Int. J. Nonlinear Anal. Appl.
Volume 11, Special Issue, Winter and Spring 2020, 183-193
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/IJNAA.2020.4587

A New Method for Reducing End Effects in Empirical Mode Decomposition

Javad Zarekar[a*], Gholamhasan Payganeh[b], Mehrdad Nouri Khajavi [c]

[a] *PhD Student, Department of solid Mechanic, Faculty of Mechanical Engineering, Shahid Rajaee University, Tehran ,IRAN.*
[b,c] *Associate Professor, Department of solid Mechanic, Faculty of Mechanical Engineering, Shahid Rajaee University, Tehran ,IRAN.*

Abstract

Real time vibrational signal processing is one of the fault detection methods for the mechanical systems. The Hilbert–Huang Transform (HHT) is a newly developed robust method for analyzing nonlinear and non-stationary vibrations based on time-frequency distribution. This approach is based on Empirical Mode Decomposition (EMD) and Hilbert spectral analysis. This paper presents a state-of-the-art method for decomposing a signal into a set of so-called Intrinsic Mode Functions (IMF). The proposed alternative method is based on the change in the screening algorithm. This modified method is useful to mitigate end effects and reduces the calculation load and time. The effectiveness of this method was validated by numerical simulation. The results show the accuracy and reliability of this method.

Keywords: End Effects ,Hilbert–Huang Transform (HHT), Empirical Mode Decomposition (EMD), Intrinsic Mode Functions (IMF).

1. Introduction

The development and use of signal processing and analysis methods that can detect faults in rotary machines is of great importance [1], including the existing linear and stationary signal processing techniques such as time analysis and frequency analysis. These techniques often lead to incorrect information when used for non-linear and non-stationary signals, because many mechanical defects are evidently non-linear and non-stationary and are caused by transient phenomena [4]. The Empirical Mode Decomposition (EMD) is one of the robust methods for analyzing nonlinear and non-stationary

*Corresponding Author: Javad Zarekar
Email address: jzarekar@yahoo.com, g.payganeh@sru.ac.ir,mehrdadnouri@yahoo.com (Javad Zarekar[a*], Gholamhasan Payganeh[b], Mehrdad Nouri Khajavi [c])

Received: February 2020 *Revised:* May 2020

signals and is based on the local characteristics of the time scales of a signal and decomposes the signal into a series of complete and orthogonal components called the Inherent Mode Functions (IMF). IMFs represents the natural oscillatory modes of the signal and are considered as base functions, determined by the signal itself and are calculated from pre-specified cores.

