Identification and Fine Tuning of Closed-loop Processes under Discrete EWMA and PI Adjustments

Rong Pan

Enrique Del Castillo

Industrial and Manufacturing Engineering Department

The Pennsylvania State University

University Park, PA 16802

April 2001

Abstract

Conventional process identification techniques of a open-loop process use the cross-correlation function between historical values of the process input and of the process output. If the process is operated under a linear feedback controller, however, the cross-correlation function has no information on the process transfer function because of the linear dependency of the process input on the output. In this paper, several circumstances where a closed-loop system can be identified by the autocorrelation function of the output are discussed. It is assumed that a Proportional Integral (PI) controller with known parameters is acting on the process while the output data were collected. The disturbance is assumed to be a member of a simple yet useful family of stochastic models, which is able to represent drift. It is shown that, with these general assumptions, it is possible to identify some dynamic process models commonly encountered in manufacturing. After identification, our approach suggests to tune the controller to a near-optimal setting according to a well-known performance criterion.

Keywords: ARMA processes, system identification, process adjustment

1 Introduction

Many manufacturing processes are carried out under repeated feedback adjustments. Frequently, the adjustment rule, or controller, is not optimal in any sense, but is sufficient to operate the process in a safe and stable condition that at least allows the manufacturing process to achieve positive economic returns. To tune a feedback controller in order to provide better performance, good transfer function and disturbance models are necessary. Identifying a process in open loop (i.e., when no adjustments take place) may result in low yields or unsafe manufacturing conditions, particularly if disturbances are nonstationary. Therefore, identifying, estimating and optimizing a manufacturing process under closed-loop feedback adjustments is of practical importance. In this paper, identification methods for processes operated under EWMA (Exponential Weighted Moving Average) and PI (Proportional Integral) control are proposed. These methods allow to tune optimally the feedback controller in use.

Conventional process identification techniques use the cross-correlation function between historical values of the process input, the controllable factor $\{u_t\}_{t=1}^{N}$, and of the process output, the observed deviations from target of the quality characteristics $\{e_t\}_{t=1}^{N}$. Effectiveness of this type of identification depends on the fact that the input is independent of the process disturbance [1]. In a process operating under a linear feedback controller, however, it is evident that the input depends on the output. This causes the cross-correlation to have no information on the process transfer function. The closed-loop identification method proposed by Box and MacGregor [2,3] is based on breaking the dependency between controlled input and output by adding a "dither" signal.

In this paper, several circumstances where a closed-loop system can be identified by the autocorrelation function of the output will be discussed. It is assumed that a Proportional Integral (PI) controller with known parameters is acting on the process while the output data were collected. This includes the case of the so-called EWMA (exponential weighted moving average) controller, popular in semiconductor manufacturing [4]. The disturbance is assumed to be a member of a simple yet useful family of stochastic models, which is able to represent drift [5]. As pointed out by Box and MacGregor [2], if *nothing* is known about the transfer function or about the disturbance model, closed loop identification from the autocorrelation of the output is not possible.

In the next section, the general models under our consideration are presented. The closed-loop descriptions of output deviations are then derived and their identification methods are discussed. The asymptotic mean square deviation (AMSD) of the output is used as a measure of the controller's performance and it will be minimized by tuning some parameters in the controller after the process transfer function has been identified. One simulated process is used to illustrate the procedure. It is shown that even when the process is partly mis-identified, the controller can still be tuned to a near optimal state.

2 Assumed Process and Disturbance Dynamics

Following Box *et al.* [1] modeling approach, it is assumed that an observed output deviation from target, e_t, consists of two components – a process "signal", S_t, and disturbance, N_t. That is, the signal generated by the underlying manufacturing mechanism can only be observed under the presence of a disturbance as follows:

$$e_t = S_t + N_t$$

More specifically, the process transfer function that generates S_t is written as a rational transfer function, and the disturbance N_t is modeled by an ARIMA (Autoregressive Integrated Moving Average) process. The assumed process is then

$$e_t = \alpha + \frac{(\beta_0 + \beta_1\mathcal{B} + \cdots + \beta_s\mathcal{B}^s)}{(1 - \phi_1\mathcal{B} - \cdots - \phi_r\mathcal{B}^r)}\mathcal{B}^b u_t + N_t \tag{1}$$

and

$$N_t = \delta + N_{t-1} - \theta\varepsilon_{t-1} + \varepsilon_t, \qquad |\theta| \leq 1 \tag{2}$$

where $\mathcal{B}$ is the backshift operator (defined as $\mathcal{B}e_t = e_{t-1}$), and α is a constant (not necessarily zero) representing the expected deviation from target when the input is set at a value of zero.

According to the Box-Jenkins taxonomy, the process is an (r,s,b) order transfer function plus an IMA(1,1) with drift disturbance. In practice, r, s and b are rarely larger than 2. The disturbance model contains a useful family of models. Depending on the different values δ and θ take, the disturbance is one of the processes listed on Table 1. Note that we allow θ to be equal to one; then the disturbance model is either a deterministic trend disturbance, which is useful to model wearing-off of a tool in a variety of manufacturing processes, or a white noise process in case $\delta = 0$.

