An improved genetic algorithm for optimizing ensemble empirical mode decomposition method

Dabin Zhang, Chaomin Cai, Shanying Chen and Liwen Ling

College of Mathematics and Informatics, South China Agricultural University, Guangzhou, People’s Republic of China

ABSTRACT
This paper proposes an improved ensemble empirical mode decomposition method based on genetic algorithm to solve the mode mixing problem in empirical mode decomposition (EMD) algorithm as well as the parameters selection issue in ensemble empirical mode decomposition (EEMD) algorithm. In a genetic algorithm (GA), the orthogonality index is used to formulate the fitness function and the Hamming distance is specified to design the difference selection operator. By coupling GA with EEMD algorithm, an improved decomposition method with higher efficiency is generated, namely GAEEEMD. Simulation experiment with both intermittent signals and sinusoidal signals verifies the effectiveness and robustness of the proposed GAEEEMD, compared with EMD, EEMD, and original GA algorithm.

1. Introduction
Huang et al. (1998), Huang, Shen, and Long (1999) proposed an adaptive time-frequency data analysis method, namely Empirical Mode Decomposition (EMD). By using EMD, a complex signal can be adaptively decomposed into several Intrinsic Mode Functions (IMFs) according to the local time scale features of the signal. By extracting IMFs from nonlinear and non-stationary data, EMD can be widely used in geophysical and seismic signal analysis (Huang & Wu, 2008), mechanical fault diagnosis (Mohammad, Siamak, & Mohammad, 2018; Yuan et al., 2018), bioelectrical signal analysis (Hikmat, Maha, & Enas, 2018; Kumar, Panigrahy, & Sahu, 2018), time Sequence prediction (Li, Tao, Ao, Yang, & Bai, 2018; Wei, Wei, Wang, & Lai, 2018; Xiang, Gou, He, Xia, & Wang, 2018), image processing (Qin, Qiao, Wang, Ren, & Zhu, 2018; Yu, Li, Zhang, Liu, & Max, 2015), speech signal analysis (Rajib, 2016) and so on. Although EMD performs well in analyzing non-stationary signals, it still has some annoying difficulties unresolved.

One of the major drawbacks of the original EMD is the mode mixing problem, which means that multiple decomposition components appear in the same frequency waveform or some decomposition components contains a plurality of different frequency waveforms. Huang et al. (1999) considered that the mode mixing problem of EMD is due to the presence of intermittent signals, resulting in uneven distribution of the extreme points. Therefore, Huang et al. proposed the intermittency test, which ameliorated these difficulties to some extent. However, this approach makes the EMD cease to be totally adaptive and depends on a subjectively selected scale. Wu and Huang (2009) proposed a noise-assisted data analysis method named EEMD. The EEMD suppresses the mode mixing problem of EMD by adding white noise to the initial data in many times. The EEMD approach is a major breakthrough in the development of EMD and significantly improves the stability of the EMD algorithm.

However, the performance of EEMD algorithm depends greatly on the noise amplitude and the number of trails. With some inappropriate amplitudes of the added noises and the number of trials, the IMFs generated by EEMD would be highly polluted and even yield the pseudo-components (Zheng, Cheng, & Yang, 2014). These unfavourable decomposed components might have some negative effects on the following modelling process and induce greater forecast errors. Considering the importance of the amplitude and trails of the added white noise, scholars have proposed several methods for parameters selection. For example, Hoseinzadeh, Kadem, and Sadooghi (2018) used mutual information and approximate entropy to select the appropriate amplitude for adding white noise, but the number of trails was not considered in the experiment. Dong, Dai, Tang, and Yu (2019) used the grid search approach to select the
optimal parameters which generated the minimal multi-scale complexity, but the calculation of multi-scale complexity was complicated and it was not suitable for small sample data. According to existed literature, the amplitude and the number of trials of the added white noise are subjectively defined beforehand and mainly follow the simulation results proposed by Wu and Huang (2009). Consequently, the motivation of this paper is to select the optimal parameters (i.e. the amplitude of the added white noise and the number of trials) via the genetic algorithm (GA) and improve the self-adaptiveness of the EEMD method. Specifically, the orthogonality index is used to formulate the fitness function so as to evaluate individual fitness more effectively. Hamming distance is specified to design the difference selection operator in order to figure out the individuals with greater fitness, which guarantee a better convergent result. Based on the GA, the mode mixing and parameters selection problems within the original EMD and EEMD algorithm can be effectively addressed, with the self-adaptiveness of the decomposition method. The rest of the paper is organized as follows. Section 2 introduces the EMD and EEMD algorithms and their disadvantages. Then, the proposed GAEEMD algorithm is presented in Section 3. Next, the proposed algorithm is compared with EMD, EEMD, and the original GA by analyzing experimental results in Section 4. Finally, the paper is concluded in Section 5.

