper: An outlook on the future of seismic imaging, Part I: Forward and reverse modelling: Geophysical Prospecting, **62**, 911–930.

Berkhout, A. J., G. Blacquière, and D. J. E. Verschuur, 2017, Enlarging the bandwidth of seismic images: SEG Technical Program Expanded Abstracts 2017, 180–185.

Blacquière, G., and S. Nakayama, 2019, Optimum seismic acquisition geometry design with the help of artificial intelligence: SEG Technical Program Expanded Abstracts 2019, 117–121.

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

Dellinger, J., A. Ross, D. Meaux, A. Brenders, G. Gesoff, J. Etgen, J. Naranjo, G. Openshaw, and M. Harper, 2016, Wolfspar®, an "FWI-friendly" ultralow-frequency marine

seismic source: SEG Technical Program Expanded Abstracts 2017, 4891–4895.

Dumoulin, V., and F. Visin, 2016, A guide to convolution arithmetic for deep learning: arXiv preprint arXiv:1603.07285.

Enokiya, Y., Y. Iwamoto, Y.-W. Chen, and X.-H. Han, 2018, Automatic liver segmentation using u-net with wasserstein gans: Journal of Image and Graphics, **6**.

Fang, J., H. Zhou, Y. Li, Q. Zhang, L. Wang, P. Sun, and J. Zhang, 2020, Data-driven low-frequency signal recovery using deep learning predictions in full-waveform inversion: Geophysics, **85**, 1–42.

Hahnloser, R. H., R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, 2000, Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit: Nature, **405**, 947–951.

He, K., X. Zhang, S. Ren, and J. Sun, 2016, Deep residual learning for image recognition: Proceedings of the IEEE conference on computer vision and pattern recognition, 770–778.

Hennenfent, G., and F. J. Herrmann, 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.

Hinton, G., L. Deng, D. Yu, G. Dahl, A. rahman Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. Sainath, and B. Kingsbury, 2012, Imagenet classification with deep convolutional neural networks: IEEE Signal Processing Magazine, **29**, 82–97.

Ioffe, S., and C. Szegedy, 2015, Batch normalization: Accelerating deep network training by reducing internal covariate shift: arXiv preprint arXiv:1502.03167.

Kontakis, A., and D. J. Verschuur, 2014, Deblending via sparsity-constrained inversion in the Focal domain: 76th EAGE Conference and Exhibition Extended Abstracts, **Th ELI1**

02.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2012, Imagenet classification with deep convolutional neural networks: NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, **1**, 1097–1105.

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

LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–444.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998, Gradient-based learning applied to document recognition: Proceedings of the IEEE, **86**, 2278–2324.

Leung, M. K. K., H. Y. Xiong, L. J. Lee, and B. J. Fre, 2014, Deep learning of the tissue-regulated splicing code: Bioinformatics, **30**, i121–i129.

Li, Y. E., and L. Demanet, 2015, Phase and amplitude tracking for seismic event separation: Geophysics, **80**, WD59–WD72.

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

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

Mueller, M. B., D. F. Halliday, D. J. van Manen, and J. O. A. Robertsson, 2015, The benefit of encoded source sequences for simultaneous source separation: Geophysics, **80**, V133–V143.

Nagi, J., F. Ducatelle, G. A. Di Caro, D. Cireşan, U. Meier, A. Giusti, F. Nagi, J. Schmidhuber, and L. M. Gambardella, 2011, Max-pooling convolutional neural networks for vision-based hand gesture recognition: , IEEE, 342–347.

Nakayama, S., and G. Blacquière, 2020, Machine-learning based data recovery and its ben-

efit to seismic acquisition: deblending, data reconstruction and low-frequency extrapolation in a simultaneous fashion: SEG Technical Program Expanded Abstracts 2020, Submitted.

Nakayama, S., G. Blacquière, and T. Ishiyama, 2019, Automated survey design for blended acquisition with irregular spatial sampling via the integration of a metaheuristic and deep learning: Geophysics, **84**, P47–P60.

Nakayama, S., G. Mercado, M. Benson, K. Belaid, and M. Garden, 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.

Oropeza, V., and M. Sacchi, 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: Geophysics, **76**, V25–V32.

Ovcharenko, O., V. Kazei, M. Kalita, D. Peter, and T. Alkhalifah, 2019, Deep learning for low-frequency extrapolation from multioffset seismic data: Geophysics, **84**, R989–R1001.

Ronneberger, O., P. Fischer, and T. Brox, 2015, U-net: Convolutional networks for biomedical image segmentation: International Conference on Medical image computing and computer-assisted intervention, 234–241.

Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, 2015, ImageNet Large Scale Visual Recognition Challenge: International Journal of Computer Vision (IJCV), **115**, 211–252.

Siahkoohi, A., R. Kumar, and F. Herrmann, 2019, Deep-learning based ocean bottom seismic wavefield recovery: SEG Technical Program Expanded Abstracts 2019, 2232–2237.

Sun, H., and L. Demanet, 2020, Extrapolated full waveform inversion with deep learning: Geophysics, **85**, 1–71.

