Robust Adaptive Boosted Canonical Correlation Analysis for Quality-relevant Process Monitoring of Wastewater Treatment

Hongchao Cheng[a,b], Jing Wu[a], Daoping Huang[a], Yiqi Liu[a], Qilin Wang[b]

[a]School of Automation Science and Engineering, South China University of Technology, Guangzhou, China

[b]Centre for Technology in Water and Wastewater, School of Civil and Environmental Engineering, University of Technology Sydten, NSW, Australia

Abstract

Quality-relevant process monitoring has attracted much attention for its ability to assist in maintaining efficient plant operation. However, when the process suffers from non-stationary and over-complex (with noise, multiplicative faults, etc.) characteristics, the traditional methods usually cannot be effectively applied. To this end, a novel method, termed as Robust adaptive boosted canonical correlation analysis (Rab-CCA), is proposed to monitor the wastewater treatment processes. First, a robust decomposition method is proposed to mitigate the defects of standard CCA by decomposing the corrupted matrix into a low-matrix and a sparse matrix. Second, to further improve the performance of the standard process monitoring method, a novel criterion function and control charts are reconstructed accordingly. Moreover, an adaptive statistical control limit is proposed that can adjust the thresholds according to the state of a system and can effectively reduce the missed alarms and false alarms simultaneously. The superiority of Rab-CCA is verified by Benchmark Simulation Model 1 (BSM1) and a real full-scale wastewater treatment plant (WWTP).

Keywords: Canonical correlation analysis (CCA), adaptive threshold, fault detection, quality-relevant, wastewater treatment.

1. Introduction

Safe and stable operation of WWTPs is crucial for energy conservation and environmental protection [1]. However, real WWTPs are always exposed to hostile working

environments. Sensor faults, machine faults or process faults, such as filamentous sludge bulking, are likely to frequently occur, especially for most small- and medium-sized WWTPs in rural areas [1]. Additionally, the WWTP is a complex industrial system with a mixture of physical, chemical and biological reactions [2]. If a fault is not satisfactorily recognized and removed in time, it will not only result in the huge economic losses but also cause secondary pollution of the environment. Moreover, the WWTP will be penalized by environmental protection agencies.

The data-driven process monitoring method is a popular and powerful tool for plant management because it can achieve better performance without requiring prior knowledge [3-8]. Multivariate statistical methods, such as principal component analysis (PCA) [6] and support vector machine (SVM)[9], are among the most popular data-driven methods and have been successfully used for process monitoring in industrial processes. Tian et al. proposed use of a multiblock strategy and PCA to monitor plant-wide processes [6]. Unfortunately, quality-relevant faults cannot be further identified and diagnosed satisfactorily by the above methods [3]. The multivariate statistical method, CCA, is able to represent the correlation between two groups of variables, which can facilitate effective diagnosis of the quality-relevant faults. In recent decades, the CCA-based fault diagnosis method has been applied in medicine, imaging, high-speed training and other fields [10-13]. Nonetheless, the CCA-based method is rarely used in energy-related industries. In this paper, the CCA-based method is applied to monitor the WWTP quality-relevant faults and to reduce the unnecessary energy consumption.

The CCA-based method usually focuses on extracting the correlation structure between input and output variables. Jiang et al. studied the improved CCA for fault detection in the local TE (Tennessee Eastman) process [11]. Although decomposition of a large-scale industrial system into local components can improve the diagnostic performance, the inherent system structure is also be destroyed to some extent by decoupling the interacting relationship leading to missing data structure. At the same time, the robustness of the standard method is usually not fully considered, which likely produces missed alarms or false alarms in the case in which the parameters or hyperparameters are not set up carefully. Chen et al. proposed a generalized CCA-based method to monitor the system faults with an application to the traction drive control system (TDCS). In this method, a random algorithm was used to optimize the control limit and ensure that the method can recognize the working condition [10]. However, under complex conditions, the generalized CCA-based method may not be able to effectively identify the fault condition. Additionally, when the internal data structure is corrupted by noise, the trained model deviates significantly [14]. To mitigate this issue, robust

decomposition is used to treat the data matrix and ensure a low-rank matrix containing essential information. In turn, the performance of the CCA-based method can be effectively improved by recovering an original data matrix using a low-rank matrix. Recently, some researchers have focused on the issue in which the data matrix is corrupted by anomalies or noise [14-16]. Candès et al. proposed use of the principal component pursuit (PCP) method to estimate the low-rank matrix, and subsequently applied the estimated matrix to improve the method performance [16]. Pan et al. proposed an improved PCP method to recover the low-rank matrix from the corrupted matrix [15].

The low-rank matrix estimation methods are used to improve the CCA. Branco et al. compared different low-rank matrix estimation algorithms to assess the impact on the CCA-based method. The experiment showed that the effective recovered algorithm can indeed improve CCA-based fault diagnosis performance [14]. However, the above studies cannot make full use of the sparse matrix, and the decomposed sparse matrix may be mixed with useful information. In this paper, a new criterion function is established to achieve the purpose of multiobjective optimization. The criterion function can simultaneously maximize the correlation of the low-rank matrix, and minimize the correlation of the sparse matrix. Thus, the low-rank matrix and the sparse matrix can be effectively utilized together and the corresponding robust loading matrix (RLM) can be further obtained.

T^2 control chart and SPE (square prediction error) control chart are commonly used fault detection tools [17-19] that are derived from the PCA, CCA, and PLS (partial least squares) methods. Subsequently, these tools are used to justify the system working condition. However, the traditional contribution plot (CP) always results in misidentification if the working condition is too complex [20]. To improve the diagnostic performance of CP, Alcala et al. proposed a reconstructed CP method [20]. However, the reconstruction method is only suitable for a single fault and does not properly consider fault detection accuracy [21]. To detect quality-relevant faults by CCA, Zhu et al. divided the data space into five components [22]: quality-relevant T^2-based control chart, input space quality-irrelevant T^2 and SPE control chart, output space quality-irrelevant T^2 and SPE control chart. Chen et al. constructed a

residual generator by using the input and output state space, which can divide the control chart into four components [10]. Based on the above studies, the input-output state space is divided into two components in this paper, the quality-relevant T^2 control chart and the quality-irrelevant SPE control chart. At the same time, the threshold of the standard control chart is taken as an invariant constant set up by the χ^2 distribution or F distribution. Generally, these invariant thresholds cannot properly cope with the industrial dynamics, which continuously leads to false or missed alarms [23]. Therefore, a new adaptive control limit (ACL) is proposed in this paper by considering both the dynamics of historical data and real-time data.

In this study, considering that the traditional CCA method cannot efficiently decompose the corrupted matrix, it cannot adaptively adjust the control limit of the control chart according to different working conditions. A novel fault detection method, termed Rab-CCA, is proposed to monitor wastewater treatment processes. Rab-CCA can effectively decompose the corrupted matrix. As scuh, the information of the corrupted matrix can be effectively extracted and used. In addition, a novel adaptive control limit (ACL) is proposed in this paper. The threshold of Rab-CCA can be adaptively adjusted according to different fault scenarios, which can reduce false alarms and missed alarms accordingly. Overall, the main contributions of this study can be summarized as follows:

(1) Considering that the traditional CCA method cannot handle a data matrix that is corrupted by anomalies or noise, the inexact augmented Lagrange algorithm (IALM) is introduced to recover the low-rank matrix containing essential and hidden information of a system. A multi-objective criterion function is proposed and constructed in this paper to obtain the corresponding robust loading matrix. By doing so, the Rab-CCA is able to robustly deal with anomalies or noise.

(2) Based on the robust loading matrix, a quality-relevant T^2 control chart and quality-irrelevant SPE control chart are reconstructed, and the reconstructed control chart shows better performance. Additionally, the ACL is further proposed to work together with T^2 and the SPE control plots to reduce false alarms and missed alarms by resorting to adaptive

strategies.

(3) The Rab-CCA-based method is proposed to monitor real wastewater treatment processes. Unlike traditional studies, the Rab-CCA-based method is able to monitor quality-relevant faults under different scenarios (device faults or sludge bulking).