δ	θ	Disturbance
0	0	Random Walk
$\neq 0$	0	Random Walk with Drift
0	$\neq 0$	IMA(1,1)
$\neq 0$	$\neq 0$	IMA(1,1) with Drift
$\neq 0$	1	Deterministic Trend
0	1	White Noise

Table 1: Disturbance models described by Equation (2)

3 Process Identification

3.1 ARMA modeling of output deviations e_t from closed-loop data

Suppose that for a manufacturing process operating in closed-loop, no first principle knowledge of the process dynamic mechanism is available. Therefore, it is necessary to identify an empirical transfer function model that best describes the process behaviour. On the other hand, the feedback controller functioning on the input data is intently designed and installed by control engineers, so its adjustment scheme is assumed to be known. When the controller is not optimal in the sense of not minimizing the mean square error of the process output, the output deviations will exhibit certain autocorrelation patterns that are useful for process identification. In this section, we will derive the ARMA models that describe the output deviations for some processes that are commonly encountered in manufacturing and show how to use some advanced statistical techniques to identify them. Here, we assume that one of two types of controller - either an EWMA or a PI controller - is in use in the closed-loop while the data were collected.

The EWMA controller has attracted considerable attention in recent years, especially for the run-to-run control of batch productions in semiconductor manufacturing (see references [6,7,8]). In such a closed-loop system, the effect of the process adjustments will be fully observed at the next output value, i.e., the process can be described by the following model:

$$e_t = \alpha + \beta u_{t-1} + N_t, \tag{3}$$

where α is the process offset, and β is the process gain. This is a particular case of model (1) with a (r,s,b)=(0,0,1) transfer function. In the EWMA controller, the process gain, β, is estimated by off-line experiments and is represented by b, and the initial estimate of α is a_0. The control scheme is as follows:

$$u_t = -\frac{a_t}{b}$$

and

$$a_t = \lambda(e_t - bu_{t-1}) + (1-\lambda)a_{t-1} \qquad (0 \le \lambda \le 1)$$

where λ is a parameter that can be adjusted to achieve a desired behavior. The EWMA controller updates a_t in order to reduce the estimation error of the process offset, α. In fact, it is easy to show that the adjustment at each step is proportional to the present output deviation, that is,

$$\nabla u_t = -\frac{\nabla a_t}{b} = -\frac{\lambda(e_t - bu_{t-1}) + (1-\lambda)a_{t-1} - a_{t-1}}{b} = -\frac{\lambda e_t}{b}. \tag{4}$$

One can compare this controller with a PI controller, a widely used industrial controller, which is a combination of two control schemes - proportional control and integral control:

$$u_t = k_P e_t + k_I \sum_{i=1}^{t} e_i \tag{5}$$

where k_P and k_I are the proportional and integral control constants respectively. An equivalent form of equation (5) would make the input adjustment depend linearly on the last two output deviations:

$$\nabla u_t = c_1 e_t + c_2 e_{t-1} \tag{6}$$

where $c_1 = k_P + k_I$ and $c_2 = -k_P$. As one can see from Equation (4), the EWMA controller is actually a special PI controller, i.e., a pure I controller with $c_1 = k_I = -\frac{\lambda}{b}$.

To derive the "closed-loop description" of the output deviations, we take first-order differences on the process and disturbance equations, then substitute the controller and disturbance functions into the process equation to obtain an ARMA model of the deviation. For instance, by taking first-order differences on process equation (3) and disturbance function (2), we have

$$\nabla e_t = \beta \nabla u_{t-1} + \nabla N_t, \tag{7}$$

and

$$\nabla N_t = \delta + (1 - \theta\mathcal{B})\varepsilon_t. \tag{8}$$

Substituting Equations (8) and (4) into (7), we get

$$(1-(1-\lambda\xi)\mathcal{B})e_t = \delta + (1-\theta\mathcal{B})\varepsilon_t \tag{9}$$

where, $\xi = \frac{\beta}{b}$, is a measure of the bias in the gain estimate. Therefore, under the adjustment of an EWMA controller, the sequence of deviations from this closed-loop system is an ARMA(1,1) process with an asymptotic mean value of $\frac{\delta}{\lambda\xi}$. This result was reported in del Castillo [5].

Equation (3) describes a simple manufacturing process where the process output is fully determined by the most recent value of the controllable factor. In some more complicated processes, the delay between the input adjustment and output observation could be longer than one time period and also the effect of adjustments could extend to several subsequent time periods. Therefore, it is important to study all possible ARMA models that describe the output deviations for a class of transfer functions. Commonly found process transfer functions include one-time or two-time delay and first-order dynamic models. The ARMA models that describe the output deviations can be derived by using the same procedure as we did for the transfer function of order (0,0,1). For brevity, our results for EWMA and PI controllers are summarized on Tables 2 and 3 respectively. On these tables, a transfer function is given first, followed by the ARMA model of the deviations from target and by the asymptotic process mean. Note that the order of the ARMA models does not exceed two for an EWMA controller and three for a PI controller. This implies that to identify a closed-loop process under these controllers, we should focus on searching a low order ARMA pattern from the process output data.

