Bearing failure of reciprocating compressor sub-health recognition based on CAGOA-VMD and GRCMDE

Meiping Song, Jindong Wang, Haiyang Zhao and Yanyang Li

Abstract
The bearing vibration signal of reciprocating compressor has complex, non-stationary, nonlinear, and feature coupling characteristics. A method for sub-health recognition of sliding bearings based on curve adaptive grasshopper optimization algorithm optimize the parameters of variational mode decomposition (CAGOA-VMD) and generalized refine composite multiscale dispersion entropy (GRCMDE) is used. First, the CAGOA was used to search the best influence parameter combination of the VMD algorithm, and determine the bandwidth parameters and the number of decompositions that need to be set by the VMD algorithm, decompose the bearing fault signal to obtain a series of IMF. Then, the kurtosis and correlation coefficient criteria are used to select a group of components that contain the most information, and the fault signal is reconstructed on this component, and then the reconstructed signal is analyzed by GRCMDE to form a fault eigenvector. Finally, KPCA is used for dimensionality reduction to select input features and input into KELM for classification and recognition. The experimental results show that this method can effectively extract the bearing fault features of reciprocating compressors, and the eigenvectors have good separability, and realize the sub-health recognition of bearing fault features of reciprocating compressors.

Keywords
Fault diagnosis, reciprocating compressor bearing, adaptive grasshopper optimization algorithm, generalized refine composite multi-scale dispersion entropy, Kernel principal component analysis

Date received: 17 August 2021; accepted: 7 February 2022
Handling Editor: Chenhui Liang

Introduction
Bearings are one of the important components of reciprocating compressors. They have the advantages of high operating accuracy and good replaceability. However, due to the influence of alternating loads, machining errors, improper installation, and other factors, the bearing can be damaged during operation, causing the reciprocating compressor to fail to work normally.1 In addition, due to the vibration signal of the reciprocating compressor bearing has strong non-stationary and non-linear characteristics, and early faults are always submerged by strong background noise, this will increase the difficulty of fault diagnosis. Therefore, sub-health monitoring and fault diagnosis of reciprocating compressors have aroused extensive attention and research.2

Entropy is often used in the field of mechanical fault diagnosis. Commonly used entropies include Samp En, FE3 and PE,4,5 etc., but both Samp En and FE are analyzed from a single scale of time. MPE and MSE are
the most common types of Multi-scale Entropy (MSE). However, in MSE, MSE has the disadvantages of slow operation speed and affected by mutation signals when calculating long data; Compared with MSE, the calculation of MPE is simpler, but the disadvantage of MPE is that it does not consider the relationship between signal amplitudes. In 2016, Rostaghi and Azami proposed a new kind of entropy-Dispersion Entropy (DE), which has the advantages of good stability, less affected, and fast calculation speed by mutation signal. Then Zhang et al. and Luo et al. proposed Refined Composite Multi-scale Dispersion Entropy (RCMDE) on this basis, which is superior to other types of multi-scale entropy in terms of calculation deviation and feature extraction. However, there are still some problems that need to be solved in RCMDE. As in the original DE, it does not distinguish the fluctuations of the graph. Second, the mapping method is not necessarily suitable for bearing vibration signals of reciprocating compressors. In RCMDE, the mean-average coarse-grained procedure is usually used. In order to overcome its inherent limitations, this paper proposes to use the root-mean-square coarse-grained procedure instead, that is Generalized Refined Composite Multi-scale Dispersion Entropy (GRCMDE).

The vibration signal of the reciprocating compressor bearing are easily affected by noise signals. When simply calculating the multi-scale entropy to process the vibration signal of the bearing, it will directly affect the result of the extracted fault eigenvectors. Therefore, the original signal needs to be preprocessed before further analysis to reduce or eliminate noise interference. Based on the information entropy energy vector of the EMD method as the fault eigenvector of the reciprocating compressor. Fault diagnosis of reciprocating compressor based on Hermite interpolation LMD method is proposed, etc.. However, in the calculation process of EMD and LMD methods, envelope analysis is used to estimate extreme points, the modal aliasing phenomenon is caused by the uneven distribution of extreme points, and then a series of components with unclear physical meaning are generated, which seriously affects the accuracy of fault diagnosis. In order to avoid the inherent modal aliasing problems of methods such as EMD and LMD, and to better analyze the field measured signals containing complex background noise, this paper introduces a new adaptive submethod, a variation based on Wiener filter Variational Mode Decomposition (VMD). This method has good robustness to noise. It has been successfully used to extract the fault characteristics of non-linear and non-stationary friction influence signals, and to perform instantaneous detection of voice signals and financial market trend analysis. However, in the decomposition process of the VMD method, the number of components \( K \) and the bandwidth parameter \( \alpha \) are the two main parameters. In order to improve the accuracy of VMD decomposition, it is necessary to consider how to adaptively select parameters. The grasshopper optimization algorithm (GOA) is a new swarm intelligence algorithm proposed by Shahrzad Salemi and others in 2017. The advantage of GOA lies in its strong development ability and has proved its superiority in optimization. At the same time, it is also used in other fields such as scheduling and multi-objective optimization. However, it still has some problems, such as low stability, so it is rarely used in mechanical fault diagnosis. Therefore, Curve Adaptation (CA) is proposed to replace the linear adaptation of the grasshopper algorithm parameters, which improves the convergence accuracy. And applied it to propose a curve adaptive grasshopper optimization algorithm (CAGOA) to optimize the parameters of the VMD (CAGOA-VMD) to decompose the fault signal of the reciprocating compressor bearing.

