APEM journal
Advances in Production Engineering & Management
Volume 13 | Number 3 | September 2018 | pp 307–320
https://doi.org/10.14743/apem2018.3.292
ISSN 1854-6250
Journal home: apem-journal.org
Original scientific paper

Change-point estimation for repairable systems combining bootstrap control charts and clustering analysis: Performance analysis and a case study

Yang, Z.J.[a], **Du, X.J.**[a], **Chen, F.**[a,*], **Chen, C.H.**[a,*], **Tian, H.L.**[a], **He, J.L.**[a]

[a]School of Mechanical Science and Engineering, Jilin University, Changchun, P.R. China

ABSTRACT

Complex repairable systems with bathtub-shaped failure intensity will normally go through three periods in the lifecycle, which requires maintenance policies and management decisions accordingly. Therefore, the accurate estimation of change points of different periods has great significance. This paper addresses the challenge of change-point estimation in failure processes for repairable systems, especially for sustained and gradual processes of change. The paper proposes a sectional model composed of two non-homogeneous Poisson processes (NHPPs) to describe the bathtub-shaped failure intensity. In order to obtain the accurate change-point estimator, a novel hybrid method is developed combining bootstrap control charts with the sequential clustering approach. Through Monte Carlo simulations, the proposed change-point estimation method is compared with two powerful estimation procedures in various conditions. The results suggest that the proposed method performs effective and satisfactory for failure processes with no limits of distributions, changing ranges and sampling schemes. It especially provides higher precision and lower uncertainty in detecting small shifts of change. Finally, a case study analysing real failure data from a heavy-duty CNC machine tool is presented. The parameters of the proposed NHPP model are estimated. The change point of the early failure period and the random failure period is also calculated. These findings can contribute to determining the burn-in time in order to improve the reliability of the machine tool.

ARTICLE INFO

Keywords:
Change-point estimation;
CNC machine tools;
Non-homogeneous Poisson process (NHPP);
Statistical process control (SPC);
Bathtub-shape behaviour;
Clustering

**Corresponding author:*
chenfeicn@jlu.edu.cn
(Chen, F.)
cchchina@foxmail.com
(Chen, C.H.)

Article history:
Received 17 September 2017
Revised 27 August 2018
Accepted 29 August 2018

1. Introduction

In lifecycle reliability analysis, the failure patterns of complex repairable systems can be generally categorised as early failure period, random failure period and wear-out failure period, which have various failure mechanisms and performances. The failure intensity during the lifecycle usually follows a bathtub-shaped curve in practice. Monitoring the changing trend and determining the change point for different periods will provide reasonable guidance for health management and decision-making in time, such as early failure elimination experiment, maintenance strategy, etc., which also enables the improvement of reliability and efficiency.

Statistical process control (SPC) charts have been widely used in the process monitoring. When a signal is detected to be out-of-control, it may indicate a sudden change in quality or a transition from one state to the next. Therefore, engineers could be initiated to trace back the cause based on the signal and make appropriate adjustments in time. Many researches have been done for change-point estimation considering various change patterns in multivariate en-

vironment based on SPC theory. Amiri and Allahyari [1] made a literature review for the existing methods to detect the real time of change in control charts. Pignatiello and Samuel [2] adopted the maximum likelihood estimator (MLE) for a step change in CUSUM and EWMA control charts. Perry *et al.* [3, 4] considered various quality characteristics for change point estimation of different change types involving linear trend and monotonic change. Unknown-parameter change-point models were also defined for identifying changes in multivariate process [5-7].

In addition to these traditional approaches, artificial neural network (ANN), clustering techniques and other learning based methods have been combined with SPC charts for monitoring of complicated systems [8]. As for ANN methods, Amiri *et al.* [9] developed a probabilistic ANN procedure to estimate the change points of quality characteristics following normal distributions. Ahmadzadeh *et al.* [10, 11] combined ANN with multivariate exponentially weighted moving average (MEWMA) control charts to estimate the actual change points in multivariate process. Maleki *et al.* [12] proposed an ANN-based model to detect step changes considering the correlation between multivariate or attribute quality characteristics. Change point estimators were also identified based on ANN in five change patterns [13]. When it comes to clustering techniques, fuzzy-statistical clustering approach was adopted to estimate change points in various control charts with fixed or variable sample sizes [14, 15]. The step change, linear trend change and monotonic change are identified using clustering approach by [16-18], respectively. Other learning based methods, such as support vector machines, have also been incorporated into SPC to solve the change-point estimation problems [19, 20].

Despite the literature has paid large attention to the SPC methods applying to change-point detection in quality monitoring, they are mostly designed for identifying step changes in the process with a specific distribution. The study of change-point estimation in reliability monitoring is rarely considered, especially for the failure processes of complex repairable systems. Considering the characteristics of complex repairable systems, the study of this problem is unique in that:

- Due to the long life cycle and high reliability, the sample size of failure data is usually very small. The historical data that can be used for reference are also not good enough owing to different operating environments and failure mechanisms.
- The failure process always consists of multiple periods. Accordingly, the real-time monitoring of reliability is essential in order to adjust maintenance strategies in time. Otherwise, it may result in decreased productivity, increased cost, and even some catastrophic damage.
- The change pattern prefers to be identified as a trend showing a bathtub-shaped curve instead of a step change. An improved SPC chart is required to be sensitive to the gradual and sustained change.
- Since the change point simply detected from SPC charts could be relatively rough [10], a change-point model based on MLE can be developed to help improve the accuracy and stability of the estimation. It makes the existing SPC charts subjected to some specific distribution not applicable.

