Taylor & Francis
Taylor & Francis Group

Monitoring feedback-controlled processes using adaptive T^2 schemes

KAIBO WANG† and FUGEE TSUNG*‡

†Department of Industrial Engineering, Tsinghua University, Beijing 100084, P. R. China
‡Department of Industrial Engineering and Logistics Management, The Hong Kong University of
Science and Technology, Kowloon, Hong Kong

The integration of engineering process control (EPC) and statistical process control (SPC) has drawn wide attention because of its capability to minimize long-term process variation and improve product quality. However, dynamic shift patterns are usually observed when feedback-control loops exist, which consequently lead to poor charting performance. In this paper, the dynamic patterns in mean shifts of proportional integral (PI) controlled and minimum mean square error (MMSE) controlled processes are formulated and analysed. In conjunction with a model-free forecasting algorithm, an adaptive T^2 charting procedure is applied to improve detection power. This adaptive procedure is constructed upon the uniformly most powerful (UMP) unbiased test with the consideration of time-varying patterns. Monte Carlo simulations have shown that the proposed strategy possesses significantly improved detection sensitivity over intended shift ranges.

Keywords: Adaptive chart; Automatic process control; Exponentially weighted moving average; Process dynamics; Statistical process control

1. Introduction

The necessity to apply statistical process control (SPC) to feedback-controlled processes has been widely recognized (Tucker *et al*. 1993, Capilla *et al*. 1999, Tsung 2000, Gultekin *et al*. 2002). Typically, SPC techniques are used to monitor process running status and give early alarms when assignable causes are identified, while engineering process control (EPC) is a strategy to keep a process at its designated level and make adjustment to controllable factors if any deviations are observed. Recently, substantial efforts have been devoted to the integration of these two techniques (Tsung *et al*. 1999, Tsung and Shi 1999) because of the evidence that although stationary disturbance of a process can be compensated by certain feedback control laws, SPC is still essential for long-term process stability assurance and assignable cause discovery.

In the integrated SPC/EPC framework, two adverse impacts on SPC are introduced by the feedback control laws of EPC: one is the "window of

*Corresponding author. Email: season@ust.hk

opportunity,'' and the other is the dynamic shift. ''Windows of opportunity'' refers to the short interval immediately after the occurrence of a process fault within which a signal can be detected. Once the opportunity is missed, the process shows little or no hint about the process abnormality, and it becomes difficult for a SPC chart to perceive this failure (Apley and Shi 1999, Tsung and Apley 2002). The root cause of this phenomenon is the compensatory action generated by the feedback controller when a process fault occurs. Consequently, monitoring the process output using conventional SPC charts cannot gain effective performance. Alternatively, some authors have suggested to monitoring the control action (Faltin and Tucker 1991, Box and Kramer 1992, Capilla *et al.* 1999), which is believed to carry more sustainable and useful information about the failure. Tsung and Tsui (2003) compared the performance of monitoring the output and the input streams in minimum mean squared error (MMSE) controlled processes and concluded that the performance of different charts depends heavily on the parameters of the process being monitored. No uniformly powerful chart is found. Recent research by Tsung *et al.* (1999) has shown that the joint monitoring of both the output and the input streams using a bivariate chart outperforms many conventional SPC approaches on a single variable. This method was further extended by Tsung and Apley (2002), in which lagged historical observations were taken into account for monitoring and the resulting chart was shown to be sensitive to both mean shifts and process parameter changes.

Nevertheless, the window of opportunity issue is only part of the challenges caused by the feedback controller. Due to the continuous adjustment action, a typical level shift, which otherwise would have stayed at a constant value, would keep changing over time. A dynamic shift pattern is therefore observed. As most conventional SPC schemes are designed for detecting constant level shifts, their performance has been seriously deteriorated by the dynamic pattern.

Clearly, the study of the dynamic shift patterns is essential to efficient chart design. Hu and Roan (1996) presented the dynamic shift patterns of time-series-based control charts, and Tsung and Tsui (2003) reported that similar patterns are observed in processes with MMSE controllers. In an attempt to monitor a feedback-controlled process, Jiang (2004) also noticed that the shift patterns of both MMSE-controlled and proportional integral- (PI) controlled processes are time-varying, and suggested that any efficient charting schemes should take dynamic shift information into consideration. However, the strategy to take advantage of dynamic shift information to improve monitoring sensitivity has not been widely studied. The T^2 chart proposed by Tsung *et al.* (1999) failed to consider pattern information, and the solution in Jiang (2004) took only part of the dynamics in chart design. These two methods will be reviewed in more detail in section 2. The generalized likelihood ratio test (GLRT) approach in Apley and Shi (1999) aims to consider the whole pattern by performing searches repeatedly in a window. However, this method requires the dynamic pattern being known exactly in the design phase, which may be difficult in practice. In addition, the GLRT procedure is developed for univariate process monitoring only.

The purpose of this paper is to address the issue of monitoring feedback-controlled processes with time-varying shifts in a multivariate environment. A newly emerged SPC scheme, namely the adaptive T^2 procedure by Wang and Tsung (2007), is investigated in this paper. The adaptive T^2 procedure is established upon the

uniformly most powerful (UMP) unbiased test, which utilizes an exponentially weighted moving average (EWMA)-based forecasting algorithm and possesses the highest detection power among all Shewhart-type T^2-based charts.

The rest of this paper is organized as follows. A review of conventional monitoring methods is presented in section 2. The time-varying shift patterns of feedback-controlled processes are analysed in section 3. In section 4, the adaptive T^2 procedure is presented. A new EWMA-based shift forecasting algorithm for oscillated processes is proposed. The performance adaptive T^2 when monitoring feedback – controlled processes is evaluated in section 5. The last section, section 6, concludes this paper with a summary of major findings and recommendations.

2. Review of SPC schemes for feedback-controlled processes

We start by considering a Run-to-Run (R2R) process with a single input and a single output. Without loss of generality, assume the target value of the process is zero. Let e_t be the measured deviation from the target, and x_{t-1} be the setting of the controllable factor at step $t-1$, which is usually generated by a feedback controller. In practice, e_t is the additive effect of two parts: output y_t that is dictated by the control action, and process disturbance d_t that is independent of the control action. We consider a simple dynamic model with $y_t = x_{t-1}$, where the output at step t, y_t, only depends on the input at step $t-1$. The measured output, e_t, can be modeled by the following equation:

$$e_t = y_t + d_t = x_{t-1} + d_t, \tag{1}$$

All variables in the model are subscripted by time stamps to indicate the sequential intervals at which the data are collected.

This model is rather physically representative and has been widely cited to illustrate chemical-mechanical polishing (CMP) processes, etching processes and epitaxial growth processes in semiconductor manufacturing and other discrete processes in related industries (see Sachs *et al.* 1995, DelCastillo and Hurwitz 1997, Fan *et al.* 2002, Jen *et al.* 2004, and the references therein).