Sun, J., S. Slang, T. Elboth, T. Larsen Greiner, S. McDonald, and L.-J. Gelius, 2020,

A convolutional neural network approach to deblending seismic data: Geophysics, **85**, WA13–WA26.

Sutskever, I., O. Vinyals, and Q. V. Le, 2014, Sequence to sequence learning with neural networks: NIPS'14 Proceedings of the 27th International Conference on Neural Information Processing Systems, **2**, 3104–3112.

Ten Kroode, F., S. Bergler, C. Corsten, J. W. de Maag, F. Strijbos, and H. Tijhof, 2013, Broadband seismic data—the importance of low frequencies: Geophysics, **78**, WA3–WA14.

Wang, B., N. Zhang, W. Lu, and J. Wang, 2019, Deep-learning-based seismic data interpolation: A preliminary result: Geophysics, **84**, V11–V20.

Wang, Y., B. Wang, N. Tu, and J. Geng, 2020, Seismic trace interpolation for irregularly spatial sampled data using convolutional autoencoder: Geophysics, **85**, V119–V130.

Wei, Z., J. Criss, A. Bull, F. Liang, and Y. Wu, 2018, The low-frequency seismic vibrator: design and experimental verification: First Break, **36**, 77–84.

Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, WA13–WA24.

Zhang, P., L. Han, Z. Xu, F. Zhang, and Y. Wei, 2017, Sparse blind deconvolution based low-frequency seismic data reconstruction for multiscale full waveform inversion: Journal of Applied Geophysics, **139**, 91–108.
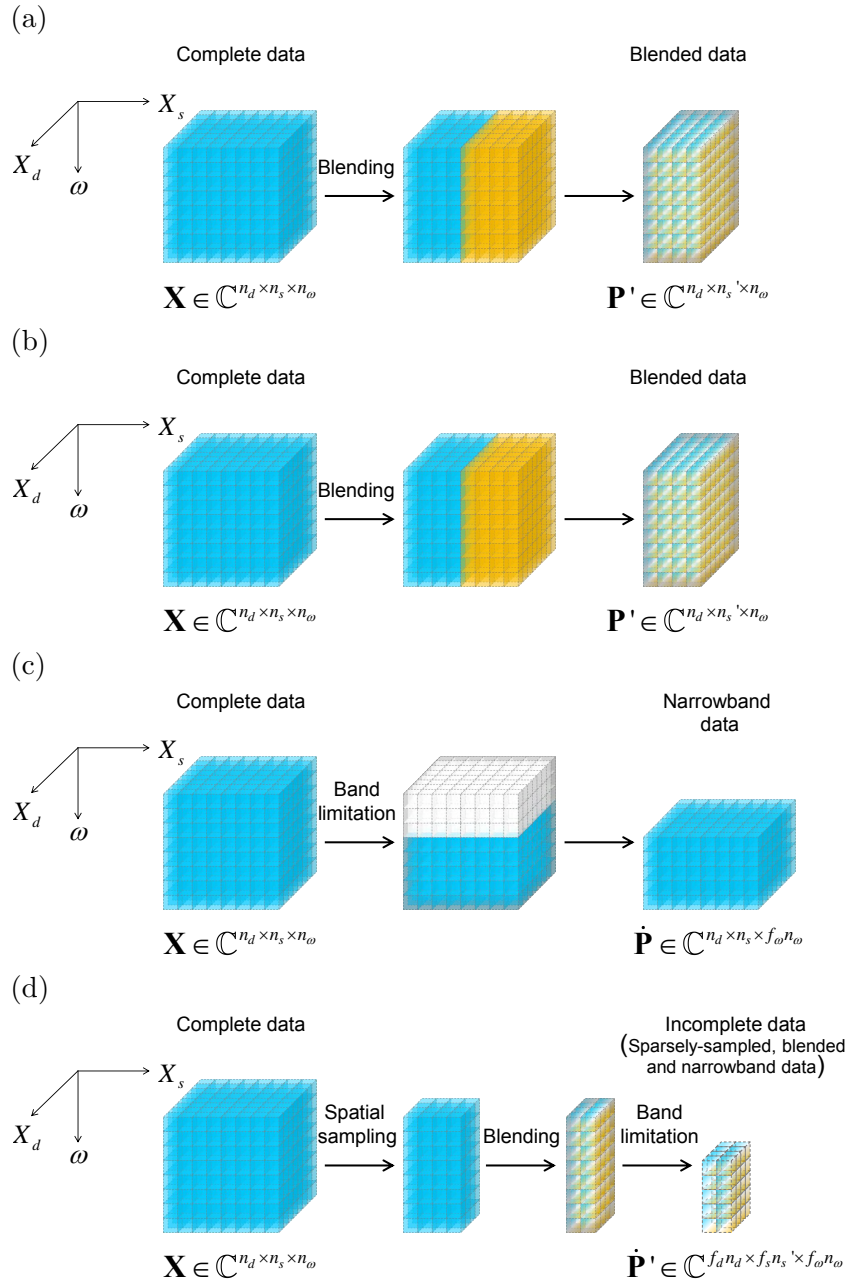
FIGURES

(a)

