Fault Diagnosis of Reciprocating Compressor Using Component Estimating Empirical Mode Decomposition and De-Dimension Template with Double-Loop Correction Algorithm

Hongyi Wang  
*Tianjin Polytechnic University*

Feng Dong  
*Tianjin University*

Xinxiu Zhou  
*Beihang University*

Hongyu Wang  
*Tianjin Polytechnic University*

Xinjun Zhu  
*Tianjin Polytechnic University*

See next page for additional authors

Publication Details

Fault Diagnosis of Reciprocating Compressor Using Component Estimating Empirical Mode Decomposition and De-Dimension Template with Double-Loop Correction Algorithm

Abstract
This paper presents an approach to implement multi-parameter (i.e., pressure, temperature, vibration, current, and liquid level) signals for fault diagnosis of the reciprocating compressor (RC). Due to the complexity of structure and motion of such compressor, the acquired signals involve transient impacts and noises. This causes the useful information to be corrupted and makes it difficult to diagnose the fault patterns accurately. A component estimating empirical mode decomposition (CEEMD) method is proposed to remove the random noise and improve data quality. Furthermore, a new template matching algorithm called de-dimension template with double-loop correction (DDT-DLC) is applied to diagnose the fault pattern contained in the time series signals. The DDT employs a judging criterion for key characterization parameters extraction and a multicellular parameter fusion method to reduce the dimension of the matching template, and then, the DLC supplies a double-loop correction algorithm to build a parameter state array computing model of the time series data by adjusting the dynamic factors. The proposed approach is validated with three fault patterns and the healthy pattern in a two-stage reciprocating air compressor. To confirm the superiority of the proposed method, its performance is compared with that of the traditional methods. The results have indicated that the proposed approach is of highly diagnostic accuracy and shortly computing time in the fault diagnosis.

Disciplines
Engineering | Science and Technology Studies

Publication Details

Authors
Hongyi Wang, Feng Dong, Xinxiu Zhou, Hongyu Wang, Xinjun Zhu, Limei Song, and Qinghua Guo

This journal article is available at Research Online: https://ro.uow.edu.au/eispapers1/3026
Fault Diagnosis of Reciprocating Compressor Using Component Estimating Empirical Mode Decomposition and De-Dimension Template With Double-Loop Correction Algorithm

HONGYI WANG1, FENG DONG2, (Senior Member, IEEE), XINXIU ZHOU3, HONGYU WANG1, XINJUN ZHU1, LIMEI SONG1, and QINGHUA GUO4, (Senior Member, IEEE)

1Key Laboratory of Advanced Electrical Engineering and Energy Technology, Tianjin Polytechnic University, Tianjin 300387, China
2Tianjin Key Laboratory of Process Measurement and Control, School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China
3School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China
4School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW 2500, Australia

Corresponding author: Hongyi Wang (wanghongyi@tjpu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51806150, in part by the Natural Science Foundation of Tianjin under Grant 18JCQNJC04400 and Grant 18JCYBJC22000, in part by the Tianjin Enterprise Science and Technology Commissioner Project under Grant 18JCTJC60600, and in part by the Tianjin Key Laboratory of Process Measurement and Control and the Program for Innovative Research Team of the University of Tianjin under Grant TD13-5036.

ABSTRACT This paper presents an approach to implement multi-parameter (i.e., pressure, temperature, vibration, current, and liquid level) signals for fault diagnosis of the reciprocating compressor (RC). Due to the complexity of structure and motion of such compressor, the acquired signals involve transient impacts and noises. This causes the useful information to be corrupted and makes it difficult to diagnose the fault patterns accurately. A component estimating empirical mode decomposition (CEEMD) method is proposed to remove the random noise and improve data quality. Furthermore, a new template matching algorithm called de-dimension template with double-loop correction (DDT-DLC) is applied to diagnose the fault pattern contained in the time series signals. The DDT employs a judging criterion for key characterization parameters extraction and a multicellular parameter fusion method to reduce the dimension of the matching template, and then, the DLC supplies a double-loop correction algorithm to build a parameter state array computing model of the time series data by adjusting the dynamic factors. The proposed approach is validated with three fault patterns and the healthy pattern in a two-stage reciprocating air compressor. To confirm the superiority of the proposed method, its performance is compared with that of the traditional methods. The results have indicated that the proposed approach is of highly diagnostic accuracy and shortly computing time in the fault diagnosis.

