Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 54 (2) (2025), 738 – 761
DOI : 10.15672/hujms.1498557

Research Article

Statistical evaluation of multiple process data in geometric processes with exponential failures

Ömer Altındağ

Department of Statistics and Computer Sciences, Bilecik Şeyh Edebali University, 11230, Gülümbe, Bilecik, Turkey

Abstract

The geometric process is a monotonic stochastic process commonly used to model some sort of processes having monotonic trend in time. The statistical inference problem for a geometric process has been well studied in the literature. However, existing studies only cover single process data obtained throughout a single realization of a geometric process. This study presents how multiple process data for a geometric process can arise and considers its statistical evaluation by assuming that all processes are homogeneous and the inter-arrival times follow an exponential distribution. Two data structures for multiple process data are introduced: one consists of complete samples, while the other includes both complete and censored samples. The maximum likelihood and modified maximum likelihood estimators for the parameters of the geometric process are derived on the basis of these data structures. The Expectation-Maximization algorithm is used to compute the maximum likelihood estimators in the case of censored data. The asymptotic properties of the estimators are also derived. Test statistics are proposed based on the asymptotic results of the estimators to distinguish a geometric process from a renewal process and to test the homogeneity of the processes. A simulation study is conducted to demonstrate the performance of the inferential procedures. Finally, both artificial and real data analyzes are presented for illustration.

Mathematics Subject Classification (2020). 62M99, 62F03, 62F10

Keywords. Geometric process, maximum likelihood, modified maximum likelihood, EM algorithm

1. Introduction

In the probabilistic modeling of consecutive events, a general approach is to use a counting process model. If the inter-arrival times of consecutive events are independent and identically distributed (iid), a homogeneous Poisson process (HPP) or its generalization, the renewal process (RP), can be used. If the inter-arrival times are not identically distributed, a non-homogeneous Poisson process (NHPP) can be used. Two important NHPP models are the Cox-Lewis process and the Weibull process. These processes are preferred when inter-arrival times follow a monotone trend. For details, see Cox and Lewis [11] and Ascher and Feingold [2]. The monotone trend of inter-arrival times can alternatively

Email addresses: omer.altindag@bilecik.edu.tr (Ö. Altındağ)
Received: 10.06.2024; Accepted: 03.03.2025

be modeled by a direct monotone counting process model. For this purpose, Lam [17] introduced a monotone counting process model, called the geometric process (GP), to model inter-arrival times following a monotone trend. Let $(X_k)_{k\in\mathbb{N}}$ be non-negative random variables representing the inter-arrival times of consecutive events. Then, the process $\{N(t),\ t \geq 0\}$, where $N(t) = \max\{n : \sum_{k=1}^{n} X_k \leq t\}$, giving the number of events that have occurred up to time t, is said to be a GP with trend parameter a if the random variables $a^{k-1}X_k$, for $k = 1, 2, \ldots$ are i.i.d. The stochastic process notation $\{X_k,\ k = 1, 2, \ldots\}$ for inter-arrival times is also said to be a GP. The GP increases stochastically when $a < 1$, and decreases stochastically when $a > 1$. Note that GP is a generalization of RP, as it reduces to RP when $a = 1$. For a comprehensive discussion of the GP, see Lam [21]. The GP has been utilized in many fields of applied probability. In particular, it has been used as a powerful model for reliability, maintenance, and warranty analysis due to its ease of implementation for monotonic processes. First, Lam [17,18] used the GP to determine the optimal replacement policy for a deteriorating system, since successive operating times of the system after repair are assumed to be stochastically decreasing, while repair times are assumed to be stochastically increasing. Lam [19] considered the optimal repairable replacement model for deteriorating systems using the GP. Since then, GP has been studied for many repair/replacement problems. Lam and Zhang [24] introduced a GP model for a two-component series system with one repairman, assuming that successive operating times of each component are exponentially distributed and form a decreasing GP, while successive repair times of each component are exponentially distributed and form an increasing GP. Lam and Zhang [25], Tang and Lam [34], and Lam [22] considered the GP for maintenance analysis of some deteriorating systems in specific cases. For other recent applications of the GP, see Chan et al. [10], Wan and Chan [41], Chan et al. [8,9], Pekalp and Aydoğdu [29,31,32], Zhang and Wang [44], Arnold et al. [1], Pekalp et al. [30], and Rasay et al. [33].

For a GP, some important parameters of the model are estimated statistically by observing the inter-arrival times of the process. Let $\{X_k,\ k = 1, 2, \ldots\}$ be the inter-arrival times of a GP with trend parameter a, where $E(X_1) = \mu$ and $Var\,(X_1) = \sigma^2$. Then, $E\,(X_k) = \mu/a^{k-1}$, and $Var\,(X_k) = \sigma^2/a^{2(k-1)}$, for $k = 1, 2, \ldots$. The inference problem for a, μ and σ^2 has been well studied in the literature. Lam [20] obtained non-parametric estimators of these parameters using the linear regression method. Parametric inference for a GP, when X_1 follows important failure distributions such as exponential, lognormal, Weibull and gamma, has been studied in detail by Kara [14], Lam and Chan [23], Aydoğdu et al. [3], Chan et al. [7], Kara et al. [15], respectively. For other inferential studies about the GP, see Biçer et al. [5], Biçer et al. [6], Lone et al. [26], Usta [39], and Yılmaz [45]. In these studies, statistical inference was concerned only with single process data. In some reliability problems, not only single process data but also multiple process data may arise. The literature includes studies on the use and statistical analysis of multiple process data for various counting process models. For example, Garmabaki et al. [12,13] studied the reliability modeling of multiple repairable systems by utilizing RP, HPP, or NHPP as models for the repair processes of multiple units, which generate multiple process data. Further, Na and Chang [27] and Wang et al. [42] considered the statistical analysis of multiple process data coming from the repair processes of multiple systems by modeling the repair processes with NHPP. However, to the best of our knowledge, there is no study in the literature that explores the usage and statistical analysis of GPs for multiple process data, despite their potential and ease of implementation for processes with a trend. In this study, we statistically evaluate multiple process data for a GP by assuming that all processes are homogeneous and inter-arrival times follow an exponential distribution. We then compare it with single process data from an effectiveness perspective and illustrate its use in modeling multiple repairable systems as a novel approach.

The rest of the paper is organized as follows. Data structures for multiple GPs are presented in Section 2. In Section 3, the maximum likelihood (ML) and modified maximum likelihood (MML) estimators for the parameters of the GP are derived, along with their asymptotic properties. Further, some test statistics are proposed to distinguish the GP from the RP, and to test the homogeneity of the samples. A Monte Carlo simulation is carried out in Section 4 to demonstrate the performance of the inferential procedures for both multiple and single process data, for comparative purposes. Data analysis examples are illustrated in Section 5. Finally, some conclusions are provided in Section 6.

2. Multiple process data for GP

Let us consider the GP as a repair model for a repairable system. Assume that the system will be repaired after each failure and that the successive operating times of the system after each repair decrease stochastically. Therefore, the GP can be utilized to model this repair process. Let the successive operating times of a repairable system be modeled by a GP with trend parameter a. To estimate the trend parameter a, as well as the mean and variance of each operating time after repair, a sufficient number of successive operating times must be observed throughout the repair process. Note that the observation process of successive operating times after each repair may take a long time. To reduce the total observation time, multiple repair processes based on identical multiple systems can be observed simultaneously, rather than just a single repair process. On the other hand, some repairable systems may become inefficient after a few repairs due to decreased operating times and increased repair times. In such cases, the repair process is terminated after some repairs. To obtain a sufficient number of successive operating times after each repair, multiple repair processes based on multiple systems should be observed simultaneously. This phenomenon suggests the use of multiple process data to make inferences for a process modeled by the GP. For examples of multiple process data modeling with some counting processes, see Garmabaki et al. [12,13], Na and Chang [27], and Wang et al. [42].

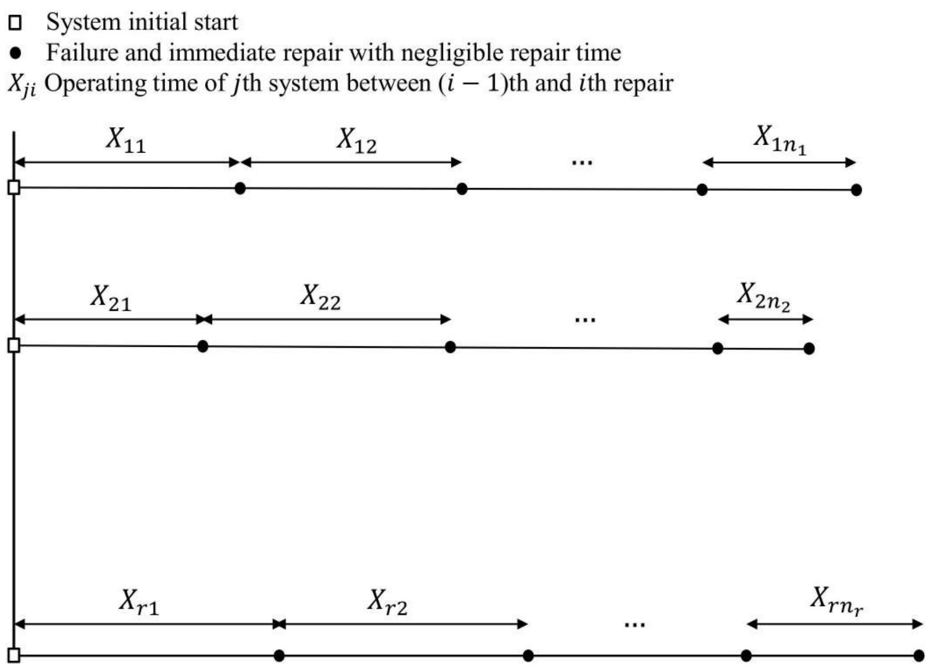

Figure 1. Repair processes of r systems with a predetermined number of failures.

Suppose that we have r units of identical repairable systems, whose repair processes are modeled with the GP. The trend parameter of the repair process, along with the mean and variance of operating times for the systems after each repair, can be estimated based on multiple process data obtained throughout the r repair processes if all repair processes are assumed to be homogeneous. When the repair processes of r systems are observed simultaneously, two observation schemes arise depending on the decision to end the observation. First, each repair process is observed until a predetermined number of failures occur for each unit. Second, all repair processes are observed until a predetermined time. In the first case, successive operating times for all units are complete. In the second case, operating times for systems functioning at the predetermined time are right-censored, whereas the previous operating times for each system are complete. Hereafter, we refer to these observation schemes as Case I and Case II, respectively. These observation schemes are illustrated in Figure 1 and Figure 2.

