Statistical Process Control Based Supervisory Generalized Predictive Control of Thin Film Deposition Processes

Jionghua Jin[1]
Department of Industrial and Operations Engineering,
The University of Michigan,
Ann Arbor, MI 48109-2117
e-mail: jhjin@umich.edu

Huairui Guo
Department of Systems and Industrial Engineering,
The University of Arizona,
Tucson, AZ 85721-0020

Shiyu Zhou
Deparment of Industrial and Systems Engineering,
University of Wisconsin,
Madison, WI 53706

This paper presents a supervisory generalized predictive control (GPC) by combining GPC with statistical process control (SPC) for the control of the thin film deposition process. In the supervised GPC, the deposition process is described as an ARMAX model for each production run and GPC is applied to the in situ thickness-sensing data for thickness control. Supervisory strategies, developed from SPC techniques, are used to monitor process changes and estimate the disturbance magnitudes during production. Based on the SPC monitoring results, different supervisory strategies are used to revise the disturbance models and the control law in the GPC to achieve a satisfactory control performance. A case study is provided to demonstrate the developed methodology.
[DOI: 10.1115/1.2114912]

Keywords: ARMAX model, predictive control, statistical process control, supervisory strategies, thin film deposition

1 Introduction

In recent years, the need of thin film deposition has increased significantly in coating manufacturing processes. Among various coating techniques, high throughput within a coating manufacturing process can be achieved by depositing a thin film into a large area of continuous roll-by-roll flexible substrates. Although the recent development of in situ sensing provides an opportunity for online measuring of thin film thickness, the complexity of the source effusion, the heat dissipation, and the source consumption within the production run, and the long dead time present in a continuous process, thickness uniformity control is still a critical challenging research issue in process development.

This paper presents a supervisory generalized predictive control (GPC) strategy by combining the GPC with statistical process control (SPC) for the thin film deposition process control. An overview of the thin film deposition process is presented in Sec. 2, where the thin film deposition chamber structure, process variables, variation sources, current control strategy, and limitations are discussed. In Sec. 3, the framework of the supervisory GPC is proposed with the detailed discussion on the process modeling, GPC design, SPC monitoring, and supervisory strategies. A case study based on production data is given in Sec. 4. Finally, the summary and recommendations are given in Sec. 5.

2 Overview of Characteristics of Thin Film Deposition Processes

2.1 Thin Film Deposition Chamber Structure. The thin film deposition process is normally conducted in a vacuum chamber. A flexible substrate is continuously transported through several deposition zones at a constant flux. With an appropriate design of each source location, the substrate motion creates a controllable flux profile at the substrate. Figure 1 shows the locations of effusion sources in the chamber. The source is filled into an insulated source container, called the source cell, with several nozzles on the top of the source container for evaporation. In general, temperature is a critical factor affecting the effusion rate and evaporation deposition in thin film deposition processes.

The requirement for a thin film deposition process is to provide a high rate, simultaneous deposition of materials to the substrate surface and to form a dense stoichiometrical, well-adherent film on the substrate. Large area deposition and fast substrate moving speed are desirable in a practical manufacturing process in order to achieve high production throughput. However, it leads to a challenging issue of reducing the loss production because of film nonuniformity as the deposition progresses across both the width and length of the substrate. The causes of film thickness variations are extremely complex, as they are affected by source design, heat dissipation to the surrounding parts, and production conditions.

2.2 Product Quality Monitoring and Process Control. For product quality control, the conventional SPC procedure is to conduct a post-quality analysis by using an off-line measurement instrument to analyze the film layer structure and thickness. Because the major process factors impacting the thin film quality are source temperature and/or pressure, an appropriate control of these factors is used to compensate for unanticipated process disturbances during production. Examples of unanticipated source variations are cell material property, rebuilding and reinstallation of source cells between runs, impurities in sources and debris in a chamber, consumption of materials over production time, changes in characteristics of source heat dissipation to substrate and chamber, degradation of source heaters, etc. As a result, the film thickness may vary even when the source temperature or pressure is maintained at the same target setting points. In addition, it is extremely difficult to model the interaction of the source evaporation process with other process variables, such as environmental temperature and various disturbances. Thus, the compensation of disturbance cannot be achieved through either an automatic feedforward control or an off-line presetting of the source temperature or pressure.

In general, it is highly desirable for a manufacturing process to have the capability of real-time sensing and detecting process changes and of making a corresponding process adjustment to prevent or reduce defective product loss over production. This

[1]To whom correspondence should be addressed.

Contributed by the Manufacturing Engineering Division of ASME for publication in the JOURNAL OF MANUFACTURING SCIENCE AND ENGINEERING. Manuscript received September 5, 2003; final manuscript received December 15, 2004. Review conducted by C. J. Li.

Copyright © 2006 by ASME

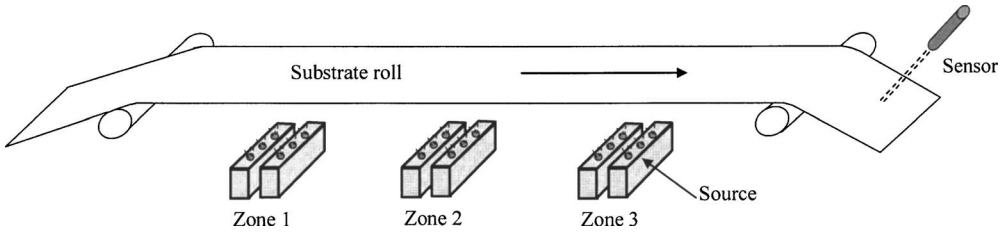


Fig. 1 Chamber structure of thin film deposition process

issue is more critical for a continuous dynamic manufacturing process when there is a need to accommodate unanticipated process disturbance, variability, or drifts during production time. The recent development of in situ thickness sensors has provided a great opportunity of achieving online monitoring and automatic process control for thin film deposition.

2.3 Process Variables and Process Variations

2.3.1 Inputs and Outputs in a Process Model. In order to develop an automatic feedback controller in a thin film process, a process model is required to describe the relationship of the in situ measurable product quality and the controllable process variables. In general, after the process setup variables (such as chamber pressure, chamber environmental temperature, and web moving speed) achieve their target values, the effusion source temperature is the most sensitive process variable directly affecting the physical effusion properties during the production run. The source temperature will be used as the controllable inputs of the process model. The output of the process model is the deposition film thickness.

2.3.2 Dead Time Variation and Uncertainty. The thin film deposition process is a continuous manufacturing process. In addition to considering the dynamic property of the source vaporization, the dead time from the temperature adjustment to the in situ thickness measurement should be considered in the process modeling and controller design. As shown in Fig. 1, the thickness measurement is generally taken at the end of the production line. The total system dead time includes the deposition time and the measurement delay due to the time needed for the substrate to move from the source location to the thickness measurement.

Based on the nominal moving speed of the substrate and the distance between source and in situ sensors, the measurement-induced dead time can be calculated for each source cell. However, the thickness measurement will reflect the compounded effect of multiple cells if one source has several cells located at different locations in a chamber, which would lead to different dead times for different cells. In addition, the substrate speed also varies around the nominal setting value over a production run, which affects both the film growth rates and substrate moving time to the thickness measurement. The inevitable variations of these factors will cause uncertainties in the dead time and estimation errors during the production, which pose a challenge in the development of a feedback controller.

2.3.3 Process Disturbances Within One Production Run. During production, there always exist significant and unanticipated disturbances leading to process variation within the run. For example, the temperature dissipation to the substrate and chamber and the consumption of source will lead to a ramp drift disturbance over the production run. A spike impulsive disturbance is normally observed because of errors in sensor measurement or data acquisition system. It is known that the control actions for a high-frequency impulsive disturbance and lower frequency linear drift disturbance should be different. An inpropriate control action will likely cause a spitting problem in the process.

2.4 Current Control Strategy and Its Limitations. The current thin film deposition control has two steps in each run: a manual process control during the initial start-up of production and an automatic feedback control during the continuous production.

In each new production run, a manual control is performed to increase the temperature through several steps. After the system response (i.e., the film thickness) is close to the target point, an automatic controller is engaged to feed back the thickness-sensing data in order to maintain the deposited film thickness at the target. The existing controller is a predesigned time-invariant controller based on a fixed model from off-line identification. Although such a controller is simple in its implementation, different disturbance structures are not considered in the controller design.

