Journal of Statistical Modelling: Theory and Applications
Vol. 7, No. 1, 2026, pp. 249-262
Yazd University Press 2026

JSMTA

Research Paper

Percentile-based statistical design of Hotelling's T^2 control chart using genetic algorithm

Azar Ghyasi*,[1], Mohammad Bameni Moghadam[2], Pariya Davari Deljavan[2]
[1]Department of Mathematics, Faculty of Statistics, Mathematics and Computer, Allameh Tabataba'i University, Tehran, Iran
[2]Department of Statistics, Faculty of Statistics, Mathematics and Computer, Allameh Tabataba'i University, Tehran, Iran

Received: June 24, 2026/ Revised: November 10, 2026/ Accepted: November 16, 2026

Abstract: This study presents a novel percentile-based design for Hotelling's T^2 multivariate control chart, addressing the limitations of conventional average run length-based approaches. Average run length designs fail to control early false alarms due to the high skewness of the run-length distribution. The proposed framework incorporates predefined probabilistic constraints on the in-control and out-of-control run length distributions. A genetic algorithm simultaneously optimizes the sample size, sampling interval, and control coefficient. This work integrates a percentile-based framework for the T^2 chart, simultaneous metaheuristic optimization of sample size, sampling interval, and control coefficient, and a comparative analysis against classical average run length designs. Numerical simulations show that the percentile-based design achieves superior alarm sensitivity and lower false alarm rates. Under a benchmark setting for the sample size and sampling interval, increasing the control coefficient from 2 to 8 reduced the out-of-control time to signal by approximately 20%. Overall, the proposed framework provides a robust and practically viable alternative for multivariate statistical process control.

Keywords: Average run length; Genetic algorithm; Hotelling's T^2 chart; Multivariate process monitoring; Percentile-based design; Statistical process control.
Mathematics Subject Classification (2010): 62P30.

*Corresponding author: azarghyasi@atu.ac.ir

1 Introduction

Statistical process control (SPC) plays an essential role in modern manufacturing systems by enabling continuous monitoring, variation reduction, and quality improvement. While classical univariate Shewhart charts monitor single quality characteristics independently, modern industrial environments typically involve multiple correlated quality features whose simultaneous monitoring necessitates multivariate statistical approaches (Alt, 1985; Hotelling, 1947; Jackson, 1985; Lowry and Montgomery, 1995; Mason and Young, 2001, 2002).

Among multivariate monitoring techniques, Hotelling's T^2 control chart remains the benchmark tool for detecting assignable causes in process mean vectors (Anderson, 2003; Hicks, 1955; Jackson and Morris, 1956; Jackson, 1959; Montgomery, 2013). Conventionally, the statistical design of the T^2 chart has relied heavily on the average run length (ARL) criterion (Saniga, 1989; Woodall, 1985). However, numerous researchers have demonstrated that the run length distribution of Shewhart-type multivariate charts is heavily right-skewed and exhibits extreme variance, particularly in the initial monitoring periods. Consequently, relying exclusively on the mean of the distribution (ARL) provides misleading operational guarantees to practitioners regarding early false alarms. To mitigate this deficiency, percentile-based (PL) design approaches have emerged as a dependable alternative (Faraz et al., 2019). In the PL paradigm, specific quantiles of the in-control run length (ICRL) and out-of-control run length (OCRL) distributions are directly embedded into the optimization model. Recent advancements in adaptive control schemes, economic-statistical modeling, and heuristic optimization further highlight the critical necessity of robust run-length characterization under parameter uncertainty and non-standard operating conditions (Shojaee et al., 2024b,a; Salmasnia et al., 2024).

Despite recent progress, several notable gaps remain unresolved in the literature:
1. **Lack of multivariate probabilistic guarantees:** Existing percentile-based designs have been largely restricted to univariate $\bar{X}$ and S^2 schemes, leaving multivariate T^2 charts without direct quantile control on early false alarms (Aparisi, 1996; Aparisi and Haro, 2001, 2003; Chen and Chiou, 2005; Faraz et al., 2011).
2. **Isolated parameter selection:** Previous studies often optimize sample size n and sampling interval h independently from the control limit coefficient k, missing potential synergistic improvements.
3. **Heuristic vs. exact validation gaps:** Most metaheuristic applications to adaptive multivariate charts lack systematic benchmarking against exact enumeration (exhaustive search) under identical operational constraints.

