

ORIGINAL ARTICLE

Open Access



# Incipient Gear Fault Detection Using Adaptive Impulsive Wavelet Filter Based on Spectral Negentropy

Mang Gao<sup>1</sup>, Gang Yu<sup>1\*</sup>  and Changning Li<sup>2</sup>

## Abstract

Adaptive wavelet filtering is a very important fault feature extraction method in the domain of condition monitoring; however, owing to the time-consuming computation and difficulty of choosing criteria used to represent incipient faults, the engineering applications are limited to some extent. To detect incipient gear faults at a fast speed, a new criterion is proposed to optimize the parameters of the modified impulsive wavelet for constructing an optimal wavelet filter to detect impulsive gear faults. First, a new criterion based on spectral negentropy is proposed. Then, a novel search strategy is applied to optimize the parameters of the impulsive wavelet based on the new criterion. Finally, envelope spectral analysis is applied to determine the incipient fault characteristic frequency. Both the simulation and experimental validation demonstrated the superiority of the proposed approach.

**Keywords:** Incipient fault diagnosis, Negentropy, Spectral kurtosis, Gear, Adaptive wavelet

## 1 Introduction

Gear and bearing faults are very common in rotary machine systems. If these faults cannot be detected in their incipient stage, catastrophic accidents and economic losses may occur in industrial plants. Numerous researchers have explored these methods without disassembling the machines to find the faults in these components, and vibration-based methods are the most widely used.

Conventional methods such as time-frequency analysis, empirical mode decomposition (EMD) [1], matching pursuit and time-scale method [2], and wavelet analysis [3] are the most common; data-driven methods such as deep learning [4] have also gained recent popularity. The spectral kurtosis (SK) method is one of the most widely used methods. Numerous researchers have attempted to apply SK to detect fault features in machinery systems.

Antoni et al. [5] explicated the definition of “kurtogram,” and developed a fast algorithm known as fast kurtogram (FK) which derives from SK. Since then, it has become a very practical fault transient extraction methodology in the fault diagnosis domain owing to its fast computation speed. Unfortunately, the application of the kurtogram is constrained in extracting transient characteristics from noisy signals under certain conditions [6], and the accuracy of the kurtogram is limited because it is based on the short-time Fourier transform (STFT) or FIR filter in extracting transient features [7]. Another limitation of the kurtogram is that it is difficult to discern whether a series of transients is repetitive [8]. If the transients are so frequent that they overlap each other, kurtosis vanishes [9]; if the transients are sufficiently spaced apart that only one can be detected, it results in the maximum value of kurtosis. This is the main reason why the kurtogram can be sensitive to impulsive noise [10]. To overcome the shortcomings of the kurtogram, Antoni [8] proposed a new approach – the infogram, which uses spectral negentropy (SN) to substitute SK in the kurtogram, which aims at capturing the signature of

\*Correspondence: 969083604@qq.com

<sup>1</sup> School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China  
Full list of author information is available at the end of the article

repetitive transients in both time and frequency domains. The experimental results show superiority compared to conventional kurtogram methods. Some researchers have attempted to extend the application of infograms, such as a combination with a Pareto-based Bayesian approach [11] and a multiscale fractional order entropy infogram [12]. Xu et al. [13] combined the fast empirical wavelet transform with spectral negentropy to filter a new measured signal, which showed better reliability for the diagnosis of compound faults. Hu et al. [14] proposed an adaptive spectral kurtosis to eliminate the fixed decomposition scheme of FK, and the results showed that the proposed method can effectively extract bearing fault features.

Wavelet analysis is an approach in the field of machinery fault diagnosis. The selection of the decomposing level and scale is empirical in traditional wavelet decomposition methods [15]. Hence, some researchers have attempted to identify faults automatically using wavelets. Lin et al. [16] first proposed a method based on an adaptive Morlet wavelet filter to detect the faults existing in a gearbox, as the Morlet wavelet is similar to an impulse component in a fault signal. In the above works, kurtosis and entropy are applied as criteria to choose the optimal parameters of the Morlet wavelet filter. Since then, many modified and integrated algorithms based on the Morlet wavelet have also been developed [17]. Morsy et al. [18] combined the optimal Morlet wavelet filter and envelope detection (ED) to denoise the background noise associated with interferential vibrations. Apart from the Morlet wavelet, other wavelets are also utilized as adaptive filters, such as the Laplace wavelet [19] and impulsive wavelet [20]. In Ref. [21], a thorough comparative analysis was carried out among different wavelet bases; thus, optimal wavelet bases are recommended for different fault types. Nevertheless, in these methods, the computation time is striking to some degree [15].

A sparse representation of transients based on wavelet bases to find fault characteristics in a gearbox was proposed by Fan et al. [22]; however, the calculation speed of the proposed method is limited because of the larger interval range and smaller step size of the parameter subset. Moreover, incipient faults are seldom considered; thus, a better criterion should be developed to represent incipient faults. Recently, a new family of models based on impulsive wavelets was proposed as a method for sparse representation by Qin [23]. However, the parameters in the impulsive wavelet were not optimized [23].

Matching pursuit, which is related to the aforementioned sparse representation approaches and adaptive wavelet filtering, has attracted a lot of researchers to explore its effect on fault feature extraction. However, an

over-complete dictionary makes the algorithm time-consuming on occasion [24].

To counteract the above issues related to adaptive wavelets applied to incipient fault detection, a new criterion based on spectral negentropy is proposed to represent the incipient faults, and an impulsive wavelet is modified to better match the fault signal. Finally, a new optimization strategy is developed to reduce the computation time for finding the optimal parameters of the impulsive wavelet filter.

The main contributions of this study are summarized as follows:

- For the first time, a new criterion based on spectral negentropy is applied to optimize the impulsive wavelet parameters.
- A comparison analysis between SN and kurtosis as criteria for the adaptive wavelet is carried out.
- A fast optimizing strategy that can achieve better results with a fast computation speed is proposed.
- The proposed method is verified by both simulation and experimental signals.

This paper is divided into five sections. Section 2 introduces the principles of the related theories and the methodology of the proposed approach. Then, in Section 3, a simulation signal is constructed to verify the proposed algorithm, which preliminarily proves its superiority. In Section 4, experimental signals are applied to further validate the efficiency and effectiveness of the proposed algorithm by comparing it with conventional methods. Finally, in Section 5, a concise conclusion is drawn.

## 2 Principle of Proposed Algorithm

In this section, the basic theories related to the proposed method are introduced, and the specific methodology is presented.

### 2.1 Wavelet Principle and Modified Impulsive Wavelet

The wavelet transform can change a time-domain signal into a time-scale domain. The wavelet transform is determined by performing dilation and translation of the mother wavelet  $\psi(t)$  [25].

$$\psi_{a,b}(t) = |s|^{-1/2} \psi((t - \tau)/s) \quad s, \tau \in R; s \neq 0, \quad (1)$$

where  $s, \tau$  represent the scale factor and time location, respectively, and the  $|s|^{-1/2}$  factor is applied for energy conservation. The scale parameters  $s > 0$  and  $\tau \in R$  can be continuous or discrete.

The wavelet transform of a finite energy signal  $x(t)$  is the convolution of  $x(t)$  and a conjugated wavelet:

$$W(s, \tau) = |s|^{-1/2} \int_{-\infty}^{\infty} x(t)\psi^*((t - \tau)/s)dt, \tag{2}$$

where  $\psi^*(t)$  is the complex conjugation of  $\psi(t)$ .

When a wavelet resembles an impulsive fault masked in a vibration signal acquired from a rotating mechanical system, resonance occurs if it is applied to convolution. Wavelets, such as the Morlet wavelet and Laplace wavelet, are widely used as adaptive filter models aiming at the impulsive faults existing in machinery vibration signals owing to this characteristic. In Refs. [20, 23], the advantages of the modified impulsive wavelet were illustrated. To obtain better filtering results, a modified impulsive wavelet [20, 23] (only the single-side part is given) is defined as

$$\psi(t) = \exp[-f_b f_b(t - \tau)] \sin[2\pi f_c(t - \tau)], \tag{3}$$

where  $f_b$  denotes the damping ratio that controls the decay rate of the impulsive wavelet,  $f_c$  represents the central frequency, and  $\tau$  is the time parameter. By dilating with  $f_c, f_b$  and moving with  $\tau$ , a series of daughter wavelets can be obtained.

In contrast to the formula given in Ref. [23], the modified impulsive wavelet has specific significance. The impulsive wavelet in the frequency domain is a filtering window, and the central frequency  $f_c$  in Eq. (3) is equivalent to the center coordinate of the window peak value, as illustrated in Ref. [20]. While in Ref. [23], the peak value was different from the central frequency.

### 2.2 Proposed Methodology

Entropy is a statistic value applied from the informatic theory (statistical thermodynamics (Boltzmann's entropy) and signal processing (Shannon's entropy)), which can be used to represent the degree of chaos in a system. The entropy is defined as Eq. (4) [26].

$$H(p) = - \sum_{i=1}^n p_i \ln p_i, \tag{4}$$

where  $\sum_{i=1}^n p_i=1$ , and  $p_i$  denotes the probabilities.  $i$  is the state index.

SK and its related methods have been thoroughly investigated by scholars as good approaches to represent impulsive components [27]. SK is clarified as the energy-normalized fourth-order spectral cumulant of a conditionally nonstationary [28, 29].

$$SK(f) = S_{4Y}(f)/S_{2Y}^2(f) - 2, \tag{5}$$

where  $SK(f)$  represents the value of SK;  $S_{2Y}(f)$  denotes the time-averaged result of  $S_{2Y}(t, f)$ , and  $S_{2Y}(t, f)$  is an

instantaneous moment and measures the energy of the complex envelope.

Based on the infogram in Ref. [8], where SN is applied as a substitution of SK in the kurtogram, it is hereby imported as a criterion to optimize the parameters of the optimal impulsive wavelet. SN is derived from the spectral entropy. According to Ref. [8], the spectral entropy shows an opposite behavior to SK; thus, to make it consistent with SK, SN is applied. The spectral entropy can be seen as a version of SK weighted by  $-\ln(SK)$ . In Ref. [8], SN is defined as follows.

$$I(f) = -H = (S_{4Y}(f)/S_{4Y}^2(f))\ln(S_{4Y}(f)/S_{4Y}^2(f)). \tag{6}$$

The purpose of the proposed method is to use SN as a criterion to evaluate the similarity between the impulsive wavelet and the transients masked in noise. In other words, the maximum value required in this method is similar to that in Ref. [16]; the maximum kurtosis is the target value. Hence  $S_{4Y}K(f)/S_{4Y}^2(f)$  in Eq. (6) can be obtained by  $SK(f)$  in Eq. (5). As a result, Eq. (6) can be simplified as

$$I(f) = -H = SK(f)\ln(SK(f)), \tag{7}$$

where  $SK(f)$  represents the spectral kurtosis of the wavelet coefficients, which originates from the convolution between the fault signal and the daughter impulsive wavelet.

The adaptive wavelet method is an approach that requires a criterion to measure the similarity between the base wavelet and the true signals collected from the mechanical system. In a fault signal, if the wavelet is similar to a transient, resonance will occur. If the criterion is more sensitive to similarity, the results are better. To further illustrate the advantage of using SN as a criterion to detect transients masked in vibration signals, a mathematical analysis is applied. According to Ref. [8], the functional form of SN can be defined as

$$N(x) = x^2 \ln(x^2), \tag{8}$$

where  $N(x)$  represents the value of SN, and  $x^2$  represents the independent variable that can denote the value of  $SK(f)$ .

Similarly, the functional forms of SK can be adapted as Eq. (9) [8]:

$$K(x) = x^2 - 2, \tag{9}$$

where  $K(x)$  represents the value of SK.  $x^2$  is same as in Eq. (8).

According to Eqs. (8) and (9), the limit of  $N(x)/K(x)$  can be obtained as follows:

$$\lim_{x^2 \rightarrow 2} \frac{N(x)}{K(x)} = \lim_{x^2 \rightarrow 2} \frac{x^2 \ln(x^2)}{x^2 - 2} = +\infty. \tag{10}$$

According to Ref. [27], as the value of  $SK(f)$ ,  $x^2$  normally ranges from 2 to 5 for a mechanical system. A larger value of  $SK(f)$  indicates that the fault degree is more severe. Therefore,  $x^2 \rightarrow 2$  means that the fault tends to be incipient, and the result of Eq. (10) indicates that the value of SN is larger than that of SK. In other words, SN is more sensitive in measuring the severity of a fault based on mathematical theory.

As clearly stated in Ref. [8], spectral negentropy can be used to measure information in energy fluctuations. Therefore, the energy increase in the resonant segments is suitable for SN to be evaluated, and SN can overcome the defect of spectral kurtosis [8]. In this study, this nature of spectral negentropy can be utilized to represent the degree of change in energy fluctuations. Specifically, a larger value of the spectral negentropy indicates a larger energy fluctuation. Normally, in stable conditions, the operation speed is almost constant, and the energy fluctuation is mainly influenced by the signal energy change, such as transients occurring in a vibration signal, which is usually viewed as a fault in the gearbox. The more intense the energy change, the more severe the fluctuating process in a system; thus, the greater the SN value. Therefore, the spectral negentropy maximum (SNM) criterion can be used to indicate the fault in a vibration signal.

