Linear filter model representations for integrated process control with repeated adjustments and monitoring

Changsoon Park *

Department of Statistics, Chung-Ang University, Dongjak-gu Hul suk-dong 221, Seoul 156-756, Republic of Korea

A R T I C L E   I N F O

Article history:
Received 28 January 2009
Accepted 26 May 2009
Available online 13 June 2009

AMS 2000 subject classifications:
primary 93E20
secondary 62N05

Keywords:
Integrated process control
Repeated adjustment
Engineering process control
Statistical process control
Expected cost per unit time
Controller
Readjustment

A B S T R A C T

An integrated process control (IPC) procedure is a scheme which combines the engineering process control (EPC) and the statistical process control (SPC) procedures for the process where the noise and a special cause are present. The most efficient way of reducing the effect of the noise is to adjust the process by its forecast, which is done by the EPC procedure. The special cause, which produces significant deviations of the process level from the target, can be detected by the monitoring scheme, which is done by the SPC procedure. The effects of special causes can be eliminated by a rectifying action. The performance of the IPC procedure is evaluated in terms of the average run length (ARL) or the expected cost per unit time (ECU). In designing the IPC procedure for practical use, it is essential to derive its properties constituting the ARL or ECU based on the proposed process model. The process is usually assumed as it starts only with noise, and special causes occur at random times afterwards. The special cause is assumed to change the mean as well as all the parameters of the in-control model. The linear filter models for the process level as well as the controller and the observed deviations for the IPC procedure are developed here.

© 2009 The Korean Statistical Society. Published by Elsevier B.V. All rights reserved.

1. Introduction

Observations taken from processes such as chemical and process industries tend to wander away from the target when the process is left by itself without proper control. The wandering behavior of the process level arises from the noise, which cannot be removed from the process due to some technical or economic reasons. The most often used model in describing the wandering behavior of the process, which comes from its inherent disturbances, is the IMA(1,1) model [see Apley and Kim (2004), Box and Kramer (1992), Box and Luceño (1997), and Vander Wiel (1996)].

Special causes occur at random times due to some assignable reasons as the process is going on. The special cause makes the process level move away from the target and the process cannot be cured until it is removed. Since there is an assignable reason for the special cause, it can be removed from the process once it is detected.

Noise and the special cause are the two major sources of variations which make the process level move away from the target. Since the noise cannot be removed from the process, the effective way of minimizing its effects is to compensate the process by the appropriate forecast of future observations. Since the special cause can be removed from the process if it is detected, on the other hand, the effective way of monitoring the process is to find it quickly and eliminate it from the process.

The activity of compensating the process with noise by its controller to make the process level close to the target is called the engineering process control (EPC) or automatic process control (APC). An efficient example of the EPC is to adjust the process by the minimum mean square estimate (MMSE) controller. Examples of the study for the EPC procedures are Castillo (2002), Jiang and Tsui (2002), and Tsung, Zhao, Xiang, and Jiang (2006).
The activity of monitoring the occurrence of a special cause is called a statistical process control (SPC). A process until the occurrence of a special cause is called an in-control (IC) process, and a process after the occurrence of a special cause is called an out-of-control (OC) process. An effective monitoring is to give a signal as early as possible when a special cause has occurred, and not to give a signal when no special cause has occurred. Examples of the study for the SPC procedures when observations have serial correlations are Castagliola and Tsung (2005), Hu and Roan (1996), Jiang and Tsui (2002), and Keats and Hubele (1989).

When noise is inherent to the process and a special cause can occur during the process, the simultaneous application of the EPC and SPC procedures is needed to control the process. Such an activity is called an integrated process control (IPC). The IPC procedure has been studied for more than two decades in the literature, such as Capilla, Ferrer, Romero, and Hualda (1999), Jiang (2004), Jiang and Tsui (2000), Montgomery, Keats, Runger, and Messina (1994), Nembhard and Chen (2007), Pan and del Castillo (2003), Park and Reynolds (2008), Runger, Testik, and Tsung (2006), Tsung, Shi, and Wu (1999), Vander Wiel (1996), and Vander Wiel, Tucker, Faltin, and Doganaksoy (1992). In the IPC procedure, we assume that special causes occur at random times as the process with noise is going on. A usual approach for the IPC procedure is, at each sampling time, to apply the EPC procedure first and then apply the SPC procedure to the result by the EPC procedure. That is, the process is adjusted by the MMSE controller first and then the adjusted process is controlled by an effective monitoring scheme.

In the IPC procedure, the observed deviations are monitored during the process where adjustments are repeatedly done by its controller. In such cases the effects of a special cause are assumed to change the process mean as well as all the parameters involved in describing the noise model. Here, the IC process means the process adjusted by the controller until a special cause is occurred, and the OC process means the adjusted process after a special cause has occurred. The controller used for the OC process will be the same as the one used for the IC process until the monitoring scheme gives a signal since the occurrence of the special cause has not been noticed yet. The controller which is optimal for the IC process is not optimal for the OC process. Thus, the main purpose of the monitoring scheme is to give a true signal as soon as possible when a special cause has occurred. Then a rectifying action to the OC process will eliminate the effect of the special cause.

The use and properties of the SPC procedure for the OC process are complicated when the adjustment designed for the IC process is made to the OC process where the effects of the noise and the special cause are mixed. In this article, statistical models have been developed by the use of linear filter expressions for considering the effects of the special cause to the process model adjusted by the controller. Since the development of the process model is essential in evaluating the properties of the IPC procedure, we focus on developing the process model expressions rather than evaluating the process control procedures. We assume that all the process parameters can be estimated accurately for the purpose of theoretical development. Although the process parameters may not be estimated accurately in practice, the effects of inaccurately estimated parameters will not be considered here.