Therefore, the paper first develops a sectional NHPP model to describe the bathtub-shaped failure intensity in the lifecycle. The model considers systems with minimal maintenance especially subject to early failures, it is able to flexibly fit time-ordered failure data in the early failure period and the random failure period. Then, a novel hybrid estimation method integrating a bootstrap control chart with a sequential clustering approach is proposed. The method is superior in real-time monitoring the gradual and sustained trend of change. Moreover, based on MLE with strict statistical deduction, the proposed change-point estimation method not only improves the calculation efficiency, but also achieves good accuracy even in the case of small samples.

This paper is structured as follows. The proposed model is established in Section 2, whose bathtub-shaped failure intensity function is also discussed in this part. The following section integrates clustering method with SPC to explain the change-point estimation procedure. In Section 4, the performance of the proposed approach is assessed by a series of simulations, and a numerical application is presented. Finally, conclusions are made in Section 5.

2. The bathtub-shaped failure intensity model

An NHPP can be adopted to describe the failures of complex repairable systems under minimal maintenance strategy. The strategy assumes that maintenance time can be ignored and the system will only return to the same state as it was right before the failure occurred, which is more rational for reliability modelling of complex repairable systems. In practical work, maintenance actions often involve only some parts of the system, so the overall state of the system is not affected. Based on this, our bathtub-shaped failure intensity model is proposed under this assumption and set up from the NHPP.

The NHPP is also called the Weibull process [21] when it has a failure intensity function as

$$\omega(t) = \lambda\beta t^{\beta-1}, t \geqslant 0, \lambda > 0, \beta > 0 \tag{1}$$

where λ is the intensity parameter, β is the shape parameter, and t is the operating time.

Following the NHPP, the number of failures during the time interval $(t_1, t_2]$ equals to

$$W(t_1, t_2) = E(N(t_2) - N(t_1)) = \int_{t_1}^{t_2} \omega(\mu)d\mu \tag{2}$$

The same can be inferred that $W(0, t) = E(N(t))$ represents the expected number of failures through time interval $[0, t]$, thereby the cumulative failure intensity function is given by

$$W(t) = \int_0^{t_0} \omega(\mu)d\mu = \lambda t^{\beta} \tag{3}$$

Then the corresponding reliability function is derived as

$$R(t) = e^{-\lambda t^{\beta}} \tag{4}$$

As well as the cumulative density function

$$F(t) = 1 - R(t) = 1 - e^{-\lambda t^{\beta}} \tag{5}$$

And the probability density function

$$f(t) = \frac{dF(t)}{dt} = \omega(t)R(t) = \lambda\beta t^{\beta-1}e^{-\lambda t^{\beta}} \tag{6}$$

However, a single NHPP model can only illustrate the situation that failure intensity and failure time are strictly monotonic, which is unable to describe the non-monotonic trends of different life stages. Therefore, a sectional model of multiple NHPPs is required to fit the bathtub-shaped curve in order to describe the rules of failures in different failure periods, which could help to obtain the change point.

2.1 Two sectional NHPP model

Experts have developed various models to describe complex and diverse data distributions [22], most of them are constructed assuming that systems are non-repairable or 'repair as new'. It is not appropriate for the concern in our case. Based on this, we propose a sectional model involving two NHPPs, representing the early failure period and the random failure period, respectively.

$$\omega(t) = \begin{cases} \omega_1(t) = \lambda_1\beta_1 t^{\beta_1-1}, & 0 \leqslant t < t_0 \\ \omega_2(t) = \lambda_2\beta_2 t^{\beta_2-1}, & t_0 \leqslant t < \infty \end{cases} \tag{7}$$

We divide the operating time t of equipment into two intervals, t_0 on behalf of the division point. Where λ_1, β_1 denote the intensity parameter and the shape parameter of the failure intensity function for early failure period, λ_2, β_2 are for random failure period, when $\beta_1 = \beta_2$, the model will degenerate into a single NHPP, that is the reason $\beta_1 \neq \beta_2$ is defined.

The sectional NHPP model has a cumulative failure intensity function with the expression

$$W(t) = \begin{cases} W_1(t) = \lambda_1 t^{\beta_1}, & 0 \leqslant t < t_0 \\ W_2(t) = \lambda_1 {t_0}^{\beta_1} + \lambda_2 t^{\beta_2} - \lambda_2 {t_0}^{\beta_2}, & t_0 \leqslant t < \infty \end{cases} \tag{8}$$

When $0 \leqslant t < t_0$, it is called the early failure period with the cumulative failure intensity $W(t) = W_1(t)$, when $t_0 \leqslant t < \infty$, it is called the random failure period with the cumulative failure intensity $W(t) = W_2(t)$, where $W_1(t)$ and $W_2(t)$ are two-parameter NHPP models, respectively.

2.2 Characterisation of bathtub-shaped failure intensity

For the failure intensity functions of the two sectional NHPP model, different values of the shape parameter β are used to denote different periods of the product's life. As shown in Fig. 1, when $\beta < 1$, the failure intensity decreases with time t, it means that failures will happen more and more infrequently with system ageing. When $\beta > 1$, the failure intensity increases as t progresses, it means that the failure rate will rise with system ageing. When $\beta = 1$, the failure intensity is a constant.

Accordingly, the cumulative failure intensity, which is also known as the cumulative failure number, has a change pattern seen in Fig. 1. Through further mathematical derivation, the conclusion agrees with the trend of failure intensity, it indicates that the growth rate of cumulative failure number will first reduce with time in early failure period, when it comes to random failure period, the growth rate will maintain a steady state, after that it will continue to rise during wear-out failure period.

