Received: 11 August 2022 | Accepted: 28 February 2023

DOI: 10.1111/1365-2478.13348

Geophysical Prospecting | EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

ORIGINAL ARTICLE

Dual-sensor wavefield separation in a compressed domain using parabolic dictionary learning

Mohammed Outhmane Faouzi Zizi[1,2] | **Pierre Turquais**[1] | **Anthony Day**[1] | **Morten W. Pedersen**[1] | **Leiv J. Gelius**[2]

[1]PGS Geophysical AS, Oslo, Norway

[2]Department of Geosciences, University of Oslo, Oslo, Norway

Correspondence
Mohammed Outhmane Faouzi Zizi, PGS Geophysical AS, Oslo, Norway.
Email: mohammed.faouzi.zizi@pgs.com

[Correction added on 30 August 2023, after first online publication: The copyright line was changed.]

Abstract

In the marine seismic industry, the size of the recorded and processed seismic data is continuously increasing and tends to become very large. Hence, applying compression algorithms specifically designed for seismic data at an early stage of the seismic processing sequence helps to save cost on storage and data transfer. Dictionary learning methods have been shown to provide state-of-the-art results for seismic data compression. These methods capture similar events from the seismic data and store them in a dictionary of atoms that can be used to represent the data in a sparse manner. However, as with conventional compression algorithms, these methods still require the data to be decompressed before a processing or imaging step is carried out. Parabolic dictionary learning is a dictionary learning method where the learned atoms follow a parabolic travel time move out and are characterized by kinematic parameters such as the slope and the curvature. In this paper, we present a novel method where such kinematic parameters are used to allow the dual-sensor (or two-components) wavefield separation processing step directly in the dictionary learning compressed domain for 2D seismic data. Based on a synthetic seismic data set, we demonstrate that our method achieves similar results as an industry-standard FK-based method for wavefield separation, with the advantage of being robust to spatial aliasing without the need for data preconditioning such as interpolation and reaching a compression rate around 13. Using a field data set of marine seismic acquisition, we observe insignificant differences on a 2D stacked seismic section between the two methods, whereas reaching a compression ratio higher than 15 when our method is used. Such a method could allow full bandwidth data transfer from vessels to onshore processing centres, where the compressed data could be used to reconstruct not only the recorded data sets, but also the up- and down-going parts of the wavefield.

KEYWORDS

compression, data processing, dictionary learning, multicomponent, seismics

INTRODUCTION

During the last decade, marine seismic acquisition capabilities have improved at a rapid pace. As a result, the size of marine seismic exploration surveys, and the number of sensors deployed have both increased. Hence, the size of the seismic data recorded during one marine seismic survey has increased significantly and often reaches several terabytes. Such large data sizes give rise to several challenges related to data transfer, storage and processing. For example: Only bandlimited data can typically be transferred from vessels to onshore processing centres because of the low-bandwidth available from satellites; storing many large seismic data sets for a long time on tapes is costly but necessary as many of them need to be backed up even after project delivery. Hence, compressing data at an early stage of the seismic processing sequence has attracted growing interest from the marine seismic community in the last few years. Indeed, designing efficient compression algorithms for seismic data is a key element to removing the barriers related to storage and data transfer. However, conventional compression processes still require the data to be decompressed before carrying out any processing or imaging steps, by transforming them back into their original time-space domain. This can limit the compression benefits related to storage or lead to accumulating compression errors if the compression process is again applied after carrying out that processing or imaging step. Moreover, conventional seismic processing and imaging is a long meticulous sequence of workflows, which generally requires transforming the data into other processing domains where data preconditioning is needed (e.g. zero padding, data interpolation, data extrapolation and decimation) and comes at a significant computational cost. In this paper, we aim to enable one seismic processing step directly in the compressed domain, namely the dual-sensor (or two-components [2C]) wavefield separation processing step. This would allow to overcome the requirements for data decompression and avoid data preconditioning with regards to this specific processing step, which is generally applied early in the seismic processing sequence.

Many compression algorithms have been designed for seismic data. Such algorithms are generally based on transforming the data to a so-called sparse domain, which is more compact than the original time-space domain. Sparse domains have been used to carry out different seismic processing and imaging applications (compressed sensing). Indeed, the compressed sensing fields have helped to tackle many difficulties related to seismic data starting from acquisition to full waveform inversion by exploiting the sparse structure of seismic data (Herrmann et al., 2013; Lin & Herrmann, 2013; Mansour et al., 2012). Conventionally, seismic compression algorithms are based on fixed sparse transforms (Averbuch et al., 2001; Duval & Rosten, 2000; Fajardo et al., 2015; Wang et al., 2004; Zheng & Liu, 2012), where the basis functions are analytically predefined and already known by the encoder and decoder, such as discrete cosines, wavelets and others (Elad, 2010; Mallat, 2008). By contrast, other seismic compression algorithms based on learned transforms have recently emerged. Schiavon et al. (2020) proposed a deep autoencoder to compress post-stack seismic data. Helal et al. (2021) proposed two convolutional autoencoders, where the first model is adapted to low compression rates (CRs), whereas the second model is more efficient when the user needs to reach high CR. These methods transform the input seismic data into feature representations which are sparse enough to allow good compression performance. Dictionary learning (DL) methods, for example *K*-mean singular value decomposition (Aharon et al., 2006) or online DL (Mairal et al., 2009), are another type of learned transform. These methods capture the similar elementary events from the seismic data, store them once in a dictionary of atoms and then express the original data as a weighted sum of the learned atoms. DL methods have been shown to provide state-of-the-art results when it comes to seismic data compression (Faouzi Zizi & Turquais, 2022). The authors have developed a compression workflow where the similarities between the different seismic events is fully exploited, and where the DL-based compression algorithm provides better compression performance compared to conventional compression methods. Moreover, different modifications of the DL methods have been shown to be suited to various seismic processing tasks, such as noise suppression (Beckouche & Ma, 2014) or interpolation (Turquais et al., 2018). For example Turquais et al. (2018) proposed a parabolic dictionary learning (PDL) method where the learned atoms represent elementary waveforms of constant amplitude along parabolic travel time move out. Hence, each atom can be characterized by a set of parameters such as the slope and the curvature, which relate to the kinematics of the wavefield (Bortfeld, 1989; Ursin, 1982). These kinematic parameters are then used to interpolate the atoms along their respective slopes, thereby reconstructing the interpolated single-component 3D streamer data in the crossline direction. These local kinematic parameters can be used not only for interpolation but also for other processing tasks such as dual-sensor wavefield separation.

In dual-sensor towed streamer acquisition, also referred as 2C streamer data, wavefield separation is the process of decomposing the data into upward and downward travelling waves using two types of sensors: hydrophones and geophones. The hydrophones record the pressure, and the geophones record the vertical component of particle velocity at the same locations. Combining both records facilitates removal of the receiver-side sea surface ghost reflection to produce data with better resolution than data with the ghost present (Söllner et al., 2008). However, for emergence angles greater than zero, the amplitudes recorded by the geophones

need to be scaled by an obliquity correction factor as only the vertical component of the particle velocity is recorded (Amundsen, 1993; Söllner et al., 2008).

In this work, we use PDL to both compress seismic data and extract the kinematic parameters from parabolic atoms such as the slope and the curvature. These kinematic parameters are further used to derive the obliquity correction factor for local events in the time domain which allows the dual-sensor wavefield separation processing step to be carried out directly in the DL compressed domain. Our PDL method for wavefield separation (WSPDL) is benchmarked against an optimized industry-standard FK method for wavefield separation (FK-WS) using a synthetic data set. Later, a field data set comprising a full 2D sail-line of marine seismic acquisition is used to assess the differences between both methods. Hence, the WSPDL method shows its robustness with respect to spatial aliasing without the need for data interpolation as for FK-WS methods. Finally, the method also succeeds in reaching high compression levels, where the compressed data can be used to reconstruct not only the recorded data sets, but also the up- and down-going parts of the wavefield.

METHODOLOGY

In order to understand how the PDL method for wavefield separation (WSPDL) enables dual-sensor wavefield separation in the compressed domain, it is first necessary to introduce the conventional dictionary learning (DL) and parabolic dictionary learning (PDL) problems.

Conventional dictionary learning and parabolic dictionary learning problems

In conventional DL problems, the aim is to represent the original data in a sparse manner with two parts: A dictionary of learned atoms representing elementary waveforms that are repeated many times in the data; and a set of sparse coefficient vectors. In the case of seismic data, M small 2D patches of size N are first extracted and then vectorized to construct a training set denoted $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M$ (Elad, 2010), which is a subset of the original data set. Then, a DL method such as the K-SVD (K-times the singular value decomposition) algorithm (Rubinstein et al., 2008) is applied to jointly: learn a redundant dictionary $\mathbf{D} \in \mathrm{R}^{N \times K}$ where $N < K$ (Donoho & Elad, 2003) of K atoms, each of size N, same size as the patches, denoted $[\mathbf{d}_1, \mathbf{d}_2 \dots \mathbf{d}_K]$, with $K < M$; and find the set of sparse coefficient vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M$ that minimize the representation error given a sparsity error ϵ imposed on the sparse coefficient vectors (Aharon et al., 2006). This approach is generally referred to as the *error constraint mode*, and this problem can be mathematically expressed by

$$\min_{\{\mathbf{x}_i\}_{i=1}^{M}, \mathbf{D}} \sum_{i=1}^{M} \| \mathbf{x}_i \|_0 \text{ subject to } \| \mathbf{y}_i - \mathbf{D}\mathbf{x}_i \|_2 \leq \epsilon, \quad i = 1, \dots M. \tag{1}$$

After learning the dictionary $\mathbf{D}$, the sparse optimization problem can be solved for patches of the original data set (Bruckstein et al., 2009). This problem is mathematically expressed as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \| \mathbf{x} \|_0 \text{ subject to } \| \mathbf{y} - \mathbf{D}\mathbf{x} \|_2 \leq \epsilon. \tag{2}$$

Equation (2) consists of finding the vector $\mathbf{x}$ of sparse coefficients that minimize the norm of the residual vector for a given patch $\mathbf{y}$ of the original data given the dictionary $\mathbf{D}$, where a sparsity error ϵ is tolerated. We can solve the problem by finding an approximate solution $\hat{\mathbf{x}}$ using orthogonal matching pursuit (Pati et al., 1993). Each patch $\mathbf{y}$ of the original data set can now be described as a linear combination of the dictionary atoms.

The PDL (Turquais et al., 2018) is a modification of the conventional DL method where a geometrical structure is imposed to the atoms while learning them. The parabolic structure was used by Turquais et al. (2018) for interpolation purposes, because it is consistent with the physics of wavefield propagation (Hoecht et al., 2009; Hubral et al., 1992; Zhang et al., 2001). The PDL problem may be mathematically expressed as follows:

$$\min_{\{\mathbf{x}_i\}_{i=1}^{M}, \{\mathbf{d}_k\}_{k=1}^{K}} \sum_{i=1}^{M} \| \mathbf{x}_i \|_0 \text{ subject to} \tag{3}$$

$$\begin{cases} \| \mathbf{y}_i - \sum_{k=1}^{K} \mathbf{d}_k \mathbf{x}_i \|_2 \leq \epsilon, \ i = 1, \dots, M \\ \mathbf{d}_k \left(t, o_k^{\mathrm{ref}}\right) = \mathbf{d}_k \left(t + c_k \Delta\mathbf{o}^2 + s_k \Delta\mathbf{o}, \ o_k^{\mathrm{ref}} + \Delta\mathbf{o}\right), \ \forall (t, \Delta\mathbf{o}) \in \mathrm{R}^2. \end{cases}$$

Equation (3) is similar to Equation (1) with an extra constraint imposed on the geometrical structure of the atoms. Here, each learned atom $\mathbf{d}_k$ is characterized by a parabolic travel time move out given by $\Delta\mathbf{t} = s_k \Delta\mathbf{o} + c_k \Delta\mathbf{o}^2$, where $\Delta\mathbf{t}$ is the time move out, s_k the atom's slope, c_k the atom's curvature and $\Delta\mathbf{o} = [o_k^1 - o_k^{\mathrm{ref}}, o_k^2 - o_k^{\mathrm{ref}}, \dots, o_k^n - o_k^{\mathrm{ref}}]$ the vector containing the displacement of a receiver location o_k^i related to each trace i of an atom k containing n traces relative to the reference receiver o_k^{ref}.

A more detailed description of the PDL problem and the method to find an approximate solution to it can be found in Turquais et al. (2018). Note that in this paper the PDL method is used with the *error constraint mode*, whereas in Turquais et al. (2018) the PDL is used with the *sparsity constraint mode*. That is because the *error constraint mode* is

more appropriate when the compressed seismic data are further processed, whereas the *sparsity constraint mode* is more appropriate for visualization purposes as explained in detail by Faouzi Zizi and Turquais (2022).

