Geophysical Prospecting

Appraisal problem in the 3D Least-Squares Fourier seismic data reconstruction

Journal:	*Geophysical Prospecting*
Manuscript ID:	GP-2013-1582.R1
Manuscript Type:	Reviews
Date Submitted by the Author:	29-Jan-2014
Complete List of Authors:	CIABARRI, FABIO; University of Pisa, Earth Sciences Department Mazzotti, Alfredo; University of Pisa, Earth Sciences Department Stucchi, Eusebio; University of Milano, Earth Sciences Department Bienati, Nicola; Eni S.p.A. - Exploration & Production Division,
Keyword:	Inversion, Seismic

Appraisal problem in the 3D Least-Squares Fourier seismic data reconstruction

Fabio Ciabarri (1), Alfredo Mazzotti (1), Eusebio Stucchi (2), Nicola Bienati (3)

(1) University of Pisa, Earth Sciences Department, Via S.Maria 53, 56126 Pisa, Italy

(2) University of Milano, Earth Sciences Department, Via Cicognara 7, 20129 Milano, Italy

(3) Eni S.p.A. - Exploration & Production Division,Via Emilia 1, 20097 San Donato Milanese (Milano) - Italy

ABSTRACT

Least-squares Fourier reconstruction is basically a discrete linear inverse problem that attempts to recover the Fourier spectrum of the seismic wave-field from irregularly sampled data along the spatial coordinates. The estimated Fourier coefficients are then used to reconstruct the data in a regular grid via a standard inverse Fourier transform (IDFT or IFFT). Unfortunately, this kind of inverse problem is usually under-determined and ill-conditioned. For this reason the LS Fourier reconstruction with minimum norm (FRMN) adopts a damped least-squares inversion to retrieve a unique and stable solution.

In this work we show how the damping can introduce artefacts on the reconstructed 3D data. To quantitatively describe this issue, we introduce the concept of "extended" model resolution matrix (EMRM) and we formulate the reconstruction problem as an appraisal problem. Through the simultaneous analysis of the EMRM and of the noise term, we discuss the limits of the FRMN reconstruction and we assess the validity of the reconstructed data and the possible bias introduced by the inversion process. Also, we can guide the parametrization of the forward problem to minimize the occurrence of unwanted artefacts. A simple synthetic example and real data from a 3D marine common shot gather are used to discuss our approach and to show the results of FRMN reconstruction.

KEYWORDS

Inverse problem, Inversion, Parameter estimation, Seismic.

INTRODUCTION

Seismic data are almost always irregularly sampled along the spatial directions due to obstacles, cable feathering, faulty equipment and logistic constraints in the acquisition phase, or due to the merging of independently acquired data with different spreads and grid orientations, such as in multi-azimuth or rich azimuth surveys.

Non-uniform and sparse spatial sampling can introduce noise and artefacts in the outcomes of

many seismic data processing algorithms (see, e.g., Aaron *et al*. 2007), especially in case of time-lapse acquisitions (see, e.g., Schonewille 2003 and Smith *et al*. 2012). Coarse sampling can also lead to an adverse effect in AVO/AVA/AVAz analysis (Downton *et al.* 2012; Sacchi and Liu 2005) and limit the effectiveness of several imaging algorithms (Downton *et al.* 2008; Gardner and Canning 1994). To overcome these problems, data regularization, or data interpolation, or even data extrapolation are commonly applied early in the processing flow. The regularization process can be described as the mapping of the seismic traces from their actual recording locations to the nodes of a regular grid. Interpolation is a process where new traces are created "in-between" existing regularly-spaced traces, while extrapolation entails the creation of new traces outside the range of input data. In this work we will use the term *reconstruction* to denote one or a combination of these processes.

In the practice of exploration seismology several reconstruction methods with different strengths and weaknesses (Abma and Kabir 2005; Mazzuchelli *et al.* 1998) have been proposed, which can be classified into five major groups: 1) binning methods, 2) wave equation methods, 3) interferometric methods, 4) filter based methods and 5) transform based methods.

Binning methods are fast and relatively simple in handling non-uniformly sampled seismic data (Yilmaz 1987). In these methods, the area of a seismic survey is divided into cells or bins with barycenters lying on the nodes of a regular grid; then the traces collected within each bin are stacked to yield an output trace for every bin. Binning methods, although widely used, are inaccurate because, neglecting the true positions of the recorded traces, they do not respect the spatial continuity of the seismic wavefield.

Wave equation methods use wavefield operators to reconstruct missing traces. Examples are the offset continuation operator (OCO) (Bagaini and Spagnolini 1996) and the shot continuation operator (SCO) (Fomel 2003). Moreover, a variety of continuation to zero offset implementations (better known as dip move-out DMO operators) have been used to reconstruct seismic data as described in Canning and Gardner (1996), while Chemingui and Biondi (2002) and Malcolm *et al*. (2005) use a similar operator called azimuth move-out (AMO) operator. Continuation operators are best applied to interpolate or extrapolate missing traces in otherwise regularly sampled data set, but they can be also used for regularization (Mazzuchelli and Rocca 1999). However, wave equation based methods require the *a-priori* knowledge of the wave velocity field and they are also fairly computationally intensive.

Interferometric methods are explicitly developed and used to interpolate and/or extrapolate missing traces. In some examples that refer to marine surveys (Dong and Schuster 2008; Wang *et al.* 2009) the original data are interferometrically cross-correlated with the surface or the seabed related multiples to estimate the kinematic of the primary events in the missing traces. The multiple events can be obtained from the data using an inverse focal transform (Berkhout and Verschuur 2006) or through a finite difference modeling of the wavefield in the water layer (Hanafy 2011). The performance of these methods is dependent on the reflector depths and on the available recording apertures which determine what events can be accurately predicted. A least-squares matching filter and amplitude corrections are needed to reconstruct the wavelet signature and to mitigate the inevitable artefacts due to a limited recording aperture.

Filter based reconstruction methods are widely used in seismic processing. The well know sinc-interpolation falls under this group and it is used to interpolate unaliased uniformly sampled data to a finer grid.

Common methods to handle alias in seismic data reconstruction use the prediction-error filter (PEF) in the frequency-space domain (Spitz 1991; Wang 2002) or in the frequency-wave-number domain (Gulunay 2003). In these methods, the low and unaliased frequencies are used to estimate the prediction filters needed to interpolate the high and aliased temporal frequency components. Similar methods were also developed to interpolate aliased data in the time-space domain or in the tau-space domain (Claerbout and Nichols 1991). The PEF based methods can also be used to fill gaps in a otherwise regularly sampled data, but PEF results degrade if the sampling is "random ". Furthermore, they assume a limited number of dipping events is present in the data.

In the *transform based* methods the assumption made is that the actual seismic observations can be suitably described in a transformed domain. Then, the inverse transform can be used to perform the synthesis of data at desired regular spatial locations. Commonly used transforms for seismic data reconstruction are: linear Radon transform (Lu 1985), hyperbolic Radon transform (Trad *et al.* 2002), parabolic Radon transform (Hugonnet and Canadas 1997; Kabir and Verschuur 1995), Fourier transform (Duijndam *et al.* 1999; Hindriks and Duijndam 2000) and "mixed" Fourier-Radon transform (Schonewille and Duijndam 1998).

Recently many other different transforms have been proposed. For instance, the pyramid transform (Kustowski 2010), and a group of digital wavelet-like transforms such as the seislet transform (Liu *et al.* 2012), the shearlet transform, the curvelet transform (Hennenfent and

Herrmann 2005; Herrmann and Hennenfent 2008), the S-transform (Naghizadeh and Innanen 2011), and the "mixed" wavelet-Radon transform (Yu *et al*. 2007).

Transform based techniques are fast when the transform can be computed efficiently and are particularly attractive because they need little user input and can be applied to any subset of seismic data. Furthermore, they do not require any *a-priori* information (that is the velocity model) and, in general, can handle either uniform grids with gaps or "random" spatial sampling.

To this group pertains the least-squares Fourier reconstruction which, since it appeared some years ago in the literature, may be considered "old" if compared to the most recent findings such as the antileakage Fourier transform ALFT (Xu *et al.* 2005; Xu *et al.* 2010) or the interpolation by matching pursuit, also knows as IMAP technique (Özdemir *et al.* 2008; Özbek *et al.* 2019). However, similar approaches are still widely used (see e.g. Jin 2010) and thus it can be of interest to quantify its limits.

Conceptually, LS Fourier reconstruction is a two step procedure (Figure 1, top row). In the first step a least-squares discrete Fourier transform (DFT_{LS}) is used to decompose a 3D seismic data set ($\mathbf{D}^{\wedge}_{\mathrm{IRR}}$), irregularly sampled along the two spatial directions, into plane wave components ($\mathbf{M}^{\wedge}_{\mathrm{EST}}$):

$$\mathbf{M}^{\wedge}_{\mathrm{EST}} = DFT_{LS}\,\mathbf{D}^{\wedge}_{\mathrm{IRR}} \qquad (1)$$

The symbol "^" indicates we are dealing with 3D matrices. In the second step the estimated Fourier coefficients $\mathbf{M}^{\wedge}_{\mathrm{EST}}$ are used to reconstruct the data ($\mathbf{D}^{\wedge}_{\mathrm{REG}}$) on a regular grid via an inverse Fourier transform (IDFT or IFFT):

$$\mathbf{D}^{\wedge}_{\mathrm{REG}} = IDFT\,\mathbf{M}^{\wedge}_{\mathrm{EST}} \qquad (2)$$

For the data matrices ($\mathbf{D}^{\wedge}_{\mathrm{IRR}}$ and $\mathbf{D}^{\wedge}_{\mathrm{REG}}$), two dimensions represent the spatial coordinates x and y and the third one represents the temporal frequency of the observed data. For the Fourier coefficients matrix $\mathbf{M}^{\wedge}_{\mathrm{EST}}$, two dimensions represent the wavenumbers k_x and k_y and the third one remains the temporal frequency.

In the first step an inversion approach is necessary because the direct transformation via standard DFT or via the Riemann Sum (NDFT) of the non-uniformly sampled data gives a poor and distorted estimation of the Fourier coefficients (spectral leakage). The inversion can be performed simultaneously on all temporal frequencies or on each frequency slice

separately.

The only requirement of this method is that the seismic data set is spatially band-limited (no aliasing). In general, this is hardly satisfied but the spatial bandwidth can be compressed by NMO correction.

LS Fourier reconstruction can indeed yield a solution but, as in many other inverse problems, its reliability is a major concern. A way to QC the estimated solution can be found by describing the inversion problem as an appraisal problem (Scales and Snieder 2000). Interesting examples of the use of model appraisal theory in geophysical inverse problems are showed by Routh and Miller (2006) for electrical and electromagnetic inversion and by Xia *et al.* (2010) for surface-wave inversion.

In this paper we apply the appraisal theory to LS Fourier reconstruction. In particular, to quantify the accuracy of the reconstructed regular data set at each spatial location, we extend the theory of model resolution analysis through the formulation of the "extended" model resolution matrix (EMRM). We start recalling the theories behind the 3D LS Fourier regularization and the appraisal problem. After, we introduce the concept of EMRM and we show how its computation is a key step to guide the forward model parametrization towards the achievement of a good and artefact free reconstruction. Finally, a simple synthetic example and the data of a common shot gather from a 3D marine survey are used to discuss the EMRM analysis and to show the results of the LS Fourier reconstruction.

THE LEAST-SQUARES FOURIER RECONSTRUCTION

The linear relation, indicated in Fig. 1 as the forward problem, between an irregularly sampled 3D seismic data set $\mathbf{D}^{\wedge}_{\mathbf{IRR}}$ and the discrete Fourier 3D spectrum of the related seismic wave-field $\mathbf{M}^{\wedge}_{\mathbf{TRUE}}$, is given by the 2D Inverse Discrete Fourier Transform (IDFT) as a double summation over the uniformly sampled k_x and k_y axis. For a given data trace $d_{IRR}(f, x_{IRR}, y_{IRR})$ we have:

$$d_{IRR}(f, x_{IRR}, y_{IRR}) = \Delta k_x \Delta k_y \sum_{mx=0}^{Mx-1} \sum_{my=0}^{My-1} m_{TRUE}(f, k_{x_{mx}}, k_{y_{my}}) e^{j2\pi(x_{mx} k_{x_{mx}} + y_{my} k_{y_{my}})} \quad (3)$$

where the actual spatial coordinates of the trace (x_{IRR}, y_{IRR}) appear in the Fourier exponential and Mx and My represent the number of Fourier coefficients along the k_x and k_y axes. If we arrange the 3D data matrix ($\mathbf{D}^{\wedge}_{\mathbf{IRR}}$) and the 3D Fourier spectrum ($\mathbf{M}^{\wedge}_{\mathbf{TRUE}}$) in a lexicographical order (Figure 2), the relation between the data space and the model space can be written in 2D matrix notation as:

$$\mathbf{D}_{\mathbf{IRR}} = \mathbf{G}\,\mathbf{M}_{\mathbf{TRUE}} + \mathbf{NOISE} \tag{4a}$$

As shown in Fig. 2 each column of the matrices $\mathbf{D}_{\mathbf{IRR}}$ and $\mathbf{M}_{\mathbf{TRUE}}$ refers to a single temporal frequency slice of $\mathbf{D}\hat{}_{\mathbf{IRR}}$ and $\mathbf{M}\hat{}_{\mathbf{TRUE}}$. Therefore, if we consider a single temporal frequency at the time, equation (4a) can be written in vector form as:

$$\mathbf{d}_{\mathbf{IRR}} = \mathbf{G}\,\mathbf{m}_{\mathbf{TRUE}} + \mathbf{noise} \tag{4b}$$

The Fourier matrix **G** is the forward operator with elements:

$$G_{nm} = \Delta k_x \Delta k_y e^{j2\pi(x_{IRR}(n)k_x(m) + y_{IRR}(n)k_y(m))} \tag{5}$$

where n and m are the trace coordinate and the wavenumber indices respectively (Fig. 2). In the case of regular spatial sampling the forward operator **G** simplifies to the Kronecker product between the 1D DFTs along the x and y axes respectively.

The sampling intervals Δk_x and Δk_y in the wavenumber domain must be chosen small enough to avoid wrap-around on the desired regular grid. The spatial bandwidth is limited to the interval $[-\frac{1}{2}M_x\Delta k_x, \frac{1}{2}M_x\Delta k_x]$ for the in-line direction and $[-\frac{1}{2}M_y\Delta k_y, \frac{1}{2}M_y\Delta k_y]$ for the cross-line direction and depends on the particular parametrization chosen. The components beyond the assigned k_xk_y-bandwidths form the "noise" in the forward model. Note that using equation (5) implies that we consider an orthogonal reference grid either for the spatial samples and for the related wave-number samples (Peterson and Middleton 1962). Other grids than the orthogonal grid can be used if e.g. the area of measurement is circular in the spatial domain, which may occur in 3D walk-away VSP acquisition. In that case a hexagonal grid in the wave-number domain is the most efficient grid because it needs the least number of Fourier coefficients (Hindriks *et al.* 1997).