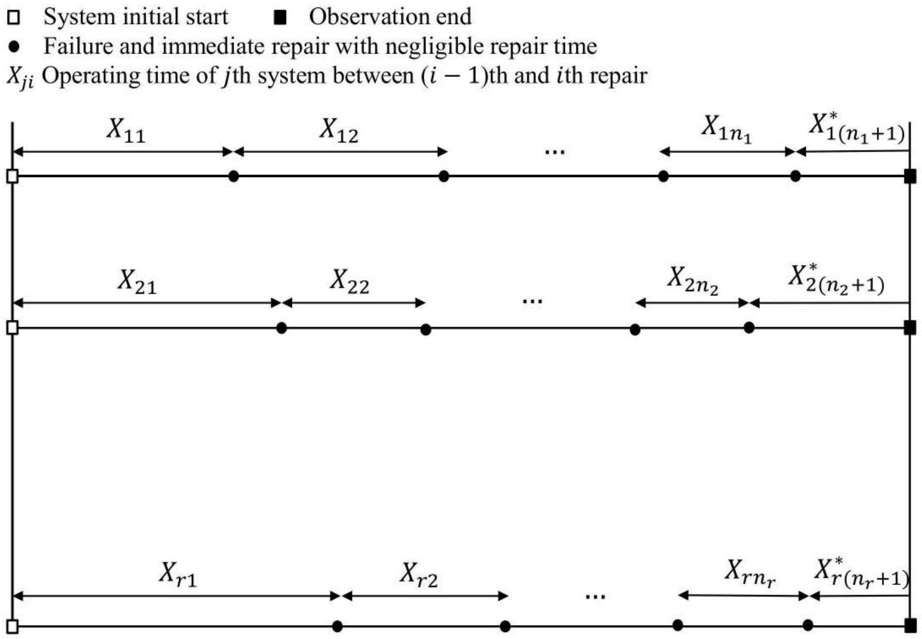

Figure 2. Repair processes of r systems with a predetermined ending time.

3. Statistical inference for multiple process data

Assume that the inter-arrival times of r independent GPs with the same trend parameter a are observed simultaneously, and that the first inter-arrival times of all processes are distributed according to an identical exponential distribution. The statistical inference problem for the trend parameter as well as the mean and variance of each inter-arrival time based on multiple process data, should be studied separately for the two different observation schemes mentioned above.

3.1. Inference for Case I

Let r independent and identical GPs with trend parameter a be observed with predetermined number of inter-arrival times such as $n_j \in \mathbb{N}, j = 1, \ldots, r$. Let $\{X_{ji}, i = 1, .., n_j\}, j = 1, \ldots, r$ be the inter-arrival times of jth process, X_{j1} has distribution function

$$F(x) = 1 - \exp(-x/\theta), x \geq 0; \theta > 0,$$

with $\mu = E(X_{j1}) = \theta$, $\sigma^2 = Var(X_{j1}) = \theta^2$ for $j = 1, \ldots, r$. The distribution function and probability density function of X_{ji} are given as

$$F_i(x) = 1 - \exp\left(-a^{i-1}x/\theta\right), x \geq 0; \theta > 0, i = 1, .., n_j,$$

$$f_i(x) = \exp\left(-a^{i-1}x/\theta\right) a^{i-1}/\theta, x \geq 0; \theta > 0, i = 1, .., n_j$$

for $j = 1, \ldots, r$. Therefore, mean and variance of X_{ji} are $\mu_i = E(X_{ji}) = \mu/a^{i-1}$, $\sigma_i^2 = Var(X_{ji}) = \sigma^2/a^{2(i-1)}$, $i = 1, .., n_j$ for $j = 1, \ldots, r$.

Let $\mathbf{X} = \{X_{ji},\ i = 1, \ldots, n_j,\ j = 1, \ldots, r\}$ denote the set of inter-arrival times, and let $\mathbf{x}$ be the sample points of $\mathbf{X}$. Then, the likelihood function for the parameters a and θ based on the multiple process data $\mathbf{x}$ is given by

$$L(a, \theta; \mathbf{x}) = \prod_{j=1}^{r} \prod_{i=1}^{n_j} f_i(x_{ji}) = \prod_{j=1}^{r} \prod_{i=1}^{n_j} \exp\left(-a^{i-1}x_{ji}/\theta\right) \frac{a^{i-1}}{\theta}.$$

The log-likelihood function is

$$\ln L(a, \theta; \mathbf{x}) = \ln a \sum_{j=1}^{r} \sum_{i=1}^{n_j} (i-1) - \ln \theta \sum_{j=1}^{r} n_j - \frac{1}{\theta} \sum_{j=1}^{r} \sum_{i=1}^{n_j} a^{i-1} x_{ji}.$$

The likelihood equations are obtained by taking partial derivatives of the log-likelihood function with respect to a and θ as

$$\frac{\partial \ln L(a, \theta; \mathbf{x})}{\partial a} = \frac{1}{a} \sum_{j=1}^{r} \sum_{i=1}^{n_j} (i-1) - \frac{1}{\theta} \sum_{j=1}^{r} \sum_{i=1}^{n_j} (i-1) a^{i-2} x_{ji} = 0, \tag{3.1}$$

$$\frac{\partial \ln L(a, \theta; \mathbf{x})}{\partial \theta} = -\frac{1}{\theta} \sum_{j=1}^{r} n_j + \frac{1}{\theta^2} \sum_{j=1}^{r} \sum_{i=1}^{n_j} a^{i-1} x_{ji} = 0. \tag{3.2}$$

The ML estimators of a and θ are derived by solving Equations (3.1) and (3.2) simultaneously as follows. First, $\widehat{a}$ is obtained by solving the equation

$$\sum_{j=1}^{r} \sum_{i=1}^{n_j} a^{i-1} x_{ji} \left[\frac{\sum_{j=1}^{r} n_j^2}{\sum_{j=1}^{r} n_j} - 2i + 1 \right] = 0, \tag{3.3}$$

numerically, and

$$\widehat{\theta} = \left[\sum_{j=1}^{r} \sum_{i=1}^{n_j} \widehat{a}^{i-1} x_{ji} \right] / \left[\sum_{j=1}^{r} n_j \right]. \tag{3.4}$$

Therefore, the ML estimators for the mean and variance of X_{ji} for $j = 1, \ldots, r$ are

$$\widehat{\mu}_i = \frac{\widehat{\theta}}{\widehat{a}^{i-1}},\ \widehat{\sigma}_i^2 = \frac{\widehat{\theta}^2}{\widehat{a}^{2(i-1)}},\ i = 1, .., n_j. \tag{3.5}$$

Considering the asymptotic properties of the ML estimators, we have the following asymptotic distributions for the estimators $\widehat{a}$ and $\widehat{\theta}$.

Theorem 3.1. *As at least one $n_j \to \infty$ for $j = 1, ..., r$;*

$$\begin{bmatrix} \widehat{a} \\ \widehat{\theta} \end{bmatrix} \sim N\left(\begin{bmatrix} a \\ \theta \end{bmatrix}, I^{-1}(a, \theta; \mathbf{X}) \right) \tag{3.6}$$

where N stands for normal distribution and $I(a, \theta; \mathbf{X})$ is Fisher information such that

$$I(a, \theta; \mathbf{X}) = \begin{bmatrix} \frac{1}{3a^2} \sum_{j=1}^{r} n_j^3 & -\frac{1}{2a\theta} \sum_{j=1}^{r} n_j^2 \\ -\frac{1}{2a\theta} \sum_{j=1}^{r} n_j^2 & \frac{1}{\theta^2} \sum_{j=1}^{r} n_j \end{bmatrix}. \tag{3.7}$$

Proof. Second partial derivatives of the log-likelihood function are

$$\frac{\partial^2 \ln L(a,\theta;\mathbf{X})}{\partial a^2} = -\frac{1}{a^2}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1) - \frac{1}{\theta}\frac{1}{a^2}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1)(i-2)a^{i-1}X_{ji},$$

$$\frac{\partial^2 \ln L(a,\theta;\mathbf{X})}{\partial \theta^2} = \frac{1}{\theta^2}\sum_{j=1}^{r} n_j - \frac{2}{\theta^3}\sum_{j=1}^{r}\sum_{i=1}^{n_j} a^{i-1}X_{ji},$$

$$\frac{\partial^2 \ln L(a,\theta;\mathbf{X})}{\partial a\partial \theta} = \frac{1}{\theta^2}\frac{1}{a}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1)a^{i-1}X_{ji}.$$

Then, the expected values of negative second partial derivatives are

$$\begin{aligned} E\left(-\frac{\partial^2 \ln L(a,\theta;\mathbf{X})}{\partial a^2}\right) &= \frac{1}{a^2}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1) + \frac{1}{\theta}\frac{1}{a^2}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1)(i-2)E\left(a^{i-1}X_{ji}\right) \\ &= \frac{1}{a^2}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1) + \frac{1}{a^2}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1)(i-2) \\ &= \frac{1}{a^2}\sum_{j=1}^{r}\frac{n_j(n_j-1)(2n_j-1)}{6} \approx \frac{1}{3a^2}\sum_{j=1}^{r} n_j^3, \end{aligned}$$

$$E\left(-\frac{\partial^2 \ln L(a,\theta;\mathbf{X})}{\partial \theta^2}\right) = -\frac{1}{\theta^2}\sum_{j=1}^{r} n_j + \frac{2}{\theta^3}\sum_{j=1}^{r}\sum_{i=1}^{n_j} E\left(a^{i-1}X_{ji}\right) = \frac{1}{\theta^2}\sum_{j=1}^{r} n_j,$$

$$\begin{aligned} E\left(-\frac{\partial^2 \ln L(a,\theta;\mathbf{X})}{\partial a\partial \theta}\right) &= -\frac{1}{\theta^2}\frac{1}{a}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1)E\left(a^{i-1}X_{ji}\right) = -\frac{1}{\theta}\frac{1}{a}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1) \\ &= -\frac{1}{\theta}\frac{1}{a}\sum_{j=1}^{r}\frac{n_j(n_j-1)}{2} \approx -\frac{1}{2a\theta}\sum_{j=1}^{r} n_j^2, \end{aligned}$$

since $E\left(a^{i-1}X_{ji}\right) = \theta$, $j = 1, \ldots, r$. The Fisher information in Equation (3.7) is readily obtained and the result is clear. □

Corollary 3.2. *As at least one* $n_j \to \infty$ *for* $j = 1, ..., r$*;*

$$\widehat{a} \sim N\left(a, \frac{12a^2\sum_{j=1}^{r} n_j}{4[\sum_{j=1}^{r} n_j^3][\sum_{j=1}^{r} n_j] - 3\left[\sum_{j=1}^{r} n_j^2\right]^2}\right), \tag{3.8}$$

$$\widehat{\theta} \sim N\left(\widehat{\theta}, \frac{4\theta^2\sum_{j=1}^{r} n_j^3}{4[\sum_{j=1}^{r} n_j^3][\sum_{j=1}^{r} n_j] - 3\left[\sum_{j=1}^{r} n_j^2\right]^2}\right). \tag{3.9}$$

The asymptotic distributions of $\widehat{\mu}_i$ and $\widehat{\sigma}_i^2$, $i = 1, .., n_j$ for $j = 1, \ldots, r$ are therefore obtained using the delta method as follows:

Corollary 3.3. *As at least one* $n_j \to \infty$ *for* $j = 1, ..., r$*;*

$$\widehat{\mu}_i \sim N\left(\frac{\theta}{a^{i-1}}, \nu\right), \; i = 1, \ldots, n_j, \tag{3.10}$$

$$\widehat{\sigma}_i^2 \sim N\left(\frac{\theta^2}{a^{2(i-1)}}, \xi\right), \; i = 1, \ldots, n_j, \tag{3.11}$$

where

$$\nu = \frac{a^{2(1-i)}\theta^2\left[12(1-i)^2\sum_{j=1}^r n_j + 12(1-i)\sum_{j=1}^r n_j^2 + 4\sum_{j=1}^r n_j^3\right]}{4[\sum_{j=1}^r n_j^3][\sum_{j=1}^r n_j] - 3\left[\sum_{j=1}^r n_j^2\right]^2},$$

$$\xi = \frac{a^{4(1-i)}\theta^4[48(1-i)^2\sum_{j=1}^r n_j + 48(1-i)\sum_{j=1}^r n_j^2 + 16\sum_{j=1}^r n_j^3]}{4[\sum_{j=1}^r n_j^3][\sum_{j=1}^r n_j] - 3\left[\sum_{j=1}^r n_j^2\right]^2}.$$

If we take $i = 1$, then it is clear that;

$$\widehat{\mu} \sim N\left(\theta, \frac{4\theta^2\sum_{j=1}^r n_j^3}{4[\sum_{j=1}^r n_j^3][\sum_{j=1}^r n_j] - 3\left[\sum_{j=1}^r n_j^2\right]^2}\right), \tag{3.12}$$

$$\widehat{\sigma}^2 \sim N\left(\theta^2, \frac{16\theta^4\sum_{j=1}^r n_j^3}{4[\sum_{j=1}^r n_j^3][\sum_{j=1}^r n_j] - 3\left[\sum_{j=1}^r n_j^2\right]^2}\right). \tag{3.13}$$

These asymptotic distributions allow to construct approximate confidence intervals for the parameters of interest, to distinguish GP from RP by testing whether $a = 1$ or $a \neq 1$, and to test the homogeneity of processes.

