Computers & Industrial Engineering 174 (2022) 108806

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

On designing an optimal SPRT control chart with estimated process parameters under guaranteed in-control performance

J.W. Teoh [a], W.L. Teoh [a,*], Michael B.C. Khoo [b], Philippe Castagliola [c], W.H. Moy [d]

[a] School of Mathematical and Computer Sciences, Heriot-Watt University Malaysia, 62200 Putrajaya, Malaysia
[b] School of Mathematical Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia
[c] Nantes Université & LS2N UMR 6004, Ave. du Professeur Jean Rouxel, BP 539, Carquefou 44475, France
[d] School of Materials & Mineral Resources Engineering, Universiti Sains Malaysia, 14300 Penang, Malaysia

ARTICLE INFO

Keywords:
Average extra quadratic loss
Guaranteed in-control performance
Markov chain
Parameter estimation
SPRT control chart
Statistical process control

ABSTRACT

Being one of the most sophisticated control charts, the sequential probability ratio test (SPRT) chart possesses fast detection ability across a broad range of process shifts. The SPRT chart has the advantage of sampling only a modest number of observations. Thus far, the SPRT chart has been developed under the assumption that the process parameters are known. As process parameters are often unknown in practice, this article reveals that parameter estimation from a limited amount of Phase-I data, leads to excessive false alarms and unstable chart's performances. To counter these problems, this article advocates the use of adjusted control limits to ensure that a sufficiently high proportion of the in-control conditional average time to signal value, is greater than a pre-specified level. This increasingly prevalent design of control charts is known as Guaranteed In-Control Performance (GICP). In this article, we propose an optimization design for the SPRT chart with estimated process parameters, by minimizing the expected value of the average extra quadratic loss under the GICP framework. Theoretical derivations by means of the Markov chain approach are developed in this article to evaluate the run-length properties of the SPRT chart with estimated process parameters. Results show that the overall performance of the proposed optimal SPRT chart is almost twice as good as the optimal CUSUM chart over a given range of process mean shifts. Finally, an implementation of the proposed optimal SPRT chart is demonstrated with real industrial data obtained from an epitaxial process.

1. Introduction

One might wonder how control charts have emerged as the go-to tool in many endeavors related to statistical process monitoring. Indubitably, the control charting method has received plenty of spotlight over the past few decades, owing to its powerful ability to reveal useful information about a process of interest. Anwar et al. (2021), Weiß et al. (2021), and Mehmood et al. (2022), to name a few, devoted substantial efforts in research work aimed at improving the control charts' performances. To date, there has been a growing emphasis on the importance of developing control charts with both high detection speed and stable in-control performance. There has also been strong consensus on the consideration of economic factors and productive efficiencies in the design of control charts, see, for example, Salmasnia et al. (2017), and Tran et al. (2021). A great deal of process monitoring literature focuses on the exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) control charts (see, for example, Jones et al., 2004; Castagliola et al., 2016). These control charts are generally favored over the traditional Shewhart $\bar{X}$ chart due to their sensitivities towards small departures in the process quality characteristic (Montgomery, 2009). In this article, we navigate our research around a powerful yet under-explored parametric control chart known as the sequential probability ratio test (SPRT) control chart.

The SPRT was first formulated by Abraham Wald for the purpose of statistical hypothesis testing (Wald, 1945). According to Wald (1947), the SPRT proves to yield around 50% reduction in the number of observations required on average compared to other test procedures based on fixed subgroups. This makes Wald's SPRT especially desirable for applications whereby data are inherently limited. Subsequently, Stoumbos and Reynolds (1997) developed the SPRT control chart

* Corresponding author.
E-mail addresses: t.jing_wei@hw.ac.uk (J.W. Teoh), wei_lin.teoh@hw.ac.uk (W.L. Teoh), mkbc@usm.my (M.B.C. Khoo), philippe.castagliola@univ-nantes.fr (P. Castagliola), whmoy@yahoo.com (W.H. Moy).

https://doi.org/10.1016/j.cie.2022.108806
Received 9 June 2022; Received in revised form 5 October 2022; Accepted 6 November 2022
Available online 10 November 2022

applying Wald's SPRT at a fixed sampling interval (FSI). They successively showed that the proposed SPRT chart has superior performance over the variable sample size and sampling interval (VSSI) $\bar{X}$ and CUSUM charts. This remarkable performance of the SPRT chart has received recognition and appreciation from many researchers, see, for example, Ou et al. (2011a, b), Willis (2011), Zhang et al. (2014), and Mahadik et al. (2021). The SPRT control chart has since been deployed in numerous quality control applications, such as those in the manufacturing industries (Stoumbos & Reynolds, 1999; Ou et al., 2011c). Since the SPRT chart expects observations to be taken in negligible time, it is very well-suited for productions in which sampling procedures can be conducted within a relatively short time. For example, Stoumbos and Reynolds (1997) employed the SPRT control chart in monitoring the thickness of thrust washers used in the transmissions of automobiles. Ou et al. (2011c) studied the FSI and variable sampling interval (VSI) SPRT charts on three different applications, i.e., monitoring the thickness of the silicon dioxide layer of a semiconductor component, the breaking strength of nylon fibers, and the diameter of a special drill. To date, all existing work related to the SPRT chart have been developed under the assumption that the mean and dispersion of the process quality characteristic are known. However, practitioners reveal that this assumption is frequently violated in quality control setting. Some baseline information about the underlying process is usually needed to implement the control chart (Chakraborti, 2006; Saleh et al., 2015a).

In manufacturing environment, true process parameters are rarely known, and any attempt to estimate them from historical data may result in extremely volatile outcomes. This issue was discussed by Jensen et al. (2006) and was subsequently addressed by several researchers, such as Chakraborti et al. (2008) and Jones-Farmer et al. (2014). Generally, they suggested to estimate process parameters from a set of in-control Phase-I samples. These Phase-I samples are used in the calibration of control charts before proceeding to Phase-II process monitoring. Many researchers have advocated the use of large Phase-I samples (see, for example, Zhang et al., 2014; Saleh et al., 2015a) to reduce practitioner-to-practitioner variability in control charts' performances. All researchers agree that, when a small number of Phase-I samples is used, the conditional average run length (CARL) obtained through different practitioners can vary by an alarmingly large degree. Zhang et al. (2014) suggested increasing the number of Phase-I samples, so that the standard deviation of the CARL falls below 10% of the desired average run length (ARL). Nonetheless, Saleh et al. (2015a) argued that such a recommendation may be unrealistic to be implemented in practice, since the collection of a large dataset can be time consuming and expensive.

Another issue arising from the use of estimated process parameters in traditional control charts is the strikingly massive false alarms obtained by practitioners. This calls for the innovation of control charts with adjusted control limits, which was initially pioneered by Albers et al. (2005), to guarantee satisfactorily low false alarm rates among practitioners. This ideology is known as Guaranteed In-Control Performance (GICP). Particularly, Albers et al. (2005) and Gandy and Kvaløy (2013) proposed the use of the Exceedance Probability Criterion (EPC) and the bootstrap approach, respectively, to ensure that the in-control CARL of a control chart is at least equal to a pre-specified level with a high probability. As the ideology grows in popularity, many researchers have endorsed the use of GICP in the design of control charts with estimated process parameters (see, for example, Goedhart et al., 2017; Capizzi & Masarotto, 2020; Jardim et al., 2020). Capizzi and Masarotto (2020) even provided a compelling statement in their article, calling GICP the "*de facto* standard design of control charts". However, Diko et al. (2019b) discovered that adjusted control limits under the GICP framework may lead to reduced sensitivity in control charts, especially when the process shifts are small. To address the trade-off between the in-control and out-of-control performances, they suggested to increase the error probability (p) and/or to introduce a tolerance term (ε) in the formulation of the EPC. Nevertheless, the proposed method has not solved the problem at its root. The chart's performances usually only stabilize when mean shifts (δ) are greater than or equal to one. Motivated by the need to devise a better solution, we carry through our attempt in combining the optimal design of a control chart with the GICP framework. We show that, without compromising the in-control performance further, the out-of-control performance of our proposed SPRT chart with estimated process parameters can be improved even for mean shifts as small as 0.6.

To the best of our knowledge, none of the existing literature has investigated the SPRT chart with estimated process parameters. Therefore, in this article, we fill the gap by assessing the impact of Phase-I parameter estimation on the performances of the SPRT chart. We also derive the conditional run-length properties of the SPRT chart with estimated process parameters, i.e., the conditional average sample number (CASN), the conditional average time to signal (CATS), and the conditional standard deviation of the time to signal (CSDTS), by means of the Markov chain approach. It is important to note that these properties are expressed as random variables of the Phase-I data set. Hence, to accurately represent these metrics, we provide their unconditional measures, i.e., the expected value of the CASN (AASN), the expected value of the CATS (AATS), the standard deviation of the CATS (SDATS), and the average of the CSDTS (ASDTS). It has been brought to our attention that the standard deviation of the time to signal (SDTS) has yet to be studied in any literature related to the SPRT chart with known process parameters. For completeness, we provide full derivations of the SDTS, CSDTS, and ASDTS formulae for both the cases of known and estimated process parameters in this article. Throughout this article, we consider the steady-state performance of the SPRT chart with estimated process parameters. This is consistent with the anticipation that, in real situations, a sustained shift occurs after a process remains in-control for a sufficiently long duration.

In the design of control charts, it is especially crucial to consider the effectiveness of the chart in detecting multiple unknown shift sizes, rather than a single deterministic shift size. Control charts that are optimally designed to safeguard against a range of process shifts are proven to be more robust towards uncertain shift distributions. Contemporarily, the average extra quadratic loss (AEQL) criterion has been widely used in the design of control charts with known process parameters (see, for example, Ou et al., 2011b; Adeoti & Malela-Majika, 2020). Hence, we adopt the expected value of the AEQL (AAEQL) criterion in the optimization design of the SPRT chart with estimated process parameters. Furthermore, we apply the GICP framework alongside the optimization design to tackle the issue of high false alarms. With an extensive literature review, we found that this concept has yet to be proposed by any researcher. Hence, by combining both the optimization design and the GICP framework, we develop a new optimization algorithm for the SPRT chart with estimated process parameters in this article. The proposed algorithm does not only render the best overall out-of-control performance in terms of the AAEQL, but also promises a satisfactory in-control performance that is customizable based on the needs of a practitioner.

The organization of this article is as follows. In the next section, we visit the SPRT chart with both known and unknown process parameters. We derive theoretical formulae for the conditional and unconditional run-length properties of the SPRT chart with estimated process parameters by means of the Markov chain approach. In Section 3, we argue that the SPRT chart with traditional control limits developed by Ou et al. (2011b) produces large numbers of false alarms when process parameters are estimated. We then present two optimization models for the SPRT chart with estimated parameters in Section 4 based on two approaches, i.e., the (i) AATS-matching and (ii) GICP approaches. Section 4 first highlights the drawback of the first approach. It is then followed by a description of how GICP solves the problem from a conditional perspective, and how the optimization design leads to neutralized out-of-control performances even for relatively small mean shifts. To

demonstrate the implementation of our proposed optimal SPRT chart with estimated process parameters, we include an application on real manufacturing data in Section 5. Concluding remarks are provided in the final section.

2. A thorough look at the SPRT chart

2.1. The SPRT chart with known process parameters

Suppose that a process quality characteristic X comes from a normal distribution with known mean μ_0 and standard deviation σ_0. Suppose further that we are interested in testing the null hypothesis $H_0 : \mu = \mu_0$ against the alternative hypothesis $H_1 : \mu = \mu_0 + \delta\sigma_0$ from a sequence of monitoring data, where μ represents the mean of the ongoing process. Note that $\delta > 0$ is a pre-specified upper-sided shift in units of standard deviation. When the process is in-control, we set $\delta = 0$ which coincides with H_0. At the outset (time $t = 0$), just after the SPRT chart is initiated, a "window period" of length $d > 0$ is allocated for potential startup intermission, such as machinery warm up, initialization of statistical software, etc. At time $t = d$, a sample of random size $(X_{1,1}, X_{1,2}, \ldots, X_{1,n})$ is taken from the process to formulate the very first SPRT, where n represents the total sample number until the termination of each SPRT (i.e., acceptance or rejection of H_0). Assume that each observation in the sample is obtained in negligible time, and each of them is sampled independently from an identical distribution.

In the upper one-sided SPRT chart, the j^{th} control statistic of the i^{th} SPRT, $U_{i,j}$, is computed recursively as follows:

$$U_{i,j} = U_{i,j-1} + Z_{i,j} - \gamma, \tag{1}$$

where

$$Z_{i,j} = \frac{X_{i,j} - \mu_0}{\sigma_0} \tag{2}$$

for $i = 1, 2, \ldots, j = 1, 2, \ldots, n$ and $\gamma > 0$ is a reference parameter. Here, the starting value of each SPRT, $U_{i,0}$, is set to be equal to 0. In a SPRT, a decision is reviewed at the end of the j^{th} sampling unit. In particular,

if $U_{i,j} > h$, the i^{th} SPRT is terminated, and an out-of-control signal is emitted,
if $g \le U_{i,j} \le h$, the i^{th} SPRT resumes sampling sequentially, and
if $U_{i,j} < g$, the i^{th} SPRT is terminated, and the process is indicated as in-control.

Here, g and h are the lower and upper control limits of the upper one-sided SPRT chart. Note that in this article, our discussion focuses mainly on the properties of an upper one-sided SPRT chart. A lower one-sided SPRT chart can be designed in a similar way by replacing $Z_{i,j}$ with $-Z_{i,j}$ in Equation (1). This is done to facilitate the detection of a downward mean shift, since a lower one-sided SPRT chart is symmetric to its upper one-sided counterpart.

After the first SPRT is concluded, sequences of SPRT's are conducted at fixed time intervals d until an out-of-control event is signaled. In other words, the first SPRT is initiated at time $t = d$, the second SPRT at $t = 2d$, the third SPRT at $t = 3d$, and so on until the chart ceases. Each SPRT computes its control statistic $U_{i,j}$ using the recursive formulae in Equations (1) and (2), with a starting value of $U_{i,0} = 0$. When an SPRT is terminated (i.e., $U_{i,j} < g$ or $U_{i,j} > h$), the contemporary sampling unit, j, is recorded as the sample number. Depending upon the process data, each SPRT is composed of a random number of observations. Hence, the average sample number (ASN) is used to effectively represent the average of these sample numbers in the long run.

The ASN, average time to signal (ATS), and SDTS of the SPRT chart with known process parameters can be formulated by means of the Markov chain approach (Stoumbos and Reynolds, 1997). Suppose that the region $[g, h]$ is partitioned into η subintervals, where each subinterval is a transient state S_k, for $k = 1, 2, \ldots, \eta$, of the Markov chain. Meanwhile, let the regions $(-\infty, g)$ and (h, ∞) represent the absorbing states of the Markov chain. We define the width of each subinterval as

$$\Delta = \frac{h - g}{\eta}, \tag{3}$$

and the midpoint of state S_k as

$$O_k = g + \Delta\cdot(k - 0.5) \tag{4}$$

Note that the ASN can be interpreted as the expected number of visits between transient states before $U_{i,j}$ is absorbed by the Markov chain. To determine the ASN, one needs to first construct a transition probability matrix $\mathbf{R} = \{r_{k,\ell}\}_{\eta\times\eta}$. Define $r_{k,\ell}$ as the one-step transition probability that the control statistic $U_{i,j}$ moves from state S_k to state S_ℓ, for $k, \ell = 1, 2, \ldots, \eta$. Then, $r_{k,\ell}$ is evaluated as

$$r_{k,\ell} = \Phi(\Delta\cdot(\ell - k + 0.5) + \gamma - \delta) - \Phi(\Delta\cdot(\ell - k - 0.5) + \gamma - \delta) \tag{5}$$

where $\Phi(\cdot)$ represents the cumulative distribution function of the standard normal distribution. Define the initial transition probability vector $\mathbf{B} = \{b_k\}_{\eta\times 1}$, where b_k represents the one-step transition probability that the initial control statistic $U_{i,0}$ transits to state S_k. The formula is given as

$$b_k = \Phi(g + \Delta\cdot k + \gamma - \delta) - \Phi(g + \Delta\cdot(k - 1) + \gamma - \delta). \tag{6}$$

Hence, the ASN can be found by the following formula:

$$\text{ASN} = 1 + \mathbf{B}^{\mathrm{T}}(\mathbf{I} - \mathbf{R})^{-1}\mathbf{1}, \tag{7}$$

where $\mathbf{I}$ is the identity matrix of size $\eta \times \eta$, and $\mathbf{1}$ is an $\eta \times 1$ vector filled with ones.

