A New Approach to Multivariate Statistical Process Control and Its Application to Wastewater Treatment Process Monitoring

Osamu Yamanaka[1], Ryo Namba[2], Takumi Obara[3], and Yukio Hiraoka[4]

[1,2,3] *Toshiba Infrastructure Systems and Solutions Corporation, Fuchu-shi, Tokyo, 183-8511, Japan*
osamu2.yamanaka@toshiba.co.jp
ryo.namba@toshiba.co.jp
takumi.obara@toshiba.co.jp

[4] *Toshiba Infrastructure Systems and Solutions Corporation, Kawasaki-shi, Kanagawa, 212-8585, Japan*
yukio.hiraoka@toshiba.co.jp

Abstract

This paper presents a new process monitoring and fault diagnosis approach based on a modified Multivariate Statistical Process Control (MSPC) and evaluates its applicability to municipal wastewater treatment process monitoring. Firstly, a conventional MSPC, based on Principal Component Analysis (PCA), is adjusted to provide an easy-to-understand user interface and then a new yet simplified reconfigurable diagnosis model is introduced. The user interface that has been developed is designed to integrate MSPC seamlessly with existing process monitoring systems that use the so-called trend graphs. The proposed diagnosis model is constructed by aggregating small models with either one or two inputs, which enhances the tractability of the diagnosis model. The effectiveness of the modified MSPC is demonstrated through a series of offline and online experiments, using a set of real multivariate process data from a municipal wastewater treatment plant.

1. Introduction

Process monitoring and fault diagnosis plays an important role for operation of social infrastructure systems such as power generation plants, water and wastewater treatment plants, railway and transportation systems, just to name a few. Supervisory Control And Data Acquisition (SCADA) system is often installed in such infrastructure systems where operators monitor time-series process data by the so-called trend graphs to keep process in control. The simplest yet often adopted fault diagnosis technique for stable plant operation is abnormality (out-of-control states) detection in process data by a pre-specified control limit for each single variable, which is similar to Statistical Process Control (SPC). To improve process performance and operational stability, however, earlier fault detection and cause localization will be important. It allows us to recognize how to improve process performance and how to avoid performance degradation.

Multivariate Statistical Process Control (MSPC) (Jackson & Mudholkar, 1979), (Kourti & MacGregor, 1996),(Kresta, Macgregor, & Marlin, 1991),(Westerhuis, Gurden, & Smilde, 2000), (Wise & Gallagher, 1996) is attractive data-driven approach for such purpose, which is suitable to monitor complex processes with high-dimensional data structure. The key idea of MSPC is subspace orthogonalization where Principal Component Analysis (PCA) is often utilized. High-dimensional process data is projected onto a subset of the subspaces and a few statistical indices for fault detection are constructed, which is effective in improving detection accuracy and cause localization ability. In addition, advanced MSPC methods have also been proposed (Choi, Morris, & Lee, 2008), (Jaffel, Taouali, Harkat, & Messaoud, 2017),(Ge, Yang, & Song, 2009), (Uchida, Fujiwara, Saito, & Osaka, 2022) for further improvement. Despite such advances, the SPC-like monitoring is still popular and widely used as real-time process monitoring in many real plants, while various applications of MSPC have been reported (AlGhazzawi & Lennox, 2008),(Camacho, Pérez-Villegas, García-Teodoro, & Maciá-Fernández, 2016), (GarcÇa-Alvarez, 2009), (Kano et al., 2002), (Lemaigre et al., 2016), (de Oliveira, Pedroza, Sousa, Lima, & de Juan, 2017), (Rosén, 2001), (Sandberg, Lennox, & Undvall, 2007),(Zhao, Yang, Yan, & Zhao, 2022). Among various possible reasons, the difficulty of intuitive understanding of diagnostic results and the difficulty of PCA model handling in MSPC will be main reasons to hinder wide applications of real-time continuous monitoring by MSPC.

Based on this motivation, this paper tries to improve the ap-

https://doi.org/10.36001/IJPHM.2024.v15i3.3854

plicability of MSPC to real-time process monitoring. To this end, this paper firstly introduces an improved user interface (UI) where the conventional trend-graph-based monitoring and MSPC-based diagnosis are combined, which will improve intuitive understanding by plant operators. Then, without changing the UI, this paper proposes a new yet simple MSPC algorithm to improve tractability and maintainability of PCA model used in MSPC. The proposed MSPC has some advantages over conventional PCA-based MSPC (PCA-MSPC hereinafter) for real-time continuous applications of MSPC to real processes. The advantages include simple and clear dependence on a specific training data in modeling and easy reconfiguration of input variables of a PCA model, which improves model tractability and possibly enables robust-and- interpretable modeling. While the proposed MSPC is simpler than the PCA-MSPC, it is confirmed that the MSPC effectively works for fault diagnosis by applying it to real process data in a municipal WasteWater Treatment Plant (WWTP).

2. CONVENTIONAL PCA-MSPC

This section overviews PCA-MSPC (Jackson & Mudholkar, 1979; Wise & Gallagher, 1996; Camacho et al., 2016) prior to presenting a modified MSPC. PCA-MSPC utilizes Hotelling's T^2 statistic (D statistic) and Q statistic (Squared Prediction Error) as fault indices together with their control limits and the variable contributions to them, which can be defined by using PCA.

PCA transforms an $n \times m$ data matrix $\boldsymbol{X}$ by combining the variables as a linear weighted sum, represented as

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{T}_{all}\boldsymbol{P}_{all}^T = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E} = \hat{\boldsymbol{X}} + \boldsymbol{E} \\ &= [\boldsymbol{t}_1\boldsymbol{p}_1^T+\cdots+\boldsymbol{t}_k\boldsymbol{p}_k^T]+[\boldsymbol{t}_{k+1}\boldsymbol{p}_{k+1}^T+\cdots+\boldsymbol{t}_m\boldsymbol{p}_m^T], \end{aligned} \quad (1)$$

where $\boldsymbol{p}_i$, $i = 1, \cdots, m$, are the principal component loadings, $\boldsymbol{P}_{all} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_m]$ and $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_k]$ are the loading matrices, $\boldsymbol{t}_i$, $i = 1, \cdots, m$, are the principal component scores, $\boldsymbol{T}_{all} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_m]$ and $\boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_k]$ are the score matrices, and $\boldsymbol{E} = [\boldsymbol{t}_{k+1}\boldsymbol{p}_{k+1}^T + \cdots + \boldsymbol{t}_m\boldsymbol{p}_m^T]$ is the residual matrix. n ,m, and k are the number of samples, that of variables, and the retained number of principal components by truncation, respectively. Superscript T denotes the transpose of a vector/matrix. It is usually assumed that the columns of $\boldsymbol{X}$ have been standardized to zero mean and unit variance by normalizing each column by its mean μ_i and standard deviation $\sigma_i, i = 1, 2, \cdots, m$. The principal component loadings are the direction vectors creating a hyperplane that is embedded inside the m-dimensional spaces and captures the maximum possible residual variance in the measured variables, while maintaining orthonormality with the other loading vectors. The loadings correspond to the eigenvectors of the covariance matrix of $\boldsymbol{X}$ and its eigenvalues indicate the variance captured by the corresponding eigenvector. The (sample) covariance matrix $\boldsymbol{\Sigma} = \frac{1}{n-1}\boldsymbol{X}^T\boldsymbol{X}$ can be described by

