Reconstruction of Signals from Highly Aliased Multichannel Samples by Generalized Matching Pursuit

Ali Özbek[1], *Massimiliano Vassallo*[2], *Kemal Özdemir*[3], *Dirk-Jan van Manen*[1], *Kurt Eggenberger*[4]

[1]Schlumberger Gould Research, Cambridge, UK
[2]WesternGeco London Technology Centre, Gatwick, UK
[3]WesternGeco Oslo Technology Center, Asker, Norway
[4]WesternGeco Houston Technology Center, Houston, TX, USA
Emails: {ozbek, mvassallo, aozdemir, dmanen, keggenberger}@slb.com

***Abstract*—This paper considers the problem of reconstructing a bandlimited signal from severely aliased multichannel samples. Multichannel sampling in this context means that the samples are available after the signal has been filtered by various linear operators. We propose the method of Generalized Matching Pursuit to solve the reconstruction problem. We illustrate the potential of the method using synthetic data that could be acquired using multimeasurement towed-streamer seismic data acquisition technology. A remarkable observation is that high-fidelity reconstruction is possible even when the data are uniformly and coarsely sampled, with the order of aliasing significantly exceeding the number of channels.**

I. Introduction

In multichannel sampling, samples of a signal that was filtered by various linear operators are available. Suppose $\boldsymbol{m}(y)=\boldsymbol{h}(y)*s(y)$, where $\boldsymbol{m}(y)=[m_1,\cdots,m_J]$ are the measurements, and $\boldsymbol{h}(y)=[h_1,\cdots,h_J]$ are the operators. The samples are available at points $y_1,\cdots,y_L$, which may be regularly or irregularly spaced. The objective is to reconstruct bandlimited signal $s(y)$ at arbitrary points y. In Figure 1, we show a slight generalization, where, for each channel j, the measurements are undersampled by a factor of R_j with respect to the bandwidth of s. In the spectral domain, we have $\boldsymbol{m}(k_y)=\boldsymbol{H}(k_y)s(k_y)$, where k_y is the wavenumber (spatial frequency).

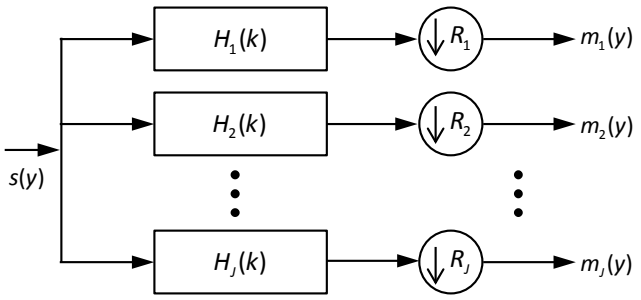

Fig. 1. Multichannel sampling.

The generalized sampling expansion proposed by Papoulis [1] implies that such a linear system, under certain conditions, allows reconstruction of the desired signal when $R_j=J$, $j=1,2,\cdots,J$. However, Papoulis [1] does not provide a readily realizable solution for the inversion of the system. Later, several articles were proposed to study the properties of the generalized sampling expansion, the well-posedness of the system, and a closed-form solution of the inverse problem [2, 3].

In some applications, such as marine seismic data acquisition, the decimation rate R_j can be significantly larger than the number of channels, J. In this case the order of aliasing significantly exceeds the number of channels. In the next section, we discuss a method that has shown promising performance in this setting.