### 2.3 Specific Algorithm Procedure

From the analysis before, SN was used as a criterion to evaluate the fault. To improve the efficiency of the algorithm, a fast computation strategy is developed to reduce the time consumption. The specific procedures of the SNM method are as follows and the flow chart is displayed in Figure 1:

- 1) First, obtain the impulsive fault signal using a data acquisition system.
- 2) Set the total number of iterations, which is equivalent to  $n$ . Initialize the search ranges of  $f_c, f_b$ . Set the initial search step size as a larger value preliminarily.
- 3) Obtain the coefficients of the impulsive wavelet at  $f_c, f_b$ .
- 4) Obtain the spectral kurtosis of the obtained impulsive wavelet coefficient.
- 5) Obtain the spectral negentropy of the spectral kurtosis at  $f_c$  and  $f_b$ .
- 6) Find the maximum spectral negentropy and the corresponding optimal  $f_c$  (represented by  $f_{co}$ ) and optimal  $f_b$  (represented by  $f_{bo}$ ).
- 7) Narrow the search range and the step size of  $f_c, f_b$  (search step size of  $f_c, f_b$  represented by  $f_c$

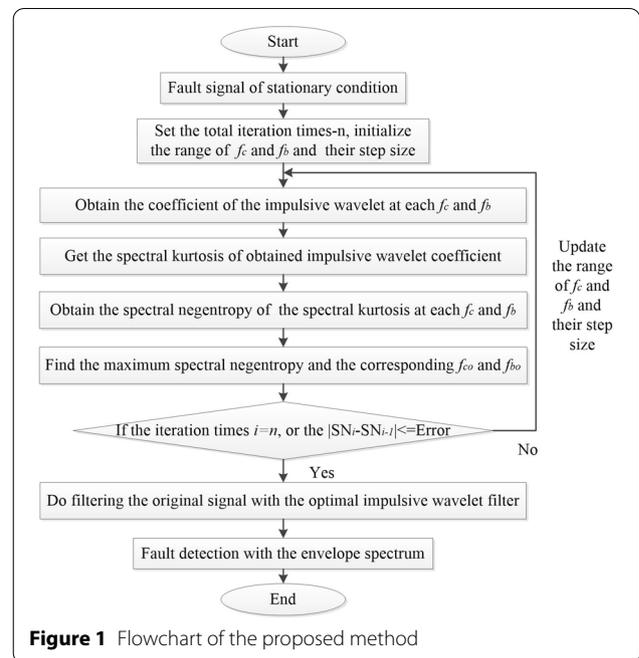


Figure 1 Flowchart of the proposed method

step and  $f_b$  step, respectively), and repeat steps 3)–7) until the resolution is small enough, which can be distinguished by the following equation:  $|SN_i - SN_{i-1}| \leq Error$ ;  $SN_i$  represents the value of the maximum spectral negentropy in the  $i$ -th iteration, and  $Error$  denotes the value set as the minimum error of SN between the neighboring iteration, which is desired in this process. Normally, the minimal  $f_c$  step equals 1 Hz, and the recommended minimal  $f_b$  step equals 0.005. This is because if  $f_c$  step and  $f_b$  step are selected as smaller values, the searching time will increase dramatically but the accuracy does not increase significantly according to the simulation test.

- 8) Filter the original signal with the optimal impulsive wavelet filter shaped by optimal  $f_c$  and  $f_b$ .
- 9) Obtain the square envelope spectrum of the filtered signal and find out the fault characteristic frequency, so as to achieve fault detection.

## 3 Simulation Analysis

### 3.1 Generation of Simulation Signals

In this section, to validate the efficiency of the proposed method, a simulation according to Refs. [23, 30] was constructed. It should be noted that the simulation signal is transient and normally exists in impulsive gear faults or bearing fault signals [31]. The simulation signal is defined as follows:

$$y(t) = x(t) + A_m n(t), \tag{11}$$

and

$$x(t) = (1 + 0.6 \cos[2\pi\omega_n(t - \tau_n)]) \sum_{m=0}^M h(t - kT_0), \tag{12}$$

where  $x(t)$  denotes the transient signal;  $n(t)$  represents a Gaussian noise with an amplitude value of 1;  $h(t)$  stands for an impact impulse function, and  $k$  and  $T_0$  are the number of transients and time period, respectively. Specifically,  $h(t)$  can be expressed as

$$h(t) = A_k \exp\left(-\zeta[2\pi\omega_0(t - \tau_0)]/\sqrt{1 - \zeta^2}\right) \cos[2\pi\omega_0(t - \tau_0)]. \tag{13}$$

When  $t \geq \tau_0 + kT_0$ , the above equation is satisfied; otherwise,  $h(t) = 0$ .  $\zeta$  is a damping ratio, and  $\omega_0$  is the transient feature frequency. The sampling frequency and time were selected as 12000 Hz and 1 s, respectively.  $A_k$  is set as a series of random numbers between 0.8 and 1.0.

All parameters in Eqs. (11) and (13) are listed in Table 1.

Then, the simulation of a transient signal can be generated according to the above equations and parameters, as shown in Figure 2.

Through calculation, the signal-to-noise ratio (SNR) of the simulation fault signal is equivalent to  $-13.4305$  dB.

### 3.2 Influence of Fault Degree on SN and Kurtosis

Kurtosis is a typical criterion for optimizing the wavelet parameters. A simulation experiment based on the MATLAB platform was conducted to investigate the effects of the fault degree on SN and kurtosis.

To explore the influence of the fault degree,  $A_k$  in Eq. (13) can be used to represent the degree of fault [32].  $A_m$  is set as 0.36 and  $A_k$  is set to increase from 0.1 to 6.5. Meanwhile, the other parameters remain the same as in Table 1 to generate 1500 series of simulation signals. The results are shown in Figure 3.

From Figure 3, it can be seen that the values of both kurtosis and SN show an increasing trend with the increase in fault degree, which means that both can indicate a fault feature. However, when the fault degree is incipient (the first 300 sets of simulation signals), the values of kurtosis remain unchanged, while the values of

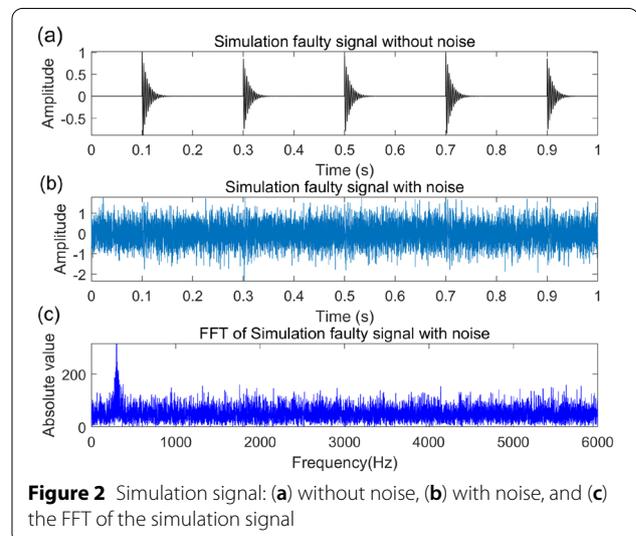
SN can still increase monotonously. In other words, for representing incipient faults, SN is superior to kurtosis.

### 3.3 Comparison Analysis

To make a fair comparison analysis, two typical related methods are utilized for comparison with the proposed method. Both qualitative and quantitative analyses were conducted to validate the superiority of the proposed method.

First, the proposed method is applied to process the simulation signal and determine the transient masked in the strong noise.

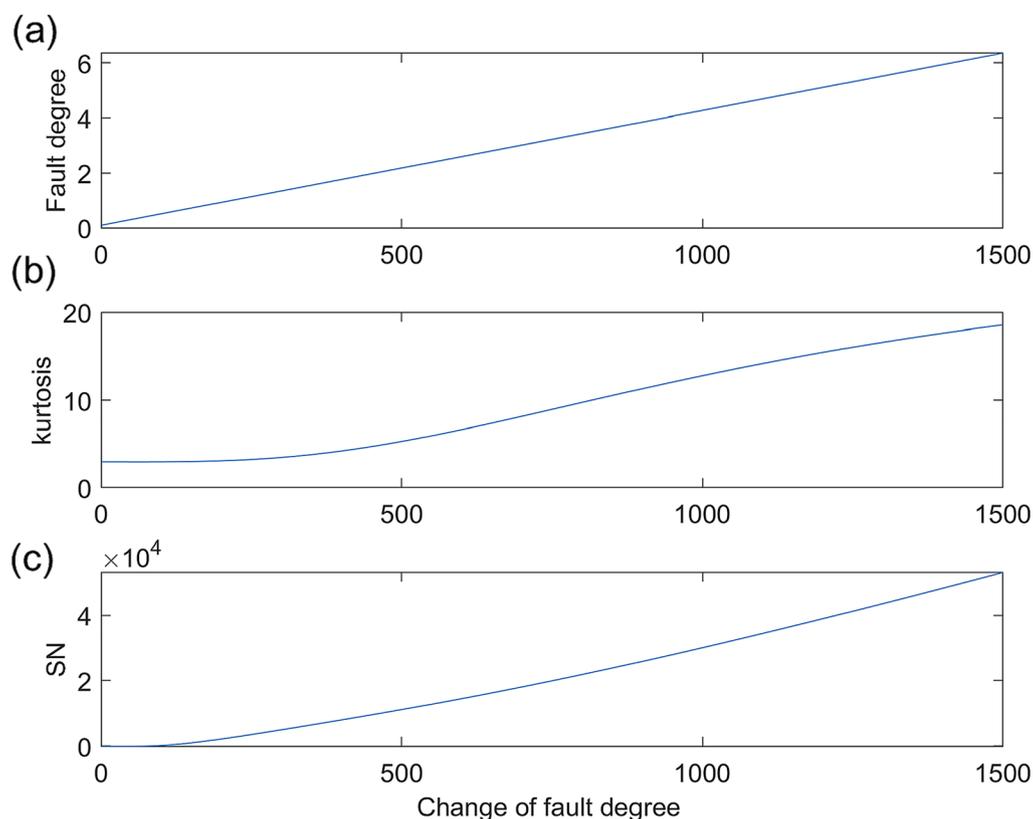
Because the selection of the search range and step length is highly related to the speed and accuracy of the method, many researchers have attempted to find a balance between them [33]. In the domain of the adaptive wavelet filter, the computation time is very long, as illustrated previously. Thus, it is very important to find a way to reduce the time consumption when searching for the optimal parameters of the wavelet filters. In this study, the new strategy used to reduce the computation time is called the gradually optimized method (described previously), which is explicated in the following section.



**Figure 2** Simulation signal: (a) without noise, (b) with noise, and (c) the FFT of the simulation signal

**Table 1** Parameters applied in the simulation signal

$\omega_n$ (Hz)	$\omega_0$ (Hz)	$\zeta$	$T_0$ (s)	$\tau_0$ (s)	$\tau_n$ (s)	$A_k$	$A_m$	$k$
0.85	300	0.05	0.2	0.1	0.1	0.8–1.0	1	0:4



**Figure 3** The change in kurtosis and SN values with the increase in fault degree: (a) the change in fault degree of 1500 simulation signals, (b) change in kurtosis value with increase in fault degree, and (c) change in SN value with increasing fault degree

According to the procedure introduced earlier, the initial central frequency  $f_c$  range is initially set from 250 to 350 Hz (As for the choice of the range of  $f_c$ , the initial value of  $f_c$  is near the frequency whose amplitude is maximum but meshing frequency, please refer to Refs. [21, 30] to obtain more details), and the search step was set as 4. As for the determination of the initial  $f_c$  step, because the bisection method is used to decrease the  $f_c$  step and the minimal  $f_c$  step is 1,  $f_c$  step is selected from 1, 2, 4, 8.... If it is selected as 8, the number of iterations is four. If the number of iterations is too large, there is a risk of missing the optimal peak without reducing the time saved so much. If it is set to 2, the iteration may end in the initial stage, and the computing time will increase as well. Therefore, the initial  $f_c$  step selection of 4 is appropriate, which is neither too large nor too small, and the number of iterations is 3 or 2. The range of  $f_b$  is selected as [0.1, 0.5] (In terms of choosing the range of  $f_b$ , it is usually near 0.3. For more details, please refer to Ref. [23]), and the initial search step is set as 0.05. After the optimal central frequency  $f_{co}$  and optimal damping ratio  $f_{bo}$  are found for the first time, the  $f_c$

range is updated to  $[f_{co} - 2f_c \text{ step}, f_{co} + 2f_c \text{ step}]$ , and the search step of  $f_c$  is set to half of the initial value. Similarly, the  $f_b$  range is updated to  $[f_{bo} - f_b \text{ step}, f_{bo} + f_b \text{ step}]$ , and the search step of  $f_b$  is one-fifth of the initial value. The total number of iterations is set to 3, and in this situation, the minimal  $f_c$  step equals 1 Hz and the minimal  $f_b$  step equals 0.005. The SN error is set to 2, and the results are shown in Figures 4, 5 and 6.

It can be observed from Figure 4 that the final number of iterations is two, which indicates that the computation is terminated ahead of time because the SN error satisfies the requirement in the second iteration.

As the optimal parameters are obtained, the impulsive wavelet with optimal parameters is used to filter the original simulation fault signal. The results are shown in Figure 5.

From Figure 5(b) and (c), it can be seen that the periodical components are quite clear. To further detect the fault characteristic frequency, the squared envelope spectrum is obtained as shown in Figure 6.