Depending on the degree of irregularity of the spatial sampling, the structure of the forward operator **G** can be strongly ill-conditioned. To compute a stable inverse and retrieve an unique and "noise" free Fourier spectrum, additional constraints or *a-priori* information on model parameters must be incorporated in the inversion scheme. Unfortunately, for this type of inverse problem there is not an *a-priori* knowledge in the Bayesian sense on either the model parameters and the noise distribution. Therefore, to overcome the combination of noise and ill-conditioning, LS Fourier reconstruction with minimum norm (FRMN) adopts one of the

most simple strategies that consists in adding a restriction on the Euclidean model norm to a standard least-squares data-misfit term (Hindriks and Duijndam 2000). If we consider a single temporal frequency at the time (that is a single column of the matrix $\mathbf{D}_{\mathbf{IRR}}$), the estimated model becomes a vector that can be written as:

$$\mathbf{m}_{\mathbf{EST}} = argmin_{m}\{\|\mathbf{d}_{\mathbf{IRR}} - \mathbf{G}\,\mathbf{m}\|_2^2 + \lambda^2\|\mathbf{m}\|_2^2\} \tag{6}$$

where $\mathbf{m}$ is a generic model and λ is a positive constant which role is to properly balance the respective influence of the data-misfit (first term of equation (6)) and of the quadratic penalty term (second term of equation (6)). If λ=0, equation (6) degenerates to the standard objective function for the linear least-squares problem. The normal equation associated to the objective function in equation (6) is:

$$\mathbf{m}_{\mathbf{EST}} = \mathbf{G}^{-g}\mathbf{d}_{\mathbf{IRR}} = (\mathbf{G}^{\mathbf{H}}\mathbf{G} + \lambda^2\mathbf{I})^{-1}\mathbf{G}^{\mathbf{H}}\mathbf{d}_{\mathbf{IRR}} \tag{7}$$

where the matrix $\mathbf{G}^{-g}$ is the generalized inverse of the direct operator $\mathbf{G}$ and the superscript "H" denotes the complex conjugate transpose. Other Fourier reconstruction schemes using sparsity constraints have been proposed by Zwartjes and Gisolf (2007) based on the high resolution Fourier transform formulated by Sacchi and Ulrich (1996).

In this study, similarly to the FRMN approach, we use a minimum energy constraint on the model but, instead of using a damped least-squares scheme as in equation (7), we compute the pseudo-inverse of the Gram matrix ($\mathbf{G}^{\mathbf{H}}\mathbf{G}$) by truncated SVD (TSVD):

$$\mathbf{m}_{\mathbf{EST}} = \mathbf{G}^{-g}\mathbf{d}_{\mathbf{IRR}} = (\mathbf{G}^{\mathbf{H}}\mathbf{G})^{\dagger}\mathbf{G}^{\mathbf{H}}\mathbf{d}_{\mathbf{IRR}} \tag{8}$$

where $\dagger$ denotes the Moore-Penrose pseudo-inverse (Aster *et al.* 2005).

The equivalence between employing a damping factor λ as in equation (7) and the application of a TSVD as in equation (8) can be easily demonstrated via the Filter Factor analysis (Aster *et al.* 2005). Using the SVD representation of the forward operator $\mathbf{G}$, both the equations (7) and (8) can be written as:

$$\mathbf{m}_{\mathbf{EST}} = \mathbf{G}^{-g}\mathbf{d}_{\mathbf{IRR}} = (\mathbf{V}_{\mathbf{P}}\mathbf{F}\mathbf{S}_{\mathbf{P}}^{-1}\mathbf{U}_{\mathbf{P}}^{\mathbf{H}})\mathbf{d}_{\mathbf{IRR}} \tag{9}$$

where the subscript "p" indicates the rank of the matrix $\mathbf{G}$ and the columns of the unitary

matrices $\mathbf{U_p}$ and $\mathbf{V_p}$ are the first p left and the first p right singular vectors of the matrix $\mathbf{G}$ respectively. $\mathbf{S_p}$ is a p by p diagonal matrix composed of the singular values of $\mathbf{G}$ and $\mathbf{F}$ is also a p by p diagonal matrix with real entries, called Filter Factor matrix, that describes how the stabilization terms (the TSVD threshold value or the λ parameter) "filter out" the small singular values of the Gram matrix ($\mathbf{G^H G}$). For a TSVD approach (equation (8)) the diagonal elements of the Filter Factor matrix are:

$$F(i,i)=\begin{cases}1 \; if \; S_P(i,i)^2 > tol_{TSVD} \\ 0 \; if \; S_P(i,i)^2 < tol_{TSVD}\end{cases} \tag{10}$$

This simply means that all the singular values (and their associated model space vectors) smaller than the TSVD threshold (tol_{TSVD}) are rejected in the estimation process. For a damped least-squares approach (equation (7)) the diagonal elements of the Filter Factor matrix are:

$$F(i,i)=\frac{S_P(i,i)^2}{S_P(i,i)^2+\lambda^2} \tag{11}$$

This produces a similar effect to that of equation (10): for $S_p(i,i)>>\lambda$, $F(i,i)=1$ and for $S_p(i,i) << \lambda$, $F(i,i)=0$.

For the LS Fourier reconstruction we have experimentally found that setting the TSVD threshold and the damping parameter λ at the same value, the results of two methods are hardly distinguishable. It is important to evidence that both the previous approaches require the setting of an adequate value for the damping term or for the TSVD threshold. In the literature many different methods exist for choosing these parameters: the most popular are the L-curve, the Generalized Cross-Validation (GCV) and the Discrepancy principle (Hansen 1998). However, in general these methods are designed for small inverse problems and for 3D LS Fourier reconstruction they become too expensive to apply.

In this work, to set an appropriate TSVD threshold we use some practical rules which are based on the singular value analysis of the Gram matrix ($\mathbf{G^H G}$). In case of regular sampling the matrix $\mathbf{G}$ is unitary and this means that the Gram matrix is:

$$(\mathbf{G^H G})=(\mathbf{G G^H})=(\Delta k_x \Delta k_y)^2 \mathbf{I} \tag{12}$$

and all the singular values of the Gram Matrix will be equal to the $(\Delta k_x \Delta k_y)^2$ product. In case of non-uniform sampling the structure of the Gram matrix becomes ill-conditioned and its

singular values decay gradually towards zero. We have experimentally found that setting a threshold value (or the damping term λ) to one or two orders of magnitude less than the $(\Delta k_x \Delta k_y)^2$ product generally brings to a noise free data reconstruction.

Note that we consider the bare FRMN technique, with no a priori assumption on the model space, such as the sparseness constraint as in Zwartjes and Gisolf (2007), and without resorting to the inclusion of weighting terms either on the model space, such as the spectral power of the estimated Fourier spectrum as in Liu and Sacchi (2004) and Trad (2009), or on the data space, such as by means of Voronoi cells as in Hindriks and Dujindam (2000).

THE "EXTENDED" APPRAISAL PROBLEM

In the ideal case, the estimated Fourier spectrum $\mathbf{M}_{\mathbf{EST}}$ equals the true Fourier spectrum $\mathbf{M}_{\mathbf{TRUE}}$ of the regularly sampled seismic wave field that we seek (equation (4a)). In general, this can never happen for an ill-posed linear system for which a generalized inverse ($\mathbf{G}^{-g}$) must be constructed incorporating some constraints (e.g., a minimum energy constraint on model parameters) to obtain a unique and stable solution. The relation between $\mathbf{M}_{\mathbf{EST}}$ and $\mathbf{M}_{\mathbf{TRUE}}$ becomes evident if we describe the inversion problem as an Appraisal problem (red arrow in Fig. 1) (Scales and Snieder 2000):

$$\mathbf{m}_{\mathbf{EST}} = \left(\mathbf{G}^{-g}\mathbf{G}\right)\mathbf{m}_{\mathbf{TRUE}} + \mathbf{G}^{-g}\,\mathbf{noise} \qquad (13)$$

The equation (13), obtained inserting equation (4b) into equation (8) (or into equation (7)), allows us to determine, for each temporal frequency, what components of the true model are recovered in the estimated model (first term on the right side of equation (13)) and what errors are attached to it (second term on the right side of equation (13)). The analysis of the square matrix ($\mathbf{G}^{-g}\mathbf{G}$) known as model resolution matrix (MRM), shows how the true model is filtered by the estimation process. The diagonal entries of the MRM quantify the degree of independent resolvability of each model parameter, while the off-diagonal elements represent the smearing between parameters in the recovered solution. The MRM depends on the forward operator ($\mathbf{G}$) in terms of forward problem parametrization and on the generalized inverse ($\mathbf{G}^{-g}$) structure.

Because $\mathbf{G}^{-g}$ is normally constructed to control the influence of the noise on the estimated model, the MRM depends indirectly on noise level. In this sense, equation (13) clearly evidences the well known dilemma in ill-posed inversion problems, that is the trade-off between the requirements imposed by noise attenuation and by resolution. How the minimum energy constraint affects the model resolution can be evidenced using SVD analysis.

Combining the SVD representation of the forward model (**G**) with the SVD representation of the generalized inverse ($\mathbf{G}^{-g}$) we can express the MRM as:

$$\mathbf{G}^{-g}\mathbf{G}=(\mathbf{V}_P\mathbf{F}\mathbf{S}_P^{-1}\mathbf{U}_P^H)(\mathbf{U}_P\mathbf{S}_P\mathbf{V}_P^H) \tag{14}$$

where **F** is the filter factor described by equation (10) for a TSVD approach or by equation (11) for a damped least-squares approach. Equation (14) shows that the introduction of a TSVD threshold (or a damping term) inevitably produces a non-identity MRM because the matrix **F** is not an identity matrix.

Therefore, the analysis of the MRM allows us to quantify the reliability of the estimation process. To QC the regular data ($\mathbf{D}^{\wedge}_{REG}$), that were reconstructed through the two steps of equations (1) and (2) (top row of Fig. 1), we extend to the data domain the relation expressed in equation (13). To this end, we recall that $\mathbf{m}_{EST}$ and $\mathbf{m}_{TRUE}$ constitute a Fourier pair with $\mathbf{d}_{REG}$ and $\mathbf{d}_{TRUE}$, respectively. Therefore, as indicated by the green arrow in Fig. 1, we may rewrite equation (13) as:

$$\mathbf{A}^H\mathbf{d}_{REG}=(\mathbf{G}^{-g}\mathbf{G})\mathbf{A}^H\mathbf{d}_{TRUE}+\mathbf{G}^{-g}\mathbf{noise}$$

that gives

$$\mathbf{d}_{REG}=\mathbf{A}(\mathbf{G}^{-g}\mathbf{G})\mathbf{A}^H\mathbf{d}_{TRUE}+\mathbf{A}\,\mathbf{G}^{-g}\mathbf{noise} \tag{15a}$$

where $\mathbf{A}^H$ implements the 2D Fourier transform in the wave-number domain.

We call $\mathbf{A}(\mathbf{G}^{-g}\mathbf{G})\mathbf{A}^H$ the "extended" model resolution matrix (EMRM). Considering all the temporal frequencies, we may rewrite equation (15a) as:

$$\mathbf{D}_{REG}=\mathbf{A}(\mathbf{G}^{-g}\mathbf{G})\mathbf{A}^H\mathbf{D}_{TRUE}+\mathbf{A}\,\mathbf{G}^{-g}\mathbf{NOISE} \tag{15b}$$

Each row of the EMRM determines one reconstructed trace at a specific (x_{REG}, y_{REG}) location by weighting the traces of the unknown regular data $\mathbf{D}_{TRUE}$. The closer the EMRM is to the identity matrix, the smaller is the bias inherent to the inversion and the higher is the fidelity of the reconstructed data to the true unknown regular data. The second term of equations (15a) and (15b) shows how the noise is mapped onto the reconstructed data set.

ANALYSIS OF THE APPRAISAL PROBLEM ON A SYNTHETIC EXAMPLE

The FRMN and EMRM analysis previously discussed are now applied to a synthetic example that consists of four plane dipping events (Figure 3(a)) randomly sampled along the in-line and cross-line coordinates. In Fig. 3(b) the desired regular grid is represented by the blue dots (in-line interval Δx: 4 m, cross-line interval Δy: 4 m) and the actual locations of the synthetic traces are shown by the red dots (in-line and cross-line intervals equal the desired regular distances randomly perturbed with deviations ranging within +/-2 meters). The non-uniformly sampled synthetic data set and its Riemann Sum (NDFT) are displayed in Figure 4(a) and 4(b), while the regular data we want reconstruct and its true Fourier spectrum are shown in Fig. 4(c), and 4(d), respectively. Note that the terms $\mathbf{D}_{IRR}$ (in time domain) and $\mathbf{M}_{TRUE}$ in equation (4a) are represented by the data in Fig. 4(a) and 4(d), respectively. Fig. 4(d) shows that the highly dipping events at shallow times (see Fig. 4(c)) are severely aliased even in the case of regular sampling and thus are outside the bandwidth considered for the reconstruction. From equation (4a), given $\mathbf{M}_{TRUE}$ and $\mathbf{D}_{IRR}$, it is possible to compute the **NOISE** component that we represent in Fig. 4(e) in time domain. Most of the **NOISE** term is due to the aliased shallow events and only the horizontal reflections, that are fully within the considered bandwidth, do not yield any contribution.

We start applying the FRMN as a simple least squares inversion without setting any threshold in the Gram matrix inversion (equation (9) with zero threshold in equation (10)). The estimated spectrum ($\mathbf{M}_{EST}$) and the regularized data ($\mathbf{D}_{REG}$) are illustrated in Figure 5(a) and 5(b). In this case, both the MRM and EMRM are identity matrices (Fig. 5(c, d)) and thus, in absence of noise, the obtained data ($\mathbf{D}_{REG}$, Fig. 5(b)) should equal the theoretical regularly sampled data ($\mathbf{D}_{TRUE}$, Fig. 4(c)). Instead, the noise (**NOISE**, Fig. 4(e)) that acts on the second term of equation (15b), or if we consider its vector form, on the second term of equation (15a), is boosted by the lowest singular values (red dots at the highest singular value numbers in Figure 6) and strongly compromises the reconstruction. The effect of noise amplification due to the ill-conditioning of the forward operator **G** is clearly evident in Fig. 5(e) which depicts the projection of the NOISE term into the reconstructed data by means of the operator $\mathbf{AG}^{-g}$ in equation (15b).

Setting an appropriate threshold is needed to prevent noise contamination. The results obtained with a threshold of $0.1(\Delta k_x \Delta k_y)^2$ are shown in Figure 7 (a, b). Although the noise projection onto the reconstructed data is strongly attenuated (Fig. 7(e)), unbalanced wavelet amplitudes and time misalignments do appear on the reconstructed data (Fig. 7(b)). The reason is found analysing the behaviour of both the MRM and EMRM. The MRM is not any more an identity matrix (Fig. 7(c)) and the EMRM shows anomalous low values along the

diagonal and non-zero off-diagonal elements (Fig. 7(d)).

Thus, it would seem we are stuck in a circular problem: if we do not apply a threshold in the matrix inversion (or, equivalently, a damping factor) the noise compromises the reconstruction (Fig. 5a, b), and conversely if we do apply a threshold, intolerable artefacts contaminate the reconstructed data (Fig. 7a, b).

A way to limit this problem is to better parametrize the forward model operator, i.e. to select the appropriate spatial bandwidths and wavenumber sampling intervals for the in-line and cross-line directions, which determine the number of Fourier coefficients to be estimated. Setting the k_x and k_y values to half the Nyquist wavenumbers of the regular grid and the Δk_x and Δk_y such that to avoid wraparound effects gives a Gram matrix with all singular values above the threshold of $0.1(\Delta k_x \Delta k_y)^2$ (Fig. 6, green dots).

The FRMN allows a stable inversion and a free noise estimation of the Fourier coefficients (Figure 8(a)). The MRM is an identity matrix (Fig. 8(c)) which indicates that we have correctly estimated the Fourier coefficients of the true *band-limited* Fourier spectrum. With this parametrization, the reconstruction of the data to a finer regular grid implies a sinc-interpolation along both the in-line and cross-line direction, as evidenced by the features along the super- and sub-diagonals of the EMRM (Fig. 8(d)). Note that the main diagonal has values around 0.5 (red curve) and that the sinc interpolation gets part of the amplitude from the sub-diagonals. The reconstructed data are shown in Fig. 8(b). Note that they are relatively noise free, with correct wavelet amplitudes and alignments but the highly dipping reflections have been almost lost. Fig. 8(e) shows the projection of the noise into the reconstructed data, that is the $\mathbf{AG}^{\#}$ **NOISE** term in equation (15b): the bandwidth limitation brings to a reduction of the disturbances into the reconstructed data. It may be of interest to know what happens to the shallow and highly dipping events present on the observed data (Fig. 4(a)). With this synthetic example it is easy to answer since $\mathbf{M}_{\mathbf{TRUE}}$ in equation (4a) is known and thus we can compute the **NOISE** term. Figure 9 shows this term transformed into the time domain: it is clear that it is dominated by the highly dipping events. As it is evident from equation (4a), it could not be otherwise since their wave-numbers have been excluded by the new parametrization and thus these reflections have to fall into the **NOISE** term.

