A preventive maintenance policy and a method to approximate the failure process for multi-component systems

Shaomin Wu*

Kent Business School, University of Kent, Canterbury, Kent CT2 7PE, UK

Majid Asadi

Department of Statistics, Faculty of Mathematics and Statistics, University of Isfahan, 81746-73441, IRAN

Abstract

Numerous maintenance policies have been proposed in the reliability mathematics and engineering literature. Nevertheless, little has been reported on their practical applications in industries. This gap is largely due to restrictive assumptions of the maintenance policies. Two of the main assumptions are that maintenance is conducted on typical components and that the reliability of an item under maintenance is known (where the item can be a component or a system composed of multiple components). These assumptions do not often hold in the real world: maintenance is often performed on a collection of components such as an integrated circuit plate and the reliability of each individual component may not be known. To reduce these gaps, this paper develops a new maintenance policy for a collection of components and an approximate method to estimate the reliability of this collection based on the failure data collected from the field. The maintenance policy considers that a system is composed of three subsystems with different levels of maintenance effectiveness (i.e, minimal, imperfect, and perfect). The approximate estimate of the reliability of each subsystem is derived based on the failure data that are time between failures of the system but not those of the components that cause the system to fail. An algorithm for simulating the superposition of generalised renewal processes is then proposed. Numerical examples are used to illustrate the proposed approximation method.

Keywords: (T) Maintenance policy, Stochastic processes; non-homogeneous Poisson process (NHPP); generalised renewal process (GRP); superposition of generalised renewal processes (SGRP).

1 Introduction

1.1 Motivation

In the literature, many papers on maintenance policy optimisation have been published (see Wu et al. (2018); de Jonge and Scarf (2020); Peng et al. (2022); Dui et al. (2023); Xu et al. (2023)

* *Corresponding author. E-mail: s.m.wu@kent.ac.uk. Telephone: +44 (0)1227 827940*

Suggested citation: Shaomin Wu, Majid Asadi, A preventive maintenance policy and a method to approximate the failure process for multi-component systems, European Journal of Operational Research, DOI: 10.1016/j.ejor.2024.05.039

for example). They normally assume that the technical system under study is composed of only *one component*. The repair effectiveness on a one-component system is then assumed one of the three possible results: minimal, imperfect, and perfect. A *minimal repair* restores the system to the status that is just immediately before the system fails, i.e., an as-bad-as-old status, a *perfect repair* brings the system back to the as-good-as-new state, and an *imperfect repair* restores the system to a status in between as-bad-as-old statuses and as-good-as-new. The assumption of the one-component system drastically simplifies real-world systems, most of which are composed of multiple components and subsystems. As such, the outcomes of the optimisation of maintenance policies based on the one-component assumption devoice from the ground truth, which can be one of the main reasons that few real-world examples of using maintenance policies derived from the literature of reliability are reported. Take cars as an example. Some components need minimal repair (e.g., repairing a punctured tyre can be regarded as a minimal repair); some components in a car need replacement upon their failures (e.g., the broken windscreen in a car must normally be replaced), and the repair on some components is imperfect (e.g., the failure of the engine in a car can be restored to status between good as new and bad as old statuses, which can be regarded as an imperfect repair). That is, a car can be regarded as a system composed of subsystems of three types, which can be restored with three types of repair effectiveness upon failures, respectively. From a mathematical modelling perspective, to model different levels of repair, stochastic processes are used. Usually, the non-homogeneous Poisson process (NHPP) models that of a system with minimal repair, the renewal process (RP) is used to model the failure process of a system with perfect repair and a stochastic process that can model all three situations is referred to as the generalised renewal process (GRP). Many GRP models have been developed in the literature, for example, the arithmetic reduction of age (ARA) model (Doyen and Gaudoin, 2004), the arithmetic reduction of intensity (ARI) model (Doyen and Gaudoin, 2004), the geometric process and its extensions (Wu, 2022a).

One can argue that the system can still be regarded as a one-component system and the repair effectiveness can be regarded as an imperfect repair. This treatment can be used only if there is little information regarding the failure process of each subsystem (Wu, 2019). Modelling the failure process on the system level will end up with a large standard error of the parameter estimates in the failure process model and can therefore cause huge uncertainty in the optimisation of maintenance policies. As such, there is a need to treat such a system composed of three subsystems and maintenance policies are developed when the three failure types are considered.

There are still two challenges. On the one hand, the items of modern times become more reliable than before and the number of failures of each individual component is so small that it cannot be sufficient for estimating its reliability. On the other hand, it is often not possible to know which component causes its system to fail, which is especially the case because many systems are composed of integrated electronic subsystems and the causes of the failures are not easy to know. For example, Jones and Hayes (2001) believes that the no-fault-found (NFF) failures contribute on average to 45% of reported service faults in electronic products. A more recent report shows that NFF observations can be high up to 85% in avionics in Boeing 777 electronics and can cost up to 90% of total maintenance cost (Singh et al., 2021). That is, the failure causes of a significant proportion of items cannot be found. As such, there is a need to develop an approach to modelling the failure process for the system with unknown causes of the system failures, as discussed below.

It is known that many real-world technical systems are multi-component systems and are structured in series. If a component fails, it is repaired immediately and the system operation resumes. The time on repair is negligible. Then the failure process of the system is the superposition of generalised renewal processes (or superimposed generalised renewal process, SGRP), which is the union of the generalised renewal processes. To estimate the parameters in the SGRP based on failure data, we need both failure data (times-between-failures) of each component and to know which components, i.e., the labels of the components, cause the system to fail. Assuming that the times between failures are available but which components cause the system to fail are unknown, then the failure data are referred to as *masked failure data*. It is not possible to obtain the SGRP for the system if only masked failure data are available. Under such a scenario, one can only develop methods to approximate the SGRP. It is often the case that only masked failure data from the fields such as warranty claim data are available for model estimation. Additionally, SGRP has many real applications, including queueing systems in the healthcare sector (He, 2020), failure data modelling in reliability engineering (Wu, 2019, 2020), and traffic analysis (Singh and Chaudhury, 2017). Hence, it is important to obtain the SGRP or develop a method to approximate the SGRP. As such, in the literature, many authors are studying the SGRP (Cox and Smith, 1954; Khinchin, 1956; Drenick, 1960), as reviewed in Section 1.2. In the literature, there is little research on the SGRP. Obviously, the superposition of renewal processes (SRP) is a special case of the SGRP. This paper therefore aims to obtain the lower and upper bounds of the rate of the SRP, extends the results to the SGRP situation, and then uses a weighted linear combination of the bounds as an approximation of the SGRP. It should be highlighted that approximating approaches to the SGRP can be utilised not only in reliability engineering but also in other areas such as inventory management and neurophysiology.

1.2 Prior work and comments

Regarding maintenance policies, the research is abundant. The reader is referred to recently published papers for more references on this topic (de Jonge and Scarf, 2020; Zhang and Li, 2022; Asadi et al., 2023; Brenière et al., 2023; Cai et al., 2023; van Staden et al., 2022), from which one can see that the development of preventive maintenance policies has been a focus in the literature on reliability and maintenance for decades. Conventionally, technical systems under the consideration for maintenance are assumed to be one-component systems or multi-component systems with a type of dependence (Nicolai and Dekker, 2008; Andersen et al., 2022). From the modelling perspective, models for describing the failure process of a multi-component system is extremely complicating. For instance, if the system under consideration is a series system, even if we assume that each failed component is repaired as good as new, the final model for the system is a superimposed renewal process. If we further consider that the costs of repairing each component (which is more realistic) are different, then the maintenance policy becomes very complicating.

It is also noted that some authors have treated with maintenance problems for modular systems (see Joo (2009); Gao et al. (2016); Putri et al. (2020); Sharifi and Taghipour (2020); Hashemi et al. (2020), for instance). However, to the best of our knowledge, little research has regarded a system composed of subsystems of three types, on which the repair effectiveness are minimal, perfect, and imperfect repairs, respectively.

Regarding the approaches to approximating the SGRP, research is relatively sparse and is reviewed below. There are some work. Cox and Smith (1954) proved that the observed failure times of an SRP tend to be distributed with a homogeneous Poisson process when the number of components goes to infinity and the time is far from the origin. Khinchin (1956) further clarified that the SRP tends to be a non-homogeneous Poisson process. Drenick (1960) proved the same property as that of Cox and Smith (1954) holds even if the failure processes of the multi-sockets follow heterogeneous renewal processes. Assume that the survival function (i.e., reliability) of each component in a system is given, then Lawrance (1973) obtains the limiting survival function of the system.

In the applications of the SRP, Kallen (2011) uses the SRP to model the effect of imperfect maintenance. Bratiychuk and Kempa (2003) applies the SRP to study batch arrival queues. Some methods have been developed to approximate the SRP. Whitt (1982) proposes a comprehensive description of two basic methods for approximating the SRP: the stationary-interval method and the asymptotic method. Both methods determine the approximating renewal process by identifying moments for intervals between successive points and fitting a convenient distribution to the moments. Albin (1984) proposes a hybrid method that combines two basic methods described by Whitt (1982) and that approximates the complex superposition process by a RP. Tortorella (1996) proposes a likelihood inference with a pooled discrete RP model. Torab and Kamen (2001) minimises the mean-squared rate error, i.e., the difference between the rate of the superposition process and that of its renewal model. Peixoto (2009) proposes a likelihood-based method with random assignments of failure times onto components. Zhang et al. (2017) enumerate all possible patterns of missing component labels and propose a likelihood-based method to estimate the component-wise renewal distribution under the assumption that all renewal distributions are identical among components. Yamamoto (2020) extend the method proposed in Zhang et al. (2017) for the cases of heterogeneous components. Li et al. (2021) proposes a nonparametric procedure to estimate the inter-occurrence time distribution by properly deconvoluting the renewal equation with the empirical renewal function.

