Management and Production Engineering Review
Volume 7 • Number 4 • December 2016 • pp. 54–64
DOI: 10.1515/mper-2016-0036

ECONOMIC STATISTICAL DESIGN OF VARIABLE SAMPLING INTERVAL $\overline{X}$ 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, Yuseong-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 $\overline{X}$ 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) $\overline{X}$ 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) $\overline{X}$ 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,
- μ_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,
- μ_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,
- σ_x – standard deviation of surrogate variable,
- σ_y – standard deviation of performance variable,
- c – magnitude of the shift in the process mean measured in σ_y unit,

λ – number of the occurrence of an assignable cause,
ρ – correlation coefficient between a surrogate variable and a performance variable,
m – number of different sampling interval length, $m \geq 2$,
h_j – j-th smallest sampling interval lengths, $j = 1, 2, \ldots, m$,
k_j – j-th threshold limit, $0 \leq k_m \leq k_j \leq \ldots \leq k_1$,
I_j – j-th sampling interval region using h_j, $j = 1, 2, \ldots, m$,
IC – in control time,
OC – out of control time,
FA – time to owing to false alarm,
AC – time to owing to assignable cause,
SN_0 – expected number of samples in in-control period,
SN_1 – expected number of samples in out-of-control period,
$q_s^{(0)}$ – probability that $\overline{X}$ fall outside the 1st control limits when $\mu_y = \mu_{y,0}$,
$q_s^{(1)}$ – probability that $\overline{X}$ fall outside the 1st control limits when $\mu_y = \mu_{y,0} + c\sigma_y$,
$q_j^{(0)}$ – probability that $\overline{X}$ belongs to I_j when $\mu_y = \mu_{y,0}$,
$q_j^{(1)}$ – probability that $\overline{X}$ belongs to I_j when $\mu_y = \mu_{y,0} + c\sigma_y$,
$q_2^{(1)}$ – probability that $\overline{X}$ fall outside the 2nd control limits when $\mu_y = \mu_{y,0} + c\sigma_y$,
R_0 – number of false alarms when the signal is in action region,
R_1 – number of false alarms when the signal is in I_1 region,
a_1 – cost for finding and eliminating an assignable cause,
a_2' – cost for identifying a false alarm,
a_2'' – cost incurred from the lost production due to a false alarm,
a_3 – fixed sampling cost,
a_4 – variable sampling cost,
b_1' – time required to find an assignable cause,
b_1'' – time required to eliminate an assignable cause,
b_2 – time required to identify a false alarm,
b_3 – time required to take and interpret a sample,
i_1 – net income per unit time in in-control state,
i_2 – net income per unit time in out-of-control state,
I – net income per cycle,
I' – total income per cycle,
C_1 – sampling cost per cycle,
C_2 – per-cycle cost for searching and eliminating an assignable cause,
τ – expected time of occurrence of an assignable cause between the two adjacent samples,
T – length of a cycle,
ARL_0 – average run length when the process is in control,
ARL_1 – average run length when the process is out of control,
ARL_L – lower bound on average run length when the process is in control,
ARL_U – 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 measuring 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 $\overline{X}$-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 μ_y and σ_y, respectively. An assignable cause occurs according to a Poisson process with an intensity of λ occurrences per unit of time. If an assignable cause of the magnitude c occurs, then the process mean shifts from μ_x to $\mu_y \pm c\sigma_x$.
- The surrogate variable X given $Y = y$ is normally distributed with the mean $\lambda_1 + \lambda_2 y$ and variance σ^2, where λ_1 and λ_2 are known constants. λ_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 $(\lambda_1 + \lambda_2\mu_y, \mu_y)$, variances $(\lambda_2\sigma_y^2 + \sigma^2, \sigma_y^2)$, and correlation coefficient $\rho = \lambda_2\sigma_y/(\lambda_2\sigma_y^2 + \sigma^2)^{1/2}$ (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_1''. The time taken to identify a false alarm is b_2, and the time required to take and interpret a sample is b_3.
- 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''. 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:

$$\begin{gathered}\left(\mu_0 - k_j\frac{\sigma_{\overline{X}}}{\sqrt{n}}, \mu_0 - k_{j+1}\frac{\sigma_{\overline{X}}}{\sqrt{n}}\right)\\ \cup\left(\mu_0 + k_{j+1}\frac{\sigma_{\overline{X}}}{\sqrt{n}}, \mu_0 + k_j\frac{\sigma_{\overline{X}}}{\sqrt{n}}\right)\\ \text{for} \quad j = 1, 2, \ldots, m-1,\\ \left(\mu_0 - k_j\frac{\sigma_{\overline{X}}}{\sqrt{n}}, \mu_0 + k_j\frac{\sigma_{\overline{X}}}{\sqrt{n}}\right) \quad \text{for} \quad j = m,\end{gathered} \tag{1}$$

where

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

$$\sigma_{\overline{X}} = \sigma_x/\sqrt{n_x}.$$

If we define $Z_{\overline{X}} = \sqrt{n_x}(\overline{X} - \mu_x)/\sigma_x$, 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 ($j >= 2$).

Step 2. If $|Z_{\overline{X}}| < k_2$ (if sampling falls in the second outermost control lines), go to step 1. Otherwise, go to step 3.

Step 3. If $|Z_{\overline{X}}| < k_1$ (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 $\overline{X}$ control chart using a surrogate variable is a renewal reward process, the long-run expected net income per unit of time is given by

$$E(A) = \lim_{t\to\infty} E\left[TI(t)\right]/t = E(I)/E(T), \tag{2}$$

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 $\overline{X}$ 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.

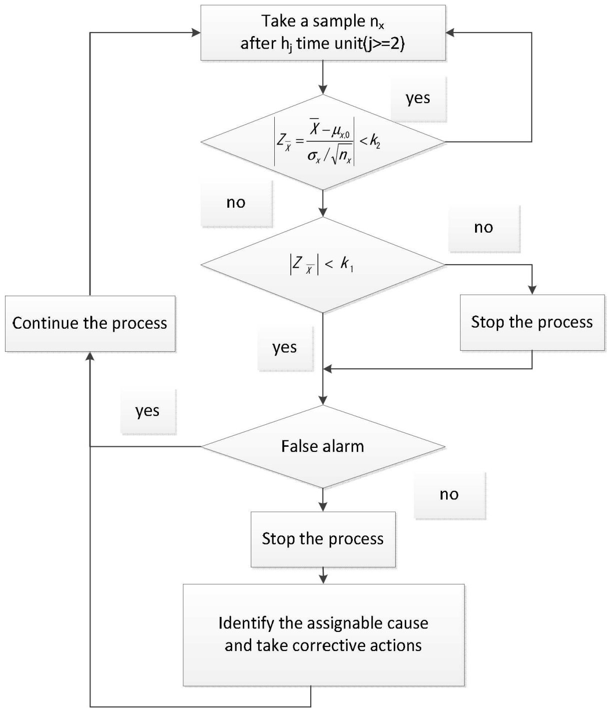

Fig. 1. The monitoring procedures.

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 (ARL_L) and an upper bound on the out-of-control ARL (ARL_U), as in Montgomery et al. [13]. A model for the economic statistical design can be formulated as follows:

$$\begin{aligned} &\text{Maximize} && \mathrm{E}\ (\mathrm{A})\\ &\text{Subject to} && \mathrm{ARL}_0 > \mathrm{ARL}_L\\ & && \mathrm{ARL}_1 < \mathrm{ARL}_U,\end{aligned} \tag{3}$$

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 $q_s^{(0)}$ and $q_s^{(1)}$, respectively. The constraints in (3) can be equivalently expressed as follows:

$$k_1 > \Phi^{-1}\left(\frac{1}{1 - 1/2\mathrm{ARL}_L}\right), \tag{4}$$

$$\frac{1}{(1-\Phi(k_1-\lambda_2 c\frac{\sigma_y}{\sigma_x}\sqrt{n_x})+\Phi(-k_1-\lambda_2 c\frac{\sigma_y}{\sigma_x}\sqrt{n_x})} < \text{ARL}_U, \quad (5)$$

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.