EMRM ANALYSIS OF A MARINE COMMON SHOT GATHER

We now apply the FRMN and EMRM analysis to a 3D marine common shot gather that was acquired by means of multiple passes of seismic vessels. Figure 10 shows its source and receivers positions. The receiver locations of the actual data are shown by the red dots and the desired regular grid of receivers is represented by the blue dots (receiver interval Δx: 12.5 m,

cross-line interval Δy: 120 m). Note the strong irregularity of the actual spatial sampling with a large gap in the cross-line direction and the asymmetric location of the source. Of the 17790 traces of the nominal grid (blue dots in Fig. 10), only 16670 (red dots in Fig. 10) have been actually acquired, of which just 1021 have coordinates within 5 m from the nominal coordinates (the few superimposed blue and red dots in Fig. 10), corresponding to about the 6% of the total nominal traces. Also, a significant number of traces are missing at either short and long offsets (543 and 354 traces, respectively) and have to be extrapolated at the two ends of the streamers in Fig. 10. Then, with respect to the previous synthetic example we now face a slightly different problem because the sampling irregularity is no more random but it is due to gaps of various size and to the streamer feathering in an otherwise regular grid. The reconstruction has been performed applying the FRMN technique to ensemble of traces pertaining to rectangular sliding windows that span all the receiver coordinates. The data were previously NMO corrected with a constant velocity of 1550 m/s and muted at offsets greater than 7600 m to prevent overstretching effects. We start describing the results for the spatial window indicated by the green rectangle in Fig. 10. The window size is X=1237.5 m along the in-line direction and Y= 3480 m along the cross-line direction.

Schonewille (2000) experimentally showed that, in case of large gaps in the data, a side effect of the damping term in FRMN is to zero the reconstructed output within the same missing data areas and he reported how difficult it is to find a compromise between the free-noise reconstruction of the more densely sampled parts of the data and the optimal reconstruction in gaps. In the following tests, through the analysis of the EMRM, we show how the side effects of minimum energy constraint can be even more disastrous.

We perform on the common shot gather the same set of experiments as done with the synthetic example. However, due to the fact that the true spectrum ($\mathbf{M}_{\mathbf{TRUE}}$) is in the actual case obviously unknown, we cannot compute the **NOISE** term as done with the synthetic data. We start applying the FRMN as a simple least squares inversion (no threshold in the Gram matrix inversion). The results are illustrated in Figure 11. The nature of the MRM and EMRM, that are both identity matrices, indicates that if we had no noise, the regularized data ($\mathbf{D}_{\mathbf{REG}}$) should exactly match the theoretical (and unknown, for this real data case) regularly sampled data ($\mathbf{D}_{\mathbf{TRUE}}$). As demonstrated in the synthetic example (Fig. (5)), the lowest singular values of the Gram matrix (red dots at the highest singular value numbers in Figure 12) amplify the noise preventing the reconstruction (Fig. 11(a, b)). The results obtained excluding the lowest singular values by setting a threshold of $0.1(\Delta k_x \Delta k_y)^2$ are shown in Figure 13 (a, b): the noise is strongly attenuated, but unacceptable artefacts do appear (see e.g. the decreased amplitudes of the reconstructed data for in-line 20). This happens notwithstanding the fair

behaviour of the MRM, which has a dominant, nearly unitary, main diagonal and few small off-diagonal elements (Fig. 13(c)). The reason is evident on the EMRM (Fig. 13(d)) which shows low values along the diagonal and non-zero off-diagonal elements.

As demonstrated with the synthetic data, a simple way to attenuate this problem is to re-parametrize the forward model operator in such a way to limit the considered spatial bandwidths and to chose appropriate wavenumber sampling intervals for the in-line and cross-line directions.

Analysis of the EMRM can guide the selection of both parameters. To avoid wrap-around effects on the reconstructed data, as occurs for the data shown in Figure 14(a), the k-sampling intervals (Δk_x and Δk_y) must be chosen small enough such that the periodicity in spatial domain is larger than the aperture (X, Y) of the final regular grid within the reconstruction window. The EMRM in Fig. 14(b) shows that if the Δk_y is selected as in the standard DFT, serious edge-effects are introduced to the reconstructed data. An optimal value of Δk_y (analogous to zero-pad the irregularly sampled data before inversion) is shown in Fig. 14(c), where the EMRM is diagonal without side-lobes. The correct choice of the spatial bandwidth for an irregularly spaced data is a highly debated issue. In the experiment we discuss, we limited the in-line bandwidth to the Nyquist of the nominal receiver interval (12.5 m) and the cross-line bandwidth to the Nyquist of the largest cross-line gap (240 m). No threshold in the Gram matrix inversion was set. The singular values of the Gram matrix after the re-parametrization are showed in Fig. 12 as green dots: all singular values fall above the threshold of $0.1(\Delta k_x \Delta k_y)^2$. The result we obtain is stable and artefacts free (Figure 15 (a, b)). The MRM is an identity matrix (Fig. 15(c)) indicating that the Fourier coefficients of the true band-limited Fourier spectrum have been correctly estimated. The reconstruction of the data to a finer regular grid implies a sinc-interpolation along the cross-line direction as the features close to the main diagonal of the EMRM indicate (Fig. 15 (d)).

Figure 16(a) shows the original, irregularly sampled, in-line 1 (Fig. 10, red dots), while Fig. 16(b), 16(c) and 16(d) illustrate the results of different regularizations that reconstruct the data along the regularly sampled in-line 1 (Fig. 10, blue dots). Note the significant deviation from the actual in-line 1 (lowest red dots in Fig. 10). The data reconstructed by FRMN imposing a threshold in the Gram matrix inversion (Fig. 16(b)) show anomalous amplitudes. The analysis of the EMRM (Fig. 13(d)) shows that the amplitude anomalies are correlated to the gaps in the main diagonal. Such errors do not appear on the data reconstructed after the mentioned re-parametrization of the forward operator (Fig. 16(c)). A comparison with the data of the actual in-line confirms the goodness of the reconstruction (Fig. 16(a)). However, as it is clearly shown in Fig. 16(c), the proposed re-parametrization is not suited to reconstruct

(extrapolate) the missing traces at short source-receiver offsets (Fig. 10, yellow window). Extrapolation is a hard task for the FRMN algorithm due to the under-determined structure of the forward operator (matrix **G**). However, guided by the EMRM analysis (not discussed here) we can set for the reconstruction of the traces included in the short offset window (yellow rectangle in Fig. 16a, 16b and 16c), an *ad-hoc* forward model parametrization which enables a proper estimation of the low wavenumber components of the true Fourier spectrum. Fig. 16(d) shows the reconstructed in-line 1 when, for the yellow window, we further limited the forward model bandwidth in the in-line and in the cross-line directions to half the Nyquist of the largest in-line gap (250 m) and to half the Nyquist of the nominal line interval (120 m), respectively. The imposed wavenumber limitations enabled a correct and noise free extrapolation of the flat and/or of the moderately dipping events. Obviously, due to the spatial bandwidth constraints, events showing significant dips (visible at the highest offsets in the yellow window) could not be extrapolated to the missing short offsets of the common shot. As demonstrated in the synthetic example, the highly dipping events are to be found in the noise component of equation (4a), being filtered out by the term $\mathbf{AG}^*$ (equation (15b)) in the reconstruction process.

As a further example, Figure 17 shows the reconstruction for the in-line 20 in the middle of the cross-line gap on the irregularly sampled grid (Fig. 10). Fig. 17(a) and 17(b) display, respectively, the irregularly sampled in-line 20 and the reconstructed in-line 20 imposing a threshold in the Gram matrix inversion. Significant artefacts do appear. Fig. 17(c) represents the FRMN reconstruction after a common re-parametrization of the forward operator for the whole shot gather, while Fig. 17(d) shows the reconstructed in-line 20 using an *ad-hoc* parametrization for the short offset traces in the yellow window. Conclusions similar to those for Fig. 16 can be drawn.

CONCLUSIONS

Fourier reconstruction based solely on the minimum energy constraint on the model not only is incapable to correctly reconstruct the signal in spatial gaps, but it may also jeopardize the reconstructed data adding unwanted spurious artefacts. This has been demonstrated both by the synthetic and actual data examples. Also, the rather severe test on the actual shot gather indicates that large data gaps are very difficult to be reconstructed because of the strong under-determinacy and ill-conditioning caused by the large number of missing or misplaced traces. This can be (partly) fixed by re-parametrizing the forward problem operator, that is by limiting the number of the Fourier coefficients to be estimated to the low wavenumbers of the spectrum. Obviously, this results in the impossibility to recover highly dipping events and in

reconstructed traces that show a distinct low frequency appearance. In other less severe situations, with irregular spatial sampling but without large data gaps, FRMN performs fairly well.

In any case, the formulation of the FRMN as an appraisal problem allows us to quantitatively describe the trade off between stability and resolution and the analysis of the EMRM helps to choose the forward problem parametrization that is crucial for a good performance of the FRMN. ~~It also enables to QC the reconstructed data.~~

~~In other words, by examining the EMRMs of different areal subsets of the entire 3D data volume, it is possible to quantitatively assess how much "large" the spatial gaps and the sampling irregularities can be for a certain model parametrization to be tolerated. This can be carried out before the actual computations are made for the whole data volume thus enabling a proper parameter selection for the various subsets.~~

ACKNOWLEDGEMENTS

The authors wish to thank ENI for the permission to publish the results. Seismic data used for the examples shown in this article are courtesy of WesternGeco. The preprocessing of the seismic data was performed by means of the Promax 3D Software of Landmark Graphics Corporation. We thank the Editor T.J Moser for his advice and the Associate Editor H. Chauris and an unknown referee for their constructive comments and suggestions.

REFERENCES

Abma R. and Kabir N. 2005. Comparisons of interpolation methods. The Leading Edge, **24**, 984-989.

Aron P., Barnes S., Schonewille M., van Borselen R. 2007. Data regularization for 3D SRME: a comparison of methods. 77th SEG meeting, San Antonio, Texas, Expanded Abstracts,1987-1991.

Aster R.C., Borchers B. and Thurber C. 2005. Parameter estimation and inverse problems. Academic Press. ISBN 9780123850485.

Bagaini C. and Spagnolini U. 1996. 2D continuation operators and their applications. Geophysics, **61**, 1846-1858.

Berkhout A. J. and Verschuur D. J. 2006. Imaging of multiple reflections. Geophysics, **71**,

SI209-SI220.

Canning A. and Gardner G. 1996. Regularizing 3D data sets with DMO. Geophysics, **61**, 1103-1114.

Chemingui N. and Biondi B. 2002. Seismic data reconstruction by inversion to common offset. Geophysics, **67**, 1575-1585.

Claerbout J. F., and Nichols D. 1991. Interpolation beyond aliasing by (tau,x) domain PEFs. 53rd EAGE meeting, Florence, Italy, Expanded Abstract, 2-3.

Dong S. and Schuster G. T. 2008. Interferometric extrapolation of OBS and SSP data. 78th SEG meeting, Las Vegas, Nevada, Expanded Abstracts, 3013-3017.

Downton J., Durrani B., Hunt L., Hadley S., Hadley M. 2008. 5D Interpolation, PSTM and AVO Inversion. 78th SEG meeting, Las Vegas, Nevada, Expanded Abstracts, 237-241.

Downton J., Hunt L., Trad D., Reynolds S., Hadley S. 2012. 5D interpolation to improve AVO and AVAz: a quantitative case history. Canadian Journal of Exploration Geophysics, **37**, 8-17.

Duijndam A. J. W., Schonewille M. A. and Hindriks C. O. H. 1999. Reconstruction of band-limited signals, irregularly sampled along one spatial direction. Geophysics, **64**, 524-538.

Fomel S. 2003. Seismic reflection data interpolation with differential offset and shot continuation. Geophysics, **68**, 733–744.

Gardner G. and Canning A. 1994. Effects of irregular sampling on 3D prestack migration. 64th SEG meeting, Los Angeles, California, Expanded Abstracts, 1553-1556.

Gulunay N. 2003. Seismic trace interpolation in the Fourier transform domain. Geophysics, **68**, 355–369.

Hanafy S. M. 2011. 3D Interferometric Extrapolation of SSP Marine Data. 73th EAGE meeting, Vienna, Austria, Expanded Abstracts, P375.

Hansen P.C. 1998. Rank-Deficent and Discrete ill-Posed Problems. Siam. ISBN 9780898714036.

Hennenfent G. and Herrmann F. J. 2005. Sparseness-constrained data continuation with frames: applications to missing traces and aliased signals in 2/3-D. 75th SEG meeting, Houston, Texas, Expanded Abstracts, 2162–2165.

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

Hindriks C.O.H., Duijndam A.J.W. and Schonewille A. 1997. Reconstruction of two-dimensional irregularly sampled wavefields. 67th SEG meeting, Dallas, Texas, Expanded Abstracts, 1163-1166.

Hindriks C.O.H. and Duijndam A.J.W. 2000. Reconstruction of 3-D seismic signals irregularly sampled along two spatial coordinates. Geophysics **65**, 253-263.

Hugonnet P. and Canadas G. 1997. Regridding of irregular data using 3D Radon decompositions. 67th SEG meeting, Dallas, Texas, Expanded Abstracts, 1111–1114.

Liu B. and Sacchi M. D. 2004. Minimum weighted norm interpolation of seismic records. Geophysics, **69**, 1560-1568.

Liu Y., Liu C., Wang D., Feng X. and Lu Q. 2012. Iterative seismic data interpolation beyond aliasing using seislet transform. 74th EAGE meeting, Copenhagen, Denmark, Expanded Abstracts, P042.

Lu L. 1985. Application of local slant stack to trace interpolation. 55th SEG meeting, Washington D.C, Expanded Abstracts, 235-238.

Kabir M. M. N. and Verschuur D. J. 1995. Restoration of missing offsets by parabolic radon transform. Geophysical Prospecting, **43**, 347–368.

Jin S. 2010. 5D seismic data regularization by a damped least-norm Fourier inversion. Geophysics, 75, WB103-WB111.

Kustowski B. 2010. Regularization of spatially aliased seismic wavefield using the Pyramid transform: a new insight. 80th SEG meeting, Denver, Colorado, Expanded Abstracts, 3574-3578.

Malcolm A. E., De Hoop M. V. and LeRousseau J. V. 2005. The applicability of dip moveout/azimuth moveout in the presence of caustics. Geophysics **70**, S1-S17.

Mazzucchelli P., Rocca F., Spagnolini U., Spitz S. 1998. Wavefield interpolation: continuation or prediction filter techniques?. 60th EAGE meeting, Leipzig, Germany, Expanded Abstracts, 2-51.

Mazzucchelli P. and Rocca F. 1999. Regularizing land acquisitions using shot continuation operators: effects on amplitudes. 69th SEG meeting, Houston, Texas, Expanded Abstracts, 1995-1998.

Naghizadeh M. and Innanen K. A. 2011. Seismic data interpolation using a fast generalized Fourier transform. Geophysics, **76**, V1-V10.

Özdemir A. K., Özbek A. and Vassallo M. 2008. Interpolation of irregularly sampled data by matching pursuit. 70th EAGE meeting, Rome, Italy, Expanded Abstracts, G025.

Özbek A., Özdemir A. K. and Vassallo M. 2009. Interpolation by matching pursuit. 79th SEG meeting, Houston, Texas, Expanded Abstracts, 3254-3275.

Peterson D. P. and Middleton D. 1962. Sampling and reconstruction of wave-number limited functions in n-dimensional Euclidean Spaces. Information and Control **5**, 279-323.

Routh P. S. and Miller C. R. 2006. Image interpretation using Appraisal analysis. 21st SAGEEP meeting, Philadelphia, Pennsylvania, Expanded Abstract, 1812-1820.

Sacchi M. D. and Ulrych T. 1996. Estimation of the discrete Fourier transform, a linear inversion approach. Geophysics **61**, 1128-1136.

Sacchi M. D. and Liu B. 2005. Minimum weighted norm wavefield reconstruction for AVA

imaging. Geophysical Prospecting, **53**, 787–801.

