Computational Geosciences (2019) 23:665–682
https://doi.org/10.1007/s10596-019-9812-6

Geostatistical interpolation of non-stationary seismic data

Francesco Turco[1,2] · **Leonardo Azevedo**[1] · **Dan Herold**[3]

Received: 25 May 2018 / Accepted: 12 February 2019 / Published online: 18 March 2019

Abstract

The problem of sparsely collected seismic data is one of the main issues in reflection seismology, because most advanced data processing techniques require a dense and regular seismic data grid. We present a geostatistical seismic data interpolation technique based on sequential stochastic simulations with local structural anisotropies. This technique, contrary to conventional existing data-driven seismic interpolation approaches based on sparsity, prediction filters, or rank-reduction, predicts the value of seismic amplitudes at non-sampled locations by exploiting the statistics of the recorded amplitudes, which are used as experimental data for the geostatistical interpolation in the original data domain. Local mean and variance are computed on-the-fly to define intervals of the global conditional distribution function, from where amplitude values are stochastically simulated. The parameters to define subsets of experimental data from which mean and variance are calculated are given by local variogram models, which in turn are obtained from a local dip and azimuth estimation in the t-x-y domain. The geostatistical seismic data interpolation technique is applied to synthetic and real 2D and 3D datasets in both post- and pre-stack domains. Besides being computationally cheaper than other methods, because the interpolation is carried out directly in the original data domain, the proposed technique provides a local quantitative analysis of the reliability of the interpolated seismic samples, which can be exploited in following processing steps.

Keywords Seismic interpolation · Stochastic simulation · Uncertainty

1 Introduction

Seismic reflection data are often irregularly collected due to economic constraint and/or physical obstacles during data acquisition. The problem of sparsely collected data and missing traces due to acquisition limitations affects negatively the quality of advanced processing techniques like multiple attenuation, pre-stack Kirchhoff migration, full waveform inversion and AVAZ/AVO analysis, which require a dense and regularly spaced sampled seismic reflection data to properly work [4, 14, 53]. Another important limitation of under-sampled seismic data sets appears during reservoir characterization. Seismic inversion methodologies, and particularly stochastic seismic inversion [2] require the knowledge of the entire cube of seismic amplitudes.

Seismic data interpolation and reconstruction aim at predicting missing seismic traces to reconstruct regularly and irregularly sampled seismic reflection data sets. Various interpolation methods have been proposed during the last decades differing mostly on the type of approach they have regarding the choice of the interpolation strategy and the mathematical operation used to fill the seismic gaps with new predicted traces. One classification divides interpolation methods into model-driven, or wave-equation-based methods, and data-driven, or signal processing-based methods [36, 49].

Model-driven techniques exploit the physics of acoustic wave propagation within the subsurface to reconstruct seismic traces: the wave field is mapped from the irregular input seismic data to another physical domain through an operator (e.g., Fourier and Radon transforms) where

✉ Francesco Turco
turfr013@student.otago.ac.nz

Leonardo Azevedo
leonardo.azevedo@tecnico.ulisboa.pt

Dan Herold
dherold@parallelgeo.com

[1] CERENA, DECivil, Instituto Superior Técnico, Lisbon University, Av. Rovisco Pais, 1049-001 Lisbon, Portugal

[2] Present address: Department of Geology, University of Otago, Dunedin 9054, New Zealand

[3] Parallel Geoscience Corporation, Crystal Bay, NV, USA

it is interpolated. The modeled physical domain is then transformed back into the original data space, obtaining a regularized grid of seismic data that would have been acquired with a properly sampled experiment [13, 27, 39, 45, 48]. These methods often require information about the spatial distribution of the subsurface velocity field [35] and can be performed on irregularly and coarsely sampled data sets with large gaps.

Interpolation approaches based on signal processing principles (data-driven approach) do not require any information on the Earth's subsurface elastic properties, but use the intrinsic properties of the signal assuming patterns in the sampled data. Many of these methods exploit the data sparsity representation in a transform domain and formulate the reconstruction as a L_1-norm minimization problem; for instance, in the Fourier [13, 32, 40, 41], Radon [8, 47], local Radon [55], or curvelet domain [19, 20, 34]. Dictionary learning methods are alternatives for predefining a transform, as they provide a dictionary that represents well the sparsity of the original data [26, 52].

Another important family of methods is based on prediction filters applied in the t-x [6, 12], f-x [28, 38, 44, 54], f-k [17, 18], or curvelet domains [33]. In this family of methods, error prediction filters are estimated from low-frequency components and are used to interpolate the higher frequency part of the data [6, 16, 32, 38, 44].

Interpolation methods based on rank-reduction rely on the assumption that fully sampled seismic data, when organized into a matrix, are of low rank-structures, and missing traces and noise increase the rank of the structures. Low-rank approximation methods have been used extensively for seismic data denoising ([50] and references therein), and were recently applied to seismic data reconstruction [23, 24, 26, 37, 51]. Yang et al. [57], Ma [30], Aravkin et al. [1], and Kumar et al. [25] address the reconstruction problem as a nuclear-norm minimization matrix completion problem. Jia and Ma [22] present a machine learning method applied to seismic interpolation, training the algorithm with data sets that display similar geometries to the one to be interpolated.

In this work, we introduce an alternative approach to the seismic interpolation problem. In our algorithm, missing traces are reconstructed through stochastic sequential simulations that are constrained by the neighbor recorded seismic amplitudes and by an estimate of the local structure variation [5, 21]. The success of sequential simulation methods relies in the ability to reproduce the statistics of the experimental data (e.g., histogram and variogram models), generating alternative interpolated traces (i.e., realizations), which are equivalent under the same set of assumptions and from where we can assess uncertainties on the estimation [9, 15]. By definition, each realization reproduces the experimental data at its location, the prior probability distribution of the simulated property and a spatial continuity pattern, estimated by a variogram in the 2-point geostatistics framework [2].

Stochastic sequential simulation techniques have been widely used in different fields related with Earth and environmental sciences [9, 15] to predict the value of a given property of interest at an unknown location based on experimental data, sampled at sparse locations. In geophysics, stochastic sequential simulation techniques have been widely used in iterative geostatistical seismic inversion methodologies [2] and in the inference of velocity models for time-to-depth conversion [10, 11]. A geostatistical method to interpolate non-stationary patterns of geological formations have been purposed by Soares [42] to model folded geological structures with kriging with local anisotropies. Other applications of kriging with local anisotropies can be found in environmental and mineral resources applications [46]. References [56] and [29] introduced the concept of local anisotropies in stochastic simulation procedures for the characterization of sedimentary sand channels.

The success of the proposed seismic interpolation approach is related with the ability to impose local variogram, or spatial covariance models inferred directly from the existing seismic records handling in this way the non-stationary behavior of seismic data (i.e., the local structure variation). The results retrieved with the proposed approach are compared against a standard f-x interpolation method [44]. Although Spitz's method is not a state-of-art seismic interpolation technique—as for example, 5D seismic interpolation [16, 33]—it is the basis of many seismic interpolation techniques [6, 17, 18, 33, 35, 38, 54]. To our knowledge, this is the first application of a stochastic sequential simulation technique to the problem of seismic data interpolation. Due to this reason, the examples shown herein aim at performing at least as good as Spitz's interpolation, opening the door for a new family of seismic interpolation techniques.

We first introduce the stochastic sequential simulation with local structural anisotropies followed by the detailed description of the geostatistical non-stationary seismic data interpolation technique. The proposed interpolation technique is then applied to both synthetic and real 2D and 3D seismic reflection data.

2 Methodology

In this section, we first introduce the stochastic sequential simulation with local structural anisotropy algorithm [5, 21] from a seismic interpolation point of view, followed by the detailed description of the proposed algorithm for geostatistical seismic data interpolation (GSDI). In

stochastic sequential simulation, the term *local structural anisotropy* refers to a local structure variation and non-stationary continuity patterns.

2.1 Direct sequential simulation with local anisotropy

From all the existing geostatistical inference methods, one can distinguish between estimation and simulation methods. The first group includes kriging, while the second group comprises stochastic sequential simulation—e.g., sequential Gaussian simulation [9] and direct sequential simulation [43]. Estimation methods infer the value of a given property at a non-sampled location, and represent the best linear unbiased estimate in a least-squares sense, with minimized local variance. Estimation algorithms generate smooth models unable to reproduce extreme values as interpreted from the experimental data: low values are overestimated and high values underestimated. Smoothing is dependent on local data configuration and is not uniform; it is minimally close to experimental data locations and increases proportionally with the distance from these locations. For this reason, these methods cannot be used to measure local uncertainty.

On the other hand, direct sequential simulations (DSS) [43], like all other stochastic simulation techniques, allows assessing the spatial uncertainty associated with the inferred property by generating multiple models (i.e., realizations) which are considered equally probable under the same a priori assumptions about a given a spatial continuity pattern and a set of experimental data. The main difference of DSS when compared with other stochastic sequential simulation algorithms is the use of the estimated local mean and variance to sample directly from the global conditional distribution function (CDF) $F_z(A)$, as estimated from the entire set of experimental data, without the need of any data transform as in stochastic Gaussian simulation [9].

The simulation grid contains both recorded seismic amplitudes (i.e., experimental data; $A(x_\alpha)$) and empty nodes (samples to be interpolated). The simulated value at x_0 is sampled by Monte Carlo from the conditional distribution function, centered in $A_{sk}(x_0)^*$ and with a range equal to σ^2_{sk}. At each location x_0 the simple kriging estimate ($A_{sk}(x_0)^*$; Eq. 1) and variance (σ^2_{sk}; Eq. 2) are computed following

$$A_{sk}(x_0)^* - m(x_0) = \sum\nolimits_\alpha \lambda_\alpha \left[A(x_\alpha) - m(x_\alpha)\right], \tag{1}$$

$$\sigma^2_{sk} = \sum\nolimits_\alpha \lambda_\alpha C(x_\alpha, x_0) \tag{2}$$

where $m(x_0)$ is the mean of existing seismic amplitudes within the pre-defined neighborhood (i.e., the seismic samples that will be used for the estimation) and $m(x_\alpha)$ is the average value of all seismic samples considered as experimental data. $A(x_\alpha)$ are the recorded or previously simulated seismic amplitudes within the searching neighborhood. λ_α are the weights attributed to each seismic sample involved in the estimation and depend on the spatial continuity pattern imposed by a variogram model ($\gamma(h)$; Eq. 3) or by a spatial covariance matrix (C; Eq. 2).