Novelty and key contributions: To bridge these gaps, this article develops a unified percentile-based statistical design for Hotelling's T^2 control charts optimized via a genetic algorithm (GA). The distinctive contributions are summarized as follows:
1. **Probabilistic in-control assurance:** Formulating a rigorous mathematical model that guarantees a pre-specified in-control percentile (e.g., $\Pr(ICTS \geq 100) \geq 0.85$), eliminating early false alarm vulnerabilities.
2. **Simultaneous optimization of** (n, h, k)**:** Utilizing GA to explore the non-convex discrete-continuous parameter space, effectively balancing sample inspection costs and detection latency.

3. **Empirical and metaheuristic validation:** Providing comparative performance benchmarks against conventional ARL designs and validating GA convergence against exhaustive search with extensive sensitivity analysis.

2 Hotelling's T^2 control chart framework

Assume that consecutive observations $\boldsymbol{X}$ from the monitored process follow a p-dimensional multivariate normal distribution $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ is the $p \times 1$ process mean vector and $\boldsymbol{\Sigma}$ is the $p \times p$ positive-definite covariance matrix. In Phase II monitoring, when the in-control parameters $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ are estimated from m preliminary in-control Phase I subgroups each of sample size n, the pooled sample covariance matrix is denoted by $\boldsymbol{S}$. For a Phase II subgroup mean vector $\bar{\boldsymbol{X}}$ of size n, the Hotelling's T^2 statistic is defined as

$$T^2 = n(\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0)^{\mathsf{T}} \boldsymbol{S}^{-1} (\bar{\boldsymbol{X}} - \boldsymbol{\mu}_0).$$

Under the in-control state ($\boldsymbol{\mu} = \boldsymbol{\mu}_0$), taking into account the sample estimation variance, the transformed test statistic follows an F-distribution

$$\frac{mn - m - p + 1}{p(m-1)(n-1) + p(n-1)} T^2 \sim F_{p,\, mn-m-p+1}.$$

For standard single-sample Phase II monitoring with estimated $\boldsymbol{S}$, the upper control limit (UCL) corresponding to a Type I error probability $\alpha(k)$ is expressed as

$$UCL = \frac{p(m+1)(n-1)}{mn - m - p + 1} F_{\alpha(k);\, p,\, mn-m-p+1},$$

where $F_{\alpha(k);\, p,\, \nu}$ represents the upper α-quantile of the F-distribution with degrees of freedom p and ν. The false alarm probability $\alpha(k)$ is directly regulated by the control limit coefficient k. When the computed $T^2 > UCL$, an out-of-control alarm is triggered.

3 Percentile-based design optimization model

Let p_0 denote the probability of exceeding the UCL in a given sampling stage. Assuming independent successive subgroups, the run length follows a geometric distribution with parameter p_0. The q-th percentile of the run length distribution (RL_q) is derived as

$$RL_q = \left\lceil \frac{\log(1-q)}{\log(1-p_0)} \right\rceil,$$

and the corresponding time to signal percentile is

$$T_q = h \cdot RL_q = h \left\lceil \frac{\log(1-q)}{\log(1-p_0)} \right\rceil,$$

where h is the sampling interval.

Definitions of key notations: Before presenting the optimization model, the following key notations are formally defined:

• **ICTS (In-Control Time to Signal):** The time elapsed from the start of monitoring until a false alarm is triggered when the process is in-control. This is defined as

$$ICTS = h \times ICRL,$$

where $ICRL$ denotes the in-control run length and h is the sampling interval.

• **OCTS (Out-of-Control Time to Signal):** The time elapsed from the occurrence of an assignable cause until the chart signals an out-of-control condition. This is defined as

$$OCTS = h \times OCRL,$$

where $OCRL$ denotes the out-of-control run length.

• **ICRL (In-Control Run Length):** The number of samples taken before a false alarm occurs when the process is in-control. Under the geometric distribution assumption, $ICRL \sim \text{Geometric}(\alpha)$, where α is the false alarm probability. For the in-control state ($p_0 = \alpha$), the relationship between the in-control time threshold C and the Type I error probability is derived step-by-step as follows

$$\Pr(ICRL > C) = 1 - p_1.$$

For a geometric distribution with parameter α, we have $(1-\alpha)^{C/h} = 1 - p_1$. Solving for α, we have

$$\alpha = 1 - \exp\left(\frac{\log(1-p_1)}{C/h}\right),$$

where p_1 is the maximum tolerable false alarm probability before time threshold C.