2. Preliminaries

2.1. Canonical correlation analysis

CCA is a multivariate statistical method available for quality-relevant fault monitoring. Let $X \in R^{n\times p}$ and $Y \in R^{n\times q}$ be the input data matrix and output matrix, respectively, where p represents the number of input variables, q represents the number of output variables, and n represents the number of samples. The correlation coefficients of $\{x_1 \cdots x_p\}_X$ and $\{y_1 \cdots y_q\}_Y$ can be solved by the corresponding linear combination:

$$\rho(\alpha X^T, \beta Y^T) = \frac{\alpha \Sigma XY \beta^T}{\sqrt{\alpha \Sigma_{XX} \alpha^T} * \sqrt{\beta \Sigma_{YY} \beta^T}}, \quad (1)$$

where $\alpha = [\alpha_1 \ \cdots \ \alpha_p]$ and $\beta = [\beta_1 \ \cdots \ \beta_q]$ represent the nonzero constant vector (the combination coefficients of X and Y, respectively), and $\sum(*)$ is the covariance matrix. To ensure the uniqueness of the result, let $\alpha \Sigma_{XX} a^T = 1$ and $\beta \Sigma_{YY} \beta^T = 1$. The following optimization formula can be derived:

$$\arg\max_{W_X, W_Y} \rho(W_X, W_Y) = \arg\max_{W_X, W_Y} W_X{}^T \Sigma_{XX}{}^{-1/2} \Sigma_{XY} \Sigma_{YY}{}^{-1/2} W_Y$$

$$\text{s.t} \quad W_X{}^T W_X = 1, \ W_Y{}^T W_Y = 1, \quad (2)$$

where $W_X = \Sigma_{XX}{}^{1/2} \alpha^T$, $W_Y = \Sigma_{YY}{}^{1/2} \beta^T$. To solve the formula (2), a singular decomposition (SVD) is performed on the matrix $Z = \Sigma_{XX}{}^{-1/2} \Sigma_{XY} \Sigma_{YY}{}^{-1/2}$.

$$Z = U \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix} V, \quad (3)$$

where U and V represent the left and right singular value vectors, respectively. $\Lambda_r = diag(\gamma_1, \ \cdots \ , \gamma_r)$ is the singular value, r=rank(Z), and $r \leq \min(p, q)$. Thus, the canonical correlation matrix can be expressed as $P = \Sigma_{XX}{}^{-1/2} U(:, 1: r)$, and $L = \Sigma_{YY}{}^{-1/2} V(:, 1: r)$. Additionally, the redundancy noncorrelation matrix can be expressed as $P_{res} = \Sigma_{XX}{}^{-1/2} U(:, r+1: p)$ and $L_{res} = \Sigma_{YY}{}^{-1/2} V(:, r+1: q)$.

2.2. CCA-based fault detection

Generally, the fault vector can be represented as follows:

$$u = u_{in} + u_{out} = u^*_{in} + \theta_{in} f_{in} + u^*_{out} + \theta_{out} f_{out}, \quad (4)$$

where u^*_{in} and u^*_{out} represent the fault-free input and output space, respectively; θ_{in} and θ_{out} represent the fault amplitude of the input space and output space, respectively; and f_{in} and f_{out} are the corresponding fault direction vectors. The fault amplitude and direction are usually not available in advance [21]. Therefore, the data space can be roughly divided into four components by combining the CCA model with T^2 and SPE control plots:

$$T^2_X = XP\Sigma^{-1}_{in}P^T X^T, \quad (5)$$

$$T^2_Y = YL\Sigma^{-1}_{out}L^T Y^T, \quad (6)$$

$$SPE^{in}_{Xres} = XP_{res}P^T_{res}X^T, \quad (7)$$

$$SPE^{out}_{Yres} = YL_{res}L^T_{res}Y^T, \quad (8)$$

where T^2_X and SPE^{in}_{Xres} are used for the input space to monitor input -output-relevant faults and input-output-irrelevant faults, respectively; T^2_Y is used for the output space to detect the output-input-relevant faults; and SPE^{out}_{Yres} is used for the output space to detect the output-input-irrelevant faults [22]. The dimension of the correlation space is usually determined by the rank(Z) [24]. Therefore, $\Sigma^{-1}_{in} = \Sigma^{-1}_{out}$.

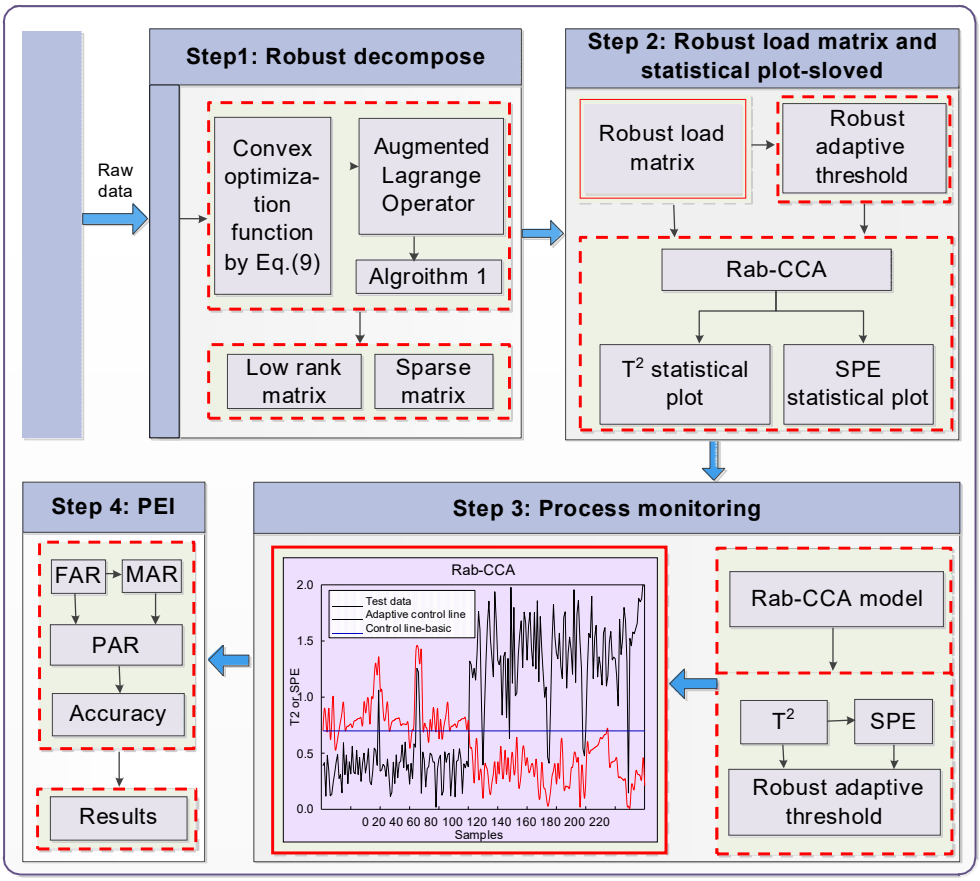

Fig 1. Schematic of the proposed method for process monitoring

3. Robust adaptive boosted canonical correlation analysis

The collected data from WWTPs are often severely corrupted by noise. Traditional CCA-based methods cannot extract the noise hidden in the data space when applied for data space decomposition. If the CCA model is trained by the contaminated data, the model performance becomes severely degraded. However, the noise matrix is usually sparse [25]. If the sparse matrix can be satisfactorily decomposed and the corresponding robust loading matrix(RLM) is accurately obtained, the model performance can be significantly improved. At the same time, because the modelling data are usually nonstationary, CCA diagnoses a fault by an invariant control limit, which cannot effectively adapt to the real working conditions. Therefore, it is imperative to develop a dynamic control limit in this paper. Based on the aforementioned considerations, we propose the following Rab-CCA method. The execution flows of Rab-CCA for process monitoring are shown in Fig. 1.

3.1. *Robust loading matrix*

To derive a robust loading matrix (RLM), the original data matrix needs to be decomposed. Suppose $X = M + E$, where M and E represent the low-rank matrix and the sparse matrix, respectively. The purpose of matrix decomposition is to solve the following convex optimization function [15, 25].

$$\begin{cases} \min \|M\|_* + \sigma \|E\|_{l_1}, \\ s.t\ X = M + E \end{cases} \tag{9}$$

where $\|M\|_*$ represents the nuclear norm of the matrix M and is used to constrain the low rank, $\sigma > 0$ is the weight parameter, and $\|E\|_{l_1}$ represents the l_1 norm of the matrix E, which is the sum of all elements in the matrix. Recently, researchers have developed many different strategies to solve the aforementioned convex optimization function. Liu et al. pointed out that the inexact augmented Lagrange algorithm (IALM) is known to generally perform well in practice [26]. Given the augmented Lagrange function, the optimization function can be formulated as follows:

$$\mathrm{L}(M, E, \Psi, \sigma, u) = \|M\|_* + \sigma \|E\|_{l_1} + \langle \Psi, X - M - E \rangle + \frac{u}{2} \|X - M - E\|_F^2, \tag{10}$$

Ψ is the Lagrange multiplier to remove the equality constraint and is bounded, and $u > 0$ is a penalty parameter. The iterative update of matrix M and E is achieved by fixing other variables. Thus, the Lagrange multiplier Ψ is updated. Finally, the optimal decomposition matrix is obtained by setting the interactive threshold. The procedure of the inexact augmented Lagrange algorithm (IALM) is shown in algorithm 1:

Algorithm 1: The iterative procedures of IALM method

1. *Input:* data matrix $\boldsymbol{X}$, weight parameter $\boldsymbol{\sigma}$, $\boldsymbol{\rho}$. Iteration upper limit $\boldsymbol{\varepsilon_{tol}}$, critical value $\mathbf{max}_{\boldsymbol{u}}$.