Proposition 3.4. *To test $H_0 : a = 1$ vs $H_1 : a \neq 1$, let us define the test statistic*

$$S_1 = \sqrt{\frac{4[\sum_{j=1}^r n_j^3][\sum_{j=1}^r n_j] - 3\left[\sum_{j=1}^r n_j^2\right]^2}{12\widehat{a}^2\sum_{j=1}^r n_j}}\,(\widehat{a} - 1)\,. \tag{3.14}$$

If $|S_1| > z_{\alpha/2}$, then, the null hypothesis is rejected at significance level α since $S_1 \sim N(0,1)$ asymptotically, under H_0. Here, $z_{\alpha/2}$ denotes the upper $\alpha/2$ quantile of the standard normal distribution.

The above inferences were obtained under the assumption that all processes are homogeneous. If multiple process data are observed on identical units, it is reasonable to assume that the random variables $X_{j1}, j = 1, 2, ..., r$ are identically distributed. However, the homogeneity of trends should be tested. The homogeneity of the trend parameters in all processes can be tested using the following proposition.

Proposition 3.5. *Let $\{X_{ji}, i = 1, 2, \ldots, n_j\}$ be the inter-arrival times of GPs with trend parameter a_j for $j = 1, 2, \ldots, r$. Assume that the random variables $X_{j1}, j = 1, 2, \ldots, r$ follow identical exponential distributions with the mean θ. To test $H_0 : a_1 = a_2 = \cdots = a_r$ against $H_1 : \exists a_i \neq a_j, i, j \in \{1, 2, \ldots, r\}, i \neq j$, let us define the test statistic*

$$T_1 = \frac{1}{r-1}\sum_{j=1}^r\left(\frac{\hat{a}_j - \bar{\hat{a}}}{\sqrt{12\hat{a}^2/n_j^3}}\right)^2. \tag{3.15}$$

Then, the null hypothesis is rejected at the significance level α if $T_1 > \chi^2_{r-1}(\alpha)$, since $T_1 \sim \chi^2_{r-1}$ asymptotically under H_0. Here, χ^2_{r-1} denotes the chi-square distribution with degrees of freedom $r-1$, and $\chi^2_{r-1}(\alpha)$ denotes its α upper quantile, $\hat{a}_j$ is the ML estimator of a_j for the jth process, $\bar{\hat{a}} = \frac{1}{r}\sum_{j=1}^r \hat{a}_j$, and $\hat{a}$ is the ML estimator based on the multiple process data for the common trend parameter a of all processes.

To compute the ML estimate of the parameter a for a given multiple process data, the Equation (3.3) must be solved numerically using numerical methods such as Newton-Raphson or bisection. However, these methods are sensitive to initial value of the algorithm and thus may lead to undesirable results such as non-convergence or convergence to an incorrect value. Tiku [35] proposed a methodology, known as the MML, where intractable terms of likelihood equations are linearized by Taylor series approximation to obtain explicit solutions. This approach is also known as the approximate ML method. For further details on the MML methodology, see Tiku [36], Tiku et al. [37], Tiku and Akkaya [38]. See also Aydoğdu et al. [3] and Kara et al. [15] for application of the MML methodology in a GP based on a single process data with Weibull and gamma inter-arrival times. To obtain MML estimators of the parameters a and θ, let us write $Y_{ji} = a^{i-1}X_{ji}$, $i = 1, \ldots, n_j$, $j = 1, \ldots, r$. From the definition of GP, the random variables Y_{ji} are iid with the distribution function $F(y) = 1 - \exp(-y/\theta), y > 0; \theta > 0$. Let $\ln Y_{ji} = (i-1)\ln a + \ln X_{ji}$, $i = 1, \ldots, n_j$, $j = 1, \ldots, r$, then the random variables $\ln Y_{ji}$ have an identical extreme value distribution with the distribution function $G(y) = 1 - \exp(-\exp(y-\delta))$, $y \in \mathbb{R}$, where $\delta = \ln\theta$. Hence, likelihood function based on logarithmically transformed data $\ln \mathbf{x}$ is

$$L(a, \delta; \ln \mathbf{x}) = \prod_{j=1}^{r}\prod_{i=1}^{n_j} \exp\left((i-1)\ln a + \ln x_{ji} - \delta - \exp\left((i-1)\ln a + \ln x_{ji} - \delta\right)\right).$$

Let $Z_{ji} = (i-1)\ln a + \ln X_{ji} - \delta$. Then, the random variables Z_{ji} have a standard extreme value distribution. If we take the logarithm of the likelihood function $L(a, \delta; \ln \mathbf{x})$ and then take partial derivatives with respect to a and δ, the likelihood equations are obtained as:

$$\frac{\partial \ln L(a, \delta; \ln \mathbf{x})}{\partial a} = \sum_{j=1}^{r}\sum_{i=1}^{n_j} c_{ji} - \sum_{j=1}^{r}\sum_{i=1}^{n_j} c_{ji}\exp(z_{ji}) = 0,$$

$$\frac{\partial \ln L(a, \delta; \ln \mathbf{x})}{\partial \delta} = \sum_{j=1}^{r}\sum_{i=1}^{n_j} \exp(z_{ji}) - \sum_{j=1}^{r} n_j = 0,$$

where $c_{ji} = (i-1)$, $z_{ji} = \ln x_{ji} + c_{ji}\beta - \delta$, $\beta = \ln a$, $i = 1, \ldots, n_j$, $j = 1, \ldots, r$. By using ordered terms and linearizing the intractable term $\exp(z_{ji})$ around the mean of the ordered standard extreme value variate, the modified likelihood equations are obtained as:

$$\frac{\partial \ln L^*(a, \delta; \ln \mathbf{x})}{\partial a} = \sum_{j=1}^{r}\sum_{i=1}^{n_j} c_{j[i]} - \sum_{j=1}^{r}\sum_{i=1}^{n_j} c_{j[i]}(a_{j(i)} + b_{j(i)}z_{j(i)}) = 0, \tag{3.16}$$

$$\frac{\partial \ln L^*(a, \delta; \ln \mathbf{x})}{\partial \delta} = \sum_{j=1}^{r}\sum_{i=1}^{n_j} (a_{j(i)} + b_{j(i)}z_{j(i)}) - \sum_{j=1}^{r} n_j = 0, \tag{3.17}$$

where $z_{j(i)} = \ln x_{j[i]} + c_{j[i]}\beta - \delta$, $a_{j(i)} = \exp\left(t_{(i)}\right)\left(1 - t_{(i)}\right)$, $b_{j(i)} = \exp\left(t_{(i)}\right)$, $t_{(i)} = E\left(Z_{j(i)}\right)$, $i = 1, \ldots, n_j$, $j = 1, \ldots, r$ and $Z_{j(i)}$ are ordered variates of Z_{ji} for each $j = 1, \ldots, r$. Note that $\left(x_{j[i]}, c_{j[i]}\right)$ are concomitants of ordered variates $z_{j(i)}$ for each $j = 1, \ldots, r$. For a detailed explanation of concomitants, see Tiku and Akkaya [38]. The expected values of the ordered standard extreme value variate are exactly calculated for each $j = 1, \ldots, r$ as

$$t_{(i)} = -\gamma - (n_j + 1 - i)\binom{n_j}{n_j + 1 - i}\sum_{k=0}^{i-1}\binom{i-1}{k}(-1)^k \ln(n_j + k + 1 - i)/(n_j + k + 1 - i),$$

$i = 1, \ldots, n_j$, where $\gamma = 0.5772$ is the Euler constant; see White [43]. In addition, it can be calculated approximately for $n_j \geq 10$, $j = 1, \ldots, r$ by inverting the distribution function of the standard extreme value variate as $t_{(i)} \cong \ln[-\ln[1 - i/(n_j + 1)]]$, $i = 1, \ldots, n_j$. Now, we are ready to explicitly obtain the MML estimators, which are given in the following proposition.

Proposition 3.6. *The MML estimators for the parameters β and δ are calculated from Equation (3.16) and (3.17) as*

$$\widetilde{\beta} = \frac{A}{C} - \frac{B}{C}, \tag{3.18}$$

$$\widetilde{\delta} = D / \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)}\right], \tag{3.19}$$

where

$$A = \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)} c_{j[i]}\right] \left\{ \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} (a_{j(i)} - 1)\right] + \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)} \ln x_{j[i]}\right] \right\},$$

$$B = \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)}\right] \left\{ \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} \left(a_{j(i)} - 1\right) c_{j[i]}\right] + \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)} c_{j[i]} \ln x_{j[i]}\right] \right\},$$

$$C = \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)}\right] \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)} c_{j[i]}^2\right] - \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)} c_{j[i]}\right]^2,$$

$$D = \sum_{j=1}^{r}\sum_{i=1}^{n_j} (a_{j(i)} - 1) + \sum_{j=1}^{r}\sum_{i=1}^{n_j} (b_{j(i)} \ln x_{j[i]}) + \widetilde{\beta} \sum_{j=1}^{r}\sum_{i=1}^{n_j} b_{j(i)} c_{j[i]}.$$

Therefore, the MML estimators for a, θ, μ_i and $\sigma_i^2, i = 1, \ldots, n_j, j = 1, \ldots, r$ are

$$\widetilde{a} = \exp\left(\widetilde{\beta}\right), \widetilde{\theta} = \exp\left(\widetilde{\delta}\right), \tag{3.20}$$

$$\widetilde{\mu}_i = \frac{\widetilde{\theta}}{\widetilde{a}^{i-1}}, \ \widetilde{\sigma}_i^2 = \frac{\widetilde{\theta}^2}{\widetilde{a}^{2(i-1)}}. \tag{3.21}$$

These estimators are asymptotically equivalent to ML estimators as $n_j \to \infty$, $j = 1, \ldots, r$. Therefore, their asymptotic distributions are the same as those of the ML estimators; see Bhattacharyya [4], Vaughan and Tiku [40]. The asymptotic variances given in Equations (3.8) and (3.9) are known as minimum variance bounds (MVB) when estimating the parameters a and θ. To illustrate the efficiency of the MML estimators, see Table 1, which presents the simulated variances for both the ML and MML estimators, as well as the MVBs. The variances and MVBs for $a = 0.95$ and $\theta = 1$ over 10^4 replications have been computed in MATLAB. Note that the ML and MML estimators for the single process data can be easily obtained by setting $r = 1$. In Table 1, it is evident that the variances of the ML and MML estimators are similar and decrease as the sample size increases. Therefore, we can conclude that the MML estimators are as efficient as the ML estimators.

