Journal of Process Control xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Journal of Process Control

journal homepage: www.elsevier.com/locate/jprocont

The optimal drift-compensatory and fault tolerant approach for mixed-product run-to-run control

Bing Ai [a], Ying Zheng [a,*], Shi-Shang Jang [b], Yanwei Wang [c], Lin Ye [a], Chunjie Zhou [a]

[a] Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan, Hubei 430074, PR China
[b] Department of Chemical Engineering, National Tsing-Hua University, Hsin-Chu, Taiwan
[c] CAD Center, Huazhong University of Science and Technology, Wuhan, Hubei 430074, PR China

ARTICLE INFO

Article history:
Received 19 October 2008
Received in revised form 25 April 2009
Accepted 25 April 2009
Available online xxxx

Keywords:
Run-to-run control
Threaded double exponentially weighted moving average (threaded-dEWMA)
"Trade-off" solution
Fault tolerant

ABSTRACT

Semiconductor manufacturing industry has elevated cost in productions. Improvement of production efficiency is always an important goal for manufacturers. Run-to-run control has been widely used in batch manufacturing processes to reduce variations. Threaded exponentially weighted moving average (threaded-EWMA) run-to-run control is an important and stable control scheme. In this paper, we study the drifted process with mixed products are manufactured in cycles on the same tool, and find that the process outputs will be off target greatly at the beginning runs of cycle 1, 2, ... if the product has a long break length. In order to reduce a possible high rework rate, a threaded double EWMA (thread-dEWMA) controller is used to handle the disturbance as well as the drift. By analysis of the system output, a drift-compensatory approach is proposed to eliminate these large deviations. In order to enhance the system performance, the well known "trade-off" solution is adopted to choose the optimal discount factors. Furthermore, how to deal with process fault is also considered in this paper. Two kinds of fault – the step fault and the ramp fault are discussed for fault tolerant approach which can reduce the large deviations of process output from the specification. Simulation study showed that the proposed approaches are effective.

1. Introduction

In the last years a great development in the field of semiconductor manufacturing has been achieved. Run-to-run control is one of the most commonly used control approaches. It is a form of discrete process and machine control in which the product recipe with respect to a particular machine process is modified ex situ, i.e., between machine "runs", so as to minimize process drift, shift and variability [1]. Exponentially weighted moving average (EWMA) algorithm is often adopted in run-to-run control. Box and Jenkins carried out pioneering work on the EWMA statistic and showed that EWMA provides the minimum mean square error forecast for an IMA(1, 1) process [2]. After Sachs and Hu introduced EWMA into semiconductor manufacturing industry, a considerable number of literatures discussed the choices of the optimal EWMA weight [3]. Ingolfson and Sachs [4] and Smith [5] analyzed stability and sensitivity of the process output for different closed-loop systems. Wang and He analyzed and compared the behaviors of EWMA controller with gain and intercept updating [6].

Many researchers showed that the use of the single EWMA (sEWMA) controller is very effective to reduce variability in the dynamics of the processes without drift. However, many applications had been shown that sometimes drift can occurs gradually due to worn-out tools and disturbances. At this time sEWMA controller will bring the process output off target. For this reason, Butler and Stefani proposed a double EWMA (dEWMA) controller for the process with deterministic linear drift [7]. It is proved that the dEWMA filter is a minimum mean square error controller for an IMA(2, 2) process [8]. Del Castillo analyzed its stability conditions, long-run behavior and transient effects, and proved that the dEWMA controller is not a minimum variance controller for processes with drift but it does provide unbiased control. In order to balance long-run variance and transient effect, he proposed a method called "trade-off" solution to find the optimal weights of the dEWMA controller [9]. Tseng et al. derived an exact expression for the process output of a dEWMA controller under the assumption that the process disturbance follows a general ARIMA(p, d, q) model, they also derived the optimal discount factors by minimizing the rework rate of the process output [10]. For the SISO system with a linear drift, Tseng et al. proposed variable EWMA controller to enhance the system performance [11]; For the first-order MIMO process with a linear drift, they obtained the stability condition for the process, and derived a formula for the adequate sample size required to achieve stability for the closed-loop system with a guaranteed probability. They also described how the offline

* Corresponding author. Tel.: +86 27 62100344; fax: +86 27 87556579.
E-mail address: zyhidy@mail.hust.edu.cn (Y. Zheng).

0959-1524/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jprocont.2009.04.008

DOE/RSM is conducted to obtain process parameter estimates [12]; For drifted MIMO systems, Lee et al. derived an analytical expression for the process output of a double multivariate EWMA controller and discussed the problems of the stability conditions of the system [13]. Chen et al. constructed the disturbance model for the aluminum sputter deposition process and derived the extending Kalman filter (EKF) controller by the time series model. Also they have applied the dEWMA controller, time-varying dEWMA controller, age-based dEWMA controller, and EKF controller to aluminum sputter deposition processes for predicting deposition rates [14].

All the aforementioned models are based on the assumption that there is only a single product in the manufacturing tool. Few researchers emphasized on mixed-product process until Edgar et al. reviewed the problems of mixed-product run-to-run control in a high-mix fab [15] and proposed just-in-time adaptive disturbance estimation (JADE) algorithm to isolate typical types of process disturbance [16]. Vanli et al. proposed a model selection approach to identify the context variables that contribute most to the process [17]. Ma et al. analyzed tool and product effects in a mixed-product and parallel tool environment [18] and proposed ANOVA approach to deal with run to run control of a high mixed operation [19]. Zheng et al. studied a drifted process with two products are manufactured on the single tool and proposed two kinds of control method – "tool-based" and "product-based" approach [20]. They found that the "tool-based" approach is unstable when the plant is non-stationary and the plant-model mismatch parameters are different for different products, while the "product-based" approach, i.e., threaded-EWMA approach is stable. Wu et al. gave further experimental study to prove the above results [21].

Giving the further study on the model proposed in [20], we found that if the product has a long break length, then at the beginning runs of each cycle the process output will be far deviated from target (see the example in Section 2). It usually takes a moderately large number of runs to bring the process output to its target and causes tremendous waste, however, it is well known that modern semiconductor manufacturing has always been an industry with high capital investment, improvement in production efficiency will potentially be very beneficial to manufacturers. In this paper, based on a threaded-dEWMA controller, the problem of large deviations at the beginning runs of each cycle is studied; a drift-compensatory approach is proposed to reduce these large deviations. In order to enhance the system performance, the "trade-off" solution is adopted to choose the optimal discount factors. Moreover, we also consider the effect of fault, two kinds of faults – a step fault and a ramp fault are discussed, and fault tolerant approach is proposed.

For case of presentation, the remainder of the paper is organized as follows: in Section 2, the problem formulation is presented and an example is provided to illustrate the main problems of threaded-dEWMA approach. In Section 3, performance analysis of the system is provided, a drift-compensatory approach is proposed and the "trade-off" solution is adopted. The step as well as the ramp faults are considered, corresponding fault tolerant approach is introduced in Section 4. In Section 5, simulation studies are provided. Conclusions are presented in Section 6. Details of mathematical derivations are given in Appendices.

2. Problem formulation

In semiconductor manufacturing batch processes, mixed products are usually manufactured on the same tool. In this paper, we consider a case that p products are manufactured on a single tool (as shown in Fig. 1). The production schedule consists of i runs, in which j_1 runs are used to manufacture product 1; j_2 runs are used to produce product 2, and so on, then j_1, is defined as the campaign length for product 1, $i-j_1$ is called break length of product 1.

Each step in semiconductor manufacturing will be a complicated physiochemical batch process; the input–output relationship is of course a nonlinear model. However, the complexity of the process and nonlinear model will make the system too complicated to analysis. Up to now, very few papers have discussed the nonlinear model, instead they choose linear model to approximate the practical system and this will help to simplify system analysis. In this paper, we also choose linear model for simplicity, and assume that the input–output relationship for the products on the given tool is linear with different intercepts $\alpha_1, \alpha_2, \ldots, \alpha_p$ and slopes $\beta_1, \beta_2, \ldots, \beta_p$, all the products share the same tool disturbance η_{it+n}, i.e.,

$$Y_{it+n} = \begin{cases} \alpha_1 + \beta_1 u_{it+n} + \eta_{it+n}, & 1 \leqslant n \leqslant j_1, \\ \alpha_2 + \beta_2 u_{it+n} + \eta_{it+n}, & j_1 + 1 \leqslant n \leqslant j_2, \\ \vdots & \\ \alpha_p + \beta_p u_{it+n} + \eta_{it+n}, & j_p + 1 \leqslant n \leqslant i. \end{cases} \tag{1}$$

where u_{it+n} is the manipulated variable at the beginning of the $(it+n)_{th}$ run, t is the number of cycle, $Y_{it+n}(n = 1, 2, \ldots, j_1)$ is the output of product 1, $Y_{it+n}(n = j_1 + 1 j_1 + 2, \ldots, j_2)$ is the output of product 2, and so on. An IMA (1, 1) disturbance with deterministic linear drift δ is used to model the change in tool condition, i.e.,

$$\eta_s - \eta_{s-1} = \varepsilon_s - \theta\varepsilon_{s-1} + \delta \tag{2}$$

where ε_s is independent identically distributed random variable and $\varepsilon_s \sim N(0, \sigma^2)$.

In threaded-EWMA run-to-run control, the EWMA filter action is performed according to the last run on which the same product is processed instead of the previous run in which a different product may have been processed. Thus, the output of product 1 is irrelevant with what is produced by the $i-j_1$ break runs, however, it is relevant with what is produced by the j_1 campaign runs. Without loss of generality, we will only discuss the performance of the product 1 in this paper; the next cases are just a generalization of the product 1's case.

Butler and Stefani proposed the double EWMA controller (predictor-corrector controller) to update recursively the estimation of the unknown parameters a, D and the manipulated variable u_{it+n} [7].

Let $u_1 = \frac{T_1}{b_1}$ be the initial value, then

$$u_{it+n} = \begin{cases} \frac{T_1 - a_0 - D_0}{b_1}, & n = 1 \quad and \quad t = 0, \\ \frac{T_1 - a_{i(t-1)+j_1} - D_{i(t-1)+j_1}}{b_1}, & n = 1 \quad and \quad t \geqslant 1, \\ \frac{T_1 - a_{it+n-1} - D_{it+n-1}}{b_1}, & n = 2, 3, \ldots, j_1 \end{cases} \tag{3}$$

where T_1 is the desired target of the product 1 and

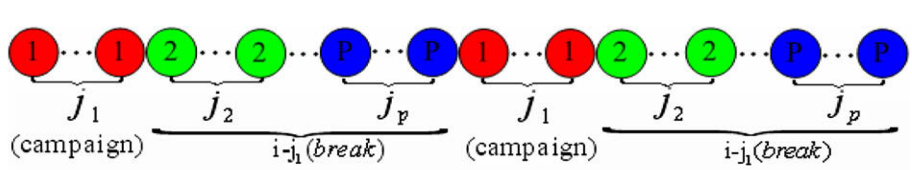

Fig. 1. An example of a tool manufacturing p products.