2. Methodology of EMD and EEMD

2.1. EMD

EMD is a self-adaptive data decomposition method, which can decompose a nonlinear and nonstationary time series into several simple but meaningful components called intrinsic mode functions (IMFs) and a residue. The basic idea of the EMD algorithm is to extract the IMFs from a given time series through a shifting process and the main process can be described as follows (Huang et al., 1998):

1. For a given original signal \( x(t) \) \((t > 0)\), set \( r_0(t) = x(t) \) and \( i = 1 \).

2. Construct the upper and lower envelopes for all extreme of the signal \( r_{i-1}(t) \) using the cubic spline interpolation method.

3. Define the baseline \( m_{i-1}(t) \) as the average of the upper and lower envelopes, namely, \( m_{i-1}(t) = \frac{[f_{\text{max}}(t_i) + f_{\text{min}}(t_i)]}{2} = 0 \), \( t_i \in [t_{a}, t_{b}] \).

4. Extract the current oscillatory component: \( IMF_i(t) = r_{i-1}(t) - m_{i-1}(t) \). If \( IMF_i(t) \) is not eligible, repeat steps (2) and (3) by setting \( r_{i-1}(t) = IMF_i(t) \) and iterate until the properties of \( IMF_i(t) \) are satisfied. Then the \( IMF_i(t) \) is treated as an IMF and set \( l_i(t) = IMF_i(t) \).

5. Compute the residue: \( r_i(t) = r_{i-1}(t) - l_i(t) \).

6. Set \( i = i + 1 \) and return to step (2) until the residue \( r_i(t) \) is a monotonic function or its extreme does not exceed three extreme points. Then, the original signal is decomposed in terms of IMFs and the residue, i.e.

\[
x(t) = \sum_{j=1}^{n} l_j(t) + r_n(t)
\]  

IMFs are simple oscillatory functions with varying amplitude and frequency, and hence have generally two following properties:

1. The number of extremes \( N_x \) and the number of zero-crossings \( N_z \) must either be equal or differ at more by one, i.e. \( N_x - 1 \leq N_z \leq N_x + 1 \);
2. For any point \( t_i \) in time series, the mean value of the upper and lower envelopes are zero, i.e. \( [f_{\text{max}}(t_i) + f_{\text{min}}(t_i)]/2 = 0 \), \( t_i \in [t_{a}, t_{b}] \).

EMD algorithm does not require a priori knowledge about the original time series and the IMFs can be directly obtained based on the local characteristic of the time series. This data-driven mechanism guarantees its superiority compared with other traditional decomposition methods, e.g. wavelet decomposition and Fourier transformation. However, the existence of intermittent signals will induce a serious mode mixing problem, which means that a single IMF component contains either signals of different scales or a signal with a similar scale existing in different IMF components. If this problem occurs, then the EMD cannot reflect the real characteristics of the original time series, which reduces the forecast accuracy. Fortunately, this problem is settled by the EEMD algorithm which adds different white noises to the target time series for many trails.

2.2. EEMD

EEMD overcomes the mode mixing issue of EMD by using the noise-added data analysis method. The added white noise is aimed to homogenize the distribution of extrema in time-frequency space, which enables the natural filter bank of EMD to extract the intrinsic local oscillations adaptively (Wu & Huang, 2009). The EEMD algorithm can be described briefly as follows.

1. For a given original signal \( x(t) \) \((t > 0)\), add random white noise \( n(t) \) to the signal \( x(t) \), namely \( S(t) = x(t) + n(t) \).
x(t) + an(t), where a is the amplitude of the added white noise.