Most of the existing research focuses on the SRP. Little research investigates the SGRP. To the best of our knowledge, there is only one paper on the SGRP: Wu (2020) proposes two methods to approximate the SGRP but does not provide the bounds of the rate of the SGRP. Although Wu and Scarf (2017), Wu (2019), and Wu (2022b) develop methods to approximate the SRP and then applies their appropriate methods to the SGRP, neither of those papers provides a rigorous derivation process in obtaining their approximation methods. As such, this paper aims to bridge this gap.

1.3 Proposed ideas and novelty

This paper derives a maintenance policy in which the system under maintenance is assumed to be composed of three subsystems with three different types of repairs, minimal, imperfect, and perfect, respectively. It then derives the lower and upper bounds of the rate of the SGRP. The failure process models in the maintenance policy are modelled by the approximate approach, which is a weighted average of the lower and upper bounds of the rate of the SGRP. The paper also proposes a method to simulate the SGRP. To the best of our knowledge, the idea of regarding a system composed of three subsystems with three different types of repair, respectively, the derivation of the lower and upper bounds of the rate of the SGRP, the approach to approximating the SGRP and the simulation approach are novel.

1.4 Overview

The remainder of the paper is structured as follows. Section 2 proposes a maintenance policy. Section 3 derives the lower and upper bounds for the SGRP. Section 3 proposes an approximation method that combines the lower and upper bounds and an algorithm to simulate the failure process of the SGRP. Section 4.1 offers numerical examples to illustrate the simulation algorithm proposed in Section 3. Section 4.2 gives a numerical example of maintenance policy optimisation. Section 5 concludes the paper.

2 Development of a block replacement policy

With the discussion in Section 1.1, apparently, different maintenance policies on such a three-subsystem system can be developed. This paper attempts to develop a method on how the three types of repair can be integrated. As such, we investigate the scenario where the simplest maintenance policy—the block replacement policy—is applied. Thus, this section focuses on the development of a block replacement policy.

The notation table is shown in Table 1.

2.1 Assumptions

To develop the maintenance policy, we make the following assumptions.

A1 Suppose a series system is composed of components that are identical and statistically independent. Each component in the system at $t = 0$ is new. Upon failures, the repair effectiveness of those components can be categorised into three types: minimal, perfect, and imperfect, respectively. As such, the system can be regarded as being composed of three subsystems, i.e., subsystems 1, 2, and 3, on which minimal repair, perfect repair, and imperfect repair are conducted upon failures, respectively. That is, each subsystem is composed of multiple components and the effectiveness of repair on each component in the same subsystem is the same, which is one of the three types.

A2 *Preventive Maintenance (PM) Policy N^*:* subsystem 3 has experienced $N^* - 1$ failures, a replacement will be conducted on the entire system at its N^*th failure.

A3 (A3.1.) The failure rate or the initial failure intensity of a component is $\lambda(t)$ before its first failure, where $\lambda(t)$ increases in t. (A3.2.) Once a failure occurs, it is immediately repaired. The time on the repair is negligible. (A3.3.) The repair effectiveness can be depicted by a model such as the virtual age model with fixed parameters, minimal, perfect, or imperfect (MPI).

Table 1: Notations

n	the total number of components in a series system;
i	index: component number;
j	index: order of memory in some models, e.g., an ARA (arithmetic reduction of age) or ARI (arithmetic reduction of intensity) model;
k	index: number of failures (or repairs) of the system or a component; $k = 1, 2, \ldots$
ℓ	index: subsystem number, where ℓ =1, 2, and 3;
t	time since the system starts;
$\mathscr{H}_{i,t-}$	history of the failures of component i up to time t;
$T_{i,k}$	time when the kth failure of component i occurs;
$T_{s,k}$	time when the kth failure of the system occurs;
$T^{(\ell,k)}$	time when the kth failure of subsystem ℓ occurs (where $\ell = 1, 2, 3$);
$X^{(\ell,k)}$	$X^{(\ell,k)} = T^{(\ell,k)} - T^{(\ell,k-1)}$, where $T^{(\ell,0)} = 0$ and $X^{(\ell,1)} = T^{(\ell,1)}$;
$N_t^{(\ell)}$	number of failures of subsystem ℓ by time t;
$N_{i,t}$	the number of failures that component i has experienced up to time t;
$N_{s,t}$	the number of failures that the system has experienced up to time t;
$\lambda_s(t)$	failure intensity function of the system at time t;
$\widetilde{\lambda}_s(t)$	approximation function to $\lambda_s(t)$;
$\lambda_{i,k}(t)$	failure intensity function of component i at t and the system has experienced k failures.
$\lambda_{s,k}(t)$	failure intensity function of the system that has experienced k failures at t;
$\lambda_k(t)$	failure intensity function of a component after the system has experienced k failures;
$\lambda(t)$	failure intensity function of a component before the first failure;

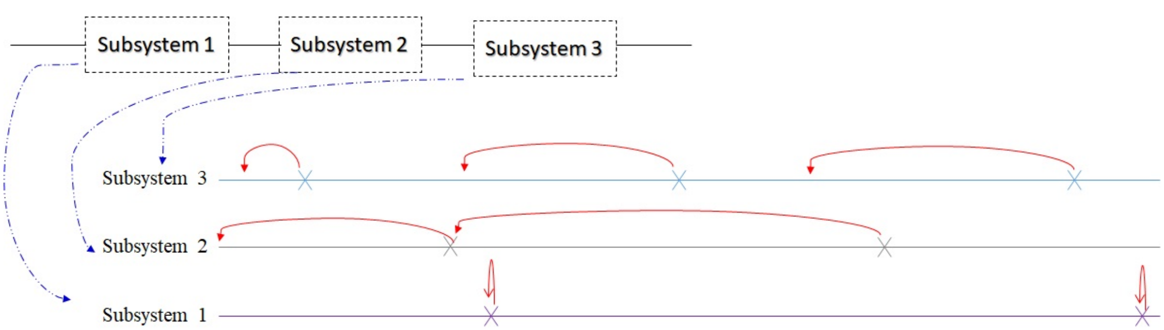

Figure 1: Illustration of a system composed of three subsystems, where the symbol $\times$ stands for a failure; lines with red (or solid) arrows denotes maintenance effectiveness.

Remark 1. *On the repair effectiveness (A3.3), let us look at an example. For example, if a virtual age model is used, we can assume that* $V_k = V_{k-1} + AX_k$*, where* V_k *is the virtual age of the item after the kth repair,* X_k *is the lifetime of the item after the kth repair and A is the degree of the repair. We assume the failure intensity function of a component after the repair is smaller than before the repair.*

2.2 A replacement policy

We assume that the costs associated with the repair of each type are Y_ℓ, with $\ell = 1, 2$, and 3, respectively, where $Y_\ell \sim H_\ell(y_\ell)$ and $H_\ell(y_\ell)$ is the cumulative distribution function of the costs associated with the minimal, perfect, and imperfect, respectively. Unlike most scenarios in the existing literature where a constant cost is assumed, the costs of maintenance considered in this paper are due to repairs for different components in a system. As such, we use the expected costs to quantify the relevant costs (instead of constants) which can facilitate the uncertainty analysis of the costs in some scenarios such as the optimisation of maintenance policies.

The cost of repairing subsystem 3 within the interval $(0, T)$, within which subsystem 3 has experienced N failures, is given by:

$$C_3(T) = \sum_{k=1}^{N_T^{(3)}} Y_3, \tag{1}$$

where $N_t^{(3)} = \sup\{m \colon T^{(3,m)} \leq t, m \geq 1\}$. That is, $N_t^{(3)}$ is the number of failures of subsystem 3 in the time interval $(0, t]$.

Assuming that Y_3 is independent of $N_t^{(3)}$, we can obtain

$$E[C_3(T)] = E[N_T^{(3)}]E[Y_3]. \tag{2}$$

Note that $\{T^{(3,m)} \leq t\} \equiv \{N_t^{(3)} \geq m\}$. Then

$$\begin{aligned}\Pr\{T^{(3,m)} \leq t\} &= \Pr\{N_t^{(3)} \geq m\} \\ &= 1 - \sum_{k=0}^{m-1} \Pr\{N_t^{(3)} = k\}.\end{aligned} \tag{3}$$

Denote the probability density function (pdf) of $T^{(3,m)}$ by $f_{(3,m)}(t)$. With Eq. (3), we obtain

$$f_{(3,m)}(t) = -\frac{d}{dt} \sum_{k=0}^{m-1} \Pr\{N_t^{(3)} = k\}. \tag{4}$$

Then the expected time to the N-th failure of subsystem 3 is given by

$$E[T^{(3,N)}] = -\sum_{k=0}^{N-1} \int_0^{+\infty} t \frac{d}{dt} \Pr\{N_t^{(3)} = k\} dt. \tag{5}$$

For the sake of simple notation, we denote $\Psi_N = E[T^{(3,N)}]$. Then the number of failures of subsystem 1, which is repaired with minimal repair, within $(0, \Psi_N)$ is given by

$$E[N_{\Psi_N}^{(1)}] = \int_0^{\Psi_N} \lambda_M(t) dt, \tag{6}$$

where $\lambda_M(t)$ is the intensity function of the failure process of subsystem 1.

