Relationships Among Control Charts Used with Feedback Control

George Runger1,*,†, Murat Caner Testik2 and Fugee Tsung3
1Department of Industrial Engineering, Arizona State University, Tempe, AZ 85287, U.S.A.
2Department of Industrial Engineering, Hacettepe University, Ankara, Turkey
3Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

Feedback control is common in modern manufacturing processes and there is a need to combine statistical process control in such systems. Typical types of assignable causes are described and fault signatures are calculated. A fault signature can be attenuated by the controller and an implicit confounding among faults of different types is discussed. Furthermore, the relationships between various control statistics are developed. Control charts have been proposed previously for deviations from target and for control adjustments. We describe why one or the other can be effective in some cases, but that neither directly incorporates the magnitude (or signature) of an assignable cause. Various disturbance models and control schemes, both optimal and non-optimal, are included in a mathematically simple model that obtains results through properties of linear filters. We provide analytical results for a widely-used model of feedback control. Copyright © 2006 John Wiley & Sons, Ltd.

Received 20 August 2005; Revised 28 October 2005

KEY WORDS: process control; deviations from target; assignable cause; closed-loop control

1. INTRODUCTION

Feedback (closed-loop) control systems are more common in manufacturing and there is an increased need to integrate them with statistical process control (SPC) methods for process improvement. On the one hand, these control schemes do not remove the sources of variation but attempt to compensate with adjustments to other variables. On the other hand, SPC schemes originated in the parts industry with the intent to eliminate assignable causes that are the sources of variation. Over the past decade, research investigated various aspects of integrating these two control schemes. For control charts, most of these studies focused on monitoring the deviations from target, i.e. the output of the controlled process (Box and Kramer1, Vander Wiel et al.2, Montgomery et al.3, and Box and Lunceño4) or the control actions (Falthin and Tucker5, MacGregor6, Tsung et al.7, Jian and Tsui8, and Tsung and Tsui9). A residual statistic was used by Capilla et al.10 and this type of statistic is discussed later in this work.
The objective of this paper is to describe general characteristics of control charts for processes with feedback control. First, we present a simple model that incorporates assignable causes of different types, and show that such types cannot easily be distinguished. There is an implicit confounding of assignable causes. We also show that for common assignable causes the signature of the fault decays over time. This makes these faults difficult to detect. Second, we develop and explain the relationships among the statistics that have been traditionally used for control charts in these systems. We show that deviations from target or control adjustments are filters of the residual series. Just as an exponentially weighted moving average (EWMA) filter can improve a control chart, these filters can be effective. However, the implicit filters applied by these statistics do not consider the assignable cause. Therefore, it is more effective to design a control chart with the assignable cause in mind. The discussion incorporates non-optimal controllers because they have considerable practical benefits. The autocorrelation in the deviations from target that results in such as case is shown by a model. Also, results are not restricted to a specific type of disturbance model, but apply to a wide class of models. The intent is a unified presentation of important concepts for control charts in the presence of feedback control.

Section 2 develops the feedback model. We present it in a manner conducive to the control charts that are later discussed. Section 3 discusses the confounding and the common decay (over time) of the signatures of assignable causes. Section 4 presents optimal controllers. Section 5 presents the general case for arbitrary controllers and disturbance models. Section 6 provides an illustrative example. Section 7 has conclusions.

2. FEEDBACK MODEL

A simple approach to model a feedback-adjustment system follows. To avoid unimportant constants we assume that the process target is the value 0. Define \( y_t \) to be the output from the process, \( d_t \) to be a disturbance to the process, and \( U_t \) be the cumulative effect on the output of previous adjustments to a manipulated variable, each at time \( t \). As the target is the value 0, \( y_t \) is also the deviation from target at time \( t \). The model is

\[
y_t = d_t + U_t
\]

We use the common backshift-operator notation defined as \( By_t = y_{t-1} \). For each of the terms in (1) some typical properties are assumed. The disturbance is usually modeled by a time series. Box et al.\(^{11}\) and other time-series books often assume a disturbance that can be modeled by a linear filter of white noise. This implies that

\[
d_t = h(B)e_t
\]

for a white-noise series \( e_t \). In particular, Box et al.\(^{11}\) and Box and Luceño\(^{4}\) state the importance of the integrated moving average (IMA) model for \( d_t \) in a physical processes. This is

\[
d_t = d_{t-1} + e_t - \theta e_{t-1}
\]

but for this work we prefer the equivalent description

\[
d_t = \frac{(1 - \theta B)}{1 - B} e_t = h(B)e_t
\]

for the function \( h(B) = (1 - \theta B)/(1 - B) \).