2. Theoretical Basics and Background

EMD is a self-adaptive method for signal analysis and is suitable for analyzing non-linear and non-stationary phenomena. It has been widely used in voice recognition, biology, and system identification among others. Signals can be divided into three general categories in terms of the type of analysis, time domain, frequency domain, and time-frequency domain. Time domain and frequency domain techniques are employed to process vibration signals and fault detection. Since the spectral distribution of an unstable signal varies over time, use of time-frequency methods are necessary for the analysis of such signals, such as Cohen's spectral time-frequency distributions [1], the short-time Fourier transform (STFT) [2], and wavelet transform-based time frequency analysis [3]. Recently, Huang et al. [4, 5] provided a new time-frequency distribution based on the Hilbert-Huang transform, which has advantages over other distributions. This method was first introduced in 1998 for the analysis of nonlinear and non-stationary data [4] and was used in various fields such as geophysical [6], climatic [7], and oceanographic studies [8] as well as engineering applications, including fault detection. The HHT is composed of two main parts: EMD and Hilbert transform. A significant amount of work has been carried out on both parts. EMD Studies are divided into two categories: those seeking to find an alternative method for EMD [12] and, those seeking modify the decomposition method [13]. The way the main signal is decomposed into the IMFs by Intrinsic Mode Decomposition (IMD) is the main advantage of HHT to other methods. Therefore, (self-adaptability), optimization and improvement of this part of HHT are very important. Studies on the analysis of experimental modes and their constraints can be dealt with in different ways. The stopping criterion for determining the IMF in the screening process has been frequently discussed. Occasionally, the proposed stopping criteria lead to non-convergence of the screening process. Thus, Huang et al. [4] have proposed stopping criteria with different theoretical foundations in subsequent studies [14]. Since there is no certain accurate criterion for choosing the best IMF, the stopping criterion is regarded as one of the crucial parameters in the screening process that can play a key role in the proper selection of the IMFs. Moreover, a great deal of studies has been conducted on the sampling frequency [15] and the method of determining the local average among others. Undoubtedly, the method of determining the local average in EMD is the main part of the EMD method. However, the signal end effects have been one of the most effective limitations [16]. In the Huang's EMD algorithm, there is a distortion problem with the cubic spline interpolation and the average envelope in the end points of the signal, called the end effect problem. In the EMD signal processing, the end effect problem is a significant oscillation, which ultimately results in undesirable signal decomposition. For separation of the IMFs, the average should be determined for several consecutive times and the change in the method of calculating the local average results in different results. Therefore, any fault, such as end effects, is repeated and adverse results are achieved. Researchers have developed a few techniques to reduce end effects, most of which focuses on the behavior of endpoints according to other points of the signal. Currently, data continuation-data generalization methods are often used to reduce end effects. A direct reversal of data [7] is an operation in which two endpoints of a signal are selected such as the position of two mirrors and extends them beyond the two mirrors in the outer directions and generates a new signal with twice the original length. However, in [8] expansion of extremums is carried out using support vector regression at both end points.In [10], the Ratio Boundary Extension method has been proposed in which a simultaneous mirror (reflective) extension and an extrapolation are used to solve the end effect problem. In this method, the ration of distance between the closest consecutive information is used to predict the signal and obtain extra points at the boundaries. This method is characterized by the simple and practical calculations. Zhang [21] proposed the use of the EEMD method for signal analysis , however, this method has only been relatively useful for solving the interference problem. Deng [22] used a single-layer and single-neuron neural network to extend the main signal to solve the boundary problem. This algorithm is relatively good for most signals, however, it has a very low speed. Huang et al. [23] proposed a closed mirror expansion method for the original data, however, it is not efficient in short time series and has a poor performance. Qi et al. [24] presented the frame window that is a classic approach based on the selection of a special window type, but this method is widely used in the Fourier spectral analysis. Furthermore, in this method, significant information are lost near the two ends of the signal in the pre-processing. Zheng et al. [25] presented a new mathematical model based on the gray prediction model for the Boundary Extension (BE), but this method does not have the ideal risk-taking and predictive accuracy. The autoregressive model [26,27] is also a new idea of the BE that is determined by the description of the input/output relationship as linear equations with variable time coefficients. The AR model requires a little prior information and uses relatively simple computations. Tang et al. [28] presented a least squares support vector regression for obtaining data.Lee et al. [29] proposed an autoregressive integrated moving average for predicting the signal time series.Ma et al. [30] used the genetic algorithm to solve the problem of selecting RBF neural network parameters to solve the end effects problem. These methods are effective in mitigating the end effect, but they are not a very optimal solution due to the complexity of the existing models and the uncertainty of their parameters. It seems that most of the above-mentioned signal processing techniques seek to change the results and manipulate the results of the EMD method. Moreover, during these generalizations and developments, the advantages of EMD such as decomposition of nonlinear and non-stationary signals as well as its self-adaptation may be affected and the results are not entirely accurate. In our proposed method, the outputs obtained by the EMD method are not manipulated and only by changing the EMD algorithm, in such a way the advantages of this method are not affected, the problem of end effects is solved. In the present study, a new method for calculating the local average is introduced and then, an EMD is carried out. Based on the proposed method, the average method is not used to obtain IMFs. In this method, by changing the EMD algorithm the problem of the final effects is solved.

3. EMD algorithm:

The EMD method is an adaptive tool for analyzing non-linear and non-stationary signals, using which, each signal is decomposed based on its local behavior and produces IMFs, each of which introduces a simple oscillation in comparison to a simple harmonic function. The data for each signal contains very different oscillatory modes that interfere with each other and generate complex data. Each intrinsic, linear, or nonlinear mode is a simple oscillation that has the extreme points (maximum and minimum) and the same zero-crossing points. In other words, the oscillations around the spatial average are symmetric. To obtain the IMFs, the following algorithm is run [6]:

1) 1. All primary signal extremes $x(t)$ are identified.
2) By the curve interpolation on the points of maximum and minimum of two curves, upper e_{max} and lower e_{min} are obtained.
3) 3. The value of $m(t)$ is calculated.

$$m(t) = \frac{e_{max} + e_{min}}{2} \tag{1}$$

4) The value of h_1 is calculated using the following equation, and this operation continues until the value of $h_1(t)$ becomes lower than the desired value of $d_1(d_1(t) <\in)$.

$$\begin{aligned} h_1(t) &= x(t) - m_1 \\ \mathrm{d}_1(t) &<\in \end{aligned} \tag{2}$$

5) If $h_1(t)$ satisfies the following two conditions, then the first IMF *or* IMF_1 is considered and otherwise the steps are repeated until the above condition is satisfied in order to obtain the first IMF.
   First condition: Over the entire length of the data, the number of extremes and zeroes of the signal must have at most one difference.
   Second condition: The average value of the local domain, the maximum and minimum of each signal segment are equal.