The expected number of failures of subsystem 2, which is repaired with perfect repair, within

$(0, \Psi_N)$ is denoted by $E[N^{(2)}_{\Psi_N}]$. Then we obtain

$$E[N^{(2)}_{\Psi_N}] = \sum_{k=1}^{\infty} k\Pr[N^{(2)}_{\Psi_N} = k] \\ = \sum_{k=1}^{\infty} \Pr[T^{(2,k)} \le \Psi_N]. \tag{7}$$

That is, $E[N^{(2)}_{\Psi_N}]$, which is a renewal function, is the expected number of repairs within the time interval $(0, \Psi_N)$.

Then the expected cost ratio of the system on the three subsystems is obtained by

$$C(N) = \frac{E[C_1(\Psi_N)] + E[C_2(\Psi_N)] + E[C_3(\Psi_N)] + C_{\mathrm{R}}}{\Psi_N} \\ = \frac{E[Y_1]E[N^{(1)}_{\Psi_N}] + E[Y_2]E[N^{(2)}_{\Psi_N}] + E[Y_3](N-1) + C_{\mathrm{R}}}{\Psi_N}, \tag{8}$$

where C_R is the cost of a replacement of the entire system and is a constant value. The first element (i.e., $E[Y_1]E[N^{(1)}_{\Psi_N}]$) on the numerator is the cost due to subsystem 1, and the second one (i.e., $E[Y_2]E[N^{(2)}_{\Psi_N}]$) is the cost incurred due to subsystem 2, and the third one (i.e., $E[Y_3](N-1)$) is due to subsystem 3, followed by Assumption A2. Normally, we assume $C_R > \max\{E[Y_1], E[Y_2], E[Y_3]\}$.

Then the optimal value of N is obtained by minimising $C(N)$, that is

$$N^* = \operatorname*{argmin}_N C(N). \tag{9}$$

According to Table 1, $T^{(3,1)}$ denotes the time when the first failure of subsystem 3 occurs. Strictly, $T^{(3,1)}$ is the quantity that satisfying $P(N^{(3)}_t < 1) = P(T^{(3,1)} > t)$. Denote $\Psi^{(1)}_N = E[T^{(3,1)}]$.

Lemma 1. *Assume that $E[Y_1]E[N^{(1)}_{\Psi^{(1)}_N}] + E[Y_2]E[N^{(2)}_{\Psi^{(1)}_N}] < C_R$, with Assumptions A1, A2, and A3, there exists an optimal solution N^* so that $N^* = \operatorname*{argmin}_N C(N)$.*

The proof of Lemma 1 and the proofs of the other propositions in this paper can be found in Appendix.

To minimise $C(N)$ in Eq. (8), we need to obtain $\lambda_M(t)$, $E[N^{(2)}_{\Psi_N}]$, and Ψ_N. In other words, we need to estimate $\Pr[T^{(2,k)} \le t]$ and $\Pr[T^{(3,k)} \le t]$, respectively, which can be done through closed-form expressions or simulation for $\Pr[T^{(2,k)} \le t]$ and $\Pr[T^{(3,k)} \le t]$. Example 1 gives the expression of $C(N)$ for special cases where the lifetime distributions of the three subsystems are available (and given). For the cases that those lifetime distributions are not available, which is often the case in the real world, Section 3 derives an approach to approximating the intensity function of the stochastic process for modelling the failure process of a multi-component system, based on which we can design simulation algorithms to estimate $E[N^{(2)}_{\Psi_N}]$ and Ψ_N.

Below we consider two scenarios: the failure intensity function after repair of subsystem i is known in Scenario 1 and their failure intensity functions need estimating from the field failure data in Scenario 2.

2.2.1 Scenario 1: the failure intensity function after repair of subsystem i is known

If the failure intensity function is deterministic, it is easy to seek N^* by minimising $C(N)$ in Eq. (8).

Example 1. *Denote the time to the first failure of subsystem* ℓ *as* $X^{(\ell,1)}$ *for* $\ell = 1, 2, 3$. *Suppose the probability distribution of* $X^{(\ell,1)}$ *as* $F_{(\ell,1)}(x) = \frac{\gamma(\alpha_\ell, x/\beta_\ell)}{\Gamma(\alpha_\ell)}$, *that is,* $X^{(\ell,1)}$ *obeys the gamma distribution, where* $\Gamma(\cdot)$ *is the gamma function:* $\Gamma(z) = \int_0^\infty u^{z-1}e^{-u}du$, *where* $\gamma(s, x) = \int_0^x t^{s-1}\, e^{-t}\, dt$.

The expected time to the N-th failure of Subsystem 3. Lam (2007, page 38) defined a stochastic process, or the geometric process (GP), to model the times-between-failures of a repairable system as *"Given a sequence of non-negative random variables* $\{X_k, k = 1, 2, \dots\}$, *if they are independent and the cdf of* X_k *is given by* $F(a^{k-1}x)$ *for* $k = 1, 2, \dots,$ *where a is a positive constant, then* $\{X_k, k = 1, 2, \cdots\}$ *is called a GP."* Assume the working times, $X^{(3,k)}$, of subsystem 3 follows the GP with $a > 1$ and the pdf of $X^{(3,1)}$ is given by $f(x; \alpha_3, \beta_3) = \frac{\beta_3^{-\alpha_3}}{\Gamma(\alpha_3)} x^{\alpha_3 - 1} e^{-x/\beta_3} 1_{\{x>0\}}$, where $1_{\{x>0\}} = 1$ if $x > 0$ and $1_{\{x>0\}} = 0$ otherwise. Then, according to Moschopoulos (1985), the pdf of $T^{(3,N)} = \sum_{k=1}^N X^{(3,k)}$ can be expressed by

$$f_{(3,N)}(t) = a^{-\frac{\alpha_3 N(N+1)}{2}} \sum_{k=0}^{\infty} \frac{\zeta_k}{\Gamma(N\alpha_3 + k)(a\beta_3)^{(N\alpha_3+k)(1-N)}} t^{N\alpha_3+k-1} e^{-t/(a^{1-N}\beta_3)}, \quad t > 0, \tag{10}$$

where ζ_{k+1} (for $k = 0, 1, 2, \dots$) is obtained in a recursive way as $\zeta_{k+1} = \frac{1}{k+1} \sum_{j_0=1}^{k+1} j_0 \eta_{j_0} \zeta_{k+1-j_0}$, with $\zeta_0 = 1$ and η_{j_0} being given by $\eta_{j_0} = \alpha_3 \sum_{j_1=1}^N (1 - a^{j_1-N})^{j_0}/j_0$.

Then, the expected length to the N-th failure of Subsystem 3 is given by

$$\begin{aligned}
\Psi_N &= \int_0^{+\infty} t f_{(3,N)}(t) dt \\
&= \int_0^{+\infty} t a^{-\frac{\alpha_3 N(N+1)}{2}} \sum_{k=0}^{\infty} \frac{\zeta_k}{\Gamma(N\alpha_3 + k)(a\beta_3)^{(N\alpha_3+k)(1-N)}} t^{N\alpha_3+k-1} e^{-t/(a^{1-N}\beta_3)} dt \\
&= a^{-\frac{\alpha_3 N(N+1)}{2}} \sum_{k=0}^{\infty} \frac{\Gamma(N\alpha_3 + k + 1)\zeta_k (a^{1-N}\beta_3)^{N\alpha_3+k+1}}{\Gamma(N\alpha_3 + k)(a\beta_3)^{(N\alpha_3+k)(1-N)}}.
\end{aligned} \tag{11}$$

The expected number of failures of Subsystem 1. According to Assumption A1, repair on failures of Subsystem 1 is minimal. Hence, the expected number of failures of Subsystem 1 within $(0, \Psi_N]$ is given by

$$E[N_{\Psi_N}^{(1)}] = \int_0^{\Psi_N} \frac{\Gamma(\alpha_1, \beta_1 u)}{\Gamma(\alpha_1)} du. \tag{12}$$

The expected number of failures of Subsystem 2. Let $T^{(2,N)} = \sum_{k=1}^N X^{(2,k)}$. Then the cumulative probability distribution of $T^{2,N}$ is given by $\Pr\{T^{(2,N)} < t\} = \frac{\gamma(N\alpha_2, \beta_2 t)}{\Gamma(N\alpha_2)}$. Then the expected number of renewals is given by

$$E[N_{\Psi_N}^{(2)}] = \sum_{k=1}^{\infty} \frac{\gamma(k\alpha_2, \beta_2 \Psi_N)}{\Gamma(k\alpha_2)}. \tag{13}$$

Plugging Ψ_N in Eq. (11), $E[N^{(2)}_{\Psi_N}]$ in Eq. (13), and $E[N^{(1)}_{\Psi_N}]$ in Eq. (12) into Eq. (8), we can obtain $C(N)$.

