Sensors & Transducers, Vol. 204, Issue 9, September 2016, pp. 21-28

Sensors & Transducers

http://www.sensorsportal.com

A Sliding Window Empirical Mode Decomposition for Long Signals Algorithm

[1] J. L. Sanchez, [2] Manuel D. Ortigueira, [3] Raul T. Rato, [4] Juan J. Trujillo
[1] Departamento de Ingeniería Informática y de sistemas,
Universidad de La Laguna 38271 La Laguna, Tenerife, Spain
[2] UNINOVA and DEE of Faculdade de Ciências e Tecnologia da UNL, Caparica, Portugal
[3] UNINOVA and Escola Superior de Tecnologia do Instituto Politécnico de Setubal,
2910-761 Setubal, Portugal
[4] Departamento de Análisis Matemático, Universidad de La Laguna 38271 La Laguna, Tenerife, Spain
[1] Tel.: 34922845043
[1] E-mail: jsanrosa@ull.edu.es

Received: 4 July 2016 /Accepted: 31 August 2016 /Published: 30 September 2016

Abstract: This document presents a sliding window algorithm for the calculation of the empirical mode decomposition for long signals. The spline calculation of very long signals requires a long computation time. Our aim is to improve the calculation time of the empirical mode decomposition for Long signals. Some authors have used sliding windows for the whole decomposition. Our main contribution is to reduce the computation time calculating each intrinsic mode function on a sliding window basis. That ensures the obtained intrinsic mode function has no discontinuities on the junction regions between consecutive windows. Moreover, the sliding window size changes adaptively according to the number of extrema in the previous intrinsic mode function. The effectiveness of the proposed method increases with the length of the signal obtaining computation times of the order of 30 % of the time required to obtain the decomposition using only a window as in the classical manner. Those results are important to apply the empirical mode decomposition to long signals. Particularly, to biomedical signals like long-term ECG or long term EEG. *Copyright © 2016 IFSA Publishing, S. L.*

Keywords: Empirical mode decomposition, Intrinsic mode function, Long signals, Sliding window.

1. Introduction

The Empirical Mode Decomposition (EMD), as was proposed initially by Huang, *et al.* [1] is a signal decomposition algorithm based on a successive removal of elemental signals: the Intrinsic Mode Functions (IMF). These are continuous functions such that at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. They are obtained through an iterative procedure called sifting that is a way of removing the dissymmetry between the upper and lower envelopes in order to transform the original signal into an amplitude modulated (AM) signal. Moreover, as the instantaneous frequency can change from instant to instant, it can be said that each IMF is a simultaneously amplitude and frequency modulated signal (AM/FM). So, the EMD is nothing else than a decomposition into a set of AM/FM modulated signals [2-5].

It must be emphasized that EMD is merely a computational algorithm that expresses a given signal as a sum of simpler components. It cannot be said that the obtained components are true parts of the signal at hand.

The original algorithm had some implicit difficulties [3, 6]: extrema location, the end effect and the stopping criterion are critical. Some solutions were proposed in [3] and implemented in an algorithm that can be found at [7] ((hereinafter referred as "the reference algorithm"). The location and amplitude of the extrema were estimated using a parabolic interpolation. To render less severe the end effect, the maxima and the minima were extrapolated by both sides. A new stopping criterion in the sifting procedure by introducing two resolution factors was defined.

In the last years, several modifications have been proposed to increase the performances of EMD, [8-18]. It is important to question if the introduced complexity compensates the quality increase. In this work, it will be preserved the simplicity of the original algorithm while it is increased the reliability and applicability of the decomposition.

In practical applications, there are several tradeoffs among resolution, signal length, the number of IMFs and running time. In fact, an increase in signal length produces two unwanted side-effects. On the one hand, it leads generally to a corresponding increase in the number of IMFs. Consequently, the running time may become so high, that the algorithm will be useless. The increase in the number of IMFs is a very important drawback because it may originate "false" components that are added to one IMF and subtracted to another one or appear isolated. So, in general, there are no guarantees to have IMFs that are really present in the original signal. On the other hand analyzing long signals with EMD is time-consuming or even impossible in a reasonable time [19] due to the fact, that spline interpolation of a large number of points takes a lot of computer resources. In applications to long signals [20, 21] the number of components and the running time would be so high that the algorithm would be almost useless. This problem was recently considered in [19], where the need of a more efficient and faster algorithm to deal with long signals was stated.