6) The remainder is calculated as follows:

$$r_1(t) = x(t) - \mathrm{IMF}_1(t) \tag{3}$$

By repeating this algorithm, the signal components are calculated, for example, the initial signal can then be calculated as follows:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{4}$$

Where r_n is the remaining component, after n of IMF and c_i is the IMF. By decomposition of the signal, components of high-frequency and low-frequency waves are obtained that their composition reconstructs the original signal. The components of the IMFs in each step have a lower frequency than the previous step. Obtaining IMFs is known as "screening process". This process continues until the standard deviation parameter limits it. The standard deviation is obtained using the following two methods:

A. following equation:

$$S_D = \sum_{t=0}^{T} \frac{(h_{k-1}(t) - h_k(t))^2}{h_{k-1}^2(t)} \tag{5}$$

B. The standard deviation value is usually selected from 0.2 to 0.3.

The EMD method can decompose a complex signal into a set of IMFs that are orthogonal to each other. It seems that the end effect is one of the most important disadvantages of EMD method.

4. Shortcoming in the screening process in the extraction of IMFs:

The screening process in the analysis of empirical signals is accompanied by shortcomings, one of the most important of which is the end effects. For example, take the following example:

1. Consider the signal with the equation $y = 3 * cos(0.5 * x) + 0.5 * sin(3 * x)$. Obviously, the signal is composed of two components that are plotted below the signal and its components in the interval [0.10 pi]. As discussed in the previous section, the EMD method is able to isolate these components from the original signal. The results of this extraction and comparison with the main signal and the main components are given below. The results show that the extracted components at the beginning and end of the IMFs have a dramatic difference with the actual signal components, which is one of the important disadvantages of HHT. This paper presents a method to solve this problem.

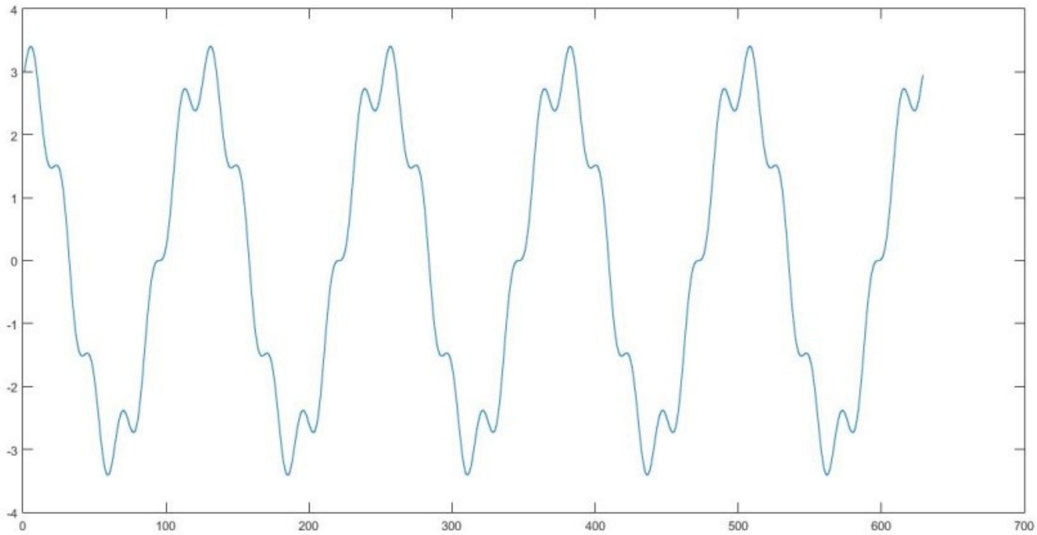


Figure 1: simulated signal [$y = 3 * cos(0.5 * x) + 0.5 * sin(3 * x)$]