Please cite this article in press as: B. Ai et al., The optimal drift-compensatory and fault tolerant approach for mixed-product run-to-run control, J. Process Contr. (2009), doi:10.1016/j.jprocont.2009.04.008

$$a_{it+n} = \begin{cases} \lambda_1(Y_1 - b_1 u_1), & n = 1 \quad and \quad t = 0, \\ \lambda_1(Y_{it+1} - b_1 u_{it+1}) + (1-\lambda_1)a_{i(t-1)+j_1}, & n = 1 \quad and \quad t \geqslant 1, \\ \lambda_1(Y_{it+n} - b_1 u_{it+n}) + (1-\lambda_1)a_{it+n-1}, & n = 2,3,\ldots,j_1. \end{cases} \tag{4}$$

$$D_{it+n} = \begin{cases} \lambda_2(Y_1 - b_1 u_1), \quad n = 1 \quad and \quad t = 0, \\ \lambda_2(Y_{it+1} - b_1 u_{it+1} - a_{i(t-1)+j_1}) + (1-\lambda_2)D_{i(t-1)+j_1}, \\ \quad n = 1 \quad and \quad t \geqslant 1, \\ \lambda_2(Y_{it+n} - b_1 u_{it+n} - a_{it+n-1}) + (1-\lambda_2)D_{it+n-1}, \quad n = 2,3,\ldots,j_1 \end{cases} \tag{5}$$

For (4) and (5), we have that $\lambda_1, \lambda_2 \in [0,1]$ are called discount factors and that b_1 is the estimate of β_1, which can be obtained by design of experiments (DOE) in a pre-control phase.

Example. Set the true parameters of product 1 to be $\alpha_1 = 2$, $\beta_1 = 1.5$, $\theta = 0.5$, $\sigma = 1$, $\delta = 0.5$, $j_1 = 100$, and $i = 200$ randomly. Assume the least square estimation (LSE) for β_1 is $b_1 = 1$, the initial value of (a, D) are $(a_0, D_0) = (0, 0)$, and the target of product 1 is $T_1 = 0$. Fig. 2 shows the outputs of product 1 by using threaded-dEWMA controller with discount factors $(\lambda_1, \lambda_2) = (0.1, 0.1)$. From the figure, it can be noticed that at the beginning runs of each cycle, especially for $t = 1, 2, 3$, the process outputs are deviated very far from the target, however after several runs of oscillation, the outputs are going to hit the target.

To better understand the performance of the threaded-dEWMA controller, the output of product 1 in cycle 0 is amplified in Fig. 3. From the Figure, it's obvious that the output converges to the

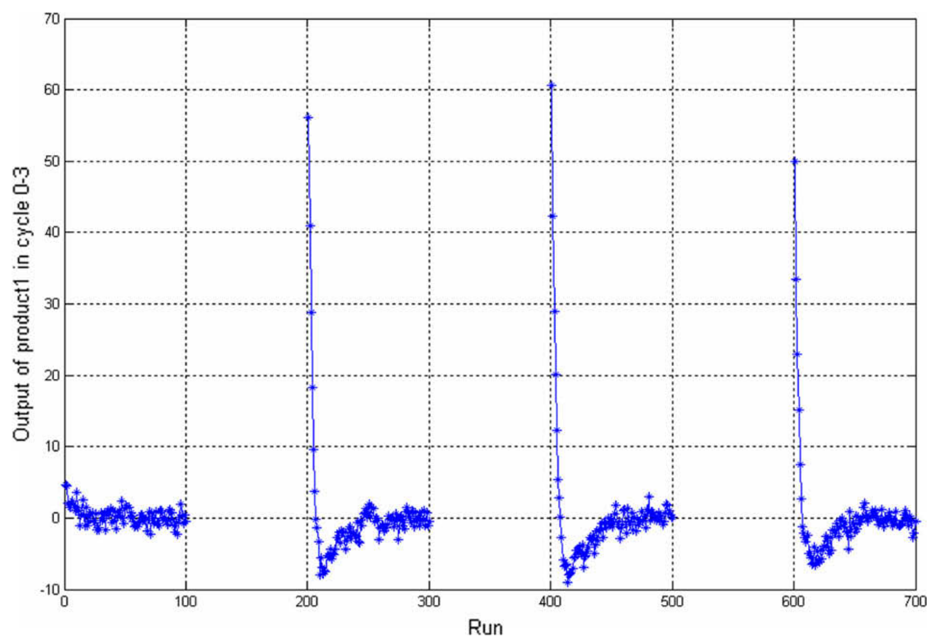

Fig. 2. Output of product 1 in cycle 0–3.

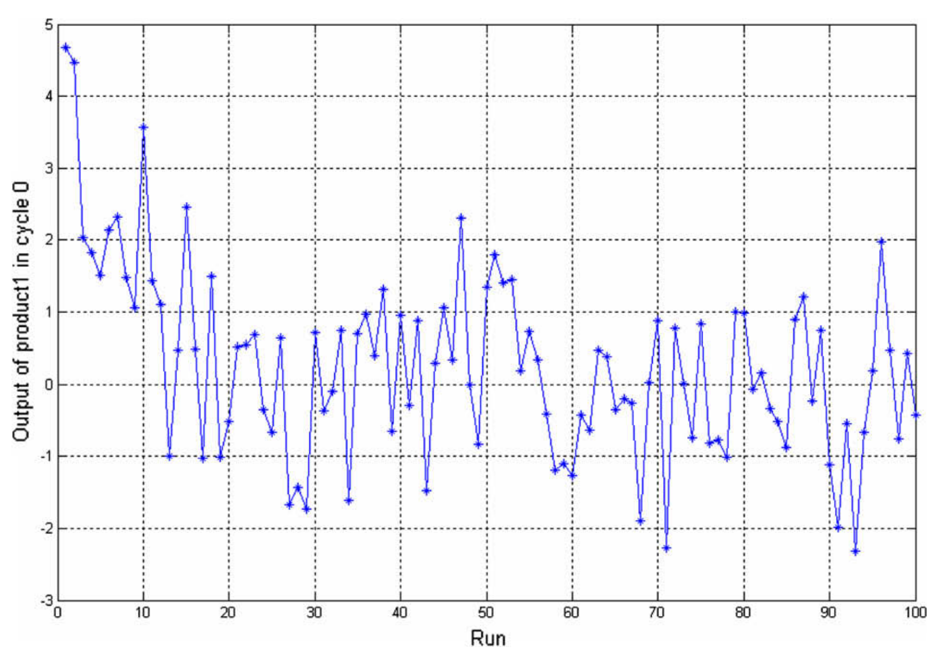

Fig. 3. Behavior of the product 1 for a specified interval of run taken from Fig. 2.

desired target slowly. The mean square error (MSE) and variance of product 1 in cycle 0 are 1.732 and 1.697 respectively, therefore for this cycle, it is necessary to repeat the production and get small MSE and variance which will bring about a high rework rate and lots of waste wafers although the process is stable at last.

In the following sections, we give answers to the related questions formulated according to our example:

(1) How to reduce large deviations at the beginning runs of cycle 1, 2, ...?
(2) How to accelerate the convergence rate and reduce the MSE and variance? i.e., how to choose the optimal discount factors?
(3) How to deal with a sudden fault?

3. The optimal drift-compensatory approach

In this section, the condition for stability of the process based on threaded-dEWMA run-to-run control will be discussed and also we will introduce the new solution called "drift-compensatory approach" to avoid the cause of large deviations in the first few runs of cycle 1, 2, ... via the analysis of the system-output.

Furthermore, in order to enhance the system performance, the "trade-off" solution is used to obtain the optimal discount factors of the controller.

3.1. Stability analysis

Stability is a fundamental requirement for any process. An unstable control scheme should not be implemented. In this subsection, we shall examine the influence of model error and the discount factors on the stability of the threaded-EWMA scheme. The output of the product 1 is expressed as follows:

Result 1. Output of the product 1 at the $n_{th}(n \geqslant 2)$ run in cycle 0 can be expressed as:

$$Y_n = \varphi_1^{n-1}(\alpha_1 + \xi_1 T_1 + \eta_1) + \sum_{k=0}^{n-2} \varphi_1^k \xi_1 (\lambda_1 + \lambda_2) T_1 - \sum_{k=0}^{n-2} \varphi_1^k D_{n-2-k} \xi_1 \lambda_1 + \sum_{k=0}^{n-2} \varphi_1^k (\eta_{n-k} - \eta_{n-k-1}) \tag{6}$$

where $\xi_1 = \frac{\beta_1}{b_1}$, $\varphi_1 = 1 - \xi_1(\lambda_1 + \lambda_2)$, and ξ_1 is defined as the model error of the process gain.

Proof. See Appendix A for details.

Result 2. Outputs of the product 1 at the first and the $n_{th}(n \geqslant 2)$ run in cycle $t(t \geqslant 1)$ are given by:

$$\begin{aligned} Y_{it+1} = {} & \varphi_1^{j_1 t-1}[\alpha_1 + \xi_1 T_1(\varphi_1 + \lambda_1 + \lambda_2) - \xi_1(\lambda_1 + \lambda_2)(\alpha_1 + \eta_1) + \eta_2] \\ & + \varphi_1^{j_1(t-1)+1} \sum_{k=0}^{j_1-3} \varphi_1^k [\xi_1(\lambda_1 + \lambda_2) T_1 + (\eta_{j_1-k} - \eta_{j_1-k-1}) \\ & - D_{j_1-k-2} \xi_1 \lambda_1] + \varphi_1^{j_1(t-1)} \xi_1 (\lambda_1 + \lambda_2) T_1 \\ & - \varphi_1^{j_1(t-1)} D_{j_1-1} \xi_1 \lambda_1 + \varphi_1^{j_1(t-1)} (\eta_{i+1} - \eta_{j_1}) \\ & + \sum_{m=0}^{t-2} \varphi_1^{m j_1} \Bigg\{ \sum_{k=1}^{j_1-1} \varphi_1^k [\xi_1(\lambda_1 + \lambda_2) T_1 + \eta_{i(t-m-1)+j_1+1-k} \\ & - \eta_{i(t-m-1)+j_1-k}] + \xi_1(\lambda_1 + \lambda_2) T_1 \\ & - \sum_{k=0}^{j_1-2} \varphi_1^k D_{i(t-m-1)+j_1-k-1} \xi_1 \lambda_1 - \varphi_1^{j_1-1} D_{i(t-m-2)+j_1} \xi_1 \lambda_1 \\ & + \eta_{i(t-m)+1} - \eta_{i(t-m-1)+j_1} \Bigg\} \end{aligned} \tag{7}$$

$$\begin{aligned} Y_{it+n} = {} & \varphi_1^{n-1} Y_{it+1} + \sum_{k=0}^{n-2} \varphi_1^k \xi_1 (\lambda_1 + \lambda_2) T_1 + \sum_{k=0}^{n-2} \varphi_1^k (\eta_{it+n-k} \\ & - \eta_{it+n-k-1}) - (\gamma_2 \varphi_1^{n-2} D_{i(t-1)+j_1} + \sum_{k=0}^{n-3} \varphi_1^k D_{it+n-k-2}) \xi_1 \lambda_1 \end{aligned} \tag{8}$$

where $\gamma_2 = \begin{cases} 0, & n < 2, \\ 1, & n \geqslant 2. \end{cases}$.

