

RESEARCH

Open Access

Extending the scope of empirical mode decomposition by smoothing

Donghoh Kim¹, Kyungmee O Kim² and Hee-Seok Oh^{3*}

Abstract

This article considers extending the scope of the empirical mode decomposition (EMD) method. The extension is aimed at noisy data and irregularly spaced data, which is necessary for widespread applicability of EMD. The proposed algorithm, called statistical EMD (SEMD), uses a smoothing technique instead of an interpolation when constructing upper and lower envelopes. Using SEMD, we discuss how to identify non-informative fluctuations such as noise, outliers, and ultra-high frequency components from the signal, and to decompose irregularly spaced data into several components without distortions.

Keywords: Empirical mode decomposition, Intrinsic mode function, Irregularly spaced data, Noisy signals, Sifting, Smoothing

Introduction

When analyzing a complex signal, we frequently decompose it into several components having simple forms and then analyze the information contained in each component to reduce the complexity and to enhance interpretability. Conventionally, decomposition is processed using a basis system. The benefits of decomposition are as follows: (1) a signal is well approximated by a finite number of basis functions, (2) information in the time (physical) domain is transformed into information in the frequency domain without losing any information, and (3) the interpretability of the signal can be enhanced by analyzing each component separately and comparing it with the other components.

Spectral analysis [1] and wavelet analysis [2-4] are popular methods for signal decomposition. However, when a signal has inherent nonstationary and nonlinear features according to the scale and time location, these methods might not be suitable. Empirical mode decomposition (EMD), developed by Huang et al. [5], provides a data-driven approach to decompose a signal into so-called intrinsic mode functions (IMFs) according to the local oscillation magnitude in the physical domain. IMFs can be considered as data-driven *empirical basis functions*. EMD

has been popularly used for analyzing nonstationary signals or nonlinear signals in many disciplines of science and engineering [6].

However, due to interpolation process in the construction of envelopes, IMFs obtained by the conventional EMD algorithm are sensitive to non-informative fluctuations such as noise, outliers, and ultra-high frequency components, and hence, the non-informative fluctuation effect distorts the subsequent decomposition results. In addition, this method focuses on a narrow scope that does not cover irregularly sampled data. These constraints of its scope strongly diminish the applicability of EMD to various signals. To extend the scope of the conventional EMD to noisy signals and irregularly spaced data, we propose a statistical EMD algorithm called SEMD that is based on a smoothing technique. This method is a fully data-adaptive algorithm as in the case of the conventional EMD. The proposed SEMD has several advantages over the conventional EMD: (1) It is robust to noise or non-informative random fluctuations such as outliers and ultra-high frequency components, and hence, SEMD can decompose such signals into appropriate IMFs without distortion caused by the above-mentioned factors. (2) It provides a reasonable boundary condition of an IMF without any boundary treatment, and therefore, SEMD can provide stable decomposition results on the entire domain including boundary regions. Furthermore,

*Correspondence: heeseok@stats.snu.ac.kr

³Department of Statistics, Seoul National University, Seoul 151-747, Korea
Full list of author information is available at the end of the article

we extend EMD to analyze irregularly spaced signals by combining SEMD with a simulation technique.

The remainder of this article is organized as follows. Section Review: empirical mode decomposition presents an overview of the conventional EMD. Section Statistical EMD describes the proposed SEMD method, and several case studies are presented to show its broad applicability. In addition, we investigate the variation diminishing property of SEMD. An extension to an irregularly spaced signal is presented in Section Extension of EMD to irregularly spaced signals. Finally, concluding remarks are presented in Section Conclusion.

Before closing this section, we note that, in the literature, there have been several attempts to enhance the performance of the conventional EMD and to extend its scope. For example, to deal with noise, Boudraa and Cexus [7] removed the high-frequency components using a filtering method, and Wu and Huang [8] used the ensemble mean approach of the simulated signal. Both methods are based on conventional sifting followed by a posterior adjustment. For applying the conventional EMD to signals with lower sampling rate, Xu et al. [9] proposed a hybrid extrema estimation algorithm based on Fourier interpolation. More recently, Diop et al. [10] suggested a PDE-based approach to compute envelopes, which is another way to use non-interpolation in construction of envelopes.

Review: empirical mode decomposition

Fourier analysis decomposes a signal into a sum of sinusoids having different frequencies. However, it is well known that for nonstationary signals, Fourier analysis does not effectively provide frequency information of the signals. Although wavelet analysis is a popular method for analyzing nonstationary signals, it suffers from a non-adaptive nature in that it applies the same type of basis functions to the entire range of data. Wavelet analysis also represents a signal by a linear combination of wavelet basis functions. Therefore, its formulation for the energy-frequency representation of nonlinear data can be misleading [5]. Thus, we require a set of flexible basis functions that reflects time-varying properties of a signal.

Huang et al. [5] proposed a data-driven algorithm for extracting an oscillatory wave from a given signal x as follows. First, we identify the local extrema and construct two functions called the upper envelope and lower envelope by interpolating the local maxima and local minima, respectively. Second, we take their average; this produces a signal with a frequency lower than that of the original signal because the main pattern of the signal is confined between the two envelopes. Third, by subtracting the envelope mean from x , the highly oscillatory wave h is separated.

Huang et al. [5] defined an oscillatory wave to be an IMF if it satisfies two conditions: (1) the number of extrema

and the number of zero-crossings should be equal or differ by one and (2) the local average should be zero, implying that the mean of the upper envelope and the lower envelope is zero. There might exist overshoots and undershoots in h after one iteration of the aforementioned procedure, in which case the two conditions are not satisfied. In such a case, until the conditions are satisfied, the procedure is repeated for h . This iterative process is called *sifting*. We may consider the IMF to be an empirical basis driven by the data-adapted process, sifting. This IMF is the mode function that has the finest resolution. By sifting, the original signal x is decomposed into the highest frequency imf_1 and a residual signal $r_1 = x - imf_1$ that is less oscillatory than the original signal x . If r_1 has signals having different frequencies, then the next IMF is obtained by considering r_1 as a new signal. The signal is sequentially decomposed into signals having different frequencies from the highest-frequency component imf_1 to the lowest-frequency component imf_n for some finite n and a residual signal r . Finally, we have n IMFs and a residual signal

$$x(t) = \sum_{i=1}^n imf_i(t) + r(t).$$

Here, index i denotes the resolution level and imf_1 is IMF at the finest level. We finally remark that Fourier analysis assumes that a signal is stationary and consists of components of a pure tone. In practice, the frequency information can evolve over time and several such frequencies can be compounded. The above EMD procedure is useful for identifying the amount of variation due to oscillation at different scale and time location and extracting an oscillatory wave from a nonstationary signal.

Statistical EMD

One of the main purposes of EMD is to decompose a signal into several components and to identify its significant frequency components. It is not uncommon for a signal to be corrupted by non-informative random fluctuations such as noise, which might consist of high frequencies and contains no interpretable information.

However, the conventional EMD algorithm cannot effectively separate noise from the signal, and hence, this algorithm does not produce stable decomposition results from noisy signals. To overcome this problem, we propose a modified sifting process based on a smoothing technique. The proposed algorithm can be easily implemented by simply replacing the interpolation with smoothing. That is, the upper and lower envelopes can be constructed by a smoothing technique. The proposed algorithm is designed for considering noisy signals that are used in the field of statistics. Thus, we call the proposed algorithm

statistical EMD (SEMD). Formally, the SEMD algorithm can be stated as follows:

- A. (Modified sifting) Take a signal x to be decomposed, and extract the first mode $h_{1,\lambda}$ by using a smoothing technique.
- (A-1) Identify the local maxima (minima) z of the signal $h_{1,\lambda}^0$ where $h_{1,\lambda}^0$ is the original signal x .
 - (A-2) Construct an upper envelope \hat{u}_λ (lower envelope $\hat{\ell}_\lambda$) by applying a smoothing technique with a smoothing parameter λ to the maxima (minima) z .
 - (A-3) Compute the local mean $m_\lambda = \frac{1}{2}(\hat{u}_\lambda + \hat{\ell}_\lambda)$ by the average of both the envelopes, and then obtain a candidate intrinsic mode $h_{1,\lambda}^1 = h_{1,\lambda}^0 - m_\lambda$.
 - (A-4) Repeat steps (A-1)–(A-3) for the signal $h_{1,\lambda}^1$ until the signal $h_{1,\lambda}^j$ at the j th iteration satisfies the IMF conditions.
 - (A-5) Decompose the signal $x = h_{1,\lambda} + r_\lambda$, where $h_{1,\lambda}$ is defined as the limit of $h_{1,\lambda}^j$ and r_λ is the remaining signal.
- B. (Conventional sifting) If the remaining signal $r_\lambda = x - h_{1,\lambda}$ has an intrinsic oscillation mode, then r_λ can be further decomposed by conventional sifting.

The only difference between SEMD algorithm and the conventional EMD is step A, where the first mode is extracted by smoothing instead of interpolation. In particular, step (A-2) in construction of \hat{u}_λ and $\hat{\ell}_\lambda$ by smoothing plays most important roles in determining the quality of the decomposition when the signal is corrupted by non-informative random fluctuations.

A key issue that needs to be considered is how to determine the degree of smoothness (i.e., smoothing parameter, λ) in the smoothing process. We propose an *automatic* selection method of λ utilizing the conventional cross-validation. The cross-validation splits observations into K roughly equal-sized parts (for example, $K = 4$). For the k th part (say, test dataset), we fit the model to the other $K - 1$ parts (say, training dataset) of the observations, and calculate the prediction error of the k th part by the fitted model. We perform this procedure for $k = 1, \dots, K$ and combine all K estimates of prediction error.

However, by omitting the test dataset, the remaining training dataset for fitting the model becomes unequally spaced data. Since the model fit is based on the decomposition, it is difficult to obtain the stable fitting results with such unequally spaced data, and hence, the conventional cross-validation method may not be directly applicable to this case.