2. *Initialize:* $\boldsymbol{\Psi_0} = \mathbf{X/J(X)}$; $\boldsymbol{M_0 = E_0 = \Psi_0 = 0}$; $\boldsymbol{u_0 > 0}$; $\boldsymbol{\rho = 1.5}$; $\boldsymbol{\varepsilon_{tol} = 10^{-7}}$; $\boldsymbol{j = 0}$; $\mathbf{max}_{\boldsymbol{u}} = \boldsymbol{10^{10}}$;

3. *While(not converged) Do*

a. Update $\boldsymbol{M_j}$: $\boldsymbol{M_{j+1} = arg \min_{M_j} \frac{1}{u_j} \|M_j\|_* + \frac{1}{2} \|M_j - (X - E_j + \frac{\Psi_j}{u_j})\|_F^2}$.

b. Update $\boldsymbol{E_j}$: $\boldsymbol{E_{j+1} = \arg\min_{E_j} L(M_{j+1}, E_j, \Psi_j, u_j)}$.

$$\boldsymbol{E_{j+1} = S_{\sigma / u_j}[X - M_{j+1} + \frac{1}{u_j} \Psi_j]}.$$

c. Update $\boldsymbol{\Psi_j}$: $\boldsymbol{\Psi_{j+1} = \Psi_j + u_j(X - M_{j+1} - E_{j+1})}$.

d. Update $\boldsymbol{u_j}$: $\boldsymbol{u_j}$ to $\boldsymbol{u_{j+1}}$, $\boldsymbol{u_{j+1} = \min(\rho u_j, \mathbf{max}_u)}$.

4. $\boldsymbol{j = j + 1}$;

5. *Until* $\boldsymbol{\|X - M_{j+1} - E_{j+1}\|_\infty \leq \varepsilon_{tol}}$.

6. *End.*

Output $(\boldsymbol{M_j}, \boldsymbol{E_j})$

where $S_{\sigma/u_j}[\mathrm{X} - M_{j+1} + \frac{1}{u_j}\Psi_j]$ represents the soft-shrinkage operator, the soft-shrinkage operator definition refers to [25], and M_j and E_j represent the low-rank matrix and the sparse kernel matrix of the algorithm 1 iterative process, respectively. However, this approach may not converge to the optimal solution in formula (9). The optimal solution can be achieved only if it obeys theorem 1.

Theorem 1. If the parameter u_j is non-decreasing in all iterative procedures and $\sum_{j=1}^{+\infty} u_j^{-1} = +\infty$, (M_j, E_j) converges to an optimal solution(M, E) by algorithm 1.

Proof: Suppose $(M^*,\ E^*,\ \Psi^*)$ is the saddle point of the Lagrange function, $X = M^* + E^*$. First, $\partial(*)$ (monotone operator) is defined as the subgradient of a convex function, if f is a convex function, then $\langle x_1 - x_2, f_1 - f_2\rangle \geq 0$, and $f_1, f_2 \in \partial(f(x_i))$[25]. According to algorithm1, $\mathrm{X} - M_{j+1} - E_{j+1} = \frac{1}{u_j}(\Psi_{j+1} - \Psi_j)$. A new bounded time series can be introduced according to lemma 1 in Lin's paper [25],

$$\hat{\Psi}_{j+1} = \Psi_j + u_j(\mathrm{X} - M_{j+1} - E_{j+1}) \qquad \hat{\Psi}_{j+1} \in \partial(\|M_j\|_*). \tag{11}$$

Assuming that $f^* = \|M^*\|_* + \sigma\|E^*\|_{l_1}$ is the corresponding optimal value, we need to prove $\|M_j\|_* + \sigma\|E_j\|_{l_1} \xrightarrow{conv} f^*$. According to algorithm 1, $0 \in \partial_M L(M^*, E^*, \Psi^*, \sigma, u)$, and $\partial_E L(M^*, E^*, \Psi^*, \sigma, u) = \partial(\|M^*\|_*) - \Psi^* - u(\mathrm{X} - M^* - E^*)$, so $\Psi^* \in \partial(\|M^*\|_*)$. Similarly, $0 \in \partial_E L(M^*, E^*, \Psi^*, \sigma, u)$, so $\Psi^* \in \partial(\|\sigma E^*\|_{l_1})$. According to the above definition, the following formula can be derived:

$$\|E_{j+1} - E^*\|_F^2 + u_j^{-2}\|\Psi_{j+1} - \Psi^*\|_F^2 = \|E_{j+1} - E_j + E_j - E^*\|_F^2 + u_j^{-2}\|\Psi_{j+1} - \Psi_j + \Psi_j - \Psi^*\|_F^2 =$$

$$\|E_j - E^*\|_F^2 + u_j^{-2}\|\Psi_j - \Psi^*\|_F^2 - \|E_{j+1} - E_j\|_F^2 - u_j^{-2}\|\Psi_{j+1} - \Psi_j\|_F^2 - 2u_j^{-2}(\langle \Psi_{j+1} - \Psi_j, E_{j+1} -$$

$$E_j\rangle + \langle M_{j+1} - M^*, \hat{\Psi}_{j+1} - \Psi^*\rangle + \langle E_{j+1} - E^*, \hat{\Psi}_{j+1} - \Psi^*\rangle). \tag{12}$$

If $\hat{\Psi}_{j+1} \in \partial(\|M_{j+1}\|_*)$, $\Psi_{j+1} \in \partial(\|\sigma E_{j+1}\|_{l_1})$, the following formula can be derived.

$$\langle \Psi_{j+1} - \Psi_j, E_{j+1} - E_j\rangle \geq 0, \tag{13a}$$

$$\langle M_{j+1} - M^*, \hat{\Psi}_{j+1} - \Psi^*\rangle \geq 0, \tag{13b}$$

$$\langle E_{j+1} - E^*, \hat{\Psi}_{j+1} - \Psi^*\rangle \geq 0. \tag{13c}$$

Because u_j is non-decreasing, $\{\|E_{j+1} - E^*\|_F^2 + u_j^{-2}\|\Psi_{j+1} - \Psi^*\|_F^2\}$ is non-increasing by Eq. (12) and Eq. (13), and $2u_j^{-2}(\langle \Psi_{j+1} - \Psi_j, E_{j+1} - E_j\rangle + \langle M_{j+1} - M^*, \hat{\Psi}_{j+1} - \Psi^*\rangle + \langle E_{j+1} - E^*, \hat{\Psi}_{j+1} - \Psi^*\rangle) \le \|E_j - E^*\|_F^2 + u_j^{-2}\|\Psi_j - \Psi^*\|_F^2 - \|E_{j+1} - E_j\|_F^2 - u_j^{-2}\|\Psi_{j+1} - \Psi_j\|_F^2 < \|E_j - E^*\|_F^2 + u_j^{-2}\|\Psi_j - \Psi^*\|_F^2$. Thus, the following formula can be obtained:

$$\sum_{j=1}^{+\infty} 2u_j^{-1}(\langle \Psi_{j+1} - \Psi_j, E_{j+1} - E_j\rangle + \langle M_{j+1} - M^*, \hat{\Psi}_{j+1} - \Psi^*\rangle + \langle E_{j+1} - E^*, \hat{\Psi}_{j+1} - \Psi^*\rangle) < +\infty. \quad (14)$$

Similarly, $\sum_{j=1}^{+\infty} u_j^{-2}\|\Psi_{j+1} - \Psi_j\|_F^2 < +\infty$ can be obtained by Eq.(12) and Eq.(14). It is known that (M_{j+1}, E_{j+1}), so if $\sum_{j=1}^{+\infty} u_j^{-1} = +\infty$ is a feasible solution, the following formula can be obtained:

$$\|X - M_{j+1} - E_{j+1}\|_F^2 = u_j^{-1}\|\Psi_{j+1} - \Psi_j\|_F^2 \to 0. \quad (15)$$

Thus,

$$\|M_j\|_* + \sigma\|E_j\|_{l_1} \le \|M^*\|_* + \sigma\|E^*\|_{l_1} - \langle \hat{\Psi}_j, M^* - M_j\rangle - \langle \Psi_j, E^* - E_j\rangle = f^* + \langle \Psi^* - \hat{\Psi}_j, M^* - M_j\rangle + \langle \Psi^* - \Psi_j, E^* - E_j\rangle - \langle \Psi^*, X - M_j - E_j\rangle. \quad (16)$$

From formulas (13) and (14), it can be known that $\sum_{j=1}^{+\infty} u_j^{-1}(\langle M_j - M^*, \hat{\Psi}_j - \Psi^*\rangle \langle E_j - E^*, \hat{\Psi}_j - \Psi^*\rangle) < +\infty$. $\sum_{j=1}^{+\infty} u_j^{-1} = +\infty$.Thus, there must be a subsequence (M_{j_h}, E_{j_h}), such that the following formula holds:

$$\langle M_{j_h} - M^*, \hat{\Psi}_{j_h} - \Psi^*\rangle + \langle E_{j_h} - E^*, \hat{\Psi}_{j_h} - \Psi^*\rangle \to 0. \quad (17).$$