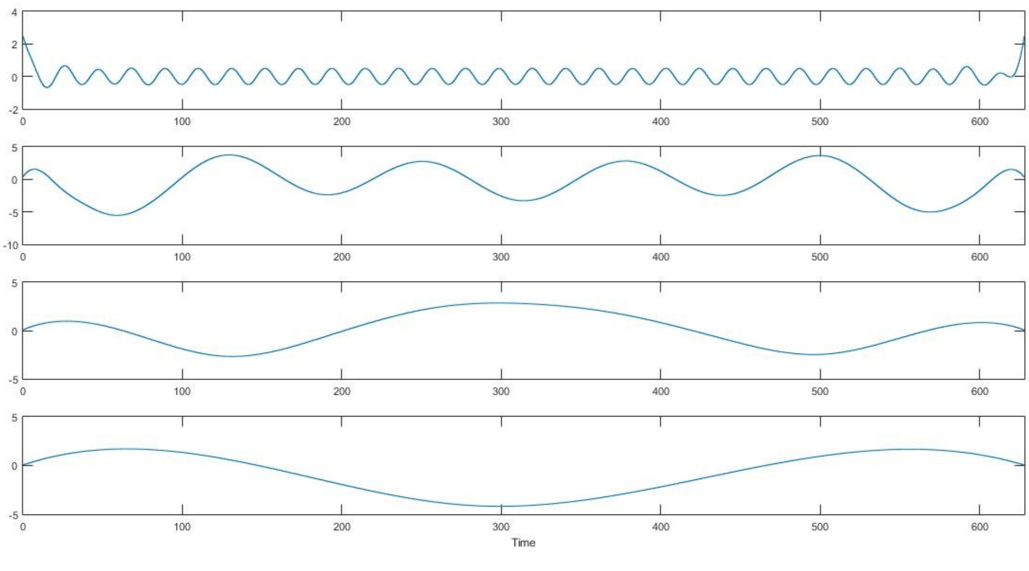


Figure 2: IMFs extracted by classic EMD

5. MEDIAN method in curve interpolation:

5.1. Introduction of the modified (new) algorithm and its effect on EMD.

The local average is defined by the conventional EMD method as the average of the curve envelops obtained from the extremums and with the Cubic spline interpolation. In the references, mathematical reasoning is not provided for this definition, and only the qualitative relationship between envelops and the energy of the signal is mentioned. This section introduces a new algorithm that has advantages over the conventional (classical) definition. Statisticians believe that if our datasets are proportional and uniform (or so-called symmetric), then the average will be good indicator for the whole set. However, if our set has heterogeneous and asymmetric numbers, then it would be better to rely on the median to represent these numbers and figures. One of the disadvantages is that it is affected by outlier data. Mathematical analyses reveal that if the mean is replaced by the mean in EMD algorithm, better results are achieved. Thus, the mean is used to extract the components at all stages of the EMD algorithm.

5.2. New EMD algorithm with MEDIAN replacement

Example 1: Comparison of the components of a signal with its extracted IMFs by the classical (average) method and the proposed new method (median):

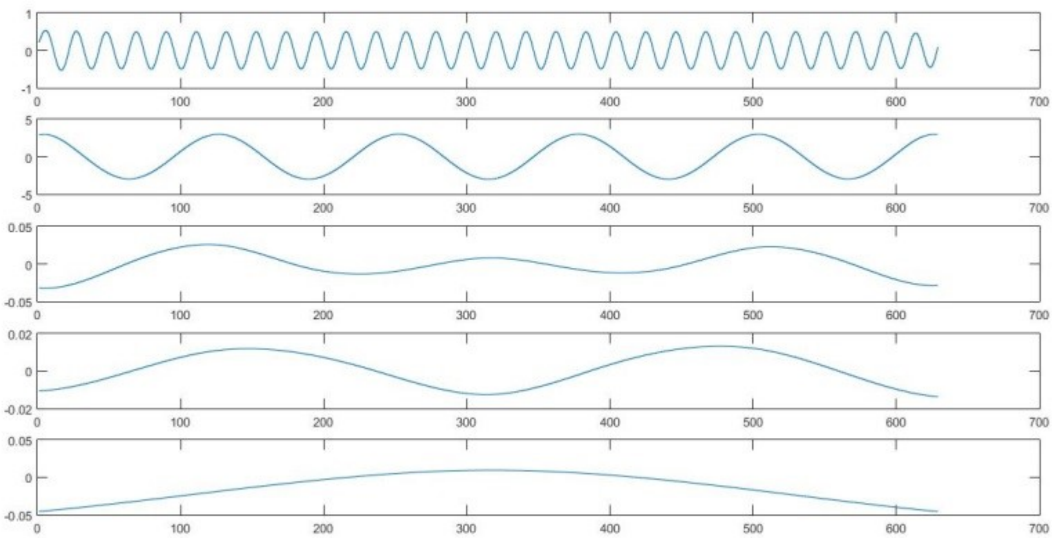


Figure 3: Extracted IMFs by proposed EMD

IMFs are extracted by three methods:original signal , the classical method and our proposed method . Obviously, the main equation consists of two parts, which are as follows:

$y = 3 * cos(0.5 * x) + 0.5 * sin(3 * x)$

Therefore, after the original signal is decomposed by signal decomposition methods, they must be decomposed into these two components. In figure4 by matching the components extracted by the classical method and the proposed method on the original signal, the error of each is visible. As can be seen, in the middle of the signal, all three items overlap, but at the sides, ie the beginning and end of the signal, the classical method has a sharp deviation from the main signal.[figure5,6] . This happens in both decomposed IMFs. The third IMF extracted by the classical method must be close to zero because it is residual. But this is not the case and the extraction and decomposition of the classical method is such that it seems that the extra IMF has been extracted while the main signal has two components. Therefore, two IMFs must be extracted for it, plus a residue that is closer to zero, indicating the accuracy of the analysis.

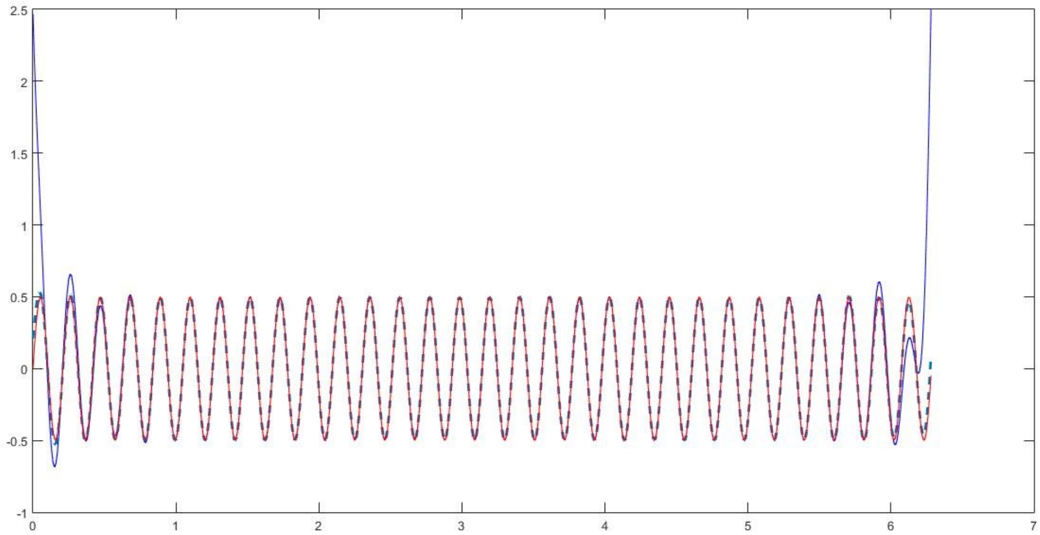
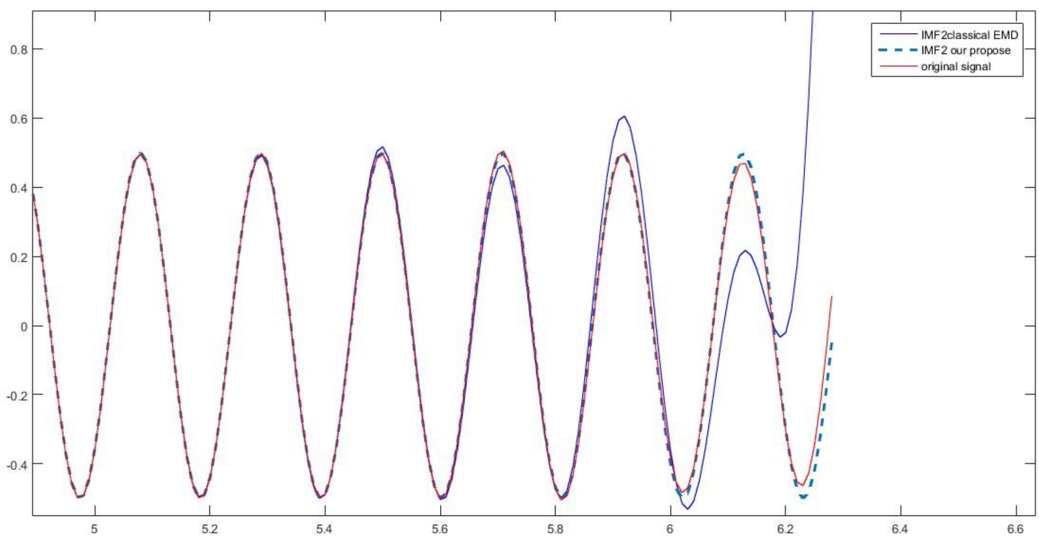
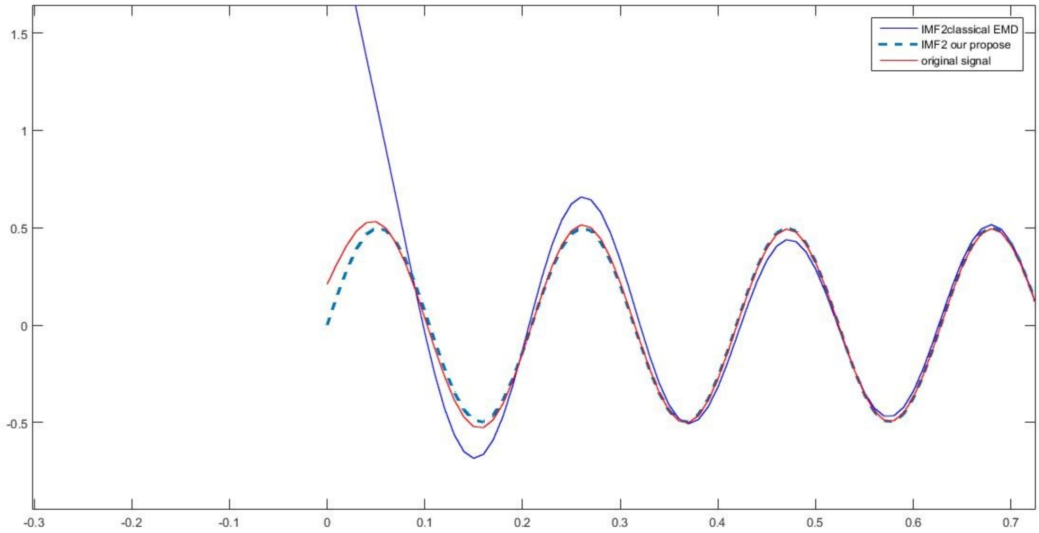