2.2.2 Scenario 2: The failure intensity function to be estimated from field failure data

As discussed in Section 1, it is not possible to know the causes of the failures of a subsystem in some cases, that is, the failure data may be masked. Hence, there is a need to develop methods to approximate the failure intensity function for each subsystem with masked failure data, which is the focus of Section 3.

3 Approximation of the SGRP

Let $\{T_{i,k} : k = 1, 2, ...\}$ be the successive failure times of component i (hence $T_{i,k+1} > T_{i,k}$), starting from $T_{i,0} = 0$, and $N_{i,t}$ be the number of failures up to time t. Let $\mathscr{H}_{i,t-}$ denote the history of the failure process up to t (exclusive of t). The failure process of the component can be defined equivalently by the random processes $\{T_{i,k}\}_{k\geq 1}$ or $\{N_{i,t}\}_{t\geq 0}$ (i.e., $\{N_i(t)\}_{t\geq 0}$) and is characterised by the intensity function,

$$\lambda_{i,N_{i,t}}(t|\mathscr{H}_{i,t-}) = \lim_{\Delta t\downarrow 0} \frac{P\{N_{i,t+\Delta t} - N_{i,t} = 1|\mathscr{H}_{i,t-}\}}{\Delta t}, \tag{14}$$

where $P\{N_{i,t+\Delta t} - N_{i,t} = 1|\mathscr{H}_{i,t-}\}$ is the probability that component i fails within the interval $(t, t + \Delta t)$, given the history of failures of the component up to time t, $\mathscr{H}_{i,t-}$. Similarly, we can define the failure intensity of the system. Although the failure intensity function of component i should be denoted by the memory of $\mathscr{H}_{i,t-}$ such as $\lambda_{i,N_{i,t}}(t|\mathscr{H}_{i,t-})$, for simplicity, this paper will use notations $\lambda_{i,N_{i,t}}(t)$. The paper uses the term the "failure intensity function" of the SGRP or SGRP exchangeably.

3.1 Lower and upper bounds of the rate of SRP with identical components

In this section, we derive the lower and upper bounds of the rate of the SRP. Suppose a system consisting of n sockets into each of which a component is inserted. If a component fails, it is replaced immediately and system operation resumes. As above discussed, the failure process of the system is the superimposed renewal process. We have the following proposition.

Proposition 1. *If replacement is performed on each failed component, or the failure process of the system is the SRP, then*

$$\sum_{j=0}^{n-1} \lambda(t - T_{s,N_{s,t}-j}) \leq \lambda_s(t) \leq (n-1)\lambda(t) + \lambda(t - T_{s,(N_{s,t})}), \tag{15}$$

where $T_{s,k} = 0$ if $k \leq 0$ and $N_{s,t} \geq 0$.

Proposition 1 can be interpreted as follows. Always replacing the oldest component with a new component upon a failure of the system results in the largest reduction in the rate of the SRP, which keeps the system at the most reliable state and therefore the smallest failure intensity. On the other hand, always replacing the youngest component with a new component upon a failure of the system results in the smallest reduction in the rate of the SRP, which keeps the system at the least reliable state and therefore the largest failure intensity.

3.2 Lower and upper bounds of the rate of SGRP with identical components

Similar to Proposition 1, for the case where failed components are repaired (instead of replacement), a similar proposition is given below.

Proposition 2. *Under assumptions A3 in Section 2.1, the failure intensity $\lambda_s(t)$ of the system after the $N_{s,t}$-th repair satisfies*

$$\sum_{j=0}^{n-1} \lambda_{j,N_{s,t}-j}(t) \leq \lambda_s(t) \leq (n-1)\lambda(t) + \lambda_{N_{s,t}}(t), \tag{16}$$

where $\lambda_{i,N_{s,t}-j}(t) = \lambda(t)$ for $N_{s,t} - j \leq 0$.

Proposition 2 can be interpreted similarly to that of Proposition 1, but interpreted in another way, as follows. The left and right terms in Inequality (16) are two extreme scenarios: the most and the least reliable situations. In the most reliable situation, the failures of the system are always due to the failure of the most anciently repaired component; in the most unreliable situation, the failures of the system are always due to the failure of the most recently repaired component. These two scenarios form the lower bound and the upper bound of the failure intensity of the system, respectively.

3.3 A series system composed of heterogeneous components

Proposition 2 derived the lower and upper bounds of the rate of the SGRP based on the assumptions A3. If we assume the components in the system are heterogeneous, then we can derive the following upper bound.

Proposition 3. *Suppose a system is composed of n components. The initial intensity functions of the components are $\xi_i(t)$ where $i = 1, 2, \ldots, n$ and $\xi_1(t) \leq \xi_2(t) \leq \cdots \leq \xi_n(t)$. Assume the maintenance effectiveness upon failures is the same for each individual component. The failure intensity $\lambda_s(t)$ of the system after the $N_{s,t}$-th repair satisfies*

$$\lambda_s(t) \leq \sum_{i=2}^{n} \xi_i(t) + \xi_{1,N_{s,t}}, \tag{17}$$

where $\xi_{1,N_{s,t}}$ is the failure intensity function of component i that has experienced $N_{s,t}$ failures.

3.4 A model for approximating the SGRP

This section aims to propose a model to approximate the SGRP and then propose an algorithm to simulate the approximation method.

An interesting question is to approximate the rate of the SGRP. Based on Proposition 2, we can use the following model, which is the weighted average of the lower bound and the upper bound of

Inequality (16), respectively, to approximate the rate of the SGRP of the n components in a series system for $N_{s,t} \geq 1$.

$$\begin{aligned}\widetilde{\lambda}_s(t) &= \delta \sum_{j=0}^{n-1} \lambda_{N_{s,t}-j}(t) + (1-\delta)\left((n-1)\lambda(t) + \lambda_{N_{s,t}}(t)\right), \\ &= (n-1)(1-\delta)\lambda(t) + (1-\delta)\lambda_{N_{s,t}}(t) + \delta \sum_{j=0}^{n-1} \lambda_{N_{s,t}-j}(t), \end{aligned} \tag{18}$$

where $\delta \in [0, 1]$.

Remark 2. *Eq.* (18) *has the following special cases.*

(a) If $\lambda(t) = \lambda$, then $\widetilde{\lambda}_s(t) = \lambda$. That is, the failure intensity of a system composed of components with a constant failure intensity, i.e., failure rate, is constant.

(b) If $n = 1$, then $\widetilde{\lambda}_s(t) = \lambda_{N_{s,t}}(t)$. That is, the system is composed of a single component, whose repair process is modelled by a GRP.

(c) If $\delta = 0$, then $\widetilde{\lambda}_s(t) = (n-1)\lambda(t) + \lambda_{N_{s,t}}(t)$. This implies that the system can be assumed to be a two-component system, in which minimal repair is applied to the component with intensity function $(n-1)\lambda(t)$ and MPI is applied to the component with failure intensity function $\lambda_{N_{s,t}}(t)$. This case is also a variant of Model I proposed in Wu and Scarf (2017), in which perfect repair is applied to the component with failure intensity function $\lambda_{N_{s,t}}(t)$.

(d) If $\delta = 1$, then model (18) *is the failure intensity of a system composed of n components in series with the failure process as an SGRP on each repair and reduces to the moving average of intensity model (MAI) proposed in Wu (2019), which shows the outstanding performance compared with ten other models on 15 real datasets and simulation data in terms of AIC (Akaike information criterion), BIC (Bayesian information criterion) and corrected AIC.*

(e) Denote $\lambda_{s,L}(t) = \sum_{j=0}^{n-1} \lambda_{N_{s,t}-j}(t)$ and $\lambda_{s,U} = (n-1)\lambda(t) + \lambda_{N_{s,t}}(t)$. $\lambda_{s,L}(t)$ is the failure intensity of the system whose failures are always due to the failures of the most anciently repaired component or the failure of the oldest component. $\lambda_{s,U}(t)$ is the failure intensity of the system whose failures are always due to the failures of the component that has been most recently repaired, or the youngest component. The approximation method in Eq. (18) *is a weighted average of two extremist scenarios. Below is an example that provides a visual interpretation.*

(f) Denote $\lambda_1(t) = (n-1)(1-\delta)\lambda(t)$, $\lambda_2(t, N_{s,t}) = (1-\delta)\lambda_{N_{s,t}}(t)$, and $\lambda_3(t, N_{s,t}) = \delta \sum_{j=0}^{n-1} \lambda_{N_{s,t}-j}(t)$. Then $\widetilde{\lambda}_s(t) = \lambda_1(t) + \lambda_2(t, N_{s,t}) + \lambda_3(t, N_{s,t})$. Then $\lambda_1(t)$ can be regarded as the failure intensity function of Subsystem 1, $\lambda_2(t, N_{s,t})$ can be regarded as the failure intensity function of Subsystem 2, and $\lambda_3(t, N_{s,t})$ can be regarded as the failure intensity function of Subsystem 3. That is, the system under study is composed of three subsystems, repair on Subsystem 1 is minimal, perfect on Subsystem 2, and imperfect on Subsystem 3.

