Advances in Adaptive Data Analysis
Vol. 4, No. 4 (2012) 1250025 (11 pages)

DOI: 10.1142/S1793536912500252

NOISE-ASSISTED EMD METHODS IN ACTION

MARCELO A. COLOMINAS[*,‡], GASTÓN SCHLOTTHAUER[*,§], MARÍA E. TORRES[*,¶] and PATRICK FLANDRIN[†,‖]

[*]*CONICET – Laboratorio de Señales y Dinámicas no Lineales*
Facultad de Ingeniería, Universidad Nacional de Entre Ríos
Ruta 11 Km 10, Oro Verde, Entre Ríos 3100, Argentina

[†]*Laboratoire de Physique (UMR 5672 CNRS)*
École Normale Supérieure de Lyon
46 allée d'Italie, 69364 Lyon Cedex 07, France
[‡]*macolominas@bioingenieria.edu.ar*
[§]*gschlotthauer@conicet.gov.ar*
[¶]*metorres@santafe-conicet.gov.ar*
[‖]*patrick.flandrin@ens-lyon.fr*

Received 9 December 2012
Revised 23 January 2013
Accepted 23 January 2013
Published 26 March 2013

In this work we explore the capabilities of two noise-assisted EMD methods: Ensemble EMD (EEMD) and the recently proposed Complete Ensemble EMD with Adaptive Noise (CEEMDAN), to recover a pure tone embedded in different kinds of noise, both stationary and nonstationary. Experiments are carried out for assessing their performances with respect to the level of the added noise and the number of realizations used for averaging. The obtained results partly support empirical recommendations reported in the literature while evidencing new distinctive features. While EEMD presents quite different behaviors for different situations, CEEMDAN evidences some robustness with an almost unaffected performance for the studied cases.

Keywords: Empirical mode decomposition (EMD); noise-assisted data analysis (NADA).

1. Introduction

Empirical Mode Decomposition (EMD) [Huang *et al.* (1998)] is an adaptive method introduced to analyze signals that may be nonstationary and/or stemming from nonlinear systems. It consists in a local and fully data-driven separation of a signal in fast and slow oscillations. At the end of the decomposition, the original signal can be expressed as a sum of amplitude and frequency modulated (AM-FM) functions called "intrinsic mode functions" (IMFs) plus a final trend either monotonic or constant.

Despite its proven usefulness, EMD may experience problems in some cases, such as the presence of oscillations of very disparate amplitude in a mode, or the presence of very similar oscillations in different modes, named as "mode mixing".

To overcome these problems, a new method was proposed: Ensemble Empirical Mode Decomposition (EEMD) [Wu and Huang (2009)]. It performs the EMD over an ensemble of the signal plus Gaussian white noise. The addition of white Gaussian noise alleviates the mode mixing problem by populating the whole time-frequency space to take advantage of the dyadic filter bank behavior of the EMD [Flandrin *et al.* (2004)]. Still, both the optimal size of the ensemble and the amplitude of the added noise are open questions. Although there are suggested values for these two key parameters [Wu and Huang (2009)], there were no extensive experiments carried out to clarify this point. In Niazy *et al.* [2009] the authors studied the denoising of an AM signal via EMD and EEMD, exploring three values for the algorithm noise and paying attention to the number of sifting iterations used for this last method, and concluding that EEMD improves the accuracy and robustness of the decomposition.

Besides the mode mixing alleviation, EEMD created some new problems. Indeed, the reconstructed signal includes residual noise and different realizations of signal plus noise may produce different number of modes, making difficult the final averaging. In order to overcome these situations, a new method was recently proposed: the "Complete Ensemble Empirical Mode Decomposition with Adaptive Noise" (CEEMDAN) [Torres *et al.* (2011)]. CEEMDAN adds a particular noise at each stage, and achieves a complete decomposition with no reconstruction error. Because of that, a smaller ensemble size can be used. Nevertheless, the amplitude of the added noise is still an open issue. In this paper, we explore the capabilities of these two noised-assisted EMD methods to recover a pure tone embedded in different kinds of noise, paying special attention to the algorithm added noise amplitude.

