Unaligned Profile Monitoring using Penalized Methods

Yangyang Zang[1], Kaibo Wang[2,*] and Ran Jin[3]

[1] *Department of Industrial Engineering, Tsinghua University, Beijing 100084, China*

E-mail: zangyy12@mails.tsinghua.edu.cn

[2] *Department of Industrial Engineering, Tsinghua University, Beijing 100084, China*

E-mail: kbwang@tsinghua.edu.cn.

[3] *Grado Department of Industrial and Systems Engineering, Virginia Tech., VA 24061, USA*

E-mail: jran5@vt.edu

In some manufacturing processes, complex profiles are collected to characterize quality status. However, some of these profiles may have unequal lengths, which makes the attempt of directly comparing them difficult. In addition, when a shift occurs in a profile, it usually affects a segment of continuously connected observations. That is, local shifts instead of global shifts are frequently seen. As shift signals are easily mixed with allowable mean trends, statistical monitoring of such unaligned profiles becomes a challenging task. In this paper, we propose a framework for monitoring profiles with unequal lengths. The profiles are first aligned using a modified robust dynamic time warping (DTW) algorithm, which is insensitive to local mean shifts. Penalization-based methods are then used to estimate profile means. Finally, mean estimates are utilized in a likelihood ratio test statistic for effective monitoring. Both simulation studies and a real example are used to demonstrate the effectiveness of the proposed monitoring procedure.

Key words: profile monitoring; dynamic time warping; penalized estimation; Fused LASSO

1 Introduction

Statistical process control (SPC) has been widely used in many industrial processes. SPC aims to improve process and product quality by reducing process variability. Control charts are one of the most important SPC tools for assignable cause identification and variation reduction. In recent years, we have seen many processes having quality characterized by complicated profiles. A profile defines a functional relationship between a response variable (Y) and one or more explanatory variables (Xs). Over the past decades, as the use of sensing technology expands, profile monitoring has become more prevalent due to the large amounts of data available in diverse manufacturing systems and service areas.

Much work has been done when profiles can be characterized by parametric regression models. Such parametric models, either linear or nonlinear, are then monitored by multivariate T^2 control charts,[1] self-

starting control charts,[2] multivariate EWMA schemes,[3] change point approaches,[4,5] *etc*. Abnormal shifts in profiles are expected to lead to model changes and be detected by such charts.

In certain applications, a profile cannot be well fitted by any parametric forms.[6] In such cases, nonparametric monitoring methods are developed. Chang and Yadama[7] used B-spline approximations to monitor non-linear profiles. The authors first applied a wavelet transformation to separate true signals from noises, after which they formulated transformed profiles using a B-spline model; a T^2 control chart was then used to identify profile mean shifts or shape changes. Shiau *et al.*[8] and Paynabar *et al.*[9] utilized principal component analysis (PCA) projection to monitor profiles, and De Ketelaere *et al.*[10] gave a detailed overview of PCA-based monitoring approaches. Jeong *et al.*[11] used wavelet transformation to handle complicated profiles first before monitoring.

Currently, most widely used parametric and nonparametric profile monitoring approaches share certain common assumptions. First, all profiles to be monitored have an equal length, that is, vectors that characterize the profiles have an equal dimension. Second, sampling points of the explanatory variable are fixed and consistent across all profiles. Although these assumptions are commonly accepted in profile monitoring, they are violated in some manufacturing processes. In certain applications, profiles collected from production cycles do not have an equal length, but only share a common variation pattern, nor do the explanatory variables have fixed sampling points.

This paper intends to develop algorithms for monitoring unaligned profiles with varying sampling points. The main contribution of this paper is twofold. First, we propose a robust dynamic time warping (robust DTW) algorithm for profile alignment. The robust DTW algorithm considers SPC-specific challenges, that is, profiles are contaminated by noises and shift signals, but can still yield robust alignment results. Second, we propose a penalization-based charting algorithm that gives more effective performance in shift detect.

The rest of this paper is organized as follows. In Section 2, we present an illustrative example to demonstrate the real needs and importance of monitoring unaligned profiles. In Section 3, a framework for unaligned profile monitoring is presented. Then, the robust DTW algorithm for profile alignment is presented in Section 4, and a new profile monitoring algorithm is introduced in Section 5. In Section 6, a real example is shown to demonstrate the use of the monitoring procedure. Finally, we conclude this work with suggestions for future research in Section 7.

2 An illustrative example

To better illustrate the importance of the profile alignment and monitoring problem, we take an ingot growth process as an example. In an ingot growth process, a collection of procedures are carefully designed to convert polycrystalline silicon into monocrystalline silicon. To maintain a constant ingot

diameter and uniform material quality, the heating power must be perfectly coordinated to ensure a desirable growth environment. Because of the complex growing mechanism, the heating power exhibits a dynamic trend in each growth cycle. Figure 1(a) shows four sample power profiles collected from different growth cycles. All the profiles share a similar trend, which is governed by the physical mechanism behind the growth process. In the initial pulling stage, the heating power has a relatively large variation; the variation of heating power gradually stabilizes, while its mean increases slowly. Another noticeable feature of the profiles is that they have different lengths. In practice, the length of a profile is determined by the amount of polycrystalline silicon deposited for that production cycle. Figure 1(b) shows aligned heating power profiles. The lengths of all profiles are unevenly stretched or compressed to the same span by an alignment algorithm that will be introduced in details in a later section.

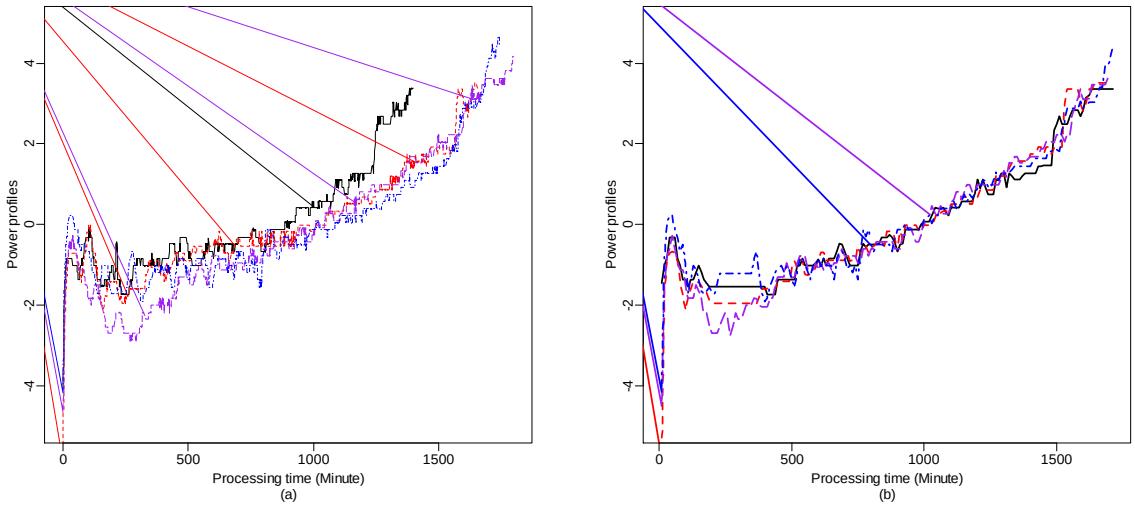


Figure 1. Heating power profiles collected from the ingot growth process: (a) raw profiles and (b) aligned profiles

In the ingot growth process, the difference in profile length is not an indication of process deterioration or failures. Instead, abnormal process changes are likely led to shocks or segments of mean shifts in profiles. To detect such shifts, we have to shrink or expand them to an equal length, with an equal number of sampling points. Profile alignment, also known as curve registration[12] in biological and medical studies, are not yet widely studied in profile monitoring. Some existing discussion on profile alignment focused on curves that are expressed in appropriate functional forms. For example, Mosesova *et al.*[13] used simple landmark registration to obtain aligned profiles for monitoring. Dai *et al.*[14] studied the monitoring of unaligned discrete profiles, and they used the conventional dynamic time warping (DTW) algorithm to obtain warped profiles for monitoring. However, as the existing DTW algorithm is not

designed for SPC purpose, important failure signatures such as local shifts may mislead the alignment operation; SPC based on wrongly aligned profiles can hardly give satisfactory performance.

3 A framework for unaligned profile monitoring

Assume the jth profile with length n_j is collected at time j, which has observation pairs $(\mathbf{x}_j, \mathbf{y}_j)$, where $\mathbf{x}_j$ and $\mathbf{y}_j$ are vectors of n_j dimensions that represent the explanatory and response variables, respectively. To facilitate discussion, we formulate the unaligned profiles using the following regression model:

$$y_{ij} = g\left(f\left(x_{ij}\right)\right) + \varepsilon_{ij}, \quad i = 1, 2, \cdots, n_j, \quad j = 1, 2, \cdots \tag{1}$$

where x_{ij} and y_{ij} denote the values of the ith element of $\mathbf{x}_j$ and $\mathbf{y}_j$ on the jth profile. Under the assumption of continuity of a profile, $x_{1j} < x_{2j} < \cdots < x_{n_j j}$. In most applications, x_{ij} are evenly distributed; while in some other cases, like power profiles in the ingot growth process, the intervals of the explanatory variables are unequally distributed. Additionally, sampling positions of different profiles may be different, which means that x_{ij} for profile j may not equal x_{il} in profile $l, l \neq j$.

In the model, ε_{ij}'s represent measurement errors, which are assumed to be independent and follow a normal distribution with mean 0 and variance σ_j^2. $g(\cdot)$ is a general function that denotes the relationship between the explanatory variables and the response variable. $f(\cdot)$ is a warping function that maps the explanatory variables of an unaligned profile to another scale so that the profiles become aligned. We set profile 0, $(\mathbf{x}_0, \mathbf{y}_0)$ with length n_0, as the reference profile to which other profiles should be aligned. Then, the warping function in Equation (1) is in fact $f(x_{ij}) = x_{k0}, k = 1, 2, \cdots, n_0$. Aligning profile $(\mathbf{x}_j, \mathbf{y}_j)$ to a reference profile means finding a series of explanatory variable values in the reference profile and linking them with their counterparts in $\mathbf{x}_j$. Details on the warping function will be introduced in the following section.

To design a proper monitoring procedure for the regression model in Equation (1), profile alignment must first be accomplished by estimating an appropriate warping function $f(\cdot)$. One of the popular methods that can be used to address the above alignment problem is dynamic time warping.[14] However, DTW was original proposed for general purposes. In the SPC scenario, a profile may be contaminated by both noises and shift signals. The existing alignment algorithms are easily misled by such signals.

