Software Impacts 15 (2023) 100455



Contents lists available at ScienceDirect

Software Impacts

journal homepage: www.journals.elsevier.com/software-impacts



Original software publication

An R package for percentile-based control charts: *pbcc*



Aamir Saghir [a,b], Zsolt T. Kosztyán [b,*]

[a] Department of Statistics, Mirpur University of Science and Technology, Mirpur 10250, Pakistan
[b] Department of Quantitative Methods, University of Pannonia, 8200 Veszprém, Hungary

ARTICLE INFO

Keywords:
Statistical quality control
Control charts
Statistical design
Optimization

ABSTRACT

Although the percentile-based control charts are very effective with guaranteed in-control and out-of-control run lengths in statistical process control, they still lack algorithmic and (especially) software support. The proposed percentile-based control charts (*pbcc*) software tool fills this gap. This software is freely available and implements the recent work on individual as well as joint percentile-based control charts.

Code metadata

Current code version	v 0.0.3
Permanent link to code/repository used for this code version	https://github.com/SoftwareImpacts/SIMPAC-2022-289
Code Ocean link	https://codeocean.com/capsule/5771454/tree/v1
Legal code license	MIT license
Code versioning system used	git
Software coding languages, tools, and services used	R
Compilation requirements, operating environments & dependencies	R 3.2 and later releases
If available, link to developer documentation/manual	https://github.com/kzst/pbcc/
Support email for questions	kzst@gtk.uni-pannon.hu

1. Motivation and significance

Statistical process control (SPC) is a quality control method that employs statistical techniques to monitor and control a process [1]. Control charts are used as online monitoring tools to establish and maintain the statistical control of a process [2]. Among others, the Shewhart control chart is the most popular SPC tool. A control chart can be implemented on a process if its design parameters are known. The design parameters are the sample size (n), the sampling interval (h) and the width of the control limits (k). The commonly known Shewhart control charts for monitoring process location/average are $\bar{X}$ and median control charts, while those for monitoring process dispersion/variation are R, S and S^2 charts. These control charts can be used either individually or jointly to monitor an industrial process parameter(s).

In the earliest works, parameters were chosen based on some heuristics that were known to perform acceptably well in practice. Later, more sophisticated methods were designed to choose good parameter values. These approaches are referred to as statistical design (SD) [3] and economic design (ED) [4]. Such methods find parameter values in ways that ensure some SD or ED cost guarantees. Average run length (ARL) metrics are used in the statistical design and performance measures of control charts. The statistical design methods for control charts can be in terms of guaranteed in-control or/and out-of-control performances classified into (i) the ARL approach and (ii) the percentile-based (PL) approach.

In the ARL approach, the control chart parameters are selected for the desired in-control ARL performance, usually 370, [5]. [1,6] criticized the use of ARL as a performance measure for designing control charts and discussed its consequences. On the other hand, the PL approach-based statistical control chart design process ensures that the desired in-control or/and out-of-control performance magnitudes can be achieved with prespecified probabilities. By using the PL approach, (i) the control limits of a control can be adjusted with the guaranteed in-control performance (see [7–11]) and (ii) the optimal parameters (n, h, k) of a control chart can be determined by minimizing

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
E-mail address: kosztyan.zsolt@gtk.uni-pannon.hu (Z.T. Kosztyán).

https://doi.org/10.1016/j.simpa.2022.100455
Received 9 December 2022; Received in revised form 14 December 2022; Accepted 14 December 2022

2665-9638/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Table 1
Comparison of statistical design methods for control charts.

Methods	Performance	
	In-control	Out-of-control
Traditional (ARL)	Not-guaranteed	Not-guaranteed
Percentile-based (PL)	Guaranteed	Guaranteed

the out-of-control performance under the desired in-control performance (see [12]). These approaches are detailed from the guaranteed performances perspective in Table 1.

