Pesquisa Operacional (2022) 42(nspe1): e262123 p.1-17
doi: 10.1590/0101-7438.2022.042nspe1.00262123

Printed version ISSN 0101-7438 / Online version ISSN 1678-5142
www.scielo.br/pope
ARTICLES

A COMPARISON OF THE PERFORMANCE OF THE GEOMETRIC PROCESS AND ITS EXTENSIONS

Jiaqi Yin[1*] and Shaomin Wu[2]

Received March 16, 2022 / Accepted July 29, 2022

ABSTRACT. The geometric process (GP) is a stochastic process that was an extension of the renewal process. It was introduced by Lam (1988) in 1988 with an intention to model the failure process of a repairable system whose the times between failures become shorter and shorter after repairs and repair times become longer and longer. The GP has been widely studied in the literature of reliability and maintenance and applied in optimisation of maintenance policies. Some authors have proposed various versions of its extensions (or the GP-like models), including the α-series process, the threshold GP, the extended Poisson process, the doubly GP, and the doubly-ratio GP. Some papers also compare the performance of the GP with that of other models, but not with the performance of the extensions of the GP. This paper therefore reviews the GP-like models, compares the performance of the GP and its extensions in terms of the Akaike information criterion (AIC), the corrected AIC (AICc) and the maximum likelihood (ML) based on 25 real-world datasets. Besides, the least square methods for estimating the parameters in some models are discussed, which is used for model performance of GP and GP-like models. The finding is useful for practitioners in their selection of the GP-like models.

Keywords: Geometric process; failure process; recurrent event data analysis

1 INTRODUCTION

Modelling the failure process of a system is an important research topic in the reliability and maintenance research community and is normally done by using stochastic processes. The geometric process (GP), which is introduced by Lam (1988), is one of such stochastic processes. The GP has been widely used for modelling system failure processes. It is able to describe a stochastically increasing or decreasing trend, which is a characteristic in many practical applications in different fields Zhang (1999); Lam et al. (2004); Wang & Yam (2017); Pekalp & Aydoğdu (2021). In fact, the GP has received a lot of attention and has been extended into different variants, such as the general repair geometric process Finkelstein (1993), the α-series process

*Corresponding author
[1]Kent Business School, University of Kent, Kent, CT2 7FS, UK – E-mail: jy236@kent.ac.uk – http://orcid.org/0000-0003-0853-7198
[2]Kent Business School, University of Kent, Kent, CT2 7FS, UK – E-mail: S.M.Wu@kent.ac.uk – http://orcid.org/0000-0001-9786-3213

Braun et al. (2005), the threshold geometric process Chan et al. (2006), the extended Possion process Wu & Clements-Croome (2006), the doubly geometric process Wu (2018) and the doubly-ratio geometric process Wu (2022). However, little research specifically compares the performance of these models based on real world datasets. This paper aims to compare the performance of the GP and its extensions.

The remainder of this paper is structured as follows. Section 2 describes the background of the GP and its extensions. Section 3 describes the least square estimation for models. Section 4 describes the test design and methods for estimating the performance of the models. Section 5 describes datasets for comparison in this paper and shows the result of the test for independence. Section 6 describes the comparison result among datasets. Section 7 concludes the findings and proposes future research.

2 GP AND ITS EXTENSIONS

This section introduces the GP and its extensions in detail. Before that, relevant concepts are introduced.

Assume that X and Y are two random variables. If for every real number u, the inequality

$$P(X \geqslant u) \geqslant P(Y \geqslant u)$$

holds, then X is stochastically greater than or equal to Y, or Y is stochastically less than or equal to X. Then, the monotonicity of a stochastic process can be defined.

Definition 1. *Ross et al. (1996) Given a stochastic process $\{X_k = 1, 2...\}$, if $X_k \leq_{\text{st}} X_{k+1}$ or $X_k \geq_{\text{st}} X_{k+1}$ for $k = 1, 2...$ then $\{X_k, k = 1, 2, ...\}$ is stochastically increasing or decreasing.*

Lam (1988) proposes the definition of the geometric process as shown below.

Definition 2. *Lam (1988) Given a sequence of non-negative random variables $\{X_k, k = 1, 2, \ldots\}$, if they are independent and the cdf of X_k is given by $F(a^{k-1}x)$ for $k = 1, 2, \ldots$, where a is a positive constant, then $\{X_k, k = 1, 2, \cdots\}$ is called a geometric process (GP).*

We refer to the random variable X_k as the kth inter-arrival time in what follows.

Remark 1

From Definitions 1 and 2, the following results have been obtained.

- If $a > 1$, then $\{X_k, k = 1, 2, \cdots\}$ is stochastically decreasing.
- If $a < 1$, then $\{X_k, k = 1, 2, \cdots\}$ is stochastically increasing.
- If $a = 1$, then $\{X_k, k = 1, 2, \cdots\}$ is a renewal process (RP).

As the GP has been extended into several variants Braun et al. (2005); Wu & Clements-Croome (2006); Chan et al. (2006); Zhang & Wang (2016); Wu (2018, 2022), they can be regarded as as a special case of the following definition Wu et al. (2020).

Definition 3. *Given a sequence of non-negative variables $\{X_k, k = 1, 2, ...\}$, if variables are independent the distribution function of X_k is given by $F(g(\delta_1, k)x^{h(\delta_2, k)})$ for $k = 1, 2, ...$, where $g(\delta_1, k)$ is a positive function of δ_1 and $h(\delta_2, k)$ is a positive function of δ_2 and k, and δ_1 and δ_2 are estimable parameter vectors, then $\{X_k, k = 1, 2, \cdots\}$ is a generalized GP.*

Based on Remark 1 and Definition 3, the following results can be obtained:

- If $g(\delta_1, k)$ increase over k and $h(\delta_2, k) = 1$, then $\{X_k, k = 1, 2, \cdots n\}$ is stochastically decreasing.
- If $g(\delta_1, k)$ decrease over k and $h(\delta_2, k) = 1$, then $\{X_k, k = 1, 2, \cdots n\}$ is stochastically increasing.
- If $a = 1$, then $\{X_k, k = 1, 2, \cdots\}$ is a (RP).