Proof. See Appendix B for the procedure.

Proposition. If $|\varphi_1| \geqslant 1$, then the process is unstable.

Proof. If $\varphi_1 \geqslant 1$ or $\varphi_1 \leqslant -1$, from (6) and (8), it can be deduced that the outputs are unstable.

3.2. Drift-compensatory control

As illustrated in Fig. 2, at the beginning runs of cycle $t(t \geqslant 1)$, the process output presents a great deviation from the desired target. In this subsection, the cause of this problem is found and an effective method is introduced to avoid the large deviation and take the process output to the desired target.

Consider, for instance, that $|\phi_1| < 1$, then substitute (2) into (6)–(8), we have

$$\begin{aligned} Y_n = {} & \varphi_1^{n-1}\alpha_1 + [\varphi_1^{n-1}(\xi_1 - 1) + 1]T_1 - \sum_{k=0}^{n-3} \varphi_1^k D_{n-2-k} \xi_1 \lambda_1 \\ & + \frac{1 - \varphi_1^n}{1 - \varphi_1} \delta + \varphi_1^{n-2}[(\varphi_1 - \theta)\varepsilon_1 + \varepsilon_2] + \sum_{k=0}^{n-3} \varphi_1^k (\varepsilon_{n-k} - \theta \varepsilon_{n-k-1}) \end{aligned} \tag{9}$$

$$\begin{aligned} Y_{it+1} = {} & T_1 + \left(\frac{\varphi_1}{1 - \varphi_1} + i - j_1 + 1 \right) \delta \\ & - \sum_{k=0}^{j_1-2} \varphi_1^k D_{i(t-1)+j_1-k-1} \xi_1 \lambda_1 + (1 - \theta) \sum_{k=j_1+1}^{i} \varepsilon_{i(t-1)+k} \\ & + (\varphi_1 - \theta) \sum_{k=0}^{j_1-2} \varphi_1^k \varepsilon_{i(t-1)+j_1-k} + \varepsilon_{it+1} \end{aligned} \tag{10}$$

$$\begin{aligned} Y_{it+n} = {} & \varphi_1^{n-1} Y_{it+1} + T_1 - \varphi_1^{n-1} T_1 \\ & - \left(\gamma_2 \varphi_1^{n-2} D_{i(t-1)+j_1} + \sum_{k=0}^{n-3} \varphi_1^k D_{it+n-2-k} \right) \xi_1 \lambda_1 \\ & + \sum_{k=0}^{n-2} \varphi_1^k (\varepsilon_{it+n-k} - \theta \varepsilon_{it+n-k-1} + \delta) \end{aligned} \tag{11}$$

where $n \geqslant 2$.

Taking the mathematical expectation for (9) and defining $n = j_1$, then

$$\begin{aligned} E(Y_{j_1}) - T_1 & = \varphi_1^{j_1-1}\alpha_1 + \varphi_1^{j_1-1}(\xi_1 - 1)T_1 - \sum_{k=0}^{j_1-3} \varphi_1^k D_{j_1-2-k} \xi_1 \lambda_1 + \frac{1 - \varphi_1^{j_1}}{1 - \varphi_1} \delta \\ & = \frac{1}{1 - \varphi_1} \delta - \sum_{k=0}^{j_1-3} \varphi_1^k D_{j_1-2-k} \xi_1 \lambda_1 \end{aligned} \tag{12}$$

Again, taking the mathematical expectation for (10) and (11), we get

$$\begin{aligned} E(Y_{it+n}) - T_1 = {} & \left[\varphi_1^{n-1} \left(\frac{\varphi_1}{1 - \varphi_1} + i - j_1 + 1 \right) + \frac{1 - \varphi_1^{n-1}}{1 - \varphi_1} \right] \delta \\ & - [\varphi_1^{n-1} \sum_{k=0}^{j_1-2} \varphi_1^k D_{i(t-1)+j_1-k-1} + \gamma_2 \varphi_1^{n-2} D_{i(t-1)+j_1} \\ & + \sum_{k=0}^{n-3} \varphi_1^k D_{it+n-2-k}] \xi_1 \lambda_1 \end{aligned} \tag{13}$$

where $n \geqslant 1$.

Please cite this article in press as: B. Ai et al., The optimal drift-compensatory and fault tolerant approach for mixed-product run-to-run control, J. Process Contr. (2009), doi:10.1016/j.jprocont.2009.04.008

Combining (12) and (13), we have the following result:

Result 3: The mainly biases of process outputs between the $(it+1)$th and $(i(t-1)+j_1)$th run, $(it+2)$th and $(it+1)$th run, $(it+n)$th and $(it+n-1)$th run are:

$$E(Y_{it+1}) - E(Y_{i(t-1)+j_1}) = (i-j_1)\delta + A \tag{14}$$

$$E(Y_{it+2}) - E(Y_{it+1}) = (\varphi_1 - 1)(i-j_1)\delta + B \tag{15}$$

$$E(Y_{it+n}) - E(Y_{it+n-1}) = \varphi_1^{n-2}(\varphi_1 - 1)(i-j_1)\delta + C \tag{16}$$

where

$$A = [(1-\varphi_1)\sum_{k=1}^{j_1-2}\varphi_1^{k-1}D_{i(t-1)+j_1-1-k} - D_{i(t-1)+j_1-1}]\xi_1\lambda_1,$$

$$B = [(1-\varphi_1)\sum_{k=0}^{j_1-2}\varphi_1^{k}D_{i(t-1)+j_1-1-k} - D_{i(t-1)+j_1}]\xi_1\lambda_1,$$

$$C = \{(1-\varphi_1)[\varphi_1^{n-2}\sum_{k=0}^{j_1-2}\varphi_1^{k}D_{i(t-1)+j_1-1-k} + \varphi_1^{n-3}D_{i(t-1)+j_1} + \gamma_4\sum_{k=1}^{n-3}\varphi_1^{k-1}D_{it+n-2-k}] - D_{it+n-2}\}\xi_1\lambda_1,$$

$$n \geqslant 3, t \geqslant 1, \quad \text{and} \quad \gamma_4 = \begin{cases} 0, & n<4,\\ 1, & n \geqslant 4. \end{cases}$$

Proof. It can be derived from (12) and (13) directly.

Remark 1. If the product has a long break length, it is obvious from (14) that the bias of the process output at the first run of cycle $t(t \geqslant 1)$ will be very large.

Remark 2. Eq. (15) shows that the bias of process output at the second run is smaller than that at the first run in cycle $t(t \geqslant 1)$.

Remark 3. It can be concluded from (16) that $\varphi_1^{n-2}(\varphi_1 - 1)(i-j_1)\delta < 0$ and $C > 0$. However, for a long break length, when n has small values, we have that $|\varphi_1^{n-2}(\varphi_1-1)(i-j_1)\delta| \gg C$, so the influence of C in the process can be ignored. In fact, this is the main reason why the output of the system presents large deviation from the desired target. On the other hand, after considerable runs, we have that $|\varphi_1^{n-2}(\varphi_1-1)(i-j_1)\delta| \leqslant C$, so the factor C starts to affect the bias and the oscillation of the output of the system starts.

According to our analysis, we found that the general threaded-dEWMA controller spends a lot of runs of trying to eliminate the large deviations. But generally the campaign length j_1 is not long enough for general threaded-dEWMA controller to do that work. Even if the campaign length is long enough, it will bring high rework rate and poor efficiency. In order to solve these problems, in what follows, we will propose the drift-compensatory approach, which will compensate the outputs at each run, i.e.,

3.2.1. Drift-compensatory approach

The output of product 1 at the nth run in cycle $t(t \geqslant 1)$ is compensated as follows:

$$(\hat{Y}_{it+n})_{nf} = Y_{it+n} + [\varphi_1^{n-1}\sum_{k=0}^{j_1-2}\varphi_1^{k}D_{i(t-1)+j_1-k-1} + \gamma_2\varphi_1^{n-2}D_{i(t-1)+j_1} + \sum_{k=0}^{n-3}\varphi_1^{k}D_{it+n-2-k}]\xi_1\lambda_1 \times [\varphi_1^{n-1}\left(\frac{\varphi_1}{1-\varphi_1} + i - j_1 + 1\right) + \frac{1-\varphi_1^{n-1}}{1-\varphi_1}]\delta \tag{17}$$

where $(\hat{Y}_{it+n})_{nf}$ is the compensated output.Eq. (17) can be concluded from (13) directly.As shown in Fig. 2, the output will reach the target after several runs, therefore we only need to compensate the beginning runs of cycle $t(t \geqslant 1)$ to get a desired output response.

3.3. Optimal discount factors selection

To further enhance the performance of the process, it is always preferable to find the optimal discount factors of the threaded-dEWMA controller. Hence, in this subsection, the "trade-off" solution proposed by Del Castillo is adopted [9] to choose such discount factors. These optimal discount factors should not only balance the transient and long-run behavior, but also minimize the total mean square error and the variance of system output.

The optimal problem proposed by Del Castillo in [9] is listed as follows:

$$\min_{\lambda_1,\lambda_2}\left\{\lim_{s\to\infty} Var(Y_s) + \frac{S_m}{m}\right\} \tag{18}$$

Constraint condition: $0 \leqslant \lambda_1 \leqslant 1, 0 \leqslant \lambda_2 \leqslant 1$.

where

$$S_m = \sum_{l=1}^{m}[E(Y_l) - T]^2 \tag{19}$$

The value of m in (19) is the run number until the transient effect is measured. By using the drift-compensatory approach, the transient response will only take effect at the beginning runs of cycle 0, therefore m will be smaller than j_1. To determinate a good value for m, we need to select this parameter according to the transient effect of the output.

In order to find the optimal discount factors, the explicit form of (18) and (19) should be derived.

The expectation and variance of Y_n can be got from (9):

$$E(Y_n) = \varphi_1^{n-1}\alpha_1 + [\varphi_1^{n-1}(\xi_1 - 1) + 1]T_1 - \sum_{k=0}^{n-3}\varphi_1^{k}D_{n-2-k}\xi_1\lambda_1 + \frac{1-\varphi_1^{n}}{1-\varphi_1}\delta \tag{20}$$

$$Var(Y_n) = \left\{[(\varphi_1-\theta)^2+1]\varphi_1^{2(n-2)} - 2\theta\varphi_1^{2(n-5)} + \frac{(1+\theta^2)(1-\varphi_1^{2(n-2)}) - 2\theta\varphi_1(1-\varphi_1^{2(n-3)})}{1-\varphi_1^2}\right\}\sigma^2 \tag{21}$$

Substituting (20) into (19), we have the following equation:

$$S_m = \sum_{l=2}^{m}\left[\varphi_1^{l-1}\alpha_1 + \varphi_1^{l-1}(\xi_1 - 1)T_1 - \sum_{k=0}^{l-3}\varphi_1^{k}D_{l-2-k}\xi_1\lambda_1 + \frac{1-\varphi_1^{l}}{1-\varphi_1}\delta\right]^2 \tag{22}$$

From (9), we know that $E(Y_1)$ is not a functions of λ_1, λ_2, so it is omitted here.