The effect of the controller is modeled through \( U_t \) and it depends on dynamics that relate the change of a manipulated variable to its effect on the output \( y_t \). The setting of the manipulated variable, denoted by \( X_t \), is a function, say \( c(B) \), of the measured outputs for a feedback system. Usually, this function is also linear in the \( y \). A common choice is integral control with

\[
X_{t-1} = -\lambda \sum_{i=1}^{t-1} y_i = -\lambda \frac{y_{t-1}}{1 - B} = c(B)y_{t-1}
\]
for an appropriately selected constant $\lambda$. If we assume a zero-order dynamics so that the full effect of a change at time $t - 1$ occurs at time $t$ then $U_t = X_{t-1}$. Note that we write $U_t = X_{t-1}$ instead of $U_t = gX_t$ for some coefficient $g$ representing system gain. The results can be easily modified from our assumed case that $g = 1$.

First-order (or higher) dynamics might also be present and a common model is

$$U_t = \delta U_{t-1} + (1 - \delta)X_{t-1}$$

where $\delta$ denotes the measure of inertia in the process dynamics and $1 - \delta$ is interpreted as the proportion of the eventual output change that occurs in one time interval after a step change has been made in the setting $X$. Zero-order dynamics is the special case $\delta = 0$ that corresponds to no inertia. Box et al.\(^\text{11}\) comment that (6) might be the most useful model for process dynamics. Equation (6) can be written as

$$U_t = \frac{(1 - \delta)}{1 - B} X_{t-1} = f(B)X_{t-1}$$

In general the process dynamics are often a linear filter such as $f(B)$ in (7) and the controller setting $X_{t-1}$ is a linear filter of $y_{t-1}$ such as (5). The net result is the composition of two linear filters so that the effect of control on the process can be written as

$$U_t = f(B)c(B)y_{t-1} = g(B)y_{t-1}$$

for a function $g(B)$ of the backshift operator. We assume that there is no pure delay in the adjustments so that (8) describes the adjustment. Therefore, we rewrite (1) to obtain the fundamental equation

$$y_t = h(B)e_t + g(B)y_{t-1}$$

In one more step we write (9) as

$$y_t - g(B)y_{t-1} = h(B)e_t$$

$$\phi(B)y_t = h(B)e_t$$

where $\phi(B)$ is defined implicitly in the equation. In the following, (9) and (10) are used to make our main points.

For example, a common case is when $d_t$ is IMA, there are zero-order dynamics, and minimum mean square error (MMSE) control is applied. See Box et al.\(^\text{11}\). In this case, the MMSE controller is integral control so that (9) is

$$y_t = \frac{(1 - \theta B)}{1 - B} e_t + \frac{\lambda B}{1 - B} y_t$$

where $\lambda = 1 - \theta$ and it is well known (or a little algebra can be used) to show that

$$y_t = e_t$$

Also, the adjustments to the manipulated variable, i.e. $x_t = X_t - X_{t-1} = (1 - B)X_t$, are

$$x_t = -\lambda y_t$$

Consequently, if an SPC monitor is used for this process, a control chart for $x_t$ or $y_t$ has the same performance. Furthermore, from (12) and (13) each of these quantities are white-noise series so that the usual control chart assumptions apply. In this example, most control chart strategies focus on the equivalent statistics $x_t$ or $y_t$. 

Copyright © 2006 John Wiley & Sons, Ltd.  
DOI: 10.1002/qre
3. ASSIGNED CAUSE CONFOUNDING

A control chart is used to detect assignable causes. Consequently, an effective strategy should consider the signal to be detected. Assignable causes may be put into three categories: actuator faults, input faults, and sensor faults. Here, an actuator fault implies a problem in control instruments, an input fault implies a change in the disturbance level, and a sensor fault implies a problem with measurement instruments. A similar characterization was used by Gertler\(^\text{12}\) and Capilla et al.\(^\text{10}\). For each category we assume a fault pattern (or signature) \(\delta_t\) that affects the model in (9) or (10). Here, \(\delta_t\) is assumed to be zero until the time of the fault and then non-zero (either positive or negative) for several (or all) time periods afterwards.