There is a substantial discussion in the literature on choosing the disturbance model, d_t, for a dynamic process. Zhang and Pollard (1994) declared that in industries, it is common for the input stream to be described by a first-order autoregressive (AR) model. Nembhard and Kao (2003) and Hwarng (2004) also adopted AR(1) models in describing process disturbances. Montgomery *et al.* (2000) used a first-order integrated moving average (IMA) model to illustrate the input stream of a tank process. Tsung *et al.* (1998) assumed that process disturbance follows either a first order autoregressive moving average (ARMA) or autoregressive integral moving average (ARIMA) model. When the stochastic behaviour of process input, d_t, follows an ARMA(1,1) model, it can be written as:

$$d_t = \phi d_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, \tag{2}$$

where ε_t is an i.i.d. white nose series, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. When $\theta \to 0$ or $\phi \to 1$, the ARMA(1,1) process reduces to an AR(1) or an IMA(1,1) process, respectively.

A white noise series is obtained when $\phi = \theta$. Therefore, the ARMA(1,1) model is a useful and representative model for describing process disturbance. The rest of the derivations in this paper will be carried out based on the ARMA(1,1) model.

In order to maintain the process output on target, it is a common practice to set up engineering process controllers to regulate the system. Among others, two most popular industrial control schemes are PI controllers (Tsung *et al.* 1998) and MMSE controllers (Box *et al.* 1994). The general form of a PI controller is given by:

$$x_t = k_P e_t + k_I \sum_{k=0}^{t} e_k = k_P e_t + k_I \frac{1}{1 - B} e_t, \tag{3}$$

where B is a backshift operator such that $Bx_t = x_{t-1}$. The proportional part of the PI controller is sensitive to sudden large shifts in the output, while the integral part is capable of compensating and eliminating sustained shifts in quality characteristics and brings the process back to its target after a transient period. The selection of the parameters, k_P and k_I, is critical to the effectiveness of the PI controller. Tsung *et al.* (1998) introduced an optimal design procedure to minimize process variations. In this paper, their design guideline is followed to generate the optimal PI controllers for all processes under investigation.

The MMSE controller is another type of automatic process control procedure, which aims to compensate predicted output deviations, and leave only unpredictable residuals in the regulated process (Pandit and Wu 1993). Given the parameter settings of the ARMA(1,1) disturbance model in (2), the MMSE controller takes the form:

$$x_t = \phi x_{t-1} + (\theta - \phi) e_t. \tag{4}$$

Substituting (4) for the control action in (1) yields $e_t = \varepsilon_t$, which suggests that the controlled output of the process is merely a white noise series. The mean square error of output sequence is therefore minimized. Interested parties are referred to Tsung *et al.* (1998) for a comparison of the PI and the MMSE controllers.

Let $\mathbf{V}_t = [e_t, x_t]^T$ be the vector of the output and the input of the latest observation, and without loss of generality, assume the mean of $\mathbf{V}_t$, $\boldsymbol{\mu}$, is zero. Tsung *et al.* (1999) defined the dynamic T^2 chart as:

$$T^2 = \mathbf{V}_t^T \boldsymbol{\Sigma}^{-1} \mathbf{V}_t > h_1, \tag{5}$$

where $\boldsymbol{\Sigma}$ is the variance-covariance matrix of vector $\mathbf{V}_t$, h_1 is a control limit that achieves a desired in-control average run length (ARL). The computation procedure of $\boldsymbol{\Sigma}$ of a process under the PI controller can be found in Tsung and Shi (1999). In this paper, we have outlined the computation for $\boldsymbol{\Sigma}$ of MMSE-control processes, which is shown in the appendix.

The T^2 chart (5) is a directionally invariant chart since it has equal sensitivity for shifts with the same magnitudes and is not influenced by shift directions (Lowry and Montgomery 1995). However, the dynamic shift feature of a feedback-controlled process, which is believed to convey informative signals about process failures, is not considered by this scheme.

Hawkins (1991) indicated that in the monitoring of multivariate processes, if prior information regarding the future mean shift direction is known to be $\mathbf{d}$, the most powerful chart in detection shifts along this direction is given by:

$$T^2 = \mathbf{d}^T \mathbf{\Sigma}^{-1} \mathbf{V}_t > h_2, \tag{6}$$

The statistic in (6) is the UMP test among all unbiased tests for the statistical hypothesis: $H_0 : \boldsymbol{\mu} = 0$ versus $H_1 : \boldsymbol{\mu} = \delta \mathbf{d}$, where δ is a scalar.

In contrast to (5), control chart (6) is a directionally variant chart, since its detection power is dictated by both the magnitude and the direction of the shift. Shifts occurring along direction $\mathbf{d}$ will be detected by chart (6) much quicker than chart (5). If unexpectedly, shifts occur in other directions, this chart may be inferior to other charts, including (5).

Two special forms of chart (6) are found in Jiang (2004). Instead of monitoring the vector, $\mathbf{V}_t$, Jiang (2004) monitored the vector of the standardized output and the one-step lagged input, $\mathbf{Z}_t = [e_t/\sigma_e, x_{t-1}/\sigma_x]^T$, where σ_e and σ_x are the standard deviations of e_t and x_t, respectively. As the shift of $\mathbf{Z}_t$ keeps changing over time, Jiang (2004) identified two snapshots of the shift direction, which are believed to be important to fault detection. The first one, denoted by $\mathbf{d}_0$, is the direction of $\mathbf{Z}_t$ immediately after the occurrence of a shift, and the second one, denoted by $\mathbf{d}_\infty$, is obtained when the process enters its steady state. Consequently, two statistics can be generated based on (6):

$$U_0 = \mathbf{d}_0 \mathbf{\Sigma}_z^{-1} \mathbf{Z}_t, \tag{7}$$

and

$$U_\infty = \mathbf{d}_\infty \mathbf{\Sigma}_z^{-1} \mathbf{Z}_t, \tag{8}$$

where $\mathbf{\Sigma}_z$ is the covariance matrix of $\mathbf{Z}_t$.

The U_0 chart and the U_∞ chart in Jiang (2004) are obtained after simple manipulation on equations (7) and (8), respectively. Because of the specific directions the charts are designed for, the U_0 chart is sensitive to large shifts, while the U_∞ chart is sensitive to small shifts.

However, it is not difficult to understand that since the real process shift changes continuously over time, both the U_0 chart and the U_∞ chart are less optimal in the sense that the real shift direction is different from any of the designated ones. In addition, any single vector is not sufficient to represent a time-varying shift pattern. Therefore, we will study the dynamic shift patterns of feedback-controlled processes in the following section, and propose a forecasting method to improve the charting performance.

3. Time-varying shift patterns in feedback-controlled processes

The special responses generated by certain types of assignable causes, which are called the signatures of the corresponding failures, are essential to quick fault identification. The original purpose of applying an automatic controller is to

maintain the process output at a designated level. Any deviation from the target should trigger an action of the controller to drag the output back to the target by adjusting controllable factors. The side effect of the controller, however, is its impact on the fault signatures. Although the process shift is constant, the observed signature may show complex patterns. As a result, the fast detection of faults becomes more difficult. In this section, we will study how the usual process fault has been changed due to the PI or the MMSE controller, and provide guidelines for subsequent chart design.