Result 4. The asymptotic variance (AV) of product 1 is

$$\lim_{\substack{t\to\infty\\ n\to\infty}} Var(Y_{it+n}) = \left(\frac{1-2\theta\varphi_1+\theta^2}{1-\varphi_1^2}\right)\sigma^2 \tag{23}$$

Proof. See Appendix C for details.

From (23) it is obvious that the AV of product 1 is not a function of δ. Substituting (22) and (23) into (18), we have the following optimal problem:

Objective function:

$$\min_{\lambda_1,\lambda_2}\left\{\frac{1}{m-1}\sum_{l=2}^{m}[\varphi_1^{l-1}\alpha_1+\varphi_1^{l-1}(\xi_1-1)T_1 - \sum_{k=0}^{l-3}\varphi_1^k D_{l-2-k}\xi_1\lambda_1+\frac{1-\varphi_1^l}{1-\varphi_1}\delta]^2\left(\frac{1-2\theta\varphi_1+\theta^2}{1-\varphi_1^2}\right)\sigma^2\right\} \tag{24}$$

subject to: $0 \leqslant \lambda_1 \leqslant 1, 0 \leqslant \lambda_2 \leqslant 1$ and $|\varphi_1| < 1$.

Solutions of this optimization problem can be achieved by the grid search method that evaluates function (24) over a grid of values in the (λ_1, λ_2) plane.

4. Fault-tolerant approach

Faults are common in semiconductor manufacturing industry. Generally when faults happen is unknown, however, the faults will bring destruction to the process. In this section, two kinds of fault – the step fault and the ramp fault are considered and fault tolerant (FT) approach is investigated.

4.1. The step fault

If a step fault f_s happens at the h_{th} run, we have

$$f_s = \begin{cases} f, & s \geqslant h_{\text{th}}, \\ 0, & s < h_{\text{th}}. \end{cases} \tag{25}$$

where f is the magnitude of the fault.

Let ψ_s be defined as the total of the process disturbance and step fault at the s_{th} run, i.e.,

$$\psi_s = \eta_s + f_s \tag{26}$$

Then, we have following approach.

4.1.1. Drift-compensatory approach in present of step fault

After a step fault happens, the compensated output of product 1 at the n_{th} run in cycle $t(t \geqslant 1)$ should be compensated as follows:

$$\begin{aligned}(\hat{Y}_{it+n})_{sf} = Y_{it+n} + \Bigg(&\varphi_1^{n-1}\sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-1)+j_1-k-1} + \gamma_2\varphi_1^{n-2}D_{i(t-1)+j_1} \\ &+\sum_{k=0}^{n-3}\varphi_1^k D_{it+n-2-k}\Bigg)\xi_1\lambda_1 - \left[\varphi_1^{n-1}\left(\frac{\varphi_1}{1-\varphi_1}+i-j_1+1\right)\right. \\ &\left.+\frac{1-\varphi_1^{n-1}}{1-\varphi_1}\right]\delta - \varphi_1^{n-1}(f_{it+1}-f_{i(t-1)+j_1}) \\ &-\sum_{k=0}^{n-2}\varphi_1^k(f_{it+n-k}-f_{it+n-k-1}) \\ &-\varphi_1^{n-1}\sum_{k=1}^{j_1-1}\varphi_1^k(f_{i(t-1)+j_1+1-k}-f_{i(t-1)+j_1-k})\end{aligned} \tag{27}$$

Proof. See Appendix D for details.

Remark. Assume that a step fault happens at the hth run in cycle $t(t \geqslant 1)$, if we still use the drift-compensatory approach which does not consider the fault (as presented in Section 3.2), it can be concluded from (27) that the process output at the $(h+k)$th run in cycle t will be $\varphi_1^k f$ away from target, and $\varphi_1^{n-1}\frac{\varphi_1^{j_1}-\varphi_1^{j_1+2-h}}{1-\varphi_1}f$ away from target at the nth run in cycle $t+1$, however, the output will hit the target after the $(t+1)$th cycle. If the value of f in (25) is unknown, we have the following theorem to estimate f.

Theorem 1. *If a step fault happens at the h_{th} run in cycle $t(t \geqslant 1)$, then the estimated magnitude of the step fault is*

$$\hat{f} = (\hat{Y}_{it+h})_{nf} - T_1 \tag{28}$$

Proof. See Appendix E for details.

Remark. $\hat{f}$ includes the output $N1_{\varepsilon_{it+h}}$, which is caused by instant noise ε_{it+h}. So the actual value of the fault is $\hat{f} - N1_{\varepsilon_{it+h}}$. Since the value of $N1_{\varepsilon_{it+h}}$ is usually small compared with the value of the fault, the influence of $N1_{\varepsilon_{it+h}}$ can be neglected.

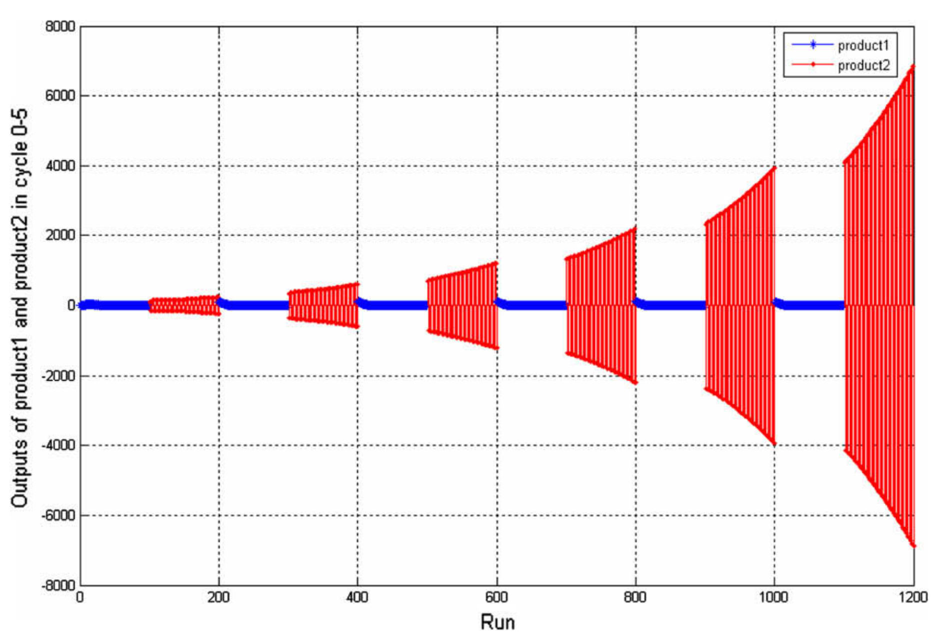

Fig. 4. Outputs of product 1 and product 2 with $(\lambda_1, \lambda_2) = (0.1, 0.1)$ and $(\omega_1, \omega_2) = (0.55, 0.9)$.

4.2. The ramp fault

If a ramp fault f_s occurs at the hth run, then its mathematical expression can be written as

$$f_s = \begin{cases} f(s-h+1), & s \geqslant h, \\ 0, & s < h. \end{cases} \tag{29}$$

where the slope of the ramp is f. The FT approach is discussed in following two cases:

Case A. f is known. We have the following FT approach, i.e.,

4.2.1. Drift-compensatory approach in present of ramp fault

If f is known, the output of product 1 at the nth run in cycle $t(t \geqslant 1)$ should be compensated as follows:

$$(\hat{Y}_{it+n})_{rf} = Y_{it+n} + \gamma_2 \varphi_1^{n-2} D_{i(t-1)+j_1} \xi_1 \lambda_1 + H - L + \varphi_1^{n-1}(G - E - F) \tag{30}$$

where $E = \sum_{k=1}^{j_1-1} \varphi_1^k g_{i(t-1)+j_1-k+1}$, $F = \sum_{k=j_1}^{i} g_{i(t-1)+k+1}$, $G = \sum_{k=0}^{j_1-2} \varphi_1^k \times D_{i(t-1)+j_1-k-1} \xi_1 \lambda_1$, $H = \sum_{k=0}^{n-3} \varphi_1^k D_{it+n-2-k} \xi_1 \lambda_1$, $L = \sum_{k=0}^{n-2} \varphi_1^k g_{it+n-k}$, $g_S = \delta + f_s - f_s - 1$.

Proof. See Appendix F for details.

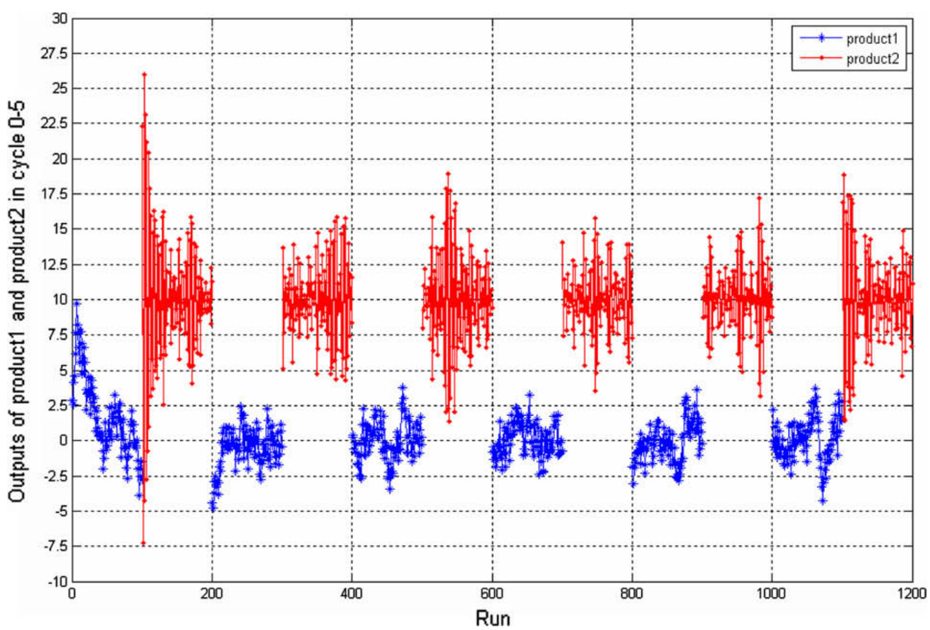

Fig. 5. Outputs of product 1 and product 2 by using the drift-compensatory approach with $(\lambda_1, \lambda_2) = (0.1, 0.1)$ and $(\omega_1, \omega_2) = (0.5, 0.5)$.