(2) Decompose the noise-added signal S(t) using EMD to obtain an IMF series c_j(t), where c_j(t) means the jth IMF of the ith noise-added signal.

(3) Repeat steps (1) and (2) until the number of white noises added is greater than or equal to the number of trials, i.e. \( i \leq Ne \).

(4) Compute the average of the sum of all IMF components to obtain the ensemble IMF, namely,

\[
l_j(t) = \frac{1}{Ne} \sum_{i=1}^{Ne} c_j(t)
\]

\[
x(t) = \sum_{j=1}^{n} l_j(t)
\]

The amplitude of white noise a in Step 1 and the ensemble number Ne in Step 3 are two important parameters for the performance of EEMD method and have to be predefined. If the parameters are set to be too small, the mode mixing problem cannot be effectively settled. In the opposite situation, if the parameters are set to be too large, the calculation cost will increase and some pseudo-components will be induced. In order to make EEMD effective, it is suggested that the number of trails is a few hundred and the amplitude can be set to be a fraction of the standard deviation of the original data. Wu and Huang (2009) found that the best result could be obtained with 100 trails and the amplitude were 0.2 of the standard deviation based on their simulation experiments. After that, scholars usually followed these empirical parameter setting in their studies (Li et al., 2018; Wang, Chau, & Qiu, 2015). To the best of our knowledge, there is no such research that focuses on how to select these two parameters more scientifically and self-adaptively. Consequently, the GA algorithm is introduced to address the parameter selecting issue in EEMD, owing to its outstanding optimization capability (Song, Zhang, Song, & Chen, 2019; Zou, Li, Kong, Ouyang, & Li, 2019).

3. GAEEMD

Coupling GA with EEMD algorithm, a novel signal decomposition method is proposed in our study, namely GAEEMD. First, a random white noise matrix is constructed for the input signal, so that each selected white noise is fixed to ensure the execution of optimization process of GA. Second, the inverse of the orthogonality index is specified as the fitness function so as to ensure the optimization efficiency of GA. Finally, considering the algorithm has difficulties to escape from local minimums owing to the large parameter searching space, Hamming distance is used to formulate the difference selection operator, in order to guarantee the diversity of the population. The overall schematic diagram of the GAEEMD algorithm is shown in Figure 1.

The algorithm is organized as follows:

(1) For a given original signal \( x(t) (t > 0) \), where \( t = 1, 2, \ldots, T \) and \( T \) is the length of the signal length.

(2) Construct a random white noise matrix \( \{n_1(t), n_2(t), \ldots, n_{Ne_{max}}(t)\} \), where \( Ne_{max} \) is the upper limit of the number of trials.

(3) Set the population \( P_i (i = 1, 2, \ldots, N_p; j = 1, 2, \ldots, M) \) and \( i = 1 \), and construct the initial population \( P_1, P_2, \ldots, P_M \), where \( N_p \) is the maximum number of iterations, \( M \) is the number of individuals in the population.

(4) Encoding the individual of the population separately, and obtain the amplitude of the added white noise \( a \) and the number of trials \( Ne \) for adding white noise. \( a \) and \( Ne \) are represented in binary coding manner, \( a \in [10^{-3}, 1] \) and \( Ne \in [1, 100] \).

(5) Add white noise series \( \{n_k(t)\} (k = 1, 2, \ldots, Ne) \) to the targeted signal \( x(t) \), i.e. \( S_k(t) = x(t) + an_k(t) \).
Table 1. Parameters of GAEEMD for the analog signals $x_i(t)$ and $y_i(t)$.

| Signal | The amplitude of the noise | The number of trials |
|--------|----------------------------|---------------------|
| $x_1(t)$ | 0.198 | 7 |
| $x_2(t)$ | 0.157 | 3 |
| $x_3(t)$ | 0.166 | 4 |
| $x_4(t)$ | 0.15 | 4 |
| $x_5(t)$ | 0.175 | 9 |
| $y_1(t)$ | 0.001 | 4 |
| $y_2(t)$ | 0.05 | 2 |
| $y_3(t)$ | 0.047 | 2 |
| $y_4(t)$ | 0.056 | 24 |
| $y_5(t)$ | 0.032 | 2 |