The paper is organized as follows. In Sec. 2 the main EEMD concepts are recalled, and the new method CEEMDAN is explained. In Sec. 3 the experiments are described and the results are presented. A discussion on them and features of the studied methods can be found in Sec. 4.

2. EEMD and CEEMDAN

EMD decomposes a signal $x(t)$ into a (usually) small number of IMFs or modes. To be considered as an IMF, a signal must satisfy two conditions: (i) the number of extrema and the number of zero crossings must be equal or differ at most by one; and (ii) the mean value of the upper and lower envelopes is zero everywhere.

EEMD defines the *true* IMF components (here notated as $\overline{\mathrm{IMF}}$ in what follows) as the mean of the corresponding IMFs obtained via EMD over an ensemble of trials, generated by adding different realizations of white noise of finite variance to the original signal $x[n]$. EEMD algorithm [Wu and Huang (2009)] can be described as:

(1) Generate $x^i[n] = x[n] + \beta\, w^i[n]$, where $w^i[n]$ $(i = 1, \ldots, I)$ are different realizations of zero mean unit variance white Gaussian noise,

(2) Fully decompose by EMD each $x^i[n]$ $(i = 1, \ldots, I)$ getting their modes $\mathrm{IMF}_k^i[n]$, where $k = 1, \ldots, K$ indicates the modes,
(3) Assign $\overline{\mathrm{IMF}}_k$ as the kth mode of $x[n]$, obtained as the average of the corresponding IMFs: $\overline{\mathrm{IMF}}_k[n] = \frac{1}{I}\sum_{i=1}^{I} \mathrm{IMF}_k^i[n]$.

Observe that in EEMD, each $x^i[n]$ is decomposed independently from the other realizations and so for each one a residue $r_k^i[n] = r_{k-1}^i[n] - \mathrm{IMF}_k^i[n]$ is obtained.

In the CEEMDAN method [Torres *et al.* (2011)], the decomposition modes will be noted as $\widetilde{\mathrm{IMF}}_k$ and, in order to obtain a complete decomposition, we propose to calculate a unique first residue as:

$$r_1[n] = x[n] - \widetilde{\mathrm{IMF}}_1[n], \tag{1}$$

where $\widetilde{\mathrm{IMF}}_1[n]$ is obtained in the same way as in EEMD. Then, we generate a new ensemble of $r_1[n]$ plus different realizations of a given noise and compute the first EMD mode of each element. By averaging of these first EMD modes, we obtain $\widetilde{\mathrm{IMF}}_2$. The next residue is defined as: $r_2[n] = r_1[n] - \widetilde{\mathrm{IMF}}_2[n]$. This procedure continues with the rest of the modes until the stopping criterion is reached.

Let us define the operator $E_j(\cdot)$ which, given a signal, produces the jth mode obtained by EMD. Let w^i be a realization of white noise with $\mathcal{N}(0,1)$, $(i = 1, \ldots, I)$. If $x[n]$ is the targeted data, we can describe our method by the following algorithm:

(1) Decompose by EMD I realizations $x[n] + \beta_0\, w^i[n]$ to obtain their first mode and compute:

$$\widetilde{\mathrm{IMF}}_1[n] = \frac{1}{I}\sum_{i=1}^{I} \mathrm{IMF}_1^i[n] = \overline{\mathrm{IMF}}_1[n].$$

(2) At the first stage $(k = 1)$ calculate the first residue as in Eq. (1): $r_1[n] = x[n] - \widetilde{\mathrm{IMF}}_1[n]$.
(3) Decompose realizations $r_1[n] + \beta_1\, E_1(w^i[n])$, $i = 1, \ldots, I$, until their first EMD mode and define the second mode:

$$\widetilde{\mathrm{IMF}}_2[n] = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1[n] + \beta_1\, E_1(w^i[n])).$$