Suppose one has observed N failures of a technical system. Denote the N observed inter-arrival times by $t^{(1)}, t^{(2)}, \ldots, t^{(N)}$. Denote the likelihood function of the model defined in Eq. (18) by $\mathscr{L}(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the set of parameters in $\widetilde{\lambda}_s(t)$, which is re-written by $\widetilde{\lambda}_s(t\colon \boldsymbol{\theta})$. Then $\widetilde{\lambda}_s(t\colon \boldsymbol{\theta})$ is given by $\mathscr{L}(\boldsymbol{\theta}) = \prod_{k=1}^{N} \frac{d\widetilde{\lambda}_s(t\colon \boldsymbol{\theta})}{dt}|_{t=t^{(k)}} \exp[-\widetilde{\lambda}_s(t^{(k)})]$. By maximising $\log(\mathscr{L}(\boldsymbol{\theta}))$, one can obtain the estimate value $\hat{\delta}$ and the estimated parameters in $\widetilde{\lambda}_s(t)$.

Example 2. *Given a series system that consists of four components, whose failures at time points are shown in the top four horizontal lines in Figure 2. The superposition of the four failure processes is shown at the bottom horizontal line: $T_{s,1}, T_{s,2}, \ldots, T_{s,11}$. As can be seen, components 1, 2, 3 and 4 have 3, 2, 2, and 4 failures, respectively. Now let us assume that only masked failure data are available.*

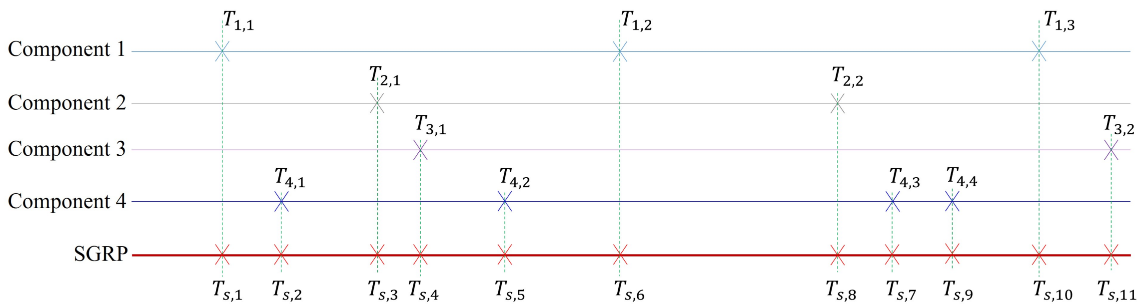

Figure 2: An example of the superposition of four imperfect repair processes.

That is, only values of $T_{s,1}, T_{s,2}, \ldots, T_{s,11}$ are available but $T_{i,k}$ are unavailable, as illustrated in Figure 3.

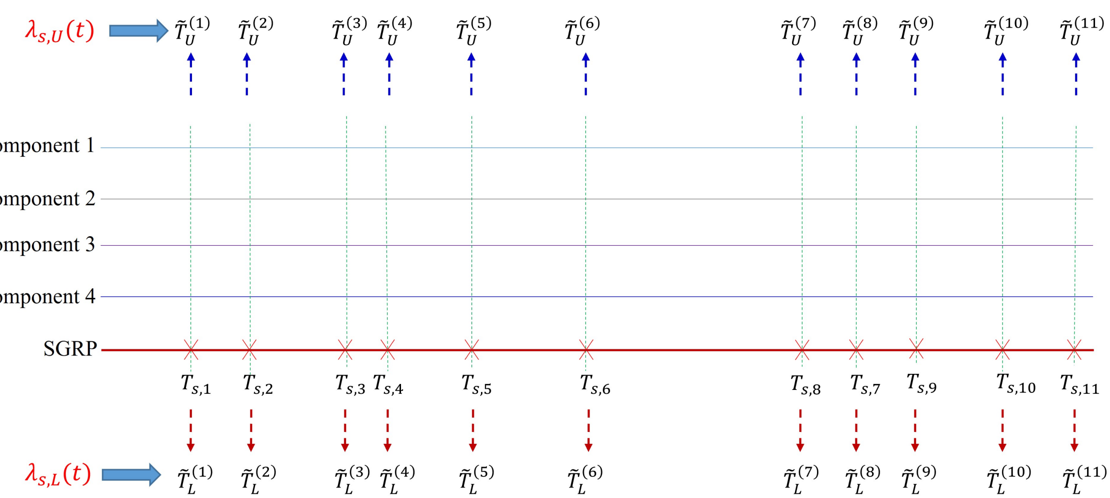

Figure 3: $\lambda_{s,L}(t)$ and $\lambda_{s,U}(t)$ assume the masked failure data to be unmasked ones, respectively.

In the literature, many GRPs have been proposed. Below we assume the ARA model is applied.

Example 3. *In case the ARA model (Doyen and Gaudoin, 2004) is used to model the repair process* *of component i, then its initial failure intensity function is given by*

$$\lambda_{i,N_{i,t}}(t) = \lambda\left(t - \rho\sum_{j=0}^{\min\{m-1,N_{i,t}-1\}}(1-\rho)^j T_{i,N_{i,t}-j}\right), \tag{19}$$

where $\rho \in [0,1]$: (a) the repair is a harmful repair if $\rho < 0$; (b) the repair is a minimal repair if $\rho = 0$; (c) the repair is an good-as-new repair (or replacement) if $\rho = 1$; and (d) the repair is efficient if $\rho \in (0,1)$. If $m = 1$, then Eq. (19) *reduces to the virtual age model I (Kijima, 1989). If the failure data are masked, the failure intensity, $\widetilde{\lambda}_s(t)$, of the superposition of the n ARA_m failure processes is therefore approximated by the expressions of the model* (18)*, where the repair process $\lambda_{i,N_{s,t}}(t)$ follows the ARA_m model. According to Eq.* (18)*, we have*

- *when* $N_{s,t} = 0$, $\widetilde{\lambda}_s(t) = \lambda(t)$;
- *when* $1 \le N_{s,t} \le n$ *and* $n \ge 2$, $\widetilde{\lambda}_s(t) = (n - N_{s,t})\delta + (n-1)(1-\delta)\lambda(t) + (1-\delta)\lambda\left(t - \rho T_{s,N_{s,t}}\right) + \delta\sum_{i=1}^{N_{s,t}}\lambda\left(t - \rho T_{s,i}\right)$; *and*
- *when* $N_{s,t} > n$, $\widetilde{\lambda}_s(t) = (n-1)(1-\delta)\lambda(t) + (1-\delta)\lambda\left(t - \rho\sum_{j=0}^{q_{N_{s,t}}}(1-\rho)^j T_{s,N_{s,t}-j}\right)$ $+\delta\sum_{j=0}^{n-1}\lambda\left(t - \rho\sum_{j=0}^{q_{N_{s,t}}}(1-\rho)^j T_{s,N_{s,t}-nj-j}\right)$, *where* $q_{N_{s,t}} = \min\{\lfloor\frac{N_{s,t}}{n}\rfloor - 1, m-1\}$.

Then we can re-write model (18) as follows.

$$\psi(N_{s,t},\lambda(t)) = (\max\{n - N_{s,t}, 0\}\delta + (n-1)(1-\delta))\lambda(t) + (1-\delta)\lambda_{s,N_{s,t}}(t) + \delta\sum_{j=0}^{\min\{N_{s,t}-1,n-1\}}\lambda_{s,N_{s,t}-j}(t). \tag{20}$$

Then, model (18) can be re-written by

$$\widetilde{\lambda}_s(t) = \begin{cases} \lambda(t) & \text{if } N_{s,t} = 0 \\ \psi(N_{s,t},\lambda(t)) & \text{if } N_{s,t} \ge 1. \end{cases} \tag{21}$$

3.5 A simulation algorithm

Developing an algorithm to simulate the behaviour of the stochastic process model is useful in real applications. This section, therefore, aims to develop an algorithm to simulate the failure process based on $\widetilde{\lambda}_s(t)$ in Eq. (18) and then shows some simulation examples. Algorithm 1 shows the procedure of simulating the conditional failure intensity of $\widetilde{\lambda}_s(t)$ given in Eq. (18), in which we assume that the failure rate of each component is $\frac{1}{n}\lambda(t)$. To simplify notations in the Algorithm, we denote $\Phi_1 = \frac{\delta}{n}\sum_{j=0}^{\min\{N_{s,t}-1,n-1\}}\lambda_{s,N_{s,t}-j}(t)$, $\Phi_2 = \frac{(1-\delta)(n-1)}{n}\lambda(t)$, and $\Phi_3 = \frac{1-\delta}{n}\lambda_{s,N_{s,t}}(t)$, which corresponds the three items in Eq. (18). Basically, the simulation method proposed in Algorithm 1 regards the model in Eq. (18) as a series system that is composed of three subsystems with failure intensity functions Φ_1, Φ_2, and Φ_3, respectively. See the descriptions after symbols '/*' and '*/' in the Algorithm.

