



A residual dictionary learning method for footprint removal from seismic data

Julián L. Gómez* (UNLP, CONICET) and Danilo R. Velis (UNLP, CONICET)

Copyright 2019, SBGf - Sociedade Brasileira de Geofísica

This paper was prepared for presentation during the 16th International Congress of the Brazilian Geophysical Society held in Rio de Janeiro, Brazil, 19-22 August 2019.

Contents of this paper were reviewed by the Technical Committee of the 16th International Congress of the Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of the Brazilian Geophysical Society is prohibited.

Abstract

We introduce a novel dictionary learning strategy for removal of footprint patterns and random noise in seismic data. To this end, we construct an augmented dictionary based solely on the atoms learned from the coherence-constrained dictionary learning (CDL), a method that is very effective on attenuating random noise. It turns out that when seismic data is contaminated with acquisition and/or processing footprint, the atoms of the learned dictionary are contaminated by coherent noise patterns. Hence, it is necessary to carry out a morphological and/or texture attribute classification of the atoms for effective footprint removal. Instead, the method that we propose relies on an augmented dictionary that is constructed using a simple data-driven empirical mode decomposition (EMD) algorithm, which leads to a dictionary that contains signal atoms and a residual dictionary that contains footprint atoms. This avoids the use of complex statistical classifications strategies to segregate the atoms of the learned dictionary. As in CDL, the proposed method does not require the user to know or adjust the noise level or the sparsity of the solution for each data set. Further, it only requires one pass of CDL dictionary learning and is shown to produce successful transfer learning results in field data. This leads to a speed-up of the denoising processing, since random and coherent noise can be removed without calculating the augmented dictionary for each time slice of the 3D data volume. Results on field data demonstrate effective footprint removal with accurate edge preservation on time slices of 3D seismic poststack data.

Introduction

The removal of unwanted noise is a crucial step for regularization, processing and interpretation of seismic data. A proper denoising facilitates the workflow that ultimately leads to decision-making and reservoir evaluation in the industry. For these reasons, the preservation of features of geological interest plays a crucial step in the process of noise removal (Fehmers and Höcker, 2003). Coherent noise patterns arise from limitations on acquisition design, acquisition equipment, and the processing workflow applied to the recorded seismic data. These patterns, known as *acquisition footprint*, have a undesired effect in seismic analysis, most noticeable by reducing the level of confidence on the identification of geological structures from co-

herence attributes (Marfurt et al., 1998). Minimizing the impact of acquisition footprint in the analysis of seismic data is therefore of interest (Alali et al., 2018).

Footprint attenuation is commonly attempted in the frequency/wavenumber domain. Al-Bannagi et al. (2004) adapt the truncated singular value decomposition algorithm to suppress random noise and footprint related to seismic acquisition. Falconer and Marfurt (2008) use wavenumber filtering followed by adaptive subtraction to detect, enhance and properly remove footprint patterns from seismic time slices. Chopra and Larsen (2000) applies a two-pass frequency-wavenumber filtering of individual time slices from 3D seismic data to suppress acquisition footprint. Drummond et al. (2000) removes periodic striations due to coherent noise leakage from 3D seismic data by wavenumber notch filtering of each time slice. Soubaras (2002) presents a procedure to design a wavenumber filter to attenuate the acquisition footprint of seismic data due to acquisition geometry. Zhang (2009) performs footprint removal with basis pursuit using Daubechies wavelets with increased symmetry. To be successful in removing the noise, most of these techniques require to carry out a quality-control (QC) of the applied muting filter prior to subtraction, and to manually repeat this QC for every time slice.

Techniques of dictionary learning (DL) borrowed from the field of sparse and redundant representation of signals (Mallat, 1999; Elad, 2010) provide several methods to tackle noise removal in seismic data (Beckouche and Ma, 2014; Turquais et al., 2017b, 2018; Zu et al., 2019). A dictionary is a collection of vectors known as atoms that represent elementary patterns of the data. The objective is to represent the data very closely using linear combinations of a few atoms. The learning process of DL involves an optimization scheme that alternatively calculates a sparse representation of the data and updates the dictionary.