According to Wu (2018), the GP has the following limitations

- Based on the definition of the GP, $\{X_k, k = 1, 2, ...\}$ always change monotonously over k. That means the GP cannot be used to describe the situation with non-monotonous failure ratio/patterns.
- Suppose that X_1 follows the Weibull distribution. The generalised GP has constant $g(\delta_1, k)$ and $h(\delta_2, k)$ over k's, which means both of scale and shape parameters of X_k are independent of k. However, this may be too stringent that the GP may not be suitable for many real-world data.

In order to improve these limitations, several GP variants have been introduced and these will be discussed in the following part.

2.1 Extensions of the GP

2.1.1 α-series process

The α-series process (α-series) is proposed by Braun et al. (2005). It considers that the expected number of counts at an arbitrary time does not exist for the decreasing geometric process. Then the decreasing version of the α-series process has a finite expected number of counts within certain conditions.

According to Braun et al. (2005), suppose $Y_1, Y_2, ...$ is a sequence of independent and identically distributed random variables, it defines

$$X_k^A = \frac{Y_k}{k^a},$$

where X_k^A is a related process to X_k^G, which is a GP. Then,

$$Y_k = X_k^A k^a, \tag{1}$$

where a is a real-valued parameter, then the sum of X_k^A is given by

$$S_n^A = \sum_{k-1}^{n} X_k^A,$$

set

$$S_0^A = 0,$$

and the point process can be defined

$$N^A(t) = \sup\{k : S_k{}^A \leqslant t\},$$

where $N^A(t)$ is an α-series process.

Besides, it indicates that for the α-series process, $E(N^A(t)) < \infty$ in the increasing case, as well as in the decreasing case, provided that the random variable Y has a sufficient number of moments. Then

- if $\alpha < 0$, then the $E(N^A(t))$ is finite for all $t > 0$.
- if $0 \leqslant \alpha \leqslant 1/2$ and $E(|Y_n|^{2+\xi}) < \infty$, then the $E(N^A(t))$ is finite for all $t > 0$.

- if $1/2 \leqslant \alpha \leqslant 1$ and $E(|Y_n|^{(1+\xi)/(1-\alpha)}) < \infty$, then the $E(N^A(t))$ is finite for all $t > 0$.

Then the definition of an α-series process can be obtained.

Definition 4. *Braun et al. (2005) Given a sequence of non-negative random variables $\{X_k, k = 1, 2, ...\}$, if they are independent and the cdf of X_k is given by $F(k^a x)$ for $k = 1, 2...$, where a is a positive constant.*

For an α-series process, based on Definitions 1 and 2, $g(\delta_1, k) = k^a$ and $h(\delta_2, k) = 1$ can be obtained.

2.1.2 Threshold geometric process

The threshold geometric process (TGP) is proposed by Chan et al. (2006). The main characterise of TGP is that it can describe non-monotonous trends in a failure process. According to Chan et al. (2006), it has separate GPs and the kth GP is denoted

$$\mathrm{GP}_k = \{X_i, T_k \leqslant i < T_{k+1}\}, k = 1, ..., K,$$

to each trend with turning point T_k. Then each GP has its a and the following definition can be introduced

Definition 5. *Chan et al. (2006) Given a sequence of non-negative random variables $\{X_k, k = 1, 2, ...\}$, if they are independent and the cdf of X_k is given by $F(a_k^{i-T_k} x)$ for $k = 1, 2...$, where a is a positive constant, then $\{X_k, k = 1, 2, ...\}$ is called a threshold geometric process.*

Then, based on Definitions 1 and 2, $g(\delta_1, k) = a_k^{i-T_k}$ and $h(\delta_2, k) = 1$ can be obtained.

2.1.3 Extended Possion Process

The Extended Possion Process (EPP) is proposed by Wu & Clements-Croome (2006). Similar to the TGP and DPG, the EPP can model a failure process with non-monotonous trends.

Definition 6. *Wu & Clements-Croome (2006) Given a sequence of non-negative random variables $\{X_k, k = 1, 2, ...\}$, if they are independent and the cdf of X_k is given by $F((\alpha a^{k-1} + \beta b^{k-1})x)$ for $k = 1, 2...$, where $\alpha + \beta \neq 0$, $\alpha, \beta \geq 0$, $a \geq 0$ and $0 < b \leq 1$, then $\{X_k, k = 1, 2, ...\}$ is called as an extended possion process (EPP).*

Besides, a EPP has these two main properties:

- if $\alpha a^{k-1} \neq 0$ and $b = 1$,
  - (a) if $a > 1$, then the EPP can model a failure process with decreasing failure intensity functions with respect to k;
  - (b) if $a < 1$, then the EPP can model a failure process with increasing failure intensity functions with respect to k;
- if $a = 1, b < 1$ and $\beta b^{k-1} \neq 1$, then $\{X_k, k = 1, 2, ...\}$ can model the process with increasing increasing failure intensities over k.

Then, based on Definitions 1 and 2, $g(\delta_1, k) = (\alpha a^{k-1} + \beta b^{k-1})$ and $h(\delta_2, k) = 1$ can be obtained.