Most of the algorithms present in the literature are not prepared to deal with long signals as they use only one window with the length of the signal. In Section 2.1 it is shown the increase in processing time with increasing lengths. In some applications like EEG or ECG processing we may have to process signals with lengths above 10^6. The processing time makes the existing algorithms almost useless. This suggests it is important to have an algorithm with the same characteristics but faster. In order to reach such goal, a sliding window EMD is proposed where consecutive windows overlap over a pre-specified amount. While [19] proposes to obtain a full EMD at every window, this approach can cause errors when dealing with signals that have fast changes in their frequency composition. In such case, it could be impossible to obtain decomposition with the same number of IMFs in all the segments. Our solution follows a different approach. Each IMF is calculated in sliding windows to ensure the number of obtained IMFs components is the same for the whole length of the signal. While [19] proposes to increase the size of the sliding window by a fixed quantity, here it is proposed to duplicate the size of the window when necessary. In that manner, fewer steps are involved for very long signals. Such implies that the length of the signal must be a power of two of the initial window size. Finally, a sliding window algorithm has an overload in the computation of the EMD for signals with a small number of samples. So, the minimum window size must have a lower threshold, corresponding to the length for which the sliding window EMD is slower than the whole length EMD. All these factors together led to an adaptive algorithm. Both, the length of the sliding windows and the length of the overlapping region depend on the length of the whole signal. Also, the overlapping region is tapered to improve the junction between adjacent windows.

The paper outlines as follows. In Section 2, the EMD and some of its drawbacks are discussed. An example is shown in Section 2.1. In Section 3.1, the problems that arise in analyzing long signals are presented and in Section 3.2, a new method, to analyze long signals, based on a sliding window is proposed. In Section 4, some illustrating results of the new method are presented. An application to the fan heater example and a comparison with the method in [3] is presented.

2. The EMD and its Drawbacks

A large number of papers published in the last years remarked the usefulness of the EMD. One of the most important advantages of EMD is the ability to decompose a complex signal into a finite set of narrowband signals without introducing any particular constraint on its characteristics. This makes easier the spectral estimation and creation of simple models. Some important aspects to be considered are:

Firstly, it is necessary to reflect about the meaning of the IMFs: In general, it is not possible to establish any special connection between a given IMF and the structure (eventually tied with the underlying physics) of the original signal. This does not mean that it cannot be done in some particular situations.

Secondly, another problem is the existence of false components in the IMFs. This is a consequence of the numerical errors in sifting: one component is added in one IMF and subtracted in another one.

Finally, the number of IMFs depends on the length of the signal. In fact, the number of components increases with the length of the signal. This may be an unwanted feature of the algorithm that is connected with the false component generation and increases the time required to do the decomposition.

It can be concluded that the main drawbacks of EMD are the false components and the large computational time for long signals. In the following, we will propose a solution for the second problem that alleviates the first.

2.1. An Example

In a search for long range processes, an experiment with the electric circuit of a heater fan was carried out. The signal was sampled during two hours with a sampling interval of 10 ms. The EMD for increasing length segments, using the reference algorithm, is computed and the results for 2 different resolutions (45 and 50) are shown in Table 1.

Table 1. Number of IMF and computational time vs. length of the signal for a heater fan signal.

Resol	Length	L Factor	IMFs	Time (s)	T Factor
45	11.400		14	28	
45	27.600	2.42	15	41	1.46
45	114.000	4.13	18	363	8.85
45	340.800	2.98	20	1.057	2.91
45	691.800	2.03	22	3.075	2.91
50	11.400		13	29	
50	27.600	2.42	16	69	2.37
50	114.000	4.13	19	602	8.72
50	340.800	2.98	22	1.774	2.94
50	691.800	2.03	22	5.195	2.92

Computations were carried out on a PC using Matlab. It is possible to reduce the computational time by compiling the Matlab code. However, we will make all the measurements without compilation.