Similarly, $X - M_{j+1} - E_{j+1} = u_j^{-1}(\Psi_{j+1} - \Psi_j)$. From eq. (15), it can be observed that $\langle \Psi^*, X - M_j - E_j\rangle \to 0$.

$$\lim_{h\to+\infty} \|M_{j_h}\|_* + \sigma\|E_{j_h}\|_{l_1} \le f^*, \quad (18)$$

where (M_{j_h}, E_{j_h}) is an optimal solution (M^*, E^*), and $\hat{\Psi}_{j+1}$ is bounded. At the same time, $\{\|E_{j+1} - E^*\|_F^2 + u_j^{-2}\|\Psi_{j+1} - \Psi^*\|_F^2\}$ is non-increasing. The following formulation can be derived:

$$\|E_{j+1} - E^*\|_F^2 + u_j^{-2}\|\Psi_{j+1} - \Psi^*\|_F^2 \to 0. \quad (19)$$

Thus, $\lim_{j\to+\infty} E_{j+1} - E^* = 0$. According to formula (15), $\lim_{j\to+\infty} X - M_{j+1} - E_{j+1} = 0$. Since $X = M^* + E^*$, $\lim_{j\to+\infty} M_{j+1} - M^* = 0$. When the low-rank matrix M and sparse matrix E can be

obtained according to the above formulas, let $X_M = M$, $X_E = E$. Thus, $X = X_M + X_E$ and $Y = Y_M + Y_E$. X_M and Y_M represent the low-rank matrix containing essential information of the system. From Eq. (2), a new canonical correlation criterion function, termed the low-rank criterion function, can be established as follows:

$$\max_{W_{X_M}, W_{Y_M}} \rho_M(W_{X_M}, W_{Y_M}) = \max_{W_{X_M}, W_{Y_M}} (W_{X_M}{}^T \Sigma_{X_M X_M}{}^{-1/2} \Sigma_{X_M Y_M} \Sigma_{Y_M Y_M}{}^{-1/2} W_{Y_M}), \quad (20)$$

where X_M and Y_M represent the sparse matrix. Because the sparse matrix E also contains some useful information [27], the sparse matrix is used to establish the sparse-criterion function.

$$\min_{W_{X_E}, W_{Y_E}} \rho_E(W_{X_E}, W_{Y_E}) = \min_{W_{X_E}, W_{Y_E}} (W_{X_E}{}^T \Sigma_{X_E X_E}{}^{-1/2} \Sigma_{X_E Y_E} \Sigma_{Y_E Y_E}{}^{-1/2} W_{Y_E}). \quad (21)$$

To make full use of the low-rank matrix and the sparse matrix, we must find an optimal (W_X^*, W_Y^*) so that formulas (20) and (21) are simultaneously established to maximize the essential feature correlation (low-rank matrix) and to minimize the generalized noise correlation (sparse matrix) at the same time. Therefore, the following comprehensive criterion function can be established:

$$\max_{W_X^*, W_Y^*} \rho_*(W_X^*, W_Y^*) = \max_{W_X^*, W_Y^*} (W_X^{*T} \Sigma_{XX^*}{}^{-1/2} (\Sigma_{XY^*}) \Sigma_{YY^*}{}^{-1/2} W^*)_Y \quad (22)$$

where $\Sigma_{XX^*} = \Sigma_{X_M X_M} - \Sigma_{X_M X_M}$, $\Sigma_{YY^*_E} = \Sigma_{Y_M Y_M} - \Sigma_{Y_M Y_M}$, $\Sigma_{XY^*} = \Sigma_{X_E Y_E} - \Sigma_{X_M Y_M}$. To ensure the uniqueness of the result, let $W_X^* = (\Sigma_{X_M X_M} - \Sigma_{X_E X_E})^{1/2} \alpha^T$, $W_Y^* = (\Sigma_{Y_M Y_M} - \Sigma_{Y_E Y_E})^{1/2} \beta^T$. The following formula can be easily obtained:

$$\begin{cases} \max_{W_X^*, W_Y^*} \rho_*(W_X^*, W_Y^*) \\ s.t\ W_X^{*T} W_X^* = 1, W_Y^{*T} W_Y^* = 1 \end{cases}. \quad (23)$$

By performing a singular decomposition (SVD) on the matrix $Z = \Sigma_{XX^*}{}^{-1/2} (\Sigma_{XY^*}) \Sigma_{YY^*}{}^{-1/2}$, the corresponding left singular value vector $U^*(:, 1:r)$ and right vector $V^*(:, 1:r)$ can be obtained respectively. Then, RLM of the input space ($P^{rob} = \Sigma_{XX^*}{}^{-1/2} U^*(:, 1:r)$) and the output space ($L^{rob} = \Sigma_{YY^*}{}^{-1/2} V^*(:, 1:r)$) can be further derived, respectively.

Remark 1: In this paper, the loading matrix is solved under the assumption that Σ^* and Σ can be inverted. However, the sparse matrix is usually not available. When the matrix is irreversible, the corresponding generalized Moore-Penrose inverse matrix is used instead.

3.2 *Reconstruction of the* control chart *and adaptive thresholds*

The CCA-based method can effectively explore the internal structure of the input space and the output space. However, in the fault detection process, the core of the detection method can judge if a system is out-of-control. Thus, the number of control charts can be reduced while ensuring acceptable accuracy. In this section, the four control charts in section 2.2 are reconstructed into two control charts: one is the quality-relevant T^2 control chart, and the other is the quality-irrelevant SPE control chart.

First, the robust score matrix is $\Phi_X = XP^{rob}$. Similarly, the counterpart of the output space $\Phi_Y = YL^{rob}$. In actual industrial processes, state changes occur between the input space and the output space. Based on [10], the following formula can be established:

$$R_{rob} = \Phi_Y - \Phi_X \Lambda_r = X\Sigma_{XX^{*-1/2}} U^*(:,1:r) - Y\Sigma_{YY^{*-1/2}} V^*(:,1:r)\Lambda_{r^*}, \tag{24}$$

where R_{rob} contains the state mutation signal with correlation. Thus, we can use the above formula to establish a T^2 control chart with the correlation between input and output. The control chart is shown as follows:

$$T^2_{cca} = R_{rob} \Sigma^{-1}_{R_{rob}} R_{rob}^{T}, \tag{25}$$

based on the above knowledge, $\Phi_X{}^T\Phi_X = I$, $\Phi_Y{}^T\Phi_Y = I$, and thus

$$\Sigma_{R_{rob}} = \frac{1}{n-1} R_{rob}{}^T R_{rob} = \frac{1}{n-} (X\Sigma_{XX^{*-1/2}} U^*(:,1:r) - Y\Sigma_{YY^{*-1/2}} V^*(:,1:r)\Lambda_{r^*})^T$$

$$*(X\Sigma_{XX^{*-1/2}} U^*(:,1:r) - Y\Sigma_{YY^{*-1/2}} V^*(:,1:r)\Lambda_{r^*}) = (I - \Lambda_{r^{*2}}), \tag{26}$$

T^2_{cc} control chart is used to detect the fault between the input and output, which has a correlation. Therefore, when the value of T^2_{cc} exceeds the control limit, it indicates that the fault severely affects the quality-relevant variables. Also, we need to establish an SPE statistic to monitor the irrelevant input-output space faults. Let $Q = \begin{bmatrix} c_i P_{res}{}^* P_{res}{}^{*T} & 0 \\ 0 & (1-c_i) L_{res}{}^* L_{res}{}^{*T} \end{bmatrix}$. According to the formulas (7) and (8), the SPE control chart can be derived as follows:

$$\mathrm{SPE}_{cca} = [X,Y] Q \begin{bmatrix} X^T \\ Y^T \end{bmatrix} = [X,Y] \begin{bmatrix} P_{res}{}^* P_{res}{}^{*T} & 0 \\ 0 & L_{res}{}^* L_{res}{}^{*T} \end{bmatrix} \begin{bmatrix} X^T \\ Y^T \end{bmatrix}. \tag{27}$$

$c_i = 0.5$ represents the weight coefficient, which means that the input residual space and output residual space have the same importance.