II. Generalized Matching Pursuit

In this section, we describe a parametric matching pursuit method to solve the reconstruction problem that arises in multichannel sampling; we call it *Generalized Matching Pursuit (GMP)*, as its aim is to reconstruct a signal of which no direct samples may be available. Suppose that the unknown signal $s(y)$ is modeled as a sum of parametric basis functions $\beta(y;\boldsymbol{\theta}_n)$ with parameter set $\boldsymbol{\theta}_n$:

$$s(y)=\sum_n \beta(y;\boldsymbol{\theta}_n)\,. \tag{1}$$

There are various basis functions that can be considered; one possibility that is especially convenient for seismic applications is

$$\beta(y;\boldsymbol{\theta}_n)=A_n\exp\left[j\left(k_{y,n}y+\phi_n\right)\right], \tag{2}$$

where the parameter set $\boldsymbol{\theta}_n$ consists of amplitude A_n, phase ϕ_n, and wavenumber $k_{y,n}$. The corresponding measurements would then be

$$\boldsymbol{m}(y)=\sum_n \boldsymbol{H}(k_{y,n})\beta(y;\boldsymbol{\theta}_n)\,. \tag{3}$$

In GMP, the forward linear filters $H_j(k_y)$ are applied to each basis function; the filtered basis functions are then iteratively matched to the multichannel measurements. Iteratively, the basis function that, once forward filtered, jointly best matches all the input signals is used to reconstruct the desired output, with or without the forward filter applied. At the N-th iteration, i.e., after N-1 basis functions have been determined previously, the residual in the measurements is given by

$$\boldsymbol{r}^{N-1}(y)=\boldsymbol{m}(y)-\sum_{n=1}^{N-1}\boldsymbol{H}\left(k_{y,n}\right)\beta(y;\boldsymbol{\theta}_n)\,. \tag{4}$$

If a new term $\beta(y;\boldsymbol{\theta}_N)$ is added to the existing representation of the signal, the residual becomes $\boldsymbol{r}^N(y;\boldsymbol{\theta}_N)=\boldsymbol{r}^{N-1}(y)-\boldsymbol{H}(k_{y,N})\,\beta(y;\boldsymbol{\theta}_N)$, where the parameters of the new term, i.e., $\boldsymbol{\theta}_N$, are to be

determined by minimizing a metric of the residual calculated over measurement locations. One such metric is

$$\mu(\boldsymbol{\theta}_N) = \sum_l \left[\boldsymbol{r}^N(y_l;\boldsymbol{\theta}_N)\right]^H \boldsymbol{C}^{-1}\boldsymbol{r}^N(y_l;\boldsymbol{\theta}_N), \quad (5)$$

where the superscript H represents the Hermitian operator, $\boldsymbol{C}$ is a positive definite matrix, and y_l represents the sensor locations in the y direction. These locations can, in general, be irregularly spaced. The role of matrix $\boldsymbol{C}$ is to weight the contributions of different measurements to the cost function to be minimized. This can take into account the difference of energy content due to the different physics of the input measurements, as well as the signal-to-noise ratio that can vary in time, space, and frequency [10].

For basis functions chosen as in (2), it can be shown that the optimal A_N and ϕ_N can be analytically related to the residuals $\boldsymbol{r}^{N-1}$, the input sample positions y_l, and the optimal wavenumber $k_{y,N}$. Hence, the only remaining parameter to select is

$$k_{y,N} = \arg\max_k \mathcal{L}\left\{A_N\left(\boldsymbol{r}^{N-1}(y_l),k\right), \phi_N\left(\boldsymbol{r}^{N-1}(y_l),k\right)\right\}. \quad (6)$$

We call the objective function $\mathcal{L}$ the generalized Lomb spectrum, in analogy with the single-channel interpolation problem. There, in the case of sinusoidal basis functions, the objective function generated by *Interpolation by Matching Pursuit (IMAP)* with optimal amplitudes in the least-squares sense corresponds to the Lomb spectrum [4, 5, 6].

The GMP iterations can be terminated once the residual energy falls below a predetermined fraction of the input energy.