We consider only for the repeated adjustment, which adjusts the process at every sampling time, and the process adjustment dynamics is of first order with delay time 1, that is the observed deviation is fully affected by the immediately previous time adjustment. The use of a linear filter model makes the derivations of the expressions of the process model for the various cases much simpler than the use of the explicit forms. Examples of using the linear filter model are Apley and Chin (2007), Castillo (2002), and Chin and Apley (2006).

2. The IC process model

Suppose that the deviation of the process level from the target, when it is not controlled appropriately, is represented by the noise, and it is well approximated by the IMA(1,1) model. Then the IMA(1,1) model for the noise is expressed as

\[ Z_t = \frac{1 - \theta B}{1 - B} \epsilon_t, \]

for the backshift operator \( B \), where \( \{\epsilon_t, t = 1, 2, \ldots\} \) is a sequence of white noises with variance \( \sigma^2 \) and \( \theta \) is the moving average parameter. Since there is always a starting time in the process model, we assume that \( Z_t = 0 \) and \( \epsilon_t = 0 \) for \( t \leq 0 \).

The model in Eq. (1) is used to represent the amount of deviation from the target for the IC process level.

The random shock form of Eq. (1) is

\[ Z_t = \epsilon_t + \lambda \sum_{j=1}^{t-1} \epsilon_{t-j}, \]

where \( \lambda = 1 - \theta \). The linear filter used in Eq. (1) is an infinite summation as

\[ Z_t = (1 + \lambda B + \lambda^2 B^2 + \cdots) \epsilon_t. \]

If the process starts from time 1, the above expression is a finite summation as

\[ Z_t = (1 + \lambda B + \lambda^2 B^2 + \cdots + \lambda B^{t-1}) \epsilon_t \]

\[ = \frac{1 - \theta B - \lambda B^t}{1 - B} \epsilon_t. \]

However, such expression used for the finite summation is more complicated than Eq. (1). Thus we use the infinite summation notation by giving restrictions on \( Z_t \) and \( \epsilon_t \), that is \( Z_t = \epsilon_t = 0 \) for \( t \leq 0 \), rather than using the finite summation notation.
The random shock, $\varepsilon_t$, can be expressed as

$$\varepsilon_t = \frac{1 - B}{1 - \theta B} Z_t.$$  (2)

Since Eq. (2) can be rewritten as

$$Z_t = \varepsilon_t + \frac{\lambda}{1 - \theta B} Z_{t-1},$$

we have the expression of the MMSE forecast for $Z_t$ as

$$\hat{Z}_t = \frac{\lambda}{1 - \theta B} Z_{t-1},$$

and by Eq. (1),

$$\hat{Z}_t = \frac{\lambda}{1 - B} \varepsilon_{t-1}.$$  

Let $A_t$ be the total adjustment made up to time $t$ from the beginning of the process and $a_t$ be the one time adjustment made at time $t$ additional to the adjustment made up to time $t - 1$. Then the MMSE controllers for $Z_t$ are

$$A_t = - \frac{\lambda}{1 - B} \varepsilon_{t-1},$$

and

$$a_t = A_t - A_{t-1} = -\lambda \varepsilon_{t-1}.$$  (3)

The observed deviation when the adjustment $A_t$ is made to the process is

$$e_t = Z_t + A_t = \frac{1 - \theta B}{1 - B} \varepsilon_t - \frac{\lambda B}{1 - B} \varepsilon_t = \varepsilon_t.$$  

That is, the result of the MMSE controller is the current random shock alone.

In the repeated adjustment scheme, we observe the actual deviation $e_t$ rather than $Z_t$ and adjust the current process by the one time adjustment, $a_t$, rather than the total adjustment, $A_t$. Thus the adjusted IMA(1,1) model by the MMSE controller, that is the observed deviation, is expressed as

$$e_t = \varepsilon_t$$

with the one time adjustment at time $t$,

$$a_t = -\lambda \varepsilon_{t-1}.$$  (4)

Therefore, the total adjustment and the one time adjustment are expressed by the observed deviations as

$$A_t = - \frac{\lambda}{1 - B} \varepsilon_{t-1},$$

and

$$a_t = -\lambda \varepsilon_{t-1}.$$  (5)

3. The OC process model

Suppose that a special cause occurs at a random time between times $\tau$ and $\tau + 1$, $\tau \geq 0$, and its effect appears from time $\tau + 1$. The effect of a special cause can be any of the change in the mean of $Z_{\tau+k}$, the change in the variance of $\varepsilon_{\tau+k}$, the change in $\theta$, and the change in $a_{\tau+k}$, for $k = 1, 2, \ldots$. At time $\tau + k$, $(k = 1, 2, \ldots)$ the mean of $Z_{\tau+k}$ is assumed to shift to a nonzero value, $\varepsilon_{\tau+k}$ is assumed to be $\bar{\sigma} \varepsilon_{\tau+k}$, $(\bar{\sigma} > 1)$, $\theta$ is assumed to be $\tilde{\theta} \theta$, $(\tilde{\theta} \neq 1)$, and $a_{\tau+k}$ is assumed to be $\tilde{\lambda} a_{\tau+k}$, $(\tilde{\lambda} \neq 1)$. In deriving the properties of the IPC procedure, we assume that all these changes in the parameters can be estimated accurately.

For cases where all these changes have occurred to the OC process, we develop a new process model, \(Z'_k = Z_{\tau+k} + A_k\), $(k = 1, 2, \ldots)$, which denotes the process level adjusted up to time $\tau$. The shifted mean of the process, $Z'_\tau$, is denoted by $M_{Z'}(k)$.

