Journal of Quality Technology

A Quarterly Journal of Methods, Applications and Related Topics

ISSN: 0022-4065 (Print) 2575-6230 (Online) Journal homepage: https://www.tandfonline.com/loi/ujqt20

Nonparametric control of the conditional performance in statistical process monitoring

Rob Goedhart, Marit Schoonhoven & Ronald J.M.M. Does

To cite this article: Rob Goedhart, Marit Schoonhoven & Ronald J.M.M. Does (2020) Nonparametric control of the conditional performance in statistical process monitoring, Journal of Quality Technology, 52:4, 355-369, DOI: 10.1080/00224065.2019.1611352

To link to this article: https://doi.org/10.1080/00224065.2019.1611352

Published online: 12 Jun 2019.

JOURNAL OF QUALITY TECHNOLOGY
2020, VOL. 52, NO. 4, 355–369
https://doi.org/10.1080/00224065.2019.1611352

Nonparametric control of the conditional performance in statistical process monitoring

Rob Goedhart, Marit Schoonhoven, and Ronald J.M.M. Does

Department of Operations Management, University of Amsterdam, Amsterdam, The Netherlands

ABSTRACT
Because the in-control distribution and parameters are generally unknown, control limits have to be estimated using a Phase I reference sample. Because different practitioners obtain different samples, their control limit estimates will vary and, consequently, also their control chart performance. We propose the use of nonparametric tolerance intervals in statistical process monitoring to guarantee a minimum control chart performance with a prespecified probability. We evaluate the performance of the proposed limits for various distributions and sample sizes. Note that this nonparametric set-up includes control charts for location and dispersion. Moreover, we compare the performance with other existing methods involving data transformations and a bootstrap procedure. It turns out that the use of nonparametric tolerance intervals performs very well in statistical process monitoring, especially when moderately large sample sizes are available in Phase I.

KEYWORDS
control charts; nonparametric; parameter estimation; SPM; statistical process monitoring; tolerance intervals

1. Introduction

The goal of statistical process monitoring (SPM) is to make a distinction between common and special causes of variation. Generally, it is unknown in advance which variation is common. To this end, various control chart techniques have been developed that estimate the in-control behavior of the data. The first and probably easiest example of such a technique is the Shewhart control chart based on 3-sigma limits. Under the assumptions of normally distributed data and known in-control parameters, this control chart yields a false-alarm rate (FAR) of 0.27 percent or, equivalently, an in-control average run length (ARL) of 370.4.

For a long time, these numbers were commonly accepted as performance indicators of the Shewhart control chart. However, the parameters μ and σ of a normal distribution are generally unknown and need to be estimated using a Phase I reference sample. Quesenberry (1993) recognized this and investigated the effect of sample size on estimated control chart limits. However, an important effect of parameter estimation was neglected there. As different practitioners use different Phase I data, the estimated control limits will vary across practitioners. Consequently, the performance of the control chart in Phase II in terms of FAR and ARL will vary across practitioners as well. This variation in control chart performance becomes smaller as more data are collected, as the estimates then become more accurate (cf., Saleh, Mahmoud, Keefe, et al. 2015). For other research on control chart performance when parameters are estimated, we refer to the literature overviews of Jensen et al. (2006) and Psarakis, Vyniou, and Castagliola (2014). Next to parameter estimation, there are many other important aspects and considerations regarding Phase I data collection and analysis that have to be taken into account. An overview of these issues is given by Jones-Farmer et al. (2014).

In order to compensate for the effect of parameter estimation, several researchers have proposed the use of adjusted control limits. Albers and Kallenberg (2004a, 2004b, 2005) provide two different approaches to adjust the limits, which they name the *bias criterion* and the *exceedance probability criterion*. The bias criterion aims to provide a predefined in-control performance (often in terms of ARL or FAR) in expectation. This approach is also used in Goedhart, Schoonhoven, and Does (2016) and Diko et al. (2017). Although this does account for some aspects of the estimation uncertainty, it still does

CONTACT Rob Goedhart R.Goedhart2@uva.nl Department of Operations Management, University of Amsterdam, Plantage Muidergracht 12, 1018TV, Amsterdam, The Netherlands.
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ujqt.
This article has been republished with minor changes. These changes do not impact the academic content of the article.

not compensate for poor control chart performance caused by the large variation between practitioners.

The exceedance probability criterion provides a more suitable alternative. Rather than providing a prespecified in-control performance in expectation, it aims to guarantee a minimum performance with a specified probability. In terms of the ARL (FAR), this means that the in-control ARL (FAR) is at least (most) equal to a minimum performance threshold with a prespecified probability. This criterion has been rapidly adopted in various research. In the case of the Shewhart control chart for individual observations, constructing control limits based on this criterion is equivalent to the construction of a tolerance interval for the population (cf., Goedhart, Schoonhoven, and Does 2017, 2018). However, the exceedance probability criterion can also be applied in other settings, such as the bootstrap method described in Jones and Steiner (2012), Gandy and Kvaløy (2013), and Saleh, Mahmoud, Jones-Farmer, et al. (2015).

When the data are normally distributed, the previously mentioned methods provide a suitable adjustment to compensate for the effect of parameter estimation. Although tolerance intervals have been derived for various distributions, the distribution of a data set under consideration is in general unknown and has to be estimated along with its parameters. When the distributional assumptions are violated, these adjustments are no longer appropriate and yield unsatisfactory control chart performance.

In this article, we estimate the control limits based on nonparametric tolerance intervals and determine the corresponding exceedance probabilities. We compare the results with several options that make use of adjustments developed for normally distributed data and transformations to normality, as well as more general parametric methods involving bootstrap. Although nonparametric methods typically require somewhat larger sample sizes, they prevent performance issues caused by a violation of the underlying distributional assumptions.

This article is organized as follows. In Section 2, we elaborate further on the effect of parameter estimation. Next, in Section 3, we illustrate the proposed nonparametric control chart design. In Section 4, we discuss some alternative methods and make a comparison. In Section 5, we demonstrate the application of the proposed procedure with a practical example. Finally, in Section 6, we provide some concluding remarks.

2. Parameter estimation and tolerance intervals

The upper control limit (UCL) and lower control limit (LCL) for the original Shewhart control chart for individual observations are based on the process parameters μ and σ and are of the form

$$\begin{aligned} UCL &= \mu + K\sigma \\ LCL &= \mu - K\sigma \end{aligned} \qquad [1]$$

where $K = \Phi^{-1}(1-\alpha/2)$. In the original setting $K=3$, which yields a FAR of $\alpha = 0.0027$ or, equivalently, an in-control ARL of $ARL_0 = 370.4$ for normally distributed data. In practice, μ and σ are generally unknown and have to be estimated by some estimates $\hat{\mu}$ and $\hat{\sigma}$ respectively. As each Phase I sample yields different estimates, the estimated control limits and their corresponding performance actually become random variables (see also Saleh, Mahmoud, Keefe, et al. 2015). We denote estimated control limits in general by $\hat{UCL}$ and $\hat{LCL}$.

In order to compensate for the random character of the control chart performance, Albers and Kallenberg (2004b) proposed the exceedance probability criterion. This states that there should be only a small probability p for some performance measure Y to be worse than a prespecified minimum performance threshold Y_0. Examples of such performance measures are the conditional FAR and ARL, conditioned on the Phase I parameter estimates and denoted by CFAR and CARL, respectively. If we consider $Y = CFAR$ and $Y_0 = \alpha_{tol}$, then, for some random variable X, this criterion implies

$$P(CFAR > \alpha_{tol}) = P\left(1 - P\left(\hat{LCL} \leq X \leq \hat{UCL}\right) > \alpha_{tol}\right) = p. \qquad [2]$$

For a random variable X, this criterion is equivalent to the construction of a tolerance interval with a coverage of at least $1-\alpha_{tol}$ with probability $1-p$. Note also that CARL equals the reciprocal of CFAR, so that using $Y = CARL$ and $Y_0 = CARL_0$ is equivalent to using $Y = CFAR$ and $Y_0 = \alpha_{tol} = 1/CARL_0$. For more details on this, we refer to Goedhart, Schoonhoven, and Does (2017, 218). In these articles, as well as in the tolerance interval literature (e.g. Krishnamoorthy and Mathew 2009), the criterion is satisfied by adjusting the control limit constant K, taking into account the sample size and the estimators being used. These adjustments are shown to work well for normally distributed data in the corresponding literature.

However, it is obvious from Eq. [1] that this form of control limits will not work as well for other distributions, as it is designed under normal theory. Especially for skewed data, the symmetrical character of these control limits will cause issues with control chart performance. A first way to overcome this is by introducing probability limits, as is done, for example, for S-charts when the data are normally distributed (e.g., Montgomery 2013). These limits are based on the distribution of the characteristic of interest. It is also possible to derive limits based on the exceedance probability criterion for these statistics, as is done, for example, in Goedhart, da Silva, et al. (2017). Moreover, Krishnamoorthy and Mathew (2009) derive parametric tolerance intervals for a wide range of distributions, which can be applied in a similar fashion.

