

ECONOMIC DESIGN OF MEWMA VSSI CONTROL CHARTS FOR MULTIATTRIBUTE PROCESSES

Seyed Taghi Akhavan Niaki¹ and Paravaneh Jahani²

¹ Sharif University of Technology, P.O. Box 11155-9414 Azadi Ave., Tehran 1458889694, Iran

² Department of Industrial Engineering, Sharif University of Technology, Tehran, Iran

Keywords: Multiattribute processes, Economic design, Multivariate exponentially weighted moving average chart, Variable Sample Size, Variable Sampling Interval, Genetic Algorithm.

Abstract: In this research, a new methodology is developed to economically design a multivariate exponentially weighted moving average (MEWMA) control chart for multiattribute processes. The optimum design parameters of the chart, i.e., the sample size, the sampling interval, and the warning/action limit coefficients, are obtained using a genetic algorithm to minimize the expected total cost per hour. A sensitivity analysis has also been carried out to investigate the effects of the cost and model parameters on the solutions obtained.

1 INTRODUCTION

In many real-world manufacturing environments the quality of products are of multiattribute type, where multiattribute control charting methods are recommended to deal with the existing correlations between the attributes. Although there are many applications for multiattribute control charts in industries and service sectors, there exists only a little research on this type of control chart in the literature. To name a few, Jolayemi (2000) proposed a multiattribute control chart based on both the J-approximation and the Gibra's model to monitor processes following multivariate binomial distribution in which there were multiple assignable causes. In a more recent research in this area, Niaki and Abbasi (2008) introduced a new method to monitor multiattribute processes and developed a multi-attribute C control chart, where a transformation was first proposed to eliminate the correlation between the attributes, and then the symmetric control limits were found.

In the designing process of a control chart, three parameters are involved; the sample size, the sampling interval, and the control limit coefficient. Economic and/or statistical designs are the two common practices in this regards. In a statistical design, the design parameters are determined based on the statistical performances of the chart. These performances are measured either in terms of type-I

and II errors or in terms of average run lengths (ARL) or average time to signal (ATS). Meanwhile, in an economic design, the design parameters are selected based on minimizing a cost model or a loss function.

Duncan (1956) proposed the first economic design of the X-bar chart to show how cost factors affect the optimality. Moreover, the Lorenzen-Vance (1986) cost model is a widely used function in determining the costs of implementing a control chart.

Following the investigation of the variable sampling interval (VSI) EWMA chart by Saccucci et al. (1992), Reynold and Arnold (2001) developed the variable sample size EWMA (VSSI EWMA) control chart to improve the performance of the fixed sample size charts on the speed of detecting small changes in the mean vector.

For the first time a methodology based on the skewness reduction approach (Niaki and Abbasi, 2008) and Lorenzen and Vance (1986) cost function is developed in this paper to economically design a multiattribute VSSI MEWMA control chart.

2 THE VSSI MEWMA CHART

Assuming vector \bar{X}_i follows a p -variate normal distribution with mean vector μ_0 and covariance

matrix Σ_0 , in a sample of size n the sample mean \overline{X}_i follows a multivariate normal distribution with mean vector μ_0 and covariance matrix Σ_0/n . In the MEWMA control chart, the vector \overline{Z}_i is first expressed by:

$$\overline{Z}_i = \lambda \overline{X}_i + (1-\lambda)\overline{Z}_{i-1} \quad ; \quad i = 1, 2, \dots \quad (1)$$

where $0 \leq \lambda \leq 1$ is the smoothing parameter and $\overline{Z}_0 = \overline{\mu}_0$. Then, the plotted values on the MEWMA chart are:

$$T_i^2 = \overline{Z}_i' \Sigma_{Z_i}^{-1} \overline{Z}_i \quad (2)$$

in which Σ_{Z_i} and $\Sigma_{Z_i}^{-1}$ are the covariance matrix and the inverse of the covariance matrix of \overline{Z}_i , respectively.