2.1.4 Doubly geometric process

The doubly geometric process (DGP) is proposed by Wu (2018). Different from the GP, the DGP can model a situation that the shape parameters of the lifetime distributions of inter-arrival times X_k changes k. Besides, it can model not only monotonously increasing or decreasing stochastic processes, but also a failure process with different trends (stochastically increasing or decreasing), which is similar to TGP. The definition of a DPG is given in the following definition.

Definition 7. *wu2017doubly Given a sequence of non-negative random variables $\{X_k, k = 1,2,...\}$, if they are independent and the cdf of X_k is given by $F(a^{k-1}x^{h(k)})$ for $k = 1,2...$, where a is a positive constant, $h(k)$ is a function of k and the likelihood of the parameters in $h(k)$ has a known closed form, and $h(k) > 0$, then $\{X_k, k = 1,2,...\}$ is called as a doubly geometric process (DGP). The a^{k-1} is refereed as the scale impact factor and $h(k)$ as the shape impact factor.*

The series $\{a^{k-1}, k = 1,2...\}$ and $\{h(k), k = 1,2...\}$ can be two different geometric series. According to Definition 7, if $h(k)$ equals to 1, then the shape parameter becomes constant over k if X_1 follows a Weibull distribution. If $h(k)$ is not 1, Wu (2018) uses

$$h(k) = (1+\log(k))^b$$

where log is the logarithm with base 10 and b is a parameter. Based on Definition 7, a DGP can model more flexible processes and it has the following properties

- if $0 < a < 1$ and $b < 0$, then $\{X_k, k = 1,2,...\}$ is stochastically increasing.
- if $a > 1$ and $b < 0$, then $\{X_k, k = 1,2,...\}$ is stochastically decreasing.
- if $0 < a < 1$ and $0 < b < 4.898226$, then $\{X_k, k = 1,2,...\}$ is stochastically increasing.
- if $a > 1$ and $0 < b < 4.898226$, then $\{X_k, k = 1,2,...\}$ is stochastically decreasing.

Then, based on Definitions 1 and 2, $g(\delta_1, k) = a^{k-1}$ and $h(\delta_2, k) = (1+\log(k))^b$ can be obtained.

2.1.5 Doubly-ratio geometric process

The doubly-ratio geometric process (DRGP) is proposed by Wu (2022). Suppose that the hazard function of X_k is denoted by $r_k(t)$ and that $\{X_k, k = 1,2,\ldots\}$ follows the GP, then

$$h_k(t) = ah_{k-1}(at), \tag{2}$$

where the two a's play different roles in and have different implications in describing maintenance effectiveness: the first a describes the effectiveness on how the hazard function is affected and the second a (i.e., the one multiplying t in the parentheses) describes the effectiveness on how the age of the item under maintenance is affected. Therefore, the following definition of the DRGP can be given

Definition 8. *(Wu, 2022) Given a sequence of non-negative random variables $\{X_k, k = 1,2,...\}$, if they are independent and the cdf of X_k is given by $F_k(t) = 1-(1-F_1(a_k t))^{b_k/a_k}$ for $k = 1,2...$, where a_k and b_k are positive parameters and $a_1 = b_1 = 1$. Then $\{X_k, k = 1,2,...\}$ is a double-ratio geometric process.*

It has the following proprieties

- If both a_k and b_k are increasing in k, then the DRGP is stochastically decreasing,

- If both a_k and b_k are decreasing in k, then the DRGP is stochastically increasing, and
- If a_k or b_k is increasing in k and b_k or a_k is decreasing in k, then the DRGP may not be stochastically monotonic.

Based on Definitions 4, 5, 6, 7 and 8, the following table summarizes the characterises of the GP and its extensions.

Table 1 – GP and its extensions.

$g(\delta_1,k)$	$h(\delta_2,k)$	Cdf	Model	Reference
1	1	$F(x)$	Renewal Process (RP)	Lam (1988)
a^{k-1}	1	$F(a^{k-1}x)$	Geometric process (GP)	Lam (1988)
k^a	1	$F(k^a x)$	α-series process (α-series)	Braun et al. (2005)
a^{i-M_k}	1	$F(a^{i-M_k}x_i)$	Threshold geometric process (TGP)	Chan et al. (2006)
$\alpha a^{k-1}+\beta b^{k-1}$	1	$F((\alpha a^{k-1}+\beta b^{k-1})x)$	Extended Possion process (EPP)	Wu & Clements-Croome (2006)
a^{k-1}	$(1+\log(n))^b$	$F(a^{k-1}x^{(1+\log(k))^b})$	Doubly geometric process (DGP)	Wu (2018)
a_k	b_k	$1-(1-F(a_k x))^{b_k/a_k}$	Doubly ratio geometric process (DRGP)	Wu (2022)

3 LEAST SQUARE ESTIMATION

In the Section 4.2 in Lam (2007), the least squares estimators of a GP is given. Similarly, the least square method can be applied to other GP's extensions.

3.1 Least square estimation of α-series

For the α-series process, let

$$Y_k = k^a X_k, k = 1,2,....n. \tag{3}$$

Then $\{Y_k, k = 1,2,...\}$ is a sequence of i.i.d random variables, so is $\{\ln Y_k, k = 1,2,...\}$. Taking logarithm on the both sides of Eq (3) give