Although these parametric tolerance intervals provide a solution for various data distributions, in practice, the problem is that the distribution of the data is generally unknown and has to be estimated along with its parameters. As mentioned in Albers, Kallenberg, and Nurdiati (2004), the total estimation error can be split up in two different distinct errors, the *model error* (ME) and the *stochastic error* (SE). The first is caused by incorrect assumptions on the distributional form, while the latter is the error resulting from parameter estimation. With nonparametric methods, the ME vanishes and the variation is caused by the SE only. While the first is strongly dependent on the distribution under consideration, the latter can be reduced by collecting larger samples.

Most nonparametric methods revolve around the use of order statistics, such as the nonparametric tolerance intervals described in Krishnamoorthy and Mathew (2009). Other proposed methods, such as the bootstrap procedure of Gandy and Kvaløy (2013), are often only partially nonparametric. Although bootstrapping the Phase I sample may be performed in a nonparametric way, a key aspect of this method is to determine the required limits for each bootstrap replication. This, in turn, requires some distributional assumptions in order to be accurate for small samples. For this reason, Gandy and Kvaløy (2013) advised using a parametric instead of nonparametric bootstrap procedure for Shewhart-type control charts. We will elaborate on this further in Section 4.2.

A major advantage of nonparametric control charts is that they can be applied to individual observations as well as subgroup statistics. For example, when treating subgroup standard deviations as individual observations, one can apply a nonparametric control chart to monitor the standard deviation. The same can be done for various other statistics as well (e.g., the average, range, or other robust estimators for location or dispersion), regardless of the distribution under consideration. Because the nonparametric setup does not give severe restrictions for the charting statistic X in [2], this makes the approach very general.

3. Nonparametric control limits

There is a wide range of literature available on nonparametric control charts; see, for example, Qiu (2018). However, most of these control charts do not incorporate the exceedance probability criterion. For this reason, we consider nonparametric tolerance intervals in this section. For more information on nonparametric statistical process control in general, we refer to Chakraborti, Van der Laan, and Bakir (2001), Chakraborti, Qiu, and Mukherjee (2015), and Qiu (2018).

The nonparametric tolerance intervals as described in Krishnamoorthy and Mathew (2009) are constructed using order statistics. In particular, each limit is determined by a single order statistic. Although this method entirely rules out the ME caused by model misspecification, it is obvious that this increases the SE compared with parametric methods. Other disadvantages of using only a single order statistic for each limit is that this approach generally leads to rather conservative estimates, or to cases where the desired coverage probability can't be guaranteed because of small sample sizes. These issues are addressed in Young and Mathew (2014), who suggested constructing tolerance limits based on interpolated and extrapolated order statistics, depending on the sample size. In this section, we explore their procedure for determining the required limits and evaluate its performance.

3.1. Estimation of control limits

The first step is to determine whether one should interpolate or extrapolate. This is done by determining the minimum Phase I sample size requirement for the construction of a two-sided nonparametric tolerance interval based on unweighted order statistics, as in Krishnamoorthy and Mathew (2009). Note that, in this article, we consider a Phase I sample consisting of m individual observations. However, as mentioned earlier, one could also apply the same procedure to subgroup statistics from a sample of m subgroups by treating them as m individual variables. If the sample size available is large enough, one should interpolate to make the estimated interval less conservative. If the sample size is not large enough, one has to extrapolate to reach a desired exceedance probability. The sample size m is sufficient when the following equation holds:

$$(m-1)(1-\alpha_{tol})^{m}-m(1-\alpha_{tol})^{m-1}+1 \geq 1-p \qquad [3]$$

where α_{tol} is the performance threshold for CFAR, and $1-p$ is the prespecified probability of achieving at least this performance. For a derivation of Eq. [3], we refer to Section 8.6.1 of Krishnamoorthy and Mathew (2009). Solutions to this equation can, for example, be determined relatively easily with the function distree.-est() included in the R-package tolerance (see also Young and Mathew 2014). In Table 1, we provide an overview of $m_2(\alpha_{tol}, p)$, the minimum sample size requirement for a two-sided tolerance interval based on unweighted order statistics, for several values of α_{tol} and p. We now make a distinction between the limits when interpolating ($m \geq m_2(\alpha_{tol}, p)$) and when extrapolating ($m < m_2(\alpha_{tol}, p)$).

3.1.1. Interpolated control limits

When $m \geq m_2(\alpha_{tol}, p)$, consider a two-sided tolerance interval $[X_{(r)}, X_{(s)}]$ as a starting point, where $X_{(j)}$ denotes the j-th order statistic from a Phase I sample X_i, with $i = 1, .., m$, and with r and s yet to be determined. This interval yields a coverage probability of

$$P(B \leq k-1) \geq 1-p \qquad [4]$$

where $B \sim Bin(m, 1-\alpha_{tol})$ and where $k = s-r$ is the smallest integer for which Eq. [4] holds. In terms of Eq. [2], this is equivalent to $P(CFAR > \alpha_{tol}) \leq p$ for $\hat{LCL} = X_{(r)}$ and $\hat{UCL} = X_{(s)}$. Next, the two intervals $[X_{(r+1)}, X_{(s)}]$ and $[X_{(r)}, X_{(s-1)}]$ are considered, which yield a coverage probability of $P(B \leq k-2) < 1-p$. Linear interpolation is then used at both sides of the original tolerance interval to obtain

$$\begin{aligned} \lambda_1 &= \frac{(1-p)-P(B \leq k-2)}{P(B \leq k-1) - P(B \leq k-2)} = \frac{(1-p)-P(B \leq k-2)}{P(B = k-1)} \\ X_{(r^*)} &= \lambda_1 X_{(r)} + (1-\lambda_1) X_{(r+1)} \\ X_{(s^*)} &= \lambda_1 X_{(s)} + (1-\lambda_1) X_{(s-1)} \end{aligned} \qquad [5]$$

Note that $\lambda_1 \in [0, 1]$, such that $X_{(r^*)} \in [X_{(r)}, X_{(r+1)}]$ and $X_{(s^*)} \in [X_{(s-1)}, X_{(s)}]$. The proposed two-sided nonparametric tolerance interval is then given by the shortest interval of $[X_{(r^*)}, X_{(s)}]$ and $[X_{(r)}, X_{(s^*)}]$. $\hat{LCL}$ and $\hat{UCL}$ are then set equal to the lower and upper limits of this interval, respectively.

When choosing s and r, it is common to set $s = m-r+1$ so that the tolerance interval $[X_{(r)}, X_{(s)}]$ corresponds to the truncated sample range (see, e.g., Wilks 1941; Krishnamoorthy and Mathew 2009; and Young and Mathew 2014). This interval starts at the r-th smallest and ends at r-th largest observations, which means that $r - 1$ observations are being trimmed on both sides of the ordered data. The value r should thus be taken equal to the maximum integer value such that Eq. [4] holds for $k = s-r$ and $s = m-r+1$.

However, for some parameter combinations of α_{tol}, p, and m, this does not lead to the shortest possible tolerance interval with at least the required coverage probability. In particular, recall that the coverage probability depends only on $k = s-r$ and not on the absolute values of s and r. Furthermore, note that the total number of observations that are trimmed from both sides equals $m-k-1$. When the restriction $s = m-r+1$ is dropped, and solving for k to be the minimum integer value that satisfies Eq. [4], this number could be odd. This means that it is possible to obtain a tolerance interval with a smaller coverage probability than the truncated sample range, but still greater than the nominal coverage probability. The two most logical choices in this case are $[X_{(r+1)}, X_{(s)}]$ and $[X_{(r)}, X_{(s-1)}]$. For each of these two intervals, the interpolation procedure is performed, resulting in four possible interpolated intervals. The proposed two-sided nonparametric tolerance interval is then given by the shortest of these four intervals. The control limits are then set equal to the lower and upper limits of this interval, respectively. In Section 5, we use a real data set to illustrate the described procedure.

Note that, as discussed, when the sample size increases, there will be a point where an additional observation can be trimmed from the ordered data in order to obtain the required intervals for interpolation. This causes a jump in the coverage probability of these intervals, that is, $P(B \leq k-1)$ and $P(B \leq k-2)$, which is inherent to the discrete nature of the binomial distribution. This has the consequence that the coverage probability does not always converge monotonically to p when m increases. Instead, the path of convergence looks more like a sawtooth. This is also observed in the results in Section 3.2.

