ECONOMIC STATISTICAL DESIGN OF VARIABLE SAMPLING INTERVAL $\bar{X}$ CONTROL CHART BASED ON SURROGATE VARIABLE USING GENETIC ALGORITHMS

Tae-Hoon Lee$^1$, Sung-Hoon Hong$^2$, Hyuck-Moo Kwon$^3$, Minkoo Lee$^4$

$^1$ Korea Atomic Energy Research Institute, VHTR Technology Development Division, Korea
$^2$ Chonbuk National University, Department of Industrial & Information Systems Engineering, Korea
$^3$ Pukyong National University, Department of Systems and Management Engineering, Korea
$^4$ Chungnam National University, Department of Information and Statistics, Korea

Corresponding author:
Minkoo Lee
Chungnam National University
Department of Information and Statistics
Daehak-ro 99, Yusong-gu, Daejeon, 34134, Korea
phone: (+82)-42-8215409
e-mail: sixsigma@cnu.ac.kr

Received: 1 July 2016
Accepted: 28 July 2016

Abstract
In many cases, a $\bar{X}$ control chart based on a performance variable is used in industrial fields. Typically, the control chart monitors the measurements of a performance variable itself. However, if the performance variable is too costly or impossible to measure, and a less expensive surrogate variable is available, the process may be more efficiently controlled using surrogate variables. In this paper, we present a model for the economic statistical design of a VSI (Variable Sampling Interval) $\bar{X}$ control chart using a surrogate variable that is linearly correlated with the performance variable. We derive the total average profit model from an economic viewpoint and apply the model to a Very High Temperature Reactor (VHTR) nuclear fuel measurement system and derive the optimal result using genetic algorithms.

Compared with the control chart based on a performance variable, the proposed model gives a larger expected net income per unit of time in the long-run if the correlation between the performance variable and the surrogate variable is relatively high. The proposed model was confined to the sample mean control chart under the assumption that a single assignable cause occurs according to the Poisson process. However, the model may also be extended to other types of control charts using a single or multiple assignable cause assumptions such as VSS (Variable Sample Size) $\bar{X}$ control chart, EWMA, CUSUM charts and so on.

Keywords
economic design, surrogate variable, variable sampling interval, TRISO Fuel, genetic algorithms.

Notations

$X$ – surrogate variable,
$Y$ – performance variable,
$n_x$ – sample size for surrogate variable,
$n_y$ – sample size for performance variable,
$\mu_x$ – mean of surrogate variable,
$\mu_{x,0}$ – mean of surrogate variable when the process is in control,
$\mu_{x,1}$ – mean of surrogate variable when the process is out of control,
$\mu_y$ – mean of performance variable,
$\mu_{y,0}$ – mean of performance variable when the process is in control,
$\mu_{y,1}$ – mean of performance variable when the process is out of control,
$\sigma_x$ – standard deviation of surrogate variable,
$\sigma_y$ – standard deviation of performance variable,
c – magnitude of the shift in the process mean measured in $\sigma_y$ unit,
\[ \lambda \quad \text{number of the occurrence of an assignable cause,} \\
\rho \quad \text{correlation coefficient between a surrogate variable and a performance variable,} \\
m \quad \text{number of different sampling interval length, } m \geq 2, \\
h_j \quad j\text{-th smallest sampling interval lengths, } j = 1, 2, \ldots, m, \\
k_j \quad j\text{-th threshold limit, } 0 \leq k_m \leq k_j \leq \ldots \leq k_1, \\
I_j \quad j\text{-th sampling interval region using } h_j, \\
IC \quad \text{in control time,} \\
OC \quad \text{out of control time,} \\
FA \quad \text{time to owing to false alarm,} \\
AC \quad \text{time to owing to assignable cause,} \\
SN_0 \quad \text{expected number of samples in in-control period,} \\
SN_1 \quad \text{expected number of samples in out-of-control period,} \\
q_s^{(0)} \quad \text{probability that } X \text{ fall outside the 1st control limits when } \mu_y = \mu_{y,0}, \\
q_s^{(1)} \quad \text{probability that } X \text{ fall outside the 1st control limits when } \mu_y = \mu_{y,0} + c\sigma_y, \\
q_j^{(0)} \quad \text{probability that } X \text{ belongs to } I_j \text{ when } \mu_y = \mu_{y,0}, \\
q_j^{(1)} \quad \text{probability that } X \text{ belongs to } I_j \text{ when } \mu_y = \mu_{y,0} + c\sigma_y, \\
q_2^{(1)} \quad \text{probability that } X \text{ fall outside the 2nd control limits when } \mu_y = \mu_{y,0} + c\sigma_y, \\
R_0 \quad \text{number of false alarms when the signal is in action region,} \\
R_1 \quad \text{number of false alarms when the signal is in } I_1 \text{ region,} \\
a_1 \quad \text{cost for finding and eliminating an assignable cause,} \\
a_j' \quad \text{cost for identifying a false alarm,} \\
a_j'' \quad \text{cost incurred from the lost production due to a false alarm,} \\
a_3 \quad \text{fixed sampling cost,} \\
a_4 \quad \text{variable sampling cost,} \\
b_j' \quad \text{time required to find an assignable cause,} \\
b_j'' \quad \text{time required to eliminate an assignable cause,} \\
b_2 \quad \text{time required to identify a false alarm,} \\
b_3 \quad \text{time required to take and interpret a sample,} \\
i_1 \quad \text{net income per unit time in in-control state,} \\
i_2 \quad \text{net income per unit time in out-of-control state,} \\
I \quad \text{net income per cycle,} \\
I' \quad \text{total income per cycle,} \\
C_1 \quad \text{sampling cost per cycle,} \\
C_2 \quad \text{per-cycle cost for searching and eliminating an assignable cause,} \\
\tau \quad \text{expected time of occurrence of an assignable cause between the two adjacent samples,} \\
T \quad \text{length of a cycle,} \\
ARL_0 \quad \text{average run length when the process is in control,} \\
ARL_1 \quad \text{average run length when the process is out of control,} \\
ARL_L \quad \text{lower bound on average run length when the process is in control,} \\
ARL_U \quad \text{upper bound on average run length when the process is out of control.} \\
\]