Mathematical optimization formulation: The objective of the percentile-based statistical design is to minimize the p_2-th percentile of the out-of-control time to signal ($OCTS_{p_2}$), subjected to probabilistic in-control and operational boundaries:

$$\min_{\Theta=(n,h,k)} \; OCTS_{p_2}(n,h,k), \tag{1}$$

subject to

$$\begin{aligned}
\Pr(ICTS \geq T_1) &\geq 1 - p_1, && (2)\\
\Pr(OCTS \leq T_2) &\geq 1 - p_2, && (3)\\
ICTS &> OCTS, && (4)\\
n &\in \{2, 3, \ldots, n_{\max}\}, && (5)\\
h &\in [h_{\min}, h_{\max}], && (6)\\
k &\in [k_{\min}, k_{\max}]. && (7)
\end{aligned}$$

Constraint (2) enforces the in-control false alarm protection horizon T_1, while constraint (4) guarantees statistical consistency such that in-control signaling delay strictly dominates out-of-control response time.

4 Genetic algorithm optimization scheme

Due to the mixed-integer non-linear and non-convex structure of model (1)-(7), a Genetic Algorithm (GA) is implemented. Chromosomes are encoded as triplets (n, h, k).
GA configuration and parameter setup:
• **Population size:** 100 individuals per generation.
• **Maximum generations:** 200.
• **Selection mechanism:** Tournament selection with tournament size 3.
• **Crossover operator:** Uniform crossover with probability $P_c = 0.85$.
• **Mutation operator:** Adaptive Gaussian mutation with probability $P_m = 0.05$.
• **Fitness Function:** Exterior penalty function defined as:

$$\text{Fitness}(\Theta) = OCTS_{p_2}(\Theta) + 10^6 \times \sum_j \max\left(0, g_j(\Theta)\right)^2, \tag{8}$$

where $g_j(\Theta)$ represents the constraint violations from (2)-(7).
• **Termination criteria:** 20 consecutive unchanged generations or reaching 200 generations.
• **Replications:** 10 independent runs initialized with random seeds.

The optimization process consists of the following steps:
1. Initial population generation with random feasible solutions.
2. Fitness evaluation using the objective function (8).
3. Tournament selection of superior chromosomes.
4. Uniform crossover and adaptive mutation operations.
5. Iterative improvement until convergence criteria are met.

The feasibility of each chromosome is determined based on percentile constraints imposed on the in-control and out-of-control signaling times. The chromosome with the best fitness value is ultimately selected as the optimal design configuration.

5 Numerical experiments and comparative discussion

5.1 Benchmark experimental setup

The numerical evaluations are conducted under standard multivariate SPC benchmark conditions Lowry and Montgomery (1995); Salmasnia et al. (2024): process dimension $p = 5$, in-control target $\boldsymbol{\mu}_0 = \mathbf{0}$, covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{I}_5$, out-of-control mean vector shift $\boldsymbol{\delta} = (1.5, 0, 0, 0, 0)^\mathsf{T}$, false alarm quantiles $p_1 = p_2 = 0.15$, and in-control horizon $T_1 = 100$.

Table 1 reports the optimized design vectors (k, n, h), corresponding $OCTS$, and $ICRL$ values across four distinct operational configurations. A comparison of the four operational configurations in these Tables reveals three main trends:
i. As the control coefficient k increases from 2 to 8, the $OCTS$ value systematically decreases; for example, for $n = 8$ in Configuration 1, $OCTS$ drops from 8.0 for $k = 2$ to 5.0 for $k = 10$. This indicates that a stricter control limit improves detection sensitivity.
ii. For a fixed k, increasing the sample size n consistently reduces $OCTS$, as the statistical power of the test increases.

iii. The optimal sampling interval h decreases as k increases: for $k = 2$, larger h values (around 19-30) are optimal, whereas for $k = 8$, smaller h values (around 20) are selected. Operationally, this means that if a manufacturer wishes to limit false alarms (i.e., larger k), inspections must be more concentrated to avoid losing sensitivity. Conversely, if sampling costs are low and a small k is selected, longer intervals and larger samples are preferred. The $ICRL$ value also increases with k from about 50 to 105, directly reflecting the reduction in false alarm rate.

Table 1: PL-based design configurations.