Proposition 3.7. *Let* $\mathbf{X} = \{X_{ji},\ i = 1, \ldots, n_j,\ j = 1, \ldots, r\}$ *be the inter-arrival times of multiple GPs with trend parameter a, $X = \{X_i, i = 1, \ldots, n\}$ be inter-arrival times of a single GP with the same trend parameter a, and $X_1, X_{j1},\ j = 1, \ldots, r$ have exponential distribution with mean θ. Assume that $\sum_{j=1}^{r} n_j = n$. Then, relation for Fisher information of a and θ contained in* $\mathbf{X}$ *and X are*

$$I\left(a; \mathbf{X}\right) < I(a; X),$$
$$I\left(\theta; \mathbf{X}\right) = I(\theta; X).$$

Proposition 3.7 emphasizes that the observation of multiple GPs, compared to a single GP with an equal number of inter-arrival times, leads to information loss for a but does not affect the information for θ.

Table 1. Simulated variances for the ML and MML estimators, and the MVBs.

	Simulated variances for ML estimators		Simulated variances for MML estimators		MVBs for ML estimators	
	a	θ	a	θ	a	θ
$r=1,\ n=30$	0.0004	0.1374	0.0005	0.1382	0.0004	0.1269
$r=2,\ n=(15,15)$	0.0017	0.1377	0.0021	0.1364	0.0016	0.1208
$r=3,\ n=(10,10,10)$	0.0040	0.1382	0.0053	0.1355	0.0036	0.1152
$r=1,\ n=50$	0.0001	0.0891	0.0001	0.0889	0.0001	0.0776
$r=2,\ n=(25,25)$	0.0004	0.0822	0.0004	0.0828	0.0003	0.0754
$r=3,\ n=(17,17,17)$	0.0008	0.0772	0.0009	0.0770	0.0007	0.0719
$r=1,\ n=100$	0.0000	0.0404	0.0000	0.0404	0.0000	0.0394
$r=2,\ n=(50,50)$	0.0000	0.0409	0.0000	0.0409	0.0000	0.0388
$r=3,\ n=(33,33,33)$	0.0001	0.0408	0.0001	0.0407	0.0001	0.0386

3.2. Inference for Case II

Let r independent and identical GPs with trend parameter a be observed until a predetermined time $T>0$. Let $\{X_{ji},\ i=1,..,n_j+1\}$, $j=1,\dots,r$ be the inter-arrival times of the jth process, X_{j1} has the distribution function $F(x)=1-e^{-x/\theta}, x\geq 0; \theta>0$, $\mu=E\left(X_{j1}\right)=\theta$, $\sigma^2=Var\left(X_{j1}\right)=\theta^2$ for $j=1,\dots,r$. The distribution function, probability density function, mean and variance of X_{ji} for $i=1,\dots,n_j+1$, $j=1,\dots,r$ can be found as in Case I. Note that the inter-arrival times $\{X_{ji},\ i=1,..,n_j\}$ are complete, while $\{X_{j(n_j+1)}\}$ are right-censored for each $j=1,\dots,r$ as demonstrated in Figure 2.

Let us denote $\mathbf{X}_{com}=\{X_{ji},\ i=1,\dots,n_j,\ j=1,\dots,r\}$, $\mathbf{X}_{cens}=\{X_{j(n_j+1)},\ j=1,\dots,r\}$, $\mathbf{X}_t=(\mathbf{X}_{com},\ \mathbf{X}_{cens})$, and $\mathbf{x}_t$ as the sample points of $\mathbf{X}_t$. Then, the likelihood function for the parameters a and θ based on the multiple process data $\mathbf{x}_t$ is given by

$$\begin{aligned} L\left(a,\theta;\mathbf{x}_t\right) &= \prod_{j=1}^{r}\left[\prod_{i=1}^{n_j} f_i\left(x_{ji}\right)\right]\left[1-F_{n_j+1}\left(x_{j(n_j+1)}\right)\right] \\ &= \prod_{j=1}^{r}\left[\prod_{i=1}^{n_j}\exp\left(-a^{i-1}x_{ji}/\theta\right)\frac{a^{i-1}}{\theta}\right]\left[\exp\left(-a^{i-1}x_{j(n_j+1)}/\theta\right)\frac{a^{i-1}}{\theta}\right]. \end{aligned}$$

The log-likelihood function is derived as

$$\ln L\left(a,\theta;\mathbf{x}_t\right)=\ln a\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1)-\ln\theta\sum_{j=1}^{r}n_j-\frac{1}{\theta}\sum_{j=1}^{r}\sum_{i=1}^{n_j+1}a^{i-1}x_{ji}.$$

By taking partial derivatives of the log-likelihood function with respect to a and θ, the likelihood equations are obtained as follows:

$$\frac{\partial\ln L\left(a,\theta;\mathbf{x}_t\right)}{\partial a}=\frac{1}{a}\sum_{j=1}^{r}\sum_{i=1}^{n_j}(i-1)-\frac{1}{\theta}\sum_{j=1}^{r}\sum_{i=1}^{n_j+1}(i-1)a^{i-2}x_{ji}=0,$$

$$\frac{\partial\ln L\left(a,\theta;\mathbf{x}_t\right)}{\partial\theta}=-\frac{1}{\theta}\sum_{j=1}^{r}n_j+\frac{1}{\theta^2}\sum_{j=1}^{r}\sum_{i=1}^{n_j+1}a^{i-1}x_{ji}=0.$$

The ML estimator $\widehat{a}$ is first obtained as numerical solution of the equation

$$\sum_{j=1}^{r}\sum_{i=1}^{n_j+1} a^{i-1}x_{ji}\left[\frac{\sum_{j=1}^{r} n_j^2}{\sum_{j=1}^{r} n_j} - 2i + 1\right] = 0, \tag{3.22}$$

then, the ML estimator $\widehat{\theta}$ is

$$\widehat{\theta} = \left[\sum_{j=1}^{r}\sum_{i=1}^{n_j+1} \widehat{a}^{i-1}x_{ji}\right] / \left[\sum_{j=1}^{r} n_j\right]. \tag{3.23}$$

Hence, the ML estimators for the mean and variance of X_{ji} for $j = 1, \dots, r$ are

$$\widehat{\mu}_i = \frac{\widehat{\theta}}{\widehat{a}^{i-1}},\ \widehat{\sigma}_i^2 = \frac{\widehat{\theta}^2}{\widehat{a}^{2(i-1)}},\ i = 1, .., n_j + 1. \tag{3.24}$$

The asymptotic distributions for the estimators $\widehat{a}$ and $\widehat{\theta}$ are given below.

Theorem 3.8. *As $T \to \infty$;*

$$\begin{bmatrix}\widehat{a}\\ \widehat{\theta}\end{bmatrix} \sim N\left(\begin{bmatrix}a\\ \theta\end{bmatrix}, I^{-1}(a,\theta;\mathbf{X}_t)\right), \tag{3.25}$$

where

$$I(a,\theta;\mathbf{X}_t) = \begin{bmatrix} \frac{1}{a^2}\sum_{j=1}^{r}\left(n_j^3/3 + n_j^2F_{n_j+1}(t_j)\right) & -\frac{1}{a\theta}\sum_{j=1}^{r}\left(n_j^2/2 + n_jF_{n_j+1}(t_j)\right)\\ -\frac{1}{a\theta}\sum_{j=1}^{r}\left(n_j^2/2 + n_jF_{n_j+1}(t_j)\right) & \frac{1}{\theta^2}\sum_{j=1}^{r}\left(n_j + F_{n_j+1}(t_j)\right)\end{bmatrix},$$

t_j is censoring time of $X_{j(n_j+1)}$ such that $t_j = x_{j(n_j+1)} = T - \sum_{i=1}^{n_j} x_{ji}$, $F_{n_j+1}(t_j) = 1 - \exp(-a^{n_j}t_j/\theta)$ for $j = 1, \dots, r$.

Proof. The Fisher information can be written as

$$I(a,\theta;\mathbf{X}_t) = I(a,\theta;\mathbf{X}_{com}) + I(a,\theta;\mathbf{X}_{cens})$$

since $\mathbf{X}_t = (\mathbf{X}_{com}, \mathbf{X}_{cens})$. The Fisher information based on complete observations $I(a,\theta;\mathbf{X}_{com})$ is equal to one given in Equation (3.6). Then, we only need to find $I(a,\theta;\mathbf{X}_{cens})$ to obtain $I(a,\theta;\mathbf{X}_t)$. For a non-negative censored variable Y with censoring time L, the Fisher information can be computed as

$$I(\theta;Y) = \int_0^L \left(\frac{\partial}{\partial\theta}\ln h(x)\right)^2 f(x)dx,$$

where f is pdf and h is hazard function of Y, see Zheng and Gastwirth [46] and Park et al. [28]. Then, for censored variates $X_{j(n_j+1)}$ with censoring times t_j, $j = 1, \dots, r$, we obtain;

$$\int_0^{t_j}\left(\frac{\partial}{\partial a}\ln h_{n_j+1}(x)\right)^2 f_{n_j+1}(x)dx = \frac{n_j^2}{a^2}F_{n_j+1}(t_j),$$

$$\int_0^{t_j}\left(\frac{\partial}{\partial\theta}\ln h_{n_j+1}(x)\right)^2 f_{n_j+1}(x)dx = \frac{1}{\theta^2}F_{n_j+1}(t_j),$$

$$\int_0^{t_j}\left(\frac{\partial}{\partial a}\ln h_{n_j+1}(x)\right)\left(\frac{\partial}{\partial\theta}\ln h_{n_j+1}(x)\right) f_{n_j+1}(x)dx = -\frac{n_j}{a\theta}F_{n_j+1}(t_j).$$

Therefore;

$$I(a,\theta;\mathbf{X}_{cens}) = \begin{bmatrix} \frac{1}{a^2}\sum_{j=1}^{r} n_j^2F_{n_j+1}(t_j) & -\frac{1}{a\theta}\sum_{j=1}^{r} n_jF_{n_j+1}(t_j)\\ -\frac{1}{a\theta}\sum_{j=1}^{r} n_jF_{n_j+1}(t_j) & \frac{1}{\theta^2}\sum_{j=1}^{r} F_{n_j+1}(t_j)\end{bmatrix},$$

and the result is obtained. Note that censoring time for the variable $X_{j(n_j+1)}$ is in fact random and equal to $T-\sum_{i=1}^{n_j} X_{ji}$. For the sake of simplicity, we consider it as constant such that $t_j = T-\sum_{i=1}^{n_j} x_{ji}$ given $X_{ji}=x_{ji}$.