$$\ln Y_k = a\ln k + \ln X_k, k = 1,2,...n. \tag{4}$$

Let

$$\ln Y_k = \mu_\alpha + e_k, k = 1,2,...n, \tag{5}$$

then,

$$\ln X_k = \mu_\alpha - a\ln k + e_k, k = 1,2,...n. \tag{6}$$

Lemma 1. *Denote Q_α as the sum squares of errors*

$$Q_\alpha = \sum_{k=1}^{n} (\ln X_k - \mu_\alpha + a\ln k)^2, \tag{7}$$

minimising Q_α then the least square estimator of μ_α is given by

$$\mu_\alpha = \frac{1}{n}\sum_{k=1}^{n} (\ln X_k + a\ln k). \tag{8}$$

Proof. Set the derivatives of Q_α to 0,

$$\frac{\partial Q_\alpha}{\partial \mu_\alpha} = 0,$$

then

$$-2\sum_{k=1}^{n} (\ln X_k - \mu_\alpha + a\ln(k)) = 0,$$

$$\sum_{k=1}^{n} \mu_\alpha = \sum_{k=1}^{n} (\ln X_k + a\ln(k)),$$

$$n\mu_\alpha = \sum_{k=1}^{n} (\ln X_k + a\ln(k)),$$

$$\mu_\alpha = \frac{1}{n}\sum_{k=1}^{n} (\ln X_k + a\ln k).$$

This completes the proof. □

3.2 Least square estimation of TGP

According to Chan et al. (2006), for the TGP, let

$$Y_j = a_j^{1-k} X_{j+T_{k-1}-1}, k = 1,2,\ldots n, j = T_{k-1}, T_{k-1}+1, \ldots T_k - 1, \tag{9}$$

where $T_0 = 0$. Then taking logarithm on the both sides of Eq. (9) gives

$$\ln Y_j = (1-k)\ln a + \ln X_{j+T_{k-1}-1}, k = 1,2,\ldots n, j = T_{k-1}, T_{k-1}+1, \ldots T_k - 1. \tag{10}$$

Let

$$\ln Y_j = \mu_j + e_{j+T_{k-1}-1}, k = 1,2,\ldots n, j = T_{k-1}, T_{k-1}+1, \ldots T_k - 1, \tag{11}$$

then,

$$\ln X_{j+T_{k-1}-1} = \mu_j - (1-k)\ln a + e_{j+T_{k-1}-1}, k = 1,2,\ldots n, j = T_{k-1}, T_{k-1}+1, \ldots T_k - 1. \tag{12}$$

Lemma 2. *Denote Q_{tgp} as the sum squares of errors*

$$Q_{tgp} = \sum_{j=T_{k-1}}^{T_k-1} Q_j, \tag{13}$$

and

$$\sum_{j=T_{k-1}}^{T_k-1} Q_j = \sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} (X_{j+T_{k-1}-1} - \mu_j a_j^{1-k})^2, \tag{14}$$

minimising Q_{tgp} then the least square estimator of μ_j is given by

$$\mu_j = \frac{\sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} (X_{j+T_{k-1}-1})}{\sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} a_j^{1-k}}. \tag{15}$$

Proof. Set the derivative of Q_{tgp} to 0,

$$\begin{cases} \dfrac{\partial Q_0}{\partial \mu_0} = 0, \\ \dfrac{\partial Q_1}{\partial \mu_1} = 0, \\ \vdots \\ \dfrac{\partial Q_j}{\partial \mu_j} = 0, \end{cases}$$

then

$$\begin{aligned} -2a_j^{1-k} \sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} (X_{j+T_{k-1}-1} - \mu_j a_j^{1-k}) &= 0, \\ \sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} (X_{j+T_{k-1}-1} - \mu_j a_j^{1-k}) &= 0, \\ \sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} (X_{j+T_{k-1}-1}) - \mu_j \sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} a_j^{1-k} &= 0, \\ \mu_j &= \frac{\sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} (X_{j+T_{k-1}-1})}{\sum_{k=1}^{n} \sum_{j=T_{k-1}}^{T_k-1} a_j^{1-k}}. \end{aligned}$$

This completes the proof. □

3.3 Least square estimation of EPP

Similarly, for the EPP, let

$$Y_k = (\alpha a^{k-1} + \beta b^{k-1}) X_k, k = 1, 2, ...n. \tag{16}$$

Then $\{Y_k, k = 1, 2, ...\}$ is a sequence of i.i.d random variables, so is $\{\ln Y_k, k = 1, 2, ...\}$. Taking logarithm on the both sides of Eq. (16) give

$$\ln Y_k = \ln(\alpha a^{k-1} + \beta b^{k-1}) + \ln X_k, k = 1, 2, ...n. \tag{17}$$

Let

$$\ln Y_k = \mu_{epp} + e_k, k = 1, 2, ...n, \tag{18}$$

then,

$$\ln X_k = \mu_{epp} - \ln(\alpha a^{k-1} + \beta b^{k-1}) + e_k, k = 1, 2, ...n. \tag{19}$$

Lemma 3. *Denote Q_{epp} as the sum squares of errors*

$$Q_{epp} = \sum_{k=1}^{n} (\ln X_k - \mu_{epp} + \ln(\alpha a^{k-1} + \beta b^{k-1}))^2, \tag{20}$$

minimising Q_{epp} then the least square estimator of μ_{epp} is given by

$$\mu_{epp} = \frac{1}{n} \sum_{k=1}^{n} \Big(\ln X_k + \ln(\alpha a^{k-1} + \beta b^{k-1})) \tag{21}$$

Proof. Set the derivatives of Q_{epp} to 0,

$$\frac{\partial Q_{epp}}{\partial \mu_{epp}} = 0,$$

then

$$\begin{aligned} -2\sum_{k=1}^{n} (\ln X_k - \mu_{epp} + \ln(\alpha a^{k-1} + \beta b^{k-1})) &= 0, \\ \sum_{k=1}^{n} \mu_{epp} &= \sum_{k=1}^{n} (\ln X_k + \ln(\alpha a^{k-1} + \beta b^{k-1})), \\ n\mu_{epp} &= \sum_{k=1}^{n} (\ln X_k + \ln(\alpha a^{k-1} + \beta b^{k-1})), \\ \mu_{epp} &= \frac{1}{n} \sum_{k=1}^{n} (\ln X_k + \ln(\alpha a^{k-1} + \beta b^{k-1})). \end{aligned}$$