Here, we propose an imputation-based cross-validation method for selecting the smoothing parameter. For the k th test dataset, we impute it by an imputation method, apply the SEMD with a given smoothing parameter λ to a new composite data which consists of the imputed k th dataset and the remaining training dataset, and calculate the prediction error for the k th test dataset. More specifically, with defining an indexing function $\kappa : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, K\}$ that indicates the partition to which observation $x(t)$ is allocated, we obtain K partitioned datasets, $T_k = \{x(t) : k = \kappa(t), t = 1, 2, \dots, n\}$ ($k = 1, \dots, K$). We then define the prediction error as follows:

- (i) Split a signal x into K test datasets $T_1, \dots, T_k, \dots, T_K$.
- (ii) Impute the k th test dataset by local average of two neighboring points and obtain \tilde{T}_k .
- (iii) With a given smoothing parameter λ , apply the SEMD algorithm to decompose the composite signal $T_1, \dots, T_{k-1}, \tilde{T}_k, T_{k+1}, \dots, T_K$ into an $h_{1,\lambda}$ and the remaining signal r_λ .
- (iv) Obtain the predicted values of remaining signal evaluated at the k th part, say $r_\lambda^k(t)$.
- (v) Repeat steps (ii)–(iv) for $k = 1, \dots, K$, and define the prediction error as

$$PE(\lambda) = \frac{1}{n} \sum_{t=1}^n \{x(t) - r_\lambda^k(t)\}^2.$$

Finally, by using an optimization algorithm such as golden section search algorithm, we select the smoothing parameter λ value that minimizes the prediction error $PE(\lambda)$. By considering each test dataset as new observations, it can be shown that the expectation of $PE(\lambda)$ is close to true prediction error [11]. Thus, the above procedure is widely used for estimating true prediction error.

We have some remarks regarding the SEMD algorithm.

- **The first mode $h_{1,\lambda}$:** The $h_{1,\lambda}$ might not contain a meaningful mode when non-informative fluctuations such as noise are present, and hence, it may not be appropriate to define the extracted mode as IMF. However, the extracted mode can be considered as IMF in the case of noiseless signals.
- **Modified sifting at the first level:** The modified sifting can be employed to extract further IMFs beyond the firstly extracted mode $h_{1,\lambda}$. However, from our experience based on extensive simulation studies, SEMD effectively extracts noise from a noisy signal x at the first level. Furthermore, Rilling and Flandrin [12] and Park et al. [13] investigated that when there exists a big discrepancy between the frequencies of two components of a signal, ordinary sifting process cannot correctly estimate the relatively low frequency component, which results in misidentifying the

relatively high frequency component. Since noise acts as a high frequency component and the modified sifting utilizing smoothing effectively estimates the low frequency component, SEMD seems to effectively extract noise at the first level.

- **Smoothing technique:** Several smoothing techniques including kernel smoothing, smoothing splines, and local polynomial method have been well developed. In this study, we use kernel smoothing with Gaussian kernel. In practice, any smoothing method can be adopted for SEMD algorithm.
- **The role of the smoothing parameter λ :** The performance of the modified sifting depends on the choice of λ . We now consider two special cases: (1) $\lambda = 0$ —both envelopes \hat{u}_λ and $\hat{\ell}_\lambda$ are constructed by interpolation, and hence, the extracted results are identical to those by the conventional EMD, and (2) $\lambda = \infty$ —both envelopes \hat{u}_λ and $\hat{\ell}_\lambda$ are the weighted averages of local extrema, so that the extracted mode becomes a over-smoothed function which might not be suitable to represent any frequency patterns of the original signal. It implies that any meaningful modes can not be extracted further. Therefore, to overcome the above problem, we propose the data-driven cross-validation approach to select an optimal λ . Finally, we remark that since the $PE(\lambda)$ is a reasonable estimate of true prediction error, the resultant λ should be close to 0 when the signal is noise-free. Thus, the resulting fitting is almost identical to interpolation result in the case of noise-free signal. In summary, SEMD can be applicable to both noisy and noise-free signals.
- **The number of K :** Through this article, we use $K = 4$, so that the entire signal is divided into four parts. The K can be chosen to be any number less than n . The case $K = n$ is known as ‘leave-one-out’ cross-validation, where $\kappa(t) = t$, and the predicted value for the t th observation is evaluated using all the data except the t th observation. Thus, the leave-one-out cross-validation is computationally intensive.
- **Sensitivity of imputation method:** An imputation method is required for the derivation of $PE(\lambda)$. In this study, we use local average of two neighboring points for imputation which is simple and fast. It can be also adapted by an advanced imputation technique such as EM algorithm. However, for all cases in the article, we observe that the selection results for the smoothing parameter are almost identical.
- **Computation cost:** Compared to the conventional EMD, the SEMD algorithm requires a longer computational time due mainly to the smoothing parameter selection. However, once the smoothing parameter for the first mode is selected, the computation time of the proposed algorithm is even

faster than that of the conventional methods when a signal is contaminated by random fluctuations, because the remaining steps are almost identical to those of conventional EMD and in this case, the conventional EMD tends to produce extra artificial modes (this observation will be shown in subsequent sections). In addition, the computational burden of the above K -fold cross-validation procedure is not considerable at all.

c

Here, we discuss a theoretical property of SEMD, namely, the variation diminishing property of envelopes. It implies that, as the value of smoothing parameter increases, variation of envelopes decreases *monotonically*. In other words, structures of envelopes such as peaks and valleys disappear monotonically as the level of smoothing increases. Thus, lower and upper envelopes generated by a certain level of smoothing parameter should not contain some artifacts due to noise. This fact has been known as causality in the scale space literature (see, e.g., [14]).

Proposition 1. (variation diminishing property of envelopes) Let $\{z_i\}_{i=1}^m$ be a sequence of the centered local maxima (or minima) of the original signal $\{x(t)\}_{t=1}^n$ ($m < n$) and $v(\lambda)$ be the number of sign changes in the Gaussian (or one-sided exponential) kernel estimate \hat{u}_λ for $\{z_i\}_{i=1}^m$. It follows that

$$v(\lambda) \leq v(\lambda')$$

for any positive value $\lambda' \leq \lambda$.

Proof. : Suppose that we observe the data $(t_1, z_1), \dots, (t_m, z_m)$. Give a smoothing parameter λ , a kernel smoothing estimate is defined as

$$\hat{u}_\lambda(t) = \frac{1}{m\lambda} \sum_{i=1}^m z_i K\{(t - t_i)/\lambda\},$$

where the kernel function is $K(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$. Then, using the fact that $K(t/\lambda_1) * K(t/\lambda_2) = K(t/\sqrt{\lambda_1^2 + \lambda_2^2})$, we obtain

$$\hat{u}_{\lambda_1}(t) * K(t/\lambda_2) = \hat{u}_{\sqrt{\lambda_1^2 + \lambda_2^2}}(t)$$

for all $\lambda_1, \lambda_2 > 0$. By Theorem of [15], it follows that the number of sign changes in $\hat{u}_\lambda(t)$ is monotonically decreasing function of λ . \square

To conclude this section, we note that the benefits of adopting a smoothing process for sifting can be summarized as a few items and these will be investigated empirically through simulation studies in the subsequent sections.

- (a) Extrema from a signal contaminated by noise are sensitive to noise or outliers. Thus, it is necessary to filter out such insignificant terms when constructing the upper and lower envelopes, and hence, the sifting process using filtered envelopes can produce stable IMFs.
- (b) As to be discussed below, the conventional EMD cannot properly handle a signal containing an ultra-high frequency component because it is difficult to obtain the desired upper and lower envelopes by using interpolation. This case can be solved by employing a smoothing approach.
- (c) Inadequate information is available on the modulation of two boundaries before the first extremum and after the last extremum when constructing envelopes. Thus, using smoothing instead of interpolation for extrema can alleviate the boundary problem.

SEMD for noisy signals

To observe the effect of smoothing in the sifting process, we consider the following synthetic signal $x(t)$, $0 < t < 9$ with frequencies f_1, f_2 , and f_3 :

$$x(t) = 0.5t + \sin(f_1\pi t) + \sin(f_2\pi t) + \sin(f_3\pi t). \tag{1}$$

The panels in Figure 1 show signals obtained from model (1) with $f_1 = 6, f_2 = 2$, and $f_3 = 1$, and the decomposition results by the conventional EMD and SEMD, respectively. For the noise-free signal, the performance of SEMD for decomposition is comparable to that of the conventional EMD, which properly produces three IMFs and the residue. We also consider monthly average surface air temperatures of Northern Hemisphere at surface level 700 hPa for the period January 1961 – August 2003 in Figure 2. It is available from the website <http://www.cdc.noaa.gov/>. It seems that the temperature data is noise-free signal. The panels in Figure 2 show the first two IMFs and the remaining signal obtained by EMD and SEMD are almost identical.