3.2 Stationarity of ARMA(2,q) process

Identifying a closed-loop process only based on the ARMA model of the output deviations could be ambiguous, because, as found in Table 2, there exist more than one transfer function corresponding to ARMA models of the

EWMA Controller		
Transfer function	r=0, s=0, b=1	$e_t = \alpha + \beta u_{t-1}$
Output deviation from target	ARMA(1,1)	$(1-(1-\lambda\xi)\mathcal{B})e_t = \delta + (1-\theta\mathcal{B})\varepsilon_t$
Process mean	$\mu_e = \frac{\delta}{\lambda\xi}$	
Transfer function	r=1, s=0, b=1	$(1-\phi\mathcal{B})e_t = \alpha + \beta u_{t-1}$
Output deviation	ARMA(2,2)	$(1-(1+\phi-\lambda\xi)\mathcal{B}-(-\phi)\mathcal{B}^2)e_t =$ $(1-\phi)\delta + (1-(\theta+\phi)\mathcal{B}-(-\theta\phi)\mathcal{B}^2)\varepsilon_t$
Process mean	$\mu_e = \frac{(1-\phi)\delta}{\lambda\xi}$	
Transfer function	r=0, s=1, b=1	$e_t = \alpha + \beta_1 u_{t-1} + \beta_2 u_{t-2}$
Output deviation	ARMA(2,1)	$(1-(1-\lambda\xi_1)\mathcal{B}-(-\lambda\xi_2)\mathcal{B}^2)e_t =$ $\delta + (1-\theta\mathcal{B})\varepsilon_t$
Process mean	$\mu_e = \frac{\delta}{\lambda(\xi_1+\xi_2)}$	
Transfer function	r=0, s=0, b=2	$e_t = \alpha + \beta u_{t-2}$
Output deviation	ARMA(2,1)	$(1-\mathcal{B}-(-\lambda\xi)\mathcal{B}^2)e_t = \delta + (1-\theta\mathcal{B})\varepsilon_t$
Process mean	$\mu_e = \frac{\delta}{\lambda\xi}$	
Transfer function	r=1, s=0, b=2	$(1-\phi\mathcal{B})e_t = \alpha + \beta u_{t-2}$
Output deviation	ARMA(2,2)	$(1-(1+\phi)\mathcal{B}-(-\lambda\xi-\phi)\mathcal{B}^2)e_t =$ $(1-\phi)\delta + (1-(\theta+\phi)\mathcal{B}-(-\theta\phi)\mathcal{B}^2)\varepsilon_t$
Process mean	$\mu_e = \frac{(1-\phi)\delta}{\lambda\xi}$	

Table 2: ARMA models describing the deviations from target from different EWMA controlled processes. In all cases, the disturbance is $N_t = \delta + N_t - \theta\varepsilon_{t-1} + \varepsilon_t$.

PI Controller		
Transfer function	r=0, s=0, b=1	$e_t = \alpha + \beta u_{t-1}$
Output deviation from target	ARMA(2,1)	$(1-(1+c_1\beta)\mathcal{B} - c_2\beta\mathcal{B}^2)e_t = \delta + (1-\theta\mathcal{B})\varepsilon_t$
Process mean	$\mu_e = \frac{\delta}{(c_1+c_2)\beta}$	
Transfer function	r=1, s=0, b=1	$(1-\phi\mathcal{B})e_t = \alpha + \beta u_{t-1}$
Output deviation	ARMA(2,2)	$(1-(1+\phi+c_1\beta)\mathcal{B} - (c_2\beta-\phi)\mathcal{B}^2)e_t =$ $(1-\phi)\delta + (1-(\theta+\phi)\mathcal{B} - (-\theta\phi)\mathcal{B}^2)\varepsilon_t$
Process mean	$\mu_e = \frac{(1-\phi)\delta}{(c_1+c_2)\beta}$	
Transfer function	r=0, s=1, b=1	$e_t = \alpha + \beta_1 u_{t-1} + \beta_2 u_{t-2}$
Output deviation	ARMA(3,1)	$(1-(1+c_1\beta_1)\mathcal{B} - (c_2\beta_1 + c_1\beta_2)\mathcal{B}^2 - c_2\beta_2\mathcal{B}^3)e_t =$ $\delta + (1-\theta\mathcal{B})\varepsilon_t$
Process mean	$\mu_e = \frac{\delta}{(c_1+c_2)(\beta_1+\beta_2)}$	
Transfer function	r=0, s=0, b=2	$e_t = \alpha + \beta u_{t-2}$
Output deviation	ARMA(3,1)	$(1-\mathcal{B} - c_1\beta\mathcal{B}^2 - c_2\beta\mathcal{B}^3)e_t = \delta + (1-\theta\mathcal{B})\varepsilon_t$
Process mean	$\mu_e = \frac{\delta}{(c_1+c_2)\beta}$	
Transfer function	r=1, s=0, b=2	$(1-\phi\mathcal{B})e_t = \alpha + \beta u_{t-2}$
Output deviation	ARMA(3,2)	$(1-(1+\phi)\mathcal{B} - (c_1\beta-\phi)\mathcal{B}^2 - c_2\beta\mathcal{B}^3)e_t =$ $(1-\phi)\delta + (1-(\theta+\phi)\mathcal{B} - (-\theta\phi)\mathcal{B}^2)\varepsilon_t$
Process mean	$\mu_e = \frac{(1-\phi)\delta}{(c_1+c_2)\beta}$	