(4) For $k = 2, \ldots, K$ calculate the kth residue:

$$r_k[n] = r_{(k-1)}[n] - \widetilde{\mathrm{IMF}}_k[n]. \tag{2}$$

(5) Decompose realizations $r_k[n] + \beta_k\, E_k(w^i[n])$, $i = 1, \ldots, I$, until their first EMD mode and define the $(k+1)$th mode as

$$\widetilde{\mathrm{IMF}}_{(k+1)}[n] = \frac{1}{I}\sum_{i=1}^{I} E_1(r_k[n] + \beta_k\, E_k(w^i[n])). \tag{3}$$

(6) Go to step 4 for next k.

Steps 4 to 6 are performed until the obtained residue can no longer be decomposed (the residue does not have at least two extrema). The final residue satisfies:

$$R[n] = x[n] - \sum_{k=1}^{K} \widetilde{\mathrm{IMF}}_k, \tag{4}$$

with K being the total number of modes. Therefore, the given signal $x[n]$ can be expressed as:

$$x[n] = \sum_{k=1}^{K} \widetilde{\mathrm{IMF}}_k + R[n]. \tag{5}$$

Equation (5) makes the proposed decomposition complete and provides an exact reconstruction of the original data. Observe that the coefficients $\beta_k = \varepsilon_k \,\mathrm{std}(r_k)$ allow to select the SNR at each stage, but this possibility will not be exploited in this study. It must be taken into account the fact that at stage k we generate the ensemble by adding the kth mode of white Gaussian noise $(E_k(w^i[n]))$ to the kth residue $r_k[n]$; then we always compute the first mode of each element of the ensemble $(E_1(\cdot))$. Figure 1 summarizes the proposed algorithm.

Concerning the amplitude of the added noise, Wu and Huang suggested to use small amplitude values for data dominated by high-frequency signals, and *vice versa* [Wu and Huang (2009)]. In the next section, we will explore different values for given ε parameters and fix the same SNR for all the stages. In all the implementations we use the EMD toolbox available at: http://perso.ens-lyon.fr/patrick.flandrin/emd.html. CEEMDAN implementation is available at: http://www.bioingenieria.edu.ar/grupos/ldnlys/.

3. Experiments and Results

In order to better understand the influence of the amplitude of the added noise in both EEMD and CEEMDAN, a pure tone of the form

$$s[n] = \sin(0.2\,n), \quad n = 1, 2, \ldots, 256, \tag{6}$$

was corrupted with three different kinds of noise (which will be described in their corresponding subsections). Each noise realization was built upon one same realization for these three cases, and the same realizations were used for both methods. Figure 2 summarizes this approach.

Different values for the SNR of the signal plus corrupting noise (referred to as "input SNR" in what follows) were used, and also different values for the algorithms parameters. Given a set of (input and algorithm) parameters, $J = 100$ noisy tones were decomposed by the two methods and a distance e was computed for each of them as follows:

$$e(\eta, \varepsilon, I)_j = \|s[n] - \mathrm{IMF}_{k_j^*}(\eta, \varepsilon, I)[n]\|_2, \tag{7}$$

