

The Adaptive EEMD Denoising Method based on Normalized Index Optimization for Near Infrared Spectra

Tong Liang, Zhao Xiaoyu, Sha Lijuan, Shao Hongbo and Zhao Ming

Communication and Electronic engineering Institute, Qiqihar University, Qiqihar 161006, China

*College of Information Technology Heilongjiang Bayi Agriculture University
Communication and Electronic engineering Institute, Qiqihar University, Qiqihar
Computer and Control engineering Institute, Qiqihar University, Qiqihar 161006,
China*

*Chemistry and Chemical Engineering, Qiqihar University, Qiqihar 161006, China
qqhrutl@sohu.com, xy_zhao77@163.com, qqhruslj@sohu.com,
qqhrushb@sohu.com, qqhruzm@sohu.com*

Abstract

Firstly, modes overlapping occurs easily between Near Infrared spectral high frequency signal and noise, so the Ensemble Empirical Mode Decomposition (EEMD) method was used to remove the noise. According to the approximate completeness of Near Infrared spectrum, an EEMD denoising method based on normalized index optimization. This method views the global optimal point of normalized denoising index as stop condition, and completes itself adaptively with no parameters setting during the decomposition. It can be concluded by Near Infrared spectrum denoising data that this method improves the denoising accuracy and decreases the difficulty in operation in an easy and efficient way.

Keywords: EEMD; Near Infrared spectra; denoise; adaptivity

1. Introduction

There are three types of noise in Near Infrared spectra, the electronic thermal motion noise of instruments, the interference from stray light and the outside interference from communication system, which have nothing to do with components of the sample and have changed the location, height, energy of spectral peaks badly. It is necessary to denoise Near Infrared spectral signals to reduce the interference from noise and power the effective information.

The method of Empirical Mode Decomposition (EMD) could compose the signal into IMF components with different amplitude and frequency, and then signals are separated from noise. The EMD method has been applied to process biomedical data, fault signal, and earthquake signal (see [1-5]) and so on. In 2010, Near Infrared first order derivative spectra were denoised by EMD, which raised the decision coefficient of forecasting model from 0.9705 to 0.9832, and declined the root mean square error from 0.5606 to 0.3310 (see [6]). In [7], EMD was combined with successive projections algorithm to extract spectral characteristic wavelengths, which reduced the total wavelengths for model to 3.6%. Later, Li [8] decomposed Raman spectra on EMD, and denoised its data by multi scale filter.

2. The Deficiency in Decomposition of Soil Near Infrared Spectra on EMD

Empirical Mode Decomposition has been proposed by N.E.Huang in 1998, in which the signal is decomposed into finite Intrinsic Mode Functions (IMF). For any IMF, there are the global condition and the local condition to be met,

The number of extreme points is equal to that of zero crossing or the biggest difference is one in the global scope;

(2) The mean of enveloping line between local maximum and minimum values is zero at any point.

The signal $x(t)$ is decomposed on EMD as follows,

Step 1: Enveloping line between local maximum and minimum values is calculated by three spline function, and its mean is expressed as $m(t)$,

$$m(t) = \frac{1}{2}[u(t) + v(t)] \quad (1)$$

Step 2: Let $R(t) = x(t) - m(t)$, if $R(t)$ can not meet with the condition of IMF, and then assign it to $x(t)$. Next return to step 1, and then calculate circularly,

$$R_{1k}(t) = h_{1k-1}(t) - m_{1k}(t) \quad (2)$$

Until the condition of IMF in (1) and (2) is satisfied.

Step 3: Then the first IMF $c_1 = R_{1k}(t)$ and the residual component $r_1(t) = x(t) - c_1$ are get.

Step 4: Return to the step 1 to decompose the residual component, we can get $c_2(t)$, $c_3(t)$, ..., $c_k(t)$ (or $IMF_1, IMF_2, \dots, IMF_k$) successively. The decomposition will end when the residual component meets the stopping rules.

Near Infrared spectra of the soil were decomposed by EMD in the paper.

The Near infrared spectrometer is WQF-600N, and its source voltage is 5V. The scanning scope is from 10000 cm^{-1} to 3500 cm^{-1} , and the background is scanned 32 times and samples are scanned 16 times with the resolution of 8 cm^{-1} in the experiment.

There are 200 samples, each 1g, which have been grinded and sieved with 60 meshes every inch. The Figure 1 is one of transmitted Near Infrared spectrum.

