International Journal of Statistics and Probability; Vol. 7, No. 6; November 2018
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Comparative Study of the Non-Homogeneous Poisson Process Type-I Generalized Half-Logistic Distribution

Lutfiah Ismail Al turk[1] & Wejdan Saleem Al ahmadi[2]

[1] Statistics Department, Faculty of Sciences, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

[2] Mathematics Department, Taibah University, Kingdom of Saudi Arabia

Correspondence: Wejdan Saleem Al ahmadi, Mathematics Department, Taibah University, Kingdom of Saudi Arabia

Received: September 16, 2018 Accepted: October 15, 2018 Online Published: October 24, 2018

doi:10.5539/ijsp.v7n6p132 URL: https://doi.org/10.5539/ijsp.v7n6p132

Abstract

In this paper, Non-Homogeneous Poisson Process (NHPP) model are created based on Type-I Generalized Half-Logistic Distribution (GHLD- I). Three methods for estimating the parameters of the NHPP GHLD- I model are considered in the case of failure-occurrence time data, for this purpose the necessary likelihood equations are obtained. Confidence intervals are studied, the upper and the lower bounds of the parameters are constructed. An application based on the NHPP GHLD-I and using four published data sets are conducted. The performance of NHPP GHLD-I model is checked based on three evaluation criteria and useful results are obtained.

Keywords: software reliability growth models, non-homogeneous poisson process, generalized half-logistic model, maximum likelihood estimation, non-linear least square estimation, weighted non-linear least square estimation

1. Introduction

Nowadays, technological achievement needs effective and high accuracy hardware and software in order to make dramatic improvements and reach the expected goals. Statistical reliability modelling is an approach that has been intensively used in quantifying the software system reliability, which describe software failures behavior based on different basic assumptions. The main issue of the traditional reliability models approach which are applied in the testing phase of software development cycle is: to find theoretical distribution that able to well fit the failure time data, to assess the future behavior of time between software failure, to predict software system reliability, and to determine when the software product becomes mature and ready to be released to the user [see: Lai and Garg (2012) and Barraza (2010)].

Several reliability models based on Non-Homogeneous Poisson Processes (NHPP) have been suggested during the past years, they are extensively and successfully used to describe the software failure process [see: Goel and Okumoto (1979), Yamada et al. (1983), Zhang et al. (2003), and Teng and Pham (2004)]. In this paper, a NHPP software reliability model is created, which optimistically will give a good representation of the uncertainty of the software system in the field of software reliability modeling belonging to the NHPP. Our suggested NHPP model is configured using Type-I Generalized Half-Logistic Distribution (GHLD-I) that proposed by [Kantam et al. (2013)] based on failure-occurrence time data. Our proposed model is expected to be flexible, able to well describe the growth phenomena, and useful for modeling life data, it offers several sub-models by changing the shape parameter, so the best fit model could be found easier and faster.

The parameters of the NHPP GHLD-I Model are estimated using Maximum Likelihood (ML), Non-Linear Least Squares (NLS), Weighted Non-Linear Least Squares (WNLS) estimation methods and the confidence intervals are constructed.

The Cumulative Distribution Function (CDF) of the GHLD-I with scale parameter σ and shape parameter θ is given by:

$$F(t) = \left[\frac{1-e^{-\frac{t}{\sigma}}}{1+e^{-\frac{t}{\sigma}}}\right]^{\theta},\ 0 < t < \infty, \sigma > 0, \theta > 0 \tag{1}$$

Using Equation (1), the Probability Density Function (pdf) can be obtained as follows:

$$f(t) = \frac{dF(t)}{dt} = \frac{2\theta e^{-\frac{t}{\sigma}}(1-e^{-\frac{t}{\sigma}})^{\theta-1}}{\sigma(1+e^{-\frac{t}{\sigma}})^{\theta+1}}. \tag{2}$$

Also, the reliability function can be obtained from Equation (1) as follows:

$$\mathrm{R(t)} = 1 - \mathrm{F(t)}$$
$$= 1 - [\frac{1-e^{-\frac{t}{\sigma}}}{1+e^{-\frac{t}{\sigma}}}]^{\theta}. \tag{3}$$

While the hazard function can be found using Equations (2) and (3) as follows:

$$\mathrm{h(t)} = \frac{\mathrm{f(t)}}{\mathrm{R(t)}}$$

$$= \frac{2\theta e^{-\frac{t}{\sigma}}(1-e^{-\frac{t}{\sigma}})^{\theta-1}}{\sigma(1+e^{-\frac{t}{\sigma}})(\left(1+e^{-\frac{t}{\sigma}}\right)^{\theta}-\left(1-e^{-\frac{t}{\sigma}}\right)^{\theta})}\ . \tag{4}$$

The rest of this paper is arranged as follows: Section 2 displays the formulation of the NHPP GHLD-I model and describes the ML, NLS, and WNLS estimation approaches for this model. Section 3 illustrates the evaluation criteria that will be used in our evaluation study. Real data application will be shown in Sections 4. In the end, Section 5 is the conclusion of this paper.

2. NHPP GHLD-I Model

The section illustrates the suggested model formulation and the computation of the necessary mathematical equations for point and interval estimation.

2.1 Model Formulation and Characteristics

In this section, the NHPP GHLD-I model will be constructed by following [Lyu (2002)]:

$$\mathrm{m(t)} = \mathrm{aF(t)}, \tag{5}$$

$$\lambda(\mathrm{t}) = \mathrm{af(t)}, \tag{6}$$

where $\mathrm{F(t)}, \mathrm{f(t)}$ are respectively the CDF and PDF of the time to failure of an individual failure, a. From this, if we consider also distributions that belong to the finite failure type, i.e., $\lim_{t\to\infty} m(t) < \infty$ we have that $\lim_{t\to\infty} m(t) = a$ since $\lim_{t\to\infty} \mathrm{F}(t) = 1$. Thus a represents the eventual number of failures observed in the system if it could have been detected over an infinite amount of time. Then by using Equations (1), (2), (5) and (6) the mean value and failure intensity functions of the NHPP GHLD-I model can be obtained respectively as follows:

$$\mathrm{m(t)} = \mathrm{a}\left(\frac{1-e^{-\frac{t}{\sigma}}}{1+e^{-\frac{t}{\sigma}}}\right)^{\theta}, \tag{7}$$

$$\lambda(\mathrm{t}) = \mathrm{a}[\frac{2\theta e^{-\frac{t}{\sigma}}\left(1-e^{-\frac{t}{\sigma}}\right)^{\theta-1}}{\sigma\left(1+e^{-\frac{t}{\sigma}}\right)^{\theta+1}}], \tag{8}$$

where the parameter a is interpreted as the number of initial faults in the software, σ is the scale parameter and θ is shape parameter of the NHPP GHLD-I model. By using Equation (7) the number of remaining errors of this model can be written as follows:

$$\mathrm{n(t)} = [\mathrm{a} - \mathrm{m(t)}]$$
$$= \mathrm{a}[1 - \left(\frac{1-e^{-\frac{t}{\sigma}}}{1+e^{-\frac{t}{\sigma}}}\right)^{\theta}] \tag{9}$$

By using Equations (8) and (9), the error detection rate can be obtained as:

$$\mathrm{d(t)} = \frac{\lambda\,\mathrm{(t)}}{\mathrm{n(t)}}$$

$$= \frac{2\theta e^{-\frac{t}{\sigma}}\left(1-e^{-\frac{t}{\sigma}}\right)^{\theta-1}}{\sigma\left(1+e^{-\frac{t}{\sigma}}\right)[\left(1+e^{-\frac{t}{\sigma}}\right)^{\theta}-\left(1-e^{-\frac{t}{\sigma}}\right)^{\theta}]} \tag{10}$$

We can obtain the mean time between failures (MTBF) of our suggested model using Equation (8) as:

$$\mathrm{MTBF} = \frac{1}{\lambda(t)}$$
$$= \frac{\sigma\left(1+e^{-\frac{t}{\sigma}}\right)^{\theta+1}}{a2\theta e^{-\frac{t}{\sigma}}\left(1-e^{-\frac{t}{\sigma}}\right)^{\theta-1}} \ . \tag{11}$$

According to Equation (7) the conditional reliability function is:

$$R(x|t) = \exp\left[-\left(m\ (t\ +\ x) - m(t)\right)\right]$$
$$= \exp[-a(\left(\frac{1-e^{-\frac{t+x}{\sigma}}}{1+e^{-\frac{t+x}{\sigma}}}\right)^{\theta} - \left(\frac{1-e^{-\frac{t}{\sigma}}}{1+e^{-\frac{t}{\sigma}}}\right)^{\theta})] \tag{12}$$

2.2 Maximum Likelihood Estimation (MLE) Method

Parameter estimation is of primary importance in software reliability prediction. The ML estimation method is the most important traditional and widely used estimation technique. This technique has several properties including consistency, efficiency and asymptotic normality.