Several quality control chart packages/algorithms have been extensively studied in recent years [[13–17]], but these packages have two limitations: (i) they statistically design the control charts using the ARL approach and do not guarantee the in-control performance, particularly when parameter(s) of the process are not known and estimated based on past/preliminary information/data, and (ii) they can only design individual control charts for a single process parameter, either mean or dispersion, while quality practitioners recommend using a single control chart for detecting joint changes in the parameters of a process (see [18] for more details). Therefore, there is a dire need for a software/package that can implement the ARL- and PL-based control charts individually as well as jointly as per the requirements of users and quality practitioners. However, to the best of our knowledge, no such software/package is available thus far.

The first goal of the proposed software tool (pbcc) is to design either individually or jointly the traditional (ARL-based) and the percentile-based (PL) control charts with guaranteed in-control performance to solve statistical quality control problems.

In addition to the guaranteed in-control performance, the guaranteed out-of-control must be addressed in the statistical design of a control chart. The optimal statistical design (OSD) of a control chart can be determined by constraining the in-control and out-of-control run lengths with their prespecified probabilities in the design of any control chart.[12] proposed a method using a genetic algorithm to determine an OSD of any control chart with guaranteed in-control and out-of-control performances. The ODS parameters of a percentile-based control chart can be determined by using the following optimization function:

$$\begin{aligned} Min \quad & OCTS_{p_2}; \Theta = \{k, n; h\} \\ & w.r.t: \\ & Pr(ICTS \geq T_1) = 1 - p_1 \\ & Pr(OCTS \leq T_2) = 1 - p_2 \\ & k > 0; ICTS > OCTS \\ & n \in \{2, 3, \ldots, n_{max}\} \\ & h \in \{h_{\min}, 2h_{\min}, 3h_{\min} \ldots, T_2\} \end{aligned} \tag{1}$$

where ICTS and OCTS denote the in-control and out-of-control times to the signals of a process, respectively. The $100(1-p_2)$th percentile of the out-of-control time to a signal can be found by $OCTS_{p_2} = \frac{\ln p_2}{\ln \beta} \times h$ for the defined fitness function in (1). Note that $OCTS_{p_2}$ is a function of the control chart design parameters because β is also a function of the control chart design parameters. Therefore, the optimization problem is a search for the optimal values of the control chart design parameters for which $OCTS_{p_2}$ is the minimum.

The second goal of the proposed software tool (pbcc) is to determine the OSD parameters (n, h, charting constants) of either the individual or the joint ARL and PL control charts with the guaranteed in-control as well out-of-control performances.

This paper proposes a new software application for quality practitioners to statistically design either individual or joint ARL-based and PL-based control charts with guaranteed in-control performance. In addition, the quality practitioner can determine the OSD of either individual or joint PL-based control charts with guaranteed in-control as well as out-of-control performances. Furthermore, the plot and summary of the group statistics of a given data set can be analyzed using this software. The *pbcc* proposed an SPC package that:

- design either individual or joint both the ARL-based and the percentile-based control charts for a given data set such that in-control run lengths are guaranteed with prespecified probability.
- determine the OSD parameters of either individual or joint percentile-based control charts for a given data set such that in-control and out-of-control run lengths are guaranteed with prespecified probabilities.
- plot the percentile-based and ARL-based control charts to visualize the process status.
- summarize the results of the optimal statistical design parameters of the ARL-based and PL control charts.

2. Software description

The software, pbcc, is available from the Comprehensive R Archive Network at https://cran.r-project.org/web/packages/pbcc.

The following subsection explains how to (i) determine the OSD parameters of either individual or joint Shewhart type control charts based on the PL approach (see Section 2.1), (ii) summarize the results (see Section 2.2), (iii) calculate both ARL-based and PL-based control limits of any chart with guaranteed in-control run length performance (see Section 2.3) and (iv) display/plot the optimal statistical design of a control chart (see Section 2.4) for a given data set.