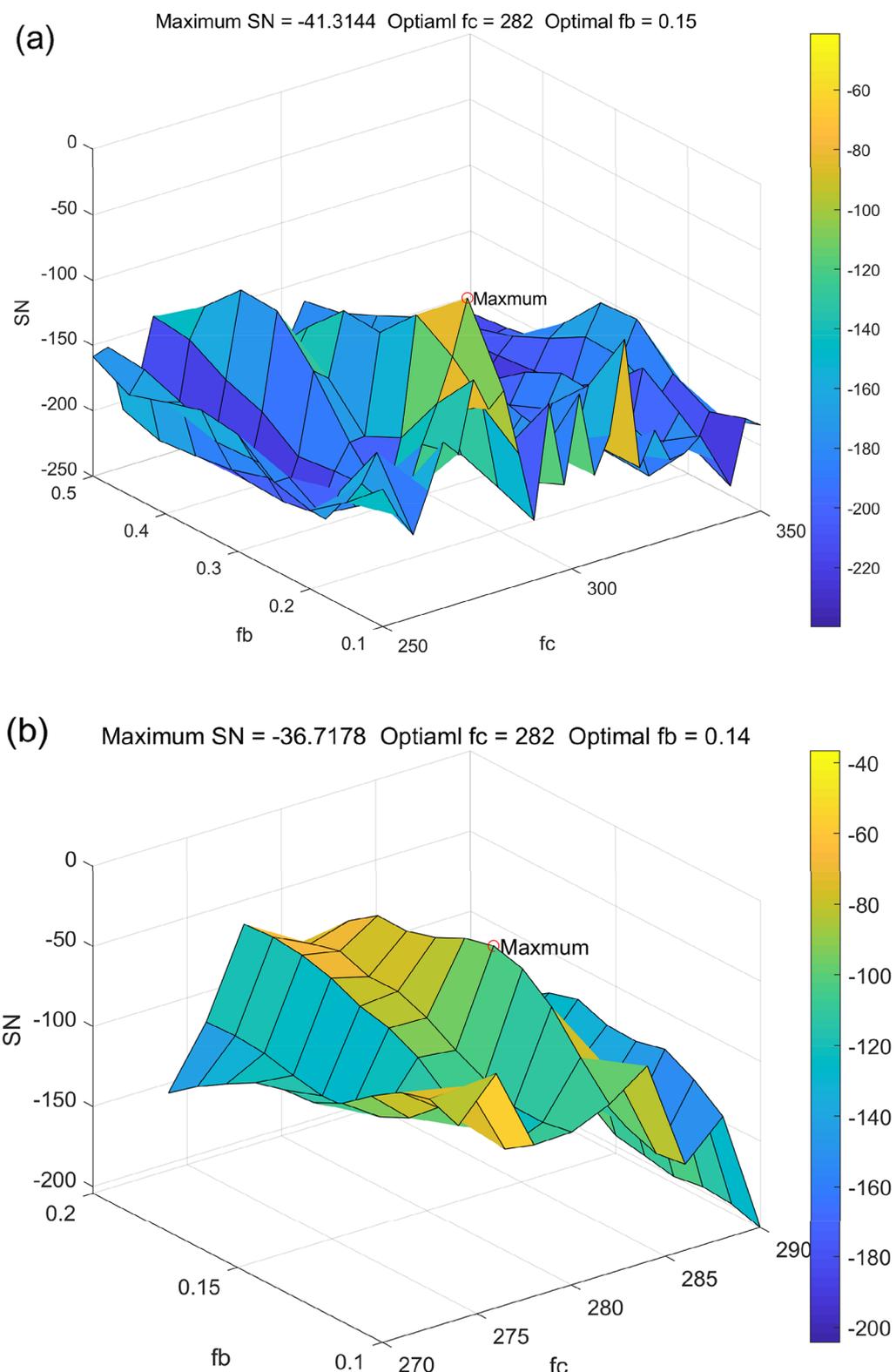
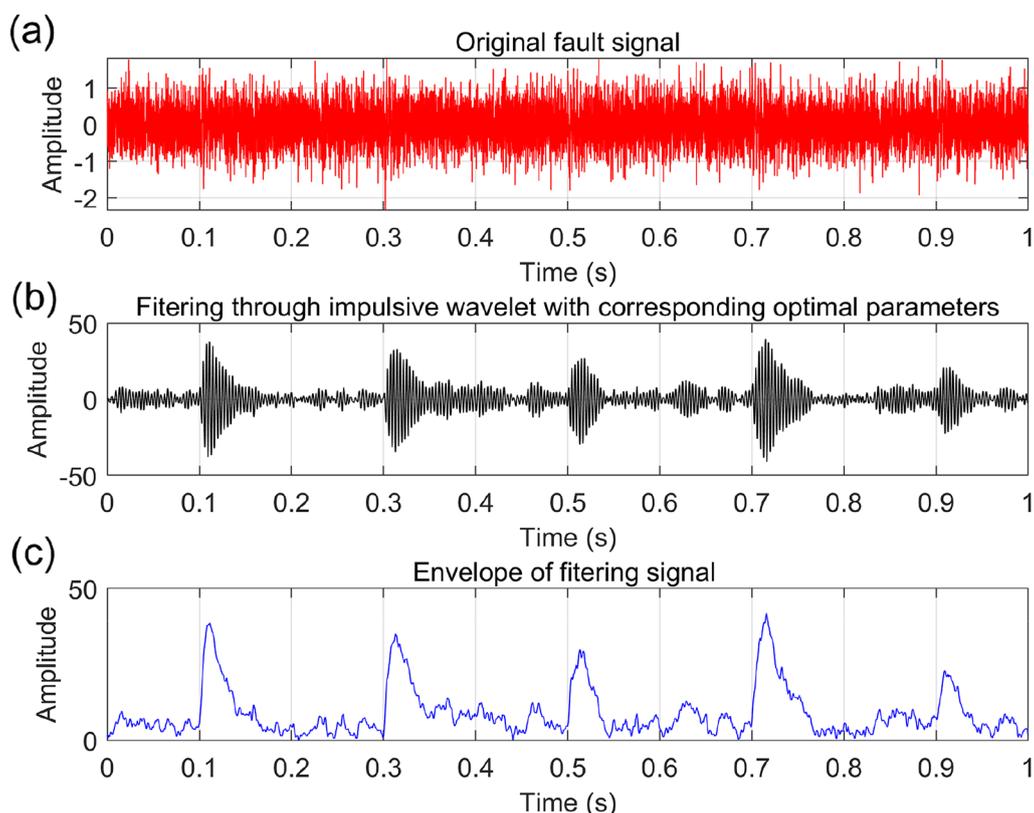
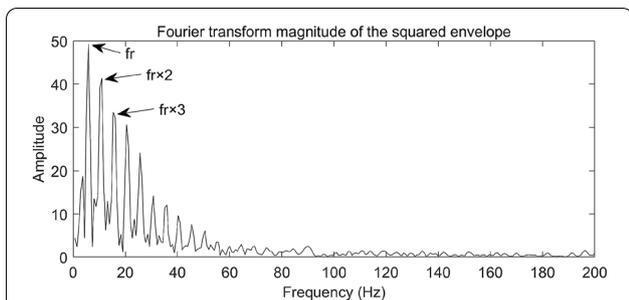


Figure 4 Optimal central frequency and damping ratio obtained by the proposed criterion: (a) first iteration results and (b) second iteration results



**Figure 5** Simulation signal processed by SNM: (a) original signal, (b) signal after filtering by impulsive wavelet, and (c) the envelope of the processed signal



**Figure 6** The squared envelope spectrum of the processed signal is based on the proposed criterion

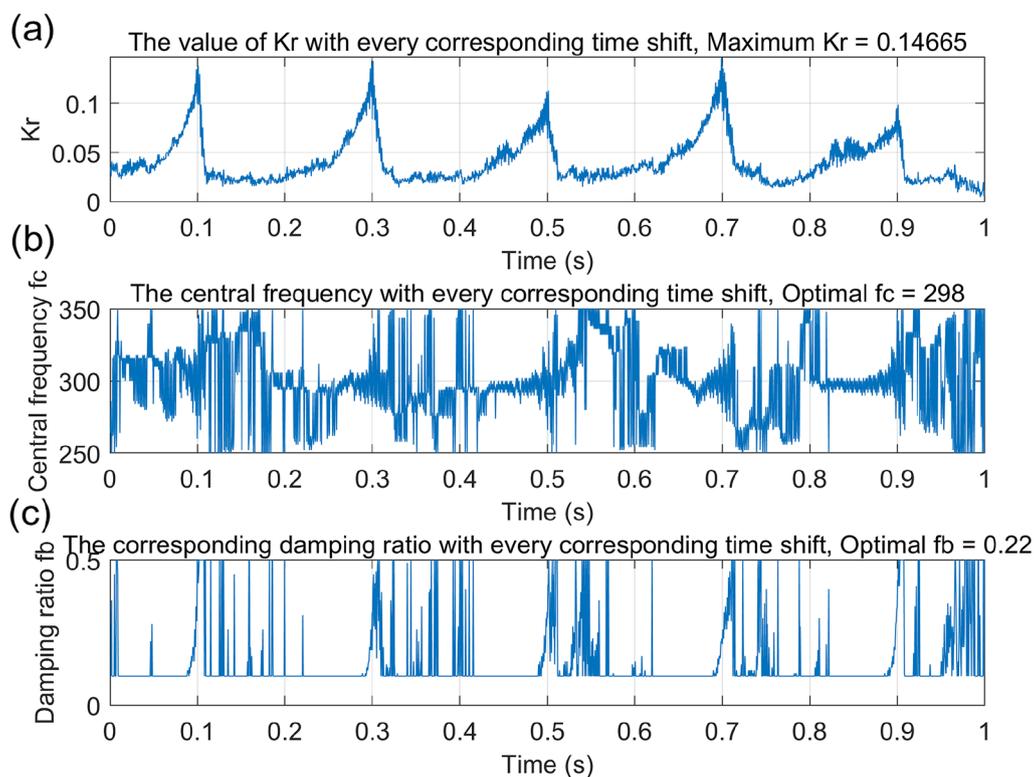
From Figure 6, the fault characteristic frequency  $f_r$  and its frequency multiplications are quite obvious. The fault characteristic frequency is 5 Hz, indicating that periodical transients appear in the simulation signal. Thus, a conclusion can be drawn that the fault features existing in the simulation signal can be detected.

First, the correlation coefficient maximum method (CCM) [21] is imported to deal with simulation signals in contrast to the proposed method. This method was applied to process the simulation signals, and the results are presented in Figures 7, 8 and 9:

From Figure 8(b), it can be observed that the periodical components are identified, but the impulse amplitude is lower than that in Figure 5(b). In Figure 9, the amplitude of the fault characteristic frequency is also not as high as that in Figure 6. Thus, the proposed method is more efficient than the CCM.

Then, the kurtosis maximum method (KM) [16] was imported to detect the fault in the simulation signal, and the results are shown in Figures 10, 11 and 12.

From Figure 11(b) and (c), it is obvious that the periodical components are discerned, but the same problem occurs in that the impulse amplitude is lower than that in Figure 5(b) and (c). In Figure 12, the amplitude of the fault characteristic frequency is not as high as that in Figure 6. Thus, the proposed method is more efficient than KM at the same resolution as the squared envelope spectrum.



**Figure 7** Optimal central frequency and damping ratio obtained by CCM: **(a)** the value of correlation coefficient with every corresponding time shift, **(b)** the central frequency with every corresponding time shift, and **(c)** the damping ratio with every corresponding time shift

The main initial parameters of these methods and the results of the optimal parameters of the corresponding impulsive wavelets are listed in Table 2.

From Table 2, it can be seen that the initial search range and step size are the same, but the filtering results of the proposed method are better through qualitative analysis.

To analyze the difference between these approaches quantitatively, the time consumption and characteristic power ratio (CPR) [32] were employed, and the definition of CPR is as follows:

$$\text{CPR} = p_f / p_r, \quad (14)$$

where  $p_f$  denotes the fault frequency power and  $p_r$  is the residual power. A higher value of the CPR indicates a more obvious fault characteristic frequency. In particular, the performance is better.

By using Eq. (14), the CPRs of the different methods were calculated and are listed in Table 3. The computational environment used a Windows 10 operation system, the CPU is i7-8700K at 3.7 GHz.

From Table 3, it is apparent that the proposed method is far better in terms of both time consumption and CPR.

Therefore, in the case of simulation fault signals, the superiority of the proposed method was verified.

