Nuclear Engineering and Design 323 (2017) 133–141

Contents lists available at ScienceDirect

Nuclear Engineering and Design

journal homepage: www.elsevier.com/locate/nucengdes

An advanced statistical method to analyze condition monitoring data collected from nuclear plant systems

Bo Li, Mahesh. D. Pandey*

Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada

ARTICLE INFO

Keywords:
Time series analysis
Noise separation
Trend detection
Statistical significance test
Empirical Mode Decomposition
Condition monitoring data

ABSTRACT

Condition monitoring data are routinely collected from various nuclear plant systems to ensure they are operating within an acceptable envelope, and to detect any potential onset of degradation in the system condition. The condition monitoring includes periodic monitoring of not only physical variables, such as temperature and vibration, but also chemical properties of lubricants, oils, and other control fluids. The time series of such monitoring data tend to exhibit non-stationary nature and complex correlation structure, as they consist of fluctuations of different time scales and noise. Since standard text-book methods of stationary time series analysis are not applicable to such data sets, the paper presents an advanced method of Empirical Mode Decomposition (EMD) to filter out the noise and identify the long-term trend, i.e., a likely indicator of degradation, in condition monitoring data. The proposed method is verified by a simulation example and then applied to a real data set obtained from an operating nuclear plant.

1. Introduction

1.1. Background

There are various kinds of small or large electrical, mechanical, chemical and structural systems that are employed for safe and reliable power generation at a nuclear station. In order to maintain the reliability of these systems and the plant as a whole, wide ranging performance indicators, such as temperature, pressure, and key chemical or physical properties, are carefully monitored and maintained within acceptable limits. In order to control adverse effects of degradation in a given plant system, it is of interest to examine the condition monitoring data and detect a trend that is reflective of aging. The detected trend can be used to assess the effectiveness of ongoing maintenance program. Once the onset of a degradation process becomes evident, planning mitigating actions, e.g., improvement of maintenance program or overhaul of the system, can be duly initiated. An accurrate method for detection of the trend function is especially important, since it determines the time of preventive maintenance before system's performance can become unacceptable. Naturally, an overestimation of the trend will lead to premature maintenance work whereas an underestimation of the trend can have adverse safety consequences.

1.2. An illustrative example

An example of such condition monitoring data comes from the electro–hydraulic control (EHC) system used in the operation of turbines in a nuclear plant. In an EHC system, phosphate esters are commonly used as control fluid due to their favourable fire retardant property. The degradation of control fluid with ongoing usage tends to produce acids which can lead to many adverse effects, such as corrosion of components, reduction in resistivity, acceleration of further degradation, and development of other unfavourable chemical reactions (e.g., formation of soap). The most critical fluid property affected by degradation is the total acid number (unit: mg KOH/g). In order to maintain the effectiveness of the control fluid, the total acid number should be kept below 0.2 mg KOH/g. For this reason, the chemistry of the control fluid system is closely monitored by collecting and examining samples taken on a weekly basis from the EHC system.

A sample time series of the total acid number collected from an EHC system is presented in Fig. 1, which shows the presence of short-term fluctuations, noise and trends in some segments of the data. While the short-term fluctuations can be caused by variations in chemical processes and the environment, any trends may be reflective of degradation in the condition of the control fluid that accumulates over time.

Given that the raw data consist of non-stationary trends, short-term cycles, and noise, a suitable statistical method is required to extract the long-term trend, which would be indicative of degradation in the

* Corresponding author.
E-mail address: mdpandey@uwaterloo.ca (M.D. Pandey).

http://dx.doi.org/10.1016/j.nucengdes.2017.08.003
Received 22 March 2017; Received in revised form 27 July 2017; Accepted 7 August 2017
Available online 30 August 2017

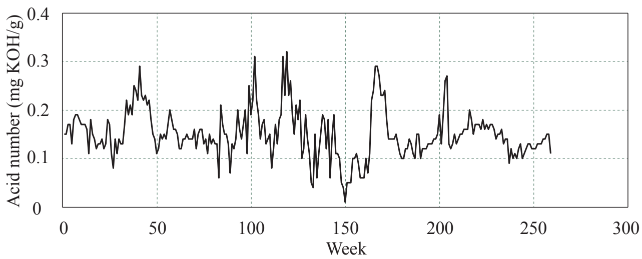

Fig. 1. A sample time series of total acid number data from a turbine EHC system.

condition of the control fluid. Based on this long-term trend, maintenance actions can be planned on a more informative basis. The ordinary least-squares regression method is generally inadequate for this type of data, since the noise term can have a complex time-dependent auto-correlation structure.

1.3. Objective and organization

The main objective of this paper is to present a highly flexible and efficient method to extract the long-term trend from a non-stationary and noisy time series of condition monitoring data, such as the total acid number of EHC control fluid. Recognizing a complex interaction of short-term cycles, correlation, and random noise in the data, the paper proposes a more advanced method of time series analysis, i.e., the EMD method. Although this method has been used in climate and geophysical sciences, its engineering applications have been quite limited. Therefore, the presentation of this method to nuclear engineering community is expected to promote applications to other problems and data sets.

The paper is organized as follows. The basic overview of the EMD method along with an illustrative simulation example are provided in Section 2. Section 3 presents an elaborate case study based on condition monitoring data collected from a turbine EHC system, and an example to further validate the EMD method. Conclusions of this study are summarized in the last section of the paper.

2. Empirical Mode Decomposition (EMD) method

2.1. Rationale for using the EMD method

Conceptually, a time series, $x(t)$, of the condition monitoring data can be represented by a sum of three basic components:

$$x(t) = r(t) + x_S(t) + y(t), \quad (1)$$

where $r(t)$ represents the long-term trend in the data, $x_S(t)$ represents the actual data without any trend or noise, and $y(t)$ represents the random noise. The main goal of the data analysis is to separate the noise and isolate the long-term trend in the actual data.

The classical models of time series data, such as autoregressive (AR) models, moving average (MA) models, and their various combinations are applicable to stationary time series without any trend, i.e., zero mean data. However, these models are not directly applicable to data with non-stationary and nonlinear nature, such as the acid number data in the turbine EHC system.

