An effective statistical process control scheme for industrial environmental monitoring

Author(s): Shamsuzzaman, Mohammad; Haridy, Salah; Maged, Ahmed; Bashir, Hamdi; Shamsuzzoha, Ahm; Ali, Ahad

Title: An effective statistical process control scheme for industrial environmental monitoring

Year: 2022

Version: Accepted manuscript

Please cite the original version:

Shamsuzzaman, M., Haridy, S., Maged, A., Bashir, H., Shamsuzzoha, A. & Ali, A. (2022). An effective statistical process control scheme for industrial environmental monitoring. *Chemometrics and Intelligent Laboratory Systems* 229, 104651. https://doi.org/10.1016/j.chemolab.2022.104651

An effective statistical process control scheme for industrial environmental monitoring

Mohammad Shamsuzzaman[a,*], Salah Haridy[a,b], Ahmed Maged[b,c], Hamdi Bashir[a], Ahm Shamsuzzoha[d], and Ahad Ali[e]

[a]*Department of Industrial Engineering and Engineering Management, Sustainable Engineering Asset Management (SEAM) Research Group, College of Engineering, University of Sharjah, United Arab Emirates,*
[b]*Department of Mechanical Engineering, Benha Faculty of Engineering, Benha University, Egypt*
[c]*Department of Systems Engineering and Engineering Management, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong*
[d]*Digital Economy Research Platform, School of Technology and Innovations, University of Vaasa, Vaasa 65101, Finland*
[e]*Department of Mechanical, Robotics, and Industrial Engineering, Lawrence Technological University, Southfield, Michigan, MI 48075, USA*

[*]Mohammad Shamsuzzaman (**Corresponding author**)
Department of Industrial Engineering and Engineering Management,
College of Engineering,
University of Sharjah,
United Arab Emirates
Email: mshamsuzzaman@sharjah.ac.ae
Phone: +971-6-5053974
Fax: +971-6-5053963
Email: mshamsuzzaman@sharjah.ac.ae
ORCID: https://orcid.org/0000-0002-1242-9627

Salah Haridy
Department of Industrial Engineering and Engineering Management,
College of Engineering,
University of Sharjah,
United Arab Emirates,
Mechanical Engineering Department,
Benha Faculty of Engineering, Benha University, Benha, Egypt
Email: sharidy@sharjah.ac.ae

Ahmed Maged
Department of Systems Engineering and Engineering Management,
City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong,
Mechanical Engineering Department,

Benha Faculty of Engineering, Benha University, Benha, Egypt
Email: amaged2-c@my.cityu.edu.hk

Hamdi Bashir
Department of Industrial Engineering and Engineering Management,
College of Engineering,
University of Sharjah,
United Arab Emirates
Email: hbashir@sharjah.ac.ae

Ahm Shamsuzzoha
Digital Economy Research Platform,
School of Technology and Innovations,
University of Vaasa, PO Box 700,
FI-65101, Vaasa, Finland
Email: ahm.shamsuzzoha@uwasa.fi

Ahad Ali
Department of Mechanical, Robotics, and Industrial Engineering
Lawrence Technological University
Southfield, Michigan, MI 48075, USA
Email: sali@ltu.edu

An effective statistical process control scheme for industrial environmental monitoring

ABSTRACT

One major source of environmental pollution is industrial effluents, especially the heavy metal pollutants, including zinc, in the wastewater discharged from manufacturing industries (hereinafter known as an environmental process). The safe disposal of such effluents into aquatic environments is very crucial and of great health and environmental concerns worldwide. To safeguard the quality of the receiving waters, continuous and effective monitoring of zinc in industrial effluents before discharging it into the environment is essential. However, the real-life environmental data are often skewed and do not follow the normal distribution. Therefore, classical statistical process monitoring techniques based on normality assumptions cannot be used directly to monitor the environmental processes. To overcome this limitation, this study designs an optimal $\bar{X}$ scheme for monitoring non-normal zinc data in industrial effluents. This scheme minimizes the total cost, including quality and manpower costs, based on optimal deployment of manpower. The optimization process ensures that the false alarm rate of the monitoring scheme meets the allowable value. A fractional factorial experiment considering different design specifications shows that the proposed optimal $\bar{X}$ scheme reduces the associated costs by about 69% and 36% compared to the basic $\bar{X}$ and improved $\bar{X}$ schemes, respectively. The robustness of the performance of the optimal $\bar{X}$ scheme is further investigated under different non-normal, and process shift distributions. A sensitivity analysis is also conducted to explore the impact of the input parameters on the associated cost of the optimal $\bar{X}$ scheme. Lastly, the design and application of the proposed optimal $\bar{X}$ scheme are presented using real data on zinc content in the wastewater emitted from a steel production plant.

Keywords: Industrial effluent, Environmental quality monitoring, Non-normal environmental data, Statistical process control, Quality cost, Manpower cost

1. Introduction

Environmental pollution is one of the major problems facing humanity. It is considered as one of the greatest threats to natural resources including water bodies. One key contributor to the pollution issue is rapid industrialization [1,2]. Industries consume more freshwater than households, and consequently produce sizeable wastewater discharges that commonly contain large effluents of contaminations. The disposal of such industrial wastewater into aquatic environments is of great environmental and health concern worldwide. In order to safeguard public health, protect the wastewater collection and deposition systems, and comprehend the problem before the quality of the receiving water depreciates beyond natural recovery, continuous monitoring of effluents from industries is extremely important before discharging the wastewater into an aquatic environment [3,4].

Water and wastewater quality monitoring or environmental performance assessment, in general, is an ongoing area of research that has seen continual development over the last decade [5,6]. To this end, many researchers developed statistical process monitoring (SPM) schemes for monitoring and controlling environmental performance. However, most of the available SPM schemes (for instance, see [7-16]) are either designed based on nonparametric methods or assuming that the quality characteristics of interest (e.g. environmental data) are normally and independently distributed. It is well known that the non-parametric approaches are somehow less efficient than their parametric counterparts [17-18], and they need special tables for designing the charting scheme with smaller sample sizes [19]. The real-life environmental data are often skewed and do not usually follow a normal distribution [20,21]. Thus, the SPM schemes designed based on normality assumption are not recommended for monitoring the non-normal environmental processes as they might lead to misleading conclusions and wrong decisions.

The in-control (IC) performance of the monitoring scheme designed based on normality assumption is significantly affected by the non-normal data, and the false alarm probability of the scheme increases as the skewness of the data increases [20-24]. An effective way to minimize the consequences of violating the normality assumption is to modify the control limits of the monitoring scheme using good approximation. This approach is comparatively easier as the original (untransformed) data can be directly plotted on the monitoring scheme. Many researchers (for example, see [21,23,25,26]) followed this approach and designed the monitoring scheme based on Burr distribution approximation [27]. Yourstone and Zimmer [23] modified the control limits of the monitoring scheme by approximating the underlying distribution using the Burr distribution with estimated skewness and kurtosis. Chou et al. [25] designed the economic $\bar{X}$ scheme for monitoring non-normal data based on Burr distribution. Similarly, Chen [26] used Burr distribution in designing an economic $\bar{X}$ scheme for monitoring non-normal data with variable sampling policy. However, these economic schemes are meant to monitor a single, or a specific number of shifts in the process mean. Consequently, the random characteristics of the process shifts are not reasonably captured by the designed monitoring schemes [28]. Even though the traditional economic schemes are useful as they help in minimizing the total costs including quality cost and other relevant costs, researchers usually criticize the economic design of the monitoring schemes as it requires estimating many costs and other design parameters, in addition to its complex mathematical formulation [29]. Recently, Liu and Xue [21] developed a simple cost-based EWMA scheme considering random shifts in the non-normal industrial pollution process.

The design of the monitoring scheme optimizes only the associated quality loss. However, the main objective of employing SPM is to optimize the total quality cost resulted from not only the poor quality but also the manpower cost [30-32]. Wu et al. [30], Wu et al. [31] and Shamsuzzaman et al. [32] designed $\bar{X}$ and $\bar{X}\,\&\,S$ monitoring schemes based on manpower and quality costs; however, these designs assumed that the quality characteristics are normally distributed, and thus, not applicable for monitoring non-normal environmental processes.

The present study develops a model for the optimization design of the $\bar{X}$ scheme for effective monitoring and controlling of non-normal environmental processes. It is noted that the objective of this research is not to identify the best statistical process control (SPC) scheme for monitoring non-normal environmental process, instead it only focuses on the optimization design of the $\bar{X}$ scheme. This is because $\bar{X}$ scheme is one of the most widely used monitoring schemes for monitoring variables in practice [29]. The contributions of this study and merits of the proposed scheme compared with the traditional economic models are summarized as follows: (i) the proposed model optimizes the charting parameters (sample size, *n*; sampling interval, *h*; and lower and upper control limits, *LCL* and *UCL*) of the $\bar{X}$ scheme to minimize the total cost, comprising manpower and quality costs. The optimization process considers random shifts (δ) in the process, which is more realistic from a practical viewpoint, (ii) the proposed model needs few design parameters that can be quantified easily, (iii) the proposed model considers the manpower as a design variable, and searches the optimal amount of manpower that should be deployed to the $\bar{X}$ scheme to minimize the total cost, (iv) the design of the proposed $\bar{X}$ scheme is automated by developing a computer program coded in C programming language that can be obtained from the author upon request, and (v) the design and application of the proposed $\bar{X}$ scheme are demonstrated using real-life wastewater data to promote its usage in practice.

2. Design of the proposed monitoring scheme

This section describes the formulation and rationale of the optimization model to monitor the non-normal environmental processes.

2.1 Model development

2.1.1 Assumptions

The optimization model is developed using the following assumptions:

(1) The environmental process parameter X is independent. Its distribution is non-normal with known in-control (IC) mean, μ_0 and a standard deviation, σ_0. The existence of an assignable cause will change the IC mean, μ_0 to an out-of-control (OOC) mean, μ_1.

$$\mu_1 = \mu_0 + \delta\sigma , \tag{1}$$

where, δ is the shift size in the mean value of the environmental process encountered by an assignable cause. If the environmental process is in control, then $\delta = 0$. For the sake of simplicity, the shift in the standard deviation of the environmental process is not taken into consideration in this research (i.e., $\sigma \equiv \sigma_0$).

(2) A Rayleigh distribution captures the random character of shifts in most environmental processes. It is widely used in the SPM literature as a realistic candidate for the process shift distribution [33,34].

(3) The IC mean (μ_0) of the environmental process meets the center value (target) between the lower specification limit (LSL) and upper specification limit (USL).

2.1.2 Input parameters

The input parameters used in designing the proposed optimization model and their definitions are shown in Table 1.

Table 1. Input parameters used in designing the model

Input parameter	Definition
c, r	Burr distribution parameters
M, S	Mean and standard deviation of the Burr distribution, respectively
μ_0, σ_0	In-control mean and standard deviation of the environmental process, respectively
τ	Minimum allowable ATS_0 of the monitoring scheme
LSL, USL	Lower and upper specification limits (allowable quantity of the chemical elements in the effluent emitted from the environmental process)
Q	The average quantity of the chemical elements in the effluent emitted from the environmental process per unit time (i.e., the overall effluent discharge rate)
m	The maximum allowable quantity of the environmental data inspected per unit time (i.e., maximum allowable inspection rate)
C_m	Manpower cost (including depreciation cost of the measuring instruments) for an inspector per unit time
C_K	Overall penalty cost for an out-of-specification quantity of the chemical elements in the effluent emitted from the environmental process
$MTBO$	Mean time between out-of-control cases of the environmental process
μ_δ	The mean value of the mean shifts δ in the environmental process

The values of Burr distribution parameters (c, r, M, S) can be obtained from [27] or [35] concerning the estimated values of skewness (α_3) and kurtosis (α_4) of the individual measurement. A brief description of the Burr distribution has been included in the Appendix. The IC process parameters (μ_0, σ_0) can be obtained during the phase I design of the monitoring scheme. The estimation of μ_0 and σ_0 is explained in the section of the example (Section 4). The value of the parameter τ can be decided based on the trade-off between the false alarm rate and the detection power. To satisfy the constraint on the false alarm rate and make full use of the potential power of the monitoring scheme, the optimization design always makes the actual (or resultant) IC average time to signal, ATS_0 (the average time required by the monitoring scheme to produce a signal when the environmental process is in an IC state) equal to τ. The specification limits (LSL, USL) can be determined based on the regulatory limits specified by the concerned authority. The values of parameters (Q, m) can be determined based on the industrial records on the effluent data of the environmental process and field test on the inspection study. Information on the manpower cost (C_m) can be obtained from a company's finance department, and the cost component (C_K) can be estimated based on the penalty cost as specified by the concerned authority. Historical data of the OOC cases can be used to estimate the μ_δ and $MTBO$ (interested readers can see [30]). As mentioned earlier, the randomness of the shifts in the environmental process is characterized by Rayleigh distribution in this study. For any given value of μ_δ, the probability density function of the Rayleigh distribution can be obtained by,

$$f_\delta(\delta) = \frac{\pi\delta}{2\mu_\delta^2}\exp\left(-\frac{\pi\delta^2}{4\mu_\delta^2}\right) \tag{2}$$

Figure 1 illustrates the probability density function of Rayleigh distribution for different values of μ_δ.