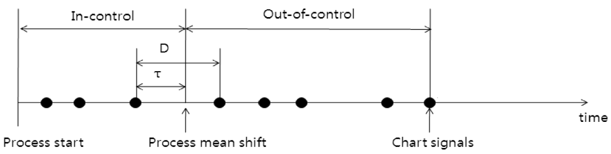

Fig. 2. The process cycle.

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

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

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 $\overline{X}$ 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 τ, 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], τ is well approximated as

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

(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

$$E(R_0)=q_s^{(0)}\frac{\sum\limits_{j=1}^{m} q_j^{(0)}e^{-\lambda h_j}}{\left(1-q_s^{(0)}-\sum\limits_{k=1}^{m} q_k^{(0)}e^{-\lambda h_k}\right)^2} \cdot\sum_{l=1}^{\eta} q_l^{(0)}(1-e^{\lambda h_j}). \quad (10)$$

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,

$$E(FA) = b_2 q_s^{(0)} \frac{\sum_{j=1}^{m} q_j^{(0)} e^{-\lambda h_j}}{\left(1 - q_s^{(0)} - \sum_{k=1}^{m} q_k^{(0)} e^{-h_k}\right)^2} \cdot \sum_{l=1}^{m} q_l^{(0)} (1 - e^{-\lambda h_l}). \tag{11}$$

In this model, we investigate the process state by performing the process if $\overline{X}$ 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, $b_1' + b_1''$ is taken to find and eliminate an assignable cause. The expected duration for finding and eliminating an assignable cause is therefore

$$E(AC) = b_3 n_x + b_1' + b_1''. \tag{12}$$

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

$$E(T) = E(IC) + E(OC) + E(FA) + E(AC). \tag{13}$$

Expected net income

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

$$E(I) = E(I') - E(C_1) - E(C_2), \tag{14}$$

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/λ, 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

$$E(I') = i_1 * E(IC) + i_2 * E(OC) = \frac{i_1}{\lambda} + i_2 \left(\sum_{j=2}^{m} h_j \left(\frac{1}{1 - r_2^{(1)}} - 1 \right) \cdot \frac{q_j^{(1)}}{1 - r_2^{(1)}} + E(D) - \tau \right). \tag{15}$$

(2) Expected sampling cost per cycle

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

$$SN_0 = \frac{\sum_{j=1}^{m} q_j^{(0)} e^{-\lambda h_j}}{\left(1 - q_s^{(0)} - \sum_{k=1}^{m} q_k^{(0)} e^{-\lambda h_k}\right)^2} \cdot \sum_{l=1}^{\eta} q_l^{(0)} (1 - e^{-\lambda h_j}). \tag{16}$$

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

$$SN_1 = \frac{1}{\left(1 - r_2^{(1)}\right)}. \tag{17}$$

Therefore, the expected sampling cost per cycle is given by

$$E(C_1) = (a_3 + a_4 n_x)(SN + SN_1) = (a_3 + a_4 n_x) \left(\frac{\sum_{j=1}^{m} q_j^{(0)} e^{-\lambda h_j}}{\left(1 - q_s^{(0)} - \sum_{k=1}^{m} q_k^{(0)} e^{-\lambda h_k}\right)^2} \cdot \sum_{l=1}^{\eta} q_l^{(0)} (1 - e^{-\lambda h_j}) + \frac{1}{(1 - r_2^{(1)})} \right). \tag{18}$$

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

The expected cost incurred from false alarms is given by

$$(a_2' + a_2'')\, E(R_0) + (a_2') E(R_1), \tag{19}$$

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

$$E(C_2) = (a_2' + a_2'')\, E(R_0) + (a_2')\, E(R_1) + \frac{q_s^{(1)}}{r_2^{(1)}} a_1 + \frac{q_1^{(1)}}{r_1^{(1)}} a_1 = (a_2' + a_2'')\, E(R_0) + (a_2')\, E(R_1) + a_1, \tag{20}$$

where

$$E(R_1) = q_1^{(0)} \frac{\sum_{j=1}^{m} q_j^{(0)} e^{-\lambda h_j}}{\left(1 - q_s^{(0)} - \sum_{k=1}^{m} q_k^{(0)} e^{-\lambda h_k}\right)^2} \cdot \sum_{l=1}^{\eta} q_l^{(0)} (1 - e^{\lambda h_j}).$$

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 ρ 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, comb ined 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 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 1
Cost and process parameters for the test examples.