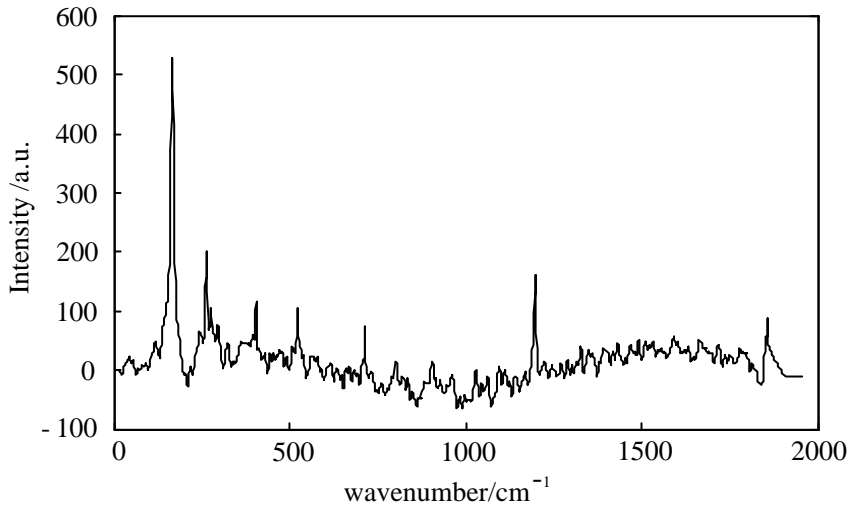


Figure 1. Transmitted Near Infrared Spectrum of the Soil

In Figure 1, the wave is sharp and dramatic changed and the frequency of effective information is so high that data intertwines with noise. So it's necessary to concern the denoising method of high frequency signals in substantial transformation.

The Near Infrared spectrum is decomposed into 9 IMF components and a residual (r) by EMD in Matlab2011b, see Figure 2.

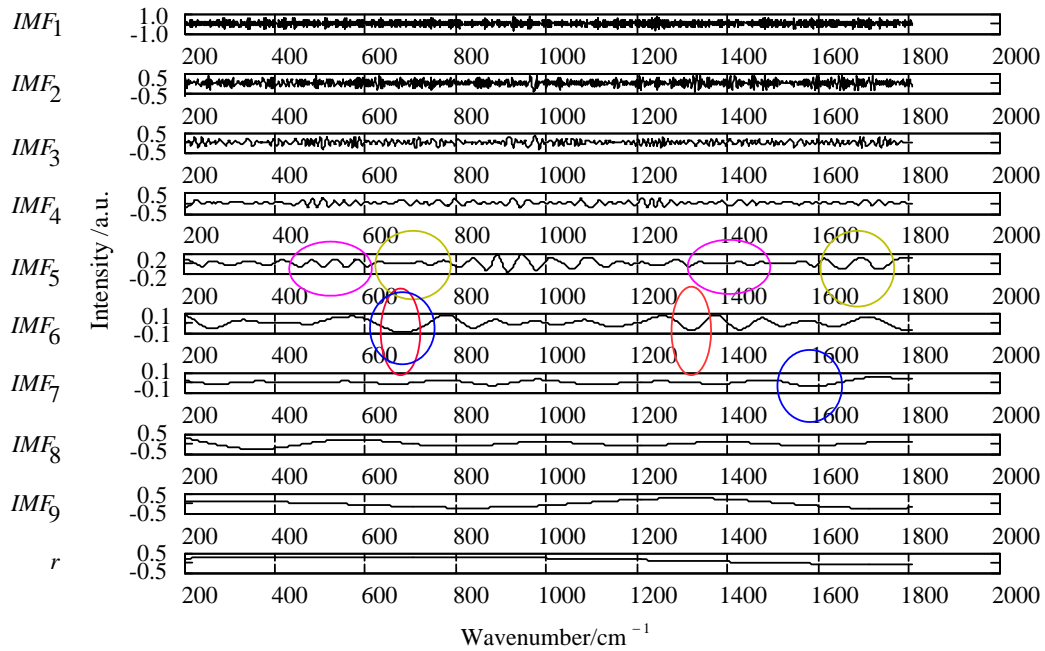


Figure 2. IMFs and r from EMD Decomposition of the Soil Near Infrared Spectrum

There are different characteristic time scales in one IMF, such as in IMF5 and IMF6, and the similar time scale is in IMF6 and IMF7, which is called modal aliasing. High frequency of soil spectral signal is part of problem, so the energy of signal and noise are overlapping. And abnormal events can not be smoothed for absence of extreme points near the characteristic peaks of high amplitude. So EMD low pass filter can not denoise Near Infrared spectra of soil.

3. The Adaptive EEMD Denoising Method based on Normalized Index Optimization for Near Infrared Spectra

3.1. Decomposition of Near Infrared Spectra by EEMD

Wu and Huang have proposed Ensemble Empirical Mode Decomposition (EEMD) to overcome overlapping in EMD, see [9]. For statistical characteristics of Gauss white noise is frequency distribution, that it can provide extreme points enough to smooth abnormal events, such as characteristic peaks of high amplitude. Therefore add Gauss white noise into spectral signals to be processed, and then average IMFs from EMD to overcome mode overlapping, which is known as EEMD.

EEMD key steps are as follows,

Step 1: Gauss white noise with the same length of signals and different amplitude is added into signals to be processed, and then composite signals are decomposed k times by EMD, hence IMFs and residual component r_{ik} have been got.