Complete data

X_s

X_d ω

Blending →

Blended data

→

$\mathbf{X} \in \mathbb{C}^{n_d \times n_s \times n_\omega}$

$\mathbf{P}' \in \mathbb{C}^{n_d \times n_s ' \times n_\omega}$

(b)

Complete data

X_s

X_d ω

Blending →

Blended data

→

$\mathbf{X} \in \mathbb{C}^{n_d \times n_s \times n_\omega}$

$\mathbf{P}' \in \mathbb{C}^{n_d \times n_s ' \times n_\omega}$

(c)

Complete data

X_s

X_d ω

Band
limitation →

Narrowband
data

→

$\mathbf{X} \in \mathbb{C}^{n_d \times n_s \times n_\omega}$

$\dot{\mathbf{P}} \in \mathbb{C}^{n_d \times n_s \times f_\omega n_\omega}$

(d)

Complete data

X_s

X_d ω

Spatial
sampling →

Blending →

Band
limitation →

Incomplete data
(Sparsely-sampled, blended
and narrowband data)

$\mathbf{X} \in \mathbb{C}^{n_d \times n_s \times n_\omega}$

$\dot{\mathbf{P}}' \in \mathbb{C}^{f_d n_d \times f_s n_s ' \times f_\omega n_\omega}$

Figure 1: Matrix representations indicating the effect of different acquisition strategies on the size of the data matrix in the frequency-space domain. The two horizontal axes (X_d and X_s) correspond to detector and source coordinates, while the vertical axis (ω) corresponds to frequency components. Each element indicates monochromatic seismic data acquired by a given detector-source pair. White-colored elements contain no information due to absence of detectors or sources or both, or band limitation. Different colors indicate data acquired by different sources that are activated simultaneously. Elements with mixed colors indicate blended data. The combined implementation of sparse geometry, blending and frequency limitation leads to a significant data compression.
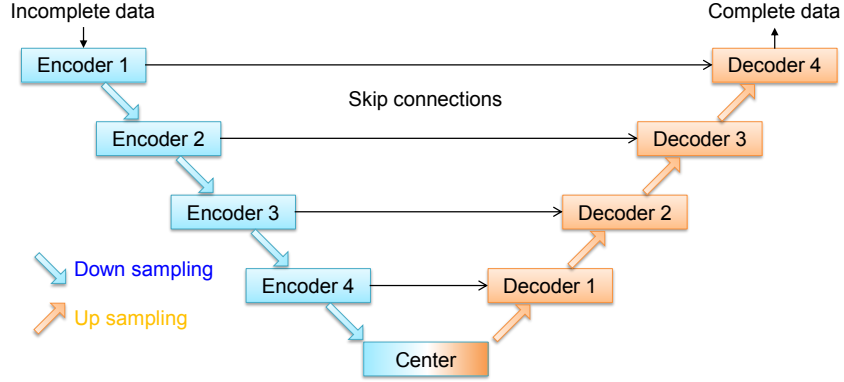
25

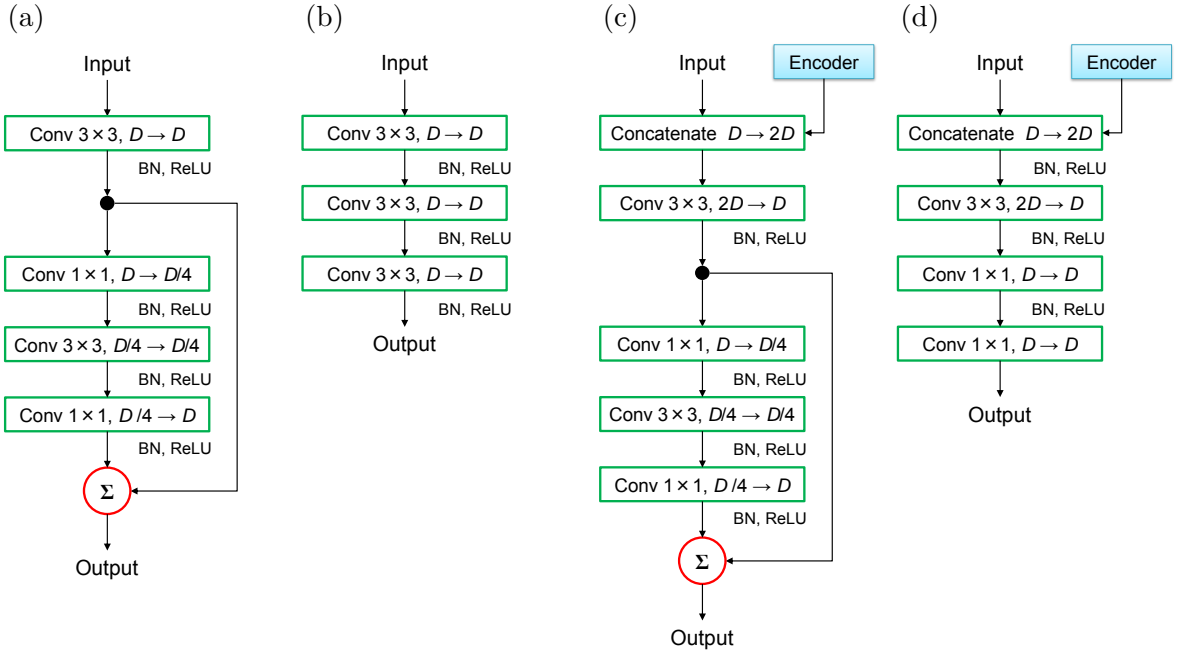Figure 2: Network architecture applied in the field data example.