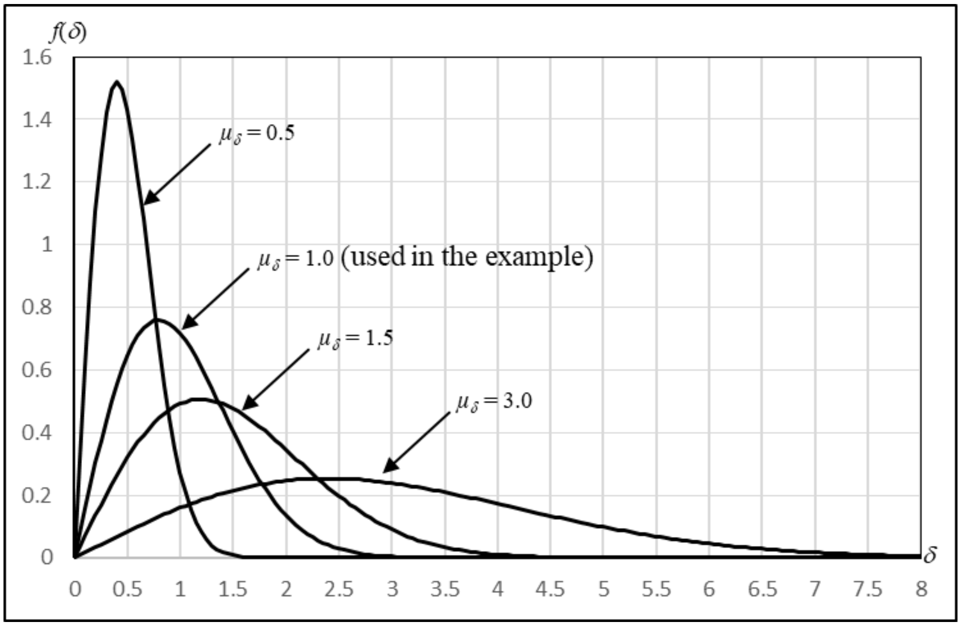


Fig. 1. Rayleigh distribution for different values of μ_δ

2.1.3 Optimization model

The following optimization model implements the design of the proposed $\bar{X}$ scheme for monitoring a non-normal environmental process.

Minimize: C_{total} (3)

Subject to: $ATS_0 \geq \tau$ (4)

Design parameters: L, n, h, LCL, UCL

The model optimizes manpower (L), sample size (n), sampling interval (h), lower control limit (LCL), and upper control limit (UCL) of the monitoring scheme in order to minimize the total cost per OOC cases, C_{total}. The optimization design confirms that ATS_0 is at least equal to the given value τ (i.e., constraint (4) on ATS_0 is satisfied).

Of the five design parameters (L, n, h, LCL and UCL), L and n are independent. For a given set of values of L and n, the other three charting parameters (h, LCL and UCL) are found as follows:

(1) Sampling interval h

Because the number of environmental data measured per unit of time is equal to either ($L \times m$) or (n / h), the sampling interval, h can be calculated by

$$h = \frac{n}{L \times m}. \tag{5}$$

This means that h should be frequent if a small n is used and *vice versa*.

(2) Control limits, LCL and UCL

To satisfy the constraint (4) on ATS_0 and make the monitoring scheme the most potent, having ATS_0 equal to τ is ideal. As a result, the overall type I error probability (α) of the $\bar{X}$ scheme is,

$$\alpha = \frac{h}{ATS_0} = \frac{h}{\tau} \tag{6}$$

Now, the control limits of a symmetric $\bar{X}$ scheme monitoring a non-normal environmental process can be calculated by,

$$\begin{aligned} UCL &= \mu_0 + k\frac{\sigma_0}{\sqrt{n}} \\ LCL &= \mu_0 - k\frac{\sigma_0}{\sqrt{n}} \end{aligned} \tag{7}$$

The parameter k in Eq. (7) is the control limit coefficient and can be determined based on the type I error probability (α) (Eq. (6)). Based on Eq. (7), the α value of the $\bar{X}$ scheme can also be determined by,

$$\alpha = P\left(\bar{X} > UCL\,|_{\mu=\mu_0}\right) + P\left(\bar{X} < LCL\,|_{\mu=\mu_0}\right) = P\left(\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} > k\,|_{\mu=\mu_0}\right) + P\left(\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}} < -k\,|_{\mu=\mu_0}\right) \tag{8}$$

According to the standardized transformation (explained in Eq. (A.3) in the Appendix) and cumulative distribution function of the Burr distribution (explained in Eq. (A.1) in the Appendix), the Eq. (8) can be rewritten as,

$$\begin{aligned}\alpha &= P\left(\frac{Y-M}{S} > k\right) + P\left(\frac{Y-M}{S} < -k\right) = P\left(Y > M + kS\right) + P\left(Y < M - kS\right) \\ &= 1 + \frac{1}{\left[1+\max(0, M+kS)^c\right]^r} - \frac{1}{\left[1+\max(0, M-kS)^c\right]^r}\end{aligned} \tag{9}$$

Considering UCL and the associated α of the $\bar{X}$ scheme, the value of k can be obtained by,

$$k = \frac{\left[\dfrac{1}{(0.5\alpha)^{1/r}} - 1\right]^{1/c} - M}{S}. \tag{10}$$

(3) The objective function

Based on the estimated L, n, h, LCL and UCL, the objective function (C_{total}) is calculated by,

$$C_{total} = C_{quality} + C_{manpower} \tag{11}$$

In this design, the overhead costs, and the quality cost under IC condition (depending on σ_0) have not been considered in calculating C_{total} as these costs are constant and have no effect on the optimal outcomes. The manpower cost, $C_{manpower}$ can be calculated by

$$C_{manpower} = C_m \times L \tag{12}$$

The quality cost, $C_{quality}$ is the cost experienced by the OOC cases. The quality cost, $C_{quality}$ per OOC cases, is given by,

$$C_{quality} = \int_0^{\delta_{\max}} \left[\frac{1}{MTBO} \cdot \left(Q \cdot ATS_1(\delta) \right) \cdot L(\delta) \cdot f(\delta) \right] d\delta$$
$$\delta_{\max} = 2\mu_\delta \sqrt{\frac{-\log(1-0.9999)}{\pi}} \quad (13)$$

where, the value of $\delta_{\max}$ is estimated based on the cumulative distribution function of the Rayleigh distribution assuming that the probability of ($\delta > \delta_{\max}$) is 0.0001. The ratio ($1/MTBO$) is the frequency of the OOC cases and ($Q \cdot ATS_1(\delta)$) is the average amount of effluent discharge from the environmental process during an OOC case for a given δ. The calculation of $ATS_1(\delta)$ will be explained shortly. The expected loss, $L(\delta)$ per OOC cases for a given δ, can be calculated by [36],

$$L(\delta) = K\sigma_0^2(1+\delta^2)$$
$$K = \frac{C_K}{(USL-\mu_0)^2} \quad (14)$$

where, the parameter K is determined based on the cost component, C_K related to the allowable specification limits (regulatory limits). Finally, the value of $C_{quality}$ is calculated by,

$$C_{quality} = \frac{Q \cdot C_K \cdot \sigma_0^2}{MTBO \cdot \left(USL - \mu_0 \right)^2} \int_0^{\delta_{\max}} \left[ATS_1(\delta) \cdot \left(1+\delta^2 \right) \cdot f(\delta) \right] d\delta . \quad (15)$$

To simplify the design process, the integration (15) only considers the upward shift δ in this study, as only the increasing shift in the effluent process is critical from a practical viewpoint. The $ATS_1(\delta)$ for a given δ (in Eq. (15)) can be calculated by,

$$ATS_1(\delta) = ARL_1(\delta) \cdot h = \left(\frac{1}{1-\beta(\delta)} \right) \cdot h . \quad (16)$$

where, $ARL_1(\delta)$ is the OOC average run length and $\beta(\delta)$ is the probability of a type II error of the monitoring scheme for a given δ. The value of $\beta(\delta)$ can be calculated by using standardized transformation (Eq. (A.3)) and cumulative distribution function of the Burr distribution (Eq. (A.1)),

$$\begin{aligned}\beta(\delta) &= P\left(LCL \le \bar{X} \le UCL\,|_{\mu=\mu_0+\delta\sigma_0}\right) = P\left(-k-\delta\sqrt{n} \le \frac{\bar{X}-(\mu_0+\delta\sigma_0)}{\sigma_0/\sqrt{n}} \le k-\delta\sqrt{n}\right)\\ &= P\left(-k-\delta\sqrt{n} \le \frac{Y-M}{S} \le k-\delta\sqrt{n}\right)\\ &= P\left(M-kS-S\delta\sqrt{n} \le Y \le M+kS-S\delta\sqrt{n}\right)\\ &= \frac{1}{\left[1+\max\left(0, M-kS-S\delta\sqrt{n}\right)^c\right]^r} - \frac{1}{\left[1+\max\left(0, M+kS-S\delta\sqrt{n}\right)^c\right]^r}\end{aligned} \quad . \tag{17}$$

Overall, based on the input parameters (c, r, M, S, μ_0, σ_0, USL, Q, m, C_m, C_k, $MTBO$, and μ_δ) and the given values of the design parameters (L, n, h, LCL, and UCL), the objective function C_{total} (the total cost per OOC cases) can be calculated using following steps:

1. Calculate the manpower cost, $C_{manpower}$ by Eq. (12).
2. Calculate the quality cost, $C_{quality}$ by Eq. (15), in which the density function, $f(\delta)$ is calculated by Eq. (2) and the $ATS_1(\delta)$ is determined by Eq.s (16) and (17) for every given value of δ.
3. Lastly, calculate the total cost, C_{total} using Eq. (11).

2.1.4 Optimization algorithm

The optimal values of the design parameters of the proposed $\bar{X}$ scheme are searched in two levels. The top level searches the optimal value of the manpower, L using a Successive Quadratic Estimation Method [37]. In the low level, for any given value of L found in the top level, the sample size, n, is searched from one with a step size of one until the total cost, C_{total}, cannot be further reduced. Both levels jointly minimize C_{total}. The overall optimization process is illustrated in Fig. 2. At the end of the whole optimization process, the optimal values of the design parameters (L, n, h, LCL, and UCL) that warrant the minimum C_{total} and satisfy the constraint ($ATS_0 \geq \tau$), are identified. The design of the proposed $\bar{X}$ scheme is automated by developing a computer program coded using C programming language; the program can be acquired from the author upon request.