A two-pass DL strategy is applied by Beckouche and Ma (2014) to denoise seismic data. In the first pass, learning and denoising is achieved on the original data for an assumed noise variance. In the second pass, using a smaller noise variance, the whole process of learning and denoising is applied to the previous processed data to yield the final result. This DL method yields results which can be superior to wavelets, curvelets, and total variation alternatives. Turquais et al. (2017b) introduces the coherence-constrained dictionary learning (CDL) for random noise denoising of seismic data. The CDL can be applied to data with noise of variable variance, and does not require the interpreter to set thresholds for the noise variance nor the sparsity of the data representation, another key parameter required by most DL methods. In order to remove coherent noise, atom classification strategies are introduced by Turquais et al. (2017a) to differentiate between struc-

tured noise and seismic signal in marine data. The training and denoising in DL is done on patches selected from the seismic input. Different patching strategies can be implemented in order to accelerate the learning process (Zu et al., 2019) and improve random noise removal (Yu et al., 2016).

With the motivation to reduce the imprint of footprint from seismic data using a DL strategy, we propose a novel adaptive DL algorithm for the removal of random and coherent noise in seismic data. To this end, we introduce the residual dictionary denoising (RDD) algorithm. We show that RDD allows to remove random noise and footprint signatures from field data. Also, we demonstrate that the proposed method is capable of transfer learning, where the augmented dictionary that is learned in a given time slice can be applied, automatically, for noise attenuation to other time slices. The method does not involve more than one cycle of DL to denoise an entire seismic subset. By using an empirical mode decomposition (EMD) algorithm (Huang et al., 1998), very competitive results, with respect to strategies based on source separation and atom classification are shown. The method we describe can be viewed as a data-driven source separation algorithm.

Background theory

The basic idea of the method to be described stems from the observation that, although the removal of random noise is achieved very successfully by CDL, the learned dictionary is often a mixture of data and coherent noise patterns. The EMD step is an alternative way to remove the overlapping textures from each vector in the learned dictionary, from which an augmented dictionary can then be defined. The augmented dictionary is applied to remove not only random, but also coherent noise in an automatic, data-driven manner.

Following Mallat (1999), the coherence of a vector \mathbf{r} relative to a dictionary \mathbf{D} of K atoms is defined by

$$\mu(\mathbf{r}, \mathbf{D}) = \max_j \left| \left\langle \frac{\mathbf{r}}{\|\mathbf{r}\|}, \mathbf{d}_j \right\rangle \right|, \quad (1)$$

where $\langle \cdot \rangle$ means inner product, \mathbf{d}_j is the j -th atom of the dictionary, and $1 \leq j \leq K$. A coherence matching pursuit denoising decomposes a signal as long as the coherence of the residue (difference between the recording \mathbf{z} and the sparse approximation $\mathbf{D}\hat{\mathbf{x}}$) is above $\hat{\mu}$ and stops when $\mu(\mathbf{r}, \mathbf{D}) < \hat{\mu}$ is satisfied, for a given coherence threshold $\hat{\mu}$. At this point, the residue is considered to be uncorrelated with respect to any $\mathbf{d}_j \in \mathbf{D}$. The recording \mathbf{z} is a 2D patch of size $\sqrt{N} \times \sqrt{N}$ taken from the seismic image which is vectorized into a $N \times 1$ column vector.

In this context, a coherence-constrained dictionary learning scheme (Turquais et al., 2017b) can be obtained by solving the following constrained optimization problem:

$$\{\hat{\mathbf{X}}, \hat{\mathbf{D}}\} = \arg \min_{\mathbf{X}, \mathbf{D}} \|\mathbf{X}\|_0 \text{ s.t. } \min_{\mathbf{D}} \|\mathbf{R}\|_F^2 \text{ and } \mu(\mathbf{R}, \mathbf{D}) \leq \hat{\mu}, \quad (2)$$

where $\mathbf{R} = \mathbf{Z} - \mathbf{D}\mathbf{X}$ is the residue matrix, whose columns are given by $\mathbf{R}_i = \mathbf{z}_i - \mathbf{D}\mathbf{x}_i$, $\|\cdot\|_F$ is the Frobenius norm,

