An adaptive confidence limit for periodic non-steady conditions fault detection

Tianzhen Wang, Hao Wu, Mengqi Ni, Milu Zhang, Jingjing Dong, Mohamed El Hachemi Benbouzid, Xiong Hu

Shanghai Maritime University, China
University of Brest, France

Abstract

System monitoring has become a major concern in batch process due to the fact that failure rate in non-steady conditions is much higher than in steady ones. A series of approaches based on PCA have already solved problems such as data dimensionality reduction, multivariable decorrelation, and processing non-changing signal. However, if the data follows non-Gaussian distribution or the variables contain some signal changes, the above approaches are not applicable. To deal with these concerns and to enhance performance in multiperiod data processing, this paper proposes a fault detection method using adaptive confidence limit (ACL) in periodic non-steady conditions. The proposed ACL method achieves four main enhancements: Longitudinal-Standardization could convert non-Gaussian sampling data to Gaussian ones; the multiperiod PCA algorithm could reduce dimensionality, remove correlation, and improve the monitoring accuracy; the adaptive confidence limit could detect faults under non-steady conditions; the fault sections determination procedure could select the appropriate parameter of the adaptive confidence limit. The achieved result analysis clearly shows that the proposed ACL method is superior to other fault detection approaches under periodic non-steady conditions.

1. Introduction

In process and manufacturing industries, the demands of efficient energy management and safety processes have stimulated research and development of process monitoring measures [1]. System monitoring on non-steady conditions has become a focal point since failure rate of non-steady conditions is much higher than the failure rate of steady conditions. The non-steady conditions refers to such working conditions as the system is not working steadily during startup, braking, and mutations of other process movements. The batch process is a typical feature of periodic non-steady conditions, such as: electrical and electronic products manufacturing, plastic products production, integrated circuit manufacturing, food processing, biopharmaceutical and fine chemical products [2,3].

Based on the great variety of processes as well as systems, fault detection measures can be classified in three approaches: namely, analytical [4,5], knowledge-based [6–8], and data-driven [9–11]. Traditional fault detection approaches [12–14]
experience great difficulties for system operating under non-steady conditions; especially when it is difficult to model the system and when the expertise assistance is not available [15]. This is due to the system high-dimension, complex correlation among variables, non-Gaussian distribution and signal change during non-steady conditions. In this paper, “change” refers to the system’s conditions abrupt change and indicates that it is fast time-varying. Therefore, the above particular context leads to the development of specific fault detection approaches for systems under non-steady conditions.

As analytical approaches are not available for large-scale systems, knowledge-based ones are not widely applied as system knowledge is difficult to obtain in some cases. Support Vector Data Description (SVDD) approach is able to achieve real-time and optimal control [16,17]. With the wide use of distribution control systems that leads to large amounts of generated and collected data, data-driven approaches where made available for monitoring. Statistical Process Control (SPC) plays an important role in monitoring and improving the manufacturing process [18,19]. One of the most common SPC approaches that are used for this purpose is Principal Component Analysis (PCA) [20]. It is well-known that the basic idea behind PCA is to reduce dimensionality and remove correlation of a data set, while retaining as much as possible the information variation [21]. However, there are two main limitations of PCA used for fault detection approaches. One is PCA model that once built is time-invariant, while most real-world processes are time-varying. The PCA is generally used for fault detection through a statistic confidence limit given by the Hotelling $T^2$ [22] or SPE statistics [23]. However, the traditional statistic confidence limit is not available for fault detection under time-varying conditions, and the rate of false alarm and missing alarm is high. In recent years, PCA development was therefore focused on achieving adaptive process monitoring [24]. Given this, Qiu et al. discussed a recursive robust PCA model to update the monitoring model [25]. Wang et al. proposed a fast moving window PCA in adaptive industrial process monitoring [26]. These approaches are better for slow time-varying stability process. If the complex system is under non-steady conditions, the above approaches would fail to effectively detect fault. The main reason is that variables do not follow Gaussian distribution under these conditions. This is another main limitation of PCA for fault detection. Namely, statistic construction and determination of statistic confidence limit are under the assumption that the process data are Gaussian. The monitoring results of fault detection approaches that are based on PCA may be misleading or causing false alarms, when the data follows non-Gaussian distribution. In order to improve the monitoring performance of non-Gaussian processes, Patil et al. have used box-cox transformation to detect fault [27].

Yu et al. proposed a Gaussian mixture model [28]. Ge et al. proposed a PCA-1-SVM model [29], and several applications of Independent Component Analysis (ICA) have been reported [30]. Nevertheless, compared with PCA, these approaches are much more complex and are not able to monitor time-varying processes. In addition, box-cox approach will be complex if the data dimension is high, and singularity problem may be caused by Gaussian Mixture Model (GMMs). The indeterminacy of the ICA algorithm itself will cause monitoring performance disturbance, and there exist problems in the PCA-1-SVM approach such as the selection of kernel parameters. Nomikos et al. proposed a multi-way PCA approach which is suitable for batch systems [31,32]. However, it is unable to detect position or quantity in every batch. Russell et al. proposed Dynamic Principal Component Analysis (DPCA) and Canonical Variate Analysis (CVA) approaches [33], which are suitable for dynamic systems but they still have some limitations as they do not considers non-Gaussian data conditions.

In conclusion, the problems of non-Gaussian distribution and signal changes are fault detection bottleneck in non-steady conditions. This paper will therefore focus on developing a fault detection method using adaptive confidence limit (ACL) for periodic non-steady conditions in batch process. Mathematical proofs and hardware in the loop simulation results are given to verify the validity of the proposed ACL method.