$$\boldsymbol{\Sigma} = \boldsymbol{P}_{all}\boldsymbol{\Lambda}_{all}\boldsymbol{P}_{all}^T = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^T + \boldsymbol{F}, \quad (2)$$

where $\boldsymbol{\Lambda}_{all} = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$ is a diagonal matrix of the eigenvalues $\lambda_i, i = 1, 2, \cdots, m$ of $\boldsymbol{\Sigma}$, $\boldsymbol{\Lambda}$ is a partial matrix of $\boldsymbol{\Lambda}_{all}$ of which elements consist of k largest eigenvalues, and $\boldsymbol{F}$ is the residual of the covariance matrix $\boldsymbol{\Sigma}$. Note that the covariance matrix $\boldsymbol{\Sigma}$ is identical to the correlation matrix (denoted by $\boldsymbol{R}$ hereinafter) if each column of $\boldsymbol{X}$ is standardized. Using these matrices, the T^2 and Q statistics (of measurements at time t) are defined by

$$\begin{aligned} T^2(t) &:= \boldsymbol{t}^T(t)\boldsymbol{\Lambda}^{-1}\boldsymbol{t}(t) = \boldsymbol{x}^T(t)\boldsymbol{P}\boldsymbol{\Lambda}^{-1}\boldsymbol{P}^T\boldsymbol{x}(t), && (3) \\ Q(t) &:= \boldsymbol{e}^T(t)\boldsymbol{e}(t) = \boldsymbol{x}^T(t)(\boldsymbol{I} - \boldsymbol{P}\boldsymbol{P}^T)\boldsymbol{x}(t), && (4) \end{aligned}$$

where $\boldsymbol{e}(t)$, $\boldsymbol{t}(t)$, and $\boldsymbol{x}(t)$ are the residual vector at time t, the score vector at time t, and the measurement data vector (sample) at time t, respectively. $\boldsymbol{I}$ is the identity matrix of size m. The T^2 statistic defined by the sum of normalized squared scores is a measure of the variation within the PCA model, while the Q statistic indicates how well each sample conforms to the PCA model and is a measure of the amount of variation not captured by the k principal components retained in the model. Note that ”PCA model” means the created hyperplane by PCA here, but it also denotes the set of loadings (eigenvectors), eigenvalues, and means and standard deviations of all input variables for PCA in the following.

The T^2 statistic and the Q statistic are complementary used for fault detection with their (upper) control limits that are often approximately expressed by

$$\begin{aligned} T_\alpha^2 &= \frac{k(n+1)(n-1)}{n(n-k)}F_\alpha(k, n-k), && (5) \\ Q_\alpha &= \theta_1\left(\frac{c_\alpha\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0(h_0-1)}{\theta_1^2}\right)^{\frac{1}{h_0}}, && (6) \end{aligned}$$

where $F_\alpha(k, n-k)$ is the value (upper limit) at $100(1-\alpha)\%$ confidence level of the F distribution with $(k, n-k)$ degrees of freedom, $\theta_i = \sum_{j=k+1}^m \lambda_j^i, i = 1, 2, 3, h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$, c_α is the $100(1-\alpha)\%$ standardized normal percentile. Note that T_α^2 can be approximated by $\chi_\alpha^2(k)$, the value at $100(1-\alpha)\%$ confidence level of the χ^2 distribution with k degrees of freedom, if n is sufficiently large since $\frac{(n+1)(n-1)}{n(n-k)} \to 1$ and $kF_\alpha(k, n-k) \to \chi_\alpha^2(k)$ as $n \to \infty$. The contributions of i_{th} component $x_i(t)$ of the data vector $\boldsymbol{x}(t)$ for the Q statistic and for the T^2 statistic could be defined as

$$\begin{aligned} Q_i(t) &:= (\boldsymbol{x}^T(t)(\boldsymbol{I} - \boldsymbol{P}\boldsymbol{P}^T)\boldsymbol{e}_i)^2, && (7) \\ T_i^2(t) &:= (\boldsymbol{x}^T(t)(\boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}^T)\boldsymbol{e}_i)^2, && (8) \end{aligned}$$

where $\boldsymbol{e}_i$ is the i_{th} column of the identity matrix $\boldsymbol{I}$ of size m. It should be noted that the variable contributions are defined so that $Q(t) = \sum_{i=1}^m Q_i(t)$ and $T^2(t) = \sum_{i=1}^m T_i^2(t)$ hold.

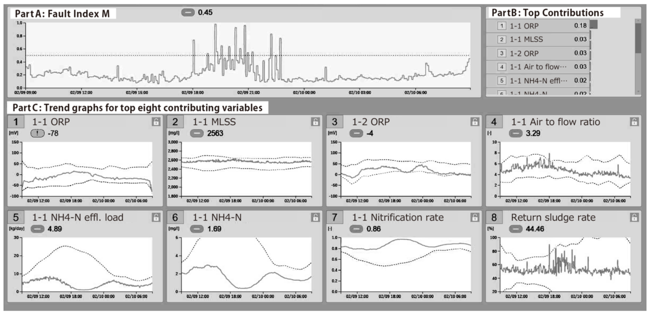
The Q and T^2 statistics with the contributions make it possible to detect faults earlier and to localize cause variables.

3. MODIFIED MSPC FOR REAL-TIME MONITORING

Intuitive design and tractability of monitoring systems will play an important role to enhance the applicability of MSPC to real-time monitoring in existing plants where SCADA is installed. To this end, this section firstly introduces an easy-to-understand UI which seamlessly connects MSPC and existing trend-graph-based monitoring. Then, we propose a novel yet simplified MSPC algorithm without changing the UI, which improves tractabilaity of MSPC model in real time and also may improve intuitive understanding by operators.

3.1. Adjusted-MSPC with Improved User Interface

Figure 1 shows an example of the developed UI consisting of three parts, which will be easy-to-understand.



Part A: Fault Index M

0.45

Part B: Top Contributions

1	1-1 ORP	0.18
2	1-1 MLSS	0.03
3	1-2 ORP	0.03
4	1-1 Air to flow···	0.03
5	1-1 NH4-N effl···	0.02
6	1-1 NH4-N	0.02

Part C: Trend graphs for top eight contributing variables

1 1-1 ORP [mV] -78

2 1-1 MLSS [mg/l] 2563

3 1-2 ORP [mV] -4

4 1-1 Air to flow ratio [-] 3.29

5 1-1 NH4-N effl. load [kg/day] 4.89

6 1-1 NH4-N [mg/l] 1.69

7 1-1 Nitrification rate [-] 0.86

8 Return sludge rate [%] 44.46

Figure 1. Developed user interface combining MSPC-based monitoring and trend-graph-based monitoring

The part A shows the time-series of a fault index M derived from the T^2 and Q statistics. The index M is defined by

$$M(t) \quad := \quad 1 - \exp(-\ln(2)C(t)) \tag{9}$$

$$C(t) \quad := \quad \frac{1}{2}(\frac{Q(t)}{Q_\alpha} + \frac{T^2(t)}{T^2_\alpha}), \tag{10}$$

where C is a scaled combined statistic of the T^2 and Q statistics with its control limit 1. The M is defined so that its range is from 0 to 1 and the control limit becomes 0.5. The reason for introducing M is as follows. Firstly, distinguishing the T^2 and the Q hampers intuitive understanding by plant operators since these two are just a fault index for operators. While introducing the combined index C is sufficient for this purpose, frequently observed outliers of process data generate unnecessary extremely large values of the C and hence the index M is introduced to bounding the range.

