frontiers

Feature-based Statistical Process Monitoring for Pressure Swing Adsorption Processes

Jangwon Lee[1], Ankur Kumar[2], Jesus Flores-Cerrillo[2], Jin Wang[1*], Q. Peter He[1*]

[1]Department of Chemical Engineering, Auburn University, AL, 36849, USA

[2]Smart Operations & Advanced Technology, Linde Technology Center, Tonawanda, NY 14150, USA

*** Correspondence:**
Q. Peter He
qhe@auburn.edu

Jin Wang
wang@auburn.edu

Keywords: pressure swing adsorption, fault detection, fault diagnosis, statistical process monitoring, statistics pattern analysis, feature space monitoring.

Abstract

Pressure swing adsorption (PSA) is a widely used technology to separate a gas product from impurities in a variety of fields. Due to the complexity of PSA operations, process and instrument faults can occur at different parts and/or steps of the process. Thus, effective process monitoring is critical for ensuring efficient and safe operations of PSA systems. However, multi-bed PSA processes present several major challenges to process monitoring. First, a PSA process is operated in a periodic or cyclic fashion and never reaches a steady state; Second, the duration of different operation cycles is dynamically controlled in response to various disturbances, which results in a wide range of normal operation trajectories. Third, there is limited data for process monitoring, and bed pressure is usually the only measured variable for process monitoring. These key characteristics of the PSA operation make process monitoring, especially early fault detection, significantly more challenging than that for a continuous process operated at a steady state. To address these challenges, we propose a feature-based statistical process monitoring (SPM) framework for PSA processes, namely feature space monitoring (FSM). Through feature engineering and feature selection, we show that FSM can naturally handle the key challenges in PSA process monitoring and achieve early detection of subtle faults from a wide range of normal operating conditions. The performance of FSM is compared to the conventional SPM methods using both simulated and real faults from an industrial PSA process. The results demonstrate FSM's superior performance in fault detection and fault diagnosis compared to the traditional SPM methods. In particular, the robust monitoring performance from FSM is achieved without any data preprocessing, trajectory alignment or synchronization required by the conventional SPM methods.

1 Introduction

The synthetic zeolites developed by Union Carbide in the 1950s enabled the development of the pressure swing adsorption (PSA) processes. The first industrial application of PSA went on stream in 1966 at a Union Carbide production facility. Since then, PSA has been widely used to separate a gas product from impurities in various fields, from traditional bulk gas separation and drying, to CO_2 sequestration, trace contaminant removal, and many others. A good review of the historical development of PSA technology can be found in (Elseviers et al., 2015). With ever-increasing product capacity and carefully designed operations, modern multi-bed PSA systems can take full advantage of the feed pressure to optimize performance and recover more product gases. For multi-bed PSA systems, adsorber vessels are connected by a complex pipe network with literally hundreds of valves to automatically switch the gas flows among the beds, which results in an intrinsically transient, cyclic, highly nonlinear, and complex dynamic process. As industrial adsorbents are usually highly efficient and stable, major production disruptions are most often caused by valve-related problems, such as internal leakage or stiction. If a potential valve problem could be detected in real-time while still in its early stage, corresponding actions can be scheduled as an online maintenance event, which can be conducted without the downtime and in coordination with other process and business considerations. In other words, if a fault can be detected early, the problem can be addressed with minimum disruption before it escalates to a highly costly emergency shutdown. Clearly, successful early fault detection and diagnosis can greatly improve the PSA process throughput, product quality, and economic performance. In addition, such a process monitoring system can serve as a remote monitoring and early warning system for unattended or autonomous PSA operations.

The intrinsically transient and cyclic operation of PSA processes renders most available fault detection and diagnosis solutions ineffective. Despite the importance and potential impact of PSA process monitoring, research in this area has been scarce. Pan et al. (2004) proposed a monitoring approach for continuous processes with periodic characteristics by identifying a stochastic state space model that captures the statistical behavior of changes occurring from one period to another. This approach was validated using a wastewater treatment process (WWTP). While there are similarities between WWTP and PSA processes, there are also major differences. Most notably, for the activated sludge process, which is the central part of a WWTP, there is a strong cycle-to-cycle dynamics due to the continuous growth of the microorganisms, which provides a "linkage" from cycle to cycle. In comparison, for PSA processes, the cycle-to-cycle dynamics is almost non-existent due to the absence of such a linkage between cycles. In addition, the activated sludge process is a natural periodic process with a somewhat constant cycle time driven by the diurnal temperature and light changes. As a result, obtaining the same number of measurements from each period can be easily achieved, which is required by the state space modeling approach. In contrast, PSA is an engineered periodic process, with cycle time dynamically controlled in response to many disturbances that affect a PSA operation, including varying customer demands, operation schedule adjustment based on electricity pricing to minimize cost, and/or raw material feed composition variations. As a result, the cycle time is frequently and often significantly adjusted, which does not satisfy the condition that each cycle contains the same number of measurements as required by the state space approach proposed in (Pan et al., 2004). In addition, the state space inferential prediction proposed by Pan et al. requires quality-relevant process output, which we do not have in this study. Recently, Wang et al. (2017) proposed a geometric framework for the monitoring and fault detection of periodic processes. The fault detection is based on the "centroids of the centroids" of the training/normal cycles and a corresponding confidence region defined based on them. The proposed approach was applied to a simulated two-bed PSA process and showed superior performance compared to the conventional

dynamic PCA (DPCA) and multi-way PCA (MPCA) methods. For the simulated PSA process, 26 variables were used for process monitoring, including feed flow rates, pressures, and concentrations in and across both beds. However, in industrial PSA processes, most of these variables are not measured, especially the concentrations in and across the beds. In fact, for almost all PSA plants, pressure is the only process variable constantly monitored. In this case, the method proposed by (Wang et al., 2017) is not applicable as there is no centroid for a single variable. Another proposed method for monitoring industrial PSA processes is a US patent (Arslan et al., 2014). This method first applied a moving window discrete Fourier transform (DFT) to convert process data (i.e., bed pressure profiles) into frequency spectra; next, a number of "relevant" peaks were identified from the frequency spectra; and finally the logarithm of the amplitude ratio of peak k between beds i and j is computed over time to capture the normal process behavior and monitored for fault detection. However, the "relevant peaks" were selected in an ad hoc manner, and there is no description on how "relevance" was evaluated or how the "relevant" peaks were identified. Recently, slow feature analysis (SFA) has been proposed to capture slowly varying dynamics in a process (e.g., dynamics across multiple cycles) (Shang et al., 2015). SFA enables separation of normal process changes due to process dynamics from process faults. SFA has been utilized for the monitoring of nonstationary chemical processes subject to time variant conditions (Zhao & Huang, 2018) and cyclic processes including WWTPs (Hong et al., 2020; K. Wang et al., 2021). However, SFA is not suitable for PSA processes due to the lack of slow cycle-to-cycle dynamics.

To develop a process monitoring solution that is suitable for PSA and other cyclic industrial processes, we present a different approach based on the feature space monitoring (FSM) framework we developed recently (He & Wang, 2018). Instead of monitoring the original pressure profile of a PSA process, we first conduct feature engineering, where statistical and shape/morphological features are computed based on the pressure profile to capture the characteristics of each step of the operation cycle. Next, these features are grouped by cycles and monitored by a linear or nonlinear MSPM method for fault detection and diagnosis. Through feature engineering and selection, we not only can readily address the unique challenges associated with cyclic processes, such as the unequal duration for different cycles/steps, but also could detect subtle changes early from a wide range of normal cycle durations. The rest of the paper is organized as follows. Sec. 2 discusses the key characteristics of the industrial PSA process and the challenges posed to the conventional MSPM methods by these characteristics. Sec. 3 briefly reviews statistics pattern analysis (SPA), which is the predecessor and a special case of FSM. Sec. 4 introduces the proposed FSM method for PSA processes. Sec. 5 presents several case studies, including simulated and real faults in an industrial PSA process, to demonstrate the performance of the proposed method, which is compared to those of the conventional MPCA-based methods. Finally, Sec. 6 discusses the results and draws some conclusions.