3 Supervisory Predictive Control Strategy

A supervisory control strategy is proposed in this paper for thin film deposition process control. The framework of the supervisory GPC is shown in Fig. 2. Detailed discussions on each module are provided in Secs. 3.1–3.3

3.1 Process Modeling Using ARMAX Model. In the start-up period of each production run, the temperature of the chamber is manually increased step by step, with each step having at least eight sampling intervals. Those inputs, together with the corre-

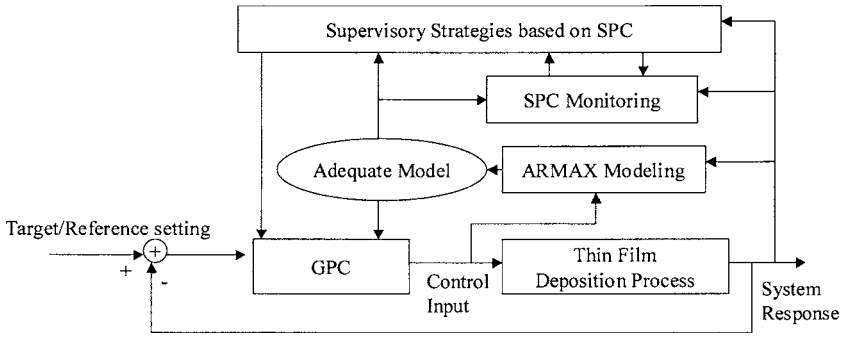


Fig. 2 The structure of the supervisory GPC

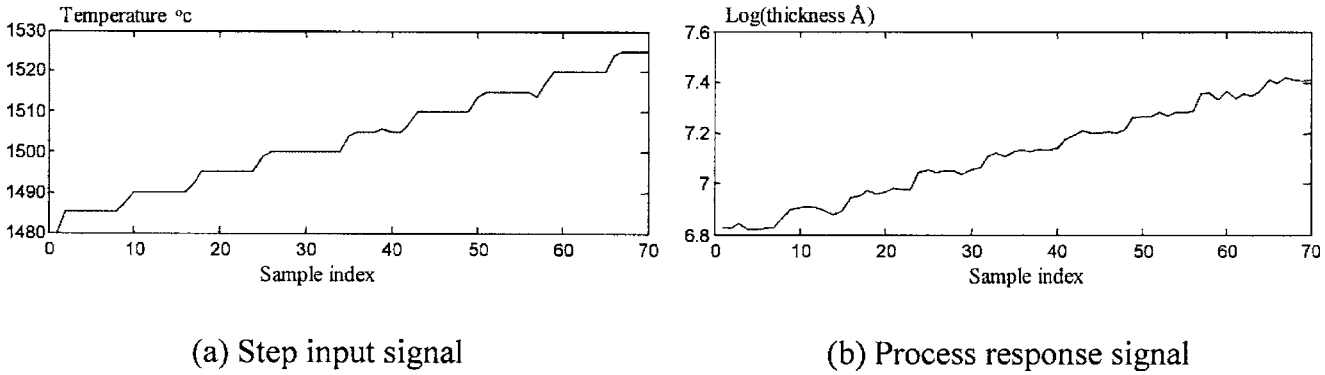


Fig. 3 Initial production run data of step input and response

sponding film thickness responses, are used to model each run of the thin film deposition process. The run-to-run variation is reflected in the model parameter change. An ARMAX (n_a, n_b, n_c, n_d, d) model with disturbance is used to describe the thin film deposition process,

$$A(q)y_t = B(q)u_{t-d} + C(q)f_t + D(q)e_t \tag{1}$$

where y_t and u_{t-d} are the system response of the film thickness at time t and control input of the corresponding source temperature at time t-d, respectively, and d is the dead time of the process, $A(q)$, $B(q)$, $C(q)$, and $D(q)$ are polynomial functions with orders n_a, n_b, n_c, and n_d, respectively, and q^{-1} is the backshift operator with $q^{-1}y_t = y_{t-1}$. Here, $A(q) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}$, $B(q) = b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}$, $C(q) = c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}$, and $D(q) = 1 + d_1 q^{-1} + \ldots + d_{n_d} q^{-n_d}$. f_t represents the disturbances that are possibly existing in the process as either a spike, a mean shift, or a linear drift function; e_t is an independent identically distributed (IID) zero-mean sequence of modeling residuals with $e_t \sim N(0, \sigma_e^2)$. The model parameters in Eq. (1) can be estimated using the least-squares method if $D(q)$ is known. In the case of unknown $D(q)$, the two-stage methods described in Sec. 10.4 of Ljung [1] should be used. In our case study discussed in Sec. 4, the ARX model structure with $D(q) = 1$ is estimated by using the least-squares method. Equation (1) can also be rewritten as

$$y_t = \frac{B(q)}{A(q)}u_{t-d} + \frac{C(q)}{A(q)}f_t + \frac{D(q)}{A(q)}e_t \tag{2}$$

Denote $B(q)/A(q) = \Sigma_{j=0}^{m_u} g_j^u q^{-j}$, $C(q)/A(q) = \Sigma_{j=0}^{m_f} g_j^f q^{-j}$, and $D(q)/A(q) = \Sigma_{j=0}^{m_e} g_j^e q^{-j}$; and m_u, m_f, and m_e are the maximum order of each corresponding item. They are determined based on the impulse response function g_j^s from each input term to the output y. That is, when $j > m_s$, $|g_j^s| < \eta_0$ $(s = u, f, e)$. An example of determining m_u is given in the case study in Sec. 4. Equation (2) can be rewritten as

$$y_t = \sum_{j=0}^{\min(t-d-1, m_u)} g_j^u u_{t-j-d} + \sum_{j=0}^{\min(t-1, m_f)} g_j^f f_{t-j} + \sum_{j=0}^{\min(t-1, m_e)} g_j^e e_{t-j} \tag{3}$$

3.2 Generalized Predictive Control. A generalized predictive control (GPC) was first proposed by Clark et al. [2], and the properties of GPC are further studied by Clark and Mohtadi [3] for a set of continuous chemical process control problems. Recently, GPC has also been used in a semiconductor control applications [4]. The first reason to choose GPC for the thin film deposition process control in this paper is the uncertainties in dead-time estimation. As described in the process overview, the single thickness output y_t is a compounded effect of several source cells located at different places. Since there is no sensor to measure each cell effusion, the dead time cannot be determined for each cell separately. As a result, the dead time used in the controller design is the combined effect of several sources. Moreover, the web speed variation also leads to dead-time variation. Without accurately knowing the dead time, it is difficult to design a controller using the proportional, integral, and derivative (PID) or minimum variance control strategy. Thus, it is desirable to design a generalized predictive controller to consider the possible range of multiple dead-time values. The second reason of using GPC is that the thin film deposition process is generally a nonminimum phase system described by an ARMAX model. Thus, significant challenges exist in the controller design for PID or minimum variance control strategies. The third reason of using GPC is its capability of integrating supervisory strategies for compensating for unknown and time-varying disturbance patterns: During the production, different disturbance patterns (such as a mean shift, a slow drift, or a spike) may occur in the thin film deposition process. An adaptive control strategy should be designed to reflect the difference in disturbance models. In recent years, research achievements on using an exponential weighed moving average (EWMA) method to estimate the time-varying mean shift or slow linear drift have demonstrated great success for a run-to-run controller design [5,6]. However, most of these methodologies assume that the disturbances exist during all production time with predefined disturbance model structures. Thus, EWMA models are used to continuously estimate the unknown or time-varying disturbance model parameters. Sachs et al. [7] investigated the run-to-run controller design by characterizing the run-to-run variations as two different modes: a slow mode and a fast mode. In contrast to other models, Sachs et al. assumed that a slow mode disturbance exists over all production time, which is estimated by EWMA and continuously compensated in the controller. The compensation of fast mode disturbance is only conducted when a large

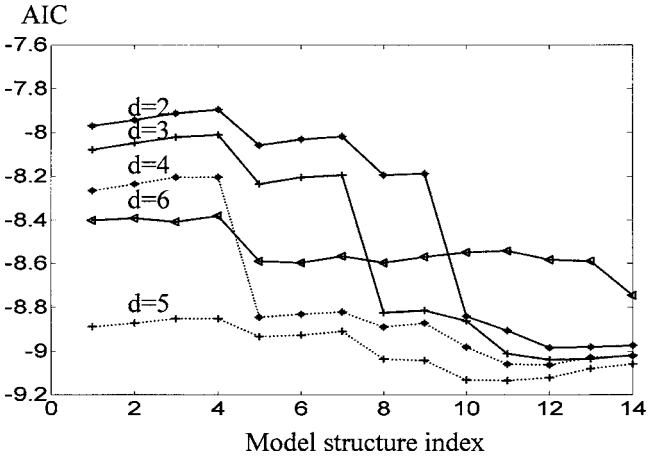
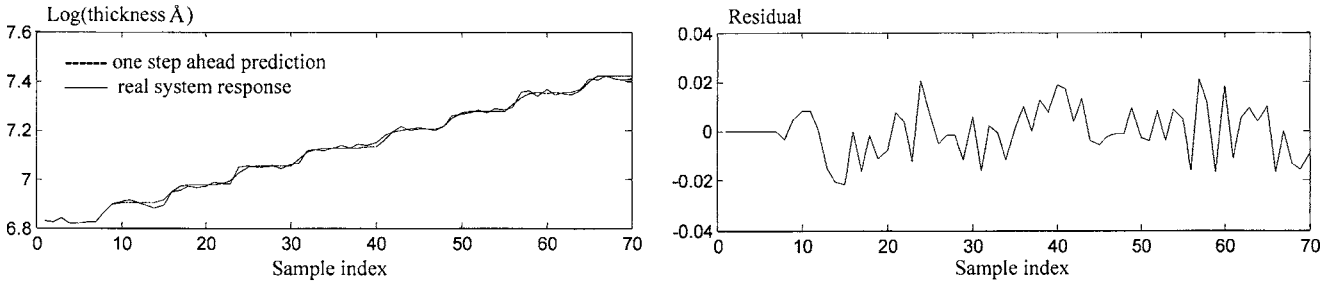


Fig. 4 AIC under different model structures



(a) Real output and one-step ahead prediction (b) One-step ahead prediction error

Fig. 5 Comparison of the modeling results with the real process response

shift is detected using an X-bar monitoring control chart. Therefore, the SPC monitoring technique provides a supervisory strategy to online determine the disturbance model structure. However, all of these run-to-run control methodologies are developed for a minimal variance controller, which is limited by assuming that the process has a fixed dead time. Therefore, they cannot be applied in the thin film deposition process in which a large and uncertain dead time exists.