2.1. Determine the OSD parameters

The first function shows how to obtain OSD parameters.
Usage: $odpbc(nmax, T_1, T_2, hv, mat, p_1, p_2, d, delta, type, pop.size, sided)$
Input:

- **nmax**: The maximum possible sample size in each sampling interval (a numeric value).
- $\boldsymbol{T_1}$: The desired in-control time to signal (a numeric value). For the traditional (ARL-based) approach, set $T_1 = 256$.
- $\boldsymbol{T_2}$: The desired out-of-control time to signal (a numeric value).
- $\boldsymbol{h_v}$: The vector of intersample interval up to maximum T2 (a numeric vector of possible sample intervals).
- **mat**: The matrix of minimum and maximum bounds for optimum parameters sample size and sample interval. The minimum values of n and h are (2,0.5), and the maximum values are (nmax, T_2).
- $\boldsymbol{p_1}$. The probability to signal in-control from a specified number (default value is 5%). For the traditional (ARL-based) approach, $p_1 = 50\%$.
- $\boldsymbol{p_2}$: The probability to signal out-of-control from a specified number (default value is 5%). For the traditional (ARL-based) approach, $p_2 = 50\%$.
- **d**: The expected shift size in the process average in terms of standard deviation, i.e., $d = |u1 - uo|/delta$ (default value is 1.0). When the process is in control with respect to the process average, set d = 0.
- **delta**: The expected shift size in the process variance (default value is 1.5). When the process is in control with respect to process variation, set delta = 1.0.
- **type**: A character string specifying the type of Shewhart control chart either individual or joint. The available types are "Xbar", "R", "S", "S2", "Xbar-R", "Xbar-S" and "Xbar-S2".
- **pop.size**: The number of rounds used to solve the optimization problem for the genetic algorithm (default value is 1000).
- **sided**: A character string is used to design either one-sided or two-sided dispersion chart(s) based on the PL approach (the default type is two-sided).

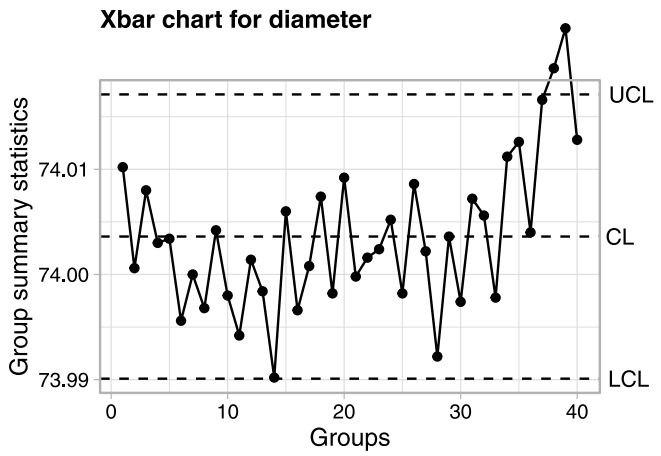


Fig. 1. $\bar{X}$ chart for the diameter data using the ARL approach.

Output:

- ***n***: The optimal sample size to design the percentile-based either individual or joint control chart.
- ***h***: The optimal intersampling interval to design the percentile-based either individual or joint control chart.
- ***k***: The optimal control chart constant/multiplier to design the percentile-based $\bar{X}$ control chart either individually or jointly.
- l_p: The optimal lower percentile point to calculate the lower control limits of the percentile-based dispersion control chart. or/and
- u_p: The optimal upper percentile point to calculate the lower control limits of the percentile-based dispersion control chart.

Example 1 Code:

Q1 = odpbc(10, 100, 3, seq(0.5, 3, by=0.5), matrix(c(2, 10, 1, 6), 2,2, byrow=TRUE), 0.05, 0.05, 3,type= "Xbar")

In this case, the maximum sample size is 10, the in-control time to signal is 100 with 5% probability, the out-of-control time to signal is 3 with 5% prespecified probability, and 3 is an expected shift in the individual mean ($\bar{X}$) chart (see Ref. [12]).