The WSPDL method

Now, we will describe and illustrate with examples the different stages of the PDL method for wavefield separation (WSPDL), where the dual-sensor wavefield separation process is applied in the DL compressed domain. First, we show how the transformation to the PDL domain is carried out. Then, we derive the obliquity correction factor for each trace of the velocity dictionary atoms. Next, we combine the corrected velocity and pressure dictionaries to obtain the up-going pressure dictionary and do the same with the velocity and pressure sparse representations to obtain the up-going pressure sparse representation. Further, compression techniques called entropy coding are used in the DL domain to encode the data with a low number of bits. Finally, we reconstruct the up-going pressure data.

Transformation stage

To describe the transformation stage of the WSPDL method, we need to consider a data set example of many shots where both the total vertical components of particle velocity and the total pressure wavefield were recorded by the geophones and the hydrophones, respectively. For reasons of simplicity, we will refer to these data sets as velocity and pressure data sets and will denote them with the indices V_z and P, respectively. The extraction and PDL stages are illustrated in Figure 1. First, a large number M of non-overlapping time-space 2D patches of size N are randomly extracted from both seismic data sets, and vectorized to obtain the training sets denoted $\mathbf{Y}_{V_z}$ and $\mathbf{Y}_P$. Then, for each training set PDL is applied as described earlier to obtain the two parabolic dictionaries $\mathbf{D}_{V_z} = [\mathbf{d}_1^{V_z}, \ldots, \mathbf{d}_K^{V_z}]$ and $\mathbf{D}_P = [\mathbf{d}_1^{P}, \ldots, \mathbf{d}_K^{P}]$ that optimally represent the training data sets in a sparse manner, where K is the number of learned atoms per dictionary. In this figure, we show only few atoms of both dictionaries. Each atom $\mathbf{d}_K$ is characterized by a set of local kinematic parameters o_k^{ref}, s_k and c_k. Further, both the velocity and pressure data sets are split into L number of 2D overlapping patches of size N as shown in Figure 2. The sparse optimization problem represented in Equation (2) can now be solved for the L patches of each of the velocity and the pressure data sets to obtain the sparse representations $\mathbf{X}_{V_z} = [\mathbf{x}_1^{V_z}, \ldots, \mathbf{x}_L^{V_z}]$ and $\mathbf{X}_P = [\mathbf{x}_1^{P}, \ldots, \mathbf{x}_L^{P}]$, given a sparsity error ϵ. Figure 2 represents the relationship between the time-space domain and the DL compressed domain. For example multiplying the sparse representation $\mathbf{x}_3^{V_z}$ (yellow frame) denoted 'Sparse 3' in the figure and the dictionary $\mathbf{D}_{V_z}$ will result in the reconstruction of the velocity patch number 3 represented with the yellow frame in the time-space domain. The atom frames in yellow, red and blue in this figure correspond to the atoms used in the different linear combinations to reconstruct the patches 3, 7065 and L, respectively. The different small squares in the sparse representations (represented with the grey scale) denote different values of non-zero coefficients used in the linear combinations. Note that the dictionary and the set of sparse representations are different for the velocity and pressure data sets.

The sparsity error ϵ is generated from a desired level of signal-to-residual ratio (SRR) of 30 dB following Equation (4), as described by Faouzi Zizi and Turquais (2022)

$$\epsilon \approx 10^{-\left(\frac{\text{SRR}}{20}\right)} \times \|\mathbf{d}_{\text{orig}}\|_2 \times \sqrt{\frac{1}{L}}, \tag{4}$$

where

$$\text{SRR} = 20\log_{10}\frac{\|\mathbf{d}_{\text{orig}}\|_2}{\|\mathbf{d}_{\text{orig}}\| - \|\mathbf{d}_{\text{rec}}\|_2}\ (\text{dB}), \tag{5}$$

with $\mathbf{d}_{\text{orig}}$ and $\mathbf{d}_{\text{rec}}$ denoting original and reconstructed data, respectively. The SRR presented in Equation (5) is a metric defining the level of desired signal distortion on the reconstructed signal. A high level of SRR = 30 dB means that a very small energy loss of 0.1% is expected between the original velocity and pressure data sets and the reconstructed ones after multiplying the dictionaries $\mathbf{D}_{V_z}$ and $\mathbf{D}_P$ with their corresponding set of sparse coefficient vectors $\mathbf{X}_{V_z}$ and $\mathbf{X}_P$. However, higher SRR leads to lower compression rate (CR), which is defined as

$$\text{CR} = \frac{\text{number of bits before compression}}{\text{number of bits after compression}}. \tag{6}$$

Explanations regarding how the data are compressed come in a later section. We will now discuss the choices for the different parameters specified in the WSPDL method based on the 100 shots data set example illustrated in Figures 1 and 2. The data are sampled at 2 ms in time and 12.5 m in space. The number of time samples (Nt) and the number of channels (Nx) are $Nt = 1850$ and $Nx = 475$, respectively. M, the number of extracted patches used to learn the dictionary, was set to 10,000 because that was found sufficient to capture most of the similar features in a data set comprising 100 shots. Although the number of atoms is generally set to be at least five times smaller than the number of training patches in conventional DL methods, here we have set K, the number of atoms per dictionary, to 6000 to enforce redundancy in the dictionary to ensure that a high level of sparsity is reached.

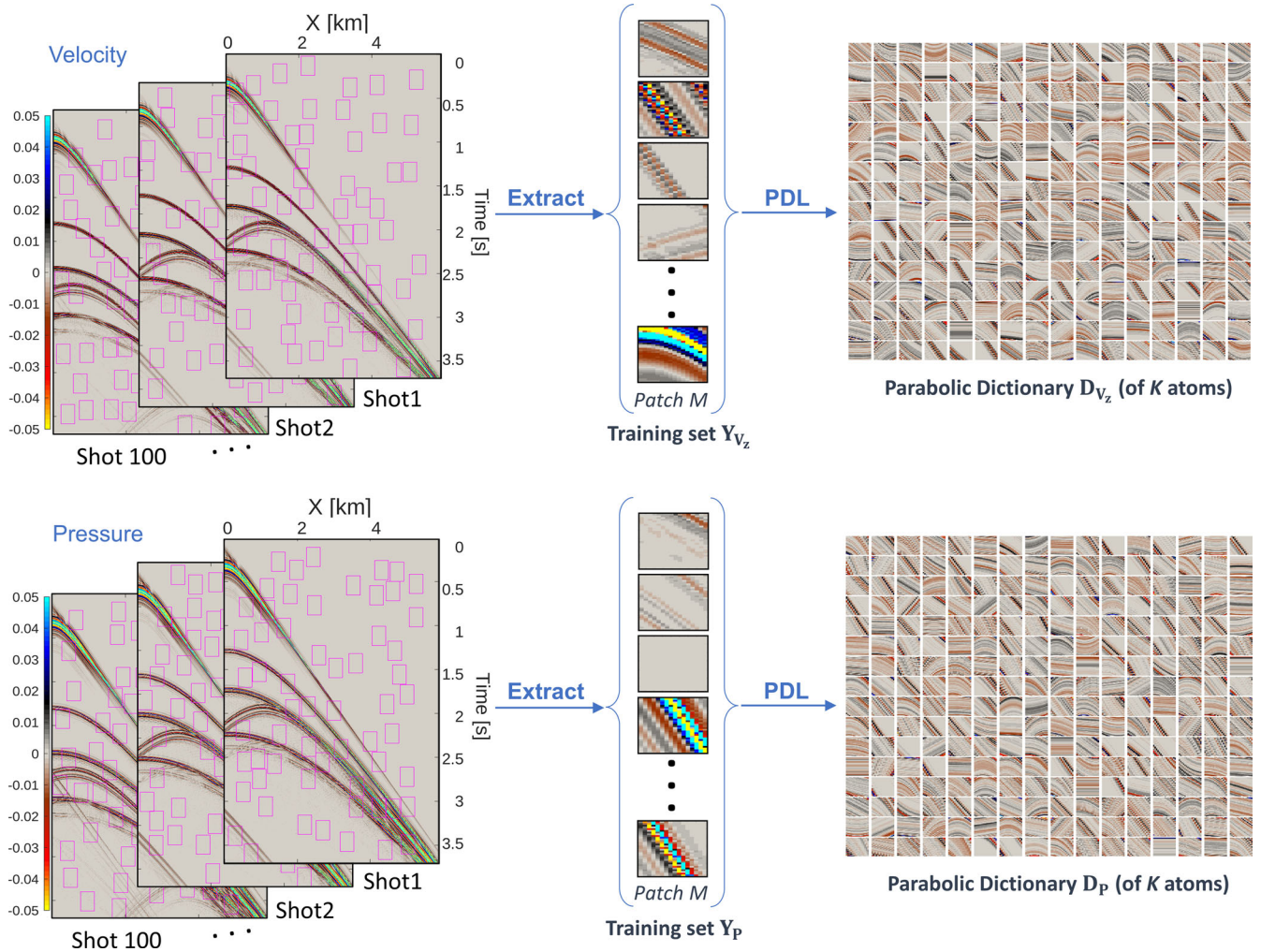

FIGURE 1 Random extraction of M non-overlapping patches from the original data sets recorded by the geophones and the hydrophones to construct the training sets $\mathbf{Y}_{V_z}$ and $\mathbf{Y}_P$, respectively. Then, parabolic dictionary learning (PDL) is applied to obtain the corresponding parabolic dictionaries $\mathbf{D}_{V_z}$ and $\mathbf{D}_P$. Only the first 256 atoms of each dictionary are represented here. The display size of the non-overlapping patches was increased for better visualization.

The greater the redundancy in a dictionary, the sparser and more accurate will be the representations (Donoho, 2006). However, increasing the size of the dictionary makes the learning stage computationally more expensive. The size of the atoms was set to 64 samples by 8 traces because it was large enough to capture the parabolic move out given the data sampling of 2 ms and 12.5 m. Note that the display of the patch size was increased in Figures 1 and 2 for better visualization. Moreover, the PDL was slightly modified here to better suit our wavefield separation application. An additional constraint was imposed where only atoms with a slope value lower than 1/1500 s/m were learned, as the apparent velocity cannot be smaller than the water velocity. This was done to avoid learning non-physical events, given that the slopes of the different atoms will later be used to correct for the obliquity scaling problem in the velocity data.

Obliquity correction

The amplitudes recorded by the geophones need to be scaled by an obliquity correction factor F, which can be expressed as follows in the following equation:

$$F = \frac{1}{\cos\theta}, \tag{7}$$

where θ is the emergence angle of each single event. This scaling is needed because the geophones record only the vertical component of particle velocity in dual-sensor acquisition. Such approach assumes that the cable is horizontal, and the propagation velocity is constant at the cable depth (Söllner et al., 2008). In the time domain, the equations for the up- and down-going pressure fields (P^{up} and P^{down}, respectively) can