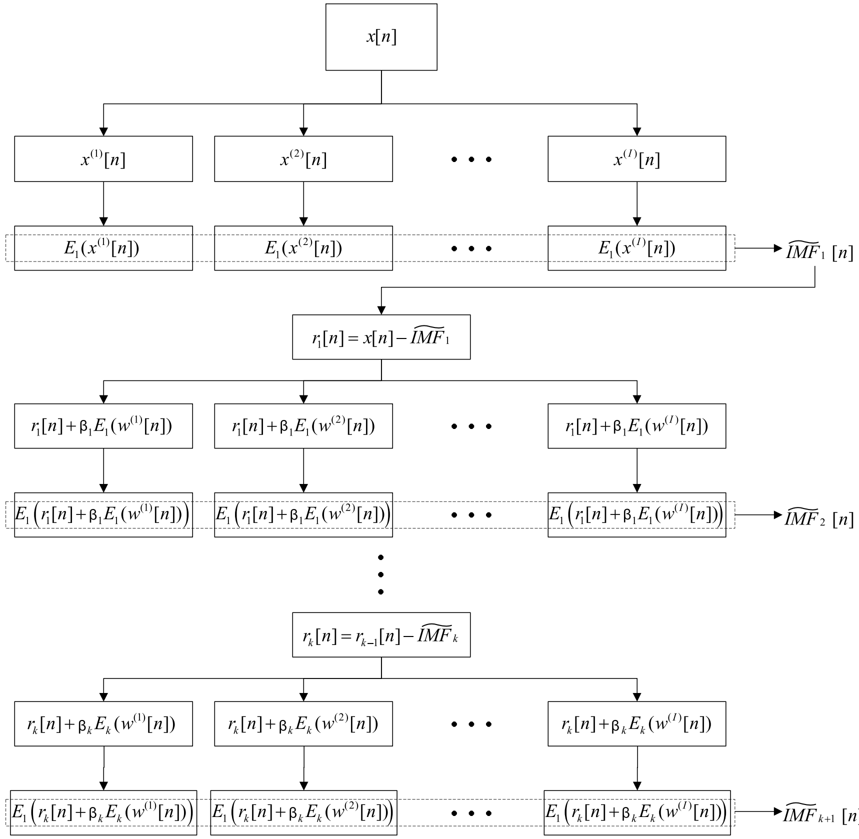

Fig. 1. CEEMDAN algorithm. $\widetilde{\mathrm{IMF}}_1$ is obtained in the same way as in EEMD. A unique first residue $r_1[n]$ is computed and a new ensemble of $r_1[n]+\beta_1 E_1(w[n])$ is generated. $\widetilde{\mathrm{IMF}}_2$ is obtained by averaging the first EMD modes of those elements. The procedure continues with the rest of the modes until the stopping criterion is reached.

where $j = 1, \ldots, J$ stands for the input noise realization, η is the input SNR, ε the algorithm noise amplitude, I the ensemble size, $\mathrm{IMF}_{k_j^*}$ the k^*th mode obtained from the jth noisy tone (by either of the two methods), and

$$k_j^* = \arg\min_k \|s[n] - \mathrm{IMF}_{k_j}(\eta, \varepsilon, I)[n]\|_2 \tag{8}$$

indicates the closest (in distance) mode to the original noiseless tone.

Final results were obtained by averaging over the J noisy tone decomposition distances: $\overline{e}(\eta, \varepsilon, I) = \langle e(\eta, \varepsilon, I)_j \rangle$. A small value for $\overline{e}$ indicates a good tone recovering and thus an avoidance of mode mixing. On the other hand, if $\overline{e}$ has a large value, this means that mode mixing was strong and the algorithm was not able to catch the tone with enough energy in only one mode.

Exploiting the no reconstruction error property of CEEMDAN, a small ensemble size (up to 20) was performed. To have a fair comparison between the two methods,