Table 3: ARMA models describing the deviations from target of different PI controlled processes. In all cases, the disturbance model is $N_t = \delta + N_t - \theta\varepsilon_{t-1} + \varepsilon_t$

same order, like in the case of an ARMA(2,1) or an ARMA(2,2). Carefully comparing the estimated values of parameters in these models may help to distinguish different processes if reliable parameter estimates are available. Since it is assumed that the closed-loop process has been stabilized by the adjustments of a suboptimal controller, the stationarity conditions of an ARMA(2,q) provide some additional constraints for the process parameters that are useful for identification purposes.

As it is well-known, a stationary ARMA(2,q) process $(1-a_1\mathcal{B}-a_2\mathcal{B}^2)z_t = \Theta(\mathcal{B})\varepsilon_t$ must satisfy the following conditions:

$$a_1 + a_2 < 1 \qquad a_2 - a_1 < 1 \qquad |a_2| < 1.$$

Applying these conditions to the two ARMA(2,2) processes listed on Table 2, for the transfer function of order (1,0,1), we get

$$0 < \lambda\xi < 2(1+\phi) \qquad -1 < \phi < 1.$$

For the transfer function of order (1,0,2), we have

$$0 < \lambda\xi \qquad 2(1+\phi) < \lambda\xi \qquad -1-\phi < \lambda\xi < 1-\phi.$$

The feasible regions of ϕ and $\lambda\xi$ for these two processes are shown in Figure 1. These two regions nicely separate from each other, which indicates that one would be able to select a correct transfer function after comparing the estimated process parameters with their feasible regions. For example, if ϕ and $\lambda\xi$ estimated from an ARMA(2,2) process are 0.5 and 1 in an EWMA controlled process, then the transfer function of order (1,0,1) is the only one that should be accepted. Similarly, the feasible regions for the two ARMA(2,1) models on Table 2 are drawn in Figure 2. Again, there is little difficulty in identifying these two processes from their parameter estimates, because one process has the coefficient of $\mathcal{B}$ equal to one unit and the other does not.

Insert Figure 1 here.

Insert Figure 2 here.

As shown in Table 3, a process controlled by a PI controller may have a closed-loop description equal to an ARMA process of order as high as 3. It is difficult to draw the feasible parameter regions for ARMA(3,q), since there are three parameters involved. But one can see that for the two ARMA(3,1) processes, one has the coefficient of $\mathcal{B}$ equal to one and the other does not. Therefore, distinguishing between these alternatives based on the parameter estimates is possible in principle.

In practice, when the manufacturing process is exposed to substantial random noise, precise estimation of parameters in ARMA models is unrealistic. In addition, an accurate form of the model is not guaranteed and there is the possibility of model bias. However, as long as the model is a reasonable approximation of the true process, closed-loop identification will provide useful information for tuning the controller and improving performance as desired.

3.3 Techniques for identification of ARMA processes

Identifying an ARMA model from the sequence of output deviations is an essential procedure for obtaining a good process model. However, it is frequently difficult to identify a mixed ARMA process. The traditional approach of model identification utilizes the autocorrelation and partial autocorrelation functions of the process data and compares them with some theoretical patterns. This approach is only effective when dealing with pure AR or MA processes; but it is very difficult to determine a mixed ARMA model this way. Since the late 1980's, more identification techniques have been developed to handle this problem (see references [9,10,11,12]). Most of them compare the process data with a series of tentative models and select the one that fits best. The extended sample autocorrelation function (ESACF) and smallest canonical correlation (SCAN) proposed by Tsay and

Tiao [11,12] are two identification techniques that have been implemented in some statistical analysis program, such as SAS. They will be used for closed-loop identification purposes.

In the ESACF method, data are filtered through an AR model, whose autocorrelation coefficients are determined by a candidate ARMA model. The residuals of this filter are called the extended samples. It has been shown that the autocorrelation of these extended samples follows an MA(q) model if the true model is an ARMA(p,q), and the ARMA candidate we entertain has a MA polynomial of order higher or equal to q. More specifically, Tsay and Tiao had shown how to arrange the different condidate information in a table. The ESACF table will exhibit a triangular pattern of zeros when the candidate ARMA model has higher order than the true ARMA model. The SCAN method is based on some consistency properties of suitably normalized second-order sample moment equations and make use of the method of canonical correlation in standard multivariate analysis. The zeroes pattern on a SCAN table will be rectangular for any candidate model having higher order than the true model. These methods will be illustrated in Section 5.

4 Tuning the controller

After the process transfer function has been identified and its parameters have been estimated, it is possible to tune the controller to a near-optimal control state. By tuning a closed-loop process, it is meant to adjust the parameters of the controller to improve its performance. Usually, a controller has some parameters that can be reset by process engineers, such as λ in an EWMA controller or $(c_1,\ c_2)$ in a PI controller. The performance of the controller can be evaluated by the asymptotic mean square deviation (AMSD) of the process outputs. This criterion is valid when there is little cost associated with changing the controllable factor; otherwise, the variance of the controllable factor needs to be taken into consideration. Furthermore, process control is often required to react to output errors rapidly. For such reason, in order to tune a controller, the transient effect of the control scheme needs to be studied.