Fig. 1 shows a tidal signal and it's EMD. Regarding the meaning of the IMFs, it can be observed that the most important IMFs are the two upper ones. The Fourier transform confirms such assumption since the peak frequencies of such IMFs correspond to the frequencies of the main components in the tidal signal: the positions of the Moon and the Sun relative to Earth and the Earth's rotation. The first has a period of about 12 hours and 25 minutes and the second has a period of 24 hours. These are clearly identified in the pictures. Even with a careful study it would be more difficult to give some meaning to some of the other components.

Regarding the existence of false components, it can be observed in Fig. 2 that the strips 5 and 6 share an important part of their spectrum. That is an indication than they are not really independent components.

Regarding the number of IMFs, it depends on the length of the signal. In fact, the number of components and the computation time increases with the length of the signal. This may be an unwanted feature of the algorithm that is connected with the false component generation.

The increase in time computation can be observed in Table 1. Resol represents the minimum resolution to obtain an IMF. Length represents the length of the signal. L Factor is the rate between the actual length and the previous one. IMFs is the number of IMFs for each trial. Time is the time in seconds. The "L Factor" represents the increase in length of the signal, while the "T factor" represents the factor by which the previous time must be multiplied for.

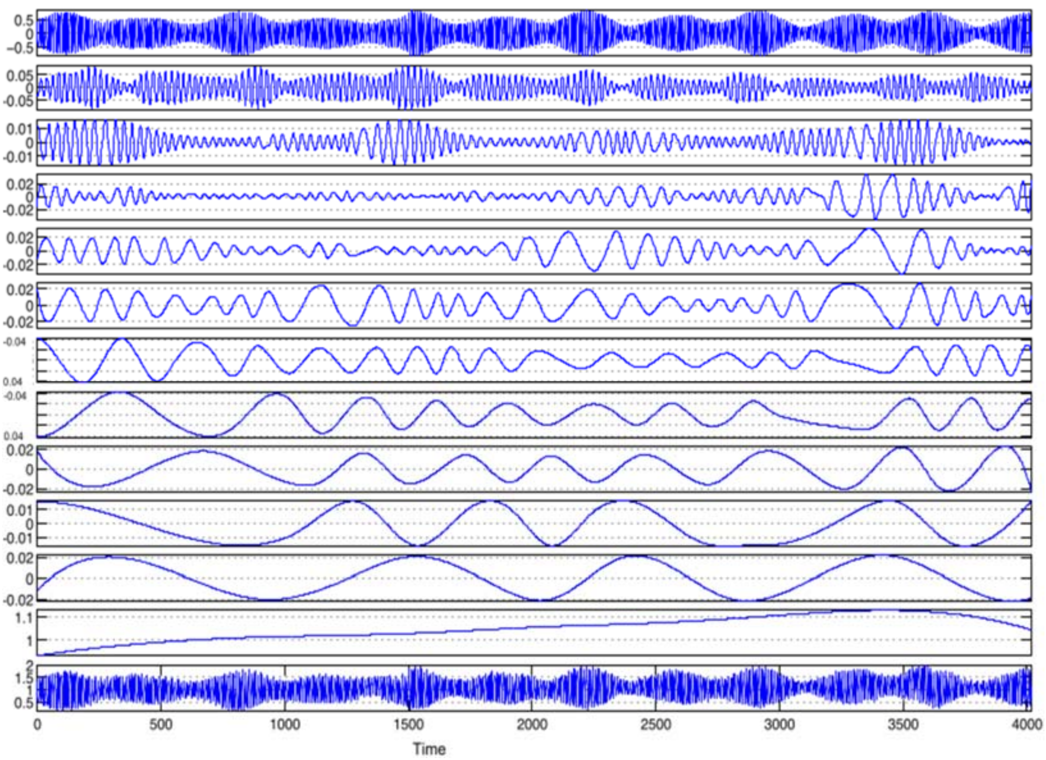

Fig. 1. EMD using the reference algorithm for a tidal signal.