Example	Λ	c	i_1	i_2	a_1	a_2'	a_2''	a_3	a_4	b_1'	b_1''	b_2	b_3
1	0.01	1	50		45	25	25	0.5	0.1	1.53	1.53	4.05	0.05
2	0.01	1	150	50	350	250	250	5.0	1.0	1.53	1.53	4.05	0.05
3	0.01	2	50	1	260	25	25	5.0	1.0	2.00	2.00	41.00	0.05
4	0.01	2	150	50	135	250	250	0.5	0.1	2.00	2.00	41.00	0.05
5	0.05	1	150	100	45	250	250	5.0	0.1	2.00	2.00	41.00	0.05
6	0.05	1	50	−50	350	25	25	0.5	1.0	2.00	2.00	41.00	0.05
7	0.05	2	150	100	260	250	250	0.5	1.0	1.53	1.53	4.05	0.05
8	0.05	2	50	−50	135	25	25	5.0	0.1	1.53	1.53	4.05	0.05
9	0.01	1	50		135	250	250	0.5	1.0	2.00	2.00	5.00	0.50
10	0.01	1	150	50	260	25	25	5.0	0.1	2.00	2.00	5.00	0.50

Table 2
Optimum designs for the test examples.

Example	The PHM Model					Proposed Economic Model						Economic Statistical Model					
	n_y	h	k	$E(A)$	ρ	n_x	h_1	h_2	k_2	k_1	$E(A)$	n_x	h_1	h_2	k_2	k_1	$E(A)$
1	17	2.75	3.14	45.91	0.6	3	1.2260	1.5702	3.4577	4.7195	47.3098	No optimal solutions to satisfy the statistical constraints					
					0.9	3	1.1677	1.4638	4.5006	5.0000	47.3190	15	2.3632	2.6993	4.0520	4.1177	46.8807
2	17	6.33	2.95	134.11	0.6	5	4.9267	5.3084	3.1699	3.5047	138.3531	No optimal solutions to satisfy the statistical constraints					
					0.9	3	3.0265	3.3633	4.4980	4.9346	138.9331	10	4.0523	4.5867	3.5890	3.5894	138.1326
3	6	6.46	3.46	42.08	0.6	2	4.5738	4.9087	3.9109	5.0000	43.4797	14	4.7446	5.4723	4.9185	5.0000	43.0161
					0.9	1	4.5467	4.8543	4.2178	4.9999	43.5203	5	4.3012	5.0266	4.7329	4.7967	43.3611
4	8	1.54	4.31	140.89	0.6	3	2.8098	3.0576	3.4878	4.5272	155.2805	14	0.8120	1.1877	4.9606	4.9822	140.9394
					0.9	1	0.7591	1.2287	3.8348	4.8121	156.0762	3	2.5539	3.1589	3.8639	3.8673	140.9695
5	26	2.41	3.76	117.73	0.6	2	2.7464	3.0693	3.3833	5.0000	143.7728	No optimal solutions to satisfy the statistical constraints					
					0.9	1	2.7759	3.0776	4.6407	4.9998	144.1059	9	4.4247	4.6776	3.4390	3.4398	117.6276
6	12	2.48	2.75	14.04	0.6	9	0.1338	0.5577	3.3537	4.9849	23.6282	No optimal solutions to satisfy the statistical constraints					
					0.9	4	1.1914	1.4230	3.3796	4.9733	24.1031	10	5.21824	5.4498	1.6464	3.3440	17.9919
7	5	2.01	3.32	114.93	0.6	3	5.0897	5.3126	2.6068	3.5264	127.2332	14	1.6124	2.2941	5.0000	5.0000	113.8980
					0.9	4	2.6827	3.0249	4.5342	4.6057	128.0696	2	4.4780	5.0428	2.8505	3.1483	116.0133
8	6	1.50	3.12	30.04	0.6	2	1.2096	1.5477	2.9989	4.9942	31.6383	14	2.9391	3.1433	2.7453	4.4531	29.4865
					0.9	1	1.0666	1.5449	4.1875	4.5799	31.7448	9	1.3453	1.6045	2.9197	4.5305	31.1335
9	9	5.03	2.72	39.96	0.6	8	1.3102	1.5591	3.7949	4.6062	43.9274	No optimal solutions to satisfy the statistical constraints					
					0.9	3	1.3313	1.6128	3.7694	4.9975	45.2688	10	4.0351	4.4225	3.5884	3.5893	43.3801
1	10	2.79	2.90	132.08	0.6	2	3.4543	3.8727	2.9144	4.7249	137.4769	No optimal solutions to satisfy the statistical constraints					
					0.9	2	3.1157	3.3624	3.8695	4.9988	137.5249	9	4.8482	5.1698	3.4006	3.4398	132.8578