In this paper, our objective is to integrate SPC with APC (automatic process control) to develop a supervisory predictive controller. The need of integrating a supervisory strategy with a predictive controller is more critical than in a minimal variance controller because more steps of disturbance prediction will be used in the predictive controller. A larger prediction error is usually generated when the prediction steps are increased because of modeling uncertainty. Therefore, accurate identification of the disturbance model structure is extremely important in the design of a predictive controller, especially for a process with large dead time.

Considering all these aspects, a GPC strategy is adopted for the thin film deposition process control. It is robust to the dead-time variability and estimation uncertainty, is effective in handling nonminimum phase systems, and has the ability to combine SPC with supervisory strategies.

The objective function of the GPC control is defined as

$$J=\sum_{i=d}^{N+d} q(i)(\hat{y}_{t+i|t}-y^*_{t+i})^2+\sum_{i=0}^{N} r(i)u^2_{t+i} \tag{4}$$

where y^*_{t+i} is the reference (or desired target) output at time $t+i$, $q(i)$ and $r(i)$ are the weighting coefficients, N is the sliding horizon for the output prediction and input series, and $\hat{y}_{t+i|t}$ is the ith step-ahead prediction made at time t. This is obtained from Eq. (3) by taking the conditional expectation

$$\hat{y}_{t+i|t}=E(y_{t+i|t})=\sum_{j=0}^{i-d} g^u_j u_{t+i-d-j}+\sum_{j=i-d+1}^{\min(t+i-d-1,m_u)} g^u_j u_{t+i-d-j}+\sum_{j=0}^{\min(t+i-1,m_f)} g^f_j f_{t+i-j}+\sum_{j=i}^{\min(t+i-1,m_e)} g^e_j e_{t+i-j} \tag{5}$$

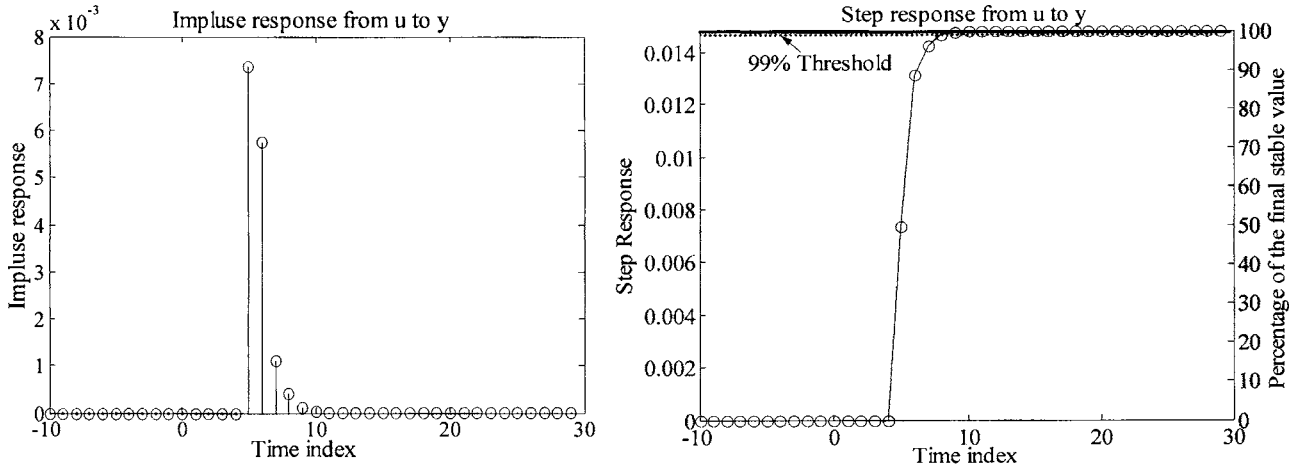
f_{t+i-j} $(j=0,1,\ldots,i-1)$ is the prediction of the deterministic disturbance function made at time t according to the identified disturbance function f_t at time t. For the random error e_t, the conditional expectation under the current time t is $E(e_{t+i|t})=0$ and $E(e_{t-i|t})=e_{t-i}$ $(i>0)$. Thus, we have $E\{\Sigma_{j=0}^{t+i-1} g^e_j e_{t+i-j}\}=\Sigma_{j=i}^{t+i-1} g^e_j e_{t+i-j}$.

The basic concepts of the GPC strategy can be summarized as follows: At current time t, an optimal control law u_t is obtained by solving the optimization problem defined in Eq. (4). By using the sliding horizon window of each of the $N+1$ step predictions, the control problem is simplified from a dynamic programing problem to a static optimization problem [8].

In order to solve the predictive control problem, the objective function in Eq. (4) is equivalently represented by a vector and matrix forms as

$$J=(\hat{Y}-Y^*)^T Q(\hat{Y}-Y^*)+U_1^T R U_1 \tag{6}$$

where $Q=\mathrm{diag}[q(N+d),q(N+d-1),\ldots,q(d)]$ and $R=\mathrm{diag}[r(N),r(N-1),\ldots,r(0)]$ are diagonal matrices of weighting coefficients. The vector $\hat{Y}$ represents all $N+1$ step predictions in



(a) (b)

Fig. 6 Impulse and step response of the thin film deposition process model

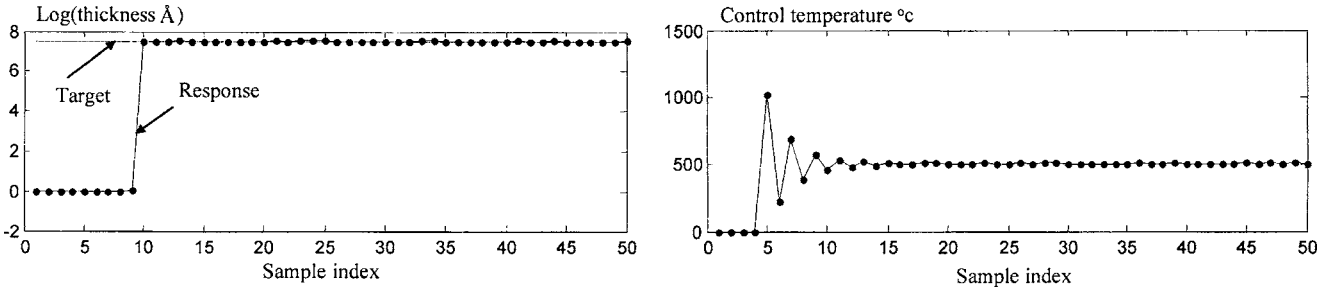


Fig. 7 A simulated control performance

the sliding window, which is represented based on Eq. (5) as

$$\hat{Y} = G_1 U_1 + G_0 U_0 + G_f F + G_e E \tag{7}$$

In this representation, vector $\hat{Y}=[\hat{y}_{t+d+N|t}, \hat{y}_{t+d+N-1|t}, \ldots, \hat{y}_{t+d+i|t}, \ldots, \hat{y}_{t+d|t}]^T$ represents all the predicted outputs based on time t within the sliding window. Corresponding to $\hat{Y}$, the optimal future control input vector U_1 is denoted as $U_1^{op}=[u_{t+N}, u_{t+N-1}, \ldots, u_{t+i}, \ldots, u_t]^T$ obtained in Eq. (9). It should be noted that only a single u_t in $\hat{U}_1^{op}$ will be used to control the process at time t. Similarly, from Eq. (5), the previously used control input vector is denoted as $U_0=[u_{t-1}, \ldots, u_i, \ldots, u_{t-\min(t-1,m_u)}]^T$, which is ordered backward from time $t-1$ to the maximum steps of m_u as defined in Eq. (3). The effect of these control inputs on the future system outputs is determined by the dynamic system memory on the historical control inputs. Vector $\hat{F}=[f_{t+d+N|t}, f_{t+d+N-1|t}, \ldots, f_{t+d+N-\min(t+d+N-1,m_f)}]^T$ is used to represent the deterministic disturbances from time $t+d+N$ backward to the maximum steps of m_f as defined in Eq. (3). Vector $\hat{E}=[e_t, e_{t-1}, \ldots, e_{t-\min(t-1,m_e-d)}]^T$ is used to denote the random residuals. These residuals can be calculated up to the current time t using Eqs. (32), (34), and (35), which correspond to a mean shift, a linear drift, and a spike disturbance, respectively. These equations will be discussed in Sec. 3.3.3. In the case where $m_e-d<0$, we set $E=0$, which means the historical random disturbances do not influence the future output. Corresponding to each of the vectors U_1, U_0, $\hat{F}$, and $\hat{E}$ used in Eq. (7), G_1, G_0, G_f, and G_e are used to reflect the different dynamic response weights, which are defined based on Eqs. (5) and (7)