Since, for $k = 1$,

$$Z_{\tau+1} = Z_{\tau} + \bar{\sigma} \varepsilon_{\tau+1} - \tilde{\theta} \theta \varepsilon_{\tau} + M_{Z'}(1)$$

$$= \bar{\sigma} \varepsilon_{\tau+1} + (1 - \tilde{\theta} \theta) \varepsilon_{\tau} + \lambda \sum_{j=1}^{\tau} \varepsilon_{\tau-j} + M_{Z'}(1),$$
and for \( k = 2, 3, \ldots \),

\[
Z_{t+k} = \bar{\sigma} e_{t+k} + (1-\tilde{\theta}) \sum_{j=1}^{k-1} \bar{\sigma} e_{t+k-j} + (1-\tilde{\theta}) e_t + \lambda \sum_{j=1}^{t-1} e_{t-j} + M_{Z'}(k),
\]

we have, for \( k = 1, 2, \ldots \),

\[
Z'_{k} = (1 + \lambda_1 B + \lambda_1 B^2 + \cdots + \lambda_1 B^{k-1}) e'_k + M_{Z'}(k).
\]

where \( e'_0 = e_t, \ e'_k = \bar{\sigma} e_{t+k} \) for \( k > 0 \), and \( \lambda_1 = 1 - \theta_1 \) for \( \theta_1 = \tilde{\theta} \).

The new model \( Z'_k \) in Eq. (6) can be expressed as the sum of the two parts: the random part, \( P_{Z'}(k) \), and the mean part, \( M_{Z'}(k) \), that is,

\[
Z'_k = P_{Z'}(k) + M_{Z'}(k).
\]

The random part is a part of \( Z'_k \) which involves all the random shocks, while the mean part is the remaining part of \( Z'_k \) which does not involve the random shock. Then the random part explains all the changes except the change in the mean of \( Z'_k \) and can be written by Eq. (6) as, for \( k = 1, 2, \ldots \),

\[
P_{Z'}(k) = \frac{1 - \theta_1 B}{1 - B} \varepsilon_k
\]

with \( P_{Z'}(k) = \varepsilon'_k = 0 \) for \( k < 0 \), which is an IMA(1,1) model with parameter \( \theta_1 \). Note that an IMA(1,1) model for the noise is generated as the random part of the OC process adjusted up to \( \tau \) due to the effects of the special cause.

The mean part is classified into the three different types of change in the mean, which are the sustained shift (SS), the sustained drift (SD), and the transient shift (TS) in the mean. These terms concerning the changes in the mean have been studied by Reynolds and Stoumbos (2004a,b), and Tsung and Tsui (2003).

The effect of the SS in the mean is the addition of a constant, \( \bar{\mu} \sigma_e \), to the random part, that is, the mean part is

\[
M_{Z'}^{SS}(k) = \bar{\mu} \sigma_e, \quad k = 1, 2, \ldots
\]

This is the same as writing

\[
Z'_1 = Z'_0 + \varepsilon'_1 - \theta_1 \varepsilon'_0 + \bar{\mu} \sigma_e
\]

\[
Z'_k = Z'_{k-1} + \varepsilon'_k - \theta_1 \varepsilon'_{k-1}, \quad k = 2, 3, \ldots
\]

The effect of the SD in the mean is the addition of \( kr \sigma_e \) to the random part for some constant \( r \), that is, the mean part is

\[
M_{Z'}^{SD}(k) = kr \sigma_e, \quad k = 1, 2, \ldots
\]

This is the same as writing

\[
Z'_1 = Z'_0 + \varepsilon'_1 - \theta_1 \varepsilon'_0 + r \sigma_e
\]

\[
Z'_k = Z'_{k-1} + \varepsilon'_k - \theta_1 \varepsilon'_{k-1} + r \sigma_e, \quad k = 1, 2, \ldots
\]

The effect of the TS in the mean is the addition of constant, \( \bar{\mu} \sigma_e \), to the random part for \( k = 1, 2, \ldots, l \) only, that is, the mean part is

\[
M_{Z'}^{TS}(k) = \begin{cases} 
\bar{\mu} \sigma_e, & k = 1, 2, \ldots, l \\
0, & k = l + 1, l + 2, \ldots
\end{cases}
\]

This is the same as writing

\[
Z'_1 = Z'_0 + \varepsilon'_1 - \theta_1 \varepsilon'_0 + \bar{\mu} \sigma_e
\]

\[
Z'_k = Z'_{k-1} + \varepsilon'_k - \theta_1 \varepsilon'_{k-1}, \quad k = 2, 3, \ldots, l,
\]

\[
Z'_{k+1} = Z'_{k} + \varepsilon'_{k+1} - \theta_1 \varepsilon'_k - \bar{\mu} \sigma_e,
\]

\[
Z'_{k} = Z'_{k-1} + \varepsilon'_k - \theta_1 \varepsilon'_{k-1}, \quad k = l + 2, l + 3, \ldots
\]

Now, consider the change in the attempted adjustment, \( \bar{\lambda} \neq 1 \), the actual adjustment realized to the process will be \( \bar{\lambda} \) times the attempted adjustment calculated from the MMSE controller. Now let \( A'_k = A_{t+k} - A_t \) be the actual total adjustment made to the process after time \( \tau \), and \( e'_k \) be the observed deviation of the process level when adjusted by \( A'_k \). The attempted one time adjustment for the OC process until a signal will be \( -\lambda \varepsilon'_{k-1} \), which is the same controller used for the IC process. Thus the actual total adjustment for the OC process will be updated as

\[
A'_k = A'_{k-1} - \bar{\lambda} (\lambda \varepsilon'_{k-1}).
\]

Inserting the relation \( e'_k = Z'_k + A'_k \) to the above equation gives

\[
A'_k = \frac{-\lambda' B}{1 - \lambda_2 B} \varepsilon'_{k-1}.
\]
Table 1
Summary of the mean parts (units are listed in parentheses).