In view of our proposed sectional model, it is aimed at describing the changing trend of the early failure period and the random failure period. Through setting the constraints that failure intensity and cumulative failure intensity are both continuous at t_0, the accurate change point t_0 could be obtained. which is expressed as

$$\begin{cases} \omega_1(t_0) = \omega_2(t_0) \\ W_1(t_0) = W_2(t_0) \end{cases} \tag{9}$$

Substitute Eq. 7 and Eq. 8 in Eq. 9

$$\begin{cases} \lambda_1\beta_1 {t_0}^{\beta_1 - 1} = \lambda_2\beta_2 {t_0}^{\beta_2 - 1} \\ \lambda_1 {t_0}^{\beta_1} = \lambda_1 {t_0}^{\beta_1} + \lambda_2 {t_0}^{\beta_2} - \lambda_2 {t_0}^{\beta_2} \end{cases} \tag{10}$$

Since the cumulative failure intensity function is much in evidence to be continuous at t_0 with the sectional two NHPP modeling, the change point t_0 can be obtained as

$$t_0 = \left(\frac{\lambda_1\beta_1}{\lambda_2\beta_2}\right)^{\frac{1}{\beta_2 - \beta_1}} \tag{11}$$

At this point, the bathtub-shaped failure intensity and its change point can be achieved with our proposed sectional model, which can be adopted to assist the failure process monitoring.

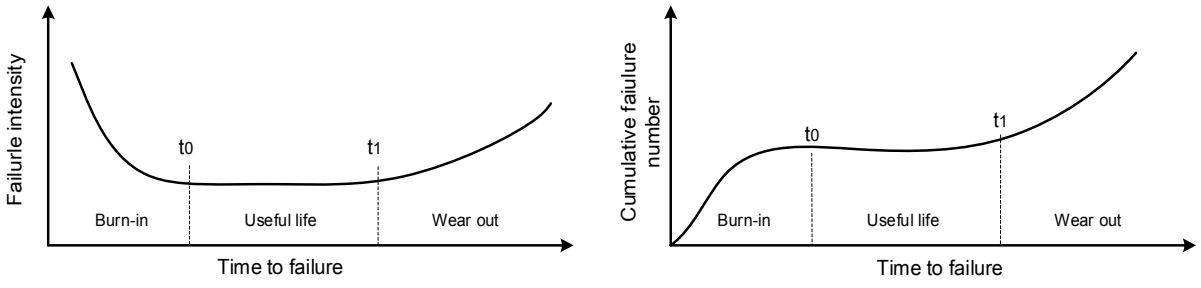

Fig. 1 Bathtub-shaped curves of the failure intensity and corresponding cumulative failure number

3. Change-point estimation combined SPC with sequential clustering

For the purpose of estimating the exact change point, a specific bootstrap control chart to monitor the performance of repairable systems is established. The objective is to track possible transition points by detecting all out-of-control signals during the failure process. These candidate points help to classify data patterns into the two phases of early failure period and random failure period, respectively. Then, the clustering techniques are adopted to eliminate interference data within out-of-control signals in order to identify the optimum transition point. Finally, the accurate change point can be estimated by fitting the two sections of our proposed failure intensity model with the segmented observation data.

The proposed approach integrates the advantages of SPC method and clustering analysis. Considering the characteristics of SPC, two possible states are first predetermined as in-control and out-of-control, it will simplify the clustering model with a known number of clusters. Meanwhile, the data is monitored in time series, the order of points are preserved for the iteration to search for the optimal clusters. At last, cluster settings will only be examined when an out-of-control signal is detected. The number of iterations is dramatically reduced to save computing time. Therefore, the SPC method is beneficial for improving efficiency for change-point estimation. Additionally, the intervention of sequential clustering approach is for benefits of lower interference and better accuracy, because the SPC chart in this study is required to be more sensitive to sustained shifts and gradual changing trends rather than sudden changes or random noises.

3.1 The bootstrap control chart

Traditional control charts [23], like standard Shewhart charts, can only be strictly applied to the normal distribution by monitoring the shifts of mean and variance. As to those modified CUSUM charts and EWMA charts [24], they are always set up for some specific distributions. In our case, the failure process is modelled with a sectional NHPP, it's difficult to find the applicable control charts and corresponding methods to establish its control limits. Therefore, a bootstrap control chart with no limits of any specific distribution is developed.

The particular SPC chart is established based on Monte Carlo simulation [25], which can be implemented as the following steps shown in Fig. 2.

Step 1: Determine the stable values of parameter λ_s and β_s. They can be defined by experience or MLEs with observations collected in Phase I, which represents a stable state in SPC theory.

Step 2: Generate bootstrap random variables $t_{i1}, t_{i2}, \dots, t_{in}$ from the NHPP model with parameters λ_s and β_s in size n. n equals to the data volume of Phase II, which is the subsequent monitoring process based on the estimated control limits. The method from [26] is improved to generate random variables of failure times $t_{i1}, t_{i2}, \dots, t_{in}$ following the NHPP sequentially.

Step 3: Calculate the MLEs from $t_{i1}, t_{i2}, \dots, t_{in}$ to obtain λ_i and β_i of the NHPP models.

Step 4: Obtain the p-quantile W_{pi} with $W_{pi} = (-(1/\lambda_i)\ ln\ p)^{1/\beta_i}$, where W_{pi} represents 100 p-th percentile of interest.