and $\mu(\mathbf{R}, \mathbf{D})$ is a short-hand to denote $\mu(\mathbf{z}_i - \mathbf{D}\mathbf{x}_i, \mathbf{D})$ for $i = 1, \dots, M$. The problem posed by equation 2 is numerically approximated by an alternating two-step process. The alternating optimization scheme entails a *sparse coding step* (where orthogonal matching pursuit can be used to find the sparse coefficients $\hat{\mathbf{x}}_i$), and a *dictionary update step* (where approximate k-SVD can be applied to find a new dictionary). Please, refer to Turquais et al. (2017b) for details.

For Gaussian noise, Turquais et al. (2017b) demonstrates that

$$\hat{\mu} = \sqrt{2 \log K/N}. \quad (3)$$

This important result means that the coherence constrained dictionary learning allows for random noise attenuation without the need to know data priors such as the noise variance or the cardinality of the sparse approximation $\hat{\mathbf{x}}$.

The empirical mode decomposition (EMD) proposed by Huang et al. (1998) decomposes a given 1D signal \mathbf{d} into a series of intrinsic mode functions (IMFs) plus a residue,

$$\mathbf{d} = \sum_{i=1}^{i=l} \text{IMF}^i + \mathbf{R}_d. \quad (4)$$

The IMFs are amplitude- and frequency-modulated signal components, which satisfy the conditions of having zero mean and a number of extrema equal (or different at most by one) to the number of zero crossings (Huang et al., 1998). The number i of IMFs used to decompose the input signal depends on the data itself. The EMD stands out as a complete data-driven strategy to separate 1D signal into separate spectral bands. In particular, the first intrinsic mode, IMF^1 , is usually removed to perform denoising because it is the most oscillating component of the signal (Battista et al., 2007; Macelloni et al., 2011). The first IMF is obtained by an iterative process called *sifting*, a process that requires the calculation of the mean envelope from the extrema points of the input signal, usually by means of spline interpolation. Then, this envelope is subtracted from the signal to form the residue. The sifting process is repeated until the residue satisfies the conditions of being an IMF. At this point, the resulting residue is regarded to be the first IMF of the data. Subsequent IMFs are derived by subtracting the previous IMF from the original signal \mathbf{d} and repeating the iterative sifting process just described. The sifting iterations can be stopped, for example, after a number of iterations is reached. The EMD algorithm we apply in the tests is based on the sifting strategy proposed by He et al. (2017), which is a recent and fast approach to EMD using separable operators and does not involve splines interpolation.

Method

Given the input time slice \mathbf{y} of size $N_x \times N_y$, we consider all the patches \mathbf{z} as columns of the matrix \mathbf{Y} . Both the initial dictionary \mathbf{D}_0 and the training set \mathbf{Y}_M to learn the dictionary \mathbf{D} can be determined by selecting $M + K$ random columns from \mathbf{Y} ; M for the training set, and K for the initial dictionary.

Other possibilities do exist to select the training data set and the initial dictionary (Yu et al., 2016). After CDL, we filter each atom of the learned dictionary \mathbf{D} by EMD removing its first intrinsic mode function (IMF¹),

$$\mathbf{d}_j^{\text{EMD}} = \mathbf{d}_j - \text{IMF}_j^1, \quad (5)$$

which leads to the filtered dictionary \mathbf{D}_{EMD} . Having obtained the filtered dictionary \mathbf{D}_{EMD} , we obtain the residual dictionary $\mathbf{D}_r = \mathbf{D} - \mathbf{D}_{\text{EMD}}$. With these dictionary components, we build the augmented dictionary $\mathbf{D}_f = \mathbf{D}_{\text{EMD}} \cup \mathbf{D}_r$, which has $2K$ atoms. Having set \mathbf{D}_f , the sparse representation of each patch is obtained, producing the matrix \mathbf{X} where each column contains the sparse representation of each column of \mathbf{Y} . Finally, the original data is reconstructed using only the atoms not related to footprint signatures. To that end, once the sparse coefficients \mathbf{X} are obtained by means of a greedy algorithm from \mathbf{D}_f , the removal of random and coherent noise is given by