One of the frequently encountered process failures is a mean shift, which may be caused by various reasons, such as a sudden change in raw material, the break of a machinery component, and the malfunctioning of a sensor. Suppose that a sustained mean shift of magnitude δ is introduced into process (1) starting from $t = 0$. The observed process output is the sum of the normal output and the shift, which takes the form:

$$e_t = \mu_t + x_{t-1} + d_t, \tag{9}$$

where

$$\mu_t = \begin{cases} 0 & t < 0 \\ \delta\sigma_d & t \geq 0. \end{cases} \tag{10}$$

The quantity σ_d is the standard deviation of d_t in (2) and $\sigma_d^2 = (1 - 2\phi\theta + \theta^2/1 - \phi^2)\sigma_\varepsilon^2$. In the following, the behaviour of the shifted process is studied with the presence of either PI or MMSE controller.

3.1 *Dynamic shift zones of PI-controlled processes*

We first apply PI controller (3) to shifted process (9). Substituting the controller for the control action, x_{t-1}, in process (9), and taking expectations on both sides of the equation yields:

$$\begin{aligned} E[e_t] = {}& (\phi + 1 + k_P + k_I)E[e_{t-1}] - (k_P + \phi + \phi k_P + \phi k_I)E[e_{t-2}] \\ & + \phi k_P E[e_{t-3}] + \mu_t - (\phi + 1)\mu_{t-1} + \phi\mu_{t-2}. \end{aligned} \tag{11}$$

In particular, the transient and steady state behaviour of e_t satisfy:

$$\begin{cases} E[e_0] = \delta\sigma_d \\ E[e_1] = (1 + k_P + k_I)\delta\sigma_d \\ \lim_{t\to\infty} E[e_t] \to 0. \end{cases} \tag{12}$$

By substituting the PI controller for e_t in (9), the following mean shift pattern of x_t is obtained:

$$\begin{aligned} E[x_t] = {}& (\phi + 1 + k_P + k_I)E[x_{t-1}] - (\phi + \phi k_P + k_P + \phi k_I)E[x_{t-2}] \\ & + \phi k_P E[x_{t-3}] + (k_P + k_I)\mu_t - (\phi k_P + \phi k_I + k_P)\mu_{t-1} + \phi k_P \mu_{t-2}, \end{aligned} \tag{13}$$

and the following equations regarding the transient and steady states of x_t hold:

$$\begin{cases} E[x_0] = (k_P + k_I)\delta\sigma_d \\ E[x_1] = (k_P^2 + k_I^2 + k_P + 2k_I)\delta\sigma_d \\ \lim_{t\to\infty} E[x_t] \to -\delta\sigma_d. \end{cases} \quad (14)$$

Equations (11) and (13) suggest that in the event of a sustained shift, both e_t and x_t exhibit certain types of dynamic patterns. In specific, e_t increases to δ immediately after the occurrence of the shift, while the signal reaches x_t one step later. When time goes to infinity, the information carried by e_t vanishes as e_t approaches zero, while the signal in x_t becomes more noticeable. Equation (12) also suggests that the PI controller is capable of eliminating sustained shifts by adjusting control factors.

In this section, processes with different parameter combination will be studied. Each process is equipped with a PI controller with parameters suggested by Tsung *et al.* (1998). Since with a PI controller, the steady state of e_t always approaches zero, processes can only be differentiated by their transient stage behaviour. In figure 1, the whole parameter space of the process is classified into four zones. In zone 1, as shown in figure 2(a), strong oscillations are observed. The mean of the process output varies alternatively from positive to negative, then back to positive again, until it enters its steady state. In figure 2(b), an example from zone 2 is illustrated. The oscillations in this area are much weaker, and last a very short period of time before the process becomes stable. Processes in zone 4 fall smoothly from a sudden shift, and approach their steady states gradually. A smoothed falling trend is clearly observed in figure 2(d). Zone 3 is analogous to the fourth zone, except for a weak spike in the falling stage. The trend is less smooth compared with the previous zone, as shown in figure 2(c).

The lower-right area of figure 1 is not labeled since the PI controller reduces to a pure P controller in this area. Consequently, x_t becomes proportional to e_t, the dynamic pattern will be much simplified and the conventional endeavors of process monitoring should be sufficient. In addition, a pure P controller is rarely used independently because of the existence of model estimation uncertainty. Therefore, we will consider the cases in which both the proportional term and the integral term of the PI controller are functioning in this paper.

3.2 *Dynamic shift zones of MMSE-controlled processes*

Showing time-varying shifts is not a unique feature of the PI controller. If instead, the process is adjusted by the MMSE controller, (4), the output of process (9) will take the form:

$$e_t = \mu_t \frac{1 - \phi B}{1 - \theta B} + \varepsilon_t. \quad (15)$$

The deterministic part of (15) follows an ARMA(1,1) model with parameter (ϕ, θ). Since ε_t is a white noise series, the mean of the process output is:

$$E[e_t] = \mu_t \left(1 + (\theta - \phi) \sum_{k=0}^{\infty} \theta^k B^{k+1} \right). \quad (16)$$
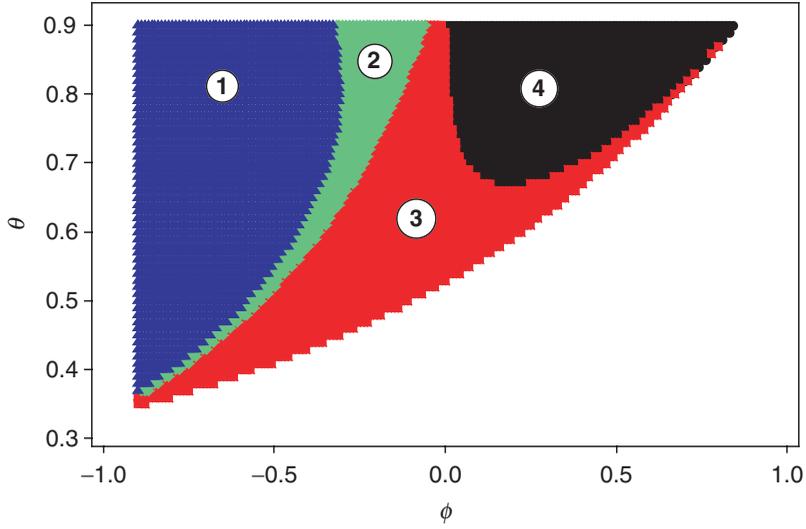
Figure 1. Zones of PI-control processes.
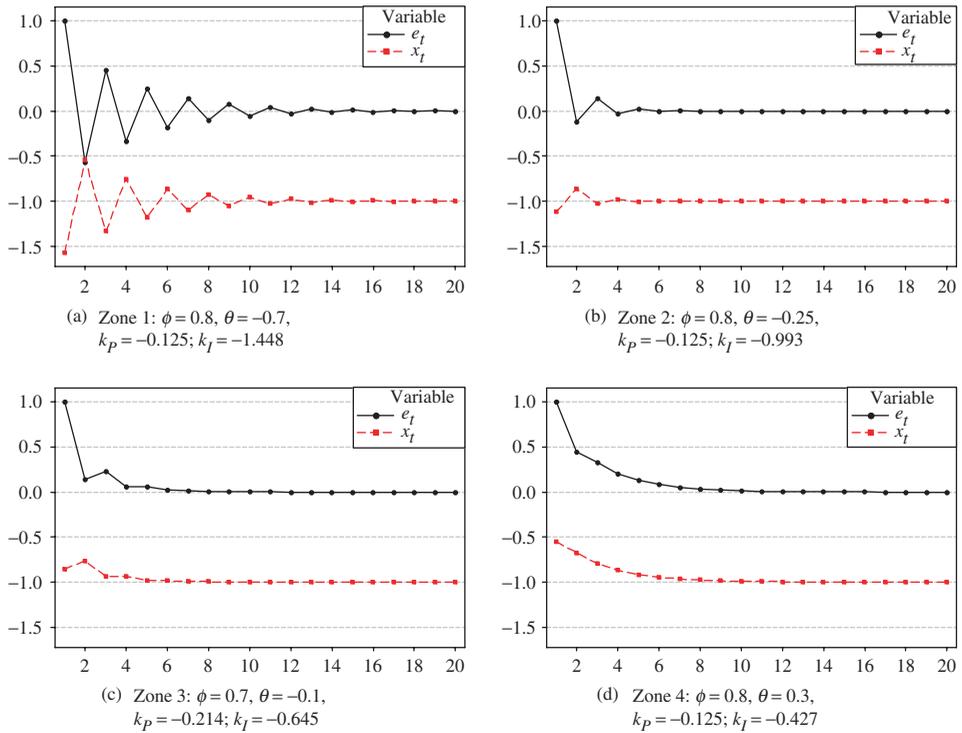
Figure 2. Dynamic shift patterns of PI-controlled processes in different zones.

Note that $\mu_t B^k = 0$ for $k \geq 1$, the output, e_t, follows

$$\begin{cases} E[e_0] = \delta \sigma_d \\ E[e_1] = (1 + \theta - \phi)\delta \sigma_d \\ \lim_{t \to \infty} E[e_t] \to \delta \sigma_d (1 - \phi)/(1 - \theta). \end{cases} \quad (17)$$

By the same token, the mean of the control action is obtained:

$$E[x_t] = \mu_t(\theta - \phi) \sum_{k=0}^{\infty} \theta^k B^k. \quad (18)$$

In particular, the transient and steady states of x_t are shown to be:

$$\begin{cases} E[x_0] = (\theta - \phi)\delta \sigma_d \\ E[x_1] = (\theta - \phi)(1 + \theta)\delta \sigma_d \\ \lim_{t \to \infty} E[x_t] \to \delta \sigma_d(\theta - \phi)/(1 - \theta). \end{cases} \quad (19)$$

The analysis clearly characterizes the time-varying shifts of the process under the MMSE controller. Compared to the PI-controlled output in (12), one significant difference is found in (17). The limiting output does not equal zero, which means the MMSE controller fails to compensate a sustained level shift in the process mean.

Likewise, the parameter space of processes under MMSE controllers can be further classified. Based on the value of the first transient response and the value of the steady state response, the stability region of process (1) is divided into three zones in Hu and Roan (1996) and Tsung and Tsui (2003). In particular, within the stability region $|\phi| < 1$, $|\theta| < 1$, the zones are defined by the following model:

$$\begin{cases} \text{zone 1}: & \theta < \phi - 1 \\ \text{zone 2}: & \theta > \phi \text{ and } \theta > \phi - 1 \\ \text{zone 3}: & \theta > \phi. \end{cases} \quad (20)$$

An example from each zone is shown in figure 3. In zone 1, strong oscillations are observed from both e_t and x_t. In zone 2, the process decays gradually until it reaches its steady state; while in zone 3, both e_t and x_t increase smoothly toward their respective steady state levels.

Henceforth, we use PI-zone i to denote the ith zone in figure 1, and MMSE-zone i to represent the ith zone given by equation (20). Examples will be taken from different zones for further investigation.

It is worth noting that taking expectations on both sides of equation (9) for $t \geq 0$ and making a simple transformation yields:

$$E[e_t] - E[x_{t-1}] = \delta, \quad (21)$$

which shows that the shift magnitudes of the output and the one-time lagged input are in parallel to each other with a constant distance, δ. When the signal in the output stream becomes weaker, the signal in the input side will be stronger, and vice versa. This equation holds independent of any particular algorithms. It partially supports the finding that the joint monitoring of e_t and x_t is favoured over the monitoring of
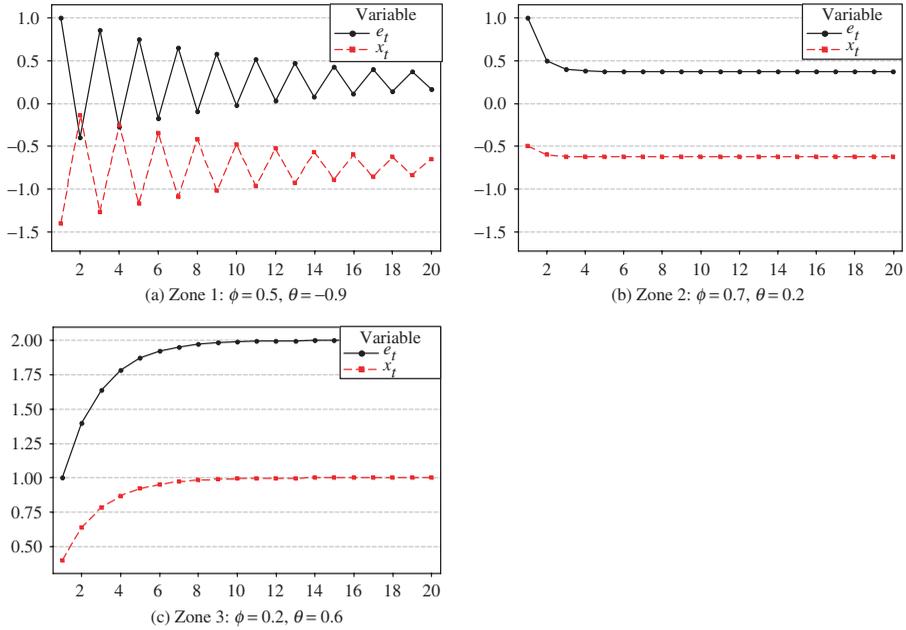
Figure 3.   Dynamic shift patterns of MMSE-controlled processes in different zones.

any individual variable, since the overall signals are not lost if both variables are being monitored.