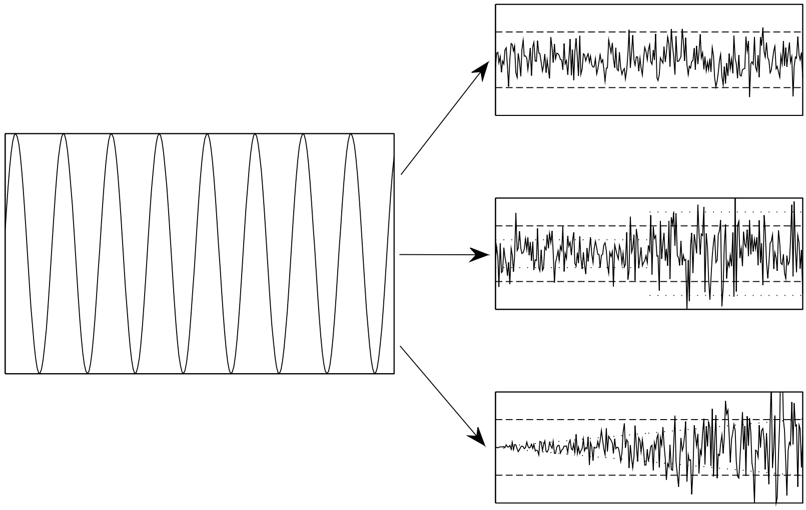

Fig. 2. Experimental design. The test signal is a pure tone (left) embedded in three different kinds of noise (right). For a same total energy, the standard deviation of those noises is supposed either constant (top), abruptly doubling its value at the mid-time (middle), or linearly increasing (bottom).

the exact same set of noise realizations was used in both of them. Sifting iterations were not limited and the default stopping criterion from Rilling *et al.* [2003] was used. Algorithm noise parameter ε values belonging to the interval $[0.025; 0.5]$ with a 0.025 step and to the interval $[0.6; 2.0]$ with a 0.1 step have been used.

3.1. *Constant amplitude noise*

As a first experiment, the noise was added with an equal standard deviation throughout the duration of the signal. Then, 100 noisy tones of the form

$$s_{\sigma_0}[n] = s[n] + \sigma_0\,\mathrm{std}(s)\,w^j[n] \tag{9}$$

were decomposed using different algorithm parameters, where $w^j[n]$ is a zero mean unit variance Gaussian noise realization. Constant σ_0 was fixed in order to obtain the following SNRs: 9 dB, 6 dB, 3 dB, and 0 dB.

The results of the average distances computed from decompositions via both methods are presented on Fig. 3. In these plots, the input SNR was kept constant, and results for different number of realizations were superimposed. On the first two rows, EEMD presents a two-regime situation with a local minimum around 0.2 for the algorithm noise. This situation disappears in the third and fourth rows, with a minimum at an intriguingly high level of algorithm noise. Instead, CEEMDAN displays almost the same behavior in all cases: a smaller local minimum around 0.2, with an expected vertical shifting corresponding to the performance improvement when the input SNR increases.