Introduction

Control charts are widely used to monitor and detect different process variations. Controlling the process variations prevents the manufacturing of poor products, the need to rework products, and waste. Shewhart control charts with fixed parameters are typically slow to detect signals of an assignable cause. Consequently, new alternatives to the Shewhart charts have been proposed to make up for this weakness. However, a traditional approach to sampling for control charts is taken from a process with a fixed size and fixed time interval between samples, and the resulting control chart is called a fixed sampling interval (FSI) control chart. Recent studies have shown that adaptive charts have superior statistical and economic performance compared to FSI control charts. Recent studies have shown that adaptive charts have superior statistical and economic performance compared to FSI control charts.

Reynolds et al. [1] introduced the idea of a variable sampling interval during the production process based on recent data obtained from this process. VSI control charts have received much attention, for example, Reynolds and Arnold, Runger and Montgomery, Reynolds, Costa and Rahim, Lin et al., Chew et al. and Zhang Min et al. [2–8].

In all of these studies, an inspection is performed on the quality characteristic of interest (performance variable). In some situations, it is impossible or not economical to directly inspect the performance variable. In such cases, the use of a surrogate variable that is highly correlated with a performance variable is an attractive alternative, especially when inspecting the surrogate variable is relatively less expensive than inspecting the performance variable. In a measurement system for VHTR tristructural-isotropic (TRISO) fuel, for example, a direct method for mea-
suring the TRISO fuel diameter, which is essential for the fabrication of the reactor fuel, requires very sophisticated equipment and high costs. When a direct measurement of the performance variable is too expensive, a surrogate variable that is linearly correlated with the performance variable but less expensive to measure may be considered instead of the performance variable. In the measurement system example, a Particle Size and Shape Analyzer (PSA) measurement method is generally less accurate but much less expensive than the X-ray measurement method and does not produce radiation because PSA uses laser diffraction to analyze the size grade and shape of the particles. Thus, the PSA measurement value might be used as a surrogate variable. Lee et al. [9] considered an economic design for the $X$-control chart using a surrogate variable under the assumption that a performance variable might not be used. Additionally, Lee et al. [10] proposed an economic design idea for the VSI $\overline{X}$ control chart using a surrogate variable.

In this paper, we developed an economic statistical design for a VSI $\overline{X}$ control chart using a surrogate variable with a genetic algorithm under the assumption that the performance variable might not be used and compared its results with a fixed $\overline{X}$ control chart design. In the next section, an outline of the economic statistical model for the $\overline{X}$ control chart and its underlying assumptions are described. In the development of profit function section, the expected cycle time and expected net income per cycle for the proposed model are derived. The cost function is then obtained as the ratio of these two quantities. In the numerical comparisons using genetic algorithms section, the proposed model is used to obtain the optimum designs using the test input values in Panagos et al. [11, PHM hereafter], and is then applied to the measurement system for the nuclear fuel. In the sensitivity analysis section, a sensitivity analysis of the proposed model is presented for changes in the input parameters. The final section briefly summarizes the results obtained in this paper.

Assumptions and model specifications

Assumptions

Assume that each product possesses a continuous performance variable $Y$ that measures the degree to which a product satisfies the stated or implied expectations of customers. If we consider a situation in which measuring the performance variable is expensive, time-consuming, or even destructive, it is attractive to use a surrogate variable $X$ that is linearly correlated with the performance variable but less expensive to measure in this situation. In such cases, we can monitor the production process using a control chart for only the surrogate variable, based on the correlation between the performance and surrogate variables obtained from the experiment performed prior to production.

With this situation in mind, we propose an $\overline{X}$ control chart based on only the surrogate variable. For a more specific presentation of the model, we describe the nature of the process conditions and the underlying statistical assumptions as follows:

- The process begins in the in-control state, with the mean and variance of the performance variable being $\mu_y$ and $\sigma_y$, respectively. An assignable cause occurs according to a Poisson process with an intensity of $\lambda$ occurrences per unit of time. If an assignable cause of the magnitude $c$ occurs, then the process mean shifts from $\mu_x$ to $\mu_y \pm c\sigma_y$.