Algorithm 1: Simulation of Model shown in Eq. (18)

Data:

- Given $\Lambda_0(v) = \int_0^v \lambda(t)dt$ and $\Lambda_1(v) = \frac{(1-\delta)(n-1)}{n}\Lambda_0(v)$.
- Given $\tau'_{i_0,k}$; /* which are $n+1$ series of successive failure times that have already been generated based on a given base repair model $\lambda_{s,N_{s,t}}(t)$ with initial failure intensity function $\frac{\delta}{n}\lambda(t)$ and each series have N data points, where $i_0 = 1, 2, \ldots, n+1, n+2$ and $k = 1, 2, \ldots, N$. That is, $\tau'_{i_0,k}$ is the kth failure time point of component i. */

Result: Simulated data: $t^{(1)}, t^{(2)}, \ldots, t^{(N)}$. /* $t^{(k)}$ is the time to the kth failure of the system. */

Sort $\tau'_{i,k}$ according to the first column (i.e., $k = 1$ and $i \in \{1, 2, \ldots, n+1, n+2\}$) in ascending order and denote the sorted matrix as $\tau_{i,k}$;
$s \leftarrow 0$;
$\tau^*_0 \leftarrow 0$;
$\tau_0 \leftarrow 0$;
/* The For-loop below is to generate data based on Φ_3, or simulates an NHPP with cumulative intensity function $\Lambda_1(t)$ */
for $k = 0; k \leq N$ **do**
    Draw $s_{k+1} \sim U(0,1)$;
    $\tau^*_{k+1} \leftarrow \tau^*_k - \log(s_{k+1})$;
    $\tau_{n+1,k+1} \leftarrow \inf\{v : \Lambda_1(v) \geq \tau^*_{k+1}\}$;
end
$i_1 \leftarrow 1$;
$i_2 \leftarrow 1$;
$t^{(1)} \leftarrow \min\{\tau_{1,1}, \tau_{n+1,1}, \tau_{n+2,1}\}$;
/* The following For-loop generates random numbers based on $\Phi_1 + \Phi_2 + \Phi_3$ for $N_{s,t} \leq n$ */
for $2 \leq i_1 \leq n$ **do**
    if $t^{(i_1)} = \tau_{n+1,i_2}$ **then**
        $i_2 \leftarrow i_2 + 1$;
        $t^{(i_1)} \leftarrow \min\{\tau_{i_1,1}, \tau_{n+1,i_2}, \tau_{n+2,i_1}\}$;
    else
        $t^{(i_1)} \leftarrow \min\{\tau_{i_1,1}, \tau_{n+1,i_2}, \tau_{n+2,i_1}\}$;
    end
end
/* The following For-loop generates random numbers based on $\Phi_1 + \Phi_2 + \Phi_3$ for $N_{s,t} > n$ */
for $i_1 \in \{n+1, \ldots, N\}$ **do**
    $S_0 \leftarrow \{1, 2, \ldots, n\}$;
    for $i_0 \in S_0$ **do**
        $t'_{i_0} \leftarrow \tau_{i_0, i_1 - i_0}$;
        $i_2 \leftarrow i_2 + 1$;
        if $t^{(i_1)} = \tau_{n+1,i_2}$ **then**
            $t^{(i_1)} \leftarrow \min_{i_0 \in S_0}\{t'_{i_0}, \tau_{n+1,i_2}, \tau_{n+2,i_1}\}$;
        else
            $t^{(i_1)} \leftarrow \min_{i_0 \in S_0}\{t'_{i_0}, \tau_{n+1,i_2}, \tau_{n+2,i_1}\}$;
        end
    end
end

4 Numerical examples

4.1 Simulation examples on the approximate model

Suppose a series system is composed of n identical components, each of which has the failure intensity function $\lambda(t) = \frac{1.3}{40}\left(\frac{t}{40}\right)^{0.3}$, and the number of failures is set to be $N = 200,000$. Each curve in the figures illustrates the pattern of failure rates. The failure rate is defined as the ratio of the number of failures in 1000 units of time to 1000. In all of the figures, the failure process of each component is assumed to be based on ARA_1 from Eq. (19). Figure 4 shows three cases when the causes of failures are available, i.e., the failure data are unmasked. It is constructed based on Eq. (18) in which $n = 100$ and the values of ρ are 0.3, 0.6, and 0.9, respectively. Figure 5 shows six cases that are generated based on the algorithm in Table 1. $n = 100$ in all curves in this figure. By comparing Figures 4 and 5, one can see that they share similar patterns. From Figure 5, it can be seen that the failure rates increase when δ decreases. The two upper curves and the two lower curves in Fig. 6 and Fig. 7 are generated by setting $\delta = 0$ and $\delta = 1$ in Eq. (18), respectively, and they correspond to the lower bounds and upper bounds, respectively. The two curves between the lower and upper curves in Figure 6 and Figure 7 are generated by the SGRP, based on Eq. (19). $\rho = 0.3$ in Figure 6 and $\rho = 0.4$ in Figure 7. It can be seen that the rates from Eq. (19) are smaller than the upper bounds and larger than the lower bounds, respectively, which confirms Proposition 2, respectively.

Figure 4: The three curves are generated based on the SGRP model with its failure intensity function shown in Eq. (19), where the X-axis represents time whose unit is 1000 units of time.

4.2 Numerical examples on maintenance policy optimisation

Section 2.2.2 discusses the expected maintenance cost ratio for the scenario when the failure intensities are estimated from the field data. This subsection provides a numerical example to illustrate the application of the optimal problem shown in Eq. (9). Suppose the number of components, $n = 20$, for the three subsystems, 1, 2, and 3, respectively. We make the following assumptions:

- *Subsystem 1.* We assume the failure intensity function of subsystem 1 is given by $\lambda(t) = \frac{2}{45}\left(\frac{t}{45}\right)$. Then the expected number of failures is given by $\Lambda(t) = \left(\frac{t}{45}\right)^2$;

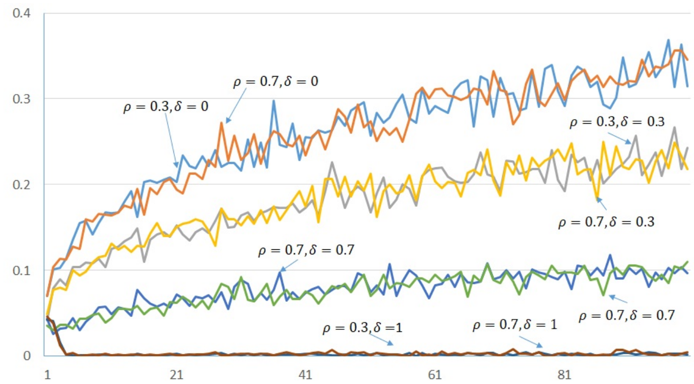

Figure 5: Failure intensity curves generated based on the failure intensity functions in Eq. (18), where the X-axis represents time whose unit is 1000 units of time.

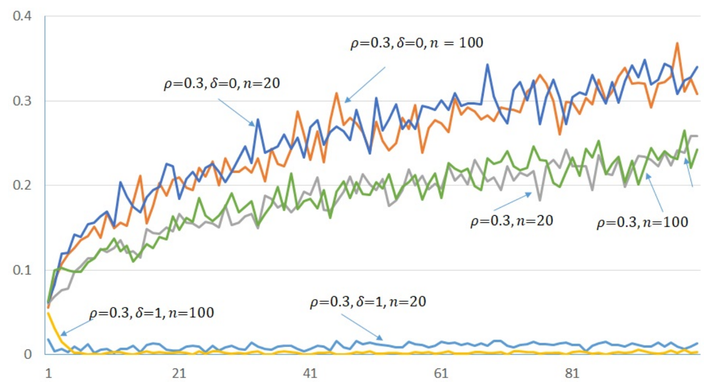

Figure 6: The curves with $\rho = 0.3$.

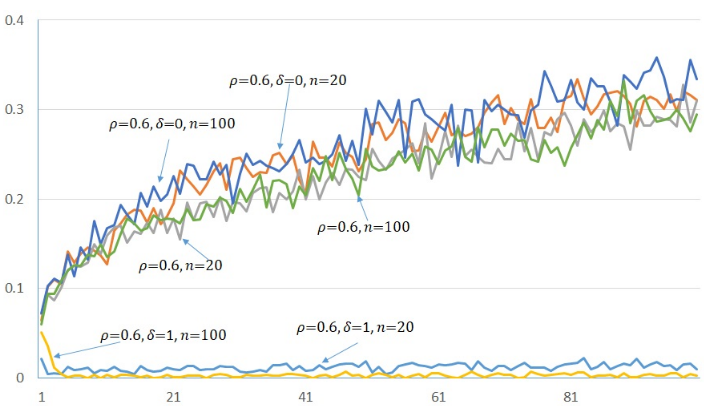

Figure 7: The curves with $\rho = 0.6$, where the X-axis represents time whose unit is 1000 units of time.

- *Subsystem 2.* δ in Eq. (18) is assumed to be $\delta = 0.4$, ρ and m in Eq. (19) are assumed to be $\rho = 1$ and $m = 1$, respectively. $\lambda(t)$ in Eq. (18) for subsystem 2 is assumed to be $\lambda(t) = \frac{1.5}{30}\left(\frac{t}{30}\right)^{0.5}$;

Table 2: The expected times $E[T]$, $E[N_{\Psi_N}^{(1)}]$, and $E[N_{\Psi_N}^{(2)}]$, and the expected cost per unit of time with different N under the above parameter setting.