The variogram $\gamma(h)$(Eq. 3) represents a measure of the spatial continuity (or variability) of the property of interest; it describes the correlation between pairs of amplitudes separated by an arbitrary distance h ($A(x_0)$ and $A(x_0 + h)$) (Fig. 1). For each simulation, the neighborhood of conditioning data is selected using an elliptical search radius parameterized by the local variogram model (Fig. 1a, b).

$$\gamma(h) = \sum\nolimits_{\alpha=1}^{N(h)} [A(x_0) - A(x_0 + h)]^2. \tag{3}$$

DSS with local anisotropy (DSS-LA) is a particular case of direct sequential simulation where different variogram models are assigned to each grid node individually contrary to the conventional DSS approach where a global variogram model is used for the entire simulation grid. In Eq. 2, the spatial covariance matrix, C, changes locally at each seismic sample location. The definition of local varying spatial continuity patterns allows handling non-stationary spatial phenomema as, for instance, hyperbolic events in the shot or common mid-point (CMP) gathers. This stochastic sequential simulation algorithm may be summarized in the following sequence of steps [5, 21]:

1. Define a random path that visits all the simulation grid nodes x_0 corresponding to the seismic samples to be interpolated;
2. At each node location, calculate simple kriging estimate $A_{sk}(x_0)^*$ (Eq. 1) and variance σ^2_{sk} (Eq. 2) conditioned by neighbor seismic samples, including previously simulated values following the local variogram models;
3. Define a local CDF from the global CDF $F_z(A)$, built from all the existing recorded seismic samples, centered in $A_{sk}(x_0)^*$ with an interval equal to σ^2_{sk};
4. Draw a simulated value $A_{sk}(x_0)$ from the selected interval of the global CDF $F_z(A)$;
5. Visit the remaining simulation grid nodes until all nodes have been visited and simulated.

As all nodes of the simulation grid are visited following a random path, and at each node local kriging estimate and variance are computed based on both experimental data and previously simulated values, the conditioning data for the stochastic simulation changes accordingly to the random path. Hence different runs generate different models.

The key point for the success of this algorithm when dealing with seismic interpolation is the reliability of the local variogram models in describing the spatial behavior

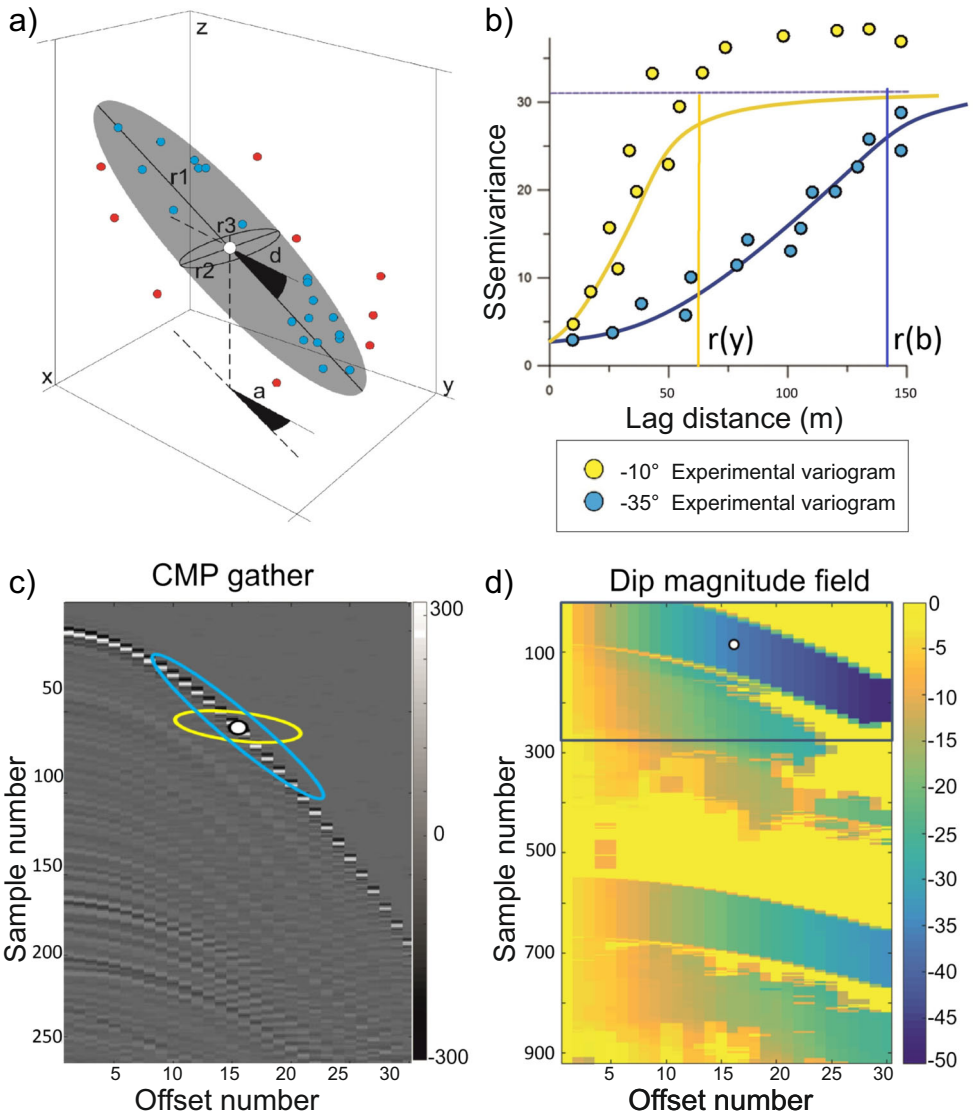

Fig. 1 Spatial representativeness of the variogram in the seismic interpolation application. **a** Spatial representation of the variogram parameters **b** Experimental variogram model along the directions represented by the two ellipses (**c**); $r(y)$ and $r(b)$ represent the horizontal ranges of the yellow and blue ellipsoids (**c**), respectively **c** Zoom of the pre-stack CMP gather (location in the blue rectangle (**d**)); the white dot represents the seismic sample considered for the experimental variogram calculation with other seismic samples within the blue and yellow ellipses **d** Dip magnitude field calculated from a pre-stack CMP gather

of seismic amplitudes within the simulation grid. The definition of the local variogram model parameters (ranges in the three dimensions (*r1–r3*), dip (*d*), and azimuth (*a*), Fig. 1a) is fundamental, since it provides information about the spatial continuity (or variability) of the seismic amplitudes Ideally, the conditioning data for the simulation of a seismic sample lying on a specific seismic event should be the recorded amplitudes of that seismic event in the neighbor traces (i.e., blue dots in Fig. 1a). Consider the seismic sample marked by the white dot in Fig. 1a, c, d; an elliptical search radius with dip magnitude of −35° (blue ellipsoid in Fig. 1c) will constrain the simulation in order to give more weight to the samples lying along the same seismic event in the neighbor traces (i.e., the seafloor). In the proposed technique, these variogram parameters are estimated directly from the seismic record as they represent the geometry of existing seismic events in the data (Fig. 1), as discussed in the next section.

2.2 Geostatistical seismic data interpolation

The idea of applying geostatistical simulation for interpolating missing seismic traces is not trivial, as seismic events are non-linear and change in space and time. However, they do form a spatial continuity pattern, which is dependent on the underlying geology. This pattern is exploited by the proposed GSDI technique using local variogram models to describe the structural variability of the data. In the proposed algorithm, the local variogram parameters are computed automatically from a local waveform similarity estimation between the neighbor seismic traces around the location of the missing trace to be interpolated.

The waveform similarity is calculated through a cross-correlation based algorithm analogous to the one presented by Bahorich and Farmer [3]. These parameters comprise direction of maximum coherency in terms of dip (d) and azimuth (a), and the ranges in the three direction of space ($r1$–$r3$) (Fig. 1). The more similar two traces are, the larger the variogram ranges and, consequently, the larger the kriging weights (λ_α) of the conditioning amplitude values along the direction of maximum coherence. When two adjacent traces are less similar, the importance of neighbor samples for the kriging estimate decreases according to the variogram ranges. It is important to note that the dip, azimuth, and coherence value calculated between neighbor traces are linearly interpolated at missing trace locations. Therefore, the performance of GSDI is better when interpolating regularly sampled data sets.

After simulating N realizations, it is possible to explore the variability related with the interpolation, assigning a value of uncertainties to every interpolated sample: well constrained areas of the simulation matrix will display low variability within the ensemble of simulated traces, whereas areas of poor coherency will have higher variability. In the proposed method, the mean model of the simulated ensemble (expected value for each interpolated seismic sample) is assigned to the location of interest. By calculating the mean of the ensemble of realization, possible random noise introduced by outlier simulated values is reduced.

The GSDI method is described in Algorithm 1, and summarized in Fig. 2.

Algorithm 1 GSDI.

1. Input: simulation grid with recorded and missing traces around the location of interest
2. User-defined parameters: number of realizations (N); temporal and spatial sampling interval (ds, dt); and initial variogram parameters ($r1$–$r3$, d, a)
3. Initialize $r1$–$r3$, d, and a cubes
4. Compute dip/azimuth and coherency cubes: cross-correlation between recorded seismic traces within a neighborhood around the location to be interpolated. Interpolate at missing traces columns
5. Update $r1$–$r3$, d, and a cubes: scaling horizontal ranges, allocate angles to d and a
6. For $i = 1{:}N$ run DSS-LA
7. Output: average of N realizations, assign trace to seismic grid

3 Examples

We show the application of GSDI technique in both synthetic and real data sets considering post- and pre-stack domains. Thus, we aim at assessing the performance of this approach in the presence of noise and complex geometries. All data sets used in these application examples are complete and fully sampled; therefore, we randomly zeroed entire seismic traces before interpolation. In this way, we are mimicking an irregular input seismic record while being able to compare the interpolated against the real data. The coherency field used to update the local variogram parameters (step 5 in Algorithm 1) are computed in different ways according to the application example. For the sake of comparison, we carried out the same experiments using an interpolation algorithm based on f-x prediction filters as in [44].

3.1 Post-stack domain interpolation

We first tested the proposed geostatistical seismic interpolation technique in the post-stack domain. The aim of these application examples is to simultaneously increase lateral

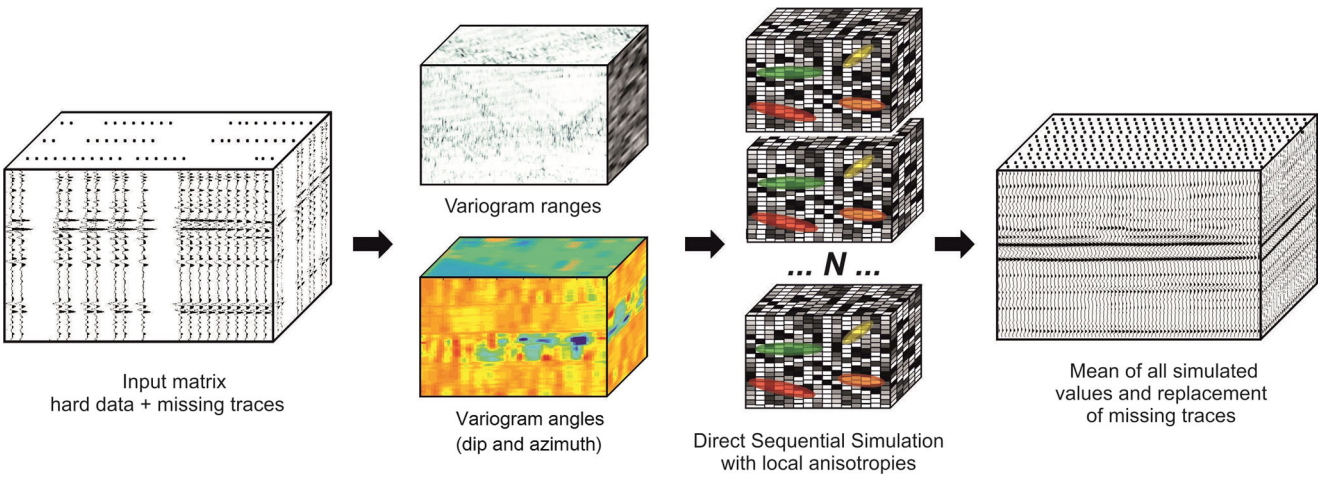

Fig. 2 Schematic representation of the workflow for geostatistical seismic data interpolation

resolution and reconstruct missing seismic traces at specific CMP locations. We show examples in 2D and 3D.