## 4 Experiment Validation

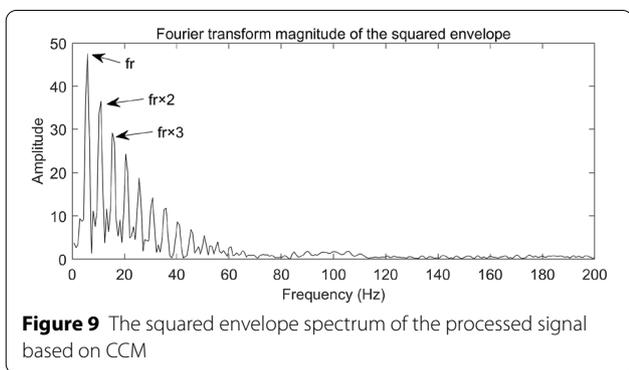
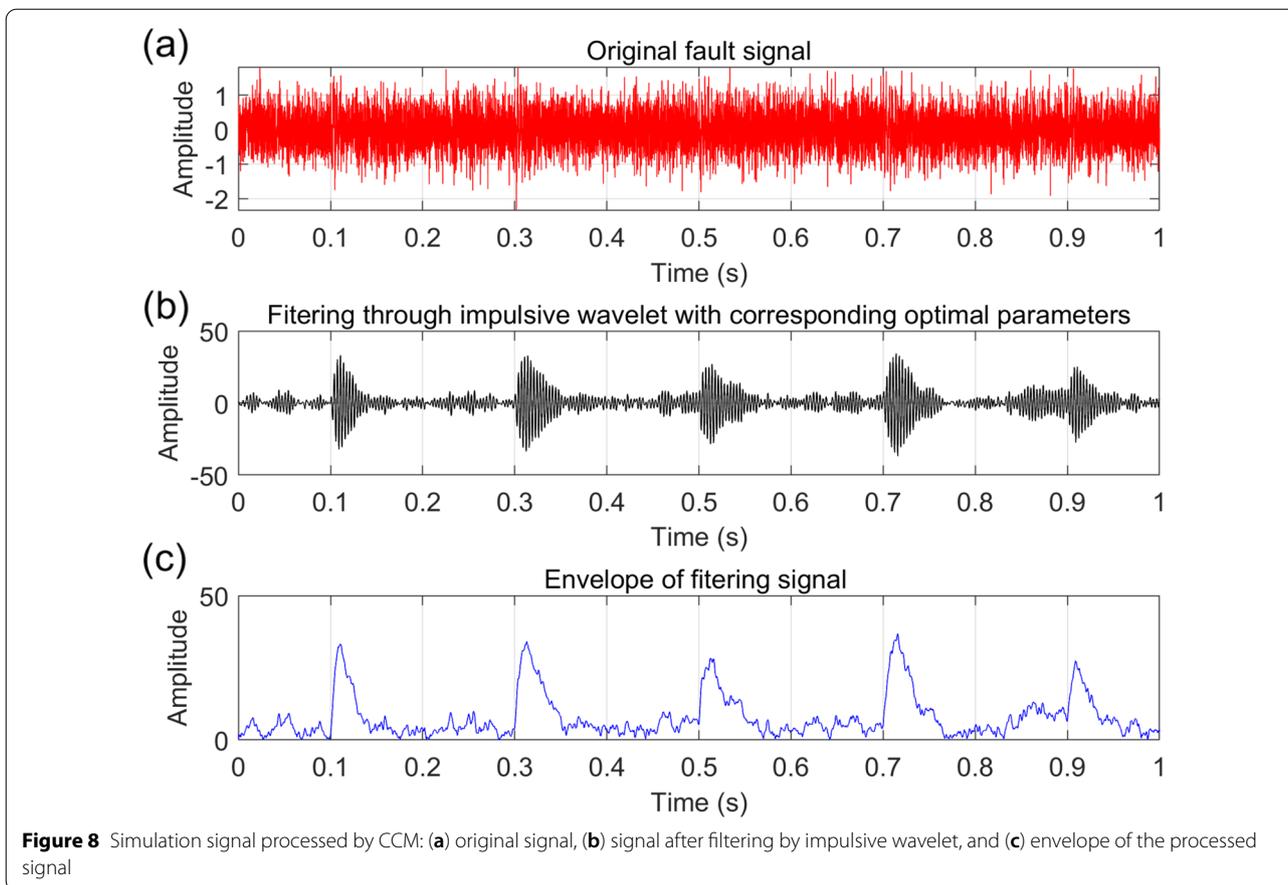
### 4.1 Case Study 1: Tooth Broken Fault

The test rig, Gearbox Dynamics Simulator (GDS) (from Spectra Quest Company), as shown in Figure 13, is a gearbox with three parallel shafts. An experimental study was conducted using a tooth broken fault. The inner structure of the gearbox and gear fault are shown in Figure 14. The rotation speed was equivalent to 30 Hz.

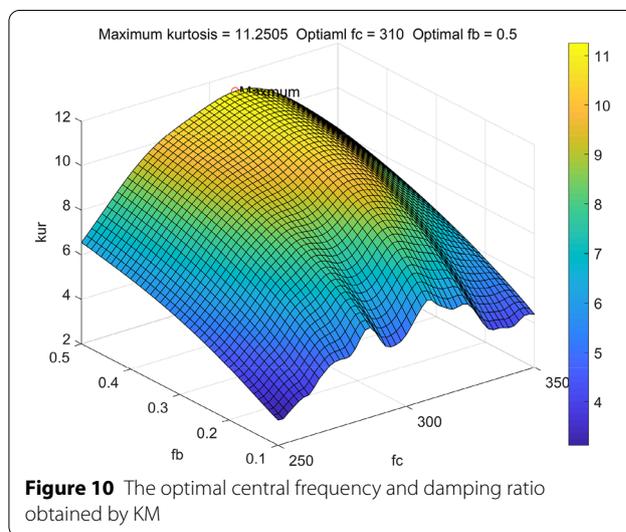
The number of teeth in the faulted gear is 36, and the meshing frequency of the faulted gear is 305.9 Hz. Therefore, the periodical transient frequency of the faulted gear can be easily obtained, which is equivalent to 8.497 Hz (0.12 s). The sampling frequency was set to 12000 Hz.

The time waveform and corresponding frequency spectrum of the original gear fault vibration signal are shown in Figure 15.

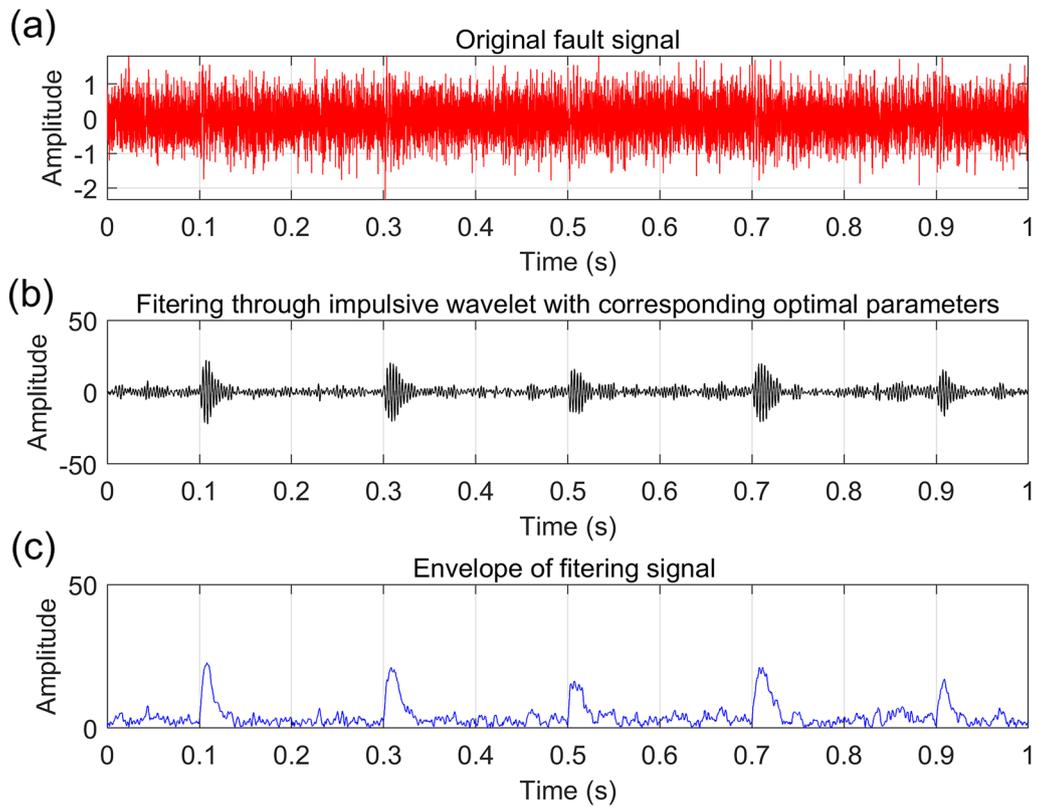
From Figure 15(b) and (c), it is difficult to detect the fault information. The proposed method was used to resolve this problem.



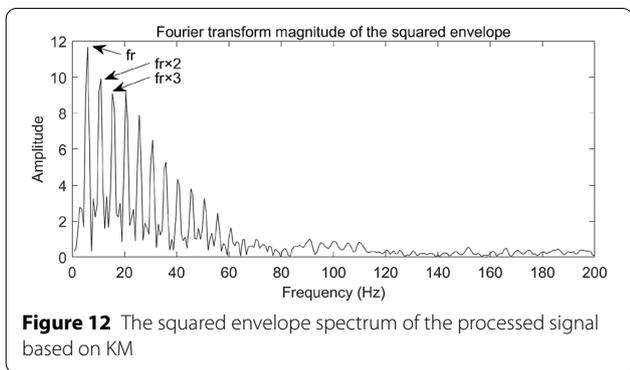
The initial central frequency  $f_c$  range was initially set as 400–500 Hz, and the initial search step size was set to 4. The range of  $f_b$  was selected as [0.1, 0.5], and the initial search step size was set to 0.05. The total number of iterations was 3, and in this situation, the minimal step is 1 Hz and the minimal step is 0.005, which are the recommended values in the algorithm procedure. The SN error was set to 2.



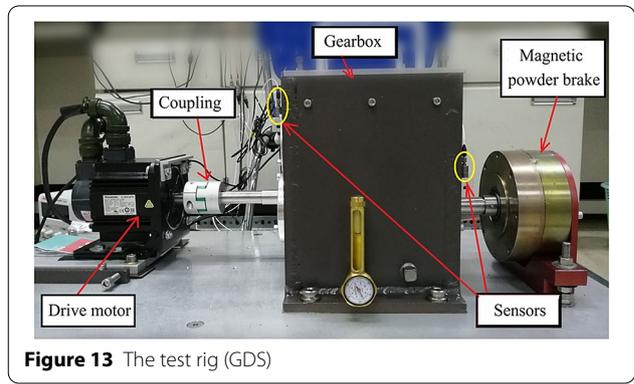
The procedure was also conducted according to the flow chart. The results from the proposed algorithm are illustrated in Figures 16, 17 and 18.



**Figure 11** Simulation signal processed by KM: (a) original signal, (b) signal after filtering by impulsive wavelet, and (c) envelope of the processed signal



**Figure 12** The squared envelope spectrum of the processed signal based on KM



**Figure 13** The test rig (GDS)

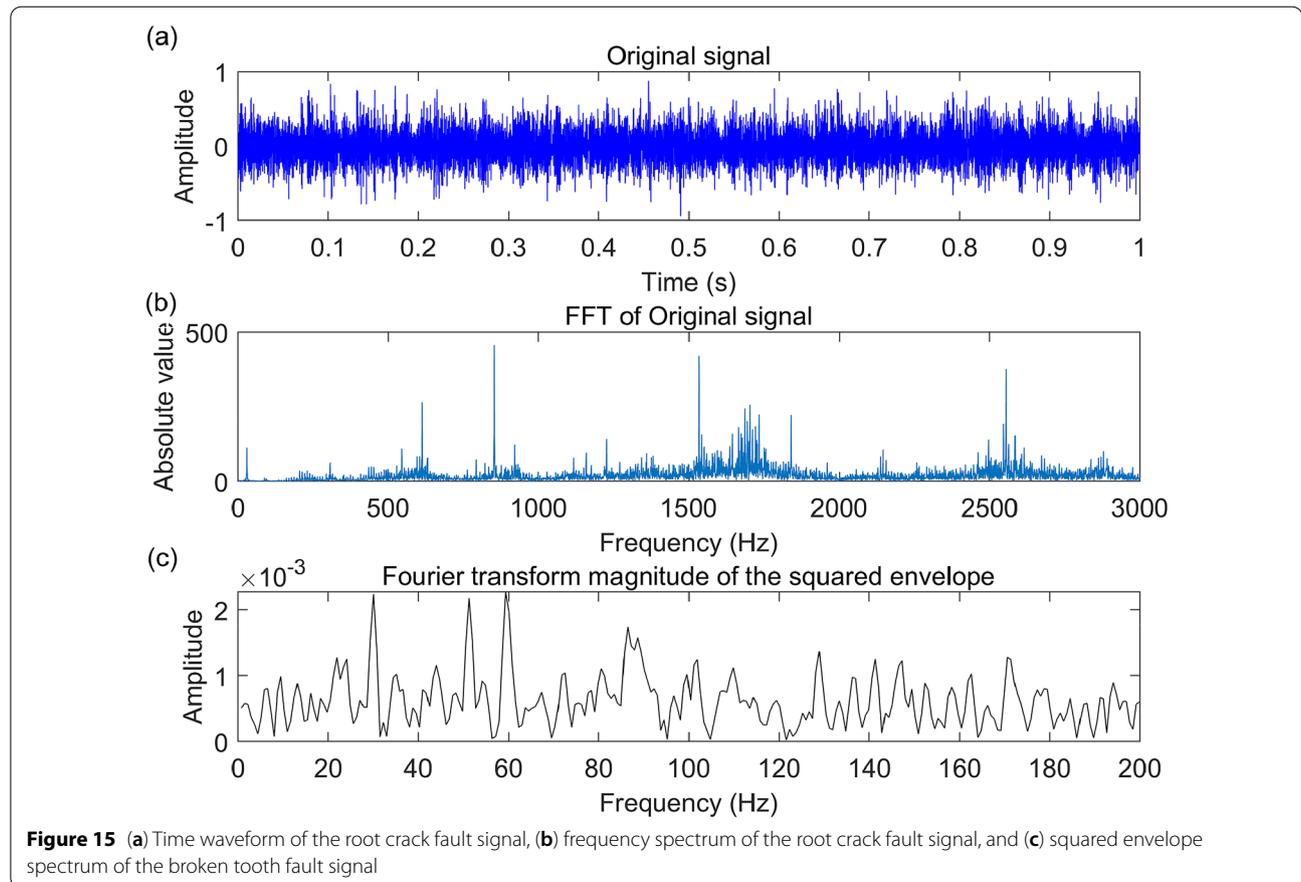
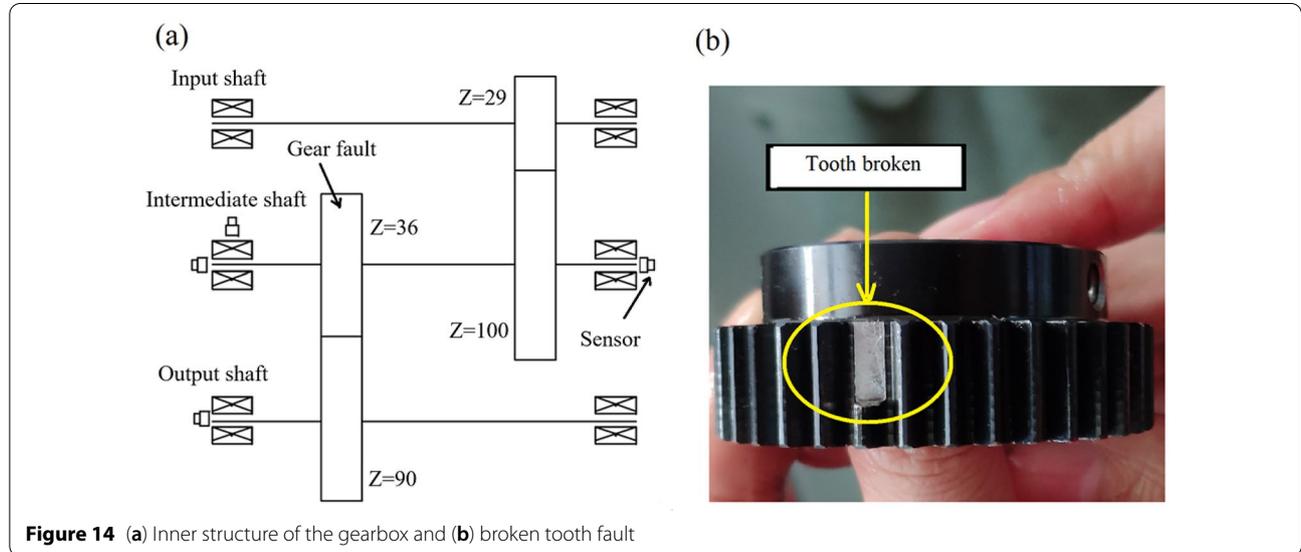
**Table 2** Searching parameters