For GRCMDE eigenvalues, the classification sensitivity and classification level are greatly affected by the fault type and data size, and the accuracy and precision of fault classification will decrease, so further optimization of the GRCMDE eigenvalues is needed. Therefore, the recently developed reduction dimension method, Kernel Principal Component Analysis (KCPA) is a good choice. It can well retain the useful information of the original high-dimensional eigenvector and obtain a low-dimensional eigenvector. The eigenvector can best retain the clustering structure of the data, so in order to improve the failure mode recognition rate, this paper uses KPCA to reduce the dimensionality of the eigenvector. Finally, KELM is used to diagnose the fault mode of the reciprocating compressor bearing.

In summary, for the characteristics of fault information in complex mechanical systems that are often implied at different scales, CAGOA-VMD and GRCMDE are combined with the KELM algorithm for fault diagnosis of sliding bearings of reciprocating compressors. It is expected to achieve higher failure recognition rate.

**CAGOA-VMD method**

The VMD method is composed of two steps: building and solving the variational model. The core idea is to apply the new screening iteration principle in the frequency domain to adaptively search for the center frequency and bandwidth of each component of the variational model to achieve signal separation. However, using VMD to process signals requires presetting the number of decomposition components \( K \) and the bandwidth parameter \( \alpha \). The difference in the settings of these two parameters has a greater impact on the final decomposition results. Moreover, the
measured signal is relatively complex and changeable, and it is difficult to determine the two parameters. Therefore, selecting the appropriate $K$ and $\alpha$ is the key to the decomposition result of the VMD algorithm.

The grasshopper optimization algorithm (GOA) is an intelligent optimization algorithm with strong global nonlinear optimization capabilities. The GOA algorithm is used to optimize the number of components $K$ and the bandwidth parameter $\alpha$ of the VMD algorithm, and the objective function is solved by the fitness function. The space performs a global parallel random search to optimize the parameters.

According to the grasshopper optimization algorithm, the population migration and foraging behavior of grasshopper in nature are simulated for mathematical modeling. The mathematical formula is:

$$X_i = \sum_{j=1}^{N} s(\|x_j - x_i\|) \frac{x_j - x_i}{d_{ij}} \cdot g^e + u^e$$  \hspace{1cm} (1)

However, the mathematical model does not solve the optimization problem, so the global and local optimization processes are coordinated by introducing the decreasing coefficient $c$ which determines the size of comfort zone, repulsion zone, and attraction zone. The improved mathematical model is as follows:

$$X_i^d = c \left( \sum_{j=1}^{N} c \frac{|u_b^d - l_b^d|}{2} s(\|x_j^d - x_i^d\|) \frac{x_j^d - x_i^d}{d_{ij}} + T_d \right)$$  \hspace{1cm} (2)

In the equation (2), the influence of gravity is not considered. The expressions of $c$ is as follows:

$$c = c_{\text{max}} - \frac{l c_{\text{max}} - c_{\text{min}}}{M}$$  \hspace{1cm} (3)

Where $l$ represents the current number of iteration, $M$ is the maximum number of iterations, $c_{\text{max}}$ is the maximum value, and $c_{\text{min}}$ is the minimum value.