3.1.1 2D examples: synthetic and real field seismic sections

The first example comprises a 2D synthetic seismic section, modeled with four distinct reflectors with dips ranging from 4 ms/trace to horizontal and dipping in both directions. There are two areas of crossing events and a small percentage of random white noise was added to the section (Fig. 3a). The temporal and spatial sampling intervals are, respectively, 1 ms and 10. The initial horizontal maximum range *r1* is set to 180 m. The coherency-based dip field was computed over moving vertical windows between

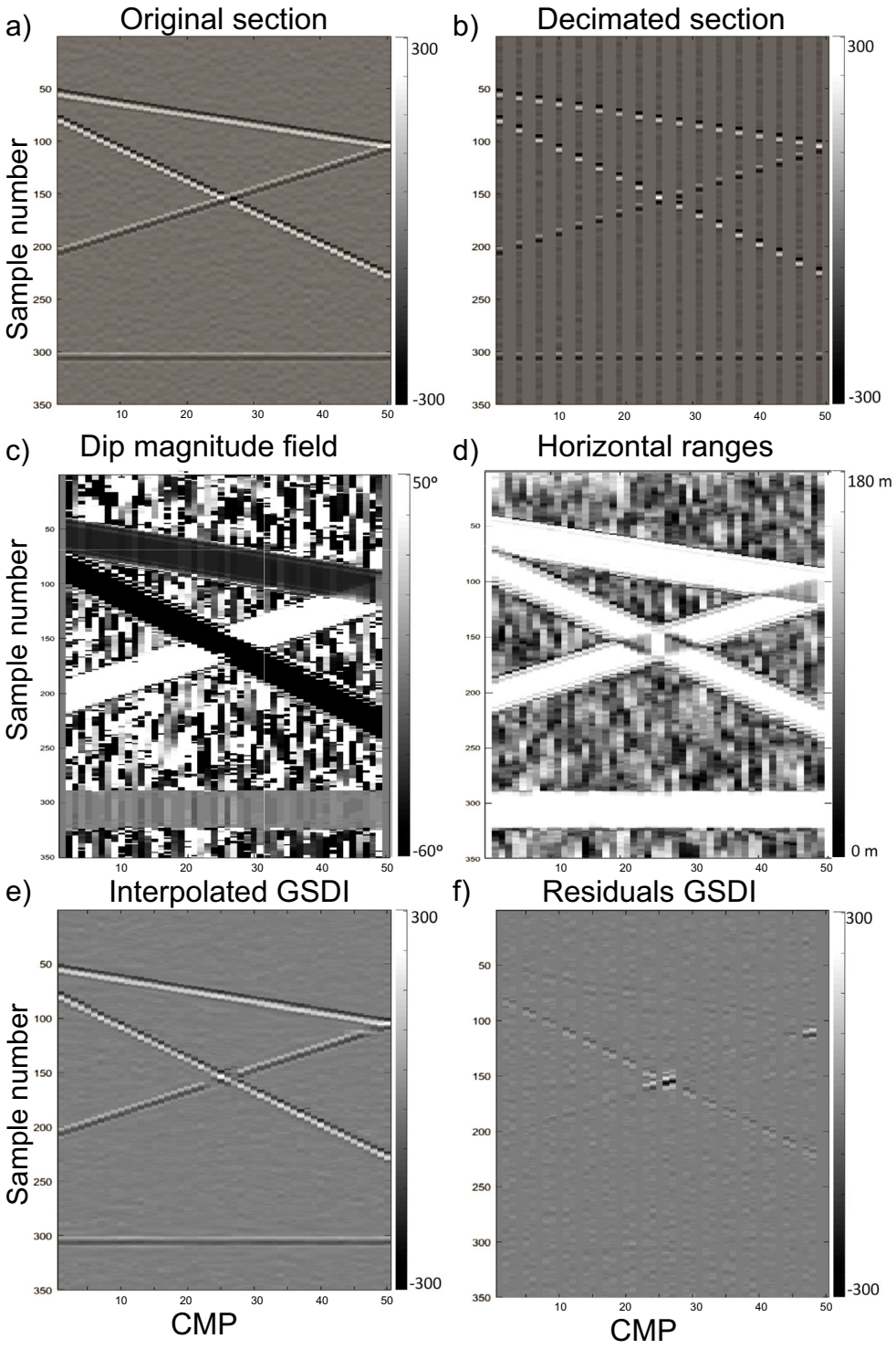

Fig. 3 Post-stack synthetic seismic section with dipping reflectors. **a** Original section. **b** Decimated by 66% (only 33% of the original traces were kept for the interpolation). **c**, **d** Dip magnitude field and the horizontal ranges field, respectively. **e** Interpolated section with GSDI method. **f** Residuals between the results of the interpolation and the true section

neighbor-recorded traces around the location to be interpolated and linearly interpolated at missing trace locations. Due to the small size of the data and for illustration of the method, we computed the cross-correlation between all existing traces. In real datasets, we just need to compute the cross-correlation between traces around the location to be interpolated. The resulting coherency dip field was used to update the local variogram model (Fig. 3c). The horizontal range is automatically reduced in regions of low correlation between adjacent traces (Fig. 3d), for instance in the area of crossing events and in absence of distinct reflectors. Thirty-three of the 50 traces were removed before coherency estimation, leaving every original third trace (Fig. 3b). Figure 3e shows the reconstructed seismic record. Given the strong dependence on the dip estimation, the reconstruction of the seismic traces is less accurate in cases of large gaps or conflicting dips, as can be noticed in Fig. 3f. Nevertheless the amount of random noise is kept low in the GSDIinterpolated section (Fig. 3e). The f-k spectra for the original and reconstructed seismic section are shown in Fig. 4. As expected, the decimated section is strongly affected by spatial aliasing (Fig. 4b); the f-k spectra of the GSDI section (Fig. 4c) reproduces well the f-k spectrum calculated from the original section and shows the ability of the proposed technique to successfully attenuate spatial aliasing.

Figure 5a shows an inline extracted from a real 3D post-stack cube showing complex geology and conflicting dips. Sixty-five percent of the traces were randomly zeroed, leaving all the remaining original traces as conditioning data for the simulation (Fig. 5b). Spacing between the crosslines is 10 m and the sampling interval is 1 ms. *r1* is initialized at 180 m. The dip magnitude field and the coherency matrix (Fig. 5c, d) calculated between recorded traces and linearly interpolated at missing trace locations are used to parameterized local variograms: when the coherency value is 1, *r1* is maximum. In absence of clear seismic events, for instance where faults are present, it is smaller due to low similarity between waveforms (Fig. 5d). This leads to a higher variability of the simulated amplitude values for grid nodes in these areas, which results in effective white noise attenuation when computing the mean of all the simulated values.

The interpolated section reproduces considerably well the original seismic record (Fig. 5e), maintaining important structural and sedimentary features such as faults, strata, and high amplitude zones. The differences with respect to the original section are shown in Fig. 5f.

The f-k spectra of the original, decimated and interpolated sections can be seen in Fig. 6a–c respectively. The f-k spectrum computed over the interpolated section is able to properly reproduce the original one.

3.1.2 3D example: field data seismic cube

Figure 7a shows the original seismic volume where a gap of 3 × 3 missing traces was created (Fig. 7b). With this simple experiment, we want to highlight an advantage of the GSDI method, which is the possibility to assess the reliability of the interpolated traces, at a sample scale, based on the variability of the simulation matrixes at given sample locations for different realizations. In this experiment, spatial and temporal intervals were set to 10 m and 1 ms, respectively. The simulation grid is a subset of 10 × 10 traces of the original seismic volume (Fig. 7b). The traces lying within this grid were used to estimate the coherency dip field. Local 3D variograms were initialized with horizontal ranges *r1* and *r2* equal to 180 m.

The correlation coefficients between the entire interpolated and the original traces are higher than 95% for all the nine traces, although the ensemble of realizations of traces 5–8 display a larger spread, with a broader range of values,

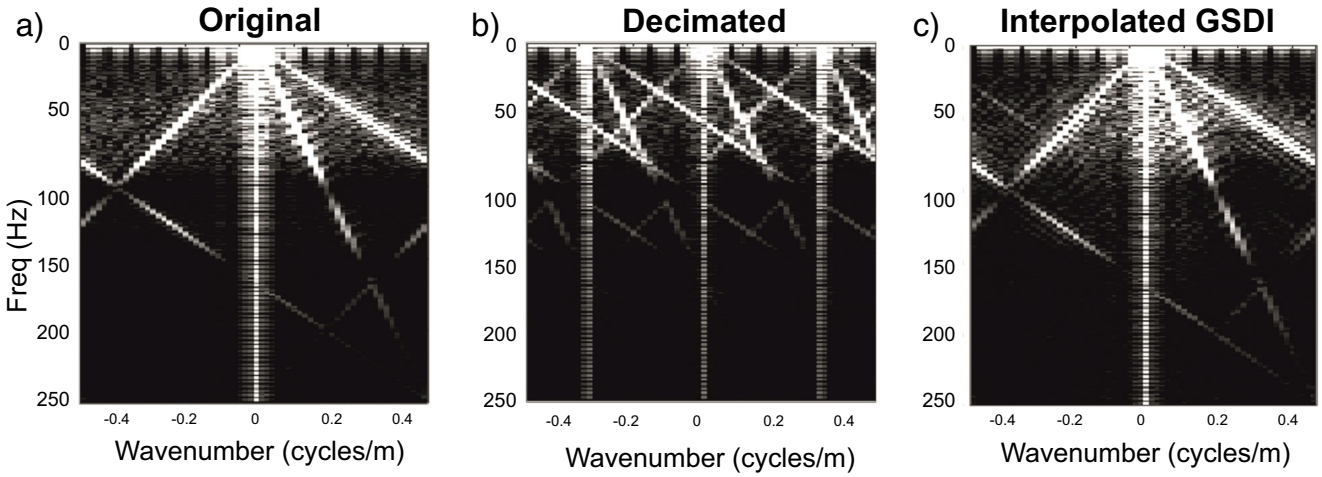

Fig. 4 F-k spectra of the **a** original seismic section, **b** decimated section, and **c** interpolated section with GSDI