Table 2. IOs of EEMD, GAEEMD for the analog signals $x_i(t)$ and $y_i(t)$.

| Signal | EMD | EEMD | GAEEMD |
|--------|-----|------|--------|
| $x_1(t)$ | 0.047 | 0.095 | 0.002 |
| $x_2(t)$ | 1.406 | 0.125 | 0.007 |
| $x_3(t)$ | 0.426 | 0.136 | 0.037 |
| $x_4(t)$ | 5.445 | 0.163 | 0.035 |
| $x_5(t)$ | 0.500 | 0.175 | 0.026 |
| $y_1(t)$ | 0.111 | 0.149 | 0.005 |
| $y_2(t)$ | 0.028 | 0.180 | 0.001 |
| $y_3(t)$ | 0.025 | 0.241 | 0.008 |
| $y_4(t)$ | 0.017 | 0.221 | 0.003 |

Note: '-' means the method cannot decompose the signal.

(6) Decompose the noise-added signal series $\{S_k(t)\}$ using EMD to obtain the IMF series $\{c_{ku}(t)\}$, where $c_{ku}(t)$ means the $u$th IMF component of the $k$th noise-added signal.

(7) Compute the average of the sum of all IMF components to obtain the ensemble IMF, i.e.

$$I_u(t) = \frac{1}{Ne} \sum_{k=1}^{Ne} c_{ku}(t)$$

(8) Compute the Orthogonality Index of the all ensemble IMFs $\{I_u(t)\}$, and the fitness function of the population is defined as $1/\text{IO}$, which are presented in Section 3.1.

(9) Calculate the next generation of the population using genetic operations of selection, crossover and mutation. Difference selection operator is defined as the selection operator and discussed in Section 3.2. One-point crossover and binary mutation method are used for the formulation of the crossover operator and the mutation operator, respectively.

(10) Set $i = i + 1$, return to step (4) until $i \geq N$. If $i = N$, the IMF components with the optimal fitness function are reported at the end of the algorithm.

### 3.1. Fitness function

Considering that the orthogonality index (IO) can measure the orthogonality of the EMD numerically, it is specified to formulate the fitness function of GA. The value of IO is zero when the IMF components are orthogonal to each other. A smaller IO indicates a better decomposition result which has a favourable frequency resolution (Huang et al., 1998). For the original signal $x(t)$ and its decomposed IMF components $I_i(t)(i = 1, 2, \ldots, n)$, $x(t)$ can be expressed as

$$x(t) = \sum_{j=1}^{n} I_j(t)$$

Square both sides of Equation (5) can be obtained as follow:

$$x^2(t) = \sum_{j=1}^{n} I_j^2(t) + 2 \sum_{i=1}^{n} \sum_{j=1}^{n} I_i(t)I_j(t)$$

If the IMF components are orthogonal with each other, the cross term on the right side of Equation (6) should be zero, so the IO can be defined as

$$\text{IO} = \sum_{t=0}^{T} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} I_i(t)I_j(t)/x^2(t) \right)$$

where $T$ is the length of the original signal $x(t)$. Hence, the fitness function can be defined as the inverse of IO and shown as follow:

$$\text{Fitness}(x(t), I(t)) = \frac{1}{\sum_{t=0}^{T} \left( \sum_{i=1}^{n} \sum_{j=1}^{n} I_i(t)I_j(t)/x^2(t) \right)}$$

### 3.2. Difference selection operator

Traditional selection operators are determined by the roulette selection and the tournament selection (Song et al., 2019). Owing to the randomness property of these methods, the results are usually unstable and the individual with high fitness will be missed sometimes (Zou et al., 2019). Some studies demonstrated that the
evolution process of the population was based on the diversity of the individual, and the decrease of individual diversity often led to the convergence to local optima. Therefore, we propose a difference selection operator whose main principle is to compare individuals with small differences and eliminate individuals with poor fitness, so as to maintain the diversity of the individual. There are many ways for calculating the differences between the individuals in a population, such as Hamming distance, Euclidian distance, Cosine Similarity and Jaccard Similarity. Since the gene coding method used in this paper is binary code, it does not require additional numerical transformation to meet the calculation conditions of Hamming distance, while other methods require additional gene decoding or some form of conversion, so Hamming distance was used for the difference selection. The difference selection operator can be described as follows.