N	1	2	3	4	5	6	7	8	9	10
$E[T](i.e., \Psi_N)$	36.64	74.59	108.93	140.77	173.59	205.31	236.11	266.91	295.75	324.52
$E[N_{\Psi_N}^{(1)}]$	0.66	2.75	5.86	9.79	14.88	20.82	27.53	35.18	43.19	52.01
$E[N_{\Psi_N}^{(2)}]$	1.69	3.51	5.47	7.41	9.68	12.05	14.40	16.89	19.45	22.14
$C(N)$	2.064	1.458	1.326	1.279	1.274	1.286	1.300	1.320	1.349	1.380
N	11	12	13	14	15	16	17	18	19	20
Ψ_N	352.66	379.45	404.64	432.21	458.26	482.39	507.82	533.70	557.30	583.04
$E[N_{\Psi_N}^{(1)}]$	61.42	71.10	80.86	92.25	103.70	114.91	127.35	140.66	153.37	167.87
$E[N_{\Psi_N}^{(2)}]$	24.91	27.60	30.21	33.20	36.09	38.83	41.78	44.91	47.74	50.90
$C(N)$	1.413	1.443	1.472	1.504	1.534	1.562	1.590	1.621	1.646	1.673

- *Subsystem 3.* δ in Eq. (18) is assumed to be $\delta = 0.4$, ρ and m in Eq. (19) are assumed to be $\rho = 0.3$ and $m = 1$, respectively. $\lambda(t)$ in Eq. (18) for subsystem 3 is assumed to be $\lambda(t) = \frac{1.3}{40}\left(\frac{t}{40}\right)^{0.3}$.

Table 2 shows the values of Ψ_N (see Eq. (5)), $E[N_{\Psi_N}^{(1)}]$ (see Eq (6)), and $E[N_{\Psi_N}^{(2)}]$ (see Eq (7)), respectively. We use Algorithm shown in Table 1 to find the expected values $E[N_{\Psi_N}^{(1)}]$, $E[N_{\Psi_N}^{(2)}]$, and $E[N_{(3)}(\Psi_N)]$, which are calculated based on 1,000 repetitions, respectively. Let $E[Y_1] = 0.4, E[Y_2] = 15, E[Y_3] = 5, C_R = 50$. Thus, when $N^* = 5$, $C(N^*) = 1.274$, as shown in Fig 8.

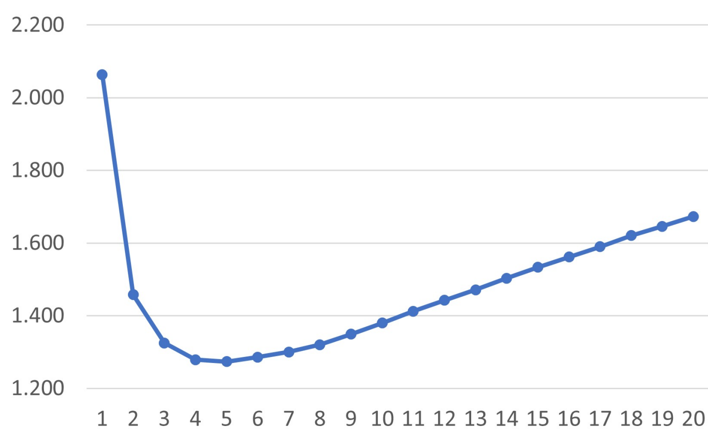

Figure 8: The curve $C(N^*)$, where the X-axis represents N and the Y-axis represents $C(N)$.

5 Conclusions

This paper regarded a multi-component system composed of three subsystems: the one that can be repaired as good as new (i.e., perfect repair), the one that can be repaired as bad as old (i.e, minimal repair), and the one that can be restored to status between perfect and minimal repairs. A maintenance policy was therefore developed by considering the effectiveness of maintenance. The reliability indexes (e.g., failure intensity functions and the survival function) of each individual subsystem is assumed to be estimated based on the field data, on which the most challenging scenario is that the components that cause the subsystems to fail are unknown. This paper then derived the lower and upper bounds of the superimposed renewal process (SRP) rate and then extended to the superposition of the generalised renewal processes (SGRP). It proposed a linear combination of the bounds to approximate the SGRP. The method of simulating SGRP was also proposed. A numerical example on optimisation of maintenance policies is presented. The weighting factor (i.e., δ) was assumed to be deterministic in this paper. A possible extension is to assume it to be a random variable with its probability distribution supported on a bounded interval. This assumption can be interesting in the development of maintenance policies. But the approximation method given in $\widetilde{\lambda}_s(t)$ in Eq. (18) will have two random variables (i.e., $N_{s,t}$ and δ) and therefore become more complicating, which can hinder the method from a wider application. But this problem will be explored in our future research.

In practice, there may be some information on which components cause the system to fail. For instance, one may know that a component is more likely to cause the system to fail than other components. Such information was not considered in this paper and will be considered in our future research.

It is noted that we did not consider the worse-than-minimal-repair scenario (i.e., the repair effectiveness is worse than that of the minimal repair) and the better-than-perfect scenario (i.e., the repair effectiveness is better than that of the perfect repair). The former rarely happens in practice as a repair team with consistent worse-than-minimal-repair outputs will not be employed for a long time. The latter can happen due to technological advance and new modelling methods should be applied. Both scenarios are considered out of the scope of this paper.

In our future work, we will also develop methods to measure the quality of the approximation and explore how the quality is affected by the number of components.

References

Albin, S. L. (1984). Approximating a point process by a renewal process, ii: Superposition arrival processes to queues. *Operations Research*, 32(5):1133–1162.

Andersen, J. F., Andersen, A. R., Kulahci, M., and Nielsen, B. F. (2022). A numerical study of markov decision process algorithms for multi-component replacement problems. *European Journal of Operational Research*, 299(3):898–909.

Asadi, M., Hashemi, M., and Balakrishnan, N. (2023). An overview of some classical models and discussion of the signature-based models of preventive maintenance. *Applied Stochastic Models in Business and Industry*, 39(1):4–53.

Bratiychuk, M. and Kempa, W. (2003). Application of the superposition of renewal processes to the study of batch arrival queues. *Queueing systems*, 44(1):51–67.

Brenière, L., Doyen, L., and Bérenguer, C. (2023). Optimization of preventive replacements dates and covariate inspections for repairable systems in varying environments. *European Journal of Operational Research*, 308(3):1126–1141.

Cai, Y., Teunter, R. H., and de Jonge, B. (2023). A data-driven approach for condition-based maintenance optimization. *European Journal of Operational Research*.

Cox, D. R. and Smith, W. L. (1954). On the superposition of renewal processes. *Biometrika*, 41(1-2):91–99.

de Jonge, B. and Scarf, P. A. (2020). A review on maintenance optimization. *European Journal of Operational Research*, 285(3):805–824.

Doyen, L. and Gaudoin, O. (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety*, 84(1):45–56.

Drenick, R. (1960). The failure law of complex equipment. *Journal of the Society for Industrial and Applied Mathematics*, 8(4):680–690.

Dui, H., Yang, X., and Fang, Y. (2023). Evaluation methodology for preventive maintenance in multi-state manufacturing systems considering different costs. *International Journal of Production Research*, 61(23):8309–8324.

Gao, Y., Feng, Y., and Tan, J. (2016). Product modular design incorporating preventive maintenance issues. *Chinese journal of mechanical engineering*, 29(2):406–420.

Hashemi, M., Asadi, M., and Zarezadeh, S. (2020). Optimal maintenance policies for coherent systems with multi-type components. *Reliability Engineering & System Safety*, 195:106674.

He, S. (2020). Diffusion approximation for efficiency-driven queues when customers are patient. *Operations Research*, 68(4):1265–1284.

Jones, J. and Hayes, J. (2001). Investigation of the occurrence of: no-faults-found in electronic equipment. *IEEE Transactions on Reliability*, 50(3):289–292.

Joo, S.-J. (2009). Scheduling preventive maintenance for modular designed components: A dynamic approach. *European Journal of Operational Research*, 192(2):512–520.

Kallen, M.-J. (2011). Modelling imperfect maintenance and the reliability of complex systems using superposed renewal processes. *Reliability Engineering & System Safety*, 96(6):636–641.

Khinchin, A. (1956). On poisson streams of random events. *Theory of Probability and Its Applications*, 1:248–255.

Kijima, M. (1989). Some results for repairable systems with general repair. *Journal of Applied probability*, 26(1):89–102.

Lam, Y. (2007). *The geometric process and its applications*. World Scientific.

Lawrance, A. J. (1973). Dependency of intervals between events in superposition processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 35(2):306–315.

Li, X.-Y., Ye, Z.-S., and Tang, C. Y. (2021). Estimating the inter-occurrence time distribution from superposed renewal processes. *Bernoulli*, 27(4):2804–2826.

Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics*, 37(1):541–544.

Nicolai, R. P. and Dekker, R. (2008). *Optimal maintenance of multi-component systems: a review*. Springer.

Peixoto, J. (2009). Estimating lifetimes when several unidentified components are reported. *Proceedings of the Physical and Engineering Sciences Section of the American Statistical Association*, pages 5078–5092.

Peng, R., He, X., Zhong, C., Kou, G., and Xiao, H. (2022). Preventive maintenance for heterogeneous parallel systems with two failure modes. *Reliability Engineering & System Safety*, 220:108310.

Putri, N. T., Buana, F. S., et al. (2020). Preventive maintenance scheduling by modularity design applied to limestone crusher machine. *Procedia Manufacturing*, 43:682–687.

Ross, S. M. (1995). *Stochastic processes*. John Wiley & Sons.

Sharifi, M. and Taghipour, S. (2020). Optimal inspection interval for a k-out-of-n system with non-identical components. *Journal of Manufacturing Systems*, 55:233–247.