- The surrogate variable $X$ given $Y = y$ is normally distributed with the mean $\lambda_1 + \lambda_2 y$ and variance $\sigma^2$, where $\lambda_1$ and $\lambda_2$ are known constants. $\lambda_2$ is assumed to be positive so that $X$ and $Y$ have a positive linear relationship. It can be easily shown that $(X, Y)$ follows a bivariate normal distribution with the means $(\lambda_1 + \lambda_2 y, \mu_y)$, variances $(\lambda_2 \sigma^2 + \sigma^2, \sigma^2)$, and correlation coefficient $\rho = \lambda_2 \sigma_y / (\lambda_2 \sigma^2 + \sigma^2)^{1/2}$ (see Tang and Lo [12]).

- The time taken to find an assignable cause is $b_1$, and the time required to eliminate it is $b_2$. The time taken to identify a false alarm is $b_3$, and the time required to take and interpret a sample is $b_4$.

- The cost for finding and eliminating an assignable cause is $a_1$, while the cost for identifying a false alarm is $a_2$ and the cost incurred from the lost production due to a false alarm is $a_2^\prime$. In addition, the cost of sampling and testing for $X$ variables is $a_3 + a_4 n_x$, where $a_3$ and $a_4$ are fixed and variable sampling costs, respectively, and $n_x$ is the sample size. The net incomes per unit time of operation in the in-control and out-of-control states are $i_1$ and $i_2$, respectively.

Model specification

Based on the previous assumptions, we developed a model for an $\overline{X}$ control chart using a surrogate variable. Let $h_j$ denote the time interval between $X$ samples, where the sampling interval is varied based on the value of the preceding sample mean. In this article, we assume that the VSI $\overline{X}$ control charts use a finite number of interval lengths $h_1, \ldots, h_m$, where $h_1 \geq \ldots \geq h_m$ and $m \geq 2$. The choice of a sampling interval length can be represented by a sampling interval, and let $k_j$ denote the $j$-th threshold limit factors for an $\overline{X}$ control chart. Let the region
between the two control limits be portioned into \( m \) sub-regions as follows:

\[
\begin{align*}
&\left(\mu_0 - k_j \frac{\sigma}{\sqrt{n}}, \mu_0 - k_{j+1} \frac{\sigma}{\sqrt{n}}\right) \\
&\cup \left(\mu_0 + k_j \frac{\sigma}{\sqrt{n}}, \mu_0 + k_{j+1} \frac{\sigma}{\sqrt{n}}\right)
\end{align*}
\]

(1)

for \( j = 1, 2, \ldots, m - 1, \)

\[
\left(\mu_0 - k_m \frac{\sigma}{\sqrt{n}}, \mu_0 + k_m \frac{\sigma}{\sqrt{n}}\right) \quad \text{for} \quad j = m,
\]

where

\[
0 \leq k_m \leq k_{m-1} \leq \ldots \leq k_1,
\]

\[
\sigma_x = \frac{\sigma}{\sqrt{n}}.
\]

If we define \( Z_x = \frac{\sqrt{n}(\bar{X} - \mu_x)}{\sigma_x} \), then the proposed model can be summarized as follows:

Step 1. Take a sample of size \( n_x \) after an interval of \( h_j \) time units (\( j \geq 2 \)).

Step 2. If \( |Z_x| < k_2 \) (if sampling falls in the second outermost control lines), go to step 1. Otherwise, go to step 3.

Step 3. If \( |Z_x| < k_1 \) (if sampling falls between the first outermost control lines and second outermost control lines), go to step 4. Otherwise, stop the process and go to step 4.

Step 4. If the alarm is false, go to step 1. Otherwise, go to step 5.

Step 5. Identify and eliminate the assignable cause. Go to step 1.

Figure 1 shows the monitoring procedures of the proposed model.

An economic model can be formulated by introducing the total cost function, which reflects the relationships between the design parameters of the control charts and the several types of costs previously discussed. Because the underlying process for the \( X \) control chart using a surrogate variable is a renewal reward process, the long-run expected net income per unit of time is given by

\[
E(A) = \lim_{t \to \infty} \frac{E[TI(t)]}{t} = \frac{E(I)}{E(T)},
\]

(2)

where \( TI(t) \) is the total net income until time \( t \); \( I \) is the net income per cycle, and \( T \) is the length of the cycle. Thus, an economic design of the \( X \) control chart using a surrogate variable is to determine the values of \( h_j, n_x, \) and \( k_j \) \( (j = 1, \ldots, m) \) such that they maximize the long-run expected net income per unit of time.

Although the economic design is most effective from a purely economic point of view, it might have undesirable statistical properties such as high type I and/or type II error probabilities. To make the resulting design satisfy both the cost effectiveness and certain statistical requirements, we may add the desired statistical constraints to the optimization procedure for the design parameters of the proposed model. We now define the economic statistical design of the control charts as the design in which the long-run expected net income per unit time is maximized subject to a lower bound on the in-control ARL (\( L_{ARL} \)) and an upper bound on the out-of-control ARL (\( L_{ARL} \)), as in Montgomery et al. [13]. A model for the economic statistical design can be formulated as follows:

Maximize \( E(A) \)

Subject to \( ARL_0 > ARL_L \), \( ARL_1 < ARL_U \),

(3)

where \( ARL_0 \) and \( ARL_1 \) are the average run lengths while in control and out of control, respectively. These constraints on ARL can add sensitivity to the shifts in the process mean to the economic model. Note that \( ARL_0 \) and \( ARL_1 \) are the means of the geometric distributions with parameters \( q_0 \) and \( q_1 \), respectively. The constraints in (3) can be equivalently expressed as follows:

\[
k_1 > \Phi^{-1} \left( \frac{1}{1 - 1/2ARL_L} \right),
\]

(4)
where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ denote the standard normal distribution function and the inverse standard normal distribution function, respectively.