4. An adaptive T^2 chart for monitoring feedback-controlled processes

The previous section has revealed that the dynamic shift patterns of feedback-controlled processes carry rich information about process faults, and hence, are critical to SPC. The major pitfall of chart (5) is its ignorance of dynamic shift patterns. Charts (7) and (8), although in which two static snapshots are considered, fail to take advantage of the whole pattern of information.

Wang and Tsung (2007) proposed an adaptive T^2 chart for monitoring dynamic shifts. In this section, we firstly introduce this adaptive T^2 chart and then applied it to monitor feedback-controlled processes. Based on the specific pattern information derived in section 3, a new forecasting algorithm is proposed to enhance forecasting accuracy and improve charting performance.

Instead of forcing a constant shift direction in (6), we treat the shift direction as dynamic and allow it to change over time. Assume at time t, the shift direction is known to be $\boldsymbol{\mu}_t$. The adaptive T^2 chart is given by:

$$T^2 = \boldsymbol{\mu}_t^T \boldsymbol{\Sigma}^{-1} \mathbf{V}_t - \frac{1}{2} \boldsymbol{\mu}_t^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_t > h_3. \qquad (22)$$

This charting statistic is the UMP unbiased test for detecting the existence of shifts in the particular direction, $\boldsymbol{\mu}_t$. The adaptive procedure is analogous to the

directionally variant T^2 chart (6), except that the shift direction vector is subscripted by time t, and is updated continuously in each step. In addition, an additional component derived from the log-likelihood ratio related to the hypothesis $E[\mathbf{V}_t] = 0$ *vs* $E[\mathbf{V}_t] = \boldsymbol{\mu}_t$ is maintained since $\boldsymbol{\mu}_t$ is now not a constant. More properties of the adaptive T^2 chart are given in Wang (2006).

One main challenge in implementing the adaptive T^2 chart is the determination of $\boldsymbol{\mu}_t$. The performance of the chart is believed to be heavily dependent on the accuracy in estimating this statistic. Although theoretical shift patterns of PI- or MMSE-controlled processes have been given in (11) and (13), or (16) and (18), the occurrence time of a process failure, which is also called the change-point of a process, is rarely known. Therefore, forecasting algorithms will be employed to obtain an estimation of $\boldsymbol{\mu}_t$.

In general, a forecasting algorithm is either model based or model free. A model-based algorithm involves fitting an ARMA(p, q) model to the dataset for each stream first, then the one-step-ahead estimation of the model is used as the true value (Pandit and Wu 1993, Apley and Tsung 2002). However, the performance of a model-based method depends heavily on the accuracy of the time-series modeling. Poor estimation of model parameters will lead to poor forecasting accuracy. A model-free method does not require an explicit fitting of any specific models. As suggested by Alwan and Roberts (1989), the EWMA statistic is, in many cases, a good approximation of time-series models. Nembhard and Kao (2003) successfully utilized an EWMA-base method in intensity forecasting in a plastic product manufacturing station. In addition, the EWMA forecasting procedure has a simple form and good interpretability. Therefore, the EWMA procedure will be utilized in this paper for dynamic shift forecasting.

Let $\mu_{e,t}$ be the estimated dynamic shift sequence for the process output. It can be obtained by an EWMA recursive procedure:

$$\mu_{e,t} = \lambda e_t + (1 - \lambda)\mu_{e,t-1}, \tag{23}$$

where λ is a smoothing parameter that satisfies $0 \leq \lambda \leq 1$. The larger the λ is, the faster the trend is followed. However, if the raw data contain large variations, a smaller value of λ will provide a more robust forecasting against noise signals.

The shift of x_t, $\mu_{x,t}$ can be obtained by the same token. Furthermore, let $\mathbf{V}_t = [e_t, x_t]^T$ be a vector of the latest observations, and $\boldsymbol{\mu}_t = [\mu_{e,t}, \mu_{x,t}]^T$ be the simultaneous forecasting of the input and the output shift directions. The recursive updating equation of $\boldsymbol{\mu}_t$ is obtained by plugging in the forecasting of each variable:

$$\boldsymbol{\mu}_t = \lambda \mathbf{V}_t + (1 - \lambda)\boldsymbol{\mu}_{t-1}, \tag{24}$$

where λ is the common smoothing parameter. The above modeling procedure is analogous to the multivariate EWMA procedure for monitoring multivariate applications (see Lowry and Montgomery 1995 and the references therein). The limiting form of the variance of $\boldsymbol{\mu}_t$ is given by:

$$\boldsymbol{\Sigma}_\mu = \frac{\lambda}{2 - \lambda}\boldsymbol{\Sigma}, \tag{25}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of $\mathbf{V}_t$.

(a) PI-zone 1: $\phi = 0.8$, $\theta = -0.7$, $k_P = -0.125$; $k_I = -1.448$, $\lambda = 0.3$

(b) MMSE-zone 3: $\phi = 0.2$, $\theta = 0.6$, $\lambda = 0.3$
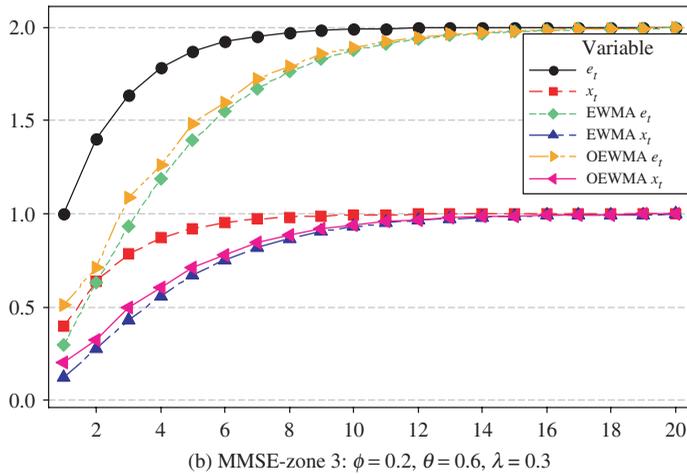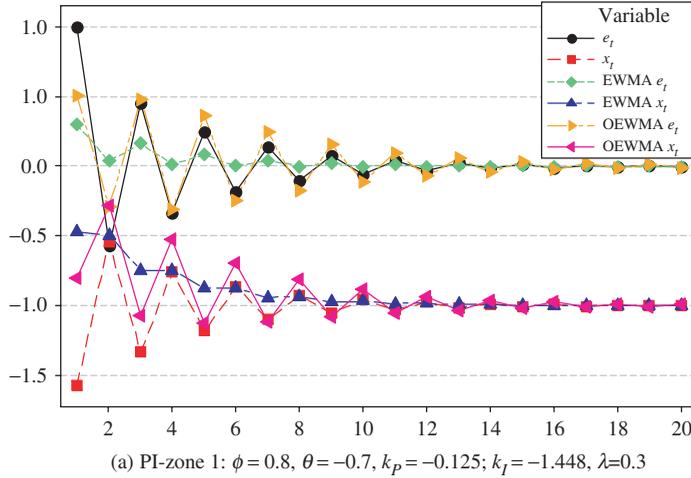
Figure 4. Forecasting of feedback-controlled processes using EWMA and OEWMA.