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 $\sigma_y^2 = (32.18\ \mu\mathrm{m})^2$, respectively, and the variance of X is $\sigma_x^2 = (34.75\ \mu\mathrm{m})^2$. 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 $(12.57\ \mu\mathrm{m})^2$ 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 $\lambda = 0.05$, $c = 2$, $i_1 = 100$, $i_2 = 0$, $a_1 = 200$, $a_2' = 200$, $a_2'' = 200$, $a_3 = 3$, $a_4 = 1.5$, $b_1' = 10$, $b_1'' = 10$, $b_2 = 25$, 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_U and ARL_L.

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

Obs	PSA	X-rays	Obs	PSA	X-rays	Obs	PSA	X-rays
1	1081	1047	18	1105	1047	35	1104	1042
2	1156	1116	19	1088	1027	36	1023	979
3	1094	1046	20	1073	1019	37	1129	1090
4	1084	1037	21	1102	1066	38	1116	1063
5	1072	1021	22	1090	1041	39	1192	1138
6	1133	1089	23	1122	1066	40	1169	1123
7	1116	1078	24	1061	1024	41	1093	1048
8	1091	988	25	1090	1050	42	1116	1083
9	1029	987	26	1117	1075	43	1106	1057
10	1114	1086	27	1082	1029	44	1105	1037
11	1119	1074	28	1069	1007	45	1103	1068
12	1082	1049	29	1066	1022	46	1116	1090
13	1118	1070	30	1070	1039	47	1063	1011
14	1113	1071	31	1043	1009	48	1111	1001
15	1100	1063	32	1104	1048	49	1158	1005
16	1126	1072	33	1060	999	50	1104	1052
17	1083	1039	34	1105	1064			

Table 4
Results of the VSI $\overline{X}$ control chart using surrogate variables.

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

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 λ and c: Fig. 3 shows that the optimum values of h_1 and h_2 decrease as the failure rate (λ) increases. This conforms to our intuition that the process needs to be controlled more tightly because it tends to fail more frequently as λ 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.

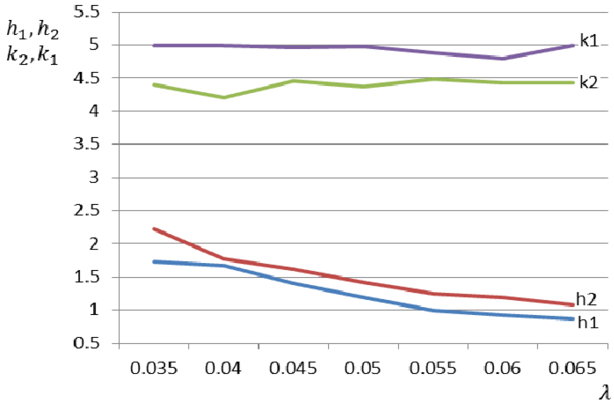

Fig. 3. Sensitivity of h_1, h_2, k_2, and k_1 to λ.

