# Fault diagnosis using an improved fusion feature based on manifold learning for wind turbine transmission system

**Ping Ma[1], Hongli Zhang[2], Wenhui Fan[3], Cong Wang[4]**

[1, 2, 4]College of Electrical Engineering, Xinjiang University, Urumqi, China
[3]Department of Automation, Tsinghua University, Beijing, China
[1]Corresponding author
**E-mail:** [1]*694073078@qq.com*, [2]*xjdq_mp@foxmail.com*, [3]*maping_0000@sina.com*, [4]*2896438968@qq.com*

Check for updates

**Abstract.** In this paper, a novel fault diagnosis method based on vibration signal analysis is proposed for fault diagnosis of bearings and gears. Firstly, the ensemble empirical mode decomposition (EEMD) is used to decompose the vibration signal into several subsequences, and a multi-entropy (ME) is proposed to make up the fusion features of the vibration signal. Secondly, an improved manifold learning algorithm, local and global preserving embedding (LGPE), is applied to compress the high-dimensional fusion feature set into a two-dimension feature set. Finally, according to the clustering accuracy of different feature set, the fault classification and diagnosis can be performed in the reduced two-dimension space. The performance of the proposed technique is tested on the fault of wind turbine transmission system. The application results indicate that the proposed method can achieve high accuracy of fault diagnosis.

**Keywords:** fault detection and diagnosis, ensemble empirical mode decomposition, multi-entropy, local and global preserving embedding.

## 1. Introduction

Wind energy as a new clean energy source is the fastest-growing energy source in the world, this trend should endure for some time. However, wind turbines are long-term operated in harsher environments, have relatively higher failure rates. Fig. 1 shows the percentage of downtime with different failure per year at the Dutch wind farm during 3 years operation [1]. It can be seen that the primary causes of failure are damage to the gearbox, generator and faults with the control and pitch systems. The bearings and gears are important parts of those systems, which faults will inevitably cause a long downtimes and increase operating costs. Therefore, proposing a timely and accurate diagnosis method to detect those faults are extremely available, which will reduces energy losses, improves productively and increases safety of such systems [2, 3].

The techniques based on vibration signal analysis are the most popular and useful approaches applied in fault diagnosis of rotating machinery [4-6]. In general, the processing of fault diagnosis can be divided into three steps: data collection, feature extraction and classification. Feature extraction is a key step, which can extract more useful information of the vibration signals from their original measured space for fault diagnosis. Some self-adaptive time-frequency analysis methods have been proposed for fault diagnosis. Empirical mode decomposition (EMD) [7], which were carried out to decompose vibration signals into the sum of intrinsic mode functions (IMF) given different frequencies, and some IMFs containing fault feature frequencies are selected to reconstruct a new signal for outstanding fault feature for further analysis [8, 9]. But the decomposed process of EMD is sensitive to noise and has exist mode mixing problem. Ensemble empirical mode decomposition (EEMD) is a newly improved version of EMD, which can eliminate the mode mixing problem of EMD automatically [10]. This approach has been applied in signal analysis of rotating machinery [11-14]. In conclusion, the self-adaptive time-frequency analysis method EEMD can extract sensitive features and improve the fault diagnosis accuracy.
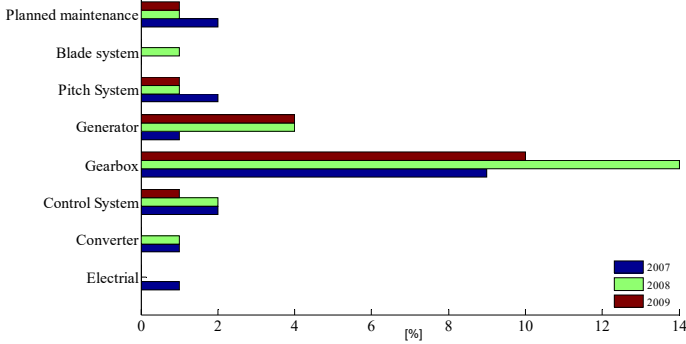
**Fig. 1.** Downtimes for the Dutch wind farm Egmond aan Zee

The vibration signals collected from wind turbine transmission systems are complex and non-stationary, which are submerged in noise under the variable speeds conditions. Therefore, the fault features from the original signal cannot be comprehensively extracted using single or single-domain processing methods [15, 16]. Therefore, the mix-domain features fusion is proposed to construct high-dimensional feature set. However, the high-dimensional feature set of mixed-domain inevitably contains redundant and disturbed information, and the accuracy of fault diagnosis can be reduced using the high dimensional feature set for pattern recognition. Using an appropriate dimensionality reduction method to extract the main eigenvectors with low dimension, high sensitivity and good clustering characteristics from the high-dimensional feature set for fault recognition is necessary.

Manifold learning [17-20], an effective nonlinear dimension reduction method has gained attention in various research fields. The idea of manifold learning is to project the original high dimensional data into a lower dimension feature space with the local neighborhood structure preserved. This technique including several mainstream algorithms such as locally linear embedding (LLE) [18], locality preserving projection (LPP) and kernel locality preserving projection (KLPP) [20]. Manifold learning is widely used in the filed of image recognition and data classification [21, 22], and there are few researches in fault diagnosis [23, 24]. Due to the advantages of the time-frequency analysis and manifold learning techniques in dealing with non-linear signal, some research works combined fault diagnosis models with time-frequency analysis techniques and dimensionality reduction techniques. Su et al. [25] combined EMD, incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier to build a new method for gearbox fault diagnosis. Ding et al. [26] presented a fusion feature extraction method on rolling bearings based on wavelet packet transform and LPP. The classification result shows that this method can enhance the discrimination between all fault classes for fault classification. Huang et al. [27] used a new technique for dimensionality reduction called the discriminate diffusion maps analysis, compared with other manifold learning algorithms, which has high classification accuracy of the bearings faults.

In the above literatures, at the process of feature extraction, the dimensionality reduction methods of manifold learning, which embed the original high-dimensional data into a lower dimension feature space with considering the local nonlinear characteristics preserved. But undeniably, in the process of reducing dimensionality, in order to extract more abundant information, it is necessary to preserve the intrinsic structure and extract the structure of the high-dimensional set.

In this paper, a novel fault diagnosis model based on fusion feature and manifold learning is proposed. Firstly, EEMD is used to decompose the vibration signals into a set of intrinsic mode functions (IMFs) with different characteristics scales, and several IMFs that containing more sensitive information are selected. Secondly, envelope entropy, permutation entropy and energy entropy of those selected IMFs are extracted as representative fusion features. Thirdly, local and

global preserving embedding (LGPE), is applied to extract the main eigenvectors from the high-dimensional fusion feature set. This method considers the local nonlinear characteristics and global external structure of the fusion feature set. The feature space dimension is optimally reduce to a low dimensional space by using LGPE, achieving the classification and identification of the fault. The proposed method is verified on the fault diagnosis of the bearing's and gearbox's signals.