Suppose we have "n" time instants at which the first, second, third..., n^{th} failures of a software are experienced. In other words, if S_k is the total time to the k^{th} failure, s_k is an observation of random variable S_k and "n" such failures are successively recorded. The joint probability of such failure time realizations $s_1, s_2, s_3, \dots, s_n$ is:

$$L = e^{-m(s_n)} \prod_{i=1}^{n} \lambda(s_i) \tag{13}$$

The function given in Equation (13) is also called the likelihood function of the given failure data. Values of the parameters of NHPP models that would maximize L are called maximum likelihood estimators and the method is called maximum likelihood (ML) estimation method [Prasad et al. (2011)].

For the purpose of estimating the unknown three parameters a, θ and σ of the NHPP GHLM-I model using ML estimation method and based on the data on failure occurrence time s_k $(k = 1, 2, \dots, n;\ 0 \le s_1 \le s_2 \le \cdots \le s_n)$, we substitute Equations (7) and (8) in Equation (13) so we obtain the likelihood function as follows:

$$L = e^{-a\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta}} \prod_{i=1}^{n} a\left(\frac{2\theta e^{-\frac{s_i}{\sigma}}\left(1-e^{-\frac{s_i}{\sigma}}\right)^{\theta-1}}{\sigma\left(1+e^{-\frac{s_i}{\sigma}}\right)^{\theta+1}}\right) \tag{14}$$

By taking the natural lnarithm of Equation (14) we obtain:

$$\ln L = -a\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta} + n\ln a + n\ln 2 + n\ln\theta - \sum\frac{s_i}{\sigma} + (\theta-1)\sum\ln\left(1-e^{-\frac{s_i}{\sigma}}\right) - n\ln\sigma -$$
$$(\theta+1)\sum\ln\left(1+e^{-\frac{s_i}{\sigma}}\right). \tag{15}$$

In order to estimate the parameters a, θ and σ, the derivatives of Equation (15) with respect to a, θ and σ will be obtained as follows:

$$\begin{cases} \frac{\partial \ln L}{\partial a} = -\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta} + \frac{n}{a}. \\ \frac{\partial \ln L}{\partial \theta} = -a\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta} \ln\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right) + \frac{n}{\theta} + \sum\ln\left(1-e^{-\frac{s_i}{\sigma}}\right) - \sum\ln\left(1+e^{-\frac{s_i}{\sigma}}\right). \\ \frac{\partial \ln L}{\partial \sigma} = -a\theta\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta-1}\left(\frac{-2e^{-\frac{s_n}{\sigma}}s_n}{\sigma^2\left(1+e^{-\frac{s_n}{\sigma}}\right)^2}\right) + \sum\frac{s_i}{\sigma^2} - (\theta-1)\sum\frac{s_i e^{-\frac{s_i}{\sigma}}}{\sigma^2\left(1-e^{-\frac{s_i}{\sigma}}\right)} - \frac{n}{\sigma} - (\theta+1)\sum\frac{s_i e^{-\frac{s_i}{\sigma}}}{\sigma^2\left(1+e^{-\frac{s_i}{\sigma}}\right)}. \end{cases} \tag{16}$$

By equating the previous equations to zero, the equation becomes:

$$\begin{cases} \hat{a} = n\left(\dfrac{1+e^{-\frac{s_n}{\hat{\sigma}}}}{1-e^{-\frac{s_n}{\hat{\sigma}}}}\right)^{\hat{\theta}}. \\ -a\left(\dfrac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta} \ln\left(\dfrac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right) + \dfrac{n}{\theta} + \sum \ln\left(1-e^{-\frac{s_i}{\sigma}}\right) - \sum \ln\left(1+e^{-\frac{s_i}{\sigma}}\right) = 0. \\ \dfrac{2s_n n\hat{\theta} e^{-\frac{s_n}{\hat{\sigma}}}}{\hat{\sigma}^2(1+e^{-\frac{s_n}{\hat{\sigma}}})(1-e^{-\frac{s_n}{\sigma}})} + \sum \dfrac{s_i}{\hat{\sigma}^2} - (\hat{\theta}-1)\sum \dfrac{s_i e^{-\frac{s_i}{\hat{\sigma}}}}{\hat{\sigma}^2\left(1-e^{-\frac{s_i}{\hat{\sigma}}}\right)} - \dfrac{n}{\hat{\sigma}} - (\hat{\theta}+1)\sum \dfrac{s_i e^{-\frac{s_i}{\hat{\sigma}}}}{\hat{\sigma}^2\left(1+e^{-\frac{s_i}{\hat{\sigma}}}\right)} = 0. \end{cases} \quad (17)$$

The value of the parameter *a* can be obtained using the first expression of Equation (17) after getting the estimate of the parameters θ and σ. Since the second and third expression of Equation (17) are nonlinear, we can not find an analytic solution and must be obtained numerically, to facilitate this R programing language are used.

2.3 Non-Linear Least Square Estimation (NLSE) Method

The least squared sum in the Least Square Estimation (LSE) method is defined by:

$$\mathrm{LSE}(\Theta) = \sum_{i=1}^{n}[i - m(t_i)]^2 \quad (18)$$

where Θ is the parameters of the NHPP model, and $m(t)$ is its mean value function. The resulting estimates of Θ which is obtained by minimizing $\mathrm{LSE}(\Theta)$ is called the OLS estimates and can be calculated by using any non-linear regression technique. Usually, Gauss-Newton method or Lenvenberg-Marquardt algorithm is used to solve the minimization problem arg $\min_\Theta \mathrm{LSE}(\Theta)$. More specifically, we consider formally the above optimization problem from the viewpoint of regression analysis.

For the time epochs $\{\tau_i: i = 1,2,\ldots,N(t)\}$, which are the i.i.d. random variables with realizations $\{t_i: i = 1,2,\ldots,n\}$, N(t) denote the cumulative number of the faults detected by time t. it is easily shown from the time-scale transform of the NHPP that $\tau_i^* = m(\tau_i)$ is a Homogeneous Poisson Process (HPP) with rate 1 and that

$$E[\tau_i^*] = E[m(\tau_i)] = I, \qquad \text{for } i = 1,2,\ldots \quad (19)$$

When one sees the relationship between the random variable $m(\tau_i)$ and its realization $m(t_i)$, it may be straightforward to consider the following regression model:

$$m(\tau_i) = m(t_i) + \varepsilon_i, \qquad \text{for } i = 1,2,\ldots \quad (20)$$

Where $\{\varepsilon_i: i = 1,2,\ldots\}$ are the error terms. In a fashion similar to the usual regression analysis, if the errror terms ε_i are the i.i.d. random variables with mean 0, the OLS is formulated to minimize:

$$\mathrm{LSE}(\Theta) = \sum_{i=1}^{n} E[m(\tau_i) - m(t_i)]^2$$

$$= \sum_{i=1}^{n}[i - m(t_i)]^2 \quad (21)$$

Unfortunately, it is worth mentioning in the NHPP that the error terms in Equation (20), $\{\varepsilon_i: i = 1,2,\ldots\}$, are neither independent nor identically distributed. This fact tells us that the OLS estimation may be irrelevant to the common linear regression analysis. It is evident that the resulting OLS estimates in the LSE can not be representative [Ishii et al (2012)] and the Non-Linear Least Square Estimation (NLSE) is needed.

