

# Multiple fault diagnosis of down-hole conditions of sucker-rod pumping wells based on Freeman chain code and DCA

LI Kun, GAO Xian-wen\*, YANG Wei-bing, DAI Ying-long and TIAN Zhong-da

College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China

© China University of Petroleum (Beijing) and Springer-Verlag Berlin Heidelberg 2013

**Abstract:** It is important to achieve continuous, stable and efficient pumping well operation in actual oilfield operation. Down-hole pumping well working conditions can be monitored in real-time and a reasonable production scheme can be designed when computer diagnosis is used. However, it is difficult to make a comprehensive analysis to supply efficient technical guidance for operation of the pumping well with multiple faults of down-hole conditions, which cannot be effectively dealt with by the common methods. To solve this problem, a method based on designated component analysis (DCA) is used in this paper. Freeman chain code is used to represent the down-hole dynamometer card whose important characteristics are extracted to construct a designated mode set. A control chart is used as a basis for fault detection. The upper and lower control lines on the control chart are determined from standard samples in normal working conditions. In an incompletely orthogonal mode, the designated mode set could be divided into some subsets in which the modes are completely orthogonal. The observed data is projected into each designated mode to realize fault detection according to the upper and lower control lines. The examples show that the proposed method can effectively diagnose multiple faults of down-hole conditions.

**Key words:** Sucker-rod pumping wells, multiple faults, designated component analysis, control chart, Freeman chain code, dynamometer card

## 1 Introduction

With the development of computer and artificial intelligence technology, down-hole working conditions of sucker-rod pumping wells can be monitored quickly and accurately by computer diagnosis. Reasonable technical measures can be developed to make the pumping wells achieve continuous and stable operation to improve working efficiency and oil production, which has very important significance to oilfield enterprises.

Many advanced analytical methods have been used in diagnosis of down-hole conditions in sucker-rod pumping wells. These include expert systems (Derek et al, 1988; Martinez et al, 1993), rough set theory (Wang and Bao, 2008), artificial neural networks (Rogers et al, 1990; Xu et al, 2007; De Souza et al, 2009; Wu et al, 2011), supported vector machines (Shi et al, 2004; Tian et al, 2007a; 2007b; Li et al, 2006; Li et al, 2013), spectrum analysis (He et al, 2008) and filter techniques (Li et al, 2010). However, in practical oilfield production, rational production programs are harder to develop as two or more abnormal down-hole conditions can

occur at the same time. This problem cannot be adequately solved by the existing methods. So, it of important scientific significance and practical value to develop an approach to diagnose down-hole conditions in sucker-rod pumping wells with multiple faults.

Feature extraction and the pattern classification are two important factors for the intelligent diagnosis of sucker-rod pumping wells based on dynamometer cards. In this paper, a dynamometer card is first represented by Freeman chain code (Freeman, 1961) and twelve important features of it are then extracted. Freeman chain code is a good method of compact representation of the contour of the curves and has been used in many research fields (Lu et al, 2007; Hasan et al, 2009a; 2009b; Li et al, 2009; Sanchez-Cruz and Bribiesca, 2009; Sanchez-Cruz, 2010; SiddiqiI and Vincent, 2010; Jusoh and Zain, 2011; Li et al, 2011; Brlek et al, 2011; Lu and Dai, 2011; Nawab and Hassan, 2012; Li et al, 2012). Then designated component analysis (DCA) (Liu, 2002; Camelio and Hu, 2004; Liu and Hu, 2005) is introduced to diagnose multiple faults of down-hole conditions in sucker-rod pumping wells. The observed data is projected into the designated fault modes which are defined according to the analysis of the typical dynamometer card. The upper and lower control limits can be determined from the statistical

\*Corresponding author. email: gaoxianwen@ise.neu.edu.cn  
Received August 8, 2012

data of the standard samples in normal working conditions. Then the fault detection for diagnostic samples can be done according to the upper and lower control limits.

## 2 Representation of the down-hole dynamometer card by the Freeman chain code

The surface dynamometer card is first transformed into a down-hole dynamometer card as the down-hole dynamometer card can truly reflect working conditions of the subsurface pump. Our study uses the Fourier coefficient method (Chen, 1988; Li et al, 2013) to solve the one-dimensional wave equation proposed by Gibbs in order to eliminate effects of the deformation, viscous resistance, vibration and inertia of the sucker rod string (Gibbs and Neely, 1966). The following discussion is based on the analysis of the down-hole dynamometer card.

The boundary of the down-hole dynamometer card can be represented by the Freeman chain code. Assuming that the down-hole dynamometer card is represented by  $n$  direction chain codes that are “ $c_1c_2...c_n$ ”, and the starting point of it is the starting point of load line and coding sequences transcribed in the clockwise direction. The entire route of chain code is: “loading line” → “upstroke line” → “unloading line” → “downstroke line”. Eight slopes of the short line segments are given in this paper, that are  $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$  and  $315^\circ$ , which are indicated by 0-7 respectively, shown in Fig. 1. The lengths of the horizontal and vertical segments are both  $d$  ( $d$  is decided by the sampling interval of the discrete grid sampling graphics), and the length of the diagonal segment is  $\sqrt{2}d$ .

In Fig. 2, the chain codes obtained in a counterclockwise sequence with the beginning point of  $S$  are marked as

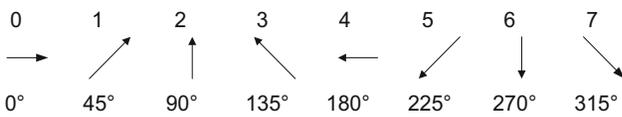


Fig. 1 Chain codes of eight directions

“01234567”, where even numbers indicate horizontal or vertical segments and odd numbers indicate diagonal segments.

It is common that different dynamometer cards have different sizes which could be influenced by many factors, such as type of pumping well and data acquisition devices, etc. Therefore, normalization of the down-hole dynamometer cards is necessary to compress the distance between data points, which will be suit to the calculation of slopes to obtain the boundary chain code with a fixed length. The normalization formulas are defined as follows:

$$\bar{x} = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

$$\bar{y} = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \tag{2}$$

where  $x_i, y_i$  represent the displacement and load data;  $\bar{x}, \bar{y}$  represent the normalized data. Normalization of one down-hole dynamometer card is given in Fig. 3.

In Fig. 4, the down-hole dynamometer card after normalization is then represented by Freeman chain code.

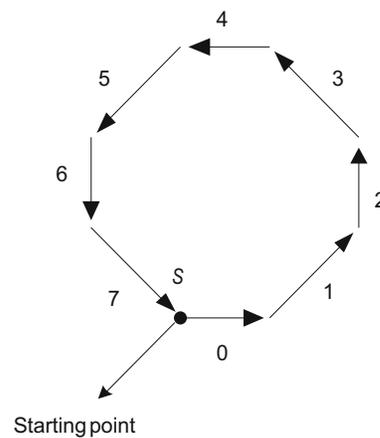
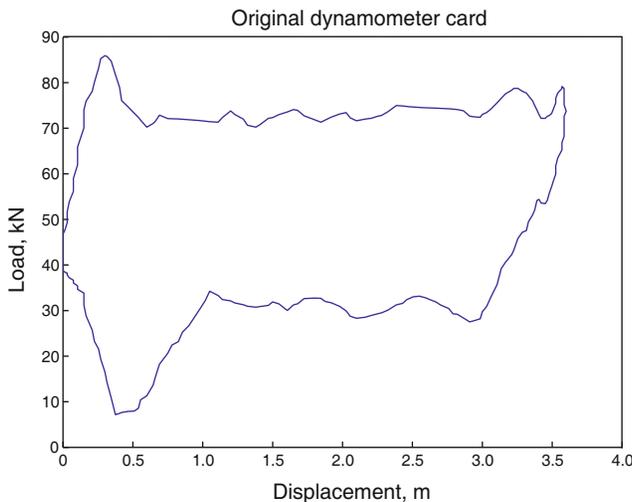
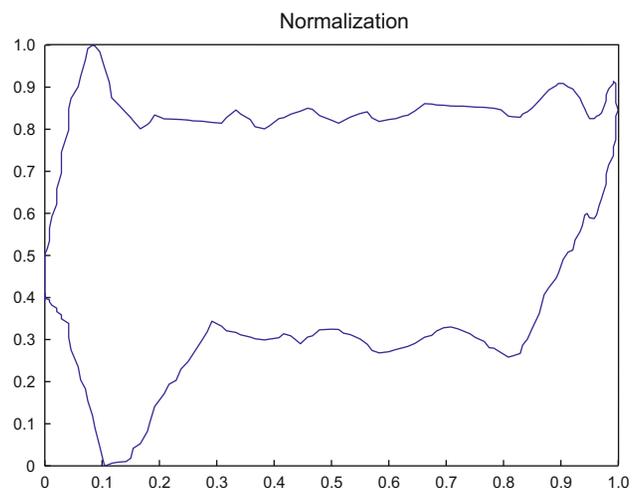


Fig. 2 Example of the Freeman chain code



(a) Original dynamometer card



(b) Normalized dynamometer card

Fig. 3 Normalization of one down-hole dynamometer card

The boundary chain code of the graph uses growth mode of square format whose lengths of the vertical and horizontal coordinates are both 1. The length of direction chain code of the axis is 1 (here, the direction chain code is considered equivalent to the horizontal and vertical segment), and the lengths of the other oblique direction chain codes are both  $\sqrt{2}$  (here, the oblique direction chain code is considered equivalent to the diagonal segment).

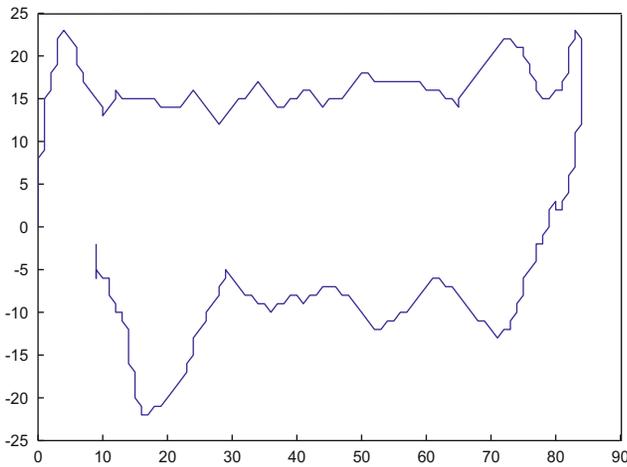


Fig. 4 Redrawn using Freeman chain code

A down-hole dynamometer card is formed by the data set  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$  whose direction degree set after transformation is  $\{\theta_1, \theta_2, \dots, \theta_n\}$ , where  $\theta_i$  ( $i = 1, 2, \dots, n$ ) is the direction degree from  $(x_i, y_i)$  to  $(x_{i+1}, y_{i+1})$  (the end point of former one is the starting point of next one). Then, the boundary chain code after transformation is “ $c_1c_2 \dots c_n$ ”.

$$c_i = \arg \min_l \left| \left( \theta_i + \theta_{i-1} - c_{i-1} \right) - \frac{\pi}{4} l \right| \quad (l = 0, 1, \dots, 7) \quad (3)$$

where  $\arg \min(\cdot)$  denotes the label  $l$  ( $l$  is one variable), which is the minimum value of a formula.

### 3 Feature extraction of the down-hole dynamometer card

In this section, 12 eigenvector parameters which can describe typical characteristics of the dynamometer card are extracted, where  $n_0-n_7$  indicate the number of chain codes 0-7 and  $n_{all}$  indicates the total number of chain codes in the whole line.

#### 1) Degree of zigzag (Zigzag)

The zigzag of the graph increases when changes of the direction are more than one angle. The factor of  $c_i \rightarrow c_{i-1}$  ( $1 < i < n$ ) is defined as:

$$b_i = \begin{cases} 1 & |c_i - c_{i-1}| \geq x \\ 0 & |c_i - c_{i-1}| < x \end{cases} \quad (4)$$

where  $b_i \in (0, 1)$ ,  $1 \leq i \leq n$ ,  $x \in (1, 2, \dots, 7)$ .

Thus, the degree of zigzag is defined as:

$$Zigzag = \frac{\sum_{i=1}^n b_i}{n_{all}} \quad (5)$$

The closer the *Zigzag* is to 1, the bigger the degree of zigzag is. Otherwise, the closer the *Zigzag* is to 0, the smoother the graph is.

#### 2) Degree of bulge of the left-bottom corner ( $B_1$ )

This feature is determined by the slope of the end of the dynamometer card during the downstroke. Assuming that chain code from  $225^\circ$  to  $270^\circ$  of the end of downstroke is a steep slope, that is, the chain code  $c_i \geq 5$ . Then  $B_1$  can be represented by the span size of the tail slope of the partial steep portion of the downstroke. That is,  $B_1$  is calculated by the span of the chain code  $c_i \geq 5$  at the end of the whole chain code.

#### 3) Degree of bulge of the right-top corner ( $B_2$ )

The maximum point of the horizontal axis of the graph is considered as the end point of the upstroke. Then the trend of the front of the upstroke is detected. The value of  $B_2$  is determined by the span size of the tail slope of the partial steep portion before the endpoint of the upstroke.

There are two steps for detection. The first step is to find the end point of the upstroke in data set of the dynamometer card; and the second step is to judge the chain code before the end point of the upstroke. The slope of the chain code is considered to be partially steep if  $1 \leq c_i \leq 2$ .  $B_2$  is calculated according to the number of continuous chain codes whose slopes are considered to be partially steep.

#### 4) Degree of flatness (Flatness)

When a sucker-rod pumping well runs normally, the sucker-rod string will be subjected to a maximum load ( $L_{max}$ ) during the upstroke and a minimum load ( $L_{min}$ ) during the downstroke.  $L'_{max}$  indicates the average maximum load (is defined as the average of a group of maximum load);  $L'_{min}$  indicates the average minimum load (is defined as the average of a group of minimum load). The *Flatness* is defined as:

$$Flatness = 1 - \frac{L'_{max} - L'_{min}}{L_{max} - L_{min}} \quad (6)$$

The graph is flat when *Flatness* is close to 1. In a limiting case, *Flatness* is equal to 1 if the graph becomes one parallel line. It is important to note that *Flatness* could be less than 0 as  $L'_{max} - L'_{min}$  is bigger than  $L_{max} - L_{min}$ , which could be ignored as the graph has no characteristic of flatness.

#### 5) Degree of lack of the left-top corner ( $A_1$ )

If the loading is slow, the posterior segment of the loading line rises in an arc-shape, which causes the left-top of the dynamometer card to be lower than in normal conditions. The chain code is represented in Fig. 5. The solid line arrows mean the normal loading line, and the dotted line arrows

mean a slow loading line.

The extraction of this feature can be transformed into calculation of the upward curvature of the posterior segment of the loading line.  $A_1$  is defined as:

$$A_1 = \frac{n_0 + 0.5n_1 - 0.5n_2}{n_{all}} \tag{7}$$

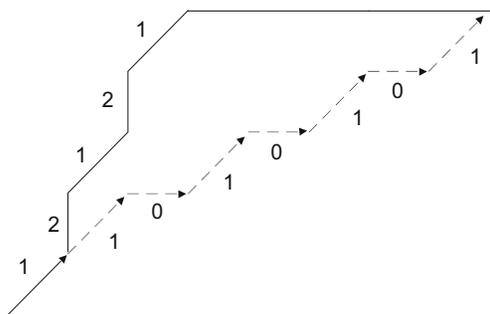


Fig. 5 Chain code of degree of lack of the left-top corner

**6) Degree of lack of the right-top corner ( $A_2$ )**

When premature unloading occurs, the posterior segment of upstroke begins to turn down, which causes lack in the right-top corner of the dynamometer card. In Fig. 6, the solid line arrows represent a normal unloading line, and the dotted line arrows show a premature unloading line.

$A_2$  is defined as:

$$A_2 = \frac{1.5n_6 + n_7 + 0.5n_0 - n_1 - 1.5n_2}{n_{all}} \tag{8}$$

The bigger  $A_2$ , the more sharp the unloading line of the posterior segment of upstroke is, and vice versa.

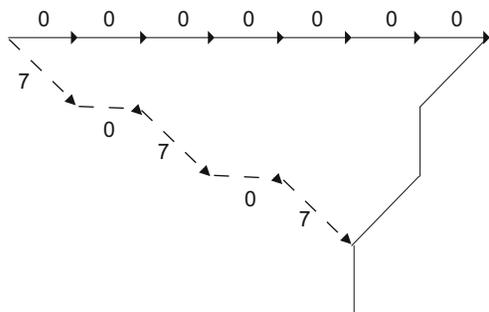


Fig. 6 Chain code of degree of lack of the right-top corner

**7) Degree of lack of the right-bottom corner ( $A_3$ )**

If the unloading is slow, the posterior segment of the unloading line begins to turn down in an arc-shape, which causes lack in the right-bottom corner of the dynamometer card. In Fig. 7, the solid line arrows show the normal unloading line, and the dotted line arrows indicate a premature unloading line.

The extraction of this feature can be transformed into calculation of the curve slope of the posterior segment of the unloading line, that is, proportion of chain code 4 in the posterior segment of the unloading line. Apparently, the more

the chain code 4, the gentler the curve slope declines.  $A_3$  is defined as:

$$A_3 = \frac{n_4 + 0.5n_5 - 0.5n_6 - n_2 - 0.5n_3}{n_{all}} \tag{9}$$

The bigger  $A_3$ , the gentler the unloading line is, and vice versa.

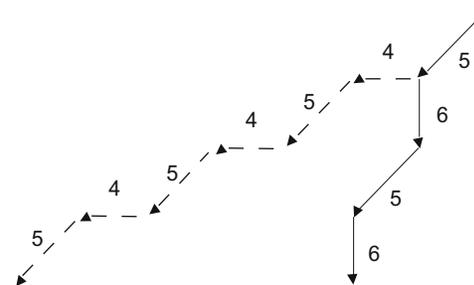


Fig. 7 Chain code of degree of lack of the right-bottom corner

**8) Degree of lack of the left-bottom corner ( $A_4$ )**

If the loading occurs in advance, the posterior segment of the downstroke begins to rise, resulting in a lack of the left-bottom of the dynamometer card. In Fig. 8, the solid line arrows show a normal loading line, and the dotted line arrows show an advanced loading line.

The extraction of this feature can be calculated by detecting whether the posterior segment of downstroke would rise.  $A_4$  is defined as:

$$A_4 = \frac{1.5n_2 + n_3 + 0.5n_4 - n_5 - 1.5n_6}{n_{all}} \tag{10}$$

The bigger  $A_4$ , the gentler the loading line is, and vice versa.

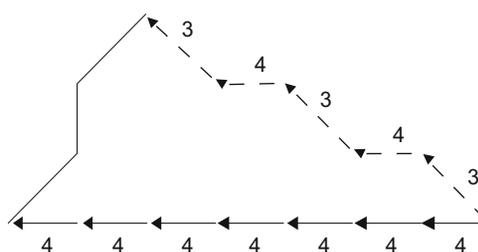


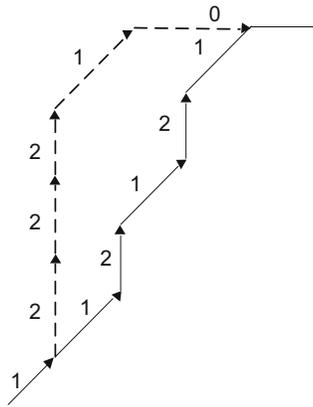
Fig. 8 Chain code of degree of lack of the left-bottom corner

**9) Degree of sharp-load of the left-top corner ( $A_5$ )**

If the load changes sharply, the front segment of the loading line rises quickly. The chain code is represented in Fig. 9. The solid line arrows are the normal loading line, and the dotted line arrows show a sharp loading line.

The extraction of this feature can be transformed into calculation of upward curvature of the loading line, and more chain code 2 represents that the curve rises steeply.  $A_5$  is defined as:

$$A_5 = \frac{1.5n_2 + n_1 - 0.5n_0}{n_{all}} \tag{11}$$



**Fig. 9** Chain code of the left-top corner with sharp loading

The bigger  $A_5$ , the greater the sharp-load level of the front segment of upstroke, and vice versa.

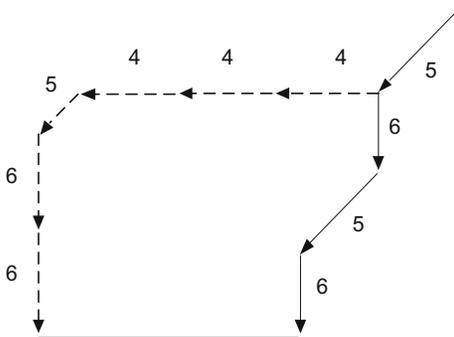
**10) Degree of sharp-unloading of the right-bottom corner ( $A_6$ )**

If the unloading changes sharply, the curvature of the whole unloading line changes substantially. The chain code is represented in Fig. 10. The solid line arrows are the normal unloading line, and the dotted line arrows show a sharp unloading line.

The extraction of this feature can be transformed into calculation of the curvature of the unloading line.  $A_6$  is defined as:

$$A_6 = \frac{0.5n_4 + n_5 + 1.5n_6 - n_2 - 0.5n_3}{n_{all}} \tag{12}$$

The bigger  $A_6$  indicates high level of sharp-unloading, and vice versa.



**Fig. 10** Chain code of the right-bottom with sharp unloading

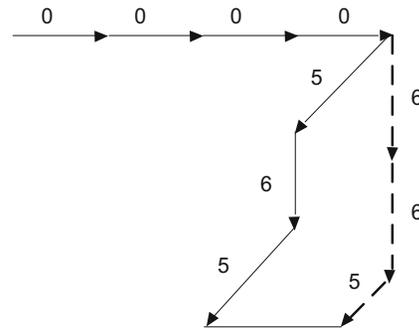
**11) Degree of rapid-unloading of the right-top corner ( $A_7$ )**

Unloading happens suddenly at the top dead center position when the end of the upstroke appears rapid unloading. The chain code is represented in Fig. 11. The solid line arrows is the normal unloading line, and the dotted line arrows is a rapid unloading line.

The extraction of this feature can be transformed into calculation of the curvature of the unloading line at the top dead center position.  $A_7$  is defined as:

$$A_7 = \frac{1.5n_6 + 0.5n_5 - 0.5n_3 - n_4}{n_{all}} \tag{13}$$

The bigger  $A_7$  indicates a high level of rapid-unloading at the top dead center position, and vice versa.



**Fig. 11** Chain code graph of the right-top corner with rapid unloading

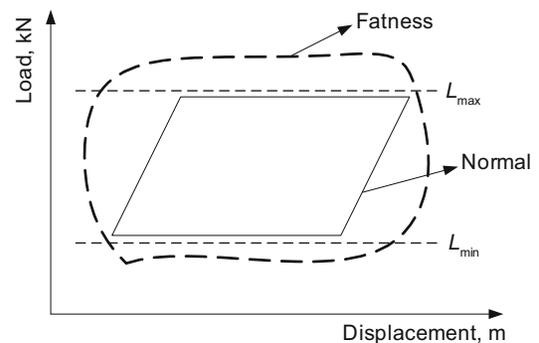
**12) Degree of fatness (*Fatness*)**

When the sucker-rod pumping well is in a normal condition, the difference of the measured maximum and minimum loads should be close to the difference of the theoretical maximum and minimum loads. However, if the dynamometer card is “fat”, difference of the measured maximum and minimum loads would become larger. A diagram is used to explain “fatness”, shown in Fig. 12.

The *Fatness* is defined as:

$$Fatness = \frac{L'_{max} - L'_{min}}{L_{max} - L_{min}} - 1 \tag{14}$$

The graph is normal if *Fatness* is equal to 0 and it is fatter when *Fatness* is close to 1. And the characteristic of the “fatness” will become more apparent if *Fatness* is more than 1. It can be ignored when *Fatness* is smaller than 0 as the graph has been no characteristic of “fatness”.



**Fig. 12** Explanation of the “fatness”

**4 DCA method**

DCA is used to project the observed data  $Y$  into designated fault mode  $d_i \in R^{p \times 1}$  (having clear physical meaning). Then

the corresponding designated component  $\omega_i$  can be obtained to determine whether the working conditions are normal or not. According to the relationship between typical faults and symptoms, fault mode  $d_i$  can be defined as:

$$d_i = [d_{i1}, d_{i2}, \dots, d_{ip}]^T, \quad i = 1, 2, \dots, l \tag{15}$$

where  $p$  means the number of symptoms;  $l$  is the number of typical faults.

$\omega_i$  is obtained when observed variable  $y$  is projected into  $d_i$ ,

$$\omega_i = d_i^T y \tag{16}$$

The following formula is obtained when sampling  $n$  times.

$$[\omega_{11}, \omega_{12}, \dots, \omega_{m}] = [d_{11}, d_{12}, \dots, d_{ip}] \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{p1} & y_{p2} & \dots & y_{pn} \end{bmatrix} \tag{17}$$

where  $i=1, 2, \dots, l$ . Then, Eq. (17) can be expressed in matrix form as:

$$W = D^T Y \tag{18}$$

where  $D$  is a matrix formed by orthogonal modes  $d_i$ , satisfying  $D^T D = I$ . Hence,

$$Y = DW \tag{19}$$

Therefore, the observed data  $Y$  can be explained by the following formula,

$$Y = \sum_{i=1}^l d_i \omega_i \tag{20}$$

where  $l$  is the number of designated modes.

Then, the fault detection can be implemented by the significant information of each designated component.

### 5 Definition of the designated mode

Designated modes can reflect working conditions of equipment, including normal mode and fault mode, which is defined by the relationship between modes and characteristics. If there are  $l$  typical faults  $d_1, d_2, \dots, d_l$ , and  $p$  eigenvectors  $u_r$  ( $r=1, 2, \dots, p$ ) can be extracted as  $d_i$  property index.  $u_r$  is defined according to the following rule,

$$u_r = \begin{cases} 1 & \text{the characteristic is obvious} \\ 0 & \text{the characteristic is not obvious} \end{cases}, \quad r = 1, 2, \dots, p \tag{21}$$

Then the mode  $d_i$  can be represented by  $p$ -dimension vector whose value is 0 or 1.

$$d_i = [d_{i1}, d_{i2}, \dots, d_{ip}]^T \tag{22}$$

12 characteristic parameters of the dynamometer card, “degree of zigzag”, “degree of bulge of the left-bottom corner”, “degree of bulge of the right-top corner”, “degree of flatness”, “degree of lack of the left-top corner”, “degree of lack of the right-top corner”, “degree of lack of the right-bottom corner”, “degree of lack of the left-bottom corner”, “degree of sharp-load of the left-top corner”, “degree of sharp-unloading of the right-bottom corner”, “degree of rapid-unloading of the right-top corner” and “degree of fatness” can be represented with  $u_1$ - $u_{12}$ ; 14 typical faults of sucker-rod pumping wells, “gas interference”, “insufficient liquid supply”, “parted rod”, “oil of high viscosity”, “leaking travelling valve”, “leaking standing valve”, “leaking travelling and standing valves”, “pump bumping (upstroke)”, “pump bumping (downstroke)”, “sand production”, “pump blocked”, “wax deposition”, “piston goes outside of cylinder”, and “large anti-impact stroke”, represented with  $d_1$ - $d_{14}$ . According to theoretical knowledge and production experience, the designated mode set is obtain in Table 1.

**Table 1** Designated mode set of the down-hole dynamometer card

	$u_1$	$u_2$	$u_3$	$u_4$	$u_5$	$u_6$	$u_7$	$u_8$	$u_9$	$u_{10}$	$u_{11}$	$u_{12}$
$d_1$	0	0	0	0	1	0	1	0	0	0	0	0
$d_2$	1	0	0	1	0	0	1	0	0	1	0	0
$d_3$	0	0	0	1	0	1	0	1	1	0	0	0
$d_4$	0	0	0	0	1	0	0	1	0	0	0	1
$d_5$	0	0	0	0	1	1	0	0	0	0	1	0
$d_6$	0	0	0	0	0	0	0	1	1	0	0	0
$d_7$	0	0	0	0	1	1	1	1	0	0	0	0
$d_8$	0	0	1	0	0	0	0	0	0	0	0	0
$d_9$	0	1	0	0	0	0	0	0	0	0	0	0
$d_{10}$	1	0	0	0	0	0	0	0	0	0	0	0
$d_{11}$	0	0	0	1	1	0	1	0	0	0	1	0
$d_{12}$	0	0	0	0	1	0	1	1	0	0	0	1
$d_{13}$	0	0	0	0	0	1	0	0	0	0	0	0
$d_{14}$	0	0	0	0	0	0	1	0	0	1	1	0

### 6 Fault detection in the projection domain based on control charts

Use of control charts is a statistical approach which can measure and evaluate the process state, and determine whether the production process is normal or not. It includes central line (CL), lower control line (LCL) and upper control line (UCL), shown in Fig. 13.

The variables would usually fluctuate over a range with the normal fluctuation of parameters and system error in the production process. The production process is steady or under

control when fluctuation of parameters is within the normal range. However, the corresponding system variables would fluctuate significantly beyond upper and lower control lines if some manmade faults and device faults occur.

The upper and lower control lines can be determined by down-hole dynamometer cards in normal working conditions. According to these control lines, fault detection of each designated mode will be implemented.

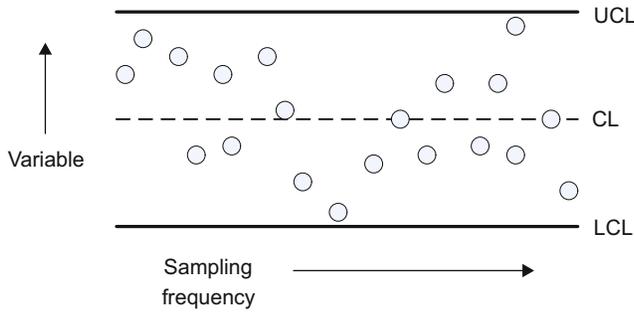


Fig. 13 Statistical control chart

6.1 Interval estimation of parameter

Given a parameter  $\theta$ ,  $0 < \alpha < 1$ . Two statistical variables  $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$  and  $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$  are determined by samples  $X_1, X_2, \dots, X_n$ , satisfying:

$$P(\underline{\theta}(X_1, X_2, \dots, X_n) < \theta < \bar{\theta}(X_1, X_2, \dots, X_n)) = 1 - \alpha \tag{23}$$

Random interval  $(\underline{\theta}, \bar{\theta})$  is the confidence interval of  $\theta$  whose confidence level is  $1 - \alpha$ .  $\underline{\theta}$  and  $\bar{\theta}$  are the lower and upper limits of the two-side confidence interval whose confidence level is  $1 - \alpha$ .

If  $\underline{\theta} = \underline{\theta}(X_1, X_2, \dots, X_n)$  and  $\bar{\theta} = \bar{\theta}(X_1, X_2, \dots, X_n)$  satisfying:

$$P(\underline{\theta}(X_1, X_2, \dots, X_n) < \theta) = 1 - \alpha \tag{24}$$

$$P(\theta < \bar{\theta}(X_1, X_2, \dots, X_n)) = 1 - \alpha \tag{25}$$

Then  $(\underline{\theta}, +\infty)$  and  $(-\infty, \bar{\theta})$  are called one-side confidence interval of  $\theta$  whose confidence level is  $1 - \alpha$ .

In this paper, the upper and lower control lines of symptom parameters of the down-hole dynamometer card are determined by the estimation of the average value  $\mu$  of parameter  $X$  in the normal condition when variation  $\sigma^2$  is unknown. The corresponding fault symptom occurs when  $X$  is beyond the confidence interval.

Although variation  $\sigma^2$  of the samples are usually unknown, the average value  $\bar{X}$  and variation  $S^2$  can be calculated.  $S^2$  is the unbiased estimation of  $\sigma^2$ , so  $\sigma$  can be replaced by  $S$ . The new sample function will not follow the normal distribution, but conform to the following distribution:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} t(n-1) \tag{26}$$

The distribution  $t(n-1)$  in Eq. (26) is independent of any unknown parameter, and the distribution function is shown in Fig. 14.

In Fig. 14, we have

$$P(-t_{\alpha/2}(n-1) < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1)) = 1 - \alpha \tag{27}$$

Then the two-sided confidence whose confidence level is  $1 - \alpha$  is as follows:

$$(\bar{X} - \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1), \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)) \tag{28}$$

The upper and lower confidence limits are:

$$\bar{\mu} = \bar{X} + \frac{S}{\sqrt{n}} t_{\alpha}(n-1) \tag{29}$$

$$\underline{\mu} = \bar{X} - \frac{S}{\sqrt{n}} t_{\alpha}(n-1) \tag{30}$$

The upper and lower confidence limits determined by Eqs. (29) and (30) can be used as the upper and lower control lines of each designated component.

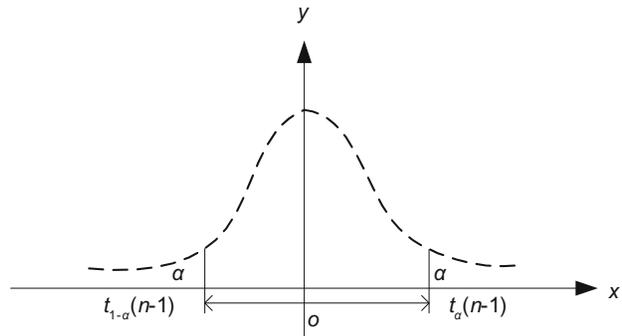


Fig. 14 Confidence interval of t distribution

6.2 Calculation of the control limit of the designated component

From Section 6.1, the upper and lower confidence limits can be used as the upper and lower control lines of each designated component according to the estimation of  $\mu$ . How to calculate the control lines of each designated component will be introduced in the following section.

It is seen that the range between the upper and lower control lines is considered to be the normal interval. When value of the observed data is beyond the control range, abnormal working conditions are considered to have occurred. In practical oil production, the pumping well usually works in normal conditions, so the control limits of each designated component can be calculated according to the down-hole dynamometer cards in normal working conditions.

From Eq. (31), the corresponding designated component  $\omega_i$  can be obtained when the observed data  $y$  is projected into the designated mode  $d_i$ ,

$$[\omega_{i1}, \omega_{i2}, \dots, \omega_{in}] = [d_{i1}, d_{i2}, \dots, d_{in}] \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{121} & y_{122} & \dots & y_{12n} \end{bmatrix} \quad (31)$$

where  $n$  indicates the number of down-hole dynamometer cards in normal working conditions;  $[y_{11}, y_{21}, \dots, y_{121}]^T$  denotes the 12 eigenvectors of the first dynamometer card;  $\omega_{in}$  denotes the mapping value of the  $n$ th normal sample in  $i$ th designated component. The mapped vector is defined as:

$$W_{d_i} = (\omega_{i1}, \omega_{i2}, \dots, \omega_{in}) \quad (32)$$

Then, the control lines of each designated component can be calculated as:

$$\bar{\mu}_i = \bar{\omega}_i + \frac{S_i}{\sqrt{n}} t_\alpha (n-1) \quad (33)$$

$$\underline{\mu}_i = \bar{\omega}_i - \frac{S_i}{\sqrt{n}} t_\alpha (n-1) \quad (34)$$

where  $\bar{\omega}_i$  denotes the average value in the  $i$ th designated component, and  $S_i$  is the variation.

### 6.3 Fault detection method of the incompletely orthogonal mode

For the fault diagnosis based on DCA, the designated component is obtained when the observed data is projected into the designated mode, which requires that all the designated modes are orthogonal to each other. However, this condition is difficult to satisfy in practical application. Therefore, it is necessary to discuss a fault detection method when the designated modes are incompletely orthogonal.

If number  $l$  of the designated orthogonal modes is equal to dimension  $p$  of the observed data, namely, a completely orthogonal mode, the observed data  $Y$  is denoted as:

$$Y = \sum_{i=1}^l d_i \omega_i \quad (35)$$

where  $l$  is the number of designated modes;  $Y$  is made up of normal subspaces and fault subspaces.

If  $l$  is smaller than  $p$ , namely, an incompletely orthogonal mode,  $Y$  can be denoted as:

$$Y = \sum_{i=1}^l d_i \omega_i + E \quad (36)$$

where  $E$  is the residual space;  $Y$  is made up of normal subspace, fault subspace and residual subspace.

For an incompletely orthogonal mode, the designated mode set could be divided into some subsets in which the modes are completely orthogonal. The observed data  $Y$  can be denoted as:

$$Y = D_1 W_1 + D_2 W_2 + \dots + D_l W_l + E \quad (37)$$

where  $D_i$  denotes the orthogonal subsets; and  $W_i$  denotes the

corresponding designated mode subsets.

According to the method proposed by Zhou et al (2009), the observed data is firstly projected into  $D_i$  to do fault detection. Then  $(Y - D_1 W_1)$  is projected into  $D_2$  to do fault detection, likewise, it is over when all orthogonal subsets are detected. It is proved that under certain conditions, the residual matrix  $E \rightarrow 0$  when  $l \rightarrow p$  (Zhou et al, 2009). Yet, there might be a situation that the orthogonal subsets  $D_i$  are composed of different  $d_i$  which are orthogonal in the group. So, the selection criteria of the best  $D_i$  should be given. The residual matrix  $E$  is denoted as:

$$E = \begin{bmatrix} \tilde{y}_{11} & \tilde{y}_{12} & \dots & \tilde{y}_{1n} \\ \tilde{y}_{21} & \tilde{y}_{22} & \dots & \tilde{y}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{y}_{p1} & \tilde{y}_{p2} & \dots & \tilde{y}_{pn} \end{bmatrix} = [\tilde{y}_1 \quad \tilde{y}_2 \quad \dots \quad \tilde{y}_n] \quad (38)$$

with

$$EE^T = \sum_{i=1}^n \|\tilde{y}_i\|^2$$

Then, the fault detection of the observed data will be carried out according to the  $D_i$  which has minimum  $EE^T$ .

From Table 1, it is an incompletely orthogonal mode as the number of orthogonal designated modes is smaller than the dimension of the observed data ( $p=14$ ). The designated mode set  $D = \{d_1, d_2, \dots, d_{14}\}$  can be represented by several completely orthogonal subsets:

$$D_1 = \{d_1, d_6\}, D_2 = \{d_2, d_4\}, D_3 = \{d_3, d_{14}\}, \\ D_4 = \{d_5, d_9\}, D_5 = \{d_7, d_{10}\}, D_6 = \{d_8, d_{12}\}, \\ D_7 = \{d_{11}, d_{13}\}$$

Then, the observed data  $Y$  can be denoted as:

$$Y = D_1 W_1 + D_2 W_2 + D_3 W_3 + D_4 W_4 + D_5 W_5 + D_6 W_6 + D_7 W_7 + E \quad (39)$$

The observed data can be fully expressed by these projection spaces when the designated mode set  $D$  is divided into seven completely orthogonal subsets. Then fault detection of the observed data in each designated component can be carried out.

### 7 Example

According to Eqs. (23)-(30), the upper and lower control lines of each designated component can be calculated from the down-hole dynamometer cards in normal working conditions. Each eigenvector is firstly modified in order to avoid calculation error caused by different measurement of them. The modified formula is given as:

$$\varphi' = |\lg|\varphi|| \quad (40)$$

The upper and lower control lines of each designated component (the confidence level is 95%) are as follows:

- $d_1$ : {LCL=0.7143, UCL=2.6819}
- $d_2$ : {LCL=1.6043, UCL=4.2547}
- $d_3$ : {LCL=1.4725, UCL=4.7032}
- $d_4$ : {LCL=1.3842, UCL=4.2348}
- $d_5$ : {LCL=0.8723, UCL=4.1336}
- $d_6$ : {LCL=1.0243, UCL=2.7455}
- $d_7$ : {LCL=1.7331, UCL=5.5017}
- $d_8$ : {LCL=0.7871, UCL=2.4090}
- $d_9$ : {LCL=0.9367, UCL=2.6333}

- $d_{10}$ : {LCL=0.5836, UCL=1.1744}
- $d_{11}$ : {LCL=1.2017, UCL=3.3985}
- $d_{12}$ : {LCL=2.2014, UCL=5.3217}
- $d_{13}$ : {LCL=0.2734, UCL=1.7964}
- $d_{14}$ : {LCL=0.7655, UCL=4.8132}

Next, we use the proposed method to analyze the fault types of 21 down-hole dynamometer cards. All diagnostic samples are shown in Fig. 15.

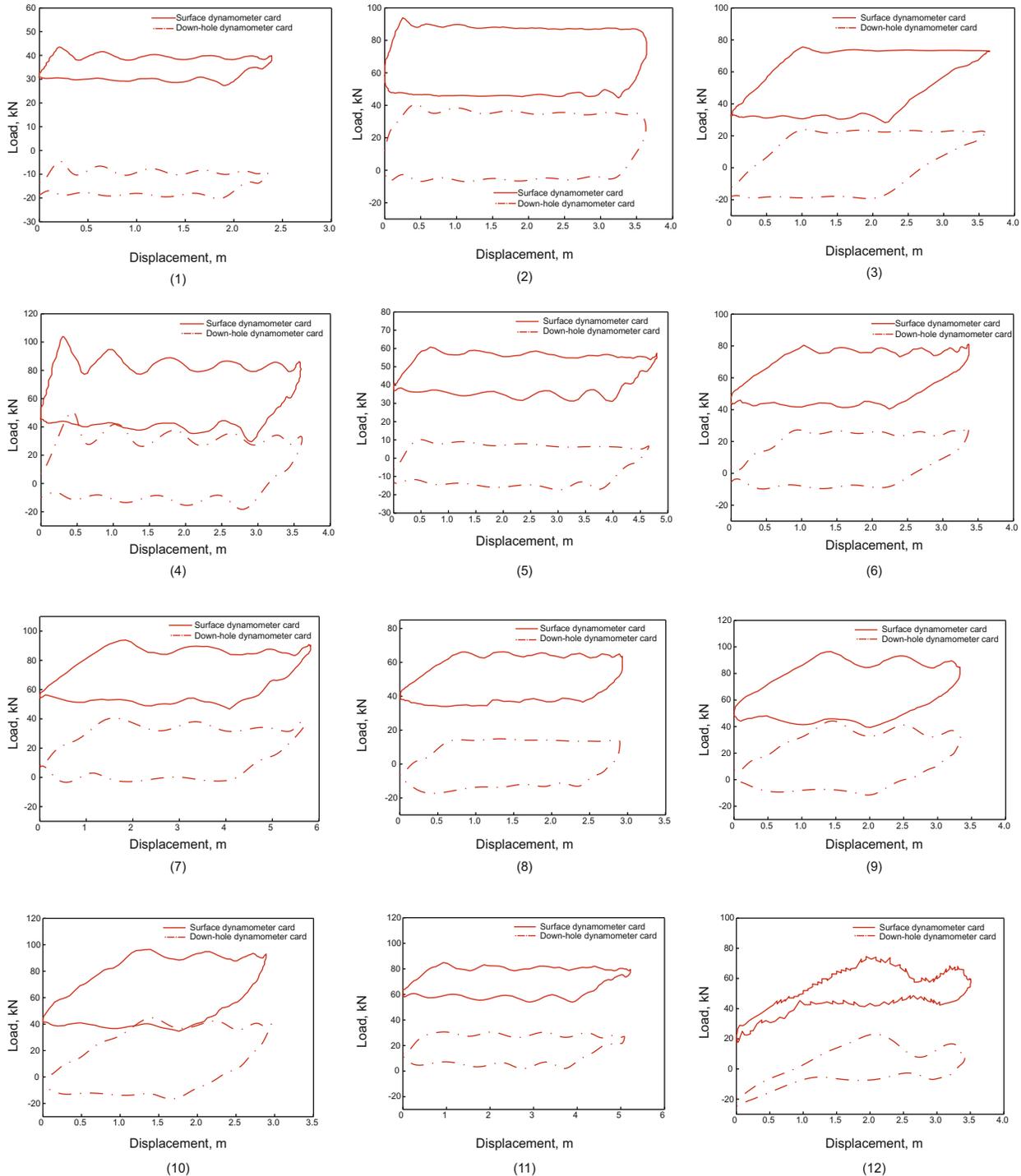
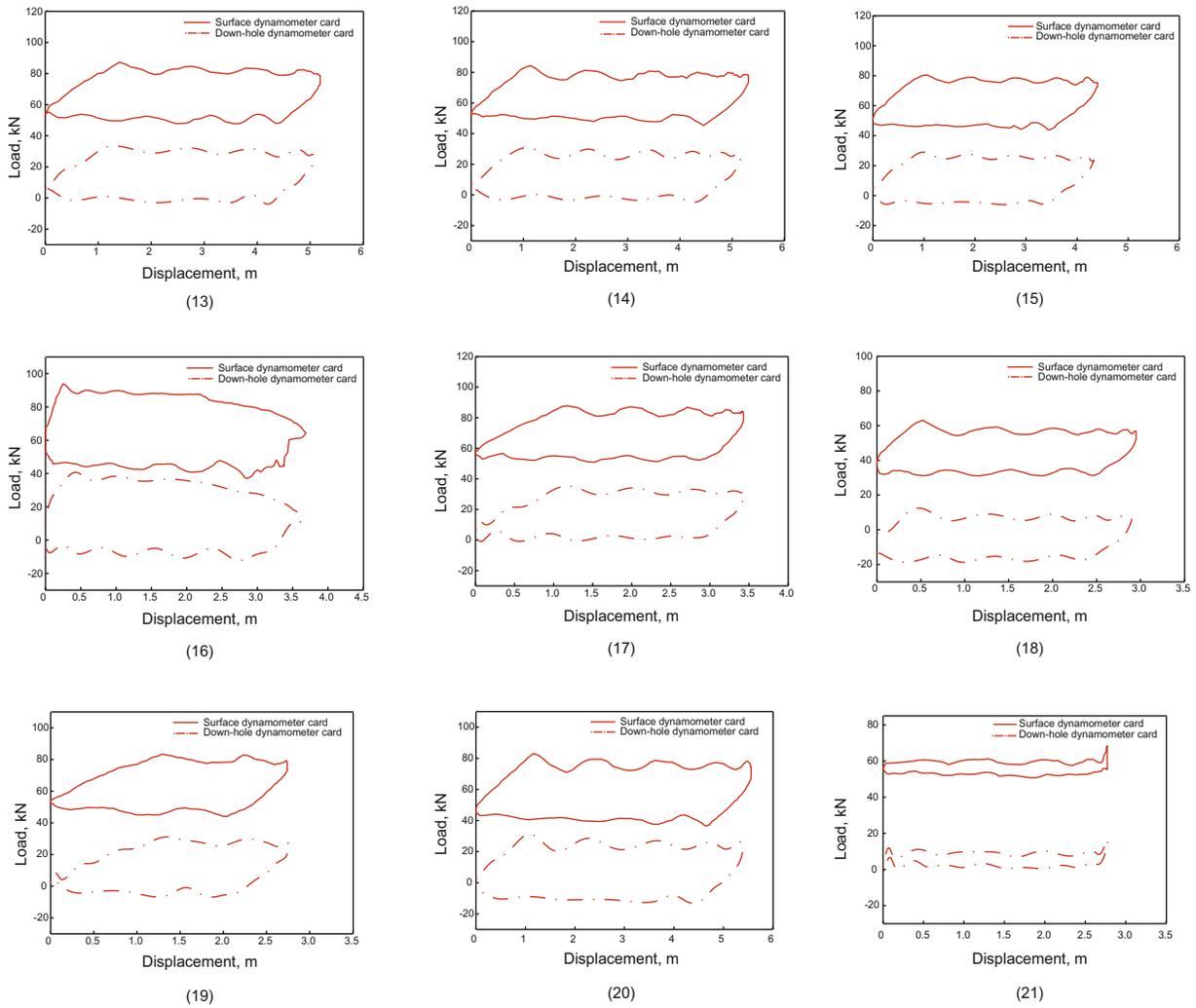


Fig. 15 Diagnostic dynamometer cards

(to be continued)

(Continued)



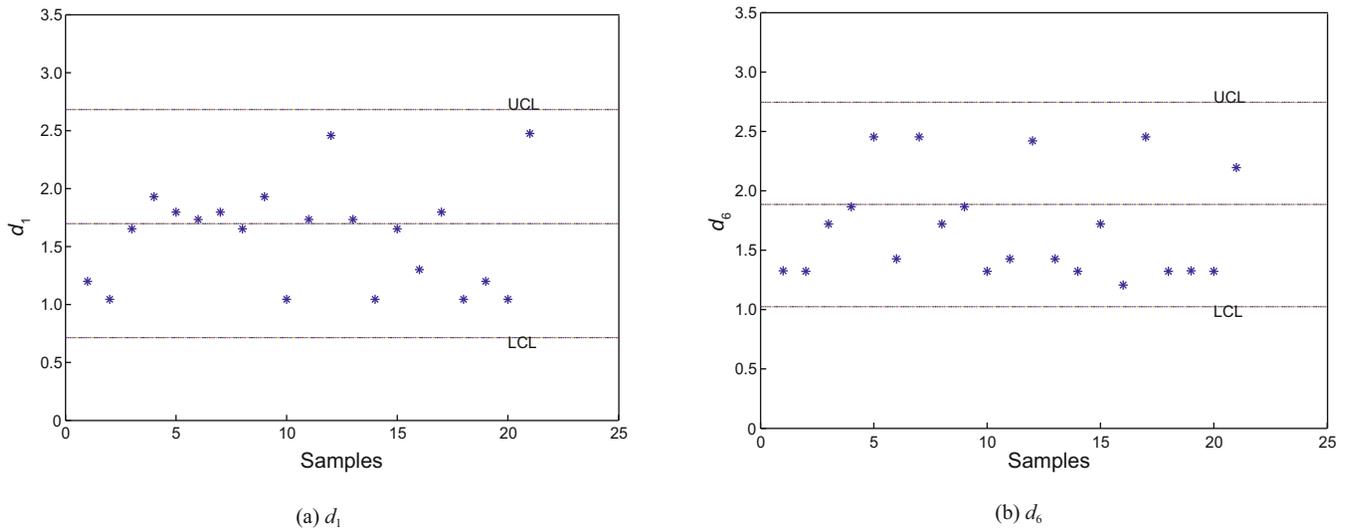
**Fig. 15** Diagnostic dynamometer cards

The observed data is projected into the first orthogonal subset  $D_1 = \{d_1, d_6\}$ . The results are shown in Fig. 16.

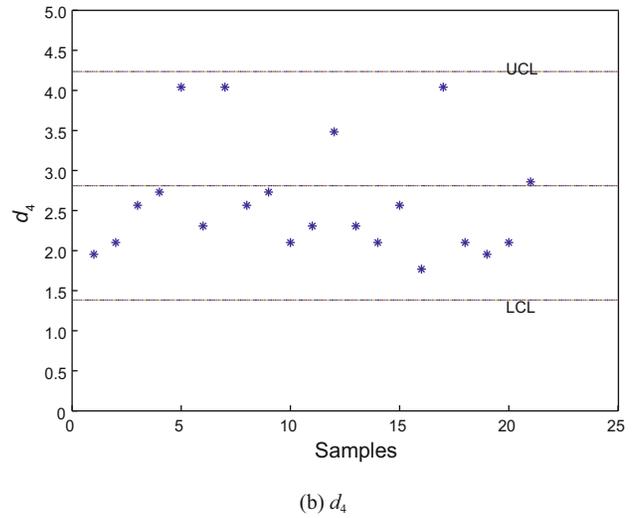
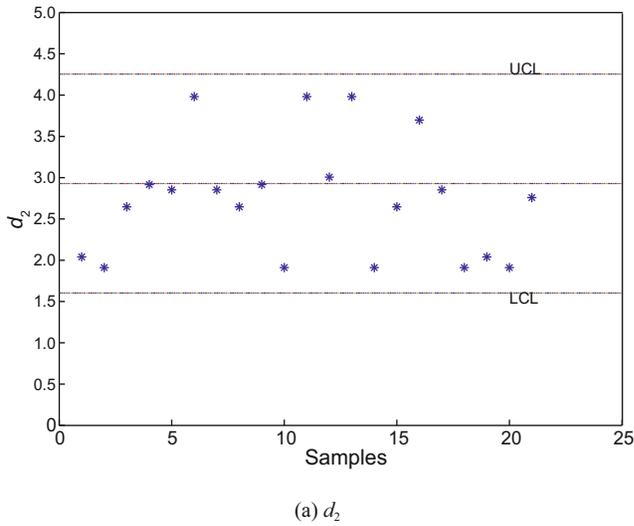
Fig. 16 indicates that none of samples are abnormal in the designated components  $d_1$  and  $d_6$ . Then the observed data is

projected into other orthogonal subspaces in order to detect whether other faults exist. The fault detection results of the second orthogonal subset  $D_2 = \{d_2, d_4\}$  are shown in Fig. 17.

Fig. 17 indicates that none of samples is abnormal in



**Fig. 16** DCA of orthogonal mode  $D_1$



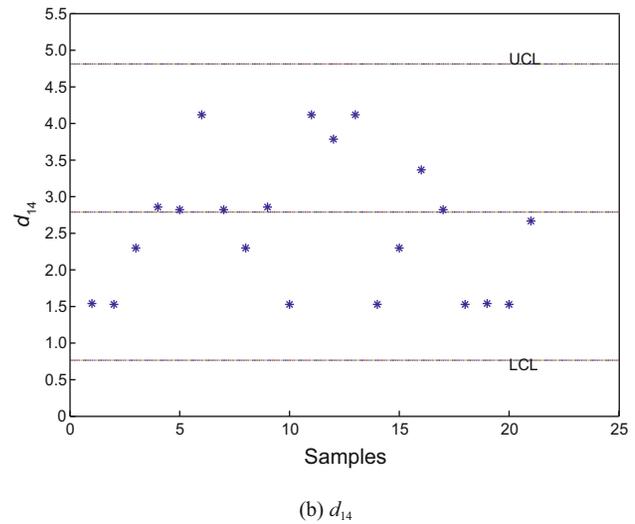
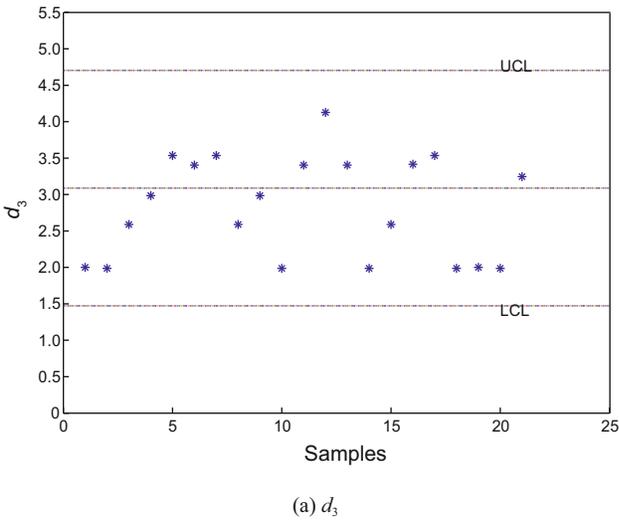
**Fig. 17** DCA of orthogonal mode  $D_2$

designated components  $d_2$  and  $d_4$ . The fault detection results of the third orthogonal subset  $D_3 = \{d_3, d_{14}\}$  are shown in Fig. 18.

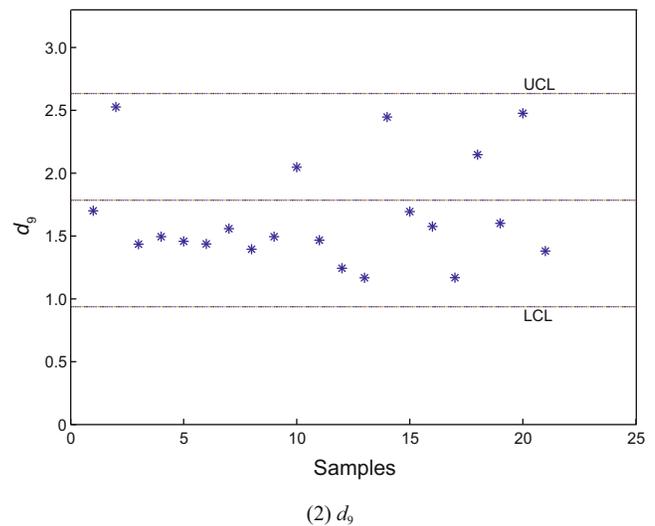
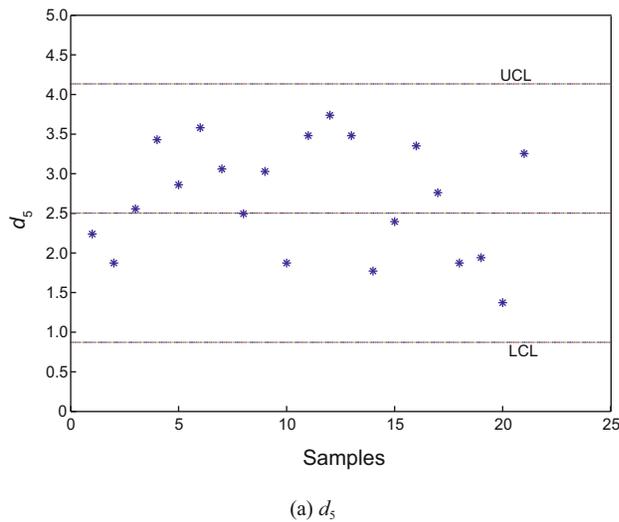
Fig. 18 indicates that sample 21 is abnormal in the designated component  $d_8$ . The fault detection results of the

fourth orthogonal mode subset  $D_4 = \{d_5, d_9\}$  are shown in Fig. 19.

Fig. 19 indicates that none of samples is abnormal in designated components  $d_5$  and  $d_9$ . The fault detection results of fifth orthogonal subset  $D_5 = \{d_7, d_{10}\}$  are shown in Fig.



**Fig. 18** DCA of orthogonal mode  $D_3$

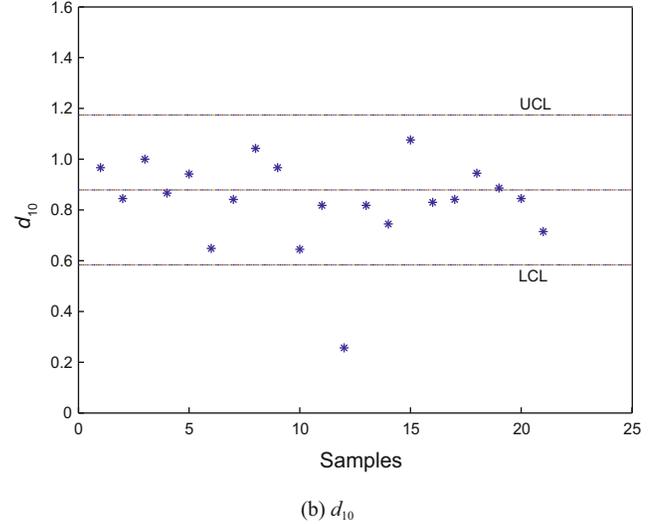
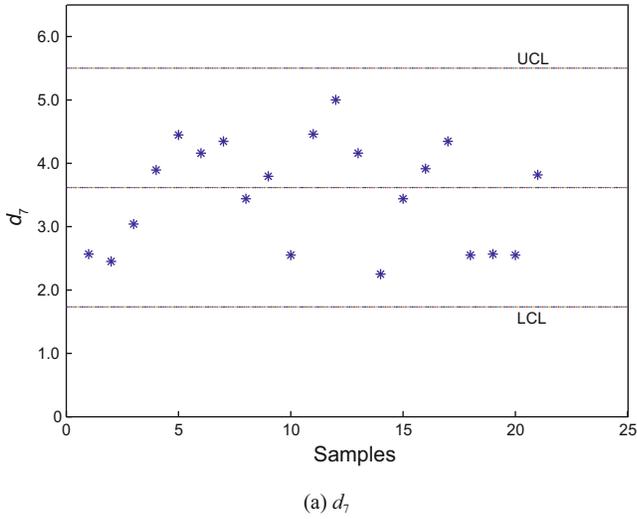


**Fig. 19** DCA of orthogonal mode  $D_4$

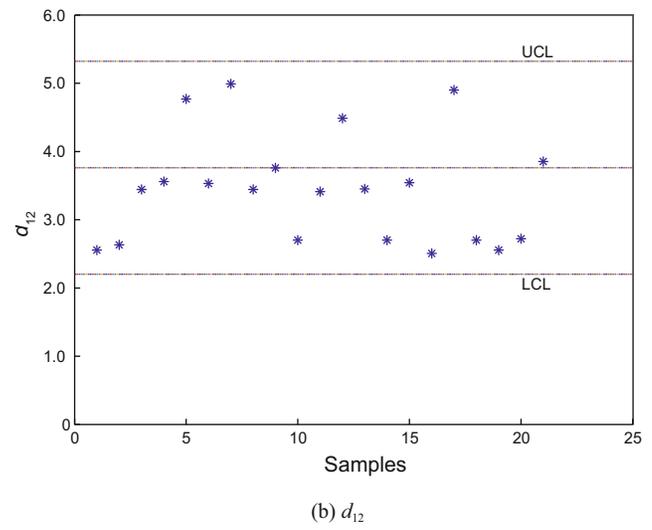
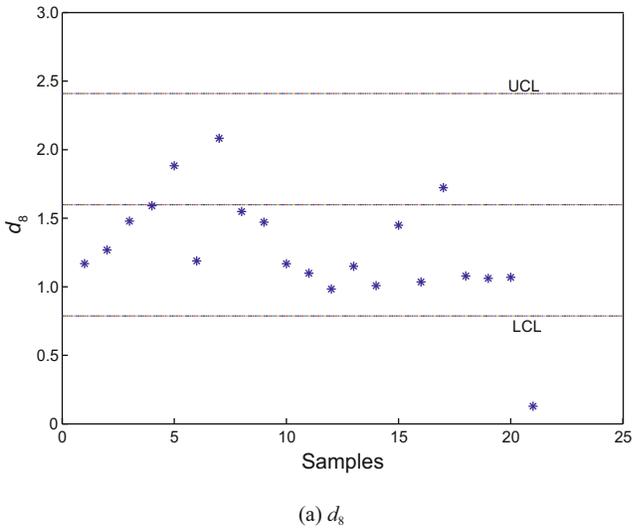
20.

Fig. 20 indicates that sample 12 is abnormal in the designated component  $d_{10}$ . The fault detection results of the sixth orthogonal subset  $D_6 = \{d_8, d_{12}\}$  are shown in Fig. 21.

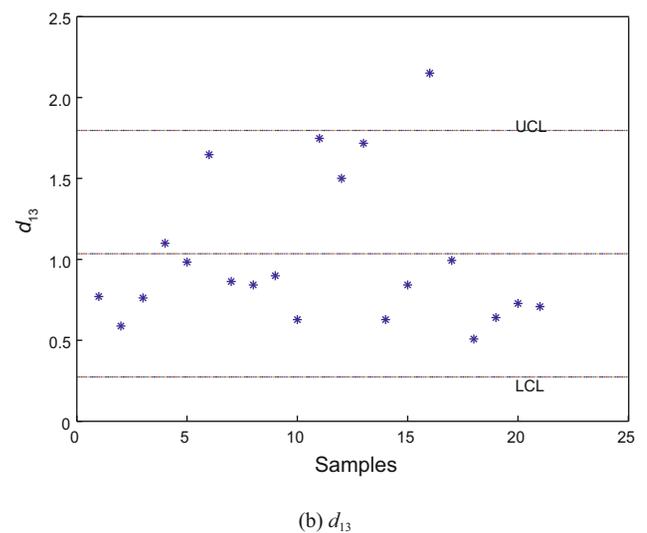
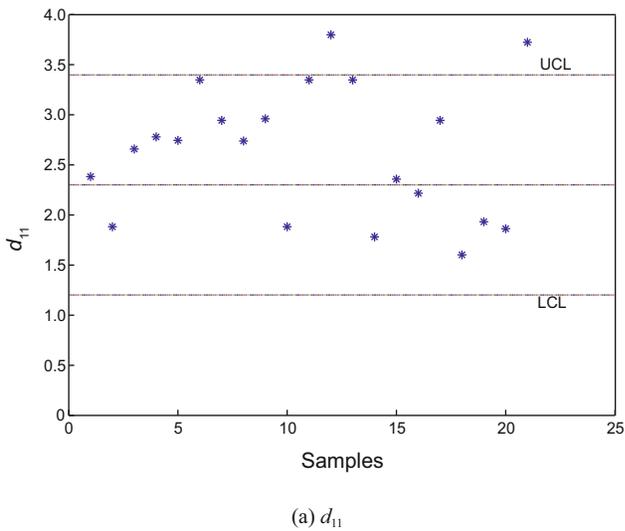
Fig. 21 indicates that sample 16 is abnormal in the designated component  $d_{15}$ . The fault detection results of the seventh orthogonal subset  $D_7 = \{d_{11}, d_{13}\}$  are shown in Fig. 22.



**Fig. 20** DCA of orthogonal mode  $D_5$



**Fig. 21** DCA of orthogonal mode  $D_6$



**Fig. 22** DCA of orthogonal mode  $D_7$

Fig. 22 indicates that samples 12 and 21 are abnormal in the designated component  $d_{11}$ , and sample 16 is abnormal in the designated component  $d_{13}$ .

According to the Figs. 16-22, the observed data is projected into seven orthogonal modes. Samples 1-11, 13-15 and 17-20 are considered to be in normal working conditions in all fault modes. Compared with the actual working conditions, they work in normal conditions based on the on-site analysis of the technical personnel. Therefore, the diagnostic conclusions are consistent with the actual

operating conditions.

Sample 12 is abnormal in the designated components  $d_{10}$  and  $d_{11}$ , which has the characteristics of “sand production” and “pump blocked”; Sample 16 is abnormal in the designated component  $d_{13}$ , which has the characteristics of “piston goes out of cylinder”; Sample 21 is abnormal in the designated component  $d_8$  and  $d_{11}$ , which has the characteristics of “pump bumping (upstroke)” and “pump blocked”. Table 2 gives the diagnostic conclusions of the proposed method and the actual working conditions for abnormal samples.

**Table 2** Conclusions of the proposed method and the actual working conditions for abnormal samples

Sample	Diagnostic conclusions	Actual working conditions
Sample 12	“Sand production” “Pump blocked”	Zigzag pattern of dynamometer card with crest tips and rapid changes; when the plunger piston moves on the upstroke, it meets resistance as the sand enters the wellbore and accumulates at the top of the plunger piston.
Sample 16	“Piston goes out of cylinder”	Relative to the requirement of long stroke in actual operation, the pump cylinder is not long enough, which causes the piston to pull out of cylinder on the upstroke; leak of oil tube, but not serious.
Sample 21	“Pump bumping (upstroke)” “Pump blocked”	When plunger piston moves on the upstroke, it meets resistance from the scaling in the oil tube; the load suddenly increases at the top stop, which is caused by eccentric wear between sucker rods and tubes.

## 8 Conclusions

A dynamometer card can effectively reflect the down-hole working conditions of sucker-rod pumping wells. However, if two or more abnormal down-hole conditions occur at the same time, there are no effective computer diagnosis methods to deal with it in current research. So in this paper work has been done to solve this problem. Freeman chain code is used to represent the down-hole dynamometer card, and 12 important eigenvectors are extracted according to analysis of the typical dynamometer card. Ten eigenvectors are calculated by changes of the curvature, and the others are calculated by the maximum and minimum load. Each fault type can be expressed by one or more eigenvectors. So, according to the DCA method, the designated mode set including 14 typical faults is constructed by the 12 eigenvectors. For each designated component, the upper and lower control lines are determined by the down-hole dynamometer cards in normal working conditions. Then the fault detection of the diagnostic samples can be implemented. When the designated modes are incompletely orthogonal, the designated mode set should be divided into some subsets in which the modes are completely orthogonal. Then the observed data is projected into each subset gradually to complete fault detection. Case studies show that the proposed method in this paper can effectively diagnose the working conditions including multiple faults.

## Acknowledgements

This research was supported by the Key Program of National Natural Science Foundation of China (61034005), Postgraduate Scientific Research and Innovation Projects of Basic Scientific Research Operating Expenses of Ministry of Education (N100604001), Excellent Doctoral Dissertations Cultivation Project of Northeastern University. We also

thank Liaohe Oilfield of China for providing us research and experimental conditions.

## References

Brek S, Koskas M and Koskas M. A linear time and space algorithm for detecting path intersection in  $Z(d)$ . *Theoretical Computer Science*. 2011. 412(36): 4841-4850

Camelio J A and Hu J S. Multiple fault diagnosis for sheet metal fixtures using designated component analysis. *Journal of Manufacturing Science and Engineering*. 2004. 126(1): 91-97

Chen J L. A fast algorithm for down-hole dynamograms in sucker rod pumping wells. *Acta Petrolei Sinica*. 1988. 9(3): 105-113 (in Chinese)

Derek H J, Jennings J W and Morgan S M. Sucker rod pumping unit diagnostics using an expert system. *Permian Basin Oil and Gas Recovery Conference*, 10-11 March 1988, Midland, Texas (paper SPE 17318)

De Souza A, Bezerra M and Filho M, et al. Using artificial neural networks for pattern recognition of downhole dynamometer card in oil rod pump system. *Proceedings of the 8th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering and Data Bases*, February, 2009, Cambridge, England, United Kingdom

Freeman H. On the encoding of arbitrary geometric configurations. *IRE Transactions on Electronics Computers*. 1961. 10: 260-268

Gibbs S G and Neely A B. Computer diagnosis of down-hole conditions in sucker rod pumping wells. *Journal of Petroleum Technology*. 1966. 18(1): 91-98

Hasan H, Haron H, Hashim S Z, et al. Logical heuristic algorithm in extracting 2D structure thinned binary image into freeman chain code (FCC). *1st International Visual Informatics Conference*, 11-13 November 2009, Kuala Lumpur, Malaysia

Hasan H, Haron H and Hashim S Z. Freeman chain code extraction using differential evolution (DE) and particle swarm optimization (PSO). *International Conference of Soft Computing and Pattern Recognition*, 4-7 December 2009, Malacca, Malaysia

He Y F, Wu X D, Han G Q, et al. Frequency spectrum analysis method

- for recognition of dynamometer card. *Acta Petrolei Sinica*. 2008. 29(4): 619-624 (in Chinese)
- Jusoh N A and Zain J M. Malaysian car plates recognition using freeman chain codes and characters' features. 2nd International Conference on Software Engineering and Computing Systems (ICSECS 2011), June 27-29, 2011, Kuantan, Malaysia
- Li C, Wang Z and Li L. An improved HT algorithm on straight line detection based on Freeman chain code. 2nd International Congress on Image and Signal Processing, 17-19 October 2009, Tianjin, China
- Li L H, Liu Y N and Liu Y K. Research on the technologies of enhancing the efficiency of Freeman direction chain code. International Conference on Electronic Information and Electrical Engineering, 15-17 June 2012, Changsha, China
- Li J, Guo S and Ye F. Shape recognition based on Freeman chain code. International Conference on Advanced Design and Manufacturing Engineering (ADME 2011), 16-18 September 2011, Guangzhou, China
- Li J Y, Chen G C, Li Z F, et al. Application of a filter technique to diagnosis of sucker-rod pumping system. *Acta Petrolei Sinica*. 2010. 31(1): 144-147 (in Chinese)
- Li K, Gao X W, Tian Z D, et al. Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a suckerrod pumping unit. *Petroleum Science*. 2013. 10(1): 73-80
- Li Z, Liu B, Liu T N, et al. The research on fault diagnoses of oil pump based on support vector machine. *Control & Automation*. 2006. 22(3): 1-3
- Liu Y G. Statistical Control of Multivariate Processes with Applications to Automobile Body Assembly. Ph.D. Thesis. University of Michigan, USA, 2002
- Liu Y G and Hu J S. Assembly fixture fault diagnosis using designated component analysis. *Journal of Manufacturing Science and Engineering*. 2005. 127(2): 358-368
- Lu D H and Dai C L. Method of recognition and feature extraction about round and rectangular in mechanical process. 4th International Conference on Machine Vision (ICMV) - Machine Vision, Image Processing and Pattern Analysis, 9-10 December 2011, Singapore
- Lu Y, Xiao C Y and Tan C L. Constructing area voronoi diagram based on direct calculation of the freeman code of expanded contours. *International Journal of Pattern Recognition and Artificial Intelligence*. 2007. 21(5): 947-960
- Martinez E R, Moreno W J, Castillo V J, et al. Rod pumping expert system. SPE Petroleum Computer Conference, 11-14 July 1993, New Orleans, Louisiana, USA (paper SPE 26246)
- Nawab N B and Hassan M M. Optical Bangla character recognition using chain-code. International Conference on Informatics, Electronics and Vision (ICIEV), 18-19 May 2012, Dhaka, Bangladesh
- Rogers J D, Guffey C G and Oldham W J B. Artificial neural networks for identification of beam pump dynamometer load cards. Annual Technical Conference and Exhibition, 23-26 September 1990, New Orleans, Louisiana (paper SPE 20651)
- Sanchez-Cruz H. Proposing a new code by considering pieces of discrete straight lines in contour shapes. *Journal of Visual Communication and Image Representation*. 2010. 21(4): 311-324
- Sanchez-Cruz H and Bribiesca E. Polygonal approximation of contour shapes using corner detectors. *Journal of Applied Research and Technology*. 2009. 7(3): 275-291
- Shi W G, Dai J J and Wang Y. FSVM based recognition approach for downhole pump dynamometer cards. *Oil Field Equipment*. 2004. 33(4): 46-48 (in Chinese)
- Siddiqi I and Vincent N. Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. *Pattern Recognition*. 2010. 43(11): 3853-3865
- Tian J W, Gao M J, Li K, et al. Fault detection of oil pump based on classify support vector machine. 2007 IEEE International Conference on Control and Automation, 30 May 30-1 June 2007, Guangzhou, China
- Tian J W, Gao M J, Liu Y X, et al. The fault diagnosis system with self-repair function for screw oil pump based on support vector machine. Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics, 15-18 December 2007, Sanya, China
- Wang J P and Bao Z F. Study of pump fault diagnosis based on rough sets theory. 2008 3rd International Conference on Innovative Computing Information and Control (ICICIC), 18-20 June, 2008, Dalian, Liaoning, China
- Wu W, Sun W L and Wei H X. A fault diagnosis of sucker rod pumping system based on wavelet packet and RBF network. *Advanced Materials Research*. 2011. 189-193: 2665-2669
- Xu P, Xu S J and Yin H W. Application of self-organizing competitive neural network in fault diagnosis of sucker rod pumping system. *Journal of Petroleum Science and Engineering*. 2007. 58(1-2): 43-48
- Zhou F N, Wen C L, Tang T H, et al. DCA Based multiple faults diagnosis method. *Acta Automatica Sinica*. 2009. 35(7): 971-982 (in Chinese)

(Edited by Sun Yanhua)