In this section, the ACL coefficient is constructed, and the coefficient extracts the system

dynamic trend according to the data change. Suppose $H_{train} = [X_{train}, Y_{train}] \in R^{n\times(p+q)}$ and $H_{test} = [X_{test}, Y_{test}] \in R^{m\times(p+q)}$ represent the training set and online testing data, respectively, n and m represent the number of samples, and p and q are the number of monitored variables. First, the difference norm of the online data and the off-line data can be obtained as follows:

$$H_{norm}(t) = \|H_{test}(t) - H_{mean}\|_{p_{norm}}, \tag{28}$$

where $\|*\|_{p_{norm}}$ is the p_{norm} norm, t represents the sampling point, $(t \in \{1 \quad \cdots \quad m\}$, and $H_{mean} = mean(H_{train})$. Assume $I_h = [1 \quad \cdots 1] \in R^{1\times(p+q)}$, so

$$H_{ad}(t) = \frac{1}{p+q}(I_h - \frac{H_{test}(t)-H_{mean}}{H_{norm}(t)}), \tag{29}$$

where $H_{ad}(t) \in R^{1\times(p+q)}$ represents the fluctuation value of the testing data relative to the historical data at the tth sampling. Due to the fluctuation information including the $p+q$ variables, when the abrupt fault occurs, $H_{test} \quad (t) \gg H_{mean}$; thus, $0 < \frac{Htest(t)-Hmean}{H_{norm}(t)}(i) < 1$, where i represents the ith variable, and thus, $H_{line}(t) = sum(H_{ad}(t)) < 1)$. Similarly, when the testing data are collected under the normal conditions, $H_{test}(t) \ll H_{mean}$. Thus $H_{line}(t) > 1$. The adaptive control limit (ACL) can be set according to the adaptive coefficient, and based on [23, 24], $g\chi^2_1(*)$ is selected as the basic statistic. The ACL of T^2 can be expressed as follows:

$$T^2_{line} = H_{line} * g_T\chi^2_{1-a}(r_T), \tag{30}$$

where $H_{line} = [H_{line}(1) \quad \cdots \quad H_{line}(n)]$, a is the confidence level, $g_T = S_T/2u_T$, and $r_T = \Theta u_T{}^2/S_T$. Typically, $\Theta = 2$, where u_T and S_T are estimated as $u_T = {}^1\sum^n \quad _{n\ i=1} T^2_{cca}(i)$, and $S_T = \frac{1}{n-1}\sum_{i=1}^{n}(T^2_{cc}(i) - u_T \ )^2$. Thus, the ACL of SPE_{cca} can be constructed as follows:

$$SPE_{line} = H_{line}SPE_H, \tag{31}$$

SPE_H can be expressed as follows:

$$SPE_H = g_{spe}\chi^2_{1-}(r_T), \tag{32}$$

where $g_{spe} = S_{spe}/2u_{spe}$, and $r_T = \Theta u_{spe}{}^2/S_{spe}$. u_T and S_T are estimated as $u_{spe} = \frac{1}{n}\sum_{i=1}^{n} SPE_{cca}(i)$ and $S_{spe} = {}^1\frac{}{n-1}\sum_{i=1}^{n}(SPE_{cca}(i) - u_{spe})^2$, respectively.

4. CASE STUDIES

4.1 Process monitoring for BSM1

(1) **Background:** BSM1 (Benchmark Simulation Model 1, BSM1) was developed by the International Water Association(IWA)[28]. This model is based on the real WWTP predenitrification process and is mainly motived by the removal of *C* and *N*. The average daily sewage treatment capacity of BSM1 is 20,000 m^3, which is a widely accepted water treatment simulation platform. The wastewater treatment process of BSM1 is shown in Fig. 2, and is mainly composed of a biochemical reaction tank (volume 5,999 m^3) and a secondary sedimentation tank (volume 6,000 m^3). The biochemical reaction tank is simulated by the activated sludge model No. 1 (ASM1) of IAWQ, including two anoxic tanks and three aerated tanks. The secondary settler tank has ten layers, and each layer is 0.4 meters. The detailed introduction can be found on the website (http://www.benchmarkWWTP.org). First, BSM1 is simulated over 150 days with constant flow and composition. second, the input data of the sunny day are used as the dynamic input of the WWTPs, and the simulation lasts for two weeks. During this period, the data are collected every 15 min. Twenty-six important variables were selected for monitoring, including 16 input variables (including process variables) and 10 quality-relevant output variables. The monitoring variables covered all the processes of WWTP to monitor the entire process globally. In this section, the abrupt fault and drift fault are defined and monitored by the proposed Rab-CCA method. At the same time, to further verify the robustness of the Rab-CCA method, the collected data are added to Gaussian noise.

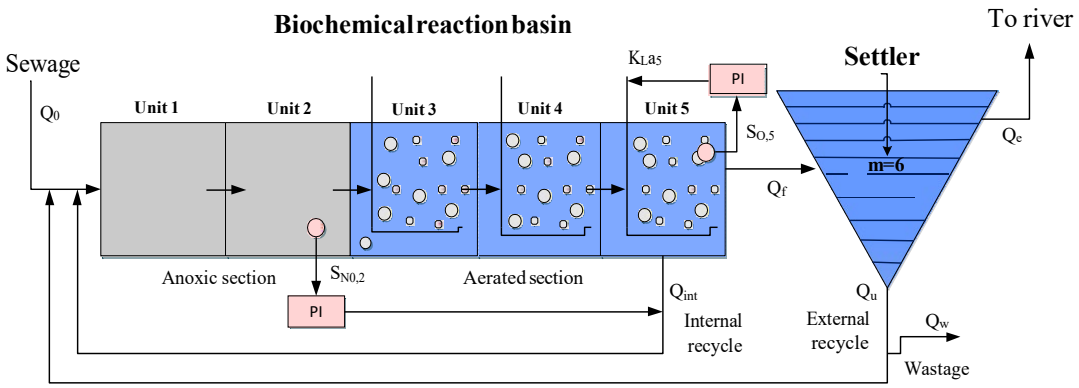

Fig 2. Schematic diagram of a wastewater treatment plant under BSM1

(2) **Diagnostic results of the proposed method:** In this section, the faults are defined as follows: Scenario a) Fault 1 is the S_O with abrupt changes at 8:00 during the 9th day in the

4th reaction tank; Scenario b) Fault 2 represents the abrupt fault S_{NO}, which occurs in the 5th reaction tank; Scenario c) Fault 3 is the abnormal events of water five-day biochemical oxygen demand (BOD_5) in the effluent and occurred at 8:00 during the 9th day. The Rab-CCA model is trained with the first 700 data samples, and the remaining 644 samples are used for testing. When the singular values are small, the extracted variables are weakly correlated. Thus, the RLM with a strong correlation is selected by adjusting the cumulative percent pariance (CPV=95%).

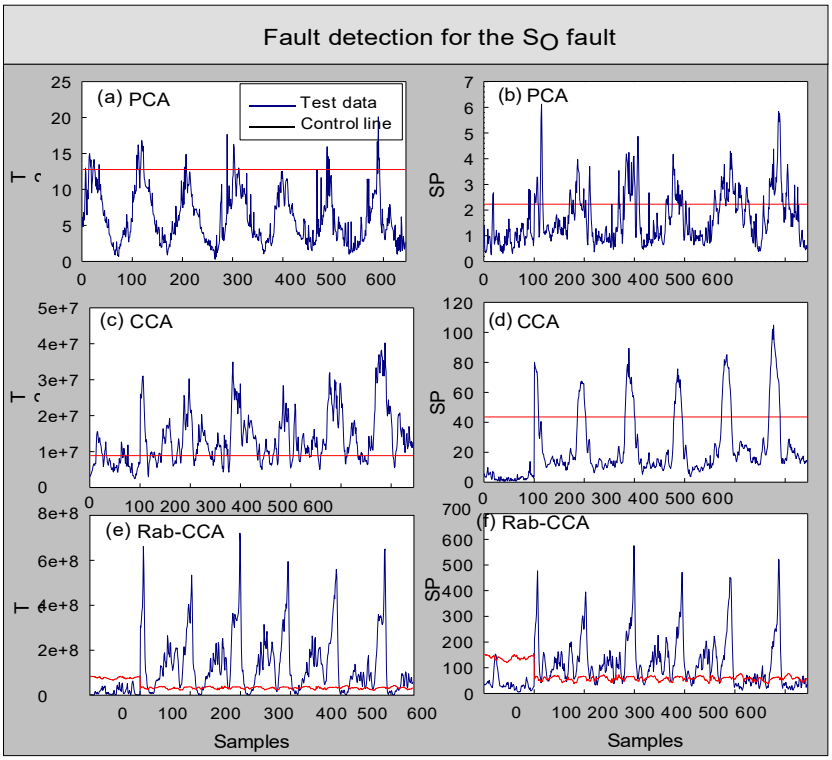

Fig 3. Detection results of fault 1 based on the proposed method and two other methods