Table 1. Minimum required sample size $m_2(\alpha_{tol}, p)$ for two-sided tolerance intervals based on unweighted order statistics.

	p		
α_{tol}	0.2	0.1	0.05
0.05	59	77	93
0.01	299	388	473
0.005	598	777	947
0.0027	1109	1440	1756

3.1.2. Extrapolated control limits

When $m < m_2(\alpha_{tol}, p)$, the coverage probability of $[X_{(1)}, X_{(m)}]$ is not sufficient and one has to extrapolate. Note that the coverage probability of $[X_{(1)}, X_{(m)}]$ equals $P(B \leq m-2)$ and that the coverage probability of the intervals $[X_{(1)}, X_{(m-1)}]$ and $[X_{(2)}, X_{(m)}]$ equals $P(B \leq m-3)$. Then, with linear extrapolation, the following results are obtained

$$\begin{aligned} \lambda_2 &= -\frac{(1-p)-P(B \le m-2)}{P(B \le m-2) - P(B \le m-3)} \\ &= -\frac{(1-p)-P(B \le m-2)}{P(B = m-2)} \\ X_{(1^*)} &= \lambda_2 X_{(2)} + (1-\lambda_2) X_{(1)} \\ X_{(m^*)} &= \lambda_2 X_{(m-1)} + (1-\lambda_2) X_{(m)} \end{aligned} \quad [6]$$

Note that, in this case, $\lambda_2<0$, such that $X_{(1^*)}<X_{(1)}$ and $X_{(m^*)}>X_{(m)}$. The two-sided nonparametric tolerance interval proposed by Young and Mathew (2014) is then equal to $[X_{(1^*)}, X_{(m^*)}]$. For a nonparametric control chart, one can thus set the $\hat{LCL}$ and $\hat{UCL}$ equal to the lower and upper limits of this interval, respectively.

3.2. Performance of the proposed limits

In this section, we evaluate the performance of these methods for various distributions. Although Young and Mathew (2014) already provide an extensive evaluation, their results are focused on theory on tolerance intervals. In that setting, values for α_{tol} of interest are generally in the order of 0.05 to 0.5 while, in statistical process monitoring, the interest lies much further in the tails. For that reason, we focus mainly on the results for $\alpha_{tol} = 0.0027$. However, we have also included the values 0.05, 0.01, and 0.005 because it is sometimes advisable to be more lenient with the demands placed on the control chart.

The evaluation of the proposed control limits is done for various distributions. In particular, we consider the standard normal distribution, a lognormal distribution with $\mu = 0$ and $\sigma = 1$, a chi-square distribution with 4 degrees of freedom (χ^2_4), and a t-distribution with 4 degrees of freedom (t_4). The standard normal distribution illustrates the performance when normal theory is applicable. The lognormal distribution provides a common skewed alternative. The χ^2_4 provides another skewed alternative and gives a good indication of the performance of these control limits for various estimators of dispersion. Note that, if the data are normally distributed, many estimators of dispersion are (approximately) distributed according to a scaled chi-square or chi distribution, as also discussed in Goedhart, da Silva, et al. (2017). In that case, for the chosen degrees of freedom ($df=4$), this distribution corresponds to that of scaled standard deviations of subgroups of size 5. The t_4 provides a symmetrical alternative to the normal distribution, but with heavier tails. Also, similarly to the χ^2_4, various test statistics for location are based on the t-distribution, such that the chosen $df=4$ could represent various statistics in subgroups of size 5.

In order to assess the performance of the proposed limits, we have performed a simulation study. In particular, for various combinations of m and α_{tol} together with $p=0.1$ and $p=0.2$, we have applied the following procedure

1. A data set consisting of m observations is drawn from the specified distribution.
2. The control limits $\hat{LCL}$ and $\hat{UCL}$ are estimated according to Eqs. [5] and [6], depending on the value of m, as described in the previous section. This method is available in the R-package tolerance with the function nptol.int().
3. The probability $P(\hat{LCL} \le X \le \hat{UCL})$, where X is a future (Phase II) in-control observation, is determined using the original distribution of the data.
4. Steps 1 to 3 are repeated for 10,000 different Phase I samples and the proportion for which $1-P(\hat{LCL} \le X \le \hat{UCL})>\alpha_{tol}$ is calculated. This proportion should be approximately equal to p according to the criterion described in Eq. [2].

The results of the described procedure are given in Table 2 for $p=0.1$ and Table 3 for $p=0.2$. Recall that the values displayed should be approximately equal to p. The values indicated in italic font type indicate that the result is based on extrapolated control limits, as not enough samples were available in that case for

Table 2. Exceedance probabilities for proposed control limits for $p=0.1$.

		α_{tol}			
	m	0.05	0.01	0.005	0.0027
Normal	100	0.0932	*0.2374*	*0.1672*	*0.0927*
	250	0.1124	*0.1612*	*0.2459*	*0.2146*
	500	0.0733	0.0906	*0.1646*	*0.2477*
	1000	0.0820	0.0972	0.0872	*0.1526*
	1500	0.1127	0.1089	0.0948	0.0988
	2500	0.0915	0.0802	0.1088	0.1034
χ^2_4	100	0.0836	*0.2236*	*0.1579*	*0.0814*
	250	0.1028	*0.1630*	*0.2261*	*0.1862*
	500	0.0990	0.0808	*0.1609*	*0.2299*
	1000	0.0903	0.0871	0.0799	*0.1494*
	1500	0.1037	0.0973	0.0879	0.0884
	2500	0.0966	0.0939	0.0954	0.0987
Lognormal	100	0.0728	*0.2355*	*0.1809*	*0.0963*
	250	0.0942	*0.1705*	*0.2428*	*0.2200*
	500	0.0958	0.0738	*0.1646*	*0.2401*
	1000	0.1000	0.0900	0.0725	*0.1613*
	1500	0.0983	0.0900	0.0850	0.0912
	2500	0.1032	0.1002	0.0872	0.0943
t_4	100	0.0904	*0.2756*	*0.2165*	*0.1243*
	250	0.1013	*0.1794*	*0.2867*	*0.2617*
	500	0.0769	0.0803	*0.1775*	*0.2744*
	1000	0.0745	0.1022	0.0798	*0.1705*
	1500	0.1094	0.1082	0.0914	0.0909
	2500	0.0845	0.0781	0.1013	0.0999

Table 3: Exceedance probabilities for proposed control limits for $p = 0.2$.

	m	α_{tol} 0.05	0.01	0.005	0.0027
Normal	100	0.1124	*0.2664*	*0.1855*	*0.1056*
	250	0.1954	*0.2192*	*0.2714*	*0.2368*
	500	0.1612	0.1239	*0.2174*	*0.2665*
	1000	0.2146	0.1325	0.1218	*0.2129*
	1500	0.2066	0.2039	0.1228	0.2015
	2500	0.2028	0.1732	0.1397	0.1435
χ^2_4	100	0.1654	*0.2484*	*0.1728*	*0.0927*
	250	0.1973	*0.2127*	*0.2605*	*0.2179*
	500	0.1887	0.1703	*0.2132*	*0.2613*
	1000	0.1958	0.1856	0.1721	*0.2014*
	1500	0.1931	0.1920	0.1853	0.1677
	2500	0.1919	0.1890	0.1968	0.1996
Lognormal	100	0.1835	*0.2747*	*0.1923*	*0.1107*
	250	0.1979	*0.2235*	*0.2679*	*0.2379*
	500	0.1964	0.1765	*0.2165*	*0.2725*
	1000	0.2016	0.1944	0.1769	*0.2089*
	1500	0.2016	0.1928	0.1962	0.1582
	2500	0.1967	0.1989	0.1967	0.2015
t_4	100	0.1338	*0.3115*	*0.2441*	*0.1397*
	250	0.1944	*0.2350*	*0.3113*	*0.2883*
	500	0.1657	0.1331	*0.2247*	*0.3114*
	1000	0.2138	0.1432	0.1397	*0.2075*
	1500	0.2022	0.1998	0.1412	0.1775
	2500	0.2037	0.1748	0.1485	0.1538

interpolation. As can be seen in the tables, the obtained values are, in general, close to their desired value. However, one can still observe a deterioration of the performance when α_{tol} becomes smaller, given the sample sizes considered. While we observe only small deviations from p when $\alpha_{tol} = 0.05$, these deviations are substantially larger for $\alpha_{tol} = 0.0027$. Also, for $\alpha_{tol} = 0.0027$, the performance is not entirely consistent for the smallest sample sizes under consideration, as we observe quite some different values for the sample sizes $m = 100$, 250, and 500. As we increase the sample size to larger values, such as $m = 1{,}500$, the performance moves closer to the desired level. This is, of course, not entirely unexpected because of the sample sizes and α_{tol} values under consideration. For example, estimating a tolerance interval with a coverage of $1-0.0027 = 0.9973$ based on order statistics from a sample of $m = 250$ observations is not as accurate as we would like. Note also that $m_2(\alpha_{tol}, p)$, the minimum sample size requirement when using Krishnamoorthy and Mathew (2009), equals 1,440 for $\alpha_{tol} = 0.0027$ and $p = 0.1$, as can be seen in Table 1. When the interpolation Eq. [5] is applied, the results for $p = 0.1$ are very accurate.