An input fault can be modeled as a change from \(d_t\) to \(d_t + \delta_t\) in (4). For example, Vander Wiel\(^\text{13}\) added an assignable cause in an IMA disturbance in the same manner. This assignable cause adds the term \(\delta_t\) to the right-hand side of the model in (10).

Similarly, a sensor that measures the output under an assignable cause changes \(y_t\) to faulted readings denoted as \(y^f_t\) where \(y^f_t = y_t + \delta_t\). Only the series \(y^f_t\) is observed. Therefore, the controller term on the right-hand side of (9) uses \(y^f_{t-1}\). Furthermore, \(y_t = y^f_t - \delta_t\) can be substituted into the left-hand side of (9) to obtain the equation for the faulted (measured) readings to be

\[
y^f_t - \delta_t = h(B) e_t + g(B) y^f_{t-1}
\]

Because \(y^f_t\) is the measured output from a sensor fault, this series can be relabeled simply as \(y_t\) in (14). Then (14) is similar to (9) except the term \(-\delta_t\) is added to the left-hand side.

An actuator fault can be modeled as a change from \(X_t\) to \(X_t + \delta_t\) or \(x_t\) to \(x_t + \delta_t\), the former case being an error in the setting of actuator and the latter being an error in control adjustments. Note that a change of \(x_t\) to \(x_t + \delta_t\) is equivalent to a change of \(X_t\) to \(X_t + \delta_t/(1 - B)\) so that an adjustment assignable cause can be represented as a setting assignable cause with a modification of the fault signature. Consequently, using (7), the term \(f(B) \delta_{t-1}\) is added to the right-hand side of (10) to represent actuator faults.

The result of any of these assignable causes is a term added to the right-hand side of the model in (10). That is,

\[
\phi(B) y_t = h(B) e_t + \Delta_t
\]

In general the effect of a fault on (15) is written as \(\Delta_t\) where it is implied that \(\Delta_t\) is either \(\delta_t\), \(\delta_t\), or \(f(B) \delta_{t-1}\), for an input, sensor, or actuator fault, respectively. Our point here is the confounding among these faults. A sensor fault has exactly the same effect as an input fault. Furthermore, an actuator fault for zero-order process dynamics is clearly confounded with the other two types of faults but for higher-order dynamics the effect is a transformed fault sequence. A signal from a control scheme applied to the model in (10) cannot distinguish the cause of the fault without additional information that is not included in the model.

4. OPTIMAL CONTROLLER

In general, the controller for (8) that provides MMSE control can be determined algebraically as the function \(g(B)\), which reduces (9) to \(y_t = e_t\) assuming that the process dead-time is zero (if the dead-time is, say \(b\), then deviations from target follows a \(b\) order moving average process). This is

\[
g(B) = B^{-1} [1 - h(B)]
\]

but the details are not important. The main point here is that for MMSE control \(y_t = e_t\) and this implies that a control chart for the deviations from target \(y_t\) meets the usual assumptions when the process has no assignable causes. If the MMSE controller is used for \(g(B)\) then (15) becomes

\[
y_t = e_t + h^{-1}(B) \Delta_t
\]
The mean for the series \( y_t \) when an assignable cause is present is modified by the disturbance term \( h^{-1}(B) \Delta_t \).

For an input fault \( \Delta_t = \delta_t \) the model has only \( h^{-1}(B) \delta_t \) added to it. For a sensor fault the mean series of \( y_t \) is again \( h^{-1}(B) \delta_t \). For an actuator (setting) fault the mean series is \( h^{-1}(B) f(B) \delta_{t-1} \). Therefore, even for an optimal controller the signature in the deviations from target can decay so that it is difficult to detect. An example is provided in a following section. In general, faults with signature \( \delta_t \) are modified.