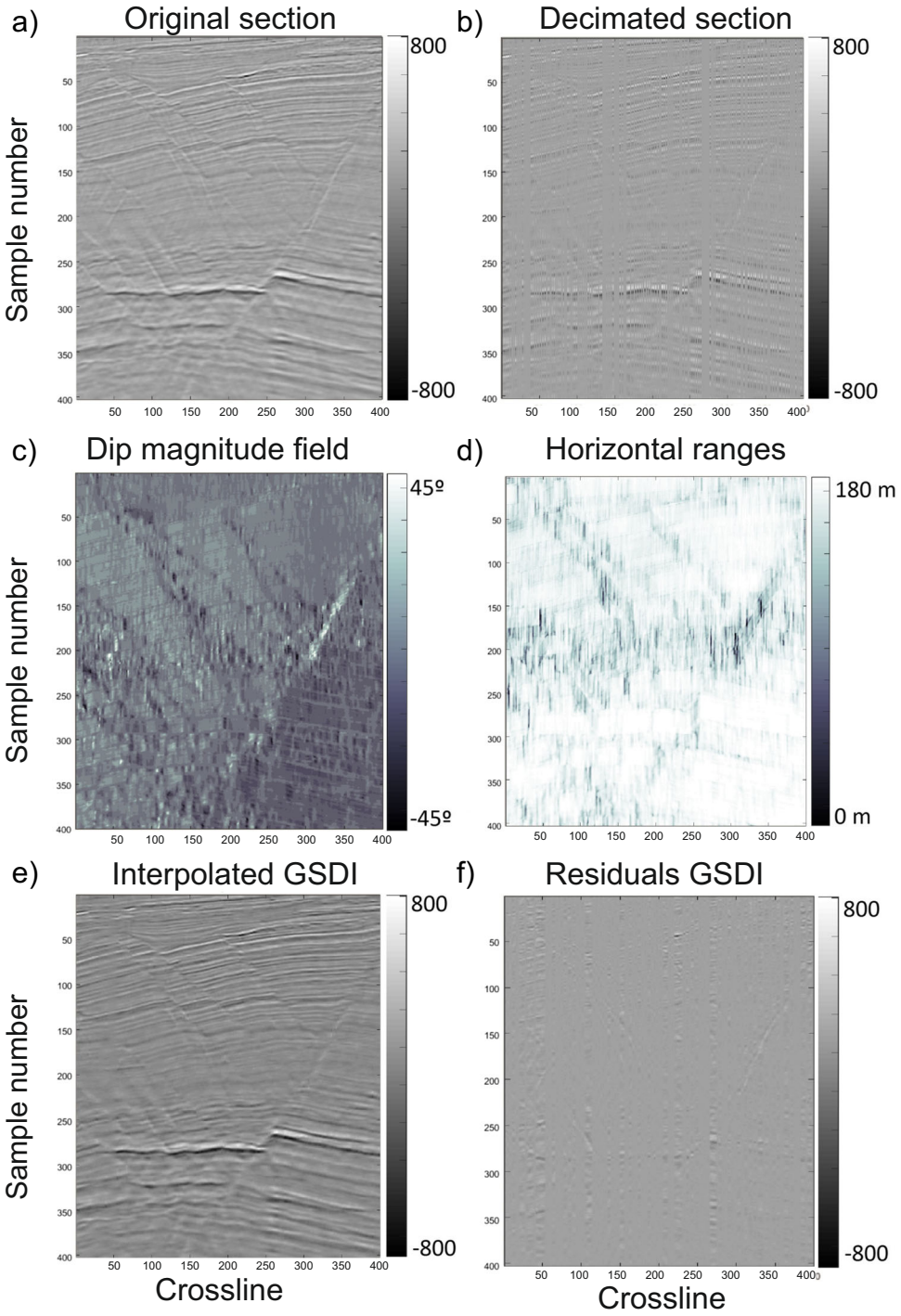

Fig. 5 Real case application to a post-stack seismic record. **a** Original section. **b** Randomly decimated by 65% (only 140 of 400 original traces left. **c**, **d** Dip magnitude field and the horizontal ranges field, respectively. **e**, **f** Reconstructed section with GSDI method and the residuals, respectively

with respect to the other interpolated traces. The dotted lines in Fig. 7c (traces 1 and 6) represent the boundaries within which all the simulated values for that portion of the traces fall. It can be noticed that the bounds are larger for trace 6 than for trace 1, meaning a lower level of confidence for the interpolated samples in that specific interval. The reason of the higher variability in the simulated values of traces 5, 6, 8, and 9 is an abrupt change in the seismic patterns (for instance, reflector steepness or thickness) in the lower-right corner of the simulation matrix (red triangle in Fig. 7a).

As a quality control of the interpolated traces, we show a comparison of amplitude spectra of the real and

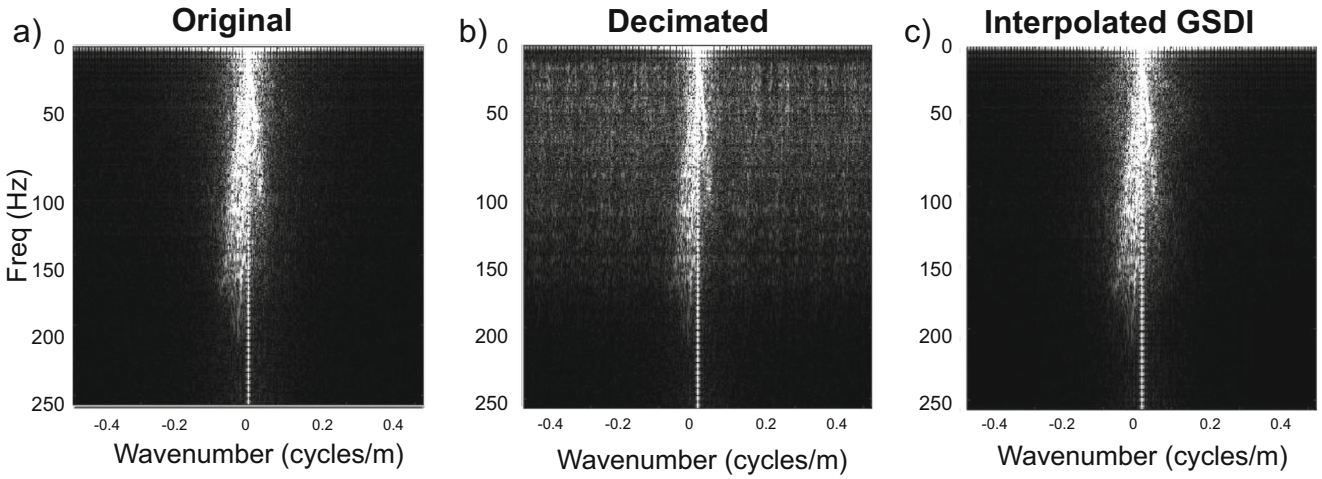

Fig. 6 *F-k* spectra of the **a** original seismic section, **b** decimated section, and **c** interpolated section with GSDI

interpolated seismic traces for the nine considered locations (Fig. 7d). The reproduction of high frequencies is ensured using stochastic sequential simulation as the interpolation technique.

3.2 Pre-stack domain interpolation

In this section, we show two application examples of GSDI in the pre-stack domain. The first example comprises the interpolation of seismic traces within an under-sampled pre-stack gather, resulting in a higher fold with respect to the original data. The second illustrates the ability of the GSDI technique to be used in interpolating in-between inlines or crosslines. Interpolation in the midpoint-offset domain improves the signal to noise ratio, while it allows decreasing the bin size (i.e., increase spatial resolution) when it is carried on in the shot point, inline, or crossline domains. It must be stressed that although in the application examples presented in this paper the interpolation is carried on in the midpoint-offset and shot-point coordinate spaces, the same method is applicable in the midpoint coordinate, inline, or crossline domains. The 2D coherency-based dip estimation and successive interpolation are carried out along the offset direction in the first example, and along the shot point direction in the second example.

3.2.1 Interpolation in offset domain

Figure 8a shows a CMP gather from a real marine 2D seismic survey. We created a decimated gather by removing every second trace (Fig. 8b); in this way, we simulate the case of an under-sampled data set that one desires to improve by increasing the fold and thus the signal-to-noise ratio. In this case, fold increases from 15 to 30 after interpolation. We also generated an irregularly sampled gather by zeroing 11 traces resulting in gaps of 3, 2, and 1 seismic traces (Fig. 8c), in order to test GSDI method in the presence of both larger gaps of missing seismic traces and non-linear events. The spatial and temporal sampling intervals were set to 10 m and 1 ms, respectively. The initial maximum horizontal variogram range *r1* was set to 120. The coherency based dip field for the decimated gather was calculated between the remaining traces after every second offset was removed, and it was linearly interpolated at missing trace locations. In the case of the gather with traces gaps (Fig. 8c), the dip field was calculated from the nearest fully sampled CMP gather. In real case, examples, the nearest fully sampled CMP gather can either be too far or be inexistent; in this case the dip field can be computed between adjacent traces and interpolated at missing trace locations. As discussed in Section 2.1, the use of a linear interpolator is only valid under the assumption that the number of consecutive missing traces is small enough compared to the curvature of seismic events. In presence of high curvature, a spline interpolator would be more appropriate.

The trace reconstruction is good both in the decimated gather (Fig. 8e, g) and in the gather with gaps (Fig. 8h, l). Despite small amplitude, decreases can be observed in Fig. 8l close to the reflection corresponding to the seafloor, the residuals for both of the interpolated gathers are considerably small (Fig. 8f, i) and the overall results show a good fitting between the original traces and the reconstructed ones. Figure 9a, b illustrates zoomed portions of five of the reconstructed traces from near, middle, and far offset regions of the gather and their amplitude

