Fast empirical seismic denoising

Julián L. Gómez* (UNLP/CONICET), Danilo R. Velis (UNLP/CONICET), and Marcelo Roizman (GEONODOS)

Abstract

In this work we present a new strategy to accelerate the computation of the so-called complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN), a data-driven technique that can be used to denoise seismic data. The new implementation replaces the use of the cubic interpolation scheme, which is required to calculate the signal and residual envelopes, by a simple window averaging. In addition, and due to the fact that the energy distribution of the modes acquires a different behavior, the new strategy facilitates the mode selection needed to carry out the denoising. As a consequence, the low and high frequency content of the signals can be efficiently and effectively isolated in a fraction of the time consumed by the standard CEEMDAN approach for the same amount of noise attenuation. We compare the performance of both approaches and apply them to successfully denoise various microseismic and seismic reflection field records.

Introduction

The CEEMDAN (also known as CEEMD) is a data-driven and noise-assisted algorithm proposed by Torres et al., 2011, that decomposes an input signal s_n of length N into a finite set of K intrinsic mode functions (IMFs) or modes and a unique residue R_n. The CEEMDAN is a improvement over the ensemble empirical mode decomposition EEMD of Wu et al. (2009), which in turn is based on the empirical mode decomposition EMD of Huang et al. (1998). The CEEMDAN decomposition can be described by

$$s_n = \sum_{k=1}^{K} IMF_n^k + R_n \quad . \tag{1}$$

The modes IMF^k_n, where subscript $k = 1,...,K$ denotes the order, have zero mean and a number of extrema equal or different at most by one to the number of zero crossings (Huang et al., 1998). Each k-th mode is obtained in a sequential order by the algorithm and they exhibit a decreasing frequency content; being the high frequencies captured by the first modes and the low frequencies by the final modes.

The CEEMDAN algorithm requires, a sifting stoppage criterion, the number of noise realizations and the injected noise amplitude to initiate the decomposition of the signal s_n by averaging over its noise-assisted realizations. The necessary steps to implement this empirical decomposition are provided in Torres et al. (2011). The sifting is crucial in the CEEMDAN decomposition since its purpose it to obtain the modes from the input signal. The sifting aims to calculate and subtract the mean envelope from the signal and its residues until a stoppage criterion, which is related to the IMF condition of zero local mean, is met. A simple and fast stoppage criterion is to sift a low but fixed number of times (Wu et al., 2009). To improve the cancellation property of the injected white noise realizations, following Lin 2012, the noise realizations consist of pairs of opposite sign. This implies that half of the noise realizations have to be actually calculated, being also an improvement over the original algorithm speed.

The standard seismic empirical denoising consists in the subtraction of the low order IMFs from the input signal; the denoised counterpart s^d_n of each seismic trace s_n is obtained then by

$$s_n^d = s_n - \sum_{k=1}^{M_1-1} IMF_n^k \quad . \tag{2}$$

The number M_1 in principle is a user-defined parameter, usually it is $M_1 = 2$ for most applications.

A refinement of the standard denoising consists in removing also high order modes from the signal,

$$s_n^d = s_n - \sum_{k=1}^{M_1-1} IMF_n^k - \sum_{k=M_2}^{K} IMF_n^k \quad . \tag{3}$$

The parameter M_2 can be selected by inspection of the distribution of the energy of each decomposed mode. As a measure of the energy of the IMF of order k, we propose the following normalized median absolute deviation:

$$E_k = median\left(|IMF_n^k|\right) / max\left(E_k\right) \quad , \tag{4}$$

where $max(E_k)$ means to divide by the maximum of all the obtained mode energies. Once an energy map is constructed, a detection of the noise-related energies at higher modes is necessary to appropriately set the value of M_2 for each seismic trace.

Fast CEEMDAN

The CEEMDAN obtains the mean envelope required for the sifting process as the average of the upper and lower envelopes of the data. These envelopes are computed by natural spline interpolation of the maxima and the minima of the signal and its residues. When faced with an important seismic data volume, this way of obtaining the

envelopes turns the CEEMDAN denoising a very time-consuming process (Battista et al., 2007).

To speed up the denoising process while keeping the implementation as simple as the original CEEMDAN, we propose to use a simple window average to obtain directly the mean local envelope required in the sifting loop. The window average is calculated by convolving the signal with a normalized Hanning window of length M_w. This window averaging is devised to denoise and not for interpretation of the individual modes. The window length can be obtained by estimating an effective period T from the average of the number of time samples between all consecutive local maxima and minima of each seismic trace.Then,

$$M_w = C\,T \quad , \qquad (4)$$

where C is a user-defined variable that accounts for the number of periods spanned by the window. We also choose as a final stage the average of the resulting M_w of each trace in order to have one parameter for the whole seismic dataset.