2 PSA Process Characteristics and Their Challenges to Conventional MSPM

In this section, we discuss the unique characteristics of PSA processes and how these characteristics pose challenges to process monitoring.

2.1 PSA process characteristics

PSA processes are operated on repeated cycles of adsorption and regeneration. As shown in Figure 1, the bed pressure is raised during the adsorption step and the impurities are adsorbed by the adsorbent, providing the high-purity product gas. During the regeneration step, the bed pressure is lowered and the impurities are cleaned or purged from the adsorbent, allowing the adsorption-regeneration cycle to be repeated. Therefore, a PSA process is a continuous process but never operates at any single

steady state. Instead, it repeats a sequence of operation steps over and over. This is usually termed a cyclic steady-state process, where cycles are very similar to each other, and a whole cycle is considered a "steady-state".

[Figure 1 here]

To take full advantage of the feed pressure and to recover more product gas, multi-bed multi-step PSA systems have been widely applied in industrial applications. In terms of process monitoring, the bed pressure is always measured and is often the only variable constantly measured for PSA processes. The industrial data utilized in this work were collected from one of Linde's 12-bed 15-step PSA systems. Figure 2 shows two common ways to visualize pressure trajectories in a multi-bed PSA process. Due to the sensitivity of the process's actual operation and production data, all axis tick labels in this and other figures are omitted when real operation data are used. Figure 2 (A) shows time-series pressure profiles of multiple beds (only three out of twelve beds are shown here to reduce clutter). This type of pressure time-series plot is useful for visualizing and observing between-bed variations. However, only severe faults that significantly deviate from the nominal trajectory can be detected by the naked eyes using this type of plot; in addition, it becomes very cluttered and difficult to read if all beds were plotted on the same figure. Another way to visualize the pressure profile within a bed over multiple cycles is to overlay cycles based on the start of each cycle, as illustrated in Figure 2 (B). This type of plot can be used to visualize within-bed variations. However, due to the variable duration of cycles, again, only severe faults that show significant deviations from the normal operation can be detected directly by the naked eyes from this type of plot.

[Figure 2 here]

In terms of process monitoring, PSA processes share more similarities with batch processes than with continuous processes. For example, PSA and batch processes can both have variable batch/cycle duration and step durations; they are often dynamic transient processes and do not have a steady state. The variable nature of the PSA cycle duration is demonstrated in Figure 3 (A), which plots the durations of different cycles from one PSA bed. For the PSA process studied in this work, each cycle consists of 15 steps, as illustrated in Figure 2. For the step durations, about half of the steps follow similar trends as the cycle duration, while the remaining steps have relatively constant durations. Figure 3 (B) plots the variable step duration of the adsorption step across different cycles, and Figure 3 (C) plots the relatively constant step duration of an equalization step across different cycles. For the PSA process studied in this work, the cycle duration is in the order of tens of minutes and the step durations vary from seconds to minutes.

[Figure 3 here]

Several observations can be made from these plots. First, the cycles are asynchronous across different beds; for the same bed, the cycles do not exactly overlap with each other either. Second, despite the overall highly nonlinear behavior for each cycle, the pressure profile for each individual step is usually much simpler and can be approximated by a simple linear or polynomial function. Finally, not only the cycle durations but also the step durations vary from cycle to cycle. It is important to note that the variations in cycle/step duration is not caused by unmeasured normal process variations, instead, it is a result of deliberate control of cycle and step durations to ensure product quality in response to dynamic scheduling and/or measured disturbances such as demand change and weather conditions. In addition, these characteristics are not unique to PSA processes but are rather common to other cyclic steady-state processes, such as heat exchanger networks under fouling with cleaning-

in-place (CIP) operations (Georgiadis & Papageorgiou, 2000), and catalytic conversion processes where the catalyst undergoes periodic deactivation and activation.(Jain & Grossmann, 1998)

2.2 Challenges in monitoring PSA processes

As discussed above, normal PSA operations cover a wide range of pressure trajectories, due to the dynamically controlled step/cycle durations in response to external disturbances. It is clearly a highly challenging task to detect a subtle fault early from a wide range of normal cycle/step durations with the bed pressure as the only monitored variable. In addition, the characteristics of the PSA processes (and cyclic steady-state processes in general), including asynchronous trajectories, variable cycle/step durations, and nonlinear dynamics, present significant challenges to process monitoring. These challenges cannot be effectively addressed by commonly used multivariate statistical monitoring (MSPM) methods, including both conventional MSPM methods such as MPCA, trilinear decomposition (TLD), and parallel factor analysis (PARAFAC) (Wise et al., 1999), and more recent methods such as multi-way independent component analysis (MICA) (Yoo et al., 2004) and kernel PCA (KPCA) (Choi et al., 2005). These methods assume that the normal process data follow the same distribution and require the construction of a two-dimensional (2-D) data matrix (for data unfolding approaches) or a 3-D data array (for multi-way approaches). In other words, they require synchronization of all steps within a cycle to achieve equal step and cycle durations. Trajectory synchronization can be done through different ways, including simple cut, interpolation, dynamic time warping (DTW), etc. However, these preprocessing steps have their drawbacks, including trajectory distortion, information loss, etc. (He & Wang, 2007, 2018). In particular, synchronization is undesirable for PSA processes because the step durations are dynamically controlled and may contain important information on the state of the process operation. Artificially changing the step/cycle durations may distort the contained information and negatively affect the fault detection and diagnosis performance.

In this work, built upon our work in batch process monitoring that can naturally handle variable batch/step durations, we develop an FSM approach for PSA processes. We show that a balance between sensitivity and robustness of the FSM approach can be achieved through feature engineering and selection, which enables early detection of subtle faults with very low false alarm rate.

3 A Brief Review of Statistics Pattern Analysis

In traditional MSPM approaches for process monitoring, such as PCA and PLS-based approaches, it is inexplicitly assumed that normal process data (or scores in principal component subspace) follow a multivariate Gaussian distribution. However, this assumption is usually not satisfied in industrial applications, especially for batch processes whose data are often highly non-Gaussian. Statistics pattern analysis (SPA) was proposed to address the non-Gaussian process data commonly seen in industrial processes. In SPA, various statistics of process variables, instead of process variables themselves, are modeled for process monitoring. A statistics pattern (SP) is a collection of various statistics calculated using process data, which captures the characteristics of individual variables (e.g., mean and variance), the interactions among different variables (e.g., covariance), the dynamics (e.g., auto-, cross-correlations), as well as process nonlinearity and process data non-Gaussianity (e.g., skewness, kurtosis, and other higher-order statistics or HOS). SPA has been implemented for both continuous and batch process monitoring. For continuous processes, SPs corresponding to different time periods are computed using a moving window approach. For batch processes, the SP for each batch (or each step in a batch) is computed using all measurements from the batch (or step). In this way, the variable batch/step duration can be naturally handled without any data preprocessing.

For process monitoring, SPA assumes that the SPs of normal operations follow a similar pattern (i.e., normal pattern), while the SPs of abnormal or faulty operations must show some deviation from the normal pattern. A multivariate statistical model can be developed for the normal SPs, which enables the determination of a boundary for normal operation or threshold for fault detection. The implementation of SPA can be simplified by assuming that the normal SPs follow a multivariate Gaussian distribution. Although this assumption appears to be the same as the traditional MSPM methods, it is important to note that this assumption (i.e., normal SPs follow a Gaussian distribution) is usually satisfied to a much better degree for SPs than for the measured process variables themselves. As different statistics are the averages of different functions of the variable measurements in a window/batch/step, the distribution of SPs is asymptotically Gaussian. This argument is supported by the central limit theorem (CLT) under weak dependencies, which relaxes the requirement on the independency among different random variables (Dedecker & Rio, 2008). It was further shown that the CLT applies to sums of bounded random variables generated from stationary dynamic systems (Pène, 2005), which applies to different statistics computed using measurements collected from stable processes. The assumption was also validated in (He & Wang, 2011) for batch process monitoring. With this simplification, the characteristics of normal SPs can be captured by the covariance structure of SPs, similar to PCA, and a threshold can be defined (e.g., based on Hotelling's T^2 or squared prediction error (SPE)). The test SPs can then be projected onto the model and the obtained metric such as T^2 or SPE is compared to the threshold for fault detection. More details on batch-based SPA can be found in (He & Wang, 2011). Since the introduction of SPA, several variations and extensions of SPA have been reported in the literature for process monitoring (He & Xu, 2016; Yang et al., 2018; Zhang et al., 2018; Zhou & Gu, 2019).