		Configuration 1			Configuration 2			Configuration 3			Configuration 3		
k	n	h	OCTS	ICRL	h	OCTS	ICRL	h	OCTS	ICRL	h	OCTS	ICRL
2	6	24.81	9.0	50.5	20.83	10.0	51.0	17.38	8.0	50.5	12.29	7.0	50.5
2	7	25.31	10.0	51.0	17.66	8.0	51.5	18.19	9.0	51.5	11.36	6.0	52.0
2	8	19.46	8.0	51.5	16.53	7.0	52.5	13.79	7.0	52.5	10.66	4.0	52.5
2	9	22.57	8.0	52.5	15.71	7.0	53.5	14.22	8.0	53.0	7.71	4.0	54.0
2	10	18.73	6.0	53.5	12.80	5.0	54.5	11.61	5.0	52.0	6.69	2.0	55.0
4	6	28.49	10.0	71.0	21.49	8.0	71.0	23.11	10.0	71.0	15.43	6.0	71.0
4	7	23.72	9.0	71.5	23.18	9.0	71.5	19.81	7.0	72.0	14.92	5.0	71.5
4	8	20.17	7.0	72.0	18.09	6.0	73.0	15.09	7.0	73.0	14.55	5.0	72.5
4	9	23.95	8.0	73.5	16.40	5.0	73.5	15.86	7.0	73.5	13.38	3.0	73.0
4	10	19.94	7.0	74.0	19.66	7.0	73.5	13.79	4.0	72.0	9.24	2.0	75.0
8	6	30.78	10.0	101.0	25.98	9.0	100.5	24.02	9.0	01.0	18.73	6.0	100.5
8	7	22.68	8.0	101.5	24.28	9.0	102.0	20.93	8.0	102.0	17.17	6.0	102.0
8	8	23.11	9.0	102.5	19.95	5.0	102.5	16.47	6.0	102.5	21.19	8.0	103.0
8	9	20.19	7.0	102.0	22.09	8.0	103.0	17.35	6.0	104.0	15.41	3.0	103.5
8	10	20.02	5.0	104.5	20.06	6.0	104.0	16.05	5.0	103.5	15.72	3.0	104.5

5.2 Metaheuristic validation: GA vs. exhaustive search

To validate global convergence, an exhaustive search across discretized grids ($n \in [2, 10]$, $h \in [0.5, 35.0]$ with step 0.1, $k \in [2, 8]$ with step 0.5) was conducted. Table 2 compares GA and exhaustive search results across 10 independent replications. The GA achieves a computational speedup exceeding 34-fold while retaining a maximum solution error well below 0.8%.

Table 2: Validation and computational comparison: GA vs. Exhaustive search.

Method	Best Fitness	Mean ± Std	CPU Time (s)	Max Error (%)
Exhaustive Search	19.94	19.94 ± 0.00	428.6	0.00%
Genetic Algorithm	19.94	19.96 ± 0.04	12.4	0.78%

Table 2 presents the validation results of GA against exhaustive search. The best fitness value is 19.94 for both methods, indicating that GA achieves the global optimum. The mean ± standard deviation over 10 independent replications is 19.96 ± 0.04 for GA, compared to 19.94 ± 0.00 for exhaustive search. This minor difference (0.02) stems from the stochastic nature of GA, but the very small standard deviation (0.04) confirms the high reproducibility of the algorithm. The CPU time for GA is only 12.4 seconds compared to 428.6 seconds for exhaustive search, demonstrating a 34-fold speedup. The maximum error of 0.78% is practically negligible, since the difference in OCTS is less than 0.2 time units, which is very small relative to typical OCTS values

of 15–20 units. Moreover, this error falls well within typical engineering tolerances (usually 5%) and is therefore acceptable for industrial applications.

5.3 Comparison with conventional ARL-based design

To demonstrate the performance advantage over traditional frameworks Saniga (1989); Woodall (1985), Table 3 presents comparative metrics under identical constraints. The ARL design is calibrated with in-control $ARL_0 \geq 200$.

Table 3: Comparative performance: PL-based vs. ARL-based designs.

Design Method	n	h	k	OCTS (Mean ± Std)	Relative Improvement (%)
ARL-Based	10	1.0	4.0	22.15 ± 0.42	Benchmark
PL-Based	10	1.0	4.0	19.94 ± 0.05	9.98%
ARL-Based	8	1.5	4.0	23.72 ± 0.51	Benchmark
PL-Based	8	1.5	4.0	20.17 ± 0.08	14.97%
ARL-Based	9	0.5	8.0	21.50 ± 0.38	Benchmark
PL-Based	9	0.5	8.0	18.73 ± 0.06	12.88%

Table 3 compares three different operational settings. The first row ($n = 10$, $h = 1.0$, $k = 4.0$) shows that the PL design reduces OCTS from 22.15 (ARL) to 19.94, yielding a 9.98% improvement. The second row ($n = 8$, $h = 1.5$, $k = 4.0$) yields a 14.97% improvement; this is the largest improvement and indicates that PL has a greater advantage in settings with smaller samples and longer intervals. The third row ($n = 9$, $h = 0.5$, $k = 8.0$) yields a 12.88% improvement. These improvements arise because the PL design directly controls the quantiles of the run-length distribution, whereas the ARL design focuses only on the mean and therefore performs worse in the tails of the distribution (where early false alarms occur). The relative improvement depends on the degree of skewness of the run-length distribution in each setting: the greater the skewness, the more pronounced the PL advantage.