Step 2: IMFs are averaged,

$$IMF_i = \frac{1}{n} \sum_{k=1}^n IMF_{ik} \quad (5)$$

$$r_i = \frac{1}{n} \sum_{k=1}^n r_{ik} \quad (6)$$

In which, n is the time of decomposition.

The amplitude of Gauss white noise is 0.2 times the standard deviation of Near Infrared data, and then IMFs and residual component from EEMD are compared with those from EMD, see Figure 3.

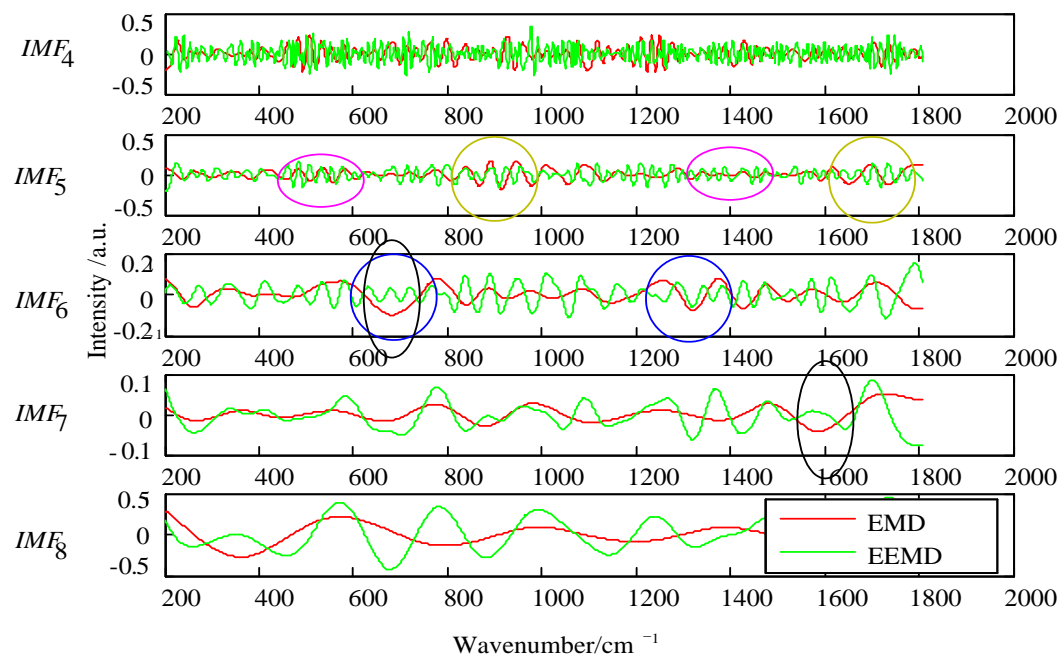


Figure 3. Contrast between EEMD and EMD

In Figure 3, there is mode overlapping in IMF5, IMF6 and IMF7 from EMD showed in red line, but frequencies of those is so clearly that modal aliasing is disappear when the Near Infrared spectrum is decomposed by EEMD, see green line. White noise provides decomposition scale in well-distributed for signal to be processed, so the error of fitting is reduced near the extreme points and EEMD fits for denoising of Near Infrared spectra of the soil.

As we known, the higher frequency is, the sooner it is decomposed in EMD or EEMD. That is to say, each IMF has its own frequency separately, and frequencies distribute according to the powder of 2 degressively. So low pass filters can denoise signals, see Formula 7.

$$x(t) = \sum_{i=1}^n IMF_i + r \quad (7)$$

In which, n is the number of IMFs and $1 < I \leq n$.

In low pass filter on the basis of EMD or EEMD, I is a important parameter which defines denosing precision. In the paper, performance benchmarks of transform, signal to noise ratio, etc. determine the parameter I . When signal and noise are in the IMF synchronously, low pass filter on EMD or EEMD will remove or keep the IMF all, and therefore the denoising is excessive or deficient. So this paper provides a decomposition and restruction method on EEMD and designs the normalized index to improve the denosing precision.

3.2. Design of the Normalized Index

Signal to Noise Ratio (SNR), Standard Error (SE), and Correlation Coefficient (R) are used to explain the denoising, and they are defined as follows,

$$SNR = 10 \log \frac{S}{N} = 10 \log \left(\frac{\sum_{i=1}^n x^2(i)}{\sum_{i=1}^n [x(i) - x'(i)]^2} \right) \quad (8)$$

$$SE = \sqrt{\frac{1}{n} \sum_{i=1}^n [x(i) - x'(i)]^2} \quad (9)$$

$$R = \frac{\sum_{i=1}^n (x(i) - \bar{x})(x'(i) - \bar{x}')}{\sqrt{\sum_{i=1}^n (x(i) - \bar{x})^2} \sqrt{\sum_{i=1}^n (x'(i) - \bar{x}')^2}} \quad (10)$$

In which, x is the original spectral signal and x' is the spectrum has been denoised,

and
$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x(i) \quad \bar{x}' = \frac{1}{n} \sum_{i=1}^n x'(i)$$

The power of signal and noise is expressed by SNR, the difference between signal denoised and original signal is calculate by SE, and the similarity between denoising and original spectrum is expressed by R . Therefore R interprets wave change and SE illustrate the drift of wave peak. For the complex transformation in power, shape, and drift, we construct the normalized index of denoising r ,

$$r = \left| \frac{SNR' - SNR}{SNR} + \frac{1}{SE} + R \right| \tag{11}$$

In which, SNR' is the signal denoised to noise ratio. The normalized index of denoising r is bigger, indexes of Signal to Noise Ratio, Standard Error SE, and Correlation Coefficient are better, and that is to say, the complex denoising effect is better.