Having a warning limit w , if the last sample statistic, T_{i-1}^2 , falls in the safe region ($0 \leq T_{i-1}^2 \leq w$), the next sample is taken using the minimum sample size n_1 and the long sampling interval h_1 . However, if T_{i-1}^2 falls in the warning region ($w \leq T_{i-1}^2 \leq l$), where l is the control limit, then the next sample is taken using the maximum sample size n_2 and the short sampling interval h_2 . Nonetheless, searching for an assignable cause is started when T_{i-1}^2 falls above the out-of-control limit ($T_{i-1}^2 \geq l$).

For a sample size of n , Lowry et al. (1992) obtained the asymptotic covariance matrix of \overline{Z}_i , Σ_{Z_i} , as:

$$\Sigma_{Z_i} = \lim_{i \rightarrow \infty} \Sigma_{Z_i} = \left(\frac{\lambda}{2-\lambda} \right) \Sigma_0 \quad (3)$$

They also showed that $\lambda = 0.1$ is effective in detecting small shift in the process mean vector and that the ARL performance of the MEWMA chart depends only on the noncentrality parameter, $\gamma = n\delta^2$, and the direction of the shift, where

$$\delta = \sqrt{(\mu_i - \mu_0)' \Sigma_0^{-1} (\mu_i - \mu_0)} \quad (4)$$

and Σ_0^{-1} is the inverse of the covariance matrix of \overline{X}_i . The smoothing parameter of the MEWMA chart of this research is taken 0.1 as well.

3 NORMALIZING TRANSFORMATION ON MULTIATTRIBUTE PROCESS DATA

Suppose the observations of the p -variate multiattribute process under consideration $\overline{Y} = [y_1, y_2, \dots, y_p]^T$; $i = 1, 2, \dots, p$ follows a multivariate binomial distribution with the parameters N_i, Pr_i , and Σ_b where the mean vector is $\overline{V} = [v_1, v_2, \dots, v_p]^T$; $v_i = N_i Pr_i$ and the covariance matrix is Σ_b , in which N_i represents the fixed sample size and Pr_i is the proportion non-conforming of the i th attribute. In order to develop a procedure to monitor this process, the inherent skewness of the data, the most important factor in the non-normality of the multiattribute process, is almost removed by employing the r^{th} root transformation method proposed by Niaki and Abbasi (2008) to the observations taken from the process at hand. Despite the fact that the zero-skewness is necessary but not sufficient condition for normality, we assume that the transformed observation vector will follow a normal distribution. To do the transformation, the proposed bisection method of Niaki and Abbasi (2008) is employed using 5000 simulated observation vector to find the root vector $\vec{r} = [r_1, r_2, \dots, r_p]^T$ in a way that the skewness of each attribute becomes almost zero. In other words, the powers in the vector $\overline{Y} = [y_1^{r_1}, y_2^{r_2}, \dots, y_p^{r_p}]^T$ are found so that the skewness of $y_i^{r_i}$; $i = 1, 2, \dots, p$ is close to zero. Having $x_i = y_i^{r_i}$, the transformed p -variate almost normal observation vector $\overline{X} = [x_1, x_2, \dots, x_p]^T$ will have a new mean vector $\overline{\mu}_0$ and covariance matrix Σ_0 . Then, to monitor the process an existing VSSI MEWMA control chart can be employed. This chart is introduced in the next Section.

4 ECONOMIC DESIGN OF THE VSSI MEWMA CHART

Since the sampling interval in the VSSI MEWMA chart is not fixed, it would be more appropriate to use average time to signal (ATS), instead of ARL

(Lin and Chou, 2005a, 2005b). To do this, Chou et al. (2006) modified the Lorenzen-Vance cost model in optimizing the expected total quality cost of the VSI EWMA chart. This approach can be applied to different type of control charts such as the VSSI MEWMA chart of this research.