Therefore, in this work, we propose a robust DTW algorithm first, then design improved methods to monitor aligned profiles.

A framework for unaligned profile monitoring is shown in Figure 2. In the first step, historical profiles that are assumed to be in control but have different lengths and observational points are analyzed, and then a robust DTW algorithm is applied to find a baseline profile. Next, other profiles are registered to the baseline, so that the in-control mean and variance of the in-control process can be estimated. The control limit is also estimated at this stage. In the second step, the designed chart is applied to online profiles. When a new profile arrives, it is first aligned to the baseline profile, after which the mean vector of the aligned profile is estimated by the proposed penalty methods. Finally, the estimated information is integrated with the control chart to detect potential shifts; a decision about process status is then made.

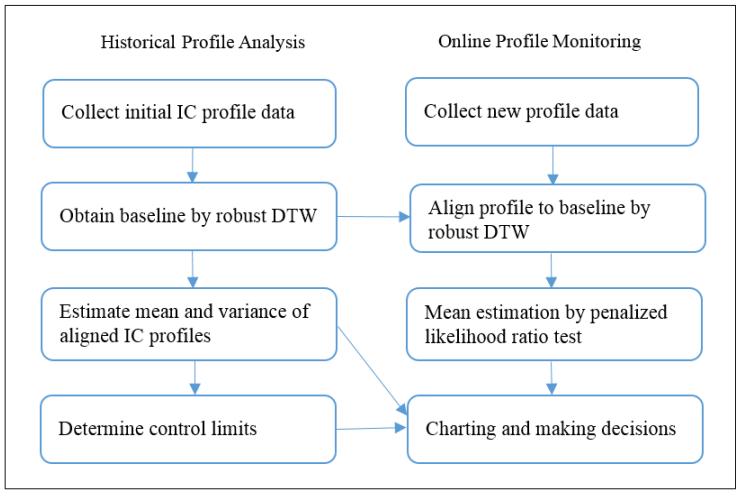


Figure 2. A framework for unaligned profile monitoring

4 Robust DTW for profile alignment

To monitor unaligned profiles, we need to align them so that an equal length and dimension are obtained. In this section, we first give a brief review of the existing DTW algorithm for curve registration. Then, taking into account the presence of possible shifts in unaligned profiles, we proposed a robust algorithm (robust DTW) for profile alignment. After that, a further discussion on parameter design for robust DTW is presented.

4.1 A brief introduction of DTW

The warping function $f(\cdot)$ in Equation (1) can be estimated using DTW. The DTW was first introduced for speech recognition, and it was later widely used in fault diagnosis, pattern recognition and cassation, etc.[15-18] Ramaker *et al.*[19] and Kassidas *et al.*[20] applied DTW for profile warping, but the authors assumed

that the profiles have an identical length. Jeong *et al.*[18] modified the traditional DTW and proposed assigning weights to different points for time series classification. To solve the problem of monitoring dynamic profiles, Dai *et al.*[14] proposed using the dynamic time warping method. However, they did not discuss the robustness of the alignment operation when profiles have local shifts.

In keeping the notations introduced earlier, let (x_{ij}, y_{ij}) be the i th sampling point on profile j. Assume there are two profiles, Q_j and R, where Q_j is the jth profile to be aligned, $\{Q_{ij} = (x_{ij}, y_{ij}), i = 1, 2, \cdots, n_j, j = 1, 2, \cdots\}$, and R is the reference curve that Q_j will be aligned to, $\{R_{k0} = (x_{k0}, y_{k0}), k = 1, 2, \cdots, n_0\}$. DTW aims to find the mapping path by minimizing the distance of these two sequences.

Let $\mathbf{c} = (c_1, c_2, \cdots, c_m, \cdots, c_h)^T$, be the alignment path, which is an index that stores the aligned point pairs from profiles Q_j and R separately, while h is the number of aligned pairs which is determined by the DTW algorithm automatically, and $\max(n_j, n_0) \le h < n_j + n_0 - 1$. Specifically, $\mathbf{c}$ defines a mapping between profiles Q_j and R, and each element of $\mathbf{c}$ is defined as $c_m = \left(Q_{ij}, R_{k0}\right) = \left((x_{ij}, y_{ij}), (x_{k0}, y_{k0})\right),\ i \in \left\{1, 2, \cdots, n_j\right\}, k \in \left\{1, 2, \cdots, n_0\right\}$, which represents that the m th aligned pair is composed of the i th sampling point on profile Q_j, Q_{ij}, and the k th sampling point on profile R, R_{k0}, and then the warping function for point Q_{ij} on profile Q_j is $f(x_{ij}) = x_{k0}$. Let $d\left(c_m\right) = d\left(Q_{ij}, R_{k0}\right) = \| y_{ij} - y_{k0} \|_p$ be a p -norm distance measure between the two aligned points Q_{ij} and R_{k0}; the distance of aligned points is defined as the p -norm difference of corresponding response variables of profiles. In practice, the Euclidian distance (p =2) is most widely used, and we also choose to use it in our work. At this point, the optimal mapping path of the two profile sequences for DTW can be found by minimizing the sum of the distances of all alignment pairs:

$$D\left(Q_j, R\right) = \min_{\mathbf{c}} \sum_{m=1}^{h} d\left(c_m\right)$$

To obtain better alignment results, the warping algorithm is often subject to certain constraints. Among other things, the boundary conditions continuity and monotonicity are widely used.[21] Dai *et al.*[14] showed that the mean of unaligned profiles should be removed before the DTW algorithm can be applied, as the DTW procedure is not invariant to location shifts in the profiles.

The optimal warping path can be efficiently found using dynamic programming. More specifically, let cumulative distance be the total distance measured from the beginning of the profiles to current point

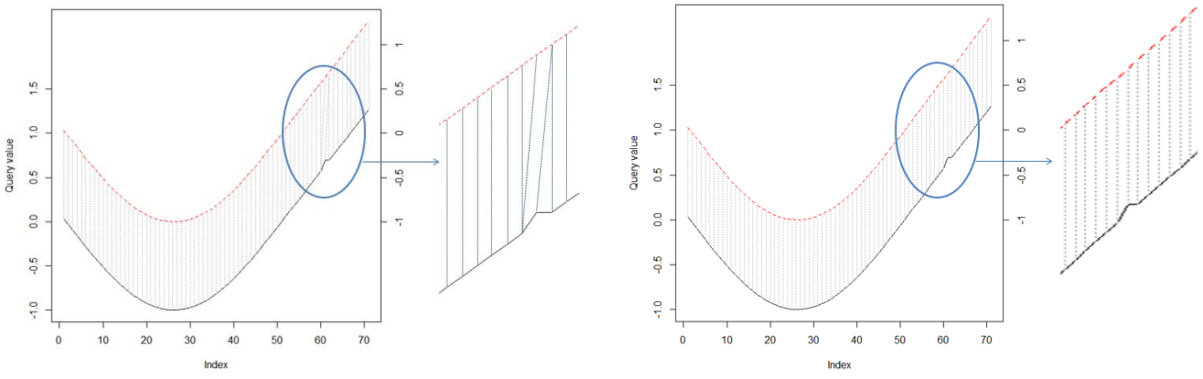
pair $\left(Q_{ij}, R_{k0}\right)$ as $D_c\left(Q_{ij}, R_{k0}\right)$, and the distances of all alignment pairs $D\left(Q_j, R\right)$ is equal to $D_c\left(Q_{n_j j}, R_{n_0 0}\right)$. DTW computes $D_c(Q_{ij}, R_{k0})$ as the sum of points' distance $d\left(c_m\right)$ and the minimum of the cumulative distances of three adjacent pairs:

$$D_c\left(Q_{ij}, R_{k0}\right)=d\left(Q_{ij}, R_{k0}\right)+\min\left\{D_c\left(Q_{(i-1)j}, R_{k0}\right), D_c\left(Q_{(i-1)j}, R_{(k-1)0}\right), D_c\left(Q_{ij}, R_{(k-1)0}\right)\right\}.$$

Following this recurrence, the optimal path $\mathbf{c}$ is found by backtracking the recurring optimal nodes staring from final point pair $\left(Q_{n_j j}, R_{n_0 0}\right)$.

4.2 Robust DTW for profile alignment

DTW is an effective algorithm that can solve the alignment problem of profiles with unequal lengths or uncertain sampling intervals. However, from the algorithm, we learn that points on the query profile are aligned to points on the reference profile one by one. That is, only a single point is considered at each iteration. If there is significant noise on any single point, the complete mapping path will be affected. However, in SPC monitoring, it is very likely that profiles are contaminated by noise signals or a segment of shift. Although we expect that two profiles should be aligned based on their overall trend patterns, such shift signals or noises inevitably affect the alignment operations locally. Figure 3 shows two profiles that have similar trends apart from a hump (a local shift signal) in one of them. If the traditional DTW algorithm is applied, as shown in Figure 3(a), the points near the shifted points are mapped incorrectly. Figure 3(b) shows the correct way to map these two profiles. It's also the case for profiles have noises on each point.



(a) (b)

Figure 3. DTW alignment for curves with a local shift signal: (a) alignment using the traditional DTW and (b) alignment using the robust DTW. Dotted: reference profile; solid: query profile.

To overcome the shortage of the existing DTW algorithm, we here propose a robust DTW algorithm. This algorithm works on a segment-by-segment basis. At each step, profile segments instead of single points from two profiles are compared and mapped. Even if one profile is contaminated by noises or shift signals at a single or even multiple points, the alignment of the two profile segments is less affected since the distances of a lot more points are considered each time.

More specifically, to find the mth alignment pair, we need to evaluate distance between the ith point on Q and the kth point on R. A robust evaluation of the distance is defined as follows:

$$d_w\left(Q_{ij}, R_{k0}\right)_m = \frac{1}{\sum_l s_l}\sum_l \left\{ s_l \left\| y_{(i+l)j} - y_{(k+l)0} \right\|_p \right\}, \quad -\infty < l < +\infty, \quad 1 \le i+l \le n_j, \quad 1 \le k+l \le n_0 \tag{2}$$

where s_l is a weighting function that weights the center the most. Here a density function of a normal distribution with mean 0 and variance ς is chosen to describe w_l, that is, $s_l = \phi_z(l), z \sim N\left(0, \varsigma^2\right)$; and ς determines the effective length of the profile segments. Because the total distance is evaluated as the weighted sum of distances of all adjacent pairs, the impact of any single or multiple points is less profound. If ς=0, it means that we only consider the center point, and thus the robust DTW reduces to the traditional DTW; the algorithm evaluates distances on a point-basis again. When ς>0, we expect that the robust DTW is less sensitive to local shifts and noises in profiles than the traditional DTW, thus can general more reliable alignment results for process monitoring.