In equation (3), the parameter $c$ is reduced the attractive, comfortable, and repelling domains among grasshopper. Therefore, $c$ is a key parameter in GOA and has a very large impact on the overall performance of GOA. The parameter $c$ in GOA adopts linear adaptive reduction (equation (3)), which can dynamically adjust the exploration and development in GOA. However, if the parameter $c$ decreases too fast in the early stage of the algorithm, the early exploration of the algorithm will be insufficient, resulting in GOA unable to search Global optimization; the reduction is too slow in the later stage of the algorithm and the later development of the algorithm is insufficient, resulting in the inability of GOA to accurately search in local areas. According to this phenomenon, the parameter $c$ adopts a curve adaptive adjustment strategy to effectively solve this problem. The parameter $c$ based on curve adaptation is as follows:

$$c = \left( 4 \cos \left( \frac{2\pi}{M} \right) + c_{\text{max}} \right) \times \frac{c_{\text{max}} + c_{\text{min}}}{2}$$  \hspace{1cm} (4)

The CAGOA (equation (4)) proposed in this paper replaces the method of linear adaptive update parameter $c$ in the grasshopper optimization algorithm, thereby forming a new CAGOA. The parameter $c$ affects the search range of the grasshopper, and the parameter from large to small corresponds to the transformation of the locust algorithm from global search to local search. The difference between curve adaptation and linear adaptation lies in the downward trend of parameter $c$. Linear adaptation decreases at the same rate, while the curve adaptation adjustment strategy has a small decrease rate at the beginning and end, and the decrease is slow, which helps enhance the grasshopper’s global search at the beginning and local search capabilities in the later stage. In addition, the value of the cosine adaptation parameter $c$ is larger than the value of linear adaptation in the early stage, which can increase the search range of locusts and enhance the ability of grasshopper exploration; the value of the latter parameter is smaller than that of linear adaptation, which can reduce the grasshoppers search scope, enhance grasshopper exploration ability.

This paper uses the CAGOA algorithm to adaptively select the $K$ and $\alpha$ of the VMD algorithm. Therefore, the types of grasshopper populations in the CAGOA algorithm are $K$ and $\alpha$. The dispersion entropy is selected as the fitness function, and the fitness value calculated by each update is compared and updated. The dispersion entropy reflects the complexity of the signal. The more complex the signal, the greater the dispersion entropy, and vice versa. For the fault signal decomposed by VMD, if the IMF component contains more noise, the signal complexity is stronger, and the dispersion entropy value will be larger; if the component contains more fault components, the stronger the regularity of the signal and the weaker the complexity, the smaller the value of dispersion entropy. Once $K$ and $\alpha$ are determined, the component dispersion entropy is obtained after VMD decomposition, and the component with the smallest entropy value is the component with the best fault characteristic information. Therefore, the minimization of dispersion entropy is
used as the fitness value and the target of parameter optimization. Figure 1 is the process of VMD optimized by CAGOA.

**Generalized refined composite multiscale dispersion entropy**

Dispersion entropy (DE) is a nonlinear dynamic method to measure the complexity of time series. Both the MDE method and the RCMDE method are multi-scaled on the original data. MDE divides the data equidistantly and then averages the data. RCMDE is mainly obtained by calculating the mean of all values in each non-overlapping segment. These two methods will cause a lot of loss of potentially useful information.

This subsection proposes the GRCMDE method to solve the above problems, and its main calculation steps are as follows.

(1) For a time series \( u \) with a length of \( L \), \( u = \{u(b), b = 1, 2, ..., L\} \), it is divided into \( \tau \) segments evenly starting from \( u_1 \), the root-mean-square of each segment is calculate, and then the root-mean-square of each segment is formed into a coarse-grained time series. The \( j \)th segment of the \( k \)th coarse-grained time series while scale factor is \( \tau \) is defined as:

\[
x_{k,j}^{(\tau)} = \sqrt{\frac{1}{\tau} \sum_{b=1}^{\tau} \left( u_{b+k-1}(j-1)\tau + k \right)^2}, \quad 1 \leq j \leq [L/\tau], \quad k = 1, 2, ..., \tau
\]  

(5)

(2) By using normal cumulative distribution function \( y_j = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_j} e^{-t^2/2} dt \) calculate the coarse-grained time series \( x_{k,j} = \{x_j, j = 1, 2, ..., N\} \) is mapped to \( y = \{y_j, j = 1, 2, ..., N\} \) which in the range of \([0, 1]\), where \( \mu \) and \( \sigma^2 \) are the expected value and variance respectively. Then use a linear algorithm to assign \( y_j \) to any integer in the range \([1, c]\), so that for each mapped signal:

\[ z^c_j = \text{round}(c \cdot y_j + 0.5) \]  

(6)