4. The Realization and Application of Adaptive EEMD Denoising Method Based on Normalized Index Optimization

Near Infrared spectra of the soil with noise are signals to be processed, and x_i is said to be the spectral signal after i times denoising and the flow chart of the denoising method based on normalized index optimization as follows,

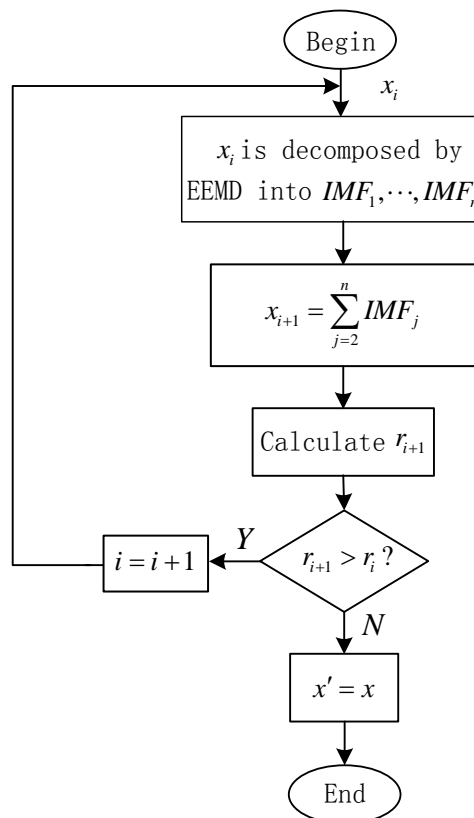


Figure 4. Adaptive Normalized index optimization de-noising flowchart based on EEMD

Step 1: The spectral signal x_i is decomposed by EEMD into n IMFs, in which the

residual component is defined as IMF_n;

Step 2: Restructure the spectral signal x_{i+1} by removing the lowest step but the highest frequency IMF1;

Step 3: Calculate r_{i+1} and compare it with r_i . If $r_{i+1} > r_i$, then the $i+1$ time denoising optimizes the normalized index and it is reasonable to decompose, filter and restructure again. So let $i = i + 1$, and return to Step 1; If $r_{i+1} < r_i$, then the filter at i time has the normalized index, and we easily obtain the signal denoised by $x = x_i$.

The Near Infrared spectrum of soil is mixed into Gauss white noise with 10 dB SNR, showed as follows,

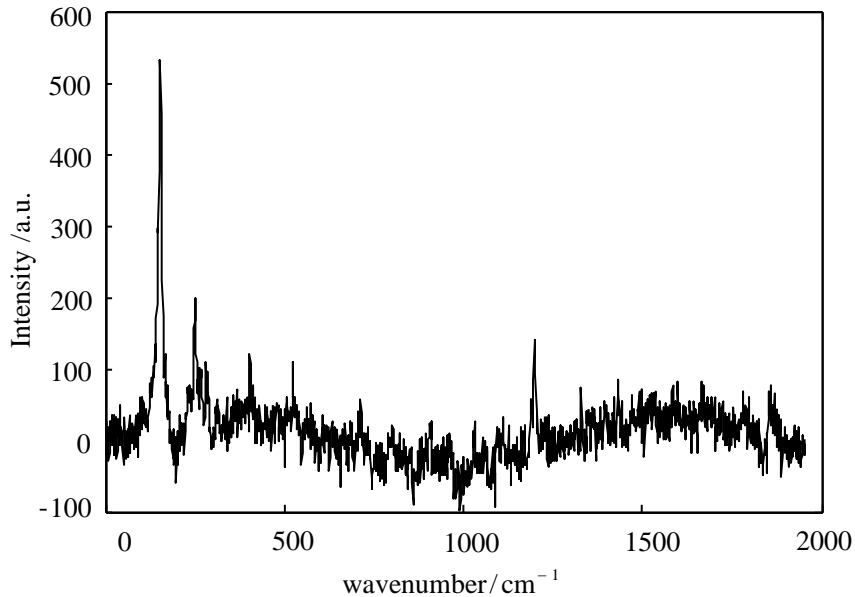


Figure 5. Near Infrared Spectrum of Soil with 10dB Noise

Near Infrared spectrum is processed according on the flow chart Figure 4, and its SNR' , SE , R and r are showed in Table 1.