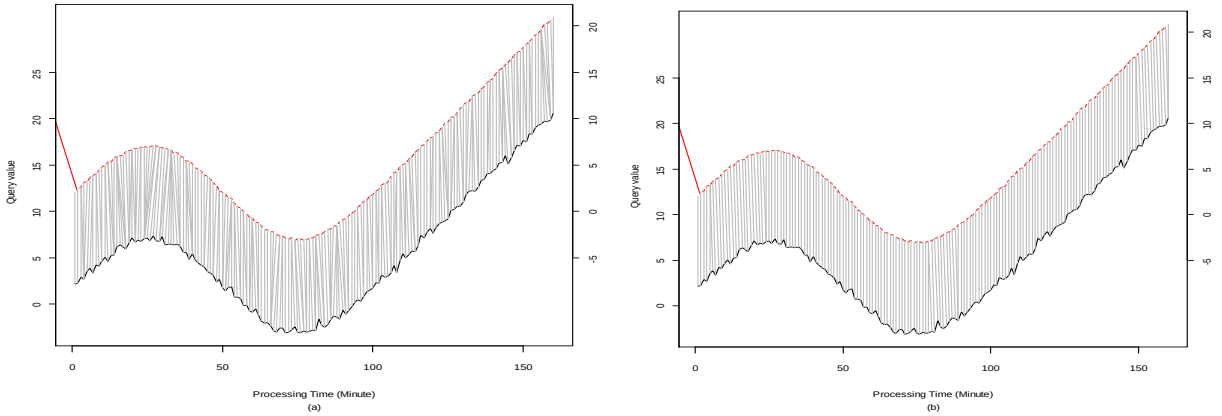


Figure 4. DTW alignment for profiles with noises: (a) alignment using the traditional DTW and (b) alignment using the robust DTW. Dotted: reference profile; solid: query profile.

It should be noted that even if the profile to be aligned is only noisy but has no shifts, the robust DTW still work effectively. Figure 4 shows the alignment of two noisy profiles (profiles contaminated by noises). If the traditional DTW algorithm is applied, as shown in Figure 4 (a), some points are mapped incorrectly, while Figure 4(b) shows that robust DTW provides better alignment performances as it is less sensitive to noise signals.

4.3 Parameter selection for the robust DTW

In robust DTW, the form of the weighting function is taken as the density function of a normal distribution. Therefore, one important parameter here is ς, as it determines the spread of the weighing function, thus determines the effective length of segments in distance calculation.

We here use numerical simulations to study the impact of this parameters. To make the study comprehensive, we choose ς from a wide range of candidates $\Theta = \{0.1, 0.2, \cdots, 1, 2, \cdots, 10\}$. As previously mentioned, the ingot growth profiles that need to be aligned have certain common trends. To mimic the real profiles seen in the ingot growth process, we assume that each simulated profile has three stages: increasing, decreasing, and increasing again. The first two stages are represented by a cycle of a sinusoidal function, and the third stage is formulated by a linear function. That is, the simulated profile is governed by the following function:

$$y_{ij} = \begin{cases} k + a\sin\left(x_{ij}/\omega\right) + \varepsilon_{ij}, & 0 < x_{ij} < 2\pi\omega, \\ y_{2\pi\omega} + b\left(x_{ij} - 2\pi\omega\right) + \varepsilon_{ij} & 2\pi\omega < x_{ij} < 2\pi\omega + 30/b \end{cases} \quad i = 1, 2, \cdots, n_j,\ j = 1, 2, \cdots$$

where $\varepsilon_{ij} \square \left(0, \sigma_j^2\right)$. Under this assumption, the profiles have similar patterns but can be stretched or compressed by changing the parameters.

To investigate the effect of ς, we set $k = 2, \mathrm{a} = 5, \omega = 5, b = 1$, with $\sigma_j = 0.2$ for the query profile and $\sigma_0 = 0$ for the reference profile R. Five types of failure signals are added to the query profile, each of which corresponds to a particular type of shift or failure in the process. Figure 5(a) shows a profile without any shifts. The five added shift signals are: (i) a sudden transient shift; the shift magnitude is set to $\Delta y_{ij} = 1, 5, 11$ at shift point $x_{ij} = 120$ in the third stage, as Figure 5(b) shows; (ii) sustained constant shifts with shift values $\Delta y_{ij} = 1, 2, 3$ after $x_{ij} = 120$ in the third stage, as shown in Figure 5(c); (iii) sustained drifts with $\Delta y_{ij} = k(x_{ij} - 120), k = 0.1, 0.5, 1.0$ after $x_{ij} = 120$ in the third stage, as shown in Figure 5(d); (iv) constant cyclical shifts with $\Delta y_{ij} = r\sin((x_{ij} - 120)\pi\omega_\mu / 2)$, $r = 0.5, 1.0, 2.0$, $\omega_\mu = 1$

after $x_{ij} = 120$ in the third stage, as shown in Figure 5(e); and (v) enlarged cyclical shifts with $\Delta y_{ij} = e^{-s((x_{ij}-120)\pi\omega_\mu/20)} \sin((x_{ij}-120)\pi\omega_\mu/20)$, $s = 0.1, 0.2, 0.3$, $\omega_\mu = 1$, as shown in Figure 5(f).

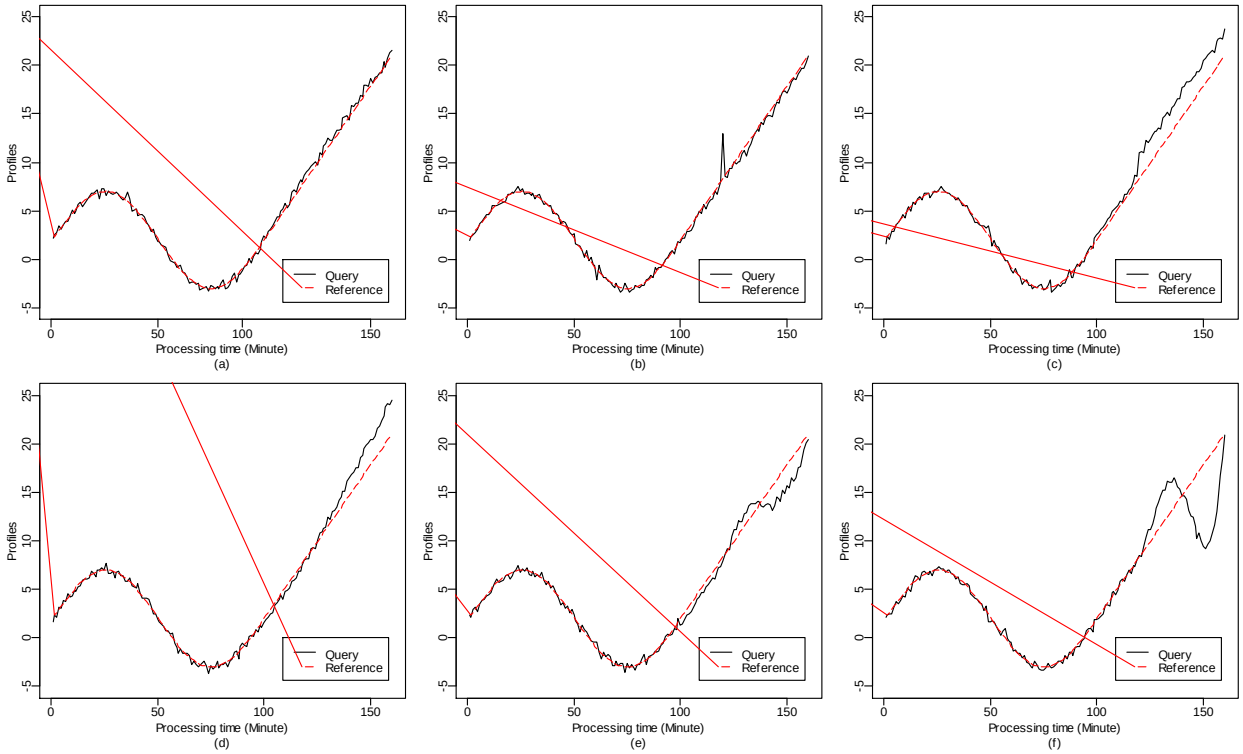


Figure 5. Profiles used for robust DTW alignment: (a) no shift, (b) sudden transient shifts, (c) sustained constant shifts, (d) sustained drift, (e) constant cyclical shifts, and (f) gradually enlarged cyclical shifts

To facilitate the evaluation of alignment performance, the number of explanatory variables for both query profile Q, and reference profile R, are equal. That is, $n_j = n_0$ holds for all profiles. $f(x_{ij})$ represents the aligned indices of query Q, while x_{i0} represents the aligned indices of reference R. Thus, the correct estimation of the warping function is $f(x_{ij}) = x_{i0}$ for every i and j. Keogh and Pazzani[22] used a misalignment index (MI) for performance evaluation by counting the differences between the correct alignment and the alignment generated by an algorithm. The MI is defined as the ratio of a measure of misalignment to a measure of profile length, as follows:

$$MI = \frac{\sum_i \left| f\left(x_{ij}\right) - x_{i0} \right|}{\frac{1}{2} n_0 \left(n_0 - 1\right)}, \; j = 1, 2, \cdots$$

where the role of the denominator is to adapt for different profile lengths.

Clearly, the value of MI is small if there a perfect alignment is achieved. Figure 6 shows the MI values for the six profiles defined above when ς varies. MI is the largest when ς =0 (robust DTW reduces to the traditional DTW). When ς increases, MI decreases slowly first, then sharply, and finally stabilizes at a certain level. Therefore, for this process, a feasible selection for ς is within $(5,9)$.