	$f_c$ step	$f_{co}$	$f_c$ range	$f_b$ step	$f_{bo}$	$f_b$ range
SNM 1st iteration	4	282	250–350	0.05	0.15	0.1–0.5
SNM 2nd iteration	2	282	278–286	0.01	0.14	0.05–0.15
CCM	2	298	250–350	0.01	0.22	0.1–0.5
KM	2	310	250–350	0.01	0.5	0.1–0.5

**Table 3** The time consumption and CPRs of different methods for simulation signals

	SNM	CCM	KM
Time Consumption(s)	174.2	1420.2	1091.5
CPRs	1.6096	1.5619	0.9287

The optimal central frequency  $f_{co}$  and optimal damping ratio  $f_{bo}$  corresponding to the maximum negentropy can be obtained in the first iteration; then, they are used to initialize the  $f_c, f_b$  range in the second iteration. Finally, the search range becomes smaller and the resolution higher. Therefore, the results are more accurate. Because



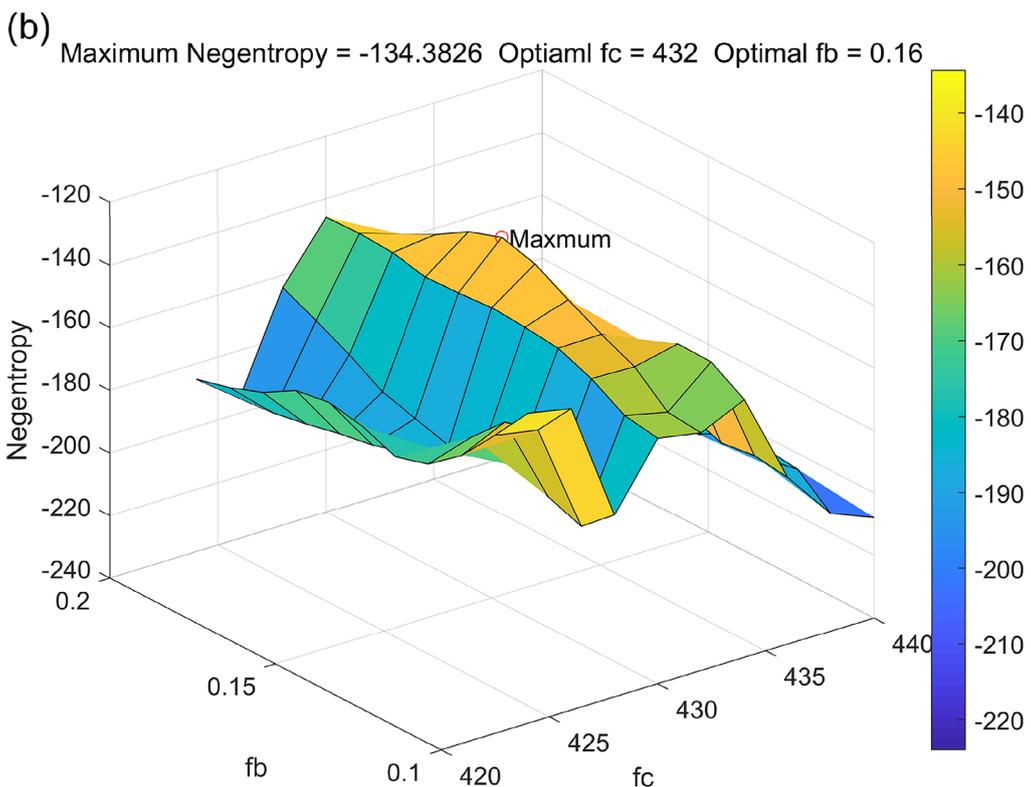
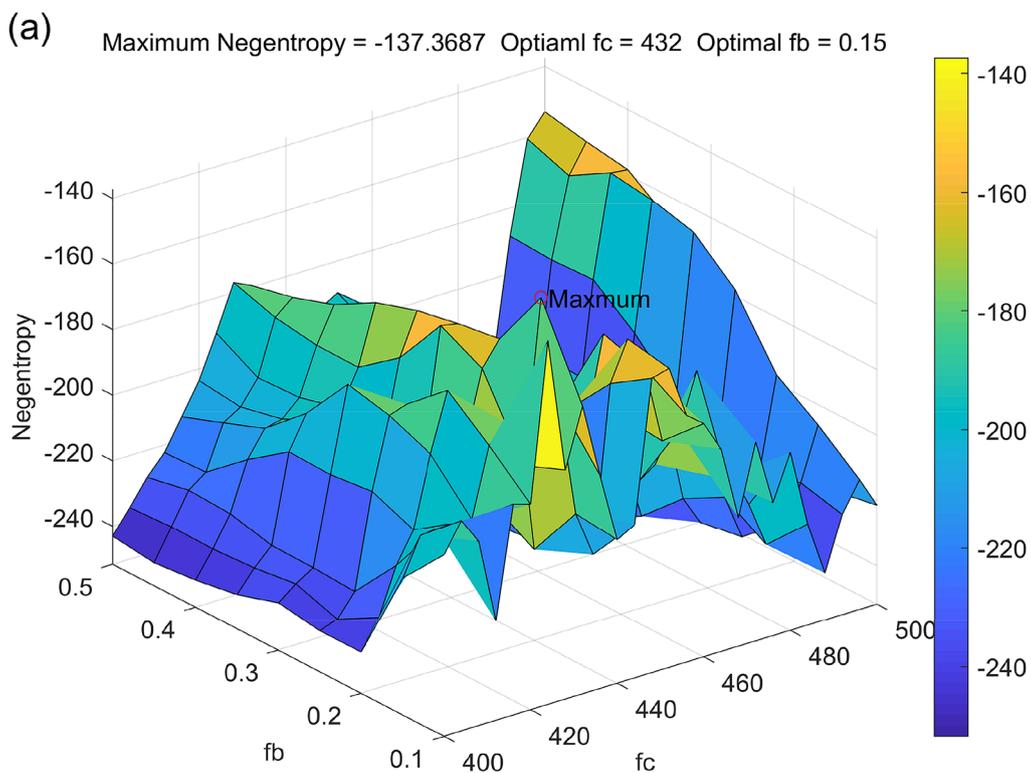
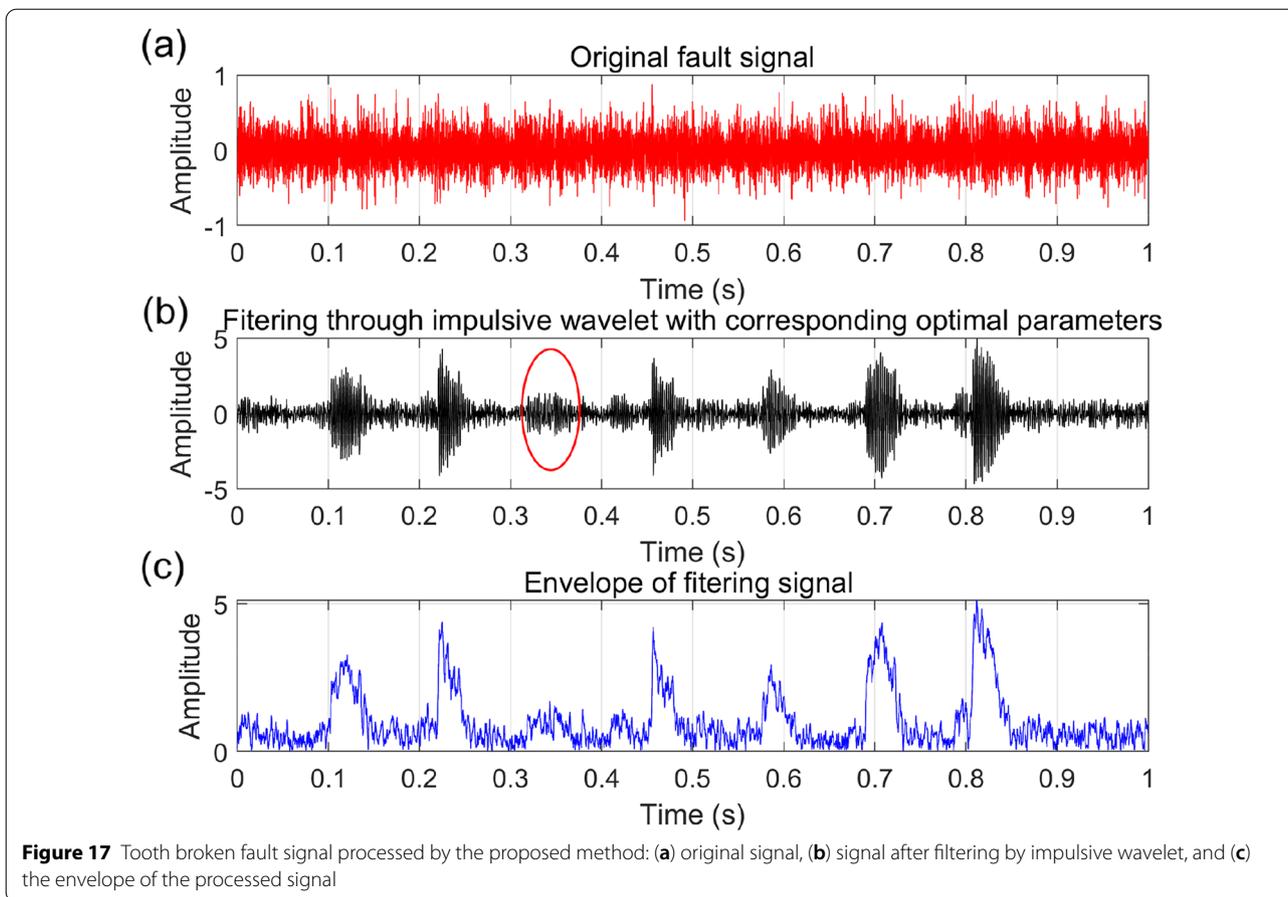
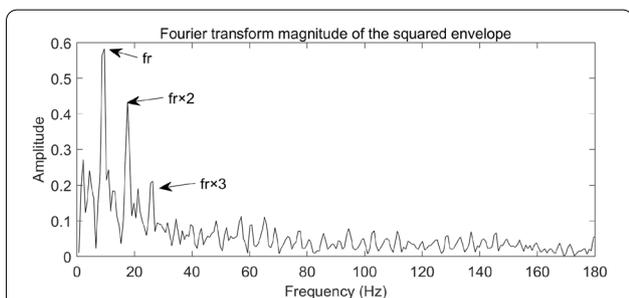


Figure 16 Optimal central frequency and damping ratio obtained by the proposed criterion: (a) first and (b) second iteration results



the search step is far less than the traditional methods, it has a faster computation speed.

It should be noted that the choice of initial step size in the first iteration is important, because too large a value may result in missing the peak SN value, while too small a value has no effect on reducing the time consumption.