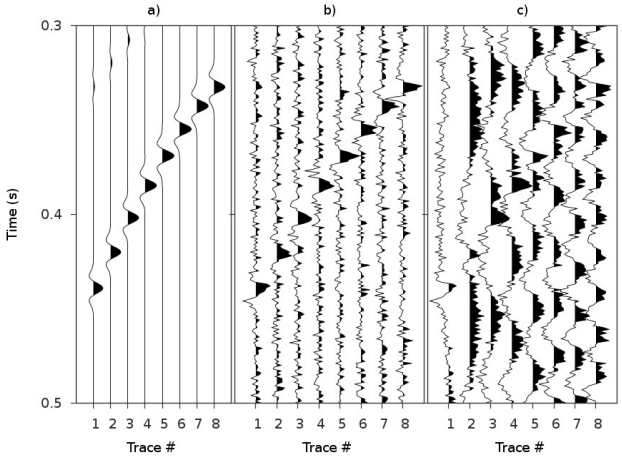

Figure 1: Synthetic microseismic examples. a) Noise-free signal. b) Signal with random noise and c) signal with random and low-frequency noise.

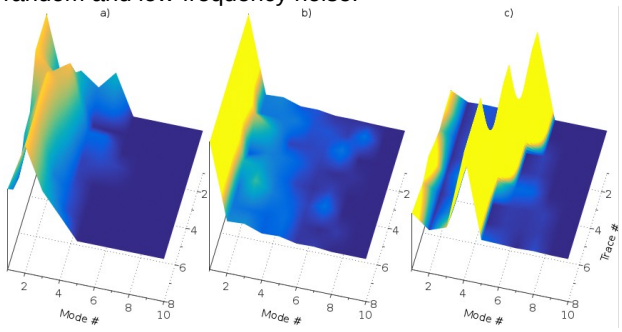

Figure 2: Energy mode map of the synthetic microseismic traces. a) Noise-free data, b) data with white noise, c) data with white and low frequency noise.The vertical scale is blue for 0 and yellow for 1.

Our proposed methodology has two advantages over the standard empirical denoising. First, the removal of random noise, that is obtained with a low value of the parameter C *and* $M_1 = 2$, is computed in a fraction of the time required by the CEEMDAN. The drop in computational time is most noticeable when the length of the data input is large. The key factor that controls this improvement resides in the convolution that replaces the interpolation scheme in the sifting process. Second, by setting a large value for C while keeping $M_1 = 2$, low frequency signal attenuation is possible without the need to estimate the M_2 parameter in equation (3). This is due to the fact that the window averaging always concentrates the energy in the first modes. The standard CEEMDAN would require an energy-based criterion for the selection and location of the higher order IMFs to estimate M_2. This means that a full decomposition in several modes for each trace has to be performed. This full mode calculation makes the removal of low-frequency content a much slower process that the fast CEEMDAN alternative. Furthermore, the criterion to detect the noisy modes may not be easily automated, requiring a visual interpretation by the user.

We consider a synthetic example to illustrate the two previous advantages. Figure 1a shows a noise-free 8-channel synthetic vertical microseismic gather. In Figure 1b, white noise is added to the data. In Figure 1c, white and low-frequency noise are both present. The mode energy of this examples is calculated for the CEEMDAN decomposition (Figure 2). The noise-free data exhibit more or less an evenly distributed energy in three consecutive modes (Figure 2a). When white noise contaminates the data, the energy shows strong amplitudes mainly concentrated at the first mode (Figure 2b). The removal of the first mode then consists in the standard empirical denoising with CEEMDAN (Figure 3a). When low frequency noise is also injected, an energy trend appears at the higher modes alongside with the random noise at the first mode (Figure 2c). To remove these low frequency components, a localization and removal of the energetic higher modes is necessary. When the energy map is taken into account in the standard CEEMDAN (Figure 3c), the noise attenuation is improved with respect (Figure 3b), but the processing is much slower. To achieve a similar result to Figure 3c, the fast CEEMDAN denoising with combined C is applied (Figure 4c). The partial outputs with C = 5 and C = 10 are shown respectively in Figures 4a and 4b. The combined result consists in subtracting the outputs of different C from the data. No energy map in necessary with the fast method and the processed signal is obtained at least ten times faster than the standard empirical denoising.