In this section, the AMSD is used as the controller performance criterion. The variance of the controllable factor will be discussed in Section 5. It is well known that the AMSD of a process consists of two parts: the variance of output deviations and the square of the bias. The calculation formula for the variance of an ARMA(2,q) process is given in the Appendix. The goal of tuning a closed-loop system is to minimize the AMSD subject to the process stationarity condition, that is,

$$Min \quad AMSD(e_t) = Var(e_t) + Bias(e_t)^2$$

$$s.t. \quad stationarity\ conditions.$$

By consulting the formula in the Appendix, exact AMSD expressions of many closed-loop processes can be obtained. For instance, for a process with the transfer function of order (1,0,1) being operated under an EWMA controller, its output deviations follow an ARMA(2,2) process with asymptotic mean value of $(1-\phi)\delta/\lambda\xi$. Therefore, its AMSD is

$$AMSD(e_t) = \frac{(1+\phi)(1-\phi)^2(1-\theta)^2 + 2\phi\lambda\xi(1-\theta)^2 + 2\theta\lambda\xi(1-\phi^2+\phi\lambda\xi)}{\lambda\xi(1-\phi)(2+2\phi-\lambda\xi)}\sigma_\varepsilon^2 + (\frac{\delta(1-\phi)}{\lambda\xi})^2. \tag{10}$$

If instead the process has a transfer function (1,0,2), the output deviations relate to another ARMA(2,2) model with different parameters and its AMSD is

$$AMSD(e_t) = \frac{(1+\phi)(1-\phi)^2(1-\theta)^2 + \lambda\xi(1+\theta^2)(1+\phi^2) - 2\theta\phi\lambda\xi(2\phi+\lambda\xi)}{\lambda\xi(1-\phi-\lambda\xi)(2+2\phi+\lambda\xi)}\sigma_\varepsilon^2 + (\frac{\delta(1-\phi)}{\lambda\xi})^2. \tag{11}$$

In a previous section, it was shown that in principle, process transfer functions can be identified correctly even when more than one closed-loop process follow an ARMA model of the same order. However, this strongly depends on the quality of our parameter estimates. Therefore, it is of interest to investigate the possibility of improving the controller's performance when the transfer function has been mis-identified. In the following example, we will show that a controller can be tuned to a near-optimal state as long as the estimated model is a reasonable approximation of the true one.

5 Example: a simulated process control with real disturbance data

Boyles [14] recently reported an uncontrolled process in which the fill weight deviation from target for a powdered food product was recorded with the controller turned off. He mentioned that the process was unstable and autocorrelated, because powder density was affected by several uncontrollable variables, such as batch-to-batch variations, and he also suggested that an integral-type controller should be used. In this section, those data of fill weight deviations are regarded as disturbance, to which a non-optimal EWMA controller (or I controller) is applied. We then apply the closed-loop identification and tuning methodology described in previous sections and optimize the controller.

First, the data set reported by Boyles is identified as an IMA(1,2) process, which is $\nabla N_t = (1 - \underset{(0.07)}{0.61} \mathcal{B} - \underset{(0.07)}{0.26} \mathcal{B}^2)\varepsilon_t$, where the numbers in paranthesis below the coefficients of $\mathcal{B}$ and $\mathcal{B}^2$ are the corresponding standard errors. The white noise sequence, $\{\varepsilon_t\}$, has an estimated variance of 207.5. This model will be used as the true disturbance in a simulated manufacturing process. Suppose the true process is repeatedly adjusted by an EWMA controller (I controller) with control parameter $\lambda = 0.4$. The adjusting action may be thought as twisting a valve that directly determines the powder volume per time unit, hence, the powder weight. Normally, the effect of this type of adjustment can be realized only partially during one time interval [1]. This results in a first order process transfer function with a one time delay. So the transfer function is characterized by the equation, $(1-\phi\mathcal{B})e_t = \beta u_{t-1}$. Here, let us assume that $\phi = 0.4$, $\beta = 1$ and $b = 0.8$ (b is the off-line estimate of β), so $\xi = \frac{\beta}{b} = 1.25$. By adding the same disturbance sequence as that in the open-loop process, we reconstruct a controlled process of fill weight deviation data as shown in Figure 3. It is evident that the process has been stabilized with mean value around 0. The estimated process variance is 292.2. We now illustrate our closed-loop identification methodology assuming the true model description is unknown.

Insert Figure 3 here.