This completes the proof. □

3.4 Least square estimation of DGP

Lemma 4. *Denote Q_{dgp} as the sum squares of errors, the Q_{dgp} is*

$$Q_{dpg} = \sum_{k=1}^{n} (\ln X_k - \frac{\mu_{dgp} - (k-1)\ln a}{h(k)})^2,$$

where $h(k) = (1 + \log(k))^b$ and minimising Q_{dpg}, then, the μ_{dgp} is given by

$$\mu_{dgp} = \frac{\sum_{k=1}^{n} (\ln X_k + \frac{k-1}{h(k)} \ln a)}{n \sum_{k=1}^{n} (\frac{1}{h(k)})}. \tag{22}$$

Proof. Set the derivatives of Q_{dgp} to 0,

$$\frac{\partial Q_{dgp}}{\partial \mu_{dgp}} = 0,$$

then

$$\begin{aligned}
-\frac{2}{h(k)}\sum_{k=1}^{n}(\ln X_k - \frac{1}{h(k)}\mu_{dgp} + \frac{k-1}{h(k)}\ln a) &= 0,\\
\sum_{k=1}^{n}(\ln X_k - \frac{1}{h(k)}\mu_{dgp} + \frac{k-1}{h(k)}\ln a) &= 0,\\
\sum_{k=1}^{n}\frac{\mu_{dgp}}{h(k)} = \sum_{k=1}^{n}\ln X_k + \sum_{k=1}^{n}\frac{k-1}{h(k)}\ln a&,\\
n\mu_{dgp}\sum_{k=1}^{n}\frac{1}{h(k)} = \sum_{k=1}^{n}\ln X_k + \sum_{k=1}^{n}\frac{k-1}{h(k)}\ln a&,\\
\mu_{epp} = \frac{1}{n}\sum_{k=1}^{n}(\ln X_k + \ln(\alpha a^{k-1} + \beta b^{k-1}))&.
\end{aligned}$$

This completes the proof. □

4 TEST DESIGN

4.1 Independent and identically distributed test

One important assumption of the GP and its extension is that $\{X_n, n = 1,2,...\}$ are independent and identically distributed (i.i.d) random variables. According to Theorem 4.2.1, Equations 4.2.1 and 4.2.2 in Lam (2007), if denote

$$Y_k = \frac{X_{2k}}{X_{2k-1}}, k = 1,2...,$$

and

$$Y_k' = \frac{X_{2k+1}}{X_{2k}}, k = 1,2...,$$

if $\{X_k, k = 1,2,...\}$ is a GP, then $\{Y_k, k = 1,2,...\}$ and $\{Y_k', k = 1,2,...\}$ are respectively two sequence of i.i.d random variable.

The statistical R package "spgs" is used in this paper for the i.i.d test (download at https://cran.r-project.org/web/packages/spgs/index.html).

4.2 Evaluation of the model performance

In this paper, the Akaike information criterion (AIC), the corrected AIC (AICc) will be used as index to compare the fitness among different GPs. The model with lower AIC score is better in this test.

$$\text{AIC} = -2\ln(L) + 2q$$

and

$$\text{AICc} = \text{AIC} + \frac{2q(q+1)}{n-q-1}$$

where q is the number of parameters in a model, L is the maximum likelihood (ML) and n is the number of observations in a dataset. Among the candidate model with the smallest AIC and AICc is regarded the best model.

The ML will be used to estimate the performance of a model. The model with the largest ML score regarded the best.

The statistical R package ”GenSA” is used in this paper for optimization strategy of parameters (download at https://cran.r-project.org/web/packages/GenSA/index.html).

5 DATASETS

The following table shows the datasets and their reference.

Table 2 – Summary of datasets.

Dataset	n	Description	Reference
1	23	Failures of a load-haul-dump (LHD) machine	Kumar & Klefsjö (1992)
2	25	Failures of a load-haul-dump (LHD) machine	Kumar & Klefsjö (1992)
3	27	Failures of a load-haul-dump (LHD) machine	Kumar & Klefsjö (1992)
4	28	Failures of a load-haul-dump (LHD) machine	Kumar & Klefsjö (1992)
5	26	Failures of a load-haul-dump (LHD) machine	Kumar & Klefsjö (1992)
6	23	Failures of a load-haul-dump (LHD) machine	Kumar & Klefsjö (1992)
7	69	Failures of the Armoured Flexible Conveyor (AFC)	Gupta et al. (2009)
8	33	Failures of the Powertrain System of bus	Guida & Pulcini (2009)
9	54	Failures of the Powertrain System of bus	Guida & Pulcini (2009)
10	30	Failures of the air conditioning system of aircraft	Proschan (2000)
11	29	Failures of the air conditioning system of aircraft	Proschan (2000)
12	24	Failures of the air conditioning system of aircraft	Proschan (2000)
13	24	Failures of the air conditioning system of aircraft	Proschan (2000)
14	65	Failures of the main pumps at oil refineries	Percy & Alkali (2007)
15	51	Failures of the main pumps at oil refineries	Percy & Alkali (2007)
16	18	Failures of material fatigue or component aging	Xie & Lai (1996)
17	128	Failures of computer break down	Chakraborty & Chakravarty (2012)
18	135	Failures of software reliability	Musa (1979)
19	190	Failures of the coal-mining disaster	Andrews & Herzberg (2012)
20	40	Failures of the car	Hand et al. (1993)
21	92	Spread of SARS epidemic	Chan et al. (2006)
22	68	Spread of SARS epidemic	Chan et al. (2006)
23	30	Arrival and inter-arrival for the valve	Modarres (2006)
24	71	Failure data related to the U.S.S. Halfbeak motor	Ascher & Feingold (1984)
25	56	Failure data of diesel engine	Lee (1980)

In Table 2, 25 datasets are used in this paper. The following Table 3 shows independence test for each dataset. The dataset with a p-value lower than 0.05 means that it does not follow the independence assumption, therefore, such datasets will not be used for further comparisons.