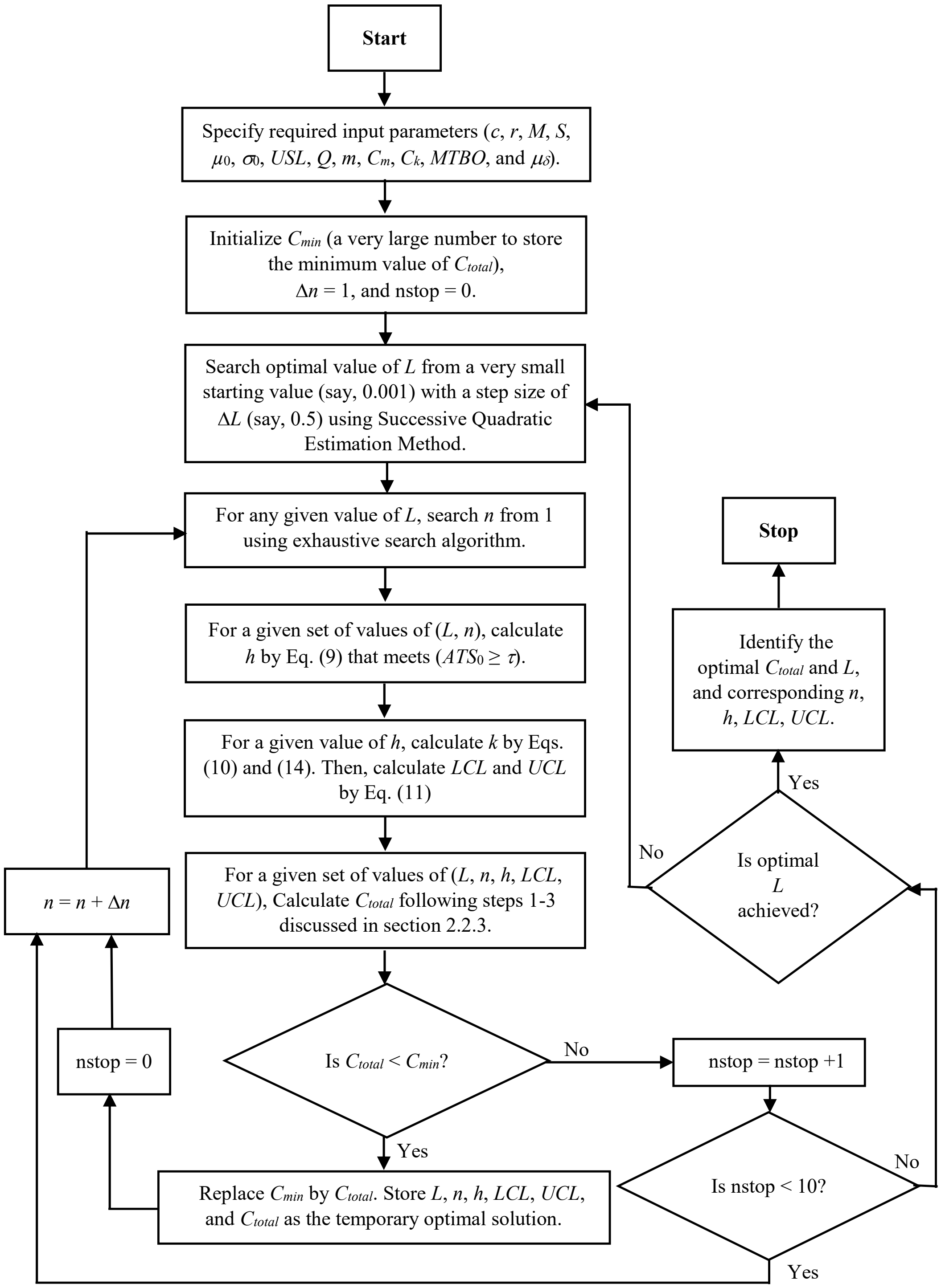


Fig. 2. Optimization algorithm of the proposed $\bar{X}$ scheme

3. Numerical studies

The following three monitoring schemes are designed, and their performances are compared in this section.

1. A basic $\bar{X}$ scheme: This is a traditional $\bar{X}$ scheme designed by considering a fixed sample size of five (n = 5) and a sampling interval of one (h = 1). The traditional $\bar{X}$ scheme is widely designed by considering a sample size of about five [29]. A fixed amount of manpower (L) is used by this scheme which is determined by Eq. (5).

2. An improved $\bar{X}$ scheme [21]. This scheme uses a fixed amount of manpower like the basic $\bar{X}$ scheme; however, the charting parameters, n, h, LCL, and UCL of this scheme are optimized to minimize the total cost, C_{total}.

3. An optimal $\bar{X}$ scheme. The manpower, L, in addition to the charting parameters, n, h, LCL, and UCL of this scheme are optimized to minimize the total cost, C_{total}.

The designs of the three monitoring schemes are conducted under standard condition (μ_0 = 0, σ_0 = 1), consequently, LSL = (-USL). Notably, the USL is expressed in terms of σ_0, and the inspection rate, m is expressed as a function of the overall effluent discharge rate, Q in this study. To facilitate the comparison study, a normalized $C_{total\ norm}$ value for each scheme is calculated by,

$$C_{total,norm} = \frac{C_{total,opt}}{C_{total}}. \tag{18}$$

where, $C_{total,opt}$ is the total cost of the optimal $\bar{X}$ scheme, and C_{total} is the total cost of other $\bar{X}$ schemes. Obviously, if the value of $C_{total,norm}$ of a scheme is smaller than one, the performance of this scheme is inferior to that of the optimal scheme, and *vice versa*.

The ratio, L_{norm} between the optimal manpower, L_{opt} of the optimal $\bar{X}$ scheme and the manpower, L of the basic (or improved) $\bar{X}$ scheme is also calculated by,

$$L_{norm} = \frac{L_{opt}}{L}. \tag{19}$$

Clearly, if the value of L_{norm} of a scheme is greater than one, the amount of manpower used by a scheme is smaller (i.e., inadequate to minimize C_{total}) than that of the optimal value, and *vice versa*.

3.1 Comparison study through factorial experiment

The performances (in terms of C_{total}) of the three schemes are compared through a 2_{IV}^{10-5} fractional factorial design with the resolution of IV [29]. In this design, all the high order interaction (three and more than three factors interaction) effects are aliased with the main effects. This means that the high order interaction effects are sacrificed in the calculation of the main effects. However, this doesn't affect the accuracy of the estimated values of the main effects as the high order effects usually have negligible impacts. The ten input parameters (c, r, τ, USL, Q, m, $MTBO$, μ_δ, C_m, and C_K) in this study are considered as the input factors and $C_{total,norm}$ is considered as the response. Each factor varies at two levels as listed in Table 2. For any given set of values of (c, r), the corresponding values (M, S) of the Burr distribution are obtained from [27].

Table 2. Factors levels

Input factor	Low level	High level
c: Burr distribution parameter	2	10
r: Burr distribution parameter	2	11
τ: Minimum allowable ATS_0	300	900
USL: Upper specification limit	$3\sigma_0$	$6\sigma_0$
Q: Overall effluent discharge rate	5	10

m: Inspection rate	$0.5Q$	$5Q$
$MTBO$: Mean time between OOC cases	100	500
μ_δ: Mean value of the mean shifts δ in the environmental process	0.5	1.5
C_m: Manpower cost for an inspector per unit time	5	20
C_K: Average penalty cost for an out-of-specification amount of the chemical elements	1000	10000

The above-mentioned three monitoring schemes are designed for each of the 32 runs resulting from the 2_{IV}^{10-5} fractional factorial experiment. The satisfaction of the constraint ($ATS_0 \geq \tau$) in designing all schemes in each run is also ensured. The resultant $C_{total,norm}$ values (see Table 3) confirm that the proposed optimal $\bar{X}$ scheme surpasses the other two $\bar{X}$ schemes for the 32 runs (except for runs 9 and 12, where the optimal manpower generated by the optimal $\bar{X}$ scheme is equal to the manpower of the improved $\bar{X}$ scheme and thus both schemes produce the same C_{total} and the value of $C_{total,norm}$ is 1). The average of the $C_{total,norm}$ values, $\bar{C}_{total,norm}$ for both basic and improved $\bar{X}$ schemes over the 32 runs (see the last row of Table 3), is found to be 0.307 and 0.640, respectively. The values of $\bar{C}_{total,norm}$ clearly demonstrates that from a holistic viewpoint (over various settings of (c, r, τ, USL, Q, m, $MTBO$, μ_δ, C_m, and C_K)), the optimal $\bar{X}$ scheme minimizes C_{total} by about 69% compared to the basic $\bar{X}$ scheme, and by about 36% compared to the improved $\bar{X}$ scheme. The reduction in C_{total} is mainly achieved by optimizing the charting parameters, n, h, LCL and UCL, and the manpower, L. Notably, the improved $\bar{X}$ scheme uses the same amount of manpower as the basic $\bar{X}$ scheme. That means the further reduction in C_{total} (69% - 36% = 33%) achieved by the optimal $\bar{X}$ scheme compared to the improved $\bar{X}$ scheme is entirely attributable to the optimization of the manpower (L), highlighting that the optimal deployment of manpower in the SPC scheme is an efficient tool to guarantee better financial results.

Finally, the grand average $\bar{L}_{norm}$ of L_{norm} values over the 32 runs is about 13.833 (see the last row of Table 3), which indicates that the basic $\bar{X}$ (or improved $\bar{X}$) scheme typically uses less manpower than the optimal value, and thus, the resulting $C_{quality}$ as well as C_{total}, is high. The allocation of optimal manpower ensures a substantial reduction in $C_{quality}$ that outweighs the increase in $C_{manpower}$ and eventually minimizes the C_{total} significantly. These findings are very important in deciding the amount of manpower utilized by an SPC scheme under limited resources.

Table 3. Comparison of the three schemes in the 2_{IV}^{10-5} experiment.

Run	Values of the input factors										$C_{total,norm}$		L_{norm}
	c	r	τ	USL	Q	m	$MTBO$	μ_d	C_m	C_K	Basic $\bar{X}$ scheme	Improved $\bar{X}$ scheme	Basic (or improved) $\bar{X}$ scheme
1	2	2	300	3σ	5	5Q	500	1.5	20	10000	0.087	0.565	8.812
2	10	2	300	3σ	5	0.5Q	100	0.5	5	10000	0.108	0.432	12.256
3	2	11	300	3σ	5	0.5Q	100	0.5	20	1000	0.309	0.991	1.221
4	10	11	300	3σ	5	5Q	500	1.5	5	1000	0.880	0.892	2.252
5	2	2	900	3σ	5	0.5Q	100	1.5	5	1000	0.046	0.728	4.443
6	10	2	900	3σ	5	5Q	500	0.5	20	1000	0.129	0.857	2.398
7	2	11	900	3σ	5	5Q	500	0.5	5	10000	0.038	0.256	33.244
8	10	11	900	3σ	5	0.5Q	100	1.5	20	10000	0.756	0.760	2.735
9	2	2	300	6σ	5	0.5Q	500	0.5	5	1000	0.237	1.000	1.000
10	10	2	300	6σ	5	5Q	100	1.5	20	1000	0.689	0.973	1.402
11	2	11	300	6σ	5	5Q	100	1.5	5	10000	0.189	0.329	18.578
12	10	11	300	6σ	5	0.5Q	500	0.5	20	10000	0.896	1.000	1.000
13	2	2	900	6σ	5	5Q	100	0.5	20	10000	0.026	0.425	19.235
14	10	2	900	6σ	5	0.5Q	500	1.5	5	10000	0.407	0.937	1.728
15	2	11	900	6σ	5	0.5Q	500	1.5	20	1000	0.379	0.462	0.213
16	10	11	900	6σ	5	5Q	100	0.5	5	1000	0.384	0.592	4.725
17	2	2	300	3σ	10	5Q	100	0.5	5	1000	0.033	0.305	45.261
18	10	2	300	3σ	10	0.5Q	500	1.5	20	1000	0.753	0.960	0.774
19	2	11	300	3σ	10	0.5Q	500	1.5	5	10000	0.276	0.478	8.774
20	10	11	300	3σ	10	5Q	100	0.5	20	10000	0.139	0.181	35.637
21	2	2	900	3σ	10	0.5Q	500	0.5	20	10000	0.036	0.582	8.536
22	10	2	900	3σ	10	5Q	100	1.5	5	10000	0.049	0.128	149.772
23	2	11	900	3σ	10	5Q	100	1.5	20	1000	0.136	0.378	14.355
24	10	11	900	3σ	10	0.5Q	500	0.5	5	1000	0.558	0.832	2.330
25	2	2	300	6σ	10	0.5Q	100	1.5	20	10000	0.107	0.679	5.369
26	10	2	300	6σ	10	5Q	500	0.5	5	10000	0.086	0.345	19.458
27	2	11	300	6σ	10	5Q	500	0.5	20	1000	0.257	0.910	1.963

28	10	11	300	6σ	10	0.5Q	100	1.5	5	1000	0.973	0.984	1.245
29	2	2	900	6σ	10	5Q	500	1.5	5	1000	0.038	0.610	7.268
30	10	2	900	6σ	10	0.5Q	100	0.5	20	1000	0.159	0.963	1.500
31	2	11	900	6σ	10	0.5Q	100	0.5	5	10000	0.048	0.322	21.004
32	10	11	900	6σ	10	5Q	500	1.5	20	10000	0.608	0.611	4.172
$\overline{C}_{total,norm}$											0.307	0.640	$\overline{L}_{norm}$ = 13.833

3.2 Sensitivity analysis of the performance of the proposed scheme

In this section, the impact of the ten factors (c, r, τ, USL, Q, m, $MTBO$, μ_δ, C_m, and C_K) on the response C_{total} of the optimal $\bar{X}$ scheme in the 2_{IV}^{10-5} factorial design conducted in Section 3.1 is investigated through analysis of variance (ANOVA). Since the replicate size is one, third-order and higher interaction effects are combined to estimate the error terms in the ANOVA analysis. A model adequacy test is performed by conducting a normality test of the C_{total} data before the ANOVA analysis. The data on C_{total} obtained from the optimal $\bar{X}$ scheme are found to be non-normal; thus, Johnson transformation is used before an ANOVA analysis is performed (see Fig. 3).

