Special Issue Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.1667

Published online 10 June 2014 in Wiley Online Library

Monitoring Profile Trajectories with Dynamic Time Warping Alignment

Chenxu Dai,[a] Kaibo Wang[a]*† and Ran Jin[b]

In conventional profile monitoring problems, profiles for products or process runs are assumed to have the same length. Statistical monitoring cannot be implemented until a complete profile is obtained. However, in certain cases, a single profile may require several days to generate, so it is important to monitor the profile trajectory to detect unexpected changes during the long processing cycle. Motivated by an ingot growth process in semiconductor manufacturing, we propose a method for monitoring growth profile trajectories of unequal lengths. The profiles are first aligned using the dynamic time warping algorithm and then averaged to generate a baseline. Online monitoring of trajectories is performed based on incomplete growth profiles. Both simulations and an actual application are used to demonstrate the use of the proposed method. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: dynamic time warping; profile monitoring; quality control; statistical process control

1. Introduction

Statistical process control (SPC) methods are extensively used to monitor and improve quality and productivity in manufacturing and service operations.[1,2] One important tool is the control chart, which is often used to monitor the key process variables and to trigger alarms when abnormal changes are detected. In the literature, various methods have been developed for processes with one variable or multiple correlated variables.[2] Recently, profile monitoring has received more attention.[3,4] A profile, which is represented in the form of a functional curve, defines the relationship between one response variable and one or more explanatory variables (e.g., time and locations).

If the profile can be fitted using a parametric model, the model parameters are usually monitored.[5,6] For example, to monitor a linear profile, Kang and Albin[7] fitted a simple linear regression model to the Phase I data and monitored all the parameters and the residual standard errors using T^2, exponentially weighted moving average (EWMA), and R charts. To monitor a more complex roundness profile, Colosimo *et al.*[8] proposed the use of a spatial autoregressive regression model, based on which a vector of parameters is estimated and used for statistical monitoring.

If the profile is so complicated that it cannot be characterized by any reasonable parametric form, nonparametric methods are sometimes employed. In nonparametric control charts for profile monitoring, the charting statistic is usually based on metrics that measure the departure of the observed profiles from a baseline.[3] Jones and Rice[9] and Nomikos and MacGregor[10] proposed to monitor the scores of principal components and the residuals obtained from principal component analysis. Jeong *et al.*[11] used wavelets to transform high-frequency signals for process monitoring. Walker and Wright[12] demonstrated the use of spline models to fit and monitor complex profiles. Zou *et al.*[13] used nonparametric regression to fit a model to a profile dataset. To filter out rotation, translation, and isometric scaling (dilation) effects, Del Castillo and Colosimo[14] proposed a generalized Procrustes algorithm that uses the full Procrustes distance as the metric after the profiles are registered or superimposed. Noorossana *et al.*[4] presented a comprehensive summary that covers a wide range of research on the statistical analysis of profile monitoring.

The conventional profile monitoring methods are usually based on the assumption that the complete profile is available so that the charting statistics can be calculated. Thus, all observations for a profile are required to construct the control chart. However, in certain manufacturing processes, the production cycle is rather long. Any faults that occur during the production cycle should be detected immediately upon appearance rather than waiting until the end of the production cycle. In such cases, a dynamically growing profile trajectory, which could be treated as a partially observable growth profile, should be monitored online.

In the ingot growth process, which is our motivating example, the ingot grows in an automated furnace (shown in Figure 1). In this process, raw polysilicon materials are melted in a quartz crucible to a temperature of more than 2000°F. A seed crystal is dipped into

[a]*Department of Industrial Engineering, Tsinghua University, Beijing 100084, China*
[b]*Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg VA 24061, USA*
**Correspondence to: Kaibo Wang, Department of Industrial Engineering, Tsinghua University, Beijing 100084, China.*
†*E-mail: kbwang@tsinghua.edu.cn*

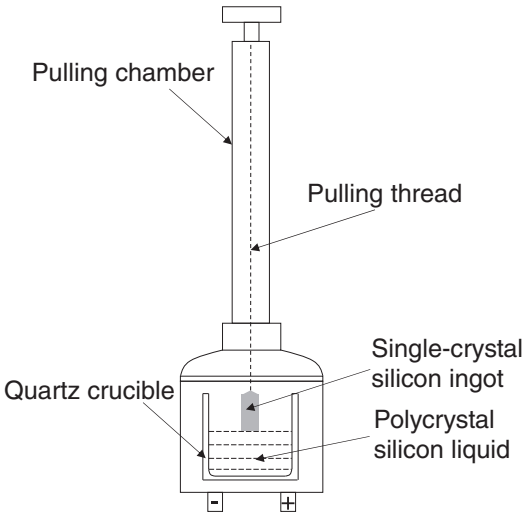

Figure 1. A schematic drawing of the single-crystal ingot growth furnace.

the molten silicon and withdrawn slowly to induce the growth of a single crystal ingot. The cycle time required to grow one ingot may be more than 50 h. In this process, all of the key process variables such as pulling speed, temperature, and heating power must be carefully coordinated to ensure a desirable growth environment. If the process deviates from the norm, corrective actions must be taken immediately to minimize the waste of material and energy. Therefore, monitoring the key variables online is of great importance for the ingot growth process to ensure quick change detection and prevent a reduction in quality.

In this study, we focus on the heating power profile, which is a key variable in the process that affects ingot quality. Because of the complex growth mechanism, the power has a dynamic profile in the growth cycle. Figure 2 shows four sample power profiles collected during the ingot growth process. Each curve represents one growth cycle, which corresponds to the growth of one ingot. Figure 3 shows additional aligned samples (the concept of profile alignment will be explained in a later section) and their mean and standard deviation curves. As can be observed from the figures, the power profiles have different and time-varying means. The trend of the profiles is governed by the physical mechanism of the growth process. An upward or downward shift of the entire profile does not imply an engineering failure. However, a significant deviation from the mean trend indicates possible process changes that should be detected. In addition, the profiles show larger variations in the early stage of the production cycle and gradually stabilize as the process evolves; that is, the inherent variation of the profile is time dependent.

Compared with the profiles that have been studied in the existing literature, the power profiles of the ingot growth process have a distinct feature. Because of the differing amounts of raw materials being used in production, the cycle time of each growth run, and thus the profiles, differs in total length. When more raw materials are used, a longer cycle time is expected. Therefore, profiles of different lengths do not indicate an out-of-control (OC; i.e., irregular) situation. Different lengths are expected as part of the inherent variability of the manufacturing process. Moreover, in online monitoring, the profile will continue to grow with time to an unknown limit. Although there are many observations of the profile when a production cycle finishes, the charting statistic must be evaluated using incomplete online profiles. This growth in the lengths of the profiles makes the online monitoring problem unique.

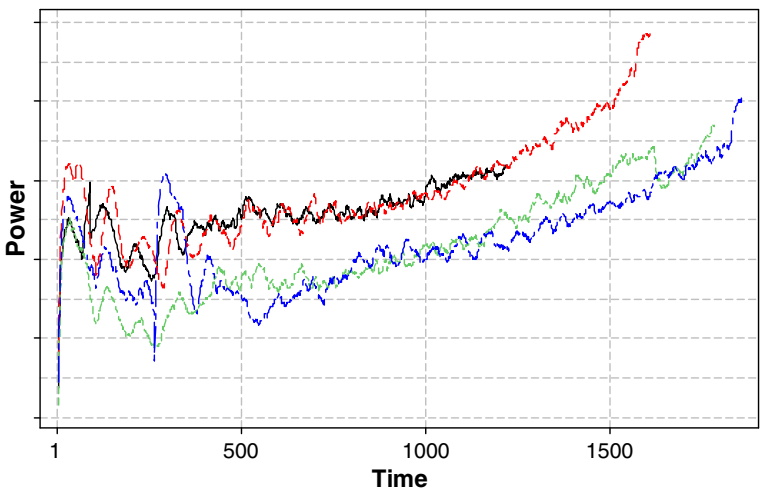

Figure 2. Examples of power profiles in the ingot growth process.