$$G_1 = \begin{bmatrix} g_0^u & g_1^u & \cdots & g_N^u \\ 0 & g_0^u & \cdots & g_{N-1}^u \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & g_0^u \end{bmatrix}_{(N+1)\times(N+1)}$$

reflects how the future control input U_1 affects the output $\hat{Y}$;

$$G_0 = \begin{bmatrix} g_{N+1}^u & g_{N+2}^u & \cdots & g_{\min(t-1,m_u)+N}^u \\ g_N^u & g_{N+1}^u & \cdots & g_{\min(t-1,m_u)+N-1}^u \\ \vdots & \vdots & \vdots & \vdots \\ g_1^u & g_2^u & \cdots & g_{\min(t-1,m_u)}^u \end{bmatrix}_{(N+1)\times[\min(t-1,m_u)]}$$

reflects the effect of the dynamic system memory of the historical control input U_0 on the outputs $\hat{Y}$;

$$G_f = \begin{bmatrix} g_0^f & g_1^f & \cdots & g_{\min(d+t+N-1,m_f)}^f \\ 0 & g_0^f & \cdots & g_{\min(d+t+N-1,m_f)-1}^f \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & g_{\min(d+t+N-1,m_f)-N}^f \end{bmatrix}_{(N+1)\times[\min(d+t+N-1,m_f)+1]}$$

reflects the effect of the deterministic disturbance F on the output $\hat{Y}$;

$$G_e = \begin{bmatrix} g_{N+d}^e & g_{N+d+1}^e & \cdots & g_{\min(t+d-1,m_e)+N}^e \\ g_{N+d-1}^e & g_{N+d}^e & \cdots & g_{\min(t+d-1,m_e)+N-1}^e \\ \vdots & \vdots & \vdots & \vdots \\ g_d^e & g_{d+1}^e & \cdots & g_{\min(t+d-1,m_e)}^e \end{bmatrix}_{(N+1)\times\{\max[0,\min(t,m_e-d+1)]\}}$$

reflects the effect of the dynamic system memory of the historical random error E on the future output $\hat{Y}$, which is effective only when $m_e-d\geq 0$ holds.

It should also be noted that when the production time is large enough, the dimension of the matrices G_0, G_f, and G_e will be constrained only by the maximum order of each corresponding impulse function as defined in Eq. (3). By solving the optimization problem with $dJ/dU_1=0$, one has

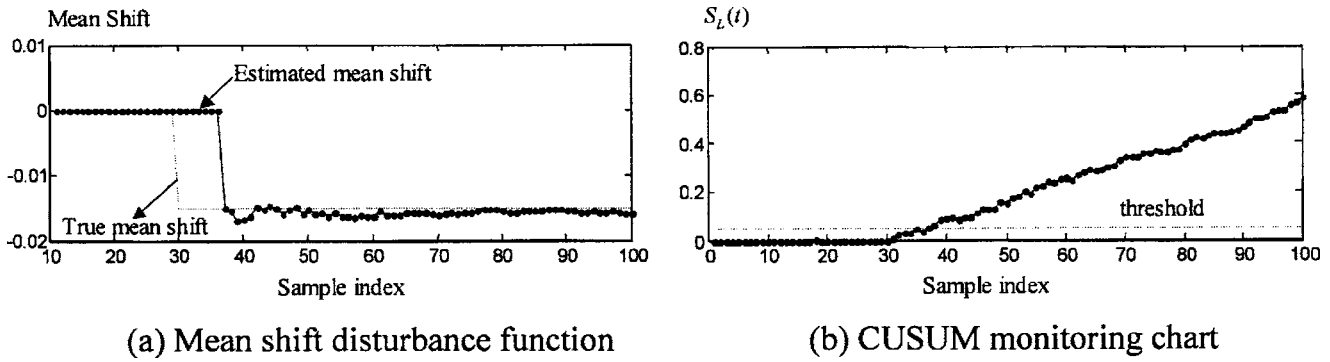
$$\frac{dJ}{dU_1} = 2[G_1^T Q(G_1 U_1 + G_0 U_0 + G_f F + G_e E - Y^*) + R U_1] = 0 \tag{8}$$

The optimal predictive control law U_1^{op}, is obtained by solving Eq. (8) as

$$U_1^{op} = -(G_1^T Q G_1 + R)^{-1} G_1^T Q(G_0 U_0 + G_f F + G_e E - Y^*) \tag{9}$$

Although the vector of all future inputs from u_t to u_{t+N} is provided in U_1^{op}, only u_t will be used to control the process at time $t+1$. At the next time step $t+1$, another optimization will be conducted based on the new observation of y_{t+1}, which is used to obtain a new optimal control vector $U_1^{op}=[u_{t+1}, u_{t+2}, \ldots, u_{t+N+1}]$. The newly obtained u_{t+1} will be used as a control input at time $t+2$. This strategy is repeated until the end of the production run.

Remark 1. Prediction Horizon N Selection. The principle in the



(a) Mean shift disturbance function (b) CUSUM monitoring chart

Fig. 8 Mean-shift disturbance function and CUSUM monitoring control chart

sliding horizon selection is to have its lower limit be equal to the dead time of the system. If the dead time is estimated with uncertainty or the dead time varies during the production, then the sliding horizon should be set as the lower bound of the dead-time estimation. The upper limit of the sliding horizon should be smaller than, or equal to, the settling time of the system responses. The settling time T_0 is defined as the time of which the absolute value of the impulse response of the process is always below a small threshold value of η_0 as time goes to infinity. T_0 can be obtained by investigating the impulse response of the model as

$$|g_j^u| < \eta_0 \quad \forall\, j > T_0 \tag{10}$$

The value of η_0 is determined by the modeling accuracy requirement. An example of how to get T_0 is given in the case study in Sec. 4. In addition, a smaller value for N should be selected if more significant disturbances or model uncertainty exist in the process.

Remark 2. Weighting Coefficients R Selection. The parameter R influences the process stability and tracking error. Correct selection of an R value is important to reach a satisfactory performance for controlling the nonminimum phase processes. In general, a larger R value leads to a more stable system, less overshoot, slower response, and larger tracking errors. Thus, if a disturbance exists, a smaller R value is preferred under the constraints of stability and acceptable overshoot. In this way, a fast tracking performance can be achieved, especially for a fast drift disturbance.

3.3 Supervisory Strategies. Various disturbances may exist in the thin film deposition process as described in Sec. 2. Three typical disturbance patterns (mean shift, linear drift or ramp, and spike) will be studied in the development of a supervisory strategy in the GPC design.

3.3.1 CUSUM Charts for Detection and Estimation of Shift and Drift Disturbances. A statistical cumulative-sum (CUSUM) monitoring chart is used to detect and estimate the disturbance f_t when it is a mean-shift or linear-drift function. The estimated function f_t is used in Eq. (9) to revise the future control adjustment to compensate for corresponding disturbances.

In the following section, a brief review of the CUSUM chart is given. A detailed discussion of the CUSUM chart technique can be found in Montgomery [9].

A CUSUM chart is constructed by positive and negative CUSUM statistics, calculated as

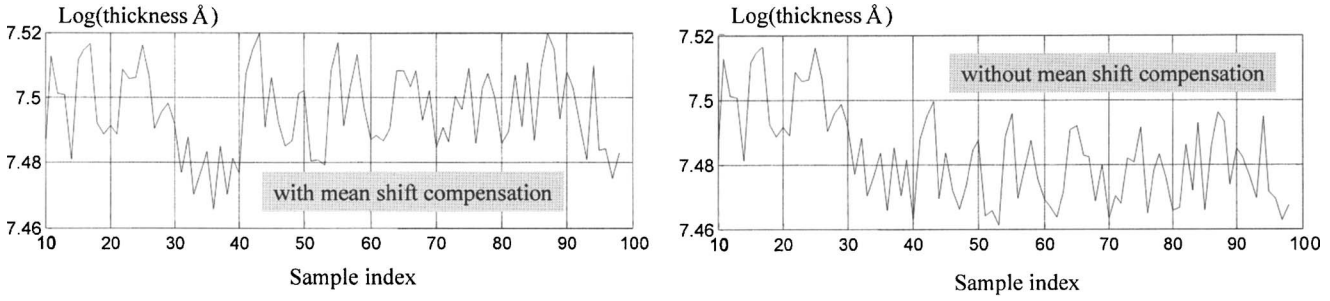
$$S_H(t) = \max[0, x_t - (\mu_0^x + K) + S_H(t-1)] \tag{11}$$

$$S_L(t) = \max[0, (\mu_0^x - K) - x_t + S_L(t-1)] \tag{12}$$

where x_t is the monitoring variable, which has a mean of μ_0^x under the in-control condition. The starting values of $S_H(t)$ and $S_L(t)$ are $S_H(0)=S_L(0)=0$. $K=|\Delta/2|$ and Δ is the mean shift of x_t to be detected.