The main contributions of the proposed method are presented as follows:

1) A fusion feature set is established based on EEMD and Mutli-entropy, in which the fault features can be comprehensively extracted from mixed time-domain signals.

2) A LGPE method is proposed to extract fault features from the high-dimensional feature set, which makes the fault features in low dimensional space has high sensitivity and good clustering characteristics.

3) A novel fault diagnosis model based on EEMD-ME and LGPE is performed to diagnose the faults of bearings and gears, which accurately achieves the identification and classification of different faults.

The organization of the rest of the paper is as follows. In Section 2, the EEMD, ME and LGPE are presented. Section 3 describes the proposed model for fault diagnosis. In Section 4, two fault diagnosis experiments are carried out to verify our proposed method, and the experimental result is discussed. The conclusions are appeared in Section 5.

## 2. The proposed fault diagnosis model

Fig. 2 shows the flow chart of the proposed fault diagnosis method. The relative methods including EEMD, ME and LGPE are presented in this section.
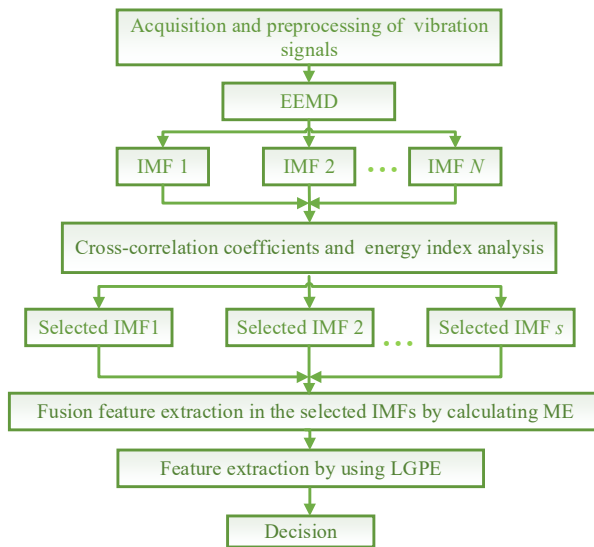


**Fig. 2.** The fault diagnosis process of the proposed method

### 2.1. Ensemble empirical mode decomposition (EEMD)

EEMD is a non-linear multi-resolution self-adaptive decomposition technique, which can adaptively decompose a complex signal into a set of IMF. The vibration signal of the rotating machinery is nonlinear, non-stationary and submerged in heavy noise, which makes it difficult to detect faults using the original signal. Therefore, EEMD is proposed to decompose the vibration signal into several IMFs with different frequencies for further analysis.

The principle of the EEMD is simple: in the process of decomposition, the frequency scales

are natural separated by adding white noise in the whole time-frequency space uniformly, which can reduce the occurrence of mode mixing. The decomposition steps of EEMD is as follows:

Step 1: Add a white noise series $n_i(t)$ with zero mean and equal variance to the original vibration signal $x(t)$:

$$x_i(t) = x(t) + n_i(t),$$ (1)

where $n_i(t)$ represents the $i$th added white noise series, and $x_i(t)$ denotes the noise-added signal of the $i$th trial, while $i = 1, 2, …, M$, $M$ is the number of ensemble.

Step 2: Each noise-added signal $x_i(t)$ is decomposes into several IMFs using EMD:

$$x_i(t) = \sum_{s=1}^{S} c_{is}(t) + r_{iS}(t),$$ (2)

where $S$ represents the number of IMF, $s = 1, 2, …, S$. $c_{is}(t)$ denotes the IMFs $(c_{i1}, c_{i2}, …, c_{iS})$ with different frequency bands, and $r_{iS}(t)$ is the residue of $x_i(t)$.

Step 3: Repeat step 1 and step 2 with M times, the different white noise series is added to the signal $x(t)$ to obtain an ensemble of IMFs $[\{c_{1s}(t), c_{2s}(t), …, c_{Ms}(t)\}]$.

Step 4: Obtain the ensemble means of the corresponding IMFs as the final IMFs:

$$c_s(t) = \frac{1}{M} \sum_{i=1}^{M} c_{is}(t),$$ (3)

where $c_s(t)$ represents the $i$th IMF decomposed by EEMD, while $s = 1, 2, …, S$.

In order select the IMF that containing useful feature information for further analysis, the cross-correlation coefficient $\rho$ and energy index $E$ are introduced to eliminate illusive IMF. Cross-correlation coefficient $a$ is common used in signals analysis and is no longer discussed in here. For a IMF obtained by EEMD, the energy index $E(s)$ is defined as follows:

$$E(s) = \sqrt{\frac{\sum_{t=1}^{N}|c_s(t)|^2}{\sum_{t=1}^{N}|x(t)|^2}},$$ (4)

where $N$ is the length of the signal, $c_s(t)$ denotes the $s$th IMF, and $E(s)$ represents the index of energy between the $s$th IMF $c_s(t)$ and the original signal $x(t)$.

## 2.2. Mutil-entropy

The entropy-based methods, such as envelope entropy, energy entropy and permutation entropy [28-30], have been applied in fault diagnosis. In this section, envelope entropy, energy entropy and permutation entropy of the selected IMFs obtained from section 2.1 are extracted as the fusion features of fault signal.

Entropy can identify nonlinear parameters and present the information of the signal, and envelope entropy can reflect the sparseness of the original signal. The envelope entropy $H_1(s)$ of a IMF $c_s(t)$ $(t = 1,2, …, N)$ can be expressed as:

$$\begin{cases} H_1(s) = -\sum_{t=1}^{N} p_t \lg p_t, \\ p_t = \dfrac{a(t)}{\sum_{t=1}^{N} a(t),} \end{cases}$$ (5)

where $a(t)$ is the envelope signal that obtained by the Hilbert demodulation of signal $c_s(t)$, $p_t$ is

the normalized form of signal $c_s(t)$.

In the same way, the energy entropy of the IMF $c_s(t)$, which can be defined as:

$$H_2(s) = -\sum_{s=1}^{S} p_s \log p_s, \tag{6}$$

where $p_s = E(s)/E$ is the proportion of the energy of the $s$th IMF in the whole signal energies.

Permutation entropy (PE) is used to analyze the data complexity. For a IMF $c_s(t)$, constructing an embedded $d$-dimensional delay embedding matrix $C_d(t) = \{x_{t+\sigma}, x_{t+2\sigma}, \cdots, x_{t+d\sigma}\}$, and arranging each vector of $C_d$ to an increasing order $C_i(t) = \{c_{i+j_1} \leq c_{i+j_2} \leq \cdots \leq c_{i+j_d}\}$. Letting $\pi_i = (j_1, j_2, \ldots, j_d)$, which is one of the $d!$ permutations of $d$ distinct symbols. Then the permutation entropy of IMF $c_s(t)$, with the probability distribution function $p(\pi_i)$ is defined as:

$$H_3(s) = -\sum_{i}^{n-\sigma d+1} p(\pi_i) \ln\big(p(\pi_i)\big). \tag{7}$$

## 2.3. Local and global preserving embedding (LGPE) algorithm

The generic reduction problem of feature space dimension is described as follows: let a $n$-dimension points $X = [x_1, x_2, \ldots, x_n] \in R^n$, which can be transformed into an $d$-dimension $Y = [y_1, y_2, \ldots, y_d] \in R^d$, $(d < n)$, where $y_i = W^T x_i$, and $W$ is an transformation matrix. In this section, a novel dimensionality reduction method, LGPE is used to project a manifold in high-dimensional space $R^n$ to a low-dimensional space $R^d$ while preserving the local neighborhood and global structure of the dataset.