We can calculate the ordinary least squares estimates of the unknown three parameters a, θ and σ of the NHPP GHLM-I model by minimizing the least squared sum, whereas by using Equation (21), the least squared sum of our model is:

$$\mathrm{NLSE}(\Theta) = \sum_{i=1}^{n}[i - a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta}]^2 \quad (22)$$

2.4 Weighted Least Square Estimation (WNLSE) Method

As notable, some of estimates based on the NLSE estimation are less precise than others in the sense that their variances are relatively larger. Since the variances of NLSE estimates are not equal, it is often necessary to adjust the NLSE estimation in such way that the mean value functions are weighted, i.e.

$$\mathrm{WNLSE}(\Theta) = \sum_{i=1}^{n} w_i\left(i - m(t_i)\right)^2 \quad (23)$$

where $w_i (i = 1,2, \dots, n)$ are positive weights to satisfy $\sum_{i=1}^{n} w_i = n$. The resulting estimates as solutions of argmin Θ WNLSE(Θ) are called the weighted non-linear leas squares (WNLS) estimates, where the NLS estimate is a special case of WNLS estimates when the weights are all the same, i.e., $w_i = 1$ for all i. By taking account of the fact that the variances of the random variable $m(\tau_i)$ in Equation (20) is unequal with respect to i, the error terms $\{\varepsilon_i : i = 1,2, \dots, n\}$ should be normalized as the random variables with variance 1.

Weighting function choice is an important concern when using the WNLSE method. Several ways of weighting techniques can be considered when using this method of estimation [Sun et al. (2012)]. From the analogy to the linear regression analysis the error terms are weighted by w_i that are inversely proportional to their variances, i.e.,

$$w_i = \frac{n}{m(t_i) \sum_{i=1}^{n} \frac{1}{m(t_i)}} \tag{24}$$

Which is due to the common property that the variance of NHPP equals its mean value. It should be noted that the error terms ε_i are not still i.i.d. random variables because our NHPP-based reliability model do not have the linear intensity. Hence, the weighted in Equation (24) is meaningful only when the error terms ε_i can be approximately i.i.d. normal random variables [Ishii et al. (2012)].

The estimates of the unknown parameters a, θ and σ of the NHPP GHLM-I model using the WNLSE method can be obtained by minimizing:

$$\mathrm{WNLSE}(\Theta) = \sum_{i=1}^{n} w_i \left(i - a \left(\frac{1 - e^{-\frac{t_i}{\sigma}}}{1 + e^{-\frac{t_i}{\sigma}}} \right)^{\theta} \right)^2 \tag{25}$$

For our application three weighting functions are considered:

$$w_{i(1)} = n \Big/ \left(\frac{1 - e^{-\frac{t_i}{\sigma}}}{1 + e^{-\frac{t_i}{\sigma}}} \right)^{\theta} \sum_{i=1}^{n} \left(\frac{1 + e^{-\frac{t_i}{\sigma}}}{1 - e^{-\frac{t_i}{\sigma}}} \right)^{\theta} \tag{26}$$

And the following two empirical weight functions are formulated as follows:

$$w_{i(2)} = \sqrt{ n \Big/ \left(\frac{1 - e^{-\frac{t_i}{\sigma}}}{1 + e^{-\frac{t_i}{\sigma}}} \right)^{\theta} \sum_{i=1}^{n} \left(\frac{1 + e^{-\frac{t_i}{\sigma}}}{1 - e^{-\frac{t_i}{\sigma}}} \right)^{\theta} } \tag{27}$$

$$w_{i(3)} = \sqrt[n]{ n \Big/ \left(\frac{1 - e^{-\frac{t_i}{\sigma}}}{1 + e^{-\frac{t_i}{\sigma}}} \right)^{\theta} \sum_{i=1}^{n} \left(\frac{1 + e^{-\frac{t_i}{\sigma}}}{1 - e^{-\frac{t_i}{\sigma}}} \right)^{\theta} } \tag{28}$$

2.5 Interval Estimation Method

The estimation of the model parameters is normally different according to different estimation methods. Furthermore, an accurate parameters estimation requires many failure data which might not be available. So, we can use interval estimation of software reliability models' parameters to solve these problems.

A confidence interval is an interval of numbers containing the most acceptable values for our distribution parameters. Let Θ symbolize the parameters of NHPP model, in order to obtain the confidence limits for parameters Θ of NHPP models we calculate the Fisher information matrix. The Fisher information is a way of measuring the amount of information that an observable random variable s_n carries about unknown parameters Θ of a NHPP models that models s_n. The inverse of the Fisher information matrix gives the asymptotic variance and covariance of the estimates of the parameters Θ of NHPP models. The two sided approximate $100\alpha\%$ confidence limits for the parameters Θ of a NHPP models are:

$$\Theta_{\mathrm{upper}} = \widehat{\Theta} + Z_\alpha [\mathrm{var}(\widehat{\Theta})]^{1l2} \tag{30}$$

$$\Theta_{\mathrm{lower}} = \widehat{\Theta} - Z_\alpha [\mathrm{var}(\widehat{\Theta})]^{1l2} \tag{31}$$

where Θ is the parameters of the NHPP model and $\widehat{\Theta}$ is the ML, NLS or WNLS estimators of these parameters. Z_α is the $(1 - \alpha)$ quartile of the standard normal distribution [Hong et al. (1997)].

In this section, we discuss the interval estimation of the parameters of the NHPP GHLD-I model of ML, NLS and WNLS estimators. In order to obtain the confidence limits for parameter a, θ and σ, we find the Fisher information matrix to obtain the asymptotic variance and covariance of the ML, NLS and WNLS estimates of the parameters. For obtaining confidence intervals of ML estimator, we define Fisher information matrix as the matrix of negative second partial derivatives of the log likelihood function that shown in Equation (15).

$$I_{MLE}(\Theta)=\begin{bmatrix}-\frac{\partial^2 \ln L}{\partial a^2} & -\frac{\partial^2 \ln L}{\partial a\,\partial\theta} & -\frac{\partial^2 \ln L}{\partial a\,\partial\sigma}\\ -\frac{\partial^2 \ln L}{\partial\theta\,\partial a} & -\frac{\partial^2 \ln L}{\partial\theta^2} & -\frac{\partial^2 \ln L}{\partial\theta\,\partial\sigma}\\ -\frac{\partial^2 \ln L}{\partial\sigma\,\partial a} & -\frac{\partial^2 \ln L}{\partial\sigma\,\partial\theta} & -\frac{\partial^2 \ln L}{\partial\sigma^2}\end{bmatrix} \tag{32}$$

such that:

$$\frac{\partial^2 \ln L}{\partial a^2}=\frac{-n}{a^2} \tag{33}$$

$$\frac{\partial^2 \ln L}{\partial a\partial\theta}=\frac{\partial^2 \ln L}{\partial\theta\partial a}=-\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta}\ln\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right) \tag{34}$$

$$\frac{\partial^2 \ln L}{\partial a\,\partial\sigma}=\frac{\partial^2 \ln L}{\partial\sigma\,\partial a}=\frac{2\theta s_n e^{-\frac{s_n}{\sigma}}\left(1-e^{-\frac{s_n}{\sigma}}\right)^{\theta-1}}{\left(1+e^{-\frac{s_n}{\sigma}}\right)^{\theta+1}} \tag{35}$$