Figure 3: Layers of different blocks within the applied network architecture (Figure 2). (a) Encoders 1-4. (b) Center block. (c) Decoders 1-3. (d) Decoder 4. Parameter D indicates the depth dimension of input data in each block. Conv, ReLU and BN mean a convolutional layer, a rectified linear unit and a batch normalization layer, respectively.
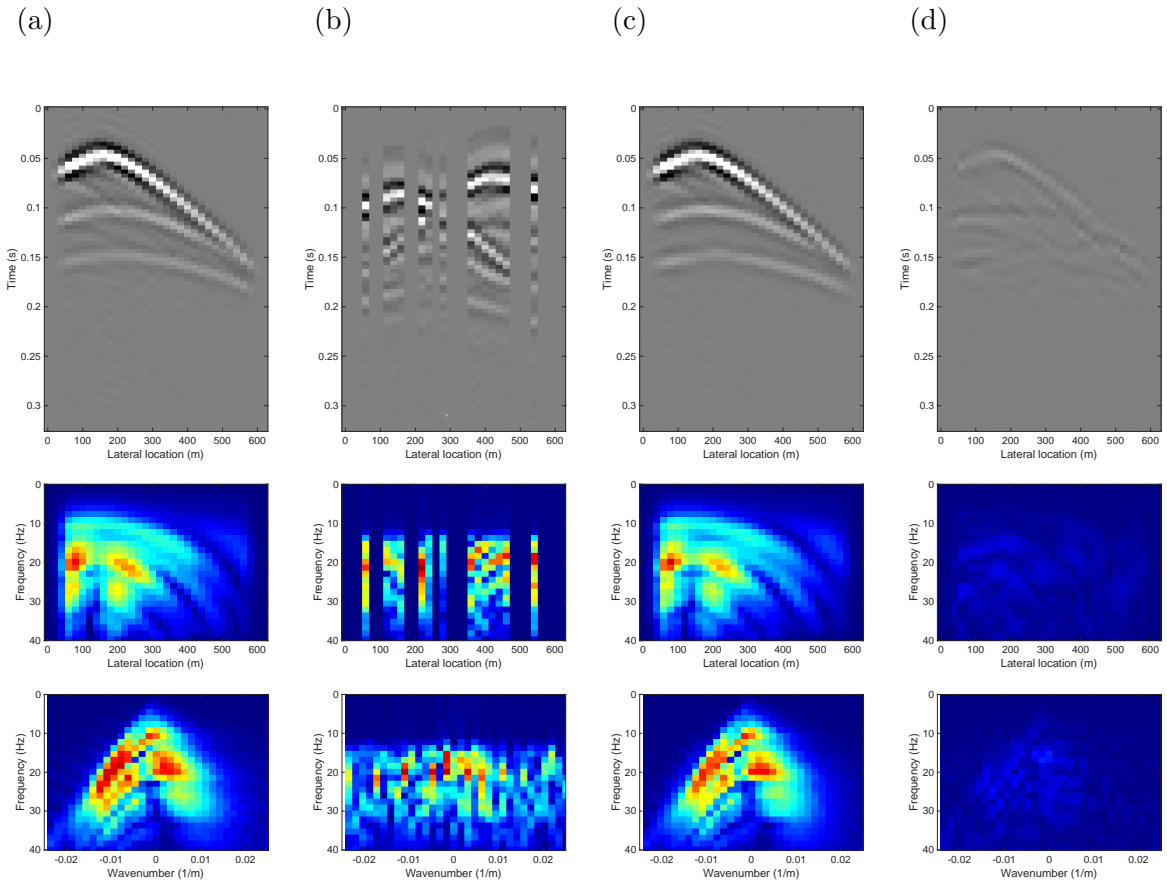
Figure 4: Data recovery results in the common shot domain (numerical example). Top, middle and bottom rows show data in the time-space, the frequency-space