Moreover, contrary to other parametric methods, the distribution under consideration only has a small impact on the performance. The differences in results between the distributions are caused by the fact that the linear interpolation and extrapolation are approximations with an accuracy that depends on the underlying distribution. However, when interpolated control limits are used, the coverage probability will be bounded between $P(B \le k-1)$ and $P(B \le k-2)$ (i.e., the coverage probability of the intervals used for interpolation) regardless of the distribution.

As observed, the proposed control limits provide a better performance in terms of exceedance probability when sample sizes are larger. Moreover, the variation between the resulting CFAR values becomes smaller as sample sizes are increased, due to a smaller SE. This is illustrated as well in Figure 1. For $\alpha_{tol} = 0.0027$, $p = 0.1$, and various sample sizes under consideration, this figure shows the variation between CFAR values obtained from the different simulated Phase I samples by means of boxplots. We have used the results from the normally distributed data, but the results for other distributions are similar. The horizontal dotted line indicates the location of α_{tol}, which should be close to the $(1-p)$-quantile of the boxplots (the actual value can be obtained from Table 2). We have left out the case $m = 100$ due to the large variation present there, which would make the rest of the boxplots more difficult to compare.

As can be observed, increasing the sample sizes results in less variation and less extreme CFAR values. Although perhaps more difficult to detect directly, these extreme CFAR values occur close to zero as well, which can be seen from the location of the bottom side of the box and/or whisker. For $m = 250$, it is obvious from the box that there are many low CFAR values. For $m = 100$, these values are even more frequent. This is mainly caused by the fact that the extrapolations will go further beyond the furthest order statistics obtained when sample sizes are small. For example, for $m = 100$ and $\alpha_{tol} = 0.0027$, the actual coverage probability of $[X_{(1)}, X_{(100)}]$ is equal to 0.03. When one desires $p = 0.1$, the required coverage probability would be 0.9, which thus requires a very large extrapolation. Because the extrapolation is linear, this may result in rather extreme estimates of the control limits. This is undesirable, as it may lead to a substantial deterioration in control chart performance in out-of-control situations. On the other hand, when sample sizes are sufficient so that the control limits can be derived by interpolation, the estimated limits will never go beyond the smallest/largest order statistics obtained from the sample, which prevents such extreme estimates. Therefore, when the Phase I sample size is small (e.g., $m = 100$), we suggest using more lenient parameter values (e.g., larger α_{tol} and/or p) rather than extrapolation. Moreover, small samples are generally a consequence of a difficult or expensive

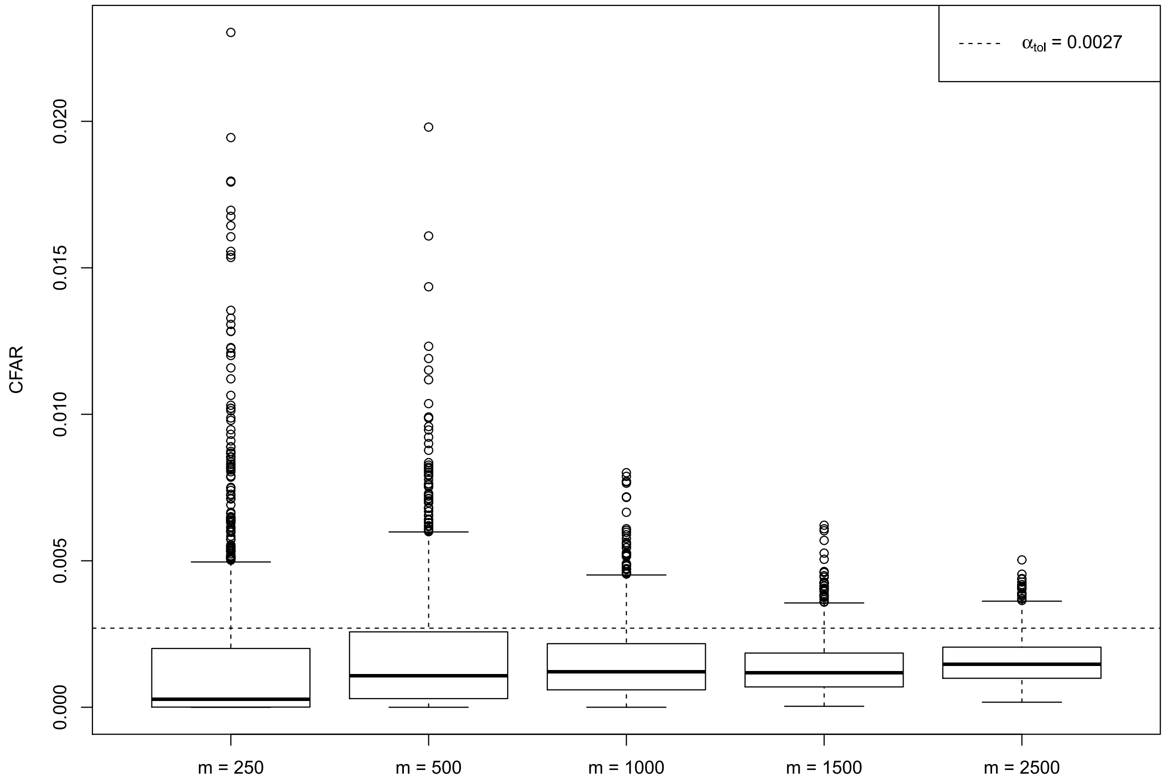


Figure 1. Boxplots of CFAR for normally distributed data with $\alpha_{tol} = 0.0027$ and $p = 0.1$, when applying nonparametric control limits.

sampling procedure or a low-frequency process. In such cases, there will most likely also be more time between consecutive Phase II observations, so that a higher false-alarm rate is acceptable. If these issues are not present, it is advisable to collect a larger Phase I sample first.

In general, the conditional performance approach considered in this article leads to slightly wider control limits. This effect is also investigated by, for example, Gandy and Kvaløy (2013) and Goedhart, Schoonhoven, and Does (2017). While this is, of course, beneficial for the in-control performance, it means that out-of-control signals are obtained less frequently. This tradeoff is inherent to control charts. However, the parameters (e.g., α_{tol} and p) can easily be adjusted to balance this tradeoff.

4. Alternative methods

In this section, we discuss alternative methods that aim to satisfy Eq. [2]. First, we discuss methods based on normal theory, after which we elaborate more on the bootstrap procedure proposed by Jones and Steiner (2012) and Gandy and Kvaløy (2013). For parametric methods, SE will be smaller compared with nonparametric methods due to better use of available information. However, ME plays a bigger role there because deviations from the model assumptions might be larger than the actual decrease in SE. Moreover, the errors can differ in size for different distributions. At the end of this section, we include a performance comparison using a numerical study.

4.1. Normal tolerance interval methods and extensions

Various publications have been devoted to tolerance intervals for normal populations; see also Krishnamoorthy and Mathew (2009). Recently, Goedhart, Schoonhoven, and Does (2018) proposed new control limits based on this theory. However, in practice, the problem is that data often aren't close to being normally distributed. To this end, we evaluate some techniques that aim to make normal theory applicable. In particular, we discuss the application of the Central Limit Theorem (CLT) on subgroup averages to create an approximately normally distributed dataset, as well as transformations to normality as proposed for SPM by Chou, Polansky, and Mason (1998). The intention of such techniques is to retain the relatively small SE compared with completely nonparametric models, but reducing the accompanying ME caused by deviations from normality.

As the use of subgroups is a common practice in SPM during data collection, one could argue using the subgroup averages as individual observations and apply the CLT. Often, subgroups of size 5 are recommended in SPM, while a sample size of 30 is generally deemed enough for the CLT to apply. Especially with the recent developments in increasing data availability, one could argue that collecting such amounts of data should not be a problem. However, although the CLT works well for the major (middle) part of the data, it is of less use when the far tails are of interest. As recently shown by Huberts et al. (2018), in some cases, it might even require subgroups of more than 1,000 observations before the control chart performance is satisfactory for application in SPM. Therefore, we will not elaborate further on this option in this article.

Another option to make the use of techniques developed for normally distributed data viable is to transform the data such that a normal distribution is appropriate. One of the most applied transformations in statistics is the Box–Cox transformation, originating from Box and Cox (1964). This method is used to transform data into a symmetrical distribution. Although this generally lowers the deviation from normality, the resulting transformed data are often far from normally distributed. Similar conclusions can be found in Sakia (1992), among others.