Next, we illustrate the antialiasing power of GMP for uniformly sampled multichannel data with a very simple multichannel sampling example. In this example, a single sinusoid signal with wavenumber 30 Km^{-1} is uniformly sampled at 25 Km^{-1}. In addition to the signal samples, the spatial gradient samples are available at the same locations. Due to uniform sampling, there is hard-aliasing, i.e., exact periodic replicas in the spectra of each channel. This is a reconstruction problem that cannot be solved by multichannel sinc interpolation [7], since the order of aliasing is greater than two. Figure 2 shows the cost function to select the optimum wavenumber (negative of the generalized Lomb spectrum) at the first iteration. The aliases of the correct wavenumber can be clearly seen. However, simultaneous use of the multichannel measurements in the optimization process results in the correct wavenumber being selected.

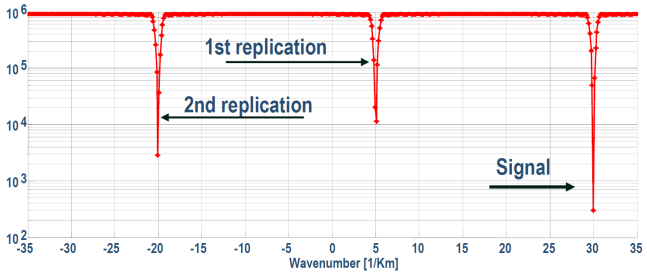

Fig. 2. Cost function for the optimum wavenumber in a hard-aliasing problem resulting from insufficient uniform sampling.

III. Application To Multichannel Sampling

Due to logistical and cost constraints, marine seismic acquisition systems can be deployed to acquire data only along a limited number of parallel lines (i.e., towed streamers) that are coarsely spaced in the crossline direction. Streamers are towed typically with crossline spacing of 75-100 m, resulting in coarse wavefield sampling that contrasts with adequate (non-aliasing) wavefield sampling of 6.25 m along the streamers (inline). Consequently, they do not adequately capture the full spatial bandwidth of the subsurface-scattered wavefield, leading to limitations in accurate subsurface imaging. Furthermore, conventional (pressure-only data) acquisition systems suffer from the ghost effect. The ghost is the reflection from the sea surface that interferes constructively or destructively with the upgoing wavefield (the signal of interest for imaging), reducing the seismic bandwidth at the low and high ends of the spectrum.

To address these critical limitations, a multimeasurement marine seismic acquisition platform was recently introduced. It is equipped with hydrophones to measure the pressure wavefield (P) and accelerometers to measure the particle acceleration vector ($\boldsymbol{A}$). The latter represents the vector spatial gradient of pressure as derived through the particle equation of motion, $\nabla P = -\rho \boldsymbol{A}$, where ρ is the fluid density [8].

A. Example: Reconstructing P from Aliased (P, P_y) Data

An important problem is to reconstruct (interpolate) the total pressure wavefield P at any desired position in the crossline direction from sparse samples of itself and its crossline gradient. P is the sum of the upgoing and downgoing (ghost) wavefields. For this problem, the unknown signal is $s(f,k_x,k_y)=P(f,k_x,k_y)$; the measurement vector is

$$\boldsymbol{m}(t,x,y_l) = \left[P(t,x,y_l) \quad P_y(t,x,y_l)\right]^T, \quad l = 1,2,\cdots,L, \quad (7)$$

where P_y is the crossline gradient of the pressure wavefield; the number of streamers (L) is typically 8-12. The forward linear operator is

$$\boldsymbol{H}(f,k_x,k_y) = \left[1 \quad jk_y\right]^T. \quad (8)$$

Here, f is the temporal frequency; k_x and k_y are the inline and crossline wavenumbers, respectively. As the data are well sampled in the temporal (t) and inline (x) coordinates, we can operate the GMP algorithm outlined in Section 2 for fixed values of f and k_x. The particular form that GMP takes for this reconstruction problem is referred to as *MIMAP (Multichannel Interpolation by Matching Pursuit)* [9].