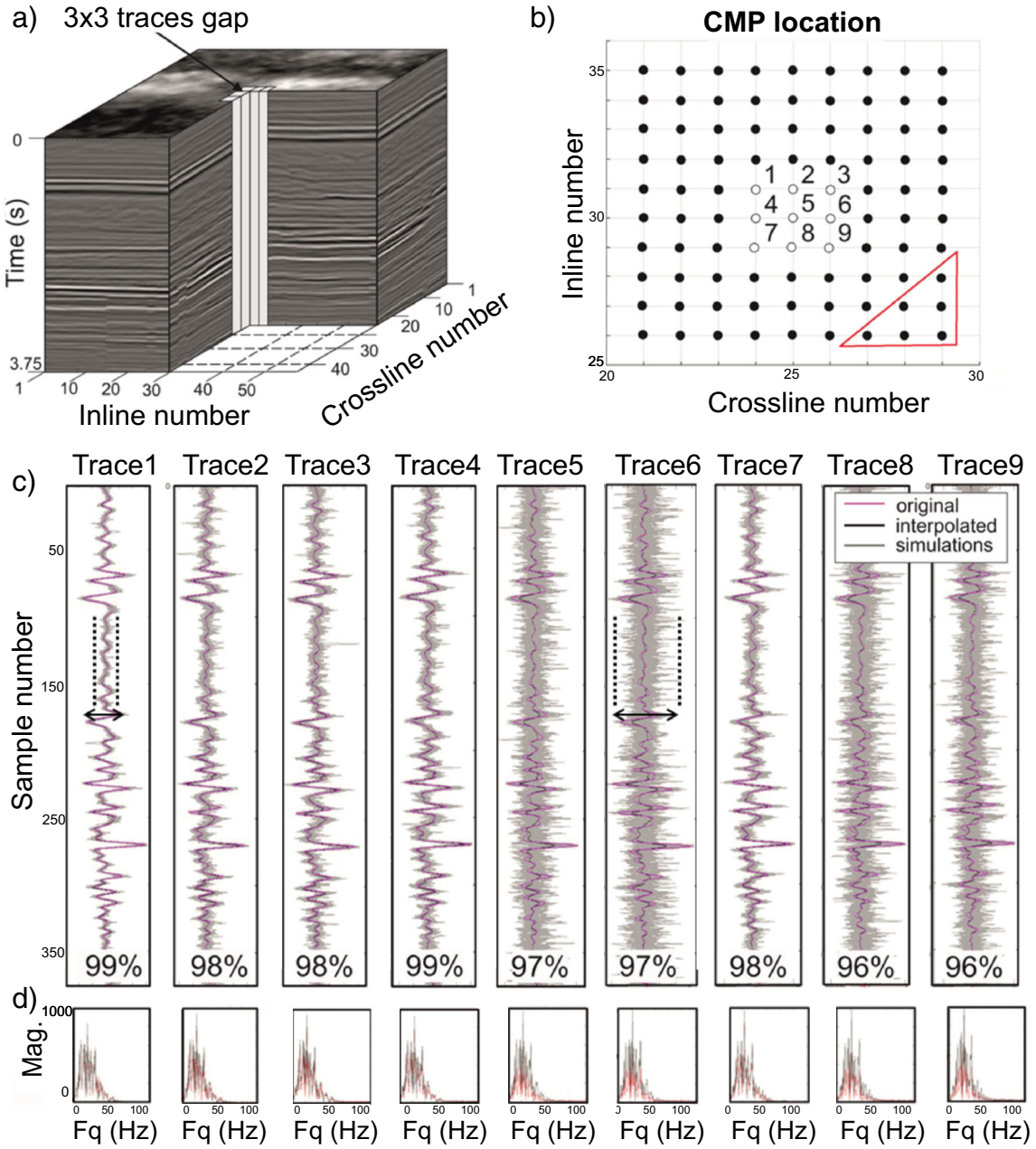

Fig. 7 **a** Seismic post-stack cube with a gap of 3 × 3 traces shown in the middle and the region of abrupt change in the geometry of the underlying geology shown with the red triangle. **b** A map view of the seismic volume indicating the nine removed traces. **c** The original (pink) and interpolated (black) traces; the predicted trace is the mean of the one hundred realizations, shown with gray lines. The larger the spread between the realization, the lower the level of reliability of the interpolated trace. **d** Amplitude spectra of each the original traces (red) and the interpolated ones (black)

spectra, respectively. Interpolated traces have correlation coefficients above 85% when compared with the true ones. This example shows that the GSDI technique can handle small gaps of missing traces even in presence of nonlinear reflection as in the case of a pre-stack CMP gather. The $f - k$ spectra for the original, decimated, and interpolated gather are shown in Fig. 10: in the decimated gather, spatial aliasing occurs at all frequencies, whereas the original and interpolated data are only aliased above 60 Hz.

To benchmark the proposed interpolation technique, f-x interpolation [44] was applied to the same CMP gather corrected for normal move-out (NMO). Results are shown in Fig. 11. It can be noticed that the GSDI introduces less amount of random noise than the f-x interpolation

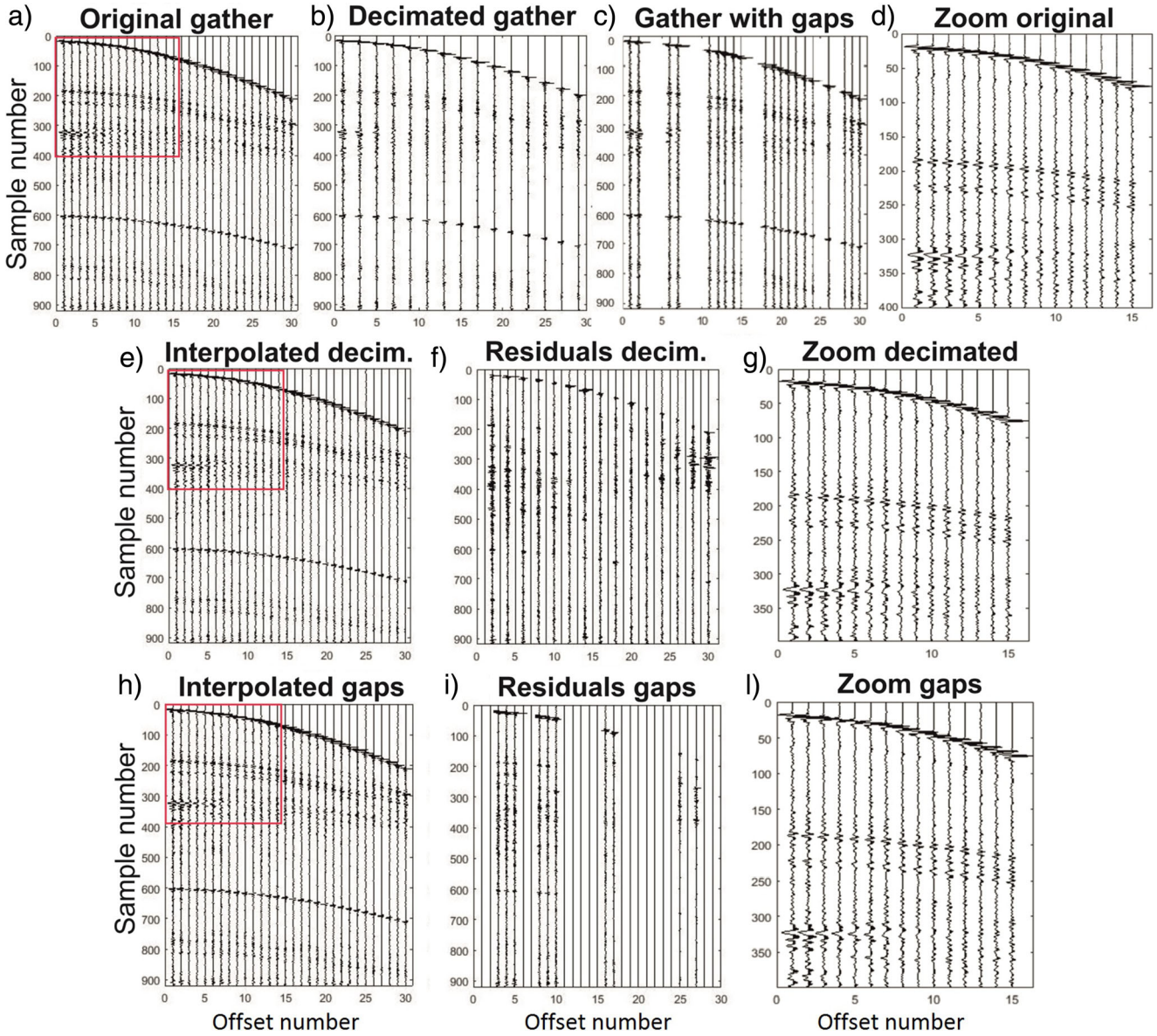

Fig. 8 2D marine CMP gather **a** original gather; **b** decimated by 50% leaving every second trace; **c** gather with random gaps; **e**, **h** show interpolation with GSDI (**b**, **c**, respectively), while **h**, **i** residuals. Note that residuals were multiplied by 4 for a better comparison. **d**, **g**, **l** Zoomed area in the red rectangles

method (Fig. 11f, i). Furthermore, curved nonlinear events that may remain even after NMO, due to inaccurate velocity model or in presence of multiples, as is this case, are handled by the GSDI technique, whereas they are poorly resolved by the f-x interpolation algorithm (Fig. 11d, g). The GSDI interpolated gathers show to minor residuals between the interpolated and the original traces (i.e., a higher percentage of zero residuals and small absolute residuals) when compared to the f-x interpolated section (Fig. 12a). The trace-by-trace signal-to-noise ratio (SNR) is also higher for the GSDI interpolated traces, supporting the performance of the proposed method (Fig. 12b). The residuals between the interpolated and the original gathers is considered noise.

3.2.2 Interpolation in shot point domain

As last example, we use a set of 2D marine shot gathers to test the GSDI method in the reconstruction of entire gathers. This procedure is particularly interesting since it can easily be extended to achieve the reconstruction of entire inlines or crosslines of an under-sampled 3D data set. The original

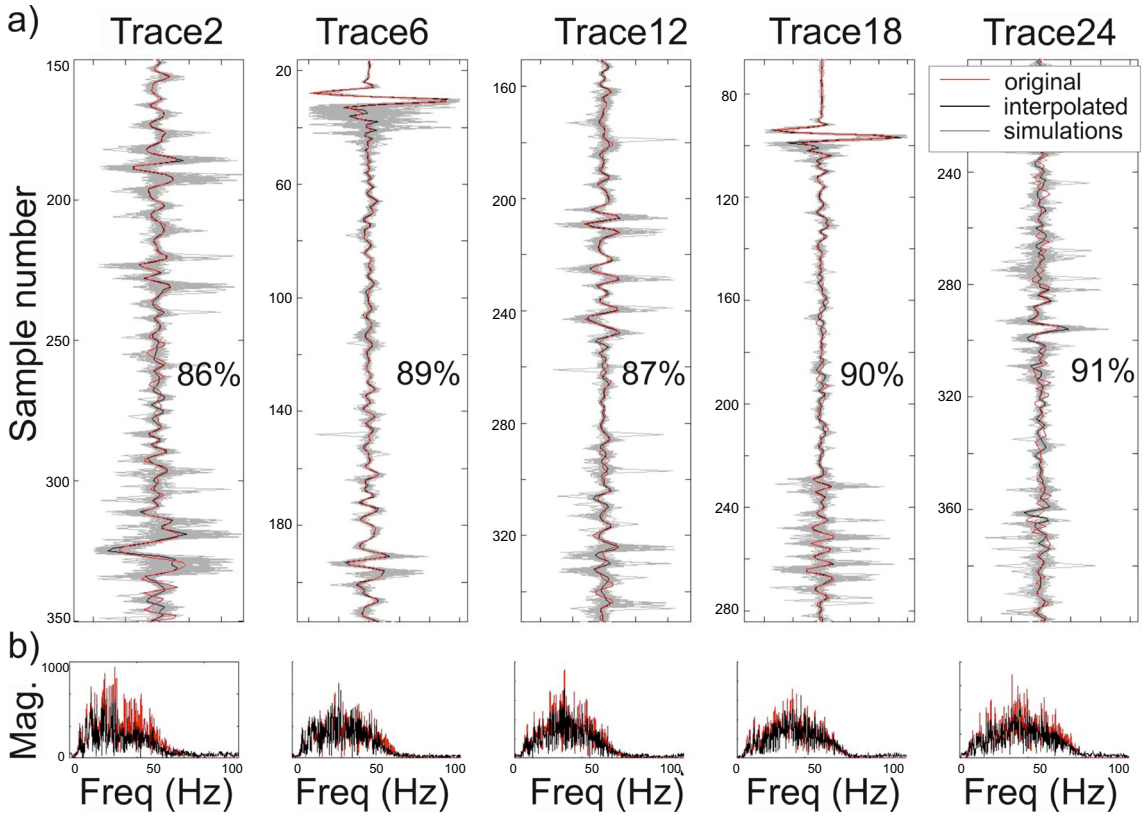

Fig. 9 **a** Comparison of reconstructed (black lines) and original traces (red lines) from the interpolated CMP gather in Fig. 8e, b relative amplitude spectra. Every gray line represents one realization of the stochastic sequential simulation, and the interpolated trace is the mean of all the realizations. Different portions of the traces are shown. Correlation coefficients are also shown

shot point spacing interval is 37.5 m, and increased to 70 m after every second shot gather was removed. Interpolation is carried out for each offset as illustrated in Fig. 13a. As can be noticed in Fig. 13b, the dip angle is close to 0° at all sample locations, due to the very smooth and regular underlying geology. The variogram horizontal range *r2* was set to 350 m and successively updated according to the coherency field (Fig. 13c). Figure 14 illustrates