and the frequency-wavenumber domain, respectively. (a) Complete data. (b) Incomplete data. (c) Recovered data. (d) Residual.
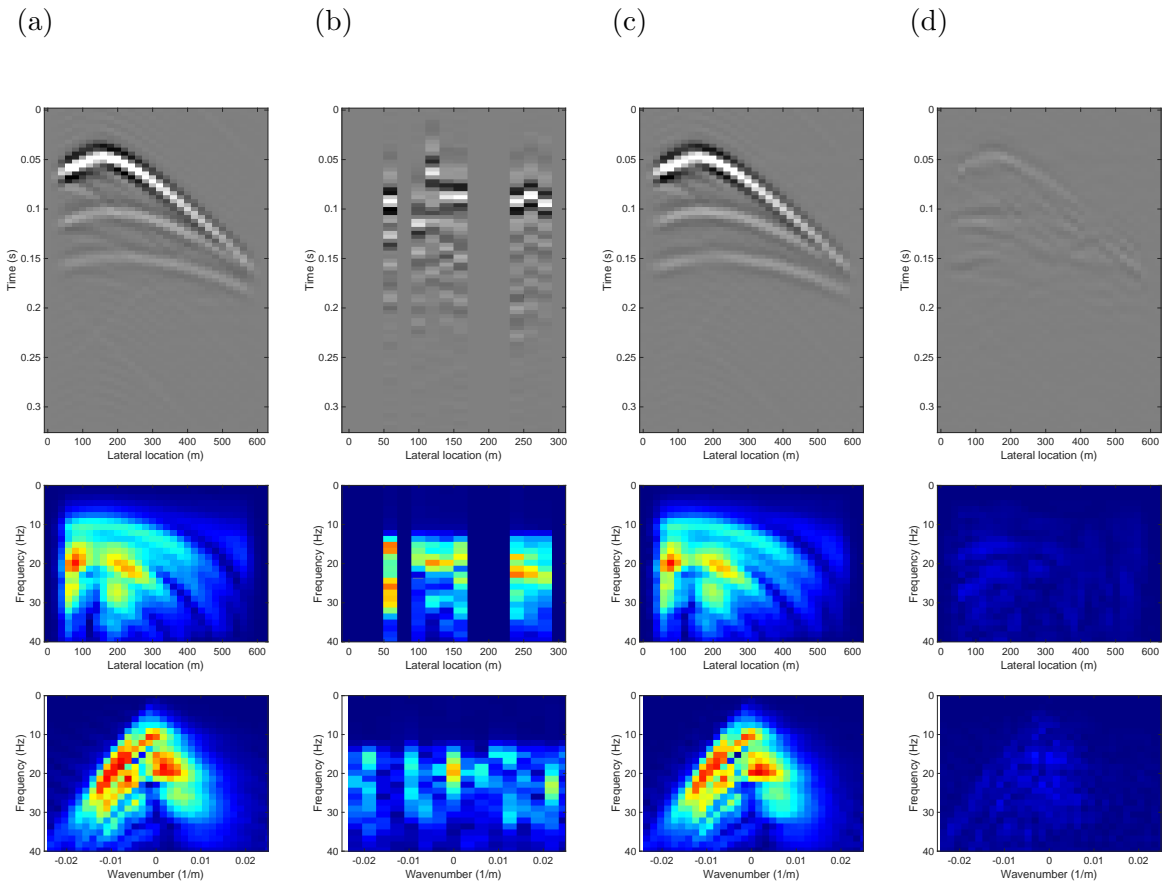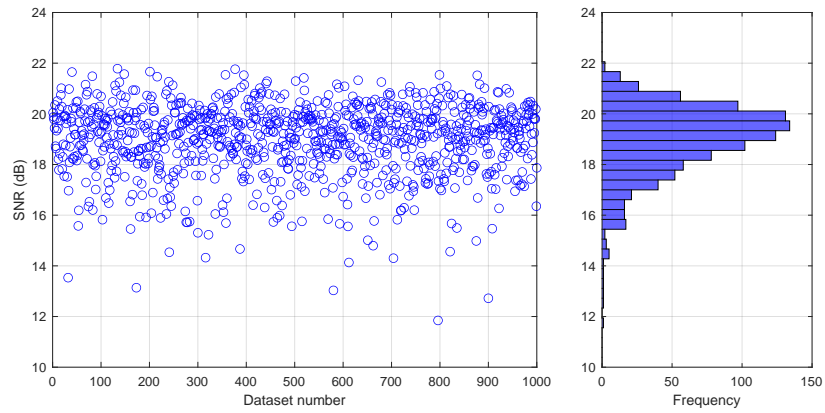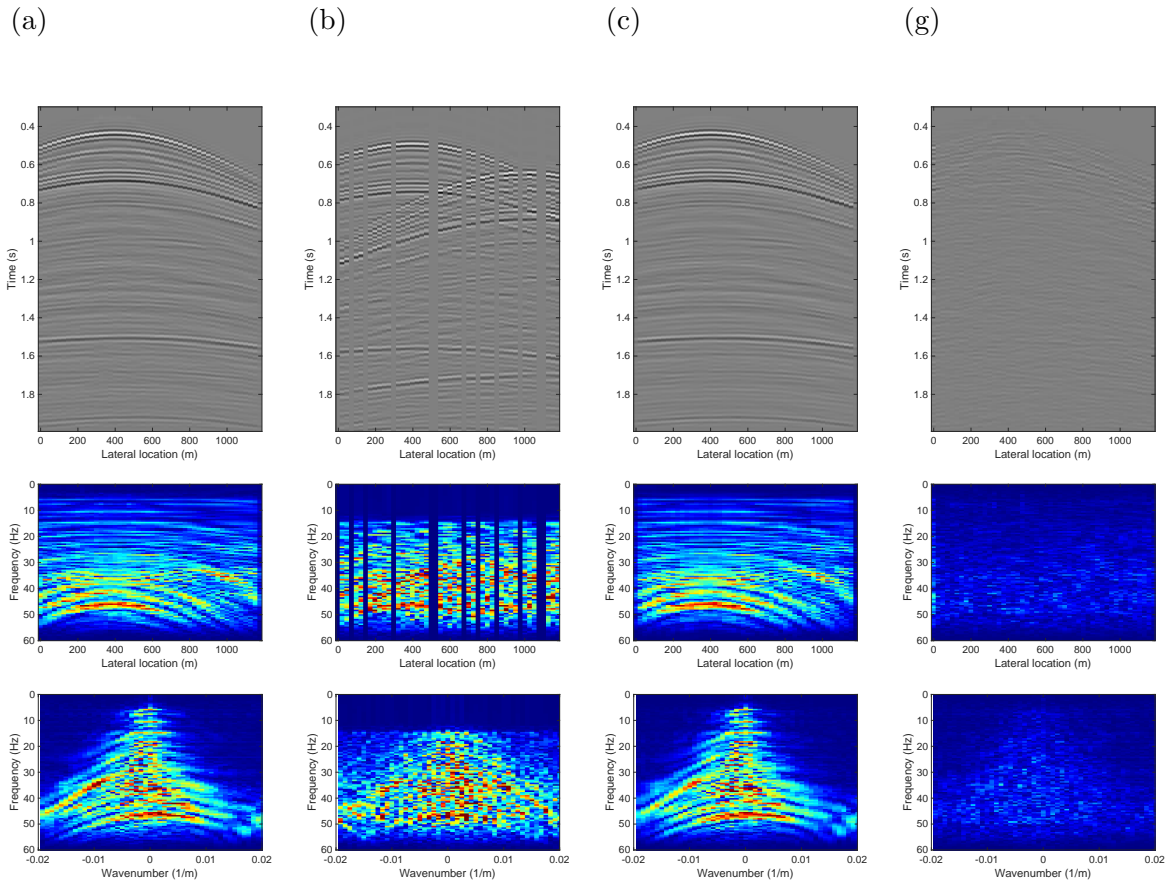
Figure 5: Data recovery results in the common detector domain (numerical example). Top, middle and bottom rows show data in the time-space, the frequency-space and the frequency-wavenumber domain, respectively. (a) Complete data. (b) Incomplete data. (c) Recovered data. (d) Residual.