(3) The time series \( z^{m,c}_i \) is composed of the embedding dimension \( m \) and the delay time \( d \):

\[
z^{m,c}_i = \{z^c_i, z^c_{i+d}, ..., z^c_{i+(m-1)d}\}, \quad i = 1, 2, ..., N - (m - 1)d
\]  

(7)

Each time series \( z^{m,c}_i \) can be mapped to a dispersion pattern \( \pi_{v_0, v_1, ..., v_n} \), where \( z^c_i = v_0, \ z^c_{i+d} = v_1, \ ... \)
\[ z_i^{m,c} = v_{m-1}. \] The number of scattered patterns assigned to each time series \( z_i^{m,c} \) is equal to \( c^m. \)

(4) Calculate the probability of each coarse-grained dispersion mode \( \pi \), and then find the average of the probabilities of all the dispersion modes.

\[
p(\pi_{v_1 \ldots v_{m-1}}) = \frac{\text{Number}\{i|i \equiv N - (m-1)d, z_i^{m,c}\text{ has type } \pi_{v_1 \ldots v_{m-1}}\}}{N - (m-1)d}
\]

where \( p(\pi_{v_1 \ldots v_{m-1}}) \) stands for the number of dispersion patterns of \( \pi_{v_1 \ldots v_{m-1}} \) assigned to \( z_i^{m,c} \) divided by the total number of embedded signals for embedding dimension \( m \).

(5) Finally, the GRCMDE of \( X = \{x(i), i = 1, 2, ..., N\} \) at the scale \( \tau \) is defined by

\[
\text{GRCMDE}(X, m, c, d, \tau) = - \sum_{\pi = 1}^{c^m} \ln(\bar{p}(\pi_{v_1 \ldots v_{m-1}}))
\]

where \( \bar{p}(\pi_{v_1 \ldots v_{m-1}}) = \frac{1}{\tau} \sum_{k=1}^{\tau} p^{(\pi)}(\pi_{v_1 \ldots v_{m-1}}) \) is the average value of the probability of the dispersion patterns \( \pi \) of the coarse-grained sequence.

**Comparison between GRCMDE and GMDE simulation**

In the GRCMDE algorithm, the values of three parameters need to be set in advance, class \( c \), embedding dimension \( m \), and time delay \( d \). According to the recommended settings of Azami et al., \( c = 5, m = 3, \) and \( d = 1 \). Select length \( N = 2000, 4000, 6000, 8000, \) and 10,000 of Gaussian white noise and 1/f noise signals to analyze and compare GRCMDE and GMDE respectively, as shown in Figure 2.

Comparing the (a) and (b) in Figure 2, it can be found that the GMDE and GRCMDE curves for Gaussian white noise decrease almost linearly with the increase of the scale factor \( \tau \). So only choosing a smaller coarse-grained sequence of Gaussian white noise will contain more important feature information. For the GMDE curve, when the \( \tau \) is large, the entropy value changes more greatly, and the GRCMDE curve changes more smoothly. Therefore, GRCMDE can obtain a more stable entropy value than GMDE. Moreover, for Gaussian white noises of different lengths, the entropy value of the GMDE or GRCMDE curve changes little.

Comparing the (c) and (d) in Figure 2, it can be found that the GMDE and GRCMDE curves for 1/f noise first accelerately decrease and then tend to a constant value with the increase of the \( \tau \). Similarly, the stability of the GRCMDE entropy value better than GMDE. Moreover, the entropy value of GRCMDE (or GMDE) of 1/f noise on most different scale factors is greater than that of Gaussian white noise, which shows that the characteristic information and structure complexity of 1/f noise are higher than Gaussian white noise.

Next, the GMDE and GRCMDE with lengths of 1000 and 200 sets of Gaussian white noise and 1/f noise are used to study their statistical stability. The result is shown in Figure 3. In terms of the average value, regardless of whether the GMDE method or the GRCMDE method is used, the average value of the Gaussian white noise and 1/f noise of the 200 sets of data is almost equal. In terms of standard deviation, the GRCMDE method is smaller than the GMDE method. By comparison, the GRCMDE method is higher than the GMDE method in statistical stability, so the GRCMDE method is more accurate in extracting noise features.

Then, for the five states of the bearing: Normal state; First-stage connecting rod large-end bearing bush clearance (FCB); First-stage connecting rod small-end bearing bush clearance (FCS); Second-stage connecting rod large-end bearing bush clearance (SCB); and Second-stage connecting rod small-end bearing bush clearance (SCS); 20 sets of two full-period data are selected, use GRCMDE and GMDE analysis, where \( m = 3; c = 4; \tau = 1 \), the results are shown in Figures 4 and 5.