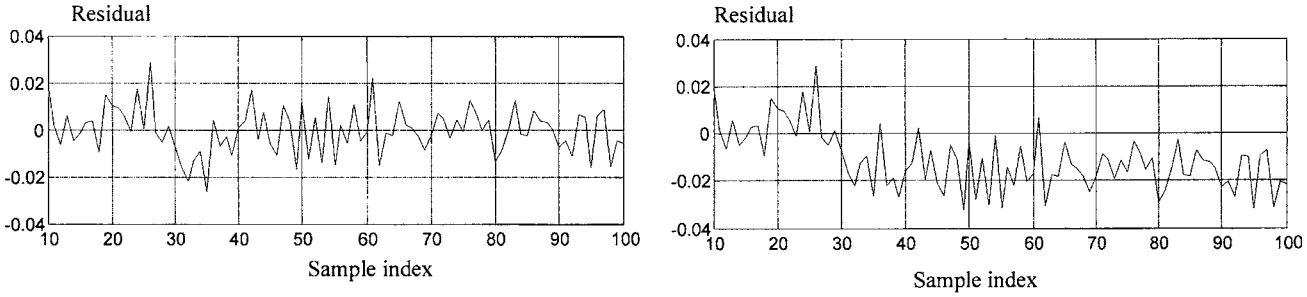
An out-of-control point is indicated at time k if $S_H(k)>h$ or $S_L(k)>h$, where h is a design parameter chosen as the control limit of a CUSUM chart. In general, the values of h and K determine CUSUM chart sensitivity to a disturbance and influence the average run length (ARL) [9].

If an out of control point is observed at time k, N^+ (or N^-) is the number of consecutive samples for which the upper-side CUSUM $S_H(k)$ (or lower-side CUSUM $S_L(k)$) has had a positive value. Thus, the time at which the mean shift is detected is $s+1$, and the occurrence time is $s=k-N^+$ for an upward shift or $s=k-N^-$ for a downward shift. The sum of mean shifts over N^+ (or N^-) steps is estimated by



(a) Response with mean shift compensation (b) Response without mean shift compensation

Fig. 9 Comparison of control performance under mean-shift disturbances



(a) Adding estimated mean shift to the model (b) Without adding mean shift to the model

Fig. 10 Comparison of model residual errors under mean shift

$$\text{if } S_H(k) > h, \quad \sum_{i=1}^{N^+} \mu_i^x = N^+(\mu_0^x + K) + S_H(k)$$
$$\text{if } S_L(k) > h \quad \sum_{i=1}^{N^-} \mu_i^x = -N^-(\mu_0^x + K) - S_L(k) \tag{13}$$

In the developed control strategy, there are two different CUSUM charts used to monitor a mean shift and a linear drift.

3.3.1.1 CUSUM chart for mean shift disturbance monitoring. In order to detect and estimate the mean shift, the monitoring variable x_t is defined as ξ_t as follows:

$$\xi_t = \frac{A(q)}{D(q)} y_t - \frac{B(q)}{D(q)} u_{t-d} = \frac{C(q)}{D(q)} f_t + e_t \tag{14}$$

This relationship follows directly from Eq. (1). Using this, ξ_t can be directly calculated for each newly observed output y_t and control input u_{t-d}. The following hypothesis test is used with a CUSUM chart to detect whether there exists a nonzero mean shift of f_t:

$$H_0: \mu^{\xi_t} = \mathrm{E}(\xi_t) = 0; \quad \text{when } f_t = 0$$
$$H_1: \mu^{\xi_t} = \mathrm{E}(\xi_t) \neq 0; \quad \text{when } f_t = \mu_f \tag{15}$$

So, taking the expectation of Eq. (14), we have:

$$\mu^{\xi_t} = \mathrm{E}[\xi_t] = \frac{C(q)}{D(q)} f_t = \sum_{i=0}^{n_f} I_i^f f_{t-i} \tag{16}$$

where n_f is the maximum order of the inverse function I_i^f. This function is obtained by expanding $C(q)/D(q)$ with the backward operator q^{-1}. By substituting the condition of Eq. (15) into Eq. (16), the test is

$$H_0, \quad \mu^{\xi_t} = 0$$
$$H_1, \quad \mu^{\xi_t} = \mu_f \left(\sum_{i=0}^{\min(N^s, n_f)} I_i^f \right) \tag{17}$$

where N^s is the number of steps of the mean shift occurring at time $t - N^s$. Thus, if either $S_H^{\xi}(k) > h$ or $S_L^{\zeta}(k) > h$ is detected in the CUSUM chart, the mean shift can be estimated based on Eqs. (13) and (17) as

$$\text{if } S_H^{\xi}(k) > h, \quad \mu_f = (N^+ \cdot K + S_H^{\xi}(k)) \left(\sum_{j=1}^{N^+} \sum_{i=0}^{\min(j,n_f)} I_i^f \right)^{-1}$$
$$\text{if } S_L^{\xi}(k) > h, \quad \mu_f = -(N^- \cdot K + S_L^{\xi}(k)) \left(\sum_{j=1}^{N^-} \sum_{i=0}^{\min(j,n_f)} I_i^f \right)^{-1} \tag{18}$$

3.3.1.2 CUSUM chart for linear-drift disturbance monitoring. When monitoring and estimating a linear-drift disturbance, x_t is defined as ς_t as

$$\varsigma_t = \xi_t - \xi_{t-1} \tag{19}$$

From Eqs. (14) and (16), it can be seen that

$$\xi_t = I_0^f f_t + I_1^f f_{t-1} + \cdots + I_{n_f}^f f_{t-n_f} + e_t \tag{20}$$

$$\xi_{t-1} = I_0^f f_{t-1} + I_1^f f_{t-2} + \cdots + I_{n_f}^f f_{t-1-n_f} + e_{t-1} \tag{21}$$

Subtracting Eq. (20) from Eq. (21), we have

$$\varsigma_t = I_0^f (f_t - f_{t-1}) + I_1^f (f_{t-1} - f_{t-2}) + \cdots + I_{n_f}^f (f_{t-n_f} - f_{t-1-n_f}) + e_t - e_{t-1} \tag{22}$$

Thus, the following hypothesis test is used with the CUSUM chart to detect whether there exists a nonzero slope of the linear drift function f_t:

$$H_0: \mu^{\varsigma_t} = \mathrm{E}(\varsigma_t) = 0; \quad \text{when } f_t - f_{t-1} = 0 \ \forall\, t > 1$$
$$H_1: \mu^{\xi_t} = \mathrm{E}(\varsigma_t) \neq 0; \quad \text{when } f_t - f_{t-1} = \beta \tag{23}$$

By taking the expectation of Eq. (22) and substituting it into the condition of Eq. (23), we have

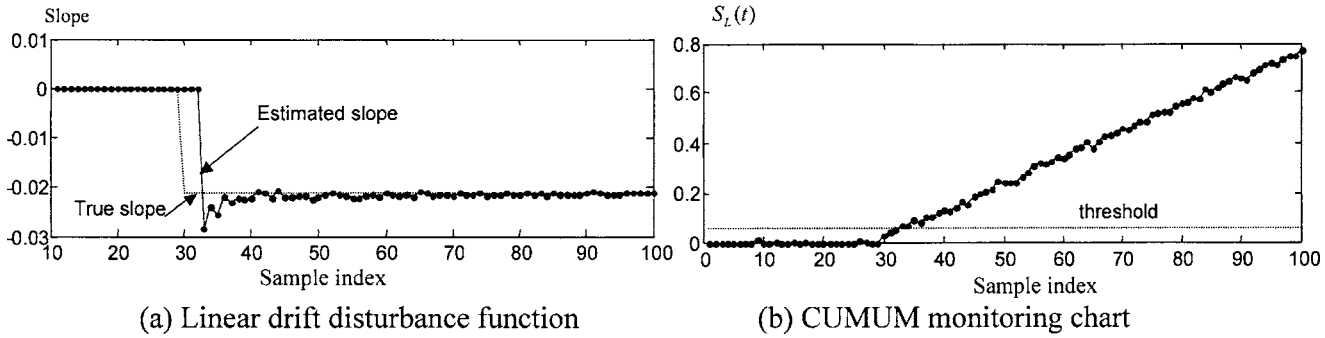
$$H_0, \quad \mu^{\varsigma_t} = 0$$
$$H_1, \quad \mu^{\varsigma_t} = \beta \left(\sum_{i=0}^{\min(N^{\beta}, n_f)} I_i^f \right) \tag{24}$$

where N^{β} is the number of steps of the linear drift occurring at time $t - N^{\beta}$. Thus, if either $S_H^{\varsigma}(k) > h$ or $S_L^{\varsigma}(k) > h$ is detected in the CUSUM chart, the slope of linear drift as

$$\text{if } S_H^{\varsigma}(k) > h, \quad \beta = \{N^+ \cdot K + S_H^{\varsigma}(k)\} \left(\sum_{j=1}^{N^+} \sum_{i=0}^{\min(j,n_f)} I_i^f \right)^{-1}$$
$$\text{if } S_L^{\varsigma}(k) > h, \quad \beta = -\{N^- \cdot K + S_L^{\varsigma}(k)\} \left(\sum_{j=1}^{N^-} \sum_{i=0}^{\min(j,n_f)} I_i^f \right)^{-1} \tag{25}$$

3.3.2 X-bar Chart for the Detection and Estimation of a Spike Disturbance. In a thin film deposition process, a spike signal may be observed from the thickness measurements. In practice, a spike signal is often due to sensor errors. Thus, it is desirable to detect and remove the sensing errors from the system response, so that the controller will not provide wrong feedback in the process control.