Geophysical Prospecting EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

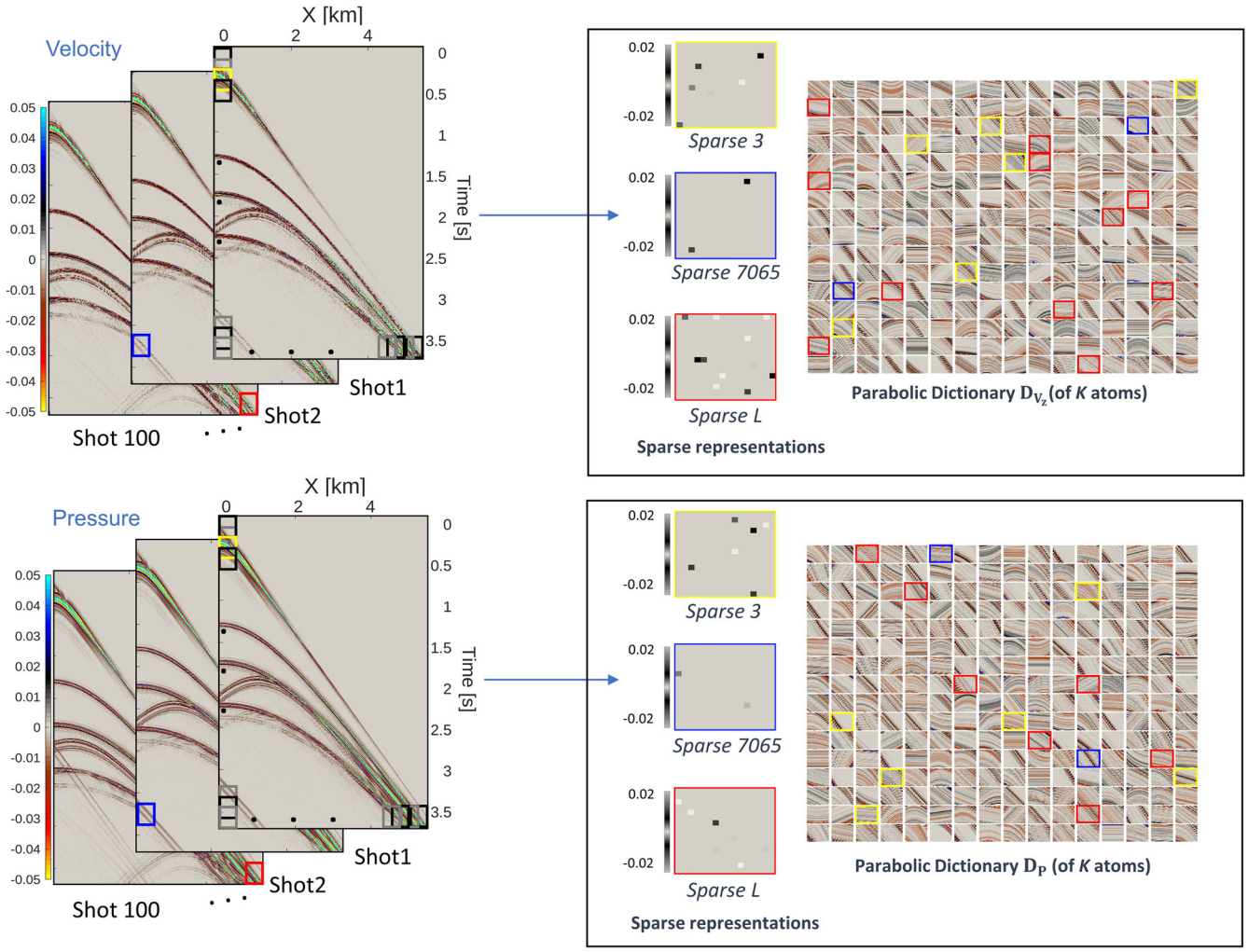

FIGURE 2 Illustration of the data sets recorded by the geophones and the hydrophones in the compressed dictionary learning (DL) domain. In this domain, each seismic data set is represented by a parabolic dictionary and a set of sparse representations. For each data set, three different sparse representations and their corresponding atoms are represented with frames of the same colour (yellow, blue and red). The display of the patch size was increased for better visualization.

be written as in the following equations:

$$P^{\text{up}} = \frac{1}{2}\left(P - \rho v F V_z\right), \tag{8}$$

$$P^{\text{down}} = \frac{1}{2}\left(P + \rho v F V_z\right), \tag{9}$$

where ρ is the water density, v is the propagation velocity in water, P is the recorded pressure and V_z is the recorded vertical particle velocity. However, it is difficult to find the obliquity correction factor F for every single event in the time domain. Hence, it is more convenient to apply the obliquity scaling after plane wave decomposition, which requires preconditioning (e.g. data interpolation, zero padding) and comes at significant computational cost. The f–k domain is conventionally used because each plane wave component maps to a particular coefficient in f–k space (Day et al., 2013).

In the f–k domain, Equations (8) and (9) can be rewritten as follows in the following equations:

$$P^{\text{up}}\left(\omega, k_x\right) = \frac{1}{2}\left(P\left(\omega, k_x\right) - \frac{\rho\omega}{k_z} V_z\left(\omega, k_x\right)\right), \tag{10}$$

$$P^{\text{down}}\left(\omega, k_x\right) = \frac{1}{2}\left(P\left(\omega, k_x\right) + \frac{\rho\omega}{k_z} V_z\left(\omega, k_x\right)\right), \tag{11}$$

where ω is the angular frequency, k_x is the horizontal angular wavenumber and k_z is the magnitude of the vertical angular wavenumber. In the WSPDL method, we use the local kinematic parameters to correct for the obliquity scaling problem in the velocity data set. Bortfeld (1989), Hubral et al. (1992) and Zhang et al. (2001) relate the parameters of the parabolic move out to the kinematics of the wavefield based on ray

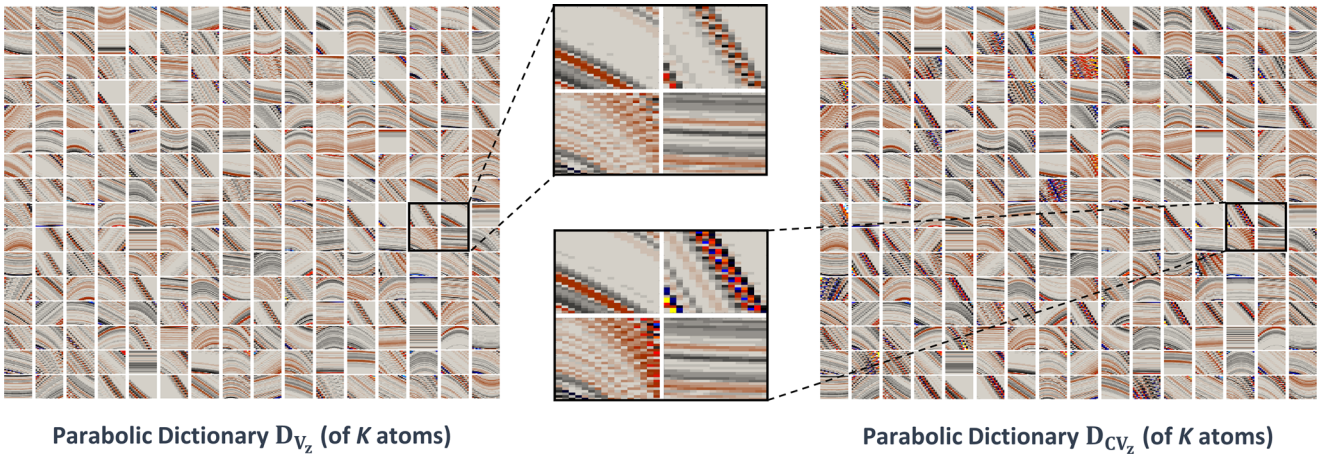

FIGURE 3 Illustration of the velocity dictionary atoms before (left) and after (right) the obliquity correction step, with close-up displays of four different atoms.

theory assumptions. Ray theory assumes the travelling event to be in the vicinity of a central ray that travelled in the subsurface, and smoothly changing amplitudes in the earth model (Bortfeld, 1989; Ursin, 1982). For example in the common shot domain, the slope s_k of an atom k can be linked to the emergence angle θ_k of the trace at the reference position o_k^{ref} as follows:

$$s_k = \frac{\sin\theta_k}{v}. \tag{12}$$

From Equations (7) and (12), we can write:

$$F_k = \frac{1}{\cos\theta_k} = \frac{1}{\sqrt{1-(vs_k)^2}}. \tag{13}$$

In Equation (13), we have found the scaling factor F_k that must be applied to each atom k of the velocity dictionary $\mathbf{D}_{V_z}$ to correct for obliquity scaling. However, the F_k factor we have found here is in fact related only to the trace located at the reference position o_k^{ref} of an atom k. Given that the atoms of our velocity dictionary contain several traces, we need to find the obliquity correction factor F_k^i related to each trace i located at each of the positions o_k^i and not only to the reference position o_k^{ref}. To do so, we consider the derivative of the parabolic time move out function $\Delta\mathbf{t} = s_k\,\Delta\mathbf{o} + c_k\Delta\mathbf{o}^2$. The obliquity correction factor for each trace i at a receiver location o_k^i in an atom k can thus be written as

$$F_k^i = \frac{1}{\sqrt{1-(vs_k^i)^2}}, \text{ where } s_k^i = 2c_k\left(o_k^i - o_k^{\text{ref}}\right) + s_k. \tag{14}$$

Figure 3 shows the velocity dictionary $\mathbf{D}_{V_z}$ on the left side as represented earlier in Figure 2. The atoms of this dictionary are characterized by a slope s_k and a curvature c_k. Hence, we can apply the obliquity correction factors F_k^i to each trace of each atom in the dictionary $\mathbf{D}_{V_z}$ following Equation (14) to obtain the corrected velocity dictionary $\mathbf{D}_{CV_z}$ represented on the right-hand side of Figure 3. The figure also shows the detail of four different atoms before and after applying the obliquity correction. These four atoms displays show that for each single trace, the higher the slope, the higher the angle of incidence and thus the higher the obliquity correction factor F_k^i, which corresponds to the dependency relationship among the obliquity correction factor, the slope and the angle of incidence described in Equations (12)–(14). Moreover, Equation (14) shows that when dealing with high slopes values, small errors in the approximation of the slope might lead to big errors in the estimation of the obliquity correction factor. Note that both dictionaries have the same set of sparse representations. Multiplying these sparse representations by the dictionary $\mathbf{D}_{CV_z}$ gives the reconstruction of the velocity data set after correcting for obliquity scaling.

Wavefield separation

After scaling the atoms of the velocity dictionary by the different obliquity correction factors following Equation (14), we end up with two parabolic dictionaries: the corrected velocity and the pressure dictionaries denoted $\mathbf{D}_{CV_z}$ and $\mathbf{D}_P$, respectively; and two sets of sparse representations denoted $\mathbf{X}_{V_z}$ and $\mathbf{X}_P$. To allow the dual-sensor wavefield separation process in the DL compressed domain, we need to express the up-going pressure field data set in terms of one dictionary and one set of sparse representations. Thus, we combine the two dictionaries $\mathbf{D}_{CV_z}$ and $\mathbf{D}_P$ into a common up-going pressure field dictionary of $2K$ atoms denoted $\mathbf{D}_{\mathbf{Pup}} \in \mathrm{R}^{N\times 2K}$

such that

$$\mathbf{D}_{\mathbf{Pup}} = \left[\mathbf{D}_{CV_z}\ \mathbf{D}_P\right] = \left[\mathbf{d}_1^{CV_z},\ \ldots,\ \mathbf{d}_K^{CV_z}, \mathbf{d}_1^{P},\ \ldots,\ \mathbf{d}_K^{P}\right], \tag{15}$$

where the first K atoms belong to the corrected velocity dictionary and the last K atoms belong to the pressure dictionary. We also combine the two sets of sparse representations $\mathbf{X}_{V_z} \in \mathrm{R}^{K\times L}$ and $\mathbf{X}_P \in \mathrm{R}^{K\times L}$ into a common up-going pressure field set of sparse representations denoted $\mathbf{X}_{\mathbf{Pup}} \in \mathrm{R}^{2K\times L}$ such that

$$\mathbf{X}_{\mathbf{Pup}} = \begin{bmatrix} -\frac{1}{2}\times\mathbf{X}_{Vz} \\ \frac{1}{2}\times\mathbf{X}_P \end{bmatrix} = \left[\mathbf{x}_1^{\mathbf{Pup}},\ \ldots,\ \mathbf{x}_L^{\mathbf{Pup}}\right], \tag{16}$$

where each sparse representation $\mathbf{x}_i^{\mathbf{Pup}}$ is a sparse vector containing $2K$ coefficients. The rescaling of $\mathbf{X}_{Vz}$ by $\frac{-1}{2}$ and $\mathbf{X}_P$ by $\frac{1}{2}$ is necessary following Equation (8). Now, each patch of the up-going pressure field data set can be described as a linear combination of the dictionary atoms stored in $\mathbf{D}_{\mathbf{Pup}}$, where coefficients of each linear combination are stored in $\mathbf{X}_{\mathbf{Pup}}$. Hence, one can stay in the DL compressed domain and further processing steps (if enabled) can directly be applied on $\mathbf{X}_{\mathbf{Pup}}$ and $\mathbf{D}_{\mathbf{Pup}}$ before having to decompress or reconstruct the data.

Data compression

Even though we are already in the DL compressed domain, we further apply mathematical compression techniques commonly known as entropy coding techniques to $\mathbf{D}_{\mathbf{Pup}}$ and $\mathbf{X}_{\mathbf{Pup}}$ as described in detail by Faouzi Zizi and Turquais (2022). The entropy techniques are applied in two main stages: the quantization and coding stages. The quantization stage is the process of mapping floating point sparse coefficients within a range $[a, b]$ into a finite set of output levels; that is the discretization of a continuous amplitude scale. We adopt the same approach as Faouzi Zizi and Turquais (2022) and rescale the dynamic range of the sparse coefficients by multiplying them by a defined scaling coefficient Sc, which is automatically generated from the level of desired SRR and the norm of the original data $\mathbf{d}_{\mathrm{orig}_2}$. Sc can be mathematically expressed as follows:

$$\mathrm{Sc} = \frac{0.25\times\sqrt{Nb}}{0.2\times 10^{-\left(\frac{\mathrm{SRR}}{20}\right)}\times\left\|\mathbf{d}_{\mathrm{orig}}\right\|_2}, \tag{17}$$

where Nb refers to the total number of non-zero coefficients in the sparse representations. In our method we do the same for the set of sparse representations $\mathbf{X}_{\mathbf{Pup}}$. A fixed rescaling factor $\mathrm{Sc} = 10^3$ is also applied to all atoms of the dictionary $\mathbf{D}_{\mathbf{Pup}}$ because the dictionary atoms are normalized and do not depend on the scale of the data. Then, we apply a rounding to obtain integer values that can be coded using a small number of bits. For the coding stage, the set of sparse representations $\mathbf{X}_{\mathbf{Pup}}$ is coded alongside the dictionary $\mathbf{D}_{\mathbf{Pup}}$ using a Huffman coding scheme (Huffman, 1952). The coding stage is a lossless compression strategy that seeks to represent the data with the lowest number of bits per symbol. More details on applying entropy coding techniques in the DL domain can be found in Faouzi Zizi and Turquais (2022).