5. Arbitrary Controllers

The general case of arbitrary disturbance and arbitrary controller is considered in this section. An MMSE controller has the disadvantage that it generates excessive variability in the controlled variable. Therefore, alternatives are frequently used in practice, even if the IMA model for disturbances is reasonable. For example, Box et al.\(^\text{11}\) described a constrained controller

\[
X_{t-1} = -k_P y_{t-1} - k_I \frac{y_{t-1}}{1 - B} = -k_P y_{t-1} - k_I \sum_{i=1}^{t-1} y_i
\]

where \( k_P \) and \( k_I \) are appropriately chosen constants. This controller can decrease the variability in the controlled variable to a much lower level than MMSE control. This is called a proportional-integral (PI) controller because the control setting is proportional to the deviation and the sum (discrete integral) of the previous deviations from target. Also, one might want to move from the IMA disturbance model to some other model for the disturbances. If so, how does a control chart monitor perform for cases such as these?

The recommended control statistic for such a model is the deviation (residual) from the predicted output. This is shown in Basseville and Nikiforov\(^\text{14}\) and can easily be derived from a likelihood approach. We derive it simply from (15). Solve for \( e_t \) to obtain

\[
e_t = \frac{\phi(B)}{h(B)} y_t - \frac{\Delta_t}{h(B)}
\]

The first term on the right-hand side of (19) is defined to be the residual series \( r_t \) from the feedback system. That is,

\[
r_t = \frac{\phi(B)}{h(B)} y_t
\]

When no fault is present \( r_t = e_t \) so that \( r_t \) is a white-noise series and it meets the usual control chart assumptions. When a fault is present,

\[
r_t = e_t + \frac{\Delta_t}{h(B)}
\]

so that the mean of \( r_t \) is changed to \( h^{-1}(B) \Delta_t \) and this provides the effect of a fault on a control chart for \( r_t \). Note that neither \( \Delta_t \) nor \( h(B) \) depends on the controller. Consequently, the distribution of \( r_t \) is the same as the deviations from target under MMSE control as in (17) when the process dead-time is zero. Given a process model and a controller, \( r_t \) can be computed. Therefore, it is not necessary to analyze the performance of control charts under various control strategies. The statistical properties of the control statistic \( r_t \) is invariant to type of controller used.

Suppose that a control chart is based on \( y_t \) in (10). Solve for \( y_t \) in (20) to obtain

\[
y_t = \frac{h(B)}{\phi(B)} r_t
\]
Therefore, the deviation series $y_t$ is a filter of the residual series $r_t$. Filters can improve the performance of control charts. For example, an EWMA control chart is a filter denoted as $z_t$ of inputs $e_t$

$$z_t = \lambda e_t + (1 - \lambda)z_{t-1}$$

$$z_t = \frac{\lambda}{1 - (1 - \lambda)B}e_t$$

(23)

The EWMA filter when applied to white noise should use a smaller value of $\lambda$ to detect a smaller shift in the mean of $e_t$ and a larger value of $\lambda$ to detect a larger shift in $e_t$. That is, the filter needs to be designed for the type of assignable cause that is to be detected. Or, at least, some recommended, general design guidelines need to be used. However, $y_t$ in (22) does not depend on the assignable cause. A filter, and therefore, a control chart design is imposed once the deviations from target are selected. Although the design can be good in some cases, a better approach is to base a control chart on $r_t$ and choose a filter, as needed, to detect a specific magnitude of assignable cause. Furthermore, the $r_t$ series is white noise so that control chart properties are well understood.

The terms in the $y_t$ series are autocorrelated. Substitute for $r_t$ from (21) into (22) to obtain the distribution of $y_t$

$$y_t = \frac{h(B)}{\phi(B)} e_t + \frac{\Delta_t}{\phi(B)}$$

(24)

If the $y_t$ series is whitened, this is a simple inverse of the filter in (22). The result is a full-circle return to a control chart based on $r_t$.

Similar conclusions apply to a control chart for control adjustments. The control adjustments comprise the series $x_{t-1} = (1 - B)X_{t-1}$. Therefore,

$$x_{t-1} = (1 - B)X_{t-1} = (1 - B)c(B)y_{t-1}$$

(25)

Solve for $y_t$ in (22) and substitute into (25) to obtain

$$x_{t-1} = \frac{(1 - B)c(B)h(B)}{1 - g(B)B} r_{t-1}$$

(26)

Therefore, similar to the case for deviations, the adjustment series $x_{t-1}$ is a filter of the residual series $r_t$. As before, the benefits from this filter depend on its form relative to the assignable cause to be detected. The distribution of the adjustment series can be obtained from a substitute for $r_t$ from (21) into (26).