INDEX TERMS Fault diagnosis, component estimating empirical mode decomposition (CEEMD), de-dimension template (DDT), double-loop correction algorithm (DLC).

I. INTRODUCTION Reciprocating compressor (RC) plays a very important role in industrial production line, such as petrochemical plants, refineries, gas storage and transport, and so on. Due to the complexity of structure and the close links between components, the fault accrued in any part of the compressor often causes the breakdown of the whole equipment. More seriously, fault accident will bring significant economic losses and threaten the safety of person. Therefore, in modern industry, higher reliability, usability, maintainability, economy and safety of mechanical equipment are required, and people pay more attention to efficiency, energy consumption and environmental protection than before. In the actual operation of the equipment, fault warning and fault pattern diagnosis are helpful for reducing and preventing the
occurrence of accidents, as well as promoting the economic benefit of production.

Several researches for RC fault diagnosis have been recently proposed in literature. Among these studies, the vibration signals have aroused widespread interest. Base on vibration signal analysis, fault detection methods of RC valves, bearing clearance, RC leakage are studied separately [1]–[4]. Acoustic emission signal is another popular parameter for RC fault analysis. Several researchers investigated the application of acoustic emission parameters for valve faults detection in RC [5], [6]. In addition, the pressure and temperature are also play a key role in RC fault diagnosis. The vibration, pressure, and current signals have already been used together for fault diagnosis of RC valves [7]. Obviously, multi-parameter contains more monitoring information and better reflects the operation status of machine. However, the more parameters there are, the harder the data to process. Developing a suitable data processing method is necessary to improve the data quality and extract fault features. In addition, most of the existing researches focused on the diagnosis of a particular type of fault, and few studies were for the diagnosis of multiple faults. Therefore, another challenge for fault diagnosis is accurately categorize the fault occurrences and their severities. It is imperative to develop a new method for simultaneous diagnosis of multiple faults.

Considering that the multiple signals acquired from the industry is complicated and mixed with noises, data processing method is important for data quality improvement. Empirical mode decomposition (EMD) is an effective way in time series signal decomposition [8]. Data features could be extracted intrinsic mode functions (IMFs) based on EMD [9]. Previous researches have employed EMD to decompose a complex time series into a number of IMFs and one residual series for forecasting model establishment [10], [11]. Statistical fault characteristics of equipment could be extracted from the IMFs based on EMD [12]–[13]. In recent years, many improved methods based on EMD were proposed to meet different needs of data pre-processing. Multivariate Empirical Mode Decomposition (MEMD) is an extension of the EMD algorithm to multivariate channels without any dimension restrictions [14], which is widely used in image processing [15]. Noteworthy, EMD is also used in noises removing area. Several improved approaches based on EMD have been studied to solve the trend filtering problem [16], [17]. A minimum arclength EMD (MA-EMD) is proposed to remove impulse-like noises in time series data [18]. In this work, a novel method based on EMD is proposed to remove the random noises in a time series data.

Fault diagnosis is usually based on mechanism analysis, prior knowledge and data analysis. For RC, it's difficult to obtain accurate mathematical models to realize fault diagnosis as a result of the complexity structure and motion characteristics. Prior knowledge model is also of great limitations in practical industrial application because of the need for a large number of prior knowledge [19].

Data-driven model uses equipment operating data, and obtains hidden information in data through intelligent and statistical algorithms, so as to diagnose the fault. It relies less on mechanism and prior knowledge and has high feasibility in the actual industrial equipment and operation. Several machine learning algorithms are widely used in fault diagnosis and pattern recognition, i.e. support vector machines (SVM) [20], neural networks (NN) [21], fuzzy clustering (FC) [22], [23], convolutional neural networks (CNN) [24], [25], so on. However, as well known, the machine learning method always requires a large number of data samples. With small sample data or scarce sample data, it is difficult to get high accuracy of machine learning model. The statistical algorithm of template matching, which is widely used in pattern recognition of image processing, could solve this problem effectively [26], [27]. Template matching method could be employed for fault sensitivity analysis [28]. Several representation methods of equipment fault feature based on template matching has also been proposed [29], [30]. However, for RC fault diagnosis, the template of the fault is difficult to calculate because the data samples from industrial platform are complex and unstable. In order to recognize the fault patterns based on time series data from industrial platform, the template matching algorithm needs to be improved.