Stoumbos and Reynolds (1997) ascertained that the run length (RL) of the SPRT chart with known process parameters has a geometric distribution with probability $1 - \text{OC}(\delta)$. Here, the operating characteristic function $\text{OC}(\delta)$ is the probability that the SPRT accepts H_0, conditional on the mean shift δ. Define the acceptance transition probability vector $\mathbf{Q} = \{q_k\}_{\eta\times 1}$, where q_k gives the one-step transition probability from state S_k to the acceptance state $(-\infty, g)$. Then

$$q_k = \Phi(\Delta\cdot(0.5 - k) + \gamma - \delta). \tag{8}$$

Let P_0 be the probability that the SPRT accepts the null hypothesis, conditional on the first observation, i.e.,

$$P_0 = \Phi(g + \gamma - \delta). \tag{9}$$

Then, $\text{OC}(\delta)$ is obtained as

$$\text{OC}(\delta) = P_0 + \mathbf{B}^{\mathrm{T}}(\mathbf{I} - \mathbf{R})^{-1}\mathbf{Q}. \tag{10}$$

It is customary to compute the ATS in performance evaluation. The in-control ATS (ATS_0) is given by

$$\text{ATS}_0 = \frac{d}{1 - \text{OC}(0)}, \tag{11}$$

where $\text{OC}(0)$ is computed from Equation (10) by setting $\delta = 0$. For the steady-state out-of-control ATS (ATS_δ), assume that the mean shift occurs uniformly over any sampling interval d, and that the control statistic has reached its stationary distribution at the time when the process shifts. It follows that the steady-state ATS_δ can be expressed as

$$\text{ATS}_\delta = d \times \left(\frac{1}{1 - \text{OC}(\delta)} - \frac{1}{2}\right). \tag{12}$$

It is important to note that the formula in Equation (12) should only be used in cases where $\delta > 0$.

Apart from the ATS, practitioners might be interested in understanding the variability of the time to signal of the SPRT chart. To the

best of our knowledge, none of the literature has provided theoretical formulae for the in-control SDTS ($SDTS_0$) and steady-state out-of-control SDTS ($SDTS_\delta$) of the SPRT chart with known process parameters. Therefore, we derive the formulae for SDTS in this article. The in-control time to signal of an SPRT chart can be expressed as $d \times RL$, where $RL \sim$ Geometric $(1 - OC(\delta))$. Then, the $SDTS_0$ is

$$SDTS_0 = d\sqrt{\mathrm{Var}[RL]} = d\sqrt{\frac{OC(0)}{[1-OC(0)]^2}}. \tag{13}$$

On the other hand, the steady-state out-of-control time to signal of the SPRT chart can be expressed as $d \times (T + RL - 1)$, where T is a uniform random variable from 0 to 1. Assuming independence between run lengths and uniformly distributed time of process shift, we evaluate the $SDTS_\delta$ as

$$SDTS_\delta = d\sqrt{\mathrm{Var}[T] + \mathrm{Var}[RL-1]} = d\sqrt{\frac{1}{12} + \frac{OC(\delta)}{[1-OC(\delta)]^2}}. \tag{14}$$

The accuracy of both Equations (13) and (14), has been verified with Monte Carlo simulations.

Another credible measure of the overall performance for a control chart is the AEQL (Reynolds & Stoumbos, 2004; Ou et al., 2011a, b). The formula for the steady-state AEQL is provided as

$$AEQL = \frac{1}{\delta_{max} - \delta_{min}} \int_{\delta_{min}}^{\delta_{max}} \delta^2 ATS_\delta \, d\delta, \tag{15}$$

where ATS_δ can be computed from Equation (12), while δ_{min} and δ_{max} are pre-defined minimum and maximum allowable shift sizes, respectively. The use of quadratic weights δ^2 in Equation (15) signifies disproportionate weights distributed among ATSs at various mean shifts. In particular, the AEQL amplifies the ATSs at large mean shifts, while depreciating those evaluated at small mean shifts. This is sensible as practitioners are often more interested in understanding how control charts perform in relatively extreme situations, rather than situations where the process shifts are small. This rationale was also justified by Montgomery (2009), who claimed that quality is inversely proportional to the square of deviations from the true mean.

2.2. The SPRT chart with estimated process parameters

In practice, the process parameters μ_0 and σ_0 are rarely known, and they need to be estimated from a verified in-control Phase-I data set. Assume that a Phase-I data set of size m is available, and the observations $(Y_1, Y_2, \ldots, Y_m)$ are independent and normally distributed with mean μ_0 and standard deviation σ_0. We estimate μ_0 using the sample mean

$$\hat{\mu}_0 = \frac{1}{m}\sum_{\theta=1}^{m} Y_\theta, \tag{16}$$

and estimate σ_0 using the sample standard deviation

$$\hat{\sigma}_0 = \sqrt{\frac{1}{m-1}\sum_{\theta=1}^{m}(Y_\theta - \hat{\mu}_0)^2}. \tag{17}$$

It is well-known that the estimators $\hat{\mu}_0$ and $\hat{\sigma}_0$ can be transformed to construct pivotal quantities whose distributions do not depend on the true parameters μ_0 and σ_0.

The control statistic $\hat{U}_{i,j}$ of the upper one-sided SPRT chart with estimated process parameters is revised with $\hat{Z}_{i,j}$, $\hat{\mu}_0$, and $\hat{\sigma}_0$ in place of $Z_{i,j}$, μ_0, and σ_0, respectively, in Equations (1) and (2). Then $\hat{U}_{i,j}$ becomes

$$\hat{U}_{i,j} = \hat{U}_{i,j-1} + \hat{Z}_{i,j} - \gamma, \tag{18}$$

where

$$\hat{Z}_{i,j} = \frac{X_{i,j} - \hat{\mu}_0}{\hat{\sigma}_0}. \tag{19}$$

To evaluate the performances of the SPRT chart with estimated process parameters, we derive a new expression for the transition probability matrix $\hat{\mathbf{R}} = \{\hat{r}_{k,\ell}\}_{\eta\times\eta}$. Note that the matrix $\hat{\mathbf{R}}$ now depends upon the pivotal quantities that involve $\hat{\mu}_0$ and $\hat{\sigma}_0$. The transition probability $\hat{r}_{k,\ell}$ from state S_k to S_ℓ is

$$\hat{r}_{k,\ell} = \Phi\left[V(\Delta\cdot(\ell - k + 0.5) + \gamma) + \frac{W}{\sqrt{m}} - \delta\right] - \Phi\left[V(\Delta\cdot(\ell - k - 0.5) + \gamma) + \frac{W}{\sqrt{m}} - \delta\right], \tag{20}$$

where the random variables $V = \hat{\sigma}_0/\sigma_0$ and $W = (\hat{\mu}_0 - \mu_0)/(\sigma_0/\sqrt{m})$ are pivotal quantities. The detailed derivations of $\hat{r}_{k,\ell}$ can be found in Appendix A.

The pivotal quantity W is the well-known z-transform of the sample mean of a normally distributed random variable. Hence, W follows the standard normal distribution $N(0,1)$. Also, it can be shown that $(m-1)V^2$ follows a chi-squared distribution with $m-1$ degrees of freedom. Therefore, the probability density function (pdf) of the pivotal quantity V can be derived by transformation of variables, and is given as

$$f_V(v) = 2(m-1)v\, f_{\chi^2}\left((m-1)v^2 | m-1\right), \tag{21}$$

where $f_{\chi^2}(\cdot|m-1)$ is the pdf of the chi-squared distribution with $m-1$ degrees of freedom.

The expression for the CASN of the SPRT chart with estimated process parameters can be determined in a similar fashion as in Equation (7), i.e.,

$$CASN = 1 + \hat{\mathbf{B}}^{\mathrm{T}}(\mathbf{I} - \hat{\mathbf{R}})^{-1}\mathbf{1}. \tag{22}$$

Here, $\hat{\mathbf{B}}$ is the $\eta \times 1$ vector whose entry $\hat{b}_k$ is revised as

$$\hat{b}_k = \Phi\left[V(g + \Delta\cdot k + \gamma) + \frac{W}{\sqrt{m}} - \delta\right] - \Phi\left[V(g + \Delta\cdot(k-1) + \gamma) + \frac{W}{\sqrt{m}} - \delta\right]. \tag{23}$$

The in-control CATS ($CATS_0$), steady-state out-of-control CATS ($CATS_\delta$), in-control CSDTS ($CSDTS_0$), and steady-state out-of-control CSDTS ($CSDTS_\delta$) can then be obtained as

$$CATS_0 = \frac{d}{1 - \widehat{OC}(0)}, \tag{24}$$

$$CATS_\delta = d \times \left(\frac{1}{1 - \widehat{OC}(\delta)} - \frac{1}{2}\right), \tag{25}$$

$$CSDTS_0 = d\sqrt{\frac{1 + \widehat{OC}(0)}{[1 - \widehat{OC}(0)]^2} - \left(\frac{1}{1 - \widehat{OC}(0)}\right)^2}, \tag{26}$$

and

$$CSDTS_\delta = d\sqrt{\frac{1 + \widehat{OC}^2(\delta)}{[1 - \widehat{OC}(\delta)]^2} - \frac{2}{3} - \left(\frac{1}{1 - \widehat{OC}(\delta)} - \frac{1}{2}\right)^2}, \tag{27}$$

respectively. The $CATS_0$ and $CATS_\delta$ are deduced using a similar method as in Equations (11) and (12), respectively. The detailed derivations of $CSDTS_0$ and $CSDTS_\delta$ are available in Appendix B. The $\widehat{OC}(\delta)$ in Equations (24) - (27) can be modified from Equation (10) as

$$\widehat{\mathrm{OC}}(\delta) = \widehat{P}_0 + \widehat{\mathbf{B}}^{\mathrm{T}}(\mathbf{I} - \widehat{\mathbf{R}})^{-1}\widehat{\mathbf{Q}}, \tag{28}$$

where $\widehat{\mathbf{Q}}$ is the $\eta \times 1$ vector whose entry $\widehat{q}_k$ is given as

$$\widehat{q}_k = \Phi\left[V(\Delta\cdot(0.5-k)+\gamma) + \frac{W}{\sqrt{m}} - \delta\right], \tag{29}$$

and $\widehat{P}_0$ can be written as

$$\widehat{P}_0 = \Phi\left[V(g+\gamma) + \frac{W}{\sqrt{m}} - \delta\right]. \tag{30}$$

The $\widehat{\mathrm{OC}}(0)$ in Equations (24) and (26) can be computed from Equation (28) by substituting $\delta = 0$. Note that all derivations of $\widehat{b}_k$, $\widehat{q}_k$, and $\widehat{P}_0$ are provided in Appendix A.

It is important to note that the CASN, CATS and CSDTS are evaluated as functions of the random variables $\widehat{\mu}_0$ and $\widehat{\sigma}_0$, or equivalently, the random variables W and V. In this respect, it is possible to derive the unconditional measures of these properties to facilitate better understanding of the chart's performances across practitioners. The unconditional measures are computed by averaging all the possible values of parameter estimates, i.e., the values of V and W. A common measure used to represent the CASN of the SPRT chart with estimated process parameters is the unconditional AASN. The expression for AASN is given as

$$\mathrm{AASN} = \mathrm{E}[\mathrm{CASN}] = \int_{-\infty}^{\infty}\int_{0}^{\infty} \mathrm{CASN} f_V(v) f_W(w)\ dvdw, \tag{31}$$

where $f_V(\cdot)$ and $f_W(\cdot)$ are the pdfs of the pivotal quantities V and W, respectively, and CASN can be quoted directly from Equation (22).

There are two commonly used measures for the CATS of the control charts with estimated parameters, i.e., the AATS and SDATS. The expression for the unconditional AATS is

$$\mathrm{AATS} = \mathrm{E}[\mathrm{CATS}] = \int_{-\infty}^{\infty}\int_{0}^{\infty} \mathrm{CATS} f_V(v) f_W(w)\ dvdw. \tag{32}$$

The in-control AATS (AATS_0) and steady-state out-of-control AATS (AATS_δ) can be computed from Equation (32), by using CATS_0 and CATS_δ in Equations (24) and (25), respectively. The expression for the unconditional SDATS is given as

$$\begin{aligned}\mathrm{SDATS} &= \sqrt{\mathrm{E}\left[\mathrm{CATS}^2\right] - (\mathrm{E}[\mathrm{CATS}])^2} \\ &= \sqrt{\int_{-\infty}^{\infty}\int_{0}^{\infty} \mathrm{CATS}^2 f_V(v) f_W(w) dvdw - \mathrm{AATS}^2}.\end{aligned} \tag{33}$$

Note that the AATS provides an average value of all the CATSs realized by practitioners, whereas the SDATS captures the practitioner-to-practitioner variability in the CATS performance.

Another meaningful performance metric is the unconditional ASDTS. The formulae for both in-control ASDTS (ASDTS_0) and steady-state out-of-control ASDTS (ASDTS_δ) are presented as

$$\mathrm{ASDTS}_0 = d\sqrt{\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1+\widehat{\mathrm{OC}}(0)}{[1-\widehat{\mathrm{OC}}(0)]^2} f_V(v) f_W(w) dvdw - \left(\frac{\mathrm{AATS}_0}{d}\right)^2} \tag{34}$$

and

$$\mathrm{ASDTS}_\delta = d\sqrt{\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{1+\widehat{\mathrm{OC}}^2(\delta)}{[1-\widehat{\mathrm{OC}}(\delta)]^2} f_V(v) f_W(w) dvdw - \frac{2}{3} - \left(\frac{\mathrm{AATS}_\delta}{d}\right)^2}, \tag{35}$$

where AATS_0 and AATS_δ can be computed from Equation (32). Detailed derivations of the formulae for ASDTS can be found in Appendix B. Similarly, the unconditional steady-state AAEQL becomes

$$\mathrm{AAEQL} = \frac{1}{\delta_{\max} - \delta_{\min}} \int_{-\infty}^{\infty}\int_{0}^{\infty}\int_{\delta_{\min}}^{\delta_{\max}} \delta^2 \mathrm{CATS}_\delta f_V(v) f_W(w)\ d\delta dvdw. \tag{36}$$

The integrations in Equations (31) - (36) are approximated with the Gauss-Legendre Quadrature method. One key difference between the SDATS and ASDTS is their functionalities. The SDATS measures the variability of the CATS values obtained across different practitioners, whereas the ASDTS measures the average variability of the time to signal acquired by the same practitioner over an infinite time horizon. Both measures deliver different information to a practitioner, and they can be used simultaneously in performance evaluations to aid multi-factor decision making.