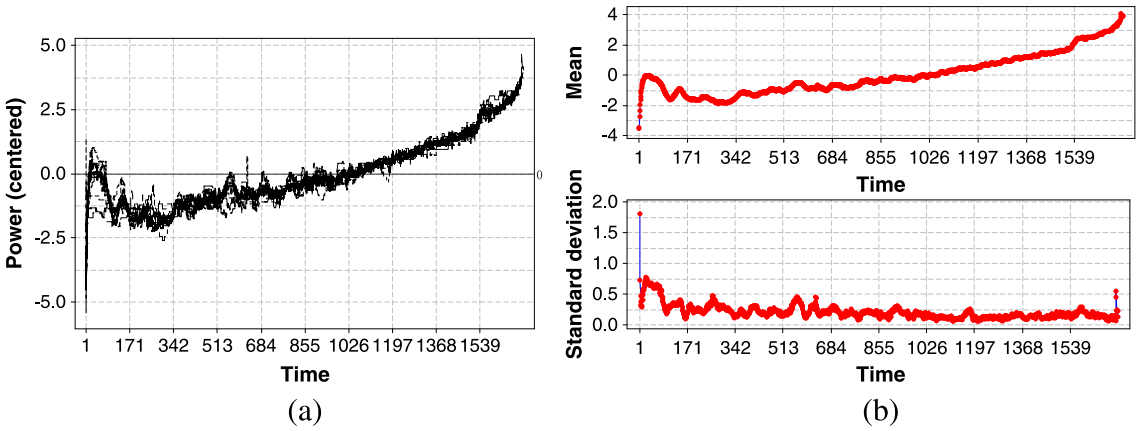

Figure 3. Plots of aligned samples and pointwise sample mean and standard deviation. (a) A collection of aligned samples. (b) The mean and the standard deviation of the aligned samples.

The behavior of the ingot growth process resembles that of a batch process. Nomikos and Macgregor[10] studied the monitoring of batch processes that have multiple variables being collected continuously, and they proposed to extract the main features from the profiles using projection-based methods. In their application, the entire profile of each batch is used for monitoring. In addition, the profile data collected from the batch process are equal in length (i.e., the same number of data points). Therefore, there is no need to align the profiles. Certain other studies have focused on a quantitative trend analysis of the profiles. Various algorithms have been proposed to identify known trends or patterns from online, noisy profile observations.[15–17]

The growth of the profiles as in this study is also found in the dynamic signals in surveillance applications. Krieger *et al.*[18] proposed a sequential method for detecting a gradual change in a profile that can be characterized by a linear regression model. Frisen and Wessman[19] reviewed surveillance methods that are being used in diverse applications. Although the variables that are monitored in surveillance study also evolve with time, unlike the ingot growth process, the surveillance variables do not require time alignment.

The profile trajectories collected from the ingot growth process convey important engineering knowledge about process failures. Given a baseline profile representing an in-control process, any deviation of an online profile from the baseline may indicate a failure in the growth process, usually caused by the process hardware or the raw materials. At that point, the growth process should be stopped to minimize material and energy losses. The profile-to-profile variation may represent either slow or abrupt changes in equipment conditions. For example, the resistivity of the heater will increase, as more production runs are completed, and the efficiency of the thermal field gradually decreases, as the number of production cycles increases. Therefore, practitioners must monitor this process and detect unexpected process shifts early.

In this study, we propose a method for the monitoring of profile trajectories. More specifically, we propose to use a time-warping technique to align the raw profiles and build a baseline model and then use the generalized likelihood ratio for online monitoring based on partially observed profile trajectories.

The paper is organized as follows. The details of the charting strategy are presented in Section 2. In Section 3, we study the performance of the proposed method and compare this method with a benchmark method. In Section 4, we present a case study of ingot growth processes to illustrate the use of the proposed method. Finally, we conclude this paper with suggestions for future research in Section 5.

2. Profile monitoring based on dynamic time warping

To monitor the profile trajectories collected from the ingot growth process, we propose a framework based on dynamic time warping (DTW), as illustrated in Figure 4. Starting with historical profile samples, we must identify the in-control behavior of the profile variable. Therefore, profiles with different lengths and locations are first aligned, and then, a baseline profile is calculated, which serves as the in-control benchmark for online monitoring. During the online monitoring stage, the observed trajectory of each incomplete profile is first aligned with the in-control baseline. Then, a charting statistic is evaluated to make a decision about the status of the process. The alignment and online monitoring calculations are repeated at each step, as a new observation for the profile becomes available and the trajectory progresses.

2.1. Profile alignment

To construct a reliable baseline for online monitoring, all profiles must be aligned to have equal lengths. All in-control profiles will exhibit a similar pattern reflecting the physics of the process. Thus, the alignment operation will use the similarities in the patterns to expand or compress various segments of one profile to match another one. The DTW algorithm is suitable for this purpose.

Dynamic time warping was first proposed in the context of speech recognition to account for the differences in speaking rates among speakers and utterances. The rationale behind DTW is that we can locally expand or compress any two profiles to make

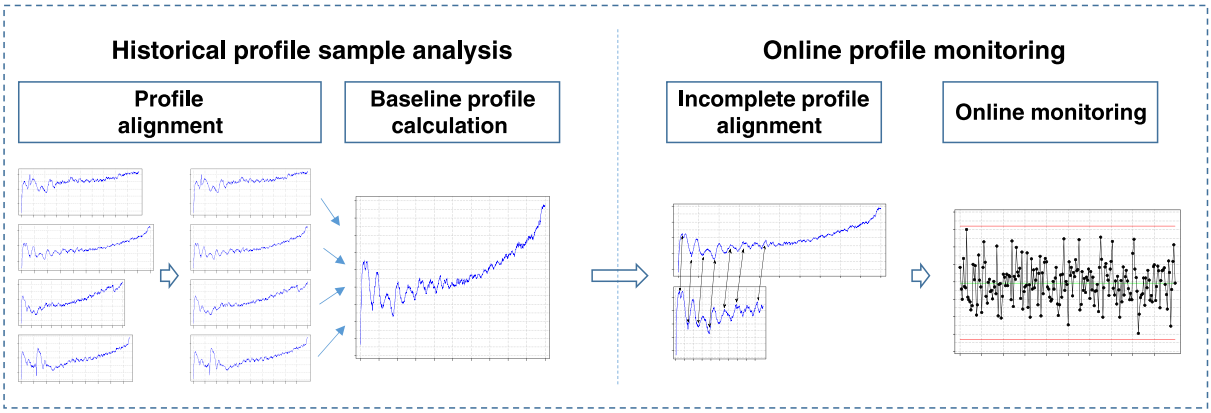

Figure 4. The framework for ingot growth profile monitoring.

one resemble the other as much as possible.[20,21] Gupta *et al.*[22] and Dai and Zhao[23] used DTW for fault diagnosis. Nelson and Runger[24] used DTW for process behavior prediction. DTW was first used for pattern matching in historical data. When the features of the current process match any of the known patterns, the future behavior of the process is predicted by referring to the historical data. Kassidas *et al.*[25] employed DTW to synchronize historical profiles and created a database of historical time-aligned in-control profiles, but they did not explain the methods for choosing a standard profile or how to monitor a new profile.