(1) For the binary gene strings \( \{x_{ij}\} \) of the individual, where \( i = 1, 2, \ldots, N \) and \( j = 1, 2, \ldots, M \). \( i \) and \( j \) are the number of individual and the position of binary gene strings, respectively.

(2) Calculate the Hamming distance between different individuals, excluding the individuals themselves and the same genes, i.e.

\[
d_{ij} = \left\| x_i' - x_j' \right\| = \sqrt{\sum_{k=1}^{M} (x_{ik}' - x_{jk}')^2}
\]

where \( i, j \in \{1, 2, \ldots, N\}, i \neq j \).

(3) For the \( i \) individual, select the top \( s \) individuals with the smallest difference from the population for pairing. The individual with the highest fitness is then selected according to the difference selection probability.
4. Experiments

In order to verify the robustness and effectiveness of the GAEEMD algorithm, the experiment is implemented by using analog signals. Considering that data with the intermittent signal is prone to induce mode mixing problem and affect the decomposition performance, two groups of intermittent signals with different amplitudes and sinusoidal signals of different frequencies are generated to demonstrate the ability of solving that problem between EMD, EEMD, and GAEEMD. The detailed description of the experimental signal and the specific parameters settings are described in Section 4.1. The evaluation criteria in the experiment are given in Sections 4.2, and the experimental results are analyzed in Sections 4.3 and 4.4, respectively. In addition, the performance comparison between the difference selection operator and the roulette selection operator which uses one of the above analog signals, and the analysis of the comparison results is described in Section 4.5. The experiment is implemented with MatlabR2016a and the system environment is a desktop computer with an operating system of Window10 and a processor of Intel(R) Xeon(R) E5-1620 v3 with a frequency of 3.50 GHz and a memory of 32 GB.

4.1. Data description and parameter settings

Given \( g(t) \) is the target signal needs to be decomposed with the expression of \( g(t) = 2\sin(30\pi t + \pi) \). The intermittent noise of different amplitudes and sinusoidal noise of different frequencies are then added to the target signal, respectively. The formulas are shown as follows:

\[
\begin{align*}
  x_i(t) &= g(t) + n_i(t) \\
  y_i(t) &= g(t) + u_i(t)
\end{align*}
\]

where \( i \) indicates the experimental times and \( i \in [1, 5] \). \( n_i(t) \) is an intermittent signal contains three white noises with the expression \( n_i(t) = 0.2i \times n(t) \). \( u_i(t) \) is a sinusoidal signal with the amplitude of one and expressed as \( u_i(t) = \sin((i+5) \times 10\pi t) \). As for the parameters selected in EEMD algorithm, the number of trials is set to be 100 and the amplitude of the added white noise is set to be one-fifth of the standard deviation of the original signal, according to Wu and Huang (2009). One-fifth of the standard deviation of the original signals \( x_i(t) \) are 0.2836, 0.284, 0.2898, 0.2954 and 0.3083, respectively, while those of the original signals \( y_i(t) \) are always 0.3162. As for the parameters in GAEEMD algorithm, the population size is set to be 100, the probability of difference selection is 0.6,
the number of difference selection pairing is 4. The probability of crossover and the probability of mutation are set to 0.6 and 0.001, respectively. The maximum number of iterations is 100. The sampling frequency of the signals is 250 Hz and the sampling period is 2 seconds. Taking $i$ for 3 as an example, the simulation signals are shown in Figure 2.

4.2. Evaluation criteria

(1) Correlation coefficient

The correlation coefficient (CC) is a measurement of the similarity between the original signal and the decomposed signal (Jie, Wen-Wen, Hao, & Zhen-Zhen, 2013). For the two series $x(t)$ and $y(t)$, the calculation of CC is shown as:

$$CC = \frac{\sum_{t=1}^{T} (x(t) - \bar{x})(y(t) - \bar{y})}{\sqrt{\sum_{t=1}^{T} (x(t) - \bar{x})^2} \sqrt{\sum_{t=1}^{T} (y(t) - \bar{y})^2}}$$  (12)

where $T$ is the length of the series, $\bar{x}$ and $\bar{y}$ are the mean of series $x(t)$ and $y(t)$, respectively. The correlation coefficient matrix is composed of the correlation coefficients of each pair of decomposition components and actual components. An optimal noises separation effect can be defined as there is only one correlation coefficient in each column or each column close to one, and the others are close to zero. A larger CC means a more accurate separation.