$$\frac{\partial^2 \ln L}{\partial\theta^2}=-a\left(\ln\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)\right)^2\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta}-\frac{n}{\theta^2} \tag{36}$$

$$\frac{\partial^2 \ln L}{\partial\theta\,\partial\sigma}=\frac{\partial^2 \ln L}{\partial\sigma\,\partial\theta}=\frac{2ae^{-\frac{s_n}{\sigma}}s_n\left(1-e^{-\frac{s_i}{\sigma}}\right)^{\theta-1}}{\sigma^2\left(1+e^{-\frac{s_n}{\sigma}}\right)^{\theta+1}}\left[\frac{1}{\left(1+e^{-\frac{s_n}{\sigma}}\right)^2}+\theta\ln\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)\right]-\sum\frac{s_i e^{-\frac{s_i}{\sigma}}}{\sigma^2\left(1-e^{-\frac{s_i}{\sigma}}\right)}-\sum\frac{s_i e^{-\frac{s_i}{\sigma}}}{\sigma^2\left(1+e^{-\frac{s_i}{\sigma}}\right)}-$$
$$a\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)^{\theta}\ln\left(\frac{1-e^{-\frac{s_n}{\sigma}}}{1+e^{-\frac{s_n}{\sigma}}}\right)+\frac{n}{\theta}+\sum\ln\left(1-e^{-\frac{s_i}{\sigma}}\right)-\sum\ln\left(1+e^{-\frac{s_i}{\sigma}}\right) \tag{37}$$

$$\frac{\partial^2 \ln L}{\partial\sigma^2}=\frac{2s_n a\theta e^{-\frac{s_n}{\sigma}}\left(1-e^{-\frac{s_n}{\sigma}}\right)^{\theta-1}\left(1+e^{-\frac{s_n}{\sigma}}\right)^{\theta}}{\sigma^4\left(1+e^{-\frac{s_n}{\sigma}}\right)^{\theta+2}}\left[s_n\left(1+e^{-\frac{s_n}{\sigma}}\right)\left(1-\frac{\theta e^{-\frac{s_n}{\sigma}}-e^{-\frac{s_n}{\sigma}}}{1-e^{-\frac{s_n}{\sigma}}}\right)-(\theta+1)e^{-\frac{s_n}{\sigma}}s_n-2\sigma\left(1+e^{-\frac{s_n}{\sigma}}\right)\right]-$$
$$\sum\frac{2s_i}{\sigma^3}-(\theta-1)\sum\frac{s_i e^{-\frac{s_i}{\sigma}}[s_i-s_i e^{-\frac{s_i}{\sigma}}+s_i e^{-\frac{s_i}{\sigma}}-2\sigma\left(1-e^{-\frac{s_i}{\sigma}}\right)}{\sigma^4\left(1-e^{-\frac{s_i}{\sigma}}\right)^2}+\frac{n}{\sigma^2}-(\theta+1)\sum\frac{s_i e^{-\frac{s_i}{\sigma}}[s_i+s_i e^{-\frac{s_i}{\sigma}}-s_i e^{-\frac{s_i}{\sigma}}-2\sigma\left(1+e^{-\frac{s_i}{\sigma}}\right)}{\sigma^4\left(1+e^{-\frac{s_i}{\sigma}}\right)^2} \tag{38}$$

Aimed at obtaining confidence intervals for the NLS estimators, we define the Fisher information matrix as the matrix of negative second partial derivatives of least squared sum function that shown in Equation (22):

$$I_{NLSE}(\Theta)=\begin{bmatrix}-\frac{\partial^2 NLSE}{\partial a^2} & -\frac{\partial^2 NLSE}{\partial a\,\partial\theta} & -\frac{\partial^2 NLSE}{\partial a\,\partial\sigma}\\ -\frac{\partial^2 NLSE}{\partial\theta\,\partial a} & -\frac{\partial^2 NLSE}{\partial\theta^2} & -\frac{\partial^2 NLSE}{\partial\theta\,\partial\sigma}\\ -\frac{\partial^2 NLSE}{\partial\sigma\,\partial a} & -\frac{\partial^2 NLSE}{\partial\sigma\,\partial\theta} & -\frac{\partial^2 NLSE}{\partial\sigma^2}\end{bmatrix} \tag{39}$$

such that:

$$\frac{\partial \mathrm{NLSE}}{\partial \mathrm{a}} = 2\sum_{i=1}^{n}\left[-i\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} + a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{2\theta}\right] \quad (40)$$

$$\frac{\partial^2 \mathrm{NLSE}}{\partial \mathrm{a}^2} = 2\sum_{i=1}^{n}\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{2\theta} \quad (41)$$

$$\frac{\partial^2 \mathrm{NLSE}}{\partial \mathrm{a}\,\partial\theta} = \frac{\partial^2 \mathrm{NLSE}}{\partial\theta\,\partial \mathrm{a}}$$

$$= 2\sum_{i=1}^{n}\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta}\ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)\left[2a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} - i\right] \quad (42)$$

$$\frac{\partial^2 \mathrm{NLSE}}{\partial \mathrm{a}\,\partial\sigma} = \frac{\partial^2 \mathrm{NLSE}}{\partial\sigma\,\partial \mathrm{a}}$$

$$= 4\sum_{i=1}^{n}\frac{-\theta\, t\, e^{-\frac{t_i}{\sigma}}\left((1-e^{-\frac{t_i}{\sigma}})\right)^{\theta-1}}{(1+e^{-\frac{t_i}{\sigma}})^{\theta+1}}\left[2a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} - i\right] \quad (43)$$

$$\frac{\partial \mathrm{NLSE}}{\partial\theta} = 2a\sum_{i=1}^{n}\ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)\left[-i\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} + a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{2\theta}\right] \quad (44)$$

$$\frac{\partial^2 \mathrm{NLSE}}{\partial\vartheta^2} = 2a\sum_{i=1}^{n}\left(\ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)\right)^{2}\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta}\left[2a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} - i\right] \quad (45)$$

$$\frac{\partial \mathrm{NLSE}}{\partial\sigma} = 4a\theta\sum_{i=1}^{n}\frac{t_i e^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{\theta-1}}{\left(1+e^{-\frac{t_i}{\sigma}}\right)^{\theta+1}}\left[i - a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta}\right]. \quad (46)$$

$$\frac{\partial^2 \mathrm{NLSE}}{\partial\sigma^2} = 4a\theta\sum_{i=1}^{n} t_i e^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{\theta}\left[2ae^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{\theta-2} + \frac{\left(1+e^{-\frac{t_i}{\sigma}}\right)^{\theta}}{\sigma^2} + \left(2e^{-2\frac{t_i}{\sigma}}\theta - 2e^{-\frac{t_i}{\sigma}} + 1 - e^{-2\frac{t_i}{\sigma}}\right)\left(i - a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta}\right)\right] \quad (47)$$

$$\frac{\partial^2 \mathrm{NLSE}}{\partial\sigma\,\partial\theta} = \frac{\partial^2 \mathrm{NLSE}}{\partial\theta\,\partial\sigma}$$

$$= 4a\sum_{i=1}^{n}\left[\frac{-a\theta t_i e^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{2\theta-1}}{\left(1+e^{-\frac{t_i}{\sigma}}\right)^{2\theta+1}}\ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right) + \frac{t_i e^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{\theta-1}}{\left(1+e^{-\frac{t_i}{\sigma}}\right)^{\theta+1}}\left(i - a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta}\right)\left(1 + \theta\ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)\right)\right] \quad (48)$$