The objective function of the LGPE can be divided into two parts: the local nonlinear characteristics preservation objective function and the global variance maximum objective function. The local nonlinear characteristics preservation objective function makes the low-dimension feature space has the similar neighborhood structure with high-dimensional feature space. The global variance maximum objective function extracts the maximize variance of the data during the dimensionality reduction process.

### 2.3.1. The local nonlinear characteristics preservation objective function

Given a set of data $X = [x_1, x_2, \ldots, x_n] \in R^n$, $x_i$ is a $n$-dimensional feature vector. Firstly, we use a possibly nonlinear function $\phi$ to map the data into a high-dimension feature space $H$: $\phi(X) = [\phi(x_1), \phi(x_2), \ldots, \phi(x_n)]$. Inspired by the idea of kernel locality preserving projection (KLPP), we seek a projecting transformation $W$, which can preserve the local nonlinear structure of the data $\phi(x)$ by minimizing the sum of the weighted distance of samples. The minimization problem of local nonlinear characteristics preservation objective function can be expressed as:

$$J_{local}(W) = \min_{W} \sum_{i,j=1}^{N} \|y_i - y_j\|^2 S_{i,j}, \tag{8}$$

where $y_i = (\phi(x))^T W$ is the low-dimension projection of $\phi(x)$ onto $W$. $S$ represents a weight matrix, which is constructed through the nearest-neighbor graph. It is defined as follows:

$$S_{i,j} = \begin{cases} \dfrac{\exp\left(-\|\phi(x_i) - \phi(x_j)\|^2\right)}{t} \phi(x_i), & \phi(x_j) \text{ as neighbors}, \\ 0, & \text{other}, \end{cases} \tag{9}$$

where $t$ is a suitable constant. $S_{i,j}$ denotes the relationship of $\phi(x_i)$ and $\phi(x_j)$. The objective

Eq. (7) can be transformed as:

$$J_{local}(W) = \min_W \sum_{i,j=1}^N \left\| y_i - y_j \right\|^2 S_{i,j} = \min_W \sum_{i,j=1}^N \left\| (\phi(x_i))^T W - (\phi(x_j))^T W \right\|^2 S_{i,j}$$
$$= \min_W \{ W^T (\phi(X))^T (D - S)(\phi(X)) W \},$$

(10)

where $D_{ii} = \sum_j S_{i,j}$ is a diagonal matrix. Because the projecting transformation $W$ must be in the span of $\phi(x_1), \phi(x_2), \dots, \phi(x_n)$, there have a coefficient vector $A = (a_1, a_2, \dots, a_n)^T$ to satisfy the equation $W = \sum_{j=1}^n a_j \phi(x_j) = \phi(X)^T A$. Then the local nonlinear characteristics preservation objective function can be expressed as:

$$J_{local}(A) = \min_A \{ A^T \phi(X)(\phi(X))^T (D - S)\phi(X)(\phi(X))^T A \}$$
$$= \min_A \{ A^T \phi(X)(\phi(X))^T L \phi(X)(\phi(X))^T A \},$$

(11)

where $L = D - S$ is a Laplacian matrix. In order to solve this nonlinear problem, a positive definite and symmetric kernel matrix $K(i,j) = \phi(x_i) \cdot \phi(x_j)^T$ is introduced to the Eq. (11). The objective Eq. (11) can be calculated as:

$$J_{local}(A) = \min_A AKLKA = \min_\alpha A^T L_l A, \quad (L_1 = KLK).$$

(12)

The local nonlinear structure of the high-dimension dataset is preserved by keeping the nearest neighbor relation of the dataset in the kernel space. It is obvious that, the solution of local nonlinear characteristics preservation objective function will keep the local structural characteristics of dataset in the process of dimensionality reduction.

### 2.3.2. The global variance maximum objective function

Given a set of data $X = [x_1, x_2, \dots, x_n] \in R^n$, and using a possibly nonlinear function $\phi$ to map the data into a high-dimension feature space $H: \phi(X) = [\phi(x_1), \phi(x_2), \dots, \phi(x_n)]$. Inspired by the idea of kernel principal component analysis (KPCA), the global variance maximum objective function $J_{global}(W)$ is defined as: seeking a projecting transformation $W$, which makes the matrix $y_i = (\phi(x_i))^T W$ after projection preserves the maximum variance information for the high-dimension data $\phi(x)$. Then the optimization task can be expressed as:

$$J_{global}(W) = \max_W \sum_{i=1}^n y_i^2 = \max_W \sum_{i=1}^n ((\phi(x_i))^T W)^2,$$
$$s.t. \ W^T W = 1.$$

(13)

It is well known that exist a coefficient vector $A = (a_1, a_2, \dots, a_n)^T$ to satisfy the equation $W = \sum_{j=1}^n a_j \phi(x_j) = \phi(X)^T A$. Substituting this equation into Eq. (13), the optimization problem transformed into:

$$J_{global}(A) = \max_W \sum_{i=1}^n \left( (\phi(x_i))^T \sum_{j=1}^n a_j \phi(x_j) \right)^2,$$
$$s.t. \ \left( \sum_{j=1}^n a_j \phi(x_j) \right)^T \left( \sum_{j=1}^n a_j \phi(x_j) \right) = 1,$$

(14)

with the introduction of the kernel function $K(i,j) = \phi(x_i) \cdot \phi(x_j)^T$, the Eq. (14) can be expressed as:

$$J_{global}(A) = \max_{W} \sum_{i=1}^{n} \left( \sum_{j=1}^{n} a_j K(i,j) \right)^2,$$
$$s.t. \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j K(i,j) = 1, \tag{15}$$

we can further express Eq. (15) as:

$$J_{global}(A) = \max_{W} A^T K K A = \max_{W} A^T C A,$$
$$s.t. \ A^T K A = 1. \tag{16}$$

### 2.3.3. The objective function of LGPE

In order to preserve the local nonlinear structure between neighboring data points and extract the variance of the maximal high-dimension data, the objective function of the LGPE is to minimize $A^T L_l A$ (for local structure preserving) and maximize $A^T K K A$ (for global variance extracting). The objective function of LGPE can be transformed to the following optimization problem:

$$J(A) = \max_{A} \left( J_{global}(A) - J_{local}(A) \right) = \max_{A} (A^T C A - A^T L_1 A) = \max_{A} (A^T (C - L_1) A),$$
$$s.t. \ A^T K A = 1. \tag{17}$$