Scales J.A. and Snieder R. 2000. The Anatomy of Inverse Problem. Geophysics **65**, 6, 1708-1710.

Schonewille M. A. and Duijndam A. J. W. 1998. Efficient non-uniform Fourier and Radon filtering and reconstruction. 68th SEG meeting, New Orleans, Louisiana, Expanded Abstracts, 1692-1695.

Schonewille M. A. 2000. Fourier reconstruction of irregularly sampled seismic data: Ph.D. Dissertation. Delft University of Technology.

Schonewille M. A. 2003. A modeling study on seismic data regularization for time-lapse applications. 73rd SEG meeting, Dallas, Texas, Expanded Abstracts, 1537-1540.

Smith P. J., Scott I. and Traylen T. 2012. Simultaneous time-lapse binning and regularization of 4D data. 74th EAGE meeting, Copenhagen, Denmark, Expanded Abstracts, E009.

Spitz S. 1991. Seismic trace interpolation in the F-X domain. Geophysics **56**, 785-794.

Trad D., Ulrych T. J. and Sacchi M. D. 2002. Accurate interpolation with high-resolution time-variant Radon transforms. Geophysics **67**, 644-656.

Trad D. 2009. Five-dimensional interpolation: Recovering from acquisition constraints: Geophysics, 74, V123–132.

Wang Y. 2002. Seismic trace interpolation in the f-x-y domain. Geophysics **67**, 1232– 1239.

Wang Y., Luo Y. and Schuster G. T. 2009. Interferometric interpolation of missing seismic data. Geophysics **74**, SI37-SI45.

Xia J., Xu Y., Miller R. D. and Zeng C. 2010. A trade-off solution between model resolution and covariance in surface-wave inversion. Pure and Applied Geophysics **167**, 1537-1547.

Xu S., Zhang Y., Pham D., and Lambaré G. 2005. Antileakage Fourier transfrom for seismic

data regularization. Geophysics, **70**, V87-V95.

Xu S., Zhang Y. and Lambaré G. 2010. Antileakage Fourier transform for seismic data regularization in higher dimensions. Geophysics, **75**, WB113-WB12.

Yilmaz O. 2001. Seismic data analysis: processing, inversion, and interpretation of seismic data. Society of Exploration Geophysicists, Tulsa. ISBN: 1560800941, 9781560800941

Yu Z., Ferguson J., McMechan G. and Anno P. 2007. Wavelet-Radon domain de-aliasing and interpolation of seismic data. Geophysics **72**, V41-V49.

Zwartjes P. M. and Gisolf A. 2007. Fourier reconstruction with sparse inversion. Geophysical Prospecting **55**, 199-221.

FIGURE CAPTIONS

Figure 1. Least-squares Fourier reconstruction problem viewed as an Appraisal problem.

Figure 2. Schematic representation of the data and the model matrices before and after the lexicographical arrangement: (a) the original 3D data volume $\mathbf{D}^{\wedge}_{\mathrm{IRR}}$. The dots indicate the actual spatial positions of the traces already transformed in the temporal frequency domain (vertical axis of the volume); (b) lexicographical arrangement of the 3D data volume into a 2D matrix $\mathbf{D}_{\mathrm{IRR}}$. Each column of $\mathbf{D}_{\mathrm{IRR}}$, that is a vector $\mathbf{d}_{\mathrm{IRR}}$, corresponds to a frequency slice of $\mathbf{D}^{\wedge}_{\mathrm{IRR}}$. Each row of $\mathbf{D}_{\mathrm{IRR}}$ corresponds to a data trace of $\mathbf{D}^{\wedge}_{\mathrm{IRR}}$; (c) same graphical representation as in (a) for the 3D Fourier spectrum $\mathbf{M}^{\wedge}_{\mathrm{TRUE}}$. Note that the Fourier elements fall on a regular $k_x - k_y$ grid; (d) lexicographical arrangement of the 3D Fourier volume $\mathbf{M}^{\wedge}_{\mathrm{TRUE}}$ into a 2D matrix $\mathbf{M}_{\mathrm{TRUE}}$. Each column of $\mathbf{M}_{\mathrm{TRUE}}$, that is a vector $\mathbf{m}_{\mathrm{TRUE}}$, corresponds to a frequency slice of $\mathbf{M}^{\wedge}_{\mathrm{TRUE}}$. Each row of $\mathbf{M}_{\mathrm{TRUE}}$ corresponds to a temporal frequency spectrum in $\mathbf{M}^{\wedge}_{\mathrm{TRUE}}$.

Figure 3. (a) The four dipping planes in the in-line, cross-line and time dimensions used to create the synthetic trace gathers. (b) Sampling geometry: the red dots indicate the actual trace locations while the blue dots represent the desired (regular) grid (4 m by 4 m).

Figure 4. (a) Irregularly sampled synthetic data ($\mathbf{D}_{\mathrm{IRR}}$, in time) and (b) its Riemann Sum (NDFT) spectrum. (c) True, regularly sampled synthetic data ($\mathbf{D}_{\mathrm{TRUE}}$, in time) and (d) its Fourier spectrum ($\mathbf{M}_{\mathrm{TRUE}}$). (e) Noise term (**NOISE**, in time) computed from equation (4a) as discussed in the text.

Figure 5. (a) Estimated 3D Fourier spectra ($\mathbf{M}_{\mathrm{EST}}$) and (b) reconstructed data set ($\mathbf{D}_{\mathrm{REG}}$, in time). No threshold value was used in the inversion and thus the lowest singular values boost the noise. Only the horizontal event is properly reconstructed because it is fully band-limited and not affected by the spectral leakage. (c) Full MRM and its main diagonal and (d) full EMRM and its main diagonal. Note that the two matrices are identity matrices. (e) Projection of the noise ($\mathbf{AG}^{*}$ **NOISE** in time, see equation (15b)) into the reconstructed data.

Figure 6. Singular value analysis of the Gram matrix ($\mathbf{G}^{H}\mathbf{G}$). The red dots are the singular values of the whole $\mathbf{G}^{H}\mathbf{G}$ matrix (full spatial bandwidths). The green dots refer to the singular values after changing the forward model parametrization to limit the spatial bandwidths

considered in the inversion. The black horizontal dashed line indicates the singular value corresponding to the regular sampling $(\Delta k_x \Delta k_y)^2$ that we take as the reference value. The blue dashed lines correspond to the singular values 10 and 100 times less than the reference one.

Figure 7. Same as in Fig. 5 but a threshold value was used in the Gram matrix inversion. (a) Estimated 3D Fourier spectra ($\mathbf{M}_{\mathbf{EST}}$) and (b) reconstructed data set ($\mathbf{D}_{\mathbf{REG.}}$ in time). The threshold value of $0.1(\Delta k_x \Delta k_y)^2$ we used stabilizes the inversion but unbalanced wavelet amplitudes and wavelet misalignments appear. (c) Full MRM and its main diagonal and (d) full EMRM and its main diagonal. Note that the two matrices are not any more identity matrices and non null off-diagonal elements do appear. (e) Projection of the noise ($\mathbf{AG}^{-g}$ **NOISE** in time, see equation (15b)) into the reconstructed data. Note that the $\mathbf{AG}^{-g}$ operator now attenuates the noise.

Figure 8. Same as in Fig. 5 and 7 but the forward operator was changed reducing the in-line and cross-line spatial bandwidths as discussed in the text. (a) Estimated 3D Fourier spectra ($\mathbf{M}_{\mathbf{EST}}$) in the considered spatial bandwidths. (b) Reconstructed data set ($\mathbf{D}_{\mathbf{REG.}}$ in time). The events are properly reconstructed with the exception of the highly dipping shallow event. (c) Full MRM and its main diagonal and (d) full EMRM and its main diagonal (red line) . Note that considering the Fourier coefficients for limited spatial bandwidths makes the MRM an identity matrix. The EMRM shows super- and sub-diagonal elements related to the sinc interpolation needed to bring the data to a finer grid. The red curve depicts the main diagonal elements. (e) Projection of the noise ($\mathbf{AG}^{-g}$ **NOISE** in time, see equation (15b)) into the reconstructed data. Considering a limited portion of the spatial bandwidths, the noise projection into the data is effectively attenuated.

Figure 9. Representation of the **NOISE** component (in time) for the case illustrated in Fig. (8). It can be determined from equation (4a) given the knowledge of $\mathbf{M}_{\mathbf{TRUE}}$. Note that the highly dipping events, discarded by the reconstruction because they lie outside the considered spatial bandwidths, fall onto this term.

Figure 10. Shot geometry. Red dots: actual receiver locations. Blue dots: desired (regular) grid of receivers (receiver interval 12.5 m, line interval 120 m). The two rectangular windows of dimensions X= 1237.5 m and Y= 3480 m are used in the discussion to illustrate the FRMN reconstruction.

Figure 11. (a) Estimated 3D Fourier spectra within the green window of Fig. 10. (b) Close-up of the reconstructed common shot. (c) Full MRM and its main diagonal. (d) Full EMRM and its main diagonal. No threshold value was used in the inversion. Similar considerations to those for the synthetic example in Fig. 5 can be made.

Figure 12. Singular value analysis of Gram matrix ($\mathbf{G}^H\mathbf{G}$) for the green window of Fig. 10. The red dots are the singular values obtained without limiting the spatial bandwidths in the forward model parametrization. The green dots refer to the singular values after the forward model re-parametrization discussed in the text. The black horizontal dashed line indicates the singular value corresponding to the regular sampling $(\Delta k_x \Delta k_y)^2$ that we take as reference. The blue lines correspond to singular values 10 and 100 times less than the reference one.

Figure 13. Same as in Fig. 11 but a threshold value of $0.1(\Delta k_x \Delta k_y)^2$ was used in the inversion. The most evident artefacts generated by the SVD threshold are the whitening of the reconstructed data of the missing lines 19 and 20. Similar considerations to those for the synthetic example in Fig. 7 can be made.

Figure 14. EMRM (top row) and close-up (bottom row) of the reconstructed common shot within the green window in Fig. 10. (a) A wrong k-sampling interval ($\Delta k_y = 1/(Y\text{-}3\Delta y)$) in the cross-line direction causes sinc-like lobes in the EMRM (top) and severe wrap-around effects seen as event duplications (indicated by the small arrows) in the reconstructed data (bottom). (b) Δk_y is selected as in the standard DFT ($\Delta k_y = 1/(Y+\Delta y)$: serious edge effects are again introduced. (c) An optimal Δk_y value ($\Delta k_y = 1/(Y+3\Delta y)$ is used: no more artefacts are generated.

Figure 15. Same as in Fig. 11 but the forward operator was changed (the cross-line spatial bandwidth was reduced) as discussed in the text. Similar considerations to those for the synthetic example in Fig. 8 can be made.

Figure 16. In-line 1: (a) original (irregularly sampled) data; (b) reconstructed data with the application of a threshold in the Gram matrix inversion; (c) reconstructed data with no imposed threshold but with appropriately chosen spatial bandwidths and number of k-samples (Mx and My) of the forward model operator; (d) reconstructed data using an "ad-hoc" parametrization for the short source-receiver offsets included in the yellow rectangle, as discussed in the text.

Figure 17. In-line 20: (a) original (irregularly sampled data); (b) reconstructed data with the application of a threshold in the Gram matrix inversion; (c) reconstructed data with no imposed threshold but with appropriately chosen spatial bandwidths and number of k-samples (Mx and My) of the forward model operator; (d) reconstructed data using an "ad-hoc" parametrization for the short source-receiver offsets included in the yellow rectangle, as discussed in the text.

Figure 1

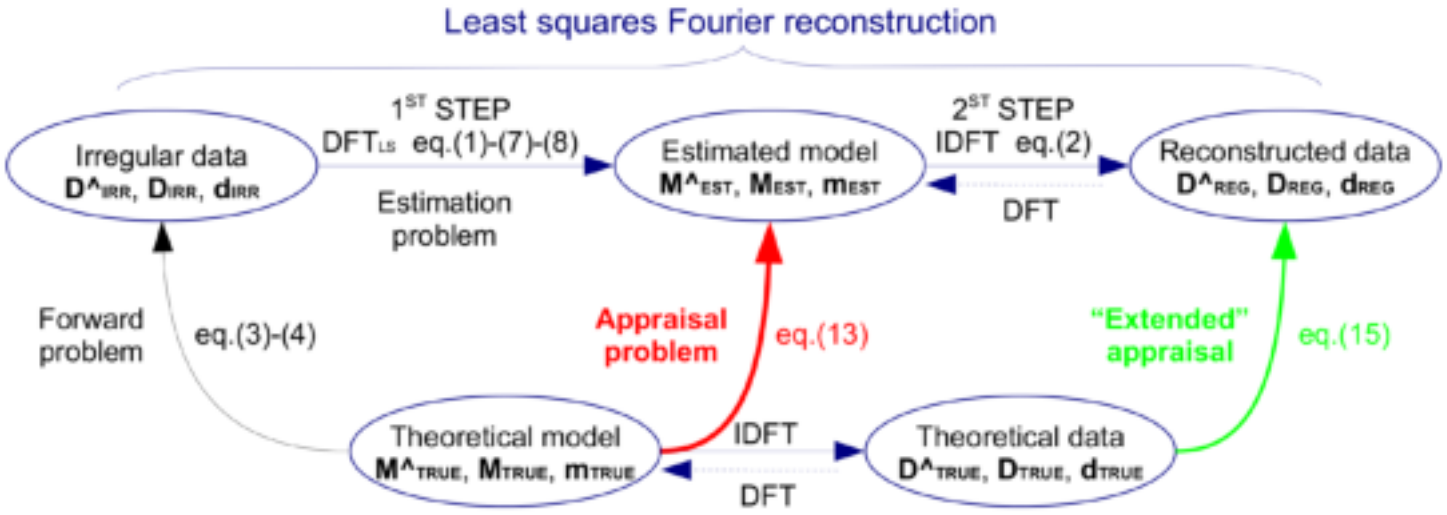

Least squares Fourier reconstruction
1ST STEP
DFTLS eq.(1)-(7)-(8)
Estimation problem
2ST STEP
IDFT eq.(2)
DFT
Irregular data
D^IRR, DIRR, dIRR
Estimated model
M^EST, MEST, mEST
Reconstructed data
D^REG, DREG, dREG
Forward problem
eq.(3)-(4)
Appraisal problem
eq.(13)
"Extended" appraisal
eq.(15)
Theoretical model
M^TRUE, MTRUE, mTRUE
IDFT
DFT
Theoretical data
D^TRUE, DTRUE, dTRUE