Therefore, the economic statistical design is one that maximizes $E(A)$ subject to the constraints given in (4) and (5). Note that $k_1$ should be determined using computational iterations because (5) cannot be explicitly solved with respect to $k_1$ because of two standard normal distribution functions.

**Development of the profit function**

In this section, we derive the expected cycle time and expected net income per cycle for the proposed model. The cost function, which is the long-run expected net income per unit of time, is then obtained as the ratio of these two quantities.

**Expected cycle time**

In the proposed model, the process begins in the in-control state and shifts from the in-control state to the out-of-control state by an assignable cause. When a signal is detected, the process is stopped immediately, and a search for an assignable cause is undertaken to see whether it really exists. If an assignable cause exists, it is eliminated and the process restarts. Figure 2 shows the process cycle assumed in the proposed model.

![Figure 2. The process cycle.](image)

The cycle time for the discontinuous process consists of four periods: an in-control period, an out-of-control period, a period during which the process is stopped due to a false alarm and a period for finding and eliminating an assignable cause. From the assumptions in the previous section, the expected length of the in-control period is $1/\lambda$. We next derive the lengths of the three other periods.

**1) Expected length of the out-of-control period**

Note that $D$ takes one of $h_1, \ldots, h_m$ and that the distribution of $D$ is determined when the process is in-control because the sampling interval is determined by a previous sampling statistics value.

Reynolds et al. [1] assumed that

$$P(D = h_j) = \frac{h_j q_j^{(0)}}{\sum_{l=1}^m h_l q_l^{(0)}}, \quad j = 1, 2, \ldots, m. \quad (6)$$

Based on the conditional probability, the expected value of $D$, given that $D = h_j$, is

$$E(D) = \sum_{j=1}^m h_j P(D = h_j). \quad (7)$$

Because we investigate the process state only if $\bar{X}$ falls in the $I_1$ region, the expected value of the out-of-control period, $E(OC)$, can be expressed as

$$E(OC) = \sum_{j=2}^m h_j \left( \frac{1}{1 - r_2^{(1)}} - 1 \right) \frac{q_j^{(1)}}{1 - r_2^{(1)}} + E(D) - \tau. \quad (8)$$

Under the assumption that an assignable cause occurs according to a Poisson process, the expected time of $\tau$, which is the time lag between the last preceding sampling point and the time at which an assignable cause occurs, can be expressed as

$$\tau = \sum_{j=1}^m P(D = h_j) E(\tau|D = h_j). \quad (9)$$

According to Duncan [14], $\tau$ is well approximated as

$$E(\tau|D = h_j) = \frac{1 - (1 + \lambda h_j) e^{-\lambda h_j}}{\lambda (1 - e^{-\lambda h_j})}, \quad j = 1, 2, \ldots, m. \quad (2)$$

**2) Expected duration elapsed due to false alarms**

Let $R_0$ be the number of false alarm signals in an action region before the process goes out of control. We then obtain the expected number of samples in the in-control period as follows

$$E(R_0) = q_j^{(0)} \sum_{j=1}^m q_j^{(0)} e^{-\lambda h_j} \left( 1 - q_k^{(0)} - \sum_{k=1}^m q_k^{(0)} e^{-\lambda h_k} \right)^2$$

$$\sum_{l=1}^m q_l^{(0)} (1 - e^{-\lambda h_l}). \quad (10)$$

A detailed derivation is given in Bai and Lee [15].

The expected duration elapsed due to false alarms is then $b_2$ times the expected number of $X$ samples taken before the shift. That is,
The expected total income per cycle is thus

$$E(FA) = b_2 q_s^{(0)} \frac{\sum_{j=1}^{m} q_j^{(0)} e^{-\lambda b_j}}{\left(1 - q_s^{(0)} - \sum_{k=1}^{m} q_k^{(0)} e^{-\lambda h_k}\right)^2} \cdot \sum_{l=1}^{m} q_l^{(0)} (1 - e^{-\lambda h_l}).$$

(11)

In this model, we investigate the process state by performing the process if $X$ falls in the $I_1$ region, and the expected duration is zero.

(3) Expected duration for finding and eliminating an assignable cause

When the process is out of control, $b_1' + b_2'$ is taken to find and eliminate an assignable cause. The expected duration for finding and eliminating an assignable cause is therefore

$$E(AC) = b_3 n_x + b_1' + b_2'.$$

(12)

Adding up these four periods, we obtain the expected cycle time for the process as


(13)

Expected net income

The expected net income per cycle for the discontinuous process can be written as

$$E(I) = E(I') - E(C_1) - E(C_2),$$

(14)

where $I'$ is the total income per cycle; $C_1$ is the sampling cost per cycle, and $C_2$ is the per-cycle cost associated with finding, investigating, and if necessary, eliminating an assignable cause when an $X$ sample mean falls outside the action or warning limits. We next derive the expressions for the expected values of these costs.