Denote any two profiles as $X = \{x_1, x_2, \ldots, x_{L_X}\}$ and $Y = \{y_1, y_2, \ldots, y_{L_Y}\}$, where L_X and L_Y are the total lengths of profiles X and Y, respectively. DTW attempts to find the best mapping relationship that matches one of the profiles (the query profile) to the other profile (the reference profile) and to increase the similarity in their variation patterns. The mapping operation may compress one segment of the profile while expanding another segment. For example, in the mapping relationship $\{(1,1),(1,2),(2,3),(3,3),\ldots,(L_X,L_Y)\}$, point x_1 is expanded to match two points, y_1 and y_1, on profile Y, and points x_2 and x_3 are compressed to match a single point y_3.

Let $(\phi_x(k), \phi_y(k))$, $k = 1, 2, \ldots, T$ be the matched pairs, meaning that the $\phi_x(k)$-th point of profile X is matched with the $\phi_y(k)$-th point of profile Y, where $\phi_x(k) \in \{1, 2, \ldots L_X\}$ and $\phi_y(k) \in \{1, 2, \ldots L_Y\}$ are the point indices of the profiles. Furthermore, denote a certain measure of the distance between a matched pair by $d(\phi_x(k), \phi_y(k))$. Then, the DTW algorithm attempts to find the best alignment by minimizing the total distance of all the matched pairs:

$$D(X,Y) = \min_{\phi} \sum_{k=1}^{T} d\left(\phi_x(k), \phi_y(k)\right) \quad . \tag{1}$$

In practice, the Euclidian distance is commonly used. The solution is usually obtained using a dynamic programming algorithm.

To meet requirements that arise in practical applications, certain additional constraints can be included in the DTW algorithm. For example, a monotonicity constraint ensures that the data points in the query profile and those in the reference profiles have the same time sequence, and a symmetric continuity constraint ensures that all data points in the query profile are mapped into the reference profile and that no points are missed. In certain applications, constraints on the initial and final conditions are imposed to ensure that the start and end points of the two profiles are exactly aligned. In other cases such as the online monitoring of the growth profile in this study, only the initial points are constrained to be aligned; the final points are free. Some examples of DTW-aligned profiles are shown in Figure 5. Details of the DTW algorithm are given in Giorgino[21] and the references therein.

It should be noted that DTW is not invariant to location shifts of the profiles. Therefore, the mean of the profiles should be removed before alignment. Otherwise, the location and the mean patterns are confounded, which may adversely affect the alignment operation. Figure 5(a) shows the two profiles with unequal means (not adjusted for mean), Figure 5(b) shows the two profiles with the means removed, and Figures 5(c) and 5(d) show the aligned profiles without and with the adjustment for the mean, respectively. In Figure 5(c), the latter segment of the query profile is incorrectly mapped to the earlier segment of the reference profile. When the means of the two profiles are removed first, this problem is resolved, as can be observed in Figure 5(d).

In addition, the aforementioned treatment of historical profiles does not involve identifying OC profiles, which would require a specially designed algorithm. Therefore, here, we assume that all historical profiles are in control and focus on baseline profile calculation for Phase II online monitoring. The Phase I identification of abnormal profiles from a collection of unaligned ones is a topic that deserves future research efforts.

2.2. *Baseline profile calculation*

Given a set of aligned profiles, we must construct a baseline profile for online monitoring. A baseline profile will be an 'average' of multiple profiles and will best capture the variations in them. During online monitoring, the baseline profile will be used as the reference for comparison and deviation identification. However, the averaging operation cannot be performed if the profiles are not aligned because they have different lengths in time. Therefore, we will need to employ the aforementioned DTW algorithm to align the profiles first before calculating their average.

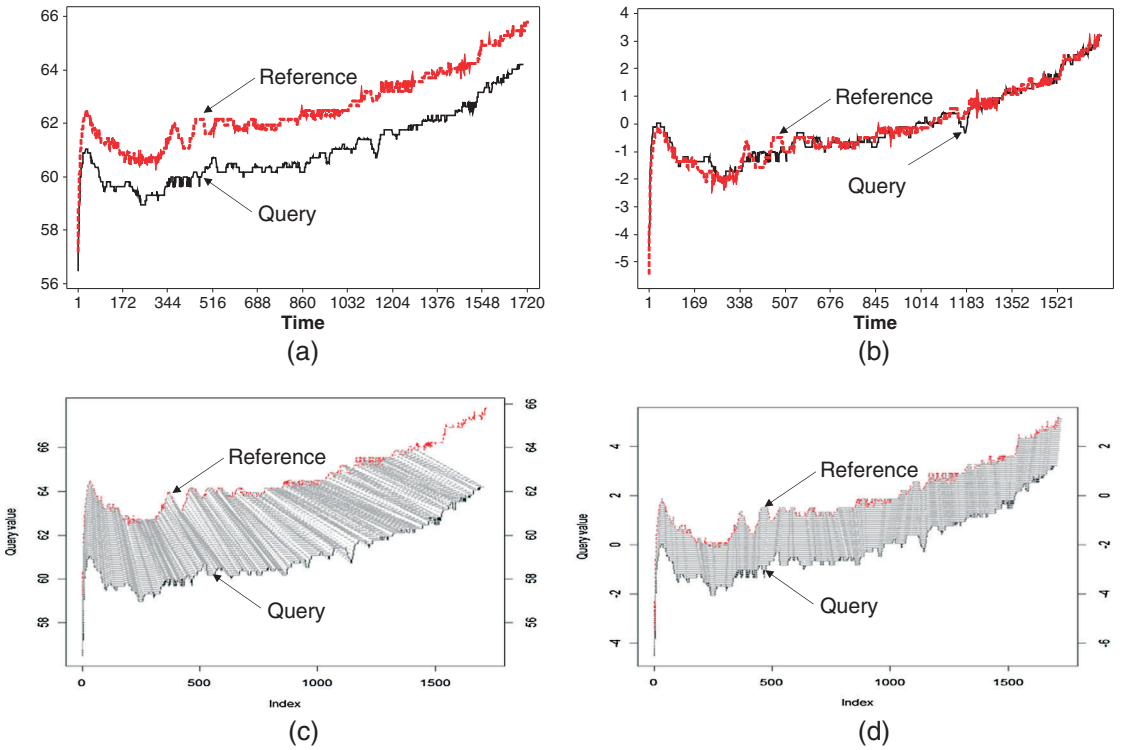

Figure 5. Profile alignment using DTW: (a) two profiles without mean adjustment, (b) two profiles with mean adjustment, (c) alignment of the unadjusted profiles, and (d) alignment of the mean-adjusted profiles.

To align multiple profiles, one additional challenge is the selection of the reference profile. Given the reference profile, all other profiles will be treated as queries and mapped to the reference profile. However, a different choice of the reference profile will lead to notably different alignment results. In this study, we propose the following iterative algorithm to identify a reference profile that minimizes the total distance:

Algorithm 1: Reference profile selection

Let P_i, $i = 1, \ldots, N$, denote N profiles.