Figure 2

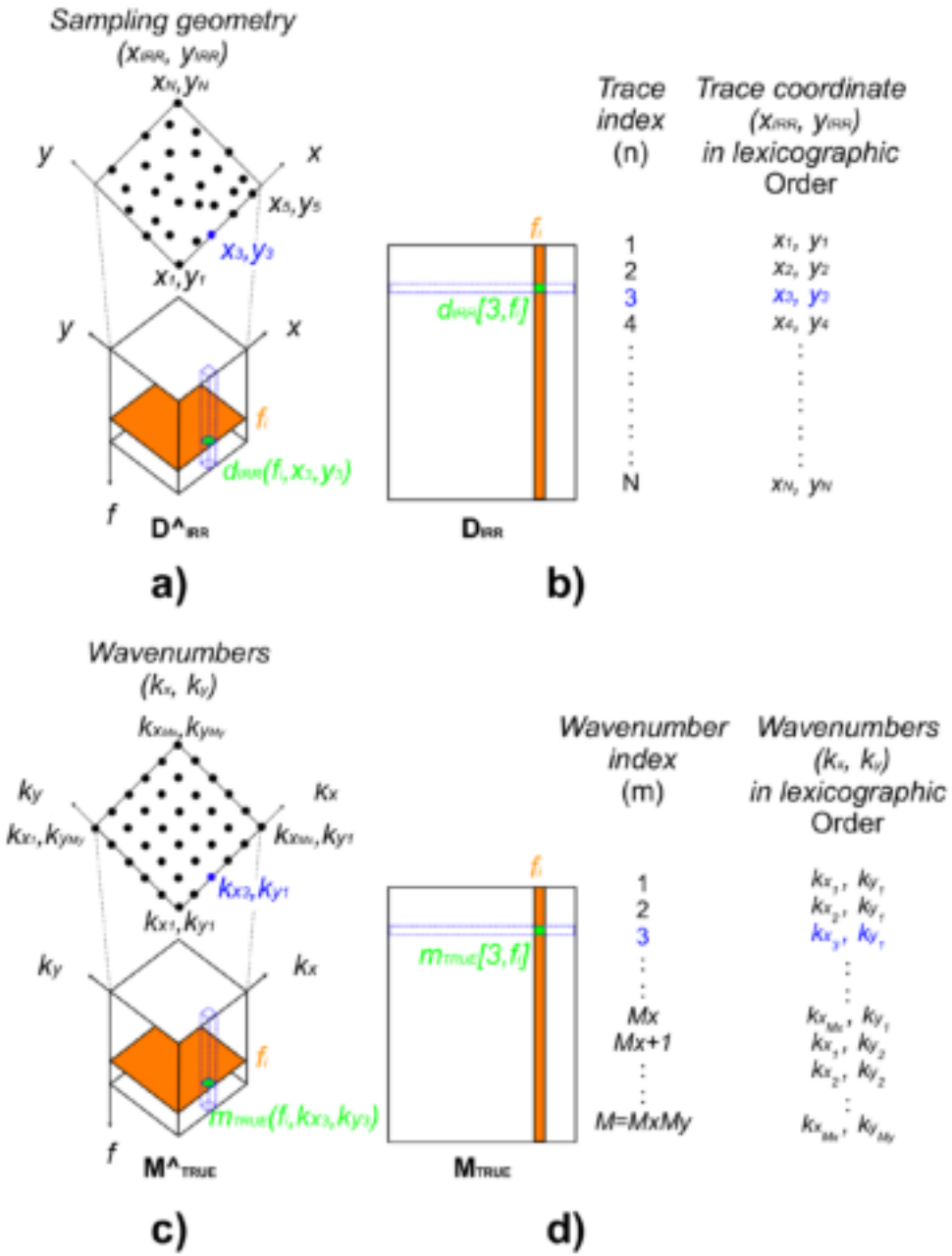

Sampling geometry
(xIRR, yIRR)
xN, yN
y
x
x5, y5
x3, y3
x1, y1
y
x
fi
dIRR(fi, x3, y3)
f
D^IRR
a)
fi
dIRR[3,fi]
DIRR
b)
Trace index (n)
Trace coordinate (xIRR, yIRR) in lexicographic Order
1
2
3
4
N
x1, y1
x2, y2
x3, y3
x4, y4
xN, yN
Wavenumbers
(kx, ky)
kxMx, kyMy
ky
kx
kx1, kyMy
kxMx, ky1
kx3, ky1
kx1, ky1
ky
kx
fi
mTRUE(fi, kx3, ky3)
f
M^TRUE
c)
fi
mTRUE[3,fi]
MTRUE
d)
Wavenumber index (m)
Wavenumbers (kx, ky) in lexicographic Order
1
2
3
Mx
Mx+1
M=MxMy

Figure 3

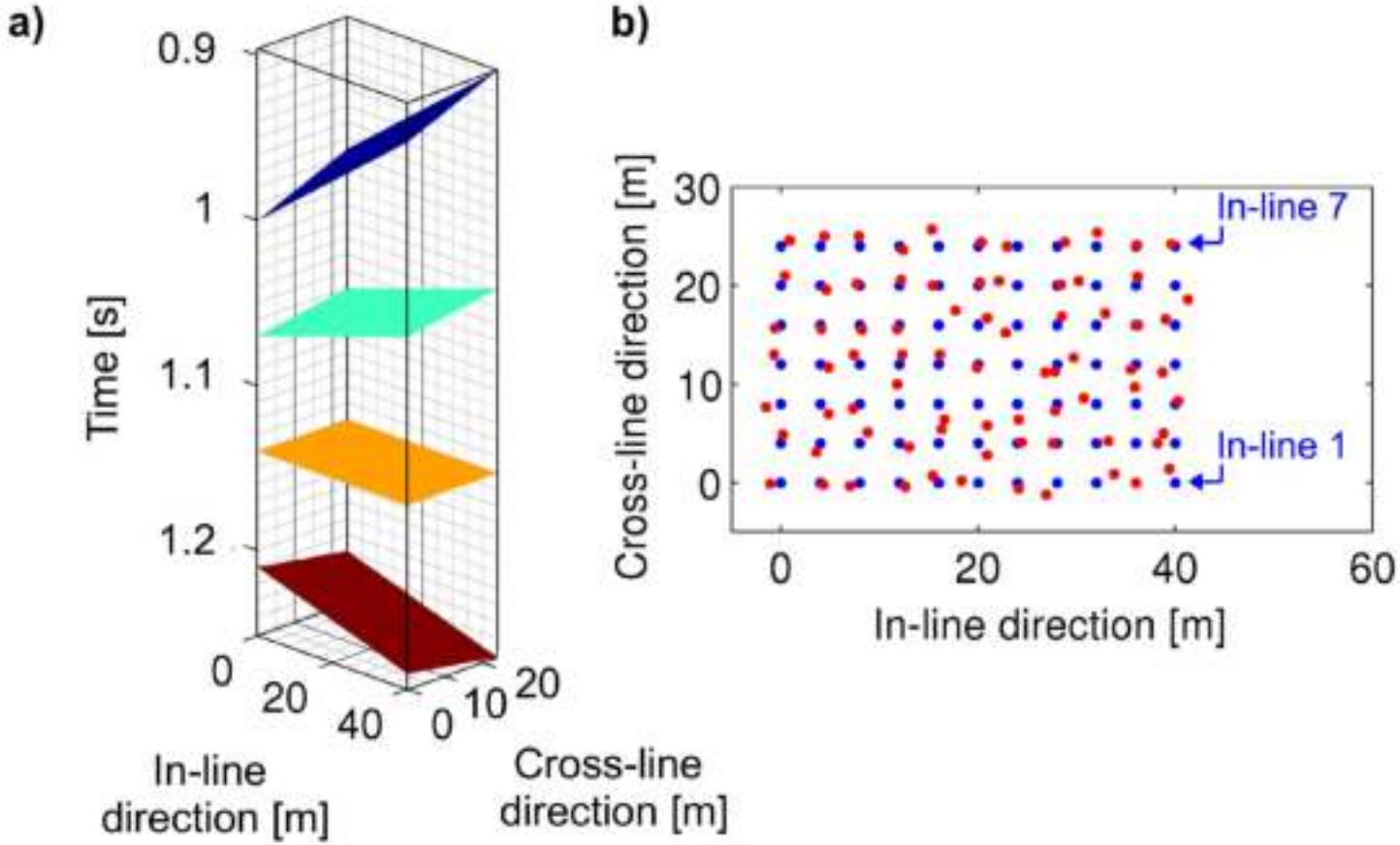

a)
0.9
1
1.1
1.2
Time [s]
0
20
40
In-line direction [m]
0
10
20
Cross-line direction [m]
b)
30
20
10
0
Cross-line direction [m]
0
20
40
60
In-line direction [m]
In-line 7
In-line 1

Figure 4

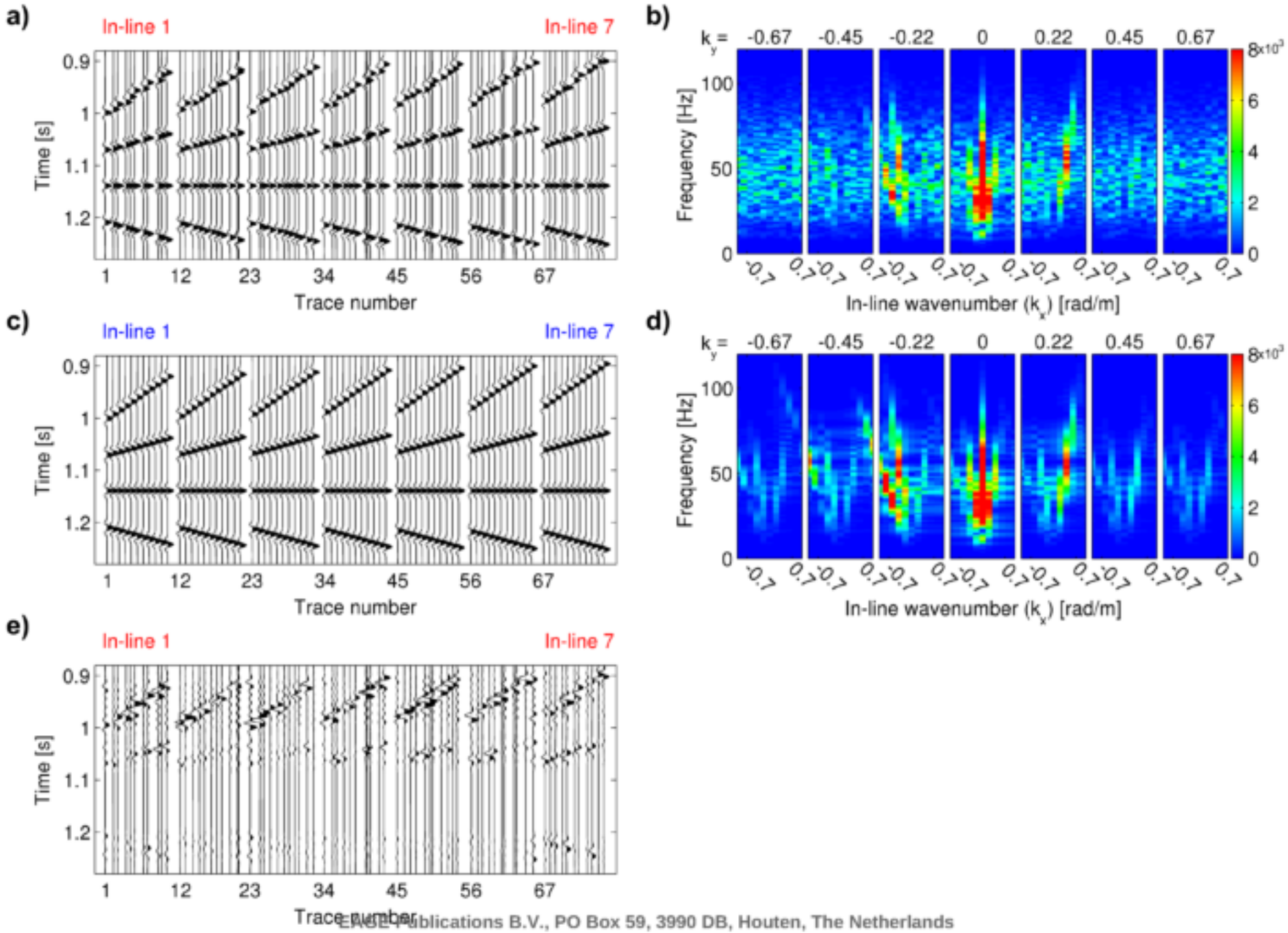

a)
In-line 1
In-line 7
Time [s]
Trace number
b)
k_y = -0.67 -0.45 -0.22 0 0.22 0.45 0.67
Frequency [Hz]
In-line wavenumber (k_x) [rad/m]
8x10^3
c)
In-line 1
In-line 7
Time [s]
Trace number
d)
k_y = -0.67 -0.45 -0.22 0 0.22 0.45 0.67
Frequency [Hz]
In-line wavenumber (k_x) [rad/m]
8x10^3
e)
In-line 1
In-line 7
Time [s]
Trace number

Figure 5

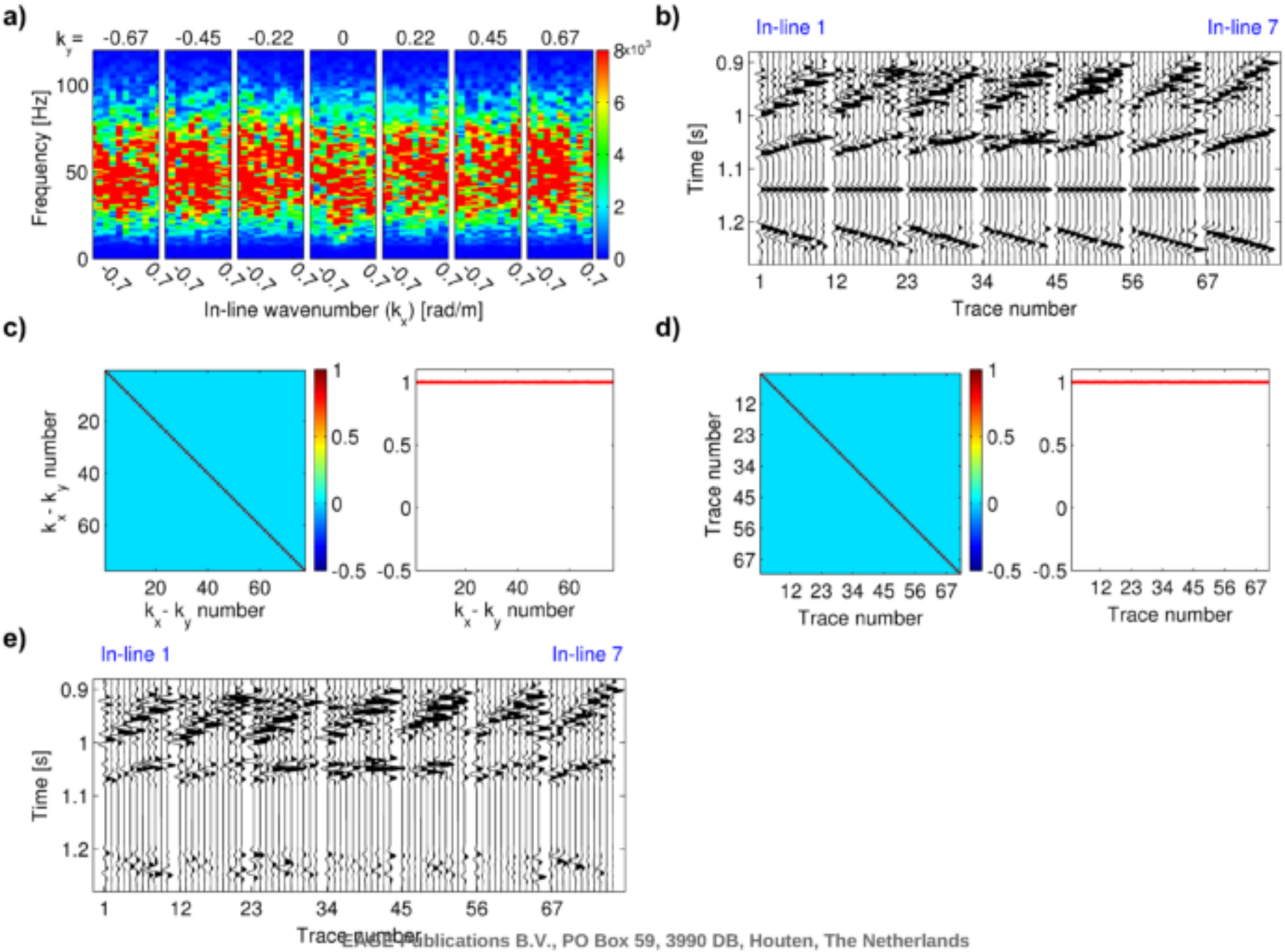

a)
k_y = -0.67 -0.45 -0.22 0 0.22 0.45 0.67
Frequency [Hz]
In-line wavenumber (k_x) [rad/m]
b)
In-line 1
In-line 7
Time [s]
Trace number
c)
k_x - k_y number
d)
Trace number
e)
In-line 1
In-line 7
Time [s]
Trace number