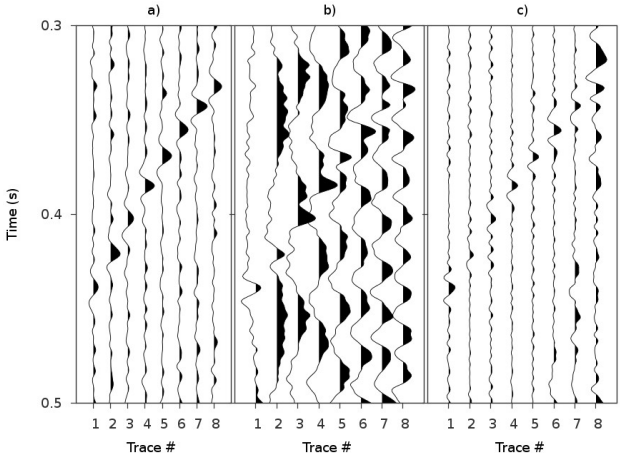

Figure 3: Standard CEEMDAN denoising: a) result from Figure 1b; b) and c) results from Figure 1c without and with energy-based criterion respectively.

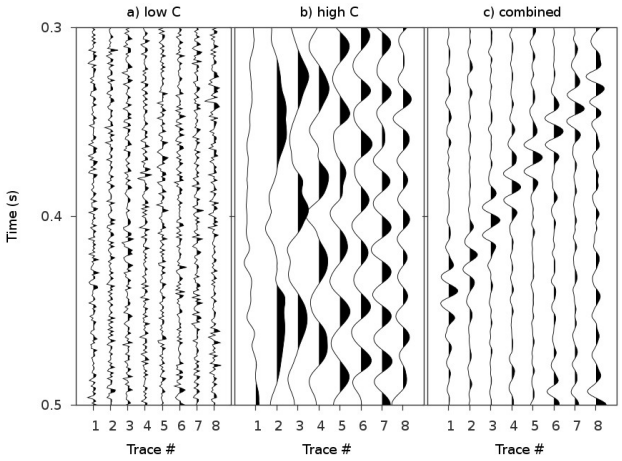

Figure 4: Denoising of the synthetic gather of Figure 2c with fast CEEMDAN. a) Noise component trapped for C=5 and b) for C=10. c) Combined result after subtracting data in a) and b) from the original signal.

Examples

Triaxial microseismic record. As a first real data example, we present a vertical triaxial microseismic record from Neuquén Basin, Argentina, that was acquired in a monitor well about 500 m away from the extraction point. Each record consists of 8 traces spaced 30 m and sampled at 1 ms. The components are shown in Figure 5. Channel 5 in the x and z components is corrupted by low-frequency noise due probably to a faulty receiver. Other channels are also severely contaminated by noise. We processed this data with the fast CEEMDAN to remove low and high frequency content (Figure 6). The low frequency content in this case is mainly due to the damaged receiver. As can be seen in Figure 7, the sinusoidal character of channel 5 is trapped and is removed in the final combined processing. The result shows that the amplitude information of channel 5 has been improved and the random noise has been also effectively attenuated. With the standard CEEMDAN a decomposition on several modes (up to ten) is necessary to trap the low-frequency signals. Then an energy-based criterion must be applied to each trace to detect and remove the modes due to the faulty trace. The fast algorithm takes 0.2 s to process each component (0.1 s for each value of the parameter C). In comparison the standard CEEMDAN takes 4.5 s for each component (Figure 8). The proposed algorithm is in this case is more than 20 times faster. If an energy-based criterion is used for the standard CEEMDAN, then the results are improved (Figure 9), but the computational time, due to a full decomposition of each trace, increases to 28 s for each component. Then the processing is 140 times slower than our fast algorithm. The denoising of low frequency noise can aid in the improvement of automatic picking algorithms by bringing amplitude information from traces that otherwise would be discarded due to their poor S/N ratio.

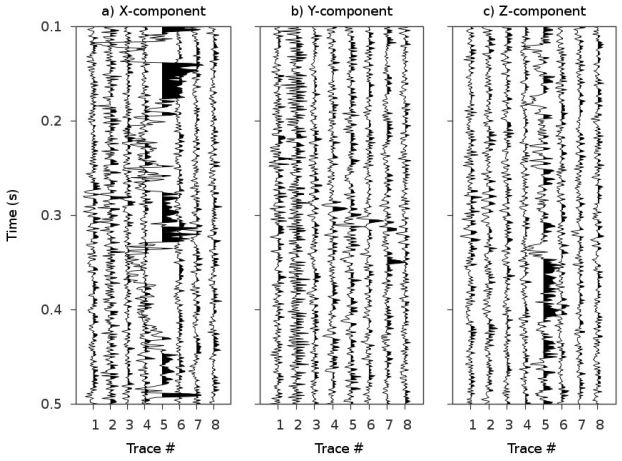

Figure 5: Three component microseismic record. a) X-component, b) y-component and c) z-component.