A common approach for identifying a long-term trend is the statistical regression method. However, the problem with this approach is that the form of the trend, e.g., linear, has to be pre-specified to the original data. This makes the approach inadequate, as it imposes a restrictive form on the data without considering the nonlinear and non-stationary nature of the data generating mechanisms. The trend is extrinsic and predetermined, and hence subject to specification errors. The least-squares regression method is also not applicable to correlated time series data. Similar to the regression method, the moving average method requires a predetermined time window to carry out the averaging, which may be difficult to identify for non-stationary data. In summary, these methods are collectively referred to as extrinsic methods, as they need to specify a priori functional form to the trend in terms of a fixed number of parameters.

Noise is an inevitable part of a real data set. It usually masks the overall trend and limits the ability to extract true information from the data. In case of linear and stationary time series, standard filtering methods, such as Fourier transforms, are effective in removing noise terms of different frequencies. However, these filtering methods are invalid for data generated by processes that are either nonlinear or non-stationary. In particular, the mixing of harmonics of real signals with the harmonics of noise, makes the filtering methods ineffective for noise separation.

The statistical process control (SPC) methods are also used to detect a change in the trend and any excursion of a statistical descriptor of the data, such as mean, beyond some control limits (Stoumbos et al., 2000; Oakland, 2007). However, the SPC suffers from all those limitations that are stated in previous paragraphs because of the following reasons: (1) data are assumed to be samples of a stationary process, (2) data are assumed to follow the normal distribution, (3) the level of noise in data is assumed to be fairly small, (4) a regression model is utilized in the analysis, (5) statistical measures of the data, e.g., mean, variance and various capability indices, are calculated for an assumed size of the window of the data. Naturally, the SPC will not be applicable to non-stationary data with noise masking the trend.

Since the introduction of the Hilbert-Huang Transform (Huang et al., 1998), the EMD method has emerged as an effective method for decomposing a time series into a set of basic constituent functions, known as intrinsic mode functions (IMFs). The IMFs are derived from the raw data without relying on any extrinsic basis functions (i.e., trigonometric series) or simplifying assumptions about the linear and/or stationary nature of the time series. For example, Wu et al. (2007) analyzed the time series of annual global surface air temperature and extracted the overall adaptive trend using the EMD method. The overall adaptive trend is the residual component of the data, which spans over the whole length of the time series. The EMD method has been successfully used in many other studies to detect trends (Boudraa et al., 2004; Capparelli et al., 2013; Liang et al., 2005; Qian et al., 2011) and other critical information in the time series (Loh et al., 2001; Li et al., 2016).

The EMD method is a data driven approach for time series analysis, which can be applied to non-stationary time series with minimal assumptions. Because it overcomes various limitations associated with the standard methods of time series analysis, it is adopted in this paper to analyze general types of condition monitoring data.

2.2. Basic approach

The EMD method is based on the premise that a time series is a superposition of different, simple, and oscillatory modes, referred to as IMFs, and a residual term or the trend, given as Huang et al. (1998):

$$x(t) = \sum_{k=1}^{n} c_k(t) + r(t). \quad (2)$$

Here, $c_1(t), c_2(t), \ldots, c_n(t)$ denote n IMFs, and $r(t)$ denotes the final residual that can be either the mean trend or a constant. An IMF satisfies the following two conditions:

1. Over the entire time series, the number of extremes and the number of zero-crossings must be equal or differ at most by one;
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

In essence, an IMF represents a simple oscillatory mode similar to a harmonic function in Fourier analysis, but possesses more general characteristics in frequencies and amplitudes of the modes. IMFs can be extracted from the raw data by a sifting process proposed in Huang's study (Huang et al., 1998), which is summarized in Appendix A. Previous studies (Huang et al., 1998; Zhang et al., 2003) have shown that IMFs may be correlated to meaningful, observable, and physical information inherent to the original data.

Since a recorded time series contains noise, some of the IMFs extracted from the raw data are likely to represent random noise. Such noisy IMFs need to be identified and separated. Thus, a time series is finally decomposed as

$$x(t) = \sum_{k=1}^{i} c_{T,k}(t) + \sum_{k=1}^{j} c_{N,k}(t) + r(t), \quad (\text{note } i + j = n), \tag{3}$$

where $c_{T,k}(t)$ and $c_{N,k}(t)$ denote a kth IMF of true data and noise, respectively.

Comparing Eq. (1) and Eq. (3), it can be concluded that by filtering out the trend, $r(t)$, and the noise term, $y(t) = \sum_{k=1}^{j} c_{N,k}(t)$, the true oscillatory signal with zero mean, $x_S(t) = \sum_{k=1}^{i} c_{T,k}(t)$, is identified.

From the details of the sifting process (presented in Appendix A), it is clearly seen that the IMFs are purely intrinsic, i.e., data based, functions without any prescribed form of basis functions or assumptions.

Traditionally, the noise is idealized as a Gaussian white noise process with zero mean and standard deviation, σ_g. Wu and Huang (2004) studied the statistical properties of IMFs extracted from a large number of simulated samples of Gaussian white noise and identified several interesting relations, which provided a basis for a statistical significance test that can identify noisy IMFs extracted from any given time series. The details of the statistical significance test are summarized in Appendix B.

This statistical significance test has been used in the EMD literature to remove noise and derive trends in nonlinear and non-stationary time series of global surface temperature (Wu et al., 2007) and the annual mean temperature (Capparelli et al., 2013).

2.3. Simulation-based illustrative example

To illustrate capability of the EMD method and the statistical significance test, this section analyzes a simulated data set with the purpose of removing the embedded noise, isolating the trend, and reconstructing the true signal of the time series. The simulated time series, as shown in Fig. 2, is generated using the following functions:

$$\begin{aligned} &x(t) = x_S(t) + r(t) + y(t),\\ &\text{where true signal, } x_S(t) = 0.5\sin\left(\frac{2\pi}{20}t\right),\\ &\text{and trend, } r(t) = 253.5 + \frac{0.3}{365}t. \end{aligned} \tag{4}$$

Note that $y(t)$ is the white noise or independent and identically distributed Gaussian random variable with zero mean and a standard deviation of 0.3.