Figure 4: EMD results of the simulated signal,Matching IMF_1's



(a)



(b)

Figure 5: EMD results of the simulated signal with (a)

Match the end of signal, (b) Match the beginning of the signa

l

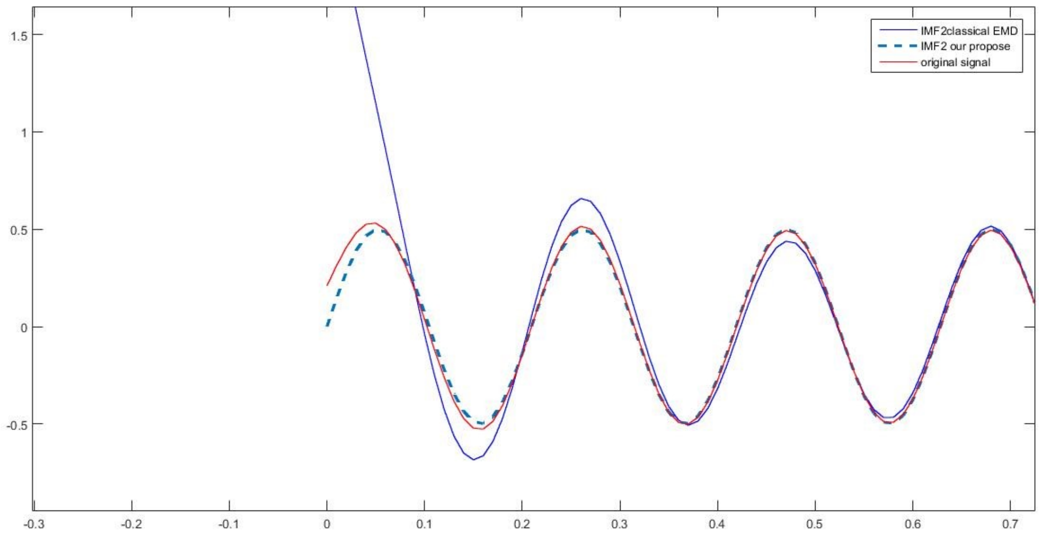


Figure 6: results of the simulated signal,Matching IMF2's

Figures [6] clearly show how the ends and the beginning of the decomposed signals overlap. The main signal (red) of the classic method (in blue) and our proposed method are plotted and displayed on each other. As can be seen on the sides of the signal, the classical method has a large deviation from the original signal. This happens in both decomposed functions. Also, the residue should be close to zero, but this is not the case in the classical method and it has been modified with our proposed method.

