Discussion Paper Series

2009 – 6

Department of Economics
Royal Holloway College
University of London
Egham TW20 0EX

USING EMPIRICAL MODE DECOMPOSITION TO ESTIMATE AMPLITUDES IN NOISY DATA

Claire Blackman

Economics Department
Royal Holloway
University of London

ABSTRACT

Empirical Mode Decomposition, an adaptive data-driven technique which can be used to extract non-stationary signals buried in noise, seldom admits theoretical calculation of the statistical properties of the extracted signals. Instead, numerical experiments are required. In this paper we use Monte Carlo simulations to investigate the accuracy of the amplitudes of sinusoids extracted from synthetic noisy signals using Empirical Mode Decomposition. We show that even for relatively low signal-to-noise data, the amplitude of the extracted signal is close to true amplitude. We also show that edge effects due to the spline curves which are used to calculate the decomposition do not affect the amplitude estimate beyond the first two oscillations.

Index Terms— Empirical mode decomposition, amplitude estimation, low signal-to-noise data

1. INTRODUCTION

Empirical Mode Decomposition (EMD) was developed by [1] to decompose non-stationary multi-component data into a set of Intrinsic Mode Functions (IMFs) for which a meaningful instantaneous frequency could be defined everywhere. Since EMD is a signal-dependent adaptive technique for which theoretical analysis is seldom possible, numerical experiments are often required to understand its properties. For example, [2] show that even for a pure tone, EMD may not extract the single mode correctly if the sampling period is insufficient. They also give an empirical relationship for the minimum distance necessary between frequencies for them to be correctly resolved. [3] show that the EMD of noisy signals can act as a dyadic filterbank, similar to wavelet analysis; based on this, [4] discuss the selection of true signal modes from noisy decompositions, and the construction of confidence intervals for the extracted modes in the case of fractional Gaussian noise.

EMD can be used to extract quasi-periodic oscillations buried in noisy data, and allow their amplitudes to be measured directly. It is important, however, to assess how accurately the amplitude of an extracted quasi-periodicity reflects the amplitude of the true signal amplitude. The purpose of this paper, therefore, is to investigate the amplitudes of signals recovered from noisy data using EMD. We create synthetic noisy signals and compare the average amplitude of the extracted signals with the (known) signal amplitudes, over a range of noise levels. This provides a simple method of testing the degree to which we can trust the amplitudes of signal extracted using EMD, when the true amplitude is unknown.

Section 2 gives a brief overview of Empirical Mode Decomposition, and discusses a few of the implementation issues encountered with this method. In section 3 we discuss the construction of the synthetic data and the Monte Carlo method we have used in our analysis. We discuss the results of our simulations in section 4, and conclude with section 5.

2. EMPIRICAL MODE DECOMPOSITION

It is assumed that the data have at least two extrema (one maximum and one minimum) and that the characteristic time scale (of each IMF) is defined by the time lapse between extrema. This does require that the data are oversampled sufficiently that the extrema are well-defined [2]. If the data contain no extrema, only inflection points, then differencing the data once or more will reveal the extrema, and the final results can be obtained by summing the components.

To extract the IMFs from the original data, $x(t)$, a sifting process is followed:

1. All local maxima are identified, and an upper envelope, $e_{max}(t)$ is constructed by interpolating between the local maxima with a cubic spline. The procedure is repeated for the local minima, forming a lower envelope, $e_{min}(t)$.

2. The mean $m_1(t)$ of the upper and lower envelopes

is calculated:

$$m_1(t) = \frac{(e_{max}(t) - e_{min}(t))}{2}$$

3. By subtracting $m_1(t)$ from $x(t)$ the first IMF, $h_1(t)$, is calculated:

$$h_1(t) = x(t) - m_1(t)$$

   $h_1(t)$ contains the finest scale or shortest period component of the data.

4. The process is repeated, using $m_1(t)$ as the new data, until the extracted mean is monotonic.

The time series is thus decomposed into n IMFs and a residual, $m_n(t)$ which is either the mean trend or a constant:

$$x(t) = \sum_{j=1}^{n} h_j(t) + m_n(t) \qquad (1)$$

Unless the spline ends are correctly constrained, they have a tendency to propagate unwanted oscillations into the envelopes, affecting the calculated mean, and hence the decomposition. To constrain the spline ends accurately, we have added ∼5 oscillations at each end of the time series, constructed to bring the splines gradually to zero. The oscillations mimic the behaviour of the data near the end, ensuring that the splines do not oscillate wildly.