A method that specifically intends to transform data to normality in the field of SPM was proposed by Chou, Polansky, and Mason (1998). They proposed transformations based on the Johnson system of distributions, and provided a step-by-step procedure revolving around the Shapiro–Wilk test for normality. Transformations are only applied when normality of the data is rejected in the first step. One can use this procedure in combination with the suggested control limits in Goedhart, Schoonhoven, and Does (2018). A more detailed description of the procedure can be found in Appendix A.

4.2. Bootstrap procedure

In the previous Subsection 4.1, we have described several methods that aim to make the normal tolerance interval theory applicable. Another option is to derive parametric limits for various other specific distributions, as is done in Krishnamoorthy and Mathew (2009). In their book, they considered the lognormal, Gamma, two-parameter exponential, and the Weibull distribution. However, in practice, the distribution of the data is unknown and still has to be estimated. To this end, we consider methods that aim to provide a more general parametric model in order to find the required limits. Such methods have to be flexible with regard to the location, shape, skewness, and kurtosis of the distribution under consideration. A well-known system of distributions that allows for such flexibility is the Pearson system, which consists of different types of distributions. An alternative that has one single distribution type based on four parameters to be estimated is given in Low (2013).

The next step is to derive tolerance limits for these general models. To do this, they can be combined with the Gandy and Kvaløy (2013) bootstrap procedure. The idea of the bootstrap procedure is that first, a distribution is fitted to the data. This fit is then assumed to be the true distribution, after which bootstrap samples are drawn from the assumed distribution. These bootstrap samples provide an estimate of the estimation uncertainty accompanying the Phase I data. Then, in each bootstrap replicate, one has to determine the required limits in order to provide a desired result in the assumed distribution of the Phase I sample. This, however, requires some distributional assumptions to be accurate, as it requires estimation of extreme (e.g., $\alpha_{tol}/2$ and $1-\alpha_{tol}/2$) quantiles of the assumed Phase I distribution. Thus, even when the bootstrap samples are drawn in a nonparametric way (using the empirical cdf), determining the required limits for each bootstrap sample is not accurate when done nonparametrically. To still remain general with the distributional assumptions, one can combine the bootstrap method with the Pearson system of distributions. A detailed description of the applied procedure can be found in Appendix B. Note that, instead of the Pearson system of distributions, one could also consider other general parametric methods such as that considered by Low (2013). However, our results in that case were similar.

4.3. Performance comparison

In order to assess the performance of the methods proposed in this section, we have evaluated the prescribed procedures for various settings. For the in-control situation, we illustrate the results for the limits from Goedhart, Schoonhoven, and Does (2018), both with (denoted Chou) and without (denoted Normal) the transformation proposed in Chou, Polansky, and Mason (1998), as well as the Gandy and Kvaløy (2013) bootstrap procedure (denoted GK) described in Section 4.2. Also, we consider the same distributions as in Section 3.2. For the out-of-control situation, we make a performance comparison for when normal theory is applicable.

4.3.1. In-control performance

For various sample sizes m together with $p = 0.1$ and $\alpha_{tol} = 0.0027$, we have assessed the exceedance probability criterion as described in Eq. [2]. Note that, in this comparison, we only consider $\alpha_{tol} = 0.0027$ because we are interested in the application of these methods in SPM. We have applied the following procedure for each parameter combination:

1. A dataset consisting of m observations is drawn from the specified distribution.
2. The methods under consideration are applied to determine the estimated control limits $\hat{LCL}$ and $\hat{UCL}$ for each method.
3. The probability $P(\hat{LCL} \leq X \leq \hat{UCL})$, where X is a future (Phase II) in-control observation, is determined using the original distribution of the data.
4. Steps 1 to 3 are repeated for 1,000 different Phase I samples, and the proportion for which $1-P(\hat{LCL} \leq X \leq \hat{UCL})>\alpha_{tol}$ is calculated. This proportion should be approximately equal to p according to the criterion described in Eq. [2].

The results of the described procedure can be found in Table 4. We have also included the results from Table 2 (denoted by YM) for comparison purposes. Recall that the values displayed in the tables should be approximately equal to p. As can be observed from the table, the performance for the parametric methods is not very stable for this small value of α_{tol}, as the outcomes vary substantially. It can also be seen that the nonparametric control limits from Section 3 perform best for every distribution except the normal distribution, as the values are in general closest to p. It is of course not unexpected that the other methods perform better for normally distributed data. The control limits of Goedhart, Schoonhoven, and Does (2018) are derived specifically for this case, while the Johnson transformation from Chou, Polansky, and Mason (1998) is only applied when the normality assumption is rejected. The Pearson system also subsumes the normal distribution as a special case in various types. Although the Pearson system also subsumes the t-distribution and the chi-square distribution as special cases, these are each only incorporated in a single type of the Pearson system. Because the type has to be estimated as well in the bootstrap procedure, this yields some extra estimation uncertainty, resulting in an unsatisfactory performance for these distributions. The lognormal distribution is not incorporated in the Pearson system and the corresponding results are very unsatisfactory for the bootstrap procedure. Due to the highly skewed character of the lognormal distribution, small deviations in the estimation of the limits lead to large changes in the CFAR or CARL.

We also observe some influence of the sample size m on the performance. However, increasing the sample size to values such as 2,500 does not seem to overcome the performance issues corresponding to small values of α_{tol} for the parametric methods. Taking these aspects in consideration, together with the fact that the distribution of the data is generally unknown, we argue that the proposed limits in Section 3 yield a more stable and satisfying performance when moderately large sample sizes are available. When sample sizes are small, one has to lower the demands placed on the control chart performance, which can be done through parameter settings (e.g., p and α_{tol}). Especially when combined with subgroup or aggregated statistics, this may not be unreasonable. For example, consider a process for which an observation is collected once every day. For a control chart with individual observations, $\alpha_{tol} = 0.0027$ corresponds to about one false signal per year. When using some weekly (aggregated) statistic, such as an average, instead of the individual daily observations, the same value of $\alpha_{tol} = 0.0027$ would lead to a false signal about once every 7 years. In order to remain with the one false signal per year, one has to increase α_{tol} accordingly.

Table 4. Comparison of exceedance probabilities for $\alpha_{tol} = 0.0027$ and $p = 0.1$.

	m	Normal	Chou	GK	YM
Normal	100	0.0990	0.0970	0.3830	0.0927
	250	0.0960	0.1180	0.1830	0.2146
	500	0.1090	0.1210	0.0900	0.2477
	1,000	0.1000	0.1150	0.0880	0.1526
	1,500	0.1000	0.1200	0.0890	0.0988
	2,500	0.0960	0.1310	0.0870	0.1034
χ^2_4	100	0.9910	0.4590	0.8630	0.0814
	250	1.0000	0.4030	0.7900	0.1862
	500	1.0000	0.3700	0.6870	0.2299
	1,000	1.0000	0.3010	0.6260	0.1494
	1,500	1.0000	0.2900	0.5500	0.0884
	2,500	1.0000	0.2590	0.4650	0.0987
Lognormal	100	0.9800	0.3860	1.0000	0.0963
	250	0.9960	0.3730	1.0000	0.2200
	500	0.9990	0.3520	1.0000	0.2401
	1,000	0.9990	0.3190	1.0000	0.1613
	1,500	1.0000	0.3220	1.0000	0.0912
	2,500	0.9990	0.3270	1.0000	0.0943
t_4	100	0.9880	0.4910	0.5550	0.1243
	250	0.9940	0.3950	0.2650	0.2617
	500	0.9980	0.4590	0.2240	0.2744
	1,000	0.9990	0.5660	0.2310	0.1705
	1,500	0.9980	0.6420	0.2610	0.0909
	2,500	1.0000	0.7360	0.2990	0.0999

This holds for every situation where observations are aggregated, regardless of the actual time difference between observations. Thus, when the Phase I sample size is small, when aggregating observations, or when using subgroup statistics, we recommend using increased values of α_{tol}.

4.3.2. Out-of-control performance

In order to compare the out-of-control performance, we have evaluated the alarm rates of the proposed methods under normality and compared this to the alarm rates of the method from Goedhart, Schoonhoven, and Does (2018), which is designed for normal theory. We consider a Phase I data set that follows a standard normal distribution and a Phase II data set where a shift of size δ in the mean has occurred, such that the latter follows a $N(\delta, 1)$ distribution. Note that $\delta = 0$ means that the process is in-control, in which case the alarm rate is actually the false-alarm rate. For $p = 0.1$ and various combinations of m, α_{tol}, and δ, we have determined the average alarm rate for each method, as an indicator for the out-of-control capability of the control chart. In particular, for all considered parameter combinations, we have applied the following procedure:

1. A dataset consisting of m observations is drawn from the specified distribution (i.e., $N(0, 1)$).
2. The methods under consideration are applied to determine the estimated control limits $\hat{LCL}$ and $\hat{UCL}$ for each method.
3. The conditional alarm rate $1-P(\hat{LCL} \leq Y \leq \hat{UCL})$ is determined, where Y is a future (Phase II) out-of-control observation drawn from a $N(\delta, 1)$ distribution.
4. Steps 1 to 3 are repeated for 10,000 different Phase I samples, after which the average alarm rate is calculated as the average over all the conditional alarm rates. The higher this value is, the better the out-of-control detection capability.