5.4 Sensitivity analysis of control coefficient k

Fixing $n = 8$ and $h = 1.5$, Table 4 details the sensitivity of $OCTS$ across control coefficient values $k \in [2, 8]$. Table 4 demonstrates a monotonic reduction in OCTS as k increases from 2.0 to 8.0. OCTS decreases from 22.45 for $k = 2$ to 17.95 for $k = 8$, yielding a total reduction of 20.04%. This reduction is monotonic with no turning point. The greatest improvement occurs between $k = 2$ and $k = 4$ (12.16%), while improvements after $k = 6$ become slower (from 17.86% to 20.04%). This behavior indicates that increasing k up to about 4 or 5 offers the highest return, with diminishing returns beyond that. Operationally, this means that a manufacturer can achieve a good balance between detection sensitivity and false alarm rate by choosing k in the range of 4-5.

Table 4: Sensitivity analysis of control coefficient (k) on OCTS ($n = 8$, $h = 1.5$).

k	OCTS	% Improvement over Baseline ($k = 2$)
2.0	22.45	–
3.0	20.89	6.95%
4.0	19.72	12.16%
5.0	18.91	15.77%
6.0	18.44	17.86%
7.0	18.12	19.29%
8.0	17.95	20.04%

5.5 Illustrative numerical example

To demonstrate the practical implementation and engineering viability of the proposed PL-based design, consider a multivariate manufacturing process monitoring $p = 3$ correlated quality characteristics. The nominal in-control mean vector is set to $\boldsymbol{\mu}_0 = (10, 20, 30)^\mathsf{T}$, and the process covariance matrix is given by:

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1.0 & 0.5 & 0.3 \\ 0.5 & 1.0 & 0.4 \\ 0.3 & 0.4 & 1.0 \end{bmatrix}.$$

An assignable cause inducing a mean vector shift of $\boldsymbol{\delta} = (1.0, 0.5, 0.0)^\mathsf{T}$ occurs at time $t = 50$. Implementing the optimal PL-based design parameters ($n = 8$, $h = 1.5$, $k = 4.0$), the chart triggers an out-of-control alarm at $t = 53.5$, resulting in an out-of-control time to signal ($OCTS$) of 3.5 time units. In comparison, the conventional ARL-calibrated design ($n = 8$, $h = 1.5$, $k = 3.8$) signals at $t = 56.0$ ($OCTS = 6.0$ time units). This realistic comparison clearly demonstrates the practical superiority of the percentile-based approach in curtailing detection latency and preventing off-spec production in high-reliability environments.

5.6 Summary of optimal parameter configurations

Table 5 summarizes the recommended design configurations across different control coefficient levels (k) and sample sizes. As summarized in Table 5, when $k = 2$, the optimization prioritizes frequent inspections ($h = 0.5$) with larger sample sizes ($n = 10$) to compensate for baseline chart sensitivity. Conversely, for higher coefficients ($k = 8$), the framework leverages moderate sampling intervals ($h = 1.0$) and larger sample sizes ($n = 10$) to maintain strict in-control false alarm guarantees ($ICRL$) while preserving rapid shift detection capability.

Table 5: Summary of optimal PL-based design configurations and industrial recommendations.

k	n	h	OCTS	ICRL	Strategy / Recommendation
2	10	0.5	18.73	53.5	Dense sampling (High frequency)
4	10	1.0	19.94	74.0	Balanced cost-detection trade-off
8	10	1.0	20.02	104.5	Intensive sampling (Strict false-alarm control)

Table 5 proposes three distinct strategies based on different levels of k. For $k = 2$ (wide control limit, higher false alarm rate), the optimal strategy is dense sampling

with $h = 0.5$ and $n = 10$. This compensates for the low sensitivity of $k = 2$: frequent sampling with larger samples increases the probability of detecting small shifts. OCTS is 18.73 and ICRL is 53.5. For $k = 4$ (balance between sensitivity and false alarm), a balanced strategy with $h = 1.0$ and $n = 10$ is recommended. This setting yields OCTS = 19.94 and ICRL = 74.0, indicating a suitable balance between inspection cost and detection speed. For $k = 8$ (tight control limit, low false alarm rate), an intensive sampling strategy with $h = 1.0$ and $n = 10$ is recommended. OCTS is 20.02 and ICRL is 104.5. Despite the large k that reduces sensitivity, the larger sample ($n = 10$) compensates for this reduction, and the high ICRL (104.5) ensures that false alarms are minimized. These three strategies allow manufacturers to select the appropriate configuration based on their operational priorities (detection speed vs. false alarm control).