Example 2 Code:

Q3 = odpbc(10, 100, 3, seq(0.5, 3, by=0.5), matrix(c(2, 10, 1, 6), 2,2, byrow=TRUE), 0.05, 0.05, 1, 2, type= "Xbar-S")

2.2. Determine the SD

The second function shows how to obtain the statistical design of a Shewhart-type individual or joint ARL and PL control charts for a given data set with guaranteed in-control run length performance.

Usage: $pbcc(data, T1, p1, type, sided)$

Input:

- **data**: A data frame, a matrix or a vector containing observed data for the variable to chart. Each row of a data frame or a matrix and each value of a vector, refers to a sample or "rationale group".
- T_1: The desired in-control time to signal (a numeric value). For the traditional (ARL-based) approach, set T_1=256.
- p_1. The probability to signal in-control from a specified number (default value is 5%). For the traditional (ARL-based) approach, p_1=50%.
- **type**: A character string specifying the type of Shewhart control chart either individual or joint. The available types are "Xbar", "R", "S", "S2", "Xbar-R", "Xbar-S" and "Xbar-S2".
- **sided**: A character string is used to design either one-sided or two-sided dispersion chart(s) based on the PL approach (default type is two-sided).

Output:

- **data.name**: Name of the input data.
- **type**: Type of control chart to be constructed on the given data.
- **sizes**: The sample size of the data.
- **statistics**: The plotting statistic according to the type of control chart.
- **std.dev**: Standard deviation of group statistics.
- **std.dev**: Control limits of the control chart.

Example 1 Code: *Q11 = pbcc(diameter,256, 0.5, type="Xbar")*

In this case, the diameter characteristic of pistonring data from library (qcc) is used to design the individual ARL-based ($\bar{X}$) chart with 50% percentile of the run length. (see Ref. [13], Fig. 1).

Example 2 Code: *Q13 = pbcc(diameter,100, 0.10, type="Xbar")*

In this case, the diameter characteristic of pistonring data from library (qcc) is used to design the individual PL-based ($\bar{X}$) chart such that no false alarm will be detected earlier than at least 100 samples with 10% probability. (see Ref. [12]).

Example 3 Code: *Q14 = pbcc(diameter,100, 0.10, type="Xbar-S2")*

In this case, the diameter characteristic of pistonring data from library (qcc) is used to design the joint PL-based $\bar{X}$ and S^2 charts such that no false alarm will be detected earlier than at least 100 samples with 10% probability. (see Ref. [12]).

2.3. Summary function

The third function summarizes the output of an object of class *pbcc* or *odpbc*.

Usage: $summary(object)$

Input:

- **object**: an object of class *pbcc* or *odpbc*.

Output: No return value, called for side effects.

Example 1 Code:

summary(Q1)

In this case, the summary of results obtained in example 1 of Section 2.1 are obtained as follows:

Summary of the Optimum Parameters:
Sample Size (n) = 3
Sampling Interval (h) = 2.5
Summary of Optimal Control Charting Constants:
Control Charting Constant (K) = 3.220.

Example 2 Code:

summary (Q11)

In this case, a summary of the results obtained in example 1 of Section 2.2 is obtained.

Chart type = Xbar
Summary of group statistics:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
73.990	73.998	74.002	74.003	74.007	74.023

Group Sample size = 5
Number of groups = 40
Center of groups statistics = 74.003
StdDev of groups statistics = 0.010
Control limits =

LCL	UCL
73.990	74.017

2.4. Plot function

The last function plots/visualizes the output of an object of class *pbcc* for a given data set.