The average sampling interval is expressed as:

$$h_0 = \left(\frac{p_1}{p_1 + p_2}\right)h_1 + \left(\frac{p_2}{p_1 + p_2}\right)h_2 \quad (5)$$

where p_1 and p_2 are respectively the probabilities that the T_i^2 falls in the safe (state 1) and warning (state 2) regions. The average time to signal when the process is in state i (ATS_i ; $i = 1, 2$) and the average time to signal when the process is out-of-control (ATS_2) can be obtained by:

$$ATS_1 = h_i(ARL_0) \quad (6)$$

$$ATS_2 = h_0(ARL_1) \quad (7)$$

In this research, the simulation method is employed to estimate ARL_0 , ARL_1 , p_1 and p_2 . As a result, the expected total cost per hour based on the Lorenzen and Vance (1986) cost model is defined as:

$$C = \sum_{i=1}^2 \frac{p_i}{p_1 + p_2} \left\{ C_1 / \theta + C_2 (-\tau + nd + ATS_2 + \gamma_1 T_1 + \gamma_2 T_2) \right\} \times \left\{ +sg / (ATS_i / h_i) + m \right\} + \left\{ 1 / \theta + (1 - \gamma_i) s T_i / (ATS_i / h_i) - \tau + nd + ATS_2 \right\}^{-1} + T_1 + T_2 \quad (8)$$

$$\left(\frac{a + bn}{h_0} \right) \left\{ 1 / \theta - \tau + nd + ATS_2 + \gamma_1 T_1 + \gamma_2 T_2 \right\} \times \left\{ 1 / \theta + (1 - \gamma_i) s T_i / (ATS_i / h_i) - \tau + nd + \right\}^{-1} + ATS_2 + T_1 + T_2$$

where

a : fixed cost per sample

b : cost per unit sampled

C_1 : quality cost per hour while process is in-control

C_2 : quality cost per hour while process is out-of-control ($> C_1$)

d : time to sample and chart one item

g : cost per false alarm

m : cost to locate and correct the assignable cause

s : expected number of samples taken while

the process is in-control $\left(= \frac{e^{-\theta h_0}}{1 - e^{-\theta h_0}} \right)$

$1/\theta$: mean time the process is in-control

τ : expected time of occurrence of assignable cause between two samples, while the process is in-control $\left(= \frac{1 - (1 + \theta h_0) e^{-\theta h_0}}{\theta (1 - e^{-\theta h_0})} \right)$

T_0 : expected search time when the signal is false alarm

T_1 : expected time to discover the assignable cause

T_2 : expected time to correct the process

$\gamma_1 = \begin{cases} 1 & ; \text{if production continues during the searches} \\ 0 & ; \text{if production ceases during searches} \end{cases}$

$\gamma_2 = \begin{cases} 1 & ; \text{if production continues during the correction} \\ 0 & ; \text{if production ceases during correction} \end{cases}$

Further, it is assumed the time between occurrences of the assignable cause is exponential with a mean of θ occurrence per hour.

Since minimizing the non-linear cost function in (8) is not straightforward and both non-linear programming techniques and traditional optimization approaches may be time consuming and inefficient, a genetic algorithm (GA) is proposed in the next Section to solve it. The objective is to find the optimal values of n_1, n_2, h_1, h_2, w , and l that minimize the expected total cost per hour given in (8).

5 THE SOLUTION METHODOLOGY

Genetic algorithm provides a convenient search procedure for minimizing the complex cost function such as (8) to provide a list of optimum design variables, without focusing on a single solution. With respect to the broad spectrum of GA parameters in relevant studies, a trial and error approach is taken in this research to improve them in generating the best results. The basic steps involved in the proposed GA of this research follow.

1. *Initialization*: The proposed GA of this research starts with an initial population of 50 chromosomes, each containing 6 genes (the design parameters) defined as $(n_1, n_2, h_1, h_2, w, l)$.

2. *Evaluation*: The fitness (the cost) associated with each chromosome is evaluated using Eq. (8).

3. *Selection*: The best 30% of the chromosomes with the lowest cost (or the highest

fitness) are selected based on the "elitism" strategy (0.3×50=15 is the number of the elites.)

4. *Crossover*: In the crossover operation that is used with a rate of 60% of the population (0.6×50=30 is the number of the parents) three chromosomes are first selected randomly. The chromosome with the lowest cost is the first parent. Then, three more chromosomes are selected, where the one with the lowest cost represents the second parent. This pair of parent chromosomes takes part in the crossover operation, where each gene of the paired parents has a 50% probability to switch between two chromosomes.

5. *Mutation*: Mutation points are randomly chosen with a rate of 10% (0.1×50=5 is the number of the muted chromosomes). The uniform selection provides the constant probability for each chromosome to enter this mechanism. Similar to the crossover operation, three chromosomes are selected randomly and the best fitness value chromosome is muted. Each gene of this chromosome has a 50% chances to mute by the mutation function that is defined as follows.