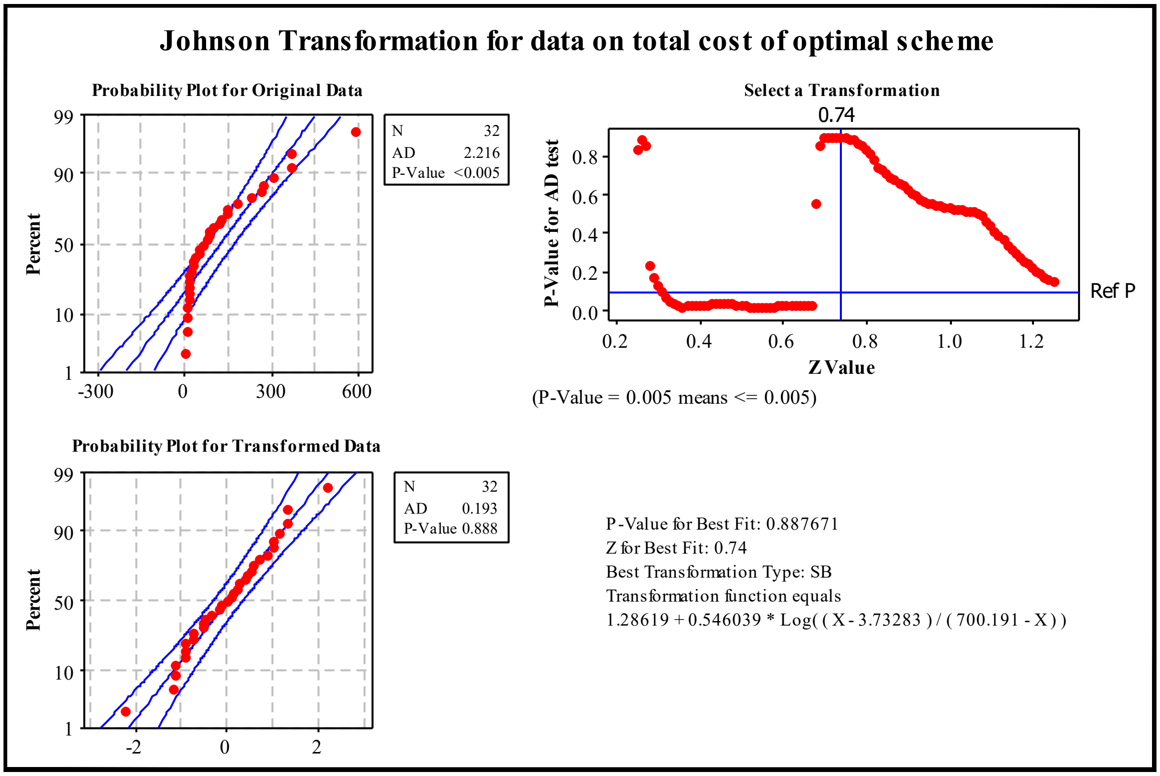


Fig. 3. Normality checks of the data on the total cost of the optimal scheme

The normal probability plot of the effects in Fig. 4 confirms that the main effects of all factors (red rectangles in Fig. 4) are statistically significant, and the two-factor interaction effects (black circles in Fig. 4) are statistically insignificant.

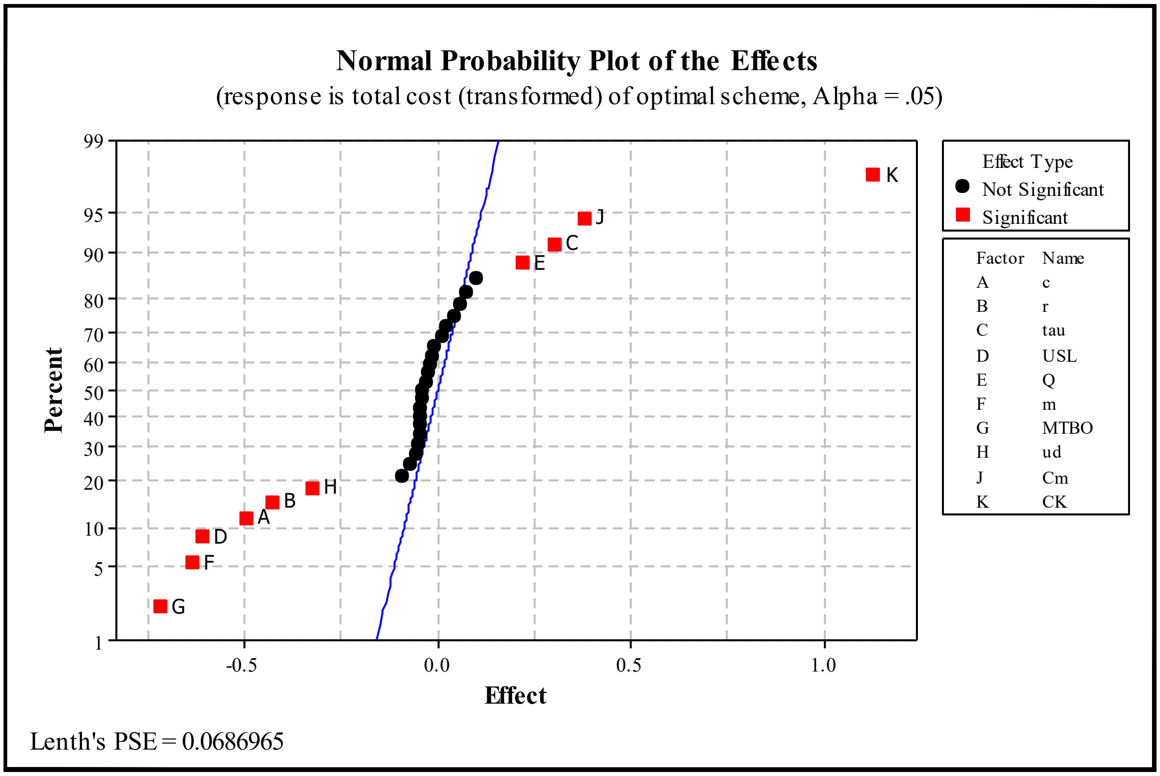


Fig. 4. Normal probability plot of the effects

The main effect plots (Fig. 5) demonstrate that the total cost, C_{total} of the optimal $\bar{X}$ scheme is negatively affected by the factors *c*, *r*, *USL*, *m*, *MTBO* and μ_d. This implies that a larger value of *c*, *r*, *USL*, *m*, *MTBO,* and/or μ_d can result in a smaller amount of C_{total}, and *vice versa*. On the contrary, the C_{total} of the optimal $\bar{X}$ scheme is positively affected by τ, *Q*, C_m, and C_K. This means that a larger value of τ, *Q*, C_m, and/or C_K produces a larger amount of C_{total}, and *vice versa*.

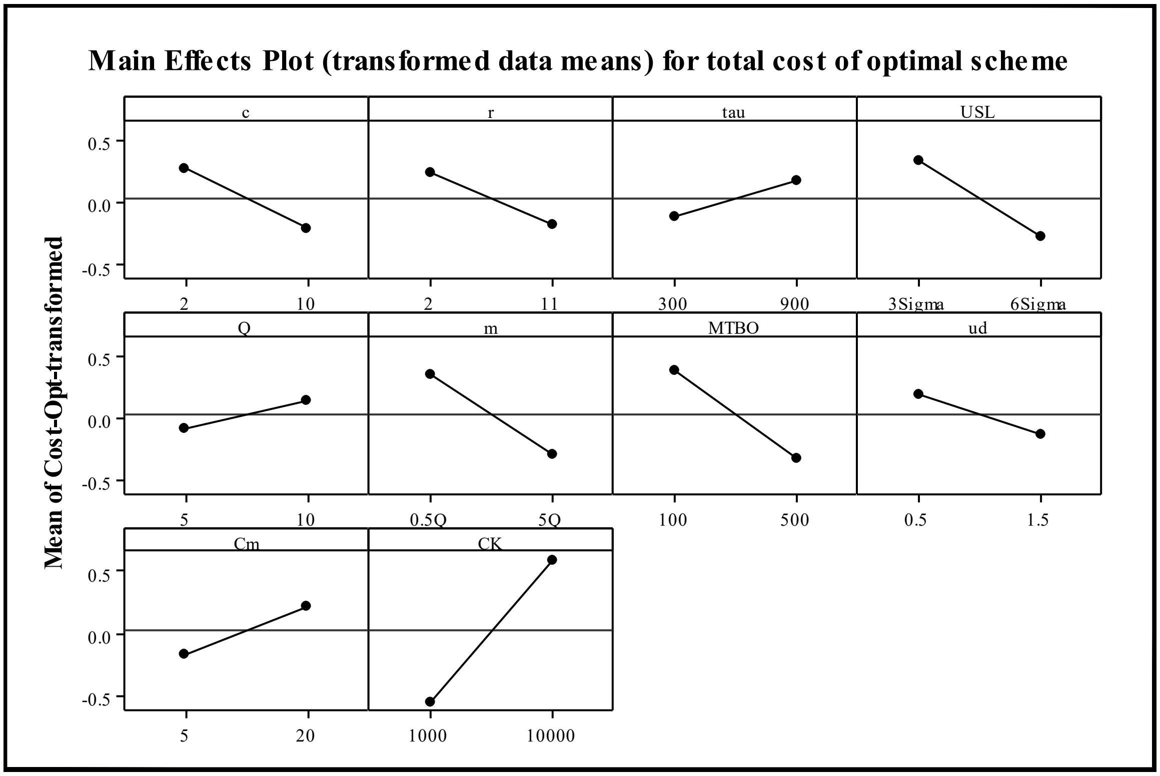


Fig. 5. Main effects plot

3.3 Evaluation of performance of the proposed scheme under different distribution

This section investigates the impact of different types of unimodal distribution (as defined by skewness and kurtosis) on the performance of the proposed optimal $\bar{X}$ scheme with respect to that of the basic and improved $\bar{X}$ schemes as explained in Section 3.1. The study is conducted for a general case, where the values of input parameters (USL, Q, m, $MTBO$, μ_δ, C_m, and C_K) are set as the mid values of the low and high levels (see Table 2) of the input parameters which gives: $USL = 4.5$, $Q = 7.5$, $m = 2.75Q$, $MTBO = 600$, $\mu_\delta = 1.0$, $C_m = 12.5$, and $C_K = 5050$. The values of the coefficient of skewness (α_3), kurtosis (α_4) and the fitted Burr distribution parameters (c, r) are decided according to [25]. The impact of different distributions on the performance of the proposed optimal $\bar{X}$ scheme is investigated under four scenarios (see Table 4). In Scenario 1, the value of α_4 is almost fixed at about 3, and the value of α_3 increases from negative to positive value. It is worthy to note that the reduction in total cost, C_{total} achieved by the optimal $\bar{X}$ scheme compared to both the basic and improved $\bar{X}$ schemes also increases with increasing α_3 under this scenario as evidenced by the ratio, $C_{total,norm}$ (Eq. (21)). The overall cost reduction obtained by the optimal $\bar{X}$ scheme compared to the basic and improved $\bar{X}$ schemes is about 71.3% and 40.5%, respectively. In Scenario 2, the value of α_3 is almost equal to zero, and the value of α_4 increases. The reduction in C_{total} achieved by the optimal $\bar{X}$ scheme compared to the basic and improved $\bar{X}$ schemes is similar to Scenario 1, and the overall cost reduction given by the optimal $\bar{X}$ scheme compared to the basic and improved $\bar{X}$ schemes is about 76.4% and 40.6%, respectively. In Scenario 3, the value of α_3 is almost equal to 1, and the value of α_4 increases from 4.443 to 7.215. A significant reduction in C_{total} attained by the optimal $\bar{X}$ scheme compared to the basic and improved $\bar{X}$ schemes is found under this scenario, where the overall cost reduction obtained by the optimal $\bar{X}$ scheme compared to the basic and improved $\bar{X}$ schemes is about 89.6% and 45.7%, respectively. In Scenario 4, the value of α_3 increases from 1.432 to 3.810, and the value of α_4 increases from 7.356 to 38.670. A further reduction in C_{total} by the optimal $\bar{X}$ scheme compared to the basic and improved $\bar{X}$ schemes is found under this scenario, where the overall cost reduction given by the optimal $\bar{X}$ scheme compared to the basic and improved $\bar{X}$ schemes is about 94.1% and 46%, respectively. It can be concluded from this study that any increase in either α_3 or α_4, or both leads to a better performance of the optimal scheme (in terms of cost reduction) compared to its competitors. However, when the values of both α_3 and α_4 simultaneously increase, the performance of the optimal $\bar{X}$ scheme is significantly superior to the situation where only one of them increases. It can also be concluded that a large amount of manpower (see L_{norm} as calculated by (Eq. (22)) should be deployed to the $\bar{X}$ scheme if α_3 and/or α_4 is high in order to maximize the benefit.