Figure 3 shows a simple example reproducing linear events with energy up to 65 Hz and various incidence angles first decimated at 75 m and then reconstructed using different techniques. At every receiver position we modeled both the synthetic signal and its horizontal gradient. For the selected geometry, an event propagating horizontally generates first order alias at 10 Hz, and second order alias at 20 Hz, as shown in 2(b). Since MIMAP does not assume that the data comprise linear events in the implementation used for this example, the presence of high orders of aliasing presents a significant challenge for reconstruction.

To show the impact of the antialiasing capabilities of MIMAP, we interpolated the data with two standard techniques in addition to MIMAP: the sinc interpolation, and the multichannel sinc interpolation [7]. Results are shown in Figure 3. In Figure 3(a) we can see a region of the input time-space gather describing the pressure synthetics, sampled at 75 m, and the frequency-wavenumber transform of the overall gather. The high order of aliasing is clearly visible in the *f-k* domain. Figures 3(c) and 3(d) show the results of the single-component conventional sinc

interpolator, bandlimited in the spatial sampling bandwidth. As expected, only frequencies up to 10Hz are not subject to aliasing, and only the events with an incident angle close to zero can be properly interpolated (e.g., the event at 2.6 s). All the rest of the reconstructed information, in fact, corresponds to aliased replicas remapped to incorrect wavenumber positions.

Figures 3(e) and 3(f) show the result of the multichannel sinc interpolation, bandlimited to twice the spatial Nyquist. In this case, we can see that more events are reconstructed correctly in the t-y plot (e.g., events at around 2.4 s, 2.5 s and 2.6 s), and that all the events are reconstructed correctly up to 20 Hz. What is also interesting is that the multicomponent sinc seems to amplify the aliased events that cannot be reconstructed, as visible in the *f-k* gather above 20 Hz. Moreover, the shape of the region not affected by the alias, or affected by a first-order alias only, is clearly recognizable as the properly reconstructed area. Finally, in Figures 3(g) and 3(h), we can observe the results produced by MIMAP, and the removal of aliasing up to very high frequencies can be appreciated. All the events are well reconstructed.

B. Example: Reconstructing P^{up} from Aliased (P, P_y, P_z) Data

Using P, P_y, and P_z data that can be recorded by a multimeasurement streamer, another and more challenging problem would be to reconstruct P^{up} at any desired position without having access to any direct samples of it. This is called the joint interpolation and deghosting problem [10], where the task of separating the wavefield into its down- and upgoing components is performed simultaneously with the task of reconstructing it at any desired position. For this problem, the unknown signal is $s(f,k_x,k_y)=P^{up}(f,k_x,k_y)$; the measurement vector is

$$\boldsymbol{m}(t,x,y)=\left[P(t,x,y)\;\; P_y(t,x,y)\;\; P_z(t,x,y)\right]^T,\;\; l=1,2,\cdots,L, \quad (9)$$

and the forward linear operator that links the measurements to the unknown signal is the ghosting operator defined by

$$\boldsymbol{H}(f,k_x,k_y)=\left[\left(1+\xi e^{j2k_zZ}\right)\;\; k_y\left(1+\xi e^{j2k_zZ}\right)\;\; jk_z\left(1-\xi e^{j2k_zZ}\right)\right]^T. \quad (10)$$

Here, k_z is the vertical wavenumber, Z is the depth of the streamer, and ξ is the reflection coefficient of the sea surface. Through the ghost model, the P_z component brings independent new information on the unknown upgoing wavefield in the crossline direction, which is crucial for this application [10].

Figure 4 shows the application of the GMP technique to solve the joint interpolation and deghosting problem in the crossline direction using synthetic data. The data set was created by finite-difference modeling and simulates a 3D multimeasurement survey over a complex geological structure. The source signature spectrum is flat up to 30 Hz. The streamer depth is 50 m; the unusual depth was chosen to place the pressure ghost notch within the 30-Hz bandwidth. Given total P, P_y and P_z data sampled at 150 m where the data are severely aliased, the reconstructed upgoing pressure wavefield sampled at the desired 25-m interval show both the dealiasing and the deghosting capabilities of this approach.