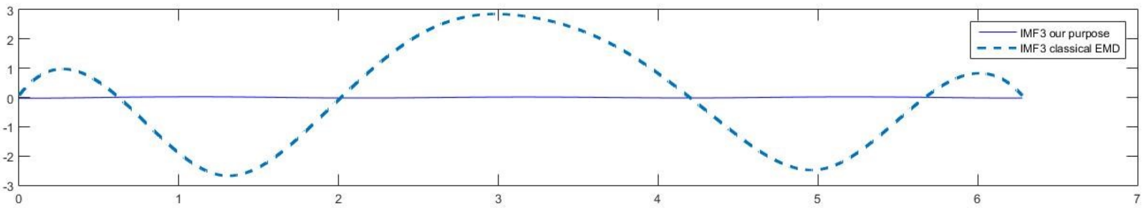


Figure 7: results of the simulated signal, Matching Residual signal(IMF_3)

Table 1: Reliability test and comparison of IMF'S with two methods

MSE	RMSE	MAPE	MAE	IMF'S	Decomposition Methods
1.4466	1.2027	9.7256	0.8981	MEAN IMF1	Classic Method (MEAN EMD)
0.0572	0.2393	0.9217	0.0509	MEAN IMF2	
0.0010	0.0329	0.0826	0.0194	MEDIAN IMF1	Proposed Method (MEDIAN EMD)
0.0003	0.0175	0.0642	0.0056	MEDIAN IMF2	

According to Table [1] Among the two methods used for signal analysis, it is observed that the method presented in this article works with much less error than the classical method. This result is obtained in each of the extracted IMFs. Especially in IMF1 according to the first row of the table, it can be seen that all error measurement methods indicate very little error in the proposed method. On the contrary, the error of the classical method is worth considering.

6. Conclusion:

For analysis of nonlinear and non-stationary signals, a number of methods are used that are based on feedback loops, but the EMD is a very robust tool, which, unlike other methods, is based on comparative analysis and certain basic functions for this method are not predefined. The signal decomposition by this method leads to the IMFs that have physical concepts. However, the end effects problem is one of the disadvantages of this method that leads to error in correct extraction of the signal components. In this study, with the change in the EMD algorithm, the EMD process has been improved. By providing examples, it was shown that by replacing the average by the mean, very good results are obtained in mitigating the end effects. The results demonstrate that the method used to mitigate the end effects works better and more accurately than the classical method.

References

[1] Cohen, L., ”Time-frequency analysis: theory and applications”, Prentice Hall, 1995.

[2] Protnoff M. R., ”Time-frequency representation of digital signals and systems based on short-time fourier analysis”, IEEE Transactions on Acoustics, Speech, Signal Processing, Vol. 28, No. 1, pp. 55-69, 1980.

[3] Qian, S. E., ”Introduction to time-frequency and wavelet transforms”, Prentice Hall, 1st ed., Nov 2001.

[4] Huang, N. E., Shen, Z., Long, S. R., Wu, M. L., Shih, H. H., Zheng, Q., Yen, N. C., Tung and C. C., Liu, H. H., ”The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis”, Proceedings of Royal Society London A, Vol. 454,pp. 903–995, 1998.

[5] Huang, N. E., and Shen, S. S. P., ”Hilbert-Huang transform and its applications”,Interdisciplinary Mathematical Science, Vol. 5, World Scientific Publication Company, 2005.

[6] Huang, N. E., Chern, C. C., Huang, Salvino, K., L., Long, W., S. R., and Fan K. L., ”A new spectral representation of earthquake data: Hilbert spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999”, Bulletin of the Seismological Society of America, Vol. 91, pp. 1310-1338, 2001.

[7] Lundquist, J. K., ”Intermittent and elliptical inertial oscillations in the atmospheric boundary layer”, Journal of the Atmospheric Sciences, Vol. 60, pp. 2661-2673, 2003.

[8] Wu, C. H., and Yao, A. F., ”Laboratory measurements of limiting freak waves on currents”, Journal of Geophysical Research, Vol. 109, 2004.

[9] Lin, S., Yang, J., and Zhou, L., ”Damage identification of a benchmark building for structural health monitoring”, Smart Materials and Structures, Vol. 14, pp. 162-169, 2005.

[10] Boualem Boashash„ ”Estimation and interpreting the instantaneous frequency of a signal - part 1: fundamentals”, Proceedings of the IEEE, Vol. 80, No. 4, 1992.