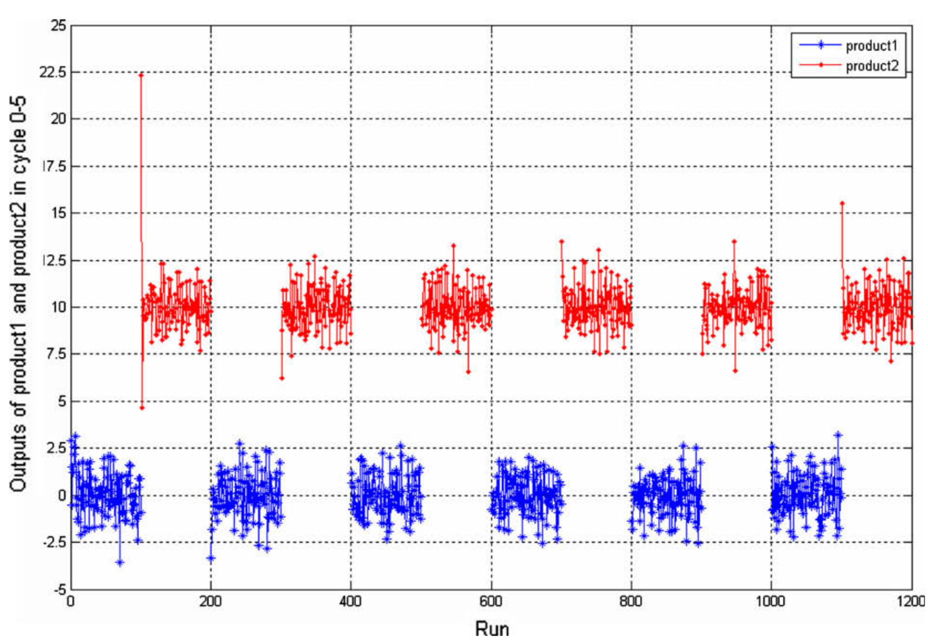

Fig. 6. Outputs of product 1 and product 2 by using the optimal drift-compensatory approach with $(\lambda_1, \lambda_2) = (0.16, 0.99)$ and $(\omega_1, \omega_2) = (0.01, 0.52)$.

Please cite this article in press as: B. Ai et al., The optimal drift-compensatory and fault tolerant approach for mixed-product run-to-run control, J. Process Contr. (2009), doi:10.1016/j.jprocont.2009.04.008

Case B. f is unknown. We can estimate f according to the following theorem.

Theorem 2. If a ramp fault happens in cycle $t(t \geqslant 1)$, then the estimated value of the slope is $\hat{f}$ and

$$\hat{f} = (\hat{Y}_{it+h})_{nf} - T_1 \tag{31}$$

Proof. See Appendix G for details.

Remark. $\hat{f}$ includes the output $N2_{\varepsilon_{it+h}}$ which is produced by instant noise ε_{it+h}. So the actual value of the fault is $\hat{f} - N2_{\varepsilon_{it+h}}$. It is clear from (31) that we have overcompensated $-[\frac{1}{1-\varphi_1} + \varphi_1^{n-1}(i-j_1)]N2_{\varepsilon_{it+h}}$ for the process outputs. The overcompensation will lead the process outputs off target at the beginning runs of cycle $t+1$, $t+2, \ldots$. However, since $N2_{\varepsilon_{it+h}}$ is small compared with the value of the fault and $|\varphi_1| < 1$, the overcompensation can be neglected with the increase of n.

5. Simulation study

In this section, several simulation examples are provided to illustrate few arguments presented above. The MSE and variance of the products outputs are used to evaluate the performance of the simulations.

We consider a case that two products, i.e., product 1 and product 2, are manufactured on the same tool with the true parameters: $(\alpha_1, \alpha_2) = (1, 0.8)$, $(\beta_1, \beta_2) = (0.6, 1)$, $\theta = 0.5$, $\sigma = 1$, $\delta = 1$, $j_1 = 100$, and $i = 200$. Assume that the least square estimate of (β_1, β_2) are $(b_1, b_2) = (1, 0.6)$, and the desired target of product 1 and product 2 are $(T_1, T_2) = (0, 10)$.

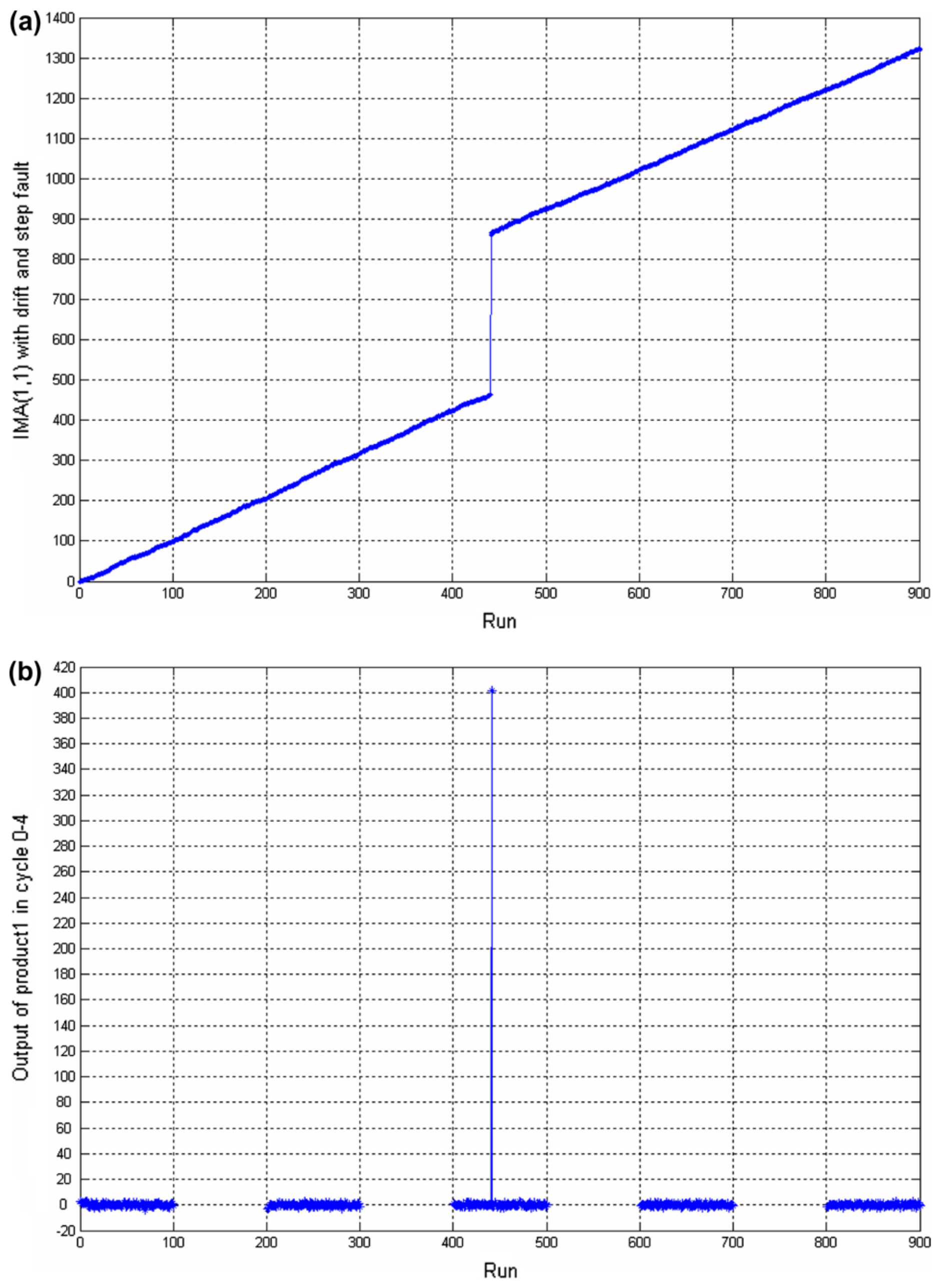

Fig. 7. Output of product 1 with a step fault happens at the 441st run.

As mentioned before $\lambda_1, \lambda_2 \in [0, 1]$, so in this example, we select (λ_1, λ_2) as $(0.1, 0.1)$, and $(\omega_1, \omega_2) = (0.55, 0.9)$, where (ω_1, ω_2) are discount factors of product 2's threaded-dEWMA controller. So the calculated $\varphi_1 = 1 - \xi_1(\lambda_1 + \lambda_2) = 0.88 < 1$, $\phi_2 = 1 - \xi_2(\omega_1 + \omega_2) = -1.4167 < -1$. Fig. 4 shows simulated outputs of product 1 and product 2 in cycle 0–5. It is obvious that the output of product 1 is convergent while the output of product 2 is divergent. The simulation result fit with our interpretation of the Proposition.

Fig. 5 shows simulated outputs of product 1 and product 2 in cycle 0–5 by using the drift-compensatory approach without optimal discount factors selection. The simulation result shows that this approach can help eliminating the large deviation at the beginning runs of cycle 1-5. However, the MSE and variance for both products in each cycle are still very large. MSE of product 1 and product 2 in each cycle are: [10.767 2.301 1.877 1.359 1.930 2.380] and [27.956 8.803 13.896 5.886 6.554 13.079]; variance for each cycle are: [7.709 2.027 1.895 1.354 1.922 2.394] and [28.238 8.892 14.036 5.946 6.620 13.211] respectively. Average MSE (AMSE) of product 1 and product 2 are 3.435 and 12.696; Average variances (AV) are 2.884 and 12.824 respectively.

If optimal discount factors selection is adopted, then we will have the simulated outputs for each product in cycle 0–5 (as illustrated in Fig. 6). By solving Eq. (24), the optimal discount factors of product 1 and product 2 are $(\lambda_1, \lambda_2) = (0.16, 0.99)$ and $(\omega_1, \omega_2) = (0.01, 0.52)$, respectively. The large deviation at the beginning runs of each cycle is also reduced. Furthermore, the MSE of product 1 and product 2 in each cycle are: [1.482 1.308 1.222 1.102 1.036 1.254] and [2.815 1.431 1.312 1.285 1.210 1.532]; variance in each cycle are: [1.490 1.317 1.234 1.113 1.043 1.262] and [2.841 1.443 1.325 1.297 1.221 1.548] respectively. AMSE for product 1 and product 2 are 1.234 and 1.597; AV are 1.243 and 1.612 respectively. MSE and variance for both products in each cycle are decreased greatly compared with that in Fig. 5. Each cycle, MSE is reduced by: [86.2% 43.2% 34.9% 18.9% 46.3% 47.3%] for product 1 and [89.9% 83.7% 90.5% 78.2% 81.5% 88.3%] for product 2; The variance is also reduced by: [80.7% 35.0% 34.9% 17.9% 45.7% 47.3%] and [89.9% 83.8% 90.6% 78.2% 81.6% 88.3%] respectively. Finally AMSE is reduced by 64.1% for product 1, and 87.4% for product 2; AV is reduced by 56.9% and 87.4% respectively.

For the drifted process, if a step fault with the magnitude equals to 400 occurs at the 441st run, the drift-compensatory approach in present of step fault is used. The simulation result of product 1 is given in Fig. 7. Fig. 7a shows the process disturbance and the step fault. Fig. 7b shows the compensated output of product 1. It is obvious that the proposed FT method is very effective to reduce this kind of fault.