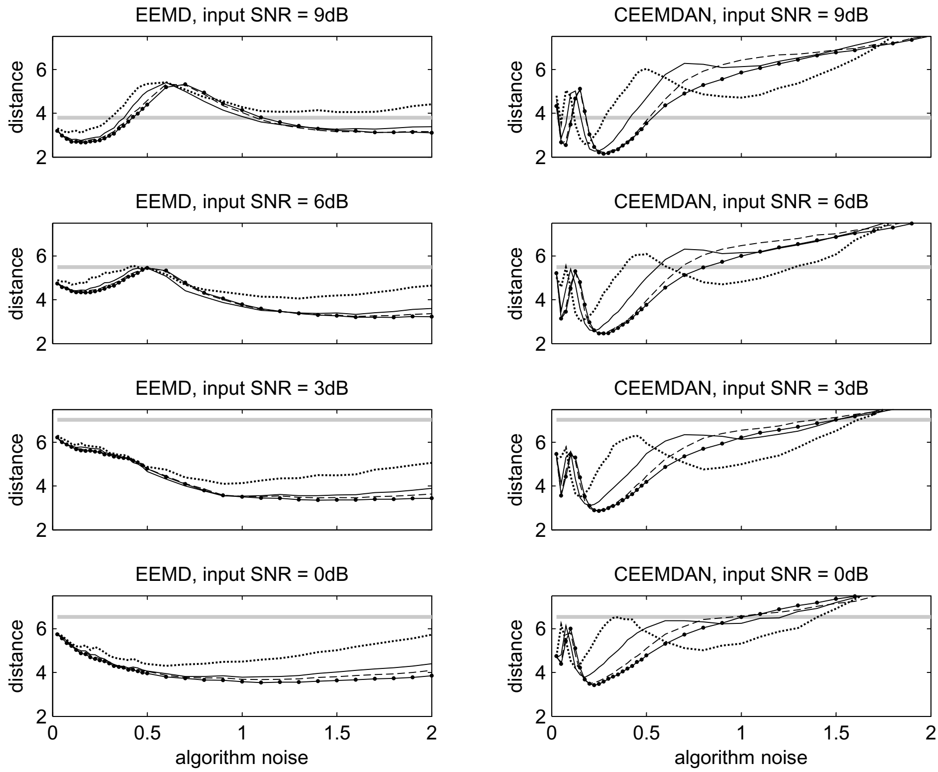

Fig. 3. Results for constant amplitude noise. No. of realizations: 5 (dotted), 10 (solid), 15 (dashed) and 20 (dashed dotted). EMD performance is shown in thick gray for comparison purposes.

3.2. *Step in the amplitude noise*

As a second case, the noise was added with a given amplitude in the first half of the signal, and with twice the amplitude in the second half. Noisy tones of the form

$$s_{\sigma_s}[n] = s[n] + \mathrm{std}(s)\, w_s^j[n] \tag{10}$$

with

$$w_s^j[n] = \begin{cases} \sigma_s\, w^j[n], & \text{if } n = 1, 2, \ldots, 128, \\ 2\,\sigma_s\, w^j[n], & \text{if } n = 129, 130, \ldots, 256, \end{cases}$$

were decomposed. In this case, $w^j[n]$ stands for the exact same jth noise realization as in the constant amplitude case. To achieve the same SNRs as in the previous example, the constants were related by $\sigma_s = \sqrt{2/5}\,\sigma_0$.

Results with a constant input SNR are depicted on Fig. 4. As in the previous case, EEMD presents two regimes in the two first rows. It is also clear how EEMD performance almost stabilizes for 10 realizations. Again CEEMDAN shows a similar

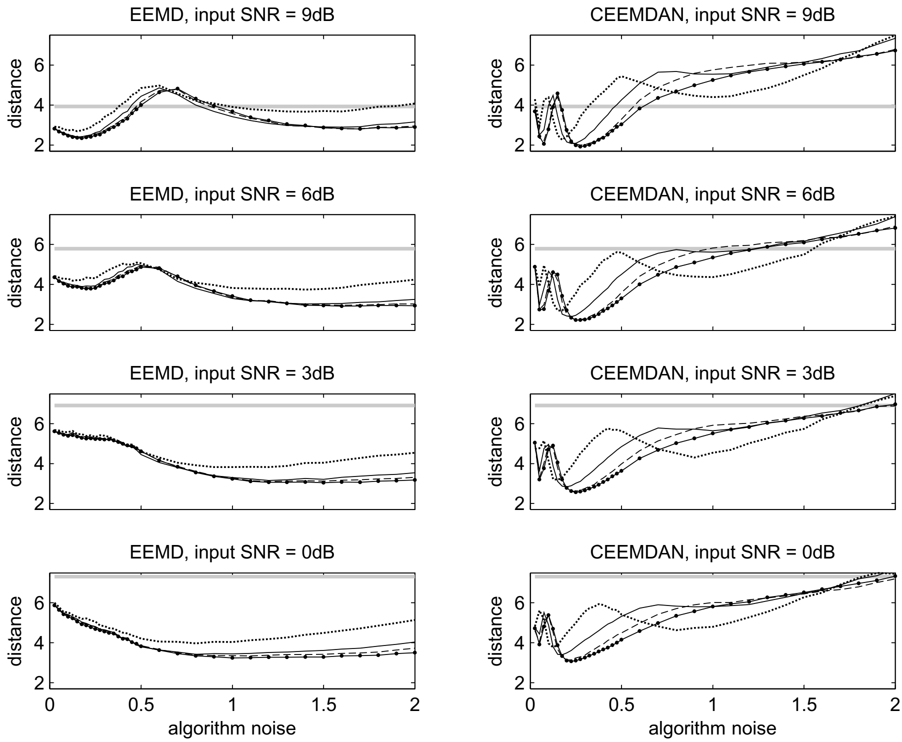

Fig. 4. Results for step in the amplitude noise. No. of realizations: 5 (dotted), 10 (solid), 15 (dashed) and 20 (dashed dotted). EMD performance is shown in thick gray for comparison purposes.

behavior for all input SNRs. Performance stabilization seems to be almost reached at 15 realizations.