(1) Expected total income per cycle

The expected total income for an in-control period is $i_1/\lambda$, while that for an out-of-control period is $i_2$ times the expected length of the out-of-control period. The expected total income per cycle is thus

$$E(I') = i_1 \ast E(IC) + i_2 \ast E(OC)$$

$$= \frac{i_1}{\lambda} + i_2 \left( \sum_{j=2}^{m} \frac{h_j}{1 - r_j^{(1)}} - 1 \right) \frac{q_j^{(1)}}{1 - r_j^{(1)}} + E(D - \tau).$$

(15)

(2) Expected sampling cost per cycle

Let $SN$ be the expected number of samples in the in-control period. We then get

$$SN_0 = \frac{\sum_{j=1}^{m} q_j^{(0)} e^{-\lambda h_j}}{\left(1 - q_s^{(0)} - \sum_{k=1}^{m} q_k^{(0)} e^{-\lambda h_k}\right)^2} \cdot \sum_{l=1}^{m} q_l^{(0)} (1 - e^{-\lambda h_l}).$$

(16)

However, when the process is out-of-control, the number of samples required to produce a signal is a geometric random variable, and the corresponding expected number of sampling is given by

$$SN_1 = \frac{1}{1 - r_1^{(1)}}.$$

(17)

Therefore, the expected sampling cost per cycle is given by

$$E(C_1) = (a_3 + a_4 n_x)(SN + SN_1)$$

$$= (a_3 + a_4 n_x) \left( \sum_{j=1}^{m} q_j^{(0)} e^{-\lambda h_j} \right) \left(1 - q_s^{(0)} - \sum_{k=1}^{m} q_k^{(0)} e^{-\lambda h_k}\right)^2$$

$$\cdot \sum_{l=1}^{m} q_l^{(0)} (1 - e^{-\lambda h_l}) + \frac{1}{(1 - r_1^{(1)})}.$$  

(18)

Expected cost associated with an assignable cause and false alarms per cycle

The expected cost incurred from false alarms is given by

$$\left( a_2' + a_2'' \right) E(R_0) \left( a_2' \right) E(R_1).$$

(19)

where $R_1$ is the number of false alarms when the signal is in the $I_1$ region. Because the expected cost incurred from an assignable cause is $a_1$, the corresponding expected cost per cycle is given by

$$E(C_2) = \left( a_2' + a_2'' \right) E(R_0) + \left( a_2' \right) E(R_1) + a_1 r_2^{(1)} a_1 + a_1.$$ 

(20)

where

$$E(R_1) = \frac{q_1^{(0)} e^{-\lambda h_j}}{\left(1 - q_s^{(0)} - \sum_{k=1}^{m} q_k^{(0)} e^{-\lambda h_k}\right)^2} \cdot \sum_{l=1}^{m} q_l^{(0)} (1 - e^{\lambda h_l}).$$
The expected net income per cycle can be obtained by subtracting (18) and (20) from (15).

Numerical comparisons using genetic algorithms

The proposed model is applied to the test example considered in PHM for comparison with the economic model based on the performance variable. It is assumed that $\rho = 0.6$ or $0.9$, and the variable sampling cost ($a_4$) is one-tenth of the original values in PHM because the process is monitored by a surrogate variable, which is usually much cheaper to measure. The values of the corresponding input parameters are given in Table 1. We perform a sensitivity analysis for $m = 2$ and $m = 3$ but describe only for $m = 2$ because we observed similar results for $m = 3$. The optimal values of the design parameters $h$, $n_x$, $k_w$, and $k_a$ for $m = 2$ that maximize $E(A)$ subject to the constraints in (4) and (5) with $\text{ARL}_L = 4$ and $\text{ARL}_U = 500$ were determined with Evolver, a genetic algorithm optimization tool. Genetic algorithms have been used in many engineering areas such as industrial engineering, mechanical engineering, aerospace engineering, etc., including statistical process control in Lin et al. [6]. The computational results are shown in Table 2. Based on the results presented in Table 2, we may observe the following:

- When the correlation is high ($\rho = 0.9$), the proposed economic model yields a higher income than that of the PHM model for all cases. However, when the correlation is low ($\rho = 0.6$), the proposed economic model yields a higher income than that of the PHM model. Thus, the proposed economic model seems to be effective when $\rho$ is relatively large.
- When the correlation is high, the sample size $n_x$ is smaller than or equal to $n_y$ for all cases.
- In all of the cases when the correlation is high, $k_1$ is higher than 4.5, which is close to the bounds set for the algorithm. This effectively means that a minimal sample needs be taken just every once in a while without stopping the process. The reason for this result might be due to higher sampling costs and a lower cost for identifying a false alarm, combined with a smaller difference in the net income per unit time between the in-control state and out-of-control state.
- The higher the correlation is, the narrower the warning region $(k_1 - k_2)$ is. This implies that the process needs to not stop to investigate a signal because the income earned from the process is larger than the cost for identifying a false alarm.
- In a high correlation case, the yield of the proposed economic model improves that of the PHM model by about 14.9%, and the yield of the economic statistical model is lower than that of the proposed economic model by about 5%. In a low correlation case, however, the proposed economic model improved that of the PHM model by about 14%. This suggests that the statistical constraints on $\text{ARL}$ tend to have a greater influence on the economic model when the correlation is high. As a result, the higher the correlation is, the more efficient the proposed economic model is.