In this paper, we introduced a novel approach to solve the multiple faults diagnosis of RC. The time series samples with various parameters of three typical faults and the healthy situation are analyzed in this work. Data random noises are removed by component estimating empirical mode decomposition (CEEMD) method. An improved template matching algorithm, de-dimension template with double-loop correction (DDT-DLC), is proposed for fault template extraction and fault diagnosis. The experiment results have verified the effectiveness of the proposed method.

II. RC FAULT DESCRIPTION

RC mainly comprises the body of machine, the driving machine, working chambers, transmission structure and Auxiliary system. The main components of RC include crankshaft, crankshaft connecting rod, crosshead, piston rod, exhausting valve, intake valve, etc. As shown in FIGURE 1, each part coordinates with each other to complete the intake, compression and exhaust of compressed medium.
The RC faults can be classified in two different ways. Firstly, it can be classified according to characteristics shown, such as leakage, wear, fracture, high temperature, and so on. Secondly, it can be classified according to the site occurred of the fault, for instance, valve, transmission structure, driving machine, seal component, etc. Obviously, it is much more convenient for equipment maintenance with the second classification method because we can accurately identify the fault location. Therefore, the fault patterns discussed in this work are related to the mechanical components of RC. According to statistics, more than 60% of the reciprocating compressor failures occur on the valve, the piston rod fracture accident accounts for about 25% of the major accidents, and about 10% for the driving motor fault. Thus, these three kinds of fault would be discussed in the following content in this work.

In industry, various sensors are installed on RC. Signals of pressure, temperature, vibration, current and liquid level in different component are measured during the production process. The changing characteristics of time series data are related to the occurred faults. Due to the complexity of structure and motion of RC the measured signals are interacted with each other. But, the relationship between measured parameters and fault reasons is not one-to-one correspondence. A fault may cause abnormal changes of several characteristic parameters, and an abnormal parameter may correspond to different fault patterns. Especially in the early stage of the fault, the corresponding relationship is not significant because the fault characteristics are weak. In this work, both the data processing method and fault diagnosis method are studied to identify fault patterns accurately with the multi-parameter data.

III. DATA PROCESSING METHOD OF CEEMD

A. F-EMD FOR DENOISING

The empirical mode decomposition (EMD) is an algorithm which decomposes a time series into a finite additive superposition of oscillatory components, each of which is called an intrinsic mode function (IMF). The extraction process of IMF was detailed in reference [8] and [16]. With EMD, the original signal of a time series \( X \) is decomposed into \( K \) IMF components and 1 residual component at different frequencies. The \( k \)-th IMF component could be written as IMF(\( k \)) or IMF\( k \). Each decomposed component contains only one mode oscillation without complex wave. The original signal could be reconstructed by summing all intrinsic mode components and residual components.

Flandrin studied the EMD in different approach to solve the trend filtering problem [16], [17]. He pointed out that the smaller order of IMF, the more noise information it contains. Since the first IMF component contains most of the noise signals, it could be viewed as a pure noise. Then, the power spectrum IMFk with noise signals shows similar characteristics to IMF1, and the noise energy of IMFk decreases with \( k \). It can be judged whether IMFk \( (k \geq 2) \) is noise or not by matching the IMFk energy and estimated noise energy as described in reference [16] and [17]. Thus, the low-frequency trend from a given time series data could be acquired by filtering the noisy IMF components and keeping the residual component.

In Flandrin’s EMD (F-EMD) filtering method, the IMF1 were viewed as noise signals directly and filtered out. However, IMF1 may be not a noisy component actually. Thus, the F-EMD may result in loss of useful information. Obviously, an evaluation method is necessary to complement his approach. In order to get a better filtering result, a new method is proposed to identify the noisy component of EMD in this work, which is called component evaluating empirical mode decomposition (CEEMD).