Also, with respect to attaining confidence intervals for the WNLS estimators, we define the Fisher information matrix as the matrix of negative second partial derivatives of weighted least squared sum function that shown in Equation (25):

$$I_{WNLSE}(\Theta) = \begin{bmatrix} -\frac{\partial^2 WNLSE}{\partial a^2} & -\frac{\partial^2 WNLSE}{\partial a\,\partial\theta} & -\frac{\partial^2 WNLSE}{\partial a\,\partial\sigma} \\ -\frac{\partial^2 WNLSE}{\partial\theta\,\partial a} & -\frac{\partial^2 WNLSE}{\partial\theta^2} & -\frac{\partial^2 WNLSE}{\partial\theta\,\partial\sigma} \\ -\frac{\partial^2 WNLSE}{\partial\sigma\,\partial a} & -\frac{\partial^2 WNLSE}{\partial\sigma\,\partial\theta} & -\frac{\partial^2 WNLSE}{\partial\sigma^2} \end{bmatrix} \tag{49}$$

such that:

$$\frac{\partial WNLSE}{\partial a} = 2\sum_{i=1}^{n} w_i \left[-i\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} + a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{2\theta} \right] \tag{50}$$

$$\frac{\partial^2 WNLSE}{\partial a^2} = 2\sum_{i=1}^{n} w_i \left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{2\theta} \tag{51}$$

$$\frac{\partial^2 WNLSE}{\partial a\,\partial\theta} = \frac{\partial^2 WNLSE}{\partial\theta\,\partial a}$$

$$= 2\sum_{i=1}^{n} w_i \left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} \ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right) \left[2a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} - i \right] \tag{52}$$

$$\frac{\partial^2 WNLSE}{\partial a\,\partial\sigma} = \frac{\partial^2 WNLSE}{\partial\sigma\,\partial a}$$

$$= 4\sum_{i=1}^{n} w_i \frac{-\theta\, t\, e^{-\frac{t_i}{\sigma}}(1-e^{-\frac{t_i}{\sigma}})^{\theta-1}}{(1+e^{-\frac{t_i}{\sigma}})^{\theta+1}} \left[2a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} - i \right] \tag{53}$$

$$\frac{\partial WNLSE}{\partial\theta} = 2a\sum_{i=1}^{n} w_i \ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right) \left[-i\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} + a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{2\theta} \right] \tag{54}$$

$$\frac{\partial^2 WNLSE}{\partial\theta^2} = 2a\sum_{i=1}^{n} w_i \left(\ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right) \right)^2 \left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} \left[2a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} - i \right] \tag{55}$$

$$\frac{\partial WNLSE}{\partial\sigma} = 4a\theta\sum_{i=1}^{n} w_i \frac{t_i e^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{\theta-1}}{\left(1+e^{-\frac{t_i}{\sigma}}\right)^{\theta+1}} \left[i - a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} \right] \tag{56}$$

$$\frac{\partial^2 WNLSE}{\partial\sigma^2} = 4a\theta\sum_{i=1}^{n} w_i t_i e^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{\theta} \left[2ae^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{\theta-2} + \frac{\left(1+e^{-\frac{t_i}{\sigma}}\right)^{\theta}}{\sigma^2} + \left(2e^{-2\frac{t_i}{\sigma}}\theta - 2e^{-\frac{t_i}{\sigma}} + 1 - e^{-2\frac{t_i}{\sigma}}\right)\left(i - a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta} \right) \right] \tag{57}$$

$$\frac{\partial^2 WNLSE}{\partial\sigma\,\partial\theta} = \frac{\partial^2 WNLSE}{\partial\theta\,\partial\sigma}$$

$$= 4a\sum_{i=1}^{n} w_i\left[\frac{-a\theta t_i e^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{2\theta-1}}{\left(1+e^{-\frac{t_i}{\sigma}}\right)^{2\theta+1}}\ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)+\frac{t_i e^{-\frac{t_i}{\sigma}}\left(1-e^{-\frac{t_i}{\sigma}}\right)^{\theta-1}}{\left(1+e^{-\frac{t_i}{\sigma}}\right)^{\theta+1}}\left(i-a\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)^{\theta}\right)\left(1+\theta\ln\left(\frac{1-e^{-\frac{t_i}{\sigma}}}{1+e^{-\frac{t_i}{\sigma}}}\right)\right)\right] \quad (58)$$

By using Equations (30) and (31), along with $I_{MLE}(\Theta), I_{NLSE}(\Theta)$ and $I_{WNLSE}(\Theta)$ matrices, the asymptotic $100\alpha\%$ two-sided confidence limits for the parameters a, θ and σ of the NHPP GHLD-I model are respectively given by:

$$\hat{a} \pm Z_\alpha[I_{aa}^{-1}(\widehat{\Theta})]^{1l2} \quad ,$$

$$\hat{\theta} \pm Z_\alpha\left[I_{\theta\theta}^{-1}(\widehat{\Theta})\right]^{1l2} \quad ,$$

$$\hat{\sigma} \pm Z_\alpha[I_{\sigma\sigma}^{-1}(\widehat{\Theta})]^{1l2} \quad ,$$

where $I_{aa}^{-1}(\widehat{\Theta}), I_{\theta\theta}^{-1}(\widehat{\Theta})$ and $I_{\sigma\sigma}^{-1}(\widehat{\Theta})$ are respectively the diagonal elements of the asymptotic variance and covariance of the estimates of the parameters a, θ and σ of the NHPP GHLD-I model.

3. Evaluation Criteria

The mean of square errors (MSE), the sum of square errors (SSE), and the variance criteria are used for the evaluation purpose in our application. These criteria illustrate the variation between the actual and predicted values. The lower the criteria value, the better model performance. The formulas of those criteria are shown in Table (1).

Table 1. Some evaluation criteria.

Criteria Name	Criteria Formula
MSE [Hwang and Pham (2009)]	$\sum_{i=1}^{n}\frac{(y_i-\widehat{m}(t_i))^2}{n-k}$,
SSE [Zhang et al. (2003)]	$\sum_{i=1}^{n}(y_i-\widehat{m}(t_i))^2$,
Variance [Huang and Kuo (2002)]	$\sqrt{(\frac{1}{n-1})\sum_{i=1}^{n}(y_i-\widehat{m}(t_i)-\mathrm{Bias})^2}$,

where

$\widehat{m}(t_i)$: The estimated cumulative number of faults at time t_i.

y_i: The total number of faults observed at time t_i.

Bias: The average of prediction error, it gotten by the relationship:

Bias=$\left|\frac{\sum_{i=1}^{n}[\widehat{m}(t_i)-y_i]}{n}\right|$.

n: The number of faults.

k: The number of model parameters.

4. Real Data Application

This section displays the real data sets that used in our application and the results obtained through estimating the unknown parameters a, θ and σ of the NHPP GHLM-I model by the three above-mentioned estimation methods as well as the obtained confidence intervals. Also, the results of the evaluation criteria are presented and discussed in this section.

4.1 Real Data Sets

In this section, four real data sets with sample sizes of (n = 34, 30, 136 and 41) are used to study the performance of the NHPP GHLM-I model. **Dataset 1:** represents the software failure data of Navel Tactical Data System (NTDS) [Prasad and Kantam (2010)]. **Dataset 2:** is from the procedure of a failures control chart for failure software process introduced by [Prasad et al. (2011)], the data consists of 30 software failures. **Dataset 3:** gives the time between 136 failures of a software product [Rao et al. (2011)]. **Dataset 4:** consists of 1197.945 time unit and 41 failures introduced by [Kim and Park (2010)]. The four data sets are summarized in Tables [(2)-(5)].

Table 2. Failure time data (DS1) [Prasad and Kantam (2010)].

Failure Number	Failure Time (days)	Failure Number	Failure Time (days)
1	9	18	3
2	12	19	6
3	11	20	1
4	4	21	11
5	7	22	33
6	2	23	7
7	5	24	91
8	8	25	2
9	5	26	1
10	7	27	78
11	1	28	47
12	6	29	12
13	1	30	9
14	9	31	135
15	4	32	258
16	1	33	16
17	3	34	35

Table 3. Failure time data (DS2) [Prasad et al. (2011)].