Step 5: Repeat the steps from Step.2 to Step.4 for $B(B > 1000)$ times to obtain B groups of p-quantiles. Order them from the smallest to the largest as $W_{p1} < W_{p2} < \cdots < W_{pB}$. Then the centre line (CL) is yielded as a mathematical expectation of acquired data, the lower control limit (LCL) is the i-th p -quantile W_{pi}, and the upper control limit (UCL) is the $(B - i)$-th value as $W_{p(B-i)}$, where $i = \alpha B/2$, representing that there're i estimators beyond the control limits. The false alarm risk α indicates the probability when the system is diagnosed to be out of control while it's actually in control.

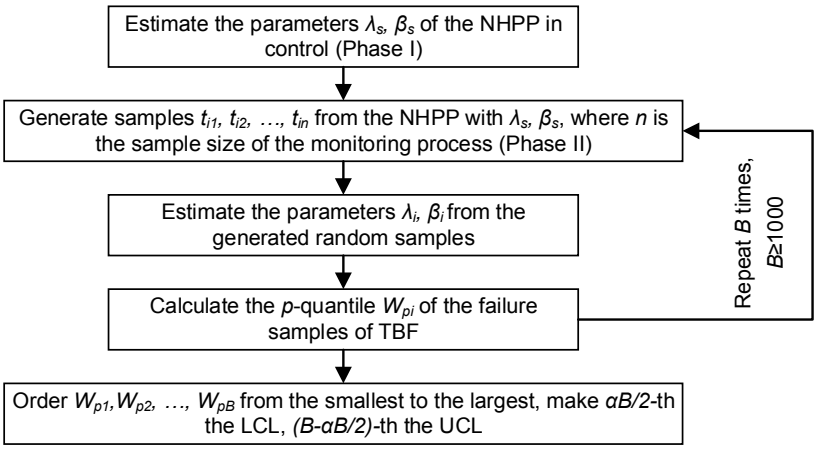

Fig. 2 Procedure to establish the bootstrap control chart

After control limits are obtained, the bootstrap control chart can be established through online monitoring. Time between failures (TBF) observations are plotted after each failure, and process shifts can be detected under some suitable runs rules.

Take an instance as shown in Fig. 3, we assume there are 200 observations in total, the first 100 points in Phase I are supposed to be in control. Then the control limits, including UCL, LCL and CL, are calculated following the steps from Step 1 to Step 5. When it's converted to Phase II, the other 100 points are plotted on the established control chart, which should have been detected to be out of control as we expected. However, the fact in Fig. 3 shows that several observations (red solid points) in Phase I are beyond the control limits, meanwhile, a large number of observations (black solid points) in Phase II are within the control limits. It infers that the proposed control chart has a problem of random interference, the accuracy of detection will be reduced and it would be hard to determine which signal is the optimal transition point. In view of this problem, the sequential clustering approach is introduced to combine with the proposed SPC method.

3.2 The sequential clustering approach

The purpose of introducing clustering approach is to remove the disturbance in SPC and extract the optimal point $\hat{\tau}$, it can detect whether systems convert to another failure period and divide observations in two phases.

In the proposed sequential clustering approach, two possible clusters are defined: for the i-th out-of-control signal O_i, all observations before O_i are classified to the in-control cluster, and all observations after O_i are automatically classified to the out-of-control cluster. In the interest of obtaining the optimal transition point $\hat{\tau}$, first of all, the clustering model with applicable validity indices is developed. Then, the objective function of each cluster setting according to time series is examined. At last, the optimal transition point $\hat{\tau}$ is obtained by optimising the objective function. Two validity indices and the objective function from [16] are illustrated and improved as follows.

Clusters within variation

A cluster within variation V_{within} is expressed as the distance between observations and their cluster centres, given by C_{in} of in-control cluster centre and C_{out} of out-of-control cluster centre, as shown in Fig. 4. In consideration of Phase I and Phase II based on the specific bootstrap control chart, C_{in} in Phase I equals to the in-control mean. C_{in} in Phase II is substituted as the expectation of the change-point model, which is also defined as the CL of established control chart.

Therefore, V_{within} in Phase I is

$$V_{within} = \sum_{i=1}^{\tau} (\overline{X}_i - C_{in})^2 + \sum_{i=\tau+1}^{n} (\overline{X}_i - C_{out})^2 \tag{12}$$

where $\overline{X}_i$ is the average value of i-th subgroup of observations, and

$$C_{in} = \sum_{i=1}^{\tau} \frac{\overline{X}_i}{\tau}, \quad C_{out} = \sum_{i=\tau+1}^{n} \frac{\overline{X}_i}{n-\tau} \tag{13}$$

What's more, V_{within} in Phase II is

$$V_{within} = \sum_{i=1}^{\tau} (\overline{X}_i - CL)^2 + \sum_{i=\tau+1}^{n} (\overline{X}_i - C_{out})^2 \tag{14}$$

Clusters between variation

A cluster between variation is expressed as the distance between cluster centres and the total cluster centre of all observations C_T, as shown in Fig. 4. Similar with the situation of clusters within variation V_{within}, C_T is substituted as the CL of bootstrap control chart in Phase II, so as to obtain the expression of $V_{between}$ in Phase I as