Table 1. Denoise Indexes of Normalized Index Optimization on EEMD

	SNR' / dB	SE	R	r
1	13.6128	2.0501	0.9668	1.8156
2	16.5629	3.3444	0.9811	1.9360
3	19.6470	7.4928	0.9494	2.0471
4	21.8718	16.4086	0.8936	2.1413
5	21.5263	21.6305	0.7426	1.9410
6	21.4841	51.7598	0.6568	1.8241
7	20.0379	46.7850	0.5979	1.6226
8	19.0549	152.6485	0.5078	1.4195
9	19.3752	13.6107	0.3441	1.3547

Numbers in first column are times of decomposition on EEMD in Table 1. From denoising normalized index r , we can see that r increases monotonously in previous four

times denoising, which indicates processes to spectrum are right, and r decreases monotonously in last five times, which indicates that next processes are spare. There is the only maximum r overall, therefore it could indicate how many times the spectrum will be decomposed, and so the adaptive denoising to Near Infrared spectra are realized. Denoising effects of decompositions are showed from Figure 6 to Figure 14.

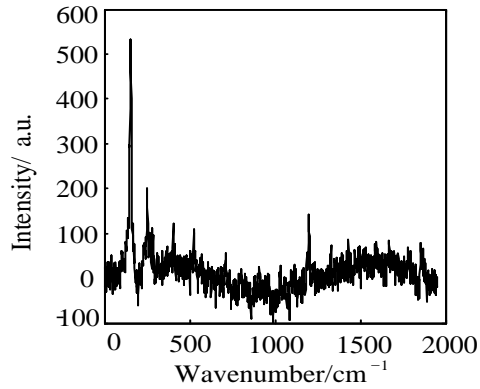


Figure 6. Denoising Effect of One Decomposition

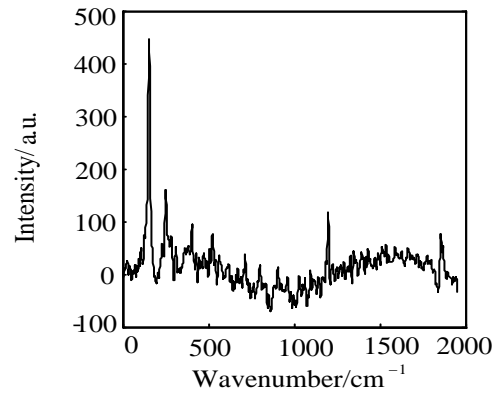


Figure 7. Denoising Effect of Two Decompositions

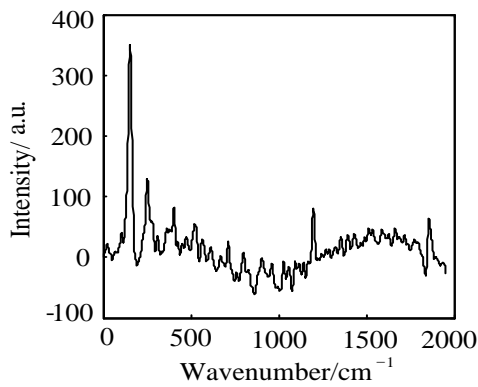


Figure 8. Denoising Effect of Three Decompositions

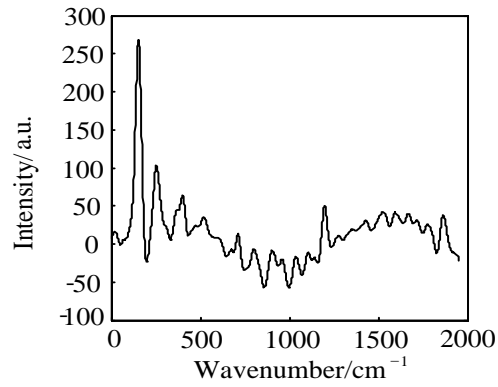


Figure 9. Denoising Effect of Four Decompositions

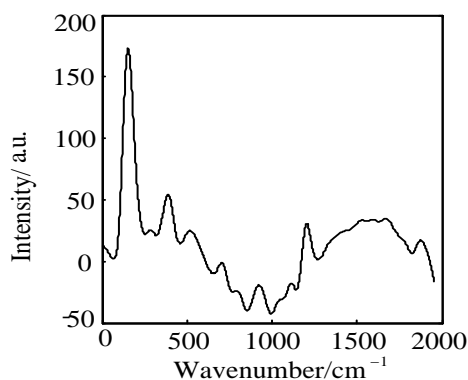


Figure 10. Denoising Effect of Five Decompositions

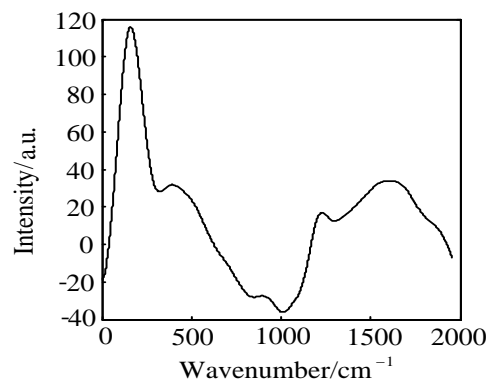


Figure 11. Denoising Effect of Six Decompositions

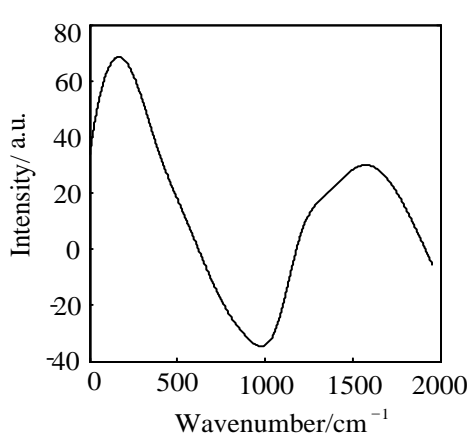


Figure 12. Denoising Effect of Seven Decompositions

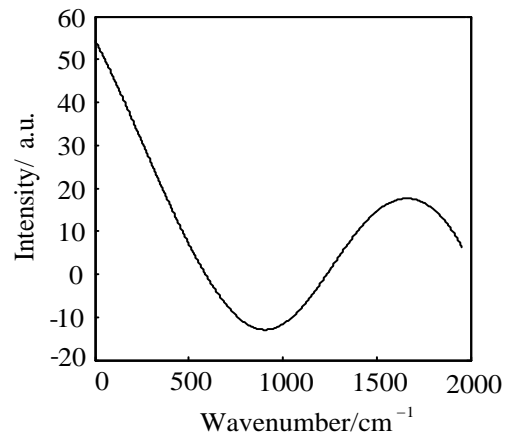


Figure 13. Denoising Effect of Eight Decompositions

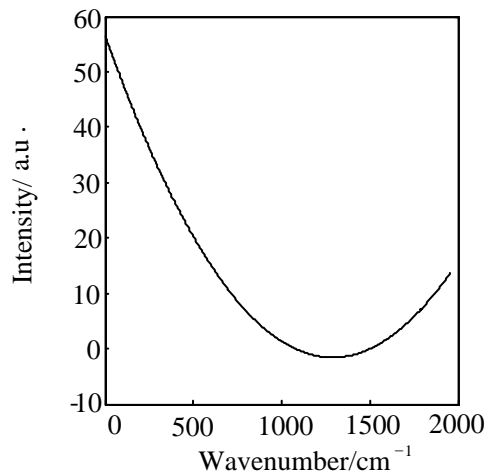


Figure 14. Denoising Effect of Nine Decompositions