□

Corollary 3.9. *As $T\to\infty$;*

$$\widehat{a} \sim N\left(a, a^2\tau_{11}\right), \tag{3.26}$$

$$\widehat{\theta} \sim N\left(\theta, \theta^2\tau_{22}\right), \tag{3.27}$$

where

$$\tau_{11} = \frac{K}{LK-M^2}, \tau_{22} = \frac{L}{LK-M^2},$$

$$K=\sum_{j=1}^{r}\left(n_j+F_{n_j+1}\left(t_j\right)\right), L=\sum_{j=1}^{r}\left(n_j^3/3+n_j^2F_{n_j+1}\left(t_j\right)\right), M=\sum_{j=1}^{r}\left(n_j^2/2+n_jF_{n_j+1}\left(t_j\right)\right).$$

The asymptotic distributions of $\widehat{\mu}_i$ and $\widehat{\sigma}_i^2$, $i=1,..,n_j+1$ for $j=1,\ldots,r$ can be obtained using the delta method similar to Case I.

Corollary 3.10. *As $T\to\infty$;*

$$\widehat{\mu}_i \sim N\left(\frac{\theta}{a^{i-1}}, \nu\right),\ i=1,\ldots,n_j+1, \tag{3.28}$$

$$\widehat{\sigma}_i^2 \sim N\left(\frac{\theta^2}{a^{2(i-1)}}, \xi\right),\ i=1,\ldots,n_j+1, \tag{3.29}$$

where

$$\nu = \frac{\theta^2(1-i)^2}{a^{2(i-1)}}\tau_{11} + 2\frac{\theta^2(1-i)}{a^{2(i-1)}}\tau_{12} + \frac{\theta^2}{a^{2(i-1)}}\tau_{22},$$

$$\xi = \frac{4\theta^4(1-i)^2}{a^{4(i-1)}}\tau_{11} + 2\frac{4\theta^4(1-i)}{a^{4(i-1)}}\tau_{12} + \frac{4\theta^4}{a^{4(i-1)}}\tau_{22},$$

$$\tau_{12} = \frac{M}{LK-M^2}.$$

If we take $i=1$, then;

$$\widehat{\mu} \sim N\left(\theta, \theta^2\tau_{22}\right), \tag{3.30}$$

$$\widehat{\sigma}^2 \sim N\left(\theta^2, 4\theta^4\tau_{22}\right). \tag{3.31}$$

The approximate confidence intervals for the parameters of interest can be constructed using these asymptotic distributions. Additionally, the GP can be distinguished from the RP using the following proposition.

Proposition 3.11. *To test $H_0: a=1$ vs $H_1: a\neq 1$, let us define the test statistic*

$$S_2 = \sqrt{\frac{1}{\widehat{a}^2\widehat{\tau}_{11}}}\left(\widehat{a}-1\right). \tag{3.32}$$

If $|S_2| > z_{\alpha/2}$, then, the null hypothesis is rejected at significance level α since $S_2\sim N(0,1)$ asymptotically, under H_0.

Similarly to Case I, the homogeneity of all processes can be tested by means of the following proposition.

Proposition 3.12. *Let* $\{X_{ji}, i = 1, 2, \ldots, n_j + 1\}$ *be the inter-arrival times of the GPs observed until a predetermined time* T*, with the trend parameter* a_j *for* $j = 1, 2, \ldots, r$*. Assume that the random variables* $X_{j1}, j = 1, 2, \ldots, r$ *have an identical exponential distribution with the mean* θ*. To test* $H_0 : a_1 = a_2 = \cdots = a_r$ *against* $H_1 : \exists a_i \neq a_j, i, j \in \{1, 2, \ldots, r\}, i \neq j$*, let us define the test statistic*

$$T_2 = \frac{1}{r-1} \sum_{j=1}^{r} \left(\frac{\hat{a}_j - \bar{\hat{a}}}{\sqrt{\hat{a}^2 \kappa_j}} \right)^2 . \tag{3.33}$$

Then, the null hypothesis is rejected at the significance level α *if* $T_2 > \chi^2_{r-1}(\alpha)$*, since* $T_2 \sim \chi^2_{r-1}$ *asymptotically under* H_0*. Here,*

$$\kappa_j = \frac{k_j}{k_j l_j - m_j^2},$$

$$k_j = n_j + \hat{F}_{n_j+1}(t_j), l_j = n_j^3/3 + n_j^2 \hat{F}_{n_j+1}(t_j), m_j = n_j^3/2 + n_j \hat{F}_{n_j+1}(t_j),$$

$$\hat{F}_{n_j+1}(t_j) = 1 - \exp\left(-\hat{a}^{n_j} t_j / \hat{\theta}\right),$$

$\hat{a}_j$ *is the ML estimator of* a_j *for the* j*th process,* $\bar{\hat{a}} = \frac{1}{r}\sum_{j=1}^{r} \hat{a}_j$*,* $\hat{a}$ *and* $\hat{\theta}$ *are the ML estimators based on multiple process data for the common trend parameter* a *and* θ*, respectively, and* t_j *is the censoring time for the censored variable* $X_{j(n_j+1)}$*.*

Note that the test statistic T_2 can be computed simply as

$$T_2 = \frac{1}{r-1} \sum_{j=1}^{r} \left(\frac{\hat{a}_j - \bar{\hat{a}}}{\sqrt{12\hat{a}^2/n_j^3}} \right)^2 ,$$

when n_j's are large enough. Because it can be shown that κ_j is asymptotically equal to $12/n_j^3$.

The ML estimate of the trend parameter a for a given multiple process data $\mathbf{x}_t$ is computed by solving the Equation (3.22) numerically. Unlike Case I, the MML methodology cannot be utilized in this case directly due to randomly censored observations. However, the MML methodology derived for Case I can be used based on only complete observations $\mathbf{x}_{com}$. Then, the MML estimators for a and θ based on $\ln \mathbf{x}_{com}$ may be considered as good initial values of the ML estimators rather than being an alternative approach. Further, the ML estimators can be alternatively computed by using the well-known Expectation-Maximization (EM) algorithm. The EM algorithm is an iterative procedure that alternates between the expectation (E) step and the maximization (M) step until a convergence criterion is met. In the E step, the expectation of the log-likelihood function conditional on the observed sample, is computed. Then, it is maximized in the M step. The log-likelihood function based on the total sample $\mathbf{x}_t$ can be written as

$$\ln L\left(a, \theta; \mathbf{x}_t\right) = \ln a \sum_{j=1}^{r} \sum_{i=1}^{n_j+1} (i-1) - \ln \theta \sum_{j=1}^{r} (n_j+1) - \frac{1}{\theta} \sum_{j=1}^{r} \left[\sum_{i=1}^{n_j} a^{i-1} x_{ji} + a^{n_j} X_{j(n_j+1)} \right]$$

by assuming censored samples as unobserved variates. Let $\lambda_t = \{\lambda_{ij},\ i = 1, \ldots, n_j+1,\ j = 1, \ldots, r\}$ denote whether the observation x_{ji} is complete or censored such that $\lambda_{ij} = 0$, $i = 1, \ldots, n_j$, $j = 1, \ldots, r$, $\lambda_{j(n_j+1)} = 1$, $j = 1, \ldots, r$. The expectation of the log-likelihood function conditional on the observed sample is obtained as

$$E\left[\ln L\left(a,\theta;\mathbf{X}_t\right)|\mathbf{X}_t=\mathbf{x}_t,\lambda_t\right] = \ln a\sum_{j=1}^{r}\sum_{i=1}^{n_j+1}(i-1)-\ln\theta\sum_{j=1}^{r}(n_j+1) - \frac{1}{\theta}\sum_{j=1}^{r}\left[\sum_{i=1}^{n_j}a^{i-1}x_{ji}+a^{n_j}E\left[X_{j(n_j+1)}\middle|X_{j(n_j+1)}>t_j\right]\right].$$

Therefore, if we consider the Equations (3.3) and (3.4) in $(h+1)$th iteration of the EM algorithm, $\widehat{a}_{(h+1)}$ is computed by solving the equation

$$\sum_{j=1}^{r}\left[\sum_{i=1}^{n_j}a^{i-1}x_{ji}+a^{n_j}E\left[X_{j(n_j+1)}\middle|X_{j(n_j+1)}>t_j;\widehat{a}_{(h)},\widehat{\theta}_{(h)}\right]\right]\left[\frac{\sum_{j=1}^{r}(n_j+1)^2}{\sum_{j=1}^{r}(n_j+1)}-2i+1\right]=0 \tag{3.34}$$

with respect to a, then

$$\widehat{\theta}_{(h+1)}=\left[\sum_{j=1}^{r}\left[\sum_{i=1}^{n_j}\widehat{a}_{(h+1)}^{i-1}x_{ji}+\widehat{a}_{(h+1)}^{n_j}E\left[X_{j(n_j+1)}\middle|X_{j(n_j+1)}>t_j;\widehat{a}_{(h)},\widehat{\theta}_{(h)}\right]\right]\right]\Big/\left[\sum_{j=1}^{r}(n_j+1)\right], \tag{3.35}$$

where $E\left[X_{j(n_j+1)}\middle|X_{j(n_j+1)}>t_j;\widehat{a}_{(h)},\widehat{\theta}_{(h)}\right]=t_j+\widehat{\theta}_{(h)}/\widehat{a}_{(h)}^{n_j}$. These iterations are repeated until $\left||(\widehat{a}_{(h+1)}-\widehat{a}_{(h)},\ \widehat{\theta}_{(h+1)}-\widehat{\theta}_{(h)})\ ||\right|<\epsilon$ is satisfied. Here $||\ .||$ stands for Euclidean norm in $\mathbb{R}^2$ and $\epsilon>0$ is a prespecified tolerance level.

Proposition 3.13. *Let* $\mathbf{X}=\{X_{ji},\ i=1,\ldots,n_j+1,\ j=1,\ldots,r\}$ *be inter-arrival times of multiple GPs with trend parameter a observed until a predetermined time,* $X=\{X_i,i=1,\ldots,n\}$ *be inter-arrival times of a single GP with same trend parameter a, and* $X_1,X_{j1},\ j=1,\ldots,r$ *have exponential distribution with mean* θ*. Then, if* $\sum_{j=1}^{r}(n_j+1)=n$*,*

$$I\left(a;\mathbf{X}\right)<I(a;X),$$

$$I\left(\theta;\mathbf{X}\right)<I(\theta;X),$$

if $\sum_{j=1}^{r}n_j=n$*,*

$$I\left(\theta;\mathbf{X}\right)>I(\theta;X).$$

The Proposition (3.13) implies that observing multiple GPs until a predetermined time, compared to a single GP with the same number of inter-arrival times, results in information loss for both a and θ. However, if the number of inter-arrival times of a single process is equal to number of completely observed inter-arrival times of multiple GPs, this leads to an information surplus for θ in $\mathbf{X}$ due to the additional information obtained from the censored observations.