With the increase of the \( \tau \), the GRCMDE and GMDE curves of the reciprocating compressor bearing in the normal clearance state gradually increase, but when \( \tau \geq 8 \), the GRCMDE and GMDE curves increase slowly. The GRCMDE and GMDE curves of the four fault states with large bearing clearances of reciprocating compressors gradually decrease with the increase of the \( \tau \). When \( \tau \geq 10 \), the GRCMDE and GMDE curves decrease slowly and the curves become more stable. Specifically, the GMDE entropy fluctuates in a larger range with the increase of the \( \tau \), while the GRCMDE curve gradually changes with the increase of the \( \tau \). The results also show that GRCMDE is better than GMDE in terms of entropy consistency and stability.

Twenty sets of data are randomly selected, and each set contains two full-period data points, and statistical stability is performed on the bearing state data of five
Figure 2. GRCMDE and GMDE curves of Gaussian white noise and $1/f$ noise: (a) GRCMDE of Gaussian white noise, (b) GMDE of Gaussian white noise, (c) GRCMDE of $1/f$ noise, and (d) GMDE of $1/f$ noise.

Figure 3. Comparative analysis of the GRCMDE and the GMDE: (a) Gaussian white noise signal and (b) $1/f$ noise signal.
reciprocating compressors, as shown in Figure 5. The average values of entropy calculated by GRCMDE and GMDE are almost equal. Especially with the increase of the \( t \), the fluctuation increases. In addition, the standard deviation of GRCMDE is smaller than the standard deviation of GMDE. Taking into account comprehensively, the GRCMDE method provides a more accurate entropy estimate.

**Fault diagnosis method of reciprocating compressor based on CAGOA-VMD and GRCMDE**

Aiming at the characteristics of complex, non-stationary, nonlinear and feature coupling of the reciprocating compressor bearing vibration signal, a feature extraction method based on CAGOA-VMD and GRCMDE is proposed. The fault diagnosis process of the reciprocating compressor bearing is shown in Figure 6. Specific steps are as follows.

1. Use the curve adaptive grasshopper algorithm to optimize the parameters of the variational modal decomposition, obtain the best optimized parameter group \([K_0, a_0]\), set it as the VMD decomposition parameter, decompose the reciprocating compressor bearing fault signal, and get \( K_0 \) IMF components;

2. Use kurtosis and correlation coefficient criteria to select a group of \( K_0 \) IMF components, select
the IMF components that can significantly represent the fault characteristics, and reconstruct the fault signal;

3) Perform GRCMDE analysis on the reconstructed fault signal, quantify the GRCMDE entropy value of the reconstructed fault signal, form the reciprocating compressor bearing fault eigenvector, and use KPCA to select the input eigenvector;

4) The kernel extreme learning machine (KELM) is used to train and test the bearing fault eigenvectors to identify different fault types and sub-health conditions.

The reciprocating compressor has been widely applied in the petroleum and chemical industry, and their operating state and security are thought to be challenging research subjects. The experimental data in this paper comes from the 2D12-70/0-13 two-stage double-acting reciprocating compressor shown in Figure 7. Its main design parameters: shaft power is 500 kW, piston stroke is 240 mm, crank speed is 496 rpm. Bearings are often subjected to alternating loads during the working process of the reciprocating compressor. Long-term service will inevitably cause bearing failures, which will cause the reciprocating compressor to stop and cause economic losses. Therefore, it highlights the necessity of condition monitoring and fault diagnosis for reciprocating compressors. Photos of the test bench and the schematic diagram of the reciprocating compressor transmission mechanism are displayed in Figures 7 and 8, respectively.

As shown in Figure 8, the connecting rod of the 2D12 reciprocating compressor is connected with the crankshaft and the crosshead by sliding bearings. In addition to its own assembly clearance, it is continuously affected by large alternating loads and friction and wear, resulting in increased local wear of the bearing. As a result, the bearing clearance is too large, which in turn causes abnormal vibration at the joints of the piston, cylinder, and crankshaft, which affects the operation of the equipment.