On the other hand, EMD distorts the decomposition when noise is present. The left-hand side panels in Figure 3 show a Gaussian noisy signal with signal-to-noise ratio (SNR) 5, the first IMF, and the remaining signal using interpolation during the sifting process. We notice that through all the experiments and simulation study of this article, $n = 1024$ data points are regularly sampled in the time domain, and SNR is defined as $SNR = \|m\|/\sigma$; m is a true function and σ is the standard deviation of noise. It is apparent that the remaining signal exhibits a noisy pattern. In contrast, the application

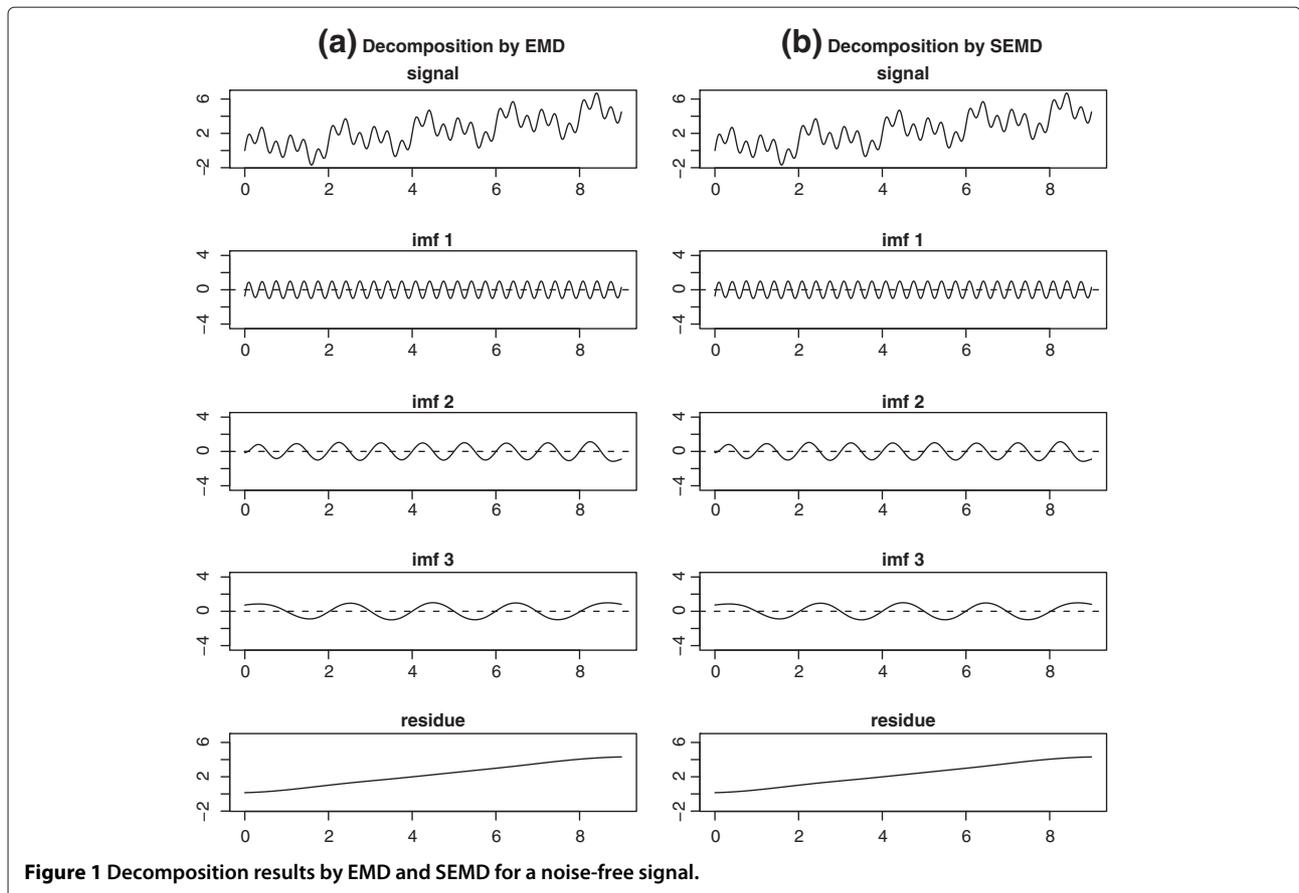


Figure 1 Decomposition results by EMD and SEMD for a noise-free signal.

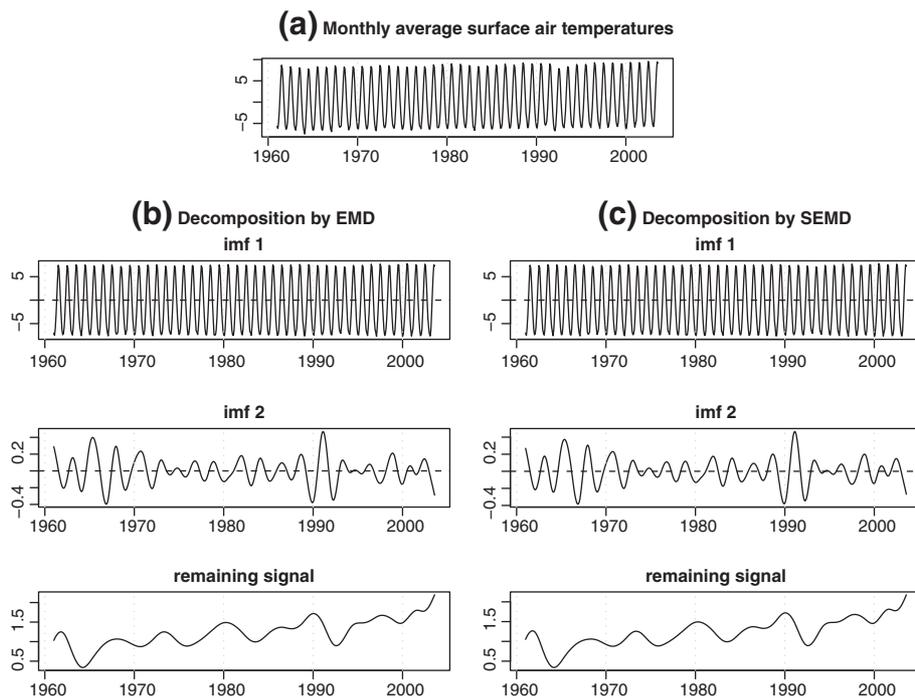


Figure 2 Monthly average surface air temperatures of Northern Hemisphere at surface level 700 hPa and its decomposition results by EMD and SEMD.

of a smoothing technique ensures that noise is effectively separated from the signal as shown in the right-hand side panels of Figure 3. This fact can be evaluated by examining the spectrum of periodogram. As shown in Figure 4, we observe that the periodogram of the first component by SEMD is almost flat, which is close to that of pure noise, while that of EMD shows a pattern which is not flat and therefore differs from that of pure noise. In

this case, the optimal smoothing parameter λ of SEMD is selected as $\lambda = 0.0523$ by minimizing the prediction error of cross-validation. Figure 5 shows the prediction errors according to the smoothing parameter, which illustrates the sensitivity to the choice of smoothing parameter. Each first component from SEMD and EMD affects the subsequent decomposition results significantly. Figure 6 shows the decomposition results for a noisy signal.

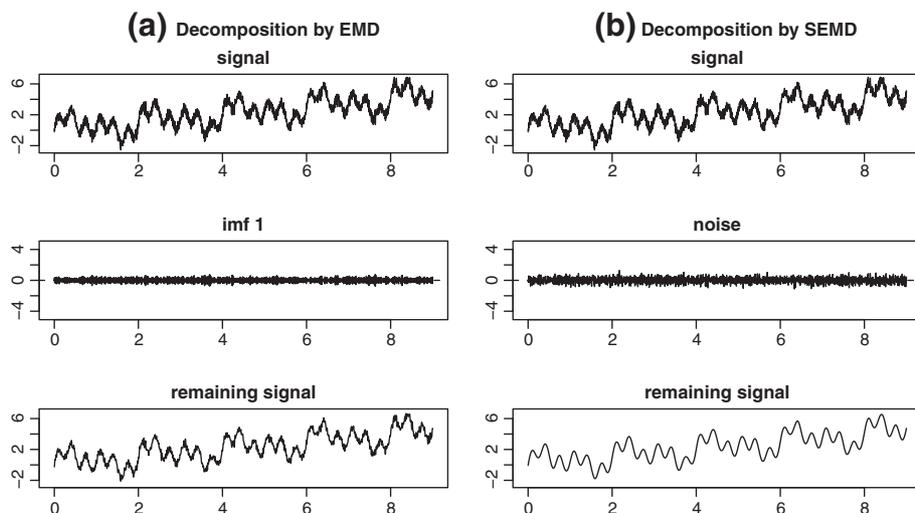


Figure 3 The first component and remaining signal by EMD and SEMD for a Gaussian noisy signal with SNR 5.

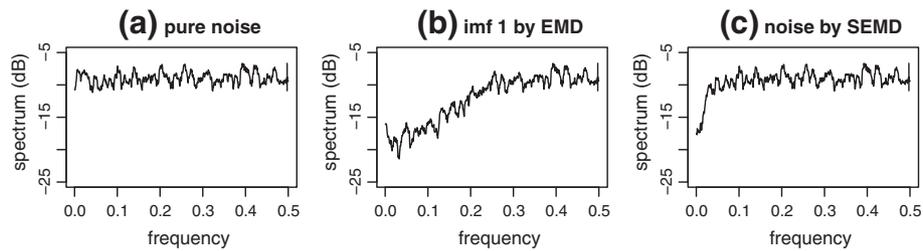


Figure 4 Periodogram of pure noise, and periodograms of the first components by EMD and SEMD for a Gaussian noisy signal with SNR 5 of Figure 3.

The conventional EMD does not produce proper IMFs, whereas SEMD extracts noise and decomposes three IMFs effectively.

SEMD for signals with outliers

Here, we consider a case in which the signal has some insignificant random fluctuations represented by outliers. Unfortunately, the conventional EMD may not be effective to decompose signals with outliers because the interpolation is sensitive to outliers. More specifically, interpolation in the conventional sifting is based on local extrema, and therefore, the upper or lower envelope moves toward the extreme values. Thus, the sequent IMFs and the residual signal are distorted to an extent that depends on the extreme values. The left-hand side panels in Figure 7 show the case with a single outlier. The EMD process cannot cope with even an outlier and produces incorrect waveforms, whereas the SEMD results shown in right-hand side panels are robust to the presence of an outlier. In this case, the optimal smoothing parameter λ of SEMD is 0.0663 which is larger than that of Gaussian noisy signal in Figure 3. In a similar manner, the first component of SEMD effectively treats a heavy-tailed noise that can produce outliers, such as the t -distribution with three degrees of freedom shown in Figure 8.

To evaluate the practical performance of SEMD and to see whether the first component effectively captures the noise or outliers from a signal, a simulation study was conducted. In the study, we compared SEMD with the conventional EMD and a wavelet shrinkage method.

1. `emd`: conventional EMD,
2. `semd`: proposed SEMD, and
3. `sure`: SURE wavelet shrinkage method of [16].

We consider the nine test functions displayed in Figure 9: Sines (`sines`) in model (1), a chirp signal (`chirp`) of the form $m(t) = \exp(-0.01t) \cos(\pi t/10)$ ($t \in [0, 500]$), Heavisine (`heav`) from [17], `fg1` (`fg1`) from [18], Wave (`wave`), Angles (`angle`), Parabolas (`para`), Time Shifted Sine (`tsine`) from [19], and Corner (`corn`) from [20].