The frequency of the lowest IMF is the highest in EEMD with no modal aliasing, so the denoising removing the lowest and the highest component is better step by step till it is too far and the spectrum is getting distorted seen from Figure 6 to Figure 14. In Figure 9, the spectrum is smooth, undistorted and keeps all characteristics acromions after 4 times decompositions and restructures, and which is as same as the denoising index showed in Table 1. Above experiments show that the best number of decompositions could be indicated by the normalized index r , so Near Infrared spectra of the soil can be denoised on EEMD.

In addition, Near Infrared spectra are decomposed by EMD into 10 IMFs (the residual component is show in IMF_{10}), and the normalized index r is the evaluation parameter too as follows,

Table 2. Denoise Indexes by Normalized Index Optimization on EMD

	SNR' /dB	SE	R	r
1	12.4979	2.2344	0.9743	1.6714
2	10.4293	4.0643	0.9821	1.2708
3	7.4593	14.3471	0.9506	0.7661
4	17.9154	15.3261	0.9295	1.7859
5	12.7006	20.3136	0.8484	1.1674
6	10.2145	59.0825	0.6928	0.7310
7	11.1018	55.2067	0.6519	0.7799
8	12.3172	63.3058	0.5815	0.8287
9	14.0887	300.0641	0.1610	0.5729

At first, the values of r in Table 2 are lower than those of EEMD. Secondly, SNR' in the second decomposition is lower than that in the first time, and it suggests that the first IMF contains effective components and parts of signal will be lost and SNR' will be lower if the first IMF is cut off directly. The index r becomes lower and then higher in the fifth, sixth and seventh decomposition similarly, that is to say, high frequency signals and noise components cannot be distinguished effectively in Near Infrared spectra decomposition by EMD, and it proves that non-uniform scales in components of IMF_5 , IMF_6 and IMF_7 are modes overlapping. The fourth filter is adopted for it's the maximal index, and its denoising result as Figure 18.

Then Near Infrared spectra of soil are denoised by EMD lowpass filter, wavelet decomposition, and EMD denoising method based on normalized index optimization, and results of denoising as Figure 15, Figure 16, Figure 17 and Figure 18.