Figure 4 shows two examples when the process has a unit mean shift starting from $t = 0$; e_t and x_t are forecasted by an EWMA procedure. The example selected from PI-zone 1 exhibits strong oscillations. Conversely, the process from MMSE-zone 3 shows smoothed increasing trend in approaching its steady state.

As the weights of the EWMA predictor fall off exponentially, the effect of applying EWMA to a stationary series is equivalent to smoothing out jagged noise while keeping the trend visible. However, oscillations in the processes from PI-zone 1 and MMSE-zone 1 are part of the fault signatures. Eliminating the oscillations will cause the information to be lost. As is shown in figure 4(a), the forecasted output stays close to zero in the transient stage of the mean shift. No doubt, the loss of transient information will decrease the possibility of detecting a fault in the earliest stage.

A careful study of the oscillated processes in figures 2 and 3 reveals that each process oscillates around its respective mean, and the cycle time of the oscillations is 1.

That is, the process will go beyond and beneath the process mean alternatively, and approach the steady state level in the end. Therefore, we propose to extend the traditional EWMA procedure to an Oscillated EWMA (OEWMA) to gain a more accurate forecasting when oscillations exhibit.

The OEWMA model is shown in equation (26). In this model, the process mean, m_t, is first predicted with an ordinary EWMA procedure. Next, the forecasting error, Δ_t, which is believed to be an oscillated process, is calculated. The residual process has a mean of zero, and changes its sign alternatively. We will therefore develop another EWMA equation, which is shown on the third line in (26) and is the key to the OEWMA, to obtain a forecasting of the oscillated residual process, p_t. Finally, the predicted mean, m_t, and the residual, p_t, are summed up to generate the one-step-ahead forecasting of the original process.

$$
\begin{cases}
m_t = (1 - \lambda)m_{t-1} + \lambda e_t \\
\Delta_t = e_t - m_t \\
p_t = -(1 - \lambda)p_{t-1} + \lambda \Delta_t \\
\mu_t = m_t + p_t.
\end{cases}
\tag{26}
$$

The equation for forecasting the residual part, p_t, resembles the regular EWMA except for a minor modification in the sign of the weights. Expanding the equation yields:

$$
p_t = \lambda \Delta_t - \lambda(1 - \lambda)\Delta_{t-1} + \lambda(1 - \lambda)^2 \Delta_{t-2} - \cdots + (-1)^k \lambda(1 - \lambda)^k \Delta_{t-k} + \cdots.
\tag{27}
$$

Compared to the regular EWMA procedure, the signs of the weights in (27) change alternatively. However, the decaying speeds of EWMA and OEWMA are equal if the same λ is used by both.

The forecasted sequences by OEWMA are also shown in figure 4. As is seen from figure 4(a), the waved trend of the oscillated process is successfully captured. For the slow-increasing process in figure 4(b), OEWMA performs closely to the ordinary EWMA. This suggests that although OEWMA is designed for oscillated processes, it exhibits a robust performance in forecasting a smoothly changing process.

It is also interesting to consider two extreme cases of the adaptive T^2 chart. If the smoothing parameter is chosen as $\lambda = 1$, then the forecasting sequence is identical to the observation sequence, $\mu_t = V_t$. The adaptive T^2 procedure (22) reduces to $T^2 = V_t^T \Sigma^{-1} V_t$, which is the conventional T^2 chart without considering dynamic shift patterns. If, instead, $\lambda = 0$, it results in $\mu_t = \mu_0$. The predicted shift direction is kept constant as the initial value. The adaptive procedure reduces to $T^2 = \mu_0^T \Sigma^{-1} V_t$ after removing a constant term, which is a directional variant T^2 chart designed for a single specific direction, μ_0. In general, the adaptive chart takes value $0 < \lambda < 1$, and it is expected to capture the shift trend, and enhance its detection performance.

The ARL of the adaptive T^2 chart is influenced by control limit h_3 in (22), which is a design parameter that helps to define a balanced tradeoff between the false alarm rate and the run length. The usual integral equation and Markov chain methods (see Lu and Reynolds 1999 and the references therein) for evaluating the properties of control charts can be extended to the adaptive T^2 chart. When the process is in control, the predicted direction vector, μ_t, follows a multivariate normal distribution with mean zero and variance given in (25). Therefore, the scaled T^2 statistic, (22),

follows a chi-square distribution with two degrees of freedom. The transition probability matrix can then be obtained based on the normal distribution of $\boldsymbol{\mu}_t$ and the chi-square distribution of the T^2 statistic. However, the study of this method deserves further research efforts. In this paper, we will compute the ARL using Monte Carlo simulations. The determination of control limit h_3 for a given in-control ARL can be calculated using a numerical method. A software program is provided to facilitate the calculation, which can be downloaded from http://qlab.ielm.ust.hk/.

5. Simulation study

In order to investigate the performance of the adaptive T^2 chart in monitoring feedback-controlled processes, we will compare it with existing schemes. The chosen competitors include the U_0 and the U_∞ charts from Jiang (2004), and the T^2 chart from Tsung *et al.* (1999). In addition, multivariate EWMA (MEWMA) charts have been shown to possess better power than Shewhart-type T^2 charts in detecting small process shifts (see Lowry and Montgomery 1995, Molnau *et al.* 2001, etc.); we will also compare the adaptive T^2 scheme with the MEWMA chart.

The adaptive T^2 chart is used in conjunction with two forecasting algorithms. The first one is the regular EWMA procedure. We denote the corresponding chart an AT2-E chart henceforth. The second one is the OEWMA procedure, and we name the chart the AT2-OE chart hereafter.

Table 1 shows the charting performance of a process under PI control. The process is taken from PI-zone 1 with $\phi = 0.8$, $\theta = -0.7$ and the controller has $k_P = -0.125$, and $k_I = -1.448$. As is shown in figure 2(a), this process shows strong oscillations in its transient stage. Two different values, $\lambda = 0.2$ and $\lambda = 0.5$, are adopted by the AT2-E and the AT2-OE charts, respectively. For the multivariate EWMA chart, the same two smoothing parameter values are applied. The shift magnitude, as shown in the first column of the table, ranges from zero to 3.0 to cover both small and large shifts. The in-control ARL of each chart is forced to be 200 for fair comparison. Each ARL value is obtained using at least 10 000 replicates in all the cases.

First, we give a general review on the competing schemes. As is seen in table 1, the T^2 chart, as expected, is sensitive to large shifts, while it is not as capable as the U_∞ and MEWMA chart in detecting small shifts. However, the large shift performance of U_∞ and MEWMA charts are rather poor, especially when a small λ value is used by the MEWMA chart. The U_0 is sensitive to large process shifts, but it is inferior to the U_∞ chart in detecting small shifts.