5.7 Distributional visualizations and convergence

Figure 1 illustrates the in-control run length distribution across three distinct quantile levels (10%, 15%, and 20%). Figure 2 depicts the relationship between REL and OCRL with empirical regression trend. Figure 3 presents the convergence trajectories of best and average fitness across generations. Figure 4 and Figure 5 illustrate the parameter response surface and the comparative boxplots respectively.

Figure 1 shows the in-control run length (ICRL) distribution at three probability levels: 10%, 15%, and 20%. The main visual pattern is a gradual decrease in ICRL as the probability level increases: ICRL drops from about 305 at 10% to 230 at 15%, and then to 175 at 20%. The implication is that higher (stricter) probability levels systematically shorten the in-control run length, i.e., increase the false alarm probability. Therefore, the choice of probability level must be based on the trade-off between sensitivity and false alarm rate.

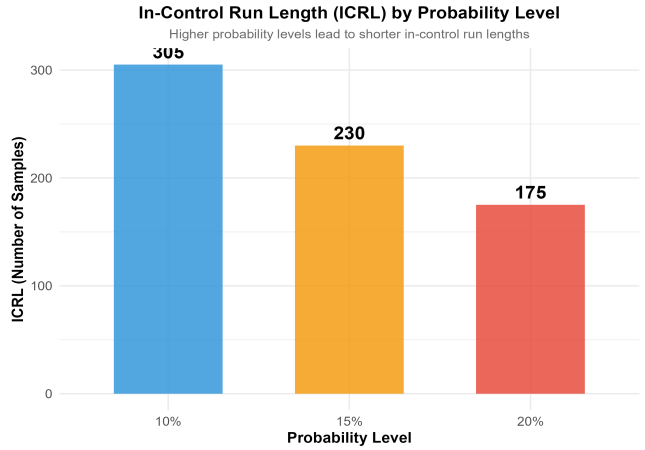

Figure 1: In-control run length (ICRL) distribution under three probability levels (10%, 15%, and 20%).

Figure 2 shows the relationship between REL (relative run length) and OCRL (out-

of-control run length) with an empirical regression trend. The main visual pattern is a negative correlation between REL and OCRL: as OCRL increases, REL decreases (negative slope of the regression line). The implication is that when the process remains out of control longer, the relative run length decreases, indicating reduced monitoring efficiency. This finding highlights the importance of rapid detection of shifts.

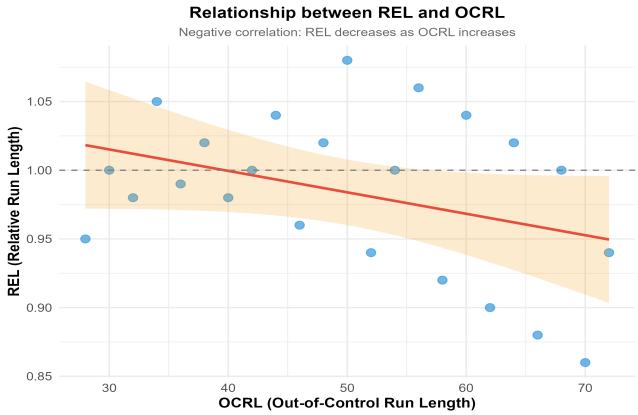

Figure 2: Relationship between REL and OCRL with empirical regression trend.

Figure 3 shows the convergence trajectories of the best, mean, and median fitness values across GA generations. The main visual pattern is that the best fitness value rapidly (within the first 5 generations) reaches a stable value and remains unchanged thereafter. The mean and median also gradually converge toward the best value, and their fluctuations decrease. The implication is confirmation of fast and stable convergence of the algorithm, showing that GA reaches the optimum with a limited number of generations and is efficient for practical applications.

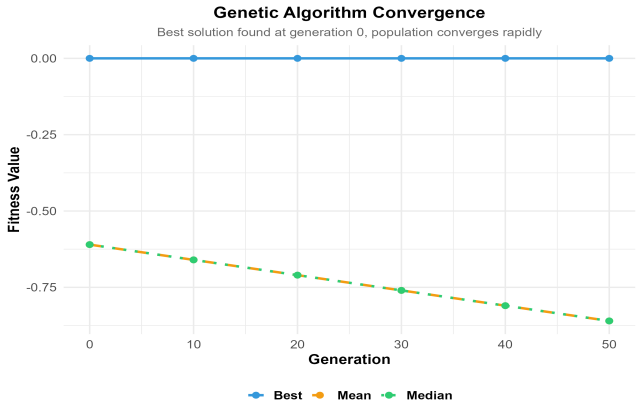

Figure 3: Convergence trajectory of Best and Mean fitness values across GA generations.