The test functions are corrupted by Gaussian and t -distribution noise with three degrees of freedom, respectively. We used seven noise levels with SNRs of 1, 2, ..., 7. For each combination of the test functions, noise types, and SNR, 100 datasets were generated with a sample size of 1024. For each generated dataset, the three methods mentioned above are applied to obtain the estimate \hat{m} of the test function m . In the EMD methods, including SEMD, the estimate of m is obtained by removing the first

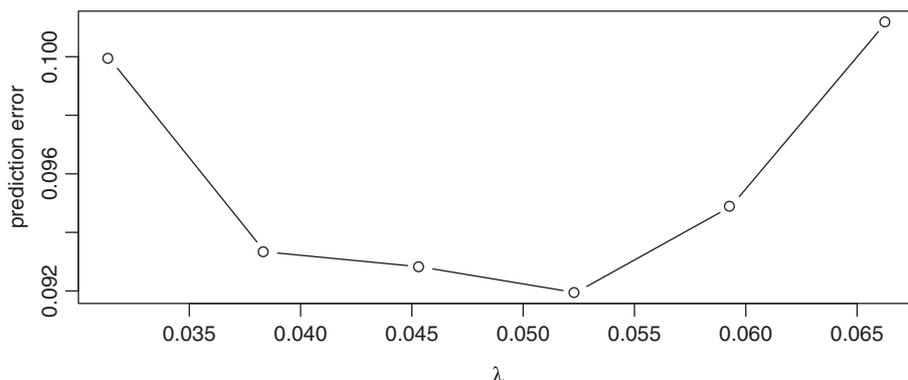


Figure 5 The prediction errors of SEMD according to smoothing parameter λ for a Gaussian noisy signal with SNR 5 of Figure 3.

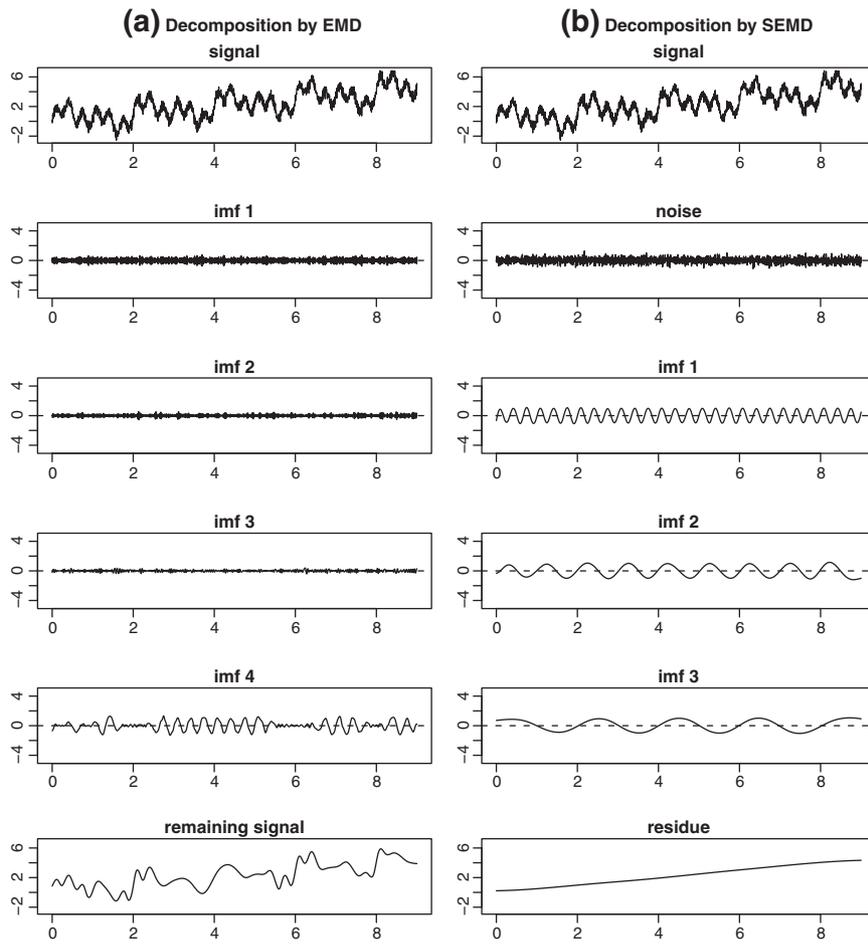


Figure 6 Decomposition results by EMD and SEMD for a Gaussian noisy signal with SNR 5.

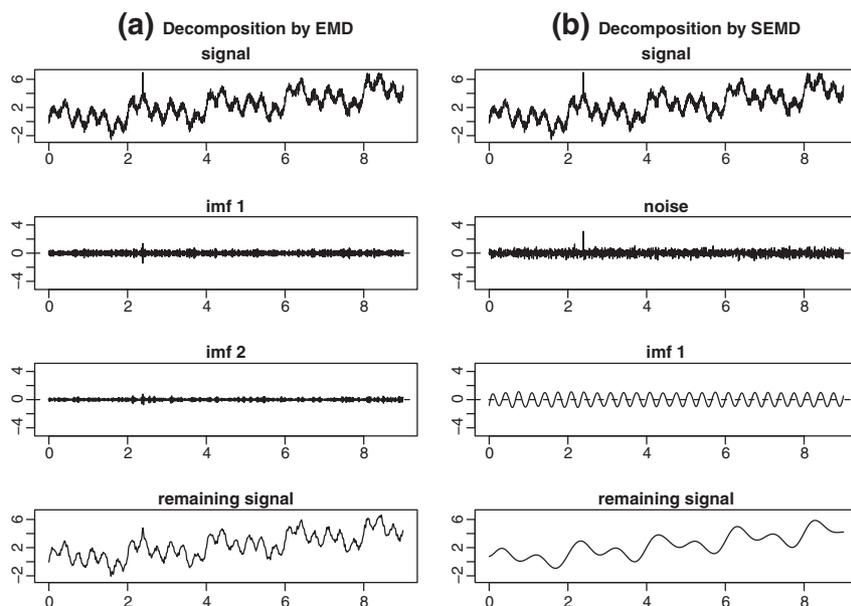


Figure 7 Decomposition results by EMD and SEMD when one extreme value is present.

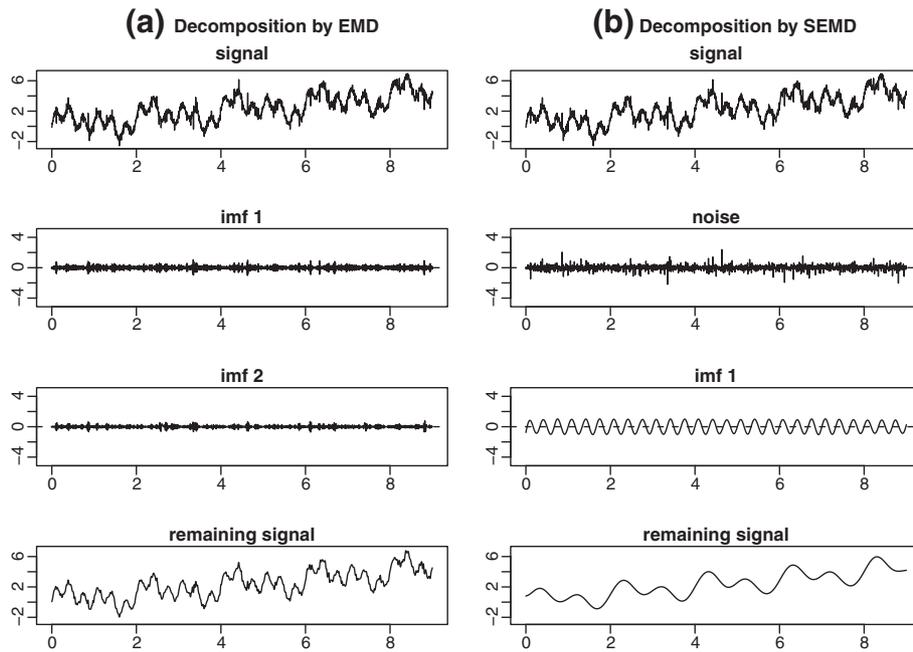


Figure 8 Decomposition results by EMD and SEMD for a heavy-tailed noise with SNR 5.

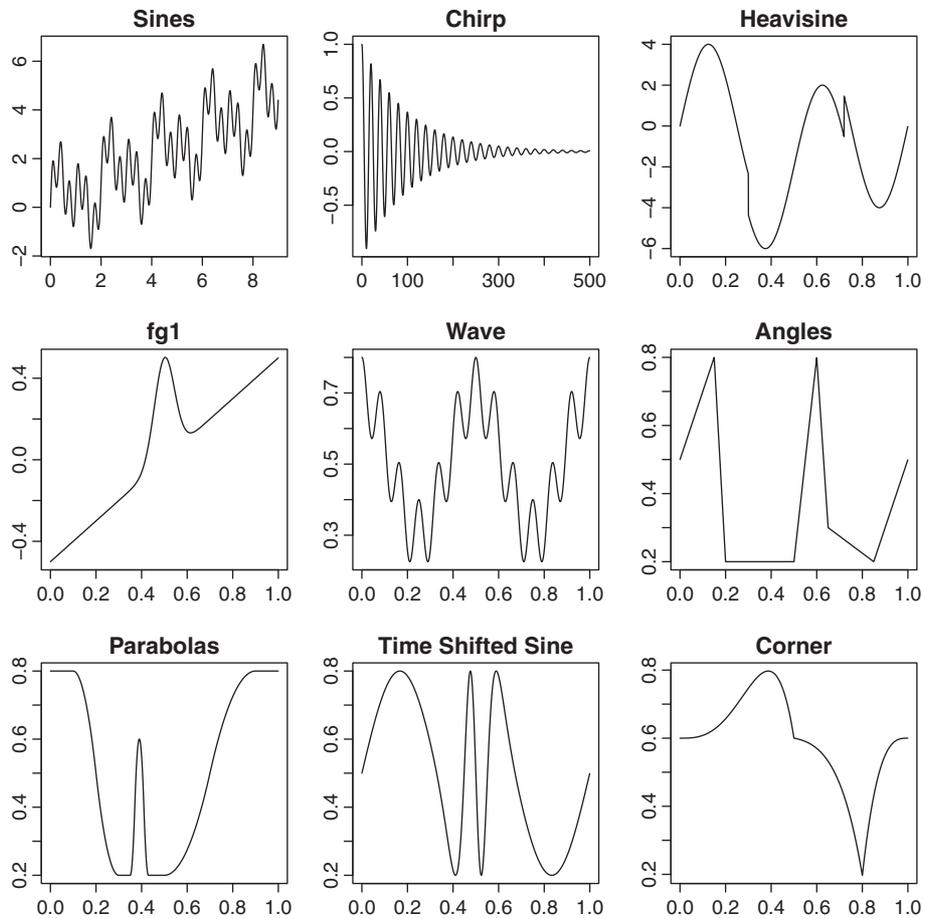


Figure 9 Nine test functions used in the simulation study.

component from the noisy signal. As a measure of performance, the mean squared error, MSE ($= \frac{1}{n} \sum \{m(t) - \hat{m}(t)\}^2$) is then calculated for each method. Figures 10 and 11 show average MSE as function of SNR, averaged over 100 datasets from two noisy types.