Step 1. Choose P_i as the reference profile, and calculate $D(P_i, P_j)$, as shown in Eqn (1) for all $j \neq i$.

Step 2. Repeat step 1 for $i = 1, \ldots, N$. Choose profile P_i, where i is the argument of $\min_i \sum_{j \neq i} D(P_i, P_j)$, as the reference profile.

For a collection of N profiles, the best choice for the reference profile is the profile that is nearest the 'center' of the samples. In step 1, all other profiles are adjusted and mapped to P_i. After alignment, the total distance of all the profiles aligned to P_i is calculated, and this distance is a measure of the deviation of the samples from the reference P_i. In step 2, we can obtain N deviation measures $D(P_i, P_j)$, $i = 1, \ldots, N$. The profile with the smallest deviation is the closest profile to the 'center' of the samples, and this profile is chosen as the final reference profile for the averaging calculation. The search procedure is graphically illustrated in Figure 6.

After the reference profile is identified and all other profiles are aligned to it, we have a collection of time-aligned samples. For each time index t, the corresponding mean m_t and the standard deviation s_t can be estimated from the N-aligned profiles. The estimated m_t will be treated as the baseline profile.

2.3. *Online profile alignment and monitoring*

During online process monitoring, the profile trajectory is compared with the baseline profile, and an alarm is triggered if a significant deviation is detected. Because the length of the profile trajectory increases with time, the time alignment should be performed before online monitoring. In addition, the profile trajectory is always incomplete during online monitoring until the process finishes. Thus, when the query profile (the trajectory) is being aligned to the reference profile (the baseline profile), we force the query profile and the reference profile to have the same starting point but leave the final point free. In this manner, the partially observed profile trajectory can be aligned with the first part of the reference profile. Detection of abnormalities will be based on the aligned segment only.

Let the profile trajectory after alignment be y_t, $t = 1, \ldots, n$, and we assume that each point y_t follows a normal distribution with an unknown but dynamically changing mean and standard deviation. Because the variation in the in-control samples also varies

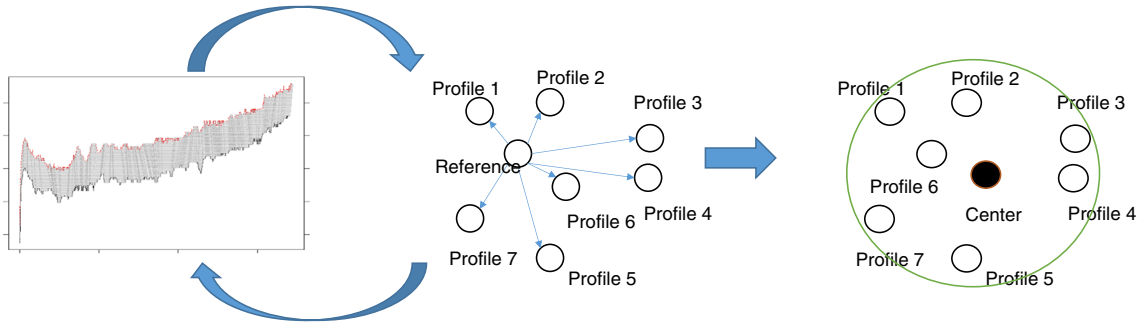

Figure 6. Distance calculation and choice of the reference profile.

over time, we calculate the difference between the online profile trajectory and the baseline profile and standardize the former using the time-varying variation:

$$y_t^{'} = (y_t - m_t)/s_t. \tag{2}$$

Hence, all points on the residual profile follow the standard normal distribution $y_t^{'} \sim N(0,1)$.

It is worth noting that in this work, we apply no limitations on the form of the baseline profile. It is assumed that the baseline profile is completely captured by m_t and s_t. Thus, the standardization and normally distributed trajectory in Eqn (2) could be monitored using existing SPC algorithms. All points on the standardized trajectory are assumed to be independent. If not, which is highly possible if the trajectory is observed with a high frequency, we suggest removing autocorrelation by fitting a time series model and monitor the uncorrelated residuals. Alternatively, SPC algorithms for autocorrelated processes[26,27] could be utilized.

Hawkins *et al.*[28] proposed a generalized likelihood ratio test (GLRT) to detect changes in a process using a change-point model. We apply this method and detect shifts in the residual profile using the following monitoring statistic:

$$T_{\max,n} = \max_j \left(\sqrt{\frac{j(n-j)}{n}} \frac{\overline{X}_{1j} - \overline{X}_{jn}}{s_{jn}} \right),$$

where

$$\overline{X}_{1j} = \sum_{t=1}^{j} y_t^{'} \Big/ j, \quad \overline{X}_{jn} = \sum_{t=j+1}^{n} y_t^{'} \Big/ (n-j), \quad s_{jn} = \sqrt{V_{jn}/(n-2)},$$

and

$$V_{jn} = \sum_{t=1}^{j} \left(y_t^{'} - \overline{X}_{1j}\right)^2 + \sum_{t=j+1}^{n} \left(y_t^{'} - \overline{X}_{jn}\right)^2.$$

Hawkins *et al.*[28] suggested that the control limits h_n should be the two-sided $\alpha/(n-1)$ fractile of a t-distribution with $n-2$ degrees of freedom, thus producing a test with a size of at most α, where α is a predefined false alarm rate. Therefore, an alarm is triggered in the residual profile if

$$\left| T_{\max,n} \right| > h_n. \tag{3}$$

If the charting statistic exceeds the control limits, we conclude that the process has shifted. As the process evolves, the monitoring statistic is evaluated when a new observation becomes available after the time alignment. In this manner, process shifts can be detected using the chart with incomplete online profile trajectories. Because this chart is constructed based on the DTW algorithm, we call it the DTW chart.

3. Performance study

In this section, we study the performance of the DTW chart and compare this method with another method. In the literature, there are few methods that can be directly used to monitor the profile trajectories that we discuss here. Zhu *et al.*[29] studied a similar problem and proposed an adaptive EWMA (AEWMA) chart to monitor profile trajectories. In the following texts, we briefly introduce the AEWMA chart and compare our proposed method with it.

The AEWMA chart was modified from existing methods to suit the time-varying feature of the growth profiles. Capizzi and Masarotto[30] proposed an AEWMA chart to monitor a univariate process with a constant mean. Because the growth profile has a

time-dependent mean trend, AEWMA uses the AEWMA algorithm of Capizzi and Masarotto[30] to capture the dynamic mean behavior of the profile as follows:

$$\hat{\mu}_t = (1 - w(e_t))\hat{\mu}_{t-1} + w(e_t)y_t,$$

where $\hat{\mu}_t$ is an estimate of the process mean at time t, $w(e_t) = \phi(e_t)/e_t$ is a weighting function, and $e_t = x_t - \hat{\mu}_{t-1}$ can be considered as the prediction error. The weighting function adaptively changes with the prediction error in the following ways:

$$\phi(e_t) = \begin{cases} e_t + (1-\lambda)k & \text{if } e_t < -R \\ \lambda e_t & \text{if } |e_t| < R \\ e_t - (1-\lambda)k & \text{if } e_t > R \end{cases},$$

where λ is a smoothing parameter and R is determined according to the variation of the profile. If the magnitude of the prediction error is less than R, then the performance of AEWMA is similar to that of conventional EWMA; if the magnitude of the prediction error is greater than R, then the effective smoothing parameter is chosen to be larger than λ. Therefore, it is expected that the AEWMA chart can capture both slow and rapid changes in the growth profile and provide a reasonable smoothing of the profile.