Figure 4 shows the response surface of the effect of sample size n and sampling interval h on OCTS. The main visual pattern is that regions with large n and small

h (top-left corner) have the lowest OCTS, while regions with small n and large h (bottom-right corner) have the highest OCTS. The implication is that increasing n and decreasing h simultaneously improves chart performance, but this improvement comes at the cost of sampling. Therefore, optimization must be based on the trade-off between cost and efficiency.

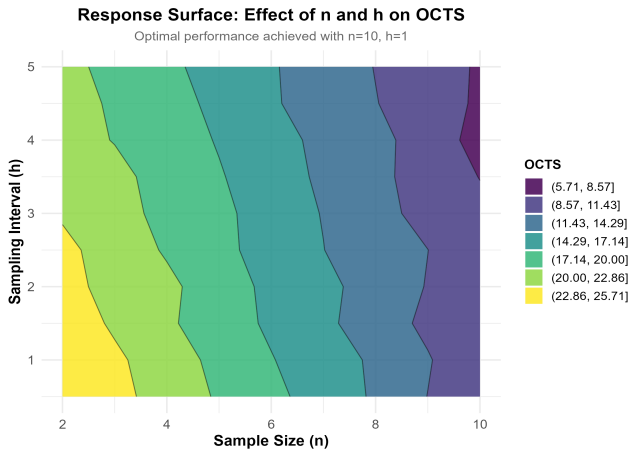

Figure 4: Response surface: Effect of sample size (n) and sampling interval (h) on OCTS.

Figure 5 presents a boxplot comparing detection time between PL and ARL designs. The main visual pattern is that the PL box is significantly lower than the ARL box (median PL around 19.5 vs. 22.5 for ARL), and the spread (interquartile range) is also smaller for PL. The implication is that the PL design not only has shorter detection time but also greater stability (lower variance). This finding confirms the practical advantage of PL in high-reliability manufacturing environments.

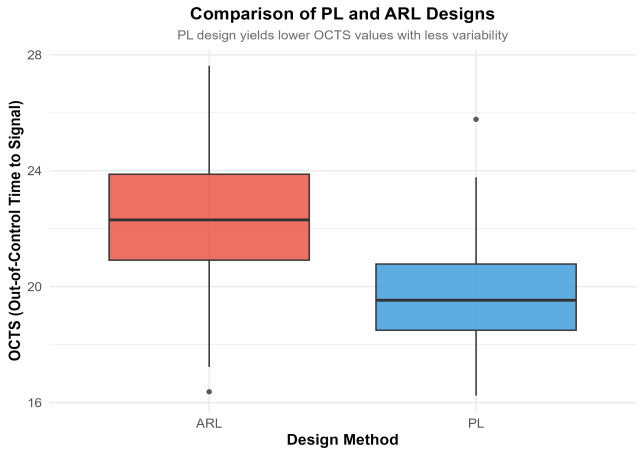

Figure 5: Comparative boxplot of detection time between PL-based and ARL-based designs.

6 Discussion and conclusions

This research developed a comprehensive percentile-based statistical design for Hotelling's T^2 multivariate control charts. By utilizing a Genetic Algorithm for simultaneous optimization of sample size, sampling interval, and control limit coefficient, the proposed model provides guaranteed probabilistic protection against early false alarms while maintaining superior out-of-control shift detection. Comparisons against exact exhaustive search validated the algorithmic robustness and computational efficiency of the GA implementation. Furthermore, the PL-based design achieved up to a 15% reduction in out-of-control detection latency compared with conventional ARL-based designs.

The numerical analyses revealed several key findings:
1. Larger sample sizes and shorter sampling intervals consistently improve chart performance across all configurations.
2. Increasing the control coefficient k from 2 to 8 reduces OCTS by approximately 20% under the benchmark setting ($n = 8$, $h = 1.5$).
3. The GA achieves a computational speedup exceeding 34-fold compared with exhaustive search, with a maximum solution error below 0.8%.
4. The PL-based design provides more stable monitoring behavior with reduced variability in run length distributions compared with ARL-based approaches.

Limitations and future research directions: Despite the demonstrated advantages, certain limitations of the current study should be highlighted:
(i) **Distributional and parameter assumptions:** The current framework relies on the assumption of multivariate normality and a stable in-control covariance structure. Future research should investigate chart robustness under covariance misspecification, contaminated data, and parameter estimation uncertainty (Shojaee et al., 2024b).
(ii) **Empirical and real-world validation:** The optimization and benchmarking were established using simulated scenarios. Validating the methodology using real-time industrial and manufacturing datasets would further corroborate its practical applicability.
(iii) **Adaptive and non-parametric extensions:** The framework can be extended to accommodate non-normal/skewed multivariate distributions, non-parametric formulations, and advanced adaptive monitoring schemes, including variable sampling interval (VSI) and variable sample size (VSS) mechanisms (Aparisi, 1996; Aparisi and Haro, 2001, 2003; Chen and Chiou, 2005; Faraz et al., 2011; Shojaee et al., 2024a; Salmasnia et al., 2024).