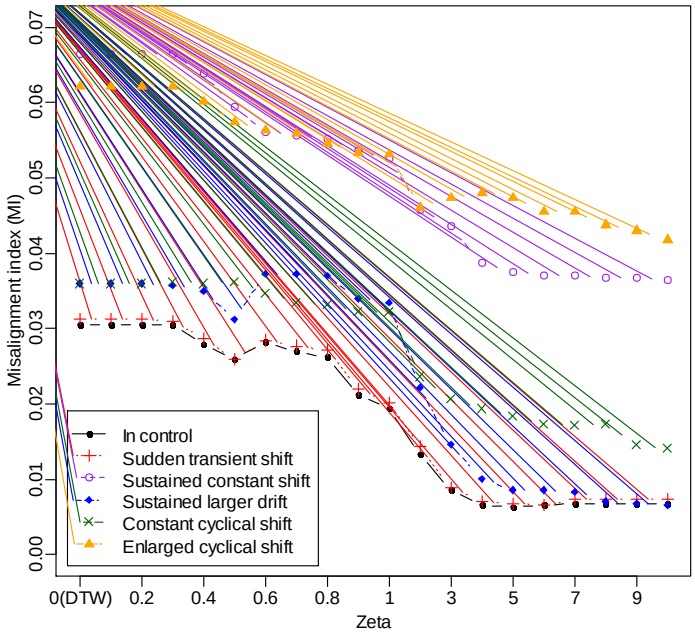


Figure 6. MI for different ς chosen for various profile shift patterns

Once the robust DTW method is developed, following the framework in Figure 2, we can align historical profiles to an equal length, then get a baseline profile for online comparison. The iterative procedure suggested by Dai *et al.*[14] could be borrowed for baseline calculation, expect that the robust distance measure proposed above should be used.

5 Profile monitoring based on penalized estimation

In process monitoring, observed samples are always contaminated by noises. If the true values of a process were known, a more efficient control chart could be designed. Similarly, in aligned profile monitoring, observed profile readings are contaminated by noises from various sources. If the true position of a profile could be obtained, such information could be utilized by a control chart to help improving its charting performance. In this section, we first introduce a GLRT statistic for shift detection, then deviate a little bit and study how to estimate shift signals that is critical to the GLRT statistic. Finally,

we come back to the GLRT statistic, and show the complete charting statistic with penalized estimates of shift signals.

5.1 A GLRT statistic for shift detection in profiles

A profile can be represented by a high-dimensional vector, with each point being one element of the vector. Let the mean of the baseline profile be $\boldsymbol{\mu}_0=(\mu_{10},\mu_{20},\cdots,\mu_{n_0 0})^T$ and the covariance matrix be $\boldsymbol{\Sigma}_0$, where n_0 represents the dimension of the vector. In online monitoring, the mean and covariance of in-control profiles are assumed to be known. Let profiles collected at time j after alignment be $\mathbf{y}_j^{'}=(y_{1j}^{'},y_{2j}^{'},\cdots,y_{n_0 j}^{'})^T$, $j=1,2,\cdots$. Here we assume that each element of $\mathbf{y}_j^{'}$ follows a normal distribution with a dynamic and unknown mean, $\mathbf{y}_j^{'} \sim MN_{n_0}(\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_0)$.

To check whether a profile is different from the baseline, we start by examining the following statistical hypothesis H_0: $\boldsymbol{\mu}_j=\boldsymbol{\mu}_0$ vs. $H_1:\boldsymbol{\mu}_j\neq\boldsymbol{\mu}_0$. Then, the logarithmic generalized likelihood ratio test (GLRT) statistic is:

$$\Lambda\left(\mathbf{y}_j^{'}\right)=\left(\mathbf{y}_j^{'}-\boldsymbol{\mu}_0\right)^T\boldsymbol{\Sigma}_0^{-1}\left(\mathbf{y}_j^{'}-\boldsymbol{\mu}_0\right)-\left(\mathbf{y}_j^{'}-\hat{\boldsymbol{\mu}}_j\right)^T\boldsymbol{\Sigma}_0^{-1}\left(\mathbf{y}_j^{'}-\hat{\boldsymbol{\mu}}_j\right) \tag{3}$$

where $\hat{\boldsymbol{\mu}}_j$ is an estimated process mean of $\boldsymbol{\mu}_j$ for the alternative hypothesis. That is,

$$\hat{\boldsymbol{\mu}}_j=\underset{\boldsymbol{\mu}_j\neq\boldsymbol{\mu}_0}{\arg\min}\left\{\left(\mathbf{y}_j^{'}-\boldsymbol{\mu}_j\right)^T\boldsymbol{\Sigma}_0^{-1}\left(\mathbf{y}_j^{'}-\boldsymbol{\mu}_j\right)\right\}. \tag{4}$$

The null hypothesis is rejected and the alternative hypothesis is favored if $\Lambda(\mathbf{y}_j^{'})>c_1$, where c_1 is a threshold value.

In Equation (3), one critical step is to estimate the profile mean $\boldsymbol{\mu}_j$ based on the profile data $\mathbf{y}_j^{'}$ collected at time j. For the traditional GLRT, the default estimates are $\hat{\boldsymbol{\mu}}_j=\mathbf{y}_j^{'}$, and thus, $\Lambda(\mathbf{y}_j^{'})=(\mathbf{y}_j^{'}-\boldsymbol{\mu}_0)^T\boldsymbol{\Sigma}_0^{-1}(\mathbf{y}_j^{'}-\boldsymbol{\mu}_0)$. The corresponding charting statistic reduces to the traditional Hotelling's T^2 control chart.

However, as $\mathbf{y}_j^{'}$ contains both useful and noise signals, $\hat{\boldsymbol{\mu}}_j=\mathbf{y}_j^{'}$ is obviously an inaccurate estimate. In the next section, we investigate different methods for profile mean estimation. For simplify of expressions, we assume $\boldsymbol{\mu}_0=\mathbf{0}$. That is, the process mean is assumed to be zero when the process is in-control. Although the following derivations are presented under this assumption, it is not difficulty to extend the results to processes with a general in-control mean vector.

5.2 Methods for profile mean estimation

Fan[23] proposed a local linear regression (LLR) method for eliminating noise information with a kernel function. Zou *et al.*[24] applied this nonparametric regression approach to estimate the profile mean estimation $\hat{\boldsymbol{\mu}}_j$, $\hat{\boldsymbol{\mu}}_j = \mathbf{W}\mathbf{y}_j^{'}$, in which $\mathbf{W}$ is a $n_0 \times n_0$ smoothing matrix. The LLR method estimates $\hat{\boldsymbol{\mu}}$ with a linear regression using a point's neighbors. Therefore, the resulting estimates are relatively smoother compared to the raw readings.

Tibshirani *et al.*[25] proposed a Fused LASSO (FLASSO) algorithm for regression estimation. The FLASSO has two penalty terms, the LASSO penalty and the fusion penalty. The LASSO penalty has the same effect as the traditional LASSO, which shrinks some elements to zero; the fusion penalty has the effect of forcing adjacent coefficients to be equal, thereby forming a spatially smoothing solution. The joint use of the FLASSO penalties helps generate smoothed solutions with sparsity, which is exactly the local shift feature we expect to see in profile mean estimation. Therefore, we introduce the use of FLASSO in profile mean estimation as follows:

$$\hat{\boldsymbol{\mu}}_j = \arg\min_{\boldsymbol{\mu}_j} \left\{ \left(\mathbf{y}_j^{'} - \boldsymbol{\mu}_j\right)^T \Sigma_0^{-1} \left(\mathbf{y}_j^{'} - \boldsymbol{\mu}_j\right) + \lambda_1 \sum_{i=1}^{n_0} \left|\mu_{ij}\right| + \lambda_2 \sum_{i=2}^{n_0} \left|\mu_{ij} - \mu_{(i-1)j}\right| \right\} , \quad (5)$$

where λ_1 and λ_2 are penalty coefficients and $|\mu_{ij} - \mu_{(i-1)j}|$ is the absolute value of the difference between the ith element and its neighboring element to the left on the profile. The first constraint $\lambda_1 \sum_{i=1}^{n_0} \left|\mu_{ij}\right|$ encourages sparsity in the coefficients, helps generate sparsity in mean estimates by setting small elements to exactly zero when the in-control mean vector is assumed to be **0**. Existing works have been seen using the LASSO penalty for mean estimation [26, 27]; the second constraint $\lambda_2 \sum_{i=2}^{n_0} \left|\mu_{ij} - \mu_{(i-1)j}\right|$ encourages sparsity in their differences (i.e., flatness of profile mean), and this Fused estimate (FE) can also be used for mean estimation. Overall, the mean estimate obtained by Equation (5) is expected to preserve useful shift information in a profile, while removing noise that may harm charting performance. The equation can be solved by a coordinate-wise optimization algorithm[28], a dynamic programming algorithm[29], or an efficient Fused LASSO algorithm[30].

The above mean estimation methods, LLR, LASSO, FLASSO and FE, emphasize different aspects of the solution, giving different results. LLR pursues a smoothing profile using weighted neighbors in a kernel function. LASSO shrinks solutions to zero to generate sparsity without any structures. FE attempts to obtain a smoothed profile estimate but is still contaminated by noise. FLASSO achieves sparsity in elements and their differences, but has the risk of being misled by noise. Figure 7 shows the estimations

of the four algorithms. In Figure 7, the dots represents the simulated data of $\mathbf{y}_j^{'}$, which are the same in these four algorithms. All of the true profile mean represented by dashed line has a same jump at the 120th point, while the estimated means shown in solid lines are apparently different from each other. LLR gives a smoothed estimate that is different from the true value at most points. LASSO give a solution that has many zeros, but occasional spikes indicate that this method is easily affected by a large amount of noise. FLASSO gives a sparse estimate, but a bias exists for most of the points. Finally, FE gives a smoothed and sparse estimate, but the shifted signal is not captured well.