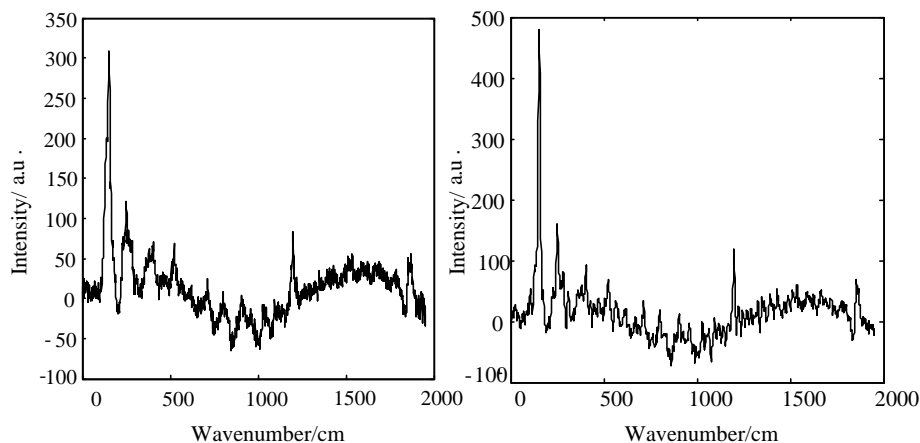


Figure 15. Denoise by EMD Low Pass Filter Figure 16. Denoise by Wavelet₁

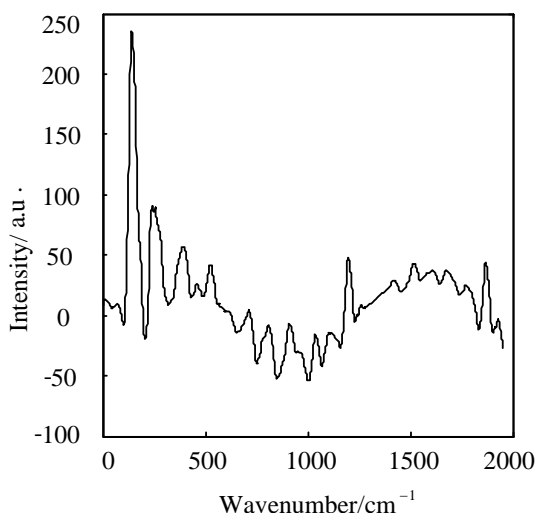


Figure 17. Denoise by Wavelet2

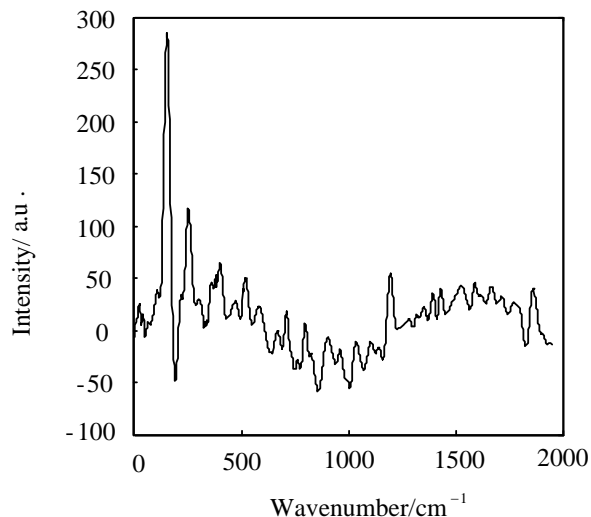


Figure 18. Denoise by Normalized Optimization EMD Four Steps Filter

There are a lot of residual noise in EMD lowpass filter with one decomposition and one IMF be cut off in Figure 15.

In 2006, Chen chen had researched wavelet decomposition on Raman spectra in [10], with conclusions that stable wavelet function is the best method for spectra denoised after contrast in 8 wavelet functions. So parameters of Wavelet1 in the paper were set according to the conclusion in [10] as follows: bior2.2 wavelet function, 5 decompositions, soft threshold, and Rigrsure for threshold made. And then from Figure 16, there are a lot of residual noise, which shows that there needs different wavelet parameters to be set for different spectra. At last, parameters of Wavelet2 are set through many combination tests between wavelet basic function, levels of decompositions, and ways for threshold made as follows: db4 wavelet function, 5 decompositions, 2/3 times smaller to detail coefficients. The denoising result of Wavelet2 is better than that of Wavelet1 obviously from Figure 17. Near Infrared spectra were denoised by the forth reconstruction after were decomposed into 10 IMFs in Figure 20, and filters are not enough on 50 cm⁻¹, 230 cm⁻¹, 270 cm⁻¹, 1420 cm⁻¹, and signal energy lost too much on 1200 cm⁻¹.

Denoising index of EMD lowpass filter, two kinds of wavelet, EMD method based on normalized index optimization and EEMD method based on normalized index optimization for Near Infrared Spectra are listed in Table 3, in which LPF represents low pass filter and NIO represents denoising method based on normalized index optimization.