The adaptive chart proposed by Capizzi and Masarotto[30] is based on the assumption that the variation in the process is fixed. However, the variation in the profile trajectories changes over time. Therefore, with our AEWMA chart, we update the standard deviation of the profile using a method developed by MacGregor and Harris,[31]

$$\hat{\sigma}_t = \sqrt{(1-\gamma)\hat{\sigma}^2_{t-1} \times + \gamma(y_t - \mu_{t-1})^2}.$$

where γ is a smoothing parameter that determines the impact of historical observations on estimating the standard deviation.

Unlike the DTW chart, the AEWMA chart estimates the trend of the profile trajectories using the current profile only. The baseline mean and standard deviation of the profile trajectory, $\hat{\mu}_t$ and $\hat{\sigma}_t$, are all recursively updated using online points. No historical profiles are required to build the AEWMA control chart.

Finally, with the AEWMA chart, one monitors the growth profile with control limits

$$\begin{cases} UCL = \mu_t + h^* \hat{\sigma}_t \\ CL = \mu_t \\ LCL = \mu_t - h^* \hat{\sigma}_t \end{cases}, \tag{4}$$

where h^* controls the width of the control limits and is identified through simulations for achieving a predetermined in-control ARL value.

3.1. Simulation settings

As previously mentioned, the profiles have certain dynamics, but the trajectories differ from one profile to another. Thus, we design a two-stage model to mimic the profile trajectories of the growth process. The first stage is from 0 to $2\pi\omega$ min, a complete cycle of a sinusoidal function. The model used is

$$y_t = K + a\sin\left(\frac{t}{\omega}\right) + \varepsilon_{1t}, \quad 0 < t \le 2\pi\,\omega,$$

where $\varepsilon_{1t} \sim N(0, \sigma_t^2)$. The second stage is from $(2\pi\,\omega + 1)$ to $(2\pi\,\omega + 30/b)$ min. The model used is

$$y_t = y_{2\pi\,\omega} + b(t - 2\pi\,\omega) + \varepsilon_{2t}, \qquad 2\pi\,\omega < t < 2\pi\,\omega + 30/b,$$

where $\varepsilon_{2t} \sim N(0, \sigma_t^2)$. In this setting, the overall trends of the profiles are similar, whereas the profiles can be extended or compressed by changing the parameters.

To simulate a misalignment of the profiles, the values of the four parameters of the model, K, α, ω, and b, are randomly generated from uniform distributions using the values shown in Table I. To simulate time-varying variations, the standard deviation σ_t was made a time-varying function. Figure 7 shows five in-control profiles generated from the aforementioned two-stage algorithm. Note that the profiles are not aligned by default, so the misalignments should not be treated as process failures.

To assess the performance of the proposed method, we studied four types of failure patterns: (i) a sudden shift in the mean in the second stage at $\tau_1 = 35$ min; (ii) a gradual drift (a mean shift with an increasing magnitude) in the second stage after $\tau_2 = 45$ min; (iii) a constant cyclical shift after $\tau_1 = 35$ min with an added signal $\delta_{t>\tau_1} = A\sin[(t-\tau_1)\pi/\omega]$ and a fixed magnitude; and (iv) an increasing cyclical shift after $\tau_1 = 35$ min with an added signal $\delta_{t>K_1} = A \times e^{0.05\times(t-\tau_1)} \times \sin[(t-\tau_1)\pi/\omega]$, which has an amplitude that increases with time after $\tau_1 = 35$ min. The magnitude of the failure increases with time, as shown in Table II. The shift patterns are also plotted in Figure 8. The sudden mean shift and the gradual drift are common failure patterns considered in conventional SPC. The cyclical shifts are used to simulate the dynamic changes caused by the vibrations in the process, which are commonly observed in complex engineering processes.[32,33]

Table I. Parameter settings to generate in-control profiles

Parameters	Effects	Values
K	Position in y-axis	Uniform (0,3)
α	Amplitude of sinusoidal function	Uniform (5,10)
ω	Length in time of the sinusoidal function	Uniform (5,7)
b	Slope of the linear function	Uniform (0.5,1)
σ_t	Standard deviation of the profile	$\sigma_t = e^{-0.01t}$

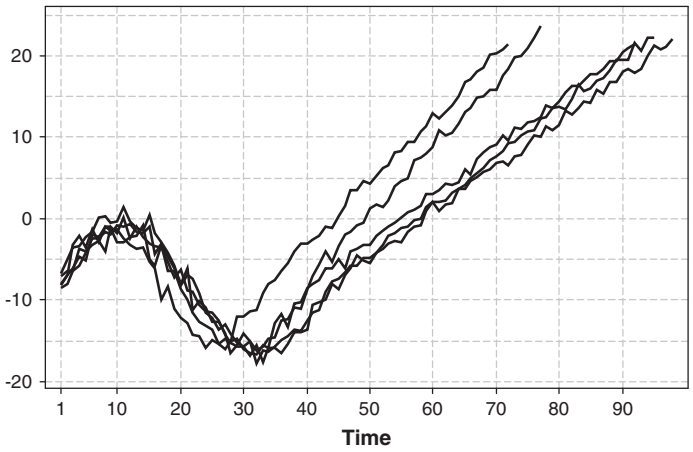

Figure 7. Plots of simulated in-control profiles.

Table II. The settings of the failure feature

	Shift pattern	Severity of failure		
Sudden shift	$K \rightarrow K + \delta_K,\ t > \tau_1$	$\delta_K = 1$	$\delta_K = 5$	$\delta_K = 11$
Gradual drift	$b \rightarrow b + \delta_b,\ t > \tau_2$	$\delta_b = 0.2$	$\delta_b = 0.4$	$\delta_b = 0.6$
Constant cyclical shift	$\delta = A\sin[(t-\tau_1)\pi/\omega],\ t > \tau_1$	$A = 1, \omega = 8$	$A = 3, \omega = 8$	$A = 5, \omega = 8$
		$A = 3, \omega = 2$	$A = 3, \omega = 4$	$A = 3, \omega = 16$
Increasing cyclical shift	$\delta = Ae^{0.05\times(t-\tau_1)}\sin[(t-\tau_1)\pi/\omega],\ t > \tau_1$	$A = 0.3, \omega = 8$	$A = 0.5, \omega = 8$	$A = 1.0, \omega = 8$

3.2. *Control chart implementation*

Based on the previous models and settings, we implemented the proposed DTW chart to monitor the simulated processes and compared its performance with that of the AEWMA chart. First, 20 profiles were generated using the parameters given in Table I without adding any shifts. Then, these profiles were aligned using DTW. A baseline profile was calculated using the methods introduced in Subsections 2.1 and 2.2. Subsequently, online profiles were generated at each time step. When a new point on the online profile became available, the incomplete profile trajectory was mapped with respect to the baseline profile using DTW. Then, the aligned profile was standardized using Eqn (2). We found that the residual profile has strong autocorrelation, which is adequately characterized by an AR(1) time series model. Therefore, an AR(1) model was fitted to the residual sequence to remove the autocorrelation; the uncorrelated residuals were monitored using the chart in Eqn (3).