Figure 6: A cross-plot (left) and a histogram (right) of SNR values from testing datasets (numerical example).

28
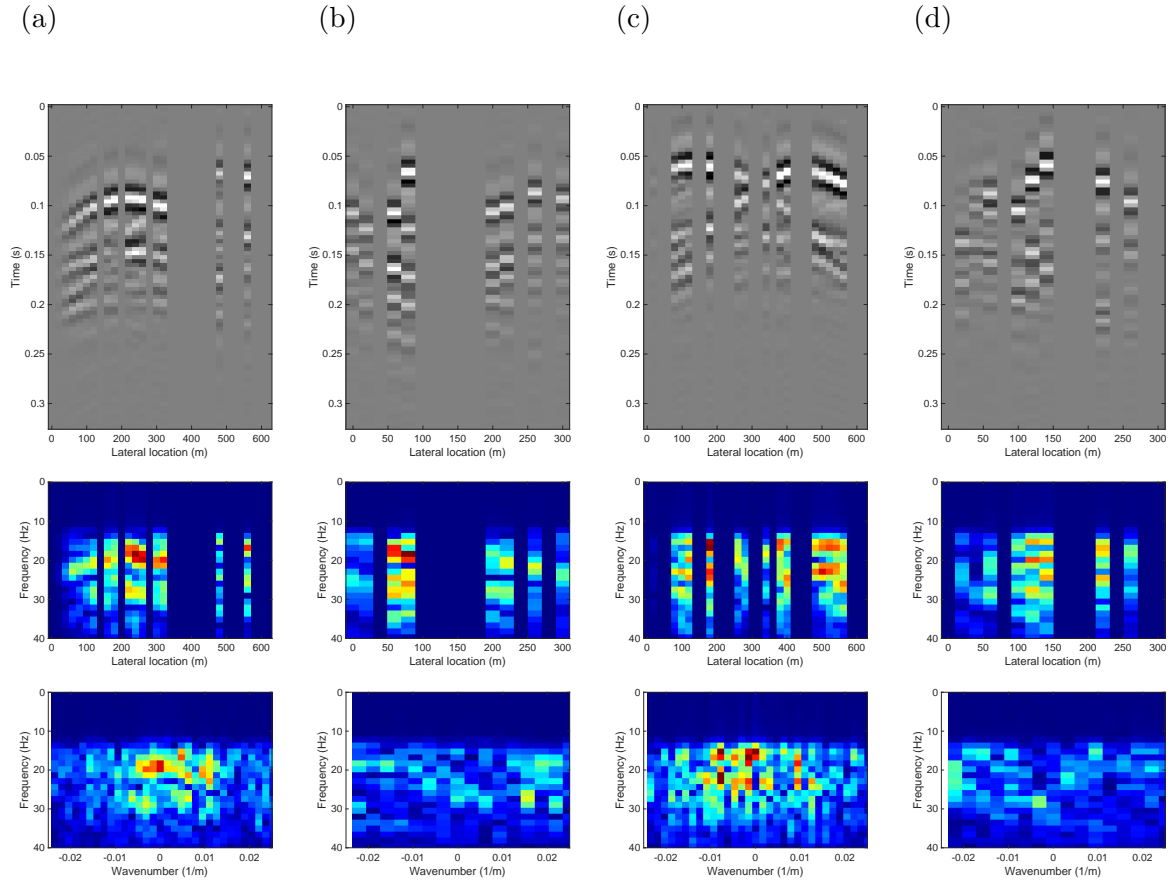
Figure 7: Incomplete data from two different acquisition scenarios in the time-space and the frequency-wavenumber domain. They are numerically simulated with the same subsurface responses. Subplots in the odd and even columns are common shot gathers and common detector gathers, respectively. They use the same numbers of detectors and sources, yet their distributions and blending codes are different. The resultant SNR values are 11.85 dB with data in (a)-(b) and 19.02 dB with data in (c)-(d), respectively. The choice of acquisition parameters accounts for this difference.