The results of this procedure are displayed in Table 5. As can be observed, the method for normal theory has slightly higher alarm rates compared with the proposed method. This holds for almost all combinations of α_{tol}, m, and δ (also in-control), but is most visible for smaller sample sizes. Moreover, the relative differences between the average alarm rates of the two methods are slightly larger when considering smaller values of α_{tol}, thus moving further in the tails of the distribution. Obviously, the method designed for normal theory yields a better performance for normally distributed data, as it is able to exploit extra information (i.e., distribution) regarding the process under consideration. However, in the cases where its performance excels (i.e., smaller values of m and α_{tol}), deviations from normality are more difficult to detect and have more impact on the control chart performance.

Table 5. Comparison of average alarm rates for $p = 0.1$.

		δ					
		0		1		2	
α_{tol}	m	YM	Normal	YM	Normal	YM	Normal
	100	0.0284	0.0333	0.1094	0.1266	0.3890	0.4347
	250	0.0354	0.0384	0.1320	0.1418	0.4434	0.4667
0.05	500	0.0376	0.0415	0.1387	0.1499	0.4604	0.4820
	1,000	0.0410	0.0439	0.1485	0.1560	0.4790	0.4928
	1,500	0.0434	0.0449	0.1544	0.1585	0.4907	0.4973
	2,500	0.0444	0.0460	0.1570	0.1610	0.4939	0.5013
	100	0.0069	0.0055	0.0321	0.0359	0.1463	0.2053
	250	0.0057	0.0067	0.0346	0.0429	0.1784	0.2329
0.01	500	0.0055	0.0075	0.0354	0.0466	0.2005	0.2474
	1,000	0.0066	0.0081	0.0421	0.0497	0.2291	0.2574
	1,500	0.0073	0.0084	0.0451	0.0510	0.2409	0.2620
	2,500	0.0075	0.0088	0.0464	0.0525	0.2458	0.2665
	100	0.0029	0.0025	0.0130	0.0204	0.0634	0.1422
	250	0.0035	0.0032	0.0211	0.0249	0.1205	0.1649
0.005	500	0.0029	0.0036	0.0206	0.0276	0.1303	0.1771
	1,000	0.0027	0.0039	0.0216	0.0297	0.1469	0.1872
	1,500	0.0031	0.0041	0.0241	0.0306	0.1591	0.1913
	2,500	0.0035	0.0043	0.0268	0.0317	0.1727	0.1953
	100	0.0010	0.0013	0.0051	0.0125	0.0240	0.1014
	250	0.0017	0.0016	0.0112	0.0153	0.0648	0.1199
0.0027	500	0.0018	0.0018	0.0139	0.0172	0.0915	0.1309
	1,000	0.0015	0.0021	0.0131	0.0186	0.0991	0.1387
	1,500	0.0014	0.0022	0.0126	0.0193	0.1015	0.1421
	2,500	0.0016	0.0023	0.0149	0.0200	0.1158	0.1457

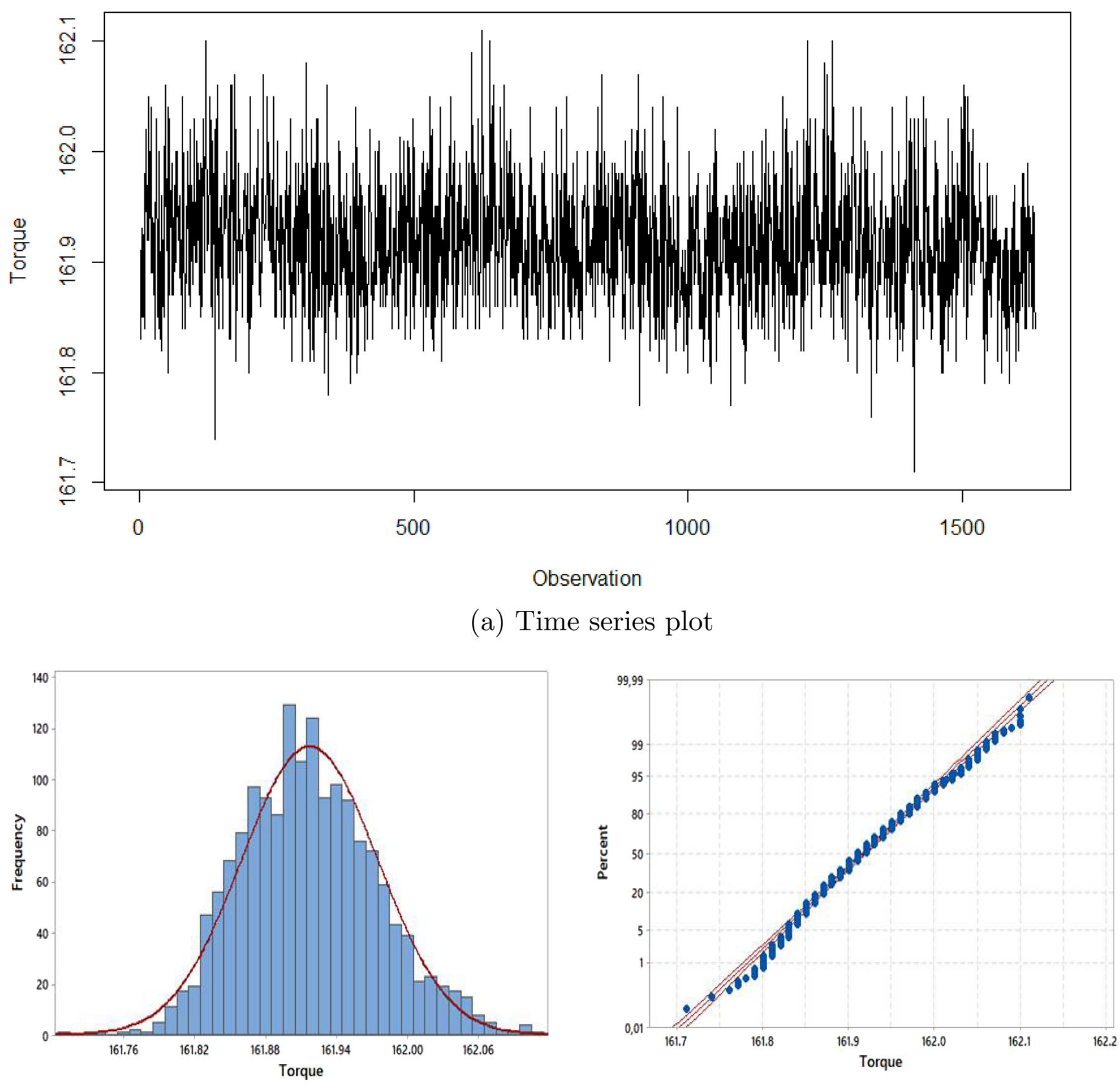
5. Application of the proposed control chart

In this section, we demonstrate the application of the proposed control chart using data from an application. A data set is available of the torque of Torque-to-Yield bolts that are used as fasteners in engines at a subsidiary of PACCAR (a global truck company). The bolts are tightened at several different positions of the engines using a specific fastening procedure. At several moments during this procedure, the torque is measured (in Newton-meter) for each bolt. We consider the first moment of measurement here. The measurements are performed by a process engineer to set up the process monitoring. It is important to detect out-of-control situations, such as trends or anomalies in the applied torque, as this can indicate problems with either the bolt(s) or the fastening procedure. For example, the performance of the used wrenches can deteriorate over time, which may result in fasteners being too tight or too loose.

To this end, an initial Phase I dataset of $m = 1{,}632$ observations is collected by the process engineer to construct the control limits. These data have been checked for possible errors and anomalies. A time series plot of the data is shown in Figure 2(a). Figure 2(b) shows a histogram of the data, with a normal distribution fit on top of it. At first glance, the normal distribution appears to be a proper fit. However, Figure 2(c) shows a normal probability plot where it can be seen that the normal distribution is still not able to model the extreme tails well. We found similar (or worse) results for other distribution functions fitted to this data. Of course, in practice, these distribution functions never provide the 'correct' model, but only adequate and useful models. While this is generally sufficient to model the bulk of the data, this does not hold for the extreme tails considered in SPM. Although the difference may not appear to be very large, minor deviations in the extreme tail have substantial consequences in terms of the ARL. As discussed in Section 4.3, deviations in model assumptions lead to a deterioration of the control chart performance. Especially for moderate to large sample sizes, such as available here, the use of the proposed nonparametric methods therefore seems more appropriate, as these are not sensitive to such model deviations.