SCAN and ESACF methods are applied to the simulated process output data to identify an ARMA model from which we can identify the process. The SCAN and ESACF tables are shown in Figure 4. One can see from the SCAN table that the pattern of rectangular zeroes starts from the AR(2) row and MA(1) column. This means that any ARMA process with order higher or equal to an ARMA(2,1) could be a candidate. The ESACF table does not shows a clear triangular pattern at low AR or MA order. Based on the parsimony principle, it is reasonable to guess that the closed-loop description of the process is ARMA(2,1). By fitting an ARMA(2,1) model to the output deviations from target, we have that the maximum likelihood estimators of the AR and MA parameters are $\hat{a}_1 = 0.688(0.179)$, $\hat{a}_2 = -0.338(0.131)$ and $\hat{b}_1 = 0.808(0.189)$, where the standard errors of these estimates appear in parenthesis. All of these estimates are significant by t-test. Therefore, the identified ARMA model is $(1-0.69\mathcal{B}+0.34\mathcal{B}^2)e_t = (1-0.8\mathcal{B})\varepsilon_t$. After consulting the list of EWMA closed-loop descriptions in Table 2, we speculate that the process transfer function is either $e_t = \beta_1 u_{t-1} + \beta_2 u_{t-2}$ or $e_t = \alpha + \beta u_{t-2}$. However, since for the latter alternative a_1 must be equal to 1 and our estimate $\hat{a}_1$ indicates this is not true, we conclude the process transfer function is $e_t = \beta_1 u_{t-1} + \beta_2 u_{t-2}$. Note that this model is a reasonable approximation of the true model, since the complete true parametric model is $(1 - 0.9\mathcal{B} + 0.4\mathcal{B}^2)e_t = (1 - \mathcal{B} - 0.02\mathcal{B}^2 - 0.10\mathcal{B}^3)\varepsilon_t$ if the controller and disturbance functions are substituted into the assumed process transfer function. Of course, the real process is never known to process engineers.

Insert Figure 4 here.

From the ARMA(2,1) model parameter estimates, the parameters in the identified transfer function and disturbance models are estimated as $\hat{\xi}_1 = 0.775$ (because, $1 - \lambda\hat{\xi}_1 = 0.69$), $\hat{\xi}_2 = 0.85$ (because, $-\lambda\hat{\xi}_2 = -0.34$), $\hat{\theta} = 0.8$, and $\hat{\delta} = 0$. Substituting them into the process equation, we can optimize the $AMSD(e_t)$ subject to the process stationarity conditions, that is,

$$Min \;\; AMSD(e_t) = \frac{1.64(1 + 0.85\lambda) - 1.6(1 - 0.775\lambda)}{1.625\lambda(1 - 0.85\lambda)(2 - 0.075\lambda)}$$

$$s.t. \;\; 0 < \lambda < 1.$$

Solution to this problem yields an optimal solution of λ equal to 0.12. By resetting λ in the EWMA controller and running the process with the same 190 disturbance data in Boyles under the re-tuned EWMA controller, we find that the estimated variance of the process output is reduced to 231.4. Note that this value is very close to the minimum variance one can achieve for the output of this process, namely, $\hat{\sigma}_e^2 = 228.0$, but the minimum variance can only be obtained when the correct ARMA model is identified and estimated perfectly.

5.1 Including the cost of adjustments

Sometimes the cost of adjustments cannot be ignored, so the objective function of the optimization model should be changed to a combination of both

ρ	λ	J	$\frac{AMSD}{\sigma_\varepsilon^2}$	$\frac{Var(\nabla u_t)}{\sigma_\varepsilon^2}$
0	0.12175	1.0122	1.0122	0.0234
0.1	0.11934	1.0145	1.0122	0.0225
0.5	0.11154	1.0229	1.0130	0.0198
1	0.10441	1.0321	1.0148	0.0173
2	0.09460	1.0477	1.0192	0.0143
5	0.07886	1.0831	1.0329	0.0100

Table 4: Optimal solutions to problem (12) in the example

output errors and adjustment efforts as proposed by Box and Luceño [13]:

$$Min \quad J = \frac{AMSD(e_t)}{\sigma_\varepsilon^2} + \rho\frac{Var(\nabla u_t)}{\sigma_\varepsilon^2} \tag{12}$$

For this example, Table 4 lists the optimal λ, the associated cost functions, the AMSD, and the adjustment variance for different values of ρ (which is a quantity defined by the user). Suppose from the table the value $\lambda = 0.08$ is chosen. If the process is again re-run with such EWMA controller and the same disturbance data as in Boyles [14], we find that $A\hat{MSD}(e_t)/\sigma_\varepsilon^2 = 1.0001$, and $V\hat{a}r(\nabla U_t)/\sigma_\varepsilon^2 = 0.0100$, which closely agree with the table.

6 Conclusion

In this paper, a method for identifying a process operating under the actions of a feedback controller was proposed. This method works for processes regulated with PI or EWMA (I) controllers under the assumption that the disturbance is IMA(1,1) with drift. We show that when the disturbance function is one from the proposed disturbance family, it is possible to identify some dynamic process models commonly encountered in manufacturing. ARMA models of the output deviations from these processes are provided. After identification, our approach suggests to tune the controller to a near-optimal setting according to a well-known performance criterion, which is either the asymptotic mean square deviation (AMSD) of the process output, or a weighted sum of AMSD of the output and variance of the adjustments.

Appendix. Variance of an ARMA(2,q) Process, $q \leq 2$.