The part B is a contribution plot of the M of which contribution M_i is defined by

$$M_i(t) \quad := \quad M(t)\frac{C_i(t)}{C(t)}, \tag{11}$$

$$C_i(t) \quad := \quad \frac{1}{2}(\frac{Q_i(t)}{Q_\alpha} + \frac{T_i^2(t)}{T^2_\alpha}), \tag{12}$$

Note that the contribution M_i is defined so that $M(t) = \sum_{i=1}^{m} M_i(t)$ holds. The variables whose contributions are in the top eight at a present time are listed from the top to the eighth so that operators can easily notice abnormal events.

The part C represents the time-series (trend graphs) of the top eight variables, with their normal operating range calculated by time-dependent standard deviations. Each trend graph is linked to a more detailed one, which can be accessed by clicking on it. An example of this is shown in Figure 2, where the third contributing variable, 1-2 ORP, is displayed.

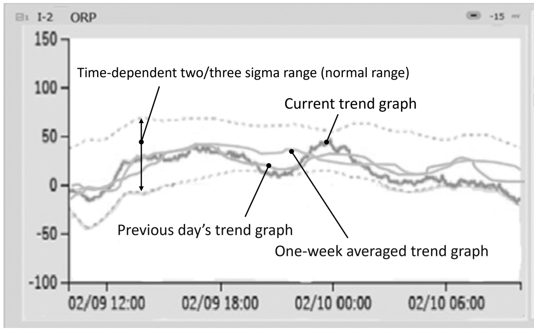


Figure 2. Example of detailed trend graph that is linked to the developed GUI in Figure 1.

As process monitoring based on trend graphs is common in reality, MSPC can be connected to conventional process monitoring in this way. The PCA-MSPC with this UI is referred to as ”Adjusted-MSPC” in the following.

3.2. Modular-MSPC for Improvement of Tractability

While defining Normal Operating Condition (NOC) is crucial for applications of MSPC to real plants, it is sometimes difficult, particularly for non-stational and/or disturbance driven processes such as WWTPs whose performance heavily depends on the ambient temperature and uncontrolled sewer and storm water. To adapt such varying conditions and disturbances, operational conditions should also be adjusted. Thus, no unique NOC can be defined in reality. In addition, sensor failures and/or replacements are the rule rather than the exception in real plants like WWTPs. Thus, PCA model should be updated appropriately in real-time monitoring but when and how to update is difficult in general. The difficulty will be partly caused by complex and strong dependence of PCA models on training data, which in turn makes diagnostic results difficult to interpret uniformly and intuitively. In addition, adding and deleting of input variables for MSPC is not easily handled since PCA treats multivariate data collectively, which also makes model update process more tedious. To cope with it, we propose a simplified reconfigurable MSPC that is referred to as "Modular-MSPC" , which is more simply and clearly dependent on training data. The main idea of Modular-MSPC is to define a new combined statistic named S, instead of the C statistic, by aggregating T^2 statistics of all m single variables and pairwise Q statistics for all possible $\binom{m}{2} = \frac{m(m-1)}{2}$ combinations of variables. In Modular-MSPC, each T^2 or pairwise Q statistic is considered as the basic building block and thus variable contributions are also defined by partly aggregating these basic building blocks, which allows us to add and delete input variables easier. In addition, dependence on training data of Modular-MSPC is clear and simple, which is illustrated below.

Firstly, the T^2 statistic for single i_{th} variable is exactly same as the square of the univariate t statistic, which is just the square of (a sample) of i_{th} variable if the data is already standardized, which is represented as

$$T_i^2(t) = (t_i(t))^2 = (x_i(t))^2, \tag{13}$$

where subscript i stands for i_{th} variable.

Then pairwise Q statistic can be derived by conducting PCA for two variables. Fortunately, $\boldsymbol{P}_{all}$ and $\boldsymbol{\Lambda}_{all}$ can be derived explicitly, which is expressed as

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{R} = \boldsymbol{P}_{all}\boldsymbol{\Lambda}_{all}\boldsymbol{P}_{all}^T \\ &= \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{-\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} 1+r & 0 \\ 0 & 1-r \end{bmatrix} \begin{bmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{-\sqrt{2}}{2} \end{bmatrix}, \end{aligned} \tag{14}$$

where r means the correlation coefficient r_{ij} of i_{th} and j_{th} variables under consideration. The equation (14) shows the following interesting properties as shown in Figure 3.

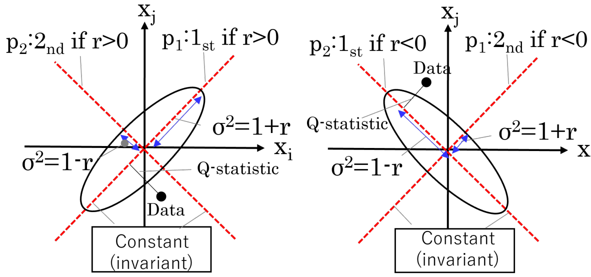


Figure 3. Two dimensional PCA to define pairwise Q

Firstly, $\boldsymbol{P}_{all}$ does not depend on training data, which implies that the direction vector never changed and can be fixed irrespective of training data. Next, diagonal elements of $\boldsymbol{\Lambda}_{all}$ can be characterized by the correlation coefficient r only, which means that the variances along the loadings are explicitly related to the correlation coefficient. Finally, the first principal component loading becomes $\boldsymbol{p}_1 = [\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]^T$ if $r > 0$, and $\boldsymbol{p}_2 = [\frac{\sqrt{2}}{2}, \frac{-\sqrt{2}}{2}]^T$ if $r < 0$, thus one can distinguish the first and the second by just checking the sign of r. By noting the nice properties and assuming that the loading $\boldsymbol{P}$ consists of the first loading, the pairwise Q statistic can be derived as

$$Q_{ij}(t)(= Q_{ji}(t)) = \frac{(x_i(t) - sign(r_{ij})x_j(t))^2}{2}, \tag{15}$$

where Q_{ij}(=Q_{ji}) denotes the pairwise Q statistic between i_{th} variable and j_{th} variable ($i \neq j$). We prefer to rescale Q_{ij} by dividing by its variance $1 - |r_{ij}|$, which is redefined as

$$Q_{ij}(t) {=} (q_{ij}(t))^2 {=} \left(\frac{(x_i(t) - sign(r_{ij})x_j(t))}{\sqrt{2(1 - |r_{ij}|)}} \right)^2, \tag{16}$$

where $q_{ij}(t)$ is defined as the square root of $Q_{ij}(t)$.This rescale allows us to consider that $Q_{ij}(t)$ and $T_i^2(t)$ (for all $i, j = 1, 2, \cdots, m, i \neq j$) follow $\chi^2(1)$ if n is sufficiently large under the standard assumption for deriving the control limits (5) and (6).