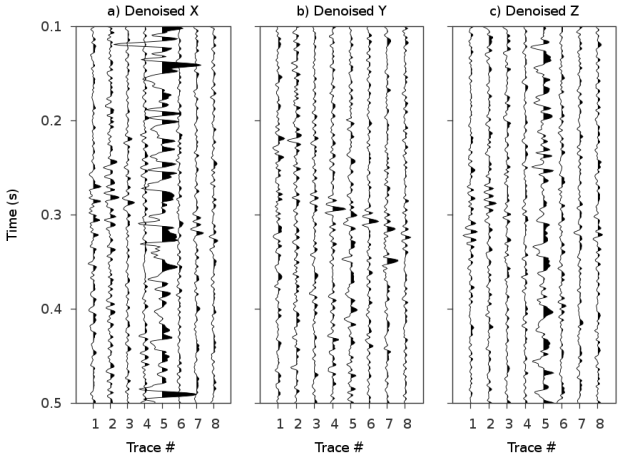

Figure 6: Denoised microseismic data by fast CEEMDAN with combined C parameter.

Seismic record with an earthquake. The second and final example is a typical seismic record acquired near the Andes mountain range. The data displays an earthquake that shows as a set. of horizontal seismic events starting

above 1 s (Figure 10a). The data consists of 31 traces sampled at 2 ms. The combined denoising with fast CEEMDAN is obtained after 8 s of computation (Figure 11a). A similar result with the standard CEEMDAN takes 40 s (5 times slower) and with removal of higher modes it consumes 233 s; which is 30 times slower (Figure 11b). The high frequency attenuation is similar in both cases, but the continuity of the horizontal events are significantly improved by the fast CEEMDAN.

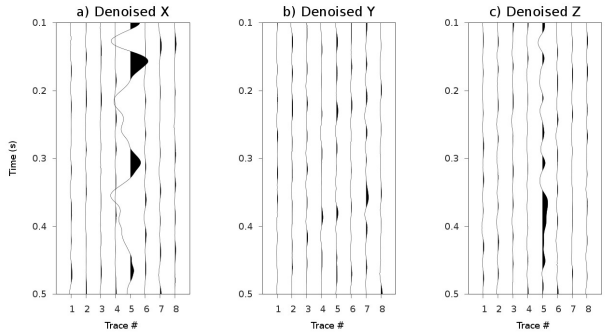

Figure 7: Low frequency amplitudes from running the fast CEEMDAN with high C.

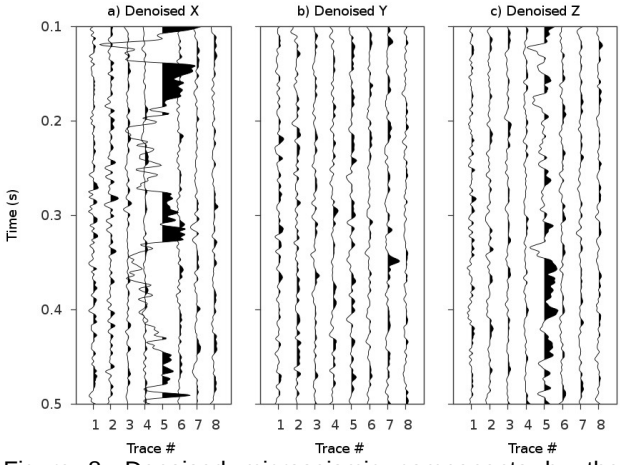

Figure 8: Denoised microseismic components by the standard CEEMDAN without energy criterion.

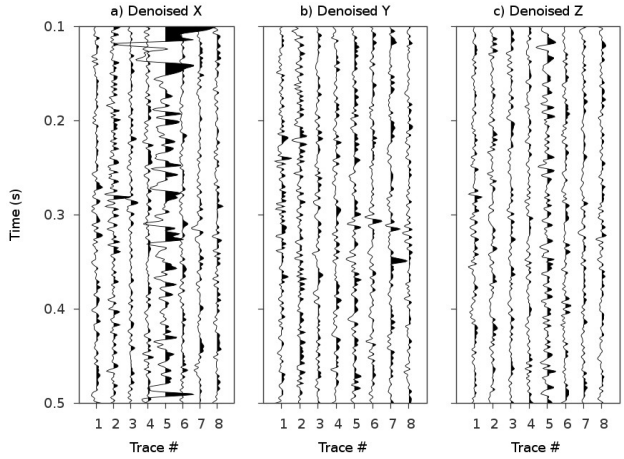

Figure 9: Denoised microseismic components by the standard CEEMDAN with energy criterion.