Ideally, $h_1(t)$ should be an IMF. However, a gentle slope may be amplified to become a local extremum in changing from rectilinear to curvilinear coordinates, so the sifting process is repeated at each step using h_1 as the input until the extracted signal is an IMF. The usual criterion for stopping is when the number of extrema equals the number of zero crossings [5]. An additional constraint sometimes used is that the mean of the upper and lower envelopes has to be 'close' to zero, for some threshold. [2] have extended this to include two thresholds: one criterion to ensure globally small fluctuations from the mean, and one to allow locally large deviations from the mean. However, we have not found this necessary for our data.

3. DATA AND METHOD

[6] use EMD to analyse rotation residuals of the solar convection zone, and estimate the error in the EMD of their signal by constructing realizations of their time series which they use for Monte Carlo simulations. Each point in each realization is drawn from a distribution with mean given by the value of the true data point, and standard deviation given by the measurement error.

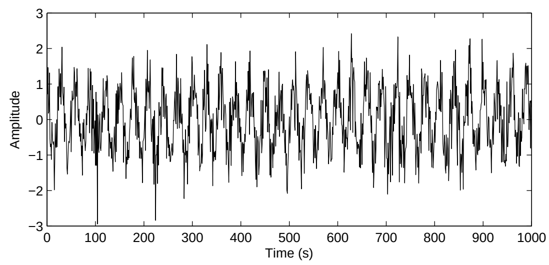

Fig. 1. A sinusoid of period 30s, with gaussian white noise of amplitude 0.6 added.

They find that the realization and the original IMF components are in good agreement. However, this method is not practical for large data sets; we propose an alternate Monte Carlo simulation using purely synthetic signals to find the average amplitude of extracted signals for various levels of noise. This has the advantage that different types of signals can be embedded in different levels of noise, and comparisons made.

3.1. Data

We consider a simple test signal consisting of a sinusoid of period 30s, length 1000s, sampled at 1s intervals, with amplitude 1. We create the 5000 noisy signals used in each Monte Carlo simulation by adding a realisation of gaussian white noise with the required amplitude to the test signal. Figure 1 shows one realisation of the noisy sinusoid, with a noise level of 0.6.

3.2. Method

Each noisy signal is decomposed into five IMF components; further IMFs could be extracted, but we found that five is sufficient to capture the IMFs that contain the signal. The left-hand panels of Figure 2 show the five IMFs of the signal shown in Figure 1. h_1 contains the highest frequency components of the noisy signal; successive IMFs contain longer period data. The IMFs h_3 and h_4 both contain regions of the sinusoidal signal. This splitting of the signal over IMFs is a consequence of the adaptive nature of EMD, and makes automated identification of IMFs containing signal components non-trivial. We use the highest peaks in the periodogram of each IMF to initially identify IMF components containing (part of) the signal component; the right-hand panels in Figure 2 show the periodograms of the calculated IMFs. It is clear that both IMFs h_3 and h_4 contain significant power at 0.03Hz, corresponding to the 30s periodic signal.

Once the signal-containing IMFs have been found, we identify the maxima and minima in the IMFs which

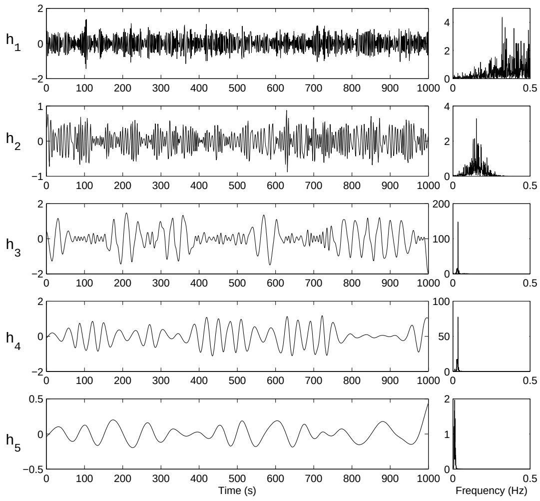

Fig. 2. The left-hand panels show successive IMFs from the Empirical Mode Decomposition of sinusoidal signal corrupted with gaussian white noise of amplitude 0.6. The panel to the right of each IMF shows the periodogram of the IMF. Both IMFs h_3 and h_4 contain regions of the underlying 30s sinusoid.

correspond to true signal maxima and minima: an extremum in an IMF is considered to correspond to a true signal extremum if it occurs within 5s of the true signal extremum, and has an amplitude greater than 0.6 times the true signal extremum. Figure 3 shows the IMF h_3. The positions of the true signal maxima are indicated by open triangles, while the IMF maxima identified as corresponding to a signal maximum are shown by open circles.

Thus from each realisation, we extract the times and amplitudes of extrema corresponding to true signal extrema. Averaging over the 5000 realisations, we are find the average amplitude and standard deviation of each extracted extremum.