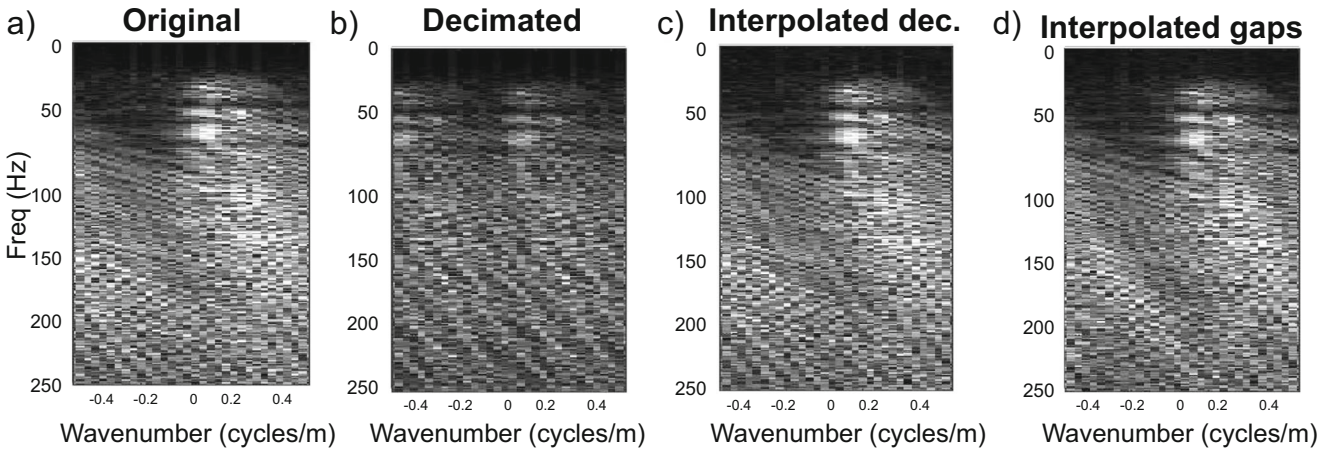

Fig. 10 *F-k* spectra of the **a** original gather, **b** decimated gather, **c** interpolated decimated gather, and **d** interpolated gather with gaps

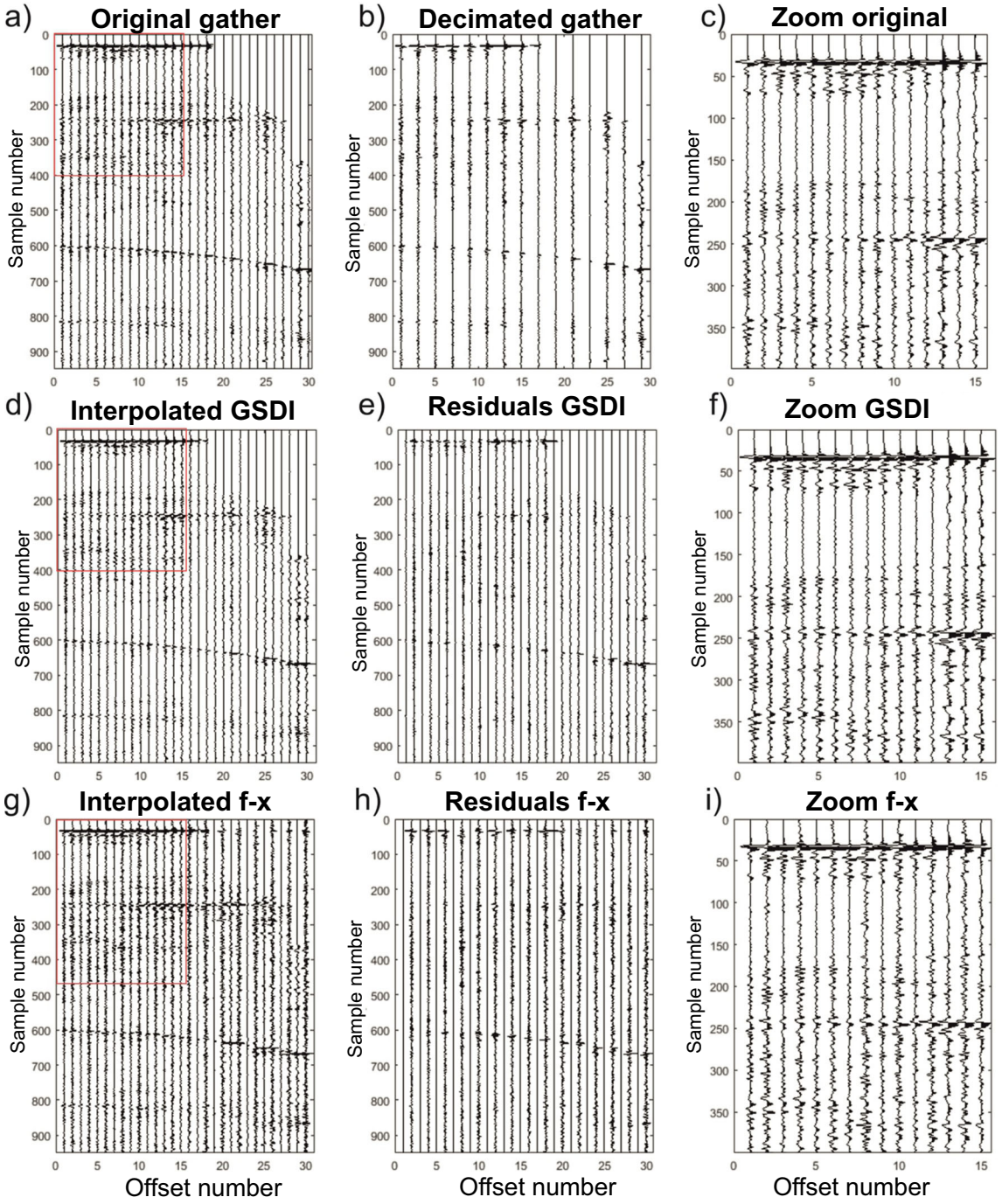

Fig. 11 NMO corrected 2D marine CMP gather. **a** Original gather. **b** Decimated by 50% leaving every second trace. **d** Interpolated gather with GSDI and **g** with f-x technique. **e**, **h** Relative residuals, multiplied by 4 for a better comparison. **c**, **f**, and **i** Zoomed areas in the red rectangles

nine of the initial gathers before and after interpolation (zoom in Fig. 15). In the deepest parts of the seismic records, where the seismic texture tends to become more chaotic and the coherency between traces is low, we can observe decrease in the interpolation accuracy, whereas in the shallower parts, where reflection events are clear and distinct, the interpolated traces match very well the original ones. Overall, despite small amplitude change, the

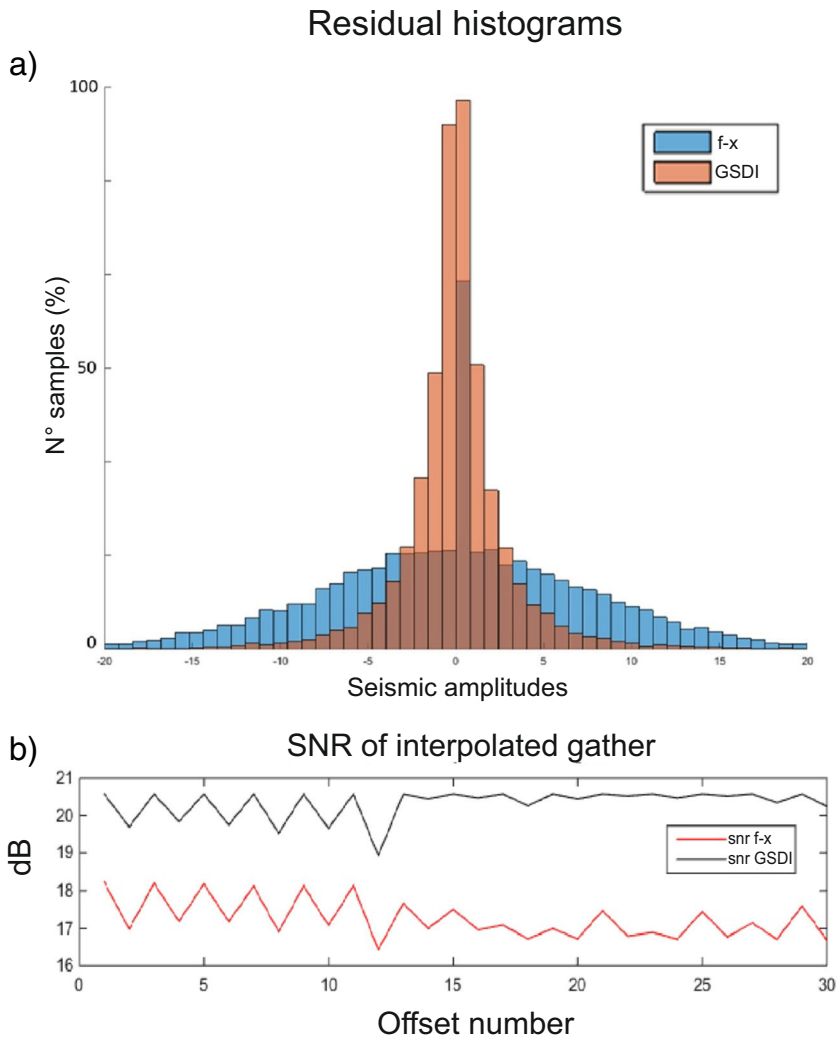

Fig. 12 **a** Histograms of the residuals for the NMO corrected pre-stack gather computed for both interpolation methods, considering only the interpolated traces (even trace numbers, Fig. 11e, h). **b** SNR of interpolated gather with f-x technique (red line) and with GSDI technique (black line). Offset numbers refer to Fig. 11d, g

reconstructed traces are hardly discernible from the input data. The f-k spectra of the original shot gather 3 and the interpolated shot gather 4 are shown in Fig. 16a, b, respectively. In this experiments, GSDI algorithm has been able to reconstruct most of the signal present in the shot gathers, despite some of the negative dip events (up dip with increasing offset) display lower energy than in the original shots (darker stripes in Fig. 16b).