After applying the EMD method to the simulated time series, eight IMFs extracted by the sifting process are shown in Fig. 3, and the extracted trend is shown Fig. 4, in comparison with the original trend function given by Eq. (4). As shown in Fig. 4, the EMD method succeeded in extracting the trend function embedded in the simulated time series.

Application of the statistical significance test is illustrated in Fig. 5, which shows the upper and lower bounds that constitute the acceptance region of the Null Hypothesis. Results of the statistical significance test indicate that, except for the third and sixth IMFs, $c_3(t)$ and $c_6(t)$, all other IMFs can be characterized as noisy components of the simulated time series. Therefore, the true signal without the trend is, $x_S(t) = c_3(t) + c_6(t)$, and the noise is the sum of the remaining six IMFs.

A comparison of the noise separated signal and the true signal is shown in Fig. 6. A comparison between mean and standard deviation of

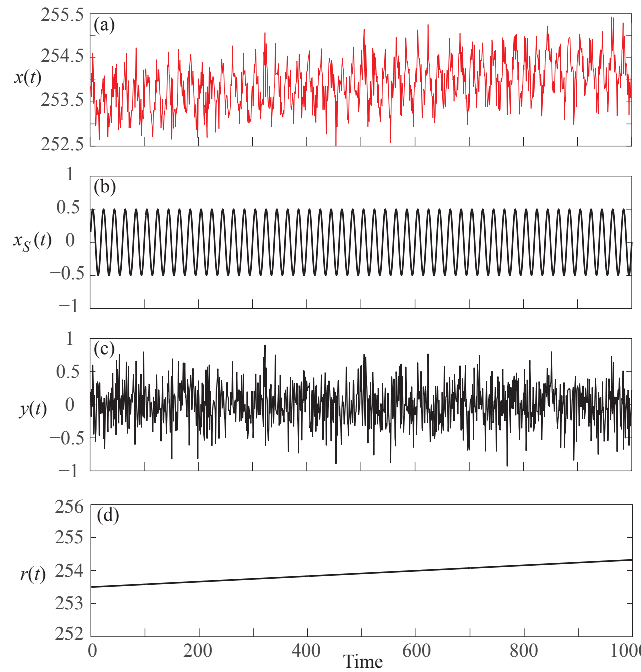

Fig. 2. (a) Time series simulated by Eq. (4), and its components (b) true signal, (c) Gaussian noise, and (d) trend function.

the noise separated signal and the true signal is shown in Table 1. It is seen that, the noise separated signal and the true signal have almost the same mean and standard deviation.

Euclidean metric is used to evaluate the similarity of the noise separated signal and the true signal. If S_n^i and S_t^i respectively represents values of the noise separated signal and the true signal at the ith point, the normalized Euclidean metric is defined as

$$d = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(\frac{S_n^i - \mu_n}{\sigma_n} - \frac{S_t^i - \mu_t}{\sigma_t}\right)^2} \tag{5}$$

where μ_n and μ_t represent mean of noise separated signal and true signal, respectively, σ_n and σ_t represent standard deviation of noise separated signal and true signal, respectively, and N represents total number of discrete data points (noise separated signal and true signal have the same discrete points).

For the noise separated signal and true signal in this example, the value of d is equal to 0.96%, which means that the difference between the noise separated signal and true signal is equal to 0.96%, a very small value. Comparison of the noise separated signal and the true signal again confirms the ability of the EMD method to filter out the noise from the data in a highly accurate manner.

In summary, this example confirms that the EMD method is able to fulfill their intended functions of removing the noise, isolating the trend, and reconstructing the true signal of the simulated time series.

3. Application: Time series of total acid number data

3.1. Time series analysis using the EMD method

Fig. 7 shows the condition monitoring data collected from the electro–hydraulic control (EHC) system of a turbine in a nuclear plant in Canada. This data set consists of the total acid number measured by chemical analysis of control fluid samples collected on a weekly basis from an EHC system.

Using the EMD method, a time series of the total acid number was decomposed into 7 IMFs and a single trend function, as shown in Fig. 8. The energy density and mean period of these IMFs, as defined in

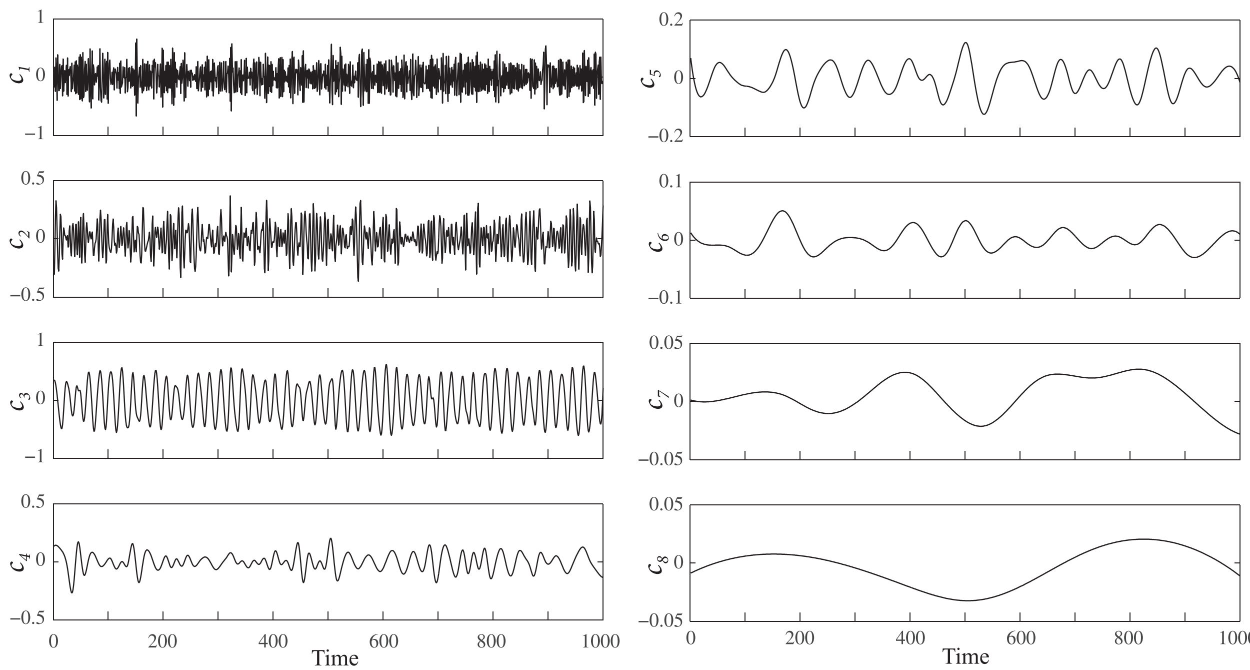

Fig. 3. Eight IMFs of the simulated time series.