Table 4. Impact of different types of distribution on the performance of the optimal $\bar{X}$ scheme

Scenarios	Coefficients of skewness (α_3) and kurtosis (α_4)		Burr distribution parameters		$C_{total,norm}$		L_{norm}
	α_3	α_4	c	r	Basic $\bar{X}$ scheme	Improved $\bar{X}$ scheme	Basic (or improved) $\bar{X}$ scheme
Scenario 1	-0.254	3.027	6	11	0.425	0.625	4.502
	-0.147	3.065	6	6	0.316	0.612	5.013
	-0.013	3.010	5	6	0.267	0.598	5.417
	0.040	3.070	5	5	0.241	0.592	5.624
	0.136	2.979	4	7	0.246	0.580	5.749
	0.329	3.006	3	11	0.230	0.565	6.018
				$\bar{C}_{total,norm}$	0.287	0.595	$\bar{L}_{norm}$ = 5.387
Scenario 2	0.050	2.866	4	11	0.290	0.590	5.406
	-0.013	3.010	5	6	0.267	0.598	5.417
	0.040	3.070	5	5	0.241	0.592	5.624
	-0.019	3.169	6	4	0.236	0.598	5.548
	0.005	3.329	7	3	0.207	0.596	5.763
	0.044	3.646	10	2	0.172	0.591	6.109
				$\bar{C}_{total,norm}$	0.236	0.594	$\bar{L}_{norm}$ = 5.645
Scenario 3	0.958	4.443	2	8	0.130	0.536	7.445
	1.014	4.707	2	7	0.121	0.536	7.586
	1.094	5.118	2	6	0.110	0.536	7.732
	0.956	5.937	4	2	0.086	0.556	7.716
	1.060	7.215	9	1	0.075	0.555	8.096
				$\bar{C}_{total,norm}$	0.104	0.543	$\bar{L}_{norm}$ = 7.715
Scenario 4	1.060	7.215	9	1	0.075	0.555	8.096
	1.432	7.356	2	4	0.080	0.537	8.319
	1.909	12.460	2	3	0.060	0.541	8.809
	2.485	29.560	5	1	0.046	0.557	9.062
	2.940	19.760	1	9	0.052	0.521	10.016
	3.810	38.670	1	6	0.043	0.532	10.259
				$\bar{C}_{total,norm}$	0.059	0.540	$\bar{L}_{norm}$ = 9.094

This section further investigates the impact of some well-known unimodal distributions such as normal, log-normal, beta, and gamma distributions on the performance of the proposed optimal $\bar{X}$ scheme when Burr distribution is used to approximate them. Based on the given parameter values of a specific distribution, the corresponding values of α_3 and α_4 are estimated, and the fitted values of the Burr distribution parameters (c, r) are approximated from [27] and [35]. The three SPM schemes are designed, and the results are shown in Table 5. It can be seen from Table 5 that the optimal $\bar{X}$ scheme consistently outperforms the basic and improved $\bar{X}$ schemes in all cases, which clearly demonstrates the robustness of the performance of the optimal $\bar{X}$ scheme under different distributions. It can also be concluded that the Burr distribution is an excellent candidate for approximating a variety of unimodal distributions.

Table 5. Impact of some specific distributions on the performance of the optimal $\bar{X}$ scheme

Distribution type	Distribution parameters	Coefficients of skewness (α_3) and kurtosis (α_4)		Burr distribution parameters		$C_{total,norm}$		L_{norm}
		α_3	α_4	c	r	Basic $\bar{X}$ scheme	Improved $\bar{X}$ scheme	Basic (or improved) $\bar{X}$ scheme
Normal (μ, σ^2)	$\mu = 0$, $\sigma = 1$	0.000	3.000	4.87372	6.15757	0.26572	0.59661	5.42737
Log-normal (μ, σ^2)	$\mu = 0$, $\sigma = 0.25$	0.778	4.096	2.25406	7.29246	0.14338	0.54218	7.18422
	$\mu = 0$, $\sigma = 0.5$	1.750	8.898	1.38966	7.59131	0.07674	0.52205	8.86727
	$\mu = 0$, $\sigma = 1.0$	3.263	26.540	1.00000	7.00000	0.04649	0.52700	10.18190
Beta (a, b)	$a = 2$, $b = 4$	0.468	2.625	-9.61779	0.08367	0.50702	0.99051	1.14892
	$a = 3$, $b = 3$	0.000	2.333	-12.13408	0.12032	0.52894	0.79146	2.05580
	$a = 4$, $b = 2$	-0.468	2.625	10.00000	7.00000	0.51348	0.64484	4.06167
Gamma (k_s, θ)	$k_s = 2$, $\theta = 2$	1.414	6.000	1.43909	14.54115	0.10606	0.51875	8.23369
	$k_s = 3$, $\theta = 2$	1.155	5.000	1.50455	24.62936	0.12799	0.51977	7.83376
	$k_s = 5$, $\theta = 1$	0.894	4.200	2.01150	9.19059	0.13975	0.53625	7.30124

3.4 Performance comparison under different distribution of the process shift

The proposed model in this study assumes a Rayleigh distribution in capturing the random characteristics of the environmental process. As mentioned earlier, Rayleigh distribution is well accepted to the research community in characterizing the randomness of a process. However, the process distribution may be quite different from the assumed Rayleigh distribution in many processes. Thus, the performance (in terms of cost reduction) of the proposed optimal $\bar{X}$ scheme is investigated under five different scenarios in this section. In Scenario 1, the optimization design is carried out assuming a Rayleigh distribution, whereas uniform distribution is used in Scenario 2. The study further conducts the optimization design assuming beta distribution with different shape parameters (Scenarios 3, 4, and 5). The skewness of a beta distribution depends on the parameters a and b [38]. The probability density functions of the uniform and beta distributions are given by Eqs. (20) and (21), respectively.

$$f(\delta) = \frac{1}{\delta_{\max}} \tag{20}$$

$$f(\delta) = \frac{\Gamma(a+b)}{\Gamma(a)\cdot\Gamma(a)} \cdot \frac{\delta^{a-1}\cdot(\delta_{\max}-\delta)^{b-1}}{\delta_{\max}^{a+b-1}} \tag{21}$$

The study is conducted for the general case used in Section 3.3 (i.e., c = 6, r = 6, USL = 4.5, Q = 7.5, m = 2.75Q, $MTBO$ = 600, μ_δ = 1.0, C_m = 12.5, and C_K = 5050). The three monitoring schemes are designed under the five scenarios and the results are listed in Table 6. The values of $\delta_{\max}$ in Eqs. (20) and (21) for the given value of (μ_δ = 1.0) is calculated by Eq. (13), which ensures the use of same $\delta_{\max}$ values in all five scenarios.

Table 6. Performance of the optimal $\bar{X}$ scheme under different distributions of δ shift

Scenarios	Distribution	Distribution parameters	$C_{total,norm}$		L_{norm}
			Basic $\bar{X}$ scheme	Improved $\bar{X}$ scheme	Basic (or improved) $\bar{X}$ scheme
Scenario 1	Rayleigh	μ_δ = 1.0	0.316	0.612	5.013
Scenario 2	Uniform	$\delta_{\max}$ = 2.97	0.512	0.660	6.928
Scenario 3	Beta skewed to right ($a < b$)	a = 2, b = 4	0.309	0.603	5.367
Scenario 4	Beta symmetrical ($a = b$)	a = 3, b = 3	0.569	0.684	3.411
Scenario 5	Beta skewed to left ($a > b$)	a = 4, b = 2	0.731	0.734	2.709

As shown in Table 6, the superiority of the optimal $\bar{X}$ scheme is more significant when the process shift distribution is skewed to the right. It is noted that Scenario 1 (Rayleigh distribution) and Scenario 3 (beta ($a < b$) distribution) both are skewed to right and show similar performance to some extent. On the other hand, Scenario 2 (uniform distribution) and Scenario 4 (beta ($a = b$) distribution) show somewhat similar performance. When the distribution is skewed to the left (beta ($a > b$) distribution), the improved $\bar{X}$ scheme provides almost no superiority over the basic $\bar{X}$ scheme. This is because both improved and basic $\bar{X}$ schemes use the same fixed amount of manpower, which is inadequate to minimize C_{total} significantly. This finding further highlights the importance of optimal deployment of manpower to the SPM scheme. Finally, it can be concluded that the significant reduction in C_{total} as evidenced by the ratio, $C_{total,norm}$ (Eq. (18)) in Table 5 indicates that under any probability distributions of the process shift δ, the proposed optimal $\bar{X}$ scheme always outperforms its competitors.

4. Illustrative example

This section illustrates the proposed monitoring scheme's design and application using real-life environmental data. The Ministry of the Environment in Canada requires industrial facilities to report wastewater discharges as part of the Effluent Monitoring and Effluent Limits Regulations under the Environmental Protection Act, known as the MISA program [39]. Many companies must pretreat their wastewater before putting it into the sewer system and they should comply with discharge requirements including reporting and monitoring procedure. At the same time, facilities that violate environmental rules might be penalized. The environmental penalty may vary from \$1,000 to \$100,000 per day based on the severity of the violation. For example, in

2019, an iron and steel manufacturing company was ordered to pay $10,000 for violating the zinc concentration limit [39].

The aim here is to screen the zinc level in the wastewater from a steel manufacturing plant. The steel industry uses a large quantity of water. Water is utilized in cooling operations and processes, such as descaling and dust scrubbing. Different types of water are utilized in steelmaking processes. Freshwater is primarily used for direct and indirect cooling processes, while seawater is typically utilized for once-through cooling after an antifouling pretreatment. The average water intake for integrated steelworks is 28.6 m^3 per ton of produced steel, with an average water discharge of 25.3 m^3 per ton of steel. Wastewaters from the steel industry usually contain zinc in high ratios. Excessive zinc content in the water may negatively affect human and animal health, and over a certain boundary concentration, zinc may even be toxic. It can cause nausea, vomiting, dizziness, fevers, and diarrhea. On top of that, zinc can be combined with sulphuric acid resulting in zinc sulfate, which causes acute toxicity leading to stomach aches and vomit. Other zinc compounds that can be highly hazardous are zinc arsenate and zinc cyanide. As the zinc level in the soil increases mainly from the disposal of zinc waste from metal manufacturing industries, it is necessary to continuously monitor the zinc content of the wastewater emitted from metal industries (a steel plant in this example) to assure that the source complies with the regulatory limits over time.

4.1 Data collection and model adequacy test

A dataset of 100 observations on the zinc content per liter (gm/L or ppm) of wastewater emitted from a steel manufacturing plant were collected, over the period from 1 Feb 2016 to 10 May 2016 to design the proposed monitoring scheme [40]. The data were then statistically analyzed and found to be non-normal (p-value < 0.005) with skewness and kurtosis of $\hat{\alpha}_3$ = 0.78022 and $\hat{\alpha}_4$ = 3.60199, respectively, as shown in Fig. 6. The data were further analyzed to check the independence (autocorrelation) assumption as any autocorrelation adversely impacts the performance of the monitoring scheme. To verify the autocorrelation among the data, an autocorrelation function (ACF) was drawn as indicated in Fig. 7. The resulting ACF confirms that the zinc content data are independent.