By using (13) and (16), non-scaled S statistic denoted S_0 and the i_{th} variable contribution S_{0i} are defined as

$$\begin{aligned} S_0(t) &:= \sum_{i=1}^{m} \left(T_i^2(t) + \frac{1}{2} \sum_{j=1, j\neq i}^{m} Q_{ij}(t) \right), \quad (17) \\ S_{0i}(t) &:= T_i^2(t) + \frac{1}{2} \sum_{j=1, j\neq i}^{m} Q_{ij}(t), \quad (18) \end{aligned}$$

where $S_0(t) = \sum_{i=1}^{m} S_{0i}(t)$ holds. Note that $S_0(t)$ can also be simply expressed as $S_0(t) = \boldsymbol{z}^T(t)\boldsymbol{z}(t)$ by defining $\boldsymbol{z}(t) = [t_1(t), \cdots, t_m(t), q_{12}(t), \cdots, q_{(m-1)m}(t)]^T$ of size $\frac{m(m+1)}{2} \times 1$ vector. To obtain the scaled S statistic, the control limit of S_0 should be decided. A proper control limit can be derived as

$$S_{0\alpha} = \sqrt{\frac{\kappa_2}{2k_0}} \left(\chi_\alpha^2(k_0) - k_0 \right) + \kappa_1, \tag{19}$$

where, $\kappa_i = 2^{i-1}(i-1)! \sum_{j=1}^{\frac{m(m+1)}{2}} \gamma_j^i, i = 1, 2, 3,\ k_0 = 8\kappa_2^3/\kappa_3^2$. $\gamma_j, j = 1, 2, \cdots, m(m+1)/2$ are the eigenvalues of $\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Z}$ where $\boldsymbol{Z}$ is the $n \times m(m+1)/2$ data matrix defined by $\boldsymbol{Z} := [\boldsymbol{z}(1), \boldsymbol{z}(2), \cdots \boldsymbol{z}(n)]^T$. The control limit can be derived under the assumption that $Q_{ij}(t)$ and $T_i^2(t)$ follow $\chi^2(1)$. The Hall-Buckley-Eagleson approximation for a weighted sum of $\chi^2(1)$ random variables (Bodenham & Adams, 2016) is applied after transforming the vectors $\boldsymbol{z}(t), t = 1, 2, \cdots, n$ to their scores. The detail of the derivation of the control limit (19) is presented in the Appendix.

By using the control limit (19), the S statistic and the variable contributions can be simply defined as $S(t) := S_0(t)/S_{0\alpha}$ and $S_i(t) := S_{0i}(t)/S_{0\alpha}$, respectively. One can adopts the S and S_i instead of using the C and C_i without changing the UI presented above. Moreover, interpretability and tractability of Modular-MSPC will be improved since it was derived in a constructive way and simply depends on training data only through the correlation coefficients r_{ij} in the correlation matrix $\boldsymbol{R}$ of all input variables.

Although Modular-MSPC is simpler than the PCA-MSPC (Adjusted-MSPC), performance of Modular-MSPC cannot be understood and thus should be compared with Adjusted-MSPC. Thus it will be compared in the next section in terms of detectability and diagnosability and will be shown that Modular-MSPC actually works well as a process monitoring technique for a real municipal WWTP.

4. APPLICATION OF MODIFIED MSPC TO WWTP

4.1. Wastewater Treatment Plant

The modified MSPC was applied to a part (called the 1_{st} train) of a municipal WWTP in Japan, while the conventional PCA-MSPC had already been evaluated at the same WWTP as a part of a national project called B-DASH (BDASH-Project, 2016). Figure 4 shows the outline of the plant layout of the train. Influent wastewater is firstly stored in the primary settler where solid wastes are removed. Then liquid wastes such as organic matters (COD), nitrogen (NH_4-N,NO_3-N), and phosphorus (PO_4-P) are treated biologically in the reactors with (partial) aeration by the blowers. It is called activated sludge process since the sludge containing microorganisms are growing while treating organic matters. Finally, the activated sludge is separated into solids (sludge) and liquid in the secondary settler. A part of sludge is removed as waste sludge and the remainder is returned to the (pair of) bioreactors connected to common primary/secondary settlers. While a large amount of data for about a thousand variables are collected every 1 minute by a SCADA, we have chosen 82 variables that are relevant to process status of the 1_{st} train as the candidates of the input variables of MSPC. The variables are influent and effluent flow-rates, water levels, flow-rates of pumps and blowers, process control indices such as SRT (Sludge Retention Time), water quality indices such as MLSS (activated sludge concentration), DO (dissolved oxygen), NH_4-N,NO_3-N,PO_4-P and so on (BDASH-Project, 2016).

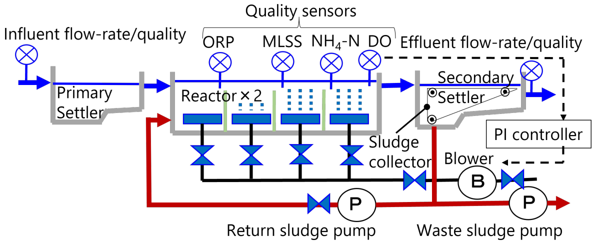


Figure 4. Plant layout for the first train of WWTP

4.2. Evaluation Method

The evaluation was conducted in two phases. In the first phase, the basic performance of Modular-MSPC was evaluated by applying it to historical data, and four diagnosis models with different input variables were adopted. Each diagnosis model has its own specific purpose, such as sensor failure detection, process fault diagnosis, and aeration control performance diagnosis, but details are omitted. Subsequently, each model is termed ModelA, ModelB, ModelC, ModelD, respectively, and 16, 18, 36, 78 variables are incorporated into them in this order. Performance was evaluated by detectability and diagnosability, with the former can be quantitatively defined but the latter is qualitative. The detectability was assessed in terms of (anomaly) detection time and distinguishability of normal and abnormal states. The detection time (DT) was defined by the elapsed time (min.) from t_0 to the detected time when the value of the index M reached the normalized control limit 0.5. The distinguishability was defined by the difference in the M-values before and after an abnormal event. The former, denoted by M_0, is the M-value at the last minute to t_0 and the latter, denoted by M_{max}, is the maximum M-value from t_0 to the time the event was noticed by an operator. Namely, the distinguishability ΔM is defined by $\Delta M := M_{max} - M_0$. The diagnosability was defined as cause localization ability that can be checked by isolating highly contributing variables and qualitatively evaluated in terms of the adequacy of process operation.