[11] Tao Qian and Sprobig, W., ”Mono components for decomposition of signals”, Mathematical Methods In Applied Science, Vol. 29, pp. 1187- 1198, 2006.

[12] Olhede, S., and Walden, A. T., ”The Hilbert spectrum via wavelet projections”, Proceedings of Royal Society London A, Vol. 460, pp. 955-975, 2004.

[13] Chen, Q., N., Huang, E. S., Riemenschneider and Y. Xu., ”A B-spline approach for empirical mode decomposition”, Advances in Computational Mathematics, Vol. 24, pp. 171–195, 2006.

[14] Huang, N. E., Wu, M. L., Long, S. R. S., Shen,S., W. D. Qu, Gloersen, P., and K., Fan, L., ”A confidence limit for the position empirical mode decomposition and Hilbert spectral analysis”, Proceedings of Royal Society London A, Vol. 459, pp. 2317–2345 2003.

[15] Gabriel Rilling and Patrick Flandrin, ”Sampling effects on the empirical mode decomposition”, Advances in Adaptive Data Analysis, Vol. 1, No. 1, pp. 43-59, 2009.

[16] Wu, F., and Qu, L., ”An improved method for restraining the end effect in empirical mode decomposition and its applications to the fault diagnosis of large rotating machinery”, Journal of Sound and Vibration, Vol. 314, pp. 586-602, 2008.

[17] .Y. S. Lee, S. Tsakirtzis, A. F. Vakakis, L. A. Bergman, D.M. McFarlan, Physics-based foundation for empirical mode decomposition. Adv. in Adaptve Data Analysis, Vol. 47, No. 12 (2009) 2938–2963.

[18] .D. Lin, Z. Guo, F. An, F. Zeng, Elimination of end effects in empirical mode decomposition by mirror image coupled with support vector regression. Mechanical Systems and Signal Processing 31 (2012) 13–28

[19] . D. Ren, S. Yang, Z. Wu, G. Yan, Evaluation of the EMD end effect and a window based method to improve EMD. International Technology and Innovation Conference, November 6-7 (2006) China.

[20] . Q. Wu, S. D. Riemenschneider, Boundary extension and stop criteria for empirical mode decomposition. Adv. in Adaptive Data Analysis, Vol. 2, No. 2 (2010) 157–169.

[21] ZHANG Meijun, TANG Jian, ZHANG Xiaoming, et al. Intelligent Diagnosis of Short Hydraulic Signal Based on Improved EEMD and SVM with Few Low-dimensional Training Samples[J]. Chinese Journal of Mechanical Engineering, 2016, 29(2): 396–405.

[22] DENG Y J, WANG W, QIAN C C. Boundary-processing-technique in EMD method and Hilbert transform[J]. Chinese Science Bulletin, 2001, 46(11): 954–961.

[23] HUANG D J, ZHAO J P, SU J L. Practical implementation of Hilbert-Huang transform algorithm[J]. Acta Oceanologica Sinica. 2003, 22(1): 1–14.

[24] QI K, HE Z, ZI Y. Cosine window-based boundary processing method for EMD and its application in rubbing fault diagnosis[J]. Mechanical Systems and Signal Processing, 2007, 21(7): 2750–2760.

[25] ZHENG K, WANG X, YANG Y, et al. Applying multiple residual error gray forecast to restrain end-points effect in HHT of network traffic[C]//1st International Conference on Pervasive Computing, Signal Processing and Applications, Harbin, China, September 17-19, IEEE Computer Society. 2010: 1057–1060.

[26] ZHANG Qingjie, ZHU Huayong, SHEN Lincheng. Semi-parametric regression model prediction method based on empirical mode decomposition[C]//2nd International Conference on Advanced Computer Control ,IEEE, Shenyang, China, Junuary 29-31, vol. 2 (2010): 330–335.

[27] AH Costa, HENGSTLER S.. Adaptive time–frequency analysis based on autoregressive modeling[J]. Signal Processing, 2011, 91 (4): 740–749.

[28] TANG Ling, YU Lean, WANG Shuai. A novel hybrid ensemble learning paradigm for nuclear energy consumption forecasting[J]. Applied Energy, 2012, (93.5): 432–443.

[29] LEE Yi-shian, TONG Lee-Ing, Forecasting time series using a methodology based onautoregressive integrated moving average and genetic programming[J]. Knowledge-Based Systems, 2011, (24): 66–72.

[30] MA Jinghua, JIANG Hong, YAO Li. End effects processing of Hilbert–Huang transform based on genetic algorithm and RBF neural network[C]//3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), IEEE, Chengdu, China, July 9–11, vol. 4 (2010): 312–316.