Data reconstruction

First, we apply the inverse operations of the quantization and coding stages as described in detail by Faouzi Zizi and Turquais (2022) to obtain the reconstructed $\hat{\mathbf{D}}_{\mathbf{Pup}}$and $\hat{\mathbf{X}}_{\mathbf{Pup}}$. Then, we can easily reconstruct the total vertical particle velocity, the total pressure, the up-going pressure or the down-going pressure seismic shot gathers. For example: multiplying $\hat{\mathbf{D}}_{\mathbf{Pup}}$ and $\hat{\mathbf{X}}_{\mathbf{Pup}}$ will reconstruct the up-going pressure seismic shot gathers; multiplying $\hat{\mathbf{D}}_{\mathbf{Pup}}$ and $\hat{\mathbf{X}}_{\mathbf{Pup}}$, after rescaling the first K coefficients of the sparse vectors in $\hat{\mathbf{X}}_{\mathbf{Pup}}$ denoted $\hat{\mathbf{X}}_{\mathbf{Pup}}[1:K,:]$ by -1 following (16), will reconstruct the down-going pressure seismic shot gathers; multiplying the first K atoms of $\hat{\mathbf{D}}_{\mathbf{Pup}}$ denoted $\hat{\mathbf{D}}_{\mathbf{Pup}}[:, 1:K]$ and the first K coefficients of the sparse vectors in $\hat{\mathbf{X}}_{\mathbf{Pup}}$ denoted $\hat{\mathbf{X}}_{\mathbf{Pup}}[1:K,:]$, after rescaling $\hat{\mathbf{X}}_{\mathbf{Pup}}[1:K,:]$ by a factor of -2 following (16), would reconstruct the corrected velocity seismic shot gathers; finally, multiplying the last K atoms of $\hat{\mathbf{D}}_{\mathbf{Pup}}$ denoted $\hat{\mathbf{D}}_{\mathbf{Pup}}[:, K+1:2K]$ and the last K coefficients of the sparse vectors in $\hat{\mathbf{X}}_{\mathbf{Pup}}$ denoted $\hat{\mathbf{X}}_{\mathbf{Pup}}[K+1:2K,:]$, after rescaling $\hat{\mathbf{X}}_{\mathbf{Pup}}[K+1:2K,:]$ by a factor of 2, would reconstruct the pressure seismic shot gathers. Figure 4 shows the reconstructed up-going pressure sparse representation number 3 denoted $\hat{\mathbf{x}}_3^{\mathbf{Pup}}$, which corresponds to the combination of the velocity and pressure sparse representations $\hat{\mathbf{x}}_3^{V_z}$ and $\hat{\mathbf{x}}_3^{P}$, respectively, and the reconstructed dictionary $\hat{\mathbf{D}}_{\mathbf{Pup}}$ corresponding to the combination of the corrected velocity dictionary $\hat{\mathbf{D}}_{CV_z}$ and the pressure dictionary $\hat{\mathbf{D}}_P$, respectively. Multiplying $\hat{\mathbf{D}}_{\mathbf{Pup}}$ and $\hat{\mathbf{x}}_3^{\mathbf{Pup}}$gives the linear combination of atoms needed to reconstruct the up-going pressure patch number 3. Note that, as only the 256 first atoms of the 6000 atoms of each dictionary are displayed in Figure 4, more atoms are in fact needed to obtain the final up-going pressure field patch displayed in the figure.

DATA APPLICATION

Synthetic data

In this section, we will assess the capability of the PDL method for wavefield separation (WSPDL) method to correct for velocity obliquity scaling and to reconstruct the

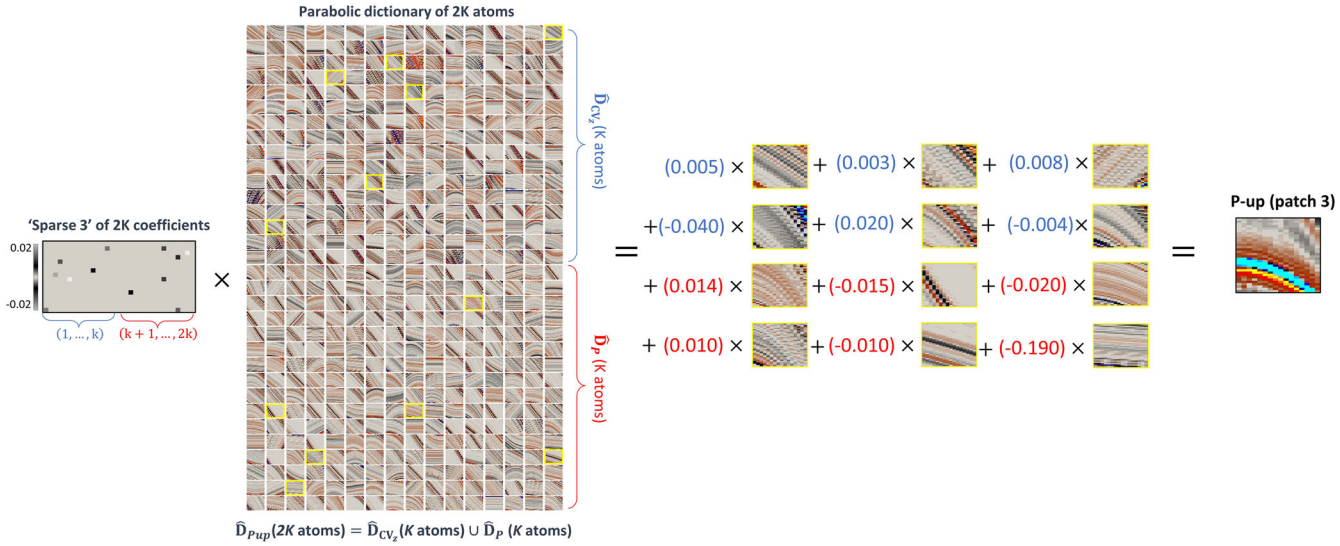

FIGURE 4 Reconstruction of the up-going pressure patch 3 by multiplying the up-going pressure dictionary, which is a combination of the corrected velocity and pressure dictionaries, and the up-going pressure sparse representation, which is a combination of the velocity and pressure sparse representations.

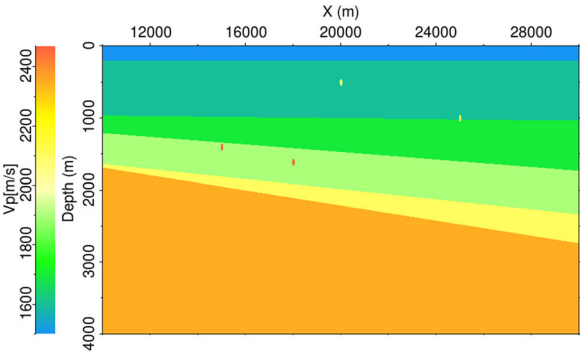

FIGURE 5 2D P-wave velocity model used to generate the synthetic data example of 100 shots. Four diffracting points are represented in the model at different depths with small oval shapes. The depth of the source is 6 m, and the depth of the cable is 20 m, where the receiver spread is fixed.

up- and down-going pressure fields. We have modelled a synthetic data set, where we have access to the true result, and compared our method to an optimized industry-standard FK method (FK-WS).

The total pressure and total vertical particle velocity wavefields were modelled using acoustic finite difference modelling for the 2D P-wave velocity model shown in Figure 5, which comprises a 200 m thick water layer and five sedimentary layers. The different layers have different boundary dips and velocities to create dipping events with different slopes. The density increases slightly with depth and goes from 1.96 g/cm^3 in the first sedimentary layer to 2.11 g/cm^3 in the fifth and last one (Figure 5). In addition, diffracting points have been added to create conflicting dips in the shot gathers which are challenging to reconstruct. The up-going parts of the pressure and vertical particle velocity fields were first modelled without free-surface effects. Then, the down-going parts of the pressure and vertical particle velocity fields, which are reflected at the sea surface, were modelled using virtual receivers located in a mirror-image position relative to the sea surface. Further, the direct arrival was also modelled and subtracted from the four data sets. We then combined the up- and down-going parts of the pressure and the up- and down-going parts of the vertical particle velocity wavefields to simulate the total pressure and the total vertical particle velocity measurements recorded by the hydrophones and the geophones, respectively. Hence, we have produced three data sets each of 100 shots: the input vertical particle velocity and pressure measurements, and their corresponding up-going pressure field which will be used as reference (true result).

The modelled data sets were sampled at 2 ms in time and 3.125 m in space. The depth of the source was 6 m, and the depth of the receivers was 20 m. Source ghosts and free-surface multiples were not modelled for the three data sets of 100 shots. Moreover, the modelled data sets underwent pre-processing, as shown in Figure 6, in preparation for the application of the wavefield separation process using the WSPDL and FK-WS methods. The performance of these methods is then evaluated and compared. Here, we consider three data sets: the recorded total vertical particle velocity, the recorded total pressure and the reference up-going pressure wavefield. First, we applied a 20 Hz low-cut filter to the three data sets because low frequencies are handled by other methods when dual-sensor wavefield separation is applied to field data examples. For low-frequencies, typically below 20 Hz, there is no decomposition of the wavefield into

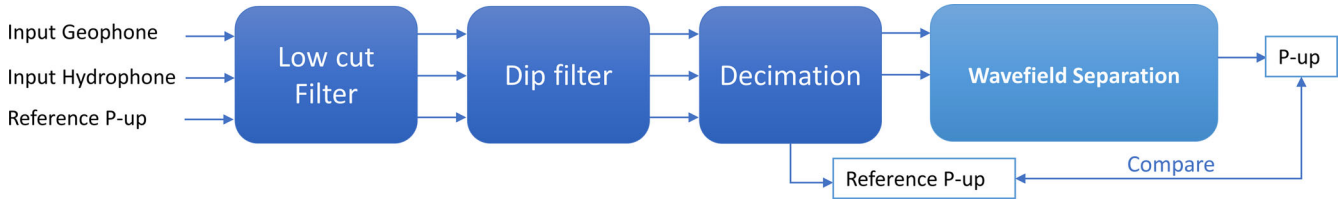

FIGURE 6 Pre-processing workflow for synthetic data before applying the dual-sensor wavefield separation process.