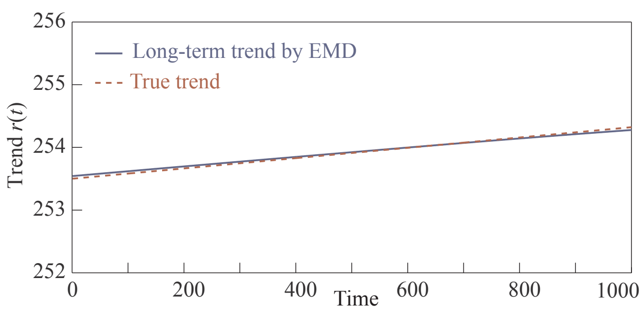

Fig. 4. Comparison of true trend and the trend extracted by EMD method: Simulation example.

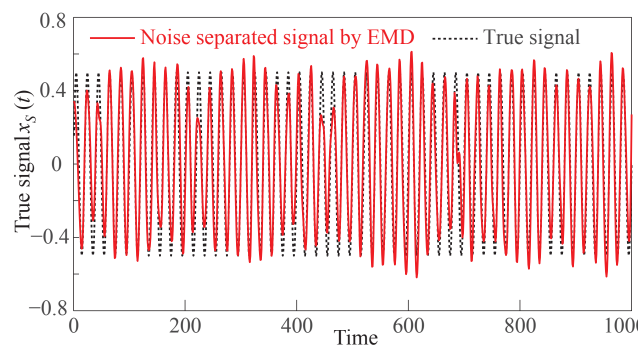

Fig. 6. Comparison of true signal with that estimated by EMD: Simulation example.

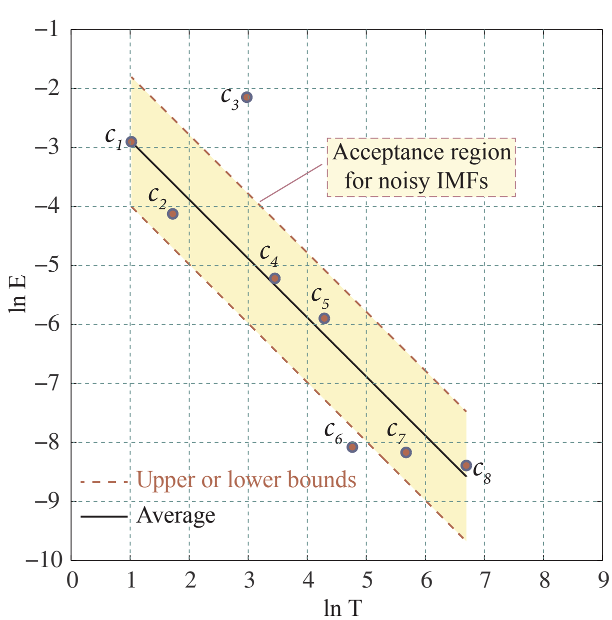

Fig. 5. Statistical significance test applied to IMFs of the simulated time series.

Table 1
Comparison between mean and standard deviation.

Signal	Mean	Standard deviation
Noise separated signal	0.0061	0.3418
True signal	0	0.3537

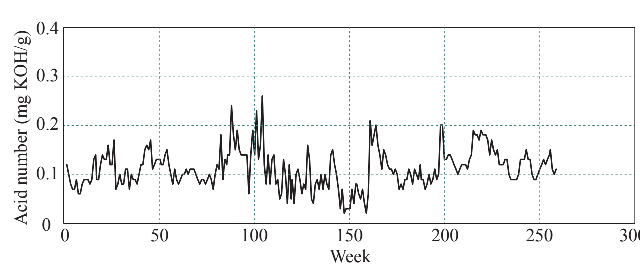

Fig. 7. The acid number data collected from a turbine EHC system.

Appendix B, are given in Table 2. Based on energy density and mean period of these IMFs, the statistical significance test was applied, which identified the first two IMFs, c_1 and c_2, as noisy IMFs (see Fig. 9). Therefore, the remaining 5 IMFs, c_3–c_7, were identified as those representing information of the actual data. As shown by Fig. 8, IMFs c_3 to