<table>
<thead>
<tr>
<th></th>
<th>SS ($\mu_\sigma$)</th>
<th>SD ($\sigma_\epsilon$)</th>
<th>TS ($\mu_\sigma$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_{e}(k)$</td>
<td>1</td>
<td>$k$</td>
<td>$1, \ k = 1, 2, \ldots, l$</td>
</tr>
<tr>
<td>$M_{e}(k)$</td>
<td>$\lambda_2^{k-1} - 1$</td>
<td>$\frac{1-\lambda_2}{\lambda_2} - k$</td>
<td>$\lambda_2^{k-1}, \ k = 1, 2, \ldots, l$</td>
</tr>
<tr>
<td>$M_{e}(k)$</td>
<td>$-\lambda_2^{k-2}$</td>
<td>$\lambda_2^{k-1} - 1$</td>
<td>$-\lambda_2^{k-2} \lambda_2^{k-1} - 1, \ k = 1, l + 1, l + 2, \ldots$</td>
</tr>
<tr>
<td>$M_{e}(k)$</td>
<td>$\lambda_2^{k-1}$</td>
<td>$\frac{1-\lambda_2}{\lambda_2} - 1$</td>
<td>$\lambda_2^{k-1} - \lambda_2^{k-1} - 1, \ k = 1, l + 1, l + 2, \ldots$</td>
</tr>
</tbody>
</table>

where $\lambda' = \tilde{\lambda} \lambda$ and $\lambda_2 = 1 - \lambda'. If we express $A'_k$ as the sum of two parts: the random part, $P_A(k)$, and the mean part, $M_A(k)$, that is, $A'_k = P_A(k) + M_A(k)$, we have the following expressions by Eq. (7) and the above equation.

$$P_A(k) = \frac{-\lambda B}{1 - \lambda_2 B} P_e(k) = \frac{-\lambda B (1 - \theta_1 B)}{(1 - \lambda_2 B)(1 - B)} e'_k,$$

(11)

$$M_A(k) = \frac{-\lambda B}{1 - \lambda_2 B} M_e(k).$$

(12)

The one time adjustment can also be expressed as the sum of the random part and the mean part, that is $A'_k = P_A(k) + M_A(k)$. From Eqs. (11) and (12), we have the following expressions.

$$P_a(k) = \frac{-\lambda B (1 - \theta_1 B)}{1 - \lambda_2 B} e'_k,$$

$$M_a(k) = \frac{-\lambda B (1 - B)}{1 - \lambda_2 B} M_e(k).$$

The observed deviation can also be written as the sum of two parts, the random part, $P_e(k)$, and the mean part, $M_e(k)$, that is, $e'_k = P_e(k) + M_e(k)$. By Eqs. (7) and (11), the random part is

$$P_e(k) = P_{e'}(k) + P_A(k)$$

$$= \frac{1 - \theta_1 B}{1 - \lambda_2 B} e'_k,$$

(13)

and by Eq. (12), the mean part is

$$M_e(k) = M_{e'}(k) + M_A(k)$$

$$= \frac{1 - B}{1 - \lambda_2 B} M_e(k).$$

(14)

Thus we see from Eq. (13) that the random part of the observed deviation is an ARMA(1,1) model with an autoregressive parameter, $\lambda_2$, and the moving average parameter $\theta_1$. This means that the observed deviations of the OC process, which is adjusted by the IC process controller, are no longer white noise as in the IC process.

Besides the linear filter model expressions, we have the following random shock model expressions.

$$P_{e'}(k) = e'_k + \lambda_3 \sum_{j=1}^{k} e'_{k-j},$$

$$P_A(k) = \sum_{j=1}^{k} (-\lambda_1 + \lambda_3 \lambda_2) e'_{k-j},$$

$$P_e(k) = e'_k + \lambda_2 \sum_{j=1}^{k} \lambda_2^{j-1} e'_{k-j},$$

where $\lambda_3 = \lambda_1 - \lambda'$. The expressions of the mean part for the three different types of mean changes are summarized in Table 1.

4. Process monitoring and efficiency

In this section, the general form of the monitoring scheme for the IPC procedure is explained. The control statistic is composed of the observed deviations after adjustment for a given type of monitoring. The general form of the frequent
control statistic used at time \( t \) using the observed deviation, \( e_t \), can be expressed as
\[
E_t = \alpha h(e_t) + \beta g(E_{t-1}),
\]
where \( E_0 = 0 \) and for some given constants \( \alpha, \beta \), and given functions \( h \) and \( g \).

Examples of the most often used monitoring schemes are the Shewhart chart, the CUSUM chart and the EWMA chart. When the Shewhart chart is applied, \( \alpha = 1, \beta = 0 \), and \( h(x) = x \). When the CUSUM chart is applied, \( \alpha = 1, \beta = 1 \), \( g(x) = \max(x, 0) \), and \( h(x) = x - \gamma \) for detecting positive shifts and \( h(x) = -x - \gamma \) for detecting negative shifts where \( \gamma \) is selected appropriately. When the EWMA chart is applied, \( \beta = 1 - \alpha \) for given \( 0 < \alpha < 1 \) and \( h(x) = g(x) = x \).