Table 3 – Summary of independence test.

Dataset	independence p-value	uniformity p-value
1	0.29	1.00
2	0.35	1.00
3	0.53	1.00
4	0.92	0.99
5	0.25	1.00
6	0.87	1.00
7	0.98	1.00
8	0.01	1.00
9	0.40	1.00
10	0.91	0.99
11	0.80	0.99
12	0.68	1.00
13	0.61	1.00
14	0.19	1.00
15	0.19	1.00
16	0.00	0.94
17	0.84	0.99
18	0.87	1.00
19	0.08	1.00
20	0.97	0.99
21	0.00	0.94
22	0.00	0.99
23	0.28	1.00
24	0.50	1.00
25	0.28	1.00

According to the above result, the following findings can be obtained:

- The size of our most dataset is not huge enough ($n \leqq 30$) so that it is impossible to use Chi-square test.
- Datasets 8, 16, 21 and 22 (BUS510, Xie, Hong Kong SARA and Singapore SARA) are not satisfied with the conditions of independence as the independence p-value is less than 0.05. Thus, there datasets will be ignored in this paper.

It is worth noticing that the Section 5 in Lam (2007) briefly studies the model performance by the GP, the RP or the homogeneous Poisson process (HPP) model, the Cox-Lewis model and the Weibull process (WP) model based on ten real datasets. Among them, six datasets, which are 11, 13, 19, 21, 22 and 24, will be used in this paper. To compare the fitness of models, the mean squared error (MSE) and the maximum percentage error (MPE) will be used to evaluate the model performance. In Lam (2007), the GP model is the best model among these four models.

In this paper, more different datasets, which have been shown in Table 2 will be used for comparison of the performance among more GP's extensions in this paper. AIC, AICc and ML will be used to evaluate the model performance. The next section describes numerical results in this paper.

6 CASE STUDIES

The following Table 4 shows the model performance based on 25 datasets. The order of data in each sheet from top to bottom is AIC, AICc and ML.

It is worth noticing that the results are shown in the sequence in the rank of model performance.

Table 4 – The values of AIC, AICc, and ML of the models.

Dataset	RP	GP	EPP	α-series	DGP	TGP	DRGP
	263.97	263.03	267.02	263.51	265.01	**260.12**	265.03
1	264.57	264.29	270.55	264.78	267.23	**261.38**	267.25
	-129.99	-128.52	-128.51	-128.75	-128.51	**-124.06**	-128.52
	301.45	303.04	307.04	301.28	299.35	**297.33**	298.23
2	301.99	304.18	310.20	302.42	301.35	**298.47**	300.23
	-148.72	-148.52	-148.52	-147.64	-145.67	**-142.66**	-145.11
	337.10	334.33	338.32	334.18	336.21	**324.12**	336.08
3	337.60	335.37	341.18	335.23	338.03	**325.16**	337.89
	-166.55	-164.16	-164.16	-164.09	-164.10	**-156.06**	-164.04
	320.10	322.06	326.06	321.24	321.97	320.27	**319.57**
4	**320.58**	323.06	328.79	322.24	323.71	321.27	321.31
	-158.05	-158.03	-158.03	-157.62	-156.99	**-154.13**	-155.79
	306.41	306.65	310.65	**306.24**	308.38	308.63	308.23
5	**306.92**	307.74	313.65	307.32	310.28	309.72	310.14
	-151.20	-150.33	-150.33	-150.12	-150.19	**-148.31**	150.11
	278.53	280.52	284.52	280.35	281.93	279.29	281.43
6	**279.13**	281.79	288.05	281.61	284.15	280.56	283.65
	-137.27	-137.26	-137.26	-137.17	-136.96	**-133.65**	-136.71
	783.58	785.21	787.59	785.13	787.06	785.18	787.13
7	**783.76**	785.58	788.54	785.50	787.69	785.55	787.76
	-389.79	-389.61	-388.80	-389.57	-389.53	**-386.59**	-389.57
	716.36	717.20	**715.78**	718.05	716.88	720.44	717.28
9	716.36	717.20	**715.78**	718.05	716.88	720.44	717.28
	-356.18	-355.60	**-352.89**	-356.03	-354.44	-354.22	-354.64
	307.87	**306.42**	309.81	306.53	308.42	308.35	308.52
10	308.32	**307.35**	312.31	307.46	310.02	309.27	310.12
	-151.94	-150.21	-149.91	-150.27	-150.21	**-148.18**	-150.26
	313.39	315.00	318.75	315.00	317.0014	320.31	**295.31**
11	313.85	315.96	321.36	315.96	318.67	321.27	**297.13**
	-154.69	-154.50	-154.38	-154.50	-154.50	-154.15	**-143.65**
	251.70	253.62	255.47	253.69	253.76	**242.03**	246.84
12	252.27	254.82	258.80	254.89	255.86	**243.24**	249.06
	-123.85	-123.81	-122.73	-123.84	-122.88	**-115.02**	-119.42
	254.73	256.32	260.04	256.53	258.25	261.65	258.11
13	**255.30**	257.52	263.38	257.73	260.35	262.85	260.21
	-125.37	-125.16	-125.02	-125.26	-125.12	**-124.82**	-125.05

Table 4 – The values of AIC, AICc, and ML of the models.