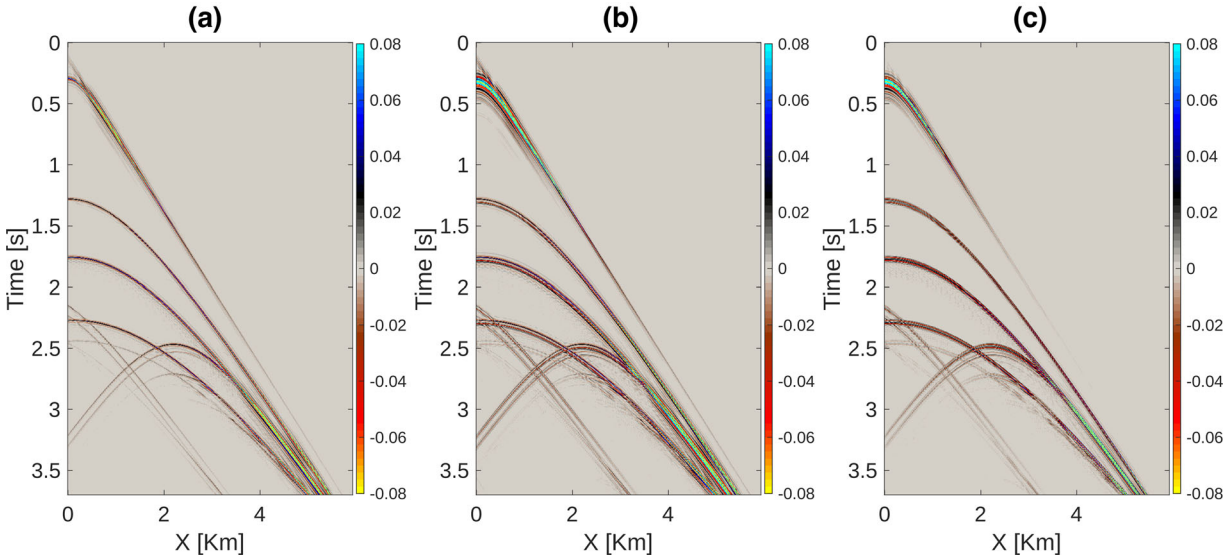

FIGURE 7 One seismic shot gather of the 100 shot synthetic data example: (a) reference up-going pressure field; (b) total pressure measured by the hydrophones and (c) total vertical particle velocity component measured by the geophones.

up- and down-going parts. Indeed, the velocity data are heavily contaminated by noise at the lowest frequencies. Thus, this problem is usually solved by the so-called low-frequency conditioning, where the noisy low frequency velocity data are reconstructed from the relatively clean pressure data (Day et al., 2013). Then, we applied a dip filter to remove steeply dipping events with apparent propagation velocities of less than 1550 m/s. This was done to allow a fair comparison of the WSPDL method and the industry-standard FK-WS method in terms of wavefield separation performance. Indeed, a tapering is generally applied at such dips in conventional FK-WS methods to avoid instability arising from the application of large, rapidly varying obliquity scaling factors. Finally, we decimate the data by a factor of 4 in the space dimension to obtain data sets sampled at 12.5 m, normally employed for field data. Figure 7 shows the same example shot gather for the three data sets: the reference up-going pressure field (Figure 7a), the input pressure (Figure 7b) and the input velocity (Figure 7c) after the decimation step but before applying the dual-sensor wavefield separation process (Figure 6).

After pre-processing the data sets, we apply the wavefield separation process to the synthetic input pressure and velocity recordings and compare the resulting up-going wavefield to the reference up-going wavefield. The wavefield separation process was carried out using the WSPDL method, then using the optimized industry-standard FK-WS method without interpolation and a third time using the same FK-WS method after interpolating the data to a 3.125 m spatial sampling interval using an industry-standard interpolation algorithm to handle the aliased events (Bekara & Robin, 2015). In real field data cases, the FK-WS method is used after interpolation as the data are more commonly acquired at 12.5 m. Figure 8 shows the residual difference between the reference up-going pressure wavefield (Figure 8a) and the combination of the input pressure (Figure 7b), and velocity (Figure 7c) when: No obliquity scaling is applied (Figure 8b), the industry-standard FK-WS method is used without interpolation (Figure 8c), the industry-standard FK-WS method is used with interpolation (Figure 8d) and the WSPDL method as described in the methodology section is used (Figure 8e). Figure 9 shows the corresponding *f–k* spectra of the shot gathers displayed in Figure 8. We compute the signal-to-residual ratio (SRR) metric for the different methods following Equation (5), to assess the quality of the wavefield separation

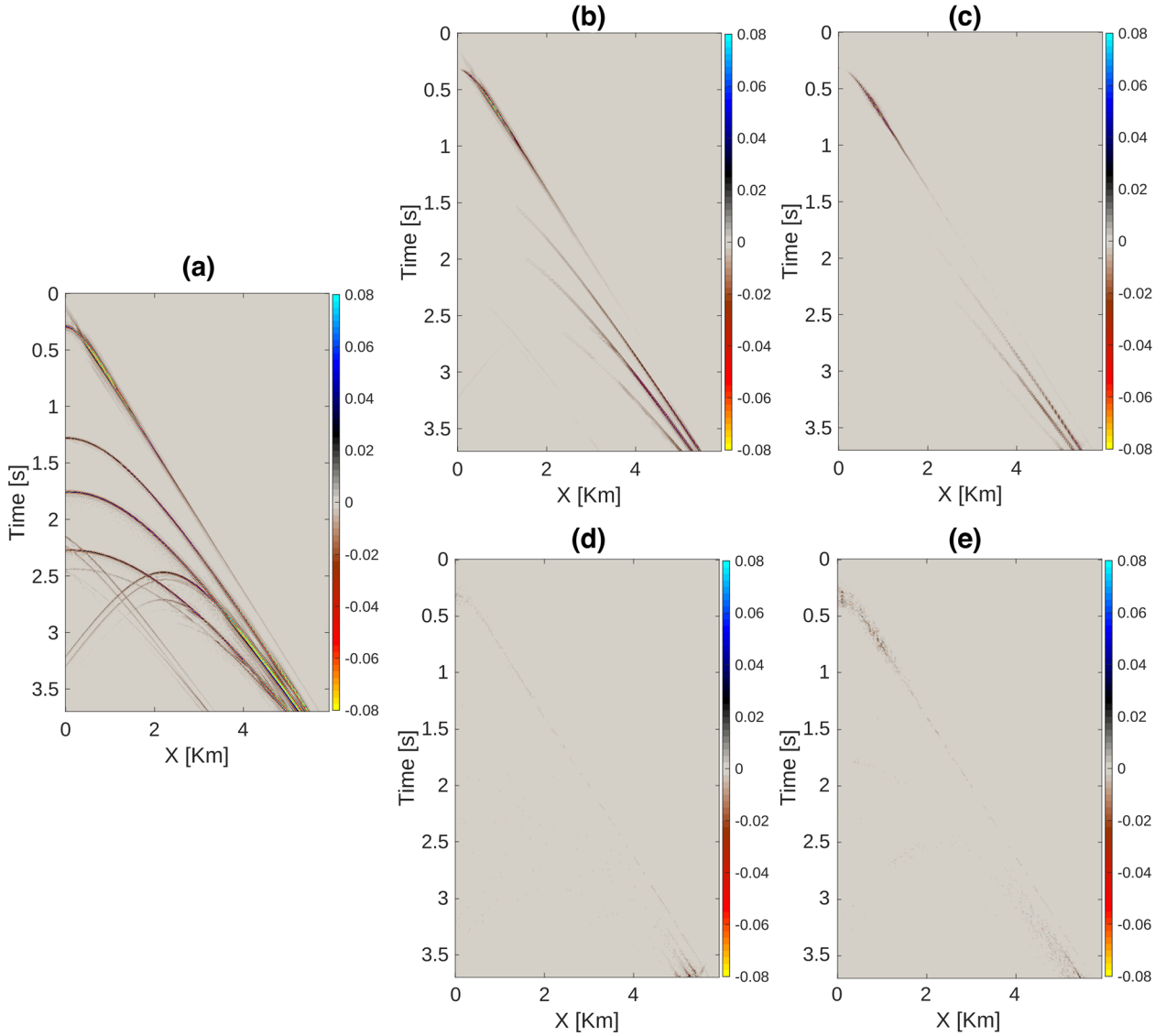

FIGURE 8 Wavefield separation results for the shot gather represented in Figure 7: (a) reference up-going pressure field; (b) residuals when no obliquity scaling is applied; (c) residuals after applying the FK-WS method without interpolation; (d) residuals after applying the FK-WS method with interpolation and (e) residuals after applying our WSPDL method.

results, with $\mathbf{d}_{\text{orig}}$ being the reference up-going pressure field (Figure 8a) and $\mathbf{d}_{\text{rec}}$ the reconstructed up-going pressure field after applying one of the wavefield separation methods. Figures 8b and 9b show visually the results obtained in case an obliquity correction factor F of 1 is employed with Equation (8). In this case, we obtain a SRR value of 10.1 dB. If no interpolation is used prior to the industry-standard FK-WS method (Figures 8c and 9c), the aliased events are not well reconstructed and a SRR of 16.54 is reached. However, when the same FK-WS method is used after interpolation (Figures 8d and 9d) the residuals are negligible and a high level of SRR = 27.64 dB is reached. From Figure 8e, we can clearly see that the WSPDL method has also succeeded in reconstructing the up-going pressure field well and reaches a high level of SRR = 24.43 dB, and with only weak residuals associated with very high amplitude events. Thus, although the industry-standard FK-WS method (Figures 8d and 9d) achieves slightly better results when interpolation is included compared to the WSPDL method (Figures 8e and 9e), both methods achieve high quality wavefield separations. Moreover, our method has shown its effectiveness in reconstructing aliased events without the need for interpolation. That is because parabolic dictionary learning (PDL) reconstructs aliased events under its own set of assumptions, namely sparsity and parabolic constraint. Under these assumptions, PDL has shown the ability to provide better interpolation results compared to a standard FK-based interpolation method (Turquais et al., 2019). In addition, our method has accomplished the dual-sensor wavefield separation process in the dictionary learning (DL) compressed

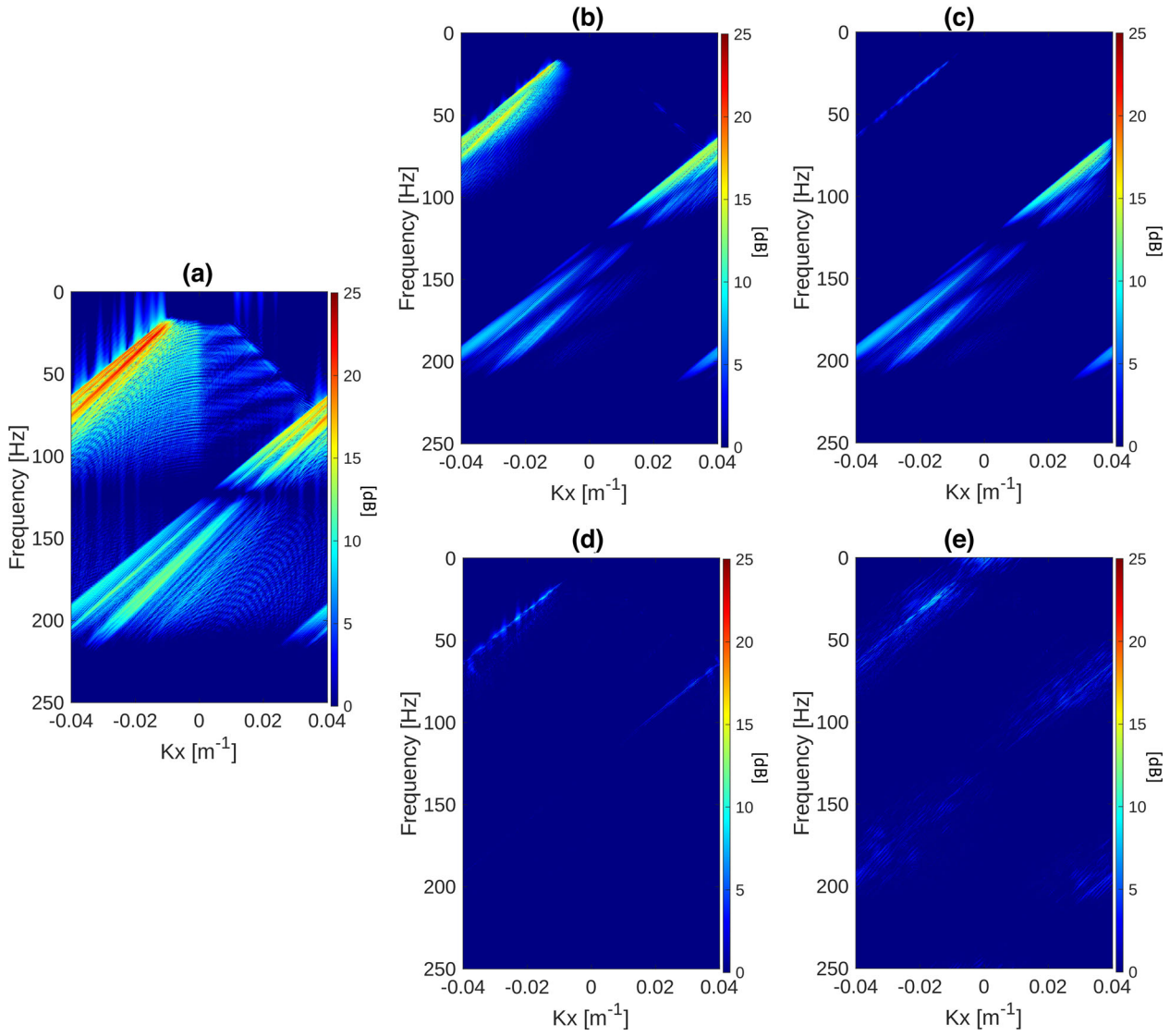

FIGURE 9 The *f–k* spectra of the wavefield separation results represented in Figure 8: (a) reference up-going pressure field; (b) residuals if no obliquity correction is applied; (c) residuals after applying the FK-WS method without interpolation; (d) residuals after applying the FK-WS method with interpolation and (e) residuals after applying our WSPDL method.

domain with CR = 13.13, that is the data used to reconstruct the up-going pressure field are 13 times smaller than the input velocity and pressure data sets. The reconstruction results obtained using the different methods are summarized in Table 1. In this synthetic data example, we have in fact compressed the input velocity and pressure data sets and then reconstructed them prior to using both FK-WS methods (with and without interpolation). This procedure was chosen to ensure that the data input to the FK-WS methods include any errors introduced by the compression step, thereby permitting a fair assessment of the effectiveness of the WSPDL method to correct for the obliquity scaling problem itself. The WSPDL method compresses the data and corrects for the obliquity scaling problem simultaneously in the compressed domain. Hence, it is not possible to apply the wavefield separation process without compressing the data when using WSPDL. If no compression and reconstruction are applied to the input velocity and pressure data sets, the results of the optimized industry-standard FK-WS method when used with interpolation reach a SRR level of 34.02 dB, and when no interpolation is used it reaches a level of SRR = 16.79 dB. The computational costs of the FK-WS method with interpolation and the WSPDL are comparable. However, when using WSPDL, the compression is the most expensive part, whereas the reconstruction is very efficient and fast as it requires only matrix multiplications as described earlier in the methodology section. This means that once the data are in the compressed domain, one can also reconstruct the down-going pressure field, the pressure or the corrected vertical particle velocity with almost no extra cost.