2. Adaptive confidence limit method

The adaptive confidence limit (ACL) method consists of two sections: the first one uses Longitudinal-Standardization (LS) to transform non-Gaussian normal data into Gaussian data, which meets the main pre-condition of Hotelling $T^2$ control chart; the second section is to construct the adaptive confidence limit to detect signals in real-time. In this paper, a multiperiod PCA (mPCA) algorithm is proposed to reduce dimensionality and remove correlation. A new fault sections determination procedure is presented to select the appropriate parameter value of the adaptive confidence limit promptly.

2.1. Longitudinal-Standardization (LS)

In error theory, errors exist in observations during practical industrial processes, thus the observed value ($x$) consists of the true value ($\psi$) and the random error ($\varsigma$), that is $x = \psi + \varsigma$. Errors are mostly random ones caused by many uncertainty factors and are mostly Gaussian [34,35]. The periodic non-steady conditions should satisfy the following constraint condition in the paper:

**Constraint condition**: the random error with zero mean follows Gaussian distribution.

The procedure of Longitudinal-Standardization is as follows:

Suppose $X^l$ is the $j$th period of sampling data with $n$ variables and $N$ samples, i.e.

$$X^l = [x_1^l, x_2^l, \ldots, x_n^l]$$  (1)
\[ x^j_i = \left[ x^j_1, x^j_2, \ldots, x^j_N \right], i = 1, 2, \ldots, n \] (2)

The observed value of the \( i \)th variable at the \( l \)th sampling point of the \( j \)th period can be shown as
\[ x^j_i(l) = \psi^j_i(l) + \varsigma^j_i(l), l = 1, 2, \ldots, N \] (3)

where \( \psi^j_i(l) \) and \( \varsigma^j_i(l) \) are the true value and the error of the \( i \)th variable at the \( l \)th sampling point of the \( j \)th period.

The true values are equal in different periods according to the periodicity knowledge, i.e.
\[ \psi^j_i(l) = \psi_i(l) \] (4)

where \( \psi_i(l) \) is the true value of the \( i \)th variable at the \( l \)th sampling point, then (4) could be transformed to
\[ x^j_i(l) = \psi_i(l) + \varsigma^j_i(l), l = 1, 2, \ldots, N \] (5)

Let \( A_i(l) = \left[ x^j_1, x^j_2, \ldots, x^j_i \right] \). And let \( \{ A_i(l), l = 1, 2, \ldots, N \} \) be a series of sampling data of the \( i \)th variable at the \( l \)th sampling point of different periods.

**Property 1.** \( \{ A_i(l), l = 1, 2, \ldots, N \} \) follows Gaussian distribution under periodic non-steady conditions.

**Proof.** Eq. (6) is transformed from (5) under periodic non-steady conditions.
\[ A_i(l) = \left[ \psi^j_i(l), \psi^j_2(l), \ldots, \psi^j_i(l) \right] = \left[ \psi_i(l) + \varsigma^j_1(l), \psi_i(l) + \varsigma^j_2(l), \ldots, \psi_i(l) + \varsigma^j_i(l) \right] \]
\[ = \psi_i(l) + \varsigma^j_i(l), \varsigma^j_2(l), \ldots, \varsigma^j_i(l) \] (6)

According to the above constraint condition, the random errors \( \{ \varsigma^j_1(l), \varsigma^j_2(l), \ldots, \varsigma^j_i(l) \} \) can be treated as sampled from Gaussian distributions \( N(\mu_i(l), \chi^2_i(l)) \), where \( \mu_i(l) \) is the mean value and \( \chi^2_i(l) \) is the standard deviation. Based on the additional properties of the multivariate Gaussian distributions, \( \{ A_i(l), l = 1, 2, \ldots, N \} \) is sampled from Gaussian distributions \( N(\psi_i(l) + \mu_i(l), \chi^2_i(l)) \), where \( \psi_i(l) + \mu_i(l) \) is the mean value and \( \chi^2_i(l) \) is the standard deviation. Thus Property 1 can be proved by the above steps.

**Definition.** Longitudinal-Standardization (LS) is defined as
\[ x^{\text{LS}}_i(l) = \frac{x^j_i(l) - \overline{A}_i(l)}{S^j_i(l)}, l = 1, 2, \ldots, N \] (7)

where \( \overline{A}_i(l) \) is the mean value and \( S^j_i(l) \) is the standard deviation of \( A_i(l) \). According to Property 1, \( \{ A_i(l), l = 1, 2, \ldots, N \} \) is sampled from Gaussian distributions \( N(\psi_i(l) + \mu_i(l), \chi^2_i(l)) \). According large numbers law, we obtained
\[ \overline{A}_i(l) = \lim_{J \to \infty} \frac{1}{J} \sum_{j = 1}^{J} x^j_i(l) = \psi_i(l) + \mu_i(l) \] (8)
\[ S^j_i(l) = \lim_{J \to \infty} \sqrt{\frac{1}{J} \sum_{j = 1}^{J} \left( x^j_i(l) - \overline{A}_i(l) \right)^2} = \chi^2_i(l) \] (9)

Therefore, LS can achieve the same performance as the standardization of \( x^j_i(l) \) by its mean and standard deviation as \( J \) tends to infinity.

**Property 2.** As \( J \) tends to infinity, the transformed data after LS follows standard normal distributions under periodic non-steady conditions.