(2) Root mean square error

The root mean square error (RMSE) measures the errors between the original signal and the decomposed signal (Wu & Huang, 2009). A smaller RMSE denotes a more accurate decomposition results. RMSE of the two series $x(t)$ and $y(t)$ is given by:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x(t) - y(t))^2}$$  (13)

where $T$ is the length of the series, $\bar{x}$ and $\bar{y}$ are the mean of series $x(t)$ and $y(t)$, respectively.

4.3. Results analysis

The amplitudes of the added white noises and the number of trials in GAEEMD are shown in Table 1. It is shown
that the settings of the two parameters do not belong to the range of Huang’s experimental suggestion, which indicates that the experimental suggestion is not comprehensive enough.

The IOs of EMD, EEMD, and GAEEMD are listed in Table 2. It is obvious that the performance of EEMD deteriorates as the amplitude and the frequency of the added white noises increases. As for some signals (i.e. $x_1(t)$, $y_2(t)$, $y_3(t)$, $y_4(t)$ and $y_5(t)$), the performance of EEMD are inferior to EMD. The possible reason is that the IMFs generated by EEMD was polluted by the inappropriate amplitudes and number of trials of the added noises, thus yielding the pseudo-components which reduced the performance. EMD cannot decompose the signal $y_1(t)$ with a sinusoidal noise frequency of 30, which indicates this algorithm is unfeasible for the mixed signal with similar frequency. In terms of intermittent signals $x(t)$, the IOs of GAEEMD are much smaller than that of EMD and EEMD, which represent a better decomposition result. As for sinusoidal signals $y(t)$, the performance of GAEEMD is again favourable between all the counterparts, which continually confirms the superiority of the proposed algorithm.

The CCs and the RMSEs obtained by EMD, EEMD, and GAEEMD in terms of $x(t)$ and $y(t)$ are shown in Figure 3 and Figure 4, respectively. The blue diamond represents for EMD, the green five-pointed star indicates EEMD, and the red dot denotes GAEEMD. As for intermittent signals $x(t)$, the CCs of all the three algorithms decreases as the amplitude of the added white noises increase. GAEEMD and EEMD have equivalent performance while EMD is much worse than the others. As for sinusoidal signals $y(t)$, the CCs are relatively stable as the noise frequency increases. In general, the IMF components decomposed by GAEEMD has the average highest correlation with the original signal, which proves its effectiveness. The results derived from Figure 4 are similar to that of Figure 3. For both intermittent and sinusoidal signals, the IMF components generated by GAEEMD have the average smallest RMSE compared with EMD and EEMD.

Based on the above analysis, three conclusions can be drawn. First, the results derived from the CC and RMSE index are generally consistent with the results obtained from the IO index, which further confirms the effectiveness of the fitness function formulated with IO. Second,
mixed signals with intermittent noise as well as similar frequency have negative effects on EMD. Third, the performance of GAEEMD has a considerable superiority compared with EMD and EEMD, in terms of all the evaluation criteria, i.e. IO, CC, and RMSE. These results strongly support the effectiveness and robustness of GAEEMD.

4.4. Spectrum analysis

Due to the space limitation, the mixed signal $y_3(t)$ is specified as an example here. $y_3(t)$ is decomposed by EMD, EEMD, and GAEEMD and their results are shown in Figure 5, respectively. $IMF_i$ represents the $i$th IMF component and $r_n$ represents the residual component. As can be seen from Figure 5(a,b), the results of both EMD and EEMD have mode mixing problem, producing the pseudo-components with no physical meanings, while GAEEMD in Figure 5(c) successfully addressed the problem. It can be seen that the decomposed component $r_n$ and $IMF_1$ correspond to the sinusoidal signal $u_3(t)$, respectively.