The performance of Modular-MSPC was compared to that of Adjusted-MSPC for abnormal events that had occurred at the WWTP in the past. An abnormal event of a sludge collector failure is highlighted here to explain the basic detectability and diagnosability of Modular-MSPC, while such comparisons were also conducted for the events reported in the B-DASH project (BDASH-Project, 2016). The failure occurred at a recorded time t_0 and was noticed by an operator about 24 hours later. During this period, the sludge concentration MLSS decreased and thereby wastewater quality such as NH_4-N was gradually deteriorating.

In the second phase, the performance of Modular-MSPC was assessed qualitatively by applying it in real-time. The plant operators were asked to monitor using a prototype system of Modular-MSPC with the developed UI for about two months in real-time. In this real-time experiment, eleven diagnosis models with different combinations of input variables were applied simultaneously and the models were updated automatically every two weeks by using the process data of the latest two weeks. After the experiment was completed, we investigated meaningful detected abnormal events by interviewing the operators about real process status and operating conditions.

4.3. Evaluation Results

As for the detectability in the first phase, the detection time, a measure of the detectability, certainly depends on the setting of the control limit and the distinguishability may also depend on it. Thus, for fair comparison, we need to set the same value for the significance level α (or equivalently $100(1-\alpha)\%$ confidence level) used in the control limit setting formulas (5),(6), and (19). It is common in industrial statistical process monitoring and control to set a confidence level of about 2σ to 3σ by assuming a normal distribution, which is equivalent to about 97.7% confidence level for one-sided 2σ value ($\alpha = 0.023$) and about 99.9% confidence level for one-sided 3σ value ($\alpha = 0.0013$). It is natural to regard anything exceeding 3σ as an outlier or fault, thus a practically acceptable largest confidence level will be considered to be around 99.9%. To seek an appropriate confidence level, we have compared the initial M-value M_0, the maximum M-value M_{max}, the distinguishability index ΔM, and the detection time DT between Modular-MSPC and Adjusted-MSPC, for all four diagnosis models, and for different settings of the significance level. Table 1 shows the result for the average values for the four models and Table 2 shows the result of all four models. In Table 1 and Table 2, 'A' and 'M' in the parentheses in the first column indicates 'Adjusted-MSPC' and 'Modular-MSPC', respectively. Also, as the threshold setting formulas do not assume a normal distribution, $k\sigma, k = 2, 3, 6$ indicates the corresponding significance level $\alpha = 0.023$, $\alpha = 0.0013$, and $\alpha = 0.00000001$, respectively. We have tested $\alpha = 0.00000001$ in addition to $\alpha = 0.023$ and $\alpha = 0.0013$, as an extreme case. If the normalized control limit 0.5 is already exceeded before the failure occurs, the detection time DT cannot be defined, indicated by'-' in Table 1 and Table 2. In Table 2, A to D in the parentheses in the first row indicate ModelA (16 variables), ModelB(18 variables), ModelC(36 variables), ModelD (78 variables), respectively.

Table 1. Comparison of average detectability indices between Modular-MSPC and Adjusted-MSPC

	2σ	3σ	6σ
M_0(A)	0.54	0.40	0.17
M_0(M)	0.32	0.22	0.09
M_{max}(A)	1.00	1.00	1.00
M_{max}(M)	1.00	1.00	0.97
ΔM(A)	0.46	0.60	0.83
ΔM(M)	0.68	0.78	0.88
DT(A) [min.]	-	-	215
DT(M) [min.]	126	179	329

As can be seen from Table 1 and Table 2, the distinguishability was superior in all cases for Modular-MSPC, and the detection time was faster in all cases for Adjusted-MSPC, provided the detection time DT can be defined. It means that Adjusted-MSPC sometimes showed abnormal M-values before the event occurred, but this phenomenon was not observed in Modular-MSPC. Also, in Modular-MSPC, if the confidence level corresponding to $2\sigma - 3\sigma$ is set as the threshold, it provides an appropriate control limit in most cases. On the contrary, in Adjusted-MSPC, as in the case of the ModelD with 78 variables, it may already show an abnormal M-value before the event occurred unless the confidence level corresponding to 6σ is set, which could be consided as an extreme case and also implies a trial and error search for an appropriate significance level setting is required.

Figure 5 shows an example of the result for the significance level $\alpha = 0.0013$, where the M-values and trend graphs for the top three contributions are displayed, as the developed UI was only available on the online prototype. As can be seen, earlier detection by Adjusted-MSPC was achieved at the cost of decreased distinguishability. In this case, the M-value was already near the normalized control limit 0.5 despite the significance level was set as $\alpha = 0.0013$ corresponding to 3σ.

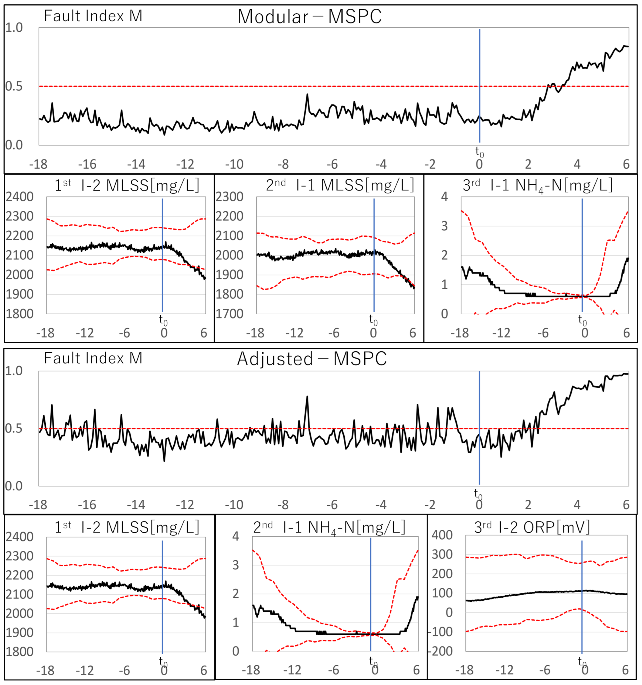


Figure 5. Example of fault diagnosis by Modular-MSPC and Adjusted-MSPC during sludge collector failure

In terms of the diagnosability, while similar contributions were obtained in both the MSPCs, Modular-MSPC seemed to be slightly better than Adjusted-MSPC. In the sludge collector failure, for example, the MLSS of the two reactors – the most directly relevant variable to the event – were ranked as the top two during almost all the time by Modular-MSPC, whereas incorrect variables probably not relevant to the event often ac-

Table 2. Comparison of detectability indices between Modular-MSPC and Adjusted MSPC for each model