Figure 6

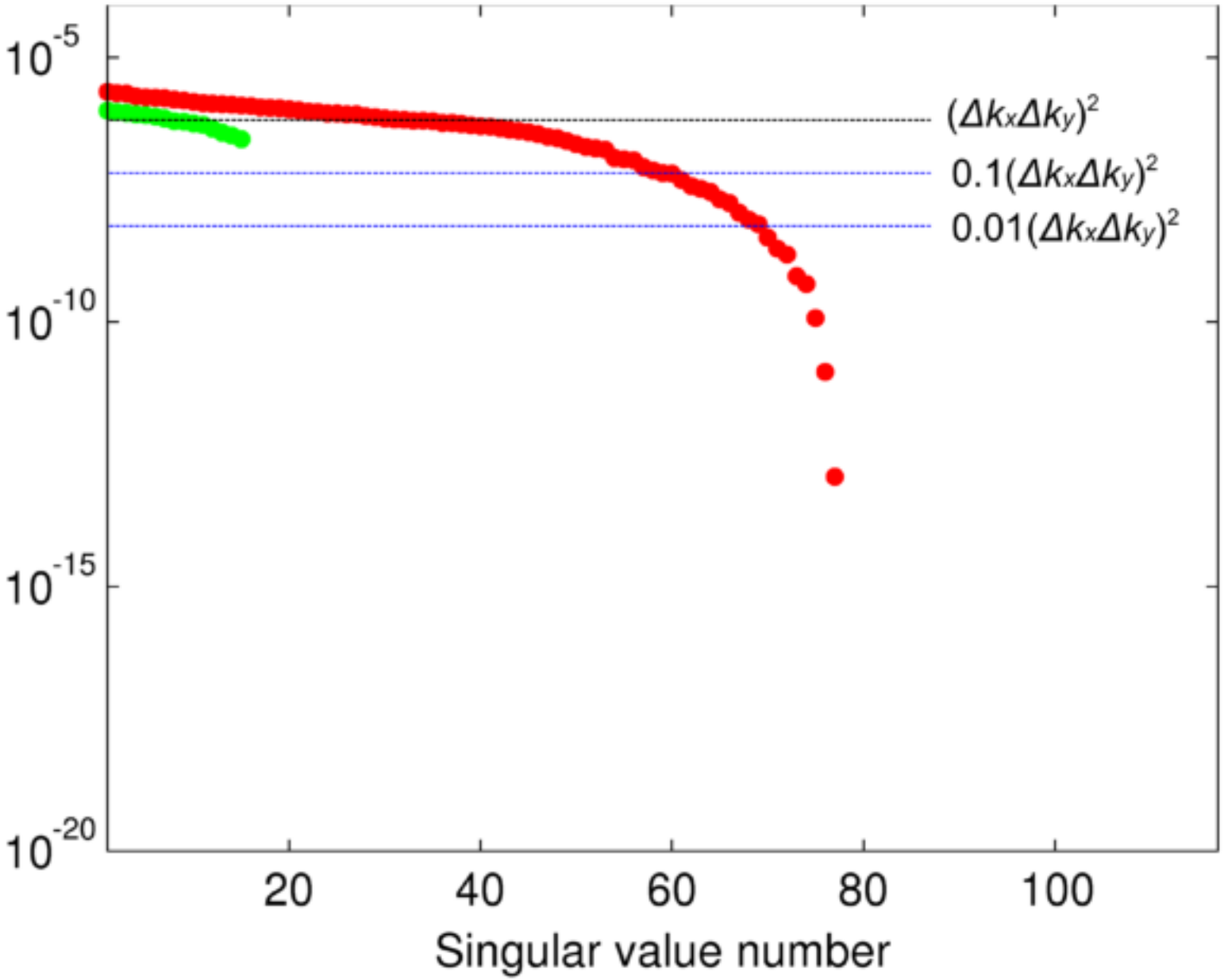

10^{-5}
10^{-10}
10^{-15}
10^{-20}
$(\Delta k_x \Delta k_y)^2$
$0.1(\Delta k_x \Delta k_y)^2$
$0.01(\Delta k_x \Delta k_y)^2$
20
40
60
80
100
Singular value number

Figure 7

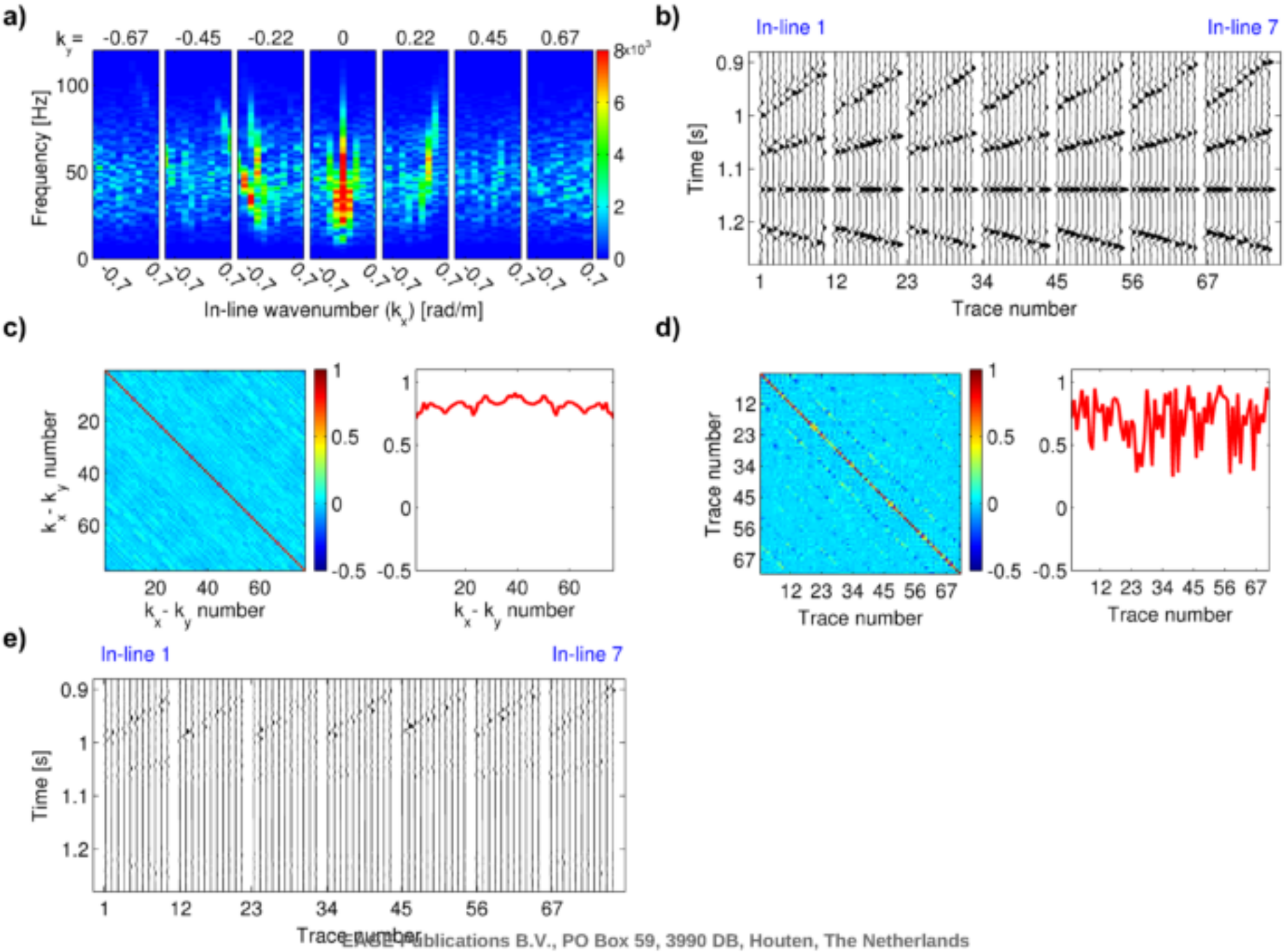

a)
k_y = -0.67 -0.45 -0.22 0 0.22 0.45 0.67
Frequency [Hz]
In-line wavenumber (k_x) [rad/m]
b)
In-line 1
In-line 7
Time [s]
Trace number
c)
k_x - k_y number
d)
Trace number
e)
In-line 1
In-line 7
Time [s]
Trace number

Figure 8

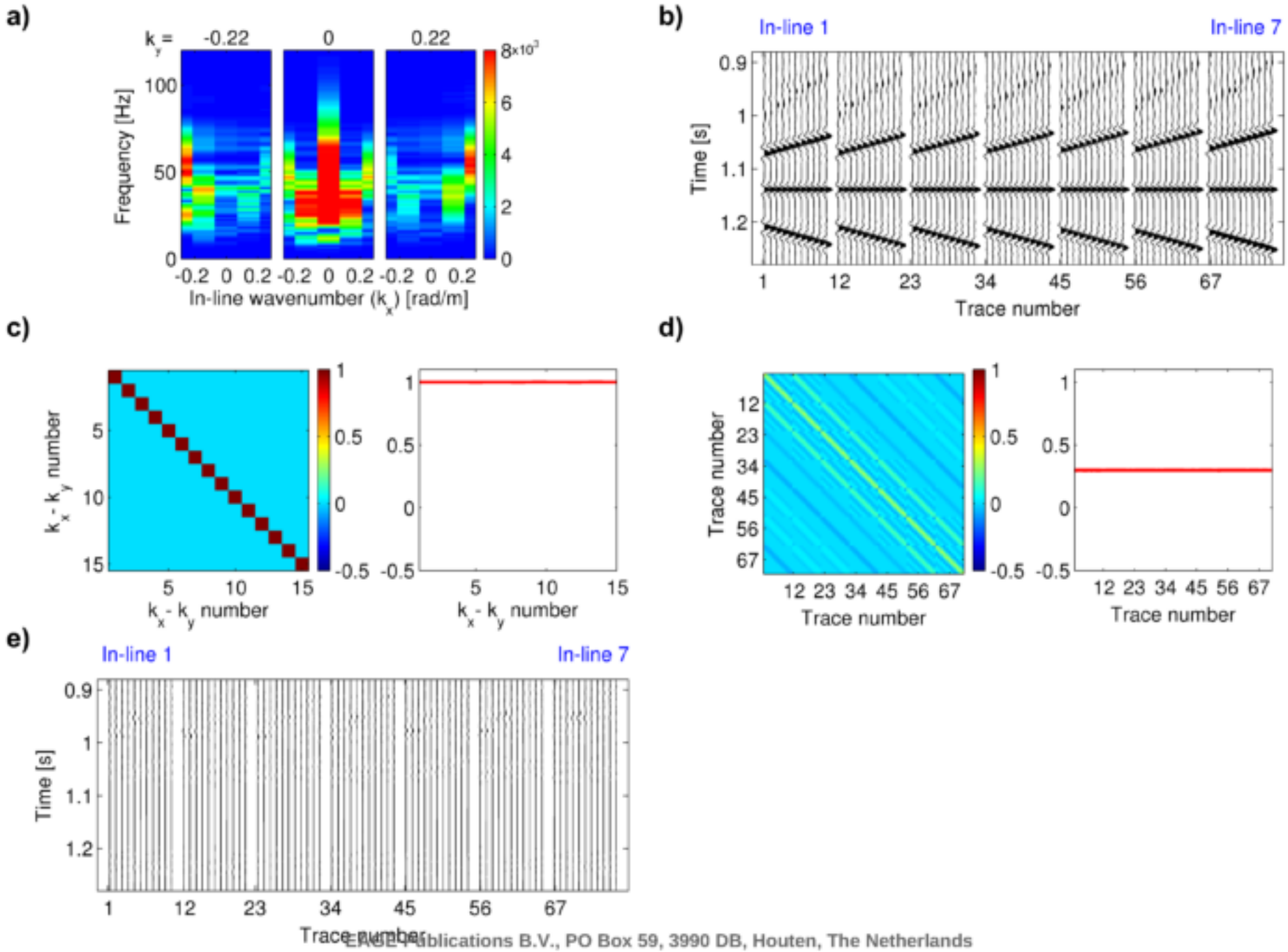

a)
k_y = -0.22 0 0.22
Frequency [Hz]
In-line wavenumber (k_x) [rad/m]
b)
In-line 1
In-line 7
Time [s]
Trace number
c)
k_x - k_y number
d)
Trace number
e)
In-line 1
In-line 7
Time [s]
Trace number