The monitoring procedure using the defined control statistic, \( E_t \), is to give a signal when
\[
q(E_t) \geq l
\]
for some simple function \( q \) and the given control limit \( l \). For the Shewhart chart and the EWMA chart, we use \( q(x) = |x| \), and for the CUSUM chart, we use \( q(x) = x \).

The properties of the monitoring scheme are usually determined by the two design categories: statistical design and economic design.

The statistical design is to design the monitoring scheme so that it gives a signal as early as possible when the process is out of control and gives a signal as late as possible when the process is in control. In statistical design, the efficiency of the control scheme is completely evaluated in terms of the average run length (ARL). The run length of the IC process, \( L_0 \), is the number of observations taken until a signal assuming that no special cause will occur. The run length of the OC process, \( L_1 \), is the number of observations taken until a signal assuming that a special cause has occurred already from the beginning of the process. In statistical design, thus, we design the control chart by selecting its parameters which minimize \( E(L_1) \) subject to \( E(L_0) \geq A_0 \) for some given constant \( A_0 \).

The economic design is to design the monitoring scheme so that it may produce the minimum cost per unit time. Thus various cost parameters are considered in the economic design. Examples of the most important cost parameters are cost per monitoring (\( C_M \)), cost per adjustment (\( C_A \)), off-target cost per squared deviation from target (\( C_T \)), and cost per false signal (\( C_F \)). The process properties considered in the economic design are the number of adjustments (\( N_A \)), the number of false signals (\( N_F \)), and the sum of the squared observed deviations from target (\( SS_e \)) as well as \( E(\tau) \) and \( E(L_1) \). In evaluating the cost per unit time, a cycle of the process is defined as the time from the start of the process to the first true signal, for the first cycle, and the time from a true signal to the next true signal, for the other cycles. We assume that a rectifying action followed by a true signal resets the process conditions the same as when the process is in control, thus a series of cycles corresponds to a renewal process.

The economic design is to determine the parameters of the control chart which minimize the expected cost per unit time (ECU) defined as
\[
ECU = C_M + \frac{C_A \cdot E(N_A) + C_T \cdot E(SS_e) + C_F \cdot E(N_F)}{E(\tau + L_1)}.
\]

Several additional parameters have been used in defining ECU in the literature [see Lorenzen and Vance (1986) and Montgomery (2005)], but only critical parameters are introduced in the above equation to explain the main idea of the ECU in a simple manner [see Jiang and Tsui (2000)]. The number of adjustments and the sum of the squared observed deviations are the sums of the corresponding numbers when the process is in control and out of control, while the number of false signals is the number only when the process is in control. When a repeated adjustment is made to the process, the number of adjustments is the same as the cycle length, that is, \( N_A = \tau + L_1 \). The sum of the squared observed deviations is calculated as
\[
SS_e = \sum_{t=1}^{\tau+L_1} e_t^2.
\]

When the process is in control, the observed deviation, \( \{e_t, t = 1, 2, \ldots, \tau\} \), is a sequence of white noises, and thus the sequence of the monitoring statistic, \( \{E_t, t = 1, 2, \ldots\} \), constitutes a Markov process. The properties of the Markov process for the IC process, such as the expected number of the false signals, \( E(N_F) \), can be derived approximately by the Markov chain approach [see Park (2001, 2007)]. When the process is out of control, however, the observed deviations are not white noises, and the sequence of the monitoring statistics does not constitute a Markov process. In deriving the properties of the IPC procedure for the OC process, the simulation method can be used using the OC process model expressions derived in the previous section.

The time, \( \tau \), that a special cause occurs corresponds to a change point in the process, and its estimation can be regarded as the change point estimation problem. An example of estimating the change point is to use the generalized likelihood ratio test proposed by Apley and Shi (1999), Capizzi (2001), Lee and Park (2007), Pignatiello and Samuel (2001), and Runger and Testik (2003).

5. Rectifying action or readjustment

When the control chart signals after the occurrence of a special cause, the special cause will be detected and eliminated from the process by a rectifying action. Elimination of the special cause from the OC process implies that the process is going back to the IC process. Thus the continuation of the process after the rectifying action corresponds to the start of a new cycle.
On the other hand, there may be cases where the special cause cannot be detected or cannot be eliminated from the process although it is detected. The case that the special cause cannot be detected may arise when the out of control action plan (OCAP) [see Montgomery (2005)] is not provided properly. The case where the special cause cannot be eliminated can be regarded as the occurrence of another noise, and thus all the changes in the process should be thought as the effects of the new noise rather than the special cause.

Whatever the reason for not being able to eliminate the special cause is, an alternative action is to readjust the process with appropriately modified adjustment scheme. For this readjustment, we develop another new process model, \( Z_\tau^\prime = Z_\tau^1 + A_{\tau}^1, (k = 1, 2, \ldots) \), which denotes the process level adjusted up to time \( \tau + L_1 \). The expression for \( Z_\tau^\prime \) can be developed as the following.