For the thin film deposition process model, the relationship between a single spike error at time s and the true system output y_t^0 without spike error can be modeled as



(a) Linear drift disturbance function (b) CUMUM monitoring chart

Fig. 11 CUSUM monitoring control chart for detecting a drift

$$y_t^0 = y_t - \rho\delta(t-s) \tag{26}$$

where $\delta(t-s)$ is the Kronecker δ function, y_t is the system output measurement including the spike error at time s, and ρ is the magnitude of the spike. So, when there is no process change f_t at $t=s$, we see from Eq. (14) that

$$\xi_{t=s} = \frac{A(q)}{D(q)}[y_t^0 + \rho\delta(t-s)] - \frac{B(q)}{D(q)}u_{t-d} = e_s + \rho \tag{27}$$

By taking the expectation on Eq. (27), we have

$$E(\xi_t) = \begin{cases} \rho & \text{when } t=s \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

From Eq. (28), An X-bar control chart can be designed to detect a large spike error greater than $3\sigma_e^2$. Thus, the control limits of X-bar chart are defined as

$$\begin{cases} \text{UCL} = 3\sigma_e^2 \\ \text{CL} = 0 \\ \text{LCL} = -3\sigma_e^2 \end{cases} \tag{29}$$

where UCL (LCL) represents the upper (lower) control limits, CL is the central line of the control chart.

Since a spike usually occurs only at one sample interval, a decision rule is needed to identify spike errors, which is "*A spike occurs at time s if $\xi(s)$ is identified as out of control, but all $\xi(s+i)$ $(i=-b,-b+1,\ldots,-1,1,2,\ldots,b)$ are identified as in control.*" Here b is the detection window determining the monitoring range. If an out of control condition is detected, both the time and magnitude of the spike will be provided to the GPC for further compensation actions. The expected magnitude of a spike is $E[\xi(t)]=\rho$ when $t=s$, and $E[\xi(t)]=0$ when $t\neq s$. So, in a given detection window with length $2b$, the estimated magnitude of the spike ρ is

$$\hat{\rho} = \xi(s) - \sum_{t=s-b,t\neq s}^{s+b} \frac{\xi(t)}{2b} \tag{30}$$

Thus, we see that this is an unbiased estimator: $E(\hat{\rho})=E[\xi(t)]-E[\Sigma_{t=s-b,t\neq s}^{s+b}\xi(t)/2b]=\rho-0=\rho$.

3.3.3 Control Law Revision for Supervisory Compensation of Detected Disturbances

3.3.3.1 Control law revision under a mean shift or linear drift disturbance. A time delay always exists in detecting a disturbance using a CUSUM chart. Assume that a mean shift or drift occurs at time index s, which is detected at time index $k(k>s)$. Thus, a detection delay of $k-s$ is experienced. In the next control law calculation for U_t, $t\geq k$, there is a need not only to revise the future prediction of the disturbance model $(f_{k+1|k},\ldots,f_{k+d+N|k})$ based on the detected disturbance function at k, but also to revise the disturbance model and residual errors during $s\leq t\leq k$ for $(f_s,f_{s+1},\ldots,f_k)$ and $(e_s,\ldots,e_k)$ in order to reflect the effect of the disturbances occurring at time s.

When a mean shift actually occurs at time s but is detected at time k, the revised deterministic disturbance model used in the control law of Eq. (9) is

$$f_t = \begin{cases} 0 & t<s \\ \mu_f & t\geq s \end{cases} \tag{31}$$

μ_f is calculated by using Eq. (18). Substituting f_t from Eq. (31) into Eqs. (14) and (16), the revised residual is obtained as

$$e_t = \begin{cases} \xi_t & t<s \\ \xi_t - \sum_{i=1}^{\min(t-s,n_f)} I_i^f \mu_f & s\leq t\leq k \end{cases} \tag{32}$$

By substituting this revised e_t from Eq. (32) and f_t from Eq. (31) into Eq. (9), the revised control law $U_1^{op}=[u_{t+N},u_{t+N-1},\ldots,u_{t+i},\ldots,u_t]^T$ can be obtained, in which the single u_t is used to control the process at time t to compensate the mean shift.

Similarly, when a linear drift actually occurs at time s but is detected at time k, the revised disturbance model used in the control law of Eq. (9) is

$$f_t = \begin{cases} 0 & t<s \\ (t-s+1)\beta & t\geq s \end{cases} \tag{33}$$

and the revised residual is obtained by substituting the f_t of Eq. (33) into Eqs. (14) and (16)

$$e_t = \begin{cases} \xi_t & t\leq s \\ \xi_t - \sum_{i=1}^{\min(t-s,n_f)} [I_i^f \beta(t-s+1)] & s<t\leq k \end{cases} \tag{34}$$

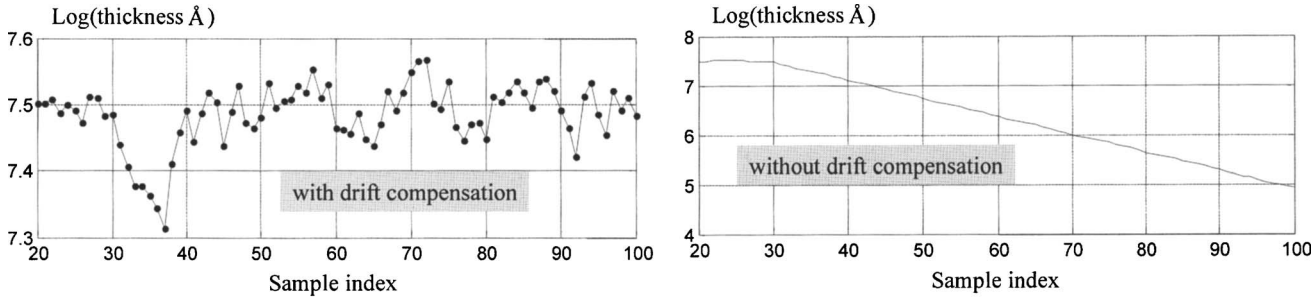
By substituting this revised e_t of Eq. (34) and the f_t of Eq. (33) into Eq. (9), the revised control law $U_1^{op}=[u_{t+N},u_{t+N-1},\ldots,u_{t+i},\ldots,u_t]^T$ can be obtained, in which the single u_t is used to control the process at time t to compensate the detected linear drift.

3.3.3.2 Control law revision under a spike disturbance. If a spike due to a sensor error is detected, then it is desirable to remove the spike data from the calculation of the feedback control. This can be achieved by recalculating the residual series e_t used in the control law of Eq. (9) when a spike is detected.

Assume that a spike occurring at time s has been detected at time $k=s+b$. In this case, the revised e_t is obtained based on Eq. (27) as

$$e_t = \begin{cases} \xi_t & t\neq s \\ \xi_t - \rho & t=s \end{cases} \tag{35}$$

By substituting the revised e_t of Eq. (35) into Eq. (9), the revised control law $U_1^{op}=[u_{t+N},u_{t+N-1},\ldots,u_{t+i},\ldots,u_t]^T$ can be obtained.



(a) Process response with drift compensation

(b) Process response without drift compensation

Fig. 12 Control performance comparison under drift disturbances

4 A Case Study

4.1 Thin Film Deposition Process Modeling. Real production data from a thin film deposition process were collected for the case study. In the data set, the step input signals of the source temperature as shown in Fig. 3(*a*) were applied during the process start-up period, and the process response is shown in Fig. 3(*b*). Seventy data points were collected for the process modeling. The sampling interval used in the data acquisition is 2 min.

Based on process engineering knowledge, the thin film process under the normal operating condition is modeled by an ARX model with $C(q)=D(q)=1$. The order of $A(q)$, $B(q)$, and delay step d, denoted as $[n_a,n_b,d]$, will be quantitatively determined using an information-theoretic approach based on AIC (Akaike's information criterion) [10,11], i.e., the smallest AIC value indicates the best-fit model among all compared model candidates. In this case study, the possible model structures include the following 14 combinations: [11 *d*], [21 *d*], [31 *d*], [41 *d*], [22 *d*], [32 *d*], [42 *d*], [33 *d*], [43 *d*], [44 *d*], [55 *d*], [66 *d*], [77 *d*], and [88 *d*]. The possible dead times are $d=2,3,4,5,6$. Figure 4 shows all AIC values of these 14 models under each dead time. It is clear that the model of [5 5 5] (model structure index=11) has the smallest AIC value, which is considered the best-fit model among these 14 models. After selecting this best-fit model, the statistical F test is used to further determine whether a lower order model ($n_a<5, n_b<5$) can be used. This statistical testing is based on whether the sum of squares of the modeling residual errors is significantly increased when a lower order is used [12]. An F test with the significance level 5% is used in the final order determination. Based on the data set, the final ARX model structure is [2 2 5] with the parameters as $A(q)=1-0.1297q^{-1}-0.04919q^{-2}$, $B(q)=0.007376+0.004798q^{-1}$. The error term series is determined with $e_t\sim N(0,10^{-4})$.