**Proof.** Owing to \( x^j_i(l) = \psi_i(l) + \varsigma^j_i(l) \), and \( \psi_i(l) \) is a deterministic term, (7) can be replaced by
\[ x^{\text{LS}}_i(l) = \frac{x^j_i(l) - \overline{A}_i(l)}{S^j_i(l)} = \frac{\varsigma^j_i(l) - \mu_i(l)}{x_i(l)} \] (10)

in probability as \( J \) tends to infinity. Obviously, \( x^{\text{LS}}_i(l) \) follows a standard Gaussian distribution since the random errors \( \{ \varsigma^j_1(l), \varsigma^j_2(l), \ldots, \varsigma^j_i(l) \} \) are independent identically distributed \( N(\mu_i(l), \chi^2_i(l)) \) random variables. The transformed data follows Gaussian distributions through LS, which meets the requirement of \( J^2 \)-chart detecting fault effectively.

When the number of cycles \( J \) does not tend to infinity, Q-Q plots can be used to assess the Gaussian assumption. When the points lie very nearly along a straight line, then the normality assumption remains tenable. Normality is suspect if the points deviate from a straight line.

There is an example to verify the two properties through Q-Q plot. The historical data \( X_{\text{test}} \) has been selected with Gaussian noise as follows. The sampling numbers are 100 during a cycle, \( X_{\text{test}} = \{X_{\text{test}1}, X_{\text{test}2}, X_{\text{test}3}, X_{\text{test}4}\} \).
1. Square Wave:

\[ X_{\text{test}1} = \begin{cases} 
10, & t \in (a, a+0.5), a \in N. \\
0, & t \in (a+0.5, a+1), a \in N. 
\end{cases} \]

2. Sine Wave:

\[ X_{\text{test}2} = 30 \sin (2\pi t) + 40, t \in (0, +\infty). \]

3. Sawtooth Wave:

\[ X_{\text{test}3} = 5(t-a), t \in (a, a+1), a \in N. \]

4. Step Wave:

\[ X_{\text{test}4} = a, a = [0, 1, \ldots, 7], t \in \left[ \frac{a}{8} + l, \frac{a+1}{8} + l \right], l \in N. \]

Q–Q plots of \( A_1(35), A_2(35), A_3(35), A_4(35) \) are shown in Fig. 1, \( q(j) \) is the quantile. The shape of the points scatter is like a straight line, which proves that \( A_1(35), A_2(35), A_3(35), A_4(35) \) are taken from Gaussian distributions. A periodic data of \( X_{\text{test}} \) is selected randomly to draw Q–Q plots before and after LS as shown in Figs. 2–5. The sampling data do not follow linearity before LS. However, the transformed data are satisfying linearity after LS. This verifies the two properties of periodic transient conditions.

2.2. Adaptive confidence limit

Because the traditional straight confidence limit of the Hotelling \( T^2 \) is difficult to apply to effectively detect faults under periodic transient conditions, the dynamic data window algorithm [36], is therefore adopted to construct the adaptive confidence limit.

The adaptive confidence limit is constructed by

\[ T_{\text{uct}} = \omega \ast T_{\text{uct}1} + (1-\omega) \ast T_{\text{uct}2} \quad (11) \]

Where \( T_{\text{uct}1} \) refers to the standard confidence limit, \( T_{\text{uct}2} \) refers to confidence limit of every sampling point, \( \omega \) \((0 < \omega < 1)\) could be selected by various users request. The above three parameters \((T_{\text{uct}1}, T_{\text{uct}2} \text{ and } \omega)\) are calculated as follows:

1. Calculating \( T_{\text{uct}1} \) and \( T_{\text{uct}2} \) based on mPCA

The mPCA algorithm is proposed to combine the detection data for the purpose of reducing dimensionality and removing correlation as well as improving the monitoring accuracy. The multiperiod test data is composed by real-time test data and historical normal data. For the sake of simplicity, only one-period random historical normal data is combined with one-period real-time measured data in this paper. The details of mPCA algorithm are as follows.

![Fig. 1. Q–Q plots of \( A_1(35), A_2(35), A_3(35), A_4(35) \).](image-url)
a. **Multiperiod data**: \[ X \text{test} = \begin{pmatrix} x_{1\text{test}}(1) & \cdots & x_{n\text{test}}(1) \\ \vdots & \ddots & \vdots \\ x_{1\text{test}}(N) & \cdots & x_{n\text{test}}(N) \end{pmatrix} \] is one period real-time test data. One-period historical normal data \[ X = \begin{pmatrix} x_{1}(1) & \cdots & x_{n}(1) \\ \vdots & \ddots & \vdots \\ x_{1}(N) & \cdots & x_{n}(N) \end{pmatrix} \] is selected stochastically, where the length of sampling data is \( N \) and \( n \) is the size of variables of sampling data.

b. **LS**: \( X_{\text{test}} \) and \( X \) are respectively transformed to \( X_{\text{test}}^* = \begin{pmatrix} x_{1\text{test}}(1) & \cdots & x_{n\text{test}}(1) \\ \vdots & \ddots & \vdots \\ x_{1\text{test}}(N) & \cdots & x_{n\text{test}}(N) \end{pmatrix} \) and \( X^* = \begin{pmatrix} x_{1}(1) & \cdots & x_{n}(1) \\ \vdots & \ddots & \vdots \\ x_{1}(N) & \cdots & x_{n}(N) \end{pmatrix} \) by LS based on (10). According to **Property 2**, \( X^* \) follows standard Gaussian distribution.
c. Combination: Let

\[
\begin{align*}
Y^* &= \left( \begin{array}{c}
X^*_{\text{test}}(1) \\
\vdots \\
X^*_{\text{test}}(N)
\end{array} \right) \\
&= \left( \begin{array}{ccc}
X_1(1) & \ldots & X_n(1) \\
\vdots & \ddots & \vdots \\
X_1(N) & \ldots & X_n(N)
\end{array} \right)
\end{align*}
\]

\[
Y^* = \left( \begin{array}{c}
y_1^*(1) \\
\vdots \\
y_1^*(2N)
\end{array} \right)
\]

\[
Y^* = \left( \begin{array}{ccc}
y_1^*(1) & \ldots & y_n^*(1) \\
\vdots & \ddots & \vdots \\
y_1^*(2N) & \ldots & y_n^*(2N)
\end{array} \right)
\]