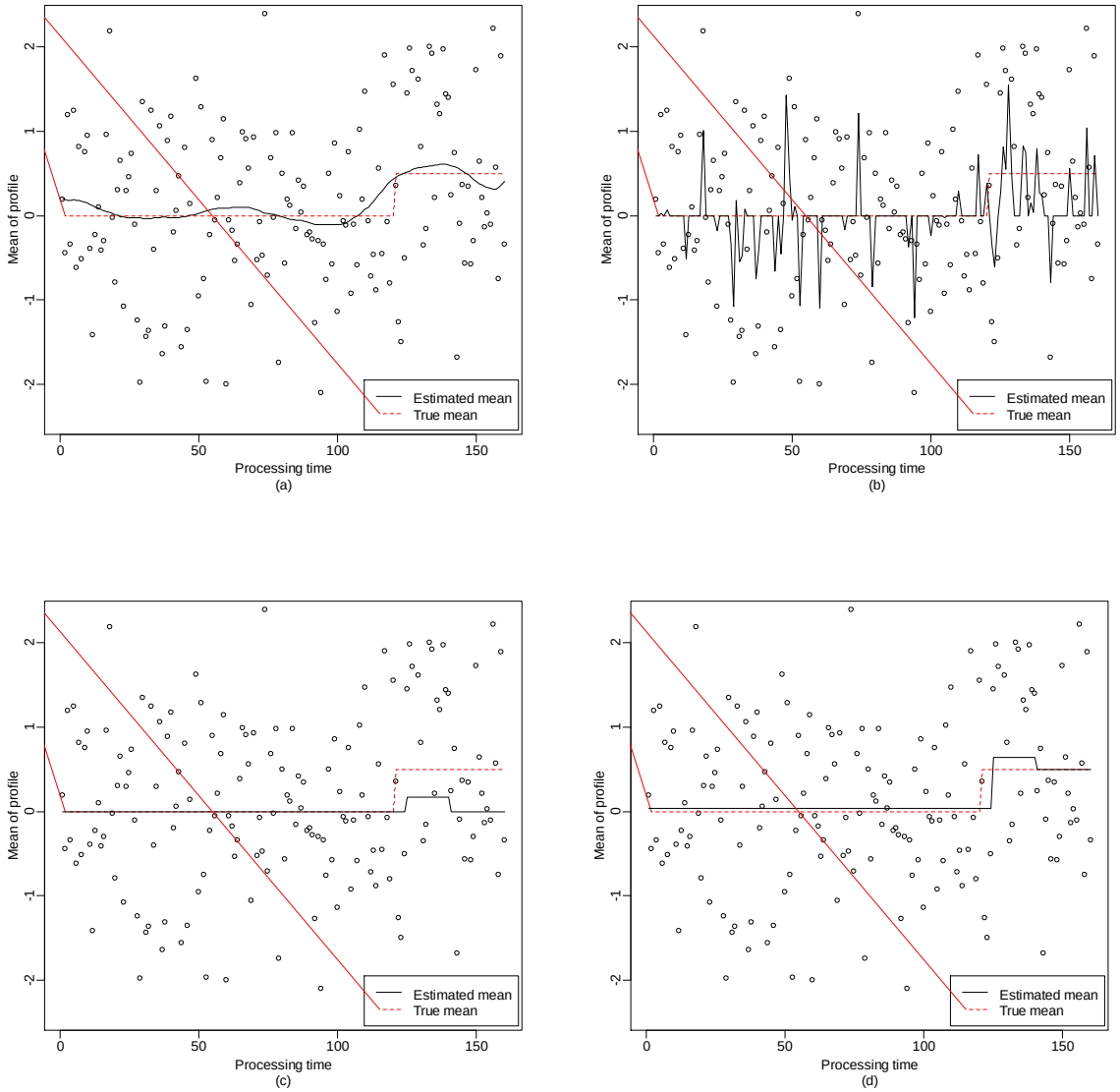


Figure 7. Profile mean estimation using different methods: (a) LLR, (b) LASSO, (c) FLASSO, and (d) FE

To tackle the drawback of the above algorithms, we incorporate the smoothing capability of LLR and the selection power of FLASSO, and propose a new way to retrieve mean estimates. The proposed method adds the Fused LASSO penalty into the LLR solution, denoted as FLASSO-LLR, which is given by:

$$\hat{\boldsymbol{\mu}}_j = \arg\min_{\boldsymbol{\mu}_j} \left\{ \left(\mathbf{W}\mathbf{y}_j^{'} - \boldsymbol{\mu}_j \right)^T \Sigma_0^{-1} \left(\mathbf{W}\mathbf{y}_j^{'} - \boldsymbol{\mu}_j \right) + \lambda_1 \sum_{i=1}^{n_0} \left| \mu_{ij} \right| + \lambda_2 \sum_{i=2}^{n_0} \left| \mu_{ij} - \mu_{(i-1)j} \right| \right\}, \tag{6}$$

where $\mathbf{W}$ is a kernel smoothing function; its bandwidth is given by $h_E = 2 \times \left(1/n_0 \sum_{i=1}^{n_0} \left(x_{ij} - \bar{x}_j \right)^2 \right)^{1/2} n_0^{-1/5}$, which is a popular choice in the literature [24,31,32]. In Equation (6), the penality is applied based on LLR-smoothed observations. Compared to Equations (5), each smoothed observation in Equation (6) is calcualted from a set of neighbours. Therefore, we expected that the estimates obtained from Equation (6) is less sensitive to noises, and thus facilities more effective monitoring of process changes. If only the Fusion penalty is used above, we name this method FE-LLR.

Figure 8 shows the performances of FLASSO-LLR and FE-LLR when applied to the simulated profile shown in Figure 7. Comparatively, FLASSO-LLR gives an estimate that is close to the true mean, while FE-LLR gives a smoother estimate than LLR.

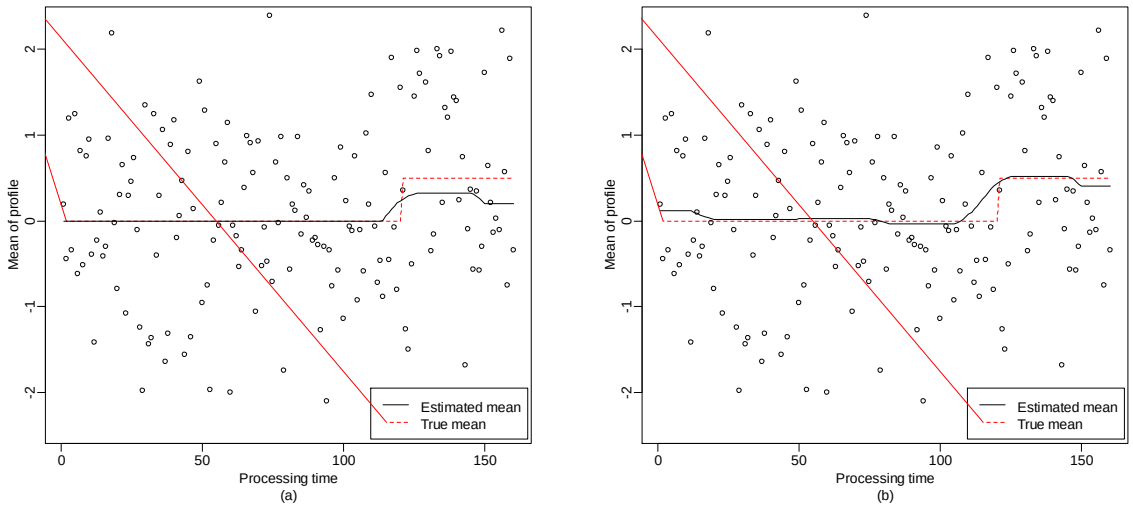


Figure 8. Profile mean estimation by different methods: (a) FLASSO-LLR and (b) FE-LLR

5.3 Profile monitoring based on penalized estimates

The mean profiles obtained by solving Equations (5) give estimates of true process status. In this section, we propose a charting scheme based on the estimates and use this information to boost charting performance.

Once a mean estimate, $\hat{\boldsymbol{\mu}}_j$, is obtained by either one of the estimation methods, the process can be monitored by plugging the estimate into Equation (3), which becomes:

$$\Lambda\left(\mathbf{y}_j^{'}\right)=2\mathbf{y}_j^{'T}\boldsymbol{\Sigma}_0^{-1}\left(\hat{\boldsymbol{\mu}}_j-\boldsymbol{\mu}_0\right)+\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0-\hat{\boldsymbol{\mu}}_j^T\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{\mu}}_j .$$

Then, an equally effective control chart based on penalized likelihood ratio is achieved:

$$\Lambda\left(\mathbf{y}_j^{'}\right)=2\mathbf{y}_j^{'T}\boldsymbol{\Sigma}_0^{-1}\left(\hat{\boldsymbol{\mu}}_j-\boldsymbol{\mu}_0\right)-\hat{\boldsymbol{\mu}}_j^T\boldsymbol{\Sigma}_0^{-1}\hat{\boldsymbol{\mu}}_j > c_2 \qquad (7)$$

where c_2 is a proper control limit determined by a predefined false alarm rate.

Before the monitoring procedure can be used, the panelized parameters λ_1 and λ_2 should be properly chosen as they determine the sparsity and smoothness of the estimated profile mean, respectively. We follow the method used by Zou *et al.*[27] and Wang *et al.*[33], in which a sequence of turning points of penalty parameters are used to obtain estimates, and a normalized statistic is used for monitoring. Following this idea, we set $\Gamma_p=\{\lambda_{1i}, i=1,2,\cdots,p\}$, and $\Gamma_q=\{\lambda_{2k}, k=1,2,\cdots,q\}$ be a set of candidates for λ_1 and λ_2, where p and q are two pre-specified constants. Thus, we define the charting statistic as follows:

$$T=\max_{\substack{i=1,\cdots,p\\k=1,\cdots,q}}\frac{\Lambda_{(\lambda_{1i},\lambda_{2k})}\left(\mathbf{y}_j^{'}\right)-E\left(\Lambda_{(\lambda_{1i},\lambda_{2k})}\left(\mathbf{y}_j^{'}\right)\right)}{\sqrt{Var\left(\Lambda_{(\lambda_{1i},\lambda_{2k})}\left(\mathbf{y}_j^{'}\right)\right)}}\ , \qquad (8)$$

where, $E(\Lambda_{(\lambda_{1i},\lambda_{2k})}(\mathbf{y}_j^{'}))$ and $Var(\Lambda_{(\lambda_{1i},\lambda_{2k})}(\mathbf{y}_j^{'}))$ are the mean and variance, respectively, of statistic $\Lambda_{(\lambda_{1i},\lambda_{2k})}(\mathbf{y}_j^{'})$ for a specific choice of (λ_1,λ_2), which are approximated via situation using the empirical expectation and variance of $\Lambda_{(\lambda_{1i},\lambda_{2k})}\left(\mathbf{y}_j^{'}\right)$ by setting $\mathbf{y}_j^{'}$ follows multivariate normal distribution with $\boldsymbol{\mu}_0=0$ and $\boldsymbol{\Sigma}_0=\mathbf{I}_{n_0}$. Similarly, profile means estimated by Equation (5) for FLASSO or Equation (6) for FLASSO-LLR could be used. This control chart triggers a signal when $T>L$, where L is a properly chosen control limit. Similarly, for the FE chart, and also the FE-LLR chart, the charting statistics and control limits can be designed and determined accordingly.