In order to verify the model accuracy, Fig. 5(*a*) shows the comparison between the real process response and its one-step-ahead prediction, and Fig. 5(*b*) shows corresponding residual errors. It is clear that the identified model has good track performance when compared to the real process output.

4.2 Predictive Control. The parameters used in the GPC are selected $Q=\text{diag}[1,1,1,1]$, $R=r\times\text{diag}[1,1,1,1]$, and $r=10^{-7}$. Generally, the weight coefficients Q of prediction error are set to 1 and the weight coefficients R of the control cost are adjusted based on the applications [8]. In order for the system to be able to quickly compensate for the process change, R is usually selected as a relatively small value under the controller stability constraint. Here, we select a reasonable value of $r=10^{-7}$ based on trial and error. In order to calculate the optimal control input U_1^{op}, the dimensions of matrixes G_0, G_f, and G_e need to be determined. These dimensions are determined by m_u, m_f, and m_e, respectively. An example of how to choose m_u based on the system step response is illustrated in Fig. 6. In Fig. 6(*b*), the settling time is determined for the step response to stay within ±1% (99% in this example) of its final stable step response [13]. Thus, the threshold η_0 in Eq. (10) is equal to 1% of its final stable step response. m_u is determined by the number of impulse responses in Fig. 6(*a*) with their absolute values larger than η_0. The values of m_f and m_e are similarly determined. The width of sliding window is selected as $N=3$. It is determined based on engineering knowledge of the delay uncertainty in the thin film processes and m_u.

In Eq. (9), the value of $-[G_1^TQG_1+R]^{-1}G_1^TQ$, which is denoted as Φ, is

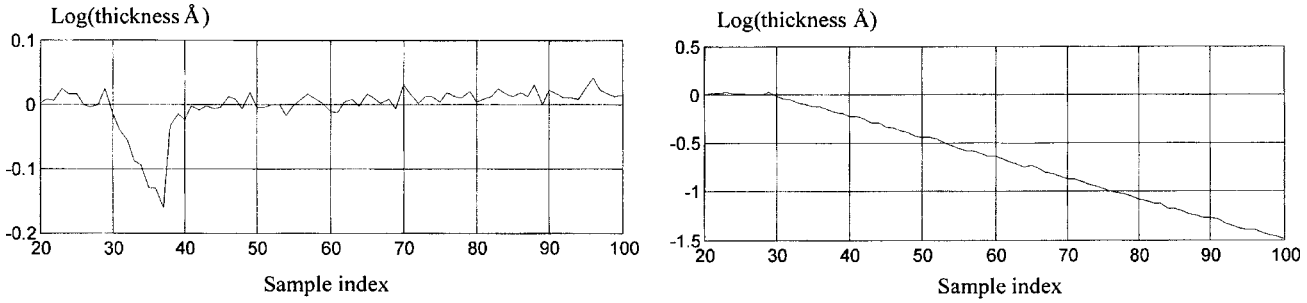
$$\Phi=\begin{bmatrix} -135.1 & 105.1 & -61.5 & 39.9 \\ -0.315 & -134.9 & 105.0 & -61.5 \\ 0.17 & -0.414 & -134.9 & 105.1 \\ -0.073 & 0.17 & -0.315 & -135.1 \end{bmatrix}$$

A simulated tracking control performance with the target of 7.5 is shown in Fig. 7. From this figure, it can be seen that there is a delay of 10 samples from the starting of the production run (at sample index 1) to the process output response. This 10-step delay comes from a 5-step dead time after the first control input was added at sample index 5. At $t\geqslant 10$, the output quickly achieves and maintains the target value with a small variation. The noise standard deviation used in this simulation is 0.01, which is equal to the estimated $\hat{\sigma}_e$ of the modeling residual e_t.

4.3 Supervisory Strategies

4.3.1 Mean-Shift Detection and Compensation. A CUSUM chart is designed to monitor a mean shift of the process. The process mean shift is simulated as $\Delta=-1.5\sigma_e=-0.015$ with $\sigma_e=0.01$, which is added to the process at $t\geqslant 30$. The design parameters of the CUSUM chart are $\mu_0^{\xi}=0$, $K=|\Delta/2|=0.0075$, and $h=5,\sigma_e=0.05$. Figure 8(*a*) shows the mean-shift disturbance function of f_t, and Fig. 8(*b*) shows the CUSUM monitoring control chart of ξ_t in the supervisory GPC. It can be seen that after the shift occurs at sample 30, the first out-of-control point is detected at sample $k=37$ with $S_L(37)=0.0522>h=0.05$. Based on Eq. (18) with $n_f=1$ and $I_o^f=1$ in the model, the shifted mean at time $t=37$ is estimated as $\mu^{\xi}(37)=-K-S_L(37)/N^-=-0.015$ with $N^-=7$. Figure 8(*a*) shows the comparison of the estimated means shift and true mean shift over different time, in which the dotted line of the estimated mean shift is very close to the dashed line of the true mean shift except for the delay period of the detection.

After detecting the mean shift, the supervisory GPC will compensate for it by using the estimated mean shift. As shown in Fig. 9(*a*), the controller can quickly compensate for the mean-shift to bring the output back to the target value. However, without compensating for the mean shift in the controller, the mean-shift disturbance will lead to a mean shift in the process response as shown in Fig. 9(*b*). It is also noted that the compensation control is not fully applied to the system until time 42. This is due to the



(a) Adding estimated drift to the model

(b) Without adding estimated drift to the model

Fig. 13 Comparison of model errors under linear drift

5-step dead time after the mean shift is detected at $t=37$. Thus, reducing the dead time in the process and having an early detection of the mean shift will effectively reduce the impact of the mean-shift disturbance.

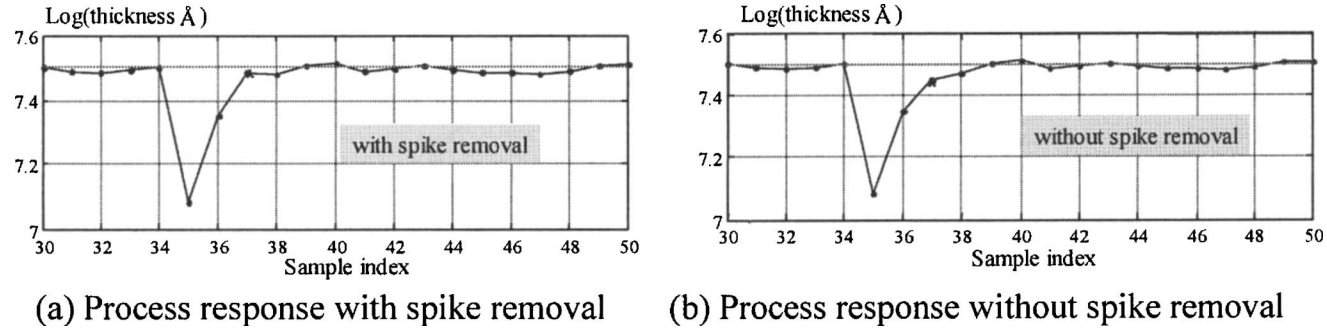
The comparison of model errors (differences between the one-step-ahead predictions and the real measurements of the process) under the mean shift was also plotted in Fig. 10.

4.3.2 Drift Disturbance Detection and Compensation. To detect a linear drift disturbance using a CUSUM chart, we need to monitor the random variable $\varsigma_t=\xi_t-\xi_{t-1}$, which has a standard deviation of $\sigma_\varsigma=\sqrt{2}\sigma_e=0.0141$ under the in-control condition. The simulated drift has a slope of $\Delta=1.5\times\sigma_\varsigma=0.0212$, which is added to the process at $t\geqslant 30$. The parameters of the CUSUM chart are defined as $\mu_0^\varsigma=0$, $K=|\Delta/2|=0.0106$, and $h=4\times\sigma_\varsigma=0.0566$. Figure 11(*a*) shows the disturbance signal of f_t, and Fig. 11(*b*) shows the CUSUM monitoring chart. It can be seen that the first out-of-control point is indicated at time $k=33$ with $S_L(33)=0.0716>0.0566$ and $N^-=4$. Similarly, based on Eq. (25) with $n_f=1$ and $I_o^f=1$ in the model, the drift slope at time $t=33$ is estimated as $\beta(33)=-0.0285$. Figure 11(*a*) shows the comparison of the estimated slope and the true slope of the linear drift disturbance over different time, in which the dotted line of the estimated slope is very close to the dashed line of the true slope except for the detection delay period.