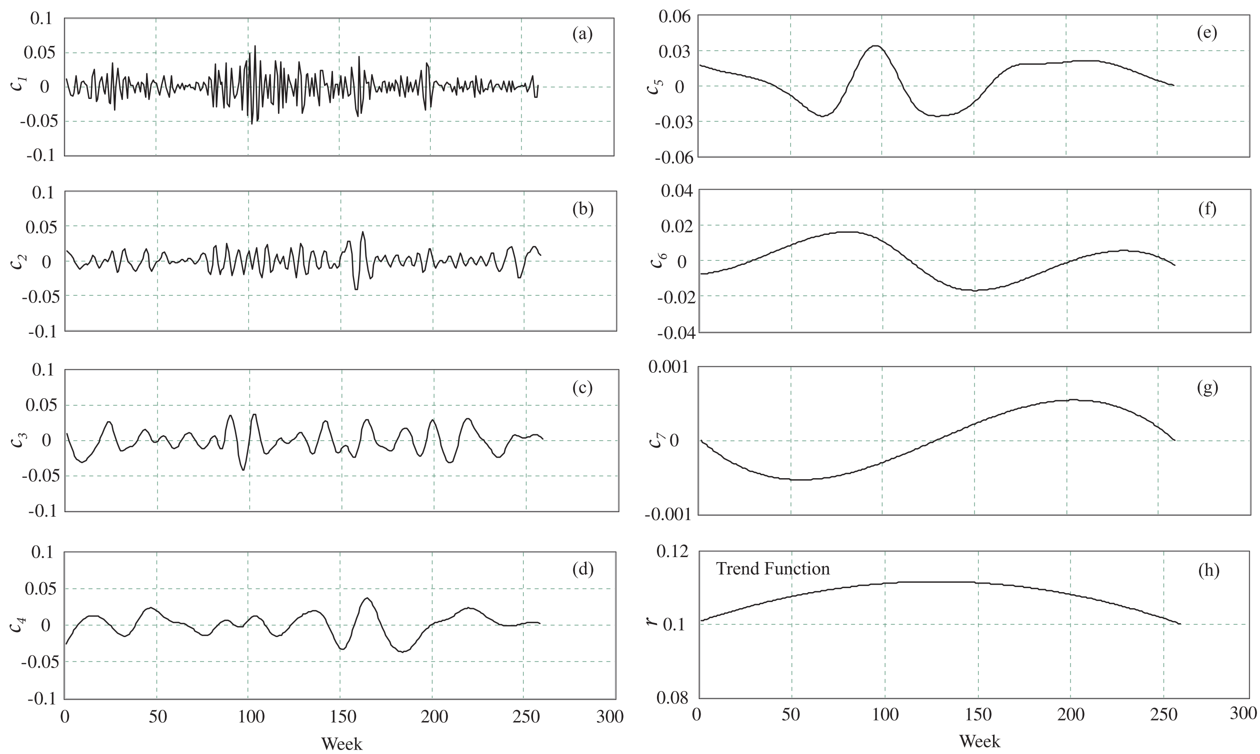

Fig. 8. IMFs and the trend function of the total acid number time series.

Table 2
Energy density (or variance) and mean period of IMFs for the acid number data.

kth IMF	$E_k(\times 10^{-4})$[(mg KOH/g)^2]	T_k(week)
1	3.235	2.943
2	1.434	6.317
3	2.559	14.389
4	2.317	34.037
5	2.982	105.516
6	0.950	144.962
7	0.001	257.395

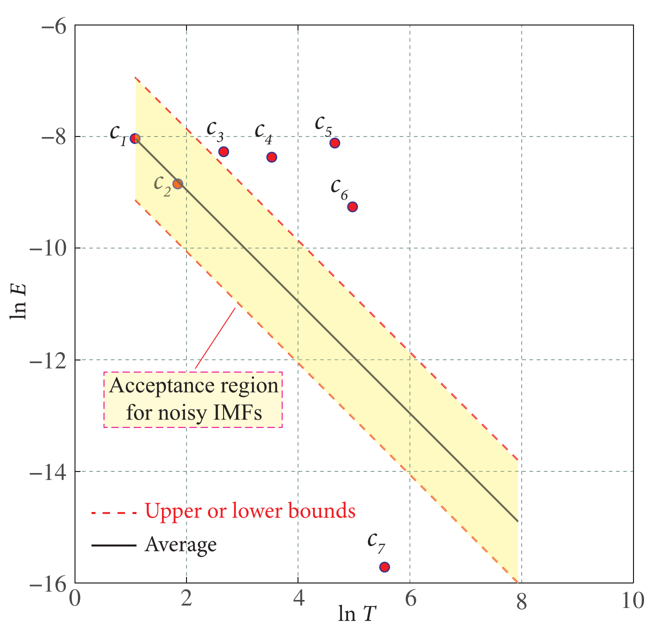

Fig. 9. Statistical significance test applied to the acid number data.

c_5 are highly non-stationary, whereas IMFs c_6 and c_7 appear to be more periodic in nature. The trend function, $r(t)$, shows that the total acid number increases initially in the first 100 weeks and then follows a declining trend.

The total noise in the data estimated by the EMD method is shown in Fig. 10. The standard deviation of the noise is estimated as 0.023 (mg KOH/g). It is also seen that, amplitudes of the noise almost uniformly fluctuate between −0.05 (mg KOH/g) and 0.05 (mg KOH/g), which are relatively large in comparison with amplitudes of the raw acid number data; the noise thus masks the overall trend and limits the ability to extract true information from the observed acid number data.

After removing the noisy IMFs, the remaining IMFs and the trend are combined to estimate the noise separated acid number, as shown in Fig. 11. Using the clean noise separated acid number as shown in Fig. 11, the operator could judge whether or not the acid number exceeded a safe threshold value. It is quite clear that the total acid number did not exceed the threshold value of 0.2 (mg KOH/g) in the sampling duration, whereas raw recorded acid number shown in Fig. 11, suggests that the total acid number did exceed the threshold value. Thus, the embedded noise indeed masks the true characteristics of the data, and the proposed EMD method provides the operator information about the real underlying condition.

In addition, Fig. 11 demonstrates that, the noise separated acid

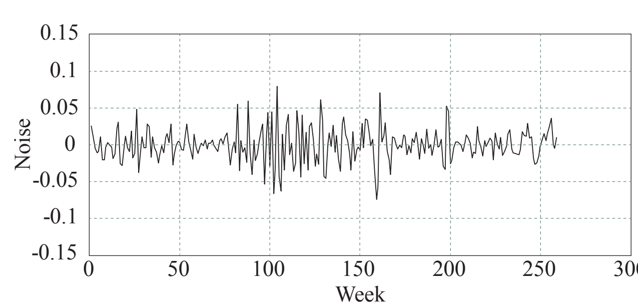

Fig. 10. Noise separated from the acid number data.