3. Effects of parameter estimation of the SPRT chart under traditional control limits

In this section, we study the effects of parameter estimation on the performances of the SPRT chart with optimal charting parameters corresponding to the case of known process parameters. Note that the main goals of this article include investigating the impact of Phase-I parameter estimation on the performance of the SPRT chart and developing the optimal AEQL-based SPRT chart under the GICP framework when the process parameters are estimated. This article should, by no means, be interpreted as providing an analysis and comparison of the performances of various control charts with estimated process parameters in literature (see, for example, Saleh et al., 2015a; Teoh et al, 2019; Jardim et al., 2020). However, to add a value to this article, we present a comparison between the performances of the SPRT and CUSUM charts with estimated process parameters in Section 4.2, as both control charts possess very similar architectures.

Generally, an effective control chart serves two main objectives in statistical process monitoring (Montgomery, 2009):

- Emit an out-of-control signal as soon as a change in the process mean takes place (i.e., $\delta > 0$), and at the same time,
- Refrain from emitting excessive false signals when the process is in-control (i.e., $\delta = 0$).

The common practice for designing control charts, assuming that the process parameters are known, is to fix a suitably large ATS_0 before tuning the charting parameters to achieve an optimum out-of-control chart's performance. Since process shifts tend to occur sparsely and randomly over time, it is extremely difficult to predict the size of a specific mean shift that may occur in the future. To address the lack of knowledge of shift size, this article proposes an optimization procedure that considers the out-of-control performances of the control chart over a range of mean shifts. See Ryu et al. (2010) and Celano et al. (2013) for discussions of the practical considerations of an unknown shift size in control chart's applications.

The optimal SPRT chart with known process parameters is developed by minimizing the AEQL over a chosen range of mean shifts, $\delta \in [\delta_{\min}, \delta_{\max}]$ (Ou et al., 2011b). In the optimization design, Ou et al. (2011b) advised setting $\delta_{\min}$ as a small value to prevent the introduction of additional variability into the process. Also, they suggested setting $\delta_{\max}$ as the maximum possible mean shift based on the practitioners' knowledge of the underlying process. To achieve the best overall results, we choose $\delta_{\min} = 0.1$ and $\delta_{\max} = 2.0$ in our optimization model throughout this article.

The SPRT chart has five charting parameters, i.e., the in-control ASN (ASN_0), γ, d, g, and h. Throughout the optimization design, all five charting parameters need to be adjusted to yield a minimum AEQL, while conforming to the user-defined specifications, i.e., the recommended value of ATS_0 (τ), the desired inspection rate ($R = \mathrm{ASN}_0/d$), and the minimum allowable sampling interval ($d_{\min}$). The inspection rate R imposes a natural restriction on the relationship between the sample

Table 1
The values of AATS, ASDTS, SDATS, $\Pr(CATS_0 \geq 370.40)$, and $\Pr(CATS_0 \geq 296.32)$ of the SPRT chart, together with the optimal charting parameters $(ASN_0, \gamma, d, g, h) = (2.132, 0.306, 0.426, 0.317, 8.388)$ obtained from minimizing the AEQL corresponding to the case of known process parameters, when $ATS_0 = \tau = 370.40$, $d_{min} = 0.25$, $\delta_{min} = 0.1$, $\delta_{max} = 2.0$, $R = 5$, and $m \in \{50, 100, 200, 400, 600, 1000, 2000, +\infty\}$.

	$m = 50$			$m = 100$		
		$\Pr(CATS_0 \geq 370.40)$	$\Pr(CATS_0 \geq 296.32)$		$\Pr(CATS_0 \geq 370.40)$	$\Pr(CATS_0 \geq 296.32)$
δ	AATS	ASDTS	SDATS	AATS	ASDTS	SDATS
		48.35%	52.18%		48.85%	54.15%
0.0	> 20,000	> 200,000	> 100,000	2,061.40	> 20,000	> 10,000
0.2	525.23	> 20,000	> 10,000	77.88	660.10	463.50
0.4	17.57	1,728.41	1,222.11	6.14	20.24	13.64
0.6	2.12	21.88	15.40	1.59	2.07	0.95
0.8	0.90	1.20	0.57	0.84	0.88	0.22
1.0	0.59	0.61	0.16	0.57	0.57	0.10
1.5	0.34	0.29	0.04	0.33	0.29	0.03
2.0	0.26	0.19	0.01	0.26	0.19	0.01
	$m = 200$			$m = 400$		
		$\Pr(CATS_0 \geq 370.40)$	$\Pr(CATS_0 \geq 296.32)$		$\Pr(CATS_0 \geq 370.40)$	$\Pr(CATS_0 \geq 296.32)$
δ	AATS	ASDTS	SDATS	AATS	ASDTS	SDATS
		49.20%	56.69%		49.37%	60.01%
0.0	809.10	2,746.86	1,856.20	538.29	997.19	593.66
0.2	40.20	97.43	62.76	30.49	46.20	24.54
0.4	4.55	6.61	3.40	4.05	4.71	1.70
0.6	1.47	1.59	0.45	1.42	1.46	0.28
0.8	0.82	0.83	0.14	0.81	0.80	0.09
1.0	0.57	0.55	0.07	0.56	0.54	0.05
1.5	0.33	0.28	0.02	0.33	0.28	0.01
2.0	0.26	0.19	0.01	0.25	0.19	0.00
	$m = 600$			$m = 1000$		
		$\Pr(CATS_0 \geq 370.40)$	$\Pr(CATS_0 \geq 296.32)$		$\Pr(CATS_0 \geq 370.40)$	$\Pr(CATS_0 \geq 296.32)$
δ	AATS	ASDTS	SDATS	AATS	ASDTS	SDATS
		49.50%	62.41%		49.59%	66.12%
0.0	473.56	720.82	384.40	428.54	555.21	249.79
0.2	28.00	36.79	16.87	26.22	30.84	11.48
0.4	3.91	4.29	1.26	3.81	4.01	0.91
0.6	1.40	1.43	0.22	1.39	1.40	0.16
0.8	0.81	0.80	0.08	0.80	0.79	0.06
1.0	0.56	0.54	0.04	0.56	0.54	0.03
1.5	0.33	0.28	0.01	0.33	0.28	0.01
2.0	0.25	0.19	0.00	0.25	0.19	0.00
	$m = 2000$			$m = +\infty$		
		$\Pr(CATS_0 \geq 370.40)$	$\Pr(CATS_0 \geq 296.32)$		$\Pr(CATS_0 \geq 370.40)$	$\Pr(CATS_0 \geq 296.32)$
δ	AATS	ASDTS	SDATS		ATS	SDTS
		49.70%	72.39%		100.00%	100.00%
0.0	398.19	454.80	155.65		370.46	370.24
0.2	24.99	27.09	7.40		23.85	23.85
0.4	3.73	3.83	0.61		3.66	3.66
0.6	1.38	1.38	0.11		1.38	1.37
0.8	0.80	0.79	0.04		0.80	0.78
1.0	0.56	0.53	0.02		0.56	0.53
1.5	0.33	0.28	0.01		0.33	0.28
2.0	0.25	0.19	0.00		0.25	0.19

number and the sampling interval. In other words, with a fixed R, it is impossible to increase ASN_0 without having to decrease d, and vice versa. See Ou et al. (2011b) for the optimization algorithm of the SPRT chart with known process parameters.

The optimal SPRT chart with known process parameters used throughout this section is constructed upon the following specifications:

$$\tau = 370.40,\ R = 5,\ d_{min} = 0.25,\ \delta_{min} = 0.1,\ \delta_{max} = 2.0. \tag{37}$$

$\tau = 370.40$ is chosen to yield a false alarm rate of 0.27%. The inspection rate R is set as 5 measurements per unit of sampling interval, and d_{min} is a quarter of the standard time unit adopted in an operation. By implementing the optimization algorithm in Ou et al. (2011b) and the formulae in Section 2.1, we obtain the following design parameters for the optimal SPRT chart with known process parameters:

$$ASN_0 = 2.132,\ \gamma = 0.306,\ d = 0.426,\ g = 0.317,\ h = 8.388,$$

which yield a minimum AEQL of 0.694. These charting parameters are used to evaluate the performances of the SPRT chart with known and estimated process parameters, which are tabulated in Table 1. Note that no further adjustments have been made to the control limits of the Phase-II control chart, hence this method is known as the traditional approach.

Table 1 displays two critical information about the SPRT chart with known and estimated process parameters, i.e., the conditional in-control performances ($\Pr(CATS_0 \geq 370.40)$ and $\Pr(CATS_0 \geq 296.32)$), as well as the unconditional in-control and out-of-control performances (AATS, ASDTS, and SDATS). Here, $\Pr(CATS_0 \geq 370.40)$ is known as the exceedance probability. It signifies the proportion of $CATS_0$ exceeding the recommended level $\tau = 370.40$. On the other hand, $\Pr(CATS_0 \geq 296.32)$ is computed to provide an inflated value of the exceedance probability, where a 20% tolerance is applied to τ, i.e., $370.40 \times 0.8 = 296.32$. Note

that these exceedance probabilities can be well approximated by constructing empirical distributions of the $CATS_0$ with 100,000 simulated values. In Table 1, the Phase-I sample sizes $m \in \{50, 100, 200, 400, 600, 1000, 2000, +\infty\}$ are given in the leading rows of each sub table, and the mean shifts $\delta \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0\}$ are given in the leftmost column. Note that $m = +\infty$ corresponds to the case of known process parameters.

As a numeric example, when $\delta = 1.0$, the charting parameters $(\gamma, d, g, h) = (0.306, 0.426, 0.317, 8.388)$ are used to compute the ATS_δ $(= 0.56)$ and $SDTS_\delta$ $(= 0.53)$ of the SPRT chart with known process parameters ($m = +\infty$). In the case where the process parameters are estimated, specifically when $m = 1000$, the $(AATS_\delta, ASDTS_\delta, SDATS_\delta)$ values are computed as (0.56, 0.54, 0.03) when $\delta = 1.0$ using the same combination of charting parameters. All the values of AATS, ASDTS, and SDATS are computed using Equations (32) to (35). The accuracy of all results in this article has been verified using Monte Carlo simulation with 100,000 runs. Besides, using the same set of charting parameters, Table 1 gives the exceedance probabilities $Pr(CATS_0 \geq 370.40)$ and $Pr(CATS_0 \geq 296.32)$ as 49.59% and 66.12%, respectively, when $m = 1000$. When $m = +\infty$ (i.e., process parameters are known), the ATS is a constant value, hence the SDATS is equal to zero. Both the exceedance probabilities ($Pr(CATS_0 \geq 370.40)$ and $Pr(CATS_0 \geq 296.32)$) are equal to 100% since $CATS_0 = \tau$.

From Table 1, it is observed that the $Pr(CATS_0 \geq 370.40)$ remains below 50% for all the cases of estimated process parameters. This indicates that more than half of the practitioners are expected to suffer higher false alarm rates than the recommended level (0.27%). When the tolerance is increased by 20%, $Pr(CATS_0 \geq 296.32)$ is still relatively low. For instance, when $m = 600$, Table 1 shows that $Pr(CATS_0 \geq 296.32) = 62.41\%$, and this probability increases to 72.39% when m is raised to 2000. These probabilities imply that at least 37% and 27% of the attained $CATS_0$ values are still lower than the discounted τ, even when $m = 600$ and a large $m = 2000$ are used, respectively. From the unconditional perspective, the AATSs and ASDTSs obtained in cases of estimated process parameters are vastly different from the ATS and SDTS obtained in the case of known process parameters, especially when m and δ are small. As a general trend, the AATS and ASDTS values converge to the ATS and SDTS values as m increases. One might have noticed that, for $m = 50$ and 100, there exists considerable volatility in the values of the CATS when δ is either small or equal to zero. For example, the SDATS value for $m = 50$ is higher than 10,000 when $\delta = 0.2$, and it exceeds 100,000 when the process is IC. These strikingly large values of SDATS indicate that, in practice, the CATSs are often very far from the AATSs when the available Phase-I sample size is small.

From the unconditional perspective, Table 1 shows that the SPRT chart presents relatively good performance for a considerably large m. However, from the conditional perspective, the in-control performance is highly unsatisfactory due to large dropouts of $CATS_0$ from the recommended region. Table 1 shows that the SPRT chart with unadjusted limits produce unpredictable results, especially when m and δ are small. Also, the traditional SPRT chart tends to generate false alarms with a high frequency. With excessive false alarms, it is very likely that businesses will suffer from enormous time and economic costs, due to unnecessary corrective actions and redirection of resources.

4. Optimization framework of the SPRT chart with estimated process parameters

In view of the undesirable in-control performance of the traditional SPRT chart, this section discusses two other approaches of designing the optimal SPRT chart when the process parameters are estimated, i.e., (1) the AATS-matching approach, and (2) the GICP approach. The control limits of the SPRT chart determined in Section 4.1 are denoted as adjusted limits 1, whereas those determined in Section 4.2 are identified as adjusted limits 2.

4.1. Optimal design of the SPRT chart with estimated process parameters using the AATS-matching approach

In this section, we outline the procedure to construct the optimal SPRT chart with estimated process parameters using the AATS-matching approach. This approach has been used broadly in the design of control charts with estimated process parameters (see, for example, Chakraborti, 2006; Goedhart et al., 2016; Teoh et al., 2019) before the conditional perspective gains prevalence. The AATS-matching method is implemented simply by equating $AATS_0$ to τ. Note that any charting parameter and specification that are random variables of the Phase-I data must be represented by their unconditional measures. In particular, the optimization model of the SPRT chart with estimated process parameters is implemented as follows:

$$\underset{(AASN_0, \gamma, d, g, h)}{\text{Minimize}} \ AAEQL, \tag{38}$$

subject to constraints

$$AATS_0 = \tau, \tag{39}$$

$$R = \frac{AASN_0}{d}, \tag{40}$$

$$AASN_0 > 1, \tag{41}$$

and

$$d > d_{\min}, \tag{42}$$

where $AASN_0$ is the in-control AASN. Generally, these specifications are set by practitioners in advance by considering current work practice and employment capacity. The optimization algorithm of the SPRT chart with estimated process parameters is outlined as follows:

Step 1: Specify a fixed set of parameters: m, τ, R, $d_{\min}$, $\delta_{\min}$, and $\delta_{\max}$.

Step 2: Perform an exhaustive grid search over the two-dimensional parameter space $(\bar{n}, \gamma)$ to locate all the possible combinations of $(\bar{n}, \gamma)$. Here, $\bar{n}$ is the value of $AASN_0$ identified in the grid search. To satisfy constraints (41) and (42), $\bar{n}$ is determined in the range of $(\max [1, d_{\min} \times R], +\infty)$, whereas γ is determined in the range of $(0, +\infty)$. Then, calculate the value of $d = \bar{n} / R$ to attain constraint (40).

Step 3: For each pair of the $(\bar{n}, \gamma)$ values obtained in Step 2, adjust the values of g and h simultaneously using the Newton-Raphson algorithm to fulfill both specifications on $AATS_0 = \tau$ (constraint (39)) and $AASN_0 = \bar{n}$ (constraint (40)). Note that $AATS_0$ and $AASN_0$ are computed from Equations (32) and (31), respectively. At the end of this step, all the possible combinations of $(AASN_0, \gamma, d, g, h)$ satisfying constraints (39) - (42) are acquired when $\delta = 0$.

Step 4: For a specific range of mean shifts $[\delta_{\min}, \delta_{\max}]$, determine the optimal charting parameters $(AASN_0, \gamma, d, g, h)$ that yield the smallest steady-state AAEQL from all the combinations of charting parameters found in Steps 2 and 3. Note that the AAEQL is computed from Equation (36) using the Gauss-Legendre Quadrature procedure.