4. Simulation study

In this section, we conduct a simulation study to evaluate the performance of inferential procedures, including the ML, the ML with the EM algorithm, and the MML methods. The simulation is carried out in MATLAB with 10^4 replications. The trend parameter a is chosen from the range $0.80:0.05:1.20$, and θ is selected from the values $1,2,3,5,10$. However, we present results only for $a=0.95$ and $\theta=1$, as the results for other parameter settings are similar. In the simulation, inter-arrival times for both single and multiple processes are randomly generated, and various sample sizes are used in both cases. The simulation is carried out for four different cases. In the first case, realizations of both

Table 2. Simulation means and variances for the ML and MML estimators, based on both single and multiple GPs data, with predetermined number of inter-arrival times.

	ML				MML			
	$\hat{a}$		$\hat{\theta}$		$\tilde{a}$		$\tilde{\theta}$	
	Mean	Var	Mean	Var	Mean	Var	Mean	Var
Sample 1	0.9474	0.0104	10.551	0.2899	0.9530	0.0047	10,129	0.2967
Sample 2	0.9536	0.0174	10.354	0.2402	0.9059	0.0196	0.7329	0.1615
Sample 3	0.9668	0.0356	10.327	0.2205	0.9755	0.0545	0.8848	0.2046
Sample 4	0.9493	0.0032	10.442	0.2050	0.9508	0.0017	10,056	0.2078
Sample 5	0.9528	0.0072	10.441	0.1979	0.9549	0.0079	0.9637	0.1944
Sample 6	0.9560	0.0136	10.331	0.1758	0.9149	0.0182	0.7594	0.1256
Sample 7	0.9500	0.0004	10.264	0.1348	0.9500	0.0005	0.9962	0.1355
Sample 8	0.9504	0.0017	10.234	0.1274	0.9505	0.0020	0.9631	0.1250
Sample 9	0.9521	0.0041	10.253	0.1237	0.9528	0.0051	0.9371	0.1202
Sample 10	0.9537	0.0069	10.220	0.1144	0.9034	0.0091	0.7066	0.0756
Sample 11	0.9501	0.0001	10.184	0.0820	0.9501	0.0001	0.9981	0.0819
Sample 12	0.9502	0.0004	10.178	0.0787	0.9503	0.0004	0.9784	0.0787
Sample 13	0.9508	0.0008	10.195	0.0766	0.9438	0.0010	0.8880	0.0675
Sample 14	0.9508	0.0015	10.154	0.0746	0.9321	0.0018	0.8031	0.0585
Sample 15	0.9515	0.0023	10.151	0.0711	0.9518	0.0029	0.9220	0.0693

single and multiple GPs are observed with a predetermined number of inter-arrival times, similar to Case I. The ML and MML estimators, as described in Section 3.1, are then evaluated based on the inter-arrival times of the GPs. It is important to note that the inter-arrival times are complete in Case I. Table 2 presents the results for this case. In Table 2, r denotes the number of processes being observed, n presents the number of inter-arrival times for each process, and Mean and Var refer to the simulation mean and variance of the corresponding estimators. The sample sizes used in the Table 2 are given in Table 3. For the computation of the ML estimator $\widehat{a}$ the function *fzero* is used to numerically solve Equation (3.3), and the MML estimator $\widetilde{a}$ is used as the starting point for the computation.

In the second case, realizations of multiple GPs are observed until a predetermined time, similar to that of Case II. The ML estimators, as described in Section 3.2, are then evaluated based on the inter-arrival times of the multiple GPs. It is important to note that the last inter-arrival times of each process are right-censored, while the previous ones are complete for Case II. The ML estimators are computed directly from Equation (3.22) and (3.23), as well as using the EM algorithm formulated in Equations (3.34) and (3.35). When employing the EM algorithm, the tolerance level ϵ is set to $5/10^{-5}$, and MML estimators based on only complete inter-arrival times are used as initial estimates for the algorithm. The results for this case are provided in Table 4, which also includes the MML estimators based solely on the complete inter-arrival times. In Table 4, r denotes the number of processes being observed, T represents the predetermined time that ends the observation period, n^* gives the average number of inter-arrival times observed for multiple GPs until

Table 3. The sample sizes given in Table 2

Sample 1	$r = 1$, $n = 15$
Sample 2	$r = 2$, $n = (8, 7)$
Sample 3	$r = 3$, $n = (5, 5, 5)$
Sample 4	$r = 1$, $n = 20$
Sample 5	$r = 2$, $n = (10, 10)$
Sample 6	$r = 3$, $n = (7, 7, 6)$
Sample 7	$r = 1$, $n = 30$
Sample 8	$r = 2$, $n = (15, 15)$
Sample 9	$r = 3$, $n = (10, 10, 10)$
Sample 10	$r = 4$, $n = (8, 8, 7, 7)$
Sample 11	$r = 1$, $n = 50$
Sample 12	$r = 2$, $n = (25, 25)$
Sample 13	$r = 3$, $n = (17, 17, 16)$
Sample 14	$r = 4$, $n = (13, 13, 12, 12)$
Sample 15	$r = 5$, $n = (10, 10, 10, 10, 10)$

Table 4. Simulation means for the ML, ML-EM and MML estimators, based on multiple GPs data, with predetermined total observation time.

		MML		ML		ML-EM		
	n^*	$\tilde{a}$	$\tilde{\theta}$	$\hat{a}$	$\hat{\theta}$	$\hat{a}$	$\hat{\theta}$	AIN
$r = 2, T = 10$	16	0.8877	0.5901	0.9115	0.9476	0.9117	0.9479	9
$r = 3, T = 10$	24	0.8501	0.4467	0.9246	0.9597	0.9246	0.9597	9
$r = 4, T = 10$	32	0.8287	0.3751	0.9311	0.9667	0.9311	0.9667	9
$r = 2, T = 20$	26	0.9212	0.6887	0.9370	0.9665	0.9370	0.9665	7
$r = 3, T = 20$	39	0.9041	0.5651	0.9412	0.9746	0.9412	0.9746	7
$r = 4, T = 20$	52	0.8927	0.4955	0.9430	0.9781	0.9430	0.9781	7
$r = 2, T = 30$	36	0.9312	0.7367	0.9423	0.9757	0.9423	0.9757	6
$r = 3, T = 30$	54	0.9207	0.6265	0.9448	0.9791	0.9448	0.9791	6
$r = 4, T = 30$	72	0.9135	0.5632	0.9461	0.9842	0.9461	0.9842	6
$r = 2, T = 50$	50	0.9395	0.7872	0.9462	0.9821	0.9462	0.9821	5
$r = 3, T = 50$	75	0.9325	0.6906	0.9472	0.9865	0.9472	0.9865	5
$r = 4, T = 50$	100	0.9288	0.6396	0.9481	0.9916	0.9481	0.9916	5

time T, AIN indicates the average number of iterations required for the EM algorithm to converge, and the values represent the simulation means of the related estimators.

In the third case, inter-arrival times of GPs from both single and multiple processes are observed to compare the effectiveness of using single versus multiple process data. In this setup, a single GP is observed with a predetermined number of inter-arrival times. Then, multiple GPs are observed until a predetermined time, which is on average equal to the total realization length of the single GP. Table 5 presents the results for the ML

Table 5. Simulation means and variances for the ML estimators, based on single and multiple GPs data, under averagely equal total observation length.

	ML			
	$\hat{a}$		$\hat{\theta}$	
	Mean	Var	Mean	Var
$r = 1, n = 8; T^* = 10$	0.9647	0.0291	11,098	0.6318
$r = 2, T = 10; n^* = 16$	0.9115	0.0096	0.9476	0.1842
$r = 3, T = 10; n^* = 24$	0.9246	0.0057	0.9597	0.1184
$r = 4, T = 10; n^* = 32$	0.9311	0.0040	0.9667	0.0876
$r = 1, n = 13; T^* = 20$	0.9525	0.0056	10,659	0.3414
$r = 2, T = 20; n^* = 26$	0.9370	0.0020	0.9665	0.1199
$r = 3, T = 20; n^* = 39$	0.9412	0.0012	0.9746	0.0760
$r = 4, T = 20; n^* = 52$	0.9430	0.0009	0.9781	0.0576
$r = 1, n = 18; T^* = 30$	0.9512	0.0021	10,524	0.2450
$r = 2, T = 30; n^* = 36$	0.9423	0.0009	0.9757	0.0926
$r = 3, T = 30; n^* = 54$	0.9448	0.0006	0.9791	0.0628
$r = 4, T = 30; n^* = 72$	0.9461	0.0004	0.9842	0.0475
$r = 1, n = 25; T^* = 50$	0.9500	0.0007	10,361	0.1691
$r = 2, T = 50; n^* = 50$	0.9462	0.0003	0.9821	0.0712
$r = 3, T = 50; n^* = 75$	0.9472	0.0002	0.9865	0.0476
$r = 4, T = 50; n^* = 100$	0.9481	0.0002	0.9916	0.0360

estimators based on single- and multiple-GP data. In Table 5, r represents the number of processes being observed, n indicates the predetermined number of inter-arrival times for the single GP, T^* is the predetermined time to observe multiple GPs, and n^* represents the average number of inter-arrival times of multiple GPs observed until time T.

In the fourth case, inter-arrival times of GPs are observed according to both Case I and Case II to evaluate the interval estimates of the ML estimators. For each case, 95% confidence intervals for the parameters are computed based on the asymptotic distributions of the ML estimators, as presented in Corollary (3.2) and (3.9). The confidence intervals are calculated as "point estimate $\pm$ 1.96 $\times$ square root of the estimated asymptotic variance". In addition, the corresponding coverage probabilities of the confidence intervals are computed to assess the convergence rate of the ML estimators. The confidence intervals and coverage probabilities for Case I are provided in Table 6, while the results for Case II are presented in Table 7.

When the results are investigated, the following comments can be made. For the first case, overall, the ML estimators $\widehat{a}$ and $\widehat{\theta}$ based on both single and multiple process data with equal inter-arrival times perform very well in terms of low bias and variance. As the number of inter-arrival times n increases, both the bias and variance of $\widehat{a}$ and $\widehat{\theta}$ decrease. The MML estimators $\widetilde{a}$ and $\widetilde{\theta}$ perform very close to the ML estimators in terms of simulation mean and variance for single process data. However, MML estimators based on multiple process data become more biased compared to those based on single process data as the number of multiple processes r increases while the total number of inter-arrival times n is fixed. Because this setting yields fewer inter-arrival times for each process and

Table 6. Confidence interval and coverage probabilities for the ML estimators, based on both single and multiple GPs data, with predetermined number of inter-arrival times.