Figure 4(a) shows the f-k_x-k_y transform of the total pressure wavefield before decimation, with pressure sampled over a 25-m x 25-m spatial grid. We can recognize the lack of energy in the low wavenumbers in the 15-Hz slice, and a circularly shaped notch in the 20-Hz and 25-Hz slices. The events that are not

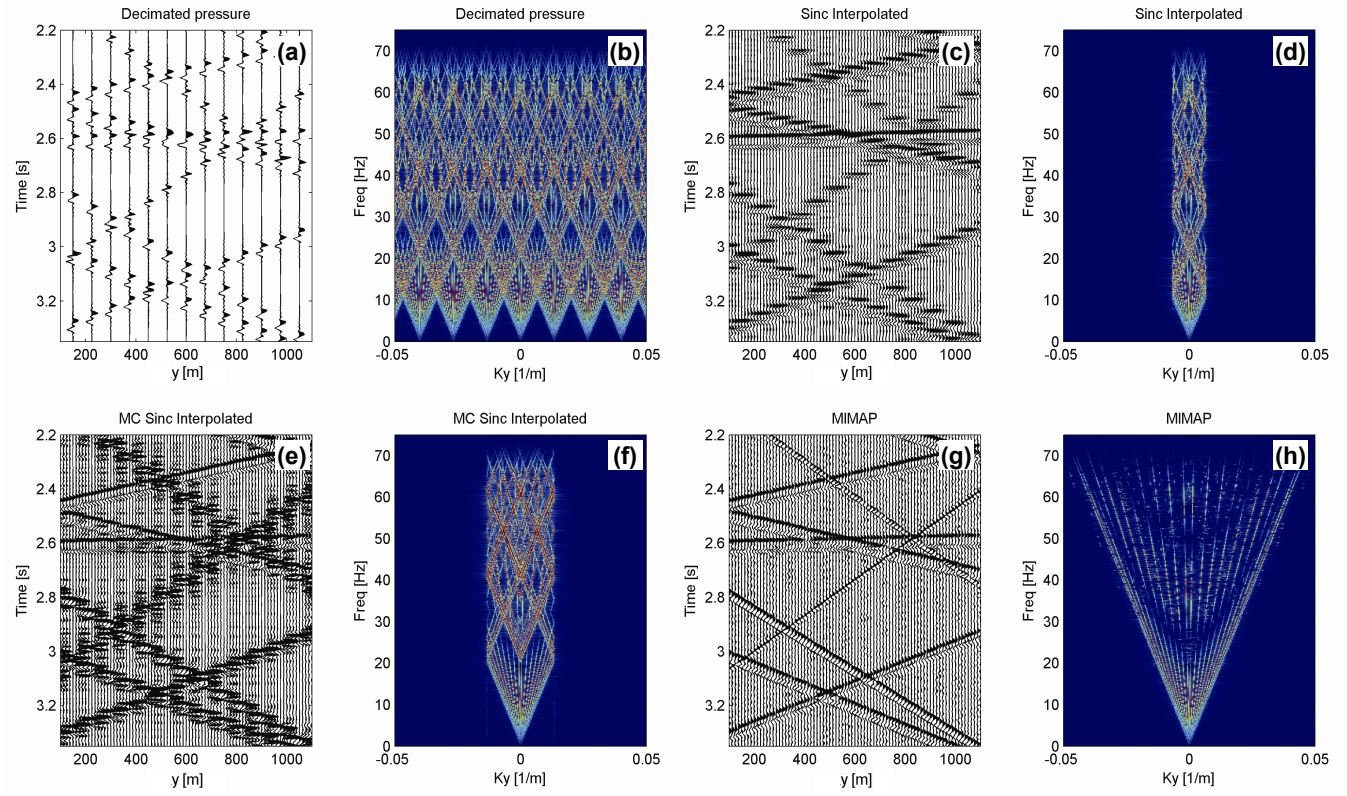

Fig. 3. Example with simple synthetics: close-up of a region of the *t-y* domain and *f-k* transforms of the whole dataset. (a, b) Input pressure, sampled at 75 m; (c, d) pressure reconstructed by using a sinc interpolator; (e, f) pressure reconstructed by using a multichannel sinc interpolator, also having as input the crossline gradients at the samples positions; (g, h) pressure reconstructed with MIMAP, also having as input the crossline gradients at the samples positions.

affected by the notch are still affected by the constructive interference of the ghost. Figure 4(b) shows the f-k_x-k_y transform of the reference upgoing pressure wavefield, sampled over a 25-m x 25-m spatial grid. The f-k_x-k_y transform of the total pressure wavefield after decimation of the data to 150 m in the crossline direction is shown in Figure 4(c). The first-order alias starts just above 5 Hz and the order of the alias grows significantly with frequency. Figure 4(d) shows the f-k_x-k_y transform of the upgoing pressure wavefield reconstructed by GMP, to a 25 m x 25 m spatial grid. The ghost notch is filled and the dealiasing impact of GMP is evident if we compare the output shown here with the spectrum of the input in the previous figures at high frequencies. Comparison of Figures 4(b) and 3(d) confirms the accuracy of joint interpolation and deghosting achieved by GMP.

IV. Summary And Conclusions

The problem of reconstructing a bandlimited signal from highly aliased multichannel samples was considered and a solution proposed in the form of Generalized Matching Pursuit. GMP proceeds by modeling the target signal as a sum of parametric basis functions that are matched to the multichannel data in a simultaneous and iterative fashion through application