In order to understand the decomposition results more thoroughly, the mixed signals $y_3(t)$ and the decomposition results yielded by GAEEMD, EMD, and EEMD are transformed from the time domain to the frequency domain with Fast Fourier Transform (FFT) method (Zhang, Wen, & Song, 2018). The corresponding spectrograms are shown in Figures 6–8, respectively. The mixed signal $y_3(t)$ contains two signals with different frequencies (i.e. 15 Hz and 40 Hz) are shown in Figure 6(a). Figures 7 and 8 reveal that there are some kinds of mode mixing problems within the EMD and EEMD method. Although EMD can isolate the sinusoidal signal $u_3(t)$ with 40 Hz frequency, some noises still exist in the $IMF_1$ component (see Figure 7(a)). Moreover, the $IMF_3$ and $r_n$ component can be regarded as pseudo-components that do not exist in the original signal (see Figure 7(c,d)). Similar issues also occur in EEMD (see Figure 8(b,d)). As for the proposed method, GAEEMD can successfully separate the signals of 40 Hz and 15 Hz frequency from the mixed signal $y_3(t)$ (see Figure 6(b,c)) without any pseudo-components. With the help of spectrograms, GAEEMD is intuitively proven to be highly effective in solving the mode mixing problem, with the comparison of EMD and EEMD.

The above results show that the way of adding white noise can improve the frequency resolution of empirical mode decomposition method, and the EEMD based on artificial experience method is not stable in the mixed signal with large frequency difference, and even gets worse than the decomposition result of EMD. As a result,
the decomposition performance of GAEEMD is more stable, and more accurate components can be derived by GAEEMD in the mixed signal of sinusoidal signals with different frequencies.

4.5. Comparison with the original GA

Here, the optimization efficiency of the proposed GAEEMD method is verified through the comparison with the original GA. The number of iterations and the calculation cost which achieves the target error and the corresponding run time, as well as the final error cost are used as indicators. The algorithm runs 20 times, the target error is set to be 0.01 and the final error indicator indicates the error obtained when the algorithm iterates 100 times, the results obtained by GA and GAEEMD are shown in Table 3.

As can be seen, in most cases, the number of iterations to achieve the target error of GAEEMD is far less than that of original GA (except for the 3rd, 7th, and 16th test). Furthermore, as for the computational cost, the run time of GAEEMD is far less than that of original GA (except for the 2nd, 3rd, 7th, and 16th test) and the average run time of GAEEMD (36.75 minutes) is far less than that of original GA (107.44 minutes), which indicates the convergence efficiency is remarkably improved. The possible reason can be ascribed to the effective difference selection operator of GAEEMD (designed by the Hamming distance), while in the original GA, it is determined by the roulette selection. In terms of the final error indicator, the original GA cannot reach the target error for 11 times in the total 20 tests, while GAEEMD generates a favourable result in each test. It is shown that the problem of local minima of GA has been successfully addressed by GAEEMD.

5. Conclusions

An improved ensemble empirical mode decomposition method based on GA is proposed in this paper. The proposed GAEEMD algorithm solves the mode mixing problem and improves the self-adaptiveness of the decomposition method. The simulation experiment reveals that (1) Using genetic algorithm to optimize the amplitude of added white noise and the number of trails, the pseudo-components caused by the improper parameters selection can be successfully addressed; (2) The orthogonality index which is used to construct the fitness function can objectively reflect the accuracy of signal decomposition; (3) The difference selection operator designed by Hamming distance maintains the diversity of the population, thus effectively overcomes the problem of local minima.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China (NSFC, 61702197), by the Natural Science Foundation of Guangdong Province (2017A030310261, 2016A030313402), by the Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education ( JYB201708).

References

Dong, J., Dai, W., Tang, L., & Yu, L. (2019). Why do EMD-based methods improve prediction? A multiscale complexity perspective. *Journal of Forecasting*. doi:10.1002/for.2593.
Hikmat, H., Maha, A., & Enas, A. (2018). Automated identification for autism severity level: EEG analysis using empirical mode decomposition and second order difference plot. *Behavioural Brain Research*, 362, 240–248.