5.4 Charting performance comparison

We now use simulations to study the performance of different charts discussed above. We assume all of the aligned profiles have a length of 160, $n_0 = 160$. Without loss of generality, we assume that all in-control aligned profiles follow the standard normal distribution, $\mathbf{y}_0^{'} \sim MN_{n_0}(\boldsymbol{\mu}_0, \mathbf{I})$, where $\boldsymbol{\mu}_0 = (\mu_{10}, \mu_{20}, \cdots, \mu_{n_0 0})^T$. In Phase II online monitoring, the aligned profile $\mathbf{y}_j^{'}$ follows a multivariate normal distribution with the same covariance structure but an unknown mean $\boldsymbol{\mu}_j = (\mu_{1j}, \mu_{2j}, \cdots \mu_{ij}, \cdots \mu_{n_0 j})^T$, $\mathbf{y}_j^{'} \sim MN_{n_0}(\boldsymbol{\mu}_j, \mathbf{I})$.

Following the framework in Figure 2, we first generate historical profiles and obtain a baseline. Next, all profiles to be monitored are standardized with respect to the baseline by calculating

$$y_{ij}^{''} = \left(y_{ij}^{'} - \mu_{i0}\right) / \sigma_{i0}, i = 1, 2, \cdots, n_0, \; j = 1, 2, \cdots .$$

When the process is in-control, the standardized profile, $\mathbf{y}_j^{''} = (y_{1j}^{''}, y_{2j}^{''}, \cdots, y_{n_0 j}^{''})^T$, follows the standard normal distribution, $\mathbf{y}_j^{''} \sim MN_{n_0}(\mathbf{0}, \mathbf{I})$. If the process is out-of-control, then $\mathbf{y}_j^{''}$ follows the shifted normal distribution, $\mathbf{y}_j^{''} \sim MN_{n_0}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_0, \mathbf{I})$, and the profile mean shift is $\boldsymbol{\mu}_j - \boldsymbol{\mu}_0$.

To mimic different shift scenarios, we consider three types of shifts for the profile mean that are frequently seen in the ingot growth process: a sustained shift, a sustained drift, and a cyclical shift (marked as (c)-(f) in Figure 5). As we have mentioned earlier, if a shift occurs in a profile, it is likely to affect segment of the profile mean vector. Hence, we assume an abnormal profile shifts only occurs starting from step τ_μ. Some shift types also have varying parameter settings, thus creating six shift patterns, OC1-OC6, as shown in Table 1. For the sustained shift, we consider different shift sizes δ_μ s (OC1) and different shift positions τ_μ s (OC2); for the sustained drift (OC3), the shift occurs at a fixed position but with different gradients k_μ s; and for the cyclical shift, we consider shifts with different sizes r_μ s (OC4) and frequencies ω_μ s (OC5), but the shift amplitude r_μ remains a constant. In practice, a cyclical shift may be amplified due to built-in feedback control mechanisms. We also observe the case where a cyclical shift's amplitude increases within the product cycle, which is described by an exponential function with parameter s_μ (OC6). The six out-of-control patterns are added to simulated profiles, and the resulting profiles are monitored by different charts.

The sparsity penalty λ_1 and the fusion penalty λ_2 are adaptively chosen to maximize the charting statistic. Based on our experience and values of the simulated data, we define the parameter spaces Φ_1, Φ_2 for the penalty parameters in Table 2. As FLASSO-LLR and FE-LLR are the penalization-based

methods based on smoothing, the smoothing estimation has decreased processes' noises, which leads to the penalty parameters for FLASSO-LLR and FE-LLR much smaller than the methods without smoothing step when to estimation the true process mean.

Table 1. Types of failures for profile mean

Shift pattern	Mean shift	Fixed parameters	Shift element
OC1	$\mu_{ij} \to \mu_{ij} + \delta_\mu, i > \tau_\mu$	$\tau_\mu = 120$	δ_μ
OC2	$\mu_{ij} \to \mu_{ij} + \delta_\mu, i > \tau_\mu$	$\delta_\mu = 0.5$	τ_μ
OC3	$\mu_{ij} \to \mu_{ij} + k_\mu \left(i - \tau_\mu\right) / \left(n_0 - \tau_\mu\right), i > \tau_\mu$	$\tau_\mu = 120$	k_μ
OC4	$\mu_{ij} \to \mu_{ij} + r_\mu \sin\left[\left(i - \tau_\mu\right)\pi\omega_\mu / 40\right], i > \tau_\mu$	$\tau_\mu = 80; \omega_\mu = 2$	r_μ
OC5	$\mu_{ij} \to \mu_{ij} + r_\mu \sin\left[\left(i - \tau_\mu\right)\pi\omega_\mu / 40\right], i > \tau_\mu$	$r_\mu = 0.7; \tau_\mu = 80$	ω_μ
OC6	$\mu_{ij} \to \mu_{ij} + r_\mu e^{s_\mu\left(i-\tau_\mu\right)} \sin\left[\left(i - \tau_\mu\right)\pi\omega_\mu / 40\right], i > \tau_\mu$	$r_\mu = 0.15; \tau_\mu = 80; \omega_\mu = 2$	s_μ

Table 2. Parameter spaces for penalty coefficients

	Φ_1	Φ_2	p	q
FLASSO	$\{0.2, 0.4, 0.6, \cdots, 2.0\}$	$\{0.5, 1.0, 1.5, \cdots, 5.0\}$	10	10
FLASSO-LLR	$\{0.02, 0.04, 0.06, \cdots, 0.20\}$	$\{0.05, 0.10, 0.15, \cdots, 0.50\}$	10	10
FE	—	$\{0.1, 0.4, 0.7, \cdots, 5.8\}$	—	20
FE-LLR	—	$\{0.01, 0.04, 0.07, \cdots, 0.58\}$	—	20

For comparison, we choose the nonparametric regression method based on LLR proposed by [34] as a benchmark. ARL is widely used in the literature for charting performance comparison. In the following example, the in-control ARLs of all control charts are set to 200, and the corresponding out-of-control ARLs are calculated and compared. Each ARL is obtained using 10,000 replicates.

5.5 ARL comparison

Table 3 shows the ARL performance of all competing charts under different types of failure patterns. For each shift scenario, the smallest ARL is shown in bold. A number of observations can be made from examining the results, as follows:

1. A general comparison between LLR and other penalization-based chart reveals that at least one penalization-based chart outperforms LLR for all of the shift patterns we considered (OC1-OC6). This shows that adding a penalty to the GLR statistic or LLR-based GLR statistic is effective in identifying shifts in a profile.

Table 3. ARLs comparison

Shift patterns	Shift values	LLR	Control charts based on penalized likelihood ratio			
			FLASSO	FE	FLASSO-LLR	FE-LLR
-	-	200	200	201	199	200
OC1	0.1	128	165	142	**125**	131
	0.2	50.7	90.4	61.4	**48.1**	51.8
	0.3	18.6	34.6	22.6	**16.8**	18.4
	0.4	7.54	12.6	8.59	**6.58**	7.21
	0.5	3.51	5.23	3.93	**3.11**	3.48
	0.6	2.06	2.55	2.19	**1.88**	2.00
	0.7	1.42	1.61	1.47	**1.33**	1.41
	0.8	1.16	1.22	1.18	**1.11**	1.15
	0.9	1.05	1.07	1.05	**1.03**	1.04
	1.0	1.01	1.02	1.01	**1.01**	1.01
OC2	10	1.01	1.28	1.03	1.02	**1.01**
	20	1.02	1.35	1.04	1.02	**1.02**
	30	1.04	1.42	1.06	1.04	**1.03**
	40	1.06	1.52	1.09	**1.05**	**1.05**
	50	1.10	1.62	1.14	**1.08**	**1.08**
	60	1.14	1.78	1.20	**1.13**	1.14
	70	1.23	1.94	1.30	**1.20**	1.22
	80	1.35	2.18	1.45	**1.31**	1.33
	90	1.55	2.54	1.69	**1.46**	1.52
	100	1.88	3.00	2.06	**1.75**	1.81
	110	2.46	3.87	2.72	**2.23**	2.38
	120	3.60	5.09	3.92	**3.14**	3.50
	130	5.95	7.72	6.63	**5.13**	5.85
	140	12.4	13.0	14.2	**9.07**	11.7
	150	38.4	29.9	47.0	**25.3**	37.0
OC3	0.1	168	182	178	**161**	168
	0.2	110	130	121	**94.6**	106
	0.3	61.0	77.4	71.0	**50.5**	58.9
	0.4	32.7	42.3	39.3	**25.3**	32.3
	0.5	18.0	21.9	21.7	**13.3**	16.9
	0.6	9.84	12.0	12.1	**7.43**	9.59
	0.7	6.10	6.78	6.97	**4.56**	5.68
	0.8	3.87	4.21	4.50	**3.05**	3.76
	0.9	2.64	2.82	3.02	**2.17**	2.59

Shift patterns	Shift values	LLR	Control charts based on penalized likelihood ratio			
			FLASSO	FE	FLASSO-LLR	FE-LLR
	1.0	1.98	2.05	2.18	**1.67**	1.94
	1.1	1.54	1.61	1.67	**1.38**	1.51
	1.2	1.33	1.33	1.39	**1.21**	1.31
	1.3	1.17	1.18	1.22	**1.10**	1.16
	1.4	1.09	1.09	1.11	**1.05**	1.08
	1.5	1.04	1.04	1.06	**1.02**	1.04
OC4	0.1	178	185	**25.9**	181	181
	0.2	125	147	**18.0**	133	130
	0.3	78.8	101	**10.9**	86.0	81.0
	0.4	44.2	58.8	**6.13**	51.2	45.9
	0.5	23.6	31.1	**3.56**	30.2	24.8
	0.6	12.8	16.0	**2.20**	17.6	13.7
	0.7	6.98	8.37	**1.58**	10.2	7.59
	0.8	4.20	4.53	**1.24**	6.25	4.49
	0.9	2.74	2.68	**1.09**	3.96	2.89
	1.0	1.93	1.79	**1.03**	2.69	1.99
	1.1	1.47	1.36	**1.01**	1.92	1.53
	1.2	1.23	1.14	**1.00**	1.48	1.26
	1.3	1.10	1.05	**1.00**	1.25	1.12
	1.4	1.04	1.01	**1.00**	1.12	1.05
	1.5	1.01	1.00	**1.00**	1.05	1.02
OC5	0.5	1.36	2.12	**1.19**	1.34	1.34
	1.0	1.49	3.42	**1.23**	1.72	1.51
	1.5	2.13	5.62	**1.39**	3.06	2.23
	2.0	7.18	8.42	**1.57**	10.2	7.50
	2.5	40.1	11.5	**1.76**	22.2	31.8
	3.0	99.5	15.2	**1.99**	34.6	64.1
	3.5	106	18.7	**2.21**	47.9	76.2
	4.0	84.8	23.2	**2.45**	61.4	76.9
	4.5	75.7	27.9	**2.67**	73.2	78.6
	5.0	86.0	32.1	**2.92**	85.2	90.0
	5.5	115	35.7	**3.16**	99.4	115
OC6	0.02	142	159	**20.7**	141	143
	0.04	124	142	**18.3**	125	123
	0.06	103	120	**15.5**	98.4	99.4
	0.08	75.7	90.7	**12.3**	71.7	74.3
	0.10	52.9	59.5	**9.18**	47.1	51.7
	0.12	31.2	33.8	**6.46**	27.5	30.5
	0.14	16.8	15.9	**4.10**	14.4	16.5
	0.16	8.20	6.91	**2.53**	6.88	7.83
	0.18	3.80	3.10	**1.62**	3.31	3.60
	0.20	1.87	1.58	**1.18**	1.72	1.81
	0.22	1.20	1.09	**1.02**	1.15	1.18
	0.24	1.01	1.00	**1.00**	1.01	1.01

2. As the shift value increases, ARL decreases except for shift scenarios OC2 and OC5. For OC2, the increase in ARL as the shift element τ_μ increases can be explained by the shift dimension of process

means. As the shift position τ_μ increases, one observes shorter shift duration in the mean vector, which leads to a lower probability of a false signal when the data are abnormal. Similarly, ω_μ controls the frequency of a cyclical signal inflating the IC profile mean for OC5. A larger ω_μ means there are more periods in the cyclical shift adding to the IC profile mean vector, which obtained a severer changes between adjacent mean's elements. The drastic undulation might lose accuracy of profile mean estimation, which makes ARL grow with ω_μ.