3.3. *Increasing amplitude noise*

As a final example, the noise was added with a linearly increasing amplitude. The decomposed noisy tones were

$$s_\alpha[n] = s[n] + \alpha\, n\, \mathrm{std}(s)\, w^j[n]. \tag{11}$$

To keep the same SNRs as in the two previous examples, the slope was $\alpha = \sqrt{\frac{N}{\sum_{n=1}^{N} n^2}}\,\sigma_0$, with $N = 256$ being the signal length. Again, $w^j[n]$ is the same jth realization as in the previous examples.

The results are presented in Fig. 5. As before, the input SNR was kept constant. In this case, the EEMD two-regime situation seems to be a little bit stronger. CEEMDAN performance is almost unaffected for the input SNR, with the same expected vertical shifting while the input SNR increases. As in the previous two examples, there is an oscillation for small levels of algorithm noise to be clarified in future works.

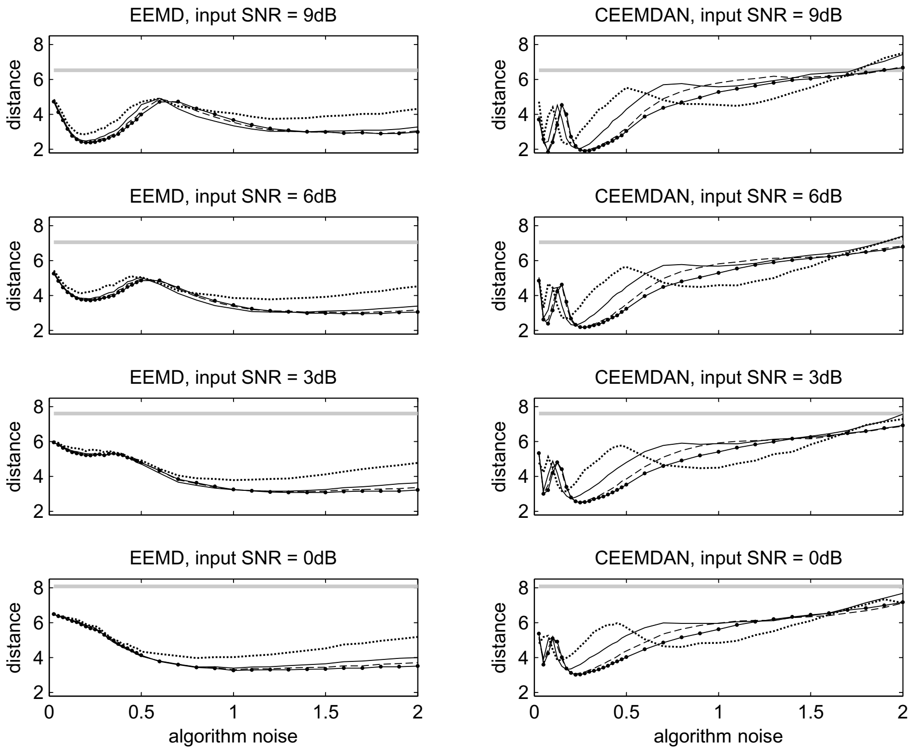

Fig. 5. Results for increasing amplitude noise. No. of realizations: 5 (dotted), 10 (thin black), 15 (dashed) and 20 (dashed dotted). EMD performance is shown in thick gray for comparison purposes.