Bearing bushes will wear out during long-term work, and if not detected in time, huge economic losses will be caused. This paper mainly takes the early weak fault signals in five states of the bearing as the research object to verify the effectiveness of the proposed method, the connecting rod and its components are shown in Figure 9. After working for a while, the normal bearing bush in Figure 9(b) will wear out and then becomes the fault state in Figure 9(c) resulting in a clearance between the journal and bearing bush, and the machine will produce abnormal sound and vibration. When the bearing clearance of the reciprocating compressor is too large, the vibration signal of the cylinder surface is mixed with the non-stationary non-linear signal of complex frequency excited by different excitation sources, and coupled with the background noise. Therefore, during the test, the sensors were installed on the upper end of the crosshead slider and the top of the crankcase as shown by the red triangle in Figure 10. As shown in Figure 7(b) and (c), the integrated circuit piezoelectric (ICP) accelerometer (CTI010LC) is installed in the red rectangular frame, the sensitivity is 100 mv/g, the measurement range is ±50 g, and the frequency range is 0.5–5 kHz. The sampling frequency and time respectively are 50 kHz and 4 s. The program written by MATLAB software analyzes the collected vibration signal. Figure 11 shows the time domain waveforms of the reciprocating compressor in two periods under five bearing states.

In this paper, first, the CAGOA algorithm is used to determine the number of components \( K \) and the bandwidth parameter \( \alpha \) of the VMD decomposition. Through the CAGOA algorithm to optimize the VMD decomposition parameters of the five bearing vibration
signals, the best optimized parameter combination can be obtained as shown in Table 1.

According to the two parameter data of VMD in Table 1, the five state signals of bearing clearance were decomposed, and based on the kurtosis value reflecting the impact of the fault and the correlation coefficient reflecting the correlation analysis between the component signal and the original signal, that is,

**Figure 7.** Two-stage double-acting reciprocating compressor of 2D12 type: (a) photos of the test bench, (b) sensor on the top of the crankcase, and (c) sensor on the top of the crosshead guide surface.

**Figure 8.** Schematic diagram of the reciprocating compressor transmission mechanism.

**Figure 9.** Reciprocating compressor connecting rod: (a) connecting rod, (b) normal state, (c) big end fault bearing, and (d) normal bearing bush.
the kurtosis-correlation coefficient criterion. Select the kurtosis value and the correlation coefficient to have the same weight to calculate the kurtosis-correlation coefficient value $Kr$ of each IMF component as shown in Table 2. Since the greater the kurtosis value, the more fault components contained in the signal, and the greater the absolute value of the correlation coefficient, the higher the degree of linear correlation between the two samples. Therefore, the first three IMF components with larger $Kr$ should be selected for signal reconstruction to perform signal analysis in each state.

It can be found from Table 2, when the bearing is in a normal state, the kurtosis-correlation coefficient value of the IMF component is about less than 2, which is

![Figure 10. The structural draw of reciprocating compressor and measuring points.](image1)

![Figure 11. Time-domain waveform diagram of reciprocating compressor under five bearing states.](image2)

| Fault type | Component numbers $K$ | Bandwidth parameter $\alpha$ |
|------------|------------------------|-----------------------------|
| Normal state | 4 | 1803 |
| FCB | 4 | 1725 |
| FCS | 5 | 1959 |
| SCB | 4 | 1612 |
| SCS | 4 | 1374 |

| Fault type | Kurtosis-correlation coefficient value of each IMF component |
|------------|------------------------------------------------------------|
| IMF1 | IMF2 | IMF3 | IMF4 | IMF5 |
| Normal state | 1.9099 | 1.3823 | 1.6958 | 1.8456 |
| FCB | 1.9106 | 3.8496 | 11.9532 | 10.7154 |
| FCS | 2.0105 | 2.7720 | 3.2771 | 3.1851 | 2.3977 |
| SCB | 2.0555 | 3.9812 | 4.8779 | 12.2285 |
| SCS | 2.0201 | 2.7055 | 3.1178 | 3.2573 |

Table 1. VMD parameter setting of bearing state vibration signal.

Table 2. The kurtosis-correlation coefficient value of each IMF component of the vibration signal of the five bearing states.
close to the normal distribution state. When the bearing is in a clearance failure state, the kurtosis-correlation coefficient value of some IMF components increases significantly. Therefore, this paper selects the first three IMF components with the largest kurtosis-correlation coefficient values in various signal states for signal reconstruction. The specific selection is shown in Table 2 for the underlined components. The reconstructed signal is shown in Figure 12.

Calculate the GRCMDE values of the reconstructed vibration signals under different bearing states, where \( m = 3 \); \( c = 4 \); \( t = 1 \), and the maximum scale factor is 20, as shown in Figure 13. In the normal bearing clearance state, GRCMDE gradually increases with the increase of the scale factor, and finally stabilizes. The other four fault signals gradually decrease with the increase of the scale factor, and tend to be stable. When \( \tau > 3 \), the GRCMDE values of these five bearing clearance states can be well distinguished. Compared with the GRCMDE value of the original signal, the reconstructed signal has good distinguishability. Considering the above factors comprehensively, in order to make the method proposed in this paper have a better fault diagnosis effect, this paper selects the multi-scale dispersion entropy value of 17 scales with \( \tau > 3 \) as the state fault eigenvector.