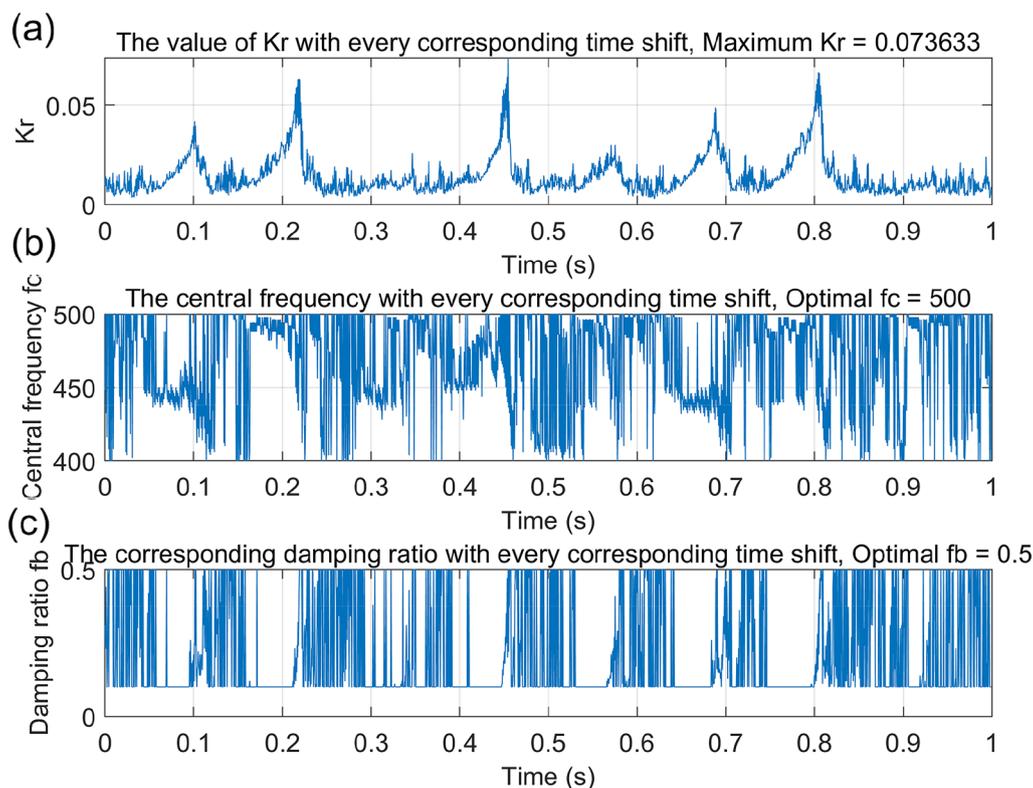


Then, the original faulty signal is processed by the optimal impulsive wavelet, which was constructed by the optimal central frequency  $f_{co}$  and optimal damping ratio  $f_{bo}$  from Figure 16(b). The envelope of the original signal after filtering is subsequently obtained, and the results are presented in Figure 17.

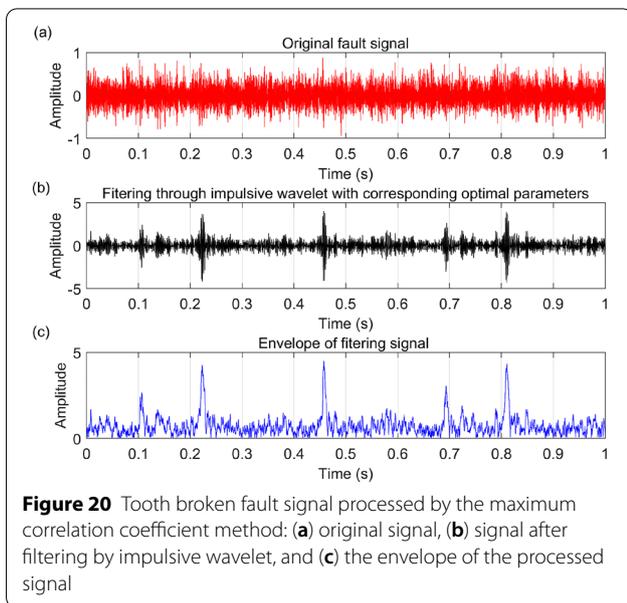
As shown in Figure 17, the periodical component is quite clear. Even weak impulsive components were detected, such as the third impulse marked in Figure 17(b). To further determine the fault characteristic frequency, the square envelope spectrum was also obtained, as shown in Figure 18.

From Figure 18, it can be seen that the fault characteristic frequency  $f_r$  is clearly identified. Therefore, the proposed method can be used to detect gear faults through the squared envelope spectrum.

To compare the proposed method with the other two typical methods referred to earlier, the same gear fault signal is also applied to be processed by the other three methods.

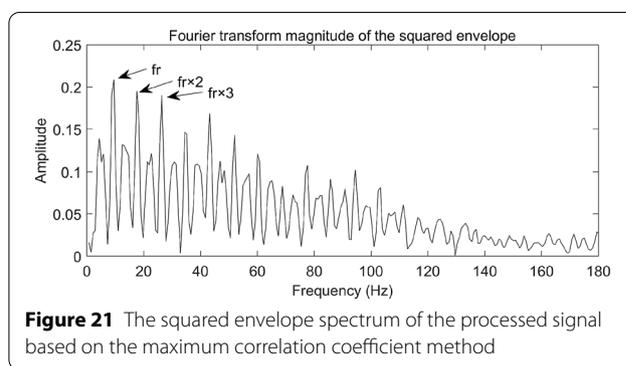


**Figure 19** The optimal central frequency and damping ratio were obtained by the maximum correlation coefficient criterion: (a) the value of the correlation coefficient with every corresponding time shift, (b) the central frequency with every corresponding time shift, and (c) the damping ratio with every corresponding time shift



**Figure 20** Tooth broken fault signal processed by the maximum correlation coefficient method: (a) original signal, (b) signal after filtering by impulsive wavelet, and (c) the envelope of the processed signal

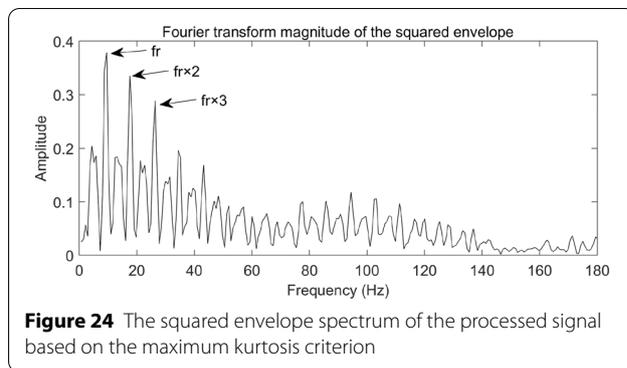
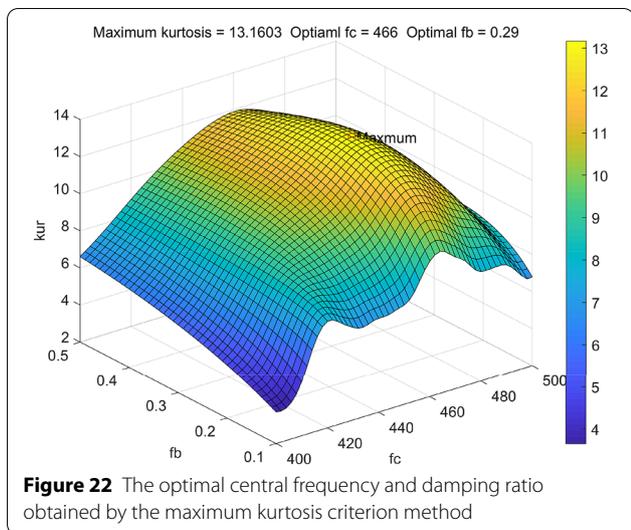
First, the correlation coefficient maximum criterion method was applied to process this signal. The results are illustrated in Figures 19, 20 and 21.



**Figure 21** The squared envelope spectrum of the processed signal based on the maximum correlation coefficient method

From Figure 20(b) and (c), the periodical components are not clearly identified; moreover, they are less obvious than those in Figure 17. The fault characteristic frequency is shown in Figure 21, but the amplitude is far lower than that shown in Figure 18.

Finally, the kurtosis maximum criterion method was applied to process the tooth broken fault signal. The results are illustrated in Figures 22, 23 and 24.

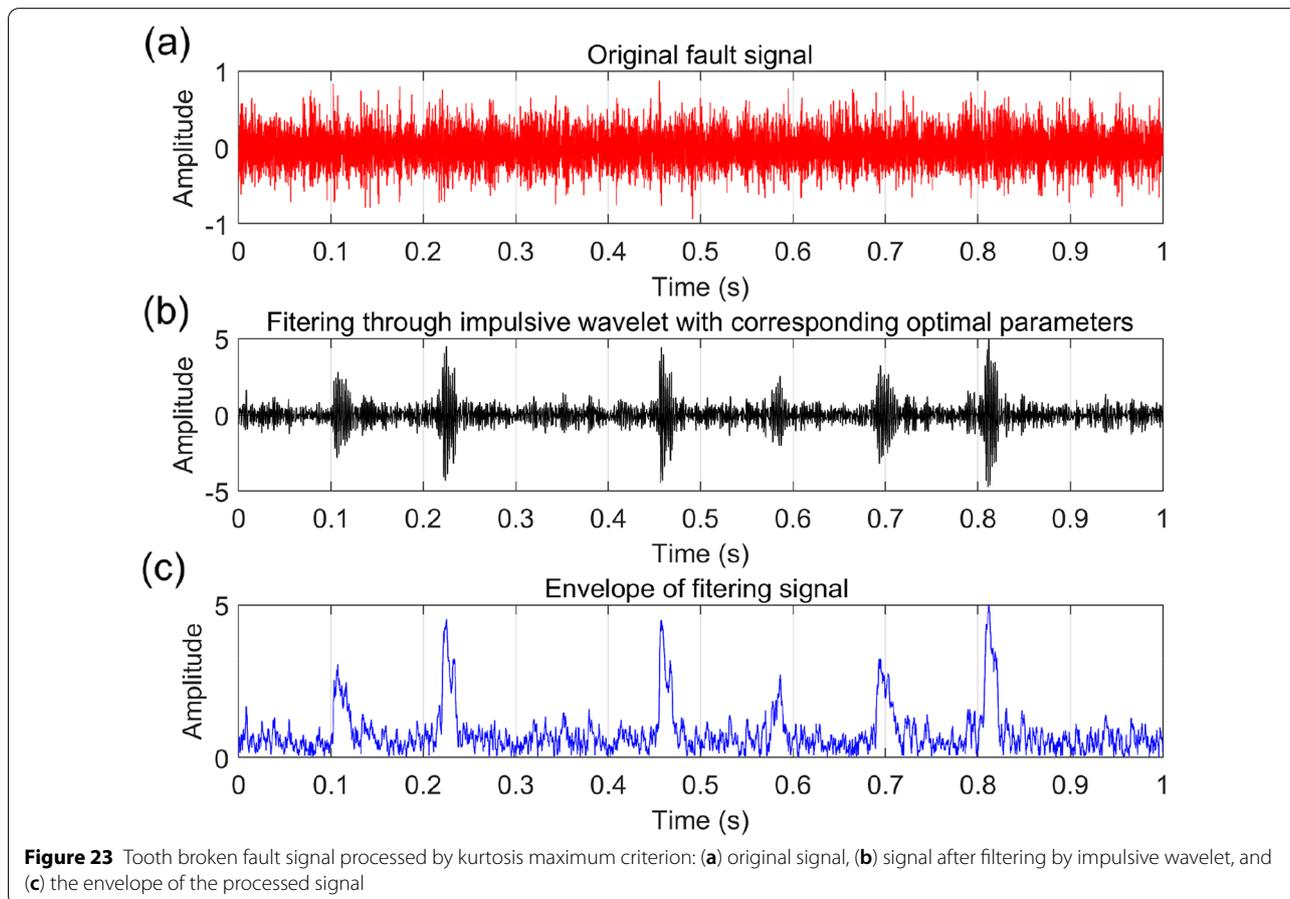


From Figure 23(b) and (c), the periodical components are also less distinct than those in Figure 17, especially in detecting the incipient component, such as the third

impulse. From Figure 24, the fault characteristic frequency is identified as well, but the amplitude is also lower than that in Figure 18.

Thus, the initial superiority of the proposed method is validated.

All the initializing parameters are listed in Table 4 to show the fairness of these methods in comparison.



**Table 4** Searching parameters

	$f_c$ step	$f_{co}$	$f_c$ range	$f_b$ step	$f_{bo}$	$f_b$ range
SNM 1st iteration	4	432	400–500	0.05	0.15	0.1–0.5
SNM 2nd iteration	2	432	436–452	0.01	0.16	0.1–0.2
CCM	2	500	250–350	0.01	0.5	0.1–0.5
KM	2	466	250–350	0.01	0.29	0.1–0.5

Similarly, the time consumption and the CPRs of the different methods were calculated and are listed in Table 5.

From Table 5, it can be seen that the time consumption of the proposed method is also far less than that of the other approaches, and the CPR of this method is larger than that of the other approaches. Therefore, the superiority of the proposed method is validated.

**4.2 Case Study 2: Root Crack Fault**

To further prove the superiority of the proposed method, another set of experimental data were employed by the

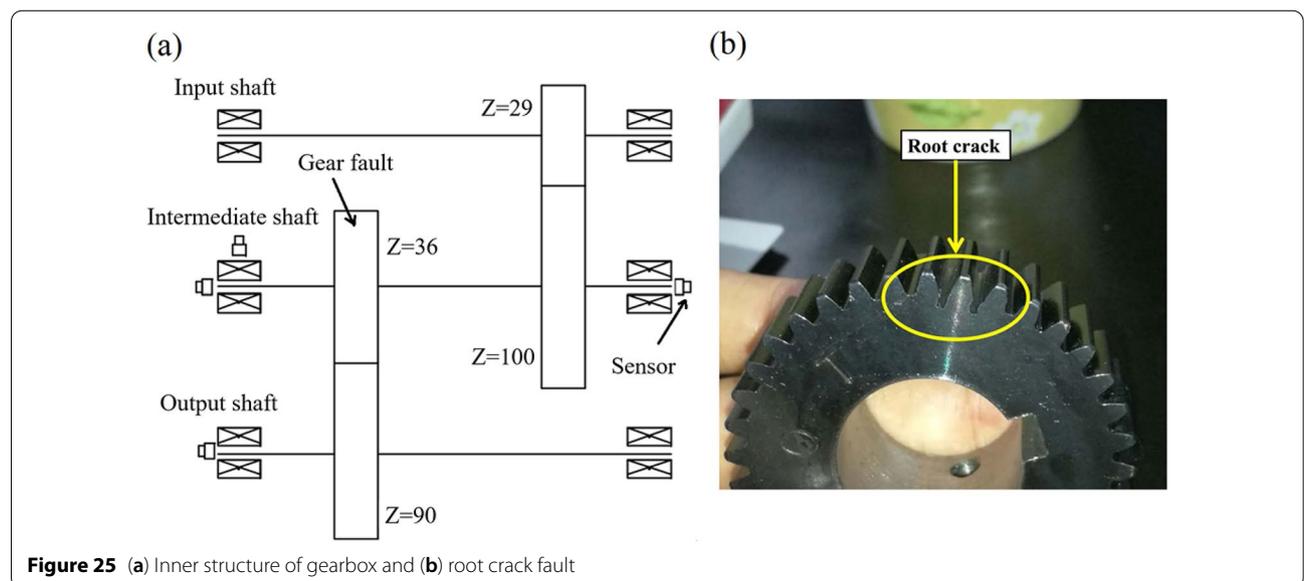
**Table 5** Time consumption and CPRs of different methods for tooth broken fault signal

	SNM	CCM	KM
Time Consumption(s)	274.8	1951.9	1726.5
CPRs	0.7178	0.3217	0.4579

tooth broken fault is substituted by a root crack, which is a crack with 1 mm deep damage. The inner structure of the gearbox and gear fault is shown in Figure 25. Normally, root crack faults are regarded as incipient faults that are difficult to discern in a rotational machine system. This experiment was conducted to further validate the efficiency of the proposed approach for detecting incipient faults.