3. When there is a sustained mean shift (OC1 and OC2) or a sustained drift (OC3), FLASSO-LLR performs the best in most cases, and FE-LLR perform quite closely to FLASSO-LLR. FLASSO and FE are relative slow in these cases, with FE showing a slightly better performance. This shows that applying penalization to the smoothed profile obtained from LLR is more likely to extract these shift patterns; the LLR smoothing is effective in removing noises and making the sustained mean shift or drift stand out.

4. When a cyclical shift with constant amplitude (OC4, OC5) or increasing amplitude (OC6) occurs, FE performs the best, and it has obvious superiority in all of these cases because of its efficient detection of shifts. Simulation results also suggest that FE control chart is quite robust even when amplitude or frequency of cyclical shift is unspecified. Performance of FLASSO differs with the values of shifts. FLASSO performs closely to FE with large value of cyclical shift, while the LLR-based method has no obvious superiority in most cases. When a process with a large cyclical shift is being monitored, applying LLR to profiles has the effect of smoothing both noise and signals, thus weakening the real cyclical signal. Therefore, in these cases, FE and FLASSO perform better. When the process undergoes a cyclical shift with a small value, FLASSO and LLR have no competitive power compared with FE. This is because the cyclical information is more important when the shift amplitude is small; the effects of FLASSO and LLR in removing noise is less prominent in such cases.

In summary, most penalization-based charts perform better than the LLR chart. This shows that the suggested penalization is effective in removing noise and screening out useful shift signals for monitoring. Among the penalized charts, the FLASSO-LLR chart best suits the cases with mean shift or drift (or cyclical shifts with relatively large amplitude), while FE best suits the cases in which cyclical features dominate.

6 **A real example**

In this section, we apply the proposed framework for unaligned profile monitoring to the ingot growth process and monitor heating power profiles. A heating power profile is formed by retrieving the power

reading with a fixed time interval within a long production cycle. Several samples of these profiles are shown in Figure 1.

To build a baseline, 20 historical profiles with different time spans that are judged by engineers as in control are aligned using robust DTW first; the baseline profile is then computed after the reference profile is determined and all other profiles are aligned to the reference. The parameter spaces Φ_1, Φ_2 for the penalization-based charts are set according to Table 2. The upper control limits (UCLs) of the charts are calculated by setting the false alarm rate to 0.005.

We then pick another 20 profiles for online monitoring. All profiles are aligned to the baseline using robust DTW. These aligned profiles are plotted in FLASSO, FE, FLASSO-LLR, FE-LLR and LLR charts. As our results have shown that FLASSO-LLR and FE-LLR have exactly the same alarm signals for all profiles, we here use one figure to represent both charts to save space. The running states of the charts are shown in Figure 9.

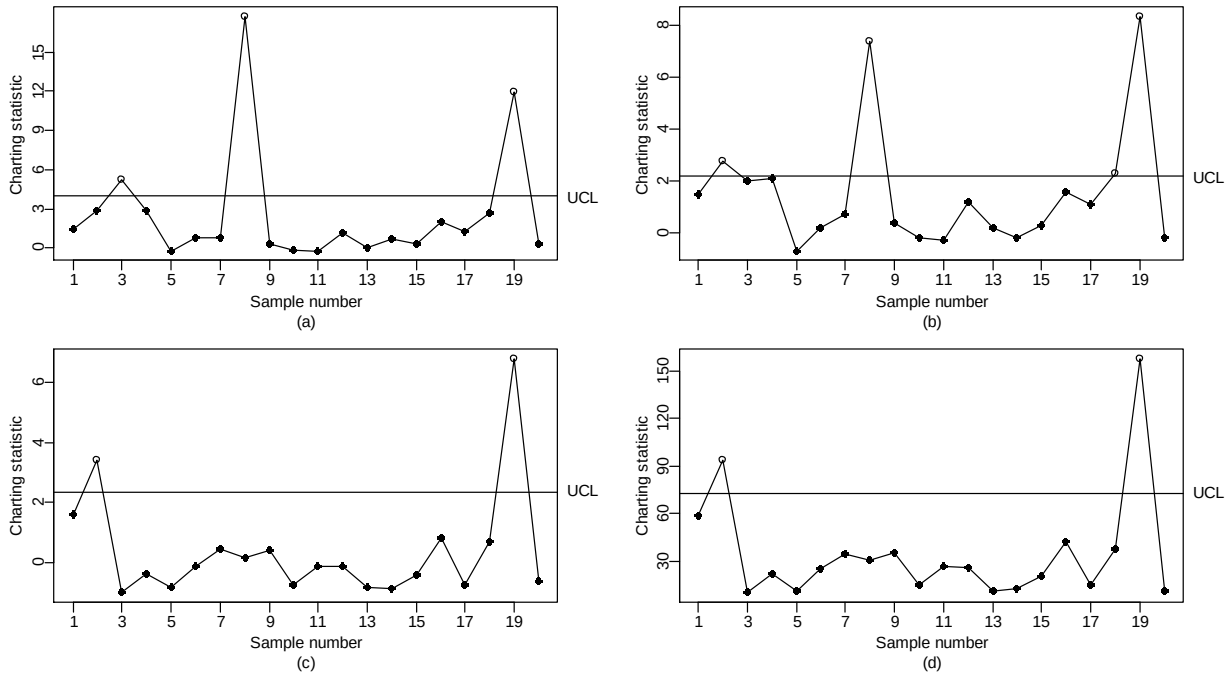


Figure 9. Phase II control charts for 20 profiles: (a) FLASSO, (b) FE, (c) FLASSO-LLR and FE-LLR, and (d) LLR.

As seen in Figure 9, these five control charts give similar monitoring signals for most of the profiles. All charts indicate that samples 1, 4-7, 9-17 and 20, are in-control, while samples 19 is out-of-control. However, profiles 2, 3, 8 and 18 are judged differently. For example, FLASSO and FE report that profile 8 is abnormal, while the other three charts related to LLR report the opposite situation. For demonstration

purpose, Figure 10(a) compares the profile with the baseline, and Figure 10(b) shows the alignment of the profile to the baseline. The mean estimates provided by all tested methods are shown in Figure 11. Profile 8 appears to fluctuate strongly starting at step 250; cyclical shifts are observed from that point on. When applying different methods to estimate the profile mean, the cyclical signals are best preserved by FE, as shown in Figure 11 (b). FLASSO in Figure 11 (a) also preserves certain shift signals. However, LLR makes the signal smooth by removing both noise and useful cyclical shift signals, and the penalization after LLR cannot further recover the cyclical shift information. Therefore, in this case, FE and FLSSO perform better. This result is consistent with the extensive study described in the previous section.

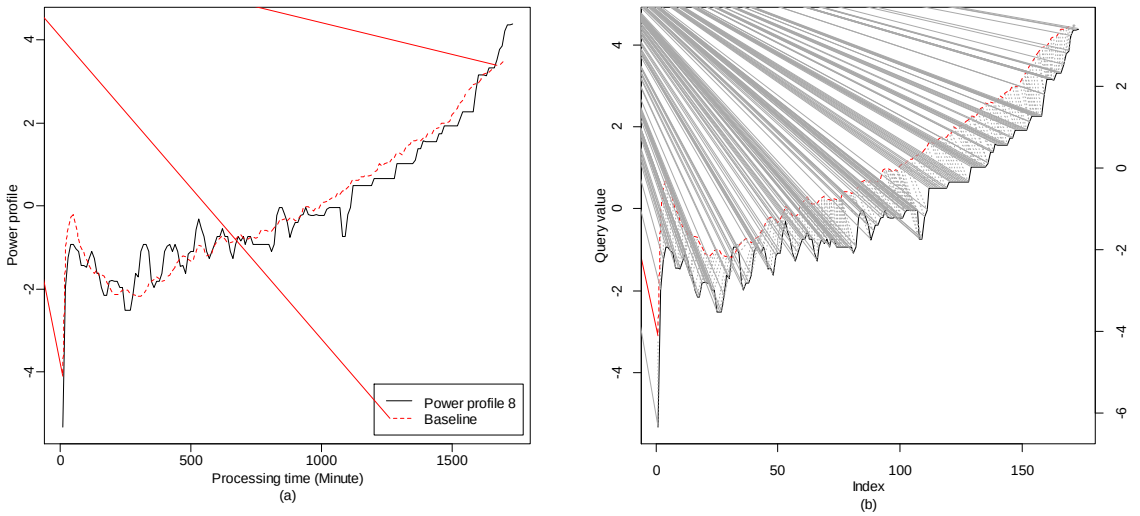


Figure 10. Details of profile 8: (a) profile curve and (b) profile alignment by robust DTW

7 Conclusions and further research

This article provides a framework for monitoring unaligned profiles based on robust DTW and penalized likelihood estimation. In the proposed procedure, a baseline profile is first calculated from aligned in-control profiles. Then, a new profile to be monitored is aligned to this baseline, and the true mean of the aligned profile is estimated using several penalization-based methods.