The KPCA algorithm with non-linear data processing capabilities is used to select the more prominent features in the feature vectors obtained in GRCMDE. The cumulative contribution rate is set to 85%, so three main eigenvalues are selected from the original 17 eigenvalues.

Through the above analysis, the method proposed in this paper is used to analyze the five state signals of the bearing to construct the reciprocating compressor bearing state eigenvector set. Corresponding to each state, 150 sets of feature vectors are extracted, 100 sets of training sets, and 50 sets of test sets are randomly selected from them, and then input the features into KELM to achieve fault classification, as shown in Table 3. The overall diagnosis rate based on CAGOAVMD and GRCMDE in this article is as high as 99.2%.

In order to verify that the method used in this article has better superiority in the fault diagnosis of reciprocating compressor bearings, respectively perform CAGOAVMD and GRCMDE, VMD and GRCMDE methods, VMD and RCMDE methods, and VMD and GMDE methods of feature vector diagnosis under the same data sample and identification method as shown in Table 3. It can be seen from the recognition results that, firstly, under the same GRCMDE feature extraction method, the recognition accuracy of the VMD method with parameters optimized by CAGOA is higher than the VMD method without optimization. Secondly, under the same VMD method, the recognition accuracy of GRCMDE is much higher than that of GMDE method, and higher than that of the currently commonly used RCMDE method. Therefore, the two improved methods proposed in this paper, namely
CAGOA-VMD and GRCMDE, are superior to the other three methods in both the overall recognition rate and the single fault recognition rate in the fault diagnosis of reciprocating compressor bearings, and the diagnosis effect is the best. However, the main disadvantage of this method is that it takes more time due to the complexity of the optimization algorithm.

Conclusion
In order to accurately and efficiently identify different fault states of the reciprocating compressor bearing and prevent the occurrence of safety accidents. This paper proposes a new method for reciprocating compressor bearing fault diagnosis based on CAGOA-VMD and GRCMDE. Experimental analysis proves the great performance of the proposed method. The main work and innovations of this article are as follows:

(1) The coarse-graining method of MDE is improved and GRCMDE is proposed, which has better fault features extraction ability and stability.

(2) CAGOA is used to optimize the parameters of the VMD method to obtain the best optimized parameter group \([K_0, \alpha_0]\), which solves the defect that the parameters need to be set manually.

(3) Combining CAGOA-VMD and GRCMDE, a new reciprocating compressor bearing fault diagnosis method is proposed. Experimental analysis shows that this method has the best fault diagnosis capability compared with other methods.

Although the proposed method can improve the accuracy of fault diagnosis, there are still some limitations. The Parameter optimization process is more time-consuming than manual setting. In future work, we will focus on researching signal denoising methods and reduce the computing time, and try to apply similar methods to fault diagnosis of other mechanical equipment, so as to expand the application scope of similar methods.

Table 3. Recognition result of bearing fault state.

| Feature extraction method | Recognition accuracy (%) of bearing state | Total accuracy (%) |
|---------------------------|-------------------------------------------|--------------------|
|                           | Normal state | FCB | FCS | SCB | SCS |                |
| CAGOA-VMD and GRCMDE      | 100 | 99 | 98 | 100 | 99 | 99.2 |
| VMD and GRCMDE            | 98 | 98 | 98 | 98 | 98 | 98   |
| VMD and RCMDE             | 97 | 96 | 95 | 96 | 97 | 96.2 |
| VMD and GMDE              | 90 | 92 | 86 | 88 | 90 | 89.2 |

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the Heilongjiang Natural Science Foundation under Grant LH2021E021, in part by the National Natural Science Foundation of China under Grant 51505079.