In order to eliminate the influence of noise in the process of dimensionality reduction, an orthogonal constraint is introduced. The derivation process is inspired by literature [31]:

$$\alpha_1^T \alpha_k = \alpha_2^T \alpha_k = \cdots = \alpha_{k-1}^T \alpha_k = 0. \tag{18}$$

In order to obtain the $k$ orthogonal basis vector $\alpha_k$, the following objective function need to be minimized:

$$J(\alpha) = \max_{\alpha} (\alpha_n{}^T (C - L_1) \alpha_n),$$
$$s.t. \ \alpha_1^T \alpha_k = \alpha_2^T \alpha_k = \cdots = \alpha_{k-1}^T \alpha_k = 0,$$
$$\alpha_1^T K \alpha_1 = \alpha_2^T K \alpha_2 = \cdots = \alpha_k^T K \alpha_k = 1. \tag{19}$$

In order to compute the $n$th discriminant vector, the lagrange multipliers in introduced to transform the $J_\alpha$. criterion including all the constraints:

$$J_\alpha = \alpha_k^T (C - L_1) \alpha_k - \lambda (\alpha_k^T K \alpha_k - 1) - \sum_{i=1}^{k-1} \mu_i \alpha_k^T \alpha_i. \tag{20}$$

The optimization is performed by setting the partial derivative of $J_\alpha$ with respect to $\alpha_k$ equal to zero:

$$\frac{\partial L_k}{\partial \alpha_k} = 0 \Rightarrow 2(C - L_1) \alpha_k - 2\lambda K \alpha_k - \sum_{i=1}^{k-1} \mu_i \alpha_i = 0. \tag{21}$$

Multiplying the left side of Eq. (21) by $\alpha_k^T$ obtained:

$$2\alpha_k^T (C - L_1) \alpha_k - 2\lambda \alpha_k^T K \alpha_k = 0 \Rightarrow \lambda = \frac{\alpha_k^T (C - L_1) \alpha_k}{\alpha_k^T K \alpha_k}. \tag{22}$$

Multiplying the left side of Eq. (21) successively by $\alpha_1^T K^{-1}$, $\alpha_2^T K^{-1}$, ..., $\alpha_{k-1}^T K^{-1}$, and obtaining a set of $k - 1$ expressions:

$$\mu_1 \alpha_1^T K^{-1} \alpha_1 + \cdots + \mu_{k-1} \alpha_1^T K^{-1} \alpha_{k-1} = 2\alpha_1^T K^{-1}(C - L_1)\alpha_k,$$
$$\mu_1 \alpha_2^T K^{-1} \alpha_1 + \cdots + \mu_{k-1} \alpha_2^T K^{-1} \alpha_{k-1} = 2\alpha_2^T K^{-1}(C - L_1)\alpha_k,$$
$$\vdots$$
$$\mu_1 \alpha_{k-1}^T K^{-1} \alpha_1 + \cdots + \mu_{k-1} \alpha_1^T K^{-1} \alpha_{k-1} = 2\alpha_{k-1}^T K^{-1}(C - L_1)\alpha_k. \tag{23}$$

Define the matrix notation:

$$U_{k-1} = [\mu_1, \mu_2, \ldots, \mu_{k-1}]^T, \quad \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{k-1})^T, \quad M_{k-1} = \left[ M_{ij}^{k-1} \right], \quad M_{ij}^{k-1} = \alpha_i^T K^{-1} \alpha_j.$$

Then the equations can be transformed in a single matrix relationship:

$$M_{k-1} U_{k-1} = 2\boldsymbol{\alpha}_{k-1}^T K^{-1}(C - L_1)\alpha_k. \tag{24}$$

Or in another form:

$$U_{k-1} = 2(M_{k-1})^{-1} \boldsymbol{\alpha}_{k-1}^T K^{-1}(C - L_1)\alpha_k. \tag{25}$$

Multiply the left side of Eq. (21) by $K^{-1}$:

$$2K^{-1}(C - L_1)\alpha_k - 2\lambda \alpha_k - K^{-1} \boldsymbol{\alpha}_{k-1} U_{k-1} = 0. \tag{26}$$

Including Eq. (26), we can obtained:

$$K^{-1}(C - L_1)\alpha_k - K^{-1} \boldsymbol{\alpha}_{k-1}(M_{k-1})^{-1} \boldsymbol{\alpha}_{k-1}^T K^{-1}(C - L_1)\alpha_k = \lambda \alpha_k. \tag{27}$$

Or in another form:

$$(I - K^{-1} \boldsymbol{\alpha}_{k-1} M_{k-1}^{-1} \boldsymbol{\alpha}_{k-1}^T)K^{-1}(C - L_1)\alpha_k = \lambda \alpha_k. \tag{28}$$

We need to maximized the criterion $\lambda$ of Eq. (28):

$$R_k = (I - K^{-1} \boldsymbol{\alpha}_{k-1} M_{k-1}^{-1} \boldsymbol{\alpha}_{k-1}^T)K^{-1}(C - L_1). \tag{29}$$

Thus, the required the $k$ orthogonal basis vector $\alpha_k$ is the eigenvector corresponding to the maximum eigenvalue of $R_k$. It is easy to verify that $\alpha_1$ is the eigenvector corresponding to the minimum eigenvalue of the generalized eigenvalue equation $(C - L_1) = \lambda K$.

## 3. Fault diagnosis process

In this study, a novel fault diagnosis model based on EEMD-ME and LGPE is proposed. The detail frameworks are shown in Fig. 3, the procedures of this model is described in successive steps:

Step 1: For a test signal $x_i = [x_1, x_2, \ldots, x_N]$, EEMD is used to decompose the signal into several IMFs. Then calculate cross-correlation coefficient and energy index of each IMF, and select the first $s$ IMFs to further analysis.

Step 2: The envelope entropy, permutation entropy and energy entropy of each selected IMFs are combined into an 3s-dimension fusion features $x_i = [x_{i1}, x_{i2}, \ldots, x_{i3s}]^T$.

Step 3: The 3s-dimension fusion feature set of test samples is input into LPGE to reduce the high-dimension feature space to two-dimension feature set $Y = [y_1, y_2]$. The dimensionality reduction steps are shown as follows:

1) For a high-dimension feature dataset $X = [x_1, x_2, \ldots, x_{3s}] \in R^{3s}$, we can construct the local nonlinear characteristics preservation objective function $J_{local}(A)$ by using Eq. (12).

2) The global variance maximum objective function $J_{global}(A)$ can be obtained by Eq. (16)

3) The objective function of proposed LGPE is defined at Eq. (19), and calculate the $k$ orthogonal basis vector $\alpha_k$ of projection matrix $\boldsymbol{\alpha} = (\alpha_1, \alpha_k, \ldots, \alpha_d)$ by iteration $\alpha_k$ is the eigenvector corresponding to the maximum eigenvalue of Eq. (29).