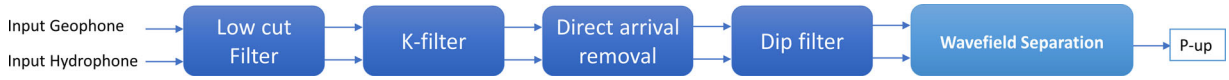

FIGURE 10 Different pre-processing steps before applying the dual-sensor wavefield separation to the field data example of 3320 shots.

TABLE 1 Comparison of the different wavefield separation methods in terms of signal-to-residual ratio (SRR), ability to reconstruct aliased data parts (+) or not (−), and compression ratio based on the synthetic data example

Method	WSPDL	Interpolation + FK	FK
SRR (dB)	24.43	27.64	16.54
Aliasing	+	+	−
CR	13.13	0.25	1

Note: We remind that compression/decompression is applied to the input data sets of all three methods to allow a fair comparison of the wavefield separation performance.
Abbreviations: CR, compression rate, SRR, signal-to-residual ratio.

Field data

We now consider a more realistic field data set that consists of 3320 shots comprising a full 2D sail-line of marine seismic acquisition. The data were acquired in the 'Nordland ridge' area offshore Norway with a time sampling of 2 ms and a spatial sampling of 12.5 m. The number of time samples (Nt) and the number of channels (Nx) are $Nt = 2050$ and $Nx = 647$, respectively. For every shot, both the total pressure and the total vertical particle velocity were recorded. Both data recordings were pre-processed before applying the dual-sensor wavefield separation process (Figure 10). Later, we will describe in detail the different steps of the pre-processing sequence. The purpose of the test is to assess the effectiveness of WSPDL when applied to field data. This is done by comparing its performance to the optimized industry-standard FK-WS method with interpolation, as this approach has been shown to provide high quality results in the synthetic data example. However, for the field data we do not have access to the true result. Moreover, we do not apply any compression to the input velocity and pressure field data sets prior to using the FK-WS method. This means that the combined effects of compression and wavefield separation are included in the quantitative assessment of the differences between the results of both methods. Furthermore, the shot gathers will be pre-stack migrated to assess the differences between both methods in the image domain.

Figure 10 shows the pre-processing steps applied to the input velocity and pressure data sets. First, a low-cut filter is applied to remove frequencies below 20 Hz for the same reason previously mentioned in the synthetic data example. Then, a *k*-filter or spatial matching filter is applied to match the responses of the geophones and hydrophones groups. This process is required because the distributions of individual sensors in the hydrophones and geophones groups are not identical. We then remove the direct arrival and apply a dip filter to remove steep events with apparent velocities less than 1550 m/s and limit our comparison to the signal cone as in the synthetic data example. Finally, the wavefield separation process is applied, once using the FK-WS method after interpolating the data to 6.25 m spatial sampling as generally done in production, and the second time using the WSPDL method. The two up-going pressure field data sets obtained are then compared. Here, we apply the same parameters described in the methodology section when using the WSPDL method. In fact, we developed our algorithm such that each group of 200 shots are processed independently. Hence, for each group of 200 shots, 2 dictionaries of 6000 atoms will be learned, one for the velocity and one for the pressure data sets, then both will be combined into one dictionary of 12,000 atoms. However, when the total number of shots cannot be divided by 200, the algorithm will automatically find the closest number to 200 that will make this division possible. For example in our case we deal with 3320 shots; therefore, groups of 208 shots will be processed independently, whereas the last set will comprise exactly 200 shots. The choice of 200 shots was found to be a good compromise between the accuracy of reconstruction and the sparsity level. Indeed, applying the same parameters previously used on a 100 shots data set to a 200 shot data set will lead to a higher level of sparsity and thus a higher compression rate (CR) as 12,000 atoms will now be used to reconstruct 200 shot gathers and not only 100. However, a very high number of shot gathers processed at the same time might lead to insufficient number of atoms to reconstruct the data with the high level of reconstruction SRR = 30 dB imposed earlier in the methodology section. Our aim is to develop a method which is transferable to different data sets without changing the DL parameters.

Figure 11 shows one shot gather of the resulting up-going pressure field after applying the wavefield separation process using both methods: the industry-standard FK-WS method (Figure 11a) and the WSPDL method (Figure 11b), respectively. Figure 11c shows the difference between the two shot gathers. The *f*–*k* spectra of Figure 11a,c are displayed in Figure 11d,f, respectively. The small differences between both methods are primarily due to locally strong amplitudes at near offsets. Indeed, this is a very challenging part of the data set for both methods where a rapid change of amplitudes occurs in both space and time. Here,

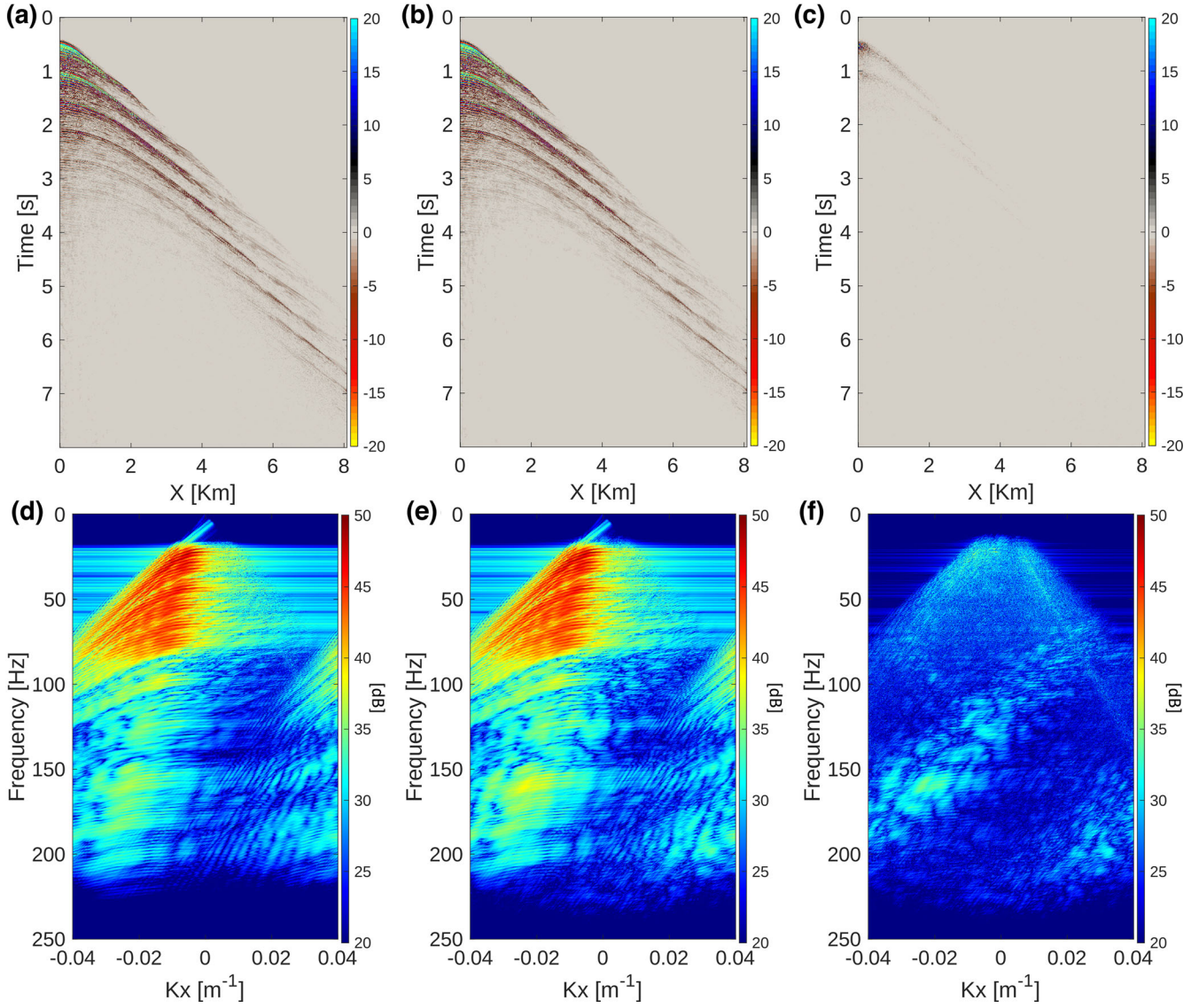

FIGURE 11 Dual-sensor wavefield separation results for one shot gather of the field data example: (a) up-going pressure field after applying the FK-WS method with interpolation; (b) up-going pressure field after applying the WSPDL method; (c) differences between (a) and (b) and (d–f) the *f–k* spectra of (a–c), respectively.

the FK-WS method suffers from the truncation effect at zero offset, whereas the WSPDL method is based on the ray theory which assumes smoothly changing amplitudes along the travel time move outs (Bortfeld, 1989; Ursin, 1982). Moreover, in Figure 11f, we see that such differences are primarily concentrated at high frequencies (around 160 Hz), whereas the differences at the low frequencies (lower than 60 Hz) and low wavenumbers are relatively weak compared to the high energy of the primary even if it is significant relatively to the rest of the energy displayed in the *f–k* spectrum of the differences (Figure 11d–f). From Figure 11d,e, we note that the WSPDL method reconstructs the high frequency events around 160 Hz with more energy than the FK-WS method. Hence, such differences are most likely not due to the compression effects, but higher values of the obliquity correction factor F applied to the velocity data for such dipping events in case of WSPDL. We recall that we spatially interpolate the data with a factor of 2 when the industry-standard FK-WS method is applied, and that we have already demonstrated the ability of the WSPDL method to work beyond aliasing in the synthetic data example. Therefore, it is reasonable to observe higher amplitudes on highly dipping events when using the WSPDL at frequencies higher than 120 Hz which is the frequency limit beyond which spatial aliasing will adversely affect the output of the FK-WS method for spatial sampling of 6.25 m. However, the FK-WS method does not show specific problems above 120 Hz on the synthetic data example, because the synthetic data are sparse in frequency, which makes it easy for frequency-based interpolation methods such as the industry-standard interpolation algorithm (Bekara & Robin, 2015) used here, to reconstruct clean aliased events (Figure 9d) compared to

FIGURE 12 Main steps of the seismic processing workflow to generate a 2D line image after applying the dual-sensor wavefield separation process using the FK and the WSPDL methods.

the real data case. Note that the WSPDL method has reconstructed the up-going pressure field in the compressed domain with CR = 15.81.

After comparing the results of both wavefield separation methods on seismic shot gathers, we will now pre-stack migrate the resulting up-going pressure field data set of 3320 shots comprising a full sail-line of marine seismic acquisition. We applied a simple seismic processing workflow with main steps as summarized in Figure 12. This workflow does not replicate a full commercial processing but is sufficient to assess the impact of both wavefield separation methods on a 2D seismic image. Figure 13 shows the part of the 2D seismic image, where most dipping events occur. Here, the focus is on the shallow part as the deep part is heavily contaminated by noise. Visually, it is hard to discern any differences between Figure 13a,b, representing the 2D seismic images after applying the industry-standard FK and the WSPDL methods, respectively. The differences between both reconstructions are displayed in Figure 13c and are mainly located close to the water bottom reflector, where amplitudes are very strong. Hence, we can conclude that the two methods provide very similar results. However, in case of WSPDL, a compression factor of almost 16 is obtained and aliased energy is handled without the need for explicit interpolation. In order to confirm the visual results, we perform an amplitude spectrum analysis of the 2D seismic images, and their differences. Figure 14 shows the cumulative amplitude spectra of the images in Figure 13a–c, denoted in blue, green and red, respectively. This figure shows that the spectrum of the differences between both images, which is 10 dB to 30 dB below signal (in red), is relatively flat and has a dissimilar shape to that of the 2D seismic sections. The spectra of the images formed from the two up-going pressure fields (in green and blue) are very similar especially below 125 Hz. This analysis confirms that the FK and WSPDL methods provide very similar results. Moreover, we do not observe any significant loss of energy due to the compression effect when we use the WSPDL method knowing that the resulting image was reconstructed from data almost 16 times smaller in size than the original data used for the FK-WS method. That is because we have targeted a high level of reconstruction of SRR = 30 dB for the input velocity and pressure data sets. We further observe that the WSPDL method recovers events between 120 and 200 Hz with higher energy than the FK-WS method. However, this does not mean that one method is more correct than the other. Indeed, the water bottom differences are located at these frequencies which is in accordance with the analysis of the shot gathers results shown in Figure 11.