ORCID iD
Meiping Song https://orcid.org/0000-0002-7290-4127

References
1. Wang J, Chen X, Zhao H, et al. Fault feature extraction for reciprocating compressors based on underdetermined blind source separation. *Entropy* 2021; 23: 1217.
2. Cerrada M, Macancela JC, Cabrera D, et al. Reciprocating compressor multi-fault classification using symbolic dynamics and complex correlation measure. *Appl Sci* 2020; 10: 2512.
3. Li Y and Zhou X. Rolling bearing fault diagnosis based on HVD algorithm and sample entropy. *J Comput Methods Sci Eng* 2019; 19: 331–340.
4. Zhang X, Liang Y, Zhou J, et al. A novel bearing fault diagnosis model integrated permutation entropy ensemble empirical mode decomposition and optimized SVM. *Measurement* 2015; 69: 164–179.
5. Li Y, Liang X, Wei Y, et al. A method based on refined composite multi-scale symbolic dynamic entropy and ISVM-BT for rotating machinery fault diagnosis. *Neurocomputing* 2018; 315: 246–260.
6. Malhotra A, Singh Minhas A, Singh S, et al. Bearing fault diagnosis based on flexible analytical wavelet transform and fuzzy entropy approach. *Mater Today Proc* 2021; 43: 629–635.
7. Rostaghi M and Azami H. Dispersion entropy: a measure for time-series analysis. *IEEE Signal Process Lett* 2016; 23: 610–614.
8. Zhang X, Zhao J, Teng H, et al. A novel faults detection method for rolling bearing based on RCMDE and ISVM. *J Vibroengineering* 2019; 21: 2148–2158.
9. Luo S, Yang W and Luo Y. Fault diagnosis of a rolling bearing based on adaptive sparest narrow-band decomposition and refined composite multiscale dispersion entropy.. *Entropy* 2020; 22: 375.
10. Zheng J and Pan H. Use of generalized refined composite multiscale fractional dispersion entropy to diagnose the faults of rolling bearing. *Nonlinear Dyn* 2020; 101: 1417–1440.
11. Rajesh Mathivanan N, Mahesh BS and Anup Shetty H. An experimental investigation on the process parameters influencing machining forces during milling of carbon and glass fiber laminates. *Measurement* 2016; 91: 39–45.
12. Wang J, Du G, Zhu Z, et al. Fault diagnosis of rotating machines based on the EMD manifold. *Mech Syst Signal Process* 2020; 135: 106443.
13. Jenarthanan M and Jeyapaul R. Evaluation of machinability index on milling of GFRP composites with different fibre orientations using solid carbide endmill with modified helix angles. *Int J Eng Sci Technol* 2014; 6: 1–10.
14. Li X, Ma Z, Kang D, et al. Fault diagnosis for rolling bearing based on VMD-FRFT. *Measurement* 2020; 155: 107554.
15. Zeng W, Yuan J, Yuan C, et al. A new approach for the detection of abnormal heart sound signals using TQWT, VMD and neural networks. *Artif Intell Rev* 2021; 54: 1613–1647.
16. Jnlu G. Research on the forecast of CSI 300 Index based on VMD-EEMD-LSTM model. *Mod Finance (J Tianjin Univ Finance Econ)* 2020; 40: 31–44.
17. Saremi S, Mirjalili S and Lewis A. Grasshopper optimization algorithm: Theory and application. *Adv Eng Softw* 2017; 105: 30–47.
18. Hamad A, Houssein EH, Hassani A, et al. Hybrid grasshopper optimization algorithm and support vector machines for automatic seizure detection in EEG signals. In: The international conference on advanced machine learning technologies and applications, Cairo, Egypt, 22–24 February 2018, vol. 723, pp. 82–91. Egypt: AMLTA.
19. Mirjalili SZ, Mirjalili S, Saremi S, et al. Grasshopper optimization algorithm for multi-objective optimization problems. *Appl Intell* 2018; 48: 805–820.
20. Zhang X, Miao Q, Zhang H, et al. A parameter-adaptive VMD method based on grasshopper optimization algorithm to analyze vibration signals from rotating machinery. *Mech Syst Signal Process* 2018; 108: 58–72.
21. Arora S and Anand P. Chaotic grasshopper optimization algorithm for global optimization. *Neural Comput Appl* 2019; 31: 4385–4405.
22. Ling L and Qing H. An feature selection method based on improved grasshopper optimization algorithm. *J Nanjing Univ (Nat Sci)* 2020; 56: 41–50.
23. Hu X, Xiao Z, Liu D, et al. KPCA and AE based local-global feature extraction method for vibration signals of rotating machinery. *Math Probl Eng* 2020; 2020: 1–17.
24. Chen P, Zhao X and Zhu Q. A novel classification method based on ICGOA-KELM for fault diagnosis of rolling bearing. *Appl Intell* 2020; 50: 2833–2847.
25. Li Y, Xu M, Wei Y, et al. A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree. *Measurement* 2016; 77: 80–94.
26. Azami H, Rostaghi M, Abasolo D, et al. Refined composite multiscale dispersion entropy and its application to biomedical signals. *IEEE Trans Biomed Eng* 2017; 64: 2872–2879.