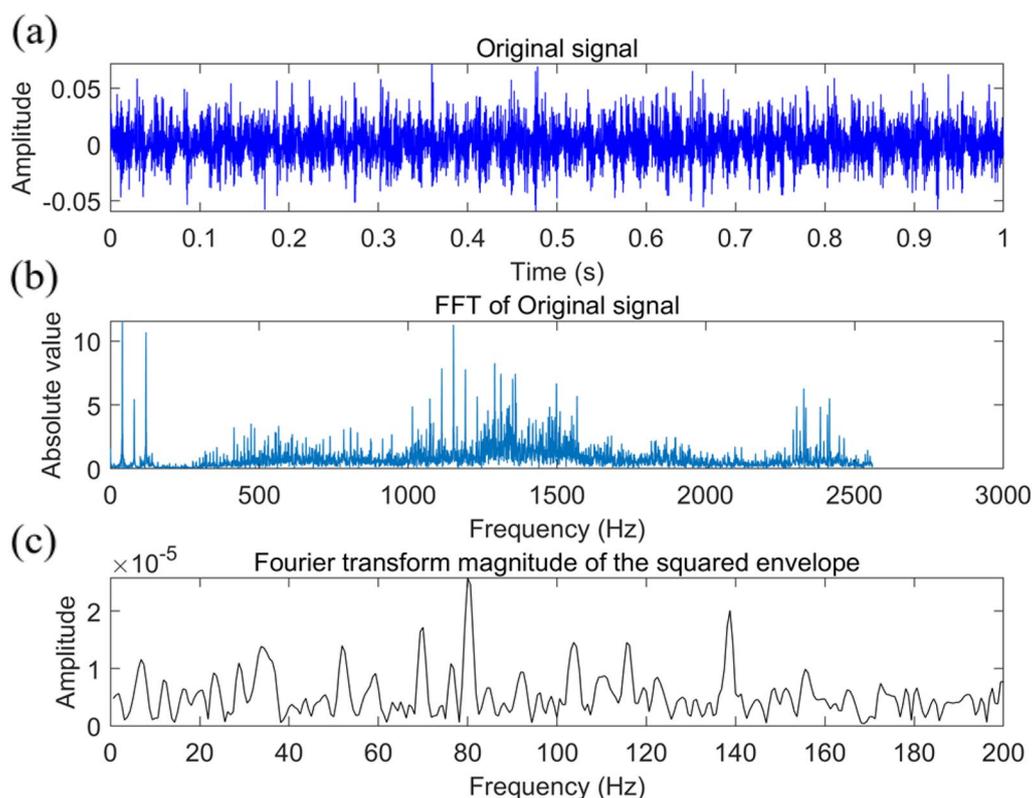
The rotational speed of the input shaft was set to 2400 r/min, and the rotational frequency of the shaft with the faulted gear was 40 Hz. The number of teeth is the same as in case study 1, and the meshing frequency of the faulted gear is 417.6 Hz. Thus, the periodical transient frequency of the faulted gear can be easily obtained, which is equivalent to 11.5 Hz (0.087 s). The sampling frequency was set to 5120 Hz.

The time waveform of the original gear fault vibration signal and its corresponding frequency spectrum are shown in Figure 26.



**Figure 25** (a) Inner structure of gearbox and (b) root crack fault

same test rig, as shown in Figure 13. The gear with a



**Figure 26** (a) Time waveform of the root crack fault signal, (b) frequency spectrum of the root crack fault signal, and (c) squared envelope spectrum of the root crack fault signal

From Figures 26(b) and (c), it is difficult to find the fault information. The proposed method is applied to resolve this problem.

As demonstrated by the SNM paradigm, the initial central frequency  $f_c$  range is set from 320 to 470. The searching parameters of  $f_b$  are set to be the same as those in case study 1. The SN error was set to 2. The results from the SNM are illustrated in Figures 27, 28.

Similarly, the SNM method was applied to determine the optimal impulsive wavelet filter that was applied to filter the raw signal, and the envelope of the original signal after filtering is obtained, and the results are presented in Figure 28.

As shown in Figure 27, the periodical component is quite clear. To further determine the fault characteristic frequency, the square envelope spectrum was also obtained, as shown in Figure 28.

From Figure 28, it can be seen that the fault characteristic frequency  $f_r$  is clearly identified. Therefore, the

proposed method can be used to identify incipient gear faults through the squared envelope spectrum.

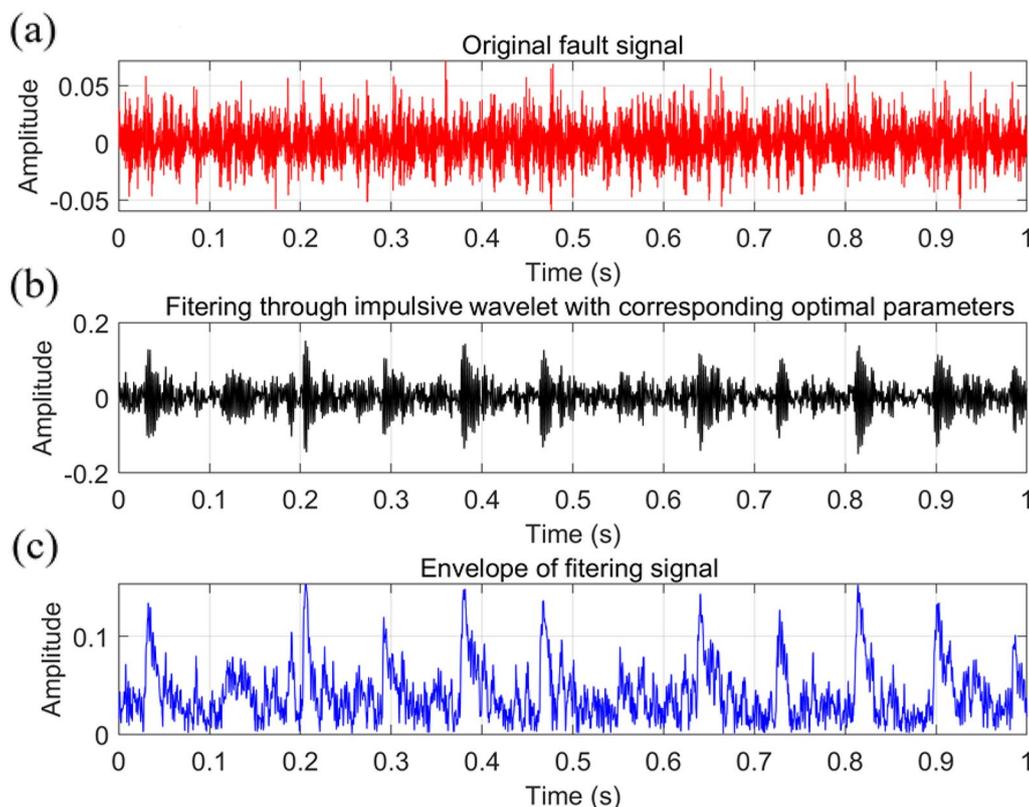
In comparison to the proposed method, the other two typical approaches referred to earlier are also employed to process the same gear fault signal.

First, the CCM was employed to process this signal. The results are illustrated in Figures 29, 30.

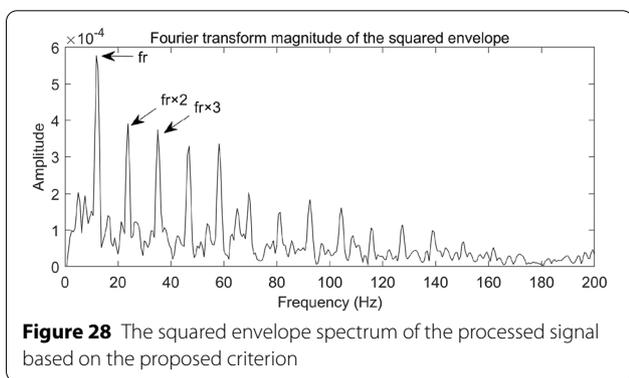
From Figure 29(b) and (c), the periodic components are identified, and the fault characteristic frequency is also found in Figure 30, but both are less obvious than those in Figures 27 and 28.

KM was then applied to process this signal as well. The results are illustrated in Figures 31, 32.

From Figure 31(b) and (c), the periodic components are also less distinct than those in Figure 27. From Figure 32, fault characteristic frequency is also found, but the amplitude is lower than that in Figure 28.



**Figure 27** Root crack fault signal processed by the proposed method: (a) original signal, (b) signal after filtering by impulsive wavelet, and (c) envelope of the processed signal



**Figure 28** The squared envelope spectrum of the processed signal based on the proposed criterion

Thus, the superiority of the proposed method is validated.

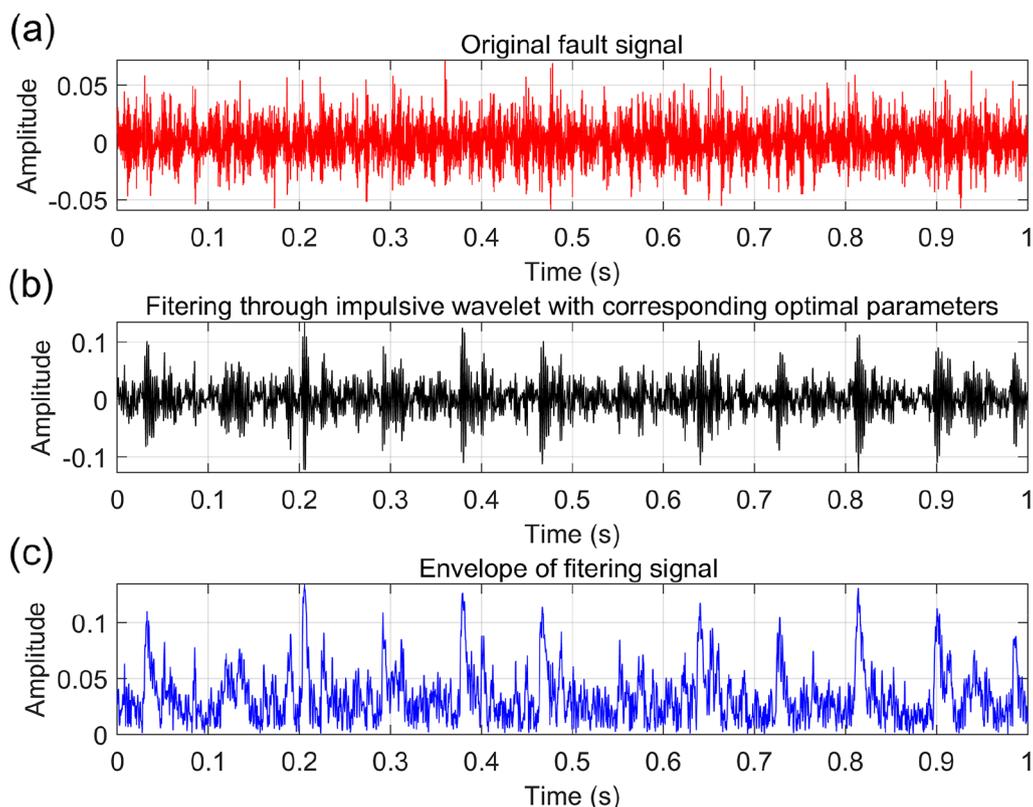
Similarly, all initializing parameters are listed in Table 6.

The time consumption and CPRs of the different methods are calculated and listed in Table 7.

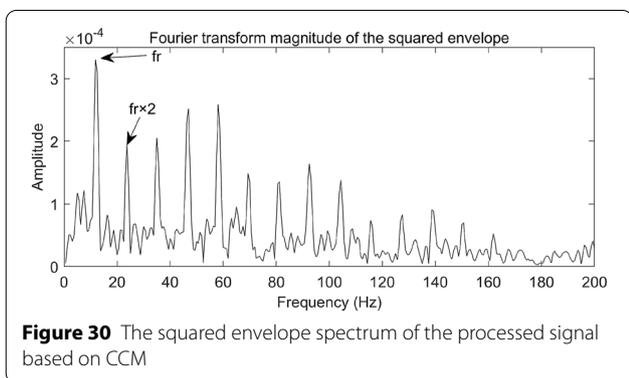
From Table 7, it can be seen that the time consumption of the proposed method is far less than that of the other approaches, and the CPR of the proposed method is also larger than that of the other approaches. Therefore, the superiority of the proposed method was validated.