<table>
<thead>
<tr>
<th>Example</th>
<th>$A$</th>
<th>$c$</th>
<th>$i_1$</th>
<th>$i_2$</th>
<th>$a_1$</th>
<th>$a_2'$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$b'_1$</th>
<th>$b'_2$</th>
<th>$b_2$</th>
<th>$b_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td></td>
<td>50</td>
<td></td>
<td>45</td>
<td>25</td>
<td>25</td>
<td>0.5</td>
<td>0.1</td>
<td>1.53</td>
<td>1.53</td>
<td>4.05</td>
</tr>
<tr>
<td>2</td>
<td>0.01</td>
<td>1</td>
<td>150</td>
<td>50</td>
<td>350</td>
<td>250</td>
<td>250</td>
<td>5.0</td>
<td>1.0</td>
<td>1.53</td>
<td>1.53</td>
<td>4.05</td>
</tr>
<tr>
<td>3</td>
<td>0.01</td>
<td>2</td>
<td>50</td>
<td>1</td>
<td>260</td>
<td>25</td>
<td>25</td>
<td>5.0</td>
<td>1.0</td>
<td>2.00</td>
<td>2.00</td>
<td>41.00</td>
</tr>
<tr>
<td>4</td>
<td>0.01</td>
<td>2</td>
<td>150</td>
<td>50</td>
<td>135</td>
<td>250</td>
<td>250</td>
<td>5.0</td>
<td>0.1</td>
<td>2.00</td>
<td>2.00</td>
<td>41.00</td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>1</td>
<td>150</td>
<td>100</td>
<td>45</td>
<td>250</td>
<td>250</td>
<td>5.0</td>
<td>0.1</td>
<td>2.00</td>
<td>2.00</td>
<td>41.00</td>
</tr>
<tr>
<td>6</td>
<td>0.05</td>
<td>1</td>
<td>50</td>
<td>−50</td>
<td>350</td>
<td>25</td>
<td>25</td>
<td>0.5</td>
<td>1.0</td>
<td>2.00</td>
<td>2.00</td>
<td>41.00</td>
</tr>
<tr>
<td>7</td>
<td>0.05</td>
<td>2</td>
<td>150</td>
<td>100</td>
<td>260</td>
<td>250</td>
<td>250</td>
<td>0.5</td>
<td>1.0</td>
<td>1.53</td>
<td>1.53</td>
<td>4.05</td>
</tr>
<tr>
<td>8</td>
<td>0.05</td>
<td>2</td>
<td>50</td>
<td>−50</td>
<td>135</td>
<td>25</td>
<td>25</td>
<td>5.0</td>
<td>0.1</td>
<td>1.53</td>
<td>1.53</td>
<td>4.05</td>
</tr>
<tr>
<td>9</td>
<td>0.01</td>
<td>1</td>
<td>150</td>
<td>50</td>
<td>135</td>
<td>250</td>
<td>250</td>
<td>0.5</td>
<td>1.0</td>
<td>2.00</td>
<td>2.00</td>
<td>5.00</td>
</tr>
<tr>
<td>10</td>
<td>0.01</td>
<td>1</td>
<td>150</td>
<td>50</td>
<td>260</td>
<td>25</td>
<td>25</td>
<td>5.0</td>
<td>0.1</td>
<td>2.00</td>
<td>2.00</td>
<td>5.00</td>
</tr>
</tbody>
</table>
Table 2
Optimum designs for the test examples.

<table>
<thead>
<tr>
<th>Example</th>
<th>The PHM Model</th>
<th>Proposed Economic Model</th>
<th>Economic Statistical Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n_0$</td>
<td>$h$</td>
<td>$k$</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>2.75</td>
<td>3.14</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>6.33</td>
<td>2.95</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>6.46</td>
<td>3.46</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>1.54</td>
<td>4.31</td>
</tr>
<tr>
<td>5</td>
<td>26</td>
<td>2.41</td>
<td>5.76</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>2.48</td>
<td>2.75</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>2.01</td>
<td>3.32</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>1.50</td>
<td>3.12</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>5.03</td>
<td>2.72</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>2.79</td>
<td>2.90</td>
</tr>
</tbody>
</table>

Application to a nuclear fuel measurement system

In this section, the proposed model is applied to the VHTR TRISO fuel measurement system, previously described by Kim et al. [16], for a comparison with the economic model based on a performance variable. TRISO fuel is a type of micro fuel particle. It consists of a kernel, low-density pyrocarbon, inner high-density pyrocarbon, silicon carbide, and outer high-density pyrocarbon. It is important to obtain the exact figure of the fuel size for uniform production. When we measure the outer diameter of the VHTR fuel, we can use methods such as a PSA, micrometer, and X-ray. Of the three methods, the use of X-ray (Y, performance variable) is the most accurate; however, the measurement equipment is too expensive to purchase. On the other hand, the PSA (X, surrogate variable) method is less precise due to a scattering of light; however, it is easier to obtain data without emitting dangerous radiation.