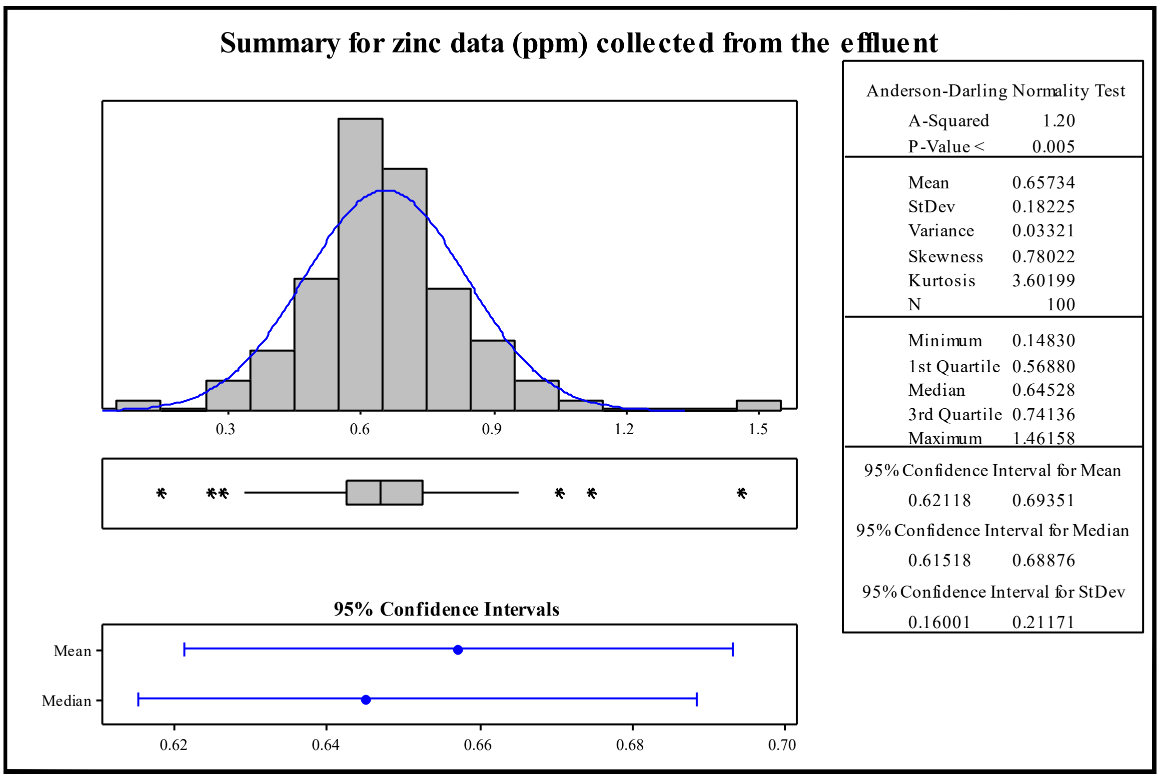


Fig. 6. Summary of statistical analysis of the zinc data

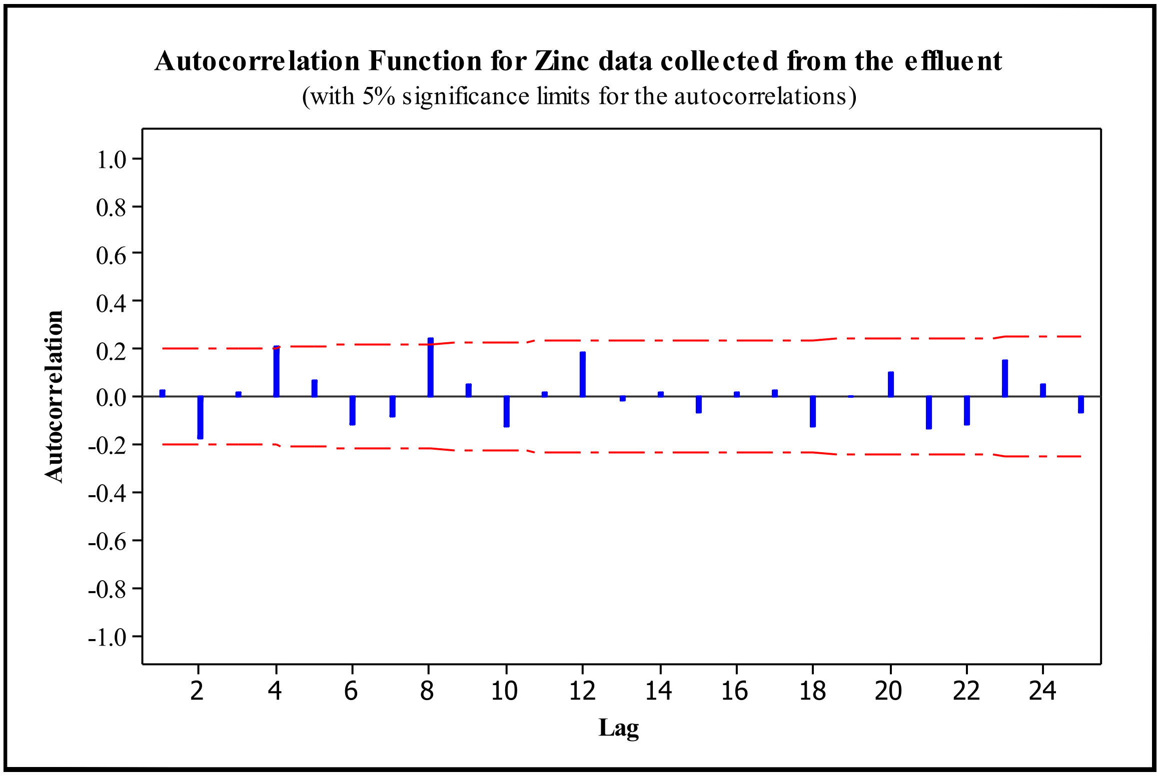


Fig. 7. Autocorrelation function of the zinc data

4.2 Design and application of the proposed monitoring scheme

The design of any monitoring scheme is usually completed over two phases: Phase I and Phase II. In Phase I design, 25 to 30 samples, each of 4 or 5 items, are typically suggested in designing a classical $\bar{X}$ scheme [29]. The objective of Phase I is to estimate the IC mean (μ_0) and standard deviation (σ_0) of the environmental process so that the monitoring scheme can be developed. In Phase II, the scheme resulting from Phase I is utilized to monitor the future environmental process.

The collected 100 observations (25 samples with a sample size of four) of zinc content data were utilized to design the basic $\bar{X}$ scheme in Phase I. The estimated skewness and kurtosis ($\hat{\alpha}_3$ = 0.78022 and $\hat{\alpha}_4$ = 3.60199) of individual data are approximately fitted with the Burr distribution with c = 1.976147, r = 23.178840, M = 0.183700, and S = 0.099616 (see [35], Table I, where the closest values are α_3 = 0.75 and α_4 = 3.6). The Burr distribution parameters along with other required design parameters are summarized below:

- c (Burr distribution parameter) = 1.976147
- r (Burr distribution parameter) = 23.178840
- M (Mean of the Burr distribution) = 0.183700
- S (Standard deviation of the Burr distribution) = 0.099616
- τ (Minimum allowable ATS_0 (day)) = 400
- USL (Upper specification limit (ppm)) = 1.0
- Q (Average amount (ppm) of zinc discharge from the plant per day) = 6.5734
- m (Maximum allowable inspection rate (per day) = 8
- C_m (Manpower cost for an inspector (\$/day)) = 100
- C_K (Average penalty cost for an out-of-specification quantity of zinc (\$)) = 10000
- μ_δ (Mean of the δ values in the wastewater discharge process) = 1.0
- $MTBO$ (Mean time between OOC cases (days)) = 90

Based on the above-mentioned design specifications, the basic $\bar{X}$ scheme is designed as shown in Fig. 8. It is clear that all the 25 sample points are plotted within the UCL and LCL of the monitoring scheme. This affirms that the zinc discharge process is statistically in control (μ_0 = 0.65734 and σ_0 = 0.18225).

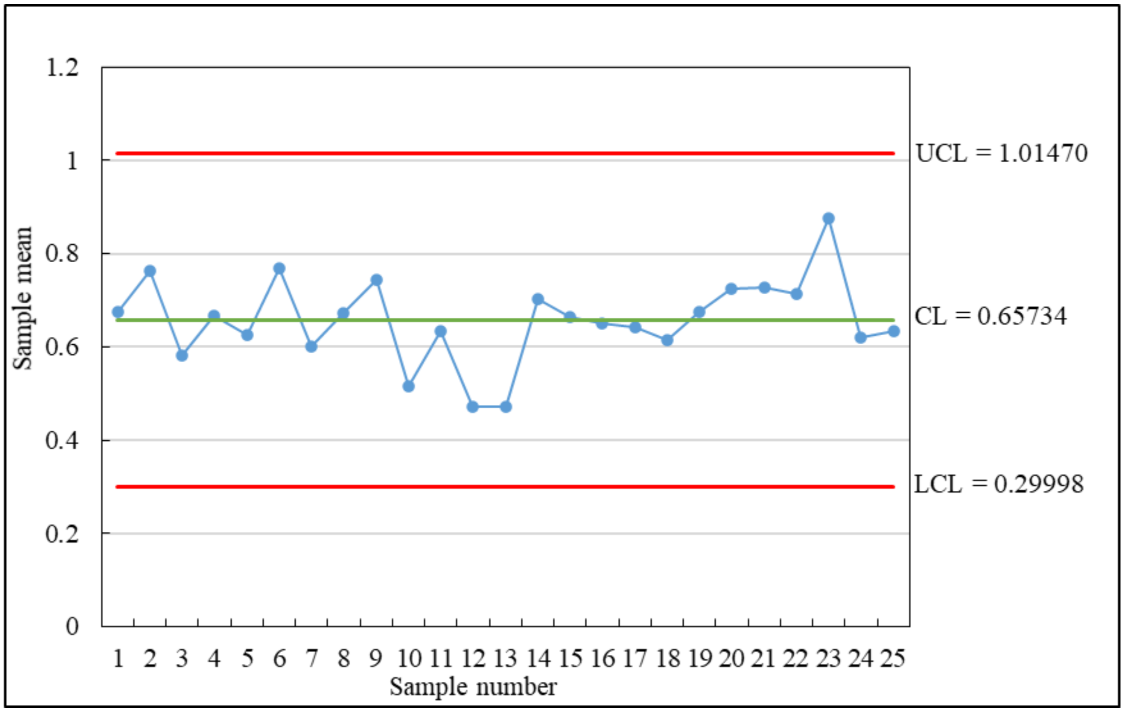


Fig. 8. Basic $\bar{X}$ scheme (Phase I)

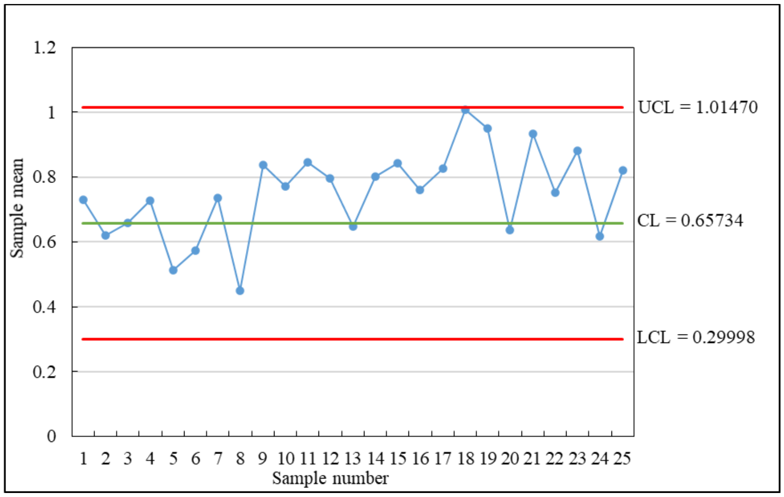
The IC process parameters values (μ_0 = 0.65734 and σ_0 = 0.18225) and other required design parameters values estimated in Phase I were then utilized in Phase II to design an improved $\bar{X}$ scheme and an optimal $\bar{X}$ scheme in addition to the basic $\bar{X}$ scheme for monitoring the quantity of zinc discharged from the environmental process. Table 7 displays the parameters of the three monitoring schemes. All three monitoring schemes satisfy the constraint on the ATS_0 value of 400.