The AEWMA chart defined in Eqn (4) was also applied to the same set of simulated profiles for comparison. To make the in-control average run length (ARL) of the two control charts identical, the parameter values for the AEWMA chart were $\lambda = 0.4$, $\gamma = 0.01$, $k = 1$, and $h = 2.97$.

3.3. *Performance comparison*

The ARL is widely used for evaluating the performance of a control chart. In traditional profile monitoring, each profile is an individual sample, and the ARL is the number of profiles inspected before an alarm is triggered. However, in monitoring profile trajectories, the

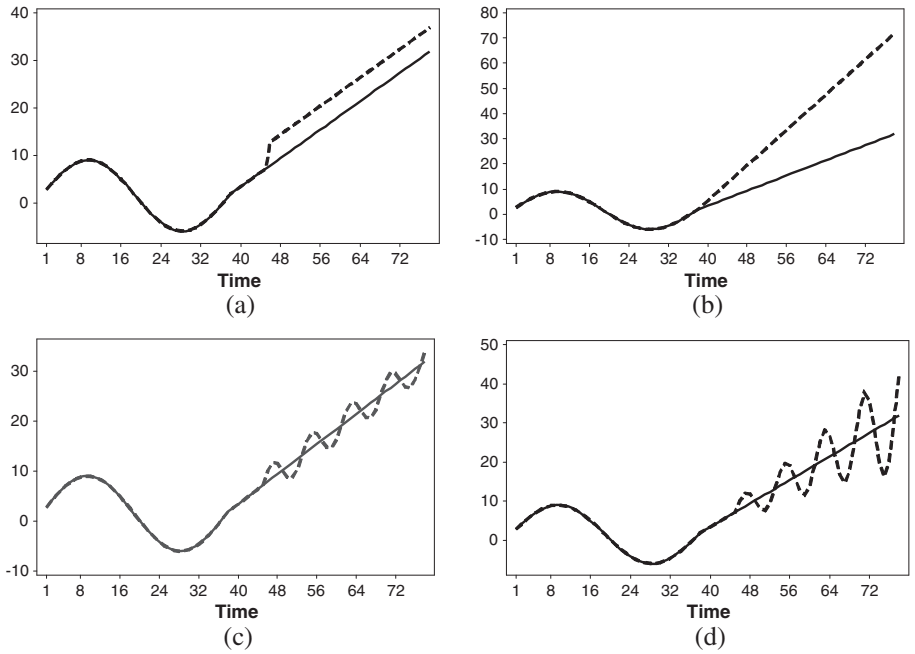

Figure 8. Plots of out-of-control patterns: (a) sudden shift, (b) gradual drift, (c) constant cyclical shift, and (d) growing cyclical shift.

charting statistic is evaluated at each step when a new point on the profile becomes available. In this situation, the ARL metric cannot be used directly.

Therefore, in this study, we used the true alarm rate and the false alarm rate to determine the sensitivity and accuracy of the charts. In addition, we calculated the average detection delay, which is defined as the number of steps that a profile continues (starting from the change point) before an alarm is triggered, as one indicator for evaluating the performance. If a shift signal is added to the growth profile but an alarm is triggered before the change point, the alarm is treated as a false alarm. If the profile terminates before an alarm is triggered, this sample is simply ignored when calculating either the false alarm rate or the average detection delay.

In fact, because the process may terminate before an alarm is triggered, the average delay could be considered as a sample from a truncated distribution, which would allow a more accurate calculation of the average delay. This approach presents other problems involving parameter estimation using truncated samples and difficulties in comparing performance, so we simply ignored those profiles without any alarms in calculating the average delay. In this manner, the average delay is not an accurate reflection of the true delay that the chart may have. As both charts were evaluated using the same standard, the average delay is nevertheless a valid measure for a fair performance comparison.

To calculate the profile-wise charting performance, we simulated 5000 profiles for each shift pattern and calculated the number of alarms. The alarm rates are calculated as the ratio of the number of true/false alarms and the total number of profiles. When a profile triggered an alarm, the process was stopped.

The simulation results are summarized in Table III. It can be observed that the DTW chart and the AEWMA chart have nearly equal level of alarm rates when the process is in control. The AEWMA chart has a smaller average detection delay than the DTW chart, which implies that with the AEWMA chart, more signals occur in the earlier stage of the growth process. As the OC signals are added to the process in a later stage of the process and alarms triggered before the occurrence of the change point are treated as false alarms, we see that the AEWMA chart also has a higher false alarm rate than the DTW chart when the process becomes OC. In addition, as increasing shift magnitude does not affect those observations collected before the change point, the false alarm rates are not affected by shift magnitude.

A number of observations can be made regarding OC processes:

1. When the process undergoes a sudden shift in the mean, the AEWMA chart reports more true alarms and exhibits a shorter average detection delay than the DTW chart, which suggests that the AEWMA chart is better at detecting a sudden mean shift. However, in this case, the AEWMA chart also produces a large proportion of false alarms.
2. When the process has a gradual drift in the mean, the DTW chart detects more OC profiles with a lower false alarm rate than the AEWMA chart. However, because the DTW chart usually requires more observations to detect a failure, it has a longer average delay time before triggering an alarm than the AEWMA chart.
3. For cyclical shifts with constant amplitudes, we can make several observations. For the case with $\omega = 8$, the DTW chart is superior to the AEWMA chart with a higher true alarm rate and lower false alarm rate, but the DTW chart exhibits greater detection delay. The detection accuracy increases when the shift magnitude increases from 1 to 5. However, for the same magnitude ($A = 3$), if the frequency of the sine wave is sufficiently high ($\omega = 2$) or low ($\omega = 16$), the performance of the DTW chart deteriorates. If the frequency is exceedingly low, the sine wave has a long period, and the failure signal becomes similar to a sustained drift. If the frequency is sufficiently high, a valid signal may be incorrectly treated as noise in the DTW alignment algorithm.

Table III. Performance comparison

		True alarm rate		False alarm rate		Average detection delay	
		DTW	AEWMA	DTW	AEWMA	DTW	AEWMA
In control			—	2.0	2.0	46	25.8
Sudden shift	$\delta_K = 1$	3.4	3.1	0.9	1.7	24.1	3.5
	$\delta_K = 5$	58.2	97.8	0.9	1.7	4.2	1.6
	$\delta_K = 11$	98.5	98.3	0.9	1.7	1.6	1.0
Gradual drift	$\delta_b = 0.2$	36.0	3.2	1.0	2.0	23.3	1.2
	$\delta_b = 0.4$	63.8	40.3	1.0	2.0	11.4	1.1
	$\delta_b = 0.6$	69.8	67.0	1.0	2.0	4.4	1.1
Constant cyclical shift	$A = 1, \omega = 8$	6.4	1.6	0.9	1.7	24.0	5.3
	$A = 3, \omega = 8$	63.9	36.2	0.9	1.7	16.8	7.2
	$A = 5, \omega = 8$	95.8	82.1	0.9	1.7	11.0	7.5
	$A = 3, \omega = 2$	31.4	87.3	0.9	1.7	17.2	5.4
	$A = 3, \omega = 4$	75.7	62.1	0.9	1.7	10.9	8.5
	$A = 3, \omega = 16$	27.9	34.6	0.9	1.7	21.2	7.3
Increasing cyclical shift	$A = 0.3, \omega = 8$	34.4	2.9	0.9	1.7	41.6	53.4
	$A = 0.5, \omega = 8$	64.3	26.1	0.9	1.7	37.2	48.8
	$A = 1.0, \omega = 8$	93.0	73.9	0.9	1.7	26.8	35.5

4. For cyclical shifts with increasing amplitudes, the DTW chart is uniformly better than the AEWMA chart and also has shorter average detection delays. This fact shows that the DTW chart is more sensitive if the process becomes unstable with increasing levels of vibration.