The next step is to determine the control limits that should be used to monitor this process using the procedure described in Section 3. For this dataset, this results in the following steps:

1. First, we determine our parameters. We have $m = 1{,}632$ and choose $\alpha = 0.0027$ and $p = 0.1$. From



(a) Time series plot

(b) Histogram

(c) Normal probability plot

Figure 2. Time series plot, histogram, and normal probability plot of the Phase I data of torque values.

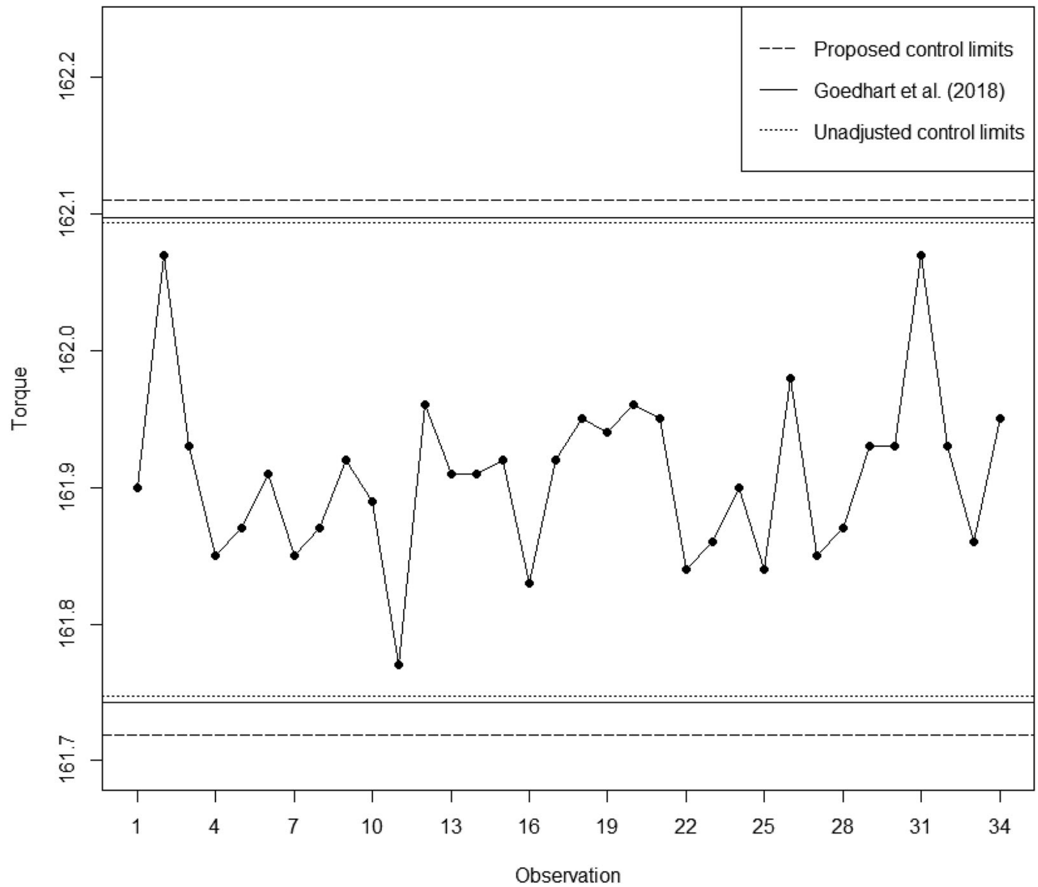


Figure 3. Control chart for Phase II monitoring of torque values. Proposed control limits (dashed), control limits adjusted for normally distributed data (straight line), and unadjusted control limits (dotted lines) are indicated.

Eq. [3] and Table 1, it follows that $m = 1632 > m_2(0.0027, 0.1) = 1440$, so that we have enough observations to construct interpolated control limits.

2. Next, given the parameters, we determine the smallest integer value k for which Eq. [4] holds. In this case, this value is equal to $k = 1{,}631$. This means that the maximum number of observations that can be trimmed from the ordered data equals $m - k - 1 = 0$. Consequently, we need $r = 1$ and $s = m = 1632$, so that we start with the tolerance interval $[X_{(1)}, X_{(1632)}]$. In this case, this interval is equal to $[161.71, 162.11]$.
3. After that, we determine λ_1, $X_{(r^*)}$, and $X_{(s^*)}$ according to Eq. [5]. Let $B \sim Bin(m, 1 - \alpha_{tol})$ as in Section 3.1.1. The coverage probability of $[X_{(1)}, X_{(1632)}]$ is then equal to $P(B \leq k - 1) = 0.9343$, and the coverage probability of $[X_{(1)}, X_{(1631)}]$ and $[X_{(2)}, X_{(1632)}]$ is equal to $P(B \leq k - 2) = 0.8160$. We then find $\lambda_1 = 0.7101$ from Eq. [5]. Next, using $X_{(2)} = 161.74$ and $X_{(1631)} = 162.10$, we find $X_{(1^*)} = 161.7187$ and $X_{(1632^*)} = 162.1071$. This means that we obtain the intervals $[X_{(1)}, X_{(1632^*)}] = [161.7100, 162.1071]$ and $[X_{(1^*)}, X_{(1632)}] = [161.7187, 162.1100]$.
4. Finally, we determine the control limits using the shortest of the two obtained intervals in the previous step. This is equal to $[161.7187, 162.1100]$, such that we find $\hat{LCL} = 161.7187$ and $\hat{UCL} = 162.1100$.

Note that, if we had chosen $p = 0.2$ instead, we would have obtained $k = 1{,}630$ in step 2 above. This would mean that the maximum number of observations that can be trimmed from the ordered data would equal 1 instead of 0. With the restriction $s = m - r + 1$, the initial tolerance limits would still be $X_{(1)}$ and $X_{(1632)}$. However, as described in Section 3.1.1, removing this restriction makes it possible to construct intervals with a smaller coverage probability, but still more than the nominal desired value $1 - p$, by trimming an observation of the ordered data on one of the two sides. This would give the intervals $[X_{(1)}, X_{(1631)}]$ and $[X_{(2)}, X_{(1632)}]$ as starting points. The next step would then be to perform the interpolation procedure for both of these intervals and choosing the shortest of the four resulting intervals.

The obtained limits can be used to monitor this process. As an illustration, we have a Phase II sample available consisting of 34 observations, as is illustrated in Figure 3. For comparison purposes, we have also indicated the unadjusted control limits according to Eq. [1] with $\alpha = 0.0027$ (resulting in the standard 3-sigma limits), as well as the adjusted control limits from Goedhart, Schoonhoven, and Does (2018) with $\alpha_{tol} = 0.0027$ and $p = 0.01$. As Phase I estimators, we consider the sample average (161.92) and standard deviation (0.0577) of the data. The resulting unadjusted lower and upper control limits then equal 161.7469 and 162.0931, respectively. Because of the large sample size, the control limit constant as calculated from Goedhart, Schoonhoven, and Does (2018) is not very different from 3, with a value of 3.07. Using this value results in lower and upper control limits of 161.7429 and 162.0971, respectively. As can be seen, the proposed control limits are slightly wider than the unadjusted limits and the Goedhart, Schoonhoven, and Does (2018) limits. However, one should note that the latter two sets of limits are only appropriate for normally distributed data. As can be observed, there are no out-of-control signals for this Phase II data set for either set of control limits.

6. Concluding remarks

To correct for the effect of parameter estimation, we propose the application of nonparametric tolerance intervals in statistical process monitoring. Because an appropriate data distribution and its corresponding parameters are generally unknown, they have to be estimated using a Phase I reference sample. Because of the variability induced by sampling (stochastic error), the estimated control limits will vary for different Phase I samples. This leads to different control chart performance for different practitioners.

We propose control limits based on nonparametric tolerance intervals in order to satisfy the *exceedance probability criterion* introduced by Albers and Kallenberg (2004a). This criterion aims to control the control chart performance by guaranteeing at least a prespecified control chart performance with a prespecified probability. The proposed control limits perform well in satisfying this criterion in general, especially when moderately large Phase I samples are available.

A major advantage of nonparametric control charts is that they can be applied to any monitoring statistic of interest, by treating subgroup statistics as individual observations. In that way, the proposed control chart can be applied to X, $\bar{X}$, R, and S control charts or even other statistics, regardless of the distribution under consideration.

About the authors

Rob Goedhart is an Assistant Professor in the Department of Operations Management and senior consultant at the Institute for Business and Industrial Statistics of the University of Amsterdam, the Netherlands. His current research topic is control charting techniques with estimated parameters.