When comparing the adaptive T^2 charts with other competing charts for small shifts, it is learned that the adaptive charts with $\lambda = 0.2$ are superior to the T^2 chart and the U_0 chart for small shifts. Although the MEWMA and U_∞ charts perform close to or better than the AT2 charts for small shifts, their large shift performance is seriously deteriorated. A close examination of the MEWMA chart with $\lambda = 0.5$ and the two adaptive T^2 charts with $\lambda = 0.2$ reveals that we should favour the latter charts even more strongly than the former one. The AT2-E and AT2-OE charts outperform the MEWMA chart over the whole range. This shows that a fine-tuned adaptive chart is more effective in detecting large scope of process shifts than the MEWMA chart.

Table 1.   Performance comparison of control charts for a process from PI-Zone 1.
$\phi = 0.8$, $\theta = -0.7$, $k_P = -0.125$; $k_I = -1.448$.

δ	AT2-E		AT2-OE		T^2	U_0	U_∞	MEWMA	
	$\lambda = 0.2$	$\lambda = 0.5$	$\lambda = 0.2$	$\lambda = 0.5$				$\lambda = 0.2$	$\lambda = 0.5$
0.5	113.51	128.39	117.69	135.66	139.36	118.19	73.00	106.46	118.86
1.0	42.39	44.20	36.09	45.24	47.82	37.16	31.35	43.14	48.52
1.5	10.82	6.81	5.06	5.67	6.00	4.05	14.90	20.56	17.15
2.0	1.91	1.17	1.10	1.07	1.07	1.04	7.04	11.32	3.71
2.5	1.01	1.00	1.00	1.00	1.00	1.00	3.35	6.87	1.12
3.0	1.00	1.00	1.00	1.00	1.00	1.00	1.73	3.77	1.00

Table 2.   ARL distributions of the AT2-E and the AT2-OE chart. $\lambda = 0.2$.

δ	AT2-E					AT2-OE				
	ARL	P(3)	P(5)	P(10)	P(20)	ARL	P(3)	P(5)	P(10)	P(20)
0.5	113.51	0.0059	0.0152	0.0561	0.1428	117.69	0.0275	0.0440	0.0873	0.1580
1.0	42.39	0.1118	0.1488	0.2492	0.3959	36.09	0.2755	0.3482	0.4349	0.5305
1.5	10.82	0.5516	0.6001	0.6884	0.8092	5.06	0.7914	0.8495	0.8869	0.9242
2.0	1.91	0.9294	0.9438	0.9625	0.9834	1.10	0.9861	0.9941	0.9967	0.9989
2.5	1.01	0.9982	0.9991	0.9995	0.9998	1.00	1.0000	1.0000	1.0000	1.0000
3.0	1.00	1.0000	1.0000	1.0000	1.0000	1.00	1.0000	1.0000	1.0000	1.0000

The large shift performance of the adaptive T^2 procedures is slightly inferior to the T^2 chart. This is a consequence of the EWMA or OEWMA smoothing procedure, which slows down the tracking of large process shifts. In the event of a large mean shift, the forecasting needs several steps to catch up due to the averaging effect of historical observations. This situation is partially alleviated by increasing the smoothing parameter and putting more weight on recent observations. As is seen from the table, the large shift performance of both the AT2-E chart and the AT2-OE chart is improved when λ increases from 0.2 to 0.5. If λ takes an extreme value 1.0, the AT2-E chart will be equivalent to the T^2 chart. However, putting more emphases on the most recent observations will make the forecasting easier to be contaminated by noises and hence deteriorates the small shift performance. The trade-off between large and small shift performance depends on the requirement from a real situation. Practitioners can adjust the smoothing parameter flexibly to achieve higher sensitivity in a designated range.

Between the AT2-E chart and the AT2-OE chart, when the same smoothing parameter is used, the AT2-E chart is slightly superior to the AT2-OE chart for small shifts, while it is inferior to AT2-OE for large shifts. We run simulations to investigate the distribution of each ARL value of the two charts. The results are illustrated in table 2. Both the AT2-E and the AT2-OE charts use $\lambda = 0.2$. Let $P(n)$ be the probability that the ARL is less than or equal to n. Obviously, the AT2-OE has a higher probability to detect shifts with shorter runs. This is explained by the effect of the modified weighting scheme. The oscillation is more easily captured by the OEWMA forecasting algorithm.

Table 3. Performance comparison of control charts for a process from MMSE- Zone 3. $\phi = 0.2, \ \theta = 0.6$.

	AT2-E		AT2-OE					MEWMA	
δ	$\lambda = 0.2$	$\lambda = 0.5$	$\lambda = 0.2$	$\lambda = 0.5$	T^2	U_0	U_∞	$\lambda = 0.2$	$\lambda = 0.5$
0.5	9.67	12.61	11.69	16.15	17.55	16.62	9.72	8.92	10.05
1.0	4.06	4.18	4.33	4.62	4.74	4.84	3.76	4.49	4.05
1.5	2.75	2.65	2.79	2.78	2.81	2.70	2.56	3.35	2.83
2.0	2.17	2.05	2.14	2.10	2.10	1.92	2.11	2.79	2.29
2.5	1.85	1.71	1.76	1.72	1.72	1.51	1.92	2.40	1.99
3.0	1.63	1.47	1.50	1.47	1.47	1.26	1.76	2.12	1.80

Table 3 studies an MMSE-controlled process with $\phi = 0.2$ and $\theta = 0.6$, which is taken from MMSE-zone 3. The process is subject to a mean shift. Similar patterns are found as in table 1. The optimality of the adaptive T^2 procedures holds for all small shifts. For large shifts, the adaptive charts perform closely to the T^2 chart, especially when a large smoothing parameter is used. The difference between the AT2-E and the AT2-OE charts are not significant, especially for moderate and large shifts, which further proves that the OEWMA forecasting method is rather robust to failure types.

More extensive simulations on examples from other zones of the PI or MMSE controller have been conducted, and similar conclusions are drawn. The superiority of the adaptive T^2 procedures is found in all cases.

6. Conclusions

It has been well recognized that SPC techniques are necessary for long-term process stability monitoring and assignable-cause discovery in feedback-controlled processes, in which dynamic shift patterns are frequently observed. However, applying conventional control charts cannot gain satisfied performance in detecting dynamic shifts, since the pattern information is not captured by these charts.

In this paper, the dynamic shifts caused by the PI and MMSE controllers have been analysed. Based on features of the dynamic shifts, the adaptive T^2 chart has been investigated. The adaptive T^2 chart features using forecasting algorithms to estimate time-varying shift directions and monitor the statistic that is optimal in detecting the existence of the predicted shift direction at each step. A model-free EWMA algorithm is used for shift forecasting. By considering the strong oscillations of some feedback-controlled processes, a modified EWMA procedure, termed OEWMA, is proposed.

Simulation results show that the adaptive T^2 chart is sensitive to both small and large shifts. Its overall performance is satisfactory. This is explained by the capability of the forecasting procedure in capturing the underlying shift trend. The adaptive T^2 chart is also flexible in design. By adjusting the smoothing parameter, its performance can be optimized for desired shift magnitudes.