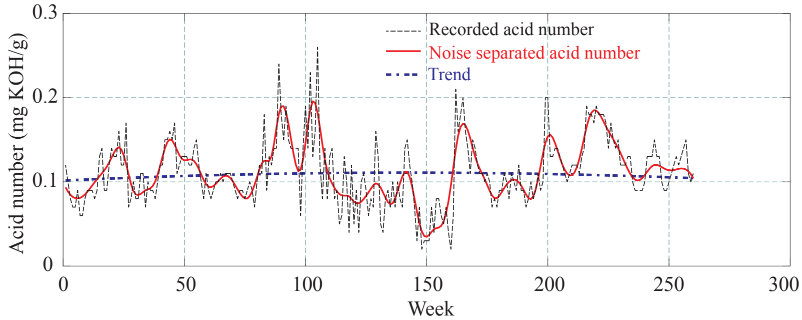

Fig. 11. Acid number data after the separation of noise.

number also presents relevant aging information, but not as clear as the trend function. Thus, it is necessary to combine the trend function and noise separated acid number together to assess performance of EHC system reflected by the acid number data: the trend function presents aging information of EHC system, and the noise separated acid number presents true oscillation amplitudes of the time series, which reflect the actual acid number (i.e., almost noise free) varying with time.

3.2. Validate the EMD method

To validate that the EMD method is also able to separate noise of real recorded data, another example is presented in this section. The acid number data set collected from another EHC system of the same nuclear plant, shown in Fig. 1, is analyzed by the EMD method; the noise separated acid number and embedded noise are obtained, as shown in Fig. 12.

The noise in Fig. 12(b) is added to the noise separated acid number in Fig. 11 to construct a new acid number time series, as shown in Fig. 13(a). The new acid number time series will be used to validate the ability of EMD method to separate noise of real recorded data. In the validation, the noise separated acid number in Fig. 11 is used as the benchmark acid number, and the noise in Fig. 12(b) is used as the benchmark noise. This example is different from the simulated-based illustrative example in Section 2.3 because components of the new acid number time series come from real recorded data.

Applying the EMD method to the new acid number time series, the noise separated acid number and noise are obtained, as shown in Fig. 13(b) and (c), respectively. Fig. 13(b) shows that the noise separated acid number by the EMD method well matches the benchmark acid number, although some minor difference exists. Fig. 13(c) shows that the noise separated by the EMD method also well matches the benchmark noise.

The example in this section demonstrates that the EMD method is also able to separate embedded noise from real recorded data, and obtain noise separated signal. Thus, the noise and noise separated acid number obtained in Section 3.1 are trustable.

4. Conclusions

This paper presented an innovative approach to statistical analysis of condition monitoring data that can be collected from any engineering systems operating in a nuclear power plant. The condition monitoring data are typically in the form of non-stationary and correlated time series with complex nonlinear trends, which cannot be analyzed by standard methods, such as statistical regression analysis or stationary time series analysis. The paper presented the EMD method to analyze such complex data sets.

The EMD method decomposes a time series into its underlying IMFs and a long-term trend function. The IMFs can be further separated into noisy modes and actual modes of the data. The trend function facilitates the detection of the onset of aging.

Ability of the EMD method in separating the noise and the trend from a time series was confirmed using a simulation-based example. The paper also presented a practical case study in which the time series of total acid number, a condition indicator of the control fluid used in a turbine electro–hydraulic system, was analyzed using the proposed

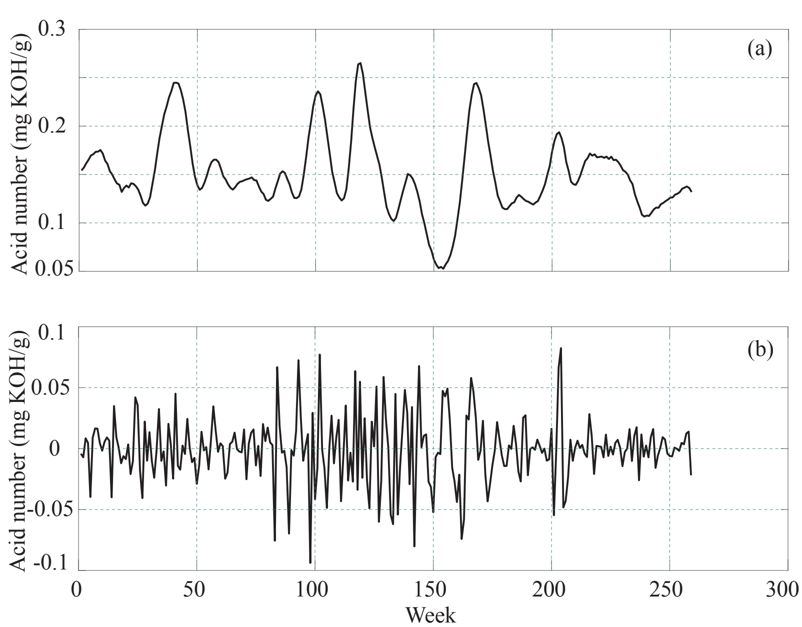

Fig. 12. Acid number data from another EHC system: (a) noise separated acid number; (b) noise.

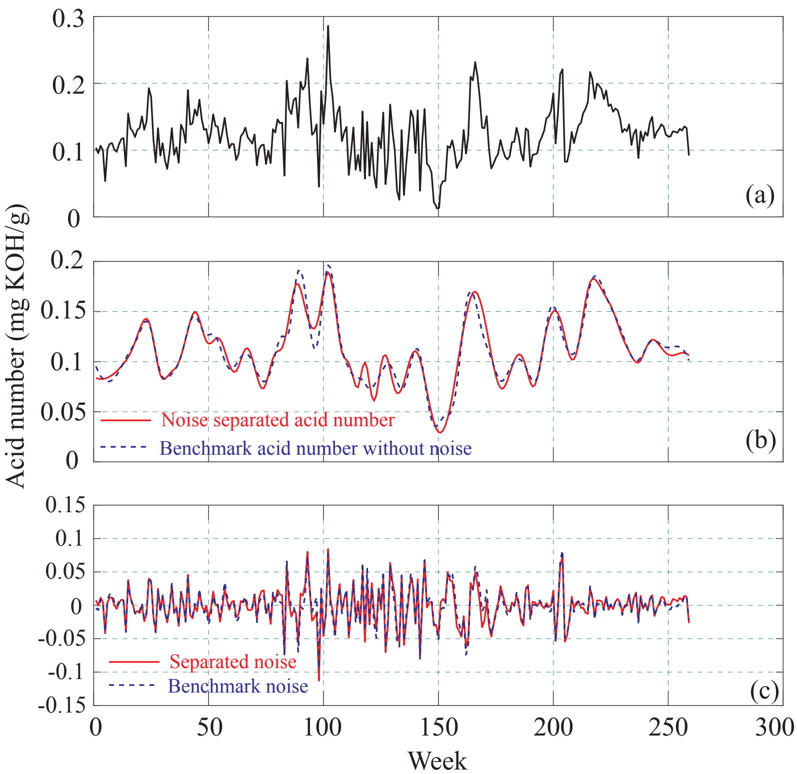

Fig. 13. Validated data: (a) new acid number; (b) noise separated acid number; (c) noise.

method. The analysis results were also validated by a subsequent example.