Dataset	RP	GP	EPP	α-series	DGP	TGP	DRGP
	607.28	606.41	610.01	**605.46**	607.94	606.46	606.92
14	607.47	606.80	611.03	**605.85**	608.61	607.70	607.59
	-301.64	-300.20	-300.01	-299.73	-299.97	**-297.23**	-299.46
	497.91	628.00	631.65	628.69	630.00	628.09	500.95
15	**498.16**	628.51	632.99	629.21	630.87	628.60	501.82
	-246.96	-311.00	-310.83	-311.35	-311.00	-308.05	**-246.48**
	615.81	617.43	616.58	617.77	618.75	617.89	617.73
17	**615.90**	617.63	617.08	617.97	619.07	618.09	618.06
	-305.90	-305.72	-303.29	-305.89	-305.37	**-299.95**	-304.87
	1981.17	**1925.02**	1928.17	1935.15	1925.49	1927.94	1926.09
18	1981.26	1925.21	1928.64	1935.33	**1925.80**	1928.12	1926.40
	-988.58	-959.51	-959.09	-964.58	-958.74	**-957.97**	-959.04
	2347.00	2324.06	2328.06	2334.13	2322.68	**2320.82**	2324.50
19	2347.06	2324.19	2328.38	2334.26	2322.89	**2320.95**	2324.72
	-1171.50	-1159.03	-1159.03	-1164.07	-1157.34	**-1151.41**	-1158.25
	385.73	381.66	384.23	**381.21**	382.340	379.50	383.39
23	386.17	382.59	386.73	382.13	384.00	**380.42**	384.99
	-190.86	-187.83	-187.12	-187.60	-187.20	**-180.75**	-187.70
	949.35	926.44	930.44	919.76	925.66	**918.82**	922.19
24	386.17	382.59	386.73	382.13	384.00	**380.42**	384.99
	-472.67	-460.22	-460.22	-456.88	-458.83	**-450.41**	-457.09
	742.58	743.53	746.61	**743.48**	745.50	744.95	745.44
25	**742.80**	743.99	747.81	743.94	746.28	745.41	746.23
	-369.29	-368.77	-368.31	-368.74	-368.75	**-366.47**	-368.72
	650.12	651.80	654.88	652. 33	652.65	**643.50**	644.9943
Average	**589.06**	590.09	594.07	591.04	589.53	589.23	623.80
	-323.06	-322.90	-322.44	-323.17	-322.32	**-315.18**	-318.50

According to the above table, the following result can be obtained:

- In terms of the AIC result
  - Based on the rank of model performance, the result shows that $\mathrm{AIC_{TGP}} > \mathrm{AIC_{RP}} > \mathrm{AIC}_{\alpha\text{-series}} > \mathrm{AIC_{GP}} = \mathrm{AIC_{DRGP}} > \mathrm{AIC_{EPP}} > \mathrm{AIC_{DGP}}$.
  - Based on the rank of the average AIC score, $\mathrm{AIC_{TGP}} > \mathrm{AIC_{DRGP}} > \mathrm{AIC_{RP}} > \mathrm{AIC_{GP}} > \mathrm{AIC}_{\alpha\text{-series}} > \mathrm{AIC_{DRGP}} > \mathrm{AIC_{EPP}}$.
  - The TGP has the best performance (6 best scores),the second one is the RP (5 best scores), then the α-series (4 best scores), EPP (2 best scores) and DRGP (2 best scores) have similar performance.
  - The EPP works better than the DGP in these 23 datasets. However, the DGP model has lower average AIC score than the EPP.
  - The DRGP has the second lowest average AIC score, its combined performance is better than other five models and is second only to the TGP.

- In terms of the AICc result
  - Based on the rank of model performance, the result shows that $AICc_{RP} > AICc_{TGP} > AICc_{\alpha\text{-series}} = AICc_{GP} = AICc_{DGP} = AICc_{DRGP} > AICc_{EPP}$.
  - Based on the rank of the average AICc score, the result shows that $AICc_{RP} > AICc_{TGP} > AICc_{DGP} > AICc_{GP} > AICc_{EPP} > AICc_{\alpha\text{-series}} > AICc_{DRGP}$.
  - The RP has the best performance (9 best scores), second one is the TGP (7 best scores). The performance of these two models is more pronounced in terms of AICc.
  - In terms of AICc, TGP has the best performance, which is different from the result of AIC.
  - The performance of the GP is better in terms of AICc (comparing to the AIC).
  - The performance of the DRGP has obvious disadvantage in terms of AICc. It is 623.80 which is much higher than other six models.
- In terms of the ML result
  - The TGP has significant advantage on model performance.
  - Based on the rank of model performance, the result shows that $ML_{TGP} > ML_{DRGP} > ML_{EPP} > ML_{RP} = ML_{GP} = ML_{\alpha\text{-series}} = ML_{DGP}$.
  - Based on the rank of the average ML score, the result shows that $ML_{TGP} > ML_{DRGP} > ML_{DGP} > ML_{EPP} > ML_{GP} > ML_{RP} > ML_{\alpha\text{-series}}$.
  - The performance of both DRGP and DGP has significant increase comparing with the result in terms of AIC and AICc.

The following table shows the best model for each dataset in terms of different index.

Table 5 – Result of the performance for datasets.