By Eq. (7), the random part of \( Z_\tau^{1+k} \) is expressed as

\[
P_{Z_\tau^\prime}(L_1 + k) = \frac{1 - \theta_1 B}{1 - B} e_{L_1 + k}^{\prime} = \frac{1 - \theta_1 B - \lambda_1 B^{k+1}}{1 - B} e_{L_1 + k}^{\prime} + \frac{\lambda_1 B^{k+1}}{1 - B} e_{L_1}^{\prime} + \frac{\lambda_1 B}{1 - B} e_{L_1 + k}^{\prime} + \frac{\lambda_1 B}{1 - B} e_{L_1},
\]

where \( e_{L_1 + k}^{\prime} = e_{L_1 + k}^{\prime} \) if \( k \geq 0 \) and 0 if \( k < 0 \). Then

\[
Z_\tau^\prime = P_{Z_\tau^\prime}(L_1 + k) + M_{Z_\tau^\prime}(L_1 + k) + P_{Z_\tau^\prime}(L_1) + M_{Z_\tau^\prime}(L_1)
= \frac{1 - \theta_1 B}{1 - B} e_k^{\prime} + \left( \frac{\lambda_1 B}{1 - B} \right) e_{L_1}^{\prime} + M_{Z_\tau^\prime}(L_1 + k) + M_{Z_\tau^\prime}(L_1)
= \frac{1 - \theta_1 B}{1 - B} e_k^{\prime} + \frac{\lambda_1 B}{1 - \theta_1 B} e_{L_1}^{\prime} + M_{Z_\tau^\prime}(k)
= \frac{1 - \theta_1 B}{1 - B} e_k^{\prime} + s_{Z_\tau^\prime} + M_{Z_\tau^\prime}(k),
\]

where \( s_{Z_\tau^\prime} = \frac{\lambda_1 B}{1 - \lambda_2 B} e_{L_1}^{\prime} \) and \( M_{Z_\tau^\prime}(k) = M_{Z_\tau^\prime}(L_1 + k) + M_{Z_\tau^\prime}(L_1) \). In Eq. (15), \( s_{Z_\tau^\prime} \) can be regarded as the starting value of the new process model, \( Z_\tau^\prime \). Since Eq. (15) can be rewritten as

\[
Z_\tau^\prime - s_{Z_\tau^\prime} - M_{Z_\tau^\prime}(k) = e_k^{\prime} + \frac{\lambda_1 B}{1 - \theta_1 B} \{ Z_\tau^\prime - s_{Z_\tau^\prime} - M_{Z_\tau^\prime}(k) \},
\]

the total adjustment of the MMSE controller is determined as

\[
A_k^{\prime} = -\frac{\lambda_1 B}{1 - \theta_1 B} Z_\tau^\prime - \frac{1 - B}{1 - \theta_1 B} \{ s_{Z_\tau^\prime} + M_{Z_\tau^\prime}(k) \}
= -\frac{\lambda_1 B}{1 - \theta_1 B} e_k^{\prime} - s_{Z_\tau^\prime} - M_{Z_\tau^\prime}(k),
\]

and the one time adjustment is

\[
a_k^{\prime} = A_k^{\prime} - A_{k-1}^{\prime}
= -\lambda_1 e_k^{\prime} - (1 - B) M_{Z_\tau^\prime}(k).
\]

The observed deviation of the adjustment by the MMSE controller in Eq. (16) is

\[
e_k^{\prime} = Z_\tau^\prime + A_k^{\prime}
= e_k^{\prime}.
\]

Thus the total adjustment in Eq. (16) and the one time adjustment can be expressed as

\[
A_k^{\prime} = -\frac{\lambda_1 B}{1 - B} e_k^{\prime} - s_{Z_\tau^\prime} - M_{Z_\tau^\prime}(k),
\]

and

\[
a_k^{\prime} = -\lambda_1 e_k^{\prime} - (1 - B) M_{Z_\tau^\prime}(k).
\]

Note that when considering the change in the attempted adjustment, \( \tilde{\lambda} \), the total attempted adjustment should be \( A_k^{\prime} / \tilde{\lambda} \), and the one time attempted adjustment should be \( a_k^{\prime} / \tilde{\lambda} \) to produce the MMSE controller the same as in Eq. (16).

From Eq. (17) we see that the readjustment of the process can cure all the changes in the parameters except the change in variance of the random shock. Since we cannot observe \( M_{Z_\tau^\prime}(k) \) in Eq. (19), the result of the adjustment is valid only when the mean part \( M_{Z_\tau^\prime}(k) \) is estimated accurately.

The linear filter expressions such as the process model, the total and one time adjustments, and the observed deviations, are summarized in Table 2.
Table 2
Summary of linear filter expressions for the process model $s.d.(e_t) = \sigma_e$, $s.d.(e_t') = s.d.(e_t'') = \sigma_e$, $\lambda = 1 - \theta$, $\lambda_1 = 1 - \theta_1$, $\lambda_2 = 1 - \lambda', \lambda_3 = \lambda_1 - \lambda'$, $s_x' = \lambda_2B/(1 - \lambda_2B)e_{k-1}'$, $M_x(k) = M_x(L_1 + k) + M_y(L_1)$.

<table>
<thead>
<tr>
<th>IC model</th>
<th>OC model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0 &lt; t \leq \tau$</td>
<td>$t = \tau + k$, $k &gt; 0$ (before signal)</td>
</tr>
<tr>
<td>$Z_t = \frac{1 - \theta}{1 - \theta_1}e_t$</td>
<td>$Z''_t = \frac{1 - \theta}{1 - \theta_1}e''_t + M_x(k)$</td>
</tr>
<tr>
<td>$A_t = -\frac{1}{1 - \theta^2}\lambda_{t-1}$</td>
<td>$A''<em>t = -\frac{1}{1 - \theta^2}\lambda</em>{t-1} - s_x - M_x(k)$</td>
</tr>
<tr>
<td>$a_t = -\lambda_2e_{t-1}$</td>
<td>$a''<em>t = -\lambda_1e</em>{t-1} - (1 - B)M_x(k)$</td>
</tr>
<tr>
<td>$e_t = e_t$</td>
<td>$e''_t = e''_t$</td>
</tr>
</tbody>
</table>

6. Interested ranges of the parameters in linear filter models

The interested ranges of the parameters appearing in the various linear filter models can be considered for practical use as follows.