If a ramp fault with the slope of 10 occurs at the 441st run for the drifted process, we have simulation results as illustrated in Fig. 8. In Fig. 8a, the ramp fault and the process disturbance are shown. When the ramp fault occurs, if we only use the optimal drift-compensatory approach which does not consider the fault, then the fault will lead the process outputs off specification for a considerable runs (as shown in Fig. 8b). If the slope of the ramp is known (Case A in Section 4), the output of product 1 is presented in Fig. 8c by using the drift-compensatory approach in present of ramp fault. It is obvious that the proposed method is very effective to reduce the ramp fault. On the other hand, if the value of the fault is unknown (Case B in Section 4), the output of product 1 is given in Fig. 8d by using the FT approach in Theorem 2. As was described in the remark of Theorem 2, the process outputs will be off target in the first few runs of cycle 3 and cycle 4. However, the number of this abnormal runs is only 3 in Fig. 8d. Therefore to reduce the loss, test wafers can be used here.

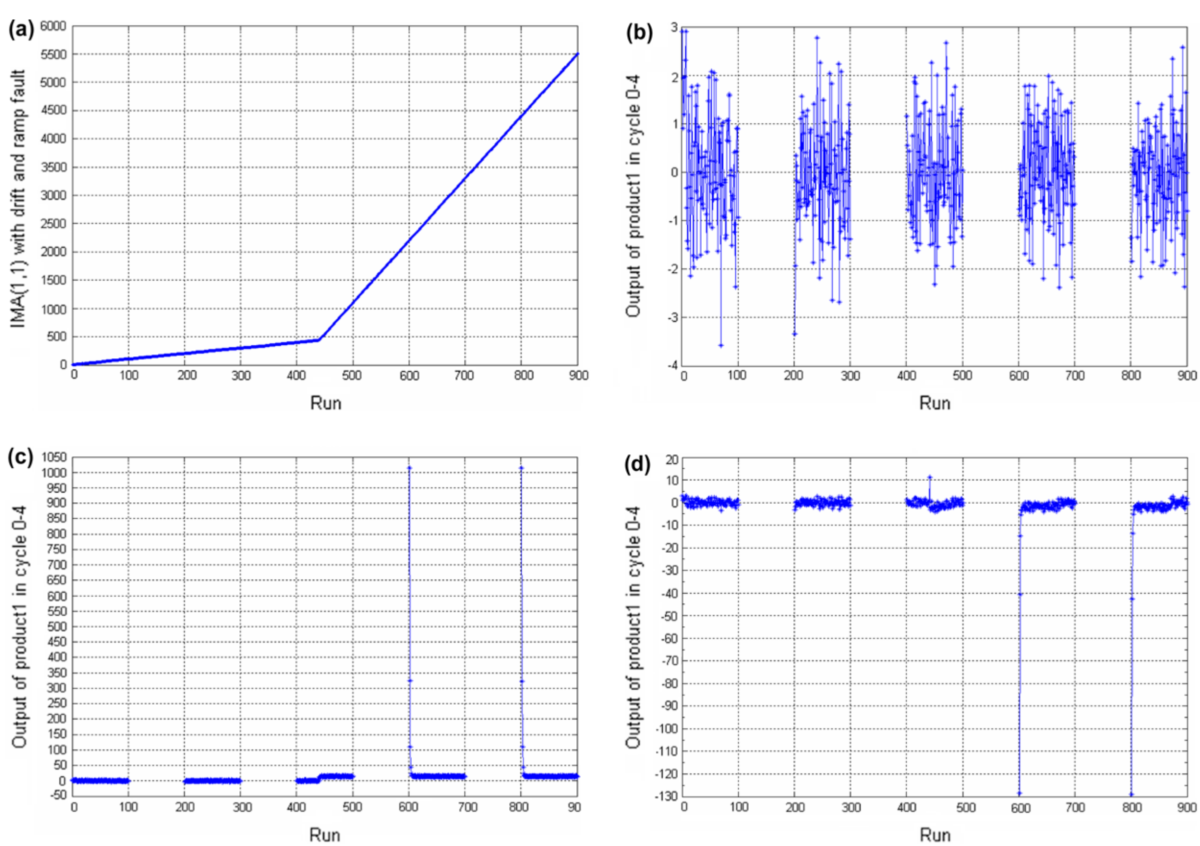

Fig. 8. Output of product 1 with a ramp fault happens at the 441st run.

Please cite this article in press as: B. Ai et al., The optimal drift-compensatory and fault tolerant approach for mixed-product run-to-run control, J. Process Contr. (2009), doi:10.1016/j.jprocont.2009.04.008

6. Conclusions

In this paper, we studied the mixed-product drifted process with IMA(1, 1) disturbance as well as the step and the ramp fault in semiconductor manufacturing batch process. Based on a threaded-dEWMA controller, the process outputs of each run in each cycle are derived. The drift-compensatory approach is proposed to deal with the large deviations at the beginning runs of cycle 1,2, …. In order to minimize the MSE and variance of the output, the "trade-off" solution is adopted. Moreover, we also considered two kinds of common fault – the step fault and the ramp fault, corresponding fault tolerant approach is introduced to deal with the fault. Simulation study shows that the approach is very effective.

Acknowledgement

The authors thank the financial support for this work from Chinese National Natural Science Foundation (60604030). The authors also thank the three referees and the editors for their helpful suggestions and comments that have significantly improved both the quality and the presentation of this paper.

Appendix A

Combining (1), (3), (4), and (5), we have

$$Y_{it+n}=\begin{cases}\alpha_1+\xi_1T_1+\eta_1, & n=1 \ \ \text{and} \ \ t=0,\\ \alpha_1+\xi_1T_1(\phi_1+\lambda_1+\lambda_2)-\xi_1(\lambda_1+\lambda_2)(\alpha_1+\eta_1)+\eta_2, & n=2 \ \ \text{and} \ \ t=0,\\ \phi_1Y_{i(t-1)+j_1}+\xi_1(\lambda_1+\lambda_2)T_1-\xi_1\lambda_1D_{i(t-1)+j_1-1}+\eta_{it+1}-\eta_{i(t-1)+j_1}, & n=1 \ \ \text{and} \ \ t\geqslant 1,\\ \phi_1Y_{it+1}+\xi_1(\lambda_1+\lambda_2)T_1-\xi_1\lambda_1D_{i(t-1)+j_1}+\eta_{it+2}-\eta_{it+1}, & n=2 \ \ \text{and} \ \ t\geqslant 1,\\ \phi_1Y_{it+n-1}+\xi_1(\lambda_1+\lambda_2)T_1-\xi_1\lambda_1D_{it+n-2}+\eta_{it+n}-\eta_{it+n-1}, & n=3,4,\ldots,j_1.\end{cases} \tag{A.1}$$

Set $t = 0$ in (A.1), and use iterative method, then

$$\begin{aligned}Y_n &= \varphi_1Y_{n-1}+\xi_1(\lambda_1+\lambda_2)T_1-D_{n-2}\xi_1\lambda_1+\eta_n-\eta_{n-1}\\ &=\varphi_1^2Y_{n-2}+\sum_{k=0}^{1}\varphi_1^k\xi_1(\lambda_1+\lambda_2)T_1-\sum_{k=0}^{1}\varphi_1^kD_{n-2-k}\xi_1\lambda_1\\ &\quad+\sum_{k=0}^{1}\varphi_1^k(\eta_{n-k}-\eta_{n-k-1})\\ &=\varphi_1^{n-2}Y_2+\sum_{k=0}^{n-3}\varphi_1^k\xi_1(\lambda_1+\lambda_2)T_1-\sum_{k=0}^{n-3}\varphi_1^kD_{n-2-k}\xi_1\lambda_1\\ &\quad+\sum_{k=0}^{n-3}\varphi_1^k(\eta_{n-k}-\eta_{n-k-1})\\ &=\varphi_1^{n-1}Y_1+\sum_{k=0}^{n-2}\varphi_1^k\xi_1(\lambda_1+\lambda_2)T_1-\sum_{k=0}^{n-2}\varphi_1^kD_{n-2-k}\xi_1\lambda_1\\ &\quad+\sum_{k=0}^{n-2}\varphi_1^k(\eta_{n-k}-\eta_{n-k-1})\\ &=\varphi_1^{n-1}(\alpha_1+\xi_1T_1+\eta_1)+\sum_{k=0}^{n-2}\varphi_1^k\xi_1(\lambda_1+\lambda_2)T_1\\ &\quad-\sum_{k=0}^{n-2}\varphi_1^kD_{n-2-k}\xi_1\lambda_1+\sum_{k=0}^{n-2}\varphi_1^k(\eta_{n-k}-\eta_{n-k-1})\end{aligned} \tag{A.2}$$

Appendix B

Setting $n = 1$ and $t = 1$ in (A.1), by the iterative method, we have

$$\begin{aligned}Y_{i+1} &= \varphi_1Y_{j_1}+\xi_1(\lambda_1+\lambda_2)T_1-D_{j_1-1}\xi_1\lambda_1+\eta_{i+1}-\eta_{j_1}\\ &=\varphi_1^{j_1-1}[\alpha_1+(\varphi_1+\lambda_1+\lambda_2)\xi_1T_1-\xi_1(\lambda_1+\lambda_2)(\alpha_1+\eta_1)+\eta_2]\\ &\quad+\sum_{k=0}^{j_1-3}\varphi_1^{k+1}\xi_1(\lambda_1+\lambda_2)T_1+\xi_1(\lambda_1+\lambda_2)T_1\\ &\quad-\sum_{k=0}^{j_1-3}\varphi_1^{k+1}D_{j_1-k-2}\xi_1\lambda_1-D_{j_1-1}\xi_1\lambda_1\\ &\quad+\sum_{k=0}^{j_1-3}\varphi_1^{k+1}(\eta_{j_1-k}-\eta_{j_1-k-1})+(\eta_{i+1}-\eta_{j_1})\end{aligned} \tag{B.1}$$

Set $n = j_1$ in (A.1), then

$$\begin{aligned}Y_{it+j_1} &= \varphi_1Y_{it+j_1-1}+\xi_1(\lambda_1+\lambda_2)T_1-D_{it+j_1-2}\xi_1\lambda_1+\eta_{it+j_1}-\eta_{it+j_1-1}\\ &=\varphi_1^{j_1-2}Y_{it+2}+\sum_{k=0}^{j_1-3}\varphi_1^k[\xi_1(\lambda_1+\lambda_2)T_1\\ &\quad-D_{it+j_1-k-2}\xi_1\lambda_1+\eta_{it+j_1-k}-\eta_{it+j_1-k-1}]\\ &=\varphi_1^{j_1-1}Y_{it+1}+\sum_{k=0}^{j_1-2}\varphi_1^k[\xi_1(\lambda_1+\lambda_2)T_1+\eta_{it+j_1-k}-\eta_{it+j_1-k-1}]\\ &\quad-(\sum_{k=0}^{j_1-3}\varphi_1^kD_{it+j_1-k-2}+\varphi_1^{j_1-2}D_{i(t-1)+j_1})\xi_1\lambda_1\end{aligned} \tag{B.2}$$

for $t \geqslant 1$.