Table 2 presents the values of AATS, ASDTS, SDATS, $Pr(CATS_0 \geq 370.40)$, and $Pr(CATS_0 \geq 296.32)$ for the optimal SPRT chart with estimated process parameters. As opposed to the traditional approach in Table 1, a customized optimal SPRT chart is designed for each $m \in \{50, 100, 200, 400, 600, 1000, 2000\}$ under the AATS-matching approach in Table 2. The first row of each sub table showcases the set of optimal charting parameters $(AASN_0, \gamma, d, g, h)$ and optimal AAEQL obtained through the optimization model in (38) - (42) and the formulae from Section 2.2. The specifications for the optimal SPRT chart with estimated process parameters in Table 2 are the same as those in (37). As a numeric example, when $m = 100$, the optimal charting parameters $(AASN_0, \gamma, d, g, h)$ are determined as (2.241, 0.289, 0.448, 0.324, 6.896), while the corresponding optimal AAEQL is given as 0.763 (see Table 2). These optimal charting parameters are then used to evaluate the

Table 2
The values of AATS, ASDTS, SDATS, $Pr(CATS_0 \geq 370.40)$, and $Pr(CATS_0 \geq 296.32)$ of the SPRT chart with estimated process parameters, together with the optimal charting parameters ($AASN_0$, γ, d, g, h) and optimal AAEQL values, when $AATS_0 = \tau = 370.40$, $d_{min} = 0.25$, $\delta_{min} = 0.1$, $\delta_{max} = 2.0$, $R = 5$, and $m \in \{50, 100, 200, 400, 600, 1000, 2000, +\infty\}$.

	$m = 50$			$m = 100$		
	($AASN_0$, γ, d, g, h, AAEQL)			($AASN_0$, γ, d, g, h, AAEQL)		
		$Pr(CATS_0 \geq 370.40)$	$Pr(CATS_0 \geq 296.32)$		$Pr(CATS_0 \geq 370.40)$	$Pr(CATS_0 \geq 296.32)$
δ	AATS	ASDTS	SDATS	AATS	ASDTS	SDATS
	(2.289, 0.250, 0.458, 0.388, 6.207, 0.805)			(2.241, 0.289, 0.448, 0.324, 6.896, 0.763)		
		14.01%	16.89%		21.02%	25.98%
0.0	369.41	> 10,000	7,906.74	370.50	2,156.02	1,501.89
0.2	32.17	433.49	305.68	29.77	108.47	73.75
0.4	4.97	23.31	16.11	4.52	8.66	5.23
0.6	1.64	2.70	1.52	1.54	1.84	0.72
0.8	0.90	1.01	0.35	0.86	0.89	0.21
1.0	0.62	0.63	0.15	0.59	0.58	0.10
1.5	0.36	0.31	0.04	0.35	0.30	0.03
2.0	0.28	0.21	0.02	0.27	0.20	0.01
	$m = 200$			$m = 400$		
	($AASN_0$, γ, d, g, h, AAEQL)			($AASN_0$, γ, d, g, h, AAEQL)		
		$Pr(CATS_0 \geq 370.40)$	$Pr(CATS_0 \geq 296.32)$		$Pr(CATS_0 \geq 370.40)$	$Pr(CATS_0 \geq 296.32)$
δ	AATS	ASDTS	SDATS	AATS	ASDTS	SDATS
	(2.211, 0.303, 0.442, 0.296, 7.446, 0.733)			(2.180, 0.303, 0.436, 0.308, 7.899, 0.715)		
		27.75%	35.25%		33.25%	44.02%
0.0	370.32	975.21	637.99	370.23	639.99	369.25
0.2	27.48	54.29	33.11	25.60	36.79	18.69
0.4	4.15	5.56	2.62	3.90	4.45	1.53
0.6	1.46	1.57	0.43	1.42	1.46	0.27
0.8	0.83	0.83	0.14	0.82	0.81	0.09
1.0	0.58	0.55	0.07	0.57	0.55	0.05
1.5	0.34	0.29	0.02	0.34	0.29	0.01
2.0	0.26	0.19	0.01	0.26	0.19	0.00
	$m = 600$			$m = 1000$		
	($AASN_0$, γ, d, g, h, AAEQL)			($AASN_0$, γ, d, g, h, AAEQL)		
		$Pr(CATS_0 \geq 370.40)$	$Pr(CATS_0 \geq 296.32)$		$Pr(CATS_0 \geq 370.40)$	$Pr(CATS_0 \geq 296.32)$
δ	AATS	ASDTS	SDATS	AATS	ASDTS	SDATS
	(2.170, 0.304, 0.434, 0.308, 8.050, 0.708)			(2.180, 0.304, 0.436, 0.299, 8.199, 0.703)		
		36.04%	49.34%		38.99%	56.03%
0.0	369.83	544.42	282.65	370.28	473.86	209.28
0.2	25.01	32.07	14.19	24.40	28.45	10.34
0.4	3.82	4.16	1.18	3.73	3.92	0.87
0.6	1.40	1.43	0.22	1.39	1.39	0.16
0.8	0.81	0.80	0.08	0.81	0.79	0.06
1.0	0.57	0.54	0.04	0.56	0.54	0.03
1.5	0.33	0.28	0.01	0.33	0.28	0.01
2.0	0.26	0.19	0.00	0.26	0.19	0.00
	$m = 2000$			$m = +\infty$		
	($AASN_0$, γ, d, g, h, AAEQL)			(ASN_0, γ, d, g, h, AEQL)		
		$Pr(CATS_0 \geq 370.40)$	$Pr(CATS_0 \geq 296.32)$		$Pr(CATS_0 \geq 370.40)$	$Pr(CATS_0 \geq 296.32)$
δ	AATS	ASDTS	SDATS		ATS	SDTS
	(2.179, 0.305, 0.436, 0.294, 8.292, 0.699)			(2.132, 0.306, 0.426, 0.317, 8.388, 0.694)		
		42.04%	65.69%		100.00%	100.00%
0.0	369.88	421.08	142.58		370.46	370.24
0.2	24.04	26.01	7.01		23.85	23.85
0.4	3.68	3.77	0.59		3.66	3.66
0.6	1.38	1.37	0.11		1.38	1.37
0.8	0.80	0.78	0.04		0.80	0.78
1.0	0.56	0.53	0.02		0.56	0.53
1.5	0.33	0.28	0.01		0.33	0.28
2.0	0.26	0.19	0.00		0.25	0.19

unconditional performance measures for each δ when $m = 100$. Note that with $m = 100$, the relatively small $AASN_0$ (=2.241) highlights the main advantage of the optimal SPRT chart in terms of the reduced number of observations compared to other optimal control charts whose subgroups are moderate and large. To facilitate comparison, the results for the case of known process parameters ($m = +\infty$) are also included in Table 2.

From Table 2, it is noticed that the unconditional metrics (AATS, ASDTS, SDATS) have improved substantially compared with those shown in Table 1, especially when m and δ are small. The values of AATS for each shift size are quite stable across the values of m. All the $AATS_\delta$ values expectedly approach the ATS_δ values in the case of known process parameters as m increases. In addition, both the within-practitioner (ASDTS) and between-practitioner (SDATS) variabilities decrease compared to those in Table 1 due to tighter control limits in Table 2. For example, when $m = 100$, the $ASDTS_0$ and $SDATS_0$ in Table 2 are equal to 2,156.02 and 1,501.89, respectively. Both variabilities are significantly lower than those shown in Table 1. In fact, these differences are

Table 3
AATS, ASDTS, and SDATS values of the SPRT chart with estimated process parameters, together with the adjusted chart's control limits (g, h) under the EPC constraint, when $p = 0.05$, $\varepsilon = 0$, $\tau = 370.40$, $AASN_0 = 2.5$, $\gamma = 0.3$, $R = 5$, $d = 0.5$, and $m \in \{50, 100, 200, 400, 600, 1000, 2000\}$.

		$\delta = 0.0$	$\delta = 0.1$	$\delta = 0.2$	$\delta = 0.3$	$\delta = 0.5$	$\delta = 0.7$	$\delta = 1.0$	$\delta = 1.5$	$\delta = 2.0$
m	(g, h)	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS
50	(0.577, 29.695)	> 10,000,000	> 5,000,000	> 1,000,000	> 500,000	> 100,000	> 10,000	0.89	0.47	0.33
		> 200,000,000	> 100,000,000	> 20,000,000	> 10,000,000	> 2,000,000	> 200,000	0.96	0.43	0.27
		> 100,000,000	> 50,000,000	> 10,000,000	> 5,000,000	> 1,000,000	> 100,000	0.27	0.07	0.03
100	(0.299, 17.450)	> 1,000,000	> 500,000	> 100,000	4,408.78	6.36	1.26	0.66	0.39	0.30
		> 20,000,000	> 10,000,000	> 2,000,000	> 200,000	1,386.21	1.56	0.64	0.33	0.22
		> 10,000,000	> 5,000,000	> 1,000,000	> 100,000	980.19	0.67	0.11	0.03	0.01
200	(0.211, 13.498)	> 60,000	4,536.58	343.02	34.69	2.53	1.09	0.60	0.37	0.29
		> 1,000,000	> 60,000	3,922.32	242.92	3.50	1.12	0.57	0.30	0.21
		> 600,000	> 45,000	2,762.88	170.01	1.72	0.23	0.07	0.02	0.01
400	(0.174, 11.609)	5,072.37	570.99	76.54	14.37	2.20	1.03	0.58	0.36	0.29
		> 14,000	1,539.21	172.40	24.54	2.38	1.03	0.55	0.29	0.20
		> 10,000	1,010.73	109.23	14.07	0.67	0.14	0.04	0.01	0.00
600	(0.161, 10.885)	2,363.06	319.44	51.78	11.63	2.11	1.02	0.57	0.36	0.29
		4,672.44	578.43	82.53	15.51	2.21	1.01	0.54	0.29	0.20
		2,850.34	340.99	45.44	7.26	0.48	0.11	0.03	0.01	0.00
1000	(0.151, 10.263)	1,317.70	206.84	38.99	10.01	2.04	1.00	0.57	0.35	0.28
		1,930.80	286.41	49.83	11.57	2.09	0.99	0.53	0.29	0.20
		998.08	140.08	21.94	4.11	0.34	0.08	0.03	0.01	0.00
2000	(0.143, 9.712)	830.70	147.77	31.48	8.96	2.00	0.99	0.56	0.35	0.28
		993.27	171.52	35.10	9.54	2.01	0.97	0.53	0.29	0.20
		385.31	61.57	10.97	2.32	0.23	0.06	0.02	0.01	0.00

especially pronounced when m and δ are both small. Note also that the values of AAEQL approach the value of AEQL in the case of known process parameters as m increases.

On the other hand, the conditional in-control performances in all cases of estimated process parameters have worsen compared to those shown in Table 1. The values of $\Pr(CATS_0 \geq 370.40)$ and $\Pr(CATS_0 \geq 296.32)$ tabulated in Table 2 are lower than those in Table 1 for all values of m. The worst performance is reported when m is the smallest. In particular, when $m = 50$, $\Pr(CATS_0 \geq 370.40)$ falls from a modest 48.35% (see Table 1) to an alarming 14.01% (see Table 2). In fact, the deterioration of in-control performance is due to the progressively low upper control limit h associated with a decrease in m. As m decreases, h is

Table 4
AATS, ASDTS, and SDATS values of the SPRT chart with estimated process parameters, together with the optimal charting parameters ($AASN_0$, γ, d, g, h) and optimal AAEQL values under the EPC constraint, when $p = 0.05$, $\varepsilon = \{0.0, 0.2\}$, $\tau = 370.40$, $d_{min} = 0.25$, $\delta_{min} = 0.1$, $\delta_{max} = 2.0$, $R = 5$, and $m \in \{200, 400, 600, 1000, 2000\}$.

			$\delta = 0.0$	$\delta = 0.2$	$\delta = 0.4$	$\delta = 0.6$	$\delta = 0.8$	$\delta = 1.0$	$\delta = 1.5$	$\delta = 2.0$
m	ε	($AASN_0$, γ, d, g, h, AAEQL)	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS	AATS ASDTS SDATS
200	0.0	(2.250, 0.420, 0.450,	> 10,000	309.66	12.91	1.90	0.83	0.54	0.31	0.25
		–0.034, 9.283, 2.111)	> 75,000	1,362.56	33.33	2.41	0.85	0.51	0.26	0.17
			> 50,000	938.27	21.72	1.06	0.18	0.07	0.02	0.00
	0.2	(2.250, 0.420, 0.450,	8,484.38	243.13	11.82	1.88	0.83	0.54	0.31	0.25
		–0.035, 8.920, 1.784)	> 45,000	954.38	28.05	2.35	0.84	0.51	0.26	0.17
			> 30,000	652.58	17.99	1.01	0.18	0.07	0.02	0.00
400	0.0	(2.251, 0.275, 0.450,	5,914.80	70.30	4.27	1.47	0.87	0.61	0.36	0.27
		0.395, 12.519, 0.983)	> 15,000	166.94	5.24	1.51	0.86	0.59	0.31	0.20
			> 10,000	107.06	2.15	0.27	0.10	0.05	0.01	0.01
	0.2	(2.251, 0.280, 0.450,	3,913.96	61.77	4.27	1.47	0.86	0.60	0.36	0.27
		0.375, 11.780, 0.928)	> 10,000	133.91	5.21	1.50	0.85	0.58	0.30	0.20
			7,881.59	84.01	2.11	0.27	0.10	0.05	0.01	0.01
600	0.0	(2.250, 0.275, 0.450,	2,634.73	47.63	3.95	1.44	0.85	0.60	0.36	0.27
		0.382, 11.721, 0.852)	5,538.31	77.53	4.40	1.46	0.84	0.58	0.30	0.20
			3,444.72	43.26	1.38	0.21	0.08	0.04	0.01	0.00
	0.2	(2.250, 0.280, 0.450,	1,833.65	42.83	3.95	1.43	0.84	0.60	0.35	0.27
		0.362, 11.015, 0.821)	3,578.25	66.27	4.39	1.45	0.83	0.57	0.30	0.20
			2,172.83	35.75	1.37	0.21	0.08	0.04	0.01	0.00
1000	0.0	(2.250, 0.279, 0.450,	1,387.01	36.50	3.76	1.41	0.84	0.59	0.35	0.27
		0.357, 10.876, 0.785)	2,084.03	46.87	3.98	1.41	0.82	0.56	0.30	0.20
			1,100.01	20.78	0.93	0.16	0.06	0.03	0.01	0.00
	0.2	(2.250, 0.275, 0.450,	1,054.39	32.30	3.68	1.40	0.84	0.59	0.35	0.27
		0.368, 10.532, 0.765)	1,541.39	40.60	3.88	1.41	0.82	0.57	0.30	0.20
			795.19	17.39	0.88	0.15	0.06	0.03	0.01	0.00
2000	0.0	(2.249, 0.275, 0.450,	858.80	29.04	3.57	1.39	0.83	0.59	0.35	0.27
		0.362, 10.379, 0.745)	1,040.87	32.41	3.65	1.38	0.81	0.56	0.30	0.20
			416.08	10.17	0.57	0.11	0.04	0.02	0.01	0.00
	0.2	(2.188, 0.290, 0.438,	641.58	28.30	3.70	1.39	0.82	0.58	0.34	0.26
		0.345, 9.495, 0.730)	757.13	31.20	3.79	1.39	0.80	0.55	0.29	0.19
			284.51	9.29	0.61	0.11	0.04	0.02	0.01	0.00

lowered to ensure that $AATS_0 = \tau$ (constraint (39)) is satisfied, hence provoking high frequency of unwanted false alarms even though the process is stable. This inevitable sacrifice of the in-control performance due to adjusted limits is often referred to as the trade-off between in-control and out-of-control performances.