4 The Proposed Feature Space Monitoring Framework

As PSA and other cyclic continuous processes share many similarities with batch processes, we expect SPA for batch monitoring can be extended to monitor PSA processes. However, major differences between PSA and regular batch processes must be considered. For the PSA process studied here, the bed pressure is the only measured variable, therefore only univariate statistics can be calculated for process monitoring. In addition, one major challenge for PSA monitoring is that although under tight process control, the normal PSA operation has a wide distribution of step/cycle durations in response to disturbances such as customer demand and scheduling based on electricity pricing. Therefore, normal PSA operation data exhibit a wide distribution of normal cycle trajectory, which makes the detection, not to mention early detection, of abnormal cycles highly challenging. To address this challenge, we explore the power of feature engineering to achieve both sensitivity and robustness in the monitoring performance, as well as minimal data preprocessing for easy practical implementation. Once a fault is detected, it is desirable to identify in which step the fault has occurred, so that the corresponding valves, bed and/or pipeline can be identified for further examination. The proposed fault detection and diagnosis framework is termed feature space monitoring or FSM. There are three steps involved in the proposed FSM framework: 1) feature engineering and selection; 2) fault detection; and 3) fault diagnosis. They are discussed in the following sections.

4.1 Feature engineering

As shown in Figure 2, although a complete cycle of a PSA process is highly nonlinear, each step is much simpler and can be described by a simple linear or polynomial model. Therefore, in this work, we compute different features for each step separately. In addition to univariate statistics, we explore morphological features to better capture the characteristics of pressure profile in each step of the

process. To handle the irregularities in industrial data, for each characteristic we consider multiple features that may exhibit different level of sensitivity to outliers. For example, to assess the dispersion of pressure measurement during a processing step, we compare features that use the mean of the pressure measurements as the reference with others that use the median as the reference. Based on the observation of the pressure profiles and discussions with process engineers, totally 12 features (as defined in the remaining section), both statistical and morphological, are examined in this work to determine if they would provide adequate process monitoring. All features are calculated for each step using raw pressure measurements without any preprocessing such as synchronization, centering, scaling or normalization.

In this work, we use a vector $\boldsymbol{x_i} \in \Re^{N_i}$ to represent the N_i pressure measurements in cycle i ($i = 1,2,\dots,\mathcal{C}$), where $\mathcal{C}$ is the number of cycles from all beds and N_i varies from cycle to cycle. A subset of $\boldsymbol{x_i}$: $\boldsymbol{x}_{i,j} \in \Re^{N_{i,j}}$ represents the pressure measurements of step j ($j = 1,2,\dots,\mathcal{S}$) during cycle i; and $x_{i,j}(t)$ $\left(t = 1,2,\dots,N_{i,j}\right)$ represents an individual pressure measurement during step j of cycle i at time t; $N_{i,j}$ is the duration of step j of cycle i. Note that total sample number in cycle i: $N_i = \sum_{j=1}^{\mathcal{S}} N_{i,j}$ and total sample number across all $\mathcal{C}$ cycles: $N = \sum_{i=1}^{\mathcal{C}} N_i = \sum_{i=1}^{\mathcal{C}}\sum_{j=1}^{\mathcal{S}} N_{i,j}$. For each step of a given cycle, the definitions of different features are given below.

1. Mean ($\mu_{i,j}$), which captures the central tendency of pressure.

$$\mu_{i,j} = \frac{1}{N_{i,j}}\sum_{t=1}^{N_{i,j}} x_{i,j}(t) \quad (1)$$

2. Standard deviation ($s_{i,j}$), which measures the spread or dispersion of pressure measurements relative to its mean.

$$s_{i,j} = \sqrt{\frac{1}{N_{i,j}}\sum_{t=1}^{N_{i,j}}\left(x_{i,j}(t) - \mu_{i,j}\right)} \quad (2)$$

3. Skewness ($\gamma_{i,j}$), which measures the asymmetry of pressure distribution about its mean.

$$\gamma_{i,j} = \frac{\frac{1}{N_{i,j}}\sum_{t=1}^{N_{i,j}}\left(x_{i,j}(t) - \mu_{i,j}\right)^3}{s_{i,j}^3} \quad (3)$$

4. Kurtosis ($\kappa_{i,j}$), which measures whether the pressure distribution is heavy-tailed or light-tailed relative to a normal distribution. The following so-called excess kurtosis is used in this work so that the standard normal distribution has a kurtosis of zero.

$$\kappa_{i,j} = \frac{\frac{1}{N_{i,j}}\sum_{t=1}^{N_{i,j}}\left(x_{i,j}(t) - \mu_{i,j}\right)^4}{s_{i,j}^4} - 3 \quad (4)$$

5. Coefficient of variation ($CV_{i,j}$), a.k.a., relative standard deviation, is the standardized measure of dispersion of pressure distribution.

$$CV_{i,j} = \frac{s_{i,j}}{\mu_{i,j}} \quad (5)$$

6. Interquartile range ($IQR_{i,j}$), which also measures spread or dispersion of pressure measurements. $IQR_{i,j}$ is based upon, and relative to, the median, instead of the mean as in $s_{i,j}$. As a result, $IQR_{i,j}$ is less sensitive to extreme measurements or outliers.

$$IQR_{i,j} = Q3_{i,j} - Q1_{i,j} \quad (6)$$

where $Q1_{i,j}$ and $Q3_{i,j}$ are the lower quartile (i.e., 25-th percentile) and upper quartile (i.e., 75-th percentile) of all pressure measurements in step j of cycle i, respectively.

7. Quartile coefficient of dispersion ($QCD_{i,j}$), which measures dispersion of pressure measurements based on its interquartile range $IQR_{i,j}$.

$$QCD_{i,j} = \frac{Q3_{i,j} - Q1_{i,j}}{Q3_{i,j} + Q1_{i,j}} = \frac{IQR_{i,j}}{Q3_{i,j} + Q1_{i,j}} \quad (7)$$

8. Mean absolute deviation ($D_{mean,i,j}$), which measures dispersion of pressure measurements in terms of the absolute deviation from their mean.

$$D_{mean,i,j} = \frac{1}{N_{i,j}} \sum_{t=1}^{N_{i,j}} \left| x_{i,j}(t) - \mu_{i,j} \right| \quad (8)$$

9. Median absolute deviation ($D_{med,i,j}$), which measures dispersion of pressure measurements in terms of the absolute deviation from their median.

$$D_{med,i,j} = median(\left| x_{i,j}(t) - median(x_{i,j}) \right|) \quad (9)$$

10. Slope ($S_{i,j}$), which measures the overall slope of change in pressure measurements.

$$S_{i,j} = \frac{x_{i,j}(N_{i,j}) - x_{i,j}(1)}{N_{i,j} - 1} \quad (10)$$

11. Slope of linear regression line ($S_{LL,i,j}$), which measures the slope of pressure change when linearity is assumed. $S_{LL,i,j}$ is determined through simple linear regression.

- $$S_{LL,i,j} = \frac{N_{i,j} \cdot \sum_{t=1}^{N_{i,j}} \left(t \cdot x_{i,j}(t) \right) - \left(\sum_{t=1}^{N_{i,j}} t \right) \cdot \left(\sum_{t=1}^{N_{i,j}} x_{i,j}(t) \right)}{N_{i,j} \cdot \sum_{t=1}^{N_{i,j}} (t^2) - \left(\sum_{t=1}^{N_{i,j}} t \right)^2} \quad (11)$$

12. Mean absolute error ($MAE_{i,j}$), which measures the errors between the expected pressure measurements based on linear regression and the observed pressure measurements.