Muted gene = gene + 0.1 × (a uniform random number between 0 and 1) × (range)

As an example, suppose the third gene (h_1) varies in the range of $3 \leq h_1 \leq 5$ and has the value of 4.351. Further, let the generated random number be 0.8147. Then, Muted $h_1 = 4.351 + 0.1 \times 0.8147 \times (5-3) = 4.51394$

6. *Stopping criterion*: After the mutation operation, the process described above is iterated until the termination condition is achieved. The stopping criterion of this research is to iterate 30 generations.

In the next Section, a simulation experiment on a process involving $p = 2$ attributes is given to demonstrate the application of the proposed methodology. The simulation experiment is based on the generated observations of a 2-variate binomial distribution and then optimization through GA using MATLAB, version R2009b. The elapsed time of each run varies between 22 and 41 minutes, depending on the starting parameters' values.

5.1 Simulation Experiment

In the simulation experiment, a process with two correlated attributes following a multivariate binomial distribution with the parameters ($N_1=20, p_1=0.2$) and ($N_2=30, p_2=0.15$) and correlation coefficient $\bar{\rho} = 0.02$ is considered. The mean vector

$\overline{\mu}_{0_{old}}$, the covariance matrix $\overline{\Sigma}_{0_{old}}$, and the skewness $\overline{skewness}_{old}$ of 5000 in-control simulation data on this process is estimated as $\overline{\mu}_{0_{old}} = [4.258, 4.5074]$, $\overline{\Sigma}_{0_{old}} = \begin{bmatrix} 3.273 & 0.0649 \\ 0.0649 & 3.8191 \end{bmatrix}$, $\overline{skewness}_{old} = [3.3155 \ 0.3894]$, respectively.

Following the described skewness reduction method in Section 3, the process mean $\overline{\mu}_0$, the covariance matrix $\overline{\Sigma}_0$, and the skewness of 5000 in-control generated observations of the process that are transformed by the root vector \overline{r} are obtained as $\overline{\mu}_0 = [2.9662, 2.9709]$, $\overline{\Sigma}_0 = \begin{bmatrix} 1.2083 & 0.0202 \\ 0.0202 & 0.9726 \end{bmatrix}$, $\overline{skewness} = [-0.0006, 0.0010]$, and $\overline{r} = [0.7942, 0.7366]$. It can be easily seen that the transformed attributes have almost zero skewness.

For an out-of-control process, let the mean of the first attribute increase by $\delta\sigma$. Hence, the observation vector of the out-of-control process follows a multivariate binomial distribution with parameters ($N_1=20, p_1'$) and ($N_2=30, p_2=0.15$) in which p_1' are derived by (9).

$$n_1 p_1' = n_1 p_1 + \delta\sigma = n_1 p_1 + \delta\sqrt{n_1 p_1 (1-p_1)} \rightarrow$$

$$p_1' = \frac{n_1 p_1 + \delta\sqrt{n_1 p_1 (1-p_1)}}{n_1} \tag{9}$$

6 SENSITIVITY ANALYSIS

A sensitivity analysis is carried out in this Section to study the effects of the cost parameters on the solution of the proposed economic design of VSSI MEWMA chart. Table (1) shows twelve cost and process parameters for the cost model detailed in equation (8). By combining the sampling cost parameters into a single parameter $Q = a + b$ and the expected times to search and discover and correct assignable cause into a single parameter $T = T_0 + T_1 + T_2$, the number of factors used in the experiment are reduced to nine. The sensitivity analysis is conducted using a 2^{9-4} fractional factorial design with five center points.

Table 1: The cost and process parameter ranges.

	a	b	C_1	C_2	g	m	d	T_0	T_1	T_2	θ	δ
low	25	5	100	250	50	25	0.05	0	16	16	0.03	0.5
center	137.5	27.5	150	375	275	137.5	0.275	1	18	18	0.04	0.75
high	250	50	200	500	500	250	0.5	2	20	20	0.05	1

Table 2: Experimental cost and process parameter values.