A comparison of the AT2-E chart and the AT2-OE chart shows that the former has a better performance for small shifts. However, the OEWMA procedure is more

robust than the EWMA procedure to strong oscillations. In addition, the AT2-OE chart has a higher probability in detecting shifts with short runs. Therefore, the AT2-OE chart is recommended for general situations.

Acknowledgment

Appendix

The variance-covariance matrix of an MMSE-controlled process

Substitute control action (4) and disturbance model (2) for corresponding terms in process (1) yields:

$$e_t = \frac{(\theta - \phi)}{1 - \phi B} e_{t-1} + \frac{1 - \theta B}{1 - \phi B} \varepsilon_t.$$

Therefore,

$$e_t = \varepsilon_t,$$
$$x_t = \phi x_{t-1} + (\theta - \phi)\varepsilon_t.$$

The variance of e_t is easily obtained as:

$$\sigma_e^2 = \sigma_\varepsilon^2.$$

Since $\text{cov}(x_t, \varepsilon_t) = (\theta - \phi)\sigma_\varepsilon^2$, the variance of x_t is obtained:

$$\sigma_x^2 = \frac{(\theta - \phi)^2}{1 - \phi^2}\sigma_\varepsilon^2.$$

The covariance between the output and the input is:

$$\text{cov}(x_t, \varepsilon_t) = (\theta - \phi)\sigma_\varepsilon^2.$$

References

Alwan, L.C. and Roberts, H.V., Time series modeling for statistical process control. *J. Bus. Econ. Stat*, 1989, **6**, 87–95.

Apley, D.W. and Shi, J., The GLRT for statistical process control of autocorrelated processes. *IIE Trans.*, 1999, **31**, 1123–1134.

Apley, D.W. and Tsung, F., The autoregressive T-2 chart for monitoring univariate autocorrelated processes. *J. Qual. Technol.*, 2002, **34**, 80–96.

Box, G. and Kramer, T., Statistical process monitoring and feedback adjustment – a discussion. *Technometrics*, 1992, **34**, 251–267.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., *Time series analysis: forecasting and control*, 1994 (Prentice Hall: Englewood Cliffs, N.J.).

Capilla, C., Ferrer, A., Romero, R. and Hualda, A., Integration of statistical and engineering process control in a continuous polymerization process. *Technometrics*, 1999, **41**, 14–28.

DelCastillo, E. and Hurwitz, A.M., Run-to-run process control: literature review and extensions. *J. Qual. Technol.*, 1997, **29**, 184–196.

Faltin, F.W. and Tucker, W.T., On-line quality control for the factory of the 1990s and beyond. In *Statistical Process Control in Manufacturing*, edited by J.B. Keats and D.C. Montgomery, 1991 (Marcel Dekker: New York).

Fan, S.K.S., Jiang, B.C., Jen, C.H. and Wang, C.C., SISO run-to-run feedback controller using triple EWMA smoothing for semiconductor manufacturing processes. *Int. J. Prod. Res.*, 2002, **40**, 3093–3120.

Gultekin, M., Elsayed, E.A., English, J.R. and Hauksdottir, A.S., Monitoring automatically controlled processes using statistical control charts. *Int. J. Prod. Res.*, 2002, **40**, 2303–2320.

Hawkins, D.M., Multivariate quality control based on regression-adjusted variables. *Technometrics*, 1991, **33**, 61–75.

Hu, S.J. and Roan, C., Change patterns of time series-based control charts. *J. Qual. Technol.*, 1996, **28**, 302–312.

Hwarng, H.B., Detecting process mean shift in the presence of autocorrelation: a neural-network based monitoring scheme. *Int. J. Prod. Res.*, 2004, **42**, 573–595.

Jen, C.H., Jiang, B.C. and Fan, S.K.S., General run-to-run (R2R) control framework using self-tuning control for multiple-input multiple-output (MIMO) processes. *Int. J. Prod. Res.*, 2004, **42**, 4249–4270.

Jiang, W., A joint monitoring scheme for automatically controlled processes. *IIE Trans.*, 2004, **36**, 1201–1210.

Lowry, C.A. and Montgomery, D.C., A review of multivariate control charts. *IIE Trans.*, 1995, **27**, 800–810.

Lu, C.W. and Reynolds, M.R., EWMA control charts for monitoring the mean of autocorrelated processes. *J. Qual. Technol.*, 1999, **31**, 166–188.

Molnau, W.E., Montgomery, D.C. and Runger, G.C., Statistically constrained economic 7design of the multivariate exponentially weighted moving average control chart. *Qual. Reliab. Eng. Int.*, 2001, **17**, 39–49.

Montgomery, D.C., Keats, J.B., Yatskievitch, M. and Messina, W.S., Integrating statistical process monitoring with feedforward control. *Qual. Reliab. Eng. Int.*, 2000, **16**, 515–525.

Nembhard, H.B. and Kao, M.S., Adaptive forecast-based monitoring for dynamic systems. *Technometrics*, 2003, **45**, 208–219.

Pandit, S.M. and Wu, S.M., *Time Series and System Analysis With Applications*, 1993 (Krieger Pub. Co.: Malabar, Florida).

Sachs, E., Hu, A. and Ingolfsson, A., Run by run process control - combining SPC and feedback-control. *IEEE Transactions on Semiconductor Manufacturing*, 1995, **8**, 26–43.

Tsung, F., Statistical monitoring and diagnosis of automatic controlled processes using dynamic PCA. *Int. J. Prod. Res.*, 2000, **38**, 625–637.

Tsung, F. and Apley, D.W., The dynamic T-2 chart for monitoring feedback-controlled processes. *IIE Trans.*, 2002, **34**, 1043–1053.

Tsung, F. and Shi, J.J., Integrated design of run-to-run PID controller and SPC monitoring for process disturbance rejection. *IIE Trans.*, 1999, **31**, 517–527.

Tsung, F., Shi, J. and Wu, C.F.J., Joint monitoring of PID-controlled processes. *J. Qual. Technol.*, 1999, **31**, 275–285.

Tsung, F. and Tsui, K.L., A mean-shift pattern study on integration of SPC and APC for process monitoring. *IIE Trans.*, 2003, **35**, 231–242.

Tsung, F., Wu, H.Q. and Nair, V.N., On the efficiency and robustness of discrete proportional-integral control schemes. *Technometrics*, 1998, **40**, 214–222.

Tucker, W.T., Faltin, F.W. and VanderWiel, S.A., Algorithmic statistical process control – an elaboration. *Technometrics*, 1993, **35**, 363–375.

Wang, K., Adaptive Charting Techniques for Multivariate and Dynamic Systems, PhD dissertation, 2006, Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Hong Kong.

Wang, K. and Tsung, F., An adaptive T^2 chart for monitoring dynamic systems. *J. Qual. Tech.*, 2007 (in press).

Zhang, N.F. and Pollard, J.F., Analysis of autocorrelations in dynamic processes. *Technometrics*, 1994, **36**, 354–368.