mathematics

MDPI

*Article*

# Nonhomogeneous Poisson Process Software Reliability Growth Model with Dependent Failures and an Exponentially Decaying Fault Detection Rate

**Kwang Yoon Song [1,2], Onon-Ujin Otgonbayar [1] and In Hong Chang [1,*]**

[1] Department of Computer Science and Statistics, Chosun University, Dong-gu, Gwangju 61452, Republic of Korea; csssig@chosun.ac.kr (K.Y.S.); ujin7709@chosun.kr (O.-U.O.)
[2] Institute of Well-Aging Medicare & CSU G-LAMP Project Group, Chosun University, Dong-gu, Gwangju 61452, Republic of Korea
* Correspondence: ihchang@chosun.ac.kr

**Abstract**

Effectively modeling software failure behavior is crucial for reliability assessment and planning of releases. However, many current software reliability growth models assume that failures are independent and fault detection mechanisms are simplified. However, these assumptions may not accurately represent real-world testing environments. This study introduces a novel Nonhomogeneous Poisson Process (NHPP)-based Software Reliability Growth Model (SRGM) that includes dependent failure behavior and exponentially decaying fault detection rates to better reflect the software debugging process. The proposed model was validated using real failure datasets and compared with 17 existing models. The performance of the model was assessed using various goodness-of-fit criteria, such as errors, prediction accuracy, and metrics based on information theory. To provide a more thorough evaluation, a multi-criteria decision-making approach was used to rank the competing models based on their overall performance. Furthermore, a one-at-a-time sensitivity analysis was conducted to examine how the initial values of the parameters affected the model's behavior. These findings indicate that the sensitivity of the model to this parameter varies depending on the dataset used. The results indicate that the proposed model achieved superior performance across multiple evaluation criteria and consistently obtained the best overall ranking under the integrated multi-criteria framework. In Dataset 1, the proposed model achieved the best performance in most goodness-of-fit criteria, whereas in Dataset 2 it produced the best results across all twelve evaluation criteria. The results show that the proposed model offers improved or competitive performance compared to existing models and provides greater flexibility in capturing complex failure processes within software systems.

**Keywords:** dependent failures; exponentially decaying fault detection rate function; multi-criteria decision-making approach; sensitivity analysis

**MSC:** 68N05; 68N30; 90B25



Academic Editors: Dahai Liu, Yongxin Liu, Hongyun Chen and Lu Gao

Received: 3 May 2026
Revised: 3 June 2026
Accepted: 12 June 2026
Published: 14 June 2026



## 1. Introduction

Software systems have become progressively more intricate and are extensively used in large-scale, industrial, and safety-critical environments. As software complexity has increased, accurate reliability evaluation and failure prediction have become crucial for

aiding testing methodologies, release decisions, and maintenance planning. To address this issue, Software Reliability Growth Models (SRGMs), particularly those based on a Nonhomogeneous Poisson Process (NHPP), have been thoroughly examined owing to their robust mathematical underpinnings and analytical manageability [1,2].

In modern software systems, reliability failures often arise in highly interconnected environments where software components continuously interact through shared libraries, cloud services, and distributed architecture. In such systems, assuming that failures occur independently may lead to biased reliability prediction and inaccurate release decisions. For example, a defect in a shared software component may simultaneously affect multiple functions. Also, the complex interactions among software components can create dependency relationships among failures, where one failure may influence the occurrence of following failures. At the same time, as software testing progresses, the remaining faults often become increasingly difficult to detect, causing changes in the fault detection behavior over time. Therefore, motivated by these practical characteristics, simultaneous consideration of failure dependency and time-dependent fault detection becomes important for more accurately representing software reliability growth processes.

Early and classical NHPP-based SRGMs, including the Goel and Okumoto model and the S-shaped model, typically assume that software failures occur independently. These models also operate under the assumption of perfect debugging, implying that the identified faults are completely removed without influencing subsequent failures [1,3]. However, these assumptions frequently do not hold true in modern software systems, where failures tend to be interrelated, rather than occurring as isolated incidents. A review by Haque and Ahmad further emphasized that unrealistic assumptions, uncertainty factors, and the lack of model flexibility remain major challenges in the application of traditional SRGMs [4].

In practical software development environments, failures often demonstrate dependent failure behaviors. Initially, multiple failures can stem from shared code components or common functional modules, leading to repeated failures until the root cause is eliminated [5]. Furthermore, fault propagation between interacting modules can result in cascading failures, in which an initial failure increases the chances of subsequent failures [6]. In addition, imperfect debugging, as studied by Pham et al. [7], suggests that detected faults may not be fully eliminated or could even create new faults, thus fostering dependency among failures. Finally, the intricate interactions among modules in large-scale systems further intensify failure dependency. Although failure propagation concepts have been extensively studied in physical and network systems, software failure dependency possesses distinct characteristics. Unlike physical systems, where failures may arise from random physical processes, software failures are generally associated with deterministic logical interactions among modules. Similar cascading behaviors have also been investigated in cyber-physical systems, where dependency structures and network interactions influence failure propagation mechanisms [8]. These studies provide useful methodological insight into dependency analysis and further highlight the importance of considering dependency effects in software reliability growth modeling.

To ease the assumption of independence, several studies have proposed modifications to classical SRGMs. Pham and Zhang [9,10] presented generalized NHPP-based SRGMs to address the variability in operational environments, and Pham et al. [7] specifically integrated the imperfect debugging processes. Although dependent failures, imperfect debugging, and delayed fault detection are related concepts, they describe different aspects of software reliability growth. Imperfect debugging primarily considers situations where debugging activities may fail to eliminate faults completely or may introduce additional faults into the system. Delayed detection models focus on time lags between fault occur-

rence and fault identification during software testing. In contrast, dependent failure models assume that software failures interact due to shared components and fault propagation effects. Therefore, the proposed model primarily addresses failure interaction characteristics rather than explicitly modeling fault introduction or delayed detection mechanisms.

In recent years, the importance of time-dependent fault detection rates has been highlighted by the increase in open-source software (OSS). OSS projects are distinct from closed industrial projects, revealing the limitations of SRGMs that assume a constant fault-detection rate. Consequently, sophisticated NHPP-based SRGMs that integrate time-dependent rates, coverage growth, imperfect debugging, and delayed fault detection have been proposed. For example, Wang et al. [11] introduced a Weibull-Weibull model specifically tailored for OSS big data projects, demonstrating significant enhancements in predictive accuracy compared with traditional models. Similarly, Wang et al. [12] developed multi-release SRGMs for OSS systems by considering the ongoing development of OSS through successive iterations.

In addition to the challenges associated with modeling, the rapid increase in the number of SRGMs has rendered model comparison and selection a significant concern in the analysis of software reliability. Numerous studies depend on a single measure of the goodness-of-fit accuracy, which can result in inconsistent conclusions when the models demonstrate varying strengths across different performance criteria. To address this limitation, multi-criteria decision-making methods offer a structured approach for incorporating multiple criteria into a cohesive assessment. Traditional multi-criteria decision-making methods have been extensively examined in the field of decision science [13,14]; however, their use in the evaluation of SRGMs remains relatively scarce.

Based on the preceding discussion, this study addresses the following research questions:

RQ1: How can dependent software failure behavior be incorporated into NHPP-based SRGMs to better represent practical software testing environments?

RQ2: Can an exponentially decaying fault detection mechanism effectively describe time-dependent fault detection behavior while maintaining model flexibility?

RQ3: To what extent can the proposed dependent failure NHPP SRGM improve fitting and predictive performance compared with traditional independent failure models?

Motivated by these findings, this study introduces a novel NHPP-based SRGM that specifically considers the dependent failure behavior. To facilitate a reliable comparison, an integrated multi-criteria evaluation framework was used, which included the Multi-Criteria Decision Method (MCDM), Multi-Criteria Decision Method using Ranking (MCDMR), Multi-Criteria Decision Method using Maximum (MCDMM), and Rank Euclidean Distance (RED). This framework allows the simultaneous assessment of various criteria related to accuracy, stability, and robustness, which offers a well-rounded foundation for model selection.

Moreover, to illustrate the general applicability of the proposed model and its evaluation, validation was performed using both a closed-type industrial software failure dataset and an OSS failure dataset. The inclusion of these two datasets guarantees that the proposed model is relevant across various software development contexts and is not confined to a particular data type.

The main contributions of this study are summarized as follows:

- A novel NHPP-based SRGM is proposed to capture the dependent failure behavior under an exponentially decaying fault-detection rate.
- A non-zero initial condition $m(0) = m_0$ was incorporated to reflect practical testing situations in which fault detection may have started before the observation period.
- An integrated multi-criteria decision-making framework was employed to provide a robust and balanced evaluation of competing SRGMs.

- Extensive analyses were conducted on both industrial and OSS failure datasets to demonstrate the effectiveness and applicability of the proposed approach.

The remainder of this paper is organized as follows. Section 2 provides a comprehensive overview of SRGMs based on the NHPP, including an in-depth derivation of the proposed model and existing SRGMs that are commonly utilized. Section 3 outlines the study criteria. Section 4 introduces the two datasets utilized in this study and presents the analysis results of the parameter estimation for the respective models, a comparison of the goodness-of-fit test outcomes, and the sensitivity analysis results for the parameter $m_0$. Finally, Section 5 concludes the study and explores potential directions for future research.

## 2. Novel SRGM

### *2.1. SRGM*

A counting process $\{N(t), t \geq 0\}$ counts the number of failures that occur up to time $t$ [5]. $N(t)$ is said to be a Poisson process with a failure intensity $\lambda > 0$ if:

1. The failure process, $N(t)$, has stationary independent increments.
2. The number of failures in any time interval of length $s$ has a Poisson distribution with mean $\lambda s$, that is:

$$P\{N(t+s) - N(t) = n\} = \frac{e^{-\lambda s}[\lambda s]^n}{n!}, n = 1,\ 2,\ 3, \cdots$$

3. The initial condition is $N(0) = 0$.

This model is called a Homogeneous Poisson Process (HPP) indicating that the constant failure rate $\lambda$ does not depend on the time $t$ but on the length of time interval s. To capture the changing aspects of software testing and fault removal, the constant failure rate $\lambda$ in the HPP is extended to a time-dependent function $\lambda(t)$, leading to the NHPP.

For the NHPP, the probability of exactly $n$ failures occurring during the time interval $(0, t)$ is given by Equation (1):

$$P\{N(t) = n\} = \frac{[m(t)]^n}{n!} e^{-m(t)}, \quad n = 0,\ 1,\ 2, \ldots \tag{1}$$

where $N(t)$ denotes a counting process that follows the Poisson distribution with a time-dependent function $m(t)$. The function $m(t)$ is a mean value function which is the integral of $\lambda(t)$ over the interval $[0, t]$, as given in Equation (2). The function $\lambda(t)$ characterizes the instantaneous failure occurrence rate at time t.

$$m(t) = E[N(t)] = \int_0^t \lambda(s) ds. \tag{2}$$

In general, the $m(t)$ value of an NHPP-based SRGM is obtained by solving the following differential equation [1]:

$$\frac{dm(t)}{dt} = b(t)[a(t) - m(t)]. \tag{3}$$

where $a(t)$ denotes the expected total number of faults present in the software system at time $t$, including both the initial faults and newly introduced faults that remain until detection. The function $b(t)$ represents the fault-detection rate per remaining fault.

The generalized solution of Equation (3) with the initial conditions $m(t_0) = m_0$ and $a(t) = a$ is given by:

$$m(t) = e^{-B(t)} \left[ m_0 + a \int_{t_0}^t b(\tau) e^{B(\tau)} d\tau \right],$$

where $B(t) = \int_{t_0}^{t} b(\tau)d\tau$ and $t_0$ denotes the time at which the testing process begins.

### *2.2. Proposed Model*

Most existing SRGMs assume that the failures occur independently. However, in practical software systems, failures may exhibit dependent behavior. In these situations, one failure can affect the chances of future failures. To address this issue, this study assumes dependent failures along with an exponentially decaying fault-detection rate. Regarding the exponentially decaying fault-detection rate function, unlike the growth-shaped function, easily detectable faults are found early, and the detection process slows as testing continues.

The mean value function of the proposed model is governed by the following differential equation [15]:

$$\frac{dm(t)}{dt} = b(t)[a(t) - m(t)]m(t). \tag{4}$$

In Equation (4), to reflect the dependent failure behavior, an additional multiplicative term $m(t)$, representing the accumulated failures, is introduced. In Equation (3), failures occur at a rate proportional only to the number of remaining faults which is $[a(t) - m(t)]$, implying that failures occur independently. However, in Equation (4), the failure rate is now proportional to both the remaining faults and accumulated failures, expressed as $[a(t) - m(t)]m(t)$. The inclusion of $m(t)$ allows previously occurring failures to influence subsequent failure behavior through cumulative dependency effects. Practically, this formulation can represent failure interactions arising from shared components, interacting modules, or fault propagation effects. Therefore, dependency is introduced through cumulative failure interactions rather than explicit correlation structures or conditional intensity functions.

The generalized solution of Equation (4) is as follows. First, the variables are separated as follows:

$$\frac{dm(t)}{[a(t) - m(t)]m(t)} = b(t)d.$$

When the fault content is assumed to be constant as $a(t) = a$, which is commonly adopted in software reliability growth modeling, the left-hand side can be decomposed using partial fractions:

$$\frac{1}{[a - m(t)]m(t)}dm(t) = \frac{1}{a}\left[\frac{1}{m(t)} + \frac{1}{a - m(t)}\right]dm(t)$$

Integrating both sides yields the following:

$$\frac{1}{a}(ln|m(t)| - ln|a - m(t)|) = \int_{t_0}^{t} b(\tau)d\tau + C,$$

where $C$ is integration constant.

Letting $B(t) = \int_{t_0}^{t} b(\tau)d\tau$ and applying the initial conditions $t_0 = 0$, $m(0) = m_0$, and $a(t) = a$, the generalized mean value function is obtained in closed form by Equation (5) as follows:

$$m(t) = \frac{a}{1 + \left(\frac{a - m_0}{m_0}\right)e^{-aB(t)}} \tag{5}$$

In this study, the fault-detection rate is represented as a time-dependent function as follows:

$$b(t) = \alpha + be^{-ct}$$

where $\alpha > 0$ denotes the baseline fault detection rate, $b > 0$ represents the initial excess detection rate at the beginning of testing, and $c > 0$ is the decay parameter that determines

how quickly the fault-detection rate declines over time. During the initial phase of testing, the exponential term is prominent, leading to a higher fault-detection rate because of active testing efforts and a large number of detectable faults. As testing continues, the exponential part diminishes, and $b(t)$ slowly approaches the baseline level $\alpha$, indicating the stabilization of the testing process and a decrease in the number of remaining detectable faults. This functional form effectively illustrates the learning effect and decreasing failure-detection behavior that are often observed in real-world software testing scenarios.

Compared with other commonly used fault detection mechanisms, the exponentially decaying function provides a relatively simple and interpretable representation of changing detection behavior. Logistic and S-shaped functions are often used to describe learning effects and delayed growth patterns during testing [3,15]. In contrast, the considered fault detection function assumes that detection efficiency gradually decreases over time while maintaining mathematical simplicity and analytical tractability.

Figure 1 shows how the time-dependent fault-detection rate $b(t)$ behaves for different decay parameter $c$ values with the other parameters fixed. The fault detection rate begins at the same starting point and steadily decreases with time. The decline rate is greatly affected by the parameter $c$. Specifically, a higher value of $c$ results in a quicker decline of $b(t)$, indicating that the ability to detect faults decreases rapidly as the test continues. In contrast, a lower $c$ value leads to a slower decline, and the fault-detection process remains stable.

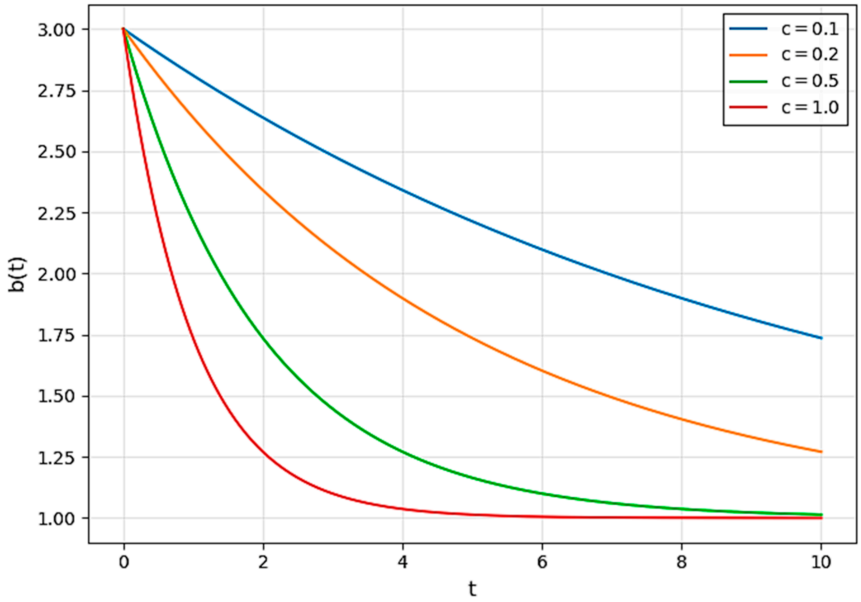



**Figure 1.** Time-dependent fault-detection rate $b(t)$ for different values of the decay parameter $c$, with $b = 2$ and $\alpha = 1$.

Finally, the newly proposed model is expressed by Equation (6) under the initial condition $m(0) = m_0$, with a constant fault content $a(t) = a$ and fault-detection rate given by $b(t) = \alpha + be^{-ct}$:

$$m(t) = \frac{a}{1 + \left(\frac{a}{m_0} - 1\right) e^{a\left(\frac{b}{c}(e^{-ct}-1)\ -\alpha t\right)}} \tag{6}$$

Table 1 lists the 18 NHPP-based SRGMs considered in this study. Models 1–15 presume that failures occur independently, whereas the last three models, which include the proposed model, specifically consider the behavior of dependent failures.

**Table 1.** Existing and proposed SRGMs.

| No. | Model | Mean Value Function |
|---|---|---|
| 1 | Goel–Okumoto (GO) [1] | $m(t) = a\left(1 - e^{-bt}\right)$ |
| 2 | Yamada et al. (DS) [16] | $m(t) = a\left(1 - (1+bt)e^{-bt}\right)$ |
| 3 | Ohba (IS) [17] | $m(t) = \frac{a\left(1-e^{-bt}\right)}{1+\beta e^{-bt}}$ |
| 4 | Yamada et al. (YID1) [18] | $m(t) = \frac{ab}{\alpha+b}\left(e^{\alpha t} - e^{-bt}\right)$ |
| 5 | Yamada et al. (YID2) [18] | $m(t) = a\left(1 - e^{-bt}\right)\left(1 - \frac{\alpha}{b}\right) + \alpha a t$ |
| 6 | Pham et al. (PNZ) [7] | $m(t) = \frac{a\left(1-e^{-bt}\right)\left(1-\frac{\alpha}{b}\right)+\alpha a t}{1+\beta e^{-bt}}$ |
| 7 | Pham–Zhang (PZ) [9] | $m(t) = \frac{\left((c+a)\left(1-e^{-bt}\right)-\left(\frac{ab}{b-\alpha}\right)\left(e^{-at}-e^{-bt}\right)\right)}{1+\beta e^{-bt}}$ |
| 8 | Yamada et al. (YE) [19] | $m(t) = a\left(1 - e^{-\gamma\alpha\left(1-e^{-\beta t}\right)}\right)$ |
| 9 | Yamada et al. (YR) [19] | $m(t) = a\left(1 - e^{-\gamma\alpha\left(1-e^{\frac{\beta t^2}{2}}\right)}\right)$ |
| 10 | Hossain–Dahiya (HDGO) [20] | $m(t) = \log\left[\frac{(e^a - c)}{\left(e^{ae^{-bt}} - c\right)}\right]$ |
| 11 | Zhang et al. (ZFR) [21] | $m(t) = \frac{a}{p-\beta}\left[1 - \left(\frac{(1+\alpha)e^{-bt}}{1+\alpha e^{-bt}}\right)^{\frac{c}{b}(p-\beta)}\right]$ |
| 12 | Teng–Pham (TP) [22] | $m(t) = \frac{a}{p-q}\left[1 - \left(\frac{\beta}{\beta+(p-q)\ln\left(\frac{c+e^{bt}}{c+1}\right)}\right)^{\alpha}\right]$ |
| 13 | Pham (IFD) [5] | $m(t) = a - ae^{-bt}\left(1 + (b+d)t + bdt^2\right)$ |
| 14 | Roy et al. (RMD) [23] | $m(t) = a\alpha\left(1 - e^{-bt}\right) - \left[\frac{ab}{b-\beta}\left(e^{-\beta t} - e^{-bt}\right)\right]$ |
| 15 | Kapur et al. (KSRGM) [24] | $m(t) = \frac{a}{1-\alpha}\left[1 - \left(\left(1 + bt + \frac{b^2t^2}{2}\right)e^{-bt}\right)^{p(1-\alpha)}\right]$ |
| 16 | Lee at al. (DPF1) [15] | $m(t) = \frac{a}{1+\frac{a}{h}\left(\frac{b+c}{c+be^{bt}}\right)^{\frac{a}{b}}}$ |
| 17 | Kim et al. (DPF2) [25] | $m(t) = \frac{a}{1+\frac{a}{h}\left(\frac{1+c}{c+e^{bt}}\right)^{a}}$ |
| 18 | Proposed | $m(t) = \frac{a}{1+\left(\frac{a}{m_0}-1\right)e^{a\left(\frac{b}{c}\left(e^{-ct}-1\right)-\alpha t\right)}}$ |

## 3. Criteria for Model Evaluation

### *3.1. Evaluation Criteria*

To assess the performance of the proposed model, it was compared to 17 well-known models using 12 performance metrics. This evaluation was based on the difference between the observed values $y_i$ and the estimated mean values $\widehat{m}(t_i)$. Here, the symbols $n$ and $m$ denote the number of observations and model parameters, respectively.

Error-based goodness-of-fit measures, such as the mean squared error (MSE), mean absolute error (MAE), and mean error of prediction (MEOP), were calculated by assessing the disparity between the actual and predicted values [26,27].

$$MSE = \frac{\sum_{i=1}^{n}\left[\widehat{m}(t_i) - y_i\right]^2}{n-m}\ ,\ MAE = \frac{\sum_{i=1}^{n}\left|\widehat{m}(t_i) - y_i\right|}{n-m}\ ,\ MEOP = \frac{\sum_{i=1}^{n}\left|\widehat{m}(t_i) - y_i\right|}{n-m+1}$$

The explained variance measures $R^2$ and $adj\_R^2$ represent the coefficients of determination and adjusted coefficient of determination of the regression equation, respectively [28].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[\widehat{m}(t_i) - y_i]^2}{\sum_{i=1}^{n}[y_i - \overline{y}_i]^2},\ adj\_R^2 = 1 - \frac{(1-R^2)(n-1)}{n-m-1}$$

The information theory model selection criteria evaluate model adequacy while accounting for model complexity. Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) evaluate models by maximizing the likelihood of applying penalties for complexity. AIC is applied to maximize the Kullback–Leibler level between the probability distribution of the model and the data, whereas BIC applies a more significant penalty based on Bayesian model selection principles [29]. Following parameter estimation using the LSE method, the corresponding log-likelihood ($logL$) values were calculated and subsequently used in the computation of AIC and BIC values. Pham's information criterion (PIC) and Pham's criterion (PC) highlight the balance between the fit quality and model simplicity by enforcing stricter penalties on extra parameters, particularly in small-sample situations [30,31].

$$AIC = -2logL + 2m,\ BIC = -2logL + mlogn$$

$$L = \prod_{i=1}^{n} \frac{(\widehat{m}(t_i) - \widehat{m}(t_{i-1}))^{y_i - y_{i-1}}}{(y_i - y_{i-1})!} e^{-(\widehat{m}(t_i) - \widehat{m}(t_{i-1}))}$$

$$logL = \sum_{i=1}^{n}\{(y_i - y_{i-1})ln(\widehat{m}(t_i) - \widehat{m}(t_{i-1})) - (\widehat{m}(t_i) - \widehat{m}(t_{i-1})) - ln((y_i - y_{i-1})!)\}$$

$$PIC = \sum_{i=1}^{n}[\widehat{m}(t_i) - y_i]^2 + m\left(\frac{n-1}{n-m}\right)$$

$$PC = \frac{n-m}{2} \times log\left[\frac{\sum_{i=1}^{n}[\widehat{m}(t_i) - y_i]^2}{n}\right] + m\left(\frac{n-1}{n-m}\right)$$

Predictive performance and stability measures assess the relative deviation and tail behavior of the predictions over time. The predicted relative variation (PRV) measures the dispersion of the prediction bias through its standard deviation [32], whereas the root mean square prediction error (RMSPE) quantifies the overall magnitude of the prediction error [33]. The tail's statistic (TS) is the average percentage deviation from the mean across all time points [34].

$$PRV = \sqrt{\frac{\sum_{i=1}^{n}[y_i - \widehat{m}(t_i) - Bias]^2}{n-1}},\ Bias = \sum_{i=1}^{n}\left[\frac{\widehat{m}(t_i) - y_i}{n}\right]$$

$$RMSPE = \sqrt{\frac{\sum_{i=1}^{n}[y_i - \widehat{m}(t_i) - Bias]^2}{n-1} + \sum_{i=1}^{n}\left[\frac{\widehat{m}(t_i) - y_i}{n}\right]^2}$$

$$TS = 100 \times \sqrt{\frac{\sum_{i=1}^{n}[y_i - \widehat{m}(t_i)]^2}{\sum_{i=1}^{n} {y_i}^2}}$$

The predicted sum squares error (preSSE) represents the sum of squared differences between the predicted values and the actual values, and it is used to evaluate the prediction accuracy of a model [35].

$$preSSE = \sum_{i=k+1}^{n}[\widehat{m}(t_i) - y_i]^2$$

These 13 criteria were used to evaluate overall model performance. For $R^2$ and adjusted $R^2$, values closer to one indicate a better model fit, whereas smaller values of the remaining 10 criteria correspond to superior performance. The model parameters were estimated using the least squares estimation (LSE) method, which was executed in MATLAB R2023b and R version 4.4.1, following which all evaluation criteria were calculated. The estimation procedure minimizes the sum of squared errors between the observed cumulative failures and the corresponding values estimated by the model mean value function.

### *3.2. Multi-Criteria Decision Framework*

To further assess the performance of the proposed model, multiple multi-criteria decision metrics such as RED, MCDM, MCDMR, and MCDMM were employed based on the individual evaluation criteria described in Section 3.1. Since different evaluation criteria reflect different aspects of model performance, inconsistencies among individual measures may occur. For example, information criteria such as AIC and BIC consider both fitting accuracy and model complexity, whereas error-based measures evaluate different characteristics of model behavior. Therefore, relying solely on a single criterion may lead to biased conclusions regarding model performance. Multi-criteria decision-making methods have been extensively used in various application domains for ranking and selection problems, including approaches such as AHP, TOPSIS, and fuzzy-based decision frameworks [36–38]. However, in software reliability analysis, the use of integrated multi-criteria approaches remains relatively limited.

#### 3.2.1. RED

The RED criterion was used to select the best model based on a set of contributing criteria. The RED function is defined as follows [39]:

$$RED_i = \sum_{j=1}^{d} w_j \sqrt{\left(\frac{C_{ij1}}{\sum_{i=1}^{s} C_{ij1}}\right)^2 + \left(\frac{C_{ij2}}{\sum_{i=1}^{s} C_{ij2}}\right)^2}$$

where $s$ denotes the total number of models, $d$ represents the number of evaluation criteria, $w_j$ is the weight associated with the $j-$th criterion, $C_{ij1}$ denotes the value of criterion $j$ for model $i$, and $C_{ij2}$ represents the rank of model $i$ with respect to criterion $j$. A lower RED value indicates better overall performance, whereas a higher RED value suggests a lower rank compared to the other models.

#### 3.2.2. MCDM

Next, the MCDM was utilized to demonstrate the superiority of the proposed model by conducting a comparative analysis of various SRGMs [40,41]. For this purpose, a decision matrix $C$ was constructed, where each element represents the value of a specific evaluation criterion for a given model. The decision matrix is defined as follows Equation (7):

$$C = \begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1k} \\ C_{21} & C_{22} & & C_{2k} \\ \vdots & & \ddots & \vdots \\ C_{s1} & C_{s2} & \cdots & C_{sk} \end{pmatrix} \tag{7}$$

where $s$ denotes the total number of models, and $k$ represents the number of evaluation criteria. To eliminate scale differences among the criteria, decision matrix $C$ was normalized

to obtain a normalized matrix $Y = (y_{ij})$. Each element of $Y$ was calculated as $y_{ij} = \frac{C_{ij}}{\sum_{i=1}^{s} C_{ij}}, i = 1, 2, \ldots, s, j = 1, 2, \ldots, k$. The normalized matrix $Y$ is given by Equation (8):

$$Y = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & & y_{2k} \\ \vdots & & \ddots & \vdots \\ y_{s1} & y_{s2} & \cdots & y_{sk} \end{pmatrix} \tag{8}$$

Based on the normalized values, the MCDM values for each model were obtained by summing the normalized criterion values across all the criteria, yielding Equation (9):

$$MCDM = \begin{pmatrix} y_{11} + y_{12} + & \cdots & + y_{1k} \\ y_{21} + y_{22} + & \cdots & + y_{12} \\ \vdots & & \vdots \\ y_{s1} + y_{s2} + & \cdots & + y_{sk} \end{pmatrix} = \begin{pmatrix} SY_1 \\ SY_2 \\ \vdots \\ SY_s \end{pmatrix} \tag{9}$$

3.2.3. MCDMR

The MCDMR reflects the ranking of each suitability among the methods used to determine weights [42]. First, the matrix R was constructed by ranking each model for each criterion using Equation (10):

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1k} \\ r_{21} & r_{22} & & r_{2k} \\ \vdots & & \ddots & \vdots \\ r_{s1} & r_{s2} & \cdots & r_{sk} \end{pmatrix} \tag{10}$$

The ranking of each criterion by $w_{ij} = \frac{r_{ij}}{\sum_{i=1}^{s} r_{ij}}, i = 1, 2, \ldots, s, j = 1, 2, \ldots, k$ was used to create a weight matrix. Matrix $V$ was then created by multiplying the normalized matrix $Y$ defined in Equation (8) by each weight $w_{ij}$:

$$V = \begin{pmatrix} w_{11}y_{11} & w_{12}y_{12} & \cdots & w_{1k}y_{1k} \\ w_{21}y_{21} & w_{22}y_{22} & & w_{2k}y_{2k} \\ \vdots & & \ddots & \vdots \\ w_{s1}y_{s1} & w_{s2}y_{s2} & \cdots & w_{sk}y_{sk} \end{pmatrix}$$

The MCDMR value for each model was obtained by summing the values across all criteria, yielding Equation (11):

$$MCDMR = \begin{pmatrix} w_{11}y_{11} + w_{12}y_{12} + & \ldots & + w_{1k}y_{1k} \\ w_{21}y_{21} + w_{22}y_{22} + & \ldots & + w_{2k}y_{2k} \\ \vdots & & \vdots \\ w_{s1}y_{s1} + w_{s2}y_{s2} + & \ldots & + w_{sk}y_{sk} \end{pmatrix} = \begin{pmatrix} V_1 \\ V_2 \\ \vdots \\ V_s \end{pmatrix} \tag{11}$$

3.2.4. MCDMM

The MCDMM was introduced to reflect the relative size of each criterion, instead of simply relying on ranking or normalization when computing the MCDM [43]. For this purpose, matrix $M$ was constructed to capture the relative size of each criterion. The elements of matrix $M$ were calculated by Equation (12) through $w_{ij} = \frac{C_{ij}}{max_i C_{ij}}, i =$

$1, 2, \ldots, s, j = 1, 2, \ldots, k$, where $max_i C_{ij}$ represents the maximum value of criterion $j$ among all models:

$$M = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1k} \\ w_{21} & w_{22} & & w_{2k} \\ \vdots & & \ddots & \vdots \\ w_{s1} & w_{s2} & \cdots & w_{sk} \end{pmatrix} \tag{12}$$

Next, the matrix $Y \odot M$ was obtained by performing an element-wise product between the matrix $\boldsymbol{Y}$ defined in Equation (8) and the matrix $M$:

$$Y \odot M = \begin{pmatrix} w_{11}y_{11} & w_{12}y_{12} & \cdots & w_{1k}y_{1k} \\ w_{21}y_{21} & w_{22}y_{22} & & w_{2k}y_{2k} \\ \vdots & & \ddots & \vdots \\ w_{s1}y_{s1} & w_{s2}y_{s2} & \cdots & w_{sk}y_{sk} \end{pmatrix}$$

The MCDMM value for each model was obtained by summing the weighted normalized values across all criteria, yielding Equation (13):

$$MCDMM = \begin{pmatrix} w_{11}y_{11} + w_{12}y_{12} + & \cdots & + w_{1k}y_{1k} \\ w_{21}y_{21} + w_{22}y_{22} + & \cdots & + w_{2k}y_{2k} \\ \vdots & & \vdots \\ w_{s1}y_{s1} + w_{s2}y_{s2} + & \cdots & + w_{sk}y_{sk} \end{pmatrix} = \begin{pmatrix} YM_1 \\ YM_2 \\ \vdots \\ YM_s \end{pmatrix} \tag{13}$$

In this framework, smaller RED, MCDM, MCDMR, and MCDMM values indicated superior overall performance. To maintain a consistent evaluation across all criteria, the goodness-of-fit measures, $R^2$ and adjusted R2 values, were transformed into $1 - \mathrm{R}^2$ and $1 - \mathrm{adjR}^2$, respectively. The remaining criteria were applied without transformation.

The weighting procedures adopted in the multi-criteria framework are defined as follows. Equal weighting was used for RED and MCDM; that is, all evaluation criteria contribute equally to model assessment. For MCDMR, weights were determined from normalized ranking values. In MCDMM, weight is defined as the ratio between the criterion value and the corresponding maximum criterion value among all models.

## 4. Numerical Example

### *4.1. Datasets*

Dataset 1 was sourced from the Apache jUDDI project, which is part of the Apache open-source projects and can be accessed via the Apache Issue Tracking System. Failure data were divided into four release groups according to the jUDDI version. Release 1 corresponds to version 3.0.1, followed by Release 2 (version 3.1.0), Release 3 (version 3.1.3), and Release 4 (version 3.1.5) [44]. Dataset 1 was recorded monthly from February 2009 to February 2014, with the total number of failures increasing from an initial count of 10 failures during t = 1, 2, . . . , 61 to a cumulative total of 393 failures.

Dataset 2, originally referred to as J3, comprises the failure data gathered during a single phase of subsystem-level integration testing for the Command and Data Subsystem (CDS) associated with project J2. The data was obtained under relatively stable testing conditions, in which the number of testing hours per week and the allocation of testing effort across major functional areas were nearly constant throughout the observation period. Owing to this controlled testing profile, the failure data in J3 were considered more accurate and homogeneous than those of the other projects. During the J3 testing phase, approximately 10.1 faults per thousand lines of code (KLOC) were detected [45]. A cumulative total of 351 failures were recorded weekly over 41 testing periods.

### 4.2. Results of Dataset 1

Table 2 lists the estimated parameter values for each model derived from Dataset 1. The proposed model contains five parameters, namely $a, b, \alpha, c$, and $m_0$. For Dataset 1, the estimated parameter values were $\hat{a} = 12,867.0328$, $\hat{b} = 0.000134$, $\hat{\alpha} = 0.0000016$, $\hat{c} = 0.2846$, and $\hat{m}_0 = 0.2144$. Figure 2 shows the actual cumulative number of failures along with the corresponding estimates from all 18 models over the entire period for Dataset 1. The black dots indicate the real cumulative failures, and the red line illustrates the estimates from the proposed model. The other colored lines represent the estimates from the existing comparison models. Figure 3 shows the relative error trajectories for all the models based on Dataset 1, further highlighting the superior estimation accuracy of the proposed model. The model maintained predicted values that were closest to the observed data in comparison with the competing models.