$$V_{between} = \tau(C_{in} - C_T)^2 + (n-\tau)(C_{out} - C_T)^2 \tag{15}$$

where $C_T = \sum_{i=1}^{n} \overline{X}_i / n$ and $V_{between}$ in Phase II is

$$V_{between} = \tau(C_{in} - CL)^2 + (n-\tau)(C_{out} - CL)^2 \tag{16}$$

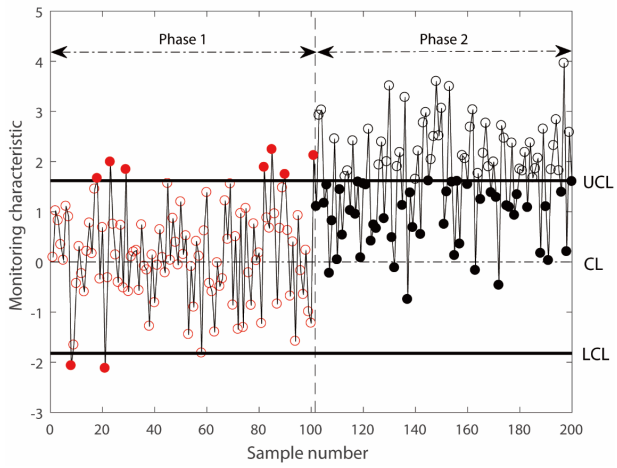

Fig. 3 An example of established bootstrap control chart

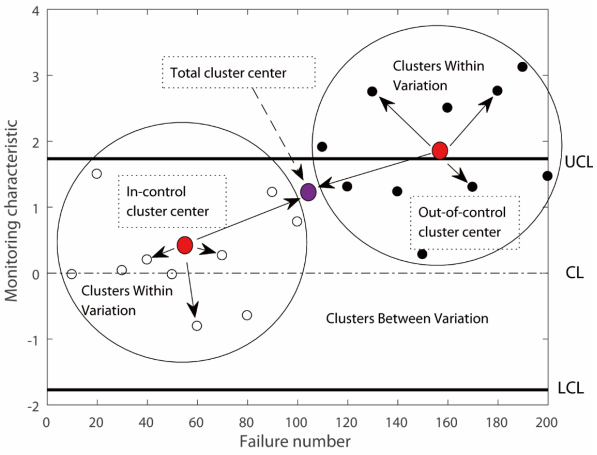

Fig. 4 Illustration of clusters within and between variation

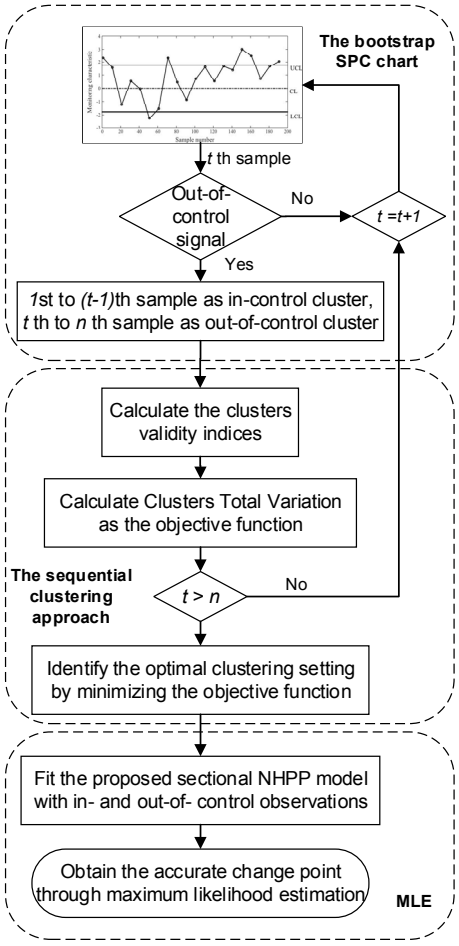

Fig. 5 Change-point estimation procedure

The objective function

Clusters total variation V_{total} is defined as the objective function. It integrates the within and between variations for the purpose of valuating different cluster settings more completely and getting the optimal clusters more efficiently. The objective function has the expression as

$$V_{total} = V_{within} + V_{between} \tag{17}$$

For the classification in Phase I, it is derived from Eq. 12 and Eq. 15

$$V_{total} = \sum_{i=1}^{\tau} (\bar{X}_i - C_{in})^2 + \sum_{i=\tau+1}^{n} (\bar{X}_i - C_{out})^2 - \tau(C_{in} - C_T)^2 + (n-\tau)(C_{out} - C_T)^2 \tag{18}$$

What's more, for the classification in Phase II, Eq. 14 and Eq. 16 is combined as

$$V_{total} = \sum_{i=1}^{\tau} (\bar{X}_i - CL)^2 + \sum_{i=\tau+1}^{n} (\bar{X}_i - C_{out})^2 - \tau(C_{in} - CL)^2 + (n-\tau)(C_{out} - CL)^2 \tag{19}$$

To find out the optimal transition point $\hat{\tau}$ and assign observations to in- and out-of-control clusters, the proposed objective function should be minimised.

$$\hat{\tau} = argmin\, V_{total}(\tau) \tag{20}$$

3.3 Change-point estimation procedure

The overall structure of proposed change-point estimation procedure is summarised as Fig. 5. First of all, the bootstrap SPC chart is constructed, possible combinations of in- and out-of-control clusters are classified by each out-of-control signal. Then, the sequential clustering approach is adopted, proposed validity indices are examined sequentially for each clustering setting and the objective function is optimised to obtain the best assignment of observations. Finally, the best assignment of observations is used to fit the proposed two sectional NHPP model, the more precise change-point estimator could be calculated through maximum likelihood estimation.