Table 3. De-Noising Index of 10 dB SNR

	SNR/dB	SE	R	r
EMD LPF	10.8964	3.5604	0.9832	1.3535
Wavelet1	16.5831	12.1722	0.7110	1.4511
Wavelet2	19.1046	11.0164	0.9230	1.9238
EMD NIO	17.9154	15.3261	0.9295	1.7859
EEMD NIO	21.8718	16.4086	0.8936	2.1413

We can conclude that EMD lowpass filter and EMD method based on normalized

index optimization have no advantage comparison with Wavelet2 and EEMD method based on normalized index optimization from Table 3. For that there are effective signal and noise jointly in the secondly IMF after the first EMD decomposition, a lot of noise remain in EMD lowpass filter. And there have no the global optimal denoising index in EMD method based on normalized index optimization, so r cannot to be the evidence to decide decompositions showed in Table 2.

Indexes of Wavelet1 and Wavelet2 are better than that of EMD lowpass filter, which indicates that wavelet can deal with non stationary and nonlinear Near Infrared spectra. Then index contrast between Wavelet1 and Wavelet2, it concludes that the effect of wavelet denoising relies on rationality of parameters. So it needs adjustment according to spectra denoised, and there is no experience to set parameters. In Table 3, it gets the biggest SNR and the normalized index and it is better than EMD lowpass filter, wavelet, EMD method based on normalized index optimization, because that the method of EEMD smooths characteristic peaks in Near Infrared spectra to overcome modes overlapping and to get high frequency signal and noise separately. And it needs no parameters in advance in EEMD method based on normalized index optimization, so it can denoise Near Infrared spectra adaptively.

5. Conclusion

At first, we introduce EMD in the paper, and then decompose Near Infrared spectra on EMD, in which mode overlapping occurs. So EEMD is used to decompose Near Infrared spectra to conquer modes overlapping. Above all, proving on approximate completeness becomes theoretical basis for Near Infrared spectra denoised on EEMD, and then design the normalized index, by which the level of EEMD decomposition circularly can be determined automatically to avoid setting parameters, such as basis function, threshold and let level *etc.* In the soil Near Infrared spectra denoised, it is better than wavelet with appropriate parameters, EMD lowpass filter, and EMD method based on normalized index optimization.

Acknowledgments

This research was supported by the Science and Technology Research Project of Heilongjiang Provincial Education Department of China (Grant No. 12541897), all support is gratefully acknowledged.

References

- [1] S. Jinan, C. Yanping and W. Lei, "Study on the Method of Extracting Fault Feature in Gearbox based on Improved EMD[J]", *Journal of Mechanical Transmission*, vol. 5, no. 38, (2014).
- [2] L. Ling, L. Wei and Z. Mei, "Application of EMD Algorithm to the Dynamic Spectrum Non-Invasive Measurement of Hemoglobin[J]", *Spectroscopy and Spectral Analysis*, vol. 8, no. 34, (2014).
- [3] X. Dan, Y. Long and W. Yumei, "Mechanical Fault Diagnosis Method for Circuit Breaker Based on the Combination of Energy Entropy and Improved EMD by Using the Least Square Method [J]", *High Voltage Apparatus*, vol. 8, no. 50, (2014).
- [4] L. Mengmeng, W. Min and X. Hui, "ECG signal in-band noises de-noising based on EMD method [J]", *Journal of Tianjin Polytechnic University*, vol. 4, no. 33, (2014).
- [5] W. Jiao, L. Zhenchun and W. Deying, "A method for wavelet threshold denoising of seismic data based on CEEMD[J]", *Geophysical Prospecting for Petroleum*, vol. 2, no. 53, (2014).
- [6] Cai Jianhua and Wang Xianchun. Near-Infrared Spectrum Pretreatment Based on Empirical Mode Decomposition [J]. *Acta Optica Sinica*. 1, 30 (2010)
- [7] Zhang Zhiyong, Li Gang and Lin Ling. Application of EMD and SPA Algorithm in the Detection of

- Benzoyl Peroxide Addition in Flour by Spectroscopy[J]. Spectroscopy and Spectral Analysis. 10, 32 (2012)
- [8] Li Qing, Zhang Guoping and Liu Yang. A Study of Raman Spectra Denoising Based on Empirical Mode Decomposition [J]. Spectroscopy and Spectral Analysis. 1, 29 (2009)
- [9] Wu Z H and Huang N E. Ensemble Empirical Mode Decomposition: A noise Assisted Data Analysis Method[J]. Advances in Adaptive Data Analysis. 1, 1 (2009)
- [10] Chen Chen, Xu Dahai and Cheng Qinghua. A De-noising Technique for Raman Spectra Based on Wavelet Transform. [J]. Journal Yangtze University(Nat Sci Edit). 3, 4 (2006)

Authors



Author's Name: Tong Liang

Author's profile: Tong Liang (1977-), male, Han nationality, native place:Heilongjiang. Associate Professor, Bachelor degree. Tel: 13904523065, E-mail: qqhrutl@sohu.com>