From the simulation results, the following main observations can be made: (i) noise distorts the decomposition results in the case of the conventional EMD, (ii) wavelet shrinkage outperforms the conventional EMD in recovering the true function, (iii) SEMD is the most effective in removing noise from a noisy signal, and (iv) SEMD is robust to the presence of extreme values. In summary, the simulation results illustrate that SEMD is an effective decomposition method for separating noise or outliers from signals.

SEMD for signals with an ultra-high frequency component

It is known that if the extrema of a signal are located near the extrema of the highest-frequency component, the conventional EMD effectively decomposes a signal [12,13]. However, when a signal contains an ultra-high

frequency component, the very small gaps between the extrema of a signal and the highest-frequency component can degrade the desired representation of the envelopes. Thus, an ultra-high frequency component may be incorrectly extracted, as shown in Figures 12, 13 and 14. A signal of Figure 12 is generated from model (1) with frequencies $f_1 = 80$, $f_2 = 2$, and $f_3 = 1$, and for generating signals in Figures 13 and 14, an ultra-high frequency component of $0.1 \sin(480\pi t)$ is added to the two signals, fg1 and Wave, respectively. Since smoothing can help to properly construct the hidden symmetric envelopes of an ultra-high frequency component, unlike the interpolation technique, we apply SEMD to the three simulated signals. As shown in the figures, SEMD successfully decomposes the ultra-high frequency IMF and the subsequent IMFs. Note that conventional EMD produces artificial IMFs.

Boundary treatment by SEMD

When constructing envelopes during the sifting process, inadequate information is available on the modulation of two boundaries before the first extremum and after the last extremum. Unless the boundaries are properly

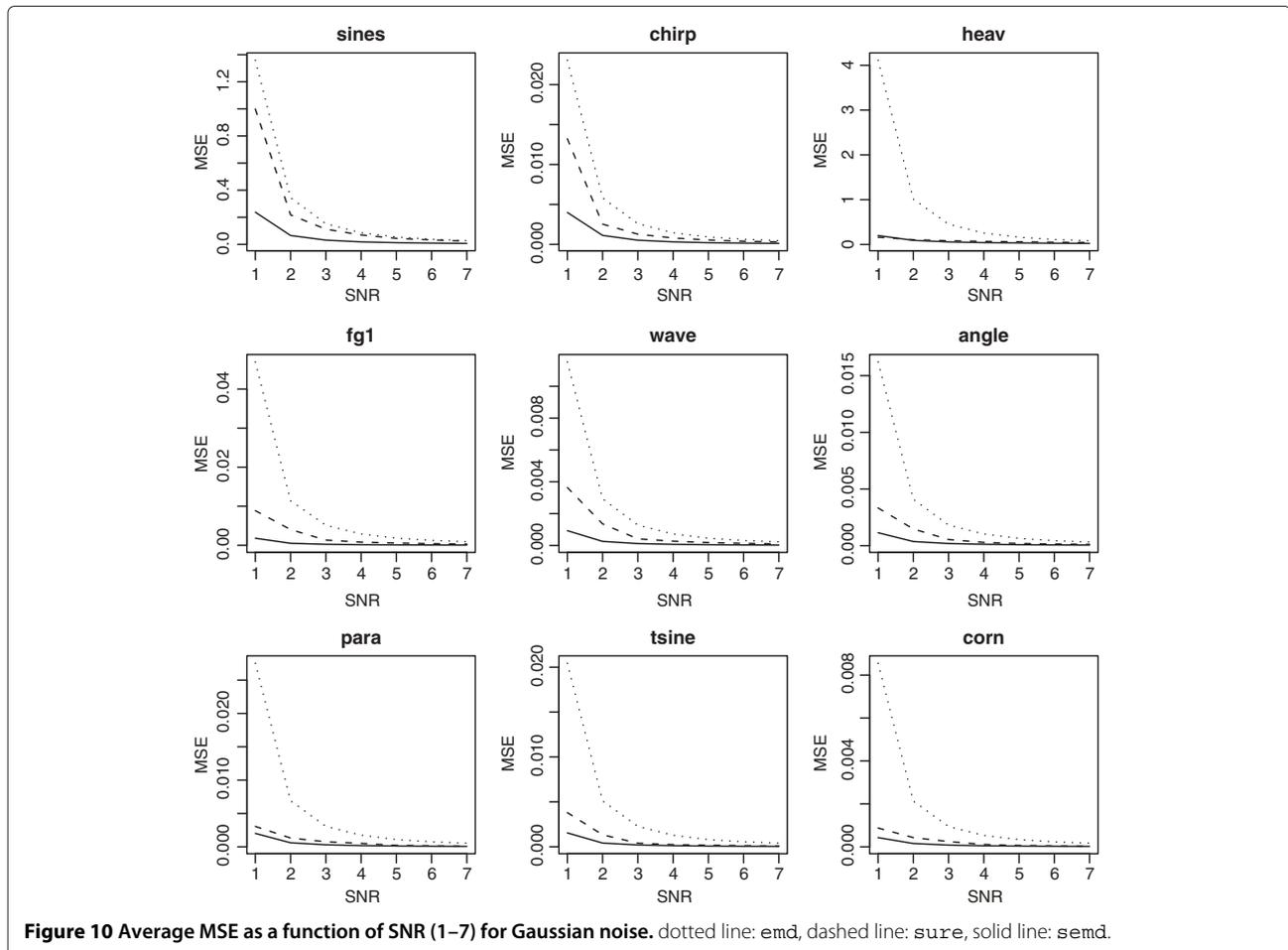


Figure 10 Average MSE as a function of SNR (1–7) for Gaussian noise. dotted line: emd, dashed line: sure, solid line: semd.

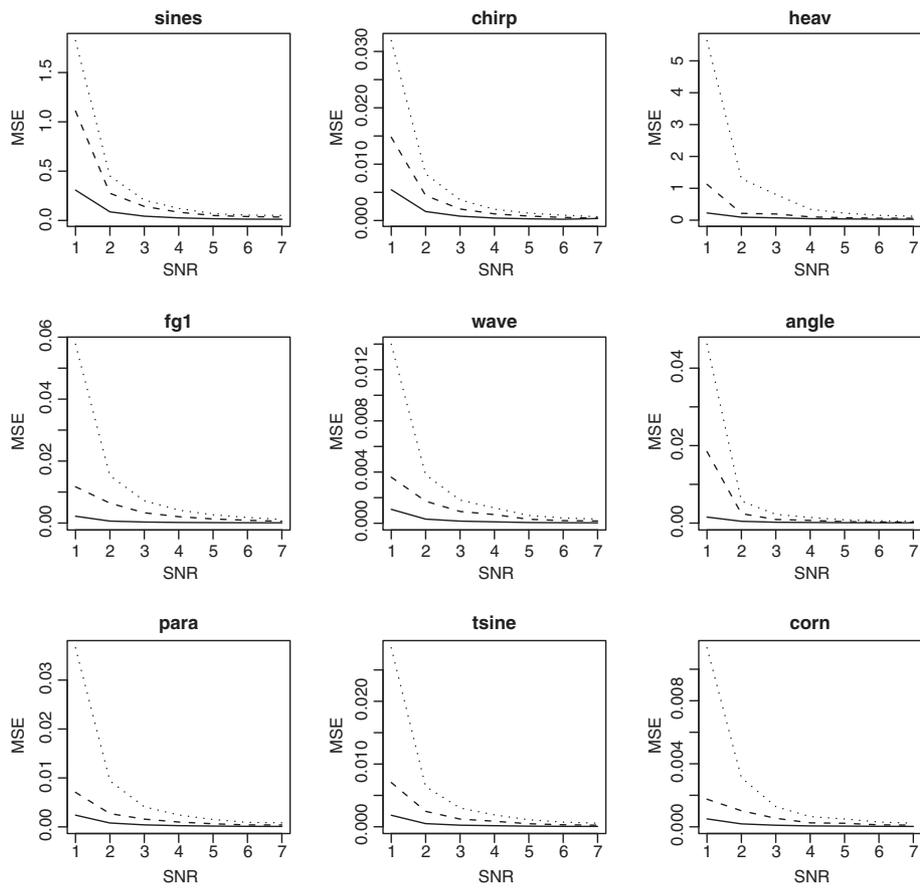


Figure 11 Average MSE as a function of SNR (1–7) for t -distribution noise. dotted line: emd, dashed line: sure, solid line: semd.