Hoseinzadeh, M. S., Khadem, S. E., & Sadooghi, M. S. (2018). Quantitative diagnosis for bearing faults by improving ensemble empirical mode decomposition. *ISA Transactions*, 83, 261–275.

Huang, N. E., Shen, Z., & Long, S. R. (1999). A new view of nonlinear water waves: The Hilbert spectrum. *Annual Review of Fluid Mechanics*, 31, 417–457.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., & H. H. Liu (1998). The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1971), 903–995.

Huang, N. E., & Wu, Z. H. (2008). A review on Hilbert-Huang transform: Method and its applications to geophysical studies. *Reviews of Geophysics*, 46(2), 1–23.

Jie, Y., Wen-Wen, Y. U., Hao, T., & Zhen-Zhen, G. (2013). Underdetermined blind source separation method based on independent component analysis. *Journal of Vibration and Shock*, 32(7), 30–33.

Kumar, S., Panigrahy, D., & Sahu, P. K. (2018). Denoising of electrocardiogram (ECG) signal by using empirical mode decomposition (EMD) with non-local mean (NLM) technique. *Biocybernetics and Biomedical Engineering*, 38(2), 297–312.

Li, C., Tao, Y., Ao, W. G., Yang, S., & Bai, Y. (2018). Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition. *Energy*, 165, 1220–1227.

Mohammad, S.H., Siamak, E.K., & Mohammad, S.S. (2018). Quantitative diagnosis for bearing faults by improving ensemble empirical mode decomposition. *ISA Transactions*, 83, 261–275.

Qin, Y., Qiao, L. H., Wang, Q. F., Ren, X. Z., & Zhu, C. H. (2018). Bidimensional empirical mode decomposition method for image processing in sensing system. *Computers & Electrical Engineering*, 68, 215–224.

Rajib, S. (2016). Mahadeva Prasanna, A better decomposition of speech obtained using modified empirical mode decomposition. *Digital Signal Processing*, 58, 26–39.

Song, Y. J., Zhang, Z. S., Song, B. Y., & Chen, Y. W. (2019). Improved genetic algorithm with local search for satellite range scheduling system and its application in environmental monitoring. *Sustainable Computing: Informatics and Systems*, 21, 19–27.

Wang, W. C., Chau, K. W., & Qiu, L. (2015). Improving forecasting accuracy of medium and long-term runoff using artificial neural network based on EEMD decomposition. *Energy*, 165, 1220–1227.

Wei, Y. J., Wei, Q., Wang, S. Y., & Lai, K. K. (2018). A hybrid approach for studying the lead-lag relationships between China’s onshore and offshore exchange rates considering the impact of extreme events. *Journal of Systems Science and Complexity*, 31(3), 734–749.

Wu, Z., & Huang, N. (2009). Ensemble empirical mode decomposition: A noise assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1), 1–41.

Xiang, Y., Gou, L., He, L. H., Xia, S. L., & Wang, W. Y. (2018). A SVR-ANN combined model based on ensemble EMD for rainfall prediction. *Applied Soft Computing*, 73, 874–883.

Yu, S. Y., Li, B. P., Zhang, Q., Liu, C. C., & Max, Q.-H. M. (2015). A novel license plate location method based on wavelet transform and EMD analysis. *Pattern Recognition*, 48, 114–125.

Yuan, J., Ji, F., Gao, Y., Zhu, J., Wei, C. J., & Zhou, Y. (2018). Integrated ensemble noise-reconstructed empirical mode decomposition for mechanical fault detection. *Mechanical Systems and Signal Processing*, 104, 323–346.

Zhang, W. K., Wen, D. S., & Song, Z. X. (2018). Spectrum reconstruction in interference spectrometer based on sparse Fourier transform. *Optik*, 154, 157–164.

Zheng, J., Cheng, J. S., & Yang, Y. (2014). Partly ensemble empirical mode decomposition: An improved noise-assisted method for eliminating mode mixing. *Signal Processing*, 96, 362–374.

Zou, D. X., Li, S., Kong, X. Y., Ouyang, H. B., & Li, Z. Y. (2019). Solving the combined heat and power economic dispatch problems by an improved genetic algorithm and a new constraint handling strategy. *Applied Energy*, 237, 646–670.