The utility of the proposed method is in assessing the effectiveness of an ongoing maintenance program by analyzing the past performance data. The EMD method allows to confirm whether or not degradation is taking place in the system, and thus provides a basis to modify the maintenance program or replacement interval.

Results of this study demonstrate that the EMD method is a highly versatile and efficient method to analyze the condition monitoring data collected from operating nuclear plants and it is also applicable to other engineering systems.

Acknowledgement

The authors gratefully acknowledge the financial support of this study provided by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University Network of Excellence in Nuclear Engineering (UNENE), which is a consortium of the Canadian nuclear utilities, Universities and the Regulator that is created to support innovative R & D in nuclear energy systems.

Appendix A. Sifting process to separate intrinsic mode functions

This section describes the sifting process proposed by Huang et al. (1998) to decompose a time series into a series of intrinsic mode functions (IMFs) and the residual known as the long-term trend function. Steps of this process are discussed as follows. Flow chart for the sifting process, i.e., EMD, is presented in Fig. 14.

First, identify all the local extrema of the time series. Then, connect all the local maxima by a cubic spline to produce the upper envelope of the data. Next, repeat the procedure for the local minima to produce the lower envelope of the data. The upper and lower envelopes should encompass all the data between them. After that, calculate the mean of these two envelopes that is designated as $m_1(t)$. Finally, calculate the difference between the time series $X(t)$ and $m_1(t)$, which forms the first component $h_1(t)$, i.e.,

$$h_1(t) = X(t) - m_1(t). \quad \text{(A.1)}$$

In the subsequent process, $h_1(t)$ is treated as the new time series, and the aforementioned process is repeated to calculate the mean of the upper and lower envelopes of $h_1(t)$, which is designated as $m_{11}(t)$. The difference between the new time series $h_1(t)$ and the mean of two envelopes $m_{11}(t)$ is:

$$h_{11}(t) = h_1(t) - m_{11}(t). \quad \text{(A.2)}$$

This process is repeated until all the conditions in the definition of an IMF are achieved. After repeated sifting, $h_{1i}(t)$ is given by

$$h_{1i}(t) = h_{1(i-1)}(t) - m_{1i}(t), \quad \text{(A.3)}$$

where $m_{1i}(t)$ is the mean of the upper and lower envelopes of $h_{1(i-1)}(t)$. $h_{1i}(t)$ is designated as the first IMF $c_1(t)$ from the time series $X(t)$, i.e.,

$$c_1(t) = h_{1i}(t). \quad \text{(A.4)}$$

The standard deviation SD, which is calculated from two consecutive sifting results, is used as the criterion to terminate the sifting process for each IMF. The SD is defined as

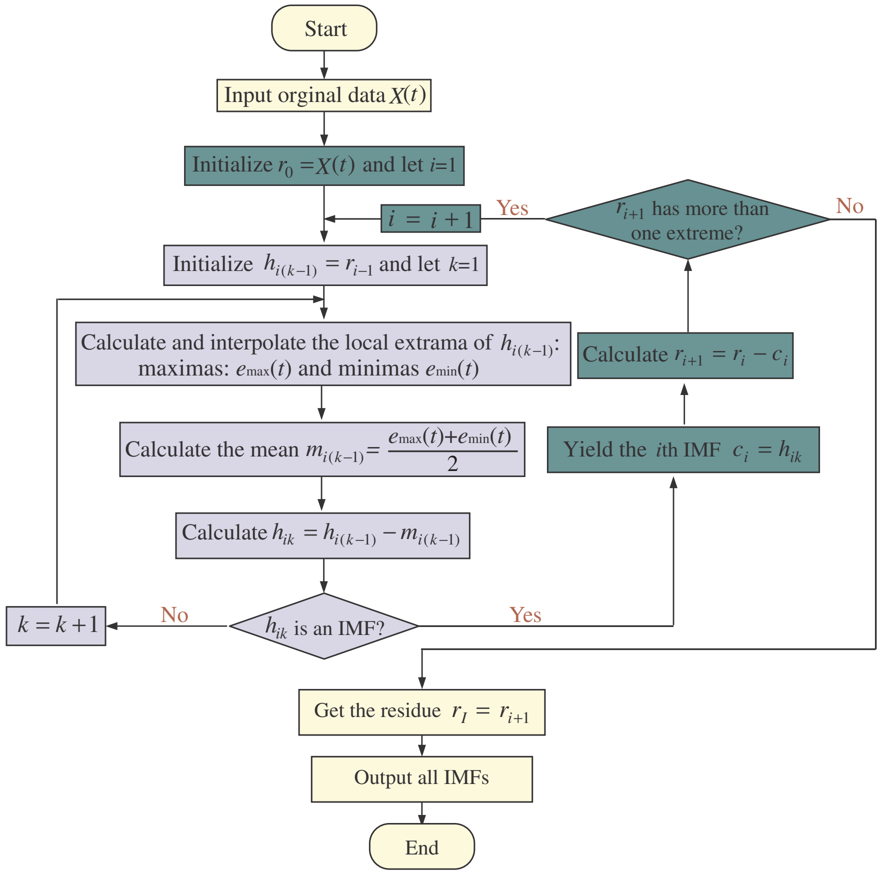

Fig. 14. Flow chart for the sifting process.

$$\mathrm{SD} = \sum_{t=0}^{T} \frac{[h_{1(i-1)}(t) - h_{1i}(t)]^2}{h_{1(i-1)}^2(t)}, \tag{A.5}$$

where T is the total time length of the discrete time series $X(t)$. When SD $\leqslant 0.2$, the sifting process for each IMF is terminated (Huang et al., 1998).