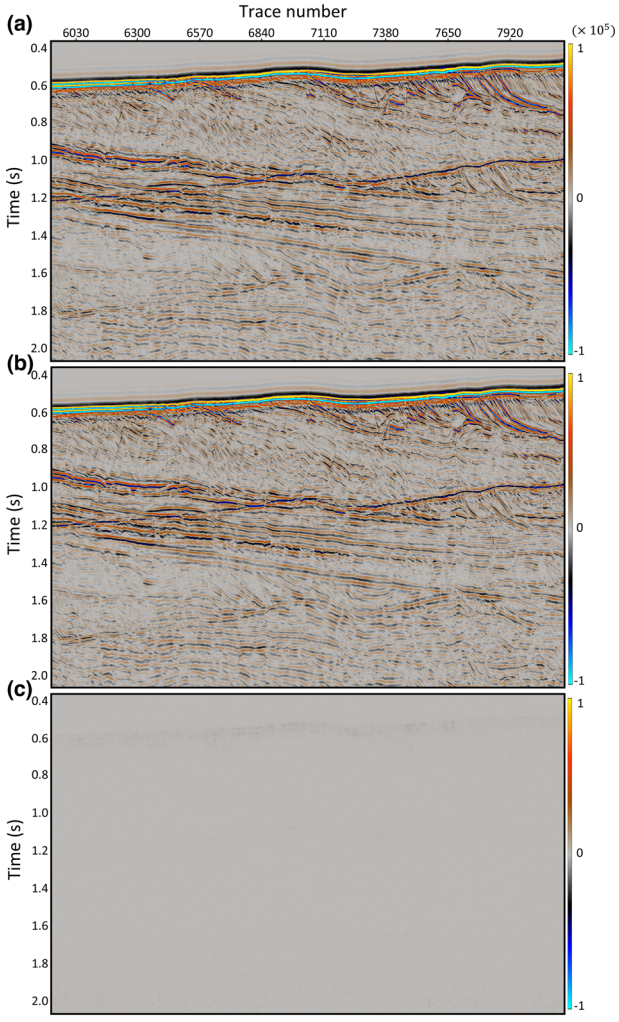

FIGURE 13 (a) 2D seismic image after applying the FK-WS method; (b) 2D seismic image after applying the WSPDL method and (c) difference between (a) and (b).

DISCUSSION

Compression sensitivity analysis

Although we have already discussed our choices for the different dictionary learning (DL) parameters, we should still discuss the compression impact on the wavefield separation

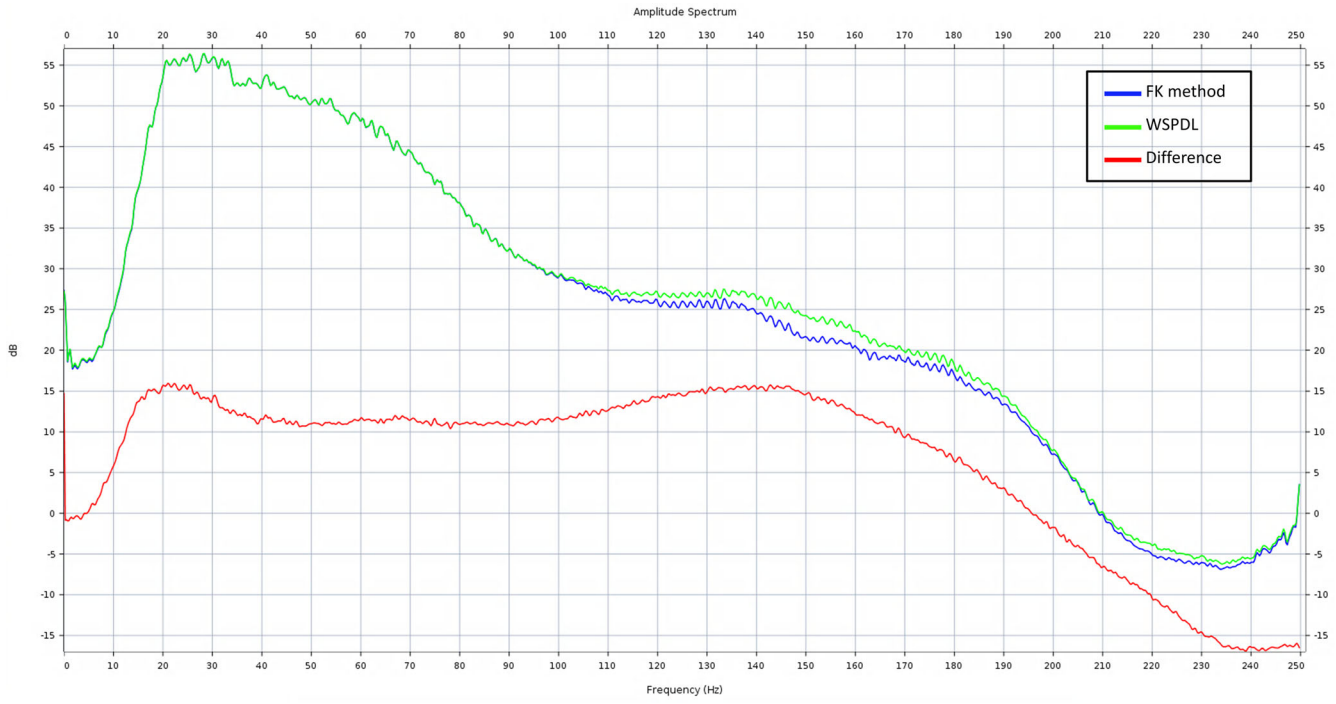

FIGURE 14 Amplitude spectrum analysis of the 2D seismic images represented in Figure 13.

results when using the PDL method for wavefield separation (WSPDL) method. Indeed, a high input signal-to-residual ratio (SRR) level of 30 dB has been imposed as a target to allow an accurate reconstruction of the input velocity and pressure data sets after compression. Such an input SRR level (that we will denote SRR_{in}) has allowed a good reconstruction of the up-going pressure field for both the synthetic data example where the SRR after wavefield separation (that we will denote SRR_{WS}) reached a high level of 24.43 dB with compression rate (CR) = 13.13 and the field data example where a 2D line section was reconstructed with insignificant energy loss while reaching a CR of almost 16. Acknowledging that as we increase the SRR_{in}, we decrease the CR, we introduce a simple experiment in which the WSPDL method is tested with different values of SRR_{in}. We then observe how the CR level impacts the wavefield separation results via the SRR_{WS} employing the 100 shot synthetic data set. The graphs displayed in Figure 15 show that the WSPDL method is stable as the CR level is decreasing when the SRR_{in} is increasing as expected. We also observe from the same figure that the SRR_{WS} level becomes relatively stable (between 24 and 25 dB) when SRR_{in} is higher than 30 dB, which shows the reconstruction limits of our method. Indeed, targeting a higher level of SRR_{in} = 36.5 dB in our method could lead to a slightly better SRR_{WS} level (Figure 15), but the CR level would drop from 13.13 to 6.35. In Figure 16 we display four different up-going pressure field reconstructions given different SRR_{in} levels. Here, we observe good performance in terms

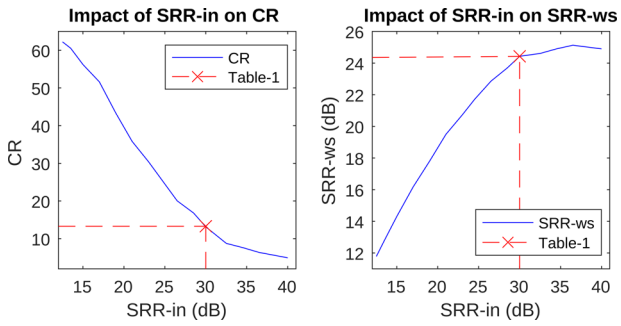

FIGURE 15 Impact of different input signal-to-residual levels (SRR_{in}) on the compression ratio (CR) and the wavefield separation reconstruction levels (SRR_{WS}). The red cross on the graphs corresponds to the case reported in Table 1, where SRR_{in} = 30 dB.

of up-going pressure field reconstructions when CR = 6.35 and CR = 13.13 (Figure 16a,b,e,f) and one can hardly see any differences between those reconstructions. When CR is around 26 (Figure 16c,g), few low amplitude events are not well reconstructed, and residuals are uniformly distributed along the different events below the first reflection. However, even when reaching such a high CR level the visual quality of the reconstruction is still acceptable. In Figure 16d,h, a high level of CR = 43.36 is reached, and many events are not well reconstructed. These displays validate our choice of targeting a high SRR_{in} level of 30 dB, as it allows to reach both high SRR_{WS} and CR levels.

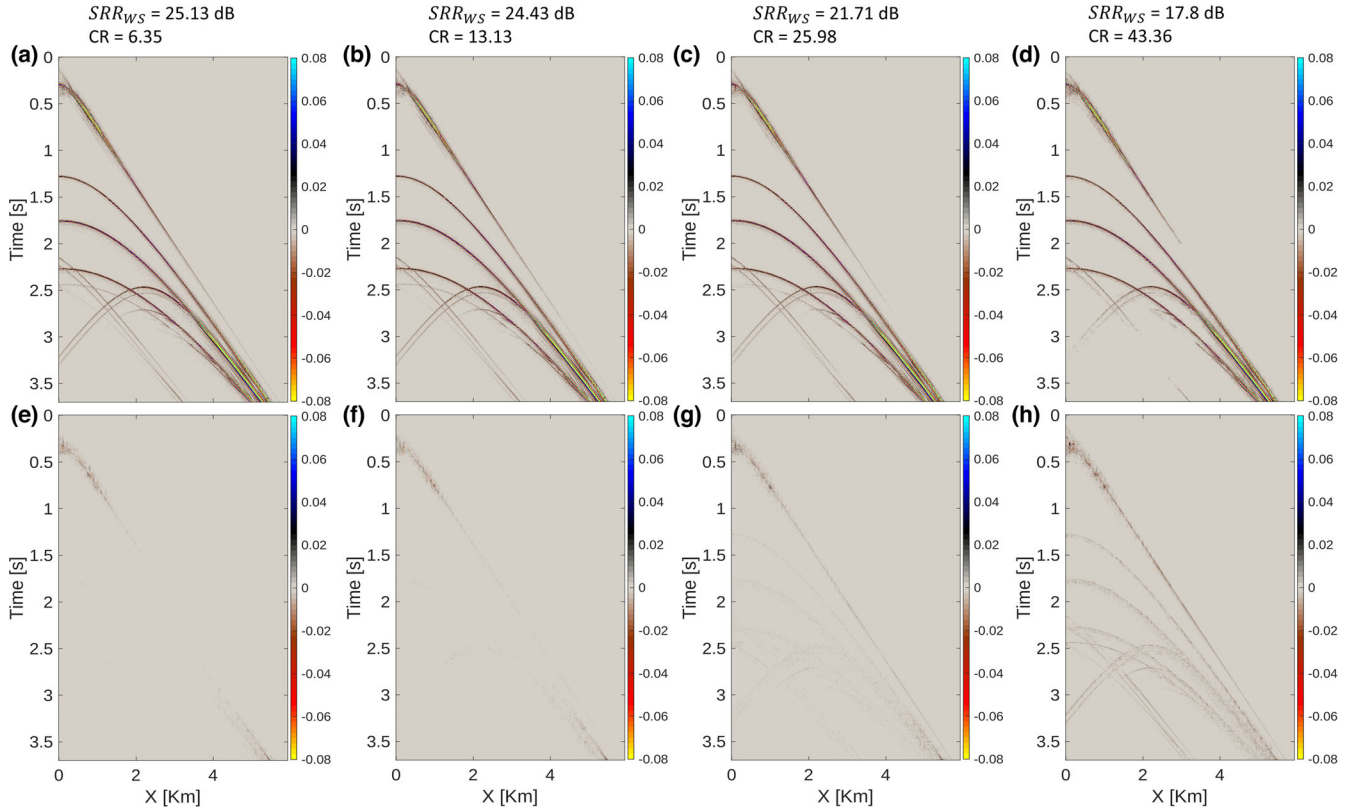

FIGURE 16 Compression effect on the 2D synthetic up-going pressure field reconstruction results: (a–d) up-going pressure field reconstructions after applying the WSPDL method with different compression levels and (e–h) the residuals corresponding to (a–d) after subtracting the reference up-going pressure field.