Run	a	b	C_1	C_2	g	m	d	T_0	T_1	T_2	θ	δ
1	25	5	100	250	50	25	0.5	2	20	20	0.05	1
2	25	5	100	250	50	250	0.05	0	16	16	0.03	0.5
3	25	5	100	250	500	25	0.05	0	16	16	0.03	1
4	25	5	100	250	500	250	0.5	2	20	20	0.05	0.5
5	25	5	100	500	50	25	0.05	0	16	16	0.05	0.5
...
33	137.5	27.5	150	375	275	137.5	0.275	1	18	18	0.04	0.75
34	137.5	27.5	150	375	275	137.5	0.275	1	18	18	0.04	0.75
35	137.5	27.5	150	375	275	137.5	0.275	1	18	18	0.04	0.75
36	137.5	27.5	150	375	275	137.5	0.275	1	18	18	0.04	0.75
37	137.5	27.5	150	375	275	137.5	0.275	1	18	18	0.04	0.75

Table 3: The results of experiments.

Run	n_1	n_2	h_1	h_2	w	l	Cost
1	5.01	11.828	4.719	1.688	6.422	14.007	208.697
2	7.29	10.377	2.594	0.819	6.454	13.856	175.809
3	7.05	10.526	2.161	1.420	7.234	13.603	143.436
4	5.27	14.155	3.963	0.602	6.810	16.249	225.998
5	7.19	10.794	2.648	0.565	6.553	13.739	382.654
...
33	5.59	11.782	4.998	2.001	7.147	15.044	346.350
34	6.78	10.177	4.998	1.947	6.342	13.954	399.155
35	6.01	12.658	4.692	1.403	7.028	14.284	351.331
36	5.90	10.716	4.768	1.558	7.925	13.676	351.194
37	5.91	14.089	4.696	1.940	6.043	13.505	348.986

Table (2) shows the experimental design with actual values used for the cost and process parameters for 32+5=37 runs. Five replicates at the design center (shown at the bottom of the table) are used to check the adequacy of the first order model (Montgomery, 1996). The final objective of fitting an adequate estimation function for the cost model is to determine the optimum values of the parameters. The independent variables of the estimation function are considered the cost parameters along with the six design parameters (n_1, n_2, h_1, h_2, w, l). The expected total cost per hour is treated as the dependent variable.

The results of the 37 experiments for the $p = 2$ attributes process are shown in Table (3), where for each cost parameter combination the proposed GA has been executed to find the near optimal values of the design parameters.

Since a single replication of the 2^{9-4} fractional factorial design was run, in order to have an estimate of the variance of the error term some interaction

terms that do not have significant effect may be pooled. A simple approach to determine these effects is to provide a normal probability plot of the effects. The points lying close to the straight line in a reasonable way indicate non-significant effects (with mean zero) (Montgomery, 1996). The normal probability plot (not shown here) indicates all interaction terms do not have significant effect and can be pooled in the error term. Moreover, the plot reveals that there is no acute symptom of nonnormality, nor is there any indication pointing to possible outliers.

The results of the analysis of variance (ANOVA) are given in Table (4) to investigate the effects of multiple model parameters on the expected total cost per hour. The results indicate that while Q, C_1, C_2, θ , and δ have significant effects, the effect of T on the expected total cost per hour is not significant. Moreover, the results show that the response surface has no significant curvature and a first order model is appropriate to estimate the

Table 4: ANOVA table for the expected total cost per hour

Resource	Sum sq.	d.f	Mean sq.	F	Prob>F
$Q(=a+b)$	136478.6	1	136478.6	197.12	< 0
C_1	18171.3	1	18171.3	26.25	< 0
C_2	128410.2	1	128410.2	185.47	< 0
g	16.1	1	16.1	0.02	0.8801
m	91.1	1	91.1	0.13	0.7197
d	17	1	17	0.02	0.8768
$T(=T_0+T_1+T_2)$	2417.1	1	2417.1	3.49	0.073
θ	27623.9	1	27623.9	39.9	< 0
δ	5432.2	1	5432.2	7.85	< 0.0095
Pure quadratic	1968.738	1	1968.738	2.84	0.1039
Error	18001.2	26	692.4		
Total	338627.4	36			

expected cost per hour in terms of the significant factors.