### 5 Conclusions

In this paper, a novel criterion based on spectral negentropy is proposed for the first time, and its effectiveness is validated by both simulation and experimental signals. From the perspective of energy fluctuations, the proposed method is more sensitive to the change in local fluctuations of the signal energy that results from a decrease in the entropy, which results in a better performance compared with conventional adaptive methods. In measuring the gear fault, kurtosis-based methods,



**Figure 29** Gear fault signal processed by CCM: (a) original signal, (b) signal after filtering by impulsive wavelet, and (c) envelope of the processed signal



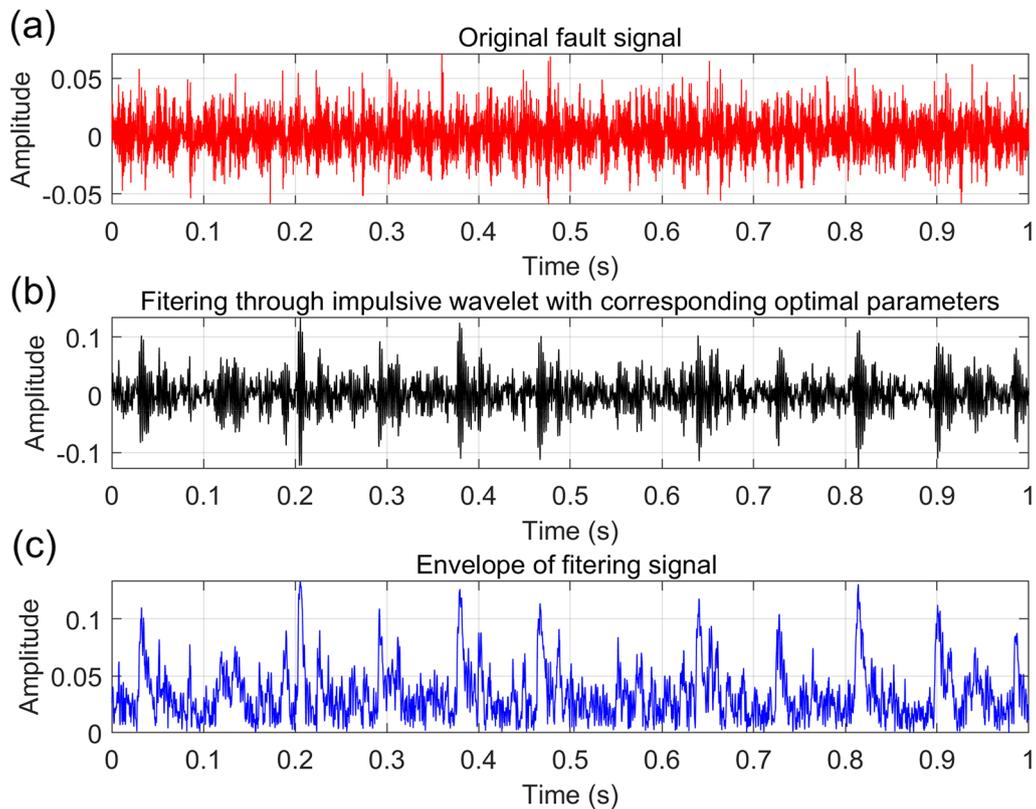
**Figure 30** The squared envelope spectrum of the processed signal based on CCM

such as FK and KM, are essentially a type of method for measuring the dispersion of the observed values. Therefore, they are limited to the distribution of the observed data (normally deemed as Gaussian distribution), which might be unsuitable for transient detection; for example,

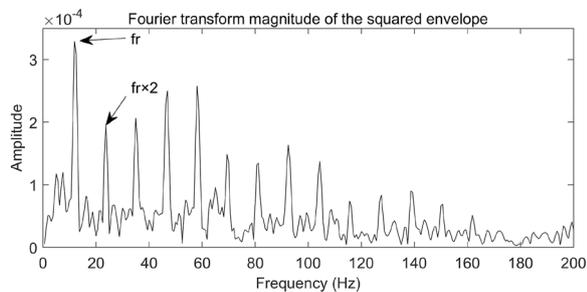
alpha-stable distribution is better than Gaussian distribution [32]. CCM can be used to measure the similarity between the transient and imported wavelets; thus, it is highly dependent on the imported wavelet. If the fault component is too weak or the imported wavelet does not match the fault transient, the application of the method is limited.

Negentropy can measure a transient’s resemblance to a Dirac comb, which is a perfect idealization of a series of transients [11]. Therefore, the performance is better than that of conventional approaches. However, this limitation is also illustrated in Ref. [11], the accuracy is highly related to the sampling frequency.

In addition, a modified impulsive wavelet is employed as a filter to process the impulsive transient masked in strong noise. The incipient fault feature was extracted using this modified wavelet. In addition, a fast optimizing strategy was developed to reduce the time required



**Figure 31** Gear fault signal processed by KM: (a) original signal, (b) signal after filtering by impulsive wavelet, and (c) envelope of the processed signal



**Figure 32** The squared envelope spectrum of the processed signal based on KM

**Table 7** The time consumption and CPRs of different methods for gear fault signal

	SNM	CCM	KM
Time Consumption(s)	113.974	1430.4	1342.1
CPRs	0.3156	0.2447	0.2454

**Table 6** Searching parameters

	$f_c$ step	$f_{co}$	$f_c$ range	$f_b$ step	$f_{bo}$	$f_b$ range
SNM 1st iteration	4	444	320–470	0.05	0.3	0.1–0.5
SNM 2nd iteration	2	444	436–452	0.01	0.31	0.25–0.35
SNM 3rd iteration	1	444	440–448	0.005	0.314	0.3–0.32
CCM	2	302	250–350	0.01	0.2	0.1–0.5
KM	2	302	250–350	0.01	0.16	0.1–0.5

to find the optimal parameters of the impulsive wavelet filter. Compared to the conventional searching methods and other adaptive wavelet approaches, the proposed method is faster and more effective; thus, its superiority was verified. Owing to the limitations of the experimental equipment, the impact of the load on the SN has not been investigated. Future research can focus on investigating the influence of varying loads on spectral negentropy.

#### Acknowledgements

Not applicable.

#### Authors' contributions

MG was in charge of the whole trial; MG, GY and CL wrote the manuscript; GY analyzed the data. All authors read and approved the final manuscript.

#### Authors' Information

Mang Gao was born in 1990. He got his B.S degree in Petroleum Engineering from Yangtze University (Jinzhou), China, in 2012. Currently, he is pursuing his Master degree in Harbin Institute of Technology at Shenzhen, Shenzhen, China. His research interests include the fault diagnosis and prognosis, condition monitoring of wind turbines, machine learning applied in complex machinery system.

Gang Yu was born in 1969. He received the B.S. and Master degrees in Mechanical Engineering from Dalian University of Technology, Dalian, China, in 1992 and 1995, respectively. He obtained his Ph.D. degree from University of Wisconsin-Milwaukee, USA. Since 2005, he has been worked in the Harbin Institute of Technology, Shenzhen, China, where he is currently an associate professor in the School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China. His current research interests include signal processing, fault prognosis and diagnosis, condition monitoring, intelligent maintenance, service robots.

Changning Li was born in Hubei, China, in 1979. He received the B.E. degree in electrical engineering from Hubei Minzu University, in 2001, and the M.S. and Ph.D. degrees in control science and engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 2005 and 2010, respectively. He is currently an Assistant Professor with the Suzhou University of Science and Technology, Suzhou, China. His current research interests include machinery fault diagnostic and prognostic, machine learning, and statistical signal analysis.

#### Funding

Supported by Shenzhen Fundamental Research (Grant No. JCYJ20190806144401666).

#### Competing Interests

The authors declare that they have no competing financial interests.

#### Author Details

<sup>1</sup>School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China. <sup>2</sup>School of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China.

Received: 24 September 2020 Revised: 18 October 2021 Accepted: 30 December 2021

Published online: 12 February 2022

#### References

- Wang, G Du, Z Zhu, et al. Fault diagnosis of rotating machines based on the EMD manifold. *Mechanical Systems and Signal Processing*, 2020, 135: 106443.
- Z Xin, Z Liu, M Qiang, et al. Bearing fault diagnosis using a whale optimization algorithm-optimized orthogonal matching pursuit with a combined time-frequency atom dictionary. *Mechanical Systems & Signal Processing*, 2018, 107: 29–42.
- J Singh, A K Darpe, S P Singh. Rolling element bearing fault diagnosis based on Over-Complete rational dilation wavelet transform and auto-correlation of analytic energy operator. *Mechanical Systems & Signal Processing*, 2018, 100: 662–693.
- R Zhao, R Q Yan, Z H Chen, et al. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, 2019, 115: 213–237.
- J Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 2007, 21(1): 108–124.
- Y Qin, J Q Zou, F L Cao. Adaptively detecting the transient feature of faulty wind turbine planetary gearboxes by the improved kurtosis and iterative thresholding algorithm. *IEEE Access*, 2018, 6: 14602–14612.
- Y Xu, W Tian, K Zhang et al. Application of an enhanced fast kurtogram based on empirical wavelet transform for bearing fault diagnosis. *Measurement Science and Technology*, 2019, 30(3): 035001.
- J Antoni. The infogram: Entropic evidence of the signature of repetitive transients. *Mechanical Systems and Signal Processing*, 2016, 74: 73–94.
- C Pachaud, R Salvétat, C Fray. Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces. *Mechanical Systems and Signal Processing*, 1997, 11(6): 903–916.
- Y Miao, M Zhao, J Lin. Improvement of kurtosis-guided-grams via Gini index for bearing fault feature identification. *Measurement Science and Technology*, 2017, 28(12): 125001.
- X H Gu, S P Yang, Y Q Liu, et al. A novel Pareto-based Bayesian approach on extension of the infogram for extracting repetitive transients. *Mechanical Systems and Signal Processing*, 2018, 106: 119–139.
- X F Xu, Z J Qiao, Y G Lei. Repetitive transient extraction for machinery fault diagnosis using multiscale fractional order entropy infogram. *Mechanical Systems and Signal Processing*, 2018, 103: 312–326.
- Y Xu, J Chen, C Ma, et al. Negentropy spectrum decomposition and its application in compound fault diagnosis of rolling bearing. *Entropy*, 2019, 21(5): 490.
- H Jiang, H Lei, Z Wen, et al. Charge-trapping-blocking layer for enhanced triboelectric nanogenerators. *Nano Energy*, 2020, 75: 105011.
- Y Qin, J Xing, Y Mao. Weak transient fault feature extraction based on an optimized Morlet wavelet and kurtosis. *Measurement Science and Technology*, 2016, 27(8): 085003.
- J Lin, M Zuo. Gearbox fault diagnosis using adaptive wavelet filter. *Mechanical Systems and Signal Processing*, 2003, 17 (6): 1259–1269.
- Y Xin, S Li, Z Zhang, et al. Adaptive reinforced empirical Morlet wavelet transform and its application in fault diagnosis of rotating machinery. *IEEE Access*, 2019: 99.
- Ei Morsy M, Achtenova G. Application of optimal morlet wavelet filter for bearing fault diagnosis. *SAE International Journal of Passenger Cars-Mechanical Systems*, 2015: 2015-01-2178.
- D Wang, C Shen, P W Tse. A novel adaptive wavelet stripping algorithm for extracting the transients caused by bearing localized faults. *Journal of Sound and Vibration*, 2013, 332(25): 6871–6890.
- G Yu, M Gao, C Jia. A fast filtering method based on adaptive impulsive wavelet for the gear fault diagnosis. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 2020: 0954406220906245.
- S Wang, W Huang, Z K Zhu. Transient modeling and parameter identification based on wavelet and correlation filtering for rotating machine fault diagnosis. *Mechanical Systems and Signal Processing*, 2011, 25(4): 1299–1320.
- W Fan, G Cai, Z K Zhu et al. Sparse representation of transients in wavelet basis and its application in gearbox fault feature extraction. *Mechanical Systems and Signal Processing*, 2015, 56-57: 230–245.
- Y Qin. A new family of model-based impulsive wavelets and their sparse representation for rolling bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, 2018, 65 (3): 2716–2726.
- Y Qin, J Zou, B Tang, et al. Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary. *IEEE Transactions on Industrial Informatics*, 2019: 1.
- S G Mallat. A theory for multiresolution signal decomposition - the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(7): 674–693.

- [26] Y Jiang, B Tang, Y Qin, et al. Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD. *Renewable Energy*, 2011, 36(8): 2146–2153.
- [27] Y X Wang, J W Xiang, R Markert, et al. Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications. *Mechanical Systems and Signal Processing*, 2016, 66-67: 679–698.
- [28] Y Lei, J Lin, Z He, et al. Application of an improved kurtogram method for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, 2011, 25 (5): 1738–1749.
- [29] J Antoni, R B Randall. The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. *Mechanical Systems and Signal Processing*, 2006, 20(2): 308–331.
- [30] H Liu, W Huang, S Wang, et al. Adaptive spectral kurtosis filtering based on Morlet wavelet and its application for signal transients detection. *Signal Processing*, 2014, 96: 118–124.
- [31] X Jiang, J Wang, J Shi, et al. A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines. *Mechanical Systems and Signal Processing*, 2019, 116: 668–692.
- [32] M Gao, G Yu, T Wang. Impulsive gear fault diagnosis using adaptive morlet wavelet filter based on Alpha-stable distribution and kurtogram. *IEEE Access*, 2019, 7: 72283–72296.
- [33] C Q Shen, F Liu, D Wang, et al. A Doppler transient model based on the laplace wavelet and spectrum correlation assessment for locomotive bearing fault diagnosis. *Sensors*, 2013, 13 (11): 15726–15746.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

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

---

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)

---