Table 7. Parameters of the three monitoring schemes

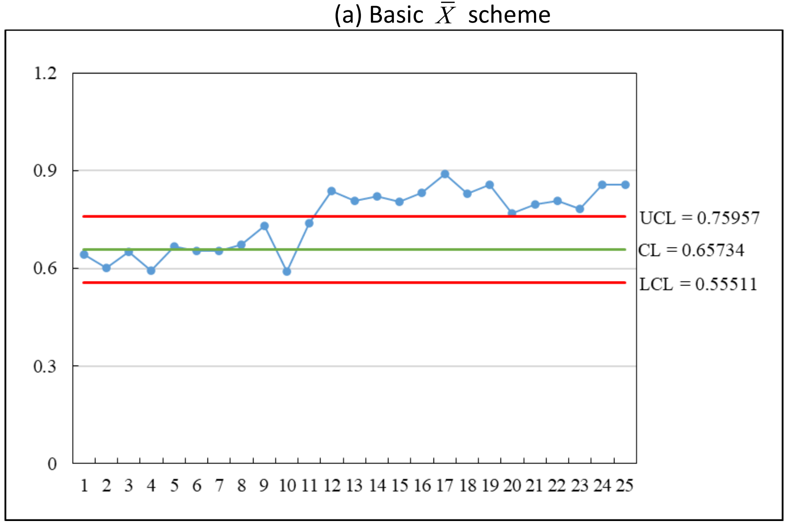
Scheme	Manpower, L	Sample size, n	Sampling interval (day), h	Control limits		Cost ($)		
				LCL	UCL	C_{man}	$C_{quality}$	C_{total}
Basic $\bar{X}$ scheme	0.5000	4	1	0.29998	1.01470	50.00	938.10	988.10
Improved $\bar{X}$ scheme	0.5000	27	6.75	0.55511	0.75957	50.00	334.39	384.39
Optimal $\bar{X}$ scheme	1.1447	40	4.37	0.56645	0.74823	114.47	215.92	330.39

As Table 7 shows, the basic $\bar{X}$ scheme's total cost (C_{total}) is quite high compared to that of the improved and optimal $\bar{X}$ schemes. This is because the amount of manpower (L) deployed to the basic $\bar{X}$ scheme is smaller than the optimal value, resulting in high quality cost ($C_{quality}$) and, consequently, high C_{total}. On the other hand, even though the improved $\bar{X}$ scheme used the same amount of manpower as the basic $\bar{X}$ scheme, it optimizes $C_{quality}$, and the resultant C_{total} is minimized to some degree. However, the optimal $\bar{X}$ scheme not only optimizes $C_{quality}$, but also utilizes the optimal amount of manpower, and thus, C_{total} is minimized further. Overall, the optimal $\bar{X}$ scheme reduces C_{total} by about 67% and 14% compared to the basic $\bar{X}$ and improved $\bar{X}$ schemes, respectively, in this particular example.

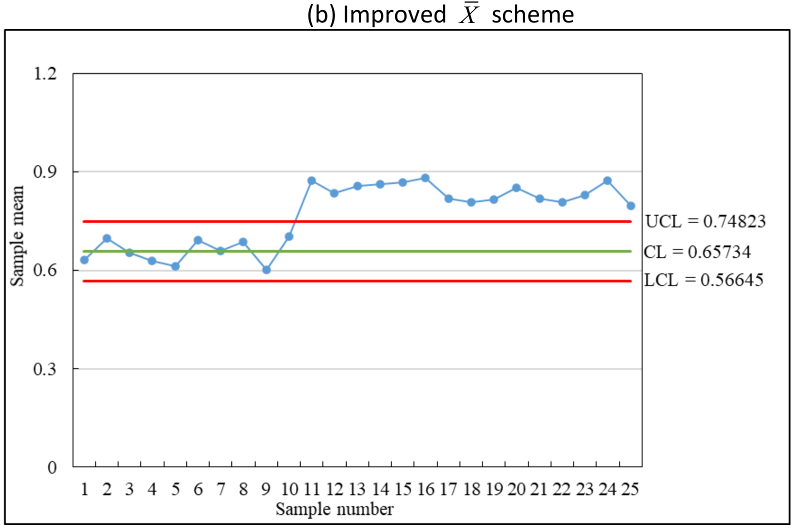
Finally, the effectiveness of the three monitoring schemes for detecting OOC cases is demonstrated through a simulation study. A simulation study is widely used in the literature to demonstrate the effectiveness of the proposed models when actual data are unavailable [10,41]. In this simulation study, 25 sample data on the quantity of zinc discharge are generated for the illustrations. The first 10 sample data are generated assuming that the process is in IC state, and the other 15 sample data are generated considering a shift of size $1.0\sigma_0$ in the mean of the environmental process. All 25 simulated sample data are plotted on the three monitoring schemes (see Fig. 9(a-c)).



(a) Basic $\overline{X}$ scheme



(b) Improved $\overline{X}$ scheme

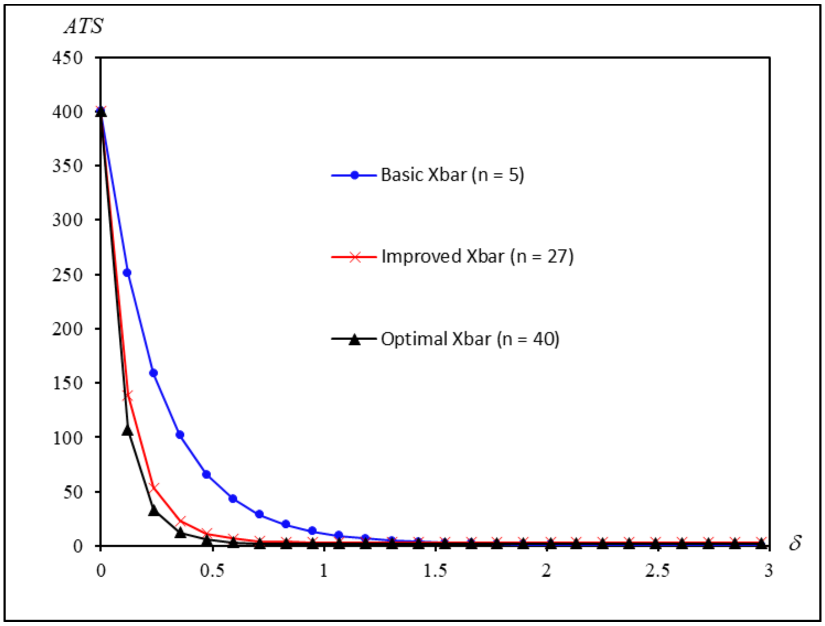


(c) Optimal $\overline{X}$ scheme

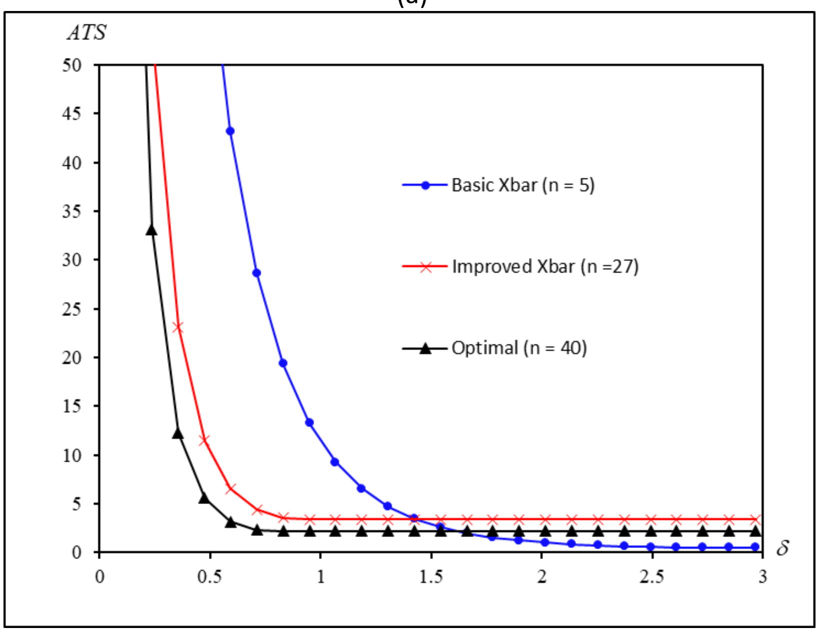
Fig. 9. The three SPM schemes in the case study (Phase II)

As shown in Fig. 9(a), all 25 sample points are plotted within the control limits of the basic $\bar{X}$ scheme, indicating that the basic $\bar{X}$ scheme is unable to identify any OOC cases. In comparison, the improved $\bar{X}$ scheme (Fig. 9(b)) detects the OOC cases by the 12th sample point, whereas the optimal $\bar{X}$ scheme (Fig. 9(c)) detects the OOC cases immediately (by the 11th sample point), which shows the superiority of the optimal $\bar{X}$ scheme over the basic and improved $\bar{X}$ schemes in identifying the OOC cases. It is worth noting that in both the improved and optimal $\bar{X}$ schemes, the charting parameters (*n, h, LCL, UCL*) are optimized to minimize the total cost (C_{total}) based on the optimization algorithm proposed in this study. However, the improved $\bar{X}$ scheme utilizes the fixed amount of manpower (*L*) like the basic $\bar{X}$ scheme, whereas the optimal $\bar{X}$ scheme has the merit of allocating the optimal amount of manpower to minimize the C_{total}. As a result, the optimal $\bar{X}$ scheme can take larger samples more frequently, and thus, the quality cost ($C_{quality}$) as well as the C_{total} are reduced substantially (see Table 7).

The *ATS* profiles of the three monitoring schemes are illustrated in Fig. 10 (a-b) to have better understanding of the effectiveness of the charts at the different shift points. Fig. 10(a) shows the *ATS* profiles over nearly the whole mean shift range, where at $\delta = 0$, the IC ATS_0 value of the three schemes are almost equal to the specified value τ (= 400). Fig. 10(b) zooms in the *ATS* profiles for moderate and large mean shifts. It can be observed that the optimal $\bar{X}$ scheme shows the best performance for $\delta \leq 1.7$.



(a)



(b)

Fig. 10. Comparison of *ATS* profile among the three charts ($\tau = 400$)

5. Conclusions and future work

One major contributor to environmental pollution is the rapid industrialization around the world. The industries consume more water compared to households and produce a considerable amount of effluent of contaminated wastewater containing toxic heavy metal pollutants like lead, chromium, mercury, thallium, and zinc. The Environmental Protection Agency (EPA) has imposed different legislation to protect the environment, and the industries nowadays are under increasing pressure to meet the stringent standards set by the EPA before discharging harmful wastewater into the environment. To avoid the penalty associated with the discharge of excessive harmful pollutants in the wastewater and to safeguard public health and the environment, continuous and effective monitoring of the pollutants in the wastewater is highly important. In such perspectives, the SPM schemes can play a vital role. However, the real-life environmental data are often skewed and do not follow normal distribution. Classical SPM schemes are designed based on normality assumptions and cannot be used directly to monitor environmental processes. The current study develops a model for the optimization design of the $\bar{X}$ scheme for monitoring non-normal environmental processes. The model identifies the optimal parameters of the monitoring scheme, including the appropriate amount of manpower that should be deployed to the SPM activities.

The proposed optimization design minimizes the total cost, including manpower and quality costs, and in the meantime ensures that the false alarm rate of the monitoring scheme is maintained at an acceptable level. The comparison studies based on a fractional factorial experiment demonstrates that the proposed $\bar{X}$ scheme reduces the total cost by about 69% compared to the basic $\bar{X}$ scheme and about 36% compared to the improved $\bar{X}$ scheme. The impact of the input parameters on the performance of the optimal $\bar{X}$ scheme is also examined through a sensitivity study. The robustness of the performance of the optimal $\bar{X}$ scheme is further investigated under different non-normal, and process shift distributions. Lastly, the design and application of the proposed $\bar{X}$ scheme are validated using actual data on the quantity of zinc in the wastewater discharged from a steel manufacturing plant. The proposed model is expected to ensure that the zinc content in the wastewater meets the quality standard over time and broaden the literature applying SPM tools to manage the quality of the environment. Even though the design and application of the proposed model are illustrated using non-normal industrial environmental data, the same monitoring scheme is equally applicable for monitoring normal data in different manufacturing processes and service industries.