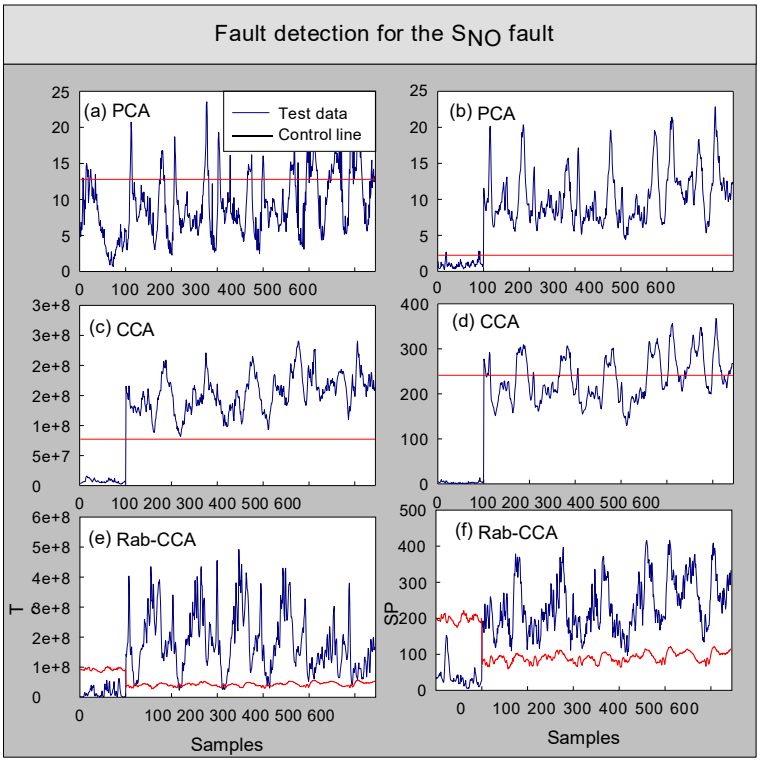

Fig 4. Detection results of fault 2 based on the proposed method and two other methods

Fault 1 is a very common sensor fault (oxygen sensor) in WWTPs due to corrosion and highly frequent signal variations. In this period, the first 100 samples of the testinging dataset are normal. As shown in Fig. 3, it can be observed that the traditional CCA-based method and PCA method have many simultaneous missed alarms and false alarms. Fig. 3(a) and Fig. 3(b) show the fault detection results of PCA-T^2 and PCA-SPE, respectively. Fig. 3(c) and Fig. 3(d) represent the CCA-T^2 and CCA-SPE, respectively. Conversely, the Rab-CCA-based method performs better, especially with fewer missed alarms. Because the inexact augmented Lagrange algorithm is used to decompose the corrupted matrix, the information of the corrupted matrix can be effectively extracted and used. In addition, Rab-CCA is configured with an adaptive control limit, and thus the Rab-CCA is able to robustly deal with anomalies or noise. The missed alarm rate of the Rab-CCA-T^2 control chart (Fig. 3(e)) is 0, and the missed alarm rate with respect to Rab-CCA-SPE (Fig. 3(e)) is only 4%. The proposed Rab-CCA-based method can effectively monitor abnormalities, with the T^2 control plot used to monitor quality-relevant faults in the wastewater treatment process. To further verify the proposed method, fault 2 (S_{NO} is the abnormal variable) is analysed as a special case. Fault 2

has a highly important influence on the biochemical reaction in the wastewater treatment process. If fault 2 is not well rectified, water qualities such as BOD and COD in the effluent can violate the standards. The process monitoring results are shown in Fig. 4, suggesting that the diagnostic accuracy of the Rab-CCA-based method for quality-relevant faults (Fig. 4(e)) is 97.2% and the accuracy of the quality-irrelevant faults (Fig. 4(f)) is 99.38%. However, because the CCA-based method cannot adaptively manipulate the control limit, CCA-T^2 (Fig. 4(d)) occurs many missed alarms. Similarly, the PCA-T^2 (Fig. 4(a)) achieved poor performance with a false alarm rate of 11% and missed alarm rate of 73.53%.

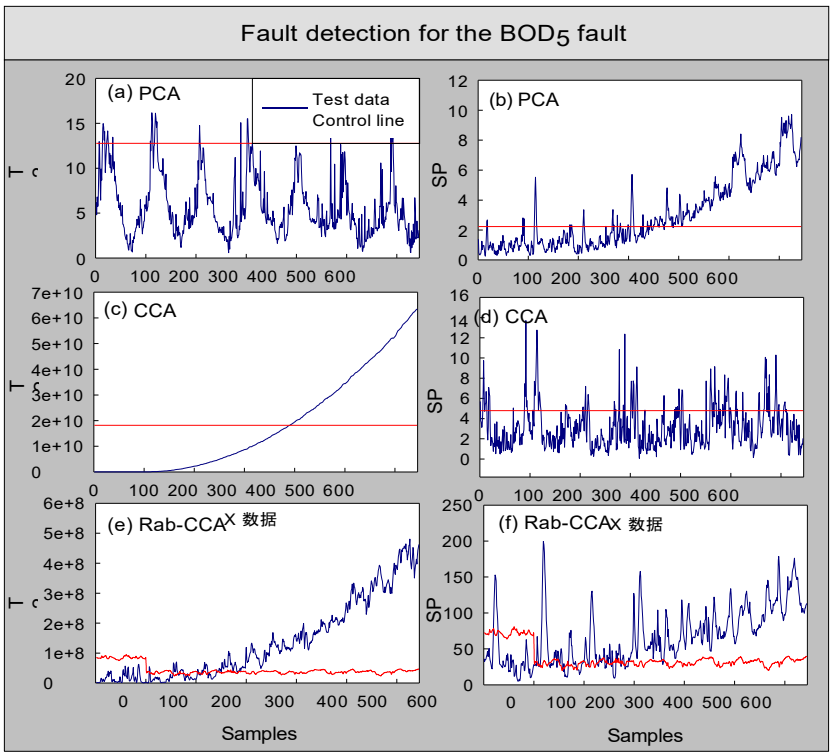

Fig 5. Detection results of fault 3 based on the proposed method and two other methods

Fault 3 represents a drifting fault of BOD5. BOD5 is an important index for measuring effluent quality, and thus, the effective monitoring of BOD5 anomalies is important for WWTPs. As shown in Fig. 5, the classic PCA (Fig. 5(a)-5(b)) and CCA (Fig. 5(c)-5(d)) methods cannot achieve the good performance, especially in the early fault stage. Because fault 3 is not obvious in the early stage (100-300 samplings), false alarms and missed alarms easily occur. However, Rab-CCA has the best performance during the entire monitoring process. Since the ACL values be adjusted according to the dynamic trend of the data, the Rab-CCA-based method can effectively reduce false alarms and missed alarms significantly.

Note that the ACL is first proposed in this paper. Although the theory of ACL is derived in section 3, it is global. Due to the faulty variables that cannot be derived in advance and exhibit non-obvious variations, the fault signal is likely to be covered by the other irrelated signal or noise. In this light, there will be occurred the false alarm.

To further verify the proposed method, three different faults are studied. Fault 4 represents the fault of S_{ALK}, which is emerges the first reaction tank. Fault 5 and Fault 6 represent the abnormal concentrations of total nitrogen (TN) and total chemical oxygen demand (COD) in the effluent, respectively. Moreover, according to the method presented in the paper [29], 1%, 5%, and 10% Gaussian noise were added to the training sets. To better and more comprehensively compare the performance of the four methods under noise interference with a different degree, the performance evaluation indices (PEI) of accuracy and pre-alarm rate are used to evaluate the monitoring model [30]. The pre-alarm rate (PAR) is a comprehensive index that combines false alarms with missed alarms [30], and the lower the pre-alarm rate, the better the performance of the method. The formulas for the accuracy and pre-alarm rate are as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}, \tag{33}$$

$$P_{AR} = \gamma M_{AR} + (1-\gamma) F_{AR}. \tag{34}$$

M_{AR} and F_{AR} are the missed alarm rate and false alarm rate, respectively, and γ is the weight coefficient. In this paper, because missed alarms have a more serious impact on the WWTP than false alarms, the coefficient $\gamma = 0.6$. When many of the missed alarms occur in the system, the engineer cannot conduct repairs in time, which can lead to serious accidents. The definitions of TP and TN can be found in [30]. The diagnostic results of the aforementioned six faults are tabulated as follows:

Table 1. Average detection results of the six types of faults

Average value of results		**PCA**		**CCA**		**Robust-PCA**		**Rab-CCA**	
		T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
Noise -free	FAR	0.11	0.04	0.042	0.095	0.787	0.842	0	0.022
	MAR	0.696	0.181	0.126	0.679	0.100	0.006	0.072	0.067
	PAR	0.462	0.125	0.092	0.445	0.375	0.340	0.043	0.049
	Accuracy	0.395	0.841	0.887	0.412	0.793	0.865	0.939	0.94

Noise -added	FAR	0.007	0.007	0.044	0.068	0.849	0.858	0.124	0.05
	MAR	0.704	0.211	0.166	0.721	0.100	0.044	0.144	0.125
	PAR	0.425	0.129	0.118	0.46	0.400	0.370	0.136	0.095
	Accuracy	0.404	0.821	0.853	0.38	0.783	0.829	0.859	0.887