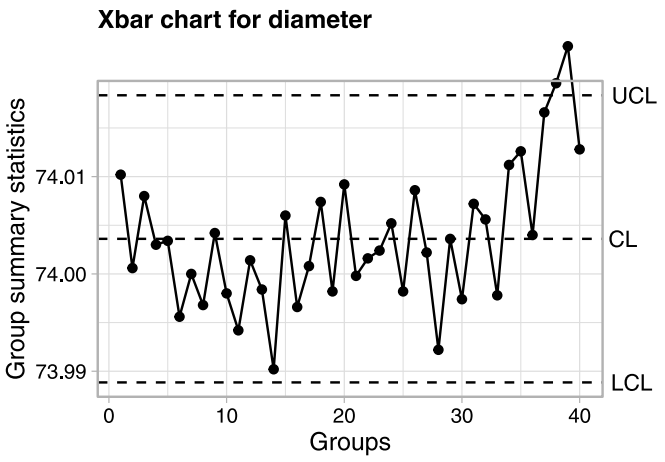


Fig. 2. $\bar{X}$ control chart for the diameter data using the PL approach.

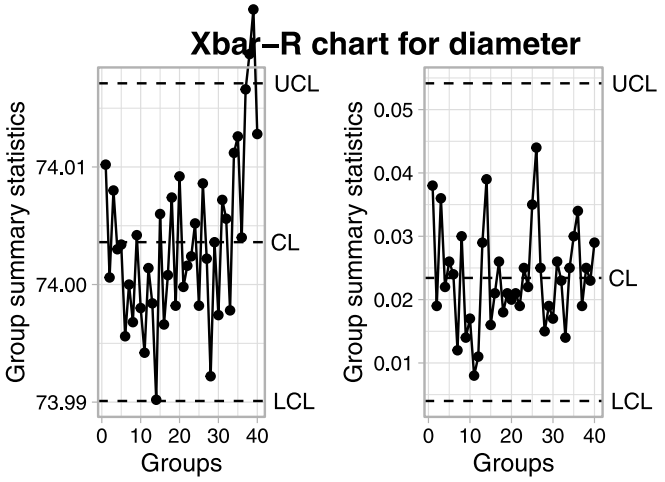


Fig. 3. Joint $\bar{X}$ and R control charts for the diameter data using the ARL approach.

Usage: *plot*(*object*)
Input:

- **object**: an object of class *pbcc*.

Output: No return value, graphical display is produced.
Examples:

Fig. 1 displays the individual ARL-based $\bar{X}$ chart, and Fig. 2 displays the individual PL-based $\bar{X}$ charts for diameter data based on the calculation of examples 1 and 2, Section 2.2.

Fig. 3 displays the joint ARL-based $\bar{X}$ and *Range* charts for diameter data. Fig. 4 shows the PL-based joint $\bar{X}$ and S^2 charts for diameter data based on the calculation of example 3, Section 2.2.

3. Software demonstration with a practical example

The following example shows the benefit of the percentile-based control chart.

We have considered a real data set consisting of the measurements of the inside diameter of the cylinder bores in an engine block, originally reported in [19]. The measurements are made to 1/10,000 of an inch. Samples of size $n = 5$ are taken roughly every half hour, and the first 35 samples are considered for process monitoring. Fig. 5 a–b represents the individual S and $\bar{X}$ charts based on ARL approach via *pbcc* package, which are exactly same as given in [19]. There are two points (samples 6 and 16) exceeding the upper control limit in

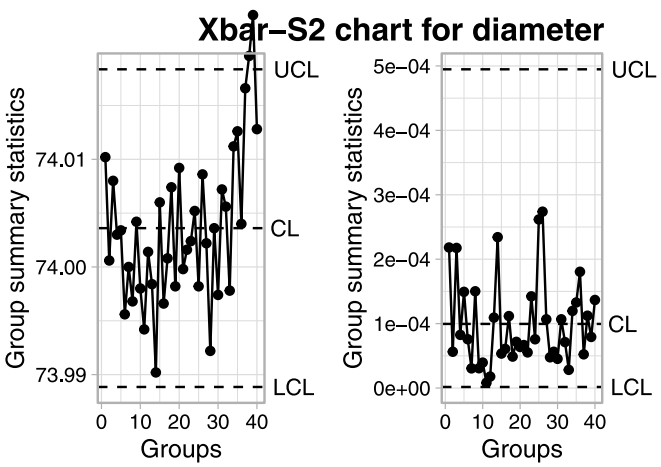


Fig. 4. Joint $\bar{X}$ and S^2 control charts for the diameter data using the PL approach.