4.2. Optimal design of the SPRT chart with estimated process parameters using the GICP approach

In the recent decade, an increasingly popular design framework, known as the GICP framework, has been used extensively by researchers. The main difference between the GICP and the AATS-matching approach is, the former ensures that the $CATS_0$ exceeds a pre-specified level (say τ) with a high probability $1 - p$, whereas the latter does not. The formulation of the GICP based on the EPC is given as

$$\Pr(CATS_0 \geq \tau) = 1 - p, \tag{43}$$

where p is a user-defined error probability.

Diko et al. (2019a, b) introduced a tolerance term $0 \leq \varepsilon < 1$, so that Equation (43) becomes

$$\Pr(CATS_0 \geq (1 - \varepsilon)\tau) = 1 - p. \tag{44}$$

The tolerance term ε is also known as the nominal percentage difference value. It serves the purpose of providing greater flexibility to the GICP design of control charts with estimated process parameters. By increasing the value of ε or p, it is possible to degrade the in-control performance slightly in contemplation of a better out-of-control performance, and vice versa.

Diko et al. (2019a, b) argued that the analytical solution for Equation (44) is difficult to obtain for control charts with correlated charting statistics, such as the EWMA and CUSUM charts. Therefore, they suggested constructing an empirical distribution for the $CATS_0$ or $CARL_0$ using the Markov chain approach, and then adjusting the charting parameters to satisfy constraint (44). Alternatively, Capizzi and Masarotto (2020) recommended a stochastic approximation algorithm to approximate the root of Equation (44) with simulated run lengths. In this article, we employ the algorithm recommended by Capizzi and Masarotto (2020). However, we replace the simulated run lengths by exact run lengths obtained using the Markov chain approach, i.e., from Equation (24).

Table 3 presents the results of the SPRT chart with estimated process parameters for $m \in \{50, 100, 200, 400, 600, 1000, 2000\}$, which is designed with the GICP approach based on the EPC formulated in Equation (44). The error probability p is set as 5% and the tolerance ε as 0%. The specifications τ and R are fixed at 370.40 and 5, respectively, whereas the design parameters ($AASN_0$, γ, d) are chosen as (2.5, 0.3, 0.5). The second column of Table 3 displays the adjusted control limits (g, h), which are computed using a stochastic approximation algorithm for each value of m. Meanwhile, the unconditional metrics (AATS, ASDTS, SDATS) are evaluated for $\delta \in \{0.0, 0.1, 0.2, 0.3, 0.5, 0.7, 1.0, 1.5, 2.0\}$ and are tabulated in subsequent columns of Table 3. As an example, for $m = 200$, the adjusted control limits (g, h) are (0.211, 13.498), whereas the corresponding values of ($AATS_\delta$, $ASDTS_\delta$, $SDATS_\delta$) are (0.60, 0.57, 0.07) when $\delta = 1.0$.

From Table 3, it is observed that the values of the upper control limit h are significantly higher compared to those in Tables 1 and 2, even though there is a trend of decreasing control limits as m increases. This is intuitive since the chart is designed to ensure that, 95% of the times, a practitioner would obtain a $CATS_0$ greater than $\tau = 370.40$ for a specific m. Nonetheless, an unsurprising drawback of such a design is the extremely huge and volatile CATS values obtained in cases where both m and δ are small. It is noticed that when $m = 50$, the values of $AATS_\delta$ for $\delta \leq 0.7$ are at least 10,000; whereas the $SDATS_\delta$ values exceed a striking value of 100,000. When m is increased from 50 to 100, the out-of-control

Table 5
AATS, ASDTS, and SDATS values of the SPRT chart with estimated process parameters, together with the optimal charting parameters ($AASN_0$, γ, d, g, h) and optimal AAEQL values under the EPC constraint, when $p = 0.10$, $\varepsilon = \{0.0, 0.2\}$, $\tau = 370.40$, $d_{min} = 0.25$, $\delta_{min} = 0.1$, $\delta_{max} = 2.0$, $R = 5$, and $m \in \{200, 400, 600, 1000, 2000\}$.

m	ε	($AASN_0$, γ, d, g, h, AAEQL)	$\delta = 0.0$ AATS ASDTS SDATS	$\delta = 0.2$ AATS ASDTS SDATS	$\delta = 0.4$ AATS ASDTS SDATS	$\delta = 0.6$ AATS ASDTS SDATS	$\delta = 0.8$ AATS ASDTS SDATS	$\delta = 1.0$ AATS ASDTS SDATS	$\delta = 1.5$ AATS ASDTS SDATS	$\delta = 2.0$ AATS ASDTS SDATS
200	0.0	(2.250, 0.380, 0.450,	7,169.16	180.04	8.86	1.70	0.82	0.55	0.32	0.26
		0.080, 9.430, 1.476)	> 38,000	725.78	19.63	2.00	0.83	0.53	0.27	0.18
			> 25,000	497.16	12.39	0.76	0.16	0.07	0.02	0.01
	0.2	(2.124, 0.350, 0.425,	5,773.75	138.32	7.38	1.66	0.85	0.57	0.33	0.25
		0.249, 9.753, 1.277)	> 30,000	550.74	15.04	1.88	0.86	0.55	0.28	0.18
			> 20,000	376.95	9.26	0.65	0.16	0.07	0.02	0.01
400	0.0	(2.250, 0.275, 0.450,	3,044.05	54.09	4.16	1.47	0.86	0.61	0.36	0.27
		0.391, 11.555, 0.890)	8,758.32	111.99	5.00	1.50	0.85	0.58	0.31	0.20
			5,807.03	69.34	1.96	0.27	0.10	0.05	0.01	0.01
	0.2	(2.250, 0.280, 0.450,	2,085.48	47.92	4.14	1.46	0.85	0.60	0.35	0.27
		0.370, 10.871, 0.851)	5,407.37	91.93	4.95	1.49	0.84	0.58	0.30	0.20
			3,527.84	55.48	1.91	0.27	0.10	0.05	0.01	0.01
600	0.0	(2.250, 0.260, 0.450,	1,780.31	37.56	3.74	1.44	0.87	0.62	0.36	0.28
		0.430, 11.538, 0.810)	3,592.42	58.25	4.10	1.46	0.86	0.59	0.31	0.20
			2,206.44	31.48	1.20	0.20	0.08	0.04	0.01	0.00
	0.2	(2.250, 0.280, 0.450,	1,204.54	36.42	3.87	1.42	0.84	0.59	0.35	0.27
		0.358, 10.362, 0.786)	2,187.60	53.42	4.27	1.44	0.83	0.57	0.30	0.20
			1,291.36	27.63	1.29	0.21	0.08	0.04	0.01	0.00
1000	0.0	(2.250, 0.279, 0.450,	1,042.82	32.84	3.71	1.40	0.83	0.59	0.35	0.27
		0.354, 10.408, 0.765)	1,516.08	41.23	3.92	1.41	0.82	0.56	0.30	0.20
			778.29	17.62	0.90	0.16	0.06	0.03	0.01	0.00
	0.2	(2.250, 0.282, 0.450,	781.90	29.91	3.68	1.39	0.83	0.59	0.35	0.27
		0.341, 9.860, 0.747)	1,097.47	36.74	3.88	1.40	0.81	0.56	0.30	0.20
			544.71	15.09	0.88	0.16	0.06	0.03	0.01	0.00
2000	0.0	(2.250, 0.275, 0.450,	711.97	27.14	3.53	1.38	0.83	0.59	0.35	0.27
		0.359, 10.056, 0.735)	852.64	30.06	3.61	1.38	0.81	0.56	0.30	0.20
			331.97	9.15	0.56	0.11	0.04	0.02	0.01	0.00
	0.2	(2.188, 0.290, 0.438,	540.10	26.44	3.65	1.39	0.82	0.58	0.34	0.26
		0.342, 9.213, 0.720)	631.08	28.98	3.74	1.38	0.80	0.55	0.29	0.19
			231.07	8.38	0.59	0.11	0.04	0.02	0.01	0.00

performances improve slightly for moderate δ (= 0.5). Even so, the chart still exhibits poor performance for relatively small values of δ ($\leq$ 0.3). The performances become better when $m \geq 200$. In fact, similar conclusions were obtained for other control charts with estimated process parameters (see Saleh et al., 2015b; Jardim et al., 2020; Weiß et al., 2021). Based on the results in Table 3, we argue that the optimization procedure based on the GICP approach may not be suitable for $m \in \{50, 100\}$, since the CATS values are too volatile even at $\delta = 0.5$. This might lead to dubious charting constants being generated. When $m = 200$, there still exists some instabilities in the out-of-control performances over $\delta \in (0, 0.2]$. However, we are less concerned about the chart's behavior around $\delta < 0.1$, since our optimization procedure only accounts for mean shifts in the range of $\delta \in [0.1, 2.0]$. Besides, we can further improve the out-of-control performances around $\delta = 0.2$ using the AAEQL optimization procedure outlined in the subsequent paragraph. Therefore, it is sensible to consider a bare minimum of $m = 200$ in the optimization design of the SPRT chart with estimated process parameters under the GICP framework.

The optimization model for the SPRT chart with estimated process parameters under the GICP framework is the same as the optimization model in (38) - (42), except that constraint (39) is replaced by constraint (44). The steps for obtaining the optimal charting parameters under the GICP framework are given as follows:

Steps 1: Specify a fixed set of parameters: m, τ, p, ε, R, $d_{\min}$, $\delta_{\min}$, and $\delta_{\max}$.

Step 2: This step is very similar to Step 2 from the optimization model in (38) - (42).

Step 3: This step is very similar to Step 3 from the optimization model in (38) - (42). In this step, the value of h is adjusted to satisfy constraint (44) using a stochastic approximation algorithm, together with Equation (24). Given the adjusted value of h, the value of g is adjusted using the Newton-Raphson algorithm to meet the specification on $\text{AASN}_0 = \bar{n}$ (constraint (40)).

Step 4: This step is very similar to Step 4 from the optimization model in (38) - (42).

Tables 4 and 5 tabulate the charting parameters together with the minimum AAEQL of the optimal SPRT chart with estimated process parameters under the GICP framework for $m \in \{200, 400, 600, 1000, 2000\}$. Their corresponding values of AATS, ASDTS and SDATS for $\delta \in \{0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 1.5, 2.0\}$ are also presented in Tables 4 and 5. Note that all the optimal charting parameters are obtained with the specifications in (37). The optimal chart in Tables 4 and 5 is designed with $p = 0.05$ and 0.10, respectively. In both tables, the EPC is implemented on two tolerance levels, i.e., $\varepsilon = 0$ (no tolerance) and $\varepsilon = 0.2$ (20% tolerance). As an example, when $p = 0.05$, $\varepsilon = 0.2$, and $m = 400$, the optimal charting parameters $(\text{AASN}_0, \gamma, d, g, h) = (2.251, 0.280, 0.450, 0.375, 11.780)$ and the optimal AAEQL = 0.928, whereas the corresponding values of $(\text{AATS}_\delta, \text{ASDTS}_\delta, \text{SDATS}_\delta)$ evaluated at $\delta = 0.6$ are computed as (1.47, 1.50, 0.27) (see Table 4).

Based on the results in Tables 4 and 5, we highlight three key points and trends as follows:

(i) First, introducing a small tolerance ε or increasing the value of p leads to better out-of-control performances. This can be observed

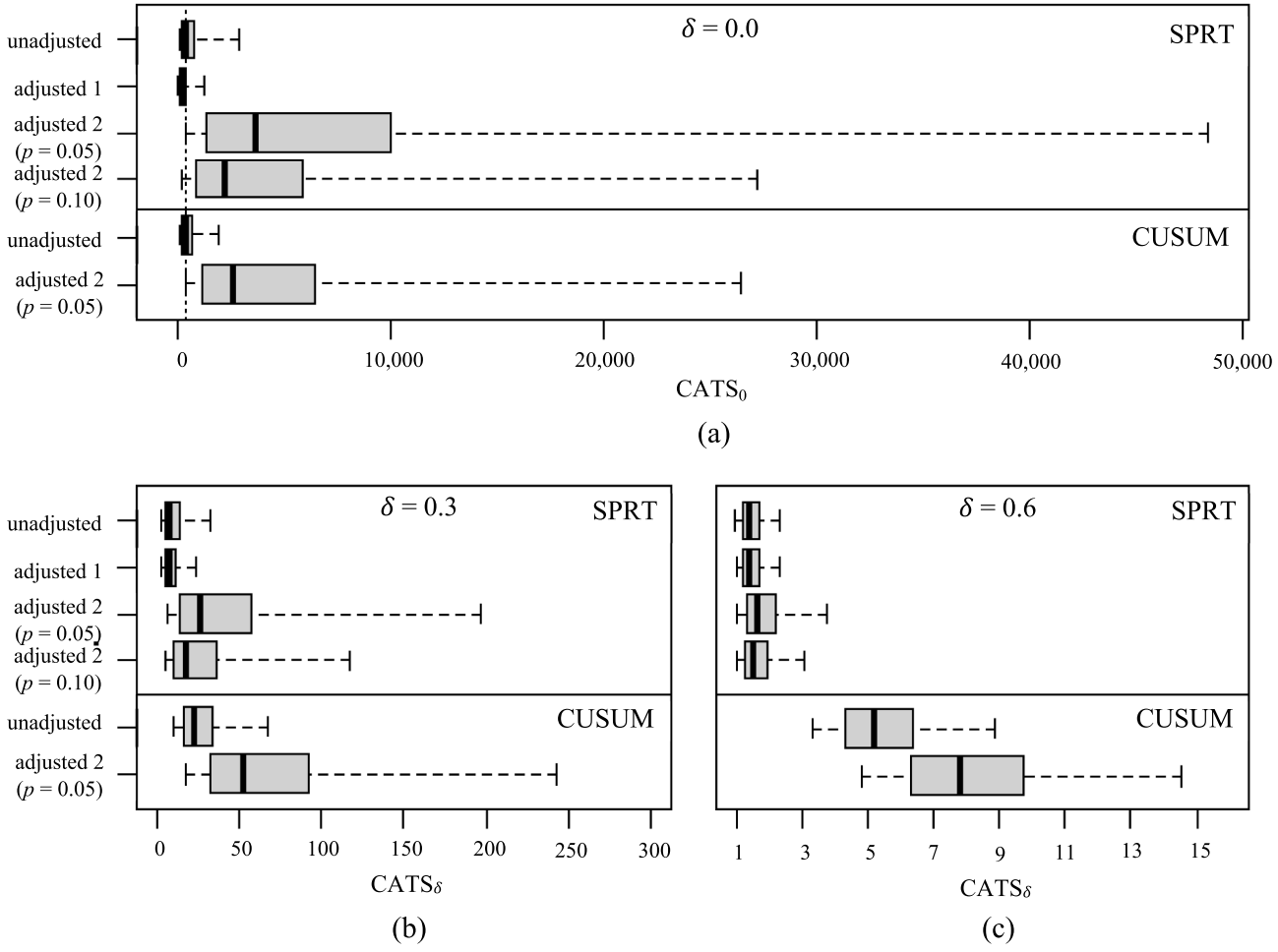

Fig. 1. Boxplots of the CATS distributions evaluated at (a) $\delta = 0.0$, (b) $\delta = 0.3$, and (c) $\delta = 0.6$ for the SPRT and CUSUM charts with estimated process parameters when $\tau = 370.40$, $R = 5$, and $m = 200$.

through a general reduction in the values of the unconditional metrics. For instance, in Table 4, when $m = 400$, the value of AAEQL drops by approximately 5.6% after a tolerance of 0.2 is introduced, whereas the values of ($AATS_\delta$, $ASDTS_\delta$, $SDATS_\delta$) evaluated at $\delta = 0.2$ fall by (12.1%, 19.8%, 21.5%) following the changes. This improvement in performances is more visible when both δ and m are small.