B. COMPONENT EVALUATING EMPIRICAL MODE DECOMPOSITION FOR DENOISING

CEEMD is an improved denoising method based on EMD, which mainly used for random noise filtering. In order to evaluate the random noise in the IMF component, an autocorrelation method is applied.

Previous studies have shown that the Gaussian noise signal has a low auto correlation. The autocorrelation coefficient \( Cor \) is calculated by the formula (1).

\[
Cor = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2}
\]

where, \( \tau \) is the interval for autocorrelation; \( X_t, E[], \mu \) and \( \sigma \) separately denote the sample point, expectations, average value and variances of time series \( X \).

A sine signal, a Gaussian noise, a noisy-adding sine signal and their autocorrelation coefficients are shown in FIGURE 2. As we can see, there is a maximum autocorrelation coefficient for noise signal, which attenuates rapidly. Although the sine signal also achieves a maximum value in its autocorrelogram, it does not show a sharp attenuation trend and has a high auto correlation.

Thus, the noise evaluation formula can be defined as:

\[
Cor(n_{max}) - Cor(n_{max} + \tau^*) \geq \eta
\]
where \( \text{Cor}(n_{\text{max}}) \) is the maximum autocorrelation coefficient, \( \text{Cor}(n_{\text{max}} + \tau^*) \) is the autocorrelation coefficient with \( \tau^* \) interval to \( n_{\text{max}} \). The sample interval \( \tau^* \) should be smaller than the sample number. Usually, \( \tau^* \in N, \ \tau^* = [2, 6] \). \( \eta \) is a threshold for noise identification, \( \eta = [0.02, 0.1] \). The smaller \( \eta \) is, the stronger the filtering effect. As can we seen, if the autocorrelation coefficients of a time series signal satisfied formula (2), the signal is a noise.

According to the above analysis, a possible strategy (CEEMD) for denoising signal corrupted by Gaussian noise is shown in FIGURE 3.

The steps of CEEMD is described as follows:
(I) Compute the EMD of the time series data \( X \), acquire the IMF components;
(II) Set \( kn = 1 \);
(III) Compute the autocorrelation coefficient of the \( kn \) th IMF;
(IV) Evaluate the component by formula (3):
If satisfied, then the component is a noise, \( kn = kn + 1 \), turn back to step (III);
If not satisfied, then then the component is not a noise, turn to step (V).
(V) Sum the non-noise components of IMF and acquire the denoised data \( X_0 \) by:
\[
X_0 = \sum_{k=kn}^{K} \text{IMF}_k + R
\]

The denoised results for sine signal and noisy sine signal with Gaussian noise based on CEEMD are compared with F-EMD, as shown in FIGURE 4. The comparing results have shown that the CEEMD is effective in removing noises and maintaining original signal information.

**FIGURE 3. Flow chart of CEEMD.**

**FIGURE 4. Signals denoised by F-EMD and CEEMD.**

## IV. FAULT DIAGNOSIS METHOD OF DDT-DLC

Template matching is a common method in pattern recognition, which is performed by calculating the similarity between testing samples and matching templates. For fault diagnosis, the type of fault is the pattern to be recognized. To ensure the production process works successfully, numerous parameters of machine would be monitored. Usually, there are three ways to build the matching template by classical method: set every sample point with all parameters as a template, set a time series with all parameters as a template, and set the mean parameter value with all parameters as a template. However, due to the complexity of industry process and irregular fluctuations of the measured signals, it is difficult to establish an accuracy template library (TL) for fault diagnosis by classical template matching method. The three classical templates introduced above are of high dimension and low accuracy for RC data. In order to establish an effective template for mechanical fault diagnosis in industry, an improved template matching method based on the time series samples, called de-dimension template with double-loop correction (DDT-DLC), is proposed in this work. The main work of DDT-DLC could be divided into three parts: build the TL of fault patterns, establish the computing model of parameter state array and match with the template for fault pattern diagnosis. For TL building, the DDT method is used to realize template dimension reduction by recognizing the characteristic parameter sets of different faults and fusing the parameters of the multicellular parameters. For parameter state array (PSA) computing model establishment, the DLC algorithm is applied by adjusting two coefficients in double closed circuits. Then, the fault pattern could be recognized by matching the PSA calculated by the computing model and the template in TL.