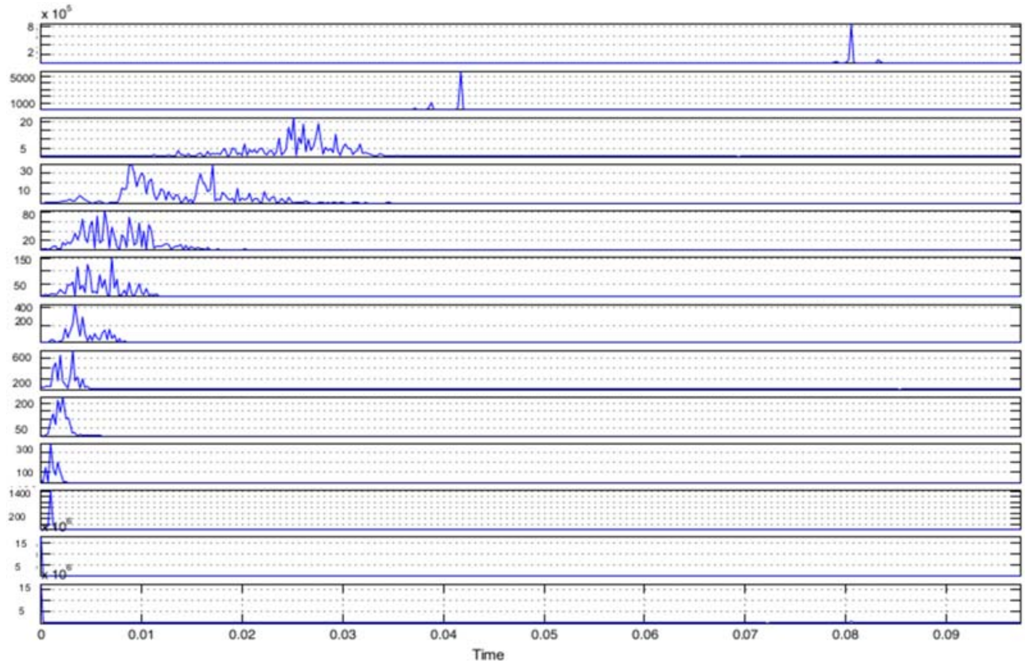

Fig. 2. Spectrum of the EMD decomposition of the tidal signal.

For 114.000 samples, a region with slow convergence appears in the signal, so the time increases much more than the number of samples. It can be observed that while for small lengths a multiplication by two, changes the computation time by approximately 1.5, for longer signals a duplication in size implies an increase factor near to 3. That is due to the fact that the detection of extrema and the spline computation increases faster than the signal length.

3. Decomposing Long Signals

The objective of this paper is to propose an algorithm that can be used with long signals with significative reduction in the processing time and eventually in the number of IMFs. In first place, the problem is raised. In second place, the solution proposed is presented.

3.1. The Problem

Let $x(t)$ be a given signal to be decomposed by EMD. The number of IMFs is not known in advance and normally grows up with increasing the length of $x(t)$. This increments the computational burden, leading in some situations to very large computational times making the algorithm useless unless suitable actions are developed – see Table 1. One obvious procedure is to cut the signal into segments. However, this can lead to poor results due to several factors. First, different segments can produce a different number of IMF in the decomposition. Secondly, the end effect introduces discontinuities at the junction points. Finally, a reduced number of extremes lead to poor quality envelopes.

3.2. The Solution: Adaptive Sliding Window EMD Algorithm

As it has been indicated one of the drawbacks in analyzing long signals is the computational load. The extrema detection and the spline interpolation of a large number of points take a lot of computer resources [20].

For this reason, the use of the sliding window EMD is proposed [22]. The underlying idea to all the algorithms that use sliding windows is to use a divide and conquer strategy. The computation time for spline interpolation is dramatically reduced using smaller segments. However, each EMD needs a minimum time to be computed. So, an excessive number of windows could have a detrimental effect. A balance must be found among the number of windows and its length. Most of the sliding window algorithms compute a whole EMD for each window. Our algorithm is based on calculating each IMF by sliding windows in order to obtain better quality decomposition. As it can lead to an excessive number of windows, three conditions are stablished. Firstly, a threshold for the window size is stablished. This threshold is set by the user. Secondly, the window length is restricted to be a power of two. Thirdly, the window length is duplicated when a lower threshold in the number of extrema is reached. In that manner, the EMD is computed in an adaptive manner where lower frequencies are analyzed with a smaller number of windows than high frequencies.