Typically, $c_1(t)$ contains the shortest-period component of the original time series $X(t)$. The residue is obtained by removing $c_1(t)$ from $X(t)$:

$$r_1(t) = X(t) - c_1(t). \tag{A.6}$$

The residue $r_1(t)$, which contains longer-period components, is treated as a new time series and subjected to the same sifting process aforementioned. This procedure is repeated to obtain all the subsequent r_k functions as follows:

$$r_k(t) = r_{k-1}(t) - c_k(t), \quad k = 2,3,...,n. \tag{A.7}$$

The sifting process is terminated by either of the following predetermined criteria:

- Either the component $c_n(t)$ or the residue $r_n(t)$ becomes so small that it is less than a predetermined value;
- The residue $r_n(t)$ becomes a monotonic function.

Appendix B. Statistical significance test to identify noisy IMFs

This section describes basic concepts underlying a statistical significance test to identify noisy IMFs, as proposed by Wu and Huang (2004).

Traditionally, the noise is idealized as a Gaussian white noise process with zero mean and some standard deviation, σ_g. Wu and Huang (2004) studied the statistical properties of IMFs extracted from a large number of simulated samples of Gaussian white noise, and discovered several interesting properties.

Consider that a white noise time series consisting of m data points is decomposed into n IMFs, $c_1(j),...,c_n(j), j = 1,...,m$. Define the mean period, T_k, and the energy density, E_k of the kth IMF as

$$T_k = \frac{m}{\text{Number of peaks in } c_k}, \tag{B.1}$$

and

$$E_k = \frac{1}{m} \sum_{j=1}^{m} |c_k(j)|^2. \tag{B.2}$$

The energy density defined by Eq. (B.2) is equivalent to the variance of the kth IMF, denoted as σ_k. For n IMFs of a sample of white noise, the following salient observations were made (Wu and Huang, 2004):

1. A set of m data points contained in any kth IMF, $c_k(j), j = 1,...,m$, follow the Gaussian distribution.
2. The Fourier amplitude spectra of all the IMFs of white noise are almost identical; they cover the same constant area on a semi-logarithmic scale of the period.
3. The product of the energy density (or the variance) and the mean period of each IMF is a constant, i.e.,

$$E_k\, T_k = \alpha \quad (\text{const.}) \Leftrightarrow \ln E_k + \ln T_k = \ln\alpha \quad (\text{const.}), \quad (k = 1,n). \tag{B.3}$$

4. The first IMF has the smallest mean period or the highest order of fluctuations. It is almost always a noisy IMF. Also, the energy density (E_1) and mean period (T_1) of the first IMF are not much affected by the sampling uncertainty, so that they can be considered as true estimates of the noise process.

Based on the last property, a hypothesis test for any kth IMF, c_k, was proposed by Wu et al. (2007) as follows. The Null Hypothesis is that an IMF, $c_k, k = 2,...,n$, is a noisy IMF. The test statistic is ($\ln E_k + \ln T_k$). The rejection region of this hypothesis was empirically defined as

$$\ln E_k + \ln T_k < \ln\left(\frac{1}{3}E_1\right) + \ln T_1, \text{ or}$$
$$\ln E_k + \ln T_k > \ln(3E_1) + \ln T_1. \tag{B.4}$$

In summary, those IMFs for which the Null Hypothesis is rejected are treated as true signals, whereas those IMFs for which the Null Hypothesis is accepted are treated as noise.

This statistical significance test has been used in EMD literature to remove noise and derive multidecadal trends of nonlinear and non-stationary time series of the global surface temperature (Wu et al., 2007) and the annual mean temperature (Capparelli et al., 2013).

References

Boudraa, A.O., Cexus, J.C., Saidi, Z., 2004. EMD-based signal noise reduction. Int. J. Signal Process. 1 (1), 33–37.

Capparelli, V., Franzke, C., Vecchio, A., et al., 2013. A spatiotemporal analysis of US station temperature trends over the last century. J. Geophys. Res.: Atmos. 118 (14), 7427–7434.

Huang, N.E., Shen, Z., Long, S.R., et al., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. London A: Math. Phys. Eng. Sci. 454 (1971), 903–995.

Liang, H.-L., Bressler, S.L., Desimone, R., et al., 2005. Empirical mode decomposition: a method for analyzing neural data. Neurocomputing 65, 801–807.

Li, B., Xie, W.C., Pandey, M.D., 2016. Generate tri-directional spectra-compatible time histories using HHT method. Nucl. Eng. Des. 308, 73–85.

Loh, C.H., Wu, T.C., Huang, N.E., 2001. Application of the empirical mode decomposition-Hilbert spectrum method to identify near-fault ground-motion characteristics and structural responses. Bull. Seismol. Soc. Am. 91 (5), 1339–1357.

Oakland, J.S., 2007. Statistical Process Control. Routledge, New York, NY.

Qian, C., Fu, C.-B., Wu, Z.-H., 2011. Changes in the amplitude of the temperature annual cycle in China and their implication for climate change research. J. Clim. 24 (20), 5292–5302.

Stoumbos, Z.G., Reynolds, M., Ryan, T.P., et al., 2000. The state of statistical process control as we proceed into the 21st century. J. Am. Stat. Assoc. 95 (451), 992–998.

Wu, Z.-H., Huang, N.E., 2004. A study of the characteristics of white noise using the empirical mode decomposition method. Proc. R. Soc. London A: Math. Phys. Eng. Sci. 460 (2046), 1597–1611.

Wu, Z.-H., Huang, N.E., Long, S.R., et al., 2007. On the trend, detrending, and variability of nonlinear and nonstationary time series. Proc. Nat. Acad. Sci. 104 (38), 14889–14894.

Zhang, R.R., Ma, S., Safak, E., et al., 2003. Hilbert-Huang Transform analysis of dynamic and earthquake motion recordings. J. Eng. Mech. 129 (8), 861–875.