As for the part of maximum likelihood estimation, let $f(t_1, t_2, \dots, t_{\hat{\tau}}, \dots, t_n)$ denotes the joint probability density of failure times $0 < t_1 < \cdots < t_{\hat{\tau}} < \cdots < t_n$, where n is the random number of failures. Among them, $t_1, t_2, \dots, t_{\hat{\tau}}$ belong to the optimal in-control cluster, and $t_{\hat{\tau}}, t_{\hat{\tau}+1}, \dots, t_n$ belong to the optimal out-of-control cluster. The joint probability density can be constructed with a failure intensity $\omega(t)$ in Eq. 7 and a cumulative failure intensity $W(t)$ in Eq. 8.

$$f(t_1, t_2, \dots t_{\hat{\tau}}, \dots, t_n) = \begin{cases} (\lambda_1\beta_1)^{\hat{\tau}} \prod_{i=1}^{\hat{\tau}} t_i^{\beta_1 - 1} e^{-\lambda_1 t_{\hat{\tau}}^{\beta_1}}, & 0 < t_1 < t_2 < \cdots < t_{\hat{\tau}} \\ (\lambda_2\beta_2)^{n-\hat{\tau}} \prod_{i=\hat{\tau}+1}^{n} t_i^{\beta_2 - 1} e^{-\lambda_2 t_n^{\beta_2} + \lambda_2 t_0^{\beta_2} - \lambda_1 t_0^{\beta_1}}, & t_{\hat{\tau}} < \cdots < t_n \end{cases} \tag{21}$$

The log-likelihood function can be obtained as

$$\begin{aligned} \ln(L) = \hat{\tau}(\ln\lambda_1 + \ln\beta_1) + (\beta_1 - 1)\sum_{i=1}^{\hat{\tau}} \ln t_i - \lambda_1 t_{\hat{\tau}}^{\beta_1} + (n - \hat{\tau})(\ln\lambda_2 + \ln\beta_2) \\ + (\beta_2 - 1)\sum_{i=\hat{\tau}+1}^{\mathrm{n}} \ln t_i - \lambda_2 t_n^{\beta_2} + \lambda_2 t_0^{\beta_2} - \lambda_1 t_0^{\beta_1} \end{aligned} \tag{22}$$

Differentiate log-likelihood functions with respect to $\lambda_1, \beta_1, \lambda_2, \beta_2$, making the equations result in zero, the MLEs of parameters will yield from

$$\begin{cases} \dfrac{\partial \ln(L)}{\partial \lambda_1} = \dfrac{\hat{\tau}}{\lambda_1} - {t_{\hat{\tau}}}^{\beta_1} - {t_0}^{\beta_1} = 0 \\ \dfrac{\partial \ln(L)}{\partial \beta_1} = \dfrac{\hat{\tau}}{\beta_1} + \sum_{i=1}^{\hat{\tau}} \ln t_i - \lambda_1 {t_{\hat{\tau}}}^{\beta_1} \ln(t_{\hat{\tau}}) - \lambda_1 {t_0}^{\beta_1} \ln(t_0) = 0 \\ \dfrac{\partial \ln(L)}{\partial \lambda_2} = \dfrac{n-\hat{\tau}}{\lambda_2} - {t_n}^{\beta_2} - {t_0}^{\beta_2} = 0 \\ \dfrac{\partial \ln(L)}{\partial \beta_2} = \dfrac{n-\hat{\tau}}{\beta_2} + \sum_{i=\hat{\tau}+1}^{n} \ln t_i - \lambda_2 {t_n}^{\beta_2} \ln(t_n) + \lambda_2 {t_0}^{\beta_2} \ln(t_0) = 0 \end{cases} \tag{23}$$

With the constraint of Eq. 11 the MLE of change point is obtained as

$$t_0 = argmax\ L(t) \tag{24}$$

4. Performance analysis and a case study

In this section, the performance of the proposed change-point estimation approach is evaluated and compared with some existing methods through Monte Carlo simulations. To demonstrate feasibility and accuracy in industrial applications, the proposed two sectional NHPP model with the estimation process is applied to analyse the field data of a heavy-duty CNC machine tool.

4.1 Performance comparison of the proposed approach with existing estimation procedures

Three series of simulations considering phases I and II of different distributions with various changing shifts are conducted. In these simulations, the evaluation results of the proposed objective function are compared with the results of methods in [16, 27].

The scheme of simulation study is constructed as follows.

Series 1 simulation

In each simulation run, samples of size $m = 1$ are generated from a normal distribution with $\mu = 0$ and $\sigma = 1$ up to the real change point $\tau = 100$ for Phase I. The bootstrap control chart is constructed following the procedure in Section 3.1 with these in-control samples. Then, for Phase II, with a shift in $\mu = 0.5, 1, 1.5, 2, 3$, respectively, samples are generated and monitored on the established control chart. At last, the proposed indices are calculated to estimate the change point. For each adjusted parameter, average change point and corresponding standard error of 10000 simulation runs are computed. The results are tabulated in Table 1 with the comparing results from [16, 27].

Conclusions can be drawn that the proposed method along with the other two comparative methods have quite close monitoring results, which are also consistent with the real change point. Meanwhile, with the increase of the parameter adjustment value, the precision of the determined out-of-control signal is improved. Among them, Ghazanfari *et al.*'s method works a little better than the other two, especially for small shifts. Despite the proposed method's slightly inferior performance with the change-point estimates, it is more stable and the results obtained are more accurate and reliable considering the much smaller standard errors.

Table 1 Average change-point estimates and associated standard errors in Series 1 simulation

Method		Shift size				
		0.5	1	1.5	2	3
Samuel *et al.* [27]	$\tilde{\tau}$	104.45	100.39	99.94	99.7	99.61
	$std(\tilde{\tau})$	23.07	7.15	3.93	3.71	3.49
Ghazanfari *et al.* [16]	$\tilde{\tau}$	103.24	100.15	99.69	99.53	99.33
	$std(\tilde{\tau})$	23.28	7.81	5.97	5.45	5.95
Proposed method	$\tilde{\tau}$	104.59	102.71	101.66	101.33	101.1
	$std(\tilde{\tau})$	17.98	3.44	1.44	0.85	0.37

Series 2 simulation

In the second series of simulation, generate samples of size $m = 1$ from an Exponential distribution with $\lambda = 2$ as the first 100 observations in Phase I. Then, define the altering parameters set as $\lambda + \delta$ ($\delta = 0.5, 1, 1.5, 2, 3$) for the next 100 samples. The change point is estimated for 10000 times in the case of each parameter. The results of average change point and standard error are listed in Table 2 to compare the performance of three methods.