6. **ILLUSTRATIVE EXAMPLE**

Suppose that the constrained controller in (18), i.e. $X_{t-1} = (-k_P - k_I (1 - B)^{-1})y_{t-1}$ is applied to the IMA disturbance model in (4). The deviations from target relate to the residuals as

$$y_t = \frac{(1 - \theta B)(1 - \delta B)}{1 + c_1 B + c_2 B^2} e_t + \frac{(1 - B)(1 - \delta B)}{1 + c_1 B + c_2 B^2} \Delta_t$$

(27)

where $c_1 = -1 - \delta + (1 - \delta)(k_P + k_I)$ and $c_2 = \delta - k_P(1 - \delta)$. The series $y_t$ can have a complex autocorrelation structure and the typical control chart assumption of white noise is not, in general, valid. In (27) the general model for $y_t$ is ARMA(2, 2) (Box et al.\textsuperscript{11}). However, $y_t$ is a filter of $r_t$ and as such it might provide appropriate smoothing in some cases. That is, a control chart based on $y_t$ might provide better performance in some cases than one naively based on $r_t$ (see Runger\textsuperscript{15} for details). Still, a better solution is to start with $r_t$ and incorporate the filtering necessary to detect a particular type of assignable cause. As $r_t$ is a white-noise sequence if no assignable causes are present, traditional rules for the design of filters for specific assignable causes can be applied. EWMA filters can be used or more specific generalized likelihood ratio (GLR) methods can target particular fault signatures.
Figure 1. Input (sensor) fault signatures in the deviations from target for three different shift magnitudes $\delta_t = 1, 2, \text{and } 3$

Entirely analogous comments apply to a control chart based on $x_t$. Here

$$x_{t-1} = -[k_P(1 - B) + k_I] \left( \frac{(1 - \theta B)(1 - \delta B)}{1 + c_1 B + c_2 B^2} e_{t-1} + \frac{(1 - B)(1 - \delta B)}{1 + c_1 B + c_2 B^2} \Delta_{t-1} \right)$$  \hspace{1cm} (28)

and again there is much complexity (autocorrelation), in general, in this series.

Suppose that for the IMA disturbance model $\theta = 0.6$, variance of $e_t$ is 1, i.e. $\text{Var}(e_t) = 1$, and for the dynamics $\delta = 0.5$. The MMSE control yields an output variance $\text{Var}(y_t) = 1$ with $\text{Var}(X_t) = 5$ (Box et al.\textsuperscript{11}). Note that MMSE control sometimes requires unacceptably large manipulations of the compensating variable. Therefore, a constrained control scheme, that trades some increase in the output variance for a reduction in the controller variance, may be useful. Box et al.\textsuperscript{11} (p. 507) showed that a constrained PI control of the form

$$U_{t-1} = 0.13e_{t-1} - 0.52 \frac{1}{1 - B} e_{t-1}$$

would make it possible to yield a 20-fold reduction in the variance of the compensating variable. Figures 1 and 2 illustrate the fault signatures for the two different types of input faults in the mean of deviations from target obtained from (17) and (24): a step-change and a trend on $d_t$. For each control strategy three different shift magnitudes of 1, 2, and 3 units (slopes for a trend) are induced at $t_0 = 10$. The values of $k_P$ and $k_I$ used for the plots are $-0.13$ and $0.52$, respectively.