Figure 8: Data recovery results in the common shot domain (field data example). Top, middle and bottom rows show data in the time-space, the frequency-space and the frequency-wavenumber domain, respectively. (a) Complete data. (b) Incomplete data. (c) Recovered data. (d) Residual.
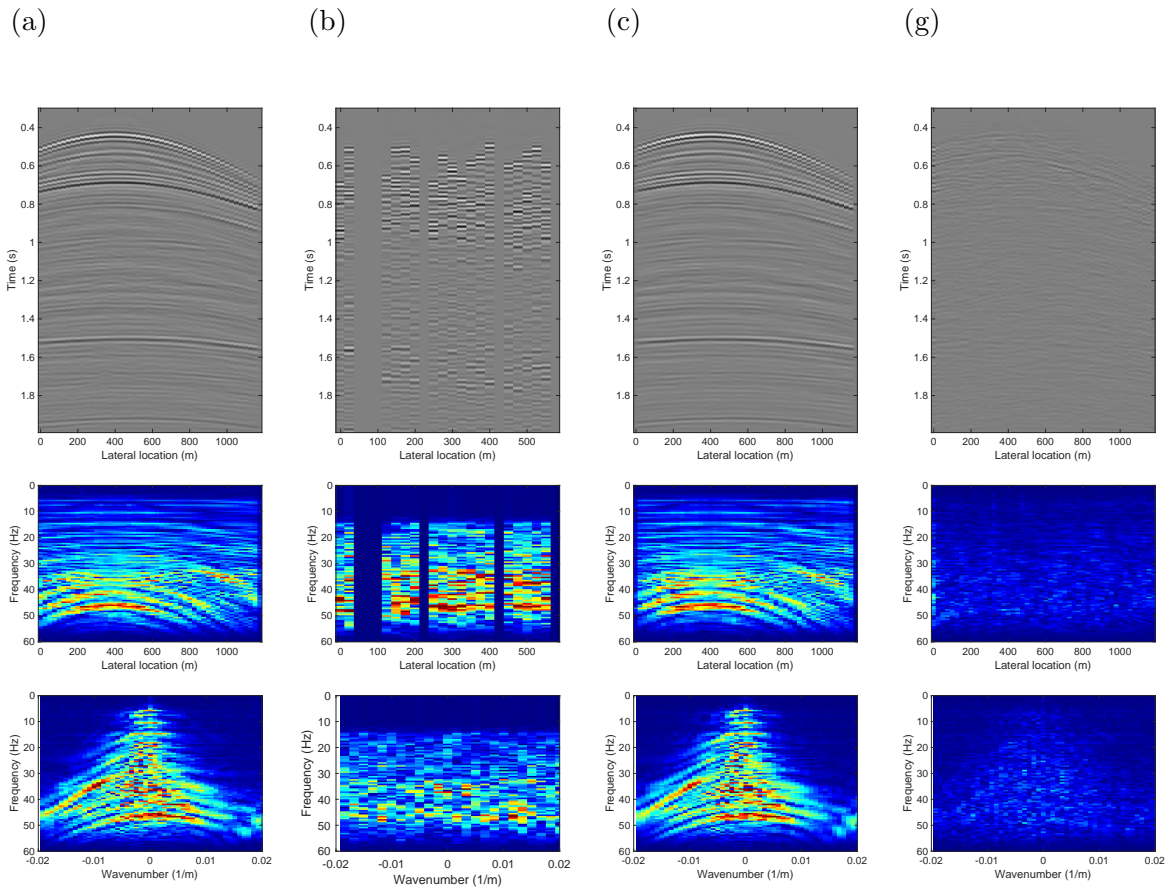
Figure 9: Data recovery results in the common shot domain (field data example). Top, middle and bottom rows show data in the time-space, the frequency-space and the frequency-wavenumber domain, respectively. (a) Complete data. (b) Incomplete data. (c) Recovered data. (d) Residual.
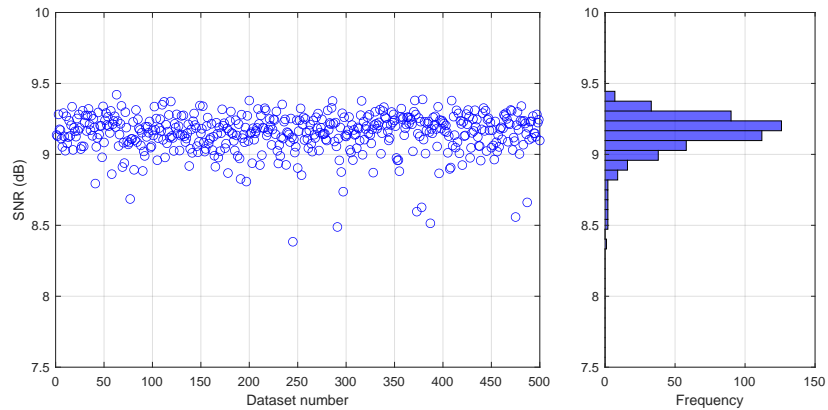