The comparison of the control performance with and without compensation of the linear drift is shown in Fig. 12. As it can be seen from Fig. 12(*a*), after a drift disturbance was introduced at $t\geqslant 30$, the process output will experience a short period of large deviations from the setting point. However, after detecting the drift at $t=33$, the supervisory GPC can quickly compensate for the disturbances at $t\geqslant 38$, but the compensation control is not fully applied to the system until five time intervals after the drift is detected. Thus, an effective way to reduce the impact of disturbance is to reduce the dead time and detect the drift earlier. Without the supervisory strategies (i.e., no detection of and compensation for the drift), the controller output has a linear drift as shown in Fig. 12(*b*). The comparison of model errors (differences between the one-step ahead predictions and the real measurements of the process) under a linear drift disturbance was also plotted in Fig. 13.

4.3.3 Spike Disturbance Detection and Control Law Revision. It is assumed that a spike, which leads to a 6% deviation in the output target value, is introduced to the system at time index 30. The control performance with the supervision in GPC is shown in Fig. 14(*a*). From this figure, it can be seen that a significant deviation from the setting point occurs in the output response at data index 35, which is due to the control reaction (or feedback) to the spike error added at $t=30$. The five-step delay is due to the dead time of the process. Under the supervision, an X-bar control chart is used to detect the spike disturbance by using the decision rule defined in Sec. 3.3.2 ($b=2$ is used in the rule). Thus, the control chart detects the spike error at $k=37$, which indicates a two-step delay after the spike is shown on the response at $t=35$. Once the spike is detected, a supervisory corrective action is taken to remove the spike effects using Eq. (35). It can be seen that the output tracking errors are reduced significantly at $t=37$ after removing the spike in the control law calculation. However, due to the two-step delay in the spike detection, there is no compensation effort for the first two time indices ($t=35,36$). Thus, a significant deviation from the target is observed at $t=35,36$. Once a spike disturbance is confirmed, the supervisory GPC will remove the influence of the spike from the process and pull the output back to the normal condition quickly. By comparing Figs. 14(*a*) and 14(*b*), it can be seen that without supervision the output response at index 37 has a larger deviation than that with supervision.

The effectiveness and importance of having this spike error removal feature depend on the dynamic characteristic of the process model. It will be more desirable to compensate for the spike error if the impulse response has a long memory of the process dynamics, which is equivalent to have the poles of the character-



(a) Process response with spike removal

(b) Process response without spike removal

Fig. 14 Control performance comparison under the spike disturbance

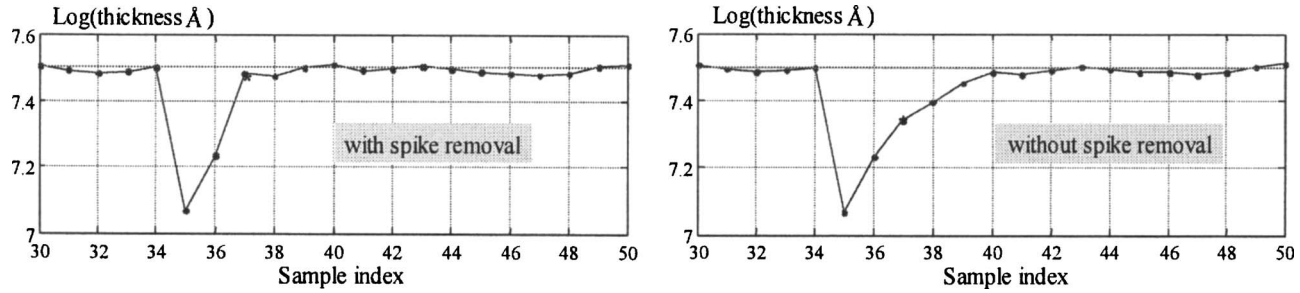


Fig. 15 Control performance under a spike disturbance for a longer memory system

istic equation close to 1. To illustrate this point, Fig. 15 gives a comparison which shows the effectiveness of the spike compensation technique based on another model. The model parameters for this model are $A(q)=1-0.39q^{-1}-0.126q^{-2}$, $B(q^{-1})=0.003645+0.004168q^{-1}$, $C(q)=D(q)=1$, the dead time of $d=5$, and the error term series is generated from $e_t \sim N(0, 10^{-4})$. The poles of the characteristic equation are 0.6 and −0.21. In this case, a similar spike, which also leads to about a 6% deviation in the output target value, is added as the sensor error to demonstrate the effectiveness of the spike error compensation. It can be seen from Fig. 15 that without the spike removal the output deviations from the target at indexes of 37–39 are larger than that with the spike removal.

5 Conclusion

The control of a thin film deposition process, which can be modeled by a single-input, single-output (SISO) ARMAX linear dynamic model, was investigated in this paper. Because of the inherent characteristics of the thin film deposition process, a GPC was adopted for the process controller design. A set of supervisory strategies is used to obtain a satisfactory control performance by compensating for the different types of disturbances. This paper emphasizes the importance of developing supervisory strategies through monitoring the process changes using SPC techniques, and then revising the controller parameters accordingly to achieve superior control performance. The integration of SPC with automatic process control (APC) provides great potential for the development of effective controllers in complex manufacturing processes. The case studies provided validate the effectiveness of the developed supervisory GPC strategy.

There are several open issues to be investigated further. One is how to determine the thresholds for the SPC control charts. Because the SPC is used to monitor a GPC-controlled process, Type I and Type II errors discussed in the conventional SPC literature cannot be used directly. Although some research has been done to investigate the SPC monitoring for a PID-controlled process [14], the SPC monitoring for the supervisory GPC control, which is more complex than the PID control strategy, deserves some further attention. Also, a cautious control strategy could be integrated to accommodate the estimation uncertainty of the process model and disturbance [15]. Another open issue worth investigating is how to systematically and simultaneously select (optimal) GPC parameters (e.g., N, R, etc.), SPC thresholds, and alternative supervisory strategies. Currently, some trial-and-error efforts are required to select those parameters [8]. Another topic related to thin film thickness control is to expand the research presented in this paper from SISO to multiple-input multiple output (MIMO) cases because there are multiple material sources as well as multiple quality indices for a thin film deposition process. Some early works in predictive control based on a state space model can be considered as a direction for further extension.

Acknowledgment

The authors would like to thank the Associate Editor and the reviewers for their insightful comments and suggestions, which have significantly improved the paper quality and readability. The authors also gratefully acknowledge the financial support of the NSF Career Award DMI-0133942 and NSF DMI-0330356, and AFOSR Grant F49620-03-1-0377.

References

[1] Ljung, L., 1999, *System Identification—Theory for the User*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.
[2] Clark, D. W., Mohtadi, C., and Tuffs, P. S., 1987 "Generalized Predictive Control—Part I. The Basic Algorithm; Part II. Extensions and Interpretations," Automatica, **23**(2), pp. 137–160.
[3] Clark, D. W., and Mohtadi, C., 1989, "Properties of Generalized Predictive Control," Automatica, **25**(6), pp. 859–875.
[4] Lee, L. L., Schaper, C. D., and Ho, W. K., 2002 "Real-Time Predictive Control of Photoresist Film Thickness Uniformity," IEEE Trans. Semicond. Manuf., **15**(1), pp. 51–59.
[5] Del Castillo, E., and Hurwitz, A. M., 1997, "Run-to-Run Process Control: Literature Review and Extensions," J. Quality Technol., **29**, pp. 184–196.
[6] Moyne, J., Del Castillo, E., and Hurwitz, A., 2000, *Run to Run Process Control in Semiconductor Manufacturing*, CRC Press, Boca Raton, FL.
[7] Sachs, E., Hu, A., and Ingolfsson, A., 1995, "Run by Run Process Control Combining SPC and Feedback Control," IEEE Trans. Semicond. Manuf., **8**(1), pp. 26–44.
[8] Astrom, K. J., and Wittenmark, B., 1995, *Adaptive Control*, 2nd ed., Addison-Wesley, Reading, MA.
[9] Mongtmoery, D., 2001, *Introduction to Statistical Quality Control*, 4th ed., Wiley, New York.
[10] Ljung, L., and Soderstrom, A., 1987, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA.
[11] Apley, D., and Shi, J., 1999 "An Order Downdating Algorithm for Tracking System Order and Parameters in Recursive Least Squares Identification," IEEE Trans. Signal Process., **47**(11), pp. 3134–3137.
[12] Box, G., Jenkins, G., and Reinsel, G., 1994, *Time Series Analysis, Forecasting, and Control*, 3rd Ed., Prentice-Hall, Englewood Cliffs, NJ.
[13] Kuo, B. C., and Golnaraghi, F., 2002, *Automatic Control Systems*, 8th ed. Wiley, New York, p. 237.
[14] Tsung, F., Wu, H., and Nair, V. N., 1998 "On the Efficiency and Robustness of Discrete Proportional-Integral Control Schemes," Technometrics, **40**, pp. 214–222.
[15] Shi, J., and Apley, D. W., 1998, "A Suboptimal N-Step-Ahead Cautious Controller for Adaptive Control Applications," ASME J. Dyn. Syst., Meas., Control, **120**, pp. 419–423.