Failure Number	Failure Time (Hours)	Failure Number	Failure Time (Hours)
1	30.02	16	15.53
2	1.44	17	25.72
3	22.47	18	2.79
4	1.36	19	1.92
5	3.43	20	4.13
6	13.2	21	70.47
7	5.15	22	17.07
8	3.83	23	3.99
9	21	24	176.06
10	12.97	25	81.07
11	0.47	26	2.27
12	6.23	27	15.63
13	3.4	28	120.78
14	9.1	29	30.81
15	2.18	30	34.19

Table 4. Failure time data (DS3) [Rao et al. (2011)].

Failure Number	Failure Time (Hours)	Failure Number	Failure Time (Hours)	Failure Number	Failure Time (Hours)	Failure Number	Failure Time (Hours)
1	3	35	227	69	529	103	108
2	30	36	65	70	379	104	0
3	113	37	176	71	44	105	3110
4	81	38	58	72	129	106	1247
5	115	39	457	73	810	107	943
6	9	40	300	74	290	108	700
7	2	41	97	75	300	109	875
8	91	42	263	76	529	110	245
9	112	43	452	77	281	111	729
10	15	44	255	78	160	112	1897
11	138	45	197	79	828	113	447
12	50	46	193	80	1011	114	386
13	77	47	6	81	445	115	446
14	24	48	79	82	296	116	122
15	108	49	816	83	1755	117	990
16	88	50	1351	84	1064	118	948
17	670	51	148	85	1783	119	1082
18	120	52	21	86	860	120	22
19	26	53	233	87	983	121	75
20	114	54	134	88	707	122	482
21	325	55	357	89	33	123	5509
22	55	56	193	90	868	124	100
23	242	57	236	91	724	125	10
24	68	58	31	92	2323	126	1071
25	422	59	369	93	2930	127	371
26	180	60	748	94	1461	128	790
27	10	61	0	95	843	129	6150
28	1146	62	232	96	12	130	3321
29	600	63	330	97	261	131	1045
30	15	64	365	98	1800	132	648
31	36	65	1222	99	865	133	5485
32	4	66	543	100	1435	134	1160
33	0	67	10	101	30	135	1864
34	8	68	16	102	143	136	4116

Table 5. Failure time data (DS4) [Kim and Park (2010)].

Failure Number	Failure Time (Hours)	Failure Number	Failure Time (Hours)
1	5.649	22	141.712
2	8.920	23	164.212
3	20.29	24	342.850
4	29.955	25	356.144
5	34.715	26	399.144
6	75.950	27	446.494
7	78.171	28	476.644
8	78.625	29	497.144
9	83.022	30	497.661
10	89.114	31	591.161
11	89.804	32	665.644
12	92.860	33	686.444
13	93.660	34	765.944
14	110.655	35	772.977
15	111.988	36	774.944
16	122.545	37	791.561
17	127.045	38	815.978
18	128.712	39	837.145
19	128.99	40	861.945
20	131.768	41	1197.945
21	131.829		

4.2 Application Results

The results of the estimation process of initial faults a, scale parameter σ and shape parameter θ of the NHPP GHLM-I model for the ML, NLS and WNLS estimation methods along with their SSE, MSE and variance using the four data sets are reported in Tables [(6)-(8)]. While the corresponding confidence intervals of the model parameters are displayed in Table 9. Tables [(10)-(13)] present the prediction results for the last 16 failures based on the three estimation methods for the four real data sets.

Table 6. Estimated parameters values using MLE method based on the NHPP GHLD-I model.

Data sets	NHPP GHLD-I					
	Parameters estimates			Model selection criteria		
	$\hat{a}_{MLE}$	$\hat{\sigma}_{MLE}$	$\hat{\theta}_{MLE}$	SSE	MSE	Variance
Ds1	34.6813	194.521	0.7763	428.8038	13.8324	5.0411
Ds2	31.6754	205.4441	0.9855	256.8063	9.5113	4.0693
Ds3	147.2553	33985.03	0.5738	1733.453	13.0335	4.4975
Ds4	**41.646**	**3414.608**	**0.5749**	**108.3965**	**2.8525**	**1.6506**

Bold values indicate the best fit model

Table 7. Estimated parameters values using NLSE method based on the NHPP GHLD-I model.

Data sets	NHPP GHLD-I					
	Parameters estimates			Model selection criteria		
	$\hat{a}_{NLSE}$	$\hat{\sigma}_{NLSE}$	$\hat{\theta}_{NLSE}$	SSE	MSE	Variance
Ds1	29.6805	59.5021	1.5316	104.7032	3.3775	1.7822
Ds2	27.0976	91.6718	1.4892	41.0961	1.5221	1.1921
Ds3	145.5919	33806.48	0.5404	1168.952	8.7891	2.9599
Ds4	**40.6382**	**4142.043**	**0.4677**	**32.0543**	**0.8435**	**0.8996**

Bold values indicate the best fit model

Table 8. Estimated parameters values using WNLSE method based on the NHPP GHLD-I model.

	NHPP GHLD-I							
Parameters estimates	Data set				Model selection criteria			
	Ds1	Ds2	Ds3	Ds4		SSE	MSE	Variance
$\hat{a}_{WLSE}(w_{i(1)})$	29.6099	26.8825	136.8682	**39.4737**	Ds1	105.0699	3.3894	1.7844
$\hat{\sigma}_{WLSE}(w_{i(1)})$	57.0869	85.2503	26899.84	**3526.323**	Ds2	42.2495	1.5648	1.207
$\hat{\theta}_{WLSE}(w_{i(1)})$	1.6252	1.6329	0.5797	**0.4959**	Ds3	1327.275	9.9795	3.1355
					Ds4	**35.0379**	**0.9221**	**0.9359**
$\hat{a}_{WLSE}(w_{i(2)})$	29.5454	26.9861	140.6833	**39.9391**	Ds1	105.3328	3.3978	1.7866
$\hat{\sigma}_{WLSE}(w_{i(2)})$	56.0506	87.847	29766.41	**3755.842**	Ds2	41.5145	1.5376	1.1965
$\hat{\theta}_{WLSE}(w_{i(2)})$	1.6602	1.5752	0.5625	**0.4854**	Ds3	1217.667	9.1554	3.0041
					Ds4	**33.2009**	**0.8737**	**0.9112**
$\hat{a}_{WLSE}(w_{i(3)})$	29.6655	27.0896	145.5158	**40.5987**	Ds1	104.7107	3.3778	1.7821
$\hat{\sigma}_{WLSE}(w_{i(3)})$	59.1144	91.3745	33741.85	**4119.637**	Ds2	41.0987	1.5222	1.1919
$\hat{\theta}_{WLSE}(w_{i(3)})$	1.545	1.4957	0.5408	**0.4687**	Ds3	1168.963	8.7892	2.9594
					Ds4	**32.0579**	**0.8436**	**0.8991**

Bold values indicate the best fit model

Table 9. 95% confidence intervals using different methods of estimation based on the NHPP GHLD-I model.