$$\hat{\mathbf{Y}} = \mathbf{D}_{\text{EMD}} \mathbf{X}[:, K, :], \quad (6)$$

where $\mathbf{X}[:, K, :]$ means taking the first K coefficients of each column in \mathbf{X} . Having obtained the sparse coefficients using \mathbf{D}_f allows for a proper separation of the energy of the seismic structures, from the energy of the coherent and undesired noise. Note that for augmenting the dictionary and performing the denoising of equation 6, no dictionary update is required. The matrix $\hat{\mathbf{Y}}$ is finally assembled to yield the denoised image \hat{y} . In practice, we reconstruct the footprint (we use the atoms from the residual dictionary \mathbf{D}_r) and subtract this result to the input data.

The augmented dictionary learned from a given time slice can be applied to their neighbor time slices with successful results, in what is known as *transfer learning* by the machine learning community (Goodfellow et al., 2016). The transfer learning gives the opportunity to exploit the obtained augmented dictionary, avoiding the need to carry out a QC of the results for each time slice in order to tune the denoising filter being applied.

The workflow of the proposed method can be summarized as follows:

1. Take all patches from the seismic input y to obtain \mathbf{Y} .
2. Select random columns from \mathbf{Y} to obtain the training set \mathbf{Y}_M and the initial dictionary \mathbf{D}_0 .
3. Apply CDL to learn the dictionary \mathbf{D} .
4. Filter each atom of \mathbf{D} with EMD to yield \mathbf{D}_{EMD} .
5. Augment the dictionary: $\mathbf{D}_f = \mathbf{D}_{\text{EMD}} \cup (\mathbf{D} - \mathbf{D}_{\text{EMD}})$.
6. Obtain the sparse coefficients \mathbf{X} using \mathbf{D}_f .
7. Obtain $\hat{\mathbf{Y}}$ by denoising with \mathbf{D}_{EMD} and the first K coefficients of \mathbf{X} .
8. Assemble the patches of $\hat{\mathbf{Y}}$ to obtain the denoised result \hat{y} .

Contrary to Turquais et al. (2017a), the proposed technique does not require to investigate appropriate texture attributes to perform atom separation. As to any DL algorithm, the proposed method is very flexible and can be

adapted to numerous strategies. For example, the augmented dictionary associated with every time slice can be used to define an initial dictionary to perform CDL and source separation. A degree of mixture between atoms can be included by filtering with more than the K initial atoms. Also, a dictionary based solely on IMFs of the atoms can be used to initialize CDL and the source separation. Although we focus the proposed scheme in processing time slices, the filtering can also be applied in 3D patches of the seismic volumes, or in the inline and crossline directions.

Results

We illustrate the method by means of a field data example. To this end we select the Penobscot (Nova Scotia Department of Energy, 1992) poststack subset from Kington (2015), since it offers the unique opportunity of a public domain data set that contains both random noise and footprint patterns. In particular, the footprint patterns of this subset, which are more evident on the shallow time slices, consist on a series of stripes of variable length that follow the crossline direction, as shown in Figure 4a.

To perform CDL, the atom size is kept to a minimum of $N = 9 \times 9$ samples, a number of $K = N$ atoms is selected, 2% of the patches are taken for training, and the dictionary update is iterated 10 times. We run the atom filtering using 5 sifting iterations of EMD and remove the first IMF from each atom, which can be regarded as the default settings for filtering the learned dictionary. Other parameter settings can improve the results presented herein.

The dictionary learned by CDL for time slice $t = 0.116$ s is shown in Figure 2a. Since the dictionary is learned on data fully contaminated by coherent noise, it is a mixture of atoms representing both signal and footprint. Figure 2b shows the filtered dictionary by EMD filtering, and Figure 2c the residual dictionary. A set of atoms before and after applying EMD can be appreciated in Figure 1. The energy from the coherent noisy pattern is strongly attenuated when taking into account the filtered and residual dictionaries together.