Combining (A.1) and (B.2), we have

$$\begin{aligned}Y_{it+i+1} &= \varphi_1Y_{it+j_1}+\xi_1(\lambda_1+\lambda_2)T_1-D_{it+j_1-1}\xi_1\lambda_1+\eta_{i(t+1)+1}-\eta_{it+j_1}\\ &=\varphi_1^{j_1}Y_{it+1}+\sum_{k=0}^{j_1-2}\varphi_1^{k+1}[\xi_1(\lambda_1+\lambda_2)T_1+\eta_{it+j_1-k}-\eta_{it+j_1-k-1}]\\ &\quad-\sum_{k=0}^{j_1-3}\varphi_1^{k+1}D_{it+j_1-k-2}\xi_1\lambda_1\xi_1(\lambda_1+\lambda_2)T_1\\ &\quad-\varphi_1^{j_1-1}D_{i(t-1)+j_1}\xi_1\lambda_1-D_{it+j_1-1}\xi_1\lambda_1+\eta_{i(t+1)+1}-\eta_{it+j_1}\\ &=\varphi_1^{j_1t}Y_{i+1}+\sum_{m=0}^{t-1}\varphi_1^{mj_1}\Bigg\{\sum_{k=1}^{j_1-1}\varphi_1^k[\xi_1(\lambda_1+\lambda_2)T_1+\eta_{i(t-m)+j_1+1-k}\\ &\quad-\eta_{i(t-m)+j_1-k}]+\xi_1(\lambda_1+\lambda_2)T_1-\sum_{k=0}^{j_1-2}\varphi_1^kD_{i(t-m)+j_1-k-1}\xi_1\lambda_1\\ &\quad-\varphi_1^{j_1-1}D_{i(t-m-1)+j_1}\xi_1\lambda_1+\eta_{i(t+1-m)+1}-\eta_{i(t-m)+j_1}\Bigg\}\\ &=\varphi_1^{j_1t+1}Y_{j_1}+\varphi_1^{j_1t}\xi_1(\lambda_1+\lambda_2)T_1-\varphi_1^{j_1t}D_{j_1-1}\xi_1\lambda_1+\varphi_1^{j_1t}(\eta_{i+1}-\eta_{j_1})\\ &\quad+\sum_{m=0}^{t-1}\varphi_1^{mj_1}\Bigg\{\sum_{k=1}^{j_1-1}\varphi_1^k[\xi_1(\lambda_1+\lambda_2)T_1+\eta_{i(t-m)+j_1+1-k}-\eta_{i(t-m)+j_1-k}]\\ &\quad+\xi_1(\lambda_1+\lambda_2)T_1-\xi_1\lambda_1\sum_{k=0}^{j_1-2}\varphi_1^kD_{i(t-m)+j_1-k-1}\\ &\quad-\varphi_1^{j_1-1}D_{i(t-m-1)+j_1}\xi_1\lambda_1+\eta_{i(t+1-m)+1}-\eta_{i(t-m)+j_1}\Bigg\}\\ &=\varphi_1^{j_1(t+1)-1}[\alpha_1+(\varphi_1+\lambda_1+\lambda_2)\xi_1T_1-\xi_1(\lambda_1+\lambda_2)(\alpha_1+\eta_1)+\eta_2]\\ &\quad+\varphi_1^{j_1t+1}\sum_{k=0}^{j_1-3}\varphi_1^k[\xi_1(\lambda_1+\lambda_2)T_1+(\eta_{j_1-k}-\eta_{j_1-k-1})-D_{j_1-k-2}\xi_1\lambda_1]\end{aligned}$$

Please cite this article in press as: B. Ai et al., The optimal drift-compensatory and fault tolerant approach for mixed-product run-to-run control, J. Process Contr. (2009), doi:10.1016/j.jprocont.2009.04.008

$$+\varphi_1^{j_1 t}\xi_1(\lambda_1+\lambda_2)T_1-\varphi_1^{j_1 t}D_{j_1-1}\xi_1\lambda_1+\varphi_1^{j_1 t}(\eta_{i+1}-\eta_{j_1}) + \sum_{m=0}^{t-1}\varphi_1^{mj_1}\left\{\sum_{k=1}^{j_1-1}\varphi_1^k[\xi_1(\lambda_1+\lambda_2)T_1+\eta_{i(t-m)+j_1+1-k}-\eta_{i(t-m)+j_1-k}] + \xi_1(\lambda_1+\lambda_2)T_1-\sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-m)+j_1-k-1}\xi_1\lambda_1 - \varphi_1^{j_1-1}D_{i(t-m-1)+j_1}\xi_1\lambda_1+\eta_{i(t+1-m)+1}-\eta_{i(t-m)+j_1}\right\} \tag{B.3}$$

From (B.1) and (B.3), we have

$$Y_{it+1}=\varphi_1^{j_1 t-1}[\alpha_1+\xi_1T_1(\varphi_1+\lambda_1+\lambda_2)-\xi_1(\lambda_1+\lambda_2)(\alpha_1+\eta_1)+\eta_2] + \varphi_1^{j_1(t-1)+1}\sum_{k=0}^{j_1-3}\varphi_1^k[\xi_1(\lambda_1+\lambda_2)T_1+(\eta_{j_1-k}-\eta_{j_1-k-1}) - D_{j_1-k-2}\xi_1\lambda_1]+\varphi_1^{j_1(t-1)}\xi_1(\lambda_1+\lambda_2)T_1 - \varphi_1^{j_1(t-1)}D_{j_1-1}\xi_1\lambda_1+\varphi_1^{j_1(t-1)}(\eta_{i+1}-\eta_{j_1}) + \sum_{m=0}^{t-2}\varphi_1^{mj_1}\left\{\sum_{k=1}^{j_1-1}\varphi_1^k[\xi_1(\lambda_1+\lambda_2)T_1+\eta_{i(t-m-1)+j_1+1-k} - \eta_{i(t-m-1)+j_1-k}]+\xi_1(\lambda_1+\lambda_2)T_1 - \sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-m-1)+j_1-k-1}\xi_1\lambda_1-\varphi_1^{j_1-1}D_{i(t-m-2)+j_1}\xi_1\lambda_1 + \eta_{i(t-m)+1}-\eta_{i(t-m-1)+j_1}\right\} \tag{B.4}$$

for $t \geqslant 1$.

Combining (A.1) and (B.4), we have

$$Y_{it+n}=\varphi_1 Y_{it+n-1}+\xi_1(\lambda_1+\lambda_2)T_1-D_{it+n-2}\xi_1\lambda_1+\eta_{it+n}-\eta_{it+n-1} = \varphi_1^{n-1}Y_{it+1}+\sum_{k=0}^{n-2}\varphi_1^k\xi_1(\lambda_1+\lambda_2)T_1 + \sum_{k=0}^{n-2}\varphi_1^k(\eta_{it+n-k}-\eta_{it+n-k-1})-(\gamma_2\varphi_1^{n-2}D_{i(t-1)+j_1} + \sum_{k=0}^{n-3}\varphi_1^k D_{it+n-k-2})\xi_1\lambda_1 \tag{B.5}$$

where $\xi_1=\frac{\beta_1}{b_1}$, $\varphi_1=1-\xi_1(\lambda_1+\lambda_2)$, $\gamma_2=\begin{cases}0, & n<2,\\ 1, & n\geqslant 2.\end{cases}$.

Appendix C

Taking the mathematical variance for (11), we have

$$Var(Y_{it+n})=Var[\varphi_1^{n-1}Y_{it+1}+\sum_{k=0}^{n-2}\varphi_1^k(\varepsilon_{it+n-k}-\theta\varepsilon_{it+n-k-1})] = Var(\varphi_1^{n-1}Y_{it+1})+Var(\sum_{k=0}^{n-2}\varphi_1^k\varepsilon_{it+n-k}) + Var(\theta\sum_{k=0}^{n-2}\varphi_1^k\varepsilon_{it+n-k-1})+2Cov(\varphi_1^{n-1}Y_{it+1},\sum_{k=0}^{n-2}\varphi_1^k\varepsilon_{it+n-k}) - 2Cov(\theta\varphi_1^{n-1}Y_{it+1},\sum_{k=0}^{n-2}\varphi_1^k\varepsilon_{it+n-k-1}) - 2Cov(\theta\sum_{k=0}^{n-2}\varphi_1^k\varepsilon_{it+n-k},\sum_{k=0}^{n-2}\varphi_1^k\varepsilon_{it+n-k-1}) = \varphi_1^{2(n-1)}Var(Y_{it+1})+[(1+\theta^2)\frac{1-\varphi_1^{2(n-1)}}{1-\varphi_1^2} - 2\varphi_1^{2n-3}\theta-2\theta\frac{\varphi_1(1-\varphi_1^{2(n-2)})}{1-\varphi_1^2}]\sigma^2 \tag{C.1}$$

Calculating limit for (C.1), we get

$$\lim_{\substack{t\to\infty\\ n\to\infty}} Var(Y_{it+n})=\left(\frac{1-2\theta\varphi_1+\theta^2}{1-\varphi_1^2}\right)\sigma^2 \tag{C.2}$$

Appendix D

Combining (1)–(5), (25), and (26) we have

$$E(Y_{it+1})=T_1+(\frac{\varphi_1}{1-\varphi_1}+i+j_1-1)\delta - \sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-1)+j_1-k-1}\xi_1\lambda_1+\sum_{k=1}^{j_1-1}\varphi_1^k(f_{i(t-1)+j_1+1-k} - f_{i(t-1)+j_1-k})+f_{it+1}-f_{i(t-1)+j_1} \tag{D.1}$$

$$E(Y_{it+n})=\varphi_1^{n-1}E(Y_{it+1})+(1-\varphi_1^{n-1})T_1 + \sum_{k=0}^{n-2}\varphi_1^k(\delta+f_{it+n-k}-f_{it+n-k-1})-(\gamma_2\varphi_1^{n-2}D_{i(t-1)+j_1} + \sum_{k=0}^{n-3}\varphi_1^k D_{it+n-2-k})\xi_1\lambda_1 = T_1-\left(\varphi_1^{n-1}\sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-1)+j_1-k-1}+\gamma_2\varphi_1^{n-2}D_{i(t-1)+j_1} + \sum_{k=0}^{n-3}\varphi_1^k D_{it+n-2-k}\right)\xi_1\lambda_1 + \left[\varphi_1^{n-1}\left(\frac{\varphi_1}{1-\varphi_1}+i-j_1+1\right)+\frac{1-\varphi_1^{n-1}}{1-\varphi_1}\right]\delta - \varphi_1^{n-1}(f_{it+1}-f_{i(t-1)+j_1})+\varphi_1^{n-1}\sum_{k=1}^{j_1-1}\varphi_1^k(f_{i(t-1)+j_1+1-k} - f_{i(t-1)+j_1-k})+\sum_{k=0}^{n-2}\varphi_1^k(f_{it+n-k}-f_{it+n-k-1}) \tag{D.2}$$