4. Discussion

The rationale behind noise-assisted variations on EMD is that some adequate level of added (algorithm) noise should improve upon the identification of well-structured signals embedded in observation (input) noise. Intuitively, a too small amount of added noise should be of no effect, while a too strong one should end up with so erratic waveforms that identification should again be very difficult. In between, it is expected that some optimum could be attained, e.g. a minimum when the performance measure is the distance between the actual waveform and the estimated one. It is therefore remarkable to see that such a local minimum is indeed observed in many of the considered examples. Moreover, this is most often so for a value of the standard deviation of the algorithm noise that is close to 0.2, the value suggested by Wu and Huang in Wu and Huang [2009].

However, some differences occur between EEMD and CEEMDAN, that can be summarized in the two main following points:

(1) As for EEMD, the observed behavior tends to differ from the expected one in two respects. For small to moderate input noise, the valley around 0.2 (where

the distance attains its minimum) is supplemented by a second valley, with a less pronounced minimum, that is obtained for much higher levels of algorithm noise. This two-regime situation tends to disappear as the input noise increases, with no local minimum around the recommended value. EEMD performance is therefore improved by adding more noise, but it is necessary in this case to perform more realizations, especially if a low reconstruction error is needed. Finally, the global minimum seems in this case to be not as low as CEEMDAN's.

(2) As for CEEMDAN, there is a similar behavior in all situations, no matter how large is the input noise: the results for CEEMDAN have almost the same "shape" for different kinds of noise. For small levels of algorithm noise, there is however an oscillation in the performance for which we have currently no explanation and which would deserve more attention in future investigations. Nevertheless, it appears that CEEMDAN seems to have some built-in robustness, the performance being not affected by the input SNR, and the global optimum being attained around the suggested value of 0.2 for standard deviation of the algorithm noise.

5. Conclusion

The present work has extensively studied the influence of the two key parameters (added noise amplitude and number of realizations) of two noise-assisted EMD methods in typical, yet not universal, situations. As expected, both of them have evidenced some form of stochastic resonance [Gammaitoni *et al.* (1998)], showing a value of algorithm noise for which performance reaches its best.

Future investigations will attempt to shed some light into the very small noise amplitude zone of CEEMDAN, where unexplained oscillations appear, as well as casting explicitly the considered noise-assisted methods in a stochastic resonance framework.

Acknowledgments

This work was supported by the Argentine National Agency for Scientific and Technological Promotion (ANPCyT-Argentina), Universidad Nacional de Entre Ríos (Argentina), the Argentine National Council for Scientific and Technical Research (CONICET-Argentina) and the Centre National de la Recherche Scientifique (CNRS-France).

References

Flandrin, P., Rilling, G. and Gonçalvès, P. (2004). Empirical mode decomposition as a filter bank. *IEEE Signal Process. Lett.*, **11**(2): 112–114.

Gammaitoni, L., Hänggi, P., Jung, P. and Marchesoni, F. (1998). Stochastic resonance. *Rev. Mod. Phys.*, **70**(1): 223–287.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, E. H., Zheng, Q., Tung, C. C. and Liu, H. H. (1998). The empirical mode decomposition method and the Hilbert spectrum for nonstationary time series analysis. *Proc. Roy. Soc. London*, **454A**: 903–995.

Niazy, R. K., Beckmann, C. F., Brady, J. M. and Smith S. M. (2009). Performance evaluation of ensemble empirical mode decomposition. *Adv. Adapt. Data Anal.*, **1**(2): 231–242.

Rilling, G., Flandrin, P. and Gonçalvès, P. (2003). On empirical mode decomposition and its algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*, Grado (I).

Torres, M. E., Colominas, M. A., Schlotthauer, G. and Flandrin, P. (2011). A complete ensemble empirical mode decomposition with adaptive noise. *IEEE Ann. International Conference on Acoustics*, *Speech and Signal Processing ICASSP'11*, 4144–4147.

Wu, Z. and Huang, N. E. (2009). Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.*, **1**(1): 1–41.