For the steps with relatively flat pressure profiles (e.g., adsorption, hold and purge steps), we first estimate the global mean of step j over all cycles under normal conditions (i.e., the training data).

$$\mu_{global,j} = \frac{1}{M} \sum_{j=1}^{M} \mu_{i,j} \quad (12)$$

where M is the total number of cycles in the training data. Then MAE can be calculated as the following

$$MAE_{i,j} = \frac{1}{N_{i,j}} \sum_{t=1}^{N_{i,j}} \left| x_{i,j}(t) - \mu_{global,j} \right| \quad (13)$$

For the steps with sloped pressure profiles (e.g., equalization, provide purge, blowdown or evacuation, and pressurization steps), the predicted pressure measurements, $\hat{x}_{i,j}(t)$ $(t = 1,2, \dots N_{i,j})$ are computed by linear regression using the linear model estimated from all training data.

$$MAE_{i,j} = \frac{1}{N_{i,j}} \sum_{t=1}^{N_{i,j}} \left| x_{i,j}(t) - \hat{x}_{i,j}(t) \right| \tag{14}$$

Finally, for each cycle, we have all the above-described features combined.

$$\boldsymbol{f}_i = \left\{ \boldsymbol{\mu}_i | \boldsymbol{s}_i | \boldsymbol{\gamma}_i | \boldsymbol{\kappa}_i | \boldsymbol{CV}_i | \boldsymbol{IRQ}_i | \boldsymbol{QCD}_i | \boldsymbol{D}_{mean,i} | \boldsymbol{D}_{med,i} | \boldsymbol{S}_i | \boldsymbol{S}_{LL,i} | \boldsymbol{MAE}_i \right\} \tag{15}$$

where $\boldsymbol{\mu}_i = \{\mu_{i,1}, \mu_{i,2}, \cdots, \mu_{i,T}\}$ is a row vector of dimension $(1 \times \mathcal{S})$. $\mathcal{S}$ is the total number of steps in a cycle. The same concatenation convention applies to all types of features. Therefore, $\boldsymbol{f}_i$ is a row vector of dimension $\left(1 \times \sum_{j=1}^{\mathcal{S}} Q_j\right)$ where Q_j is the number of features used to characterize step j of the PSA process. For simplicity, we use F to denote the total number of features included in each cycle so that $\boldsymbol{f}_i$ is a row vector containing F features. In this work, we utilize 12 different types of features, as defined in Eqs. (1)~(14), for all the 15 steps, which would result in 180 features for each cycle (*i.e.*, $F = 180$).

After the extracted features are concatenated into a row vector for each cycle following Eqn. (15), the features from multiple cycles are concatenated into a matrix as the following.

$$\boldsymbol{F}_{TR} = \begin{bmatrix} \boldsymbol{f}_1 \\ \boldsymbol{f}_2 \\ \vdots \\ \boldsymbol{f}_{\mathcal{C}} \end{bmatrix} \tag{16}$$

In this way, a training feature matrix $\boldsymbol{F}_{\mathrm{TR}}$ based on $\mathcal{C}$ normal cycles is obtained, which has a dimension of $(\mathcal{C} \times F)$, despite different step/cycle durations in the data. The features of test cycle(s) $\boldsymbol{F}_{TE}$ are extracted in the same way except that some of the features are generated with reference to the training cycles (e.g., $\mu_{global,j}$ in Eqn. (12)).

It is worth noting that the statistical and morphological features are extracted for each step of each cycle based on the raw training data without any preprocessing, without synchronization, scaling, normalization nor alignment. Since the features are calculated using all measurements from each step of the cycle, they are all scalars regardless of the step/cycle durations. Therefore, FSM naturally handles unequal step/cycle durations and asynchronous step/cycle trajectories. In addition, the structure shown in Eqn. (15) has the flexibility of allowing different number of features for different steps. In addition, cycle-based features can be conveniently added in a similar fashion.

4.2 Feature selection

Feature selection has been widely studied in supervised learning where it has been shown that including irrelevant and noisy features increases model complexity and can degrade model prediction performance (Andersen & Bro, 2010; Lee et al., 2020; Lindgren et al., 1994; Wang et al., 2015). Feature selection also has other benefits including reducing computational cost, improving interpretability of the model, etc. Since MSPM method based on PCA is a dimension reduction technique that is capable of handling collinearity in the process data, it may appear that feature selection is redundant and unnecessary. However, as shown in (Ghosh et al., 2014), feature selection

can have a significant impact on the monitoring performance of PCA-based MSPM. Specifically, experiments were conducted to show that including irrelevant and noisy features in a PCA-based MSPM model can degrade process monitoring performance (Ghosh et al., 2014).

In general, variable selection is more challenging for process monitoring as it is an unsupervised learning. Specifically, in process monitoring available data for model training are predominantly normal data. Even if fault data were available, they do not represent all possible fault scenarios. Therefore, for process monitoring, it is reasonable to assume that only normal operations data are available for feature selection, as features that are sensitive for detecting one type of fault may not be sensitive for detecting other (potentially unseen) faults. In this work we propose a new feature selection method for process monitoring that utilizes normal operation data only. We assume the true relevant features that are important for process monitoring should capture the key characteristics of the normal operation; consequently, if the true relevant features were used for process monitoring, the monitoring performance would be insensitive to the subsets of the training data used for model building. In other words, features extracted from a set of normal operation data (e.g., the training data) should show (highly) similar behavior as those extracted from another independent set of normal data (e.g., the validation data). In this work, we use false alarm rate (FAR) and false alarm magnitude (FAM) to quantify the difference between the training and validation performance, where FAM is defined as the difference between the monitoring statistic (e.g., T^2 or SPE) of the false alarm sample and the threshold of that statistic. In this work, 10-fold cross-validation is conducted using normal operation data to select features that result in similar FAR and FAM in the validation data, and feature selection is conducted through exhaustive search. A more systematic approach is under investigation. In the end, the following four features were selected: mean (μ), standard deviation (s), slope of linear regression line (S_{LL}) and mean absolute error (MAE). It is worth noting that all median or quartile based robust features were not selected in this work. The possible reason is that since the PSA process is tightly controlled, these robust (hence less sensitive) features do not offer advantage over mean-based features that are more sensitive to changes in FAR and FAM.

4.3 Model development

After feature selection, a multivariate statistical model can be developed to extract the patterns of normal cycles by examining the correlations among all features. This model enables the determination of a boundary or threshold for process monitoring. In this work, we assume that under normal operations, the features form a multivariate normal distribution. Since all features are the averages of some functions of multiple measurements in a step/cycle, their distributions are asymptotically Gaussian.(Pène, 2005) Similar to (He & Wang, 2011), here we choose PCA to capture the directions of maximum covariances among all the features. Other SPM methods, such as independent component analysis (ICA), can be applied as well.

Because the features in FSM are usually different types, it is reasonable to scale the training feature matrix $\boldsymbol{F}_{TR}$ to zero mean and unit variance for correlation based PCA as follows.

$$\boldsymbol{F}_{TR} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \widetilde{\boldsymbol{F}}_{TR} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \widetilde{\boldsymbol{T}}\widetilde{\boldsymbol{P}}^{\mathrm{T}} \tag{17}$$

where $\boldsymbol{P} \in \Re^{F\times P}$ is the loading matrix with its columns containing the directions of the first P (i.e., the number of principal components) maximum correlations among all features in descending order. $\boldsymbol{T} \in \Re^{M\times P}$ is the score matrix with its columns containing the projections of $\boldsymbol{F}_{TR}$ onto $\boldsymbol{P}$. $\widetilde{\boldsymbol{P}}$ and $\widetilde{\boldsymbol{T}}$ are the residual loading and score matrices, respectively. The principal component subspace (PCS) is $\mathcal{S}_P = span\{\boldsymbol{P}\}$, which captures the systematic variations of the normal process operation, including

measured and unmeasured disturbance, as well as set-point changes. The residual subspace (RS) is $S_R = span\{\tilde{\boldsymbol{P}}\}$, which captures the remaining variations after subtracting the systematic variations. They are the random variations of the process, including measurement noise under normal operations.