4) According to the equation $y_i = (\phi(x_i))^T W$, the low-dimension feature set $Y = [y_1, y_2]$ after projection space transformation can be obtained.

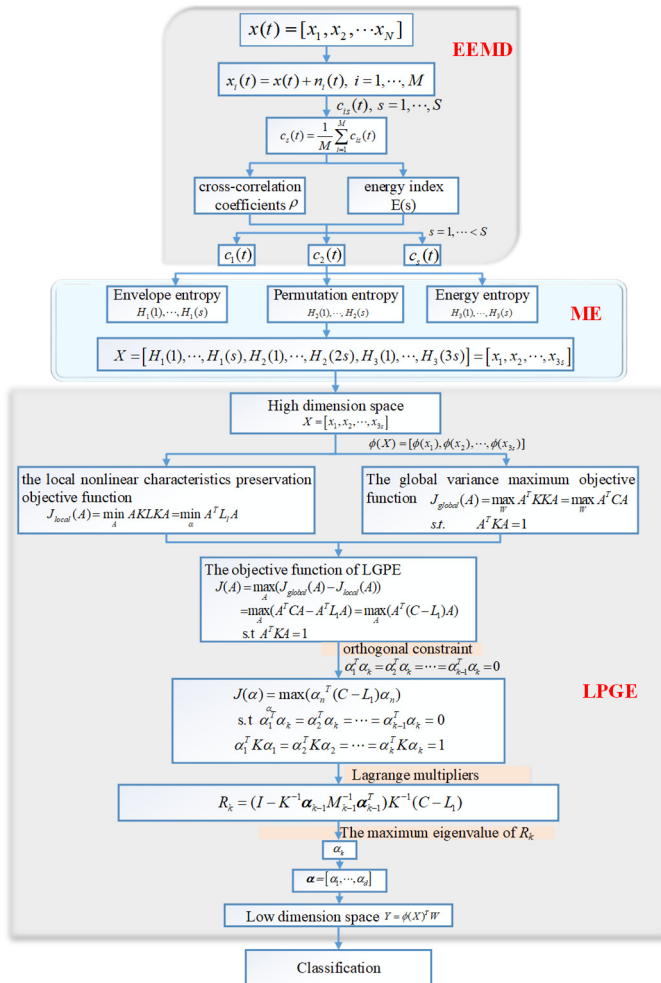Step 5: Classify and recognize the different faults by the degree of clustering in two-dimensional space.



**Fig. 3.** The details framework of the proposed model

## 4. Result and discussion

To investigate the effectiveness of the proposed technology for fault diagnosis, two experimental cases are considered. They including the bearing data obtained from the Case Western Reserve University (CWRU) Bearing Data Center and the gear data produced by QPZZ-II system.

## 4.1. The fault diagnosis of bearings

The bearing data obtained from CWRU Bearing Data Center has become a standard preference at the filed of fault diagnosis in bearings. A ball bearing as shown in Fig. 4, which was installed in a motor driven mechanical system. Vibration data is collected using accelerometers, which are attached to the housing with magnetic bases. In total four sets of data are obtained form the experiment systems, which include under normal conditions, with inner race fault, with ball fault and with outer race fault. The sampling frequency is 12 kHz for drive end bearing experiments. In this paper, the 0.007 inches fault diameter is selected for fault diagnosis. The vibration signal obtained from four different conditions are divided into 100 segments of 1024 sample each, as shown in Fig. 5.
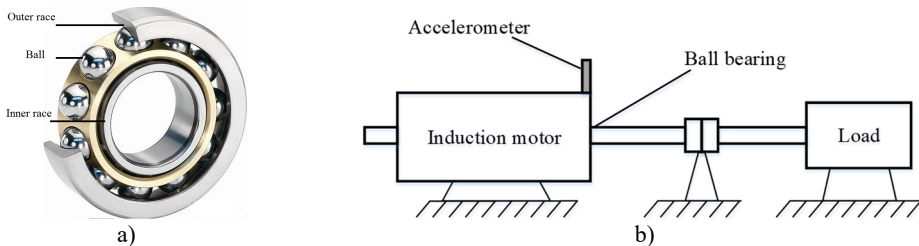


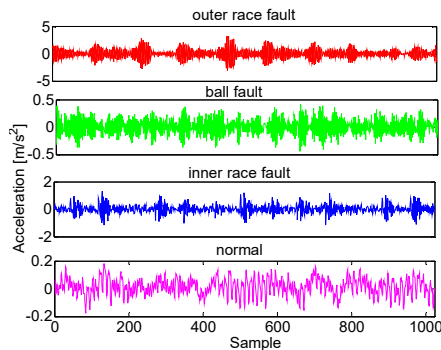Fig. 4. Experimental system



Fig. 5. Segments of the vibration signals collected from four different conditions of bearing

Each segment of vibration signal is decomposed into 10 IMFs by EEMD. For a segment, the cross-correlation coefficients $\rho$ and energy index $E$ of the former 4 IMF are presented in Table 1. It can be noticed that the first 4 IMFs have higher cross-correlation coefficients and energy index, which are chosen as the sensitive components for further analysis. The envelope entropy, permutation entropy and energy entropy of the first 4 IMFs in each segment are extracted as fusion features. In total 12 features are extracted for each segments as a point $x_i = (x_{i1}, x_{i2}, ..., x_{i12})$ in the feature space with the dimension of 12. And these points make up a fusion feature set $X = [x_1, x_2, ..., x_{100}]$ in a high-dimension space. Fig. 6(a), (b) and (c) show the envelope entropy, permutation entropy and energy entropy of the IMF1, respectively. In Fig. 6, we can see that the extracted feature have a certain potential for the fault detection at some extent but it also not for a fault diagnosis.

Since none of the extracted feature of 12 respective features is completely suitable for fault diagnosis, it is necessary to extract the features that can achieve a better classification ability between different classes of bearing condition. The idea of LGPE is to find a feature space whose dimension can be reduced without any loss information, and makes the classification process become simple. In this paper, we use the LGPE to solve the classification problem.