Based on the significant factors and using a first order regression model, the estimated response function is obtained in equation (10), where $f_{cost(2)}$ denotes the expected cost per hour for the 2-attribute process at hand.

$$f_{cost(2)} = -78.836 + 0.484Q + 0.477C_1 + 0.507C_2 + 2938.106\theta - 52.116\delta \tag{10}$$

Using equation (10), the estimated minimum cost is 146.0646 by the LINGO80 software, in which the parameter values are $Q = 30(a = 25, b = 5)$, $C_1 = 100$, $C_2 = 250$, $\theta = 0.03$, and $\delta = 1$.

The following can be inferred based on equation (10):

- The number of occurrences of the assignable cause per hour (θ) has the greatest effect on the expected total cost per hour.
- Increase in the sampling costs (a, b), in the expected costs when the process is in and out-of-control (C_1, C_2), or in the number of occurrences of the assignable cause per hour (θ) cause the expected total cost per hour to increase as well.
- A larger shift size (δ) in the process mean cause the expected total cost per hour to decrease.
- The minimum value of the expected total cost per hour is obtained based on low levels of (a, b), C_1, C_2 and θ and high level of θ given in Table (2).

7 CONCLUSIONS

An economic design of VSSI MEWMA control chart to monitor multiattribute processes was proposed in this research using skewness reduction approach, the Lorenzen-Vance cost function, and GA. Simulation experiments were performed in a sensitivity analysis of sampling cost, expected costs per hour, and the number of occurrences of the assignable cause on the total expected cost of the chart. Results showed that when these parameters increase, the total expected cost per hour increase as well. By contrast, a larger shift size in the process mean caused a decrease in the total expected total cost per hour. While these parameters were shown to significantly affect the total expected cost per hour, the other cost parameters were found insignificant. Moreover, based on the estimated multiple linear regression function of the total expected cost per hour, the number of occurrences of the assignable cause per hour was the most significant parameter that affects the expected total cost per hour.

For future research, the proposed model can be extended for a situation where the process mean may experience shifts by increase in one or more attribute means so that it will be able to first identify which attribute(s) have caused the shift. Furthermore, a DOE approach can be utilized to determine the optimum GA parameters. Moreover, the multivariate CUSUM control chart can replace the MEWMA chart for a comparison purpose.

REFERENCES

Chou, C. Y., Chen, C. H., Liu, H. R. (2006). Economic design of EWMA charts with variable sampling intervals. *Quality & Quantity* 40: 879–896.

- Duncan A. J. (1956). The economic design of X-charts used to maintain current control of a process. *Journal of the American Statistical Association* 51: 228–242.
- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.
- Jolayemi J. K. (2000). An optimal design of multiattribute control charts for processes subject to a multiplicity of assignable causes. *Applied Mathematics and Computation* 114: 187-203.
- Lin, Y. C., Chou, C. Y. (2005a). Adaptive X-bar control charts with sampling at fixed times. *Quality and Reliability Engineering International* 21: 163–175.
- Lin, Y. C., Chou, C. Y. (2005b). On the design of variable sample size and sampling intervals X-bar charts under non-normality. *International Journal of Production Economics* 96: 249–261.
- Lorenzen, T. J., Vance, L. C. (1986). The economic design of control charts: A unified approach. *Technometrics* 28: 3-10.
- Lowry, C. A., Woodall, W. H., Champ, C. W., Erigdon, S. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics* 34: 46–53.
- Montgomery, D. C. (1996). *Design and Analysis of Experiments*, 4th ed., John Wiley & Sons, New York, NY.
- Niaki, S. T. A., Abbasi, B. (2008). A transformation technique in designing multi-attribute C control charts. *Scientia Iranica* 15: 125-130.
- Reynolds, M. R., Jr., Arnold, J. C. (2001). EWMA control charts with variable sample sizes and variable sampling intervals. *IIE Transactions* 33: 511-530.
- Saccucci, M.S., Amin, R.W. Lucas, J. M. (1992). Exponentially weighted moving average control schemes with variable sampling intervals. *Communication in Statistics-Simulation and Computation* 21: 627-657.