Figure 3 displays the results of using the dictionaries \mathbf{D} and \mathbf{D}_{EMD} separately. In the first case, we clearly observe that while random noise is attenuated satisfactorily, footprint signatures are enhanced (Figure 3a). On the other hand, edges are blurred by using \mathbf{D}_{EMD} only (Figure 3c). In both cases, the residuals show the significant signal leakage and details that are lost in the processing.

Figure 4 shows the results of using the augmented dictionary \mathbf{D}_f . We can observe that for this data set, both random noise and coherent noisy patterns were successfully attenuated with edge preservation. To yield this result, a minimum user intervention was required. Namely, setting the parameters for CDL and the number of sifting iterations. Figure 5 shows the impact of the denoising on the Sobel magnitude filter. The uncertainty produced by the coherent noise has been greatly reduced, simplifying the identification and delineation of seismic structures.

Figure 6 illustrates the transfer learning capabilities of the method. In the example, the augmented dictionary for

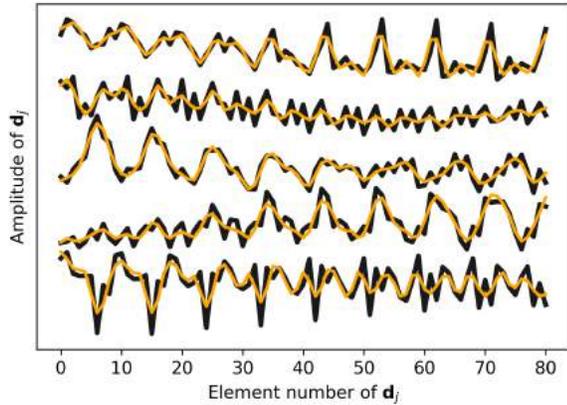


Figure 1 – Original atoms from the learned dictionary (black). Filtered atoms after removal of the first IMF (orange).

$t = 0.116$ s is applied for the time slice $t = 0.136$ s. The result is encouraging, since it proves that the method effectively reduces noise for data not from the same time slice from where the augmented dictionary was derived. This result is useful to save processing time in industrial applications, since the dictionary does not need to be learned for each time slice of a seismic 3D data to produce acceptable results.

Conclusions

We presented an alternative algorithm to perform dictionary learning for random and coherent noise removal in seismic data. The method is an alternative to the standard footprint removal in the frequency/wavenumber domain, which is widely applied in the industry. Our proposed technique uses coherence-constrained dictionary learning coupled with empirical mode decomposition. The dictionary learning provides the resource to learn a set of atoms which are a mixture of seismic structures and coherent acquisition footprint patterns which are capable of filtering out random noise. The filtering of each atom in the learned dictionary leads to an augmented set of atoms that comprises the filtered dictionary and the residual dictionary. The atoms in the residual dictionary can represent the footprint and allow for automated source separation. The introduced method does not require to perform atom selection to separate coherent noise from seismic amplitudes. Its transfer learning capabilities automates and simplifies the denoising process. Results on seismic field data are encouraging, showing the benefits of the proposed residual dictionary learning technique.

Acknowledgements

Partial support was granted by *Programa de Incentivos, Universidad Nacional de La Plata (UNLP)* and *Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, PIP 112-201201-00626-CO)*.

References

Aharon, Michal and Elad, Michael and Bruckstein, Alfred, 2006, K-SVD: An algorithm for designing overcom-