Figure 10: A cross-plot (left) and a histogram (right) of SNR values from testing datasets (field data example).
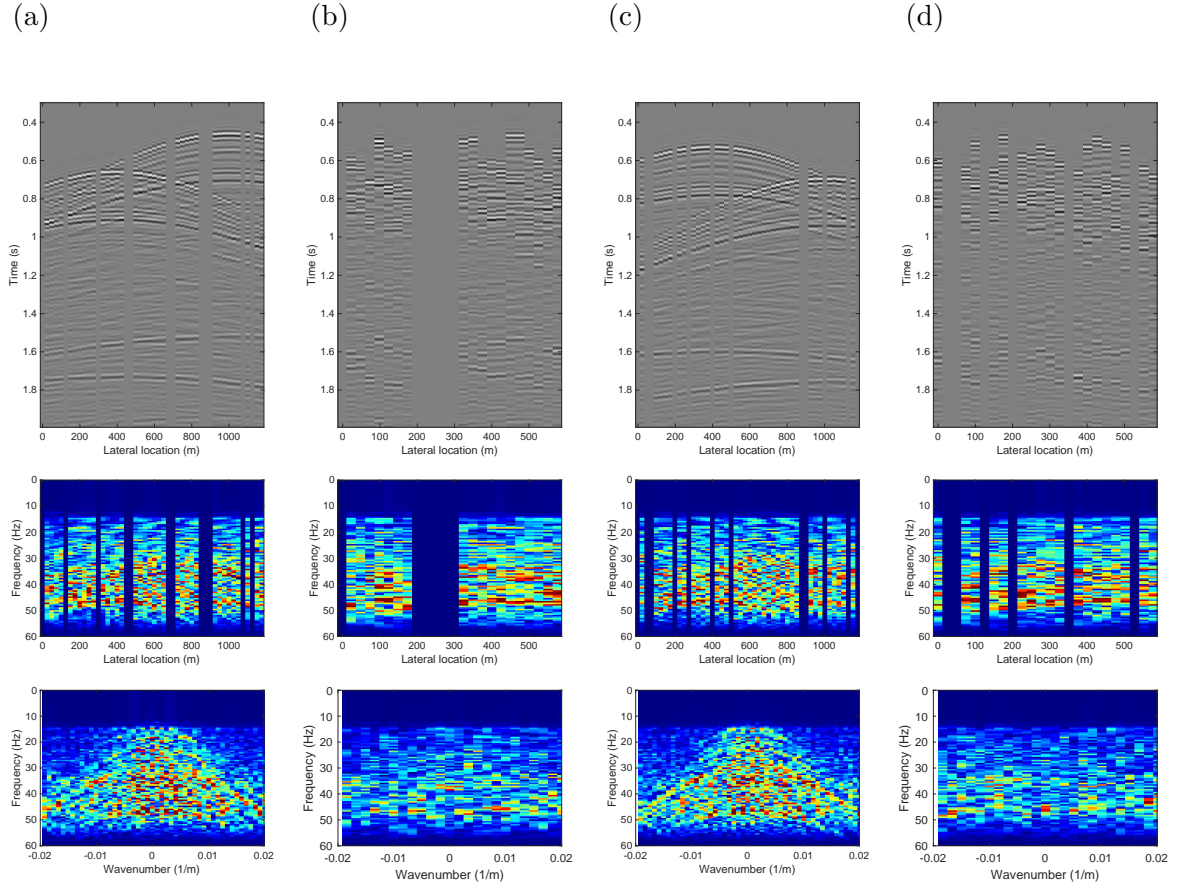
Figure 11: Incomplete data from two different acquisition scenarios in the time-space and the frequency-wavenumber domain. They are derived from the same complete data. Subplots in the odd columns and in the even columns are common shot gathers and common detector gathers, respectively. They use the same numbers of detectors and sources, yet their distributions and blending codes are different. The resultant SNR values are (a)-(b) 8.38 dB and (c)-(d) 9.42 dB, respectively. The choice of acquisition parameters accounts for this difference.