**Table 1.** Cross-correlation coefficients and energy index of each IMF

| IMFs | Outer race fault | | Ball fault | | Inner race fault | | Normal | |
|---|---|---|---|---|---|---|---|---|
| − | $\rho$ | $E$ | $\rho$ | $E$ | $\rho$ | $E$ | $\rho$ | $E$ |
| 1 | 0.8224 | 0.8094 | 0.7125 | 0.6538 | 0.7209 | 0.5833 | 0.6161 | 0.4591 |
| 2 | 0.4980 | 0.4287 | 0.5783 | 0.4735 | 0.7549 | 0.5519 | 0.6359 | 0.4370 |
| 3 | 0.3009 | 0.2190 | 0.4225 | 0.2260 | 0.4213 | 0.3071 | 0.2940 | 0.1719 |
| 4 | 0.1983 | 0.1502 | 0.3084 | 0.2440 | 0.1838 | 0.1126 | 0.4515 | 0.3830 |
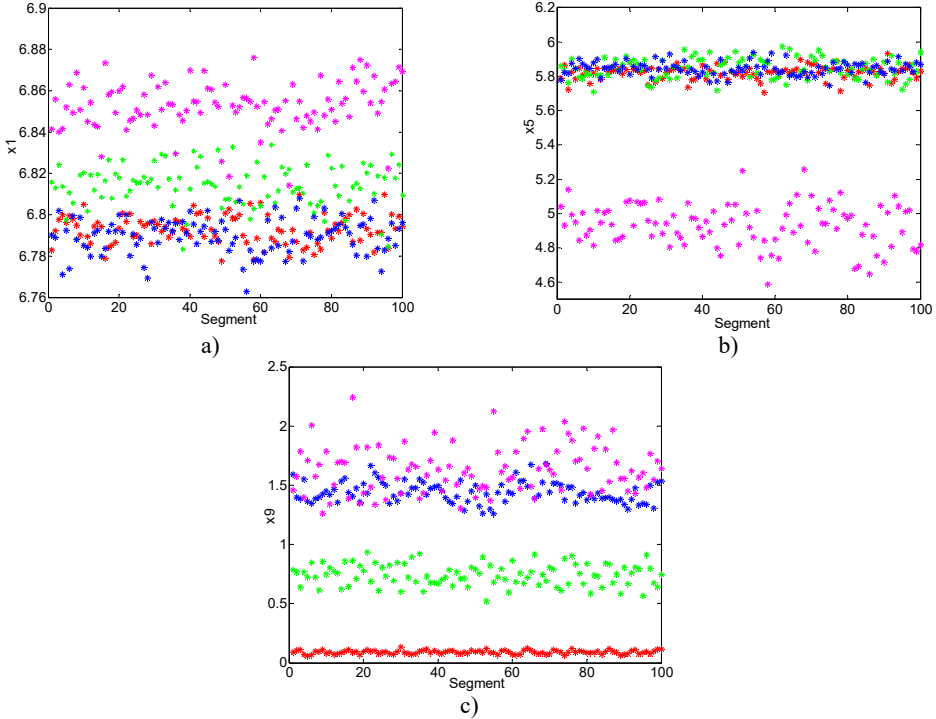


**Fig. 6.** a) Envelope entropy, b) permutation entropy, c) energy entropy of the IMF1,
outer race fault (red), ball fault (green), inner race fault (blue) and normal (magenta)

In this experiment, the reduce dimensionality of LPGE is set to $d = 2$, and the number of neighborhood points $k$ is 10. using the mixed-domain feature fusion method EEMD-ME and dimensionality reduction method LPGE, the distribution of the low-dimensional feature set after dimension reduction as shown in Fig. 7(a). As is can be seen in Fig. 7(a), all these four types of samples separated from each other in $d$-dimension space. The low dimensional feature set of the three fault sample cluster well, which are far away from the normal state. Thus, the fault diagnosis is performed.

The dimension reduction effect of LPGE is compared with other three mainstream dimension reduction algorithms KLPP, KPCA and LPP. The dimensionality reduction parameters of these algorithms are also set to $d = 2$, $k = 10$. The distributions of the low-dimensional feature sets after dimension reduction of KLPP, KPCA, LPP are shown in Fig. 7(b)-(d). Both Fig. 7(b), (c) and (d) indicate that KLPP, KPCA and LPP cannot separate the four different type of bearing effectively. As it can be notice that while using KLPP and KPCA, there are some overlaps between the outer race fault and ball fault, and outer race fault is large mixed with the inner race fault while using KPCA. The result proves that LPGE has more excellent clustering and classifying performance than KLPP, KPCA and LPP.

a) Dimensionality reduction with LPGE

b) Dimensionality reduction with KLPP

c) Dimensionality reduction with KPCA

b) Dimensionality reduction with LPP
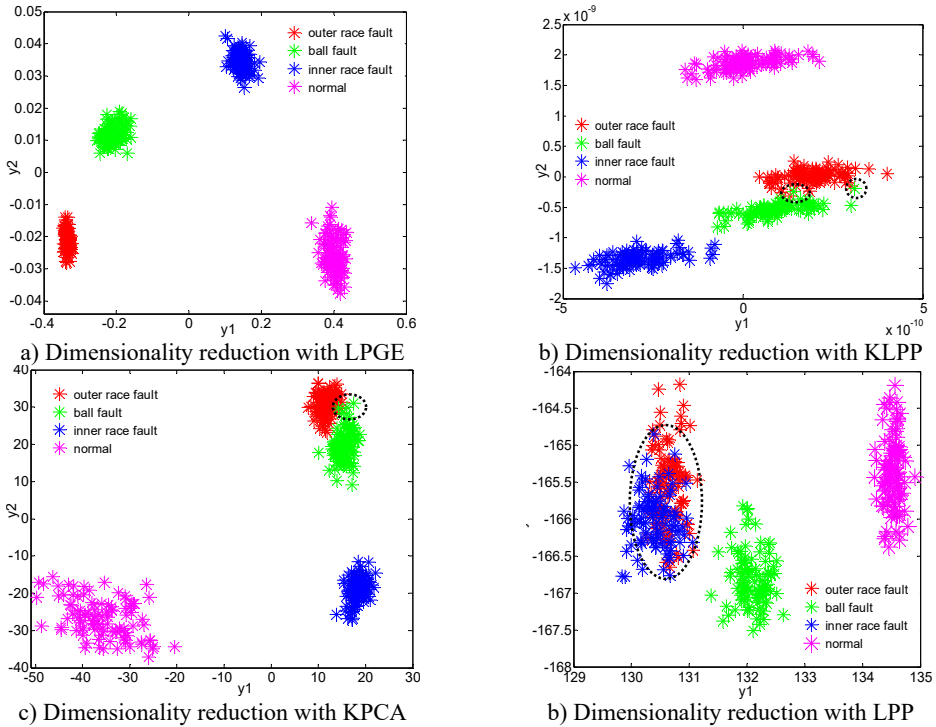
**Fig. 7.** The distribution of the low-dimensional feature sets of
bearing after dimensionality reduction of LPGE, KLPP, KPCA and LPP

## 4.2. The fault diagnosis of gear

In this experiment, the QPZZ-II test rig is designed to perform the fault test of gears. In total three types of data are obtained from the experiment systems, which include under normal conditions, with wear fault and with broken tooth fault. The sampling frequency is 5120 Hz. Each vibration signal is divided into 50 segments of 1024 sample, as shown in Fig. 8.
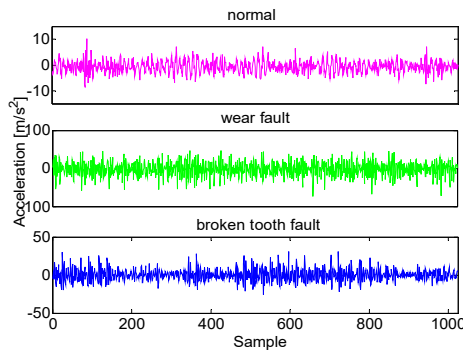


**Fig. 8.** Vibration signals collected from three different conditions of the gear

In the same way, each vibration segment of gear signal is decomposed into 10 IMFs by EEMD. Taking one segment as an example, the cross-correlation coefficients $\rho$ and energy index $E$ of each IMF of a segment are presented in Table 2. From it we can see that the first 3 IMFs have higher cross-correlation coefficients and energy index of the original vibrant signal. We select these 3 IMFs as the sensitive components for further analysis. Calculating the envelope entropy, permutation entropy and energy entropy of the first 3 IMFs in each segment. In total 9 features

are extracted for each segments as a point $x_i = (x_{i1}, x_{i2}, \dots, x_{i9})$ in the feature space with the dimension of 9. And these points make up a feature dataset $X = [x_1, x_2, \dots, x_{100}]$ in a high-dimension space.