### A. TEMPLATE LIBRARY BASED ON DDT

DDT is a way to reduce the dimension of TL by two steps: key characteristic parameter set extraction and multicellular parameter fusion.
1) CHARACTERISTIC PARAMETER SETS EXTRACTION

Key characteristic parameter set (KCPs) for each pattern of fault could be extracted by a judgment criterion (JC) method. JC is a statistical method to measure the difference between two data sets by considering their inner-distance and inter-distance. For the KCPs of each fault is different, the JC is calculated with the time series data between every fault pattern and the healthy pattern.

Suppose $X_i$ and $X_{nor}$ are the $i$th time series data for a certain fault and the healthy situation. There are $C$ patterns of fault, $N$ samples for each fault pattern, $K_p$ parameters for each sample and $n$ data points for each parameter.

The mean parameter value for sample $i$ is:

$$\bar{X}_i = \frac{1}{n} \sum_{t=1}^{n} X_i(t)$$  \hfill (4)

The mean parameter value for healthy pattern is:

$$\bar{X}_{nor} = \frac{1}{n} \sum_{t=1}^{n} X_{nor}(t)$$  \hfill (5)

The inner-distance $S_{wi}$ of the parameters for sample $i$ is

$$S_{wi} = \frac{1}{n} \sum_{t=1}^{n} (X_i(t) - \bar{X}_i)(X_i(t) - \bar{X}_i)^T$$  \hfill (6)

The total inner-distance $S_w$ of the parameters for the given fault is:

$$S_w = \frac{1}{N} \sum_{i=1}^{N} S_{wi}$$  \hfill (7)

The inter-distance $S_B$ of the parameters between the given fault and healthy data is:

$$S_B = \frac{1}{N} \sum_{i=1}^{N} (\bar{X}_i - \bar{X}_{nor})(\bar{X}_i - \bar{X}_{nor})^T$$  \hfill (8)

Then, the judgment criterion $J_c$ could be constructed by the determinant or trace with the inner-distance and the inter-distance of two different data patterns, as shown in the following formula.

$$J_c = \frac{\text{tr} S_B}{\text{tr} S_w}$$  \hfill (9)

The larger value in the matrix trace, the more significant of the corresponding parameter. The KCPs could be acquired by choosing the parameters with larger matrix trace value in $J_c$.

2) MULTICELLULAR PARAMETER FUSION

There are always some multicellular parameters to ensure the reliability of industrial machine monitoring. The multicellular parameters are the same parameters for the same machine part measured by different sensors. Any parameter of the multicellular abnormally indicates a fault occurs. In order to further compression of the initialized template library, the multicellular parameters could be fused to one parameter by formula (10).

$$P'_z = P_{z1} \cup P_{z2} \cup \cdots \cup P_{zN}$$  \hfill (10)

where $P_{z1}, P_{z2}, \ldots, P_{zN}$ are the $z$ parameters in a multicellular, $P'_z$ is the fusion parameter. When any parameter in the multicellular is abnormal, the state of the fusion parameter is abnormal. Then the final template could be acquired by merging the same templates.

FIGURE 5 shows an example of multicellular parameter fusion. FIGURE 5(a) is the templates with multicellular parameters. FIGURE 5(b) is the templates processed by formula (10). FIGURE 5(c) is the final templates after templates merging. The $P_{11}$ and $P_{12}$ are multicellular parameters, $P_{21}$ and $P_{22}$ are multicellular parameters, $P'_1$ and $P'_2$ denote the fusion parameters for the two multicellular respectively.

3) TEMPLATE LIBRARY BUILD UP

The method of template building up referred in this work does not concern the sample points at every moment, but concentrates on the characterization of the parameter for each fault pattern.

Extract the KCPs based on $J_c$. The initialized template of the fault pattern could be acquired by initializing the working state of parameters in KCPs to 1 and other parameters to 0. Obviously, the template for the healthy data is a zero array. Then, the initialized template library (ITL) could be built up by combining the initialized template of all fault patterns and health pattern. The size of the ITL based on KCPs is $M \times K_p$, which is much smaller than that of the time series template library (TSTL) with the size of $M \times K_p \times n$, where $M = (C + 1) \times N$ is the total numbers of the time series sample and $C + 1$ is the total number of the fault patterns and healthy.