For an ARMA(2,1) process such as

$$(1 - a_1\mathcal{B} - a_2\mathcal{B}^2)z_t = (1 - b_1\mathcal{B})\varepsilon_t, \tag{13}$$

its autocovariance is computed by multiplying both sides of the above equation by z_{t-k} and taking expectation:

$$(1 - a_1\mathcal{B} - a_2\mathcal{B}^2)\gamma_k = \gamma_{z\varepsilon}(k) - b_1\gamma_{z\varepsilon}(k-1)$$

where γ_k is the autocorrelation coefficient of z, and $\gamma_{z\varepsilon}$ is the cross-correlation coefficient of z and ε.

When $k = 0$, we have that

$$\gamma_0 = a_1\gamma_1 + a_2\gamma_2 + \sigma_\varepsilon^2 - b_1\gamma_{z\varepsilon}(-1).$$

When $k = 1$,

$$\gamma_1 = a_1\gamma_0 + a_2\gamma_1 - b_1\sigma_\varepsilon^2$$

and when $k \geq 2$,

$$\gamma_k = a_1\gamma_{k-1} + a_2\gamma_{k-2}.$$

Also, multiplying both sides of equation (13) by ε_{t-1} and taking expectation, we get

$$\gamma_{z\varepsilon}(-1) = (a_1 - b_1)\sigma_\varepsilon^2.$$

Therefore, variance of z_t is

$$\gamma_0 = \frac{(1 - a_2)(1 + b_1^2) - 2b_1a_1}{(1 + a_2)(1 - a_1 - a_2)(1 + a_1 - a_2)}\ \sigma_\varepsilon^2. \tag{14}$$

Following a similar derivation, the variance of an ARMA(2,2) process $(1 - a_1\mathcal{B} - a_2\mathcal{B}^2)z_t = (1 - b_1\mathcal{B} - b_1\mathcal{B}^2)\varepsilon_t$ can be shown equal to

ARMA(2,1) model	$a_1\mathcal{B}$	$a_2\mathcal{B}^2$
$b=1,\ r=0,\ s=1$	$1-\lambda\xi_1$	$-\lambda\xi_2$
$b=2,\ r=0,\ s=0$	1	$\lambda\xi$
ARMA(2,2) model	$a_1\mathcal{B}$	$a_2\mathcal{B}^2$
$b=1,\ r=1,\ s=0$	$1+\phi-\lambda\xi$	$-\phi$
$b=2,\ r=1,\ s=0$	$1+\phi$	$-\lambda\xi-\phi$

Table 5: Parameters of four ARMA models

$$\gamma_0 = \frac{(1-a_2)(1+b_1^2+b_2^2)-2b_1(1-b_2)a_1-2b_2a_2-2b_2(a_1^2-a_2^2)}{(1+a_2)(1-a_1-a_2)(1+a_1-a_2)}\,\sigma_\varepsilon^2. \quad (15)$$

As one can see from Table 2, each closed-loop process has a different expression for the coefficients of $\mathcal{B}$ and $\mathcal{B}^2$ in its ARMA model of the output deviations. The coefficients are listed in Table 5.

References

1. Box GEP, Jenkins GM, and Reinsel GC. *Time Series Analysis, Forecasting, and Control*, Englewood Cliffs: Prentice Hall, 1994.

2. Box GEP and MacGregor JF. The Analysis of Closed-Loop Dynamics-Stochastic Systems. *Technometrics* 1974 **16**: 391-398.

3. Box GEP and MacGregor JF. Parameter Estimation with Closed-Loop Operating Data. *Technometrics* 1976 **18**: 371-384.

4. Moyne J, del Castillo E, and Hurwitz AM. *Run-to-Run Control in Semiconductor Manufacturing*, CRC press, 2000.

5. Del Castillo E. Closed-Loop Disturbance Identification and Controller Tuning for Discrete Manufacturing Processes. to appear in *Technometrics* 2001.

6. Ingolfsson A and Sachs E. Stability and Sensitivity of an EWMA Controller. *Journal of Quality Technology* 1993 25 **4**: 271-287.

7. Sachs E, Hu A, Ingolfsson A. Run by Run Process Control: Combining SPC and Feedback Control. *IEEE Transactions on Semiconductor Manufacturing* 1995 8 **1**: 26-43.

8. Del Castillo E and Hurwitz A. Run to Run Process Control: a Review and Some Extensions. *Journal of Quality Technology* 1997 29 **2**: 184-196.

9. Gray HL, Kelley GD, McIntire DD. A New Approach to ARMA Modeling. *Commu. Statist. - Simula. Computa.* 1978 B7 **1**: 1-77.

10. Hannan EJ and Rissanen J. Recursive Estimation of Mixed Autoregressive-Moving Average Order. *Biometrika* 1982 69 **1**: 81-94.

11. Tsay RS and Tiao GC. Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. *Journal of the American Statistical Association* 1984 79 **385**: 84-96.

12. Tsay RS and Tiao GC. Use of Canonical Analysis in Time Series Model Identification *Biometrika* 1985 72 **2**: 299-315.

13. Box GEP and Luceño A. *Statistical Control by Monitoring and Feedback Adjustment*; John Wiley & Sons, New York, NY, 1997.