**Table 2.** Cross-correlation coefficients and energy index of each IMF

| IMFs | Outer race fault | | Ball fault | | Inner race fault | |
|------|---------|--------|---------|--------|---------|--------|
| – | $\rho$ | $E$ | $\rho$ | $E$ | $\rho$ | $E$ |
| 1 | 0.6123 | 0.5342 | 0.8597 | 0.8428 | 0.8915 | 0.8721 |
| 2 | 0.7907 | 0.6791 | 0.5518 | 0.4315 | 0.4841 | 0.3615 |
| 3 | 0.3425 | 0.2322 | 0.1907 | 0.1713 | 0.2242 | 0.1835 |

For the gear vibration signal, using EEMD-ME and LPGE, the distribution of the low-dimensional feature set is shown in Fig. 9(a). The proposed method can separate three different gear conditions from each other. At the low dimensional space, the normal sample and the wear fault sample cluster well, while the broken tooth fault sample which are also has a certain degree of dispersion. In general, these samples after dimensionality reduction can effectively achieve the different conditions classification of gears.



a) Dimensionality reduction with LPGE

b) Dimensionality reduction with KLPP

c) Dimensionality reduction with KPCA

b) Dimensionality reduction with LPP
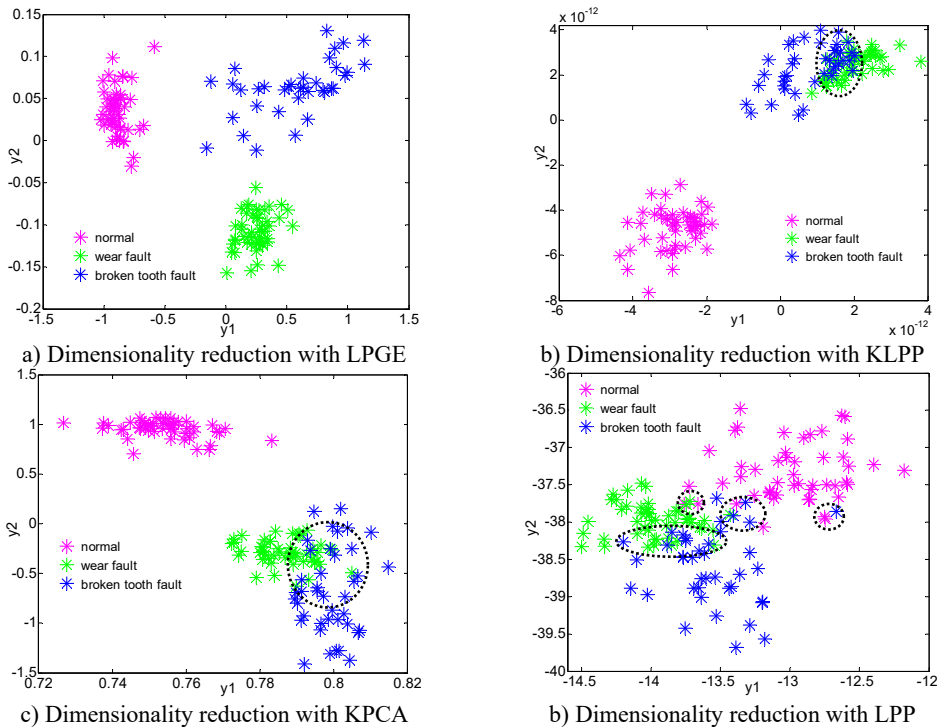
**Fig. 9.** The distribution of the low-dimensional feature sets of gear after dimensionality reduction of LPGE, KLPP, KPCA and LPP

Fig. 9(b)-(d) show the distribution of the low-dimensional feature set after the dimension reduction of KLPP, KPCA, LPP. In Fig. 9(b)-(d), it can be noticed that both the three methods cannot effectively separate the wear fault and broken tooth fault. In Fig. 9(b) and (c), when using KLPP and KPCA, there are some overlaps between the wear fault and broken fault, and the clustering of the samples with same fault is not well. In Fig. 9(d), the three type samples of gear are mixed together. The result verifies that the LPGE can achieve higher classification accuracy than other methods.

The above results demonstrate the superior performance of the proposed method.

Dimensionality reduction with LPGE is better than other classical dimension reduction algorithms. For an online vibration monitoring system, it is necessary to install an automated technique for fault diagnosis. The proposed fault diagnosis model can be used as the core diagnosis strategy of online vibration monitoring systems. Such the system enables an objective, reliable detection and diagnosis of mechanical faults. It also saves time of maintenance technicians and improves the economic performance of the equipment.

## 5. Conclusions

Bearings and gears are extensively used in wind turbine transmission systems. Defective bearings and gears cause high amplitude of vibration, which can increase power consumption and reduce the economic benefits of rotating machinery. Therefore, a reliable and faster fault diagnosis technique for bearings and gears is important for wind turbine maintenance decisions and reduce operating costs. In this paper, we proposed such a technique to be used for the key components of wind turbine transmission systems. The new technique based on the mixed-domain feature fusion (EEMD-ME) and dimensionality reduction (LPGE) demonstrates a high classification accuracy. The accuracy of the technique was tested on bearing and gear vibration signals. The four classes of the recorded bearing vibration signals, i.e. normal, outer race fault, inner race fault and ball fault operation, and three classes of the recorded gear vibration signals, i.e. normal, wear fault and broken tooth fault operation. The result demonstrates the technology achieves 100 % accuracy.

As part of further work, we plan to introduced the technique into the fault diagnosis of a wind turbine transmission system under variable speeds and alternating loads for wind turbine maintenance decision-making.