As given in Table 2, the performances of three methods are essentially in agreement with *Series 1 simulation* for Normal distributions. By comparing the results, it illustrates that the proposed method performs best when the shift size of the parameter is small, while its estimated change point tends to be stable under increasing parameter adjustments.

Table 2 Average change-point estimates and associated standard errors in Series 2 simulation

Method		Shift size				
		0.5	1	1.5	2	3
Samuel *et al.* [27]	$\tilde{\tau}$	113.33	104.69	102.54	101.73	101.05
	$std(\tilde{\tau})$	16.81	6.89	4.66	4.02	2.51
Ghazanfari *et al.* [16]	$\tilde{\tau}$	112.83	104.51	102.6	101.76	101.02
	$std(\tilde{\tau})$	15.97	6.8	4.21	3.81	3.23
Proposed method	$\tilde{\tau}$	109.15	104.95	103.96	103.64	103.2
	$std(\tilde{\tau})$	13.65	6.51	4.97	4.53	3.58

Series 3 simulation

At last, simulations are executed for the situation that failure process following Normal distributions with a small amount of samples. Subgroups of size $m = 4$ for Phase I are generated from a Normal distribution with $\mu = 100$ and $\sigma = 5$ until the real change point $\tau = 10$. Then from another Normal distribution with a shift in μ as $\mu = \mu + \sigma \times \delta$ ($\delta = 0.5, 1, 1.5, 2, 3$), samples with a size of 10 for Phase II are generated. The simulation procedure for each parameter μ is repeated over 10000 times, and the estimated results are compared with other methods [16, 27] in Table 3.

The simulation considers the situation when there is no sufficient amount of observations. We can tell from the results that our proposed method still behaves well with a small sample size for all levels of alterations. To be more specific, it is more closed to the real change point compared with the other two methods [16, 27] especially for small shift sizes of parameter μ. Even if the results for larger shift sizes are slightly rougher than the others, the standard errors of the proposed method are much smaller, which means the performance is more stable and the obtained results are considerably consistent.

Table 3 Average change-point estimates and associated standard errors in Series 3 simulation

Method		Shift size				
		0.5	1	1.5	2	3
Samuel *et al.* [27]	$\tilde{\tau}$	13.51	11.54	9.76	10.39	10.12
	$std(\tilde{\tau})$	3.78	2.37	1.43	0.9	0.39
Ghazanfari *et al.* [16]	$\tilde{\tau}$	10.38	10.04	10	10	10
	$std(\tilde{\tau})$	4.1	1.56	0.59	0.27	0.05
Proposed method	$\tilde{\tau}$	10	10.95	11.02	11.02	11
	$std(\tilde{\tau})$	3.55	1.15	0.43	0.16	0.03

4.2 Numerical application of a heavy-duty CNC machine tool

The case of a heavy-duty CNC machine tool from some factory is studied. The heavy-duty CNC machine tool is a typical kind of complex repairable system, it has the features such as small batch of production, long service life, and high maintenance cost. Therefore, the concern of the factory is focused on ensuring the reliability of the service period and reducing maintenance cost. To achieve this, in practice of engineering, a maintenance policy of 'minimal repair' is always implemented to minimize the impact of maintenance, burn-in is also a common mean to improve the reliability and maximize the useful life.

Table 4 Field data of TTF and TBF of the heavy-duty CNC machine tool (Unit: hour)

No.	TTF	TBF	No.	TTF	TBF	No.	TTF	TBF	No.	TTF	TBF
1	10	10	8	480.5	43.5	15	747.08	18.5	22	1831.17	375
2	178	160	9	523.5	38	16	848.58	99.5	23	1905.17	72
3	230	50	10	576.5	50	17	919.17	68.5	24	2060.17	149
4	276.5	44.5	11	618.5	40	18	1084.17	163	25	2336.17	262
5	297	17.5	12	649.5	29	19	1227.17	107.5	26	2429.67	71.5
6	385.5	77	13	683	24.5	20	1376.17	147	27	2548.67	111.5
7	435	42.5	14	727.08	42	21	1454.17	74	28	2745.17	182

Based on these facts, this case study aims to estimate the accurate change point of early failure period and random failure period in order to determine the optimal burn-in time of the product. Then take account of the effect of maintenance and the lack of failure data, our proposed method is particularly suitable for this situation. The field monitoring was conducted for 3000 hours since it initially finished production and came into early failure testing, the time to failure (TTF) data of the total 28 failures were collected and listed in Table 4.

Choose TBF as monitoring characteristic, the bootstrap control chart is constructed. Compared with the traditional Weibull probability plot, the SPC chart maintains a time sequence of TBFs and realizes real-time condition monitoring. As shown in Fig. 6, basically the first 15 data are below the LCL ignoring the randomness of the observed point 2 and point 6, it illustrates that failures happened frequently and the machine tool was still in the early failure phase. After the 15th data, observations mainly kept stable around the CL with no signal outside the control limit, which is interpreted as the state converted into the random phase with stable failure intensity.