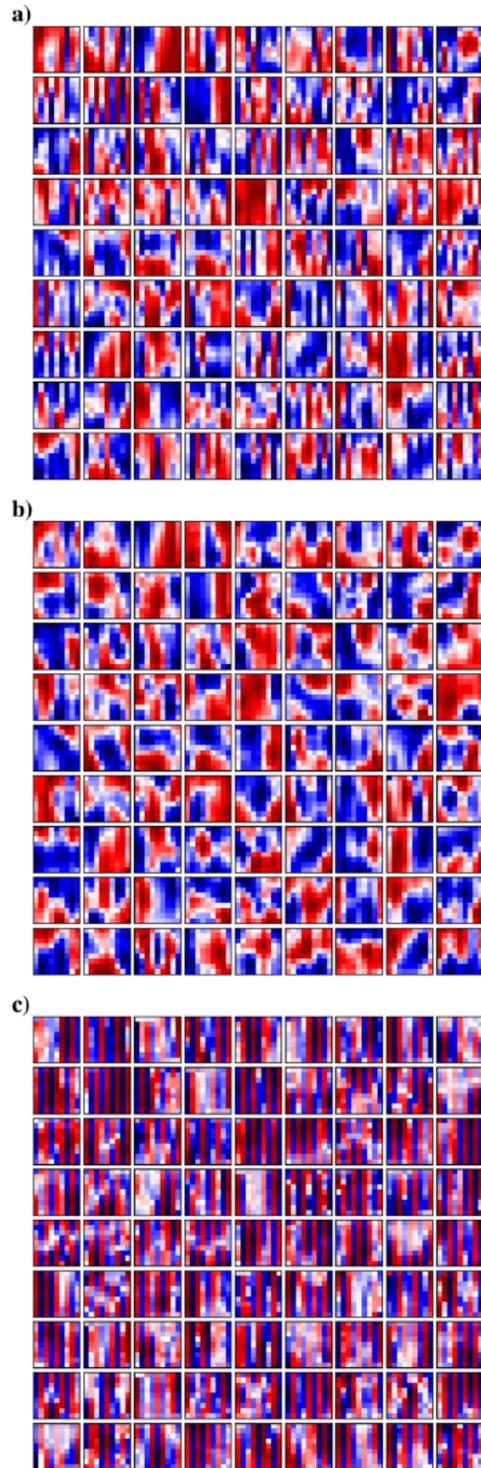


Figure 2 – a) Learned dictionary D from the time slice by CDL. The dictionary is a mixture of seismic signal and footprint patterns. The augmented dictionary D_f consists of b) the EMD filtered dictionary D_{EMD} , and c) the residual dictionary D_r . The residual dictionary will represent the footprint structures of the data.

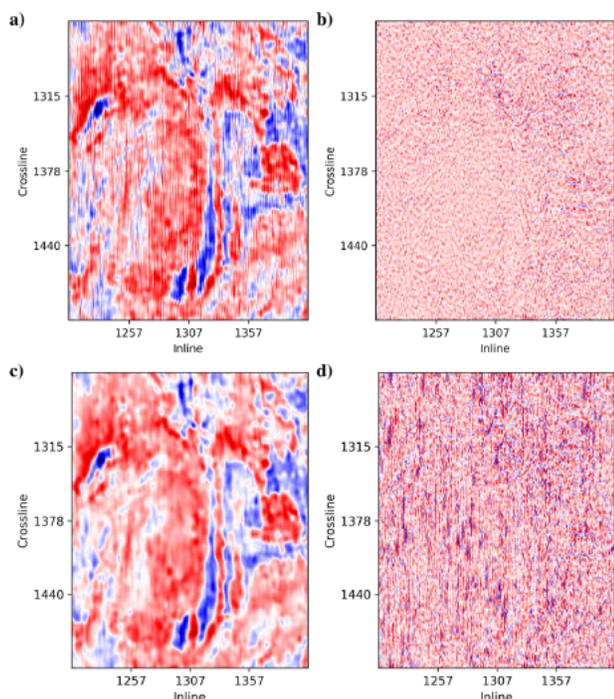


Figure 3 – a) Filtered time slice with the learned dictionary D . b) Data residual. c) Filtered time slice with the EMD filtered dictionary D_{EMD} . d) Data residual. None of these dictionaries, when used separately, can attenuate the coherent amplitudes due to footprint.

plete dictionaries for sparse representation: IEEE Transactions on signal processing, 54, 4311–4322.

M. S. Al-Bannagi and K. Fang and P. G. Kelamis and G. S. Douglass, 2004, Acquisition footprint suppression via the truncated SVD technique: SEG Technical Program Expanded Abstracts, 1957–1960.

Abdulmohsen Alali and Gabriel Machado and Kurt J. Marfurt, 2018, Attribute-assisted footprint suppression using a 2D continuous wavelet transform: Interpretation, 6, T457–T470.

Battista, Bradley Matthew and Knapp, Camelia and McGee, Tom and Goebel, Vaughn, 2007, Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data: Geophysics, 72, H29–H37.