## Acknowledgements

## References

[1]     Wind Energy Report Germany. 2012, https://www.iwes.fraunhofer.de/.
[2]     **Zhao H., Yao R., Xu L., et al.** Study on a novel fault damage degree identification method using high-order differential mathematical morphology gradient spectrum entropy. Entropy, Vol. 20, Issue 9, 2018, p. 682.
[3]     **Avendaño Valencia L.-D., Fassois S. D.** Damage/fault diagnosis in an operating wind turbine under uncertainty via a vibration response Gaussian mixture random coefficient model based framework, Mechanical System and Signal Processing, Vol. 91, 2017, p. 326-353.
[4]     **Deng W., Yao R., Zhao H., et al.** A novel intelligent diagnosis method using optimal LS-SVM with improved PSO algorithm. Soft Computing, 2017, https://doi.org/10.1007/s00500-017-2940-9.
[5]     **Hu A. J., Yan X. A., Xiang L.** A new wind turbine fault diagnosis method based on ensemble intrinsic time-scale decomposition and WPT-fractal dimension. Renewable Energy, Vol. 83, 2015, p. 767-778.
[6]     **Deng W., Zhang S., Zhao H., et al.** A novel fault diagnosis method based on integrating empirical wavelet transform and fuzzy entropy for motor bearing. IEEE Access, Vol. 6, 2018, p. 35042-35056.
[7]     **Huang N. E., Shen Z., Long S. R., et al.** The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, Vol. 454, Issue 1971, 1998, p. 903-995.
[8]     **Bin G. F., Gao J. J., Li X. J., Dhillon B. S.** Early fault diagnosis of rotating machinery based on wavelet packets – empirical mode decomposition feature extraction and neural network. Mechanical System and Signal Processing, Vol. 27, 2012, p. 696-711.
[9]     **Ricci R., Pennacchi P.** Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions. Mechanical System and Signal Processing, Vol. 25, Issue 3, 2011, p. 821-838.

[10] **Wu Z., Huang N. E.** Ensemble empirical mode decomposition: a noise-assisted data analysis method. Advances in Adaptive Data Analysis, Vol. 1, Issue 1, 2009, p. 1-41.

[11] **Lei Y. G., Li Lin et al. N. P. J.** Fault diagnosis of rotating machinery based on an adaptive ensemble empirical mode decomposition. Sensor, Vol. 13, 2013, p. 16950-16964.

[12] **Zhao H., Sun M., Deng W., et al.** A new feature extraction method based on EEMD and multi-scale fuzzy entropy for motor bearing. Entropy, Vol. 19, Issue 1, 2016, p. 14.

[13] **Wang J., Gao R. X., Yan R.** Integration of EEMD and ICA for wind turbine gearbox diagnosis. Wind Energy, Vol. 17, Issue 5, 2014, p. 757-773.

[14] **Imaouchen Y., Kedadouche M., Alkama R., et al.** A frequency-weighted energy operator and complementary ensemble empirical mode decomposition for bearing fault detection. Mechanical System and Signal Processing, Vol. 82, 2017, p. 103-116.

[15] **Chen J. L., Pan J., Li Z. P., Zi Y. Y., Chen C. F.** Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals. Renewable Energy, Vol. 89, 2016, p. 80-92.

[16] **Tang B. P., Song T., Li F., Deng L.** Fault diagnosis for a wind turbine transmission system based on manifold learning and Shannon wavelet support vector machine. Renewable Energy, Vol. 62, 2014, p. 1-9.

[17] **Tenenbaum J. B., De Silva V., Langford J. C.** A global geometric framework for nonlinear dimensionality reduction. Science, Vol. 290, Issue 5500, 2000, p. 2319-2323.

[18] **Roweis S. T., Saul L. K.** Nonlinear dimensionality reduction by locally linear embedding. Science, Vol. 290, Issue 5500, 2000, p. 2323-2326.

[19] **Seung H. S., Lee D. D.** The manifold ways of perception. Science, Vol. 290, Issue 5500, 2000, p. 2268-2269.

[20] **He X., Niyogi P.** Locality preserving projections. Advances in Neural Information Processing Systems, 2004, p. 153-160.

[21] **Bhatia K. K., Rao A., Price A. N.** Hierarchical manifold learning for regional image analysis. IEEE Transaction on Medical Imaging, Vol. 33, Issue 2, 2014, p. 444-461.

[22] **Liu W., Sawant A., Ruan D.** Prediction of high-dimensional states subject to respiratory motion: a manifold learning approach. Physica Medica-European Journal of Medical Physics, Vol. 61, Issue 13, 2016, p. 2016-4989.

[23] **Wang X., Zheng Y., Zhao Z., et al.** Bearing fault diagnosis based on statistical locally linear embedding. Sensors, Vol. 15, Issue 7, 2015, p. 16225-16247.

[24] **Chen R., Chen S., Yang L., et al.** Looseness diagnosis method for connecting bolt of fan foundation based on sensitive mixed-domain features of excitation-response and manifold learning. Neurocomputing, Vol. 219, 2017, p. 376-388.

[25] **Su Z. Q., Tang B. P., Ma J. H., Deng L.** Fault diagnosis method based on incremental enhanced supervised locally linear embedding and adaptive nearest neighbor classifier. Measurement, Vol. 48, 2014, p. 136-148.

[26] **Ding X., He Q., Luo N.** A fusion feature and its improvement based on locality preserving projections for rolling element bearing fault classification. Journal of Sound and Vibration, Vol. 335, 2015, p. 367-383.

[27] **Huang Y., Zha X. F., Lee J., et al.** Discriminant diffusion maps analysis: A robust manifold learner for dimensionality reduction and its applications in machine condition monitoring and fault diagnosis. Mechanical System and Signal Processing, Vol. 34, Issue 1, 2013, p. 277-297.

[28] **Sun J. D., Xiao Q. Y., Wen J. T., Wang F.** Natural gas pipeline small leakage feature extraction and recognition based on LMD envelope spectrum entropy and SVM. Measurement, Vol. 55, 2014, p. 434-443.

[29] **Zhang X., Zhou J.** Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. Mechanical System and Signal Processing, Vol. 41, Issue 1, 2013, p. 127-140.

[30] **Tiwari R., Gupta V. K., Kankar P. K.** Bearing fault diagnosis based on multi-scale permutation entropy and adaptive neuro fuzzy classifier. Journal of Vibration and Control, Vol. 21, Issue 3, 2015, p. 461-467.

[31] **Duchene J., Leclercq S.** An optimal transformation for discriminant and principal component analysis. IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 10, 1988, p. 978-983.

**Ping Ma** received the Ph.D. degree from the School of Electrical Engineering, Xinjiang University, Urumqi, China. She current research interests include control, dynamics and fault diagnosis.
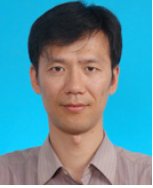
**Hongli Zhang** received the Ph.D. degree from Beijing Institute of Technology in 2009. His current research interests include control, dynamics and fault diagnosis.

**Wenhui Fan** received the Ph.D. degree from Zhejiang University in 1998. He is currently an Professor in Tsinghua University, Beijing, China. His current research interests include control, dynamics and fault diagnosis.

**Cong Wang** is currently an Professor in Xinjiang University, Urumqi, China. Her research interests include machine learning, intelligent computing, and system modeling and control.