4.4 Fault detection

PCA based fault detection is well established for monitoring multivariate processes at steady state, where Hotelling's T^2 can be employed for monitoring variations in the PCS while SPE or Q statistic can be employed for monitoring variations in the RS. By monitoring features of individual cycles, we can straightforwardly extend PCA for monitoring cyclic processes which are inherently non-steady state. Similar to PCA-based monitoring of a continuous process, T^2 can capture faults that shift away from the normal operation region without violating the covariance among measured/monitored process variables. These faults are usually large operational changes such as a change of feedstock or raw material. On the other hand, SPE are sensitive to the process faults that violate the collinear relationships among the monitored features. The control limits of T^2 and SPE can be defined theoretically based on the Gaussian assumption of the features. They can also be determined empirically, *e.g.,* by kernel density estimation (KDE). The latter is used in this work.

By design, PSA processes are tightly controlled to operate in a targeted optimal region. Therefore, we expect there are few process changes that could violate the threshold in PCS and are detectable by T^2. In addition, as multiple features included in the FS are closely related to each other, such as the same features from different steps, there could be significant collinearities among features. Therefore, we expect SPE to be sensitive to the process faults with small magnitude but violating the collinearities among features, and enable early detection of potentially catastrophic faults. As shown in Section 5.2, both the traditional multi-way PCA (MPCA) and the proposed FSM detected the faults largely through SPE, as expected.

4.5 Fault diagnosis

Once a fault is detected by SPE statistic, the contribution plot can be used for fault diagnosis. In this work, we propose a hierarchical fault diagnosis using SPE to first determine in which step the fault occurred based on the step-wise contribution plot, then postulate what type of fault occurred based on the feature-wise contribution plot.

At the step level (*i.e.*, step-wise diagnosis), SPE statistic is broken down by step:

$$SPE = \sum_{j=1}^{T} SPE_j = \sum_{j=1}^{T} \left\| \tilde{\boldsymbol{f}}_j \right\|^2 \tag{18}$$

where $SPE_j = \left\| \tilde{\boldsymbol{f}}_j \right\|^2 = \sum_{k=1}^{Q_j} \tilde{f}_k^2$ is the contribution from the j^{th} step. $\tilde{f}_k^2$ is the residual (row) vector of features extracted from step j. The bar chart of $\frac{SPE_j}{SPE} \times 100\%$ $(j = 1 \cdots T)$ provides information for the step-wise diagnosis for the faulty cycle. The step(s) with the highest contribution(s) are identified as the root cause(s). Note that if different number of features are used for different steps, a normalization (*e.g.*, dividing by number of features) can be applied. Once the faulty step is identified, a feature-wise diagnosis is performed to identify the nature of the fault. The bar chart of $\frac{\tilde{f}_k^2}{SPE_j} \times 100\%$ $(k = 1, \cdots Q_j)$ provides information for the feature-wise diagnose for step j. For example, if mean contributed significantly to SPE_j, then more likely there were step change(s) in the

pressure profile during step j. Similar diagnosis can be made for other features. It is worth noting that the features proposed in this work specifically target the PSA process. For other processes, other features, including auto- and cross-correlation coefficients and higher-order statistics (HOS), could and have been utilized for quantifying process dynamics and nonlinearity (J. Wang & He, 2010), (He & Wang, 2011).

5 An Industrial Case Study

5.1 Fault description and methods studied

In this section, we use an industrial PSA case study to demonstrate the performance of the proposed FSM method, and compare it to the traditional MPCA method. Note that the method proposed by Wang et al. (2017) is not applicable because pressure measurements are the only available measurements for process monitoring in this study. The patented method by Arslan et al. (2014) cannot be implemented either, because there are no details as how the peaks are defined or classified as "relevant", and the criteria used for peak selection and control limits determination are unknown. The state space modeling approach by Pan et al. (2004) is not applicable due to the absence of cycle-to-cycle dynamics of the PSA processes and the lack of quality-relevant measurement for the PSA process studied in this work. Because MPCA requires that each step across all cycles has the same duration, two different data preprocessing techniques are studied: one with simple cut denoted as MPCA_{SC} and the other with dynamic time warping (DTW) denoted as MPCA_{DTW}. For MPCA_{SC}, the shortest step durations across all cycles are used as the reference while the last few measurements of any cycle with longer step duration are simply removed to match the shortest, which resulted in 438 variables for the whole cycle. For MPCA_{DTW}, the number of variables after unfolding is 705. The significant difference in the number of variables for MPCA_{SC} and MPCA_{DTW} reflects the significant variation of step durations across different cycles. For FSM, four features are used for each of the 15 steps, which resulted in 60 variables. To ensure that enough number of samples are available for all methods, 2,070 cycles under normal operations are used as the training set, which is about three times the number of variables for MPCA_{DTW}. The first 1,449 cycles (70% of 2,070 cycles) are used for model training and the remaining 621 cycles for model validation.

Six fault scenarios of a PSA process are studied in this work, which are listed in Table 1. The first four are simulated faults while the last two are from real industrial data. For the simulated faults, similar faulty process behaviours have been observed in actual operations. They are reproduced based on historical plots of those faults because the historical data are no longer available. All faults are related to valve malfunctions. For example, an internally leaky valve could result in higher or lower pressure in one vessel depending on whether the vessel serves as a pressure provider or receiver, such as the fault scenarios 1, 3 and 5. A sticky valve could result in higher pressure variation, such as the fault scenario 2, sudden pressure increase or drop, such as the fault scenario 4, or a non-smooth (e.g., zig-zag) pressure profile, such as the fault scenario 6. As discussed in the Introduction section, due to the very high frequency of open and close operations of valves, it has been found that the process faults are most often caused by valve-related problems.

For each fault scenario, totally 16 cycles are used as the test set and among which 3 (deliberately arranged as cycle 4, 9 and 14 for better comparison across all scenarios) are faulty cycles. In these cases, the simulated faults are introduced by modifying a normal cycle randomly selected from the industrial data set. The pressure trajectories of the test set for fault scenarios 1 (simulated fault) and 5 (real fault) are shown in Figure 4.

[Table 1 here]

[Figure 4 here]

For all fault detection methods, the number of principal component (PCs) is selected to cover 90% of the variance of their corresponding full feature space. The control limits on Hoteling's T^2 and squared prediction error (SPE) are calculated empirically based on the kernel density estimation of the corresponding statistics (*i.e.*, T^2 or SPE) of the training dataset at confidence level 99%. The number of PCs and other information discussed above are listed in Table 2.

[Table 2 here]

5.2 Fault detection results

By considering faults detected in both residual subspace using SPE and principal subspace using T^2, the overall fault detection results are shown in Table 3. Specifically, the table lists faulty cycles detected by either SPE, or T^2, or both. The fault detection rate (FDR) and false alarm rate (FAR) of each method are also summarized in Table 3. These results show that FSM detects all faulty cycles under all fault scenarios without generating false alarms. In comparison, $MPCA_{SC}$ has missed detection under fault scenario 1, while $MPCA_{DTW}$ has missed detection under fault scenario 5. In addition, both $MPCA_{SC}$ and $MPCA_{DTW}$ have false alarms. More details of the fault detection results by T^2 and SPE are shown in Table 4. It can be seen that SPE statistic in general is more effective in detecting faults, although it also generates false alarms in the cases of $MPCA_{SC}$ and $MPCA_{DTW}$. In comparison, T^2 does not generate false alarms for all methods. However, it misses several faults in $MPCA_{SC}$ and performs even worse in $MPCA_{DTW}$. Overall, FSM performs robustly with both T^2 and SPE and is significantly better than $MPCA_{SC}$ and $MPCA_{DTW}$.

To provide additional details on the performance of different methods, we visualized some fault detection results. Due to limited space, only the detection results of fault scenarios 1 and 5 in the residual subspace (i.e., using SPE statistic) are visualized in Figure 5 and Figure 6 and discussed in detail below. Figure 5 shows that for fault scenario 1, $MPCA_{SC}$ has difficulty in detecting Fault 1: missing two out of three faulty cycles. $MPCA_{DCW}$ detects all three faulty cycles but also generates a false alarm. Only FSM detects all three faulty cycles without generating false alarms. Figure 6 shows that for fault scenario 5, $MPCA_{SC}$ detects all three faulty cycles while generating a false alarm. $MPCA_{DCW}$ failed to detect two out of three faulty cycles while generating a false alarm. Again, only FSM successfully detects all faulty cycles without generating false alarms. It is worth noting that FSM results in linear models, which have low risk of overfitting. This is demonstrated in Fig. 5 (C) and Fig. 6 (C) where the normal test samples have similar SPE values as those of the normal training samples.