	$a = 0.95,\ \theta = 10$					
	$\hat{a}$			$\hat{\theta}$		
	Confidence Interval		Coverage Probability	Confidence Interval		Coverage Probability
$r = 1, n = 15$	0.8297	1.0507	0.918	0	21.0988	0.875
$r = 2, n = (15, 15)$	0.8708	1.0277	0.956	2.8795	17.3759	0.932
$r = 3, n = (15, 15, 15)$	0.8874	1.0158	0.939	4.2321	16.1317	0.930
$r = 1, n = 20$	0.8781	1.0225	0.944	1.3167	20.0100	0.896
$r = 2, n = (20, 20)$	0.9002	1.0024	0.946	3.9201	16.7007	0.929
$r = 3, n = (20, 20, 20)$	0.9086	0.9918	0.947	5.0432	15.3771	0.939
$r = 1, n = 30$	0.9106	0.9891	0.929	2.8985	17.4901	0.901
$r = 2, n = (30, 30)$	0.9221	0.9776	0.954	4.9836	15.1953	0.932
$r = 3, n = (30, 30, 30)$	0.9276	0.9729	0.958	5.9558	14.3433	0.942
$r = 1, n = 50$	0.9317	0.9682	0.959	4.5096	15.7292	0.933
$r = 2, n = (50, 50)$	0.9372	0.9630	0.950	6.1424	14.0627	0.951
$r = 3, n = (50, 50, 50)$	0.9394	0.9604	0.949	6.7946	13.1914	0.947

Table 7. Confidence interval and coverage probabilities for the ML estimators, based on both single and multiple GPs data, with predetermined total observation time.

	$a = 0.95,\ \theta = 10$					
	$\hat{a}$			$\hat{\theta}$		
	Confidence Interval		Coverage Probability	Confidence Interval		Coverage Probability
$r = 1, T = 200, n^* = 15$	0.7958	1.0471	0.917	0	19.4345	0.837
$r = 2, T = 200, n^* = 30$	0.8569	1.0203	0.947	2.6513	16.7078	0.897
$r = 3, T = 200, n^* = 45$	0.8796	1.0109	0.935	4.1645	15.9377	0.934
$r = 1, T = 350, n^* = 20$	0.8662	1.0071	0.929	1.1671	17.7382	0.858
$r = 2, T = 350, n^* = 40$	0.8961	0.9917	0.933	3.8394	15.6600	0.894
$r = 3, T = 350, n^* = 60$	0.9072	0.9847	0.946	5.0130	14.7367	0.921
$r = 1, T = 700, n^* = 30$	0.9041	0.9826	0.940	2.7139	16.4693	0.890
$r = 2, T = 700, n^* = 60$	0.9194	0.9736	0.936	4.8489	14.5983	0.914
$r = 3, T = 700, n^* = 90$	0.9260	0.9700	0.941	5.8420	13.9020	0.933
$r = 1, T = 2000, n^* = 50$	0.9285	0.9674	0.933	4.2383	15.3055	0.907
$r = 2, T = 2000, n^* = 100$	0.9354	0.9629	0.940	5.9874	13.9592	0.939
$r = 3, T = 2000, n^* = 150$	0.9383	0.9606	0.942	6.7153	13.2039	0.941

so, the approximation for the MML estimators gets worse compared to those based on more inter-arrival times. Further, the performance of MML estimators improves when more inter-arrival times are observed for each process.

In the second case, the ML estimators give desirable results for the parameters a and θ. Both the bias and the variance of the ML estimators decrease as the observation period T and/or the number of processes observed r increase. However, MML estimators based on only complete samples are very biased, especially for θ, even with large values of T. The reason for this phenomenon is again explained by the small number of inter-arrival times, similar to the first case. When multiple GPs are observed until the predetermined time T, some of the processes produce only a few inter-arrival times in some replications due to randomness.This negatively affects the performance of the MML estimators, as their approximation worsens in the case of very small samples. Besides that, the EM algorithm converges in only a few iterations when the MML estimates are chosen as the initial values for the algorithm.

In the third case, the ML estimators based on both single and multiple process data perform close to each other. The bias of the ML estimator $\widehat{a}$ based on single process data is smaller than that based on multiple process data. However, the bias of the ML estimator $\widehat{\theta}$ based on multiple process data is less than that of the ML estimator based on single process data. Furthermore, the variance of the estimators based on multiple process data decreases as the number of processes being observed r and the total observation period T increase, since the total number of inter-arrival times being observed increases.

For the fourth case, when the confidence intervals and coverage probabilities derived from the asymptotic distributions of the ML estimators are examined, it is seen that the coverage probabilities converge to the actual level of 95% for moderate sample sizes. If the number of processes and inter-arrival times increase, the confidence intervals get tighter, as expected.

5. Data analysis

In this section, we present data analysis examples to illustrate the effectiveness and usefulness of GP multiprocess data in the context of multirepairable system modeling. First, artificially generated single- and multiple-process data of GPs are analyzed from effectiveness perspective. Then, a real data representing the failure processes two supercomputer shared memory processors (SMPs) are analyzed statistically under both the GP and the NHPP-PLP models for comparison.

5.1. Data analysis for artificially generated data

In this section, a numerical example is examined for inferential procedures based on single and multiple GPs data to compare the effectiveness of single and multiple processes. For this purpose, a single GP and multiple GPs are generated with trend parameter $a = 1.1$ and $\theta = 10$. For the single process data, the number of inter-arrival times observed is chosen as $n = 20$. For the data from multiple processes, the number of processes observed is chosen as $r = 5$, and the total observation period is limited to $T = 40$. The inter-arrival times for these processes are given in Table 8. Note that the last inter-arrival times of each process are right-censored, which are denoted with a sign of plus.

The ML estimators are computed as $\widehat{a} = 1.0857$ and $\widehat{\theta} = 9.1244$ based on single process data, and as $\widehat{a} = 1.1240$ and $\widehat{\theta} = 10.5060$ based on multiple process data. To test whether the data are compatible with the GP or RP, we test $H_0 : a = 1$ against $H_1 : a \neq 1$ for single process data with the test statistic S_1 given in Equation (3.14) by taking $r = 1$. We also test, $H_0 : a = 1$ against $H_1 : a \neq 1$ for multiple process data by utilizing the test statistic S_2 given in Equation (3.34). The results are $S_1 = 2.0379$ with p-value $p = 0.0208$ for single process data and $S_2 = 2.1468$ with p-value $p = 0.0159$ for multiple process

Table 8. Artificially generated data for single and multiple GPs with $a = 1.1$ and $\theta = 10$.

Inter-arrival times for single process with $n = 20$.													
1.84	4.38	5.40	1.10	15.89	0.60	12.56	3.39	9.07	2.46	20.77	0.93	0.52	0.25
0.03	1.63	2.84	4.54	1.22	0.87								
Inter-arrival times for multiple processes with $r = 5$ until predetermined time $T = 40$.													
2.71	22.63	3.41	4.96	6.29+									
0.63	4.79	6.77	0.45	2.88	4.93	0.99	3.23	2.76	1.64	3.86	5.01	1.74	0.32+
8.99	19.85	6.71	4.45+										
9.13	1.66	7.50	7.07	6.97	7.67+								
13.47	22.22	4.31+											

data. Both results yield $a \neq 1$ at $\alpha = 0.05$ significance level, that is the processes are distinguished from the RP. It can be concluded from this result that the model parameters of the GP, a and θ, can be estimated effectively based on both single and multiple process data. However, the total observation time for single process data is 90.30 while it is only 40 (predetermined) for multiple process data. Consequently, the model parameters of a GP can be estimated effectively based on multiple process data which can be observed in a shorter period compared to single process data.

5.2. Real data analysis

In this section, a real data set is analyzed using both the GP and the NHPP for comparative purposes. The data set covers the failure processes of two identical SMPs of the Blue Mountain supercomputer at Los Alamos National Laboratory. When an SMP fails, it is restarted. Therefore, the failure process of each system can be regarded as the repair process of a repairable system; see Wang et al. [42]. The original data present consecutive failure times of two identical SMPs, and can be found in Wang et al. [42]. The inter-arrival times of consecutive failures for each system are given in Table 9. Wang et al. [42] showed that these multiple failure process data exhibit reliability growth, i.e., the inter-arrival times between consecutive failures tend to increase after each restart/repair. Therefore, it can be modeled with a monotone counting process model. Wang et al. [42] modeled the data using the power-law process (PLP) which is a monotonic NHPP with the intensity function $\lambda(t) = \lambda\beta t^{\beta-1}$. They also showed that the failure processes of two SMPs are homogeneous. Therefore, each failure process can be analyzed using an identical intensity function of the PLP. The ML estimators of the intensity function parameters based on the failure processes of two SMPs are calculated as $\hat{\lambda} = 0.2496$ and $\hat{\beta} = 0.7794$. For ML estimation based on multiple process data in the PLP, see the appendix of Garmabaki et al.[12].

As an alternative, these data can also be modeled by the GP. To fit the data set with a GP model, it is essential to properly answer the following questions: Does the data exhibit a trend? Is the data consistent with the model? Are the trends of the multiple processes homogeneous? What are the estimates of the model parameters? Finally, how well does the model fit the data compared to other models? The examination of these questions is provided as follows. The combined Laplace test has been applied to multiple process data to determine whether the processes exhibit a trend. The test statistic has been calculated as $L_c = -15.3731$ with a p-value of $p < 0.0001$. This indicates that the multiple process data of SMPs' failures exhibit a trend. Additionally, the value of the test statistic suggests that the failures tend to decrease over time, corresponding to an

Table 9. Inter-arrival times between the consecutive failures of the two SMPs in the Blue Mountain supercomputer

j	n_j	X_{ji}									
1	31	4.74	15.49	1.98	3.81	9.07	2.25	20.70	1.67	7.96	5.22
		1.27	11.52	8.90	17.79	11.75	0.71	4.30	6.84	5.95	4.59
		16.55	38.41	13.62	24.77	4.88	0.22	15.66	7.72	24.55	7.86
		79.43									
2	23	5.21	9.28	27.82	12.21	4.57	7.00	8.90	44.28	8.20	19.01
		26.29	17.44	13.31	19.85	29.27	12.88	18.54	28.62	9.99	11.03
		14.78	62.91	23.37							

increasing trend in the inter-arrival times of consecutive failures. For more details on the combined Laplace test for multiple process data, see Kvaløy and Lindqvist [16] and Garmabaki et al. [13]. Lam [20] proposed the following auxiliary variables to test whether the data is compatible with the GP. Let $U_{ji} = X_{j(2i)}/X_{j(2i-1)}, i = 1, 2, \ldots, (n_j - 1)/2$ for $j = 1, 2$. These variables are supposed to be iid for each $j \in \{1, 2\}$, if the multiple process data follow a GP model. The iid property of the auxiliary variables has been tested by the well-known turning-point test. The test statistics have been computed as $TP_1 = -0.4354$ with p-value $p = 0.66$ for $j = 1$, and $TP_2 = -1.0885$ with p-value $p = 0.27$ for $j = 2$. Consequently, each process is compatible with the GP. It is reasonable to assume that the first inter-arrival times for each process are identically distributed, as the failure processes have been observed over identical systems. However, the homogeneity of trend parameters for each process should be tested. For this purpose, the test statistic defined in Equation (3.15) has been calculated as $T_1 = 0.4281$ with p-value $p = 0.51$. This indicates that the trend parameters of the failure processes are homogeneous. Thus, the multiple process data of SMPs' failure processes can be modeled by the GP with a common trend parameter a. Taking the assumption that X_{11} and X_{21} have an exponential distribution with mean θ, the ML estimates of the parameters are calculated as $\hat{a} = 0.9654, \hat{\theta} = 9.0295$. Note that $\hat{a} = 0.9654 < 1$, which implies that the inter-arrival times of consecutive failures tend to decrease, giving a consistent result with the PLP model. Now, we can compare the GP and the PLP models for the multiple process data of SMPs' failure processes. The ML estimates and related Akaike information criterion (AIC) values are given in Table 10. As seen in the table, the AIC value for the GP model is slightly smaller. This result suggests a better fit of the GP model to the data compared to the PLP.