Figure 9

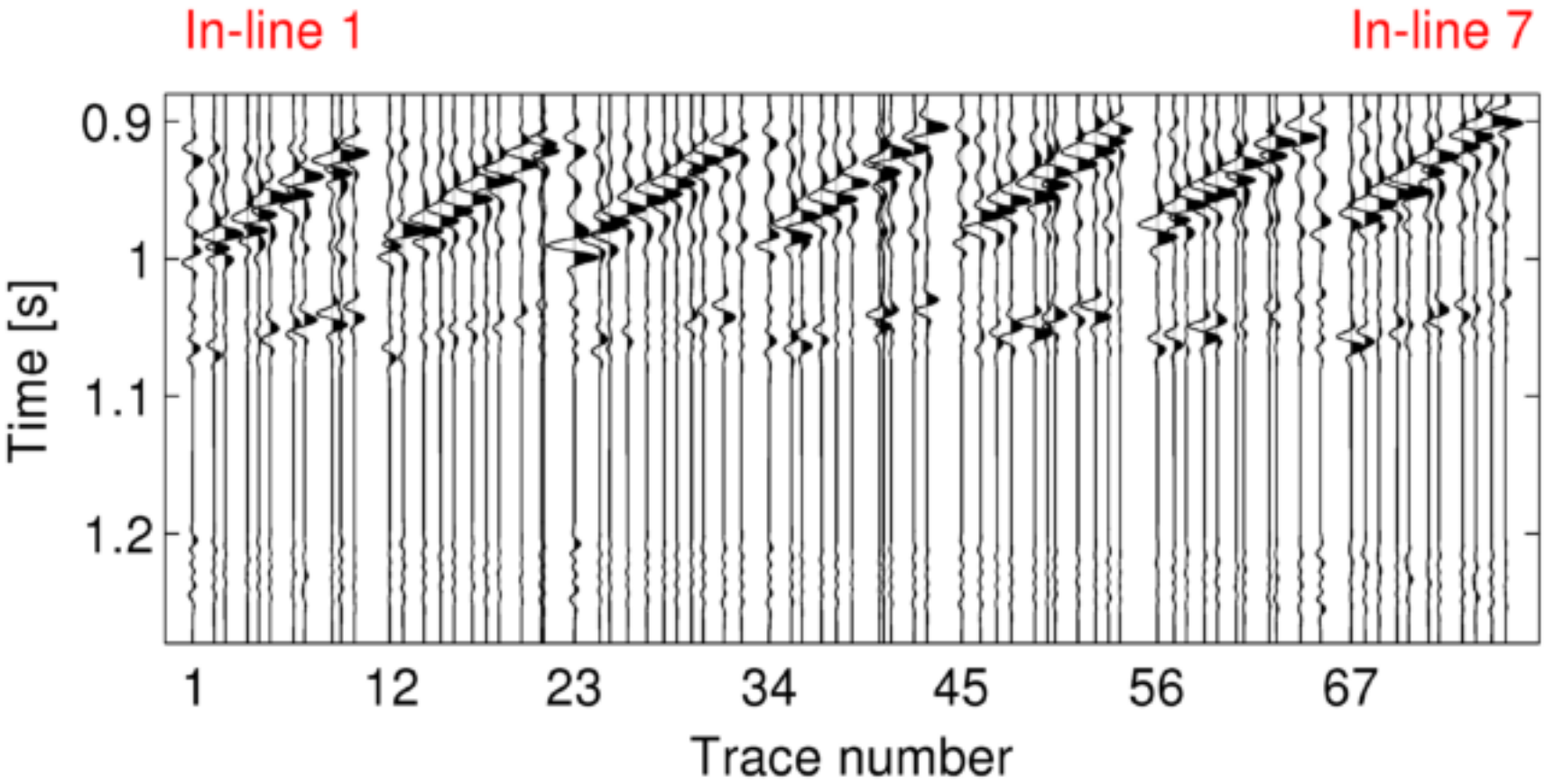

In-line 1
In-line 7
0.9
1
1.1
1.2
Time [s]
1
12
23
34
45
56
67
Trace number

Figure 10

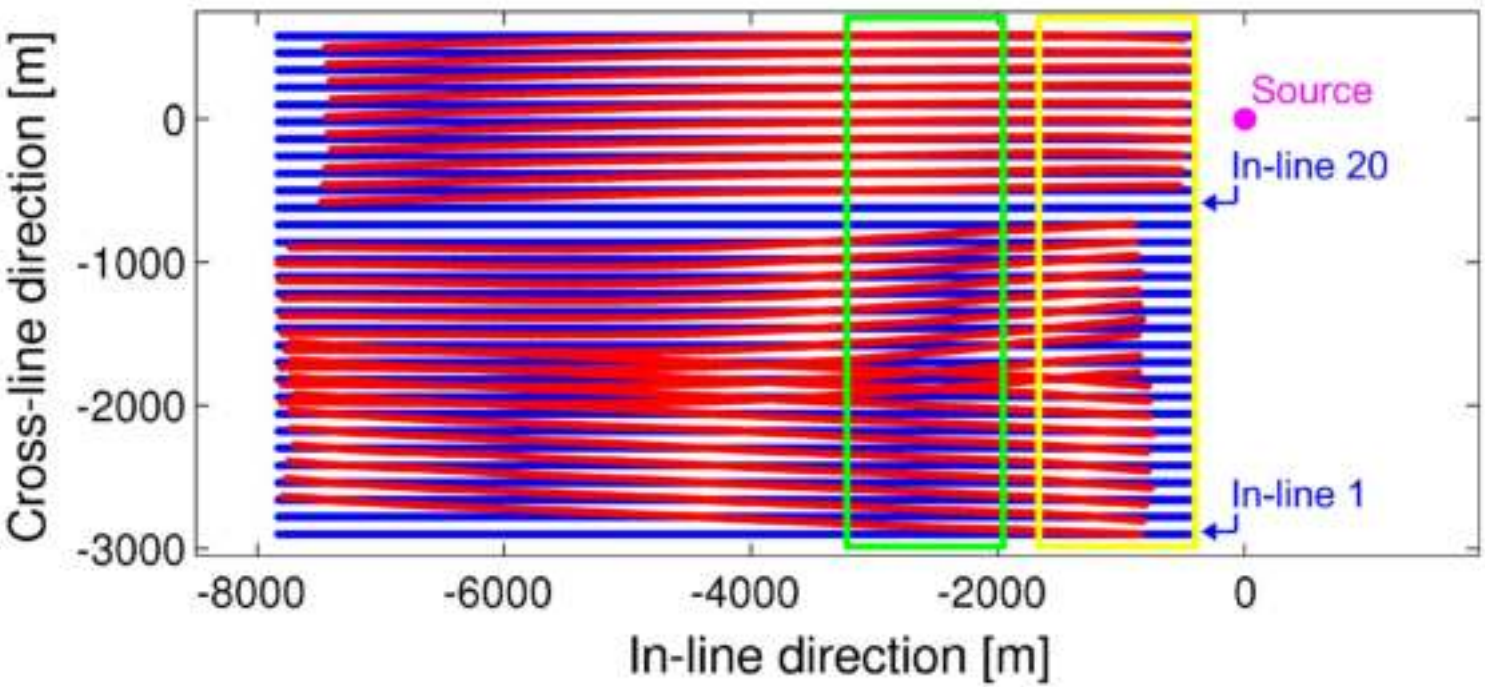

Source
In-line 20
In-line 1
0
-1000
-2000
-3000
Cross-line direction [m]
-8000
-6000
-4000
-2000
0
In-line direction [m]

Figure 11

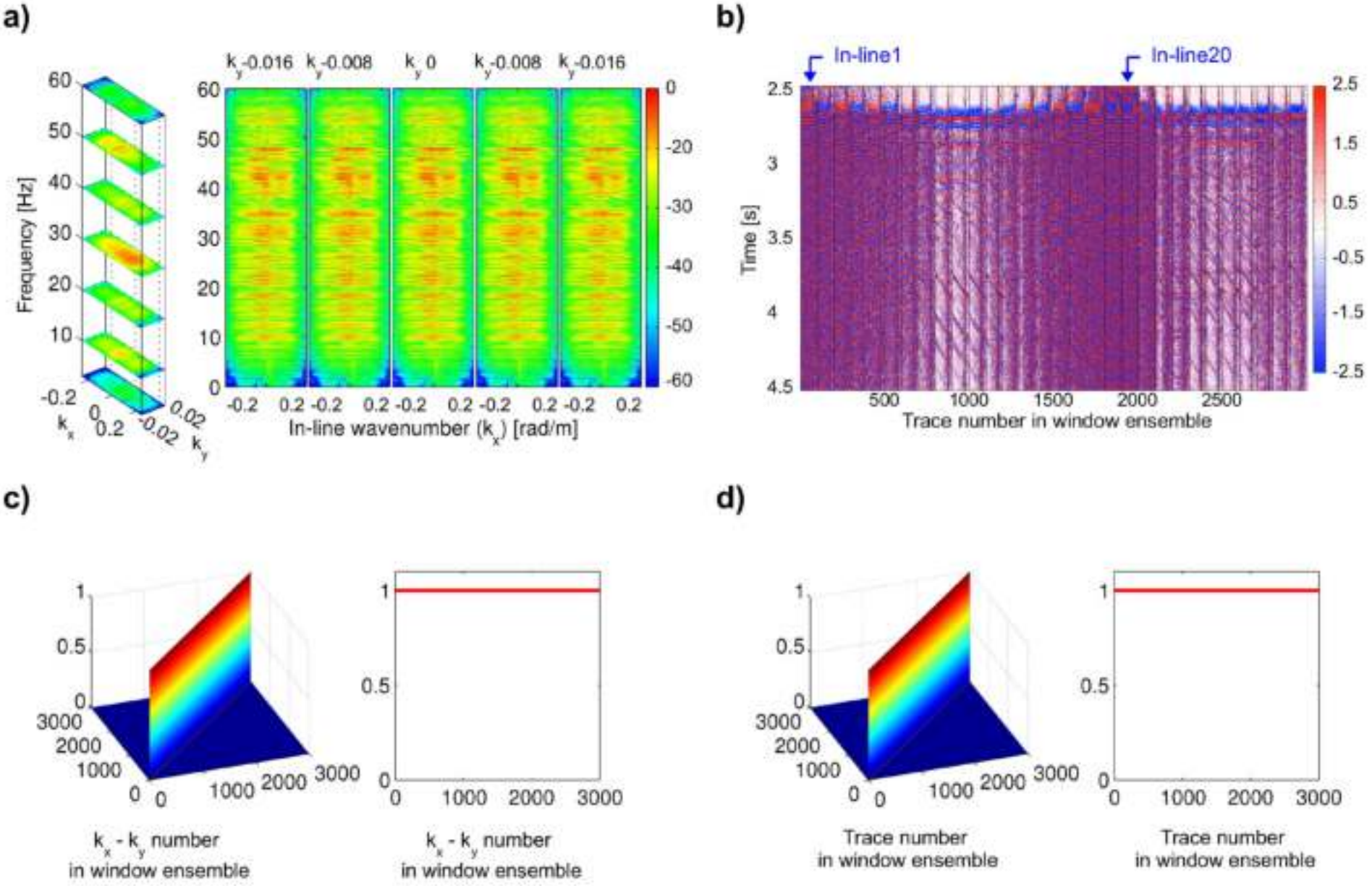

a)
Frequency [Hz]
k_x
k_y
k_y-0.016 k_y-0.008 k_y 0 k_y-0.008 k_y-0.016
In-line wavenumber (k_x) [rad/m]
b)
In-line1
In-line20
Time [s]
Trace number in window ensemble
c)
k_x - k_y number
in window ensemble
k_x - k_y number
in window ensemble
d)
Trace number
in window ensemble
Trace number
in window ensemble

Figure 12

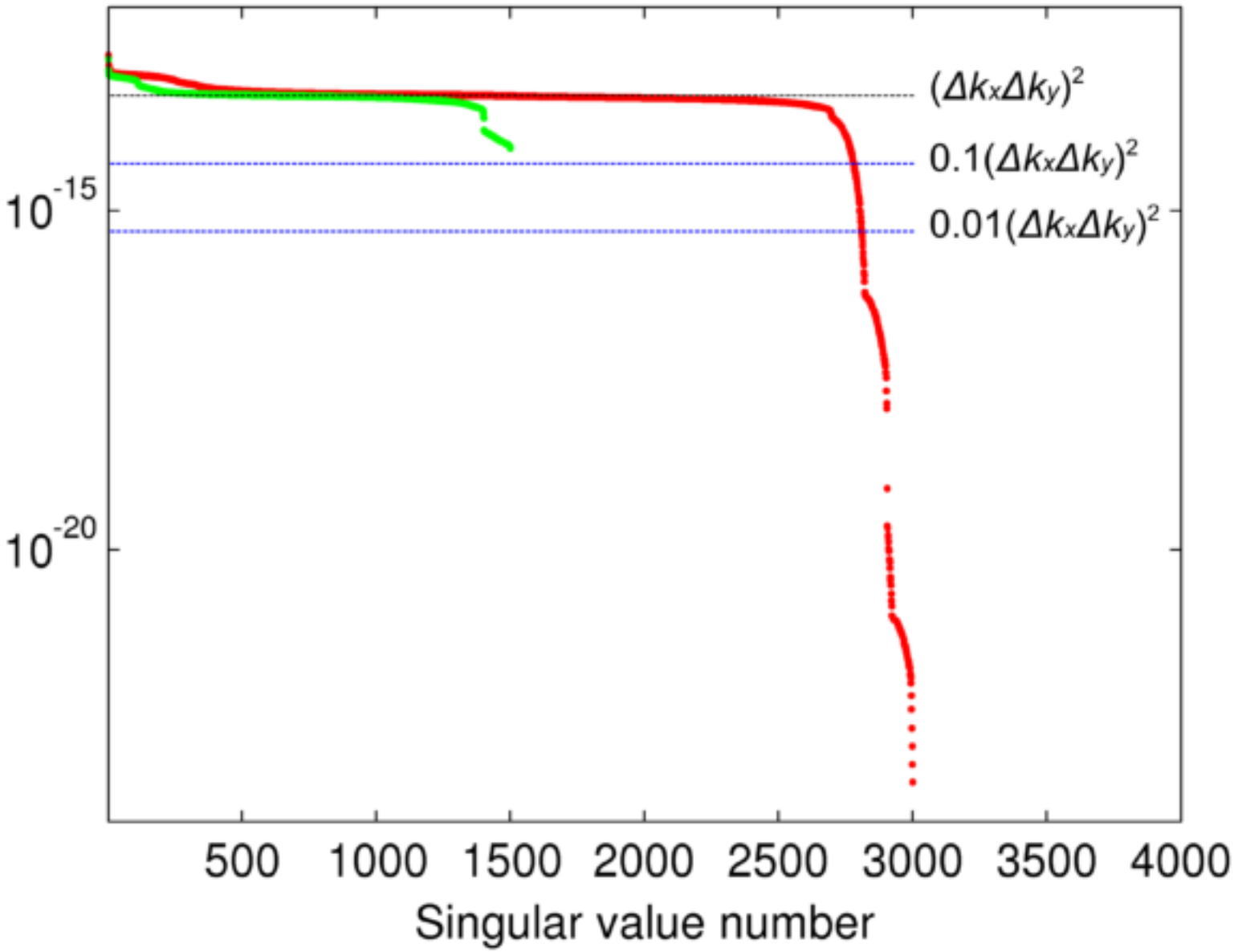

$(\Delta k_x \Delta k_y)^2$
$0.1(\Delta k_x \Delta k_y)^2$
$0.01(\Delta k_x \Delta k_y)^2$
10^{-15}
10^{-20}
500
1000
1500
2000
2500
3000
3500
4000
Singular value number

Figure 13

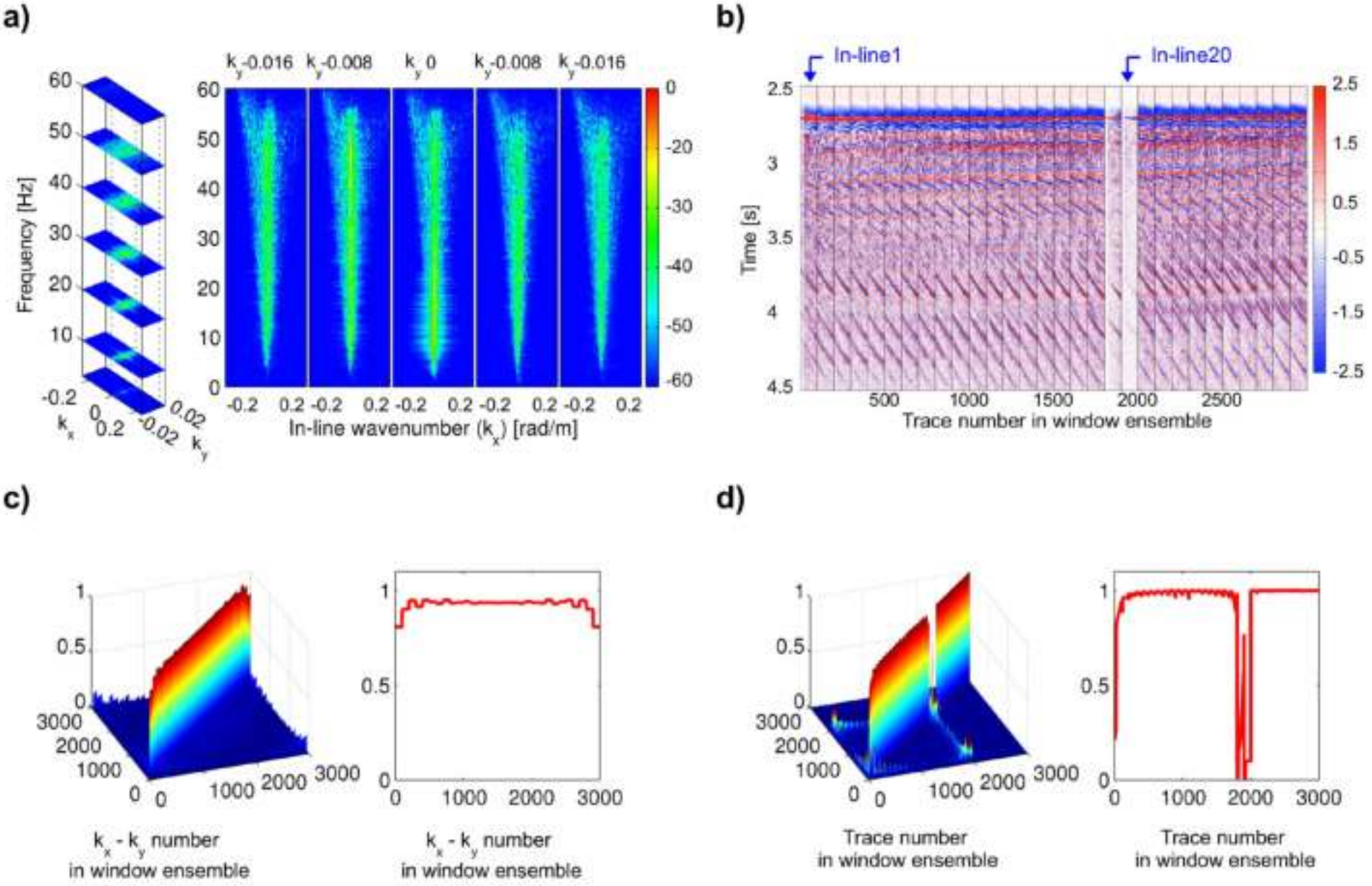

a)
Frequency [Hz]
k_x
k_y
k_y-0.016 k_y-0.008 k_y 0 k_y-0.008 k_y-0.016
In-line wavenumber (k_x) [rad/m]
b)
In-line1
In-line20
Time [s]
Trace number in window ensemble
c)
k_x - k_y number in window ensemble
k_x - k_y number in window ensemble
d)
Trace number in window ensemble
Trace number in window ensemble