Notably, the average detection delay is calculated using profiles with alarms only. Therefore, the true and false alarm rates reflect the performance of the charts on the profile level, and the average delay reflects the performance of the charts given that an alarm has been triggered. Therefore, a chart with a higher true alarm rate and a lower false alarm rate is recommended, even though the detection delay is greater.

In summary, the AEWMA chart is more successful in detecting a sudden mean shift, whereas the DTW chart can detect gradual drifts and cyclical shifts more rapidly in most cases. This result can be explained by the fact that the AEWMA chart adjusts to the dynamic profile from its past observations using EWMA smoothing, and a sudden shift or a drift signal that cannot be smoothed by EWMA smoothing is easily detected. However, as shown in Figure 5, the DTW chart can be easily distorted by sudden mean shifts. In other words, sudden mean shifts may be incorrectly treated as an increasing trend in the growth profile by the alignment algorithm. However, the DTW alignment algorithm is not adversely affected by the cyclical signals, and it is therefore more sensitive to these shifts.

4. Application to an ingot growth process

In this section, we use the DTW chart and the AEWMA chart to analyze heating power profiles collected from ingot growth cycles. Ten historical conforming profiles are used to estimate the baseline profile for the DTW chart (refer to Figures 2 and 3 for plots of sample profiles). In online monitoring, the current trajectory must be aligned with the baseline profile at each step, as a new observation becomes available. To save computational time, we sampled one point every 10 steps from the raw data for the demonstration.

Both charts were set to have a false alarm rate of 0.01. The parameters used for the AEWMA chart were $\lambda = 0.4$, $\gamma = 0.005$, $k = 0.05$, and $h = 3$. In practice, an autoregressive model is used to remove the autocorrelation in the residual profile before implementing the DTW chart. Both charts used some initial observations (15 data points) for warm-up.

Figure 9 shows two profiles, one of which was considered conforming by engineers and the other profile considered nonconforming. These profiles were monitored using the two control charts, the behaviors of which are shown in Figure 10. It can be observed that neither chart triggered any alarms for the conforming profile. For the nonconforming profile, the DTW chart triggered an alarm at step 142, and the AEWMA chart triggered an alarm at step 141. The one-step delay of the DTW chart was attributed to the AR(1) model fitted to the residuals before calculating the charting statistics. It is evident from Figure 9 that the OC profile was smooth before around step 140, but after that, a very large deviation from the process mean occurred. This shows that the power became unstable at that time. However, because this abnormal change was not identified and compensated for by the operator in practice, the process deteriorated and failed quickly. This example additionally shows the importance of timely detection of process changes in a long-duration process.

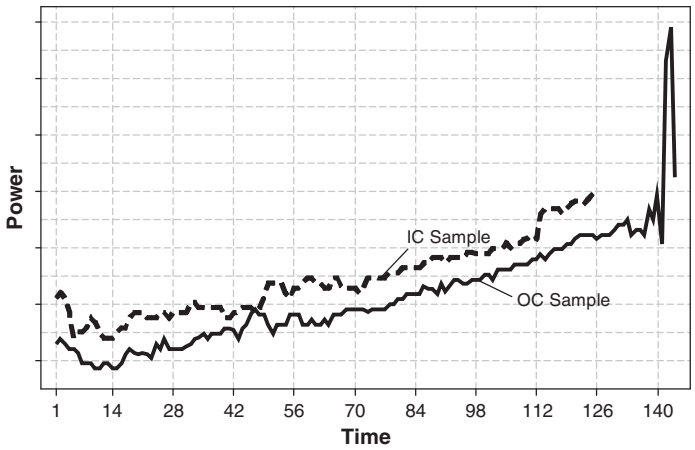

Figure 9. Sample in-control (IC) and out-of-control (OC) profiles.

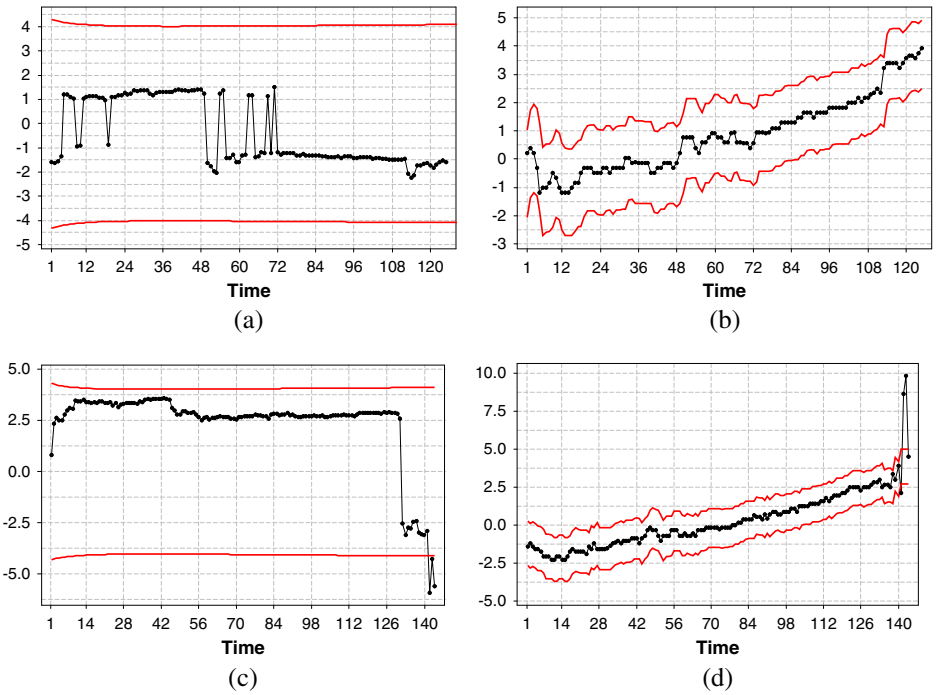

Figure 10. Monitoring of two power profiles: (a) DTW for an in-control profile, (b) AEWMA for an in-control profile, (c) DTW for an out-of-control profile, and (d) AEWMA for an out-of-control profile.

5. Conclusions

In our study, we focused on the monitoring of profile trajectories, which are time unaligned with finite but unequal lengths and are incomplete during online monitoring. Therefore, conventional SPC cannot be directly applied for online monitoring.

In this study, we proposed a method for monitoring such growth profiles with DTW-based alignment. In the proposed method, a baseline profile is calculated from aligned historical profiles. During online monitoring, incomplete profile trajectories are aligned with the baseline profile, and then, the GLRT statistic, which is derived from change-point theory, is evaluated to detect OC situations. Compared with the modified AEWMA chart, the proposed DTW chart is less sensitive to sudden mean shifts, but it has better performance in detecting gradual drifts and cyclical dynamic shifts, which are frequently observed in complex engineering processes.

This study focused on the monitoring of profile trajectories given a set of historical profiles; the baseline profile is cleated from the known in-control profiles. The identification of abnormal profiles from a historical dataset was not discussed here but is worthy of research in the future. In addition, for baseline profile calculations, an exhaustive search algorithm would require considerable computational power if the number of historical profiles is large. A computationally more efficient algorithm for baseline profile calculations would be an interesting goal for future research.