Simon Beckouche and Jianwei Ma, 2014, Simultaneous dictionary learning and denoising for seismic data: Geophysics, 79, A27–A31.

Chopra, Satinder and Larsen, Glen, 2000, Acquisition footprint—its detection and removal: CSEG Recorder, 25, 16–20.

J. M. Drummond and Arthur J. L. Budd and James W. Ryan, 2000, Adapting to noisy 3D data—attenuating the acquisition footprint: SEG Technical Program Expanded Abstracts, 9–12.

Elad, Michael, 2010, Sparse and redundant representations: from theory to applications in signal and image processing: Springer Science & Business Media.

Scott Falconer and Kurt J. Marfurt, 2008, Attribute-driven footprint suppression: SEG Technical Program Expanded

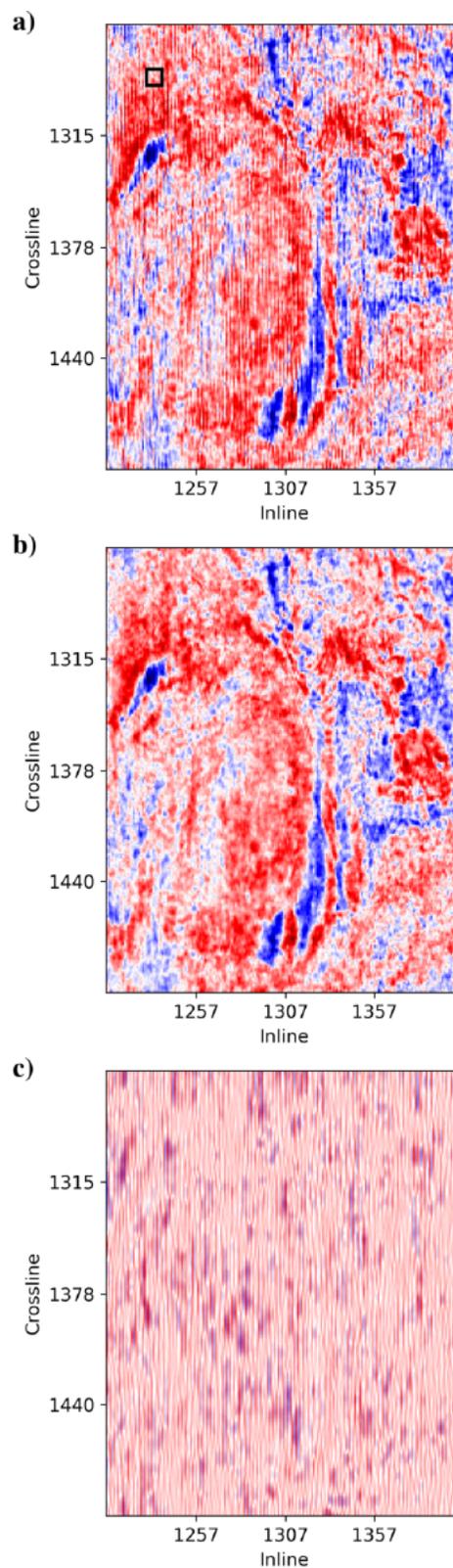


Figure 4 – a) A time slice of a poststack 3D seismic amplitude data displaying random noise and footprint patterns. The black rectangle shows the patch size. b) Filtered time slice with the augmented dictionary D_f . c) Data residual.

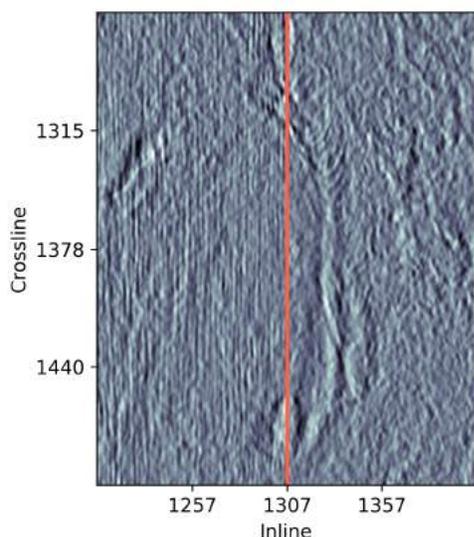


Figure 5 – Sobel magnitude attribute of the field data timeslice. Left to the red line, attribute on the original data. Right to the red line, attribute on the processed data.