Once the profile alignment based on robust DTW has been completed, control charts are then derived from the likelihood-ratio test by taking the estimated mean into consideration. We compare the proposed penalization-based charts with a traditional nonparametric regression chart via extensive simulation studies and a real example. The results show that penalization-based charts generally have superior performance. More specifically, the FLASSO-LLR chart performs the best in most traditional cases. FE chart is the best in cyclical situation which is commonly encountered in feed-back systems, and its performance is not much worse than FLASSO-LLR in other cases.

It is worth noting that the penalization-based charts have an additional merit. Whenever an out-of-control signal is triggered, the estimated mean provides direct clues for diagnosis. Thus, the use of such penalized charts is especially appealing in such applications.

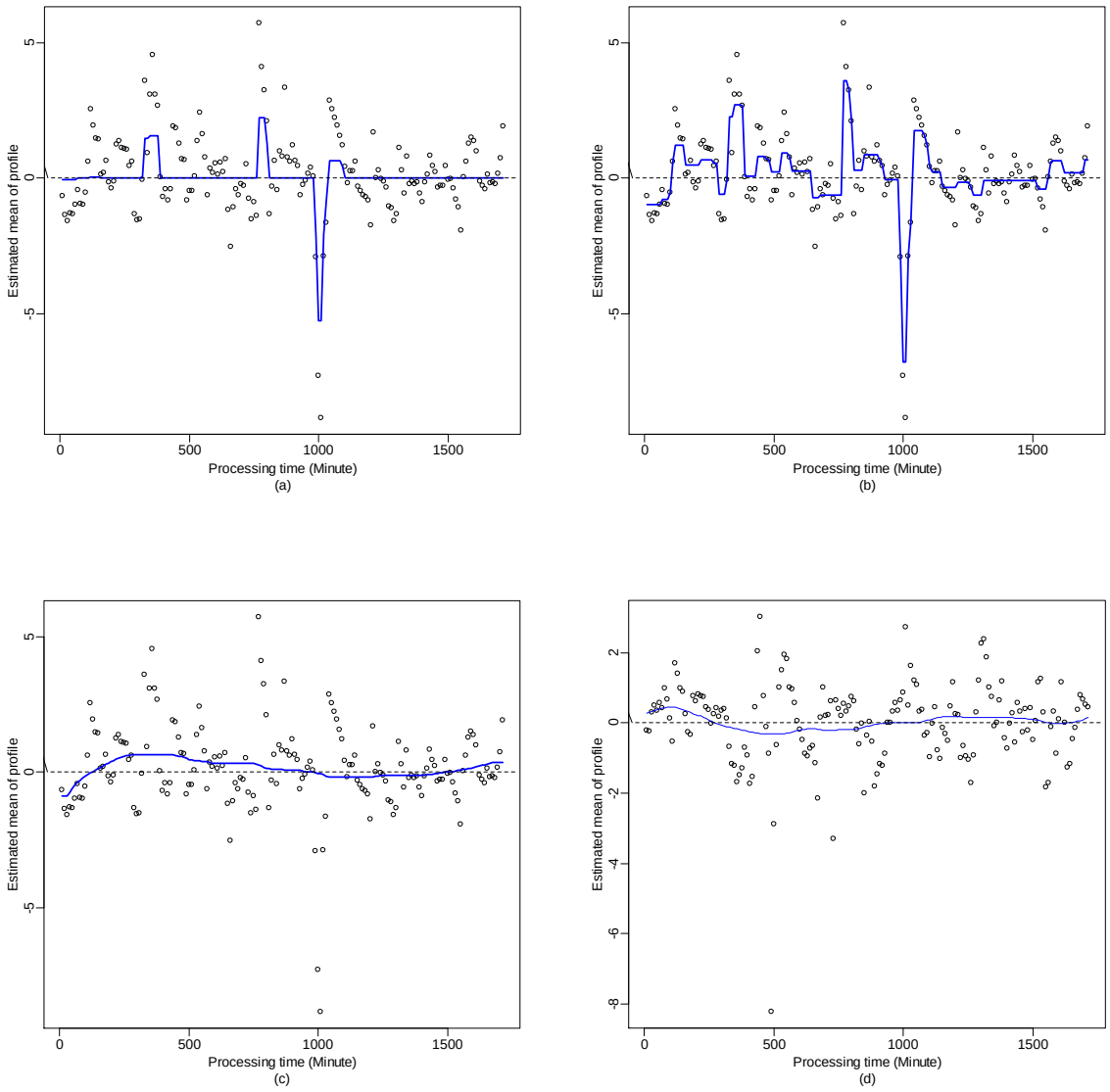


Figure 11. Mean estimation of aligned profile 8 by different methods: (a) FLASSO, (b) FE, (c) FLASSO-LLR and FE-LLR, and (d) LLR

References

1. Kang L, Albin SL (2000) On-line monitoring when the process yields a linear profile. *Journal of Quality Technology*. **32**: 418-426.

2. Zou C, Tsung F, Wang Z (2007) Monitoring general linear profiles using multivariate exponentially weighted moving average schemes. *Technometrics*. **49**: 395-408.

3. Zou C, Zhou C, Wang Z, Tsung F (2007) A self-starting control chart for linear profiles. *Journal of Quality Technology*. **39**: 364-375.

4. Zou C, Zhang Y, Wang Z (2006) A control chart based on a change-point model for monitoring linear profiles. *IIE transactions*. **38**: 1093-1103.

5. Mahmoud MA, Parker PA, Woodall WH, Hawkins DM (2007) A change point method for linear profile data. *Quality and Reliability Engineering International*. **23**: 247-268.

6. Chicken E, Pignatiello JJJR, Simpson JR (2009) Statistical Process Monitoring of Nonlinear Profiles Using Wavelets. *Journal of Quality Technology*. **41**: 198-212.

7. Chang SI, Yadama S (2010) Statistical process control for monitoring non-linear profiles using wavelet filtering and B-Spline approximation. *International Journal of Production Research*. **48**: 1049-1068.

8. Shiau J-JH, Huang H-L, Lin S-H, Tsai M-Y (2009) Monitoring nonlinear profiles with random effects by nonparametric regression. *Communications in Statistics—Theory and Methods*. **38**: 1664-1679.

9. Paynabar K, Qiu P, Zou C (2016) A Change Point Approach for Phase-I Analysis in Multivariate Profile Monitoring and Diagnosis. *Technometrics*. **58**: 191-204.

10. De Ketelaere B, Hubert M, Schmitt E (2015) Overview of PCA-Based Statistical Process-Monitoring Methods for Time-Dependent, High-Dimensional Data. *Journal of Quality Technology*. **47**: 318.

11. Jeong MK, Lu J-C, Wang N (2006) Wavelet-based SPC procedure for complicated functional data. *International Journal of Production Research*. **44**: 729-744.

12. Ramsay J, Li X (1998) Curve registration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. **60**: 351-363.

13. Mosesova SA, Chipman HA, MacKay RJ, Steiner SH (2007) Profile monitoring using mixed-effects models. *Technical Report*.

14. Dai C, Wang K, Jin R (2014) Monitoring Profile Trajectories with Dynamic Time Warping Alignment. *Quality and Reliability Engineering International*. **30**: 815-827.

15. Gupta A, Samanta A, Kulkarni B, Jayaraman V (2007) Fault diagnosis using dynamic time warping. In: Ghosh A, De RK, Pal SK, eds. *Pattern Recognition and Machine Intelligence.* Springer, Berlin Heidelberg, 57-66.

16. Dai Y, Zhao J (2011) Fault Diagnosis of Batch Chemical Processes Using a Dynamic Time Warping (DTW)-Based Artificial Immune System. *Ind Eng Chem Res*. **50**: 4534-4544.

17. Zhen D, Wang T, Gu F, Ball A (2013) Fault diagnosis of motor drives using stator current signal analysis based on dynamic time warping. *Mechanical Systems and Signal Processing*. **34**: 191-202.

18. Jeong Y-S, Jeong MK, Omitaomu OA (2011) Weighted dynamic time warping for time series classification. *Pattern Recognition*. **44**: 2231-2240.

19. Ramaker H-J, van Sprang ENM, Westerhuis JA, Smilde AK (2003) Dynamic time warping of spectroscopic BATCH data. *Analytica Chimica Acta*. **498**: 133-153.

20. Kassidas A, MacGregor JF, Taylor PA (1998) Synchronization of batch trajectories using dynamic time warping. *AIChE Journal*. **44**: 864-875.

21. Keogh E, Ratanamahatana CA (2005) Exact indexing of dynamic time warping. *Knowl Inf Syst*. **7**: 358-386.

22. Keogh EJ, Pazzani MJ (2001) Derivative Dynamic Time Warping. *SIAM*. **1**: 5-7.

23. Fan J (1993) Local Linear Regression Smoothers and Their Minimax Efficiencies. *The Annals of Statistics*. **21**: 196-216.

24. Zou C, Tsung F, Wang Z (2008) Monitoring profiles based on nonparametric regression methods. *Technometrics*. **50**: 512-526.

25. Tibshirani R, Saunders M, Rosset S, Zhu J, Knight K (2005) Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. **67**: 91-108.

26. Wang K, Jiang W (2009) High-dimensional process monitoring and fault isolation via variable selection. *Journal of Quality Technology*. **41**: 247.

27. Zou C, Ning X, Tsung F (2012) LASSO-based multivariate linear profile monitoring. *Annals of operations research*. **192**: 3-19.

28. Friedman J, Hastie T, Höfling H, Tibshirani R (2007) Pathwise coordinate optimization. *The Annals of Applied Statistics*. **1**: 302-332.

29. Johnson NA (2013) A Dynamic Programming Algorithm for the Fused Lasso and L0-segmentation. *Journal of Computational and Graphical Statistics*. **22**: 246-260.

30. Liu J, Yuan L, Ye J (2010) An efficient algorithm for a class of fused lasso problems. *ACM*. 323-332.

31. Qiu P, Zou C, Wang Z (2010) Nonparametric profile monitoring by mixed effects modeling. *Technometrics*. **52**: 265-277.

32. Zhang Y, He Z, Zhang C, Woodall WH (2014) Control charts for monitoring linear profiles with within‐profile correlation using Gaussian process models. *Quality and Reliability Engineering International*. **30**: 487-501.

33. Wang K, Yeh AB, Li B (2014) Simultaneous monitoring of process mean vector and covariance matrix via penalized likelihood estimation. *Computational Statistics & Data Analysis*. **78**: 206-217.

34. Fan J, Gijbels I (1966) Local Polynomial Modeling and Its Applications. *London: Chapman & Hall*.