Singh, D. and Chaudhury, S. (2017). A stochastic theoretical approach to study the size-dependent catalytic activity of a metal nanoparticle at the single molecule level. *Physical Chemistry Chemical Physics*, 19(13):8889–8895.

Singh, V. P., Ganguly, K., and Samad, T. A. (2021). No fault found: a systematic literature review and future research agenda. *International Journal of Quality & Reliability Management*.

Torab, P. and Kamen, E. W. (2001). On approximate renewal models for the superposition of renewal processes. In *ICC 2001. IEEE International Conference on Communications. Conference Record (Cat. No. 01CH37240)*, volume 9, pages 2901–2906. IEEE.

Tortorella, M. (1996). Life estimation from pooled discrete renewal counts. In *Lifetime Data: Models in Reliability and Survival Analysis*, pages 331–338. Springer.

van Staden, H. E., Deprez, L., and Boute, R. N. (2022). A dynamic "predict, then optimize" preventive maintenance approach using operational intervention data. *European Journal of Operational Research*, 302(3):1079–1096.

Whitt, W. (1982). Approximating a point process by a renewal process, I: Two basic methods. *Operations Research*, 30(1):125–147.

Wu, D., Xiao, H., and Peng, R. (2018). Object defense with preventive strike and false targets. *Reliability Engineering & System Safety*, 169:76–80.

Wu, S. (2019). A failure process model with the exponential smoothing of intensity functions. *European Journal of Operational Research*, 275(2):502–513.

Wu, S. (2020). Two methods to approximate the superposition of imperfect failure processes. *Reliability Engineering & System Safety*, 207:107332.

Wu, S. (2022a). The double ratio geometric process for the analysis of recurrent events. *Naval Research Logistics (NRL)*, 69(3):484–495.

Wu, S. (2022b). Lower and upper bounds of the superposition of renewal processes and extensions. *arXiv preprint arXiv:2208.11043*.

Wu, S. and Scarf, P. (2017). Two new stochastic models of the failure process of a series system. *European Journal of Operational Research*, 257(3):763–772.

Xu, J., Liu, B., Zhao, X., and Wang, X.-L. (2023). Online reinforcement learning for condition-based group maintenance using factored markov decision processes. *European Journal of Operational Research*.

Yamamoto, W. (2020). Exponential superimposed renewal processes. In *2020 Asia-Pacific International Symposium on Advanced Reliability and Maintenance Modeling (APARM)*, pages 1–4. IEEE.

Zhang, H. and Li, Y.-F. (2022). Integrated optimization of test case selection and sequencing for reliability testing of the mainboard of internet backbone routers. *European Journal of Operational Research*, 299(1):183–194.

Zhang, W., Tian, Y., Escobar, L. A., and Meeker, W. Q. (2017). Estimating a parametric component lifetime distribution from a collection of superimposed renewal processes. *Technometrics*, 59(2):202–214.

Appendix

Proof of Lemma 1

Proof. Note that the right hand side $\frac{E[Y_1]E[N_{\Psi_N}^{(1)}]+E[Y_2]E[N_{\Psi_N}^{(2)}]+E[Y_3](N-1)+C_{\mathrm{R}}}{\Psi_N}$ in Eq. (8) is the sum of two functions

$$h_1(\Psi_N) = \frac{E[Y_1]E[N_{\Psi_N}^{(1)}] + E[Y_2]E[N_{\Psi_N}^{(2)}] + E[Y_3](N-1)}{\Psi_N},$$

and

$$h_2(\Psi_N) = \frac{C_{\mathrm{R}}}{\Psi_N}.$$

When $N = 1$, with the assumption of Lemma 1, $h_1(\Psi_N^{(1)}) < h_2(\Psi_N^{(1)})$. If $\Psi_N \to +\infty$, $h_2(\Psi_N) \to 0$, but $h_1(\Psi_N) \not\to 0$ because

$$\lim_{\Psi_N \to +\infty} \frac{E[Y_2]E[N^{(2)}_{\Psi_N}]}{\Psi_N} = \frac{E[Y_2]E[N^{(2)}_{T^{(2,1)}}]}{\Psi_N}$$

(according to the elementary renewal theorem, see page 107 in Ross (1995)). Since $h_1(\Psi_N)$ is an increasing function, $h_1(\Psi_N) > h_2(\Psi_N)$ for $\Psi_N \to +\infty$. Hence, there exists an optimal solution in the sum of $h_1(\Psi_N)$ and $h_2(\Psi_N)$. □

Proof of Proposition 1

Proof. It is noted that neither $N_{i,t}$ nor $T_{i,N_{i,t}}$ is available. Since a replacement is carried out at time $T_{i,N_{i,t}}$ on component i, then

$$\lambda_{i,N_{i,t}}(t) = \lambda(t - T_{i,N_{i,t}}), \tag{22}$$

where $\lambda_{i,N_{i,t}}(t)$ is the failure rate of component i at time t, and $\lambda(t - T_{i,N_{i,t}})$ is the failure rate of component i at time t and its latest replacement occurs at time $T_{i,N_{i,t}}$. Component i becomes more reliable after a replacement at time $T_{i,N_{i,t}}$ than without the replacement at that time point, then,

$$\lambda_{i,N_{s,t}-j}(t) \le \lambda_{i,N_{i,t}-1}(t). \tag{23}$$

Since the system is structured in series, the failure intensity function of the system can be expressed by

$$\lambda_s(t) = \lambda_{s,N_{s,t}}(t) = \sum_{i=1}^{n} \lambda_{i,N_{s,t}-j}(t). \tag{24}$$

Now we are going to prove the first part of the inequality, $\sum_{j=0}^{n-1} \lambda(t - T_{s,N_{s,t}-j}) \le \lambda_s(t)$, by induction. If $N_{s,t} = 0$, $\lambda_s(t) = \sum_{j=0}^{n-1} \lambda_{i,0}(t) = n\lambda(t)$. If $N_{s,t} = 1$, which means the system has experienced one failure and the label of the component that causes the system to fail is unknown, then

$$\begin{aligned}
\lambda_s(t) &= \lambda_{s,1}(t) \\
&= \sum_{i=1,i\neq j}^{n} \lambda_{i,0}(t) + \lambda_{j,1}(t) && \text{[assuming component } j \text{ has failed]} \\
&= \sum_{i=1,i\neq j}^{n} \lambda(t) + \lambda(t - T_{j,1}) && \text{[according to Eq. (22)]} \\
&= \sum_{i=1,i\neq j}^{n} \lambda(t) + \lambda(t - T_{s,1}). && \text{[since there is only one failure, } T_{j,1} = T_1] \\
&= \sum_{j=0}^{n-1} \lambda(t - T_{s,N_{s,t}-j}). && \text{[because } T_{s,N_{s,t}-j} = 0 \text{ for} N_{s,t} - j \leq 0] \quad (25)
\end{aligned}$$

In Eq. (25), the sentence on the right-hand side of each equality is the comment on the reason that the respective equality is derived. Now assume the inequality holds when $N_{s,t} = N$. That is, the system has experienced N failures up to time t and

$$\lambda_s(t) \geq \sum_{j=0}^{n-1} \lambda(t - T_{s,N-j}) \tag{26}$$

holds. When $N_{s,t} = N + 1$, we have

$$\begin{aligned}
\sum_{j=0}^{n-1} \lambda(t - T_{s,N+1-j}) &= \lambda(t - T_{s,N+1}) + \lambda(t - T_{s,N}) + \ldots \lambda(t - T_{s,N-n+2}) \\
&= \sum_{j=0}^{n-2} \lambda(t - T_{s,N-j}) + \lambda(t - T_{s,N+1}) \\
&= \sum_{j=0}^{n-1} \lambda(t - T_{s,N-j}) - \lambda(t - T_{s,N-n+1}) + \lambda(t - T_{s,N+1}) \\
&\leq \sum_{j=0}^{n-1} \lambda(t - T_{s,N-j}) && \text{[Because } T_{s,N-n+1} \leq T_{j,N+1} \text{and } \lambda(t) \text{ increases in } t] \\
&= \lambda_s(t). && (27)
\end{aligned}$$

Now we are to prove the second part $\lambda_s(t) \leq (n-1)\lambda(t) + \lambda(t - T_{s,N_{s,t}})$ as follows. From inequality (23), we can obtain

$$\lambda_{i,N_{i,t}}(t) \leq \lambda_{i,N_{i,t}-1}(t) \leq \lambda_{i,N_{i,t}-2}(t) \leq \cdots \leq \lambda_{i,0}(t) = \lambda(t). \tag{28}$$

Hence, combining inequalities (22), (24), and (28), we obtain

$$\sum_{i=1}^{n} \lambda(t - T_{s,N_{s,t}-j}) \leq \lambda_s(t) \leq n\lambda(t). \tag{29}$$

Since the SGRP assumes that repair on a failed component is perfect, it ensures that the component with the latest repair (i.e., the replacement at time $T_{s,N_{s,t}}$) has a failure rate $\lambda(t - T_{s,N_{s,t}})$. Hence, we can obtain

$$\lambda_s(t) \leq (n-1)\lambda(t) + \lambda(t - T_{s,N_{s,t}}). \tag{30}$$

Combining inequalities (27) and (30), we establish Proposition 1. □

Proofs of Propositions 2 and 3

Similar to the proof of Proposition 1, both Propositions 2 and 3 can be established.