(12)

d. PCA: PCA is applied to reduce multivariables dimensionality of \( Y^* \), then calculate covariance matrix, eigenvalues, corresponding eigenvector, and the number \( m \) of PCs is decided by the cumulative percent variance.
The mPCA algorithm is not only used to reduce dimensionality and remove correlation but also used to reflect the difference between test data and historical normal data at the same sampling point.

Next $T_{act1}$ is calculated by (13):

$$
T_{act1} = \frac{m(2N-1)}{2N-m} F_{\alpha}(m, 2N-m)
$$

$T_{act1}$ is the standard confidence limit of $T^2$ statistics, $2N$ is the length of $\mathbf{Y}$*, and $m$ is the number of reserved PCs. $F_{\alpha}(m, 2N-m)$ is the critical value of $F$ distribution corresponding to the test level $\alpha$, the degree of freedom $m$ and $2N-m$. Confidence $1-\alpha$ can be determined by users' needs, for example, when confidence is 90% or 95%, then $\alpha = 0.10$ or $\alpha = 0.05$.

$T^2$ statistics can be calculated from:

$$
T_{(l)}^2 = \mathbf{Y}^2(l) \mathbf{P}_m \mathbf{A}_m^{-1} \mathbf{P}_m^T \mathbf{Y}(l)^T
$$

In (14), $l_0 = 1, 2, \ldots, 2N$, $\mathbf{A}_m = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$ that is the diagonal matrix composed of the top $m$ eigenvalues, $m$ is the number of PCs. $\mathbf{P}_m = [p_1, p_2, \ldots, p_m]$ is the load vector matrix.

This moment

$$
T_0^2 = [T^2(1), \ldots, T^2(N)]
$$

is the $T^2$ statistics of one-period normal data,

$$
T_0^2 = [T^2(N+1), \ldots, T^2(2N)]
$$

is the $T^2$ statistics of one-period real-time test data.

The validity and rationality of mPCA algorithm are as follows:

According to (14), $T^2$ statistics of the same $l$th sampling point in test data $\mathbf{X}_{test}^*$ and normal data $\mathbf{X}^*$ can be expressed as

$$
T_{(l)}^2 = \mathbf{Y}^2(l) \mathbf{P}_m \mathbf{A}_m^{-1} \mathbf{P}_m^T \mathbf{Y}(l)^T = \mathbf{X}^*(l) \mathbf{P}_m \mathbf{A}_m^{-1} \mathbf{P}_m^T \mathbf{X}^*(l)^T
$$

$$
T_{(N+l)}^2 = \mathbf{Y}^2(N+l) \mathbf{P}_m \mathbf{A}_m^{-1} \mathbf{P}_m^T \mathbf{Y}(N+l)^T = \mathbf{X}^*_{test}(l) \mathbf{P}_m \mathbf{A}_m^{-1} \mathbf{P}_m^T \mathbf{X}_{test}^*(l)^T
$$

The statistics differences are visible between test and normal data, due to the fact that $T^2$ statistics of test and normal data are calculated in the same principal component space. From (17) and (18), we obtain the $T^2$ statistics deviation of test data from normal data at the same $l$th sampling point:

$$
\Delta T^2(l) = T_{(l)}^2 - T_{(N+l)}^2 = (\mathbf{X}^*_{test}(l) - \mathbf{X}^*(l)) \mathbf{P}_m \mathbf{A}_m^{-1} \mathbf{P}_m^T (\mathbf{X}^*_{test}(l) - \mathbf{X}^*(l))^T
$$

In (19), $(\mathbf{X}^*_{test}(l) - \mathbf{X}^*(l))$ is the deviation of test data from normal data at $l$th sampling point, and with the increment of $(\mathbf{X}^*_{test}(l) - \mathbf{X}^*(l))$ the value of $\Delta T^2(l)$ is increased too. When $\mathbf{X}^*_{test}(l)$ is normal sampling data, the deviation $(\mathbf{X}^*_{test}(l) - \mathbf{X}^*(l))$ will be a small shift. And when $\mathbf{X}^*_{test}(l)$ is fault sampling data, the deviation $(\mathbf{X}^*_{test}(l) - \mathbf{X}^*(l))$ will increase and the value of $\Delta T^2(l)$ is increased too. Thus we can detect faults and improve monitoring accuracy by checking the statistics differences between test and normal data at every sampling point.