As referred above, the main idea of the algorithm is to apply the EMD sifting segment to segment to obtain only one IMF at a time. This procedure is done along the whole signal. This ensures that a real EMD is obtained. A pseudocode of our algorithm is shown in List 1.

List 1. Adaptive Sliding Window EMD pseudocode.

INPUT:
Filename
Startingsample
SignalLength
Resolution (dB)
OverlapPercentage
MinWindowSize
MinOscStop
MinOscEndStop

START:
OptimumSizeWindow=f1(SignalLength)
OverlapSamples=f2(MinWindowSize, OverlapPercentage)
OSC ← Inf
MINOSC ← MinOscStop

BODY:
WHILE (Osc > MINOSC)
IF (WindowSize<Length)
MINOSC ← MinOscEndStop
SWEMD(1:Length, L0, MINOSC)
ELSE
MINOSC ← MinOscEndStop
EMD(L, MINOSC)
ENDIF
ENDWHILE

For a general formulation consider a signal of length L. Select the segment length N and the overlapping M points.

1) Determine the starting window size, the number of samples in the overlap region, and the number of residual samples that do not fit in an integer number of windows.

2) Start a loop to obtain the whole set of IMF.

a) Start a loop to obtain an IMF on a sliding window basis.

b) The first window size determines the number of iterations of the sifting process for the rest of the IMF. The stopping criterion for a given IMF is the resolution in dB as proposed in [3].

c) The process stops when the last segment is processed and the whole IMF is obtained

d) Continue obtaining IMFs until the stopping criteria for small windows is reached and duplicate the window size.

3) Once the window size equals the whole length, the process continues with a fixed size until the number of obtained extrema is less or equal than two.

4) Obtain the residual of the decomposition to have the whole EMD decomposition.

The main part of the algorithm is the outer While loop. It controls the stopping criteria for the EMD. The user must set how many oscillations are allowed in the last stage of the actual level. Once it is reached, the window size is duplicated. The process enlarges the size of the window adaptively as it is needed to analyze components with bigger wavelengths. Moreover, enlarging the window size allow a better interpolation between distant extrema.

Each IMF is calculated with a sliding window if the window size is smaller than the whole signal length. This situation is evaluated in the first part of the IF statement. The situation in which there is only one window for the whole signal is evaluated in the ELSE part.

Tapering segments applying a complementary symmetrical window to avoid discontinuities at the boundaries was studied. However, averaging the samples in the overlapping region produces good enough results with lower computation times. So, the criterion adopted was to average the samples in the overlapping region by both sides.

To implement this process it is necessary to take into account the following observations:

1) A minimal window size must be imposed to avoid an excessive number of partitions. The starting window size is correlated to the length of the signal to be analyzed. A simple possibility is making the starting segment a sub-power 2^{-k} of the signal length. In this manner, the last window will cover the whole length signal. It must be taken into account that, dividing the signal length by powers of two can lead to a non-integer size of the starting window; the integer part of the quotient is used. So the signal length is covered by an integer number of windows and a residual. Depending on its size this residual can be assigned to the last window, enlarging its size or constitute a new window, usually with different size to the previous one.

2) Regarding the criterion to determine when to enlarge the signal, it must be taken into account the fact that the minimum frequency to be analyzed depends on the window size (as it has been indicated before). As the sifting process requires oscillatory signals, it must be ensured that the window contains at least a minimum number of periods that can be selected by the user. The criterion adopted to change the window size is: The window length is duplicated the window size when the number of extrema in the previous IMF is lower than a user selected threshold.

3) Concerning the overlap region, it must be taken into account that it is necessary to reduce the undesirable end effects. Extrapolation is not used, since there are enough number of samples outside the actual segment. That implies that overlapping consecutive windows solve both, boundary and end effects.

4. Results

To evaluate the performance of the algorithm, it must be compared with the decomposition obtained by the reference algorithm (shown in Fig. 3). Secondly, the same signal is decomposed using the method proposed in this paper. The result can be observed in Fig. 4. A comparison of the computation time and the number of IMF obtained with both algorithms is presented in Table 2.