The MMSE forecast of $Z_t$ in the IC process is $-A_t$, and from Eq. (1) we see that

$$-A_t = \lambda Z_{t-1} + \theta(-A_{t-1}).$$

The MMSE forecast in the above equation is an exponentially weighted moving average (EWMA) of the actual observation $Z_{t-1}$ with weight $\lambda$, when $0 < \lambda \leq 1$. The EWMA of the actual observations taken from a time series model has shown to be a successful forecast of the next observation [see Box, Jenkins, and Reinsel (1994)]. For the reason to make the MMSE forecast correspond to an EWMA, the interested range for $\theta$ may be considered as $0 \leq \theta < 1$, $0 < \lambda \leq 1$ for practical use [see Box and Luceno (1997)].

Let the MMSE controller of the random part in the OC process be $P_y(k)$, although it is not used actually, when the OC process model is assumed known. Then the MMSE forecast, $-P_y(k)$, is expressed as

$$-P_y(k) = \lambda_1P_x(k - 1) + \theta_1(-P_y(k - 1)).$$

The MMSE forecast for the OC process when $0 < \lambda_1 \leq 1$. By the same reason as in the IC process, the interested range for $\lambda_1$ may be considered as $0 < \lambda_1 \leq 1$, $0 \leq \theta_1 < 1$ for practical use.

The actual controller of the random part in the OC process until a signal is given is $P_H(k)$, and the forecast of $P_z(k)$ for this period is $-P_H(k)$. From Eq. (11) we see that

$$-P_H(k) = \lambda'P_x(k) + \lambda_2(-P_H(k - 1)).$$

The forecast used for the OC process is an EWMA when $0 < \lambda' \leq 1$. By the same reason as in the IC process, the interested range for $\lambda'$ may be considered as $0 < \lambda' \leq 1$, $0 \leq \lambda_2 < 1$ for practical use.

The correlation of the two consecutive process levels in the IC process is computed as

$$\text{Corr}(Z_t, Z_{t-1}) = \frac{\lambda + (t - 2)\lambda^2}{\sqrt{1 + (t - 1)\lambda^2}\sqrt{1 + (t - 2)\lambda^2}}.$$  

Similarly, the correlation of the two consecutive process levels in the OC process is computed as

$$\text{Corr}(Z'_t, Z'_{t-1}) = \frac{\lambda_1 + (k - 2)\lambda_1^2}{\sqrt{1 + (k - 1)\lambda_1^2}\sqrt{1 + (k - 2)\lambda_1^2}}.$$  

From the above two equations, we can see that the noises in both IC and OC processes, where the parameters, $\lambda$ and $\lambda'$, are selected from $0 < \lambda \leq 1$ and $0 < \lambda_1 \leq 1$, respectively, will have positive serial correlations when left by themselves. A positive correlation between the two consecutive observations is more realistic than a negative one in real production environments.

The next thing to consider is that the change in the attempted adjustment due to a special cause, $\tilde{\lambda}$, may not reverse the sign of the adjustment, that is $\tilde{\lambda} > 0$. Also the change in the adjustment may either increase or decrease the attempted adjustment. Thus the interested range for $\tilde{\lambda}$ can be considered as either $0 < \tilde{\lambda} < 1$ or $\tilde{\lambda} > 1$. It seems that the case of $0 < \tilde{\lambda} < 1$ is more likely to occur than the case of $\tilde{\lambda} > 1$, since the actual amount of adjustment tends to be less than the attempted amount of adjustment when the adjustment scheme is out of order or controlled poorly.

When the possible range of $\lambda_2$ is $0 \leq \lambda_2 < 1$ as considered above, we can see from the expressions in Table 1 that the effect of mean change is to make the mean part of the observed deviation, $M_x(k)$, converge to some fixed constant as time, $k$, increases. This implies the repeated adjustment of the OC process until a signal makes the mean part of the observed deviation stable. Note that when $\lambda_2$ is selected from $0 \leq \lambda_2 < 1$, the observed deviation vanishes for the SS and TS cases, but converges to a nonzero fixed level for the SD case.
7. An example

In order to illustrate how the IPC procedure will be considered in manufacturing practice, the poly silicon deposition process during the manufacturing of the very large scale integrated (VLSI) circuits is cited here [see Phadke (1989)]. Manufacturing of the VLSI circuits involves more a hundred major steps. Deposition of poly-silicon comes after about the half of the steps are complete, and, as a result, the silicon wafers used in the process have a significant amount of value added by the time they reach this step. The function of the poly-silicon deposition process is to deposit a uniform layer of a specified thickness, for example, 3600 Å. The poly-silicon deposition rate, $X_t$, works as the compensatory variable which can be manipulated to adjust the deviation of the deposition thickness, $Z_t$. It is often the case that the effect of the compensation will fully appear in one unit delayed time in the manufacturing processes. The locally linear model, $EZ_t = \beta X_{t-1}$ is employed to relate the deposition rate to the mean thickness from target. When the deposition time is fixed. The slope $\beta$ is called the process gain and a well-calibrated machine will be close to 1. Then the deviation of the deposition thickness from target can be expressed as

$$e_t = \beta X_{t-1} + Z_t.$$  

For the purpose of an example, we assume that the true parameter values are known as $\beta = 1$, $\theta = 0.7$ and $\sigma^2 = 750$ Å. We also assume that the effect of the special cause is to produce a SS with $\tilde{\mu} = 3.0$, $\tilde{\sigma} = 2$, $\tilde{\theta} = \tilde{\lambda} = 1$.