From the actual data analysis, it is known that the mean and variance of $Y$ are $\mu_Y = 1099.32$ and $\sigma_Y^2 = (32.18 \mu m)^2$, respectively, and the variance of $X$ is $\sigma_X^2 = (34.75 \mu m)^2$. It is also known that $X$ for the given $Y = y$ is normally distributed with a mean of $100.41 + 0.86y$ and a variance of $(12.57 \mu m)^2$ and that the correlation coefficient between $X$ and $Y$ is $\rho = 0.799$. Table 3 shows the TRISO fuel sizes measured using the PSA and X-ray methods.

The optimum values of the design parameters are obtained in Table 4 for both the economic model and the economic statistical model based on the surrogate variable, along with those for the economic model based on the performance variable. The following values for the cost and process parameters are assumed to be $\lambda = 0.05, c = 2, t_1 = 100, t_2 = 0, a_1 = 200, a_2' = 200, a_2 = 200, a_4 = 3, a_4 = 1.5, b_1 = 10, b_1' = 10, b_2 = 25$, and $b_3 = 0.3$.

Table 4 shows that the economic model using the surrogate variable yields a 2% higher expected net income per hour than the economic model using the performance variable. It also shows that the economic statistical model yields the optimum design, which is only slightly different from that for the economic model with a minimal decrease in the expected per-hour net income. We note that the parameter values of the economic statistical model are very robust to changes in both $ARL_Y$ and $ARL_L$.
Management and Production Engineering Review

Table 3
TRISO fuel sizes measured by the PSA method and X-ray method.

<table>
<thead>
<tr>
<th>Obs</th>
<th>PSA (Obs)</th>
<th>X-rays (Obs)</th>
<th>Obs</th>
<th>PSA (Obs)</th>
<th>X-rays (Obs)</th>
<th>Obs</th>
<th>PSA (Obs)</th>
<th>X-rays (Obs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1061</td>
<td>1047</td>
<td>18</td>
<td>1055</td>
<td>1047</td>
<td>35</td>
<td>1104</td>
<td>1042</td>
</tr>
<tr>
<td>2</td>
<td>1156</td>
<td>1116</td>
<td>19</td>
<td>1088</td>
<td>1027</td>
<td>36</td>
<td>1023</td>
<td>979</td>
</tr>
<tr>
<td>3</td>
<td>1094</td>
<td>1046</td>
<td>20</td>
<td>1073</td>
<td>1019</td>
<td>37</td>
<td>1129</td>
<td>1090</td>
</tr>
<tr>
<td>4</td>
<td>1084</td>
<td>1037</td>
<td>21</td>
<td>1102</td>
<td>1066</td>
<td>38</td>
<td>1116</td>
<td>1063</td>
</tr>
<tr>
<td>5</td>
<td>1072</td>
<td>1021</td>
<td>22</td>
<td>1090</td>
<td>1041</td>
<td>39</td>
<td>1192</td>
<td>1138</td>
</tr>
<tr>
<td>6</td>
<td>1133</td>
<td>1089</td>
<td>23</td>
<td>1122</td>
<td>1066</td>
<td>40</td>
<td>1169</td>
<td>1123</td>
</tr>
<tr>
<td>7</td>
<td>1116</td>
<td>1078</td>
<td>24</td>
<td>1061</td>
<td>1024</td>
<td>41</td>
<td>1093</td>
<td>1048</td>
</tr>
<tr>
<td>8</td>
<td>1091</td>
<td>988</td>
<td>25</td>
<td>1090</td>
<td>1050</td>
<td>42</td>
<td>1116</td>
<td>1083</td>
</tr>
<tr>
<td>9</td>
<td>1029</td>
<td>987</td>
<td>26</td>
<td>1117</td>
<td>1075</td>
<td>43</td>
<td>1106</td>
<td>1057</td>
</tr>
<tr>
<td>10</td>
<td>1114</td>
<td>1086</td>
<td>27</td>
<td>1082</td>
<td>1029</td>
<td>44</td>
<td>1105</td>
<td>1037</td>
</tr>
<tr>
<td>11</td>
<td>1119</td>
<td>1074</td>
<td>28</td>
<td>1069</td>
<td>1007</td>
<td>45</td>
<td>1103</td>
<td>1068</td>
</tr>
<tr>
<td>12</td>
<td>1082</td>
<td>1049</td>
<td>29</td>
<td>1066</td>
<td>1022</td>
<td>46</td>
<td>1116</td>
<td>1090</td>
</tr>
<tr>
<td>13</td>
<td>1118</td>
<td>1070</td>
<td>30</td>
<td>1070</td>
<td>1039</td>
<td>47</td>
<td>1063</td>
<td>1011</td>
</tr>
<tr>
<td>14</td>
<td>1113</td>
<td>1071</td>
<td>31</td>
<td>1043</td>
<td>1069</td>
<td>48</td>
<td>1111</td>
<td>1001</td>
</tr>
<tr>
<td>15</td>
<td>1100</td>
<td>1063</td>
<td>32</td>
<td>1104</td>
<td>1048</td>
<td>49</td>
<td>1158</td>
<td>1005</td>
</tr>
<tr>
<td>16</td>
<td>1126</td>
<td>1072</td>
<td>33</td>
<td>1060</td>
<td>999</td>
<td>50</td>
<td>1104</td>
<td>1052</td>
</tr>
<tr>
<td>17</td>
<td>1083</td>
<td>1039</td>
<td>34</td>
<td>1105</td>
<td>1064</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4
Results of the VSI X control chart using surrogate variables.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Economic Design (Performance var.)</th>
<th>Economic Design (Surrogate var.)</th>
<th>Statistical Economic Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n_x$ ($n_y$)</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>$h_1$</td>
<td>6.1382</td>
<td>4.6418</td>
<td>4.7150</td>
</tr>
<tr>
<td>$h_2$</td>
<td>6.3844</td>
<td>4.9145</td>
<td>5.1759</td>
</tr>
<tr>
<td>$k_2$</td>
<td>4.2050</td>
<td>4.9620</td>
<td>3.6526</td>
</tr>
<tr>
<td>$k_1$</td>
<td>4.9945</td>
<td>4.9928</td>
<td>3.6540</td>
</tr>
<tr>
<td>$E(T)$</td>
<td>107.4766</td>
<td>106.5779</td>
<td>106.9600</td>
</tr>
<tr>
<td>$E(T')/E(T)$</td>
<td>46.5518</td>
<td>46.9372</td>
<td>46.7708</td>
</tr>
<tr>
<td>$E(C_1)/E(T)$</td>
<td>1.5044</td>
<td>1.0173</td>
<td>0.9823</td>
</tr>
<tr>
<td>$E(C_2)/E(T)$</td>
<td>2.4192</td>
<td>2.4495</td>
<td>2.4331</td>
</tr>
<tr>
<td>$E(A)$</td>
<td>42.6282</td>
<td>43.4804</td>
<td>43.3554</td>
</tr>
</tbody>
</table>