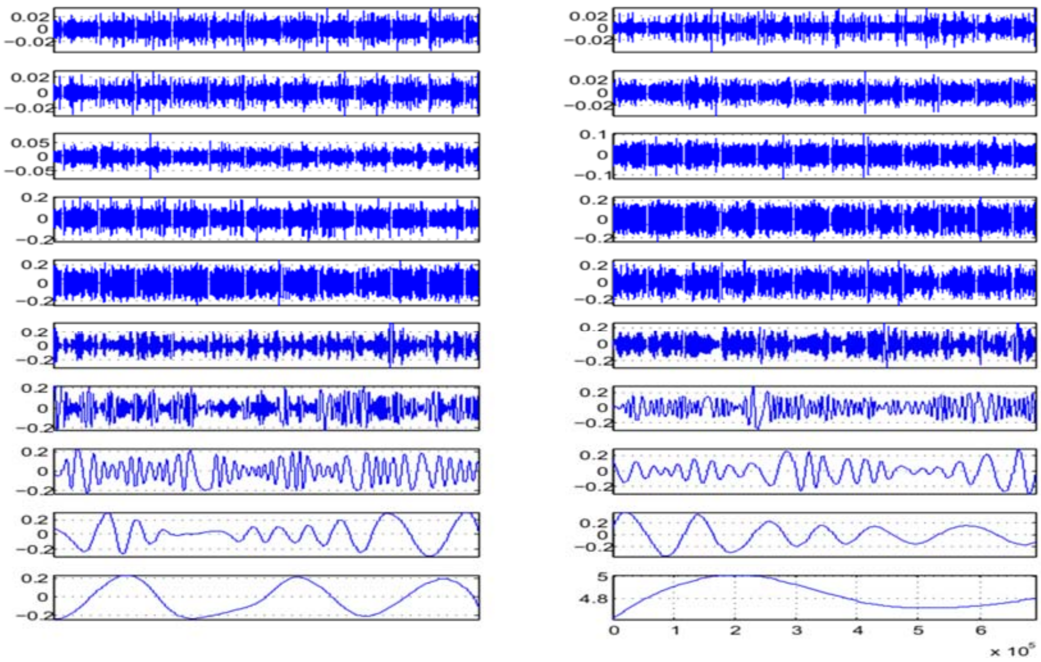

Fig. 3. EMD using reference algorithm.

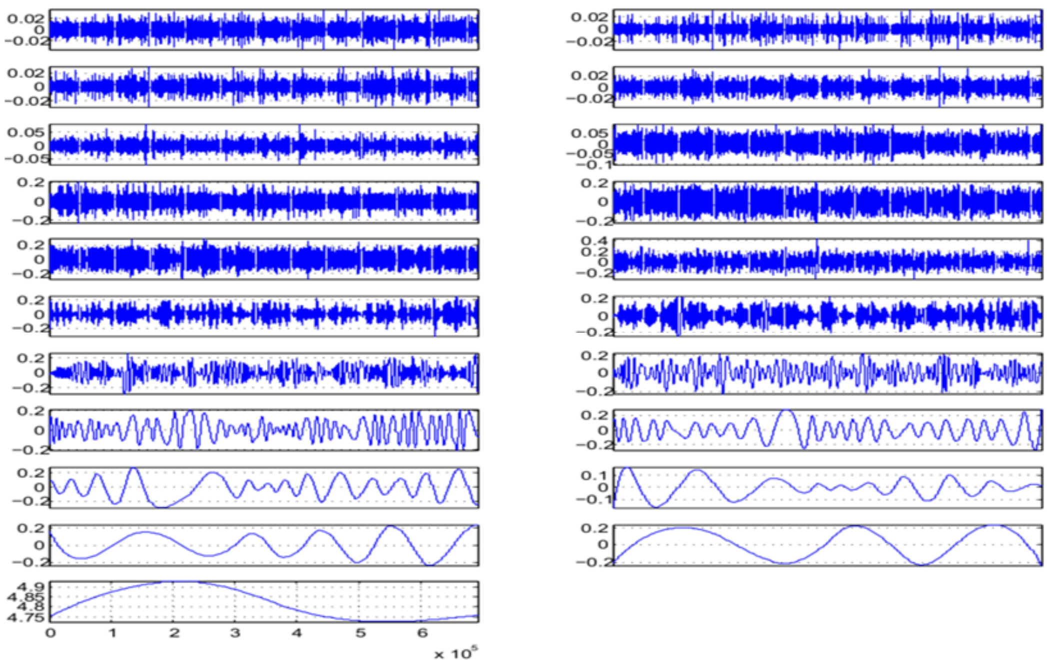

Fig. 4. EMD using algorithm proposed in this paper.