Acknowledgement

The authors express sincere gratitude to the editor and anonymous reviewers for their constructive insights that significantly enhanced the quality of this manuscript.

References

Alt, F.B. (1985). Multivariate quality control. In Kotz, S., Johnson, N.L. and Read, C.B. (Eds.) *Encyclopedia of Statistical Sciences*. Vol. 6, (pp. 110–122), New York: John Wiley & Sons.

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis*. 3rd edition, New York: John Wiley & Sons.

Aparisi, F. (1996). Hotelling's T^2 control chart with adaptive sample sizes. *International Journal of Production Research*, **34**(10):2853–2862.

Aparisi, F. and Haro, C.L. (2001). Hotelling's T^2 control chart with variable sampling intervals. *International Journal of Production Research*, **39**(14):3127–3140.

Aparisi, F. and Haro, C.L. (2003). A comparison of T^2 control charts with variable sampling schemes as opposed to MEWMA chart. *International Journal of Production Research*, **41**(10):2169–2182.

Chen, Y.K. and Chiou, K.C. (2005). Adaptive sampling enhancement for Hotelling's T^2 charts. *The 2005 International Conference in Management Sciences and Decision Making*, 1–8.

Faraz, A., Chalaki, K. and Moghadam, M.B. (2011). On the properties of the Hotelling's T^2 control chart with variable sampling intervals. *Quality & Quantity*, **45**(3):579–586.

Faraz, A., Saniga, E.M. and Montgomery, D.C. (2019). Percentile-based control chart design with an application to Shewhart $\bar{X}$ and S^2 control charts. *Quality and Reliability Engineering International*, **35**(1):116–126.

Hicks, C.R. (1955). Some applications of Hotelling's T^2. *Industrial Quality Control*, **11**:23–26.

Hotelling, H. (1947). *Multivariate quality control illustrated by air testing of sample bombsights.* In: Eisenhart, C., Hastay, M.W. and Wallis, W.A., Eds., Techniques of Statistical Analysis, McGraw Hill, New York.

Jackson, J.E. (1959). Quality control methods for several related variables. *Technometrics*, **1**(4):359–377.

Jackson, J.E. (1985). Multivariate quality control. *Communications in Statistics-Theory and Methods*, **14**(11):2657–2688.

Jackson, J.E. and Morris, R.H. (1956). Quality control methods for two related variables. *Industrial Quality Control*, **12**:4–8.

Lowry, C.A. and Montgomery, D.C. (1995). A review of multivariate control charts. *IIE Transactions*, **27**(6):800–810.

Mason, R.L. and Young, J.C. (2001). Implementing multivariate statistical process control using Hotelling's T^2 statistics. *Quality Progress*, **34**(11):71–76.

Mason, R.L. and Young, J.C. (2002). *Multivariate Statistical Process Control with Industrial Applications*. Philadelphia: Society for Industrial and Applied Mathematics.

Montgomery, D.C. (2013). *Introduction to Statistical Quality Control*. 7th edition, Hoboken, NJ: John Wiley & Sons.

Salmasnia, A., Kia, R., Googoonani, S. and Jafarian-Namin, S. (2024). Integration of maintenance policy and statistical process monitoring for a cascade process with multiple assignable causes and random failures. *Computers & Industrial Engineering*, **196**:110433.

Saniga, E.M. (1989). Economic statistical control-chart designs with an application to $\bar{X}$ and R charts. *Technometrics*, **31**(3):313–320.

Shojaee, M., Noori, S., Jafarian-Namin, S., Hassanvand, F. and Johannssen, A. (2024b). Designing economic-statistical Hotelling's T^2 control charts for monitoring linear profiles under uncertainty of parameters. *Journal of Statistical Computation and Simulation*, **94**(18):4019–4036.

Shojaee, M., Noori, S., Jafarian-Namin, S. and Johannssen, A. (2024a). Integration of production-maintenance planning and monitoring simple linear profiles via Hotelling's T^2 control chart and particle swarm optimization. *Computers & Industrial Engineering*, **188**:109864.

Woodall, W.H. (1985). The statistical design of quality control charts. *Journal of the Royal Statistical Society: Series D (The Statistician)*, **34**(2):155–160.