The MMSE controller is expressed as

$$X_t = (1 - 0.7) \sum_{j=1}^{t} e_j.$$  

Fig. 1 illustrates the behavior of the two deviations, unadjusted and adjusted, where the special cause occurs before time 71. The three sigma limits are drawn for reference. The solid line represents the unadjusted deviation from target, $Z_t$, and the dotted line the adjusted deviation, $e_t$. Up to time 70, we see that the adjusted deviation behaves as a series of random shocks while the unadjusted deviation wanders off the target. If the Shewhart chart with 3-sigma limit is implemented using the unadjusted deviation in this case, there should be many false signals because ten points exceeds the lower control limit. However all the adjusted deviations are well inside the limits. After time 70, we see that the adjusted deviations are much closer to the target compared to the unadjusted deviations, especially as the time passes. Both the adjusted and unadjusted deviations will be quickly detected by the Shewhart chart with a 3-sigma limit. Even in the adjusted case, the deviations are widely spread about the target due to the increased variation, thus it is crucial to detect the occurrence of the special cause quickly.

This example shows that the implementation of IPC in the VLSI manufacturing industry can improve the quality of the wafers by achieving a uniform thickness. The expression of the OC model makes it possible to study the properties of the IPC procedure in determining the type of monitoring scheme and its effects.

8. Summary and discussions

The statistical models for the process level, the process adjustment, and the observed deviation have been expressed as linear filter models when noise is inherent in the process and special causes have random chances to occur. The derivation of such models has been greatly simplified by use of the linear filter model expressions.
The properties of the IPC procedures can be derived by either analytic methods or the simulation methods based on the linear filter model expressions derived here. Also all the possible changes of the parameters in the process model as well as the three types of mean change have been considered in the development of the OC process model. Thus the models derived here will be particularly useful in deriving the properties of the IPC procedures when the process is out of control.

The readjustment of the OC process after having a true OC signal is considered when the special cause cannot be detected or when it is not practically possible to eliminate the special cause. The use of the readjustment may also be considered when the elimination of the special cause costs more than the future expected cost. The readjustment may also be considered when the elimination of the special cause costs more than the future expected cost. Future expected cost implies the total cost expected to incur until the end of the production process when the readjustment is made. In many production processes, a production line is planned to produce a certain limited number of items and it will be reset to produce another item afterwards. In such cases the time that the production line is reset after producing the required number of items is considered as the end of the production length. The readjustment will be particularly considerable when the remaining production length is relatively short. Based on the assumption that no special cause will occur again until the end of the production length, a decision whether to rectify or to readjust the process, even when the special cause can be removed, can be made by considering the future expected cost as follows.

Let \( C_R \) be the cost for a rectifying action to eliminate the special cause and \( L_2 \) be the remaining production length at the time of signal, that is \( t = \tau + L_1 \). The cost for the rectifying action until the end of production length is \( C_R + L_2 \cdot (C_A + C_T \sigma^2) \) since the resulting observed deviations after the rectifying action will be white noises with variance \( \sigma^2 \) by Eq. (3). On the other hand, the cost for the readjustment until the end of the production length is \( L_2 \cdot (C_A + C_T \cdot \tilde{\sigma}^2 \sigma^2) \) since the resulting observed deviations after the readjustment will be the white noises with variance \( \tilde{\sigma}^2 \sigma^2 \) by Eq. (17). Thus readjustment is a better choice than the rectifying action if

\[
C_R + L_2 \cdot (C_A + C_T \sigma^2) > L_2 \cdot (C_A + C_T \cdot \tilde{\sigma}^2 \sigma^2).
\]

Let us consider the ARMA(1,1) model in representing the noise of the IPC procedure in replacement of the IMA(1,1) model. The stationary time series model like ARMA(1,1) model would continually return by itself to the neighborhood of its mean, whereas the nonstationary model like IMA(1,1) would permanently drift away from the target if not controlled properly. In a finite time period of the process, however, the practical difference between the ARMA(1,1) and IMA(1,1) lies in the way that the controller is obtained rather than in either stationary or nonstationary properties.

Suppose that the noise follows an ARMA(1,1) model for the IC process as

\[
Z_t = \frac{1 - \theta B}{1 - \phi B} \varepsilon_t,
\]

where \( \phi \) denotes the autoregressive parameter and \( \theta \) denotes the moving average parameter. Here we assume that \( Z_t = \varepsilon_t = 0 \) if \( t \leq 0 \). The MMSE controller for the total adjustment can be derived analogously as

\[
A_t = -\frac{\phi - \theta}{1 - \phi B} \varepsilon_{t-1},
\]

and the one time adjustment corresponds to

\[
a_t = -\frac{(\phi - \theta)(1 - B)}{1 - \phi B} \varepsilon_{t-1}.
\]

From the above expression, we see that the one time adjustment uses all the previous observed deviations including the current one. When compared to the controller of the IMA(1,1) model which uses the current observed deviation only, this type of adjustment is complicated and difficult to apply in practice. Thus it does not seem to be appropriate to use the ARMA(1,1) model in representing the noise.

It has been shown in the EPC procedure that the bounded scheme, which is designed to adjust the process when the predicted deviation exceeds certain bounds, is better than the repeated adjustment in the context of the minimum cost control [see Box and Kramer (1992)]. The linear filter model expressions for the bounded adjustment scheme is left for further study.

**Acknowledgement**

This research was supposed by the Chung-Ang University Research Grants in 2008.

**References**