Consider the curves denoted by MMSE control in Figures 1 and 2. Owing to the confounding, these curves apply to either input or sensor faults. The means of the residual sequence are the same as the means of deviations from target for MMSE control. Consequently, regardless of whether MMSE control is used, the fault signatures for the residuals equal the MMSE control curves in Figures 1 and 2. Note how the step fault signature in Figure 1 is attenuated over time and becomes more difficult to detect. Only the trend fault in Figure 2 has a persistent signature.
Figure 2. Input (sensor) fault signatures in the deviations from target for three different shift slopes of $\delta_t = 1, 2, \text{ and } 3$

The PI curves in Figures 1 and 2 show the fault signatures for the deviations from target for a (non-optimal) PI controller. The fault signatures are less attenuated. However, when the control is not MMSE the deviations from target are not uncorrelated. Owing to correlation in the deviations from target, one cannot conclude that the PI fault signatures are more easily detected than the MMSE signatures (see Runger\textsuperscript{15} for a discussion). As we mentioned previously, if the deviations from target are whitened (through charts of residuals from the appropriate time series model), then the fault signature of the whitened data equals the fault signature for MMSE control when the process dead-time is zero.

Similarly, fault signatures in the mean of deviations from target are shown for the actuator faults in Figures 3 and 4 for impulse and step changes in the adjustments, respectively. Note that these changes in the adjustments imply a step and a trend change, respectively, in the controller setting.

7. CONCLUSIONS

Processes with feedback control change the practice of SPC and several important characteristics were summarized and explained in this work. There is an implicit confounding between input, sensor, and actuator faults. The fault signature commonly decays and becomes more difficult to detect over time. The mean of the residual sequence (fault signature) was determined in a straightforward manner. A series commonly used for control charts in the references cited, the deviations from target was shown to be a filter of the residuals. By analogy with autocorrelation induced in a conventional EWMA control chart, we explained why a control chart based on the deviations from target can be effective. However, we argued that a disadvantage is that the filter implicitly generated by the deviations does not depend on the fault signature to be detected. Whitening of the deviations from target produces a control chart for residuals.

The residual sequence, from a process with dead time zero, equals the deviations from target that would be obtained from a MMSE controller, even though one might not be used in a system. Consequently, a control chart
Figure 3. Actuator fault signatures in the deviations from target for an impulse in the adjustments (step in the settings) for three different shift magnitudes $\delta_t = 1, 2, \text{and} 3$

Figure 4. Actuator fault signatures in the deviations from target for a step change in the adjustments (trend in the settings) for three different shift slopes $\delta_t = 1, 2, \text{and} 3$
for residuals is identical to a control chart of deviations from target in a MMSE controller and the performance for one can be obtained from the performance of the other.

We extended the conclusions for the deviations from target to charts for controller settings and controller adjustments. Both were presented as further filters of the residuals. Some analytical formulas for these properties of a feedback system were provided for the special case of IMA disturbances, a PI controller, and first-order dynamics. This is a widely used model for feedback adjustment. We also provided some illustrative numerical calculations.

Acknowledgements
This material is based on work supported by the National Science Foundation under Grant No. 0085041. This research was also supported in part by RGC Competitive Earmarked Research Grants HKUST6232/04E and HKUST6204/05E.

REFERENCES

Authors’ biographies
George C. Runger is a Professor in the department of Industrial Engineering at Arizona State University. His research is on real-time monitoring and control, data mining, and other data-analysis methods with a focus on large, complex, multivariate data streams. His work is funded by grants from the National Science Foundation and corporations. In addition to academic work, he was a senior engineer at IBM. He holds degrees in industrial engineering and statistics.
Murat Caner Testik is an Associate Professor of Industrial Engineering at Hacettepe University, Ankara, Turkey. He has a PhD in Industrial Engineering from Arizona State University. His research interests are statistical process control and data mining for process improvement.

Fugee Tsung is an Associate Professor in the Department of Industrial Engineering and Logistics Management at the Hong Kong University of Science and Technology. He received both his MS and PhD in Industrial and Operations Engineering from the University of Michigan, Ann Arbor. He is an Associate Editor of *Technometrics*, a Department Editor of the *IIE Transactions*, and on the Editorial Boards for *Quality and Reliability Engineering International*, the *International Journal of Reliability, Quality and Safety Engineering* and the *International Journal of Six Sigma and Competitive Advantage*. He is an ASQ Certified Six Sigma Black Belt, ASQ authorized Six Sigma Master Black Belt Trainer, and former Chair of the Quality, Statistics, and Reliability (QSR) Section at INFORMS. He is also the winner of the Best Paper Award for the *IIE Transactions* focus issue on Quality and Reliability in 2003. His research interests include quality engineering and management, statistical process control, monitoring and diagnosis.