14. Boyles RA. Phase I Analysis for Autocorrelated Processes. *Journal of Quality Technology* 2000 32 **4**: 395-409.

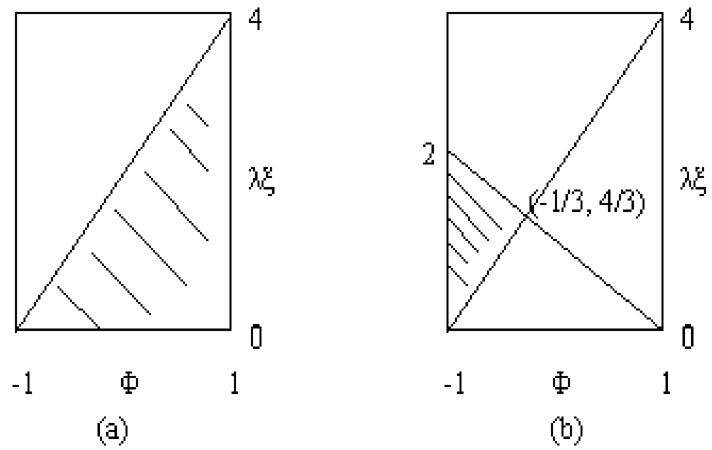

Figure 1: Stability region of two ARMA(2,2) processes. a) Transfer function is (1,0,1); b) transfer function is (1,0,2).

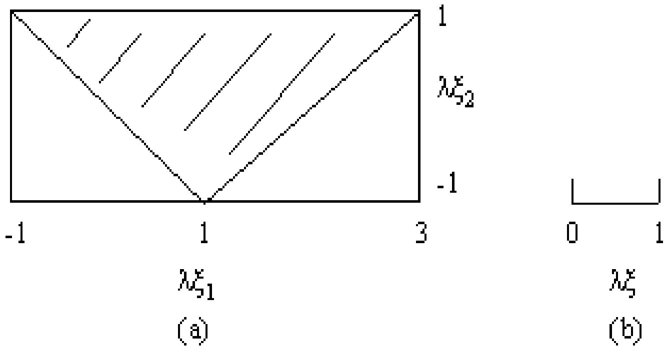

Figure 2: Stability region of two ARMA(2,1) processes. a) Transfer function is (0,0,1). Here, $\xi_i = \beta_i / b_i$. b) Transfer function is (0,0,2).

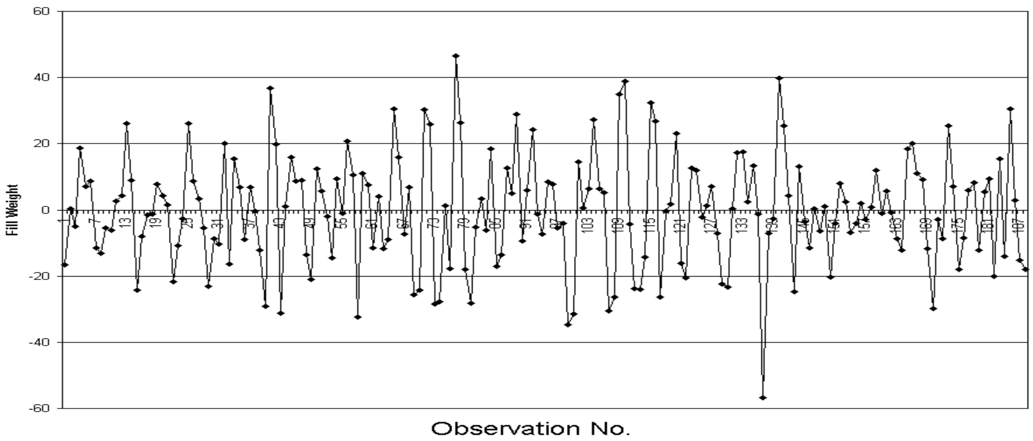

Figure 3: A controlled closed-loop process

Squared Canonical Correlation Estimates

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.0086	0.1180	0.0697	0.0295	0.0047	0.0352
AR 1	0.1253	0.1062	0.0003	0.0243	0.0375	0.0185
AR 2	0.0479	0.0120	0.0219	0.0114	0.0005	0.0003
AR 3	0.0922	0.0253	0.0229	0.0120	0.0004	0.0005
AR 4	0.0129	0.0005	0.0030	0.0092	0.0175	0.0159
AR 5	0.0042	0.0010	<.0001	0.0088	<.0001	0.0027

Extended Sample Autocorrelation Function

Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	0.0927	-0.3405	-0.2616	-0.1685	0.0670	0.1828
AR 1	0.2432	-0.2825	-0.0286	-0.2209	0.0734	0.1583
AR 2	-0.4617	-0.2453	0.0534	-0.1709	-0.0307	0.0221
AR 3	-0.4885	-0.4594	0.3032	-0.1962	-0.1993	0.0043
AR 4	-0.3301	-0.0885	-0.2734	-0.1840	0.0284	-0.0316
AR 5	-0.4328	0.0781	-0.0227	-0.2435	0.0078	-0.0696

Figure 4: SCAN and ESACF tables of the simulated example