the S chart, and one point (sample 11) exceeding the upper control limit in the X-bar chart. After, removing these points, [19] showed that the process is working in-control when points 6, 11 and 16 were removed. Thus, the application of individual S and $\bar{X}$ charts based on ARL approach increases the false alarm rate (which is $3/35 = 0.086$), while the measurement of the inside diameter of the cylinder bores was in-control.

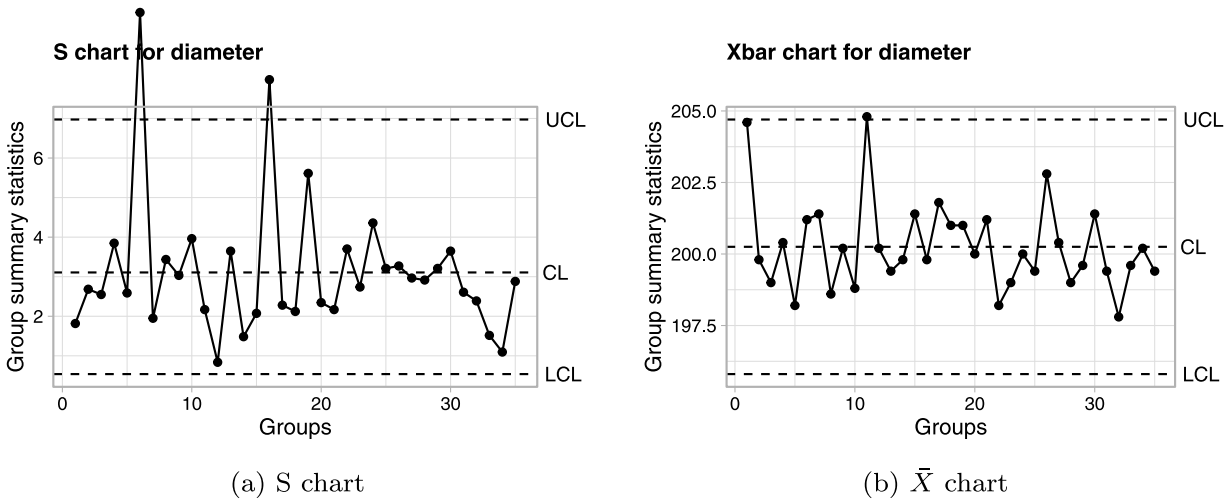
Alternatively, a joint $\bar{X}$ and S charts based on ARL approach may be applied to overcomes false alarm rate and it can be constructed via *pbcc* package. However, this function is not available in any other existing package in the literature. Fig. 6 display the joint $\bar{X}$ and S charts based on ARL approach via *pbcc* package. However, still point 1 and 12 ($2/35 = 0.06$) falls outside the control limits and does not guaranteed the in-control performances of the charts in terms of in-control status. Finally, the percentile-based joint $\bar{X}$ and S control charts are constructed with 95% guaranteed in-control performances in Fig. 7 via *pbcc* package. Note that only sample point 6 falls outside the control limits while sample point 12 lies on the line. So, the false alarm rate for the in-control PL charts for the given data is $1/35 = 0.028$. This implies that the PL approach can be implemented via individual or joint control charts with the guaranteed in-control performance using the *pbcc* package while no other existing package could offer this function.

4. Software impacts

Traditional (ARL-based) and the percentile based (PL) control charts are very effective in the industrial engineering for statistical quality control or anomaly detections. However, they still lack algorithmic and especially software support to implemented these charts in real application in the industry or in the research. Because, the implemented codes of either joint ARL-based or individual or/and joint PL-based control charts in any software are not available/accessible to a quality engineer, user and researcher.