It must be taken into account that any error in a given IMF is propagated to the rest of the decomposition. As subsequent components have smaller amplitudes, the errors have greater importance. Our algorithm has shown a good behavior as it can be observed in 4. That is due to the fact that averaging the overlapped region between two consecutive windows smooths the result. As the number of samples in the overlapping region is based on a fixed percentage of the window size, the number of samples changes with the window size.

Comparing Fig. 3 and Fig. 4, it can be observed that Fig 4. Shows one component more in the region of low frequencies. That is due to increase the window length when low frequencies are reached.

In Table 2, it can be observed that the performance of the algorithm increases with time. While for 11,400 samples the sliding window computation time is 79 % of the EMD calculated as in the reference algorithm, for 691,800 samples the computation time is only 27 %. That is due to the fact that the window size increases by powers of two, which results in smaller running times for very long signals. Despite of the fact of using many windows for the calculation, the obtained IMFs show a high quality as no discontinuities can be observed in the last IMFs for a signal with more than 600,000 samples.

Table 2. Time comparison for both algorithms.

Resol	Length	IMFs	Time (s)	IMFs	Time (s)
45	11.400	14	28	14	17
45	27.600	15	41	14	37
45	114.000	18	363	16	159
45	340.800	20	1.057	19	452
45	691.800	22	3.075	20	923
50	11.400	13	29	12	23
50	27.600	16	69	15	50
50	114.000	19	602	18	217
50	340.800	22	1.774	20	669
50	691.800	22	5.195	21	1.414

5. Conclusions

The Empirical Mode Decomposition is a technique to decompose any signal into a finite set of narrowband components, the Intrinsic Mode Functions. The number of components and computational time increase dramatically when the length of the signal becomes large. Proposals for solving this problem had been done, but without the required quality. A modified sifting algorithm to deal with long signals was proposed here. It is based on computing every IMF using a sliding window. The algorithm is adaptive as both, the length of the sliding windows and the overlapping region depends on the signal to be analyzed. The change on the length of the sliding window by powers of two has two positive consequences. On one hand, the final window will cover the whole length of the signal in a few steps. On the other hand, the effectiveness of the proposed method increases with the length of the signal.

Acknowledgements

This work was partially funded by National Funds through the Foundation for Science and Technology of Portugal, under the project PEst-UID/ EEA/00066/2013 and by project MTM2013-417-P from Government of Spain.

References

[1]. Norden E. Huang, *et al*., The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis, in *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, London, Vol. 454, Issue 1971, 1998, pp. 903-995.

[2]. S. Peng, X. Hu, W. L. Hwang, Multicomponent AM-FM signal separation and demodulation with null space pursuit, *Signal, Image and Video Processing*, Vol. 7, Issue 6, 2013, pp. 1093-1102.

[3]. R. T. Rato, M. D. Ortigueira, A. G. Batista, On the HHT, its problems, and some solutions, *Mechanical Systems and Signal Processing*, Vol. 22, Issue 6, 2008, pp. 1374-1394.

[4]. R. T. Rato, M. D. Ortigueira, A. G. Batista, The EMD and its use to identify system modes, in *Proceedings of the International Workshop on New Trends in Science and Technology [CD-ROM]*, Ankara, Turkey, Nov. 03-04, 2008.

[5]. M. K. Hasan, K. M. S. Apu, M. K. I. Molla, A robust method for parameter estimation of ar systems using empirical mode decomposition, *Signal, Image and Video Processing*, Vol. 4, Issue 4, 2010, pp. 451-461.