4 Discussion

The GSDI technique presented herein is a local interpolation method based on the assumption that there is a spatial continuity pattern within the recorded data (i.e., distinct reflections), which is expressed in terms of dip, azimuth, and coherence between portions of adjacent traces. In absence of clear events, as well as in presence of large amount of white noise, the GSDI loses precision due to the inaccuracy of the dip estimation and the decreasing number of conditioning data for the simulation as can be observed examples illustrated by Figs. 14 and 15. The reconstruction of missing seismic traces is based on the definition of local variogram models which are in turn characterized by dip, azimuth, and range parameters that vary for each seismic sample. In presence of large gaps or crossing events, the dip estimation precision decreases considerably, and will eventually partially or totally compromise the interpolation accuracy. It is hard to define the maximum number of consecutive missing traces

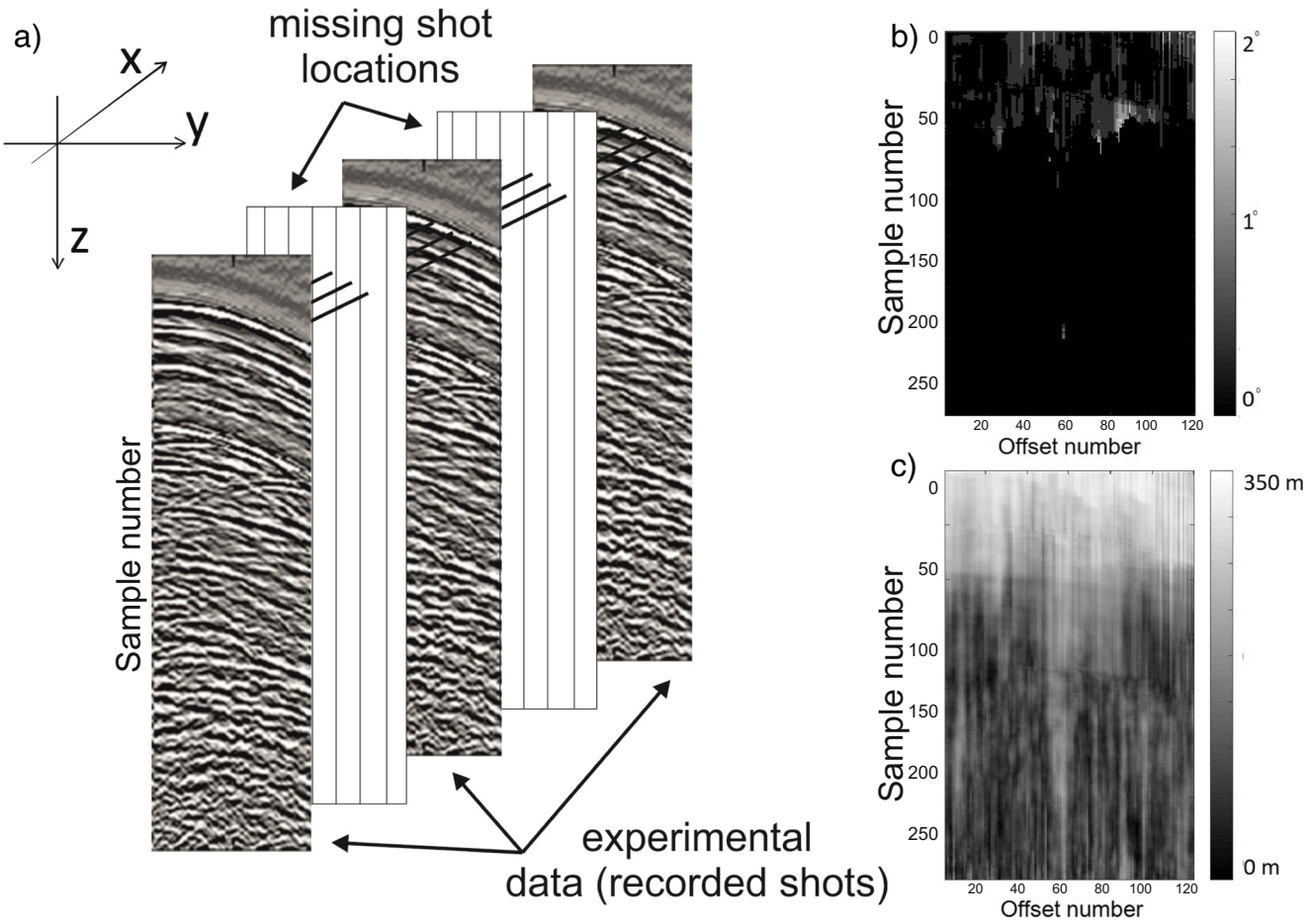

Fig. 13 **a** Dip magnitude estimation example for the reconstruction of an entire gather between recorded gathers. **b**, **c** Dip magnitude field and the horizontal ranges field, respectively, for shot gather number 4 of Fig. 14

for which the proposed method is effective for different data sets, as it depends on the complexities of the seismic events that are present in the recorded data, on the quality of the data and on the similarity between the two traces at the sides of the gap. A dip estimation based on cross-correlation does not handle this problem properly, as the wavelet signature changes when two reflectors interfere with each other; on the other hand, the use of a more accurate dip estimation technique [7, 31] could lead, in some cases, to better results than the ones presented in this paper

A significant advantage of the proposed method, compared to other interpolation methods, is the possibility to assess the reliability of the reconstructed traces at the sample scale by computing the variability of the ensemble of N realizations generated for each seismic sample. In real case examples, when we want to interpolate missing traces, we do not have access to the real seismic traces to benchmark new predicted traces and evaluate the quality of the interpolation technique. With the GSDI technique, the variance of the ensemble of realizations at an unknown location can be used as a confidence index of the predicted amplitude value. When the realizations are very similar one to another, the level of confidence is considered high, while when the variance between different realizations is higher, the level of uncertainty increases (e.g., traces 5, 6, 8, and 9 in Fig. 7c). This confidence index can be exploited in further data processing steps, for example by increasing or decreasing weights to the interpolated pre-stack traces in the stack phase.

Finally, the main issue of many seismic data interpolation techniques in the pre-stack domain is the presence of strong curvatures in the offset direction. Many interpolation

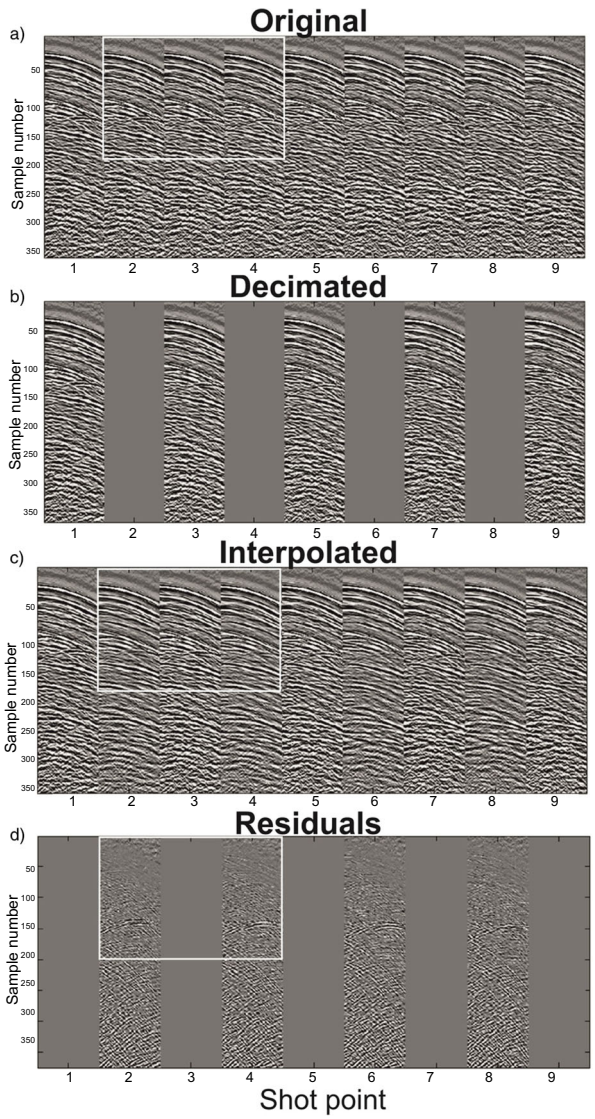

Fig. 14 **a** Original shot-point gathers. **b** Decimated gathers. **c** Interpolated gathers (even numbers). **d** residuals. White rectangle is zoomed in Fig. 15

methods assume linear events in the timespace domain [18, 27, 33, 44, 52], and must therefore reduce the curvature as much as possible before interpolation by application of NMO corrections or time/space windowing. One of the main advantages of GSDI technique relies on the fact that the method does not need any assumption about linearity of seismic events and is applicable equally to post- and pre-stack data sorted in any domain (shot number, common midpoint, inline/crossline, etc.). Moreover, by interpolating seismic traces before velocity analysis, it is possible to increase the fold and model more accurate velocity functions.

5 Conclusions

GSDI technique, introduced in this paper, shows a great potential for seismic data reconstruction based on stochastic sequential simulations. The examples show the significant versatility of this method, which is suitable for both 2D and 3D data applications, in pre- and post-stack domains. Among the main advantages of this method are the following:

- The ability to interpolate non-stationary seismic data, without making assumptions about the linearity of seismic events;
- The possibility of interpolating pre-stack seismic data without the need of a normal move-out correction;
- The possibility to assess to the reliability of the interpolated amplitude values of the reconstructed seismic traces at the sample scale, and the chance to exploit the confidence index in further signal processing steps.
- The ability to reconstruct missing seismic traces without the need to transform the data into another domain to interpolate.

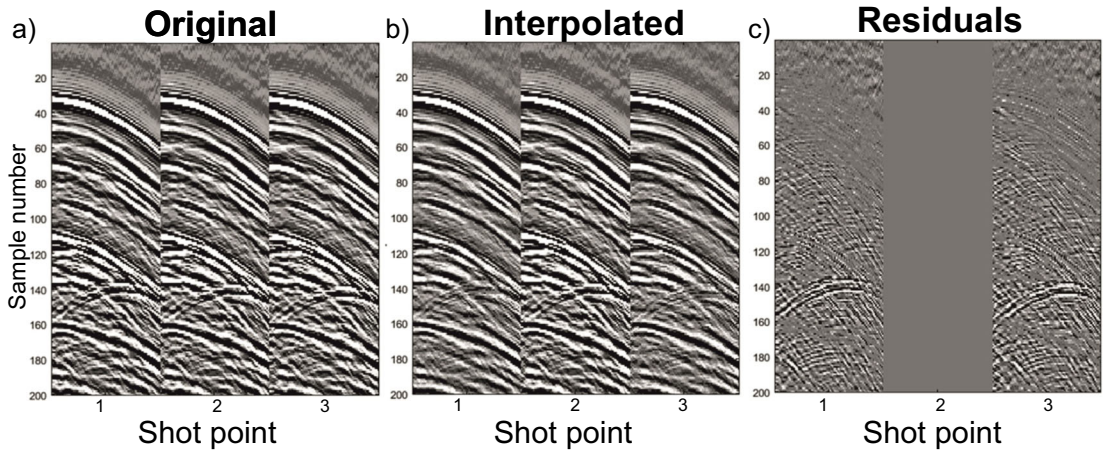

Fig. 15 Zoom of Fig. 15. **a** Original shot-point gathers. **b** Interpolated gathers. **c** Residuals

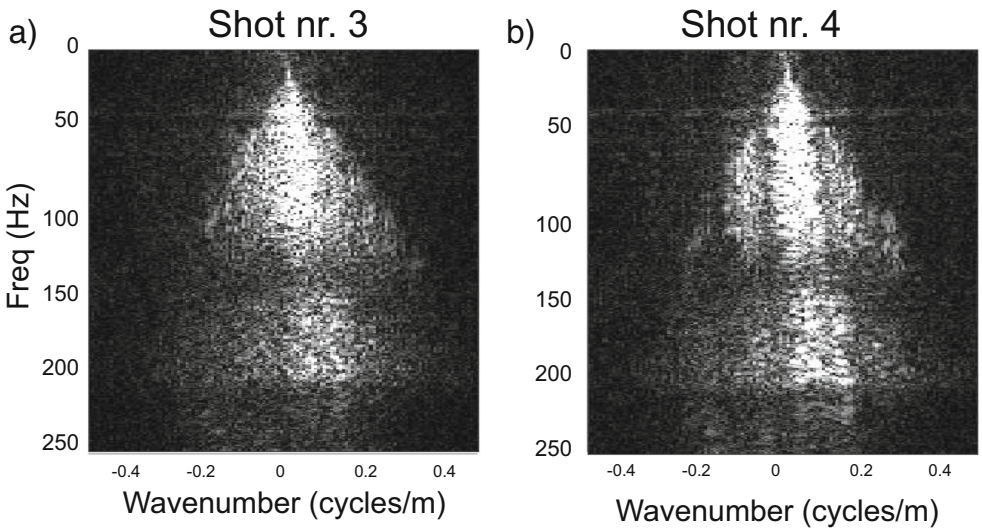

Fig. 16 $F-k$ spectra of **a** shot gather 3 and **b** shot gather 4 of Fig. 14

Acknowledgements The authors gratefully acknowledge the support of the CERENA (strategic project FCT-UID/ECI/04028/2013) and Parallel Geoscience for data availability and the use of SPW 4.0. L. M. Pinheiro for proposing the challenge. A special acknowledgement goes to Gary F. Margrave, Mostafa Naghizadeh, and the CREWES consortium for making available an exhaustive Matlab software library used to test the f-x interpolation technique.