Although *qcc* or *iqcc* packages in R software are available to construct the control charts, these are limited to only individual control charts based on traditional (ARL) approach. *pbcc* is a new and more generalized software application of quality control charts as it offers the construction of traditional individual control charts with additional features of constructing joint control charts based on the ARL approach as well as individual or/and joint control charts based on PL approach. Furthermore, it allows quality engineers and practitioners to statistically design individual or joint control charts with the following advantages:

- it is easily accessible in *R* software and *codeocean* to implement the control charts in real applications.



(a) S chart (b) $\bar{X}$ chart

Fig. 5. Individual S and $\bar{X}$ charts based on ARL approach for diameter data.

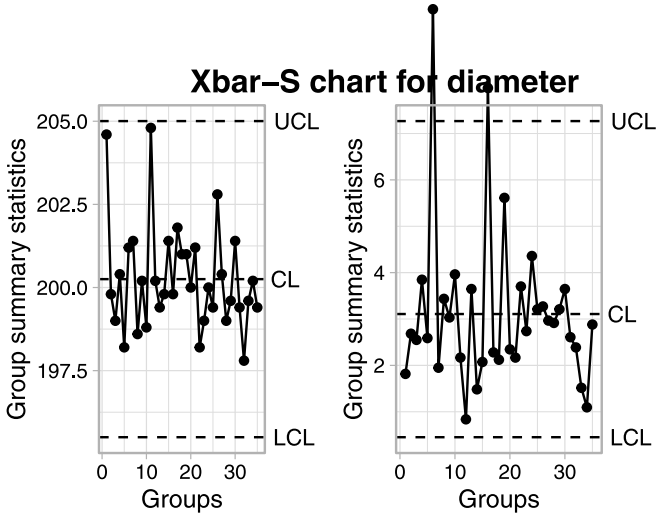


Fig. 6. A joint $\bar{X}$ and S charts based on ARL approach for diameter data.

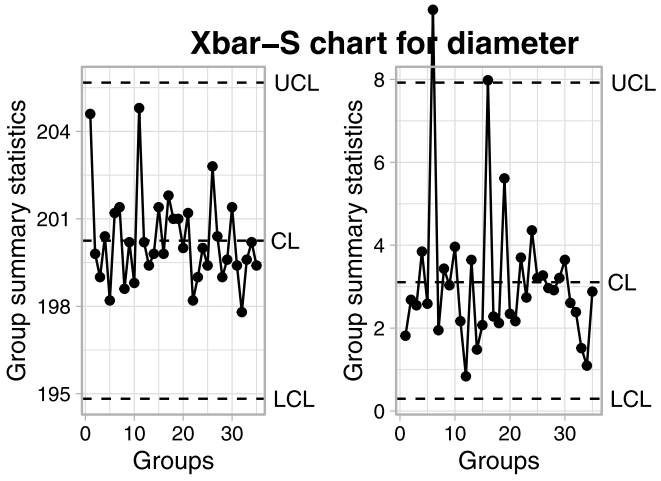


Fig. 7. A joint $\bar{X}$ and S charts based on PL approach for diameter data.

- it can be executed by a new user in the quality control to easily design individual or joint control charts based on ARL (traditional) approach.
- it helps quality engineers to statistically design individual or joint percentile based control charts with guaranteed in-control performances.
- the statistical design of individual as well as joint percentile based control charts with guaranteed in-control and out-of-control performances can be obtained via this software. This will helps in industry to design an optimal structure of the quality control processes.
- it saves time and cost and provides efficient applicability of control charts in SPC.

5. Conclusions

The proposed *pbcc* package/software determines the individual or joint statistical design as well as the optimal statistical design of Shewhart-type control charts for both traditional (ARL-based) and percentile-based approaches. The genetic algorithm is used to solve the optimal statistical design problem of control charts. In addition, joint ARL- and PL-based control charts for a given data set can be computed with prespecified consumer and producer risks.