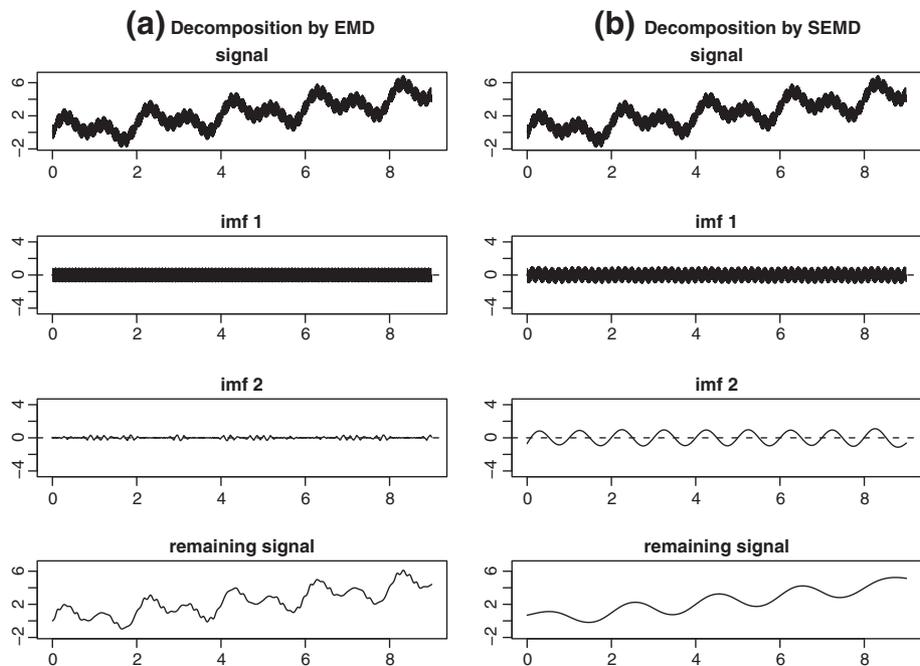


Figure 12 Decomposition results by EMD and SEMD for a signal $x(t) = 0.5t + \sin(80\pi t) + \sin(2\pi t) + \sin(\pi t)$, $0 < t < 9$.

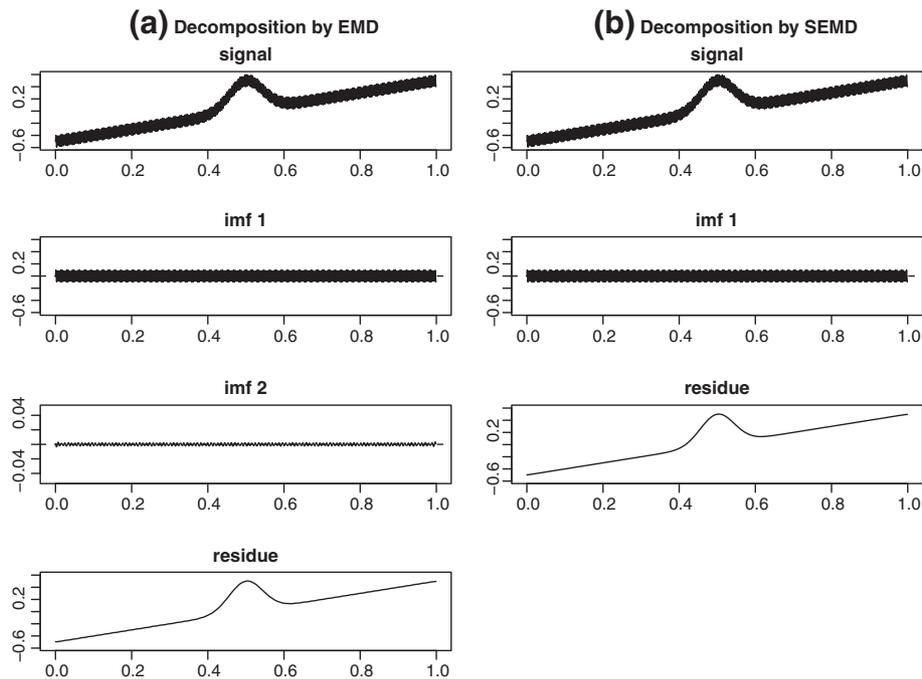


Figure 13 Decomposition results by EMD and SEMD for fg1 signal with an ultra-high frequency component $0.1 \sin(480\pi t)$.

treated, large swings occur on both sides, and these eventually distort the entire decomposition result. This phenomenon is particularly exaggerated in lower-frequency IMFs because there is inadequate information on an intrinsic mode. In addition to traditional boundary treatments such as periodic or symmetric conditions, Huang

et al. [5] extended the original signal by adding artificial waves called characteristic waves, and these can be constructed by repeating the intrinsic mode formed by extreme values nearest to the boundary.

To evaluate the performance of boundary treatment of SEMD, we consider three test signals. For a signal

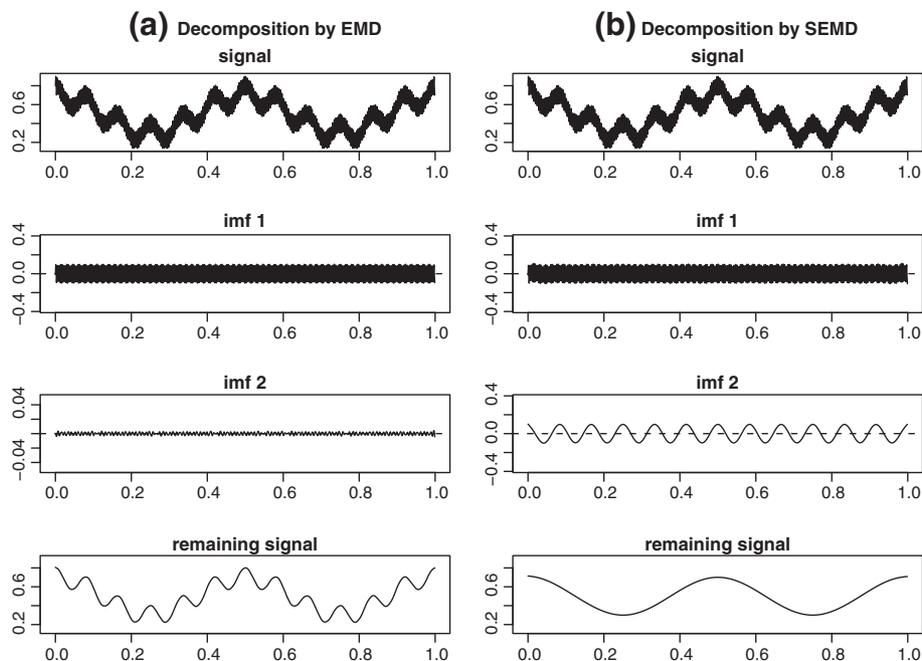


Figure 14 Decomposition results by EMD and SEMD for Wave signal with an ultra-high frequency component $0.1 \sin(480\pi t)$.

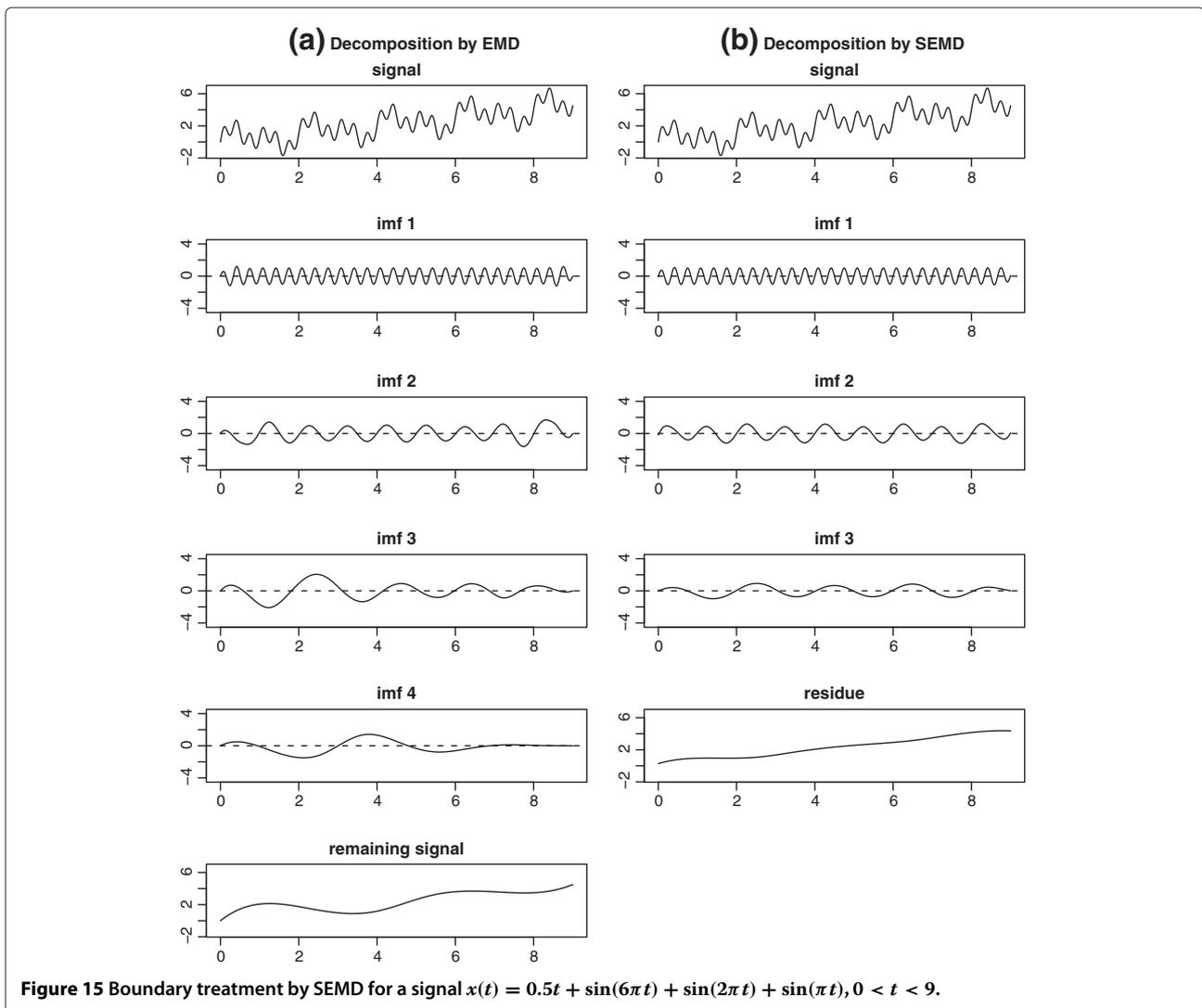
in Figure 15, frequencies of model (1) are defined as $f_1 = 6$, $f_2 = 2$, and $f_3 = 1$, and we generate signals in Figures 16 and 17 by adding a term $0.1 \sin(60\pi t)$ to the two signals, fg1 and Wave, respectively. The left-hand side panels of Figures 15, 16 and 17 show the decomposition results of the conventional EMD without any boundary treatment. The relatively large amplitude at both boundaries in the first IMF eventually has an effect on the sequential IMFs. This effect subsequently produces the artificial IMFs, which are not a component of the original signal. On the other hand, SEMD itself provides an alternative to the boundary problem treatment, that is, without using any periodic/symmetric condition or characteristic wave, SEMD can alleviate the boundary problem by applying a smoothing procedure to all levels. As shown in the right-hand side panels of Figures 15, 16 and 17, the decomposition results of SEMD provide a better boundary adjustment, and

this can produce appropriate components on the entire domain.