	2σ(A)	3σ(A)	6σ(A)	2σ(B)	3σ(B)	6σ(B)	2σ(C)	3σ(C)	6σ(C)	2σ(D)	3σ(D)	6σ(D)
M_0(A)	0.58	0.41	0.14	0.47	0.34	0.12	0.46	0.35	0.14	0.63	0.52	0.28
M_0(M)	0.34	0.23	0.09	0.30	0.19	0.07	0.29	0.21	0.09	0.35	0.26	0.12
M_{max}(A)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
M_{max}(M)	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.93
ΔM(A)	0.42	0.59	0.86	0.53	0.66	0.87	0.54	0.65	0.86	0.37	0.48	0.72
ΔM(M)	0.66	0.77	0.91	0.71	0.81	0.91	0.71	0.79	0.90	0.65	0.74	0.81
DT(A) [min.]	-	105	305	20	145	265	5	85	195	-	-	95
DT(M) [min.]	145	165	360	145	205	430	145	225	350	70	120	175

counted for a part of high ranks of Adjusted-MSPC. Taking Figure 5 as an example, the top two of Modular-MSPC are the MLSS, the most directly relevant variable, and the third is also a relevant variable to the event as the NH_4-N is a component to be removed but it will be deteriorated by this abnormal event. On the other hand, the ORP, which is not relevant to the event, was ranked as the third by Adjusted-MSPC despite it not fluctuating and one of the MLSS was not included in the top three contributors. We have observed that cause localization/isolation ability in Modular-MSPC was better than or at least almost equal to that in Adjusted-MSPC for the events reported in the B-DASH project (BDASH-Project, 2016), the reason has not yet been clarified completely. However, less cause localization ability by the contribution in the conventional PCA-MSPC has been pointed out in literature such as (Camacho et al., 2016; Alcala & Qin, 2009; Joe Qin, 2003; Westerhuis et al., 2000). It was pointed out in these papers that the reasons for the failure of cause localization will be smearing out of information due to data compression by PCA, which can lead to misdiagnosis and fail to identify the cause variables correctly even for simple sensor faults. Thus, the diagnosability of Adjusted-MSPC, a slightly modified conventional PCA-MSPC, will exhibit similar behavior. In addition, the contributions for Q statistic and T^2 statistic are combined after normalizing by control limits Q_α and T^2_α in Adjusted-MSPC, and thus the rank of contributions could fluctuate depending on the significance level α setting. Hence, cause localization by contributions in Adjusted-MSPC sometimes fail to identify correct cause variables. On the other hand, Modular-MSPC was developed in a constructive manner, it is clear under what situations the contribution of a specific variable increases, so the smearing effect observed in the contributions of conventional PCA-MSPC does not occur. We believe that this simple and explainable structure of Modular-MSPC could be related to the improvement in diagnosability.

In the second phase, the M-values sometimes exceeded the threshold 0.5. From such cases, we have identified the 26 notable events including process faults and intentional changes of operating condition by interviewing the operators. The detected notable events could be classified into the following 6 categories:

A. Sensor failures of wastewater quality and flow-rate (8 events)

B. Degradation of control performances such as the so-called controller hunting for DO concentration control and/or not-well controlled DO (6 events)

C. Large and sudden influent load variation (6 events)

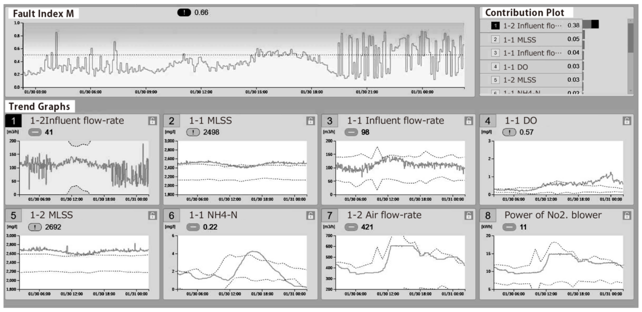
D. Treated effluent water quality problems such as degradation of NH_4-N (3 events)

E. Intentinal operational mode changes such as the set-point changes of the waste and return sludge flow-rates to adjust the MLSS (2 events)

F. Maintenance of pumps and blowers (1 event)

Among these notable events, the following are two examples from category $\mathbf{A}$ and category $\mathbf{B}$.

Figure 6 is an example of the screenshot of the developed UI when the influent flow-rate sensor failure occurred. The measured flow-rate fluctuated, which is inconsistent with the expected behavior of the real flow-rate. This phenomenon was caused by the failure of the circuit board connected to the flow-rate sensor, as confirmed in a post-experiment interview.



Fault Index M

! 0.66

Contribution Plot

1	1-2 Influent flo···	0.38
2	1-1 MLSS	0.05
3	1-1 Influent flo···	0.04
4	1-1 DO	0.03
5	1-2 MLSS	0.03
6	1-1 NH4-N	0.02

Trend Graphs

1	1-2Influent flow-rate	[m3/h]	41
2	1-1 MLSS	[mg/l]	! 2498
3	1-1 Influent flow-rate	[m3/h]	98
4	1-1 DO	[mg/l]	! 0.57
5	1-2 MLSS	[mg/l]	! 2692
6	1-1 NH4-N	[mg/l]	0.22
7	1-2 Air flow-rate	[m3/h]	421
8	Power of No2. blower	[kWh]	11

Figure 6. Screenshot of UI when flow-rate sensor failure occured

Figure 7 is another example of the screenshot of the developed UI when controller hunting of DO concentration PI control occured. As can be seen, the measured DO, air flow rate, blower motor rotational speeds (RPM:Revolutions Per

Minute) are started to oscillate near the right end of the trend graphs. This phenomenon, known as feedback controller hunting, is caused by a mismatch between the feedback PI controller parameters and the process response. This phenomenon occurs when the PI control parameter values are too strong, which may lead to poor energy efficiency and process performance and has the potential to damage equipment such as blowers as well. In this experiments, this PI controller hunting occured due to the change of the operating point.

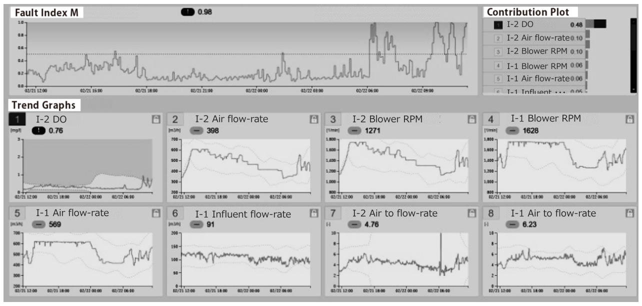


Figure 7. Screenshot of UI when PI controller huting occured

While details of other detected events are omitted here, we have confirmed through the interview that the detected 26 events provided some useful information for the operator. Also, we received operators' feedback that the developed UI is easy to understand and helpful for process monitoring. This is because a single fault indicator was used and access to the trend-graph of a variable included in top contributing variables was easy.

5. CONCLUSION

This paper has presented a practical monitoring method based on a modified MSPC and has demonstrated its effectiveness through application to a real municipal WWTP in Japan, in both offline and online manners. Firstly, a novel user interface was developed by integrating the conventional PCA-MSPC with existing SPC-like process monitoring. Subsequently, a simplified and reconfigurable PCA-MSPC, termed "Modular-MSPC", was proposed. The adoption of the Modular-MSPC not only improved model tractability, but also demonstrated good performance in terms of detectability and diagnosability, as shown through the WWTP application. Future works will involve long-term evaluation of Modular-MSPC across various real plants, and the development of an effective model update algorithm and a useful automatic input variable selection method.