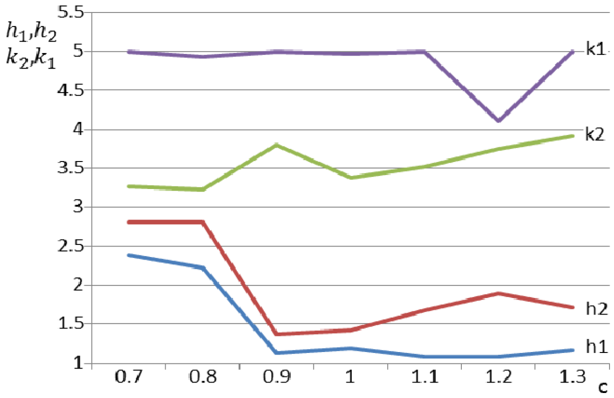

Fig. 4. Sensitivity of h_1, h_2, k_2, and k_1 to c.

- 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.

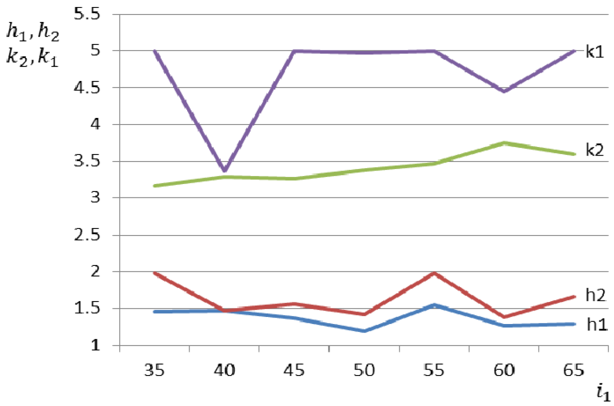

Fig. 5. Sensitivity of h_1, h_2, k_2, and k_1 to i_1.

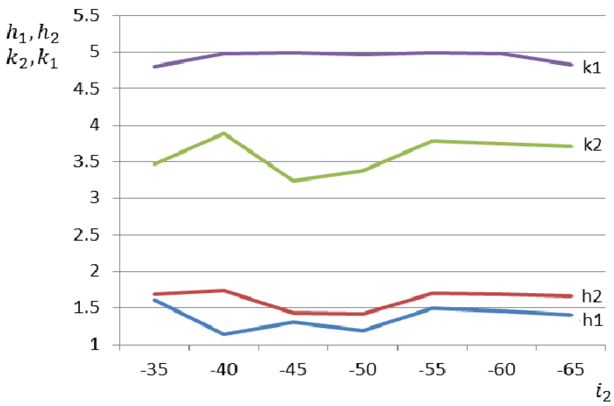

Fig. 6. Sensitivity of h_1, h_2, k_2, and k_1 to i_2.

- 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.

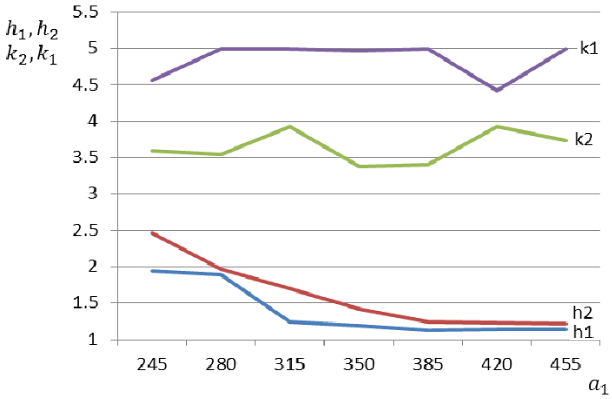

Fig. 7. Sensitivity of h_1, h_2, k_2, and k_1 to a_1.