Data set	Parameters estimates		NHPP GHLD-I				
			MLE	LSE	WNLSE $(w_{i(1)})$	WNLSE $(w_{i(2)})$	WNLSE $(w_{i(3)})$
DS1	$\hat{a}$	Lower	22.9336	28.3388	28.1842	28.2230	28.3285
		Upper	46.4289	31.2102	31.0932	30.9436	31.1770
	$\hat{\sigma}$	Lower	88.0258	45.3288	48.3797	45.8576	45.2914
		Upper	301.0162	83.4627	68.1005	70.2422	82.1209
	$\hat{\theta}$	Lower	0.4276	1.0631	1.3586	1.2910	1.0812
		Upper	1.1249	2.2698	1.9862	2.1883	2.2706
DS2	$\hat{a}$	Lower	19.8213	26.0673	25.6747	25.8990	26.0569
		Upper	43.5295	28.1835	28.1328	28.1163	28.1768
	$\hat{\sigma}$	Lower	69.9059	77.3018	75.0153	76.1114	77.2104
		Upper	340.9823	110.2253	97.3949	102.2022	109.5834
	$\hat{\theta}$	Lower	0.4703	1.2058	1.4319	1.3377	1.2153
		Upper	1.5007	1.8701	1.8795	1.8781	1.8713
DS3	$\hat{\theta}$	Lower	122.2127	140.4260	132.5460	136.1155	140.3597
		Upper	172.2979	152.1585	142.0168	146.2646	152.0659
	$\hat{\sigma}$	Lower	30207.0606	30078.4896	24368.7350	26780.1560	30025.8677
		Upper	37762.9991	38710.7828	30020.7227	33533.2323	38627.0582
	$\hat{\theta}$	Lower	0.4757	0.5227	0.5645	0.5458	0.5230
		Upper	0.6719	0.5592	0.5956	0.5800	0.5595
DS4	$\hat{a}$	Lower	28.8535	39.0661	37.9941	38.4557	39.0323
		Upper	54.4385	42.7751	41.2791	41.8290	42.7206
	$\hat{\sigma}$	Lower	2247.3104	3462.3114	3036.9827	3191.8628	3446.2786
		Upper	4581.9050	5140.5624	4178.5324	4536.2182	5105.5515
	$\hat{\theta}$	Lower	0.3840	0.4388	0.4744	0.4597	0.4399
		Upper	0.7658	0.4995	0.5191	0.5135	0.5003

Table 10. Predictions based on three estimation methods for DS1.

Actual number of failure	Prediction results by (MLE)	Prediction results by (NLSE)	Prediction results by (WNLSE) $w_{i(1)}$	Prediction results by (WNLSE) $w_{i(2)}$	Prediction results by (WNLSE) $w_{i(3)}$
19	12.2301	**17.3142**	17.4220	17.5036	17.3336
20	12.3170	**17.4736**	17.5860	17.6690	17.4936
21	13.2542	**19.1341**	19.2901	19.3854	19.1603
22	15.8709	**23.0926**	23.3051	23.4040	23.1261
23	16.3908	**23.7474**	23.9600	24.0549	23.7807
24	22.1551	**28.2826**	28.3653	28.3729	28.2935
25	22.2625	**28.3278**	28.4073	28.4132	28.3381
26	22.3160	**28.3498**	28.4278	28.4329	28.3599
27	26.2777	**29.3667**	29.3483	29.3062	29.3607
28	27.9152	**29.5376**	29.4948	29.4418	29.5275
29	28.2843	**29.5636**	29.5166	29.4617	29.5528
30	28.5491	**29.5800**	29.5302	29.4741	29.5687
31	31.4808	**29.6700**	29.6024	29.5390	29.6557
32	33.8023	**29.6803**	29.6099	29.5453	29.6654
33	33.8709	**29.6803**	29.6099	29.5454	29.6654
34	34.0031	**29.6804**	29.6099	29.5454	29.6655

Bold values indicate the predictions based on the best estimation method

Table 11. Predictions based on three estimation methods for DS2.

Actual number of failure	Prediction results by (MLE)	Prediction results by (NLSE)	Prediction results by (WNLSE) $w_{i(1)}$	Prediction results by (WNLSE) $w_{i(2)}$	Prediction results by (WNLSE) $w_{i(3)}$
15	10.3039	**13.6514**	13.7592	13.6948	13.6534
16	11.3665	**15.2355**	15.4197	15.3259	15.2411
17	13.0608	**17.5810**	17.8535	17.7282	17.5912
18	13.2395	**17.8136**	18.0927	17.9652	17.8242
19	13.3618	**17.9712**	18.2545	18.1257	17.9820
20	13.6233	**18.3032**	18.5947	18.4634	18.3145
21	17.7185	**22.6093**	22.8913	22.7785	22.6222
22	18.6029	**23.3171**	23.5695	23.4717	23.3290
23	18.8036	**23.4670**	23.7118	23.6177	23.4787
24	25.4915	**26.5328**	26.4583	26.5058	26.5316
25	27.3710	**26.8629**	26.7178	26.7942	26.8584
26	27.4151	**26.8686**	26.7221	26.7991	26.8641
27	27.7077	**26.9044**	26.7489	26.8295	26.8994
28	29.4083	**27.0457**	26.8500	26.9464	27.0388
29	29.7143	**27.0605**	26.8599	26.9581	27.0533
30	30.0070	**27.0720**	26.8673	26.9672	27.0647

Bold values indicate the predictions based on the best estimation method

Table 12. Predictions based on three estimation methods for DS3.

Actual number of failure	Prediction results by (MLE)	Prediction results by (NLSE)	Prediction results by (WNLSE) $w_{i(1)}$	Prediction results by (WNLSE) $w_{i(2)}$	Prediction results by (WNLSE) $w_{i(3)}$
121	118.1840	**118.5723**	118.7130	118.7483	118.5757
122	118.5594	**118.9256**	119.0164	119.0769	118.9286
123	122.5425	**122.6676**	122.1595	122.5177	122.6667
124	122.6098	**122.7308**	122.2114	122.5751	122.7297
125	122.6165	**122.7371**	122.2166	122.5809	122.7360
126	123.3264	**123.4029**	122.7612	123.1847	123.4010
127	123.5678	**123.6292**	122.9453	123.3894	123.6270
128	124.0743	**124.1040**	123.3297	123.8177	124.1011
129	127.6838	**127.4821**	125.9934	126.8238	127.4742
130	129.4056	**129.0904**	127.2142	128.2267	129.0795
131	129.9171	**129.5677**	127.5701	128.6393	129.5559
132	130.2272	**129.8570**	127.7844	128.8885	129.8446
133	132.6490	**132.1138**	129.4163	130.8077	132.0966
134	133.1172	**132.5497**	129.7229	131.1731	132.5315
135	133.8397	**133.2218**	130.1900	131.7329	133.2019
136	135.3117	**134.5899**	131.1182	132.8580	134.5664

Bold values indicate the predictions based on the best estimation method

Table 13. Predictions based on three estimation methods for DS4.

Actual number of failure	Prediction results by (MLE)	Prediction results by (NLSE)	Prediction results by (WNLSE) $w_{i(1)}$	Prediction results by (WNLSE) $w_{i(2)}$	Prediction results by (WNLSE) $w_{i(3)}$
26	25.7888	**25.3933**	25.7477	25.5867	25.4024
27	27.5044	**26.8313**	27.2331	27.0567	26.8423
28	29.1435	**28.2124**	28.6450	28.4604	28.2248
29	30.6648	**29.5056**	29.9503	29.7654	29.5188
30	32.0153	**30.6676**	31.1062	30.9284	30.6812
31	33.4199	**31.8962**	32.3070	32.1457	31.9094
32	34.7730	**33.1061**	33.4636	33.3294	33.1181
33	35.9485	**34.1857**	34.4697	34.3703	34.1960
34	37.0356	**35.2162**	35.4026	35.3473	35.2238
35	37.9309	**36.0957**	36.1738	36.1659	36.1003
36	38.6597	**36.8392**	36.8046	36.8448	36.8405
37	39.2607	**37.4771**	37.3276	37.4157	37.4749
38	39.7566	**38.0258**	37.7618	37.8966	38.0199
39	40.1603	**38.4921**	38.1176	38.2968	38.4826
40	40.4872	**38.8871**	38.4078	38.6282	38.8741
41	40.8268	**39.3201**	38.7119	38.9821	39.3027

Bold values indicate the predictions based on the best estimation method

4.3 Discussion of Results

The purpose of our application is to evaluate the performance of the NHPP GHLM-I model through three evaluation criteria and based on different real data sets. Three methods of estimation are used to estimate the initial faults a, scale parameter σ and shape parameter θ of the NHPP GHLM-I model. The 95% confidence intervals around all effects for the four selected real data sets are constructed. By comparing the evaluation criteria values of the four real data sets using the ML, NLS and WNLS estimation methods, the following observations can be made:

- The NHPP GHLM-I model shows good predictive ability according to our selected three criteria and using the four selected real data sets, but the model fits the fourth real data set with size 41 more than the rest of the studied real data sets since all the evaluation criteria have the lowest values for the fourth real data set compared to the other three.
- The ML estimation method shows the worst performance among the three studied estimation methods; however, its evaluation criteria values are lower for the fourth data set than the rest of the studied data sets.
- According to our studied data sets, the NHPP GHLD-I model shows the best performance when using the NLS estimation method with all the four studied real data sets. It has the lowest values for all the evaluation criteria among the others selected estimation methods.
- The evaluation criteria values of the WNLS estimation method at weighting function ($w_{i(3)}$) are approximately the same as the evaluation criteria values of the NLS estimation method. In addition, for three comparison cases it has lower criteria values than the NLS estimation method this indicates that the use of this weighting function gives better prediction results than the other considered weighting functions.
- According to confidence intervals of the model parameters the following points are concluded: For DS1, the estimator $\hat{a}$ has the shortest expected length when using the WNLS estimation method at weighting function $\left(w_{i(3)}\right)$ while the shortest expected length for $\hat{\theta}$ and $\hat{\sigma}$ estimators is obtained by using the WNLS estimation method at weighting function $\left(w_{i(1)}\right)$. For DS2, the estimator $\hat{a}$ has the shortest expected length when using the NLS estimation method while the shortest expected length for $\hat{\theta}$ and $\hat{\sigma}$ estimators is obtained by using the WNLS estimation method at weighting function $\left(w_{i(1)}\right)$. For DS3 and DS4, all parameters estimators have the shortest expected length when using the WNLS estimation method at weighting function $\left(w_{i(1)}\right)$. Based on the interval estimation the WNLS estimation method gives the best performance while the ML estimation method gives the worst performance. The first- and second-best weighting function are respectively $w_{i(1)}$ and $w_{i(3)}$.

5. Conclusion

In this paper, a software reliability growth model that belong to the NHPP type of modeling and based on GHLD-I distribution have been constructed. The estimation process of the unknown parameters of the proposed NHPP GHLD-I model have been conducted by using the ML, NLS and WNLS estimation methods. In addition, confidence intervals of the model parameters which is important to the software reliability evaluation have been obtained. An application based on the NHPP GHLD-I model and using four real data sets have been conducted to measure the performance of our proposed model based on three evaluation criteria. Our numerical study illustrates the flexibility of the NHPP GHLD-I model and presents its positive contribution to the field of software reliability modeling.

References

Barraza, N. R. (2010). Compound and Non-Homogeneous Software Reliability Models. Annals of the 11th Argentine Symposium on Software Engineering, ASSE Buenos Aires, Argentina.

Goel, A. L., & Okumoto, K. (1979). Time-dependent fault-detection rate model for software and other performance measures. *IEEE transactions on Reliability, 28*, 206–211. https://doi.org/10.1109/TR.1979.5220566

Hee-Cheul, K., & Hyoung-Keun, P. (2010). The Comparative Study of Software Optimal Release Time Based on Burr Distribution. *International Journal of Advancements in computing Technology, 2*(3), 120-129.

Hong, G. Y., Xie, M., Zhao, M., & Wohlin, C. (1997). Interval Estimation in Software Reliability Analysis. Proceedings4th International Applied Statistics in Industry Conference, 105-112, Kansas City, Missouri, USA.

Huang, C. Y., & Kuo, S. Y. (2002). Analysis of incorporating logistic testing effort function into software reliability modeling. *IEEE Trans. on Reliability, 51*(3), 261-270. https://doi.org/10.1109/TR.2002.801847

Hwang, S., & Pham, H. (2009). Quasi-renewal time-delay fault-removal consideration in software reliability modelling. *IEEE Trans. on systems, man and cybernetics-Part A:Systems and humans, 39*(1), 200-209. https://doi.org/10.1109/TSMCA.2008.2007982

Ishii, H., Dohi, T., & Okamura, H. (2012). Software reliability prediction based on Least Squares estimation. *Quality Technology and Quantitative Management Journal, 9*(3), 243-264. https://doi.org/10.1080/16843703.2012.11673290

Kantam, R. R. L., Ramakrishna, V., & Ravikumar, M. S. (2013). Estimation and Testing in Type I Generalized Half-Logistic Distribution. *Journal of Modern Applied Statistical Methods, 12*(1), 198-206. https://doi.org/10.22237/jmasm/1367382060

Lai, R., & Garg, M. (2012). A Detailed Study of NHPP Software Reliability Models. *Journal of Software, 7*(6), 1296-1306. https://doi.org/10.4304/jsw.7.6.1296-1306

Lyu, M. R. (2002). Software Reliability Theory. Encyclopedia of Software Engineering, Wiley, 1611-1630. https://doi.org/10.1002/0471028959.sof329

Prasad, R. S., Rao, K. H., & Kantha, R. L. (2011). Software Reliability Measuring Using Modified Maximum Likelihood Estimation and SPC. *International Journal of Computer Applications, 21*(7), 1-5. https://doi.org/10.5120/2527-3440

Prasad, R., & Kantam, R. (2010). Half Logistic Software Reliability Growth Model. *International Journal of Advanced Research in Computer Science, 1*(2), 31-36.

Rao, K. H., Prasad, R. S., & Kantham, R. L. (2011). Assessing Software Reliability Using SPC – An Order Statistics Approach. *International Journal of Computer Science, Engineering and Applications (IJCSEA), 1*(4), 121-131. https://doi.org/10.5121/ijcsea.2011.1411

Sun, R., Peng, W. W., Huang, H. Z., Ling, D., & Yang, J. (2012). Improved reliability data curve fitting method by considering samples distinction. *Eksploatacja i Niezawodnosc – Maintenance and Reliability, 14*(1), 62–71.

Teng, X., & Pham, H. (2004). A software cost model for quantifying the gain with considerations of random field environments. *IEEE Trans. Comput., 53*(3), 380–384. https://doi.org/10.1109/TC.2004.1261844

Yamada, S., & Ohba, M. (1983). Osaki, S.: S-shaped reliability growth modeling for software fault detection. IEEE Trans. *Reliab., 12,* 475–484. https://doi.org/10.1109/TR.1983.5221735

Zhang, X., Teng, X., & Pham, H. (2003). Considering fault removal efficiency in software reliability assessment. In *IEEE Transactions on Systems, Man, and Cy-bernetics - Part A: Systems and Humans, Jan, 33*(1), 114-120. https://doi.org/10.1109/TSMCA.2003.812597

Authors Information

Lutfiah Ismail Al turk is currently working as Associate Professor of Mathematical Statistics in Statistics Department at Faculty of Sciences, King AbdulAziz University, Jeddah, Kingdom of Saudi Arabia. Lutfiah Ismail Al turk obtained her B.Sc degree in Statistics and Computer Science from Faculty of Sciences, King AbdulAziz University in 1993 and M.Sc (Mathematical statistics) degree from Statistics Department, Faculty of Sciences, King AbdulAziz University in 1999. She received her Ph.D in Mathematical Statistics from university of Surrey, UK in 2007. Her current research interests include Software reliability modeling and Statistical Machine Learning.

Email: lturk@kau.edu.sa; URL: http://lturk.kau.edu.sa

Address: P.O. Box 42713 Jeddah 21551. Kingdom of Saudi Arabia.

Wejdan Saleem Al ahmadi is presently Teaching Assistant in Mathematics Department at Faculty of Sciences, Taibah University, Saudi Arabia. Wejdan Saleem Al ahmadi obtained her B.Sc degree in Mathematics from Faculty of Sciences, Taibah University in 2009.

Copyrights