Table 10. GP and NHPP-PLP models for the multiple SMPs failure data

Model	Maximum Likelihood Estimates	AIC
GP	$\hat{a} = 0.9654, \hat{\theta} = 9.0295$	400.2603
NHPP-PLP	$\hat{\lambda} = 0.2496, \hat{\beta} = 0.7794$	401.4636

6. Conclusion and discussion

In this paper, the multiple process data for a GP have been statistically evaluated. The motivation for utilizing the GP model in the analysis of multiple process data is explained in the context of modeling the repair processes of multiple repairable systems. In this regard, the study presents some novel results, as the literature commonly uses NHPP models to analyze multiple process data with a monotonic trend. The statistical inference problem is studied in detail under the assumption that all processes are homogeneous and that inter-arrival times follow an exponential distribution. For this purpose, the ML and MML estimators have been obtained along with their asymptotic distributions. The simulation results show that both methods effectively estimate the model parameters, showing low bias and variance. When all samples have a sufficient number of observations, the ML and MML estimators perform similarly and produce almost the same estimates. An advantage of the MML estimators is that they can be calculated analytically, whereas the ML estimators require numerical computation. When the samples include censored observations, the MML estimators cannot be directly calculated because of the presence of randomly censored observations. In such cases, we recommend using the EM algorithm to compute the ML estimators, with the MML estimates from complete observations serving as the starting points for the algorithm.

From Propositions 3.7 and 3.13, the multiple process data of the GPs contain less information about the trend parameter a than the single process data with an equal number of observations. However, the simulation study shows that statistical inference for the trend parameter a based on multiple process data is as effective as that based on single process data. Artificial data analysis demonstrates the effectiveness of multiple process data compared to single process data from a statistical perspective. Furthermore, real data analysis highlights the modeling capability of GPs for analyzing multiple process data, providing a better fit to the SMP failure data than the PLP model.

In the study, the inter-arrival times for GPs are assumed to follow an exponential distribution. However, statistical inference for the GP under multiple process data with other commonly used non-negative distributions, such as the gamma, lognormal, or Weibull distributions, should be considered to extend the present study. Furthermore, the inference problem for alternative monotonic counting processes, such as the α-series process, double GP, and others, should be addressed in multiple process data modeling.

Acknowledgements

The author is very grateful to the editor and three anonymous referees for their valuable suggestions. The author also thanks Prof. Dr. Birdal Şenoğlu for his suggestion about the MML methodology used in the article.

Author contributions. All the work has been done solely by the author.

Conflict of interest statement. The author declare that there isn't any competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding. The study is not supported by any funding.

Data availability. No data was used for the research described in the article.

References

[1] R. Arnold, S. Chukova, Y. Hayakawa and S. Marshall, *Warranty cost analysis with an alternating geometric process*, Proc. of the Inst. of Mech. Eng., Part O: J. of Risk and Reliab. **233** (4), 698-715, 2019.

[2] H. Ascher and H. Feingold, *Repairable systems reliability: modeling, inference, misconceptions and their causes*, Wiley, 1984.

[3] H. Aydoğdu, B. Şenoğlu and M. Kara, *Parameter estimation in geometric process with Weibull distribution*, Appl. Math. and Comp. **217** (6), 2657-2665, 2010.

[4] G.K. Bhattacharyya, *The asymptotics of maximum likelihood and related estimators based on type II censored data*, J. of the American Stat. Assoc. **80** (390), 398-404, 1985.
[5] C. Biçer, H.S. Bakouch and H.D. Biçer, *Inference on parameters of a geometric process with scaled Muth distribution*, Fluct. and Noise Let. **20** (01), 2150006, 2021.
[6] H.D. Biçer, C. Biçer and H.S. Bakouch, *A geometric process with Hjorth marginal: Estimation, discrimination, and reliability data modeling*, Qual. and Reliab. Eng. Int. **38** (5), 2795-2819, 2022.
[7] J.S.K. Chan, Y. Lam and D.Y. Leung, *Statistical inference for geometric processes with gamma distributions*, Comp. Stat. & Data Analy. **47** (3), 565-581, 2004.
[8] J.S.K. Chan, C.P. Lam, P.L.H. Yu, S.T.B. Choy, and C.W.S. Chen, *A Bayesian conditional autoregressive geometric process model for range data*, Comp. Stat. & Data Analy. **56** (11), 3006-3019, 2012.
[9] J.S.K. Chan, W.Y. Wan and P.L.H. Yu, *Poisson geometric process approach for predicting drop-out and committed first-time blood donors*, J. of Appl. Stat. **41** (7), 1486-1503, 2014.
[10] J.S.K. Chan, P.L. Yu, Y. Lam and A.P. Ho, *Modelling SARS data using threshold geometric process*, Stat. in Med. **25** (11), 1826-1839, 2006.
[11] D.R. Cox and P.A. Lewis, *The statistical analysis of series of events*, Springer, 1966.
[12] A.H.S Garmabaki, A. Ahmadi, Y.A. Mahmood and A. Barabadi, *Reliability modelling of multiple repairable units*, Qual. and Reliab. Eng. Int. **32** (7), 2329-2343, 2016
[13] A.H.S. Garmabaki, A. Ahmadi, J. Block, H. Pham and U. Kumar, *A reliability decision framework for multiple repairable units*, Reliab. Eng. & Syst. Saf. **150**, 78-88, 2016.
[14] M. Kara, *Parameter estimation in geometric processes*, M.Sc. Thesis, Ankara University, 2009.
[15] M. Kara, G. Güven, B. Şenoğlu and H. Aydoğdu, *Estimation of the parameters of the gamma geometric process*, J. of Stat. Comp. and Sim. **92** (12), 2525-2535, 2022.
[16] J.T. Kvaløy and B.H. Lindqvist, *TTT-based tests for trend in repairable systems data*, Reliab. Eng. & Syst. Saf. **60** (1), 13-28, 1998.
[17] Y. Lam, *Geometric processes and replacement problem*, Acta Math. Appl. Sinica **4** (4), 366-377, 1988.
[18] Y. Lam, *A note on the optimal replacement problem*, Adv. in Appl. Prob. **20** (2), 479-482, 1988.
[19] Y. Lam, *An optimal repairable replacement model for deteriorating systems*, J. of Appl. Prob. **28** (4), 843-851, 1991.
[20] Y. Lam, *Nonparametric inference for geometric processes*, Comm. in Stat. Theo. and Meth. **21** (7), 2083-2105, 1992.
[21] Y. Lam, *The geometric process and its applications*, World Scientific, 2007.
[22] Y. Lam, *A geometric process maintenance model with preventive repair*, Euro. J. of Op. Res. **182** (2), 806-819, 2007.
[23] Y. Lam and J.S.K. Chan, *Statistical inference for geometric processes with lognormal distribution*, Comp. Stat. & Data Analy. **27** (1), 99-112, 1998.
[24] Y. Lam and Y.L. Zhang, *Analysis of a two-component series system with a geometric process model*, Nav. Res. Log. **43** (4), 491-502, 1996.
[25] Y. Lam and Y.L. Zhang, *A geometric-process maintenance model for a deteriorating system under a random environment*, IEEE Trans. on Reliab. **52** (1), 83-89, 2003.
[26] S.A. Lone, I. Alam and A. Rahman, *Statistical analysis under geometric process in accelerated life testing plans for generalized exponential distribution*, Ann. of Data Sci. **10** (6), 1653-1665, 2023.

[27] I.Y. Na and W. Chang, *Multi-system reliability trend analysis model using incomplete data with application to tank maintenance*, Qual. and Reliab. Eng. Int. **33** (8),2385-2395, 2017.

[28] S. Park, N. Balakrishnan and G. Zheng, *Fisher information in hybrid censored data*, Stat. & Prob. Let. **78** (16), 2781-2786, 2008.

[29] M.H. Pekalp and H. Aydoğdu, *An integral equation for the second moment function of a geometric process and its numerical solution*, Nav. Res. Log. **65** (2), 176-184, 2018.

[30] M.H. Pekalp, H. Aydoğdu and K.F. Türkman, *Discriminating between some lifetime distributions in geometric counting processes*, Comm. in Stat. Sim. and Comp. **51** (3), 715-737, 2020.

[31] M.H. Pekalp and H. Aydoğdu, *Power series expansions for the probability distribution, mean value and variance functions of a geometric process with gamma interarrival times*, J. of Comp. and Appl. Math. **388**, 113287, 2021.

[32] M.H. Pekalp and H. Aydoğdu, *Parametric estimations of the mean value and variance functions in geometric process.* J. of Comp. and Appl. Math. **449**, 115969, 2024.

[33] H. Rasay, F. Azizi and F. Naderkhani, *A mathematical maintenance model for a production system subject to deterioration according to a stochastic geometric process.* . Ann. of Op. Res. 1-28, 2024

[34] Y.Y. Tang and Y. Lam, *A δ-shock maintenance model for a deteriorating system*, Euro. J. of Op. Res. **168** (2), 541-556, 2006.

[35] M.L. Tiku, *Estimating the mean and standard deviation from a censored normal sample*, Biometrika **54** (1/2), 155-165, 1967.

[36] M.L. Tiku, *Estimating the parameters of log-normal distribution from censored samples*, J. of the Ame. Stat. Assoc. **63** (321), 134-140, 1968.

[37] M.L. Tiku, W.Y. Tan and N. Balakrishnan, *Robust Inference*, Marcel Dekker, 1986.

[38] M.L. Tiku and A.D. Akkaya, *Robust estimation and hypothesis testing*, New Age International, 2004.

[39] I. Usta, *Bayesian estimation for geometric process with the Weibull distribution*, Comm. in Stat. Sim. and Comp. **53** (5), 2527-2553, 2024.

[40] D.C. Vaughan and M.L. Tiku, *Estimation and hypothesis testing for a nonnormal bivariate distribution with applications*, Math. and Comp. Mod. **32** (1-2), 53-67, 2000.

[41] W.Y. Wan and J.S.K. Chan, *A new approach for handling longitudinal count data with zero-inflation and overdispersion: Poisson geometric process model*, Biometrical Journal: J. of Math. Meth. in Biosci. **51** (4), 556-570, 2009.

[42] Y. Wang, Z. Lu, and S. Xiao, *Parametric bootstrap confidence interval method for the power law process with applications to multiple repairable systems*, IEEE Access **6**, 49157-49169, 2018.

[43] J.S. White, *The moments of log-Weibull order statistics*, Technometrics **11** (2), 373-386, 1969.

[44] Y.L. Zhang and G.J. Wang, *A geometric process warranty model using a combination policy*, Comm. in Stat. Theo. and Meth. **48** (6), 1493-1505, 2019.

[45] A. Yılmaz, *Bayesian Parameter Estimation for Geometric Process with Rayleigh Distribution*, Bitlis Eren Üni. Fen Bil. Der. **13** (2), 482-491, 2024.

[46] G. Zhen and J.L. Gastwirth, *On the Fisher information in randomly censored data*, Stat. & Prob. Let. **52** (4), 421-426, 2001.