However, the error caused by unsteadiness and uncertainty of samples, such as a sudden change at point 2 and point 6, may affect the accuracy and reliability of SPC monitoring. Then the sequential clustering approach is adopted to obtain the accurate transition point for the optimal clustering setting. Following the procedure in Section 3.2, the validity indices are calculated each time when the observation is above the LCL, and the proposed objective function turns out to be minimized at point 17, which has been marked with a red circle in Fig. 6.

As a result, the observations are divided into two optimal groups taking 17th point as the boundary. Substitute them into two sections of the proposed NHPP model as Eq. 21 and the MLEs of parameters are obtained as $\lambda_1 = 0.8045, \beta_1 = 0.3819, \lambda_2 = 5.7262 \times 10^{-6}, \beta_2 = 1.8648$.

Substitute the results above in Eq. 11 to calculate the change point t_0, which connects the early failure period and the random failure period. The result is $t_0 = 1.0161 \times 10^3\ h$.

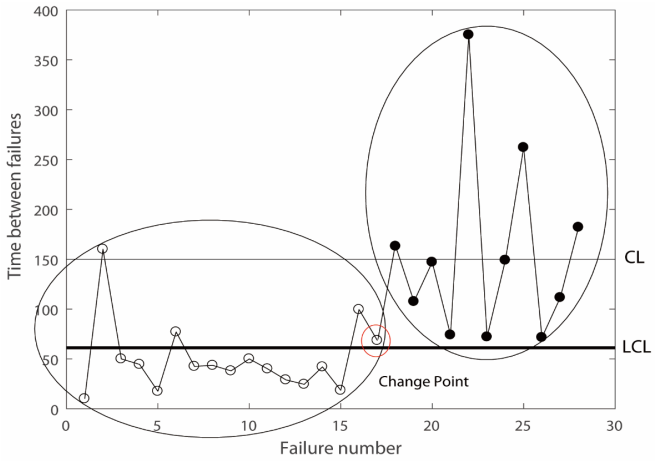

Fig. 6 Bootstrap control charts with the optimal clusters

Table 5 Definitions of goodness-of-fit indices

Index	Expression	Variable
K-S	$\max\left\{\max_{1\le i\le M}(\frac{i}{M}-\widehat{U}_i),\max_{1\le i\le M}(\widehat{U}_i-\frac{(i-1)}{M})\right\}$	L: The log-likelihood function, $t_1, t_2, \dots t_n$: The failure times, k: The number of parameters in the model, $M=n-1$, $\beta=\frac{n}{\sum_{i=1}^{n-1}\ln(\frac{t_n}{t_i})}$, $\widehat{U}_i=(\frac{t_i}{t_n})^{\beta}, i=1,\cdots,M.$
A-D	$-\frac{\{\sum_{i=1}^{M}(2i-1)[\ln\widehat{U}_i+\ln(1-\widehat{U}_{M+1-i})]\}}{M}-M$	
C-V	$\frac{1}{12M}+\sum_{i=1}^{M}[\widehat{U}_i-\frac{(2i-1)}{2M}]^2$	
AIC	$-2\ln(L)+2k$	

Table 6 Comparing results of the goodness-of-fit test

Method	Change point	Parameter estimator	K-S	A-D	C-V	AIC
Original model	-	(0.0796, 0.7405)	0.1849 (0.20)	1.0727 (1.33)	0.2196 (0.23)	313.982
Proposed model	Phase 1 (1:17)	(0.8045, 0.3819)	0.1502 (0.28)	0.3964 (1.36)	0.0452 (0.24)	191.0966
	Phase 2(18:28)	(5.7262e-06, 1.8648)	0.184 (0.34)	0.3928 (1.39)	0.0489 (0.25)	160.5379

After obtaining the estimated change point and parameters corresponding to the optimal in-control and out-of-control clusters, the goodness-of-fit test is adopted. The Kolmogorov-Smirnov (K-S) statistic, the Anderson-Darling (A-D) statistic, the Cramér–von Mises (C-V) statistic, and the Akaike information criterion (AIC) are used as goodness-of-fit indices, which are defined in Table 5. If the calculated indices have smaller values, then the model is accepted as a better fitting result of the data. The results of the proposed two sectional NHPP model and the original NHPP model are listed in Table 6 for comparison. It's easy to tell that the values of three statistics and AIC of our new model are all smaller than the original model, the results indicate that our new model together with the change-point estimation approach can hence provide a better fit to the lifetime data, the estimated change point of the early failure period and the random failure period results in more consistent with the engineering practice, the practicality and accuracy of the proposed method are verified.

For the new products just involved in manufacturing work, they are more sensitive to the early failures, and their early phase in lifecycle contains a large proportion of failures as well, for the design, manufacturing or assembling reason. Burn-in is a popular process to eliminate early failures and improve the reliability. However, if the burn-in lasts too long, the useful life of products will be reduced. Otherwise, the reliability will decrease and the maintenance cost will increase. In practical execution, our proposed approach can present a simple and effective way to detect the changing trend of reliability and identify the optimal burn-in time, which will be of great value.

5. Conclusion

The paper proposed a change-point estimation method to approximate the failure process in lifecycle and estimated the change points of different failure periods for complex repairable systems. In the proposed method, a sectional model involving two NHPP functions was first developed to describe the bathtub-shaped failure intensity. Then, a bootstrap SPC chart was proposed in order to monitor changing trends of reliability and estimate the real time of change point. By introducing a sequential clustering approach in the control chart, the random interference caused by traditional SPC is eliminated and the calculation efficiency is improved. Three series of simulation considering the processes with different distributions, changing ranges and sampling schemes were conducted. The proposed method was proved to have a considerably good per-