For further de-dimension, fusing the multicellular parameters of ITL, the TL could be acquired and downsized to $M_f \times K'_f$, where $M_f \leq M$ and $K'_f \leq K_p$. 

![FIGURE 5. Multicellular parameter fusion process. (a) Templates with multicellular parameters. (b) Templates of multicellular parameter fusion. (c) Final templates.](image-url)
Then the permitted deviation radius is:

\[ R = \alpha \cdot R_0 \]  

(12)

The adjustable range of \( \alpha \) element could range from 1 to 3.

**Step 3:** Input the data samples of fault and health. And initialize the dynamic threshold matrix \( Th = \{Th(j) = n, j = 1, 2, \ldots, K_p\} \).

**Step 4:** Initialized the binary state array of parameter.

Compute the deviation \( E_r \) of time series sample from the parameter center by

\[ E_r = |X_i - \bar{X}_{nor}| \]  

(13)

For the \( j \)th parameter, count the number of points satisfy \( E_r(j) > R(j) \), written as \( count(j) \). Then the \( j \)th parameter state \( P_0(j) \) could be acquired by

\[ P_0(j) = \begin{cases} 1 & \text{if } count(j) \geq Th(j) \\ 0 & \text{if } count(j) < Th(j) \end{cases} \]  

(14)

The initialized binary state array \( P_0 \) could be acquired by traversing all the parameters with the above operation.

**Step 5:** Inner loop correction by adjusting \( Th \).

Matching \( P_0 \) with its corresponding initialized template \( IT_i \) in TL by

\[ Dis = P_0 - IT_i \]  

(15)

If the \( j \)th element \( Dis(j) \neq 0 \), then adjust \( Th(j) \) by \( Th(j) = Th(j) - \tau_{th} \) and turn to step 4 until satisfying the following constrain:

\[ Th(j) \leq 0 || P_{nor}(j) = 1 || Dis(j) = 0 \]  

(16)

where \( P_{nor} \) denotes the template of healthy data, \( \tau_{th} \in [1, n] \) denotes the modifying factor.

**Step 6:** Outer loop correction by adjusting \( \alpha \).

Fusing the multicellular parameter states by formula (10), we could get the binary state array \( P \). It is important to note that the binary state array for healthy situation must be a zero array.

The fault pattern of a sample data could be recognized by formula (17).

\[ D_{min} = \min\{(P - T_i)(P - T_i)^T, ~i = 1, 2, \ldots, M_f\} \]  

(17)

where \( T_i \) denotes the \( i \)th template of TL, \( D_{min} \) denotes the smallest difference between the train sample and template. The pattern of the template which corresponds to \( D_{min} \) is the fault pattern of the sample. Then the recognition rate \( Q \) for all the train samples is

\[ Q = \frac{N_r}{N_s} \]  

(18)

where \( N_r \) is the number of correct recognition, \( N_s \) is the total number of the train samples. If \( Q \) achieves accuracy requirement, then the optimal \( Th \) and \( \alpha \) are found, and the binary state array of data sample could be calculated. Else if \( Q \) does not achieve the accuracy requirement, then reset \( \alpha \) by...
formula (19) and turn to step 2 until satisfying the constraint of formula (20).

\[
\alpha(j) = \alpha(j) + \Delta \alpha \tag{19}
\]

\[
Q \geq Q_{\text{set}} \quad |i \alpha | \geq i \alpha_{\text{max}} \tag{20}
\]

where \(\Delta \alpha\) is a learning pace factor. Usually, \(\Delta \alpha \in [0.1, 1]\). \(Q_{\text{set}}\) is the required accuracy, \(i \alpha\) is the iteration time, \(i \alpha_{\text{max}}\) is the maximum iteration time. If it cannot find a satisfactory \(Q\), then choose the \(Th\) and \(\alpha\) correspond to the maximum \(Q\) as the optimal coefficient for the binary state array computing.

Thus, the PSA computing model is completed with the optimal \(\alpha\) and \(Th\). And the computing process of PSA based on the model is shown in FIGURE 7. The healthy range of every parameter could be determined based on the optimal \(\alpha\). And the binary state array of the measured data could be calculated with optimal \(Th\). Finally, the fault patterns can be diagnosed by matching the binary state array to the templates in TL with formula (17).

V. APPLICATION FOR RC FAULT DIAGNOSIS

The method is evaluated by using 5-year operation data collected from an offshore oil corporation. Thirty-two parameters of a two stage reciprocating air compressor were measured with a sample time of 1 minute. In order to identify the patterns of mechanical faults, the running data, under different working situation including healthy and fault, were collected and analyzed. However, the number of fault samples is not enough for machine learning method. Therefore, the template matching method is applied in this work.

The whole process of the proposed method is executed as shown in FIGURE 8. Firstly, all the data are denoised by removing the stop points, singular points and the random noise, as described in section V.A. Secondly, the TL is built up based on the train sample by DDT algorithm. Thirdly, the PSA computing model is established based on the train sample by DLC algorithm. Fourthly, the PSA of test data is computed by the PSA computing model. Finally, the fault pattern could be diagnosed by matching the PSA and the template in TL.

A. DATA PROCESSING RESULTS

The original data normally involves a lot of noise for the harsh environment in industry field. Take the primary exit pressure (PEP) for example, as shown in FIGURE 9(a).

There are various kinds of noise in the original data, such as boot-up and shutdown data, singular points, random noise, and so on. To achieve highly diagnosis rates, appropriate denoising methods are employed in data pre-
processing. Since this work only aims at the fault diagnosis in stable operation process, the data during the downtime and unstable operation process need to be removed. Usually, data is invalid from the shutdown time to 30 minute after the next boot-up. All signals should be processed by removing invalid data. In addition, there are also some singular points in the data, which may be caused by sudden disturbances on the worksite or gross error in measurement. A clustering algorithm based on local data clustering is applied to search the singular points, and the Gaussian interpolation method is used for data interpolation. In addition, the pre-denoise process could reduce the generation of modal aliasing for CEEMD denoising in the following work. Then, the pre-denoised data after filtering the stop points and singular points could be acquired. The pre-denoised PEP is shown in FIGURE 9(b) as an example.

CEEMD method is applied for further denoising. The experimental results of denoised PEP are shown in FIGURE 10. Obviously, CEEMD eliminates random noise and retains the main information and changing trend of the signal.

B. FAULT DIAGNOSIS RESULTS

In this work, three patterns of fault and the healthy working situation are diagnosed. The three fault patterns are the value fault, the piston rod fracture accident and the driving motor fault. All of these faults are commonly happened during the industry process.

In the implementation process of fault diagnosis, there are 5 samples of each pattern applied for TL establishment and DDT-DLC model training. A number of 60 samples were applied for model testing, including 38 samples of healthy state, 11 samples of value fault, 5 samples of the piston rod fracture accident and 6 samples of driving motor fault. There are 32 parameters for each sample and 20 sample points for each parameter. Among the 32 parameters, there are 16 multicellular with 2 parameters for each because every kind of parameter was measured by two sensors.

During the process of TL establishment, the KCPs of each fault pattern could be extracted. Four parameters in KCPs with the highest JC value for the three faults are shown in TABLE 1. As can we seen, the factors in KCPs are closely related to the fault location of RC.

Due to the compressing of the TL dimension, the matching speed would be highly improved than the classical method. The dimensions of TL based on different methods are shown in TABLE 2. The consuming times and the diagnosis rates for these methods are also compared in TABLE 2.

As can we seen, the TL dimension of DDT-DLC is greatly reduced. Though the computing time depends on the speed of the CPU to some extent, the DDT-DLC cost less time due to the low dimension of TL. Although the CEEMD-DDT-DLC
method cost a little longer computing time than DDT-DLC, it is still faster than TSTL and DDT-DLC without parameter fusion. The most important thing is that the fault recognition rates based on CEEMD-DDT-DLC are much higher than the others. The results have validated that the DDT-DLC is effective in fault diagnosis. And the CEEMD is helpful in improving the fault recognition rates.

In order to further demonstrate the advantages of CEEMD-DDT-DLC, the SVM method of pattern recognition, which instead of template matching in the last step of fault diagnosis as shown in FIGURE 8, is applied for comparison. The SVM classification model could be established by training the PSA from CEEMD-DDT-DLC. And the same train samples and test samples to test template matching method are implemented for SVM. The fault diagnosis rates of every fault pattern based on SVM model trained by single sample point and the PSA are shown in TABLE 3. As can we seen, the diagnosis rates of the latter method are improved significantly. However, due to the small number of the train samples, the diagnosis rates of SVM based on CEEMD-DDT-DLC is a little lower than template matching based on CEEMD-DDT-DLC. Even so, the experimental results have shown that the CEEMD-DDT-DLC is effective in RC fault diagnosis.

VI. CONCLUSION

This paper has proposed a new approach to RC fault diagnosis using multi-parameters involving pressure, temperature, vibration, current and liquid level measured from different part of RC. To improve the data quality, an improved EMD method, called CEEMD, is developed by noise estimating of IMF component based on autocorrelation coefficient analysis. For RC fault diagnosis, DDT-DLC is proposed to solve the problem of small sample recognition with frequent and irregular changes of time series signal. To reduce the dimension of TL for RC, DDT algorithm based on KCPs extraction and multicellular parameters fusion is applied in this work. And a double-loop correction algorithm, called DLC, is proposed to establish the binary state array computing model of the time series data. Fault pattern could be diagnosed by matching the binary state array and the template in TL. The experimental results have confirmed that the proposed method can accurately diagnose the fault patterns. And this method can be extended to diagnose more fault types as long as there are suitable parameters in sample data.

REFERENCES


**HONGYI WANG** received the B.S. degree from Yanshan University, Hebei, China, in 2006, and the M.S. and Ph.D. degrees from the School of Electrical Engineering and Automation, Tianjin Polytechnic University, Tianjin, China, in 2011. She is currently a Research Assistant with Tianjin Polytechnic University. Her main research interests include multi-phase flow measurement, intelligent detection technology, pattern recognition, and fault diagnosis.

**FENG DONG** (M’03–SM’06) received the B.S., M.S., and Ph.D. degrees in control science and engineering from Tianjin University, Tianjin, China, in 1988, 1996, and 2002, respectively. He is currently a Professor with the School of Electrical and Information Engineering, Tianjin University, and the Director of the Tianjin Key Laboratory of Process Measurement and Control. His research interests include process parameter detection and control systems, industrial process flow measurement, and intelligent information processing.

**XINJUN ZHU** received the B.S. degree from Linyi University, in 2008, the M.S. degree from the Shandong University of Technology, in 2011, and the Ph.D. degree from Tianjin University, in 2015. He joined Tianjin Polytechnic University, in 2015. His research interests include fringe analysis, phase retrieval for optical 3-D measurement, and light scattering techniques for particles.

**LIMEI SONG** received the B.S., M.S., and Ph.D. degrees in measurement technology and instrumentation from Tianjin University, Tianjin, China, in 1999, 2001, and 2004, respectively. She was an Associate Professor with the Southwest University of Science Technology, from 2006 to 2008. She was a Visiting Scholar with the Hiroshima Institute of Technology, Japan, in 2006, and the University of Wollongong, Wollongong, Australia, from 2011 to 2012. In 2008, she joined Tianjin Polytechnic University, as an Associate Professor, where she is currently a Professor and the Vice Dean of the School of Electrical Engineering and Automation. Her research interests include machine vision, 3-D optical measurement, reverse engineering, and pattern recognition.

**QINGHUA GUO** (S’07–M’08–SM’18) received the B.S. degree in electronic engineering and the M.S. degree in signal and information processing from Xidian University, Xi’an, China, in 2001 and 2004, respectively, and the Ph.D. degree in electronic engineering from the City University of Hong Kong, Hong Kong, in 2008. He is currently with the School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Wollongong, NSW, Australia. His research interests include signal processing, 3-D reconstruction, and telecommunications. He was a recipient of the Australian Research Council’s Discovery Early Career Researcher Award.