Appendix E

Combine (25) and (27), and set $n = h$, then

$$T_1=E((\hat{Y}_{it+h})_{sf}) = E(Y_{it+h})-\left[\varphi_1^{h-1}\left(\frac{\varphi_1}{1-\varphi_1}+i-j_1+1\right)+\frac{1-\varphi_1^{h-1}}{1-\varphi_1}\right]\delta - \varphi_1^{h-1}(f_{it+1}-f_{i(t-1)+j_1})-\sum_{k=0}^{h-2}\varphi_1^k(f_{it+h-k}-f_{it+h-k-1}) - \varphi_1^{h-1}\sum_{k=1}^{j_1-1}\varphi_1^k(f_{i(t-1)+j_1+1-k}-f_{i(t-1)+j_1-k}) + (\varphi_1^{h-1}\sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-1)+j_1-k-1}+\gamma_2\varphi_1^{h-2}D_{i(t-1)+j_1}+\sum_{k=0}^{h-3}\varphi_1^k D_{it+h-2-k})\xi_1\lambda_1 = E(Y_{it+h})-\left[\varphi_1^{h-1}\left(\frac{\varphi_1}{1-\varphi_1}+i-j_1+1\right)+\frac{1-\varphi_1^{h-1}}{1-\varphi_1}\right]\delta-f + \left(\varphi_1^{h-1}\sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-1)+j_1-k-1}+\gamma_2\varphi_1^{h-2}D_{i(t-1)+j_1}+\sum_{k=0}^{h-3}\varphi_1^k D_{it+h-2-k}\right)\xi_1\lambda_1 \tag{E.1}$$

Subtract (E.1) from (17), and define $n = h$, then

$$(\hat{Y}_{it+h})_{nf}-T_1=Y_{it+h}-E(Y_{it+h})+f=N1_{\varepsilon_{it+h}}+f=\hat{f} \tag{E.2}$$

where $N1_{\varepsilon_{it+h}}$ is the process output which is produced by instant noise ε_{it+h}.

Appendix F

Combining (2), (26), and (29), we have

$$\psi_s - \psi_{s-1} = \varepsilon_s - \theta\varepsilon_{s-1} + \delta + f_s - f_{s-1} = \varepsilon_s - \theta\varepsilon_{s-1} + g_s \quad \text{(F.1)}$$

Combining (1)–(5), and (F.1) we get the following equation after considerable algebraic manipulation:

$$E(Y_{it+1}) - T_1 = E + F - G \quad \text{(F.2)}$$

$$E(Y_{it+2}) - T_1 = \varphi_1(E + F - G) - D_{i(t-1)+j_1}\xi_1\lambda_1 + g_{it+2} \quad \text{(F.3)}$$

$$E(Y_{it+n}) - T_1 = \varphi_1^{n-1}(E + F - G) + L - (H + \varphi_1^{n-2}D_{i(t-1)+j_1})\xi_1\lambda_1 \quad \text{(F.4)}$$

where $n \geqslant 3$.

The following equation can be derived directly from equation (F.2)–(F.4):

$$E(Y_{it+n}) - T_1 = \varphi_1^{n-1}(E + F - G) + L - \gamma_2\varphi_1^{n-2}D_{i(t-1)+j_1}\xi_1\lambda_1 - H \quad \text{(F.5)}$$

where $E = \sum_{k=1}^{j_1-1}\varphi_1^k g_{i(t-1)+j_1-k+1}, F = \sum_{k=j_1}^{i} g_{i(t-1)+k+1}, G = \sum_{k=0}^{j_1-2}\varphi_1^k \times D_{i(t-1)+j_1-k-1}\xi_1\lambda_1, H = \sum_{k=0}^{n-3}\varphi_1^k D_{it+n-2-k}\xi_1\lambda_1, L = \sum_{k=0}^{n-2}\varphi_1^k g_{it+n-k}$.Drift-compensatory approach in present of ramp fault can be concluded from (F.5).

Appendix G

If the value of the fault is known, we conclude from drift-compensatory approach in present of ramp fault that the output at the h_{th} run in cycle t is $(\hat{Y}_{it+h})_{rf}$. It considers both the drift and the ramp fault, and $(\hat{Y}_{it+h})_{rf}$ can be got by (30):

$$\begin{aligned}(\hat{Y}_{it+h})_{rf} = {} & Y_{it+h} + [\gamma_2\varphi_1^{h-2}D_{i(t-1)+j_1} + \sum_{k=0}^{h-3}\varphi_1^k D_{it+h-2-k}]\xi_1\lambda_1 \\ & - \sum_{k=0}^{h-2}\varphi_1^k g_{it+h-k} + \varphi_1^{h-1}\Bigg[\sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-1)+j_1-k-1}\xi_1\lambda_1 \\ & - \sum_{k=j_1}^{i} g_{i(t-1)+k+1} - \sum_{k=1}^{j_1-1}\varphi_1^k g_{i(t-1)+j_1-k+1}\Bigg] \end{aligned} \quad \text{(G.1)}$$

If we ignore the instant noise at hth run in cycle t, then the output should be T_1, i.e.,

$$\begin{aligned} T_1 &= E((\hat{Y}_{it+h})_{rf}) \\ &= E(Y_{it+h}) + (\gamma_2\varphi_1^{h-2}D_{i(t-1)+j_1} + \sum_{k=0}^{h-3}\varphi_1^k D_{it+h-2-k})\xi_1\lambda_1 \\ &\quad - \sum_{k=0}^{h-2}\varphi_1^k g_{it+h-k} + \varphi_1^{h-1}(\sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-1)+j_1-k-1}\xi_1\lambda_1 \\ &\quad - \sum_{k=j_1}^{i} g_{i(t-1)+k+1} - \sum_{k=1}^{j_1-1}\varphi_1^k g_{i(t-1)+j_1-k+1}) \end{aligned} \quad \text{(G.2)}$$

Substituting (29) and (26) into (G.2), we get

$$\begin{aligned} T_1 = {} & E(Y_{it+h}) + \left(\varphi_1^{h-1}\sum_{k=0}^{j_1-2}\varphi_1^k D_{i(t-1)+j_1-k-1} + \gamma_2\varphi_1^{h-2}D_{i(t-1)+j_1}\right. \\ & \left. + \sum_{k=0}^{h-3}\varphi_1^k D_{it+h-2-k}\right)\xi_1\lambda_1 - \left[\varphi_1^{h-1}\left(\frac{\varphi_1}{1-\varphi_1} + i - j_1 + 1\right)\right]\delta \\ & + \frac{\varphi_1(1-\varphi_1^{h-2})}{1-\varphi_1}\delta - g_{it+h} \end{aligned} \quad \text{(G.3)}$$

Subtract (G.3) from (17) and set $n = h$, then

$$\begin{aligned} (\hat{Y}_{it+h})_{nf} - T_1 &= Y_{it+h} - E(Y_{it+h}) + g_{it+h} - \delta \\ &= Y_{it+h} - E(Y_{it+h}) + f + \delta - \delta \\ &= N2_{\varepsilon_{it+h}} + f \\ &= \hat{f} \end{aligned} \quad \text{(G.4)}$$

References

[1] J. Moyne, E.D. Castillo, A.M. Hurwitz, Run-to-Run Control in Semiconductor Manufacturing, CRC Press, Florida, 2001.
[2] G.E.P. Box, G.M. Jenkins, Further contributions to adaptive quality control: simultaneous estimation of dynamics: nonzero costs, Bull. Int. Stat. Inst. 34 (1963) 943–974.
[3] E. Sachs, A. Hu, A. Ingolfsson, Run by run process control: combining SPC and feedback control, IEEE Trans. Semicond. Manuf. 8 (1) (1995) 26–43.
[4] A. Ingolfsson, E. Sachs, Stability and sensitivity of an EWMA controller, J. Quality Technol. 25 (4) (1993) 271–287.
[5] T. Smith, Novel techniques for the run by run process control of chemical–mechanical polishing, M.S. Thesis, Electrical Engineering and Computer Science Dept., MIT, Cambridge, MA, 1996.
[6] J. Wang, Q.P. He, EWMA run-to-run controller with gain updating: stability and sensitivity analysis, in: American Control Conference, 2007, pp. 2872–2877.
[7] S.W. Butler, J.A. Stefani, Supervisory run-to-run control of a polysilicon gate etch using in situ ellipsometry, IEEE Trans. Semicond. Manuf. 7 (2) (1994) 193–201.
[8] P.J. Brockwell, R.A. Davis, Introduction to Time Series and Forecasting, second ed., Springer, 2001.
[9] E.D. Castillo, Long run and transient analysis of a double EWMA feedback controller, IIE Trans. 31 (1999) 1157–1169.
[10] S.T. Tseng, R.J. Chou, S.P. Lee, Statistical design of a double EWMA controller, Appl. Stochastic Models Bus. Indust. 18 (3) (2002) 313–322.
[11] S.T. Tseng, F. Tsung, P.Y. Liu, Variable EWMA run-to-run controller for drifted processes, IIE Trans. 39 (2007) 291–301.
[12] S.T. Tseng, J. Tang, C.H. Lin, Sample size determination for achieving stability of double multivariate exponentially weighted moving average controller, Technometrics 49 (4) (2007) 409–419.
[13] S.P. Lee, R.J. Chou, S.T. Tseng, Stability and performance of a double EWMA controller for drifted MIMO systems, IIE Trans. 40 (2008) 690–705.
[14] J.H. Chen, T.W. Kuo, A.C. Lee, Run-by-run process control of metal sputter deposition: combining time series and extended Kalman filter, IEEE Trans. Semicond. Manuf. 20 (3) (2007) 278–285.
[15] T.F. Edgar, S.K. Firth, C. Bode, V. Martinenz, Multi-product run-to-run control for high-mix fabs, in: AEC/APC Asia, Hsinchu, Taiwan, 2004.
[16] S.K. Firth, W.J. Campbell, A. Toprac, T.F. Edgar, Just-in-time adaptive disturbance estimation for run-to-run control of semiconductor processes, IEEE Trans. Semicond. Manuf. 19 (3) (2006) 298–315.
[17] O.A. Vanli, N.S. Patel, M. Janakiram, E.D. Castillo, Model context section for run-to-run control, IEEE Trans. Semicond. Manuf. 20 (4) (2007) 506–516.
[18] M.D. Ma, C.C. Chang, D.S.H. Wong, S.S. J, Identification of tool and product effects in a mixed product and parallel tool environment, J. Process Control 16 (2008).
[19] M.D. Ma, C.C. Chang, S.S. J, D.S.H. Wong, Mixed product run-to-run process control – an ANOVA model with ARIMA disturbance approach, J. Process Control 16 (2008).
[20] Y. Zheng, Q.H. Lin, D.S.H. Wong, S.S. J, K. Hui, Stability and performance analysis of mixed product run-to-run control, J. Process Control 16 (2006) 431–443.
[21] M.F. Wu, W.K. Lin, C.L. Ho, D.S.H. Wong, S.S. J, K. Hui, A feed-forward/feedback run-to-run control of a mixed product process: simulation and experimental studies, Industrial Eng. Chem. Res. 46 (2007) 6963–6970.