(ii) Second, the upper control limit h of the optimal SPRT chart generally decreases as m increases. This is consistent with the fact that estimated process parameters eventually converge to the true process parameters as m increases. However, there is one notable exception. When $m = 200$, the resulting control limits are the lowest among all the cases. Also, the optimal γ values seem to steer away from the other cases. To illustrate the situation, we draw one's attention to the optimal γ values obtained for $m = 200$ in both Tables 4 and 5. It can be observed that the optimal γ values obtained when $m = 200$ range from 0.350 to 0.420, whereas the rest of the cases report consistent values of γ lying between 0.260 and 0.290. Referring to this unusual pattern, we have explained using Table 3 that, for small values of m, the optimization process is likely to be distorted due to unstable conditional properties of the SPRT chart with estimated process parameters. This leads to slightly different optimal charting parameters obtained when m is small. After careful assessment of the reliability of the results, at much discretion, we decide that the optimization design is still feasible and meaningful when $m = 200$.

(iii) Third, the out-of-control performances in all cases of estimated process parameters are almost the same when $\delta \geq 0.6$ even with $m = 200$. For all values of m, the $AATS_\delta$ values drop well below twice the sampling interval (i.e., less than $2d$) when $\delta \geq 0.8$. This indicates fast detection speed of the optimal SPRT chart designed under the GICP framework towards moderate and large δ. These results are undoubtedly favorable to practitioners as they prevail almost the same detection speed as those cases with known process parameters, even for moderate values of δ and m.

Fig. 1 displays the boxplots of CATS distributions obtained from the optimal SPRT chart with estimated process parameters developed in Sections 3, 4.1 and 4.2 for $m = 200$. We also include boxplots of CATS distributions obtained from the optimal upper one-sided CUSUM chart with estimated process parameters when $m = 200$ for comparison. In Fig. 1, six boxplots are presented in each subfigure. The first four boxplots represent the CATS distributions obtained using the optimal SPRT chart, whereas the last two boxplots represent those obtained using the optimal CUSUM chart. Each boxplot is constructed from one of the following designs:

(i) the traditional approach from Section 3 (unadjusted),
(ii) the AATS-matching approach from Section 4.1 (adjusted 1), and.
(iii) the GICP approach from Section 4.2 (adjusted 2), with $\varepsilon = 0$ and $p = \{0.05, 0.10\}$.

The charting parameters of the optimal SPRT chart, designed under (i), (ii), and (iii) can be found from Tables 1, 2, and 4, respectively. It is important to note that the optimal CUSUM chart are designed only under (i) and (iii), since the primary focus of this article encompasses the GICP framework. To ensure fairness across comparison, we adopt the same specification as in (37), i.e., $\tau = 370.40$, $R = 5$, $d_{\min} = 0.25$, $\delta_{\min} = 0.1$, and $\delta_{\max} = 2.0$ when designing the optimal CUSUM chart under designs (i) and (iii). The charting parameters and the resulting AEQL or AAEQL of the optimal CUSUM chart with subgroup equal to one are determined as:

unadjusted: $d = 0.2$, $\beta = 0.400$, $UCL_{CUSUM} = 6.859$, AEQL = 2.290.

adjusted 2, $p = 0.05$: $d = 0.2$, $\beta = 0.416$, $UCL_{CUSUM} = 9.023$, AAEQL = 4.412.

Here, UCL_{CUSUM} is the upper control limit and β is the reference parameter of the upper one-sided CUSUM chart with estimated process parameters. Ou et al. (2011a, b) suggested that the subgroup of the CUSUM chart is chosen as one to achieve the best performance over a wide range of mean shifts.

The traditional optimal upper-sided CUSUM chart (design (i)) can be obtained directly from Ou et al. (2011b), whereas the optimal upper-sided CUSUM chart under GICP (design (iii)) can be developed by modifying the framework provided by Diko et al. (2019a). Specifically, under design (iii), the $CATS_0$ values are calculated by considering only the upper one-sided counterpart of the optimal CUSUM chart with GICP-adjusted limits. Since the SPRT and CUSUM charts with estimated process parameters possess similar working mechanisms, we omit full descriptions of the optimization algorithm for the CUSUM chart under design (iii).

To illustrate the relationship between the in-control and out-of-control performances, the boxplots of CATS distributions are given for the in-control case ($\delta = 0.0$) and two out-of-control cases ($\delta \in \{0.3, 0.6\}$) in Fig. 1. The vertical dotted line in Fig. 1(a) indicates the recommended level $\tau = 370.40$. All the boxplots should be interpreted as showing the 5th, 25th, 50th, 75th and 95th quantiles of the CATS distribution. From Fig. 1(a), the control charts designed under GICP outperform those designed under the traditional and AATS-matching approaches in terms of the in-control performance. It is obvious that from Fig. 1(a), all the boxplots with unadjusted limits and adjusted limits 1 show that approximately 50% of the $CATS_0$ fall below the recommended level of 370.4, whereas a satisfactorily high percentage of $CATS_0 = 95\%$ (or 90%) is achieved after the control limits have been adjusted to satisfy the EPC with $p = 0.05$ (or 0.10) (i.e., adjusted limits 2). It is important to note that, under design (iii), the percentage of harmful false alarms has been reduced from approximately 50% (i.e., in designs (i) and (ii)) to only 5% or 10%. As expected, one also notices the prominently large variability present in the $CATS_0$ distribution associated with adjusted limits 2. This is due to an increase in the values of the upper control limits under the GICP design. Nonetheless, this is far less concerning, since we are assured that majority (>90%) of the control charts with estimated process parameters perform as good as the baseline τ, even though the in-control performances are quite dispersed. As a result, the out-of-control performances of the control charts with adjusted limits 2 are worse than those of the other charts with unadjusted limits and adjusted limits 1, especially for small δ ($\leq$0.3). It is interesting to note that the optimal SPRT chart shows better out-of-control performance than the optimal CUSUM chart for all δ (see Fig. 1(b) and 1(c)). In fact, similar conclusions for the in-control and out-of-control performances were also found by Saleh et al. (2015a), Goedhart et al. (2017), and Diko et al. (2019b) for other control charts with estimated process parameters. The loss in out-of-control performance can be thought of as the necessary price to pay for the "guaranteed" in-control performance under the GICP framework. To restore the out-of-control performance, one is recommended to exercise a more relaxed in-control policy, i.e., by using a large value of p and/or introducing a small tolerance term ε in the formulation of the EPC. This approach has been discussed and implemented by numerous researchers, such as Goedhart et al. (2017), Diko et al. (2019a, b), and Jardim et al. (2020). By meticulously tuning the value of p and/or ε, one can probe the "nice spot" between satisfactory in-control and out-of-control performances of the control chart with estimated process parameters. In the article, we have provided ready-to-use optimal charting parameters of the SPRT chart with GICP-adjusted limits in Tables 4 and 5, for $p \in \{0.05, 0.10\}$ and $\varepsilon \in \{0, 0.2\}$. In case a practitioner would like to design the SPRT chart using other values of p and/or ε, we direct them to the optimization algorithm outlined in Section 4.2.

One relieving fact to note here is that the conditional out-of-control

performance of the SPRT chart with estimated process parameters, under all four control limits, are not significantly different from one another when $\delta = 0.6$ (see Fig. 1(c)), even with $\varepsilon = 0$. This is by virtue of the optimization design, which greatly reduces $CATS_\delta$ at moderate and large mean shifts, as proposed in Section 4.2. On the other hand, the CATS distributions of the CUSUM chart with estimated process parameters for unadjusted and adjusted limits 2 still present considerable differences when $\delta = 0.6$. In fact, both Diko et al. (2019a) and Saleh et al. (2015a) found that the CUSUM and EWMA control charts with GICP-adjusted limits, respectively, only perform as good as their corresponding charts with known process parameters when $\delta \geq 1.0$. Hence, we deduce that our SPRT chart with GICP-adjusted limits is a better performer, since its out-of-control performances closely resemble those of the known-parameter case when $\delta \geq 0.6$. It is also unsurprising to note that the AAEQL of the optimal CUSUM chart with adjusted limits 2 ($p = 0.05$) is at least twice as large as that of the corresponding optimal SPRT chart. Ultimately, we conclude that the optimal SPRT chart with adjusted limits 2 renders the best overall performance in both the in-control (i.e., 90% or 95% good false alarm rates) and out-of-control (i.e., short $CATS_\delta$ even when δ is relatively small ($\delta = 0.6$)) situations.

5. An application on real data

In this section, we present an application of the SPRT chart with estimated process parameters in monitoring real process data obtained from a wafer substrate manufacturing company. Prior to the illustrative example, we outline a general procedure for implementing the proposed optimal SPRT chart under the GICP framework. This step-by-step procedure is devised carefully to ensure the comprehensibility of the instructions, and ease of implementation for industrial practitioners. In the following, we outline the general steps for implementing the optimal SPRT chart with estimated process parameters:

(1) Collect m Phase-I observations. We recommend choosing $m \geq 200$.
(2) Calculate the estimated mean $\hat{\mu}_0$ and standard deviation $\hat{\sigma}_0$ from the Phase-I samples obtained in Step 1, using Equations (16) and (17), respectively.
(3) Verify that the m Phase-I observations indeed come from an in-control source. See Chakraborti et al. (2008) for an overview of Phase-I parametric control charts.
(4) Define the chart's specifications required to design the optimal SPRT chart with GICP-adjusted control limits:
   (i) General: the recommended ATS_0 (τ), the inspection rate (R), the minimum allowable sampling interval ($d_{\min}$),
   (ii) AAEQL Optimization: the minimum allowable mean shift ($\delta_{\min}$), the maximum tolerable mean shift ($\delta_{\max}$), and.
   (iii) GICP: the error probability (p), the tolerance level (ε).
   Run the optimization program in Section 4.2 using specifications (i)~(iii) to build a customized optimal SPRT chart with GICP-adjusted limits. Alternatively, practitioners may select ready-to-apply optimal charting parameters ($AASN_0$, γ, d, g, h) from Tables 4 and 5 for various Phase-I sample sizes m.
(5) Sometimes, the optimal sampling interval d obtained in Step 4 may be awkward to be implemented in industrial settings. Therefore, it is possible to adjust d slightly as long as the chart's performance is not substantially compromised (see, for example, Ou et al., 2011a). Note that this step is optional.
(6) Initialize the SPRT chart for Phase-II process monitoring using the optimal charting parameters ($AASN_0$, γ, d, g, h) determined in Step 4. In the i^{th} SPRT, compute the recursive control statistic $\widehat{U}_{i,j}$ from a sequence of observations using Equations (18) and (19). If, at any time point,
   (i) the control statistic $\widehat{U}_{i,j}$ rises above h,
   - an out-of-control signal is produced, and the production line is suspended immediately. An out-of-control action plan is to be carried out to identify and eliminate the root cause.
   (ii) the control statistic $\widehat{U}_{i,j}$ lies in the region between g and h,
   - another observation is taken to compute the $(j+1)^{th}$ control statistic. The status of the SPRT chart is then re-evaluated using the new control statistic $\widehat{U}_{i,j+1}$.
   (iii) the control statistic $\widehat{U}_{i,j}$ falls below g,
   - sampling terminates and the process is accepted as in-control. The $(i+1)^{th}$ SPRT will be initiated after d time units, and the control statistic $\widehat{U}_{i+1,0}$ is reset to zero.

To demonstrate the execution of the optimal SPRT chart with estimated process parameters, we acquire real manufacturing data from an epitaxial process of silicon wafers. In manufacturing industry, silicon epitaxial wafers are widely used as semiconductors for the fabrication of integrated circuits. Particularly, epitaxial wafers are made of monocrystalline layers deposited on wafer substrates. Epitaxial layers are manufactured in a range of thicknesses and resistivities. In commercial epitaxial wafers, layer thickness is measured in micrometer (μm),

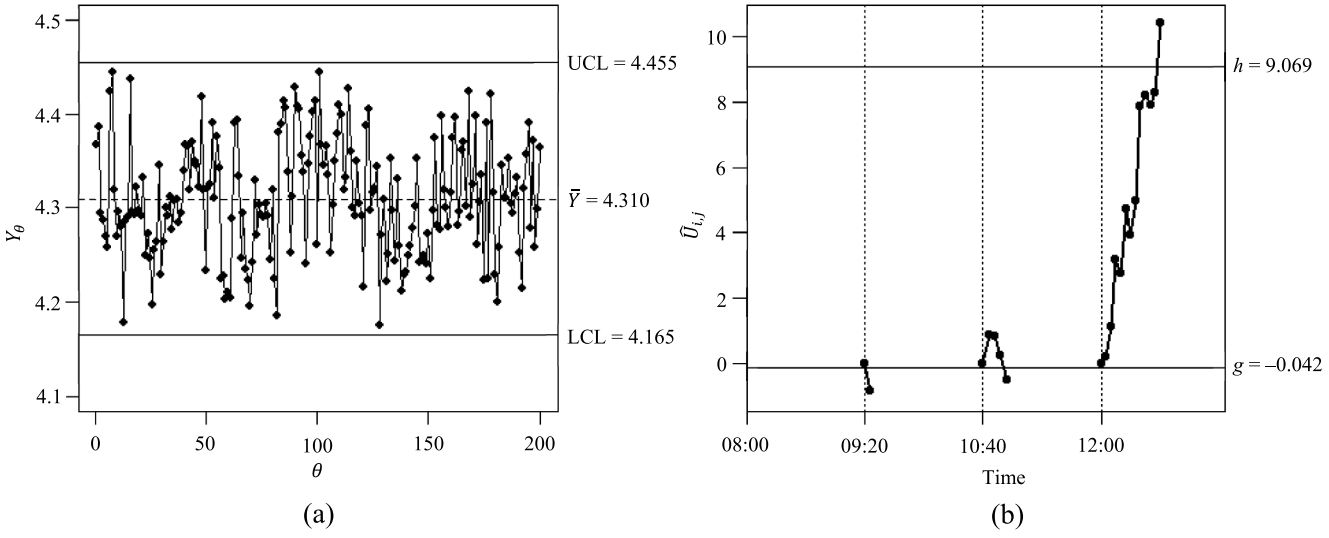

Fig. 2. (a) The Phase-I AMR chart for $m = 200$ individual observations and (b) the Phase-II optimal SPRT chart with estimated process parameters for monitoring industrial data obtained from the N-type epitaxial process.

Table 6
Phase-II data set for the illustrative example using the optimal SPRT chart with GICP-adjusted control limits.

Sampling time	Number of SPRT, i	Number of sample, j	Observation, X_{ij}	Standardized observation, $\widehat{Z}_{ij}$	Control statistic, $\widehat{U}_{ij}$	Conclusion
08:00	–	–	–	–	–	–
09:20	1	1	4.285	–0.407	–0.837	In-control
10:40	2	1	4.389	1.292	0.862	*
		2	4.334	0.390	0.821	*
		3	4.302	–0.131	0.260	*
		4	4.289	–0.344	–0.513	In-control
12:00	3	1	4.349	0.638	0.208	*
		2	4.393	1.363	1.141	*
		3	4.459	2.448	3.159	*
		4	4.311	0.027	2.755	*
		5	4.457	2.406	4.731	*
		6	4.288	–0.358	3.943	*
		7	4.399	1.457	4.970	*
		8	4.515	3.349	7.889	*
		9	4.357	0.765	8.224	*
		10	4.318	0.130	7.925	*
		11	4.358	0.792	8.287	*
		12	4.467	2.577	10.434	Out-of-control

whereas resistivity is measured with the unit ohm-centimeter (ohm-cm). Epitaxial resistivity is often the key quality characteristic to be monitored and controlled in epitaxial wafer production. Proper control of epitaxial resistivity is vital to produce devices with desired performance and high-quality integrated circuits.