The *pbcc* tool can help quality engineers and decision makers determine which approaches are best for their requirements. In our future work, the application of software *pbcc* to other control charts (see Refs. [8,20]) will be implemented.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been implemented by the TKP2021-NVA-10 project with support provided by the Ministry of Culture and Innovation of Hungary from the National Research, Development and Innovation Fund, financed under the 2021 Thematic Excellence Programme funding scheme.

References

[1] D.C. Montgomery, Introduction to Statistical Quality Control, John Wiley & Sons, 2020.
[2] E.M. Saniga, Economic statistical control-chart designs with an application to and R charts, Technometrics 31 (3) (1989) 313–320.
[3] W.H. Woodall, The statistical design of quality control charts, J. R. Stat. Soc. D 34 (2) (1985) 155–160.

[4] A.J. Duncan, The economic design of X charts used to maintain current control of a process, J. Amer. Statist. Assoc. 51 (274) (1956) 228–242.
[5] P. Castagliola, G. Celano, G. Chen, The exact run length distribution and design of the S2 chart when the in-control variance is estimated, Int. J. Reliab. Qual. Saf. Eng. 16 (01) (2009) 23–38.
[6] M.B. Khoo, V. Wong, Z. Wu, P. Castagliola, Optimal design of the synthetic chart for the process mean based on median run length, IIE Trans. 44 (9) (2012) 765–779.
[7] A. Gandy, J.T. Kvalø y, spcadjust: An R package for adjusting for estimation error in control charts, R J. 9 (1) (2017) 458.
[8] N.A. Saleh, M.A. Mahmoud, L.A. Jones-Farmer, I. Zwetsloot, W.H. Woodall, Another look at the EWMA control chart with estimated parameters, J. Qual. Technol. 47 (4) (2015) 363–382.
[9] A. Faraz, W.H. Woodall, C. Heuchenne, Guaranteed conditional performance of the S 2 control chart with estimated parameters, Int. J. Prod. Res. 53 (14) (2015) 4405–4413.
[10] A. Faraz, C. Heuchenne, E. Saniga, An exact method for designing Shewhart and S 2 control charts to guarantee in-control performance, Int. J. Prod. Res. 56 (7) (2018) 2570–2584.
[11] A. Saghir, S. Akber Abbasi, A. Faraz, The exact method for designing the Maxwell chart with estimated parameter, Comm. Statist. Simulation Comput. 50 (1) (2021) 270–281.
[12] A. Faraz, E. Saniga, D. Montgomery, Percentile-based control chart design with an application to Shewhart X?? and S2 control charts, Qual. Reliab. Eng. Int. 35 (1) (2019) 116–126.
[13] L. Scrucca, qcc: an R package for quality control charting and statistical process control, Dim (Pistonrings) 1 (200) (2004) 3.
[14] D.R. Recchia, P. Emanuel, E.d.J.G. Barbosa, M.E.P. Barbosa, Package 'iqcc', 2010.
[15] J. Anhoej, T. Roeder, M.J. Anhoej, Package 'qicharts', 2021.
[16] K. Grey, M.K. Grey, Package 'ggqc', 2018.
[17] E. Santos-Fernandez, M.E. Santos-Fernandez, Package 'msqc', 2016.
[18] Z. Jalilibal, A. Amiri, M.B. Khoo, A literature review on joint control schemes in statistical process monitoring, Qual. Reliab. Eng. Int. (2022).
[19] G. Chen, S.W. Cheng, Max chart: combining X-bar chart and S chart, Statist. Sinica (1998) 263–271.
[20] T.P. McWilliams, E.M. Saniga, D.J. Davis, Economic-statistical design of X and R or X and S charts, J. Qual. Technol. 33 (2) (2001) 234–241.