4. RESULTS

Figure 4 shows the noise-free signal (solid line), with the average amplitude and standard errors of each extremum (black dots) calculated from 5000 realisations with a noise level of 0.6. Edge effects, due to the fitting of the spline curves, are evident in the extrema at either end of the data, where the average amplitude is somewhat lower than the true value. However, these effects do not propagate inward beyond two oscillations.

Averaging the average extrema and standard deviations over all extrema gives us an average amplitude and

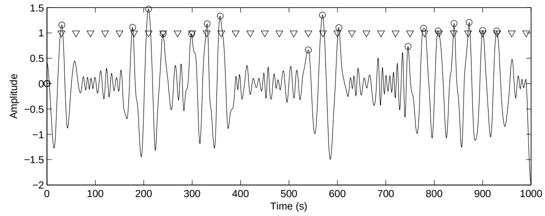

Fig. 3. The IMF h_3 (solid line) and the IMF maxima identified as corresponding to a signal maximum (open circles). The positions of the true signal maxima are indicated by open triangles.

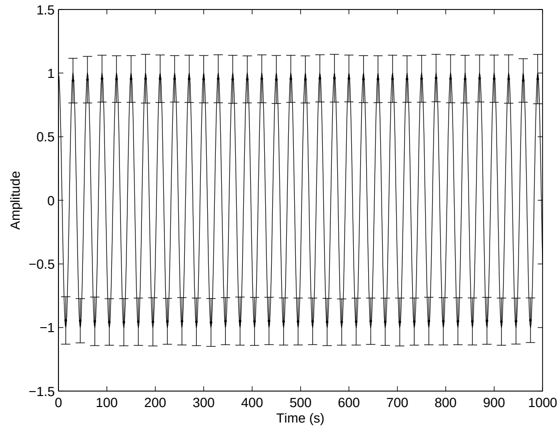

Fig. 4. The noise-free signal is indicated by the solid line. The average extrema and standard deviations found by Monte Carlo simulations (using a noise level of 0.6) are indicated by black dots and error bars.

Table 1. Average amplitudes and standard deviations from Monte Carlo simulations of signal extraction from noisy data.

Noise Level	Average IMF Amplitude	Average std
0.1	0.98	0.13
0.2	1.01	0.10
0.3	1.02	0.16
0.4	0.99	0.20
0.5	0.96	0.20
0.6	0.95	0.19
0.7	0.98	0.19

error for the extracted signal, which can be compared with the know amplitude of 1 of the noise-free signal.

Table 1 gives the average amplitudes and standard deviations found for noise levels ranging from 0.1 to 0.7. 5000 realizations were used at each noise level. Even when the noise level is 0.7, the average amplitude is close to the true amplitude, and in all cases the true amplitude (1) lies within the error bars.

5. CONCLUSIONS

We have developed a method for assessing the accuracy of the amplitude of signals extracted from composite noisy signals using Empirical Mode Decomposition. We have shown that even for data with a relatively low signal-to-noise ratio, the amplitudes of signal-containing IMFs accurately reflect the amplitude of the noise-free signal. Edge effects, due to the constraining of the spline curves used in the fitting of the envelopes, do not appear to propagate beyond the first two oscillations.

This method can easily be extended to investigate more complicated test signals containing multiple periodic components, and quasi-periodic components. Noise models other than gaussian white noise can also be investigated. Another extension of this method is to test the accuracy of the time at which extracted extrema are detected in comparison to the times at which the real signal extrema occur.

6. REFERENCES

[1] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N. Yen, C.C. Tung, and H.H. Liu, "The empirical mode decomposition and hilbert spectrum for nonlinear and nonstationary time series analysis," *Proc R. Soc. Lond. A*, vol. 454, pp. 903–995, 1998.

[2] G. Rilling, P. Flandrin, and P. Gonçalvès, "On empirical mode decomposition and its algorithms," 2003.

[3] P. Flandrin, G. Rilling, and P. Gonçalvès, "Empirical mode decomposition as a filter bank," *IEEE Signal Process. Lett.*, vol. 11, pp. 112–114, 2004.

[4] P. Flandrin, P. Gonçalvès, and G. Rilling, "Detrending and denoising with empirical mode decomposition," in *Proceedings of the 12th European Signal Processing Conference (EUSIPCO '04)*, Vienna, Austria, September 2004.

[5] N.E. Huang, Z. Shen, and S.R. Long, ," *Ann. Rev. Fluid Mech.*, vol. 31, pp. 417–457, 1999.

[6] R. Komm, F. Hill, and R. R. Howe, "Empirical mode decomposition and hilbert analysis applied to rotation residuals of the solar convection zone.," *Astrophysical Journal*, vol. 558, pp. 428–441, 2001.