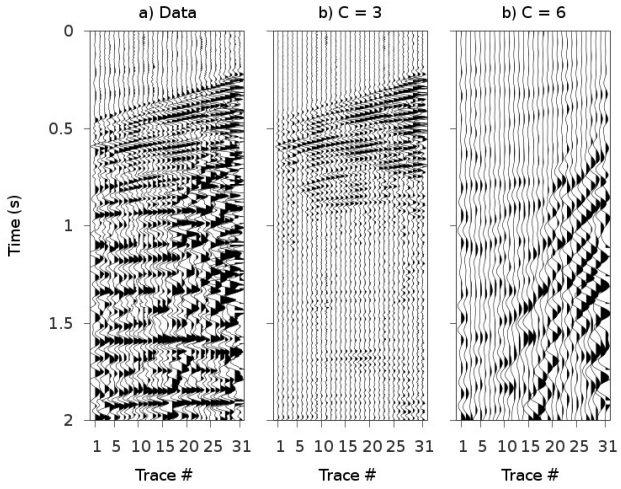

Figure 10: Portion of a seismic record contaminated with an earthquake: a) original data, b) C = 3 result and c) C = 6 result.

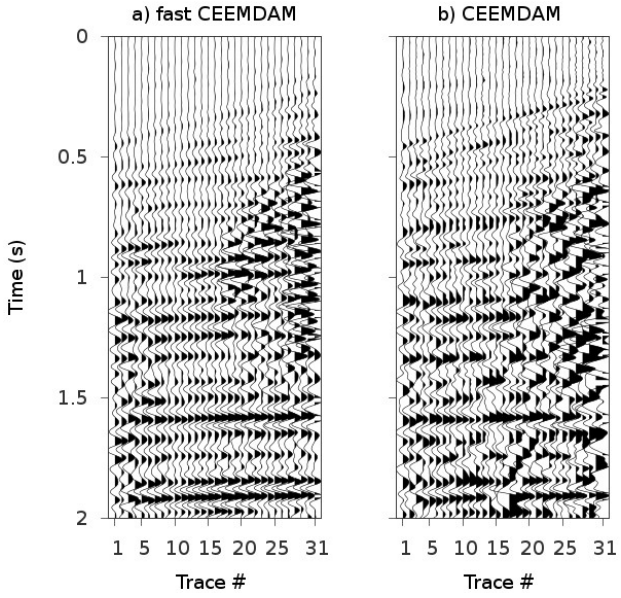

Figure 11: Results of denoising the seismic gather with a) fast CEEMDAN with combined C and b) standard CEEMDAN.

Conclusions

A fast implementation of the CEEMDAN algorithm was successfully applied for the denoising of coherent and random noise in microseismic and seismic data. The fast implementation allows for the processing of an important volume of seismic traces in an reasonable time. The CEEMDAN algorithm is modified by eliminating the need for cubic interpolation to obtain the mean local envelope. The denoising results from the fast CEEMDAN are similar to the standard algorithm with spline interpolation. However the computational cost of the proposed method is at least one order of magnitude lower.

Acknowledgments

This work was partially supported by Secretaría de Políticas Universitarias, Argentina and Programa de Incentivos, Universidad Nacional de La Plata (UNLP). We wish to extend our gratitude to Gabriela Goñi and Tilo Branca from GEONODOS. Their help and assistance was really helpful.

References

Battista, B. M., C. Knapp, T. McGee, and V. Goebel, 2007, Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data: Geophysics, **72**, H29-H37.

Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 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.

Lin, J., 2012, Improved ensemble empirical mode decomposition and its applications to gearbox fault signal processing: International Journal of Computer Science, **9**, 194-199.

Torres, M. E., M. A. Colominas, G. Schlotthauer, and P. Flandrin, 2011, A complete ensemble empirical mode decomposition with adaptive noise: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2011, IEEE, 4144-4147.

Wu, Z., and N. E. Huang, 2009, Ensemble empirical mode decomposition: a noise-assisted data analysis method: Advances in Adaptive Data Analysis, **1**, 1-41.