Figure 14

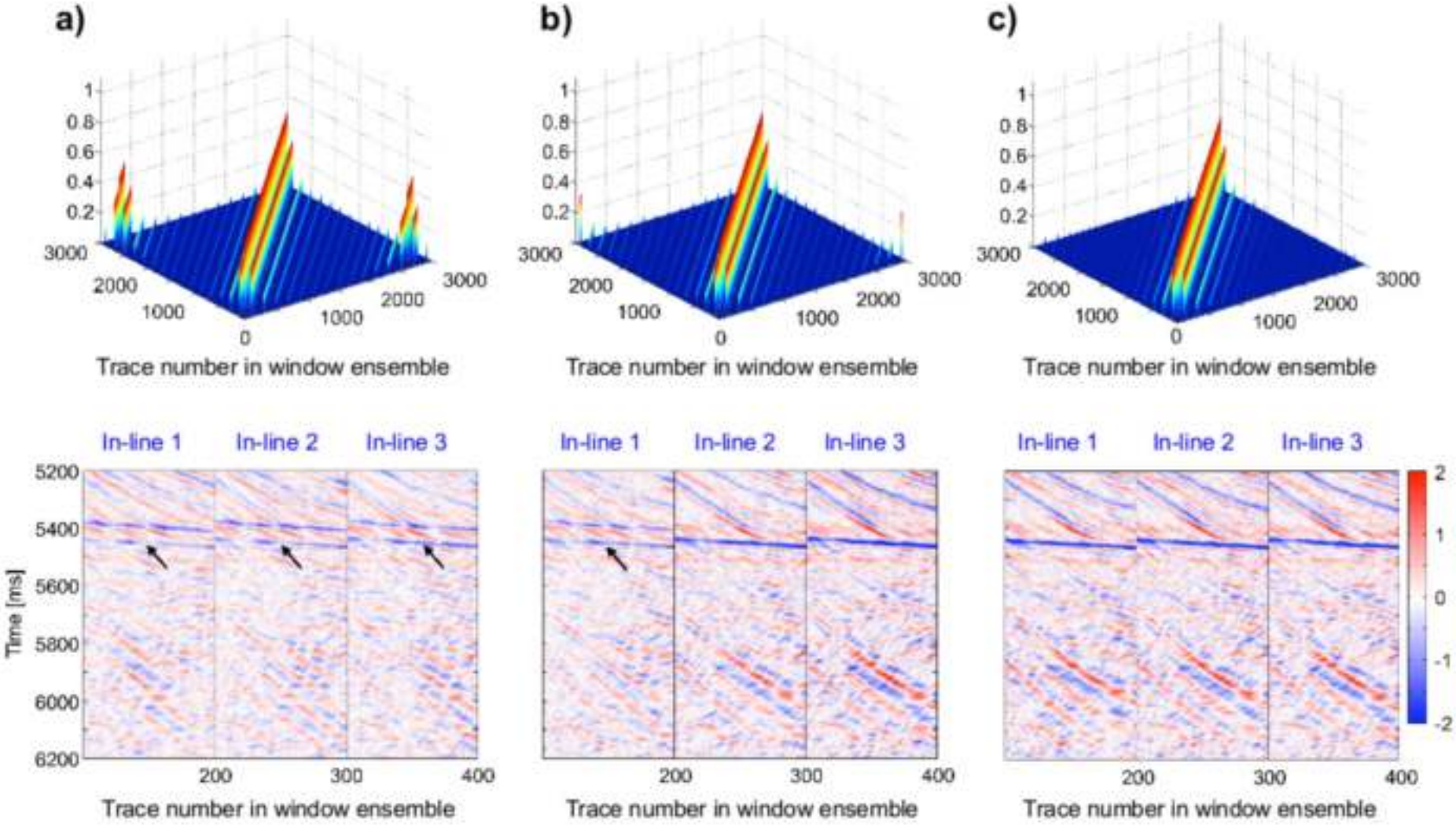

a)
b)
c)
Trace number in window ensemble
In-line 1
In-line 2
In-line 3
Time [ms]
5200
5400
5600
5800
6000
6200
200
300
400
2
1
0
-1
-2

Figure 15

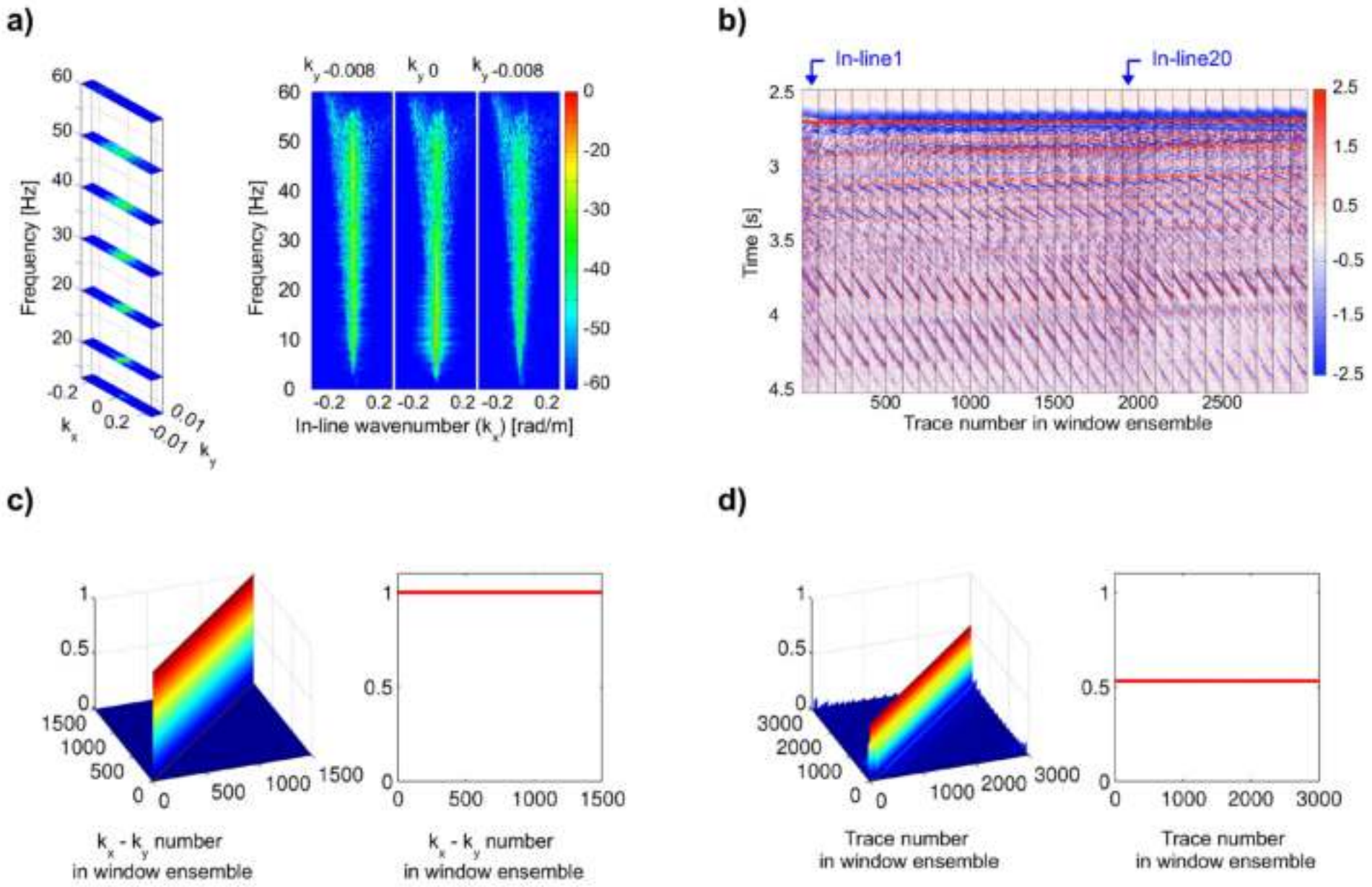

a)
Frequency [Hz]
k_x
k_y
k_y -0.008
k_y 0
k_y -0.008
In-line wavenumber (k_x) [rad/m]
b)
In-line1
In-line20
Time [s]
Trace number in window ensemble
c)
k_x - k_y number
in window ensemble
d)
Trace number
in window ensemble

Figure 16

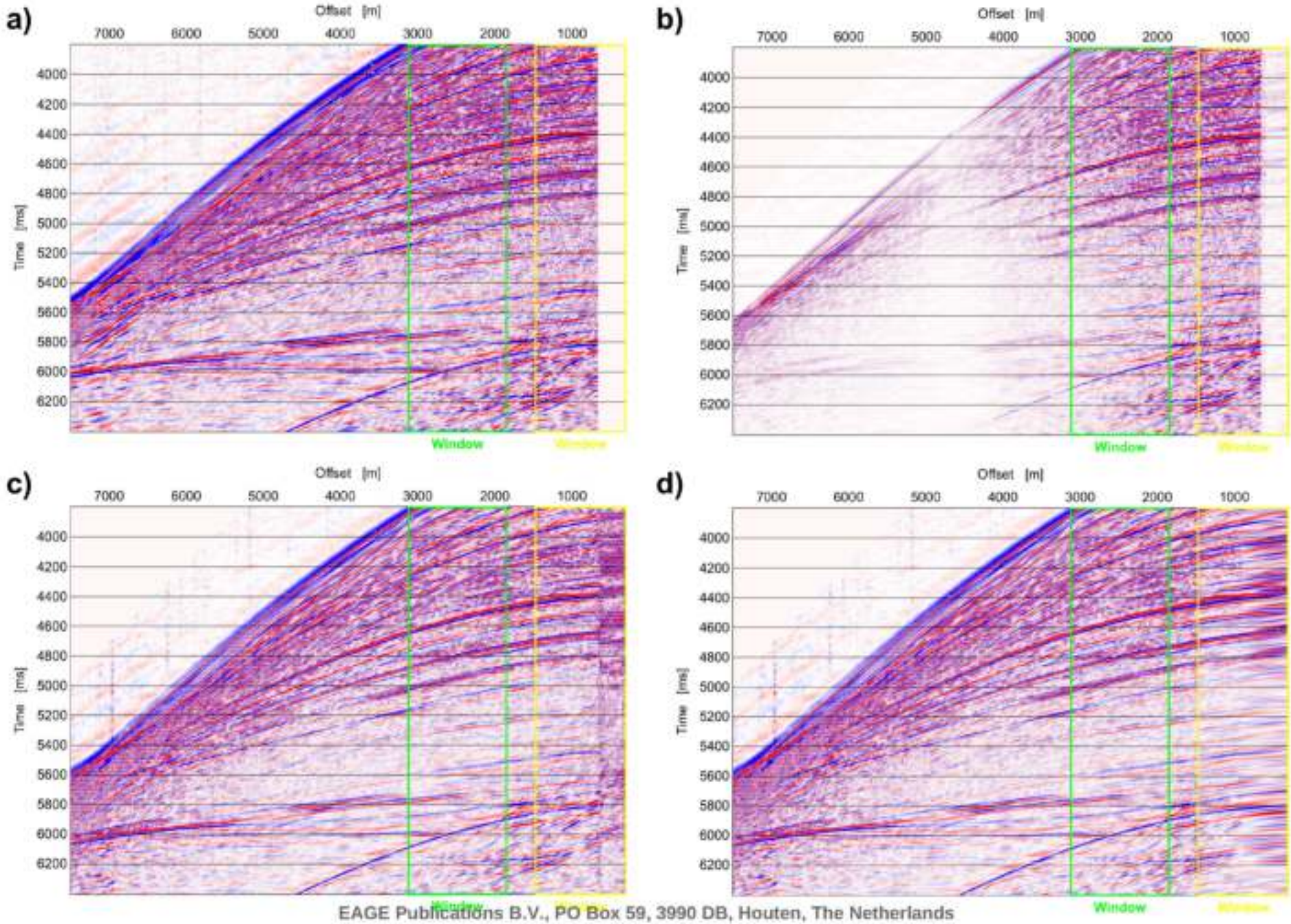
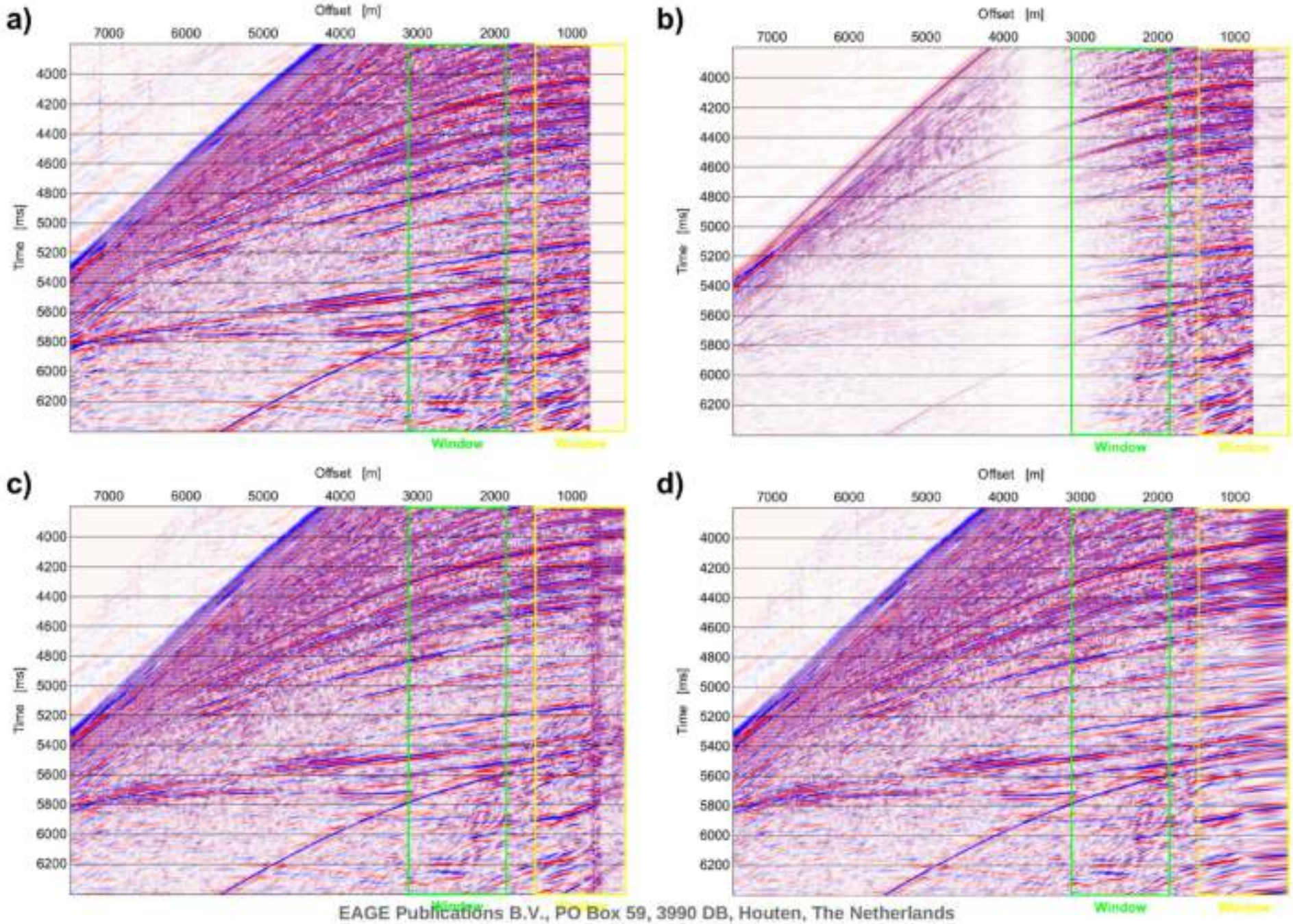

a)
b)
c)
d)
Offset [m]
7000
6000
5000
4000
3000
2000
1000
Time [ms]
4000
4200
4400
4600
4800
5000
5200
5400
5600
5800
6000
6200
Window

Figure 17


a)
b)
c)
d)
Offset [m]
7000
6000
5000
4000
3000
2000
1000
Time [ms]
4000
4200
4400
4600
4800
5000
5200
5400
5600
5800
6000
6200
Window