REFERENCES

Alcala, C. F., & Qin, S. J. (2009). Reconstruction-based contribution for process monitoring. *Automatica*, *45*(7), 1593–1600.

AlGhazzawi, A., & Lennox, B. (2008). Monitoring a complex refining process using multivariate statistics. *Control Engineering Practice*, *16*(3), 294–307.

BDASH-Project. (2016). *Guideline for introducing ict-based advanced process control and remote diagnosis technology for efficient wastewater treatment plant operation* (Tech. Rep. No. 939). National Institute for Land and Infrastructure Management. Retrieved from `http://www.nilim.go.jp/lab/bcg/siryou/tnn/tnn0939.htm` (Accessed on: 2024-02-14)

Bodenham, D. A., & Adams, N. M. (2016). A comparison of efficient approximations for a weighted sum of chi-squared random variables. *Statistics and Computing*, *26*(4), 917–928.

Camacho, J., Pérez-Villegas, A., García-Teodoro, P., & Maciá-Fernández, G. (2016). Pca-based multivariate statistical network monitoring for anomaly detection. *Computers & Security*, *59*, 118–137.

Choi, S. W., Morris, J., & Lee, I.-B. (2008). Nonlinear multiscale modelling for fault detection and identification. *Chemical engineering science*, *63*(8), 2252–2266.

de Oliveira, R. R., Pedroza, R. H., Sousa, A. O., Lima, K. M., & de Juan, A. (2017). Process modeling and control applied to real-time monitoring of distillation processes by near-infrared spectroscopy. *Analytica chimica acta*, *985*, 41–53.

GarcÇa-Alvarez, D. (2009). Fault detection using principal component analysis (pca) in a wastewater treatment plant (wwtp). *Proceedings of the International Student's Scientific Conference*, 55–60.

Ge, Z., Yang, C., & Song, Z. (2009). Improved kernel pca-based monitoring approach for nonlinear processes. *Chemical Engineering Science*, *64*(9), 2245–2255.

Jackson, J. E., & Mudholkar, G. S. (1979). Control procedures for residuals associated with principal component analysis. *Technometrics*, *21*(3), 341–349.

Jaffel, I., Taouali, O., Harkat, M. F., & Messaoud, H. (2017). Kernel principal component analysis with reduced complexity for nonlinear dynamic process monitoring. *The International Journal of Advanced Manufacturing Technology*, *88*, 3265–3279.

Joe Qin, S. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *17*(8-9), 480–502.

Kano, M., Nagao, K., Hasebe, S., Hashimoto, I., Ohno, H., Strauss, R., & Bakshi, B. R. (2002). Comparison of multivariate statistical process monitoring methods with applications to the eastman challenge problem. *Computers & chemical engineering*, *26*(2), 161–174.

Kourti, T., & MacGregor, J. F. (1996). Multivariate spc methods for process and product monitoring. *Journal of quality technology*, *28*(4), 409–428.

Kresta, J. V., Macgregor, J. F., & Marlin, T. E. (1991). Multivariate statistical monitoring of process operating performance. *The Canadian journal of chemical engineering*, *69*(1), 35–47.

Lemaigre, S., Adam, G., Goux, X., Noo, A., De Vos, B., Gerin, P. A., & Delfosse, P. (2016). Transfer of a static pca-mspc model from a steady-state anaerobic reactor to an independent anaerobic reactor exposed to organic overload. *Chemometrics and Intelligent Laboratory Systems*, *159*, 20–30.

Rosén, C. (2001). *A chemometric approach to process monitoring and control-with applications to wastewater treatment operation*. Lund University.

Sandberg, E., Lennox, B., & Undvall, P. (2007). Scrap management by statistical evaluation of eaf process data. *Control engineering practice*, *15*(9), 1063–1075.

Uchida, Y., Fujiwara, K., Saito, T., & Osaka, T. (2022). Process fault diagnosis method based on mspc and lingam and its application to tennessee eastman process. *IFAC-PapersOnLine*, *55*(2), 384–389.

Westerhuis, J. A., Gurden, S. P., & Smilde, A. K. (2000). Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and intelligent laboratory systems*, *51*(1), 95–114.

Wise, B. M., & Gallagher, N. B. (1996). The process chemometrics approach to process monitoring and fault detection. *Journal of Process Control*, *6*(6), 329–348.

Zhao, L.-T., Yang, T., Yan, R., & Zhao, H.-B. (2022). Anomaly detection of the blast furnace smelting process using an improved multivariate statistical process control model. *Process Safety and Environmental Protection*, *166*, 617–627.

APPENDIX: THRESHOLD DERIVATION

In this section, we derive the control limit (19) under some acceptable assumptions. First, we make the following two assumptions.

H1) Each univariate statistic $t_i(t), i = 1, 2, \cdots, m$ in the definition of the univariate T^2 statistic in (13) follows a standard normal distribution, namely, a normal distribution with zero mean and unit variance.

H2) Each pairwise statistic $q_{ij}(t), i, j = 1, 2, \cdots, m, i \neq j$ defined in (16) follows a standard normal distribution.

The assumption that a certain statistic follows a normal distribution is often accepted, which simplifies statistical analysis while maintaining a degree of rationality. From assumptions H1) and H2), each of the statistics $T_i^2(t), i = 1, 2, \cdots, m$ and $Q_{ij}(t), i, j = 1, 2, \cdots, m, i \neq j$ follows a χ^2 distribution with one degree of freedom. Therefore, the non-scaled S statistic S_0, which is defined as the total sum of all the statistics $T_i^2(t), i = 1, 2, \cdots, m$ and $Q_{ij}(t), i, j = 1, 2, \cdots, m, i \neq j$, is expected to follow a sum of χ^2 distributions. Although the sum of p χ^2 random variables with one degree of freedom follows a χ^2 distribution with p degrees of freedom if the component random variables are independent, this is not the case for the S_0 statistic since the component univariate statistics $T_i^2(t), i = 1, 2, \cdots, m$ and $Q_{ij}(t), i, j = 1, 2, \cdots, m, i \neq j$ are typically correlated with each other. This is because the univariate statistics $T_i^2(t), i = 1, 2, \cdots, m$ and $Q_{ij}(t), i, j = 1, 2, \cdots, m, i \neq j$ are defined using input process variables $\mathbf{x}(t)$ that are often correlated with each other in nature. To address this problem, consider a linear transformation that transforms the set of $\frac{m(m+1)}{2}$ correlated variables,

$$\boldsymbol{z}(t) = [t_1(t), \cdots, t_m(t), q_{12}(t), \cdots, q_{(m-1)m}(t)]^T \tag{20}$$

into another set of $\frac{m(m+1)}{2}$ uncorrelated variables, without changing the value of the S_0 statistic. That is, the challenge is to find a matrix $\boldsymbol{L}$ of size $\frac{m(m+1)}{2} \times \frac{m(m+1)}{2}$ and a corresponding transformed vector $\boldsymbol{w}(t)$, which satisfies the following properties.