Extension of EMD to irregularly spaced signals

We consider an extension of EMD to irregularly spaced signals. The conventional EMD interpolates in-between extrema using cubic splines; this might not be appropriate for obtaining the upper and lower envelopes when the observed data are scattered: they are not observed on regular (spatial) grids, and they have spatially inhomogeneous densities including data voids of various sizes.

Here, we propose a new method based on the combination of a simulation technique for generating random fields and the SEMD algorithm, called *simulation-based SEMD*. This method can be easily adopted for one-dimensional signals. The proposed method comprises two steps: (1) Extrema are generated on a regularly spaced domain by a simulation method. (2) The upper and lower



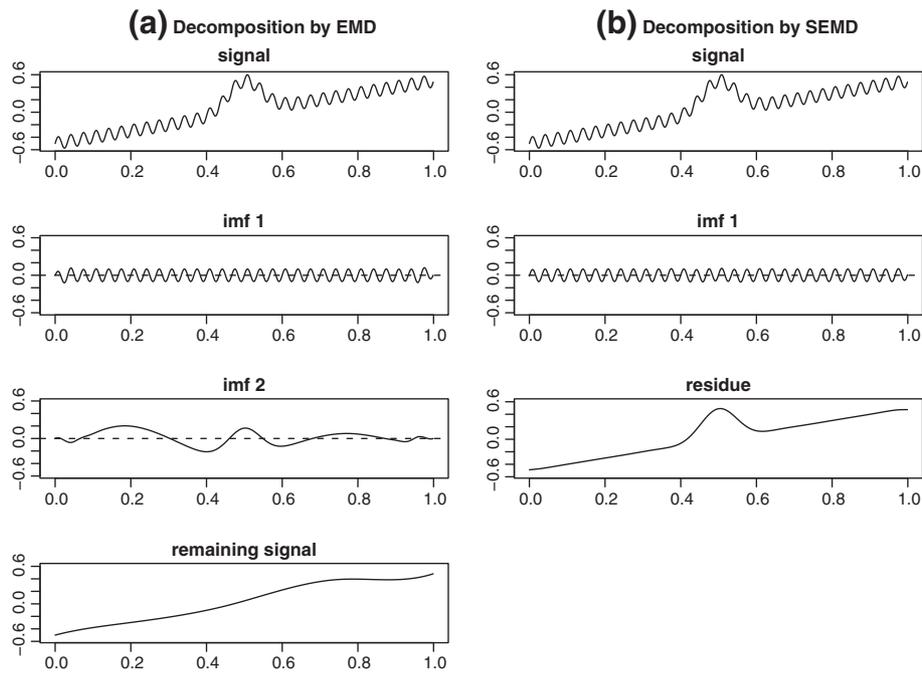


Figure 16 Boundary treatment by SEMD for a signal $x(t) = \text{Eg1}(t) + 0.1 \sin(60\pi t)$.

envelopes are constructed using the simulated extrema and the SEMD algorithm.

A key feature of the proposed simulation-based SEMD method is that it can integrate various patterns between the simulated extrema. Furthermore, the uncertainty of

the resulting IMFs can be evaluated on the basis of several sets of simulations.

To generate simulated extrema, we can use some well-studied methods for simulating random fields in spatial statistics. In this study, we employ a kriging-based

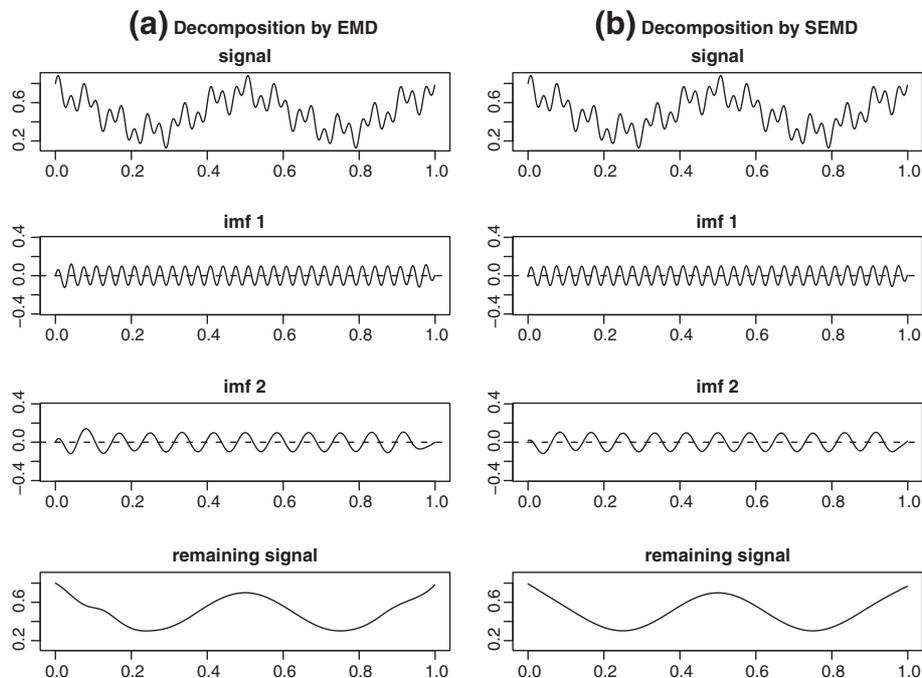


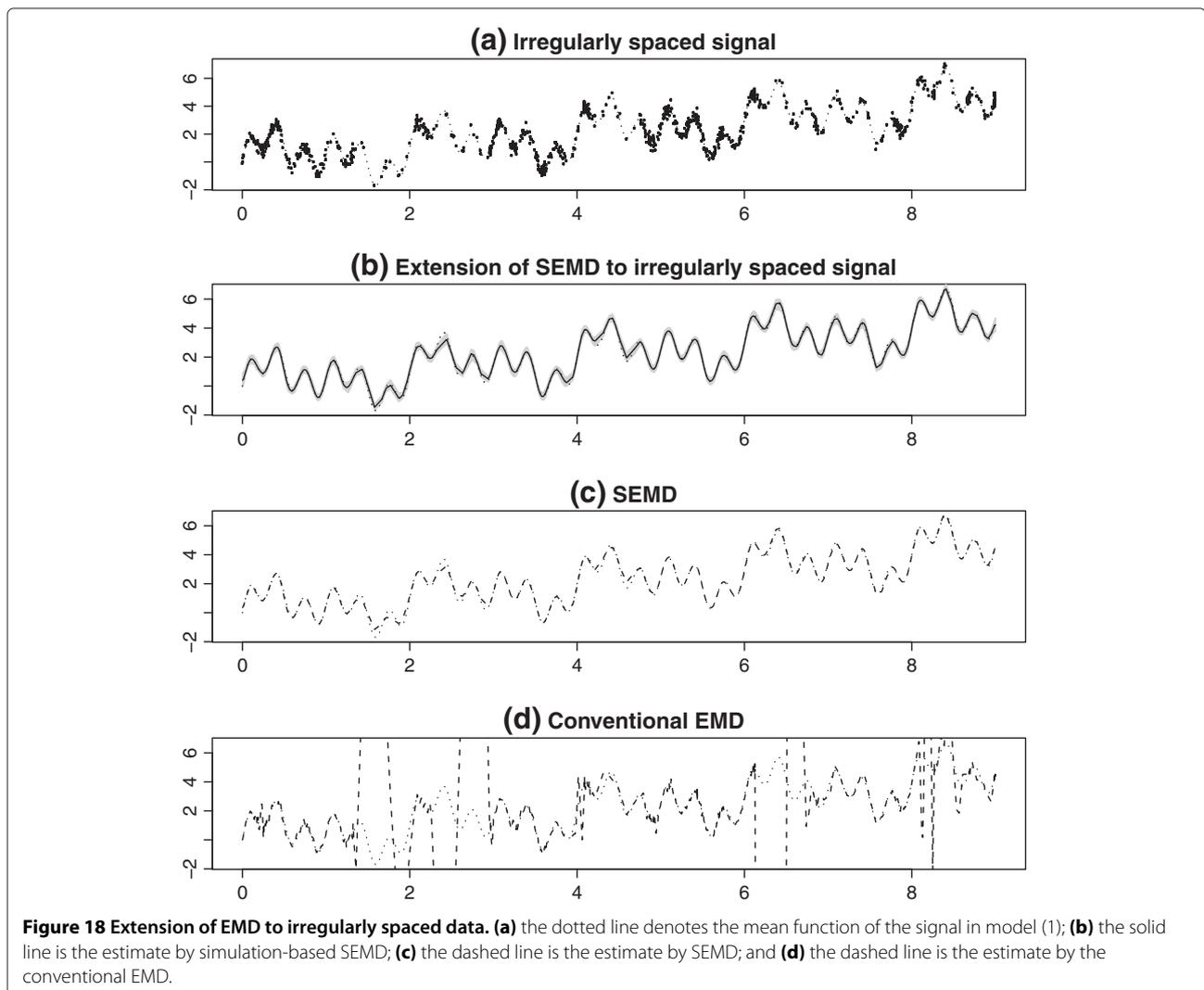
Figure 17 Boundary treatment by SEMD for a signal $x(t) = \text{Wave}(t) + 0.1 \sin(60\pi t)$.