[Figure 5 here]

[Figure 6 here]

We further investigated the false alarms (*i.e.*, cycle 10 for $MPCA_{SC}$ and cycle 15 for $MPCA_{DTW}$) to understand why those two cycles generate false alarms. Figure 7 plots the trajectory of all test cycles for the fault scenario 5. All true faulty cycles are plotted in red dashed lines, and all normal test cycles are plotted in black solid lines – except cycle 10 in the cyan dash-dotted line, and cycle 15 in the green dotted line. As described in Table 1, the pressure profiles of the faulty cycles do not follow the normal cycle pressure trajectory during the re-pressurization step (highlighted in the blue dashed-line box). It can be seen that cycles 10 and 15 both behave normally during the re-pressurization step. However, if we zoom in to visualize the pressure profile in other steps (e.g., the insert in Figure 7,

522 which is the zoom-in view of the adsorption step), it can be seen that cycles 10 and 15 are at the
523 lower boundary of all normal cycles, suggesting that the MPCA based approaches may be more
524 sensitive to mean shift. Although the mean shift in this case is within the normal operation range, the
525 cumulative effect (*i.e.*, the persistent small shift that lasts for a period of time) would likely be
526 captured by MPCA as a fault.

527 [Figure 7 here]

528 The normal samples (i.e., cycles) in the test data are industrial data collected from the same PSA
529 process and used for all the fault scenarios. For each faulty cycle, the fault occurred during different
530 steps for different fault scenarios as defined in Table 1. As a result, the false alarms for a specific
531 method across different fault scenarios are the same (e.g., cycle 15 for $MPCA_{SC}$ and cycle 10 for
532 $MPCA_{DTW}$) since they are the same normal cycles used in all fault scenarios. Clearly, the value of
533 FAR in Table 1 would be affected by the normal cycles included in the test data. In this study, the
534 normal testing samples were selected randomly to avoid any potential bias.

535 Further investigation is conducted to understand the reason for $MPCA_{DCW}$'s failure in detecting true
536 faulty cycles under fault scenario 5 (a real fault), for which only one out of three faulty cycles is
537 detected. Since $MPCA_{SC}$ is able to detect all faulty cycles, we suspect that the failure is related to
538 data preprocessing by DTW. Therefore, we plot the original pressure profiles of the 16 test cycles,
539 which are shown in Figure 8 (A) and compare them to the pressure profiles after DTW as shown in
540 Figure 8 (B). A zoom-in view of the faulty step is included for both figures. The comparison clearly
541 indicates that the irregular discrepancies of the faulty cycles shown in the original pressure profiles
542 diminished after trajectory synchronization by DTW. This case suggests that DTW could cause
543 severe information loss or distortion, which in turns affects the fault detection performance. This is
544 consistent with our previous findings that data manipulations during preprocessing, including DTW,
545 could cause information loss or distortion and should be avoided if possible (He & Wang, 2011).
546 This example further raises the alarm that the widely used DTW for batch trajectory warping or
547 alignment in process monitoring applications could potentially be a problematic practice that may
548 lead to missed detections of process faults.

549 [Figure 8 here]

550 **5.3 Fault diagnosis results**

551 After fault detection, fault diagnosis is performed for the detected faulty cycles. For FSM, the
552 procedure outlined in Sec. 4 is followed to construct step-wise and feature-wise contribution plots.
553 For $MPCA_{SC}$ and $MPCA_{DTW}$, since pressure is the only measured process variable, only step-wise
554 contribution plot is applicable. Again, we use fault scenarios 1 and 5 as examples. For fault scenario
555 1, test cycle 4 is used for illustration as the fault is detected by all methods. For the same reason, test
556 cycle 9 is used for fault scenario 5. Figure 9 (A)-(C) show the step-wise contribution plots for fault
557 diagnosis of test cycle 4 in fault scenario 1. From Table 1, we know that this is a fault occurring
558 during the adsorption step (*i.e.*, step 1), and the faulty cycles have lower pressure than normal cycles.
559 Figure 9 shows that $MPCA_{SC}$ incorrectly attributes this fault to step 10, while $MPCA_{DTW}$ and FSM
560 correctly identify the faulty step, with FSM providing the strongest conviction. For FSM, once the
561 faulty step is identified, the faulty step contribution is further broken down to the feature level
562 following the procedure outlined in Sec. 4, as shown in Figure 9(D). In this figure, MAE and the
563 mean (μ) are identified as the most significant contributor to the fault, indicating that there might be a
564 mean shift in pressure profile during step 1 of the faulty cycles. It is worth noting that MAE usually
565 contributes the most to the fault detection indices as it captures the absolute deviation of a faulty

trajectory from the nominal behavior, which includes the effects of both mean and spread (e.g., standard deviation) of the samples.

[Figure 9 here]

The fault diagnosis for cycle 9 of fault scenario 5 is conducted similarly and the results are shown in Figure 10. Here the true fault is a deviation of the pressure profile from the normal trajectories during the last step of the cycle. In this care, both $MPCA_{SC}$ and FSM correctly identify the faulty step, while $MPCA_{DTW}$ misidentifies the step 10 as the faulty step. FSM also shows the clearest diagnosis among all methods. Feature-wise diagnosis from FSM identifies MAE and the standard deviation (s) as the most significant contributors to the fault, indicating that there could be an increased variation in pressure profile during the last step of the faulty cycle. The misdiagnosis of this fault by $MPCA_{DTW}$ can be at least partially attributed to the pressure profile distortion during the DTW preprocessing step as discussed earlier and illustrated in Figure 8.

[Figure 10 here]

The overall fault diagnosis results are shown in Table 5. It can be seen that FSM correctly diagnoses all fault scenarios (*i.e.*, correctly identifies all faulty steps). $MPCA_{SC}$ has two misdiagnoses (fault scenarios 3 and 6) and one inconsistent diagnosis from T^2 and SPE (fault scenario 1). $MPCA_{DTW}$ has two misdiagnoses (fault scenarios 5 and 6). In addition, FSM provides meaningful diagnoses at the feature level for all fault scenarios.

[Table 5 here]

6 Conclusions and Discussions

In this work, we present a simple yet effective fault detection and diagnosis method, namely feature space monitoring or FSM, for PSA and other cyclic/periodic processes. Different from the conventional MSPM methods, FSM characterizes the normal operation cycle behavior with various statistical and shape/morphological features for each step. FSM naturally handles the challenges in monitoring cyclic processes without any preprocessing steps, which include variable cycle/step duration, wide range of normal cycle/step trajectories, and/or limited measurements – for PSA, bed pressure is often the only measured variable for process monitoring. To be able to detect subtle faults from a wide range of normal cycle trajectories, FSM relies on feature engineering and selection to balance the robustness and sensitivity of the fault detection performance. Finally, through a hierarchical fault diagnosis framework, once a fault is detected, the proposed FSM approach first identifies the step where the fault occurred using a step-wise contribution plot, then postulates the type of fault based on a feature-wise contribution plot.

Using an industrial case study, we demonstrate that FSM outperforms MPCA with simple cut ($MPCA_{SC}$) or dynamic time warping ($MPCA_{DTW}$) in six fault scenarios. Specifically, FSM successfully detected all three faulty cycles in every fault scenario without generating false alarms. In comparison, both $MPCA_{SC}$ and $MPCA_{DTW}$ had missed detections in some fault scenarios, and both had false alarms in all fault scenarios. In addition, the hierarchical fault diagnosis framework based on FSM correctly identified the faulty steps under all scenarios studied in this work. In comparison, $MPCA_{SC}$ and $MPCA_{DTW}$ based contribution plots all had misdiagnosis under some fault scenarios. Finally, FSM provides the feature-wise diagnosis capability, which enables a plant engineer to further determine the nature of a fault, such as whether it is a simple mean shift or an increase in

variation or a complex fault of both. These fault detection and diagnosis results demonstrate that the FSM-based linear models have a low risk of overfitting and are easy to interpret.