of the respective linear operators. It was shown that under quite general conditions GMP can achieve high-quality reconstructions of signals aliased by orders significantly higher than the number of different measurements, including the notoriously difficult case of regular undersampling, and signals for which no direct measurements are available.

We should emphasize that the results shown in this paper were obtained without using any priors (e.g., using a low-frequency solution, which is assumed to be unaliased, to constrain a high-frequency solution), which are commonly utilized to interpolate aliased data. In the same vein, the reconstructions were carried out independently at each temporal frequency, i.e., without making any assumptions on local wavefronts being planar.

During the presentation, we intend to show results obtained using real data acquired by multimeasurement towed-streamer seismic data acquisition technology; we had to omit them from this paper due to lack of space.

Acknowledgement

The synthetic data set was generated as part of collaborative projects between Schlumberger, Lawrence Livermore National Laboratory, and Statoil; we thank Shawn Larsen of Lawrence Livermore National Laboratory and Martin Musil and Clément Kostov of Schlumberger. We thank our colleagues Johan Robertsson, Tony Curtis, Smaine Zeroug, Ed Kragh, Phil Christie, Everhard Muyzert, and Ralf Ferber for stimulating discussions. We also thank Statoil for permission to show the synthetic data.

References

[1] A. Papoulis, "Generalized sampling expansion," *IEEE Trans. Circ. Syst.,* vol. 24, pp. 652-654, Nov. 1977.

[2] J. L. Brown, Jr., "Multi-channel sampling of low-pass signals," *IEEE Trans. Circ. Syst.,* vol. 28, pp. 101-106, Feb. 1981.

[3] J. L. Brown, Jr., and S. D. Cabrera, "Multi-Channel Signal Reconstruction Using Noisy Samples," *Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc.*, vol. 3, pp. 1233 – 1236, Albuquerque, NM, April 1990.

[4] N. R. Lomb, "Least squares frequency analysis of unequally spaced data," *Astrophysics and Space Science,* vol. 39, pp. 447–462, 1976.