Sensitivity analysis

The sensitivity of the design is important to find an optimal plan because it is difficult to exactly specify the values of the process parameters. In this section, we find how sensitive the optimum values of the design parameters ($h_1$, $h_2$, $k_2$, and $k_1$) are to changes in certain input parameter values using example 6 in Sec. 5. The values of the input parameters are varied ±30% from the base value. Based on the results of the sensitivity analysis, we observed the following:

- Effects of $\lambda$ and $c$: Fig. 3 shows that the optimum values of $h_1$ and $h_2$ decrease as the failure rate ($\lambda$) increases. This conforms to our intuition that the process needs to be controlled more tightly because it tends to fail more frequently as $\lambda$ increases. An increase in the shift size of the process mean ($c$) results in an increase in the warning region ($k_1 - k_2$) in the optimum value of $h$ shown in Fig. 4.

Fig. 3. Sensitivity of $h_1$, $h_2$, $k_2$, and $k_1$ to $\lambda$. 
Fig. 4. Sensitivity of $h_1$, $h_2$, $k_2$, and $k_1$ to $c$.

- Effects of $i_1$ and $i_2$: Fig. 5 shows that an increase in the per-hour net income in the in-control state ($i_1$) leads to a decrease in $h_1$ and $h_2$, whereas the optimum values of $k_2$ increase as the net income per unit time in the out-of-control state ($i_2$) increases shown in Fig. 6. The reason for the increasing trend in the warning region in Fig. 6 is that it is more efficient and more economical to check the process state instead of a process stop.

Fig. 5. Sensitivity of $h_1$, $h_2$, $k_2$, and $k_1$ to $i_1$.

Fig. 6. Sensitivity of $h_1$, $h_2$, $k_2$, and $k_1$ to $i_2$.

- Effects of $a_1$, $a'_2$ and $a''_2$: An increase in $a_1$ causes a decrease in the optimum value of $h_1$ and $h_2$ shown in Fig. 7. Figure 8 shows both an increase in $h_1$ and $h_2$ and a decrease in the warning region as the cost for identifying a false alarm ($a'_2$) increases.

Fig. 7. Sensitivity of $h_1$, $h_2$, $k_2$, and $k_1$ to $a_1$.

Fig. 8. Sensitivity of $h_1$, $h_2$, $k_2$, and $k_1$ to $a'_2$.

The rest of the parameters including $a''_2$, $b'_1$ and $b''_1$ show a similar trend to those of the above parameters in the sensitivity analysis.

Conclusions

We proposed an economic design of a VSI $\bar{X}$ control chart based on a surrogate variable for a case in which using the performance variable is impossible or inappropriate. Compared with the control chart based on a performance variable, the proposed model gives a larger expected net income per unit of time in the long-run if the correlation between the performance variable and the surrogate variable is relatively high.

When the proposed VSI $\bar{X}$ control chart was applied to a nuclear fuel measurement system, the numerical comparison results show that the VSI model using a surrogate variable is more efficient than the
VSI model using a performance variable or FSI model from a net income point of view. Additionally, if the correlation coefficient between the surrogate variable and performance variable is higher, it was found that the long-run expected net income per unit of time is also increased.

Conversely, for a low level of the correlation coefficient, it can be more useful to use a VSI model using a performance variable. The proposed model was confined to the sample mean control chart under the assumption that a single assignable cause occurs according to the Poisson process. However, the model may also be extended to other types of control charts using a single or multiple assignable cause assumptions such as VSS $\bar{X}$ control chart, EWMA, and CUSUM charts.

References