Dataset	AIC	AICc	ML
1	TGP	TGP	TGP
2	TGP	TGP	TGP
3	TGP	TGP	TGP
4	DRGP	RP	TGP
5	α-series	RP	TGP
6	RP	RP	TGP
7	RP	RP	TGP
9	EPP	EPP	EPP
10	GP	GP	TGP
11	DRGP	DRGP	DRGP
12	TGP	TGP	TGP
13	RP	RP	TGP
14	α-series	α-series	TGP
15	RP	RP	DRGP
17	RP	RP	TGP
18	GP	DGP	TGP
19	TGP	TGP	TGP
20	α-series	TGP	TGP
24	α-series	RP	TGP
25	TGP	RP	α-series
Best	TGP	RP	TGP
Second Best	RP	TGP	DRGP

7 CONCLUSIONS

This paper compared the performances of the geometric process (GP) and its extensions based on 25 real-world datasets. These extensions, the α-series process, the threshold geometric process (TGP), the extended possion process (EPP), the doubly geometric process (DGP) and the doubly-ratio geometric process (DRGP), have been used in this paper. Three performances metrics, AIC (Akaike Information Criterion), AICc (correction of AIC) and ML (Maximum Likelihood) were used. Besides, it discussed the LSE (least square estimation) as an alternative method to estimate the model parameters.

According to the comparison results, based on the AIC, the best model is the TGP and the second best is RP. Then in terms of the AICc, the best model is the DRGP and the second best is the TGP. Based on the result of the ML, the best model is the TGP and the second best is the DRGP. These results would give supports for choosing the best model in practical applications.

References

ANDREWS DF & HERZBERG AM. 2012. *Data: a collection of problems from many fields for the student and research worker*. Springer Science & Business Media.

ASCHER H & FEINGOLD H. 1984. *Repairable systems reliability: modeling, inference, misconceptions and their causes*. M. Dekker New York.

BRAUN WJ, LI W & ZHAO YQ. 2005. Properties of the geometric and related processes. *Naval Research Logistics (NRL)*, **52**(7): 607–616.

CHAKRABORTY S & CHAKRAVARTY D. 2012. Discrete gamma distributions: Properties and parameter estimations. *Communications in Statistics-Theory and Methods*, **41**(18): 3301–3324.

CHAN JS, YU PL, LAM Y & HO AP. 2006. Modelling SARS data using threshold geometric process. *Statistics in medicine*, **25**(11): 1826–1839.

FINKELSTEIN MS. 1993. A scale model of general repair. *Microelectronics Reliability*, **33**(1): 41–44.

GUIDA M & PULCINI G. 2009. Reliability analysis of mechanical systems with bounded and bathtub shaped intensity function. *IEEE Transactions on Reliability*, **58**(3): 432–443.

GUPTA S, MAITI J, KUMAR R & KUMAR U. 2009. A control chart guided maintenance policy selection. *International Journal of Mining, Reclamation and Environment*, **23**(3): 216–226.

HAND D, DALY F, MCCONWAY K, LUNN D & OSTROWSKI E. 1993. *A Handbook of Small Data Sets*. CRC Press. Available at: https://books.google.co.uk/books?id=m4O1DwAAQBAJ.

KUMAR U & KLEFSJÖ B. 1992. Reliability analysis of hydraulic systems of LHD machines using the power law process model. *Reliability Engineering & System Safety*, **35**(3): 217–224.

LAM Y. 1988. A note on the optimal replacement problem. *Advances in Applied Probability*, **20**(2): 479–482.

LAM Y. 2007. *The geometric process and its applications*. World Scientific: Singapore.

LAM Y, ZHU LX, CHAN JS & LIU Q. 2004. Analysis of data from a series of events by a geometric process model. *Acta Mathematicae Applicatae Sinica, English Series*, **20**(2): 263–282.

LEE L. 1980. Testing adequacy of the Weibull and log linear rate models for a Poisson process. *Technometrics*, **22**(2): 195–199.

MODARRES M. 2006. *Risk analysis in engineering: techniques, tools, and trends*. CRC press.

MUSA JD. 1979. Software reliability data. *Technical Report in Rome Air Development Center*, .

PEKALP MH & AYDOĞDU H. 2021. Power series expansions for the probability distribution, mean value and variance functions of a geometric process with gamma interarrival times. *Journal of Computational and Applied Mathematics*, **388**: 113287.

PERCY DF & ALKALI BM. 2007. Scheduling preventive maintenance for oil pumps using generalized proportional intensities models. *International Transactions in Operational Research*, **14**(6): 547–563.

PROSCHAN F. 2000. Theoretical explanation of observed decreasing failure rate. *Technometrics*, **42**(1): 7–11.

ROSS SM, KELLY JJ, SULLIVAN RJ, PERRY WJ, MERCER D, DAVIS RM, WASHBURN TD, SAGER EV, BOYCE JB & BRISTOW VL. 1996. *Stochastic processes*. vol. 2. Wiley New York.

WANG GJ & YAM RC. 2017. Generalized geometric process and its application in maintenance problems. *Applied Mathematical Modelling*, **49**: 554–567.

WU D, PENG R & WU S. 2020. A review of the extensions of the geometric process, applications, and challenges. *Quality and Reliability Engineering International*, **36**(2): 436–446.

WU S. 2018. Doubly geometric processes and applications. *Journal of the Operational Research Society*, pp. 1–13.

WU S. 2022. The double ratio geometric process for the analysis of recurrent events. *Naval Research Logistics (NRL)*, pp. 484–495.

WU S & CLEMENTS-CROOME D. 2006. A novel repair model for imperfect maintenance. *IMA Journal of Management Mathematics*, **17**(3): 235–243.

XIE M & LAI CD. 1996. Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, **52**(1): 87–93.

ZHANG YL. 1999. An optimal geometric process model for a cold standby repairable system. *Reliability Engineering & System Safety*, **63**(1): 107–110.

ZHANG YL & WANG GJ. 2016. An extended geometric process repair model for a cold standby repairable system with imperfect delayed repair. *International Journal of Systems Science: Operations & Logistics*, **3**(3): 163–175.

How to cite

JIAQI Y & SHAOMIN W. 2022. A Comparison of the Performance of the Geometric Process and Its Extensions. *Pesquisa Operacional*, **42** (nspe1): e262123. doi: 10.1590/0101-7438.2022.042nspe1.00262123.