The raw data provided in this study are obtained from an N-type epitaxy. Based on the experience of a process engineer, epitaxial resistivity is likely to increase due to insufficient dopant during the doping process, and vice versa. A low N-type dopant deposition will lead to layers with high resistivity, while a high N-type dopant deposition will lead to layers with low resistivity. For illustration purposes, we demonstrate the detection of an upward mean shift in the N-type epitaxial resistivity by employing an upper one-sided SPRT control chart with estimated process parameters.

The following steps demonstrate the implementation of the upper one-sided SPRT chart with estimated process parameters for monitoring N-type epitaxial resistivity:

(1) In the Phase-I process, a set of data comprising $m = 200$ individual resistivity measurements (Y_1, Y_2, …, Y_{200}) are collected.
(2) The estimated sample mean and standard deviation of the Phase-I measurements are computed as $\widehat{\mu}_0 = 4.310$ ohm-cm and $\widehat{\sigma}_0 = 0.061$ ohm-cm, respectively, using Equations (16) and (17).
(3) The stability of the Phase-I samples is assessed using a classical control chart for individual observations based on the average moving range (AMR). The AMR control chart has been extensively discussed in several literatures (see Roes et al., 1993; Vermaat et al., 2003; Montgomery, 2009). This stems from the fact that the AMR chart is one of the simplest and most effective individuals control charts in process monitoring. Roes et al. (1993) advised against the use of a moving range chart in conjunction with the AMR individuals control chart. They claimed that the interpretation of the moving range chart can be confusing to practitioners due to the presence of serial correlation between moving ranges. Therefore, in this illustration, only the AMR control chart for individual observations is used in Phase-I chart calibration.

Fig. 2(a) displays the AMR control chart for monitoring the Phase-I observations. The lower (LCL) and upper (UCL) control limits of the AMR chart are determined by $\bar{Y} \pm Z_{\alpha/2}(\sqrt{\pi}/2)\bar{MR} =$ (4.165, 4.455), where $\bar{Y} = \sum_{\theta=1}^{200} Y_\theta/200 = 4.310$ ohm-cm and $\bar{MR} = \sum_{\theta=2}^{200}|Y_\theta - Y_{\theta-1}|/199 = 0.055$ ohm-cm. Here, $Z_{\alpha/2}$ is chosen as 3 to provide a false alarm probability of 0.0027. Fig. 2(a) displays the AMR control chart for 200 Phase-I observations (Y_1, Y_2, …, Y_{200}). It is verified from Fig. 2(a) that the 200 Phase-I samples indeed come from an in-control process.
(4) The wafer company is currently inspecting five silicon wafers every three hours. This choice is elicited since from previous experience, the process has exhibited a rather stable nature. Each epitaxial wafer can be sampled and measured for its resistivity in a short time using a resistivity meter. Considering the current sampling interval (i.e., three hours) as one time unit, we design the optimal SPRT control chart with estimated process parameters based on the in-control specification, $\tau = 370.40$. We also adopt an inspection rate R equal to five, i.e., five wafers on average per time unit, where one time unit corresponds to three hours. Furthermore, we set $d_{\min} = 0.25$, $\delta_{\min} = 0.1$ and $\delta_{\max} = 2.0$. Given these specifications, we decide to employ the optimal SPRT chart with estimated process parameters designed under the GICP framework using an error probability $p = 0.05$ and $\varepsilon = 0$ as detailed in Section 4.2. From Table 4, the optimal charting parameters ($AASN_0$, γ, d, g, h) and the resulting AAEQL are (2.250, 0.420, 0.450, –0.034, 9.283) and 2.111, respectively.
(5) The optimal sampling interval of the SPRT chart with estimated process parameters is $d = 0.450$, or in calendar time, 81 min. For the convenience of implementation, we take the new sampling interval as 80 min (i.e., $d = 0.444$). The charting parameters are adjusted accordingly to satisfy $\Pr(CATS \geq 370.40) = 0.95$ and $R = 5$. The new charting parameters ($AASN_0$, γ, d, g, h) and AAEQL are found as (2.219, 0.430, 0.444, –0.042, 9.069) and 2.137, respectively. Note that the AAEQL increases by only 1.23% following these new adjustments.
(6) The optimal SPRT chart with estimated process parameters for monitoring the Phase-II epitaxial resistivity process is displayed both in tabular form (see Table 6) and graphical form (see Fig. 2(b)). Table 6 contains the summary statistics of the Phase-II SPRT chart, including the timeline, individual measurements, standardized measurements, control statistics, and decision at each sampling point. Fig. 2(b) presents a simple visualization of the SPRT chart in action, with the horizontal and vertical axes revealing precise timestamps and control statistics, respectively. Note that the first dot in each SPRT shown in Fig. 2(b) represents the initial value $\widehat{U}_{i,0} = 0$, whereas the subsequent dots in each SPRT represent the recursively computed chart statistic, $\widehat{U}_{ij}$ in Equation (18), for $i = 1, 2, \ldots$ and $j = 1, 2, \ldots, n$. It is important to note that the resistivity measurements are taken sequentially from the second dot onwards in each SPRT. Referring to Table 6 and Fig. 2(b), the following observations are obtained.

(i) At time 08:00, the SPRT control chart with estimated process parameters is initiated with an 80-minute time-period allocated for machinery start-up.
(ii) At time 09:20, the first SPRT is computed by taking the first resistivity measurement, i.e., $X_{1,1} = 4.285$. Using Equations (19) and (18), the standardized measurement $\widehat{Z}_{1,1}$ and the first control statistic $\widehat{U}_{1,1}$ are computed as –0.407 and –0.837, respectively. Since the chart statistic $\widehat{U}_{1,1} < -0.042 = g$, the process is declared as in-control and the second SPRT is to be initiated 80 minutes later.
(iii) At time 10:40, four new observations are taken sequentially to calculate the second SPRT. The SPRT chart indicates that the process is still in-control.
(iv) At 12:00, the third SPRT samples a total of 12 observations before an out-of-control signal is given. The signal is emitted since the chart statistic $\widehat{U}_{3,12}$ lies above the upper control limit h. The assignable causes leading to the process shift are immediately investigated, and the time to signal is reported to be less than four hours.

6. Conclusions

In this article, we improve the architecture of the SPRT chart by considering the estimation errors of the process parameters. By deriving the SDTS formula, we fill a gap in the run-length distribution of the SPRT chart with known process parameters. We also extend the optimization design of the SPRT chart to a more realistic ground, i.e., when the process parameters are estimated. In our endeavor, we devise an optimal SPRT chart with GICP-adjusted limits that not only guarantees a satisfactory level of in-control performances across practitioners, but also warrants little to no compromise in the out-of-control performances.

In this section, we provide a list of contributions established throughout the article and highlight some key remarks that are worth revisiting.

1. New theoretical formulae for the SPRT chart with estimated process parameters:
   The theoretical formulae for the conditional and unconditional properties are derived using the Markov chain approach. By evaluating the unconditional properties (AATS, ASDTS, SDATS), it is possible to assess the expected chart's performances across practitioners. From the conditional viewpoint, we suggest drawing inferences based on the exceedance probabilities computed using the empirical distributions of the $CATS_0$.
2. Negative impacts of using traditional and AATS-matching control limits for the SPRT chart when the process parameters are estimated:
   We discover that the traditional SPRT chart has a 50% chance of yielding unfavorable false alarm rates, even when $m = 2000$. This condition worsens when the SPRT chart is designed using the AATS-matching approach. We argue that the SPRT chart with unadjusted and AATS-matching control limits may not be suitable for industrial applications, since the massive number of false signals can endanger the efficiency of a production line.
3. AAEQL optimization design for the SPRT chart with estimated process parameters based on the GICP framework:
   By coupling the GICP framework with the AAEQL optimization design, we develop an optimal SPRT chart that is very robust towards parameter estimation. Particularly, the GICP design ensures that the conditional in-control performance is met with a very high probability (90% or 95%), whereas the AAEQL optimal design improves the out-of-control performances for $\delta \geq 0.6$. Since this concept has been proposed for the first time in literature, we provide a comprehensive optimization algorithm in Section 4.2 to encourage repeatability and reproducibility of the research work. Besides, we recommend using at least $m = 200$ Phase-I samples to ensure the reliability of the results generated.
4. Advantages of adopting AAEQL in the proposed optimization model:
   Since process shifts tend to occur sparsely in practice, the AAEQL makes a useful criterion as it averages the out-of-control performances of the SPRT chart over a pre-specified range of mean shifts. Moreover, from an economic point of view, the AAEQL optimization minimizes the average detection time based on the severity of mean shifts. This is a desirable property for industrial quality control since economic costs are often vital considerations in the design of control charts.
5. Ready-to-use optimal SPRT charting parameters for various Phase-I sample sizes:
   The results tabulated in Tables 4 and 5 allow practitioners to select the most suitable set of optimal charting parameters based on their preferences and available resources. Furthermore, we offer flexibility to industrial practitioners, who might prefer to adopt a more relaxed control policy, by providing several chart designs based on different error rates ($p \in \{0.05, 0.10\}$) and tolerance levels ($\varepsilon \in \{0, 0.2\}$). This should accelerate successful implementation of our proposed chart to a myriad of industrial applications, including the ones in the conventional manufacturing environment (for example, monitoring the quality characteristics of thrust washers, nylon fibers, and airbag in factories) and smart control systems (for example, cloud computing infrastructure to maintain quality checks or vital processes in aerospace, and pharmaceutical companies).

Considering the widespread influence of the GICP in the development of statistical control charts, future research may explore the SPRT chart with estimated process parameters under a cautious parameter learning framework proposed by Capizzi and Masarotto (2020). The cautious parameter learning framework can be used in conjunction with GICP to alleviate the huge variability present in the CATS distribution. We are embarking upon this extension of the SPRT control chart with estimated process parameters in our future research.

CRediT authorship contribution statement

J.W. Teoh: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **W.L. Teoh:** Conceptualization, Formal analysis, Funding acquisition, Supervision, Writing – review & editing. **Michael B.C. Khoo:** Supervision, Validation, Writing – review & editing. **Philippe Castagliola:** Supervision, Writing – review & editing. **W. H. Moy:** Resources, Validation, Writing – review & editing.

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.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was supported by the Ministry of Higher Education (MOHE) Malaysia and Heriot-Watt University Malaysia under Fundamental Research Grant Scheme (FRGS), no. FRGS/1/2021/STG06/HWUM/02/1.

Appendix A. Derivations of $\hat{r}_{k,\ell}$, $\hat{b}_k$, $\hat{q}_k$, and $\hat{P}_0$ to evaluate the performances of the SPRT chart with estimated process parameters

In this appendix, we provide the derivations of $\hat{r}_{k,\ell}$, $\hat{b}_k$, $\hat{q}_k$, and $\hat{P}_0$, which are used in Section 2.2 to evaluate the performances of the SPRT chart with estimated process parameters. Let $\hat{r}_{k,\ell}$ in Equation (20) denote the transition probability from state S_k to S_ℓ. We have

$$\hat{r}_{k,\ell} = \Pr\left(\hat{U}_{i,j} \in S_\ell \middle| \hat{U}_{i,j-1} \in S_k\right) = \Pr\left(g + \Delta\cdot(\ell - 1) \le O_k + \hat{Z}_{i,j} - \gamma \le g + \Delta\cdot\ell\right), \tag{A.1}$$

where O_k is the midpoint of state S_k as in Equation (4). Substituting O_k into Equation (A.1) and with some rearrangements, we obtain

$$\hat{r}_{k,\ell} = \Pr\left(\Delta\cdot(\ell - k - 0.5) + \gamma \le \hat{Z}_{i,j} \le \Delta\cdot(\ell - k + 0.5) + \gamma\right). \tag{A.2}$$

Substituting $\hat{Z}_{i,j}$ from Equation (19) into Equation (A.2) and with some rearrangements, we obtain

$$\hat{r}_{k,\ell} = \Pr\left(V(\Delta\cdot(\ell - k - 0.5) + \gamma) + \frac{W}{\sqrt{m}} - \delta \le \frac{X_{i,j} - (\mu_0 + \delta\sigma_0)}{\sigma_0} \le V(\Delta\cdot(\ell - k + 0.5) + \gamma) + \frac{W}{\sqrt{m}} - \delta\right),$$

where $V = \hat{\sigma}_0/\sigma_0$ and $W = (\hat{\mu}_0 - \mu_0)/(\sigma_0/\sqrt{m})$. Since $X_{i,j} \sim N(\mu_0 + \delta\sigma_0, \sigma_0^2)$, we have $\left[X_{i,j} - (\mu_0 + \delta\sigma_0)\right]/\sigma_0 \sim N(0,1)$. It follows that

$$\hat{r}_{k,\ell} = \Phi\left[V(\Delta\cdot(\ell - k + 0.5) + \gamma) + \frac{W}{\sqrt{m}} - \delta\right] - \Phi\left[V(\Delta\cdot(\ell - k - 0.5) + \gamma) + \frac{W}{\sqrt{m}} - \delta\right]. \tag{A.3}$$

Let $\hat{b}_k$ in Equation (23) denote the one-step transition probability that the initial control statistic $\hat{U}_{i,0} = 0$ transits to state S_k, i.e.,

$$\hat{b}_k = \Pr\left(\hat{U}_{i,1} \in S_k \middle| \hat{U}_{i,0} = 0\right) = \Pr\left(g + \Delta\cdot(k - 1) \le \hat{Z}_{i,1} - \gamma \le g + \Delta\cdot k\right).$$

Using the same approach as in the derivation of $\hat{r}_{k,\ell}$, the formula $\hat{b}_k$ can be derived as

$$\hat{b}_k = \Phi\left[V(g + \Delta\cdot k + \gamma) + \frac{W}{\sqrt{m}} - \delta\right] - \Phi\left[V(g + \Delta\cdot(k - 1) + \gamma) + \frac{W}{\sqrt{m}} - \delta\right]. \tag{A.4}$$

Let $\hat{q}_k$ in Equation (29) denote the one-step transition probability from state S_k to the acceptance state, i.e.,

$$\hat{q}_k = \Pr\left(\hat{U}_{i,j} \le g \middle| \hat{U}_{i,j-1} \in S_k\right) = \Pr\left(O_k + \hat{Z}_{i,j} - \gamma \le g\right).$$

$\hat{q}_k$ can be deduced using the same method of derivation as in $\hat{r}_{k,\ell}$, i.e.,

$$\hat{q}_k = \Phi\left[V(\Delta\cdot(0.5 - k) + \gamma) + \frac{W}{\sqrt{m}} - \delta\right]. \tag{A.5}$$

Let $\hat{P}_0$ in Equation (30) denote the one-step transition probability that the initial control statistic $\hat{U}_{i,0} = 0$ transits to the acceptance state, i.e.,

$$\hat{P}_0 = \Pr\left(\hat{U}_{i,1} \le g \middle| \hat{U}_{i,0} = 0\right) = \Pr\left(\hat{Z}_{i,1} - \gamma \le g\right).$$

Then, $\hat{P}_0$ is attained using the same method of derivation as in $\hat{r}_{k,\ell}$, i.e.,

$$\hat{P}_0 = \Phi\left[V(g + \gamma) + \frac{W}{\sqrt{m}} - \delta\right]. \tag{A.6}$$

Appendix B. Derivations of the CSDTS and ASDTS of the SPRT chart with estimated process parameters

In this appendix, we provide derivations for the CSDTS_0, steady-state CSDTS_δ, ASDTS_0 and steady-state ASDTS_δ from Equations (26), (27), (34) and (35), respectively. Note that ASDTS measures the overall standard deviation of the time to signal across all possible values of V and W. In the case of an in-control process, let RL be a geometric random variable with the probability $1 - \widehat{\text{OC}}(0)$, where $\widehat{\text{OC}}(0)$ can be computed from Equation (28) by setting $\delta = 0$. The first (E[RL]) and second ($\text{E}[\text{RL}^2]$) moments of RL can be easily obtained as

$$\text{E}[\text{RL}] = \frac{1}{1 - \widehat{\text{OC}}(0)} \tag{B.1}$$

and

$$\text{E}[\text{RL}^2] = \frac{2 - [1 - \widehat{\text{OC}}(0)]}{[1 - \widehat{\text{OC}}(0)]^2} = \frac{1 + \widehat{\text{OC}}(0)}{[1 - \widehat{\text{OC}}(0)]^2}, \tag{B.2}$$

respectively. The CSDTS_0 can be derived using Equations (B.1) and (B.2) as follows

$$\mathrm{CSDTS}_0 = d\sqrt{E\left[\mathrm{RL}^2\right] - E[\mathrm{RL}]^2} = d\sqrt{\frac{1+\widehat{\mathrm{OC}}(0)}{[1-\widehat{\mathrm{OC}}(0)]^2} - \left(\frac{1}{1-\widehat{\mathrm{OC}}(0)}\right)^2}, \tag{B.3}$$

and the unconditional ASDTS_0 is obtained by averaging all the possible values of V and W, i.e.,

$$\begin{aligned}\mathrm{ASDTS}_0 &= d\sqrt{\int_{-\infty}^{\infty}\int_0^{\infty}\frac{1+\widehat{\mathrm{OC}}(0)}{[1-\widehat{\mathrm{OC}}(0)]^2}f_V(v)f_W(w)dvdw - \left(\int_{-\infty}^{\infty}\int_0^{\infty}\frac{1}{1-\widehat{\mathrm{OC}}(0)}f_V(v)f_W(w)dvdw\right)^2} \\ &= d\sqrt{\int_{-\infty}^{\infty}\int_0^{\infty}\frac{1+\widehat{\mathrm{OC}}(0)}{[1-\widehat{\mathrm{OC}}(0)]^2}f_V(v)f_W(w)\ dvdw - \left(\frac{\mathrm{AATS}_0}{d}\right)^2},\end{aligned} \tag{B.4}$$

where $f_V(v)$ can be found in Equation (21), $f_W(w)$ represents the pdf of a standard normal distribution, and AATS_0 can be computed from Equation (32).