[5] J. D. Scargle, "Studies in astronomical time series analysis II. statistical aspects of spectral analysis of unevenly sampled data," *Astrophysical Journal*, vol. 263, pp. 835–853, 1982.

[6] K. Özdemir, A. Özbek, and M. Vassallo, "Interpolation of irregularly sampled data by matching pursuit," *Proc. EAGE Conference*, paper G025, Rome, June 2008.

[7] D. A. Linden, "A discussion of sampling theorems," *Proc. IRE*, vol. 47, pp. 1219-1226, 1959.

[8] J. Robertsson, I. Moore, M. Vassallo, A. K. Özdemir, D.-J. Van Manen, and A. Özbek, "On the use of multicomponent streamer recordings for reconstruction of pressure wavefields in the crossline direction," *Geophysics*, vol. 73, pp. A45–A49, 2008.

[9] M. Vassallo, A. Özbek, K. Özdemir, and K. Eggenberger, "Crossline wavefield reconstruction from multicomponent streamer data: Part 1 - Multichannel interpolation by matching pursuit using pressure and its crossline gradient," *Geophysics*, vol. 75, pp. WB53–WB67, 2010.

[10] A. Özbek, M. Vassallo, K. Özdemir, D.-J. Van Manen, and K. Eggenberger, "Crossline wavefield reconstruction from multicomponent streamer data: Part 2 - Joint interpolation and 3D up/down separation by generalized matching pursuit," *Geophysics*, vol. 75, pp. WB69–WB85, 2010.

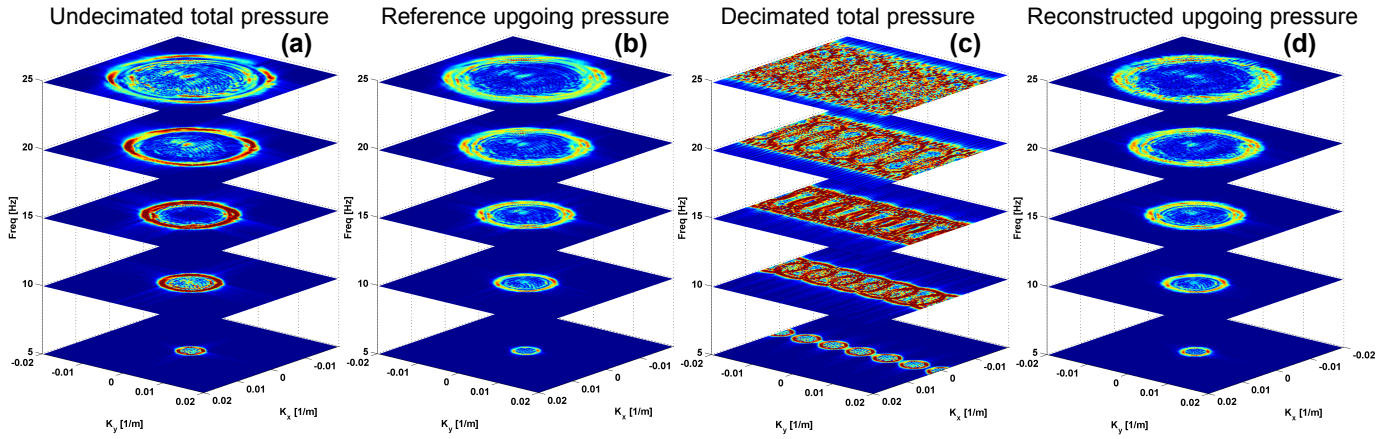

Fig. 4. Seismic synthetics in the 3D spectrum (f-k_x-k_y) domain. (a) reference total pressure wavefield, sampled over a 25- x 25-m grid; (b) reference upgoing pressure wavefield, sampled over the same grid; (c) input total pressure wavefield at 150-m crossline spacing; (d) upgoing pressure wavefield, reconstructed over a 25- x 25-m grid by GMP, processing P, P_y, and P_z at 150 m in the crossline.