[6]. P. C. Chu, C. W. Fan, N. Huang, Compact empirical mode decomposition: an algorithm to reduce mode mixing, end effect, and detrend uncertainty, in *Advances in Adaptive Data Analysis*, Vol. 4, Issue 3, 2012, pp. 1250017 (18 pages).

[7]. M. Ortigueira, Empirical Mode Decomposition [online] Available: http://www.mathworks.com/matlabcentral/fileexchange/21409-empirical-mode-decomposition. [Accessed: 18-Apr-2016].

[8]. K. M. Chang, S. H. Liu, Gaussian noise filtering from ECG by wiener filter and ensemble empirical mode decomposition, *Journal of Signal Processing Systems, Special Issue 'Signal Processing Circuits and Systems for Bio-Signals'*, Vol. 64, Issue 2, 2011, pp. 249-264.

[9]. M. A. Colominas, G. Schlotthauer, M. E. Torres, Improved complete ensemble EMD: A suitable tool for biomedical signal processing, *Biomedical Signal Processing and Control*, Vol. 14, 2014, pp. 19-29.

[10]. M. Feldman, Analytical basics of the EMD: Two harmonics decom position, *Mechanical Systems and Signal Processing*, Vol. 23, Issue 7, 2009, pp. 2059-2071.

[11]. X. Guanlei, W. Xiaotong, X. Xiaogang, Z. Lijia, Improved EMD for the analysis of FM signals, *Mechanical Systems and Signal Processing*, Vol. 33, No. 11, 2012, pp. 181-196.

[12]. H. Jiang, C. Li, H. Li, An improved EEMD with multiwavelet packet for rotating machinery multi-fault diagnosis, *Mechanical Systems and Signal Processing*, Vol. 36, Issue 2, 2013, pp. 225-239.

[13]. Z. K. Peng, P. W. Tse, F. L. Chu, An improved Hilbert-Huang transform and its application in vibration signal analysis, *Journal of Sound and Vibration*, Vol. 286, Issues 1-2, 2005, pp. 187-205.

[14]. N. U. Rehman, C. Park, N. E. Huang, D. P. Mandic, Emd via MEMD: Multivariate noise-aided computation of standard EMD, *Advances in Adaptive Data Analysis*, Vol. 5, No. 2, 2013, pp. 1-25.

[15]. P. Singh, P. K. Srivastava, R. K. Patney, S. D. Joshi, K. Saha, Nonpolynomial spline based empirical mode decomposition, in *Proceedings of the IEEE International Conference on Signal Processing and Communication (ICSC)*, Noida, India, December 2013, pp. 435-440.
[16]. Y. Yang, J. Deng, D. Kang, An improved empirical mode decomposition by using dyadic masking signals, *Signal, Image and Video Processing*, Vol. 9, Issue 6, 2015, pp. 1259-1263.
[17]. X. D. Yu, M. Y. Zhang, M. Q. Zhu, K. H. Xu, Q. C. Xiang, An improved extension method of EMD based on SVRM, *Applied Mechanics and Materials*, Vol. 543-547, 2014, pp. 2697-2701.
[18]. A. Eftekhar, C. Toumazou, E. M. Drakakis, Empirical mode decomposition: Real-time implementation and applications, *Journal of Signal Processing Systems*, Vol. 73, Issue 1, 2013, pp. 43-58.
[19]. P. Stepien, Sliding window empirical mode decomposition – its performance and quality, *EPJ Nonlinear Biomedical Physics*, Vol. 2, Issue 1, 2014, pp. 2-14.
[20]. F. Ebrahimia, S. K. Setarehdana, H. Nazeranb, Automatic sleep staging by simultaneous analysis of ECG and respiratory signals in long epochs, *Biomedical Signal Processing and Control*, Vol. 18, 2015, pp.69-79.
[21]. Md. A. Kabir, C. Shahnaz, Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains, *Biomedical Signal Processing and Control*, Vol. 7, Issue 5, 2012, pp. 481-489.
[22]. J. L. Sanchez, Manuel D. Ortigueira, Raul T. Rato, Juan J. Trujillo, An Improved Empirical Mode Decomposition for Long Signals, in *Proceedings of the First International Conference on Advances in Signal, Image and Video Processing (SIGNAL'16)*, Lisbon, Portugal, June 26-30, 2016.