simulation method that is described below. Consider a Gaussian random field with a covariance function Σ and a realization vector $\mathbf{x}(\mathbf{s}) = [x(s_1), \dots, x(s_m)]^T$ sampled at irregularly spaced locations $\mathbf{s} = [s_1, \dots, s_m]^T$. The aim is to generate a set of extrema on a regular grid with the same mean and covariance structure as $\mathbf{x}(\mathbf{s})$ and to ensure that the realization passes through the observed values. Suppose that we have the decomposition

$$\mathbf{x}(\mathbf{t}) = p(\mathbf{t}; \mathbf{x}) + \mathbf{e}(\mathbf{t}) - p(\mathbf{t}; \mathbf{x}),$$

where $p(\mathbf{t}; \mathbf{x})$ denotes the kriging predictor at regularly spaced locations $\mathbf{t} = (t_1, \dots, t_n)$ that depends on \mathbf{x} . The quantity $\mathbf{e}(\mathbf{t}) = \mathbf{x}(\mathbf{t}) - p(\mathbf{t}; \mathbf{x})$ denotes kriging residuals that are not available in practice. Therefore, we generate $\mathbf{e}(\mathbf{t})$ with an estimated covariance. Using the simulated values $p(\mathbf{t}; \mathbf{x}) + \mathbf{e}(\mathbf{t})$, we identify the extrema and use the SEMD to obtain the IMF at \mathbf{t} . Irregularly spaced IMFs are derived at \mathbf{s} .

We apply the proposed method to a signal from model (1) with frequencies $f_1 = 6, f_2 = 2$, and $f_3 = 1$. The top panel of Figure 18 shows 1,024 irregularly spaced observations contaminated by Gaussian noise with SNR = 7. The mean estimate for a signal by the simulation-based SEMD method with 200 replications and its 99% empirical confidence interval are indicated by a solid line and gray band in the second panel of Figure 18, respectively. To evaluate the performance of the proposed method, we apply SEMD and the conventional EMD to the irregularly spaced data shown in the top panel. The dashed lines in the third and fourth panels show the reconstructions of SEMD and EMD, respectively. The MSEs of the proposed simulation-based SEMD, SEMD, and EMD are 0.0059, 0.0072, and 511.603, respectively. Furthermore, Figure 19 shows the mean estimates of the IMFs and residue by the simulation-based SEMD with 200 replications. From the results of Figure 19, the simulation-based SEMD method is capable of effectively decomposing irregularly spaced data.



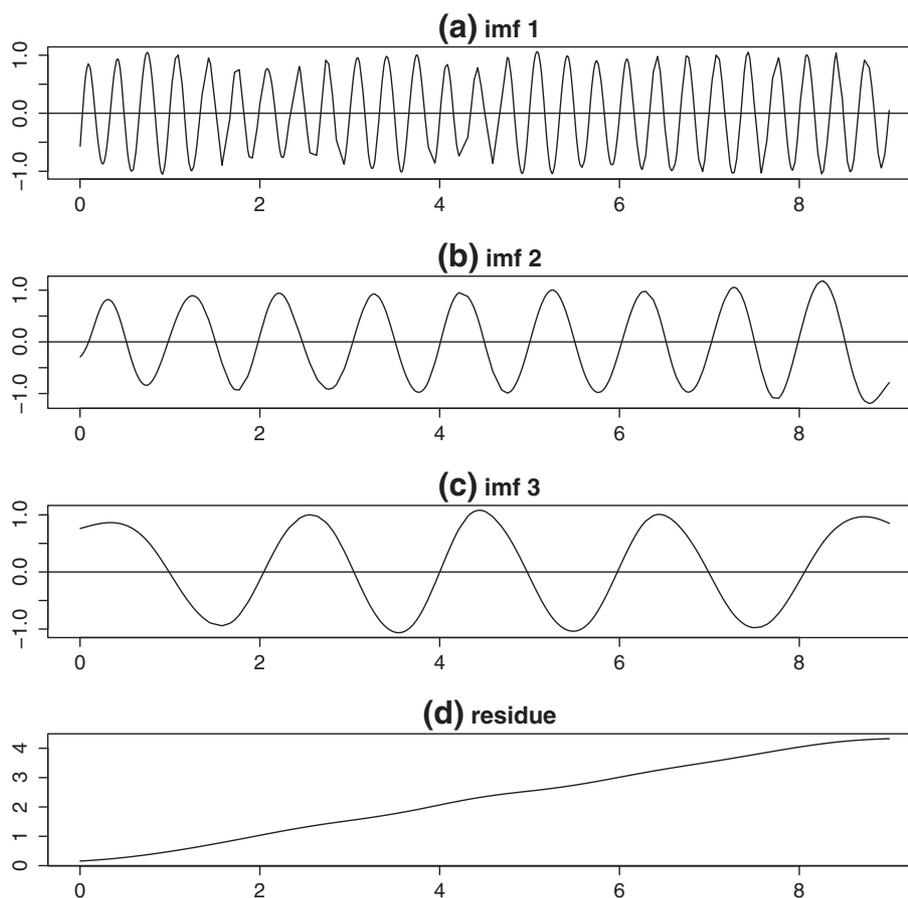


Figure 19 The mean estimates of IMFs and residue by the simulation-based SEMD with 200 replications.

Conclusion

In this article, we have proposed a statistical EMD to deal with a noisy signal by combining smoothing techniques and the conventional EMD. The results obtained from various numerical experiments confirm the effectiveness of the statistical EMD method. Furthermore, we have extended EMD to irregularly spaced signals by utilizing simulated extrema. These extensions of the conventional EMD are expected to increase the applications of EMD.

Further studies of the proposed SEMD are needed. The current algorithm of SEMD requires the selection of smoothing parameter, which is indeed computationally expensive and might be an obstacle of handling massive data. Hence, it is necessary to develop a computationally efficient method of smoothing parameter selection. As another possible refinement of SEMD, we would like to investigate intermittence problem of mode mixing, which means that different modes of oscillations coexist in a single IMF. Finally, although SEMD is relatively robust to outliers compared with the conventional EMD, a least-squared-based smoothing method such as kernel smoothing can be affected by outliers in construction of

envelopes. Therefore, it seems that a quantile-based EMD would merit further study.

Competing interest

The authors declare that they have no competing interests

Acknowledgements

This work of Hee-Seok Oh was supported by the National Research Foundation of Korea (NRF) grant (No. 2012002712) funded by the Korea government (MEST). This research of Donghoh Kim was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0076223).

Author details

¹Department of Applied Mathematics, Sejong University, Seoul 143-747, Korea. ²Department of Industrial Engineering, Konkuk University, Seoul 143-701, Korea. ³Department of Statistics, Seoul National University, Seoul 151-747, Korea.

Received: 22 February 2012 Accepted: 25 July 2012

Published: 07 August 2012

References

1. MB Priestley, *Spectral Analysis and Time Series*, vols. 1 and 2 (Academic Press, New York, 1981)
2. S Mallat, *A Wavelet Tour of Signal Processing* (Academic Press, New York, 2009)

3. I Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992)
4. B Vidakovic, *Statistical Modeling by Wavelets* (John Wiley & Sons, New York, 1999)
5. NE Huang, Z Shen, SR Long, MC Wu, HH Shih, Q Zheng, NC Yen, CC Tung, HH Liu, The empirical mode decomposition and Hilbert spectrum for nonlinear and nonstationary time series analysis. *Proc. Roy. Soc. Lond. A.* **454**, 903–995 (1998)
6. NE Huang, SSP Shen, *Hilbert-Huang Transform and Its Applications* (World Scientific, Singapore, 2005)
7. AO Boudraa, JC Cexus, EMD-based signal filtering. *IEEE Trans. Instrum. Meas.* **56**, 2196–2202 (2007)
8. Z Wu, NE Huang, Ensemble empirical mode decomposition: a noise assisted data analysis method. *Adv. Adapt. Data Anal.* **1**, 1–49 (2009)
9. Z Xu, B Huang, F Zhang, Improvement of empirical mode decomposition under low sampling rate. *Signal Process.* **89**, 2296–2303 (2009)
10. EHS Diop, R Alexandre, AO Boudraa, Analysis of intrinsic mode functions: a pde approach. *IEEE Signal Process. Lett.* **17**, 398–401 (2010)
11. T Hastie, R Tibshirani, *Generalized additive models* (Chapman and Hall, London, 1990)
12. G Rilling, P Flandrin, One or two frequencies? the empirical mode decomposition answers. *IEEE Trans. Signal Process.* **56**, 85–95 (2008)
13. M Park, D Kim, HS Oh, A reinterpretation of EMD by cubic spline interpolation. *Adv. Adapt. Data Anal.* **3**, 527–540 (2011)
14. T Lindberg, *Scale-Space Theory in Computer Vision* (Kluwer, Boston, 1994)
15. BW Silverman, Using Kernel density estimates to investigate multimodality. *J. Roy. Stat. Soc. B.* **43**, 97–99 (1981)
16. DL Donoho, IM Johnstone, Adapting to unknown smoothing via wavelet shrinkage. *J. Am. Stat. Assoc.* **90**, 1200–1224 (1995)
17. DL Donoho, IM Johnstone, Ideal spatial adaptation by wavelet shrinkage. *Biometrika.* **81**, 425–455 (1994)
18. J Fan, I Gijbels, Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. Roy. Stat. Soc. B.* **57**, 371–394 (1995)
19. JS Marron, S Adak, IM Johnstone, MH Neumann, P Patil, Exact risk analysis of wavelet regression. *J. Comput. Graph. Stat.* **7**, 278–309 (1998)
20. TT Cai, Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Ann. Stat.* **27**, 898–924 (1999)

doi:10.1186/1687-6180-2012-168

Cite this article as: Kim et al.: Extending the scope of empirical mode decomposition by smoothing. *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:168.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com