We observed that the DTW chart is less powerful than the AEWMA chart in detecting sudden mean shifts. The main reason is that a mean shift may not be properly addressed in the DTW alignment algorithm. Therefore, a more robust DTW algorithm designed specifically for SPC is a topic that deserves more research. In addition, because engineering processes are often characterized by multiple growth profiles, the monitoring of these processes using multiple misaligned profiles is also an important topic for researchers.

Acknowledgements

We are grateful to the anonymous referees for their valuable comments, which have helped us improve this paper greatly. This work was supported by the National Natural Science Foundation of China under Grant No. 71072012 and the Tsinghua University Initiative Scientific Research Program.

References

1. Stoumbos ZG, Reynolds Jr MR, Ryan TP, Woodall WH. The state of statistical process control as we proceed into the 21st century. *Journal of the American Statistical Association* 2000; **95**:992–998.
2. Montgomery DC. Introduction to Statistical Quality Control (5th edn.). John Wiley: Hoboken, N.J., 2005; 759.
3. Woodall WH, Spitzner DJ, Montgomery DC, Gupta S. Using control charts to monitor process and product quality profiles. *Journal of Quality Technology* 2004; **36**:309–320.
4. Noorossana R, Saghaei A, Amiri A. Statistical Analysis of Profile Monitoring. John Wiley & Sons: Hoboken, NJ, 2011.
5. Wang K, Tsung F. Using profile monitoring techniques for a data-rich environment with huge sample size. *Quality and Reliability Engineering International* 2005; **21**:677–688.
6. Williams JD, Woodall WH, Birch JB. Statistical monitoring of nonlinear product and process quality profiles. *Quality and Reliability Engineering International* 2007; **23**:925–941.
7. Kang L, Albin S. On-line monitoring when the process yields a linear profile. *Journal of Quality Technology* 2000; **32**:418–426.
8. Colosimo BM, Semeraro Q, Pacella M. Statistical process control for geometric specifications: on the monitoring of roundness profiles. *Journal of Quality Technology* 2008; **40**:1–18.
9. Jones M, Rice JA. Displaying the important features of large collections of similar curves. *The American Statistician* 1992; **46**:140–145.
10. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics* 1995; **37**:41–59.
11. Jeong MK, Lu JC, Huo XM, Vidakovic B, Chen D. Wavelet-based data reduction techniques for process fault detection. *Technometrics* 2006; **48**:26–40.
12. Walker E, Wright SP. Comparing curves using additive models. *Journal of Quality Technology* 2002; **34**:118–129.
13. Zou CL, Tsung FG, Wang ZJ. Monitoring profiles based on nonparametric regression methods. *Technometrics* 2008; **50**:512–526.
14. Del Castillo E, Colosimo BM. Statistical shape analysis of experiments for manufacturing processes. *Technometrics* 2011; **53**:1–15.
15. Rengaswamy R, Venkatasubramanian V. A syntactic pattern-recognition approach for process monitoring and fault diagnosis. *Engineering Applications of Artificial Intelligence* 1995; **8**:35–51.
16. Rengaswamy R, Hägglund T, Venkatasubramanian V. A qualitative shape analysis formalism for monitoring control loop performance. *Engineering Applications of Artificial Intelligence* 2001; **14**:23–33.
17. Venkatasubramanian V, Rengaswamy R, Kavuri SN, Yin K. A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering* 2003; **27**:327–346.
18. Krieger AM, Pollak M, Yakir B. Surveillance of a simple linear regression. *Journal of the American Statistical Association* 2003; **98**:456–469.
19. Frisen M, Wessman P. Evaluations of likelihood ratio methods for surveillance: differences and robustness. *Communications in Statistics - Simulation and Computation* 1999; **28**:597–622.
20. Myers C, Rabiner L. A comparative study of several dynamic time-warping algorithms for connected word recognition. *The Bell System Technical Journal* 1981; **60**:1389–1409.
21. Giorgino T. Computing and visualizing dynamic time warping alignments in r: The dtw package. *Journal of Statistical Software* 2009; **31**:1–24.
22. Gupta A, Samanta A, Kulkarni B, Jayaraman V Fault diagnosis using dynamic time warping. In: Pattern Recognition and Machine Intelligence, Ghosh A, De RK, Pal SK (eds). Springer: Berlin Heidelberg, 2007; 57–66.
23. Dai Y, Zhao J. Fault diagnosis of batch chemical processes using a dynamic time warping (DTW)-based artificial immune system. *Industrial & Engineering Chemistry Research* 2011; **50**:4534–4544.
24. Nelson BJ, Runger GC. Predicting processes when embedded events occur: dynamic time warping. *Journal of Quality Technology* 2003; **35**:213–226.
25. Kassidas A, MacGregor JF, Taylor PA. Synchronization of batch trajectories using dynamic time warping. *AIChE Journal* 1998; **44**:864–875.
26. Montgomery DC, Mastrangelo CM. Some statistical process control methods for autocorrelated data. *Journal of Quality Technology* 1991; **23**:179–193.
27. Apley DW, Shi J. The GLRT for statistical process control of autocorrelated processes. *IIE Transactions* 1999; **31**:1123–1134.
28. Hawkins DM, Qiu PH, Kang CW. The changepoint model for statistical process control. *Journal of Quality Technology* 2003; **35**:355–366.
29. Zhu L, Dai C, Sun H, Li W, Jin R, Wang K. Curve monitoring for a single-crystal ingot growth process. Department of Industrial Engineering, Tsinghua University, 2014. Avaliable at http://www.ie.tsinghua.edu.cn/kbwang/
30. Capizzi G, Masarotto G. An adaptive exponentially weighted moving average control chart. *Technometrics* 2003; **45**: 199–207.
31. MacGregor J, Harris T. The exponentially weighted moving variance. *Journal of Quality Technology* 1993; 25:106–118.
32. Wang K, Tsung F. Monitoring feedback-controlled processes using adaptive T^2 schemes. *International Journal of Production Research* 2007; **45**:5601–5619.
33. Wang K, Tsung F. An adaptive T^2 chart for monitoring dynamic systems. *Journal of Quality Technology* 2008; **40**:109–123.

Authors' biographies

Chenxu Dai is a graduate student in the Department of Industrial Engineering, Tsinghua University, Beijing, China. He received both his BS and MS degrees in Industrial Engineering from Tsinghua University, Beijing, China. His research focuses on the statistical modeling, monitoring, and control of engineering systems.

Dr. Kaibo Wang is an associate professor with the Department of Industrial Engineering, Tsinghua University, Beijing, China. His research is devoted to statistical quality control and data-driven complex system modeling, monitoring, diagnosis, and control, with a special emphasis on the integration of engineering knowledge and statistical theories for solving problems from the real industry. Dr. Wang is a senior member of ASQ and a member of INFORMS and IIE.

Dr. Ran Jin is an assistant professor at the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received his PhD degree in Industrial Engineering from Georgia Tech in 2011, his master's degree in Industrial Engineering (2007) and in Statistics (2009), both from the University of Michigan, and his bachelor's degree in Electronic Engineering from Tsinghua University in 2005. His research interests are in engineering-driven data fusion for manufacturing system modeling and performance improvements, such as the integration of data mining methods and engineering domain knowledge for multistage system modeling and variation reduction, and sensing, modeling, and optimization based on spatial correlated responses. He is a member of INFORMS, IIE, and ASME.