P1) $\boldsymbol{w}(t) = \boldsymbol{L}\boldsymbol{z}(t) = [w_1(t), w_2(t), \cdots, w_{\frac{m(m+1)}{2}}(t)]^T$ for any t, where each $w_i(t), i = 1, 2, \cdots \frac{m(m+1)}{2}$ follows a normal distribution independently.

P2) $\boldsymbol{z}^T(t)\boldsymbol{z}(t) = \boldsymbol{w}^T(t)\boldsymbol{w}(t)$ for any t.

Property P1) is necessary to apply well-known statistical techniques for the sum of $p(> 1)$ i.i.d. (independently and identically distributed) χ^2 random variables (Bodenham & Adams, 2016). Property P2) is considered because our goal is to derive the control limit of the statistic $S_0 = \boldsymbol{z}^T(t)\boldsymbol{z}(t)$.

One can obtain a candidate for $\boldsymbol{L}$ to satisfy the properties P1) and P2) by using the SVD for the data matrix $\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Z}$,

$$\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Z} = \boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{Q}^T, \tag{21}$$

where $\boldsymbol{Z}$ is the $n \times m(m+1)/2$ data matrix defined by $\boldsymbol{Z} := [\boldsymbol{z}(1), \boldsymbol{z}(2), \cdots \boldsymbol{z}(n)]^T$, n is the number of samples, m is the number of variables. In (21), $\boldsymbol{Q}$ is a unitary matrix and $\boldsymbol{\Gamma} = diag(\gamma_1, \gamma_2, \cdots, \gamma_{m(m+1)/2})$ is a diagonal matrix. It should be noted the equation (21) is the same as the PCA for the data matrix $\boldsymbol{Z}$. Thus, the score matrix $\boldsymbol{V} := [\boldsymbol{v}(1), \boldsymbol{v}(2), \cdots \boldsymbol{v}(n)]^T$ can be defined by $\boldsymbol{V} := \boldsymbol{Z}\boldsymbol{Q}$, which can be rewritten as $\boldsymbol{V}^T = \boldsymbol{Q}^T\boldsymbol{Z}^T$. Since each element of $\boldsymbol{z}(t), t = 1, 2, \cdots, n$ in the data matrix $\boldsymbol{Z}^T$ follows a standard normal distribution, then the corresponding element of the $\boldsymbol{v}(t), t = 1, 2, \cdots, n$ also follows a (non-standard) normal distribution since the score matrix $\boldsymbol{V}^T$ is a linear transformation of the data matrix $\boldsymbol{Z}^T$ by $\boldsymbol{Q}^T$. Furthermore, the $\frac{m(m+1)}{2}$ elements of $\boldsymbol{v}(t) := [v_1(t), v_2(t), \cdots, v_{\frac{m(m+1)}{2}}(t)]$, that is, $v_i(t)$, $i = 1, 2, \cdots, \frac{m(m+1)}{2}$, are uncorrelated with each other, which implies that they are independent since uncorrelation and independence are equivalent if data are assumed to follow normal distributions.

Next, it should be noted that the average of the S_0 statistic $\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{z}^T(t)\boldsymbol{z}(t)$ is equal to tr$(\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Z})$ and the value is exactly the same as the sum of the diagonal matrix $\boldsymbol{\Gamma}$, that is, $\sum_{k=1}^{\frac{m(m+1)}{2}}\gamma_k$. It is also equal to tr$(\frac{1}{n}\boldsymbol{V}^T\boldsymbol{V})$ since the equation (21) can be rewritten as $\frac{1}{n}\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{\Gamma}$.

Therefore, the properties P1) and P2) could be satisfied, if one defines $\boldsymbol{w}(t)$ and $\boldsymbol{L}$ as $\boldsymbol{w}(t) := \boldsymbol{v}(t)$ and $\boldsymbol{L} := \boldsymbol{Q}^T$, respectively. Precisely speaking, the properties P1) and P2) are satisfied only for the finite samples $\boldsymbol{w}(t)$ and $\boldsymbol{z}(t)$, $t = 1, 2, \cdots, n$. But one can expect that the properties P1) and P2) would be satisfied if sufficiently many samples are used for the data matrix $\boldsymbol{Z}$,i.e., $n \to \infty$.

Based on the above understanding, it would be reasonable to make the following additional assumption.

H3) The newly defined $\frac{m(m+1)}{2}$ variables $u_1(t), u_2(t), \cdots, u_{\frac{m(m+1)}{2}}(t)$ follow an identical standard normal distribution and are independent, where,
$u_i(t) := w_i(t)/\sqrt{\gamma_i}, i = 1, 2, \cdots \frac{m(m+1)}{2}$,
$\boldsymbol{w}(t) := \boldsymbol{L}\boldsymbol{z}(t) = \boldsymbol{Q}^T\boldsymbol{z}(t) = [w_1(t), w_2(t), \cdots, w_{\frac{m(m+1)}{2}}(t)]^T$,
$\boldsymbol{Q}$ and $\boldsymbol{\Gamma} = diag(\gamma_1, \gamma_2, \cdots, \gamma_{m(m+1)/2})$ are defined in (21).

Under the assumption H3), the S_0 statistic follows a weighted sum of $\frac{m(m+1)}{2}$ i.i.d. χ^2 distribution since the following equation holds.

$$\begin{aligned} S_0(t) &= \boldsymbol{z}^T(t)\boldsymbol{z}(t) = \boldsymbol{w}^T(t)\boldsymbol{w}(t) \\ &= \sum_{i=1}^{\frac{m(m+1)}{2}} w_i^2(t) = \sum_{i=1}^{\frac{m(m+1)}{2}} \gamma_i u_i^2(t) \end{aligned} \qquad (22)$$

Now, one can apply an efficient approximation of the cumulative distribution function (cdf) of a positively weighted sum of N i.i.d. χ^2 random variables as presented in (Bodenham & Adams, 2016). The paper (Bodenham & Adams, 2016) compared several approximation methods in terms of the computation time and the accuracy, and recommended to use the Hall-Buckley-Eagleson method to approximate the calculation of the cdf of a weighted sum of N i.i.d. χ^2 random variables. The direct application of the Hall-Buckley-Eagleson method by setting a proper confidence level leads to the control limit presented in (19).

It should be noted that one has to calculate $\frac{m(m+1)}{2}$ eigenvalues $\gamma_i, i = 1, 2, \cdots \frac{m(m+1)}{2}$ of $\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Z}$ to estimate the control limit, which may require considerable computation burden if the number of the variables m increases. However, the non-zero eigenvalues of $\frac{1}{n}\boldsymbol{Z}^T\boldsymbol{Z}$ are equal to those of $\frac{1}{n}\boldsymbol{Z}\boldsymbol{Z}^T$, thus one can also calculate the eigenvalues by using the matrix $\frac{1}{n}\boldsymbol{Z}\boldsymbol{Z}^T$instead if $\frac{m(m+1)}{2} > n$. Moreover, it is often the rule rather than the exception that the several largest $\gamma_i, i = 1, 2, \cdots$ dominate the sum of all the eigenvalues, thus it is often sufficient only to calculate the several largest eigenvalues in practice.