It can be observed from Table 1 that the Rab-CCA method has extraordinarily better performance under the noise-free condition. According to the experimental results, the average accuracies of PCA-T^2 and PCA-SPE are only 39.5% and 84.1%, respectively, and those of the corresponding robust-PCA- T^2 and robust-PCA-SPE are 79.3% and 86.5%, respectively. These results indicate that robust technology is helpful to improving the performance of the model. Moreover, the average accuracies of Rab-CCA-T^2 and Rab-CCA are 93.9% and 94%, respectively, which are significantly higher than those of the other methods. The average value of the pre-alarm rate (PAR) further confirms the superiority of the proposed method. The Rab-CCA based T^2 statistic and SPE statistic are 4.3% and 4.9%, respectively. When the data contain different proportions of Gaussian noise, i.e., the above six types of faults have added noise of 1%, 5%, and 10% respectively, and the average accuracy and average PAR are shown in Table 1. The average accuracies of Rab-CCA-T^2 and Rab-CCA- SPE are the highest among the four methods. The PCA-based T^2 statistic is improved under noise interference, mainly because the data fluctuation becomes larger due to the addition of noise, and to a certain extent, this will reduce the false alarms. It is worth noting that the special cases are the regular events of the method in the face of complex faults, which will not affect the implementation and application of the method. The Rab-CCA T^2 control chart has an average accuracy of 85.9% for six types of faults under the noise-added conditions, which decreases by 8.5% compared with the noise-free conditions. However, the accuracy of traditional CCA based method T^2 control chart only decreases by 3.8%. Because the adaptive control coefficient is global, when noise is mixed in the data, the ACL may fail to reflect the real fluctuation of the system. Despite these observations, the Rab-CCA based method T^2 control chart and SPE control chart have the highest diagnosis accuracy under the noise-added condition. Additionally, the PAR of Rab-CCA is the lowest among the four methods.

4.2. Filamentous sludge bulking monitoring for a real full-scale WWTP

(1) **Background:** In this study, the proposed method is used to monitor a full-scale WWTP with the oxidation ditch (OD) process. The plant serves a population of 480,000, with a daily treatment flow of 170,000 m^3 and a hydraulic retention time of 16.5 h, and uses a long solid residence time (SRT) to achieve good nitrogen removal performance. The SRT is typically maintained at 15–22 days. Due to some external factors such as weather, temperature and sludge activity, filamentous sludge bulking not only occurs frequently but is also difficult to monitor online in real- time. The schematic of the WWTP is shown in Fig. 6. The data were collected from 1 October to 21 March of the next year (the sampling interval is one day), and filamentous sludge bulking occurred during the sample period. The first 70 days of data are used for training, and the remaining days for testing. The 15 monitoring variables include 8 input variables and 7 output variables. In the testing data, the first 23 days of sampling are normal, whereas the rest of the sampling occurs under sludge bulking. The collected data not only contain noise but also may have sensing errors.

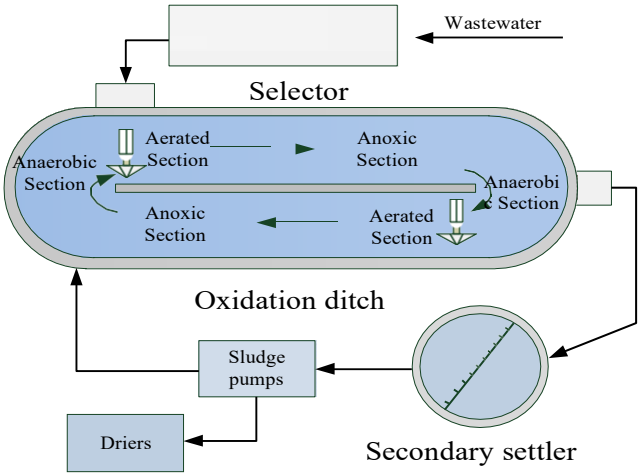

Fig 6. Schematic diagram of a real full-scale WWTP

(2) **Diagnostic results of the proposed method:** Filamentous sludge bulking is a drift fault [31]. In contrast to abrupt faults, the sludge bulking of real wastewater treatment processes may return to normal after the self-regulation of microorganisms [32, 33]. In this section, the experiment is ignores some special cases, i.e., the experiment assumes that the sludge bulking could last after the 24th day. The CPV of the Rab-CCA-based method still retains 95%, and other fixed parameters are set as in case study 1.

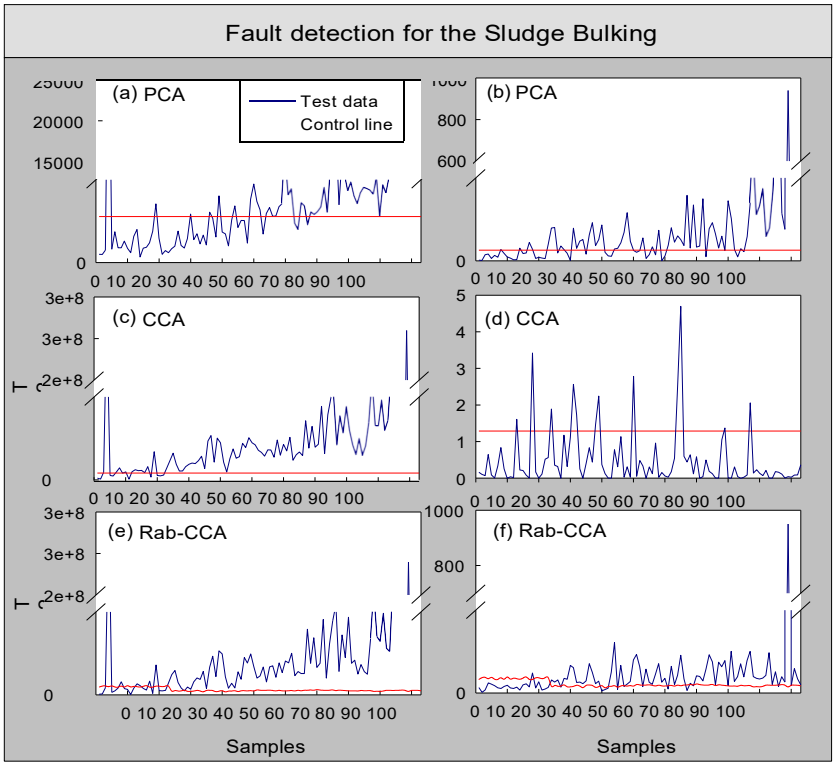

Fig 7. Detection results for sludge bulking in a real WWTP

The monitoring results are shown in Fig. 7, and Fig. 7(a) and Fig. 7(b) represent the PCA-T^2 control chart and PCA- SPE control chart, respectively. PCA-T^2 has many missed alarms. Since sludge bulking is a typical drift fault, PCA-T^2 cannot effectively identify the initial sludge bulking. The missed alarm rate of PCA- T^2 reaches 32.5%. However, CCA- T^2 has relatively few missed alarms. As shown in Fig. 7(c), the missed alarm rate is zero. However, the false alarm rate of CCA- T^2 is 84.61% higher than that of PCA- T^2, indicating that the control limit of the CCA- T^2 method is set too low, which makes the control chart appear to have more false alarms. According to the results of the SPE control chart, the false alarm in Fig. (7d) (CCA-SPE) is lower than that in Fig. (7b) (PCA-SPE). However, the missed alarm in Fig. (7d) is higher than that in Fig. (7b). Since sludge bulking is a drift fault and the traditional control chart cannot adaptively adjust the control limit, false alarms and missed alarms become more obvious. To solve the above problem, both robust-PCA and Rab-CCA are used to monitor sludge bulking in wastewater treatment plants. The false alarm rate, missed alarm rate, accuracy, and pre-alarm rate are tabulated in Table 2. Table 2 shows that the performances of robust-PCA and Rab-CCA methods are significantly better than those of PCA and CCA. The

Rab-CCA-SPE missed alarm rate is reduced by 82.86% compared to CCA-SPE. The missed alarm rate of Robust-PCA-SPE is reduced by 84.21% compared to PCA-SPE. This result fully proves the effectiveness of the robust-based method. In addition, from the comprehensive evaluation index PAR and accuracy, the PAR of the Rab-CCA method is the lowest among the four methods, e.g, Rab-CCA-T^2 is reduced by 42.9% compared with CCA- T^2. Additionally, the PAR of Rab-CCA- T^2 is 56.31% lower than that of Robust-PCA-T^2. Overall, Rab-CCA achieves the best performance in terms of the PAR. In addition, the consecutive filamentous sludge bulking causes unpermitted effluent and secondary pollution to the environment. Therefore, it is crucial to design an effective method to monitor quality-relevant faults for real WWTPs. In this section, the proposed Rab-CCA can be used to monitor quality-related faults in WWTPs. According to Fig. 7(f) and Table 2, the detection accuracy of Rab-CCA-T^2 for quality-related faults reaches 92.23%. Compared with the traditional CCA-T^2, the accuracy is improved by 5.55%. Moreover, the accuracy of Rab-CCA- T^2 is 10.46% higher than that of Roubst-PCA- T^2. The above experimental results show that the proposed Rab-CCA method can effectively monitor effluent quality-related faults. Meanwhile, the performance of Rab-CCA-SPE is still that of superior to the other methods in monitoring of quality-irrelevant faults. The corresponding fault detection accuracy on the WWTP can be ranked as follows: Rab-CCA-SPE (87.38%) > Robust-PCA (81.55%) > PCA-SPE (76.7%) > CCA-SPE (30.1%). This result fully demonstrates that the proposed Rab-CCA method can provide effective early warning for sludge bulking.