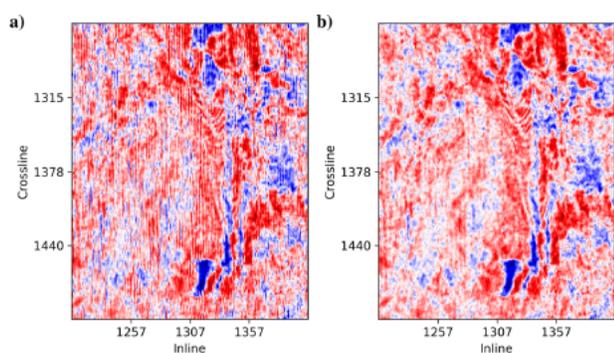


Figure 6 – a) Time slice ($t = 0.136$ s) from the poststack data set. b) Filtered time slice with transfer learning (the dictionaries were obtained for $t = 0.116$ s).

Abstracts, 2667–2671.

Gijs C. Fehmers and Christian F. W. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: *Geophysics*, 68, 1286–1293.

Goodfellow, Ian and Bengio, Yoshua and Courville, Aaron, 2016, *Deep learning*: MIT press.

He, Zhi and Li, Jun and Liu, Lin and Shen, Yi, 2017, Three-dimensional empirical mode decomposition (TEMD): A fast approach motivated by separable filters: *Signal Processing*, 131, 307–319.

Huang, Norden E. and Shen, Zheng and Long, Steven R. and Wu, Manli C. and Shih, Hsing H. and Zheng, Quan-an and Yen, Nai-Chyuan and Tung, Chi Chao and Liu, Henry H., 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454, 903–995.

Joe Kington, 2015, Semblance, coherence, and other discontinuity attributes: *The Leading Edge*, 34, 1510–1512.

Macelloni, L. and Battista, B. M. and Knapp, C.C., 2011, Optimal filtering high-resolution seismic reflection data using a weighted-mode empirical mode decomposition operator: *Journal of Applied Geophysics*, 75, 603–614.

Mallat, Stephane, 1999, *A wavelet tour of signal processing: The sparse way*, 3rd ed.: Academic Press.

Kurt J. Marfurt and Ronald M. Scheet and John A. Sharp and Mark G. Harper, 1998, Suppression of the acquisition footprint for seismic sequence attribute mapping: *Geophysics*, 63, 1024–1035.

Nova Scotia Department of Energy, 1992, SEG Wiki: Penobscot 3D, http://wiki.seg.org/wiki/penobscot_3D, accessed 5 January 2018.

Robert Soubaras, 2002, Attenuation of acquisition footprint for non-orthogonal 3D geometries: *SEG Technical Program Expanded Abstracts*, 2142–2145.

Pierre Turquais and Endrias G. Asgedom and Walter Söllner, 2017a, Coherent noise suppression by learning and analyzing the morphology of the data: *Geophysics*, 82, V397–V411.

—, 2017b, A method of combining coherence-constrained sparse coding and dictionary learning for denoising: *Geophysics*, 82, V137–V148.

Pierre Turquais and Endrias G. Asgedom and Walter Söllner and Leiv Gelius, 2018, Parabolic dictionary learning for seismic wavefield reconstruction across the streamers: *Geophysics*, 83, V263–V282.

Siwei Yu and Jianwei Ma and Stanley Osher, 2016, Monte Carlo data-driven tight frame for seismic data recovery: *Geophysics*, 81, V327–V340.

Rui Zhang, 2009, Footprint suppression with basis pursuit denoising: *SEG Technical Program Expanded Abstracts*, 3361–3366.

Shaohuan Zu and Hui Zhou and Rushan Wu and Mao-cai Jiang and Yangkang Chen, 2019, Dictionary learning based on dip patch selection training for random noise attenuation: *Geophysics*, 84, V1–V15.