This study focuses on the optimization design of the Phase II control chart and uses a fixed sample size of five in Phase I for illustration. However, it might be interesting to investigate the impact of the sample size used in Phase I on the performance of the proposed optimal $\bar{X}$ scheme. A future study can also design EWMA or CUSUM scheme to effectively monitor the non-normal environmental processes, and the detection power of these schemes can be investigated based on the statistical and economic properties.

Appendix

The Burr distribution [27] is widely used to approximate a variety of non-normal unimodal distributions. The density function, $f(y)$, and the cumulative distribution function, $F(y)$ of the Burr distribution is given by,

$$\begin{aligned} f(y) &= \frac{cry^{c-1}}{(1+y^c)^{r+1}},\ y \ge 0;\ c, r \ge 1 \\ F(y) &= 1 - \frac{1}{(1+\max\{0, y\}^c)^r}. \end{aligned} \tag{A.1}$$

The parameters (c, r) in Eq. (A.1) represent a wide range of skewness (α_3) and kurtosis (α_4) coefficients of different probability density functions, including normal distribution. Burr [27,35] tabulated different combinations of values of (c, r) for estimating α_3 and α_4 of the individual measurement. If X_1, X_2, …, X_n are the observations of a sample of random variable X of size n, then the estimated values of the coefficients of skewness ($\hat{\alpha}_3(\bar{X})$) and kurtosis ($\hat{\alpha}_4(\bar{X})$) of the sample mean, $\bar{X}$ can be obtained as follows [23]:

$$\begin{aligned} \hat{\alpha}_3(\bar{X}) &= \frac{\hat{\alpha}_3}{\sqrt{n}} \\ \hat{\alpha}_4(\bar{X}) &= \frac{\hat{\alpha}_4 - 3}{n} + 3. \end{aligned} \tag{A.2}$$

The tabulated values in [27,35] permit a user to make a standardized transformation between a Burr variable, *Y*, and the sample mean $\bar{X}$ of any random variable *X* (assuming *X* and *Y* have the same values of α_3 and α_4) as follows:

$$\frac{Y - M}{S} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \tag{A.3}$$

where *M* and *S* stand as the mean and standard deviation of the Burr variable *Y*, and μ and σ refer to the population mean and standard deviation of the random variable *X*. For any given group of values of (*c*, *r*), the values of *M* and *S* can be obtained from [27,35].

Acknowledgements

This article is a part of an ongoing research project titled "Statistical monitoring of carbon emissions in industry using the Shewhart-EWMA scheme," conducted by the Sustainable Engineering Asset Management research group at the University of Sharjah, UAE.

Funding body

This work is supported by the University of Sharjah under Competitive Research Grant (Project No. 200204051161, 2020).

Disclosure statement

No potential conflict of interest exists.

References

[1] I. Kamika, M.N. Momba, Assessing the resistance and bioremediation ability of selected bacterial and protozoan species to heavy metals in metal-rich industrial wastewater. BMC Microbio. 13 (1) (2013) 1-14, https://doi.org/10.1186/1471-2180-13-28.

[2] S. Chetty, L. Pillay, Assessing the influence of human activities on river health: a case for two South African rivers with differing pollutant sources. Environ. Monit. Assess. 191 (3) (2019) 1-11, https://doi.org/10.1007/s10661-019-7308-4.

[3] J.P. Vareda, A.J.M. Valente, L. Duraes, Assessment of heavy metal pollution from anthropogenic activities and remediation strategies: a review, J. Environ. Manage. 246 (2019) 101-118, https://doi.org/10.1016/j.jenvman.2019.05.126.

[4] E. Iloms, O.O. Ololade, H.J.O. Ogola, R. Selvarajan, R. (2020). Investigating industrial effluent impact on municipal wastewater treatment plant in Vaal, South Africa, Int. J. Environ. Res. Publ. Health 17 (3) (2020) 1096, https://doi.org/10.3390/ijerph17031096.

[5] M. Métadier, J.-L. Bertrand-Krajewski, The use of long-term on-line turbidity measurements for the calculation of urban stormwater pollutant concentrations, loads, pollutographs and intra-event fluxes, Water Res. 46 (20) (2012) 6836-6856, https://doi.org/10.1016/j.watres.2011.12.030.

[6] D. Mullins, D. Coburn, L. Hannon, E. Jones, E. Clifford, M. Glavin, A novel image processing-based system for turbidity measurement in domestic and industrial wastewater, Water Sci. Techno. 77 (5) (2018) 1469-1482, https://doi.org/10.2166/wst.2018.030.

[7] C.N. Madu, Managing green technologies for global competitiveness, Praeger Publishers Inc, Westport, United States, 1996.

[8] C.J. Corbett, J.N. Pan, Evaluating environmental performance using statistical process control techniques. Euro. J. Opera. Res. 139(1) (2002) 68-83, https://doi.org/10.1016/S0377-2217(01)00155-2.

[9] J.N. Pan, S.T. Chen, The economic design of CUSUM chart for monitoring environmental performance, Asia Pacific Manage. Rev. 10 (2005) 155-161.

[10] V. Leiva, C. Marchant, F. Ruggeri, H. Saulo, A criterion for environmental assessment using Birnbaum-Saunders attribute control charts, Environmetrics 26(7) (2015) 463-476, https://doi.org/10.1002/env.2349.

[11] C. Marchant, V. Leiva, G. Christakos, M.F. Cavieres, Monitoring urban environmental pollution by bivariate control charts: new methodology and case study in Santiago, Chile. Environmetrics 30(5) (2019), https://doi.org/10.1002/env.2551.

[12] N. Abbas, U. Saeed, M. Riaz, Assorted control charts: An efficient statistical approach to monitor p^H values in ecotoxicology lab, J. Chemometrics 33(6) (2019) 1-20, https://doi.org/10.1002/cem.3129.

[13] R.A. Sanusi, Z.L. Chong, A, Mukherjee, M. Xie, Distribution-free hybrid schemes for process surveillance with application in monitoring chlorine content of water, Chemomet. Intelli. Lab. Sys. 206 (2020) 104099, https://doi.org/10.1016/j.chemolab.2020.104099.

[14] T. Abbas, T. Mahmood, M. Riaz, M. Abid, Improved linear profiling methods under classical and Bayesian setups: An application to chemical gas sensors, Chemomet. Intelli. Lab. Sys. 196 (2020) 103908, https://doi.org/10.1016/j.chemolab.2019.103908.

[15] C. Capezza, A. Lepore, A. Menafoglio, B. Palumbo, S. Vantini, Control charts for monitoring ship operating conditions and CO_2 emissions based on scalar-on-function regression. Appl. Stochas. Mod. Busin. Indus. 36(4) (2020) 1-24. https://doi.org/10.1002/asmb.2507.

[16] M. Shamsuzzaman, A.H.M. Shamsuzzoha, A. Maged, S. Haridy, H. Bashir, A. Karim, Effective monitoring of carbon emissions from industrial sector using statistical process control, Appl. Energy 300 (2021) 117352, https://doi.org/10.1016/j.apenergy.2021.117352.

[17] P. Qiu, Z. Li, On nonparametric statistical process control of univariate processes, Technometrics 53(4) (2011) 390-405, http://dx.doi.org/10.1198/TECH.2011.10005.

[18] P. Qiu, Z. Li, Distribution-free monitoring of univariate processes. Stat. Prob. Letters, 81(12) (2011), 1833-1840. https://doi.org/10.1016/j.spl.2011.07.004.

[19] S. Chakraborti, P. Van der Laan, S.T. Bakir, Non-parametric control charts: An overview and some results, J. Qual. Technol. 33 (2001) 304-315, https://doi.org/10.1080/00224065.2001.11980081.

[20] Y.S. Chang, D.S. Bai, Control charts for positively skewed populations with weighted standard deviations, Qual. Reliab. Eng. Inter. 17 (2001) 397-406, https://doi.org/10.1002/qre.427.

[21] Y. Liu, L. Xue, The optimization design of EWMA charts for monitoring environmental performance, Ann. Oper. Res. 228(1) (2015) 113-124, doi:10.1007/s10479-012-1239-6.

[22] E.G. Schilling, P.R. Nelson, The effect of non-normality on the control limits of $\bar{X}$ charts, J. Qual. Techno. 8(4) (1976) 183-188, https://doi.org/10.1080/00224065.1976.11980743.

[23] S.A. Yourstone, W.J. Zimmer, Non-normality and design of control charts for averages, Decision Sciences 23 (1992) 1099-1113, https://doi.org/10.1111/j.15405915.1992.tb00437.x.

[24] S. Haridy, Z. Wu, P. Castagliola, Univariate and multivariate approaches for evaluating the capability of dynamic-behavior processes (case study), Stat. Method. 8 (2011) 185-203, https://doi.org/10.1016/j.stamet.2010.09.003.

[25] C.-Y. Chou, C.-H. Chen, H.-R. Liu, Economic-statistical design of $\bar{X}$ charts for non-normal data by considering quality loss, J. App. Stat. 27(8) (2000) 939-951, https://doi.org/10.1080/02664760050173274.

[26] Y.-K. Chen, Economic design of $\bar{X}$ control charts for non-normal data using variable sampling policy, Int. J. Prod. Econ. 92(1) (2004) 61-74, doi:10.1016/j.ijpe.2003.09.011.

[27] I.W. Burr, Cumulative frequency functions, The Ann. Math. Stat. 13 (1942) 215-232, doi:10.1214/aoms/1177731607.

[28] Z. Wu, M. Shamsuzzaman, E.S. Pan, Optimization design of control charts based on Taguchi's loss function and random process shifts, Int. J. Prod. Res. 42(2) (2004) 379-390, https://doi.org/10.1081/00207540310001614169.

[29] D.C. Montgomery, Introduction to statistical quality control, John Wiley & Sons, Singapore, 2013.

[30] Z. Wu, M. Shamsuzzaman, Q. Wang, Designs of control charts with optimal manpower deployment, Int. J. Prod. Res. 44(11) (2006) 2119-2132, https://doi.org/10.1080/00207540500478413.

[31] Z. Wu, M. Shamsuzzaman, Q. Wang, The cost minimization and manpower deployment to SPC in a multistage manufacturing system, Int. J. Prod. Econ. 106 (2007) 275-287, doi:10.1016/j.ijpe.2006.06.007.

[32] M. Shamsuzzaman, Z. Wu, M.R.S. Elias, Designs of $\bar{X}$ &S control charts with optimal manpower deployment, Comp. Ind. Eng. 56 (2009) 1589-1596, https://doi.org/10.1016/j.cie.2008.10.007.

[33] M. Shamsuzzaman, Z. Wu, Design of EWMA control chart for minimizing the proportion of defective units, Int. J. Qual. Reli. Manage. 29 (8) (2012) 953-969, https://doi.org/10.1108/02656711211270379.

[34] S. Haridy, A. Maged, A. Baker, M. Shamsuzzaman, H. Bashir, I. Alsyouf, M. Xie, Monitoring scheme for early detection of coronavirus and other respiratory virus outbreaks, Comp. Ind. Eng. 156, (2021) 1-10, https://doi.org/10.1016/j.cie.2021.107235.

[35] I.W. Burr, Parameters for a general system of distributions to match a grid of α_3 and α_4, Commu, Stat. 2 (1) (1973) 1-21, https://doi.org/10.1080/03610927308827052.

[36] P.J. Ross, Taguchi techniques for quality engineering, loss function, orthogonal experiments, parameter and tolerance design, McGraw-Hill, New York, 1989.

[37] M.J.D. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, The Comp. J. 7 (1964) 155-162, https://doi.org/10.1093/comjnl/7.2.155.

[38] S. Haridy, Z. Wu, M.B.C. Khoo, F.J. Yu, A combined synthetic & np Scheme for detecting increases in fraction nonconforming, Comp. Ind. Eng. 62 (2012) 979-988, https://doi.org/10.1016/j.cie.2011.12.024.

[39] Ontario Government-Ministry of the Environment and Conservation and Parks. Environmental penalty annual report, https://files.ontario.ca/moe_mapping/downloads/4Other/2019_EP_AnnualReport_EN.xlsx, 2019 (accessed 15 July 2021).

[40] Integrated Steel Producer. Compounds-wastewater, https://www.kaggle.com/noelpoirier/compoundswastewater, 2016 (accessed 14 August 2021).

[41] C. Paroissin, L. Penalva, A. Pétrau, G. Verdier, New control chart for monitoring and classification of environmental data, Environmetrics 27 (2016) 182-193, https://doi.org/10.1002/env.2382.