Marit Schoonhoven is an Associate Professor at the Department of Operations Management and senior consultant at the Institute for Business and Industrial Statistics of the University of Amsterdam, the Netherlands. Her current research interests include control charting techniques and operations management methods.

Ronald J.M.M. Does is a Professor of Industrial Statistics at the University of Amsterdam, Director of the Institute for Business and Industrial Statistics and Head of the Department of Operations Management at the University of Amsterdam. He is an associate editor of the *Journal of Quality Technology* and *Quality Engineering*. He is the recipient of the Hunter Award (2008), Shewhart Medal (2019), and Box Medal (2019). He is a fellow of the American Statistical Association, a fellow of the American Society for Quality, an Academician of the International Academy for Quality, and an elected member of the International Statistical Institute. His current research activities include the design of control charts for nonstandard situations, healthcare engineering, and operations management methods.

Acknowledgments

The authors are very grateful to the editor and the anonymous referees. Their constructive and detailed comments have contributed to substantial improvements of this article.

References

Albers, W., and W. C. M. Kallenberg. 2004a. Are estimated control charts in control? *Statistics* 38 (1):67–79.

Albers, W., and W. C. M. Kallenberg. 2004b. Estimation in Shewhart control charts: Effects and corrections. *Metrika* 59 (3):207–34.

Albers, W., W. C. M. Kallenberg, and S. Nurdiati. 2004. Parametric control charts. *Journal of Statistical Planning and Inference* 124 (1):150–84.

Albers, W., and W. C. M. Kallenberg. 2005. New corrections for old control charts. *Quality Engineering* 17 (3):467–73.

Box, G. E., and D. R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society* 26 (2): 211–52.

Chakraborti, S., P. Van der Laan, and S. T. Bakir. 2001. Nonparametric control charts: An overview and some results. *Journal of Quality Technology* 33 (3):304–15.

Chakraborti, S., P. Qiu, and A. Mukherjee. 2015. Editorial to the special issue: Nonparametric statistical process

control charts. *Quality and Reliability Engineering International* 31 (1):1–2
Chou, Y. M., A. M. Polansky, and R. L. Mason. 1998. Transforming non-normal data to normality in statistical process control. *Journal of Quality Technology* 30 (2):133–41.
Diko, M. D., R. Goedhart, S. Chakraborti, R. J. M. M. Does, and E. K. Epprecht. 2017. Phase II control charts for monitoring dispersion when parameters are estimated. *Quality Engineering* 29 (4):605–22.
Gandy, A., and J. T. Kvaløy. 2013. Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics* 40 (4):647–68.
Goedhart, R., M. Schoonhoven, and R. J. M. M. Does. 2016. Correction factors for Shewhart X and $\bar{X}$ control charts to achieve desired unconditional ARL. *International Journal of Production Research* 54 (24):7464–79.
Goedhart, R., M. Schoonhoven, and R. J. M. M. Does. 2017. Guaranteed in-control performance for the Shewhart X and $\bar{X}$ control charts. *Journal of Quality Technology* 49 (2): 155–71.
Goedhart, R., M. Schoonhoven, and R. J. M. M. Does. 2018. On guaranteed in-control performance for the Shewhart X and $\bar{X}$ control charts. *Journal of Quality Technology* 50 (1): 130–2.
Goedhart, R., M. M. da Silva, M. Schoonhoven, E. K. Epprecht, S. Chakraborti, R. J. M. M. Does, and A. Veiga. 2017. Shewhart control charts for dispersion adjusted for parameter estimation. *IISE Transactions* 49 (8):838–48.
Huberts, L. C. E., M. Schoonhoven, R. Goedhart, M. D. Diko, and R. J. M. M. Does. 2018. The performance of control charts for large non-normally distributed datasets. *Quality and Reliability Engineering International* 34 (6):979–96.
Jensen, W. A., L. A. Jones-Farmer, C. W. Champ, and W. H. Woodall. 2006. Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology* 38 (4):349–64.
Jones, M. A., and S. H. Steiner. 2012. Assessing the effect of estimation error on risk-adjusted CUSUM chart performance. *International Journal for Quality in Health Care* 24 (2):176–81.]
Jones-Farmer, L. A., W. H. Woodall, S. H. Steiner, and C. W. Champ. 2014. An overview of Phase 1 analysis for process improvement and monitoring. *Journal of Quality Technology* 46 (3):265–80.
Krishnamoorthy, K., and T. Mathew. 2009. *Statistical tolerance regions: Theory, applications, and computation*. New York, NY: John Wiley & Sons.
Low, Y. M. 2013. A new distribution for fitting four moments and its applications to reliability analysis. *Structural Safety* 42:12–25.
Montgomery, D. C. 2013. *Introduction to statistical quality control*. New York, NY: John Wiley & Sons.
Psarakis, S., A. K. Vyniou, and P. Castagliola. 2014. Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International* 30 (8):641–50.
Qiu, P. 2018. Some perspectives on nonparametric statistical process control. *Journal of Quality Technology* 50 (1):49–65.
Quesenberry, C. P. 1993. The effect of sample size on estimated limits for $\bar{X}$ and X control charts. *Journal of Quality Technology* 25 (4):237–47.
Sakia, R. M. 1992. The Box–Cox transformation technique: A review. *The Statistician* 41 (2):169–78.
Saleh, N. A., M. A. Mahmoud, M. J. Keefe, and W. H. Woodall. 2015. The difficulty in designing Shewhart $\bar{X}$ and X control charts with estimated parameters. *Journal of Quality Technology* 47 (2):127–38.
Saleh, N. A., M. A. Mahmoud, L. A. Jones-Farmer, I. M. Zwetsloot, and W. H. Woodall. 2015. Another look at the EWMA control chart with estimated parameters. *Journal of Quality Technology* 47 (4):363–82.
Wilks, S. S. 1941. Determination of sample sizes for setting tolerance limits. *The Annals of Mathematical Statistics* 12 (1):91–6.
Young, D. S., and T. Mathew. 2014. Improved nonparametric tolerance intervals based on interpolated and extrapolated order statistics. *Journal of Nonparametric Statistics* 26 (3):415–32.

Appendix A: Chou, Polansky, and Mason (1998) Transformations to normality

In this article, we have discussed the transformation to normality as described in Chou, Polansky, and Mason (1998). The procedure for this transformation is as follows:

1. First, test the data for normality using the Shapiro–Wilk test. If normality is not rejected, proceed to step 3. Otherwise, continue with step 2.
2. Transform the data as described in Chou, Polansky, and Mason (1998). This can be done, for example, with the R-package RE.Johnson.
3. Compute the average and standard deviation of the (transformed) data and use these to construct the control limits using Goedhart, Schoonhoven, and Does (2018) with the input parameters m, α_{tol} and p.
4. If the data were transformed, transform the control limits back to their original scale by inverting the transformation used in step 2. If not, no further actions are required.

Appendix B: Bootstrap procedure

In this article, we have discussed the application of the bootstrap procedure of Gandy and Kvaløy (2013) in combination with the Pearson system of distributions. The procedure for this transformation is as follows:

1. Fit a Pearson distribution on the data set based on the first four central moments. Procedures to do this are available in various statistical software programs, such as the function pearsonFitM from the R-package PearsonDS. We denote this fitted distribution as $\hat{F}$.
2. From the fitted distribution $\hat{F}$, draw m_B (e.g., 500 or 1,000) bootstrap samples of size m each. For each bootstrap sample, fit a Pearson distribution, similarly to step 1. Denote the fitted distribution of the bootstrap sample as $\tilde{F}$.
3. For each bootstrap sample $(i = 1, \ldots, m_B)$, determine the required value $\alpha_{B,i}$ such that $\hat{F}(\tilde{F}^{-1}(\alpha_{B,i}/2)) +$

$1-\hat{F}(\tilde{F}^{-1}(1-\alpha_{B,i}/2)) = \alpha_{tol}$. Note that, for two-sided control charts, determining the limits through adjusting the chosen quantiles $\alpha/2$ and $1-\alpha/2$ seems to be the most reasonable choice for correcting equally on both sides, as treating the LCL and UCL separately may lead to an infinite number of solutions. Note also that, due to the bounded character of some of the Pearson distribution types, this equation does not always have a solution. This happens in particular for highly skewed distributions. When this is the case, set $\alpha_{B,i} = \alpha_{tol}/100$ or some other small value. The reason to choose such a rather small value is that, when it occurs rarely, it won't impact the outcome of the bootstrap procedure (see the next step). However, when it's not rare, it means that such small quantiles are actually required to control the performance, but going for even smaller values would result in control limits converging to infinity on one side.

4. Determine the required adjusted value α_{adj} as the p-quantile of the vector α_B obtained in the previous step. Then construct control limits according to $\hat{LCL} = \hat{F}^{-1}(\alpha_{adj}/2)$ and $\hat{UCL} = \hat{F}^{-1}(1-\alpha_{adj}/2)$.