The proposed FSM framework can be applied to other periodic or cyclic processes. In terms of implementation, what FSM features to be included has a big impact on the fault detection and diagnosis performance. In general, the features that should be included for process monitoring depend on the process behavior, and the domain knowledge plays an important role. In addition, depending on the noise level of the process data, it may be necessary to evaluate different versions of the same features that have different degrees of sensitivity to extreme points or outliers. There is usually a trade-off between robustness and sensitivity of the monitoring performance. In this work, the list of features we evaluated were generated based on PSA process behavior and discussions with process engineers. In addition, we proposed an approach to select the relevant features using normal process data only. The variable selection for unsupervised learning was achieved based on the assumption that the truly relevant features should provide consistent monitoring performance, regardless of the training data used. In this work, feature selection was done through a manual search. It is desirable to have a systematic and automated approach for feature engineering and feature selection, which is the area that we are currently working on.

Finally, this work suggested that the widely used DTW for batch trajectory warping or alignment in process monitoring applications could potentially be a problematic practice that may lead to missed fault detections. Specifically, this work demonstrated that DTW could cause severe batch/cycle trajectory distortion, which in turns negatively affected the fault detection performance. Therefore, data preprocessing in process monitoring, including DTW, should be avoided if possible, or conducted with caution and verification if unavoidable.

7 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

8 Author Contributions

QH and JW conceived the idea and secured the funding. JL implemented and improved the methodology, conducted data analysis, generated majority of the results, and participated in the manuscript revision; QH supervised and conducted data analysis; QH and JW wrote the initial draft and revised the manuscript; AK and JFC provided PSA process knowledge and the process data, critiqued the methodology, reviewed the results, and participated in the manuscript revision. All authors read and approved the final manuscript.

9 Funding

This research is supported by the US National Science Foundation under grant CBET #1805950.

References:

Andersen, C. M., & Bro, R. (2010). Variable selection in regression---a tutorial. *Journal of Chemometrics*, *24*(11–12), 728–737.

Arslan, E., Neogi, D., Li, X. J., & Misra, P. (2014). *Apparatus and methods to monitor and control cyclic process units in a steady plant environment*. Google Patents.

646 Choi, S. W., Lee, C., Lee, J. M., Park, J. H., & Lee, I. B. (2005). Fault detection and identification of
647 nonlinear processes based on kernel PCA. *Chemometrics and Intelligent Laboratory Systems*,
648 *75*(1), 55–67. https://doi.org/10.1016/j.chemolab.2004.05.001

649 Dedecker, J., & Rio, E. (2008). On mean central limit theorems for stationary sequences. *Annales de*
650 *l'IHP Probabilités et Statistiques*, *44*(4), 693–726.

651 Elseviers, W., Hassett, P. F., Navarre, J.-L., & Whysall, M. (2015). *50 Years of PSA Technology for*
652 *H2 Purification*.

653 Georgiadis, M. C., & Papageorgiou, L. G. (2000). Optimal energy and cleaning management in heat
654 exchanger networks under fouling. *Chemical Engineering Research and Design*, *78*(2), 168–
655 179.

656 Ghosh, K., Ramteke, M., & Srinivasan, R. (2014). Optimal variable selection for effective statistical
657 process monitoring. *Computers & Chemical Engineering*, *60*, 260–276.

658 He, F., & Xu, J. (2016). A novel process monitoring and fault detection approach based on statistics
659 locality preserving projections. *Journal of Process Control*, *37*, 46–57.

660 He, Q. P., & Wang, J. (2007). Fault detection using the k-nearest neighbor rule for semiconductor
661 manufacturing processes. *Semiconductor Manufacturing, IEEE Transactions On*, *20*(4), 345–
662 354.

663 He, Q. P., & Wang, J. (2011). Statistics pattern analysis: A new process monitoring framework and
664 its application to semiconductor batch processes. *AIChE Journal*, *57*(1), 107–121.
665 https://doi.org/10.1002/aic.12247

666 He, Q. P., & Wang, J. (2018). Statistics Pattern Analysis: A Statistical Process Monitoring Tool for
667 Smart Manufacturing. *Computer Aided Chemical Engineering*, *44*, 2071–2076.
668 https://doi.org/10.1016/B978-0-444-64241-7.50340-2

669 Hong, H., Jiang, C., Peng, X., & Zhong, W. (2020). Concurrent monitoring strategy for static and
670 dynamic deviations based on selective ensemble learning using slow feature analysis. *Industrial*
671 *& Engineering Chemistry Research*, *59*(10), 4620–4635.

672 Jain, V., & Grossmann, I. E. (1998). Cyclic scheduling of continuous parallel-process units with
673 decaying performance. *AIChE Journal*, *44*(7), 1623–1636.

674 Lee, J., Flores-Cerrillo, J., Wang, J., & He, Q. P. (2020). Consistency-Enhanced Evolution for
675 Variable Selection Can Identify Key Chemical Information from Spectroscopic Data. *Industrial*
676 *& Engineering Chemistry Research*, *59*(8), 3446–3457.

677 Lindgren, F., Geladi, P., Rännar, S., & Wold, S. (1994). Interactive variable selection (IVS) for PLS.
678 Part 1: Theory and algorithms. *Journal of Chemometrics*, *8*(5), 349–363.

679 Pan, Y., Yoo, C., Lee, J. H., & Lee, I.-B. (2004). Process monitoring for continuous process with
680 periodic characteristics. *Journal of Chemometrics*, *18*(2), 69–75.

Pène, F. (2005). Rate of convergence in the multidimensional central limit theorem for stationary processes. Application to the Knudsen gas and to the Sinai billiard. *The Annals of Applied Probability*, *15*(4), 2331–2392.

Shang, C., Yang, F., Gao, X., Huang, X., Suykens, J. A. K., & Huang, D. (2015). Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis. *AIChE Journal*, *61*(11), 3666–3682.

Wang, J., & He, Q. P. (2010). Multivariate Statistical Process Monitoring Based on Statistics Pattern Analysis. *Industrial & Engineering Chemistry Research*, *49*(17), 7858–7869. https://doi.org/10.1021/ie901911p

Wang, K., Chang, P., & Meng, F. (2021). Monitoring of wastewater treatment process based on slow feature analysis variational autoencoder. *2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS)*, 495–502.

Wang, R., Edgar, T. F., & Baldea, M. (2017). A geometric framework for monitoring and fault detection for periodic processes. *AIChE Journal*, *63*(7), 2719–2730.

Wang, Z., He, Q. P., & Wang, J. (2015). Comparison of variable selection methods for PLS-based soft sensor modeling. *Journal of Process Control*, *26*(2015), 56–72. https://doi.org/10.1016/j.jprocont.2015.01.003

Wise, B. M., Gallagher, N. B., Butler, S. W., White, D. D., & Barna, G. G. (1999). A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process. *Journal of Chemometrics*, *13*(3–4), 379–396.

Yang, J., Lv, Z., Shi, H., & Tan, S. (2018). Performance monitoring method based on balanced partial least square and Statistics Pattern Analysis. *ISA Transactions*, *81*, 121–131.

Yoo, C. K., Lee, J.-M., Vanrolleghem, P. A., & Lee, I.-B. (2004). On-line monitoring of batch processes using multiway independent component analysis. *Chemometrics and Intelligent Laboratory Systems*, *71*(2), 151–163.

Zhang, H., Tian, X., Deng, X., & Cao, Y. (2018). Multiphase batch process with transitions monitoring based on global preserving statistics slow feature analysis. *Neurocomputing*, *293*, 64–86.

Zhao, C., & Huang, B. (2018). A full-condition monitoring method for nonstationary dynamic chemical processes with cointegration and slow feature analysis. *AIChE Journal*, *64*(5), 1662–1681.

Zhou, B., & Gu, X. (2019). Multi-block statistics local kernel principal component analysis algorithm and its application in nonlinear process fault detection. *Neurocomputing*.