In the case of an out-of-control process, recall that the time to signal can be expressed as $d \times (T + \mathrm{RL} - 1)$ (see Section 2.1), where T is a uniform random variable with range [0, 1]. Assuming independence between run lengths and uniformly distributed time of process shift, the first ($\mathrm{E}[T+\mathrm{RL}-1]$) and second ($\mathrm{E}\left[(T+\mathrm{RL}-1)^2\right]$) moments of $(T + \mathrm{RL} - 1)$ are evaluated as

$$\mathrm{E}[T+\mathrm{RL}-1] = \frac{1}{1-\widehat{\mathrm{OC}}(\delta)} - \frac{1}{2} \tag{B.5}$$

and

$$\begin{aligned}\mathrm{E}\left[(T+\mathrm{RL}-1)^2\right] &= \mathrm{E}\left[T^2 + 2T(\mathrm{RL}-1) + (\mathrm{RL}-1)^2\right] \\ &= \mathrm{E}\left[T^2\right] + 2\cdot\mathrm{E}[T]\cdot\mathrm{E}[\mathrm{RL}-1] + \mathrm{E}\left[(\mathrm{RL}-1)^2\right] \\ &= \frac{1}{3} + 2\cdot\frac{1}{2}\cdot\left(\frac{1}{1-\widehat{\mathrm{OC}}(\delta)} - 1\right) + \frac{\widehat{\mathrm{OC}}(\delta)[1+\widehat{\mathrm{OC}}(\delta)]}{[1-\widehat{\mathrm{OC}}(\delta)]^2} \\ &= \frac{1+\widehat{\mathrm{OC}}^2(\delta)}{[1-\widehat{\mathrm{OC}}(\delta)]^2} - \frac{2}{3},\end{aligned} \tag{B.6}$$

respectively, where $\mathrm{RL} \sim$ Geometric $(1-\widehat{\mathrm{OC}}(\delta))$ and $\widehat{\mathrm{OC}}(\delta)$ can be obtained from Equation (28). The steady-state CSDTS_δ is derived using Equations (B.5) and (B.6) as follows:

$$\mathrm{CSDTS}_\delta = d\sqrt{\mathrm{E}\left[(T+\mathrm{RL}-1)^2\right] - (\mathrm{E}[T+\mathrm{RL}-1])^2} = d\sqrt{\frac{1+\widehat{\mathrm{OC}}^2(\delta)}{[1-\widehat{\mathrm{OC}}(\delta)]^2} - \frac{2}{3} - \left(\frac{1}{1-\widehat{\mathrm{OC}}(\delta)} - \frac{1}{2}\right)^2}, \tag{B.7}$$

and the unconditional steady-state ASDTS_δ is obtained by

$$\begin{aligned}\mathrm{ASDTS}_\delta &= d\sqrt{\int_{-\infty}^{\infty}\int_0^{\infty}\left(\frac{1+\widehat{\mathrm{OC}}^2(\delta)}{[1-\widehat{\mathrm{OC}}(\delta)]^2} - \frac{2}{3}\right)f_V(v)f_W(w)dvdw - \left[\int_{-\infty}^{\infty}\int_0^{\infty}\left(\frac{1}{1-\widehat{\mathrm{OC}}(\delta)} - \frac{1}{2}\right)f_V(v)f_W(w)dvdw\right]^2} \\ &= d\sqrt{\int_{-\infty}^{\infty}\int_0^{\infty}\frac{1+\widehat{\mathrm{OC}}^2(\delta)}{[1-\widehat{\mathrm{OC}}(\delta)]^2}f_V(v)f_W(w)\ dvdw - \frac{2}{3} - \left(\frac{\mathrm{AATS}_\delta}{d}\right)^2},\end{aligned} \tag{B.8}$$

where AATS_δ can be computed from Equation (32).

Appendix C. Glossary of acronyms and notations

AEQL	Average extra quadratic loss
AMR	Average moving range
ARL	Average run length
ASN	Average sample number
ASDTS	Average standard deviation of the time to signal
ATS	Average time to signal
CARL	Conditional average run length
CASN	Conditional average sample number
CATS	Conditional average time to signal
CSDTS	Conditional standard deviation of the time to signal
$U_{i,j}$, $\widehat{U}_{i,j}$	Control statistic of the upper one-sided SPRT chart with known or estimated process parameters
CUSUM	Cumulative sum chart

(continued on next page)

(*continued*)

p	Error probability in EPC
$\widehat{\mu}_0$, $\widehat{\sigma}_0$	Estimated in-control mean and standard deviation of a quality characteristic
EPC	Expected probability criterion
AAEQL	Expected value of the average extra quadratic loss
AASN	Expected value of the average sample number
AATS	Expected value of the average time to signal
EWMA	Exponentially weighted moving average chart
FSI	Fixed sampling interval
GICP	Guaranteed in-control performance
μ_0, σ_0	In-control mean and standard deviation of a quality characteristic
R	Inspection rate
g, h	Lower and upper control limits of the SPRT chart
LCL	Lower control limit
δ	Mean shift in numbers of standard deviation
$d_{\min}$	Minimum allowable sampling interval
$\delta_{\min}$, $\delta_{\max}$	Minimum and maximum allowable mean shifts
Y_θ	Phase-I observation
m	Phase-I sample size
$X_{i,j}$	Phase-II observation
τ	Recommended in-control ATS
γ	Reference parameter of the SPRT chart
RL	Run length
d	Sampling interval
SPRT	Sequential probability ratio test
SDATS	Standard deviation of the average time to signal
SDTS	Standard deviation of the time to signal
$Z_{i,j}$, $\widehat{Z}_{i,j}$	Standardized observation of the SPRT chart with known or estimated process parameters
TS	Time to signal
ε	Tolerance term in EPC
UCL	Upper control limit
VSSI	Variable sample size and sampling interval
VSI	Variable sampling interval

References

Adeoti, O. A., & Malela-Majika, J. C. (2020). Double exponentially weighted moving average control chart with supplementary runs-rules. *Quality Technology & Quantitative Management, 17*(2), 149–172.

Albers, W., Kallenberg, W. C. M., & Nurdiati, S. (2005). Exceedance probabilities for parametric control charts. *Statistics, 39*(5), 429–443.

Anwar, S. M., Aslam, M., Zaman, B., & Riaz, M. (2021). Mixed memory control chart based on auxiliary information for simultaneously monitoring of process parameters: An application in glass field. *Computers & Industrial Engineering, 156*(6), Article 107284.

Capizzi, G., & Masarotto, G. (2020). Guaranteed in-control control chart performance with cautious parameter learning. *Journal of Quality Technology, 52*(4), 385–403.

Castagliola, P., Maravelakis, P. E., & Figueiredo, F. O. (2016). The EWMA median chart with estimated parameters. *Quality and Reliability Engineering International, 48*(1), 66–74.

Celano, G., Castagliola, P., Fichera, S., & Nenes, G. (2013). Performance of *t* control charts in short runs with unknown shift sizes. *Computers & Industrial Engineering, 64*(1), 56–68.

Chakraborti, S. (2006). Parameter estimation and design considerations in prospective applications of the X chart. *Journal of Applied Statistics, 33*(4), 439–459.

Chakraborti, S., Human, S. W., & Graham, M. A. (2008). Phase I statistical process control charts: An overview and some results. *Quality Engineering, 21*(1), 52–62.

Diko, M. D., Chakraborti, S., & Does, R. J. M. M. (2019a). An alternative design of the two-sided CUSUM chart for monitoring the mean when parameters are estimated. *Computers & Industrial Engineering, 137*(11), Article 106042.

Diko, M. D., Chakraborti, S., & Does, R. J. M. M. (2019b). Guaranteed in-control performance of the EWMA chart for monitoring the mean. *Quality and Reliability Engineering International, 35*(4), 1144–1160.

Gandy, A., & Kvaløy, J. T. (2013). Guaranteed conditional performance of control charts via bootstrap methods. *Scandinavian Journal of Statistics, 40*(4), 647–668.

Goedhart, R., Schoonhoven, M., & Does, R. J. M. M. (2016). Correction factors for Shewhart X and X control charts to achieve desired unconditional ARL. *International Journal of Production Research, 54*(24), 7464–7479.

Goedhart, R., Schoonhoven, M., & Does, R. J. M. M. (2017). Guaranteed in-control performance for the Shewhart X and X control charts. *Journal of Quality Technology, 49*(2), 155–171.

Jardim, F. S., Chakraborti, S., & Epprecht, E. K. (2020). Two perspectives for designing a Phase II control chart with estimated parameters: The case of the Shewhart X chart. *Journal of Quality Technology, 52*(2), 198–217.

Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., & Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: A literature review. *Journal of Quality Technology, 38*(4), 349–364.

Jones, L. A., Champ, C. W., & Rigdon, S. E. (2004). The run length distribution of the CUSUM with estimated parameters. *Journal of Quality Technology, 36*(1), 95–108.

Jones-Farmer, L. A., Woodall, W. H., Steiner, S. H., & Champ, C. W. (2014). An overview of Phase I analysis for process improvement and monitoring. *Journal of Quality Technology, 46*(3), 265–280.

Mahadik, S. B., Godase, D. G., & Teoh, W. L. (2021). A two-sided SPRT control chart for process dispersion. *Journal of Statistical Computation and Simulation, 91*(17), 3603–3614.

Mehmood, R., Riaz, M., Lee, M. H., Ali, I., & Gharib, M. (2022). Exact computational methods for univariate and multivariate control charts under runs rules. *Computers & Industrial Engineering, 163*(1), Article 107821.

Montgomery, D. C. (2009). *Introduction to statistical quality control*. New York: John Wiley & Sons.

Ou, Y., Wu, Z., & Goh, T. N. (2011a). A new SPRT chart for monitoring process mean and variance. *International Journal of Production Economics, 132*(2), 303–314.

Ou, Y., Wu, Z., Lee, K. M., & Chen, S. L. (2011b). *A highly effective SPRT control chart for monitoring process mean*. School of Mechanical and Aerospace Engineering, Nanyang Technological University. Technical Report,.

Ou, Y., Wu, Z., Yu, F. J., & Shamsuzzaman, M. (2011c). An SPRT control chart with variable sampling intervals. *The International Journal of Advanced Manufacturing Technology, 56*(9–12), 1149–1158.

Reynolds, M. R., Jr., & Stoumbos, Z. G. (2004). Control charts and the efficient allocation of sampling resources. *Technometrics, 46*(2), 200–214.

Roes, K. C. B., Does, R. J. M. M., & Schurink, Y. (1993). Shewhart-type control charts for individual observations. *Journal of Quality Technology, 25*(3), 188–198.

Ryu, J. H., Wan, G., & Kim, S. (2010). Optimal design of a CUSUM chart for a mean shift of unknown size. *Journal of Quality Technology, 42*(3), 311–326.

Saleh, N. A., Mahmoud, M. A., Jones-Farmer, L. A., Zwetsloot, I., & Woodall, W. H. (2015a). Another look at the EWMA control chart with estimated parameters. *Journal of Quality Technology, 47*(4), 363–382.

Saleh, N. A., Mahmoud, M. A., Keefe, M. J., & Woodall, W. H. (2015b). The difficulty in designing Shewhart X and X control charts with estimated parameters. *Journal of Quality Technology, 47*(2), 127–138.

Salmasnia, A., Abdzadeh, B., & Namdar, M. (2017). A joint design of production run length, maintenance policy and control chart with multiple assignable causes. *Journal of Manufacturing Systems, 42*(1), 44–56.

Stoumbos, Z. G., & Reynolds, M. R., Jr. (1997). Control charts applying a sequential test at fixed sampling intervals. *Journal of Quality Technology, 29*(1), 21–40.

Stoumbos, Z. G., & Reynolds, M. R., Jr. (1999). The SPRT control chart for the process mean with samples starting at fixed times. *Nonlinear Analysis: Real World Applications, 2*(1), 1–34.

Teoh, W. L., Ong, L. V., Khoo, M. B. C., Castagliola, P., & Chong, Z. L. (2019). The variable sampling interval EWMA X chart with estimated process parameters. *Journal of Testing and Evaluation, 49*(2), 1237–1265.

Tran, K. D., Nadi, A. A., Nguyen, T. H., & Tran, K. P. (2021). One-sided Shewhart control charts for monitoring the ratio of two normal variables in short production runs. *Journal of Manufacturing Processes, 69*(9), 273–289.
Vermaat, M. B., Ion, R. A., Does, R. J. M. M., & Klaassen, C. A. J. (2003). A comparison of Shewhart individuals control charts based on normal, non-parametric, and extreme-value theory. *Quality and Reliability Engineering International, 19*(4), 337–353.
Wald, A. (1945). Sequential tests of statistical hypotheses. *Annals of Mathematical Statistics, 16*(2), 117–186.
Wald, A. (1947). *Sequential analysis*. New York: John Wiley & Sons.
Weiß, C. H., Testik, M. C., & Homburg, A. (2021). On the design of Shewhart control charts for count time series under estimation uncertainty. *Computers & Industrial Engineering, 157*(7), Article 107331.
Willis, A. J. (2011). Design of a modified sequential probability ratio test (SPRT) for pipeline leak detection. *Computers and Chemical Engineering, 35*(1), 127–131.
Zhang, M., Megahed, F. M., & Woodall, W. H. (2014). Exponential CUSUM charts with estimated control limits. *Quality and Reliability Engineering International, 30*(2), 275–286.
Zhang, P., Su, Q., Li, C., & Wang, T. (2014). An economically designed sequential probability ratio test control chart for short-run production. *Computers & Industrial Engineering, 78*(12), 74–83.