Funding The authors thank the EU Youth Guarantee program, the Italian Ministry of Education and the Autonomous Region Friuli-Venezia Giulia for financial support.

References

1. Aravkin, A.Y., Kumar, R., Mansour, B., Recht, B., Herrmann, F.J.: Fast methods for denoising matrix completion formulations, with application to robust seismic data interpolation. SIAM J. Sci. Comput. **36**, S237–S266 (2014)
2. Azevedo, L., Soares, A.: Geostatistical methods for reservoir geophysics. Advances in oil and gas exploration & production. Springer International Publishing (2017)
3. Bahorich, M., Farmer, S.: 3-D seismic coherency for faults and stratigraphic features: The coherence cube. The Leading Edge **14**(10), 1053–1058 (1995). https://doi.org/10.1190/1.1437077
4. Biondi, B., Palacharla, G.: 3-D prestack migration of common-azimuth data. Geophysics **61**, 1822–1832 (1996)
5. Caeiro, M.H., Demyanov, V., Soares, A.: Optimized history matching with direct sequential image transforming for non-stationary reservoirs. Math. Geosci. https://doi.org/10.1007/s11004-015-9591-0 (2015)
6. Claerbout, J.F.: Earth Soundings Analysis: Processing Versus Inversion. Blackwell Scientific Publications, Cambridge (1992)
7. Dalley, R.M., Gevers, E.E.A., Stampli, G.M., Davies, D.J., Gastaldi, C.N., Ruijtenberg, P.R., Vermeer, G.J.D.: Dip and azimuth displays for 3D seismic interpretation. First Break **7**, 86–95 (1989)
8. Darche, G.: Spatial interpolation using a fast parabolic transform. In: 60th Annual International Meeting, SEG, Expanded Abstracts, pp. 1647–1750 (1990)
9. Deutsch, C., Journel, A.G.: GSLIB. Geostatistical Software Library and Users' Guide. Oxford University Press, Oxford (1998)
10. Doyen, P.M.: Seismic Reservoir Characterization. EAGE, Madrid (2007)
11. Dubrule, O.: Geostatistics for Seismic Data Integration in Earth Models. SEG/EAGE Distinguished Instructor Short Course Number 6, Tulsa (2003)
12. Fomel, S.: Applications of plane-wave destruction filters. Geophysics **67**, 1946–1960 (2002)
13. Fomel, S.: Seismic reflection data interpolation with differential offset and shot continuation. Geophysics **68**, 733–744 (2003). https://doi.org/10.1190/1.1567243
14. Gardner, G.H.F., Canning, A.: Effects of irregular sampling on 3-D prestack migration. In: 64th Annual International Meeting, SEG, Expanded Abstracts, pp. 1553–1556 (1994)
15. Goovaerts, P.: Geostatistics for Natural Resources Evaluation. Oxford University Press, New York (1997)
16. Guo, J., Zhou, X., Yang, H.J.: Efficient f-k domain seismic trace interpolation for spatially aliased data. In: 66th Annual International Meeting, SEG, Expanded Abstracts, pp. 1457–1460 (1996)
17. Gülünay, N.: Seismic trace interpolation in the Fourier transform domain. Geophysics **68**, 355–369 (2003). https://doi.org/10.1190/1.1543221
18. Gülünay, N., Chambers, R.E.: Unaliased f-k domain trace interpolation (UFKI). In: 66th Annual International Meeting, SEG, Expanded Abstracts, pp. 1461–1464 (1996)
19. Hennenfent, G., Herrmann, F.J.: Simply denoise: wavefield reconstruction via jittered undersampling. Geophysics **73**, V19–V28 (2008)
20. Herrmann, F.J., Hennenfent, G.: Non-parametric seismic data recovery with curvelet frames. Geophys. J. Int. **173**, 233–248 (2008)
21. Horta, A., Caeiro, M., Nunes, R., Soares, A.A.: Simulation of continuous variables at meander structures: application to contaminated sediments of a lagoon. In: Atkinson P, Lloyd C (eds.) geoENV VII—Geostatistics for Environmental Applications. Quantitative Geology and Geostatistics, pp. 161–172. Springer, The Netherlands (2009)
22. Jia, Y., Ma, J.: What can machine learning do for seismic data processing: an interpolation application? Geophysics **82**(3), V163–V177 (2017). https://doi.org/10.1190/GEO2016-0300.1

23. Kreimer, N., Stanton, A., Sacchi, M.D.: Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction. Geophysics **78**(6), 273–V284 (2013). https://doi.org/10.1190/geo2013-0022.1
24. Kumar, R., Mansour, H., Herrmann, F.J., Aravkin, A.Y.: Reconstruction of seismic wavefields via low-rank matrix factorization in the hierarchical-separable matrix representation. In: SEG Technical Program Expanded Abstracts, pp. 3628–3633 (2013)
25. Kumar, R., Da Silva, C., Akalin, O., Aravkin, A.Y., Mansour, H., Recht, B., Herrmann, F.J.: Efficient matrix completion for seismic data reconstruction. Geophysics **80**(5), V97–V113 (2015)
26. Liang, J., Ma, J., Zhang, X.: Seismic data restoration via data-driven tight frame. Geophysics **79**(3), V65–V74 (2014)
27. Liu, B., Sacchi, M.D.: Minimum weighted norm interpolation of seismic records. Geophysics **69**, 1560–1568 (2004). https://doi.org/10.1190/1.1836829
28. Liu, Y., Fomel, S.: Seismic data interpolation beyond aliasing using regularized nonstationary autoregression. Geophysics **76**, V69–V77 (2011)
29. Luis J.J., Almeida J.: Stochastic characterization of fluvial sand channels. In: E. Baafi, et al. (eds.) Geostatistics Wollogong 96, pp. 465–477, Kluwer Academic Publishers (1997)
30. Ma, J.: Three-dimensional irregular seismic data reconstruction via low-rank matrix completion. Geophysics **78**(5), V181–V192 (2013)
31. Marfurt, K.J.: Robust estimates of 3D reflector dip and azimuth. Geophysics, **71**(4). https://doi.org/10.1190/1.2213049 (2006)
32. Naghizadeh, M., Sacchi, M.D.: f-x adaptive seismic trace interpolation. Geophysics **74**(1), V9–V16 (2009). https://doi.org/10.1190/1.3008547
33. Naghizadeh, M., Sacchi, M.D.: Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data. Geophysics **75**, WB189–WB202 (2010)
34. Naghizadeh, M., Sacchi, M.D.: Hierarchical scale curvelet interpolation of aliased seismic data. In: 80th Annual International Meeting, SEG, Expanded Abstracts, pp. 3656–3661 (2010)
35. Naghizadeh, M., Innanen, K.A.: Seismic data interpolation using a fast generalized Fourier transform. Geophysics **76**(1), V1–V10 (2011)
36. Nguyen, T., Winnett, R.: Seismic interpolation by optimally matched Fourier components. In: SEG Technical Program Expanded Abstracts, pp. 3085–3089 (2011)
37. Oporeza, V., Sacchi, M.D.: Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. Geophysics **76**, V25–V32 (2011)
38. Porsani, M.J.: Seismic trace interpolation using half-step prediction filters. Geophysics **64**, 1461–1467 (1999)
39. Ronen, J.: Wave-equation trace interpolation. Geophysics **52**, 973–984 (1987)
40. Sacchi, M.D., Ulrych, T.J.: Estimation of the discrete Fourier Transform, a linear inversion approach. Geophysics **61**, 1128–1136 (1996)
41. Sacchi, M.D., Ulrych, T.J., Walker, C.J.: Interpolation and extrapolation using a high-resolution discrete Fourier transform. IEEE Trans. Signal Process. **46**, 31–38 (1998)
42. Soares, A.: Geostatistical estimation of orebody geometry: morphological kriging. Math. Geol. **22**(7), 787–802 (1990)
43. Soares, A.: Direct sequential simulation and co-simulation. Math. Geol. **33**(8), 911–926 (2001)
44. Spitz, S.: Seismic trace interpolation in the F-X domain. Geophysics **56**, 785–794 (1991). https://doi.org/10.1190/1.1443096
45. Stolt, R.H.: Seismic data mapping and reconstruction. Geophysics **67**, 890–908 (2002). https://doi.org/10.1190/1.1484532
46. Stroet, C., Judith, J.: Mapping curvilinear structures with local anisotropy kriging. Math. Geol. **37**(6) (2005)
47. Trad, D., Ulrych, T.J., Sacchi, M.D.: Accurate interpolation with high-resolution time-variant Radon transforms. Geophysics **67**(2), 644–656 (2002)
48. Trad, D.: Interpolation and multiple attenuation with migration operators. Geophysics **68**, 2043–2054 (2003). https://doi.org/10.1190/1.1635058
49. Trad, D.: Five-dimensional interpolation: recovering from acquisition constraints. Geophysics **74**(6), V123–V132 (2009). https://doi.org/10.1190/1.3245216
50. Trickett, S.R.: F-xy Eigenimage noise suppression. Geophysics **68**, 751–759 (2003)
51. Trickett, S., Burroughs, L., Milton, A., Walton, L., Dack, R.: Rank-reduction-based trace interpolation. In: SEG Technical Program Expanded Abstracts, pp. 3829–3833 (2010)
52. Turquais, P., Asgedom, E.G., Söllner, W.: Structured dictionary learning for interpolation of aliased seismic data. In: 87th Annual Meeting, SEG, pp. 4257–4261 (2017)
53. Van Dedem, E.J., Verschuur, D.J.: 3-D surface-related multiple elimination and interpolation. In: 68th Annual International Meeting, SEG, pp. 1321–1324 (1998)
54. Wang, Y.: Seismic trace interpolation in the f-x-y domain. Geophysics **67**, 1232–1239 (2002)
55. Wang, J., Ng, M., Perz, M.: Fast high resolution Radon transforms by greedy least-squares methods. In: SEG Expanded Abstracts, vol. 28, pp. 3128–3132 (2009)
56. Xu, W.: Conditional curvilinear stochastic simulation using pixel-based algorithms. Math. Geol. **28**(7), 937–949 (1996)
57. Yang, S., Ma, J., Osher, S.: Seismic data reconstruction via matrix completion. In: UCLA CAM, pp. 12–14 (2012)