Future work

In the synthetic data and field data subsections, we have applied the same pre-processing steps (Figures 6 and 10) to input data sets before applying the WSPDL or the optimized industry-standard FK-WS method. Indeed, this paper focuses only on enabling the dual-sensor wavefield separation process in the DL compressed domain. Thus, it is important to apply the same pre-processing steps before applying both methods to allow a fair comparison in terms of wavefield separation performance. However, some of these steps are more suited for FK-WS methods as they already include a transformation in the FK domain. Hence, future work can focus on enabling pre-processing or other processing steps in the DL compressed domain to avoid going back and forth between different processing domains. For example the dip filter does not necessarily need to be applied before the WSPDL method here. This was applied in synthetic and field data examples to fairly compare the impact of both methods on the signal cone as tapering is generally applied in conventional FK-WS methods as mentioned in the data application section. In practice, a taper can also be applied after reconstructing the data with the WSPDL method or a dip filter can directly be applied at the atoms level using the kinematic parameters. Another example is the *k*-filter that was not needed to compare both methods but was applied in the real data example to simulate a real data case where the responses of the geophones and hydrophones groups need to be matched. Consequently, enabling such a filter in the DL compressed domain can be very convenient. Moreover, the low-cut filter was applied because low frequency velocity data are generally reconstructed from the relatively clean pressure data (Day et al., 2013), as mentioned earlier in the data application section. Therefore, one can also investigate the possibility of using the low frequencies pressure data to reconstruct the low frequencies up-going pressure field directly in the DL compressed domain.

In this paper, we have learned the dictionaries $\mathbf{D}_{V_z}$ and $\mathbf{D}_P$ independently. This was enough to achieve good reconstruction of the up-going pressure field while reaching a high level of compression. Instead, one could use a joint DL method (Wang et al., 2022) where the same kinematic parameters are imposed to the atoms of both dictionaries. Such an approach may lead to better approximations of the obliquity correction factors, as the total pressure data will also be used in the estimation of the kinematic parameters. Furthermore, the scope of this paper focuses on applying WSPDL in the 2D case. The extension to the 3D data case can be done by

learning 3D atoms which would enable the reconstruction of the 3D wavefield. 3D atoms would be characterized by five parameters instead of only two in the 2D case, namely two slopes, two curvatures and a mixed travel time parameter correlating both directions (in-line and crossline). In the common reflection surface approach, Hoecht et al. (2009) estimated these parameters to achieve 3D interpolation. Similarly, such parameters can be learned using WSPDL. Alternatively, one can adopt a pseudo-3D approach where 2D dictionaries will be learned independently in the in-line and crossline directions. Such an approach would be computationally cheaper than learning 3D atoms but might lead to less accurate results as correlations between the in-line and crossline directions will not be used in the estimation of the kinematic parameters. In this case, one should still think about methods to correlate the in-line and crossline dictionaries such as joint DL methods (Wang et al., 2022). Moreover, the paper focuses also on the dual-sensor towed streamer acquisition (or two-components streamer data acquisition), where: One hydrophone and one geophone are used to record the total pressure and the total particle velocity motion, respectively. Hence, other acquisition settings do exist such as the 3C streamer data acquisition, where two geophones or accelerometers record the particle velocity field in the vertical and the cross-cable directions. However, the horizontal component is heavily contaminated by noise in the in-line direction. The 3C acquisition setting would theoretically provide the crossline component of the particle velocity measurement without the need to learn kinematic parameters in the crossline direction (Vassallo et al., 2014). Such information can be combined with the kinematic parameters estimated in the in-line direction via WSPDL to allow a better approximation of the reconstructed wavefield.

CONCLUSION

In this work, we have successfully demonstrated a novel method that enables dual-sensor wavefield separation in a compressed domain using a parabolic dictionary learning (PDL) algorithm. The method (WSPDL) uses PDL to transform input data sets, namely the total pressure and the total vertical particle velocity measurements, to a compressed domain composed of two main parts: a dictionary of parabolic atoms and a set of sparse coefficients. The atoms of the dictionary are characterized by kinematic parameters such as the slope and the curvature, which are used to correct for velocity obliquity scaling directly in the compressed domain, and thus allow to successfully separate the up- and down-going parts of the pressure wavefield. We have tested and validated the performance of our method by quantifying the accuracy of the up-going pressure field reconstruction using the signal-to-residual ratio metric based on a 100 shot synthetic data example, where we have access to the true result. Moreover, we observe similar wavefield separation performance when compared to an optimized industry-standard FK-WS method based on both synthetic and field data examples. Hence, we have observed only small differences between the two methods after 2D line pre-stack migration. Moreover, the WSPDL method is robust with respect to spatial aliasing without the need for data preconditioning such as interpolation and comes with the advantage of a data compression rate higher than 15. Such a method could allow full bandwidth data transfer from vessels to onshore processing centres as it reaches significant compression levels. The transmitted compressed data can be used to reconstruct not only the input data sets recorded by the hydrophones and the geophones, but also the up- and down-going parts of the wavefield without the need to run the conventional dual-sensor wavefield separation process as this reconstruction requires only a simple matrix multiplication.

In this paper, we have succeeded in applying one seismic processing step, namely the dual-senor wavefield separation, in the dictionary learning compressed domain using kinematic parameters. Enabling other seismic processing steps in the compressed domain using these kinematic parameters would provide a novel and efficient workflow to simultaneously compress and partly or fully process seismic data sets.

ACKNOWLEDGEMENTS

We would like to express our gratitude to the R&D team of PGS for the fruitful discussions on the subject and thank PGS for providing the data and giving the permission to publish this work.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are confidential, not publicly available and cannot be released due to privacy or ethical restrictions.

REFERENCES

Aharon, M., Elad, M. & Bruckstein, A. (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54, 4311–4322.

Amundsen, L. (1993) Wavenumber-based filtering of marine point-source data. *Geophysics*, 58, 1335–1348.

Averbuch, A., Coifman, R., Donoho, D., Israeli, M. & Walden, J. (2001) Fast slant stack: a notion of radon transform for data in a cartesian grid which is rapidly computible, algebraically exact, geometrically faithful and invertible. *SIAM Scientific Computing*, 37, 192–206.

Beckouche, S. & Ma, J. (2014) Simultaneous dictionary learning and denoising for seismic data. *Geophysics*, 79, A27–A31.

Bekara, M. & Robin, F. (2015) Seismic data up-sampling beyond aliasing using polynomial phase signals. In: *77th EAGE Conference and Exhibition*. pp. 1–5.

Bortfeld, R. (1989) Geometrical ray theory: rays and traveltimes in seismic systems (second-order approximations of the traveltimes). *Geophysics*, 54, 342–349.

Bruckstein, A.M., Donoho, D.L. & Elad, M. (2009) From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Review*, 51, 34–81.

Day, A., Klüver, T., Söllner, W., Tabti, H. & Carlson, D. (2013) Wavefield-separation methods for dual-sensor towed-streamer data. *Geophysics*, 78, WA55–WA70.

Donoho, D.L. & Elad, M. (2003) Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization. *Proceedings of the National Academy of Sciences of the United States of America*, 100, 2197–2203.

Donoho, D.L. (2006) Compressed sensing. *IEEE Transactions on Information Theory*, 52, 1289–1306.

Duval, L.C. & Rosten, T. (2000) Filter bank decomposition of seismic data with application to compression and denoising. In: 70th SEG Annual International Meeting, Expanded Abstracts. pp. 2055–2058.

Elad, M. (2010) *Sparse and redundant representations: from theory to applications in signal and image processing*, 1st edition, New York: Springer Publishing Company.

Fajardo, C., Reyes, O.M. & Ramirez, A. (2015) Seismic data compression using 2D lifting-wavelet algorithms. *Ingenieria y Ciencia*, 11, 221–238.

Faouzi Zizi, M.O. & Turquais, P. (2022) A dictionary learning method for seismic data compression. *Geophysics*, 87, V101–V116.

Helal, E.B., Saad, O.M., Hafez, A.G., Chen, Y. & Dousoky, G.M. (2021) Seismic data compression using deep learning. *IEEE Access: Practical Innovations, Open Solutions*, 9, 58161–58169.

Herrmann, F.J., Calvert, A.J., Hanlon, I., Javanmehri, M., Kumar, R., van Leeuwen, T. et al. (2013) Frugal full-waveform inversion: from theory to a practical algorithm. *The Leading Edge*, 32, 1082–1092.

Hoecht, G., Ricarte, P., Bergler, S. & Landa, E. (2009) Operator-oriented CRS interpolation. *Geophysical Prospecting*, 57, 957–979.

Hubral, P., Schleicher, J. & Tygel, M. (1992) Three-dimensional paraxial ray properties: Part I. Basic relations. *Journal of Seismic Exploration*, 1, 265–279.

Huffman, D.A. (1952) A method for construction of minimum redundancy codes. *Proceedings of the IRE*, 40, 1098–1101.

Lin, T.T. & Herrmann, F.J. (2013) Robust estimation of primaries by sparse inversion via one-norm minimization. *Geophysics*, 78, R133–R150.

Mairal, J., Bach, F., Ponce, J. & Sapiro, G. (2009) Online dictionary learning for sparse coding. In: *26th Annual International Conference on Machine learning*. pp. 689–696.

Mallat, S. (2008) *A wavelet tour of signal processing: the sparse way*, 3rd edition, Cambridge: Academic Press.

Mansour, H., Wason, H., Lin, T.T., T, T. & Herrmann, F.J. (2012) Randomized marine acquisition with compressive sampling matrices. *Geophysical Prospecting*, 60, 648–662.

Pati, Y.C., Rezaiifar, R. & Krishnaprasad, P.S. (1993) Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers*. pp. 40–44.

Rubinstein, R., Zibulevsky, M. & Elad, M. (2008) *Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. Technion Report, CS-2008-08*.

Schiavon, A.P., Ribeiro, K., Navarro, J.P., Vieira, M.B. & Silva, P.M.C. (2020) 3D poststack seismic data compression with a deep autoencoder. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.

Söllner, W., Day, A. & Tabti, H. (2008) *Space Frequency Domain Processing of Irregular Dual-Sensor Towed Streamer Data. SEG expanded abstracts*, vol. 27. pp. 1078–1083.

Turquais, P., Asgedom, E.G., Söllner, W. & Gelius, L.-J. (2018) Parabolic dictionary learning for seismic wavefield reconstruction across the streamers. *Geophysics*, 83, V263–V282.

Turquais, P., Pedersen, M. & Söllner, W. (2019) Parabolic dictionary learning: A method for seismic data reconstruction beyond the linearity assumption. In: *81st EAGE Conference and Exhibition*. pp. 1–5.

Ursin, B. (1982) Quadratic wavefront and traveltime approximations in inhomogeneous layered media with curved interfaces. *Geophysics*, 47, 1012–1021.

Vassallo, M., Özbek, A., Kamil, Y., van Manen, D., Eggenberger, K. (2014) Reconstruction of signals from highly aliased multichannel samples by generalized matching pursuit. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing*. pp. 2382–2385.

Wang, X., Teng, Y., Gao, M. & Jiang, H. (2004) Seismic data compression based on integer wavelet transform. *Acta Seismologica Sinica*, 17, 123–128.

Wang, Y., Zhang, G., Li, H., Yang, W. & Wang, W. (2022) The high-resolution seismic deconvolution method based on joint sparse representation using logging–seismic data. *Geophysical Prospecting*, 70, 1313–1326.

Zhang, Y., Bergler, S. & Hubral, P. (2001) Common-reflection-surface (CRS) stack for common offset. *Geophysical Prospecting*, 49, 709–718.

Zheng, F. & Liu, S. (2012) *A fast compression algorithm for seismic data from non-cable seismographs*. World Congress on Information and Communication Technologies, Trivandrum, India, 2012, pp. 1215–1219. doi: 10.1109/WICT.2012.6409260.

How to cite this article: Faouzi Zizi, M.O., Turquais, P., Day, A., Pedersen, M.W. & Gelius, L.J. (2023) Dual-sensor wavefield separation in a compressed domain using parabolic dictionary learning. *Geophysical Prospecting*, 71, 792–810. https://doi.org/10.1111/1365-2478.13348