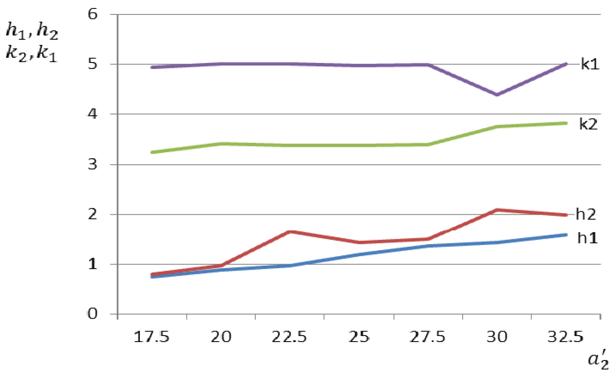

Fig. 8. Sensitivity of h_1, h_2, k_2, and k_1 to a_2'.

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 $\overline{X}$ 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 $\overline{X}$ 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 $\overline{X}$ control chart, EWMA, and CUSUM charts.

References

[1] Reynolds Jr. M.R., Amin R.W., Arnold J.C., Nachlas J.A., *$\overline{X}$ charts with variable sampling intervals*, Technometrics, 30, 181–192, 1988.

[2] Reynolds Jr. M.R., Arnold J.C., *Optimal one-sided Shewhart control charts with variable sampling intervals*, Sequential Analysis, 8, 51–77, 1989.

[3] Runger G.C., Montgomery D.C., *Adaptive sampling enhancements for Shewhart control charts*, IIE Transactions, 25, 41–51, 1993.

[4] Reynolds Jr. M.R., *Shewhart and EWMA variable sampling interval control chats with sampling at fixed times*, Journal of Quality Technology, 28, 199–212, 1996.

[5] Costa A.F.B., Rahim M.A., *Economic design of $\overline{X}$ charts with variable parameters: the Markov chain approach*, Journal of Applied Statistics, 28, 875–885, 2001.

[6] Lin Sung-Nung, Chou Chao-Yu, Wang Shu-Ling, Liu Hui-Rong, *Economic design of autoregressive moving average control chart using genetic algorithms*, Experts System with Application, 39, 1793–1798, 2012.

[7] Chew X.Y., Khoo Michael B.C., Teh S.Y., Castagliola P., *The variable sampling interval run sum $\overline{X}$ control chart*, Computer & Industrial Engineering, 90, 25–38, 2015.

[8] Zhang Min, Nie Guohua, He Zhen, *Performance of cumulative count of conforming chart of variable sampling intervals with estimated control limit*, International Journal of Production Economics, 150, 114–124, 2014.

[9] Lee T.H., Lee J.H., Lee M.K., Lee J.H., *Economic design of $\overline{X}$ control chart using a surrogate variable*, Journal of the Korean Society for Quality Management, 37, 46–57, 2009.

[10] Lee T.H., Lee J.H., Lee M.K., *Economic design of variable sampling interval control chart using a surrogate variable*, Journal of Korean Institute of Industrial Engineers, 39, 422–428, 2013.

[11] Panagos M.R., Heikes R.G., Montgomery D.C., *Economic design of $\overline{X}$ control charts for two manufacturing process models*, Naval Research Logistics, 32, 631–646, 1985.

[12] Tang K., Lo J., *Determination for the optimal process mean when inspection is based on a correlated variable*, IIE Transactions, 25, 66–7, 1993.

[13] Montgomery D.C., James C.C., Torng J.K., Frederick P.L., *Statistically Constrained Economic Design of the EWMA Control Chart*, Journal of Quality Technology, 27, 250–256, 1995.

[14] Duncan A.J., *The economic design of $\overline{X}$ charts used to maintain current control of a process*, Journal of American Statistical Association, 51, 228–242, 1956.

[15] Bai D.S., Lee K.T., *An economic design of variable sampling interval $\overline{X}$ control charts*, International Journal of Production Economics, 54, 57–64, 1998.

[16] Kim W.K, Lee Y.W., Park J.Y., Park J.B., Ra S.W., *Nondestructive measurement of the coating thickness in the simulated TRISO-Coated fuel particle using micro-focus X-Ray radiography*, Journal of the Korean Society for Nondestructive Testing, 26, 69–76, 2006.