717

718 **Tables**

Table 1. Fault scenarios studied in this work

Fault #	Description
1	During adsorption step, the faulty cycles have lower pressure than normal cycles.
2	During adsorption step, the faulty cycles have higher pressure variations than normal cycles
3	During a hold step, the pressure of the faulty cycles decreases instead of being held steady
4	During an equalization step, the pressure of the faulty cycles was held steady followed by a sudden drop instead of smooth decrease
5	During re-pressurization, the pressure of the faulty cycles does not follow the normal cycle trajectory
6	During an equalization step, the pressure of the faulty cycles follows a zig-zag or stair-like profile instead of a smooth increase

719

720

Table 2. Training, testing datasets and model parameters

	$MPCA_{SC}$	$MPCA_{DTW}$	FSM
# of features/variables	438	705	60
# of PC's	24	32	20
Training	2,070 normal cycles		
Testing	16 cycles (13 normal, 3 fault: cycle 4, 9, 16)		
Confidence level	99%		

721

722

Table 3. Fault detection results (true faulty cycles: 4, 9 & 14; FDR: fault detection rate; FAR: false alarm rate). Correctly detected cycles are in bold black. Incorrectly detected cycles are in red.

Fault Scenario	$MPCA_{SC}$			$MPCA_{DTW}$			FSM		
	Cycle detected	FDR	FAR	Cycle detected	FDR	FAR	Cycle detected	FDR	FAR
1	**4, 14,** 15	66.7%	7.7%	**4, 9,** 10**, 14**	100%	7.7%	**4, 9, 14**	100%	0%
2	**4, 9, 14,** 15	100%	7.7%	**4, 9,** 10**, 14**	100%	7.7%	**4, 9, 14**	100%	0%
3	**4, 9, 14,** 15	100%	7.7%	**4, 9,** 10**, 14**	100%	7.7%	**4, 9, 14**	100%	0%
4	**4, 9, 14,** 15	100%	7.7%	**4, 9,** 10**, 14**	100%	7.7%	**4, 9, 14**	100%	0%
5	**4, 9, 14,** 15	100%	7.7%	**9,** 10	33.3%	7.7%	**4, 9, 14**	100%	0%
6	**4, 9, 14,** 15	100%	7.7%	**4, 9,** 10**, 14**	100%	7.7%	**4, 9, 14**	100%	0%

723

Table 4. Details of the fault detection results by T^2 and SPE. Correctly detected cycles are in bold black. Incorrectly detected cycles are in red.

Fault Scenario	$MPCA_{SC}$		$MPCA_{DTW}$		FSM	
	T^2	SPE	T^2	SPE	T^2	SPE
1	**4, 14**	**4,** 15	**4, 9, 14**	**4, 9,** 10**, 14**	**4, 9, 14**	**4, 9, 14**
2	**9**	**4, 9, 14,** 15	None	**4, 9,** 10**, 14**	**4, 9, 14**	**4, 9, 14**
3	**9, 14**	**4, 9, 14,** 15	None	**4, 9,** 10**, 14**	**9, 14**	**4, 9, 14**
4	**4, 9, 14**	**4, 9, 14,** 15	None	**4, 9,** 10**, 14**	**4, 9, 14**	**4, 9, 14**
5	**4, 9, 14**	**4, 9, 14,** 15	None	**9,** 10	**4, 9, 14**	**4, 9, 14**
6	**9**	**4, 9, 14,** 15	None	**4, 9,** 10**, 14**	**9, 14**	**4, 9, 14**

Table 5. Details of the fault diagnosis results by T^2 and SPE. One faulty cycle detected by all methods is selected for each fault scenario. If the faulty step is correctly identified, it is marked as "Yes", otherwise marked as "No". If the faulty cycle is not detected as a fault, its diagnosis is marked as "NA".

Fault Scenario (cycle for diagnosis)	$MPCA_{SC}$		$MPCA_{DTW}$		FSM	
	T^2	SPE	T^2	SPE	T^2	SPE
1 (4)	**Yes**	No	**Yes**	**Yes**	**Yes**	**Yes**
2 (9)	**Yes**	**Yes**	NA	**Yes**	**Yes**	**Yes**
3 (9)	No	No	NA	**Yes**	**Yes**	**Yes**
4 (4)	**Yes**	**Yes**	NA	**Yes**	**Yes**	**Yes**
5 (9)	**Yes**	**Yes**	NA	No	**Yes**	**Yes**
6 (9)	No	No	NA	No	**Yes**	**Yes**

Figure Captions

Figure 1: Schematic illustration of the major steps involved in a PSA process: (A) Adsorption; (B) Regeneration

Figure 2: Visualization of pressure trajectories in a multi-bed PSA process: (A) A sample time-series pressure profiles of three beds in a multi-bed PSA process; (B) Overlaid pressure profiles of a single bed over multiple cycles, which illustrates the variable durations from different cycles

Figure 3: Illustrations of the variable cycle and step durations: (A) Cycle durations vary significantly from cycle to cycle; (B) Durations of the adsorption step follow a similar trend as the cycle durations; (C) durations of the equalization steps are close to constant

Figure 4: Pressure trajectories of the selected test sets: (A) Fault scenario 1; (B) Fault scenario 5. Normal cycles are plotted in black while faulty cycles are in red. The steps in which the fault occurred are marked by blue dashed-line rectangles and shown in the zoom-in views.

Figure 5: Fault scenario 1: fault detection in residual subspace (SPE) from (A) $MPCA_{SC}$, (B) $MPCA_{DTW}$ and (C) FSM. Blue circles are normal training cycles; black squares are normal testing cycles; black triangles are faulty testing cycles; horizontal red dashed lines are fault detection thresholds; vertical black dash-dotted lines separate training from testing.

Figure 6: Fault scenario 5: fault detection in residual subspace (SPE) from (A) $MPCA_{SC}$, (B) $MPCA_{DTW}$ and (C) FSM. Blue circles are normal training cycles; black squares are normal testing cycles; black triangles are faulty testing cycles; horizontal red dashed lines are fault detection thresholds; vertical black dash-dotted lines separate training from testing.

Figure 7: The false alarm cycles 10 (in cyan) and 15 (in green) behave normally during the re-pressurization step in which the true fault occurred (highlighted in the blue dashed-line rectangle with faulty cycles plotted as red dashed lines). The zoom-in view of the adsorption step indicates that cycles 10 and 15 are at the lower, but normal, boundary of all cycles.

Figure 8: Comparison between the pressure profiles of (A) the original 16 test cycles and (B) the test cycles after DTW. The normal testing cycles are plotted in black solid lines while the faulty testing cycles are plotted in red dashed lined. The irregular discrepancies among cycles shown in the original profiles (highlighted in the blue dashed-line rectangles and the zoom-in views) have diminished after DTW, indicating that DTW causes significant information loss or distortion.

Figure 9: Step-wise fault diagnosis for cycle 4 of fault scenario 1: (A) $MPCA_{SC}$, (B) $MPCA_{DTW}$, and (C) FSM, and (D) feature-wise fault diagnosis from FSM. $MPCA_{SC}$ wrongly attributes the fault to step 10, while both $MPCA_{DTW}$ and FSM correctly attribute the fault to step 1. FSM also shows the clearest diagnosis among all methods. FSM can also further drill down the fault to the feature level where the mean (μ) and MAE are identified as the most significant contributor to the fault, correctly indicating that there might be a significant shift in pressure profile during step 1 of the faulty cycles.

Figure 10: Step-wise fault diagnosis for cycle 9 of fault scenario 5: (A) $MPCA_{SC}$, (B) $MPCA_{DTW}$, and (C) FSM, and (D) feature-wise fault diagnosis from FSM. Both $MPCA_{SC}$ and FSM correctly attribute the fault to the last step of the cycle, while $MPCA_{DTW}$ wrongly attributes the fault to step 10. FSM also shows the clearest diagnosis among all methods. FSM can also further drill down the fault to the feature level where the standard deviation and MAE are identified as the most significant contributor

768 to the fault, correctly indicating that there is an increased variation in pressure profile during the last
769 step of the faulty cycles.