Table 2. Detection results of four methods for sludge bulking in a real WWTP

Results	**PCA**		**CCA**		**Robust-PCA**		**Rab-CCA**	
	T^2	SPE	T^2	SPE	T^2	SPE	T^2	SPE
$\mathbf{F_{AR}}$	0.087	0.2174	0.5652	0.087	0.7391	0.6957	0.3043	0.0435
$\mathbf{M_{AR}}$	0.325	0.2375	0	0.875	0	0.0375	0.0125	0.15
$\mathbf{P_{AR}}$	0.2298	0.2295	0.2261	0.5598	0.2957	0.3008	0.1292	0.1074
Accuracy	0.7282	0.767	0.8738	0.301	0.835	0.8155	0.9223	0.8738

5. Conclusions

In this study, the novel fault detection method Rab-CCA is proposed for monitoring of

quality-relevant faults in the effluent of WWTPs. Different from previous studies, Rab-CCA can efficiently remedy the defects of traditional methods, especially for different types of faults under complex interference. Additionally, Rab-CCA mitigates the issue in which the collected data are corrupted by anomalies or noise. Rab-CCA can effectively decompose the corrupted matrix into a low-rank matrix and a sparse matrix. Subsequently, a new criterion function is established to achieve the purpose of multi-objective optimization, in such a way that the information of the low-rank matrix and sparse matrix can be fully extracted. Moreover, the ACL is first assimilated into the proposed method. In this paper, the six types of faults imposed on BSM1 and real full-scale WWTP sludge bulking are monitored by Rab-CCA. The results show that the ACL can indeed efficiently reduce missed alarms and false alarms. Additionally, when the aforementioned faults are imposed with different proportions of noise, Rab-CCA still achieves superior performance. It should be noted that the Rab-CCA-based method focuses on fault detection. However, prediction of the fault propagation route is necessary for system maintenance. Especially when the data structure is more complex (non-Gaussian and non-linear), the quality-relevant fault root cause diagnosis and prognosis are still not well researched. In future research, we will focus on quality-relevant fault root cause diagnosis and prediction.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61873096, 61673181, 61533002), the Science and Technology Program of Guangzhou, China (201804010256). The Australian Research Council (ARC) Discovery Projects DP170102812 and DP200100933. The author thanks the anonymous referees and other people for their help to improve the article.

References:

1. Olsson G. ICA and me--a subjective review. Water Research. 2012, 46, 1585-624.
2. Wu J, Cheng H, Liu Y, Huang D, Yuan L, Yao L. Learning soft sensors using time difference–based multi-kernel relevance vector machine with applications for quality-relevant monitoring in wastewater treatment. Environmental Science and Pollution Research. 2020, 27, 28986-99.
3. Zhu Q, Liu Q, Qin SJ. Quality-relevant fault detection of nonlinear processes based on kernel concurrent canonical correlation analysis. 2017 American Control Conference (ACC)2017. p. 5404-9.
4. Qingchao, Jiang, Xuefeng, Yan. Plant-wide process monitoring based on mutual information–multiblock principal component analysis. Isa Transactions. 2014, 53, 1516-27.

5. Ge Z, Song Z, Gao F. Review of recent research on data-based process monitoring. Industrial & Engineering Chemistry Research. 2013, 52, 3543-62.
6. Tian Y, Yao H, Li Z. Plant-wide process monitoring by using weighted copula–correlation based multiblock principal component analysis approach and online-horizon Bayesian method. ISA Transactions. 2020, 96, 24-36.
7. Liu Y, Xie M. Rebooting data-driven soft-sensors in process industries: A review of kernel methods. Journal of Process Control. 2020, 89, 58-73.
8. Cheng H, Liu Y, Huang D, Xu C, Wu J. A Novel Ensemble Adaptive Sparse Bayesian Transfer Learning Machine for Nonlinear Large-Scale Process Monitoring. Sensors. 2020, 20, 6139.
9. Cheng H, Liu Y, Huang D, Pan Y, Wang Q. Adaptive Transfer Learning of Cross-Spatiotemporal Canonical Correlation Analysis for Plant-Wide Process Monitoring. Industrial & Engineering Chemistry Research. 2020, 59, 21602-14.
10. Chen Z, Ding SX, Peng T, Yang C, Gui W. Fault detection for non-Gaussian processes using generalized canonical correlation analysis and randomized algorithms. IEEE Transactions on Industrial Electronics. 2017, 65, 1559-67.
11. Jiang Q, Ding SX, Wang Y, Yan X. Data-driven distributed local fault detection for large-scale processes based on the GA-regularized canonical correlation analysis. IEEE Transactions on Industrial Electronics. 2017, 64, 8148-57.
12. Friman O, Cedefamn J, Lundberg P, Borga M, Knutsson H. Detection of neural activity in functional MRI using canonical correlation analysis. Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine. 2001, 45, 323-30.
13. Lee SH, Choi S. Two-dimensional canonical correlation analysis. IEEE Signal Processing Letters. 2007, 14, 735-8.
14. Branco JA, Croux C, Filzmoser P, Oliveira MR. Robust canonical correlations: A comparative study. Computational Statistics. 2005, 20, 203-29.
15. Pan Y, Yang C, An R, Sun Y. Robust Principal Component Pursuit for Fault Detection in a Blast Furnace Process. Industrial & Engineering Chemistry Research. 2018, 57, 283-91.
16. Candès EJ, Li X, Ma Y, Wright J. Robust principal component analysis? Journal of the ACM (JACM). 2011, 58, 1-37.
17. Liu Y, Pan Y, Sun Z, Huang D. Statistical Monitoring of Wastewater Treatment Plants Using Variational Bayesian PCA. Industrial & Engineering Chemistry Research. 2014, 53, 3272–82.
18. Jiang B, Huang D, Zhu X, Yang F, Braatz RD. Canonical variate analysis-based contributions for fault identification. Journal of Process Control. 2015, 26, 17-25.
19. Liu Q, Qin SJ, Chai T. Multiblock concurrent PLS for decentralized monitoring of continuous annealing processes. IEEE transactions on Industrial Electronics. 2014, 61, 6429-37.
20. Alcala CF, Qin SJ. Reconstruction-based contribution for process monitoring. Automatica. 2009, 45, 1593-600.
21. Cheng H, Wu J, Liu Y, Huang D. A novel fault identification and root-causality analysis of incipient faults with applications to wastewater treatment processes. Chemometrics and Intelligent Laboratory Systems. 2019, 188, 24-36.
22. Zhu Q, Qiang L, Qin SJ. Concurrent Canonical Correlation Analysis Modeling for Quality-Relevant Monitoring. 2016, 49, 1044-9.
23. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. Technometrics. 1995, 37, 41-59.

24. Liu Y, Liu B, Zhao X, Xie M. A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring. IEEE Transactions on Industrial Electronics. 2017, 65, 6478-86.
25. Lin Z, Chen M, Ma Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. arXiv preprint arXiv:10095055. 2010.
26. Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y. Robust Recovery of Subspace Structures by Low-Rank Representation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2013, 35, 171-84.
27. Tan C-H, Chen J, Chau L-P. Edge-preserving rain removal for light field images based on RPCA. 2017 22nd International Conference on Digital Signal Processing (DSP). IEEE2017. p. 1-5.
28. Xiao H, Huang D, Pan Y, Liu Y, Song K. Fault diagnosis and prognosis of wastewater processes with incomplete data by the auto-associative neural networks and ARMA model. Chemometrics and Intelligent Laboratory Systems. 2017, 161, 96-107.
29. Torres ME, Colominas MA, Schlotthauer G, Flandrin P. A complete ensemble empirical mode decomposition with adaptive noise. 2011 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE2011. p. 4144-7.
30. Cheng H, Liu Y, Huang D, Liu B. Optimized Forecast Components-SVM-Based Fault Diagnosis With Applications for Wastewater Treatment. IEEE Access. 2019, 7, 128534-43.
31. Liu Y, Guo J, Wang Q, Huang D. Prediction of Filamentous Sludge Bulking using a State-based Gaussian Processes Regression Model. Scientific Reports. 2016, 6, 31303.
32. Seka AM, Wiele TVD, Verstraete W. Feasibility of a multi-component additive for efficient control of activated sludge filamentous bulking. Water Research. 2001, 35, 2995-3003.
33. Sabliy L, Kuzminskiy Y, Zhukova V, Kozar M, Sobczuk H. New approaches in biological wastewater treatment aimed at removal of organic matter and nutrients. Ecological Chemistry and Engineering S. 2019, 26, 331-43.