The procedures for the construction of the confidence limit $T_{act2}$ are as follows:

a. One-period normal data statistics $T_0^2$ in (15) is selected and converted into $\xi$. Let

$$
S_k = \{ \xi_{k-1}, \ldots, \xi_{k-1} \}
$$

where $S_k$ represents the contiguously section determined values of $\xi$, $l$ is the length of the dynamic data window and the values of the circulation beginning $k$, $k = L+1$.

b. We rewrite $S_k$ using (20) and calculate the value of $g_kf_k$, then we obtain the confidence limit for the $k$th value $T_{act2}(k)$.

$$
g_k = \frac{\delta_1}{2\xi}
$$

$$
h_k = \frac{2\xi^2}{\delta_1}
$$

$$
T_{act2}(k) = g_k \chi^2(\alpha, f_k)
$$

where $\xi$ is the mean value and $\delta_1$ is the variance of $\xi$ on the basis of $S_k$, $\chi^2$ is the Chi-Squared distribution function [37].

c. Judging whether factors lead to circulation turnoff. If $k \leq N, k = k+1$, the working condition will return to the step a), otherwise, the circulation condition will stop.

Through the above procedures, $T_{act2}$ is obtained.

2) Setting based on the fault sections determination procedure
A new fault sections determination procedure is proposed to set $\omega$ in this section. We set
\[ X_{uct} = 3 \]  
(24)
as the standard line. The fault sections would be located only if there are some continuous sampling points after LS above the standard line $X_{uct}$. That is, $(X_{test}^*(l) > X_{uct})$ & $(X_{test}^*(l+1) > X_{uct})$ & $(X_{test}^*(l+2) > X_{uct})$ & ….. According to the experience, the number of continuous sampling points is from 3 to 6. If the number is less than 3, the false alarm rate will be increased. If the number is more than 6, the missing alarm rate will be increased, at the same time the computational time will be increased.

The following statement will explain why the standard line is $X_{uct} = 3$?

Suppose the faults begin at $l_1$th sampling point and end at $l_2$th sampling point in test data, $l_1, l_2 \in (1,N), l_1 \leq l_2$. Thus, we can obtain fault data $X_{fault} = \begin{pmatrix} x_{1fault}(l_1) & \cdots & x_{nfault}(l_1) \\ \vdots & \ddots & \vdots \\ x_{1fault}(l_2) & \cdots & x_{nfault}(l_2) \end{pmatrix}$, and $l_3$th is a sampling point $l_1 \leq l_3 \leq l_2$. Then the fault data of first variable at $l_3$th sampling point is $x_{1fault}(l_3)$, and the normal data should as
\[ x_i(l_3) = \bar{X}_i(l_3) + \varepsilon_i(l_3) \]  
(25)
$\bar{X}_i(l_3)$ is the mean value of historical normal data. $\varepsilon_i(l_3)$ is the random errors sampled from the Gaussian distribution $N(0, \sigma_i^2(l_3))$ according to (5) and (8). Based on the $3\sigma$ principle of Gaussian distributions, i.e. if $X \sim N(\mu, \sigma^2)$, then $P_{X \in [\mu - 3\sigma, \mu + 3\sigma]} > 99.7\%$, we obtain
\[ P_{\varepsilon_i(l_3) \in [-3\sigma_i(l_3), 3\sigma_i(l_3)]} > 99.7\% \]  
(26)
The normal data of first variable at $l_3$th sampling point can therefore be transformed to
\[ X_i^*(l_3) = \frac{x_i(l_3) - \bar{X}_i(l_3)}{\sigma_i(l_3)} = \frac{\varepsilon_i(l_3)}{\sigma_i(l_3)} \]  
(27)
After LS, $X_{1fault}(l_3)$ is transformed to
\[ x_{1fault}(l_3) = \frac{e}{\sigma_1(l_3)} > 3 \text{ or } \frac{e}{\sigma_1(l_3)} < -3 \]  
(28)
According to the above analysis, it is apparent to conclude that the normal data transformed to
\[ P_{X_i^*(0 \in (-3,3))} > 99.7\% \]  
(29)
the fault data are transformed to
\[ \left| x_{fault}(l_i) \right| > 3 \]  
(30)
i = 1, 2, ..., n, l = 1, 2, ..., N after LS. Thus we set $X_{uct} = 3$
as the standard line. In order to avoid misinformation, fault sections would be located only if there are some continuous sampling points above the standard line $X_{uct}$. However, the determined fault sections are not accurate, thus fault sections determination procedure is not used to fault detection directly.

In Fig. 6, $\omega$ is chosen through matching one-period real-time test data with historical database in fault sections.

a. Setting up the database of the optimal $\omega$ and its fault sections off-line. Choose one-period fault data from history database; for $\omega = 0:0.01:1$, calculate missing and false alarm rate about different $\omega$ values by ACL; select the optimal $\omega$ according to missing alarm rate and false alarm rate (or the system requirements); save the optimal $\omega$ and its fault sections; repeat the above process in different fault sections. At last, various fault sections with the corresponding optimal $\omega$ will be saved in history database.

b. Fault sections determination. Fault sections determination procedure is used to determine fault sections of the real-time test data. When there are much more variables, the most representative variable is to be selected to estimate the fault sections.

c. Set $\omega$. According to the fault sections of the real-time test data in b), select the optimal $\omega$ value corresponding to the fault sections in the $\omega$ history database.

If the test fault sections do not match the corresponding fault sections in the historical database, then $\omega$ can be set by the flow chart in step a). The false alarm and missing alarm rates are calculated based on the determined fault sections.
2.3. ACL method procedure

The ACL method procedure is illustrated by Fig. 7.

a. **Constructing the LS standards.** Historical normal data are used to establish the normalized standards $\bar{A}_i(l)$ and $S^j_0(l)$ by (8) and (9).

b. **LS.** A period of real-time test data and a period of historical normal data randomly are transformed by (10).

c. **Fault sections determination.** According to (24)–(32), the fault sections would be located only if there exist some continuous sampling points after LS above the standard line $X_{ucl}$. The number of continuous sampling points is 6 with experience in this paper. When there are much more variables, the most representative variable is to be selected to estimate the fault sections.

d. **mPCA algorithm.** Combining the real-time test data with the historical data to a two-period test data as (12), then calculating Hotelling $T^2$ and $T_{ucl}$ by (13) and (14).

e. **Constructing the adaptive confidence limit.** Selecting $T^2_0$ from one-period normal data by (15), then using the dynamic data window algorithm to construct the adaptive confidence limit based on (11).

f. **Setting $\omega, \omega.$** $\omega$ can be set by the flow chart as shown in Fig. 1.

g. **Fault detection based on $T_{ucl}.$** Sample the statistics $T^2_1$ from one-period test data by (14), if $T^2_1 > T_{ucl}$, then the system is in abnormal situation and the real-time alarm should be triggered, otherwise, the system is normal. Then sample next period of test data back to b).
10 fault detection approaches are compared. In the following, 10 fault detection approaches are compared. The external panel is connected to the platform for hardware in the loop simulation through the I/O interface as shown in Fig. 8. The DC motor basic parameters are as follows: voltage: $U = 60$, pole pair number: $P_0 = 1$, armature resistance $R_a = 25 \Omega$, armature inductance $L_a = 0.3 \, \text{H}$, rotary inertia $I = 0.0004 \, \text{kg m}^2$, rated excitation $Ce = 0.05236$. The sampling data includes three variables: speed $n_0$, load torque $T_L$, armature current $i_a$. The mean value of the load torque is set to $0.4 \, \text{N m}$, and its variance is equal to 0.005, and the load torque changeable range is $\pm 0.1 \, \text{Nm}$; the system would malfunction if it was out of this range.

The normal historical data are obtained to calculate the mean value $\bar{x}(l)$ and the standard deviation $\sigma(l)$ from the DC motor under normal situation. Here $N = 400$, $J = 500$. Faulty (test) data are issued while load torque suddenly increases $\Delta T_L$ noise is added in all the data, and SNR is $17 \, \text{dB}$. Cumulative percent variance CPV $\geq 85\%$ is selected to determine the number of the PCs in each approach. In the following, 10 fault detection approaches are compared.

Fig. 9 shows the Q–Q plots of load torque $T_L$ in one-period normal data and one-period test data with faults. The graph shows that the curves do not conform to linearity before LS; this indicates that all the normal data and test data do not follow Gaussian distribution. After LS, the normal data follow Gaussian distribution. When $T_L$ is exceeding the normal range (that is the test data with faults), which is more easy to detect the faults after LS.

Fig. 10(a) shows the fault detection results by the PCA approach with $T^2$ statistics, and the number of PCs is 1. This figure shows that when the system is working under periodic transient conditions, the change of $T^2$-chart is obvious. The missing alarm rate attains $20.25\%$ by $99\%$ confidence limit. The $90\%$ confidence limit causes false and missing alarm. Fig. 10 (b) shows the fault detection results by the PCA approach with SPE statistics. The false alarm rate is $1.5\%$, while the missing alarm rate is up to $19.25\%$. Thus PCA approach is not useful for fault detection when the system is working under periodic transient conditions. This is mainly due to the fact that the variables do not follow Gaussian distribution and the system time-varying characteristics are not taken in consideration. In addition, the same confidence limit is used to detect different sampling points under various working conditions.

Fig. 11 shows fault detection results by the Recursive PCA (RPCA) approach when the number of PCs is 1. The missing alarm rate reaches $15.5\%$ by the $99\%$ confidence limit, which is better than the PCA approach. However, the false alarm rate reaches $4.75\%$, which is higher than the PCA’s. As seen from the figure, the detection effect of the RPCA approach is not ideal when the system is working under periodic transient conditions. Indeed, the RPCA approach is more useful for processes with slow time-varying stable signals. However, RPCA is useless in processing non-Gaussian distribution data and changing signals.

Fig. 12 shows detection results of the MPCA (Multi-way PCA) approach. The statistical model is built by 7 periods' normal data, which is used to test a periods' test data. The false alarm rate attains $0\%$, and the missing alarm rate reaches $12.5\%$ by $90\%$ confidence limit. MPCA is useful in batch process. However, as the MPCA approach does not consider non-Gaussian data conditions, it still has some limitations in comparison with ACL method. First is that the data do not follow Gaussian distribution and the system time-varying characteristics are not taken in consideration. Second is that MPCA still use the traditional $T^2$ control limit in fault detection (it is a straight line), which leads to an undesired performance.

3. Case study

In this paper, the ACL method is applied to detect faults in a plastic bag making machine DC motor who works under periodic non-steady conditions and its motion includes four conditions: acceleration, constant speed, deceleration, speed recovery. The data comes from a real-time RT-Lab platform, which achieves Hardware In the Loop (HIL) simulations. The external panel is connected to the platform for hardware in the loop simulation through the I/O interface as shown in Fig. 8. The mean value of the load torque is set to $0.4 \, \text{N m}$, and its variance is equal to 0.005, and the load torque changeable range is $\pm 0.1 \, \text{Nm}$; the system would malfunction if it was out of this range.

The normal historical data are obtained to calculate the mean value $\bar{x}(l)$ and the standard deviation $\sigma(l)$ from the DC motor under normal situation. Here $N = 400$, $J = 500$. Faulty (test) data are issued while load torque suddenly increases $\Delta T_L$ noise is added in all the data, and SNR is $17 \, \text{dB}$. Cumulative percent variance CPV $\geq 85\%$ is selected to determine the number of the PCs in each approach. In the following, 10 fault detection approaches are compared.

Fig. 9 shows the Q–Q plots of load torque $T_L$ in one-period normal data and one-period test data with faults. The graph shows that the curves do not conform to linearity before LS; this indicates that all the normal data and test data do not follow Gaussian distribution. After LS, the normal data follow Gaussian distribution. When $T_L$ is exceeding the normal range (that is the test data with faults), which is more easy to detect the faults after LS.
Fig. 13 shows detection results of the DPCA approach. It is observed that detection have achieved the expected results with better performances than the previous approaches. The missing alarm rate attains 0.25% by 90% confidence limit and false alarm rate reaches 0.75% by 99% confidence limit. However, as the DPCA approach does not consider non-Gaussian data conditions, it still has some limitations in comparison with the proposed method.

Fig. 14 shows fault detection results by the ICA approach [38], and the number of ICs is 3. The confidence limits are $\alpha_2$ and $\alpha_2/C_0/C_1$ quantiles of normal data sampling distribution, the adopted confidence is $\alpha = 0.05$. This figure shows that final detection results of IC1 and IC2 are not effective, but the detection result of IC3 is fine. Therefore, ICA approach can extract fault information effectively, and IC3 can express the fault information in test data. The detection result of IC3 is better than PCA and RPCA in terms of missing alarm rate, but false alarm rate reaches 5.75%, which is the higher one. In addition, ICA approach is much more complex than PCA and RPCA approach, requiring more computational time.

Fig. 15 shows the fault detection results by the SVDD approach [16]. In this paper, Gaussian kernel function is selected as the kernel function and its parameter $\sigma$ is set as 10. Meanwhile, the regularization constant $C$ that determines the tradeoff between the empirical error and the complexity term is set as 10 experimentally. In Fig. 15, the ordinate represents the distance from each test sample to the center of the hypersphere. The confidence limit is the radius of the hypersphere. The missing alarm rate is zero, but the false alarm rate is up to 2.75%.

Fig. 16 shows fault detection results by the PCA-1-SVM approach. Gaussian kernel function is selected as kernel function. The parameters of one-class SVM model are selected as follows: $C = 1, \sigma = 1$. The ordinate of Fig. 16 represents ratios where the numerator is the distance from every test sample to the center of the hypersphere and the denominator is the radius of the hypersphere. Thus the confidence limit is equal to 1. The false alarm rate reaches 1.25%, which is less than those of RPCA.
and ICA approaches, but the missing alarm rate is not ideal and is equal to 8%. In addition, this approach is much more complex and is inapplicable for real-time detection.

Fig. 17 shows fault detection results by the local approach-based PCA [39]. The local approach is used to normalize the non-Gaussian data based on central limit theorem. However, the local approach-based PCA method still use the traditional $T^2$ control limit in fault detection (unlike the Adaptive Confidence Limit, it is a straight line), which leads to an undesired performance. The false alarm rate is zero, but the missing alarm rate is up to 1.75%, according to the 95% standard confidence limit.

Fig. 18(a) shows the adaptive confidence limit, which is constructed by (11). Fig. 18(b) shows the final detection results by ACL. Although the false alarm is equal to 0.25%, which is not the smallest one among all the above approaches, the comprehensive detection accuracy is still the best. The main reason is that the adaptive confidence limit has considered the various conditions in different sampling time. The detection sensitivity can be adjusted by $\omega_t$ to meet the demand of missing and false alarms.

All the above detection results are summarized in Table 1.

The following conclusions can be derived from Table 1 and all the above presented detection graphs:

1. The PCA approach did not consider the periodic transient conditions characteristics.
2. In the case of RPCA, the detection process is very complex and only valid for slow time-varying stable process, therefore, the detection effect of RPCA approach is not ideal to deal with signal changes under periodic transient conditions.

3. The ICA approach effectively extracts fault information and the final detection results are better than PCA and RPCA, but it is complex and causes higher false alarm rate.

4. PCA-1-SVM is one of the distribution-free approaches for process monitoring and has improved the false alarm rate, but it is very complex and is inappropriate for real-time detection. In addition, the missing alarm rate is unsatisfactory.

5. DPCA has achieved the expected result with better performances. However, as it does not consider non-Gaussian data conditions, it still has some limitations.

6. MPCA is useful in batch processes. However, as it does not consider non-Gaussian data conditions. The missing alarm rate is higher than other methods.

Based on the above studied, tested, and criticized approaches, the ACL method uses LS to transform non-Gaussian normal data into Gaussian data, and uses dynamic data window to build the adaptive control limit, then adopts mPCA to reduce dimensionality, and calculate the $T^2$ of test and historical normal data in the same principal component space, which could improve the monitoring accuracy. As seen in Table 1, the average computational time reflects the algorithm complexity of each approach. So the ACL method seems to be less complex than ICA, RPCA, MPCA, DPCA, and PCA-1-SVM.
Fig. 13. Fault detection results by the DPCA approach.

Fig. 14. Fault detection results by the ICA approach.

Fig. 15. Fault detection results by the SVDD approach.
Table 1
Detection results of different approaches.

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Confidence limit (%)</th>
<th>False alarm rate (%)</th>
<th>Missing alarm rate (%)</th>
<th>Computational time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>90</td>
<td>1.5</td>
<td>6.75</td>
<td>0.1568</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0</td>
<td>15</td>
<td>0.1568</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>0</td>
<td>20.25</td>
<td>0.1568</td>
</tr>
<tr>
<td></td>
<td>with $T^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>1.5</td>
<td>19.25</td>
<td>0.2415</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>4.75</td>
<td>4.5</td>
<td>0.3127</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>1.5</td>
<td>15.25</td>
<td>0.3127</td>
</tr>
<tr>
<td></td>
<td>with SPE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>0</td>
<td>12.5</td>
<td>0.2713</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>0</td>
<td>18.25</td>
<td>0.2713</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>0</td>
<td>20.25</td>
<td>0.2713</td>
</tr>
<tr>
<td></td>
<td>DPCA</td>
<td>90</td>
<td>1.25</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>1</td>
<td>0.25</td>
<td>0.3108</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>0.75</td>
<td>0.5</td>
<td>0.3108</td>
</tr>
<tr>
<td></td>
<td>ICA (IC3)</td>
<td>5.75</td>
<td>0</td>
<td>0.5259</td>
</tr>
<tr>
<td></td>
<td>SVDD</td>
<td>2.75</td>
<td>0</td>
<td>2.4872</td>
</tr>
<tr>
<td></td>
<td>PCA-1-SVM</td>
<td>1.25</td>
<td>8</td>
<td>235.0931</td>
</tr>
<tr>
<td></td>
<td>Local Approach-based PCA</td>
<td>0</td>
<td>1.75</td>
<td>0.9217</td>
</tr>
<tr>
<td></td>
<td>ACL</td>
<td>90</td>
<td>0.50</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>95</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>99</td>
<td>0</td>
<td>0.75</td>
</tr>
</tbody>
</table>
Some various test data with different fault sections are used to test the ACL method with 95% confidence limit. The detection results are shown in Table 2. As seen from these results, the determined fault sections are similar with the real sections and all the detection results of various test data are ideal. The ACL method, therefore, could detect various test data effectively and the influence of fault sections is negligible.

The effect of ACL method is tested with different degrees of failure. For the normal data, the load torque $\Delta T_L$ changes in the range $\pm 0.1$ N m. In the test data, the fault sampling points are 50–250, where the $\Delta T_L$ changes in varying degrees as shown in Table 3. The fault detection results are shown in Table 3. It shows that the fault detection results are ideal even for tiny degree failure. So the ACL method is also useful for tiny failure detection.

<table>
<thead>
<tr>
<th>Fault sections</th>
<th>Determined fault sections</th>
<th>False alarm rate (%)</th>
<th>Missing alarm rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–0</td>
<td>0–0</td>
<td>0.75</td>
<td>0</td>
</tr>
<tr>
<td>0–80</td>
<td>0–85</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>200–280</td>
<td>200–284</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>0–100</td>
<td>0–104</td>
<td>0.5</td>
<td>0</td>
</tr>
<tr>
<td>100–250</td>
<td>100–255</td>
<td>0.25</td>
<td>0</td>
</tr>
<tr>
<td>50–250</td>
<td>50–256</td>
<td>0</td>
<td>0.25</td>
</tr>
<tr>
<td>100–350</td>
<td>100–355</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>0–400</td>
<td>0–394</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**Table 2**

Detection results for different fault sections.

![Adaptive Confidence Limit](image1.png)

**Fig. 18.** Fault detection results by the ACL method. (a) Adaptive confidence limit of 95%. (b) Detection result.
To analysis the effectiveness of the ACL method in different SNR (Signal Noise Ratio), we have increased the Gaussian white noise intensity in the simulation model to examine the influence of noise on the final detection results, and observed the change of missing alarm rate and false alarm rate. The load torque of the test data suddenly increase to 0.3 Nm during the sampling points 50–250, exceeding the normal range. And the Gaussian white noise SNR range is 0–100 dB. Fig. 19 shows the change of missing alarm rate and false alarm rate during the different Gaussian white noise intensities. As illustrated by Fig. 19, the Gaussian white noise will change false and missing alarm rates significantly when the SNR is low. With the increment of SNR the false and missing alarm rates converge to a fixed value, the false alarm rate to 0 and the missing alarm rate to 0.25%.

4. Conclusions

This paper deals with fault detection in systems operating in periodic non-steady conditions. An ACL fault detection method is proposed, which includes LS, mPCA and fault sections determination procedure. LS allows transforming normal data from non-Gaussian distribution into standard Gaussian distribution that meets the main pre-condition of Hotelling $T^2$ control chart to effectively detect faults. To solve signal change problems, and also to reduce dimensions and to remove correlation, an mPCA algorithm and a fault sections determination procedure are adopted.

The proposed ACL method has the following advantages: 1) allows detecting non-Gaussian distribution data; 2) allows reducing dimensionality and remove correlation; 3) setting up adaptive confidence limit to detect signal changes in real-time; 4) allows adjusting the monitoring sensitivity by regulating $\omega$ to achieve a balance between missing and false alarms. 5) sensitivity to faults, even for some tiny faults.

The carried-out simulations have clearly shown that the proposed ACL method is not only suitable for fault monitoring under periodic non-steady conditions, but could also be effectively used to detect faults and reduce false alarms.

Acknowledgments

This paper was supported by the projects of National Natural Science Foundation of China (NSFC) (61203089, 61304186 and 61403229) and Innovation Key Project of Shanghai Municipal Education Commission (14ZZ141). The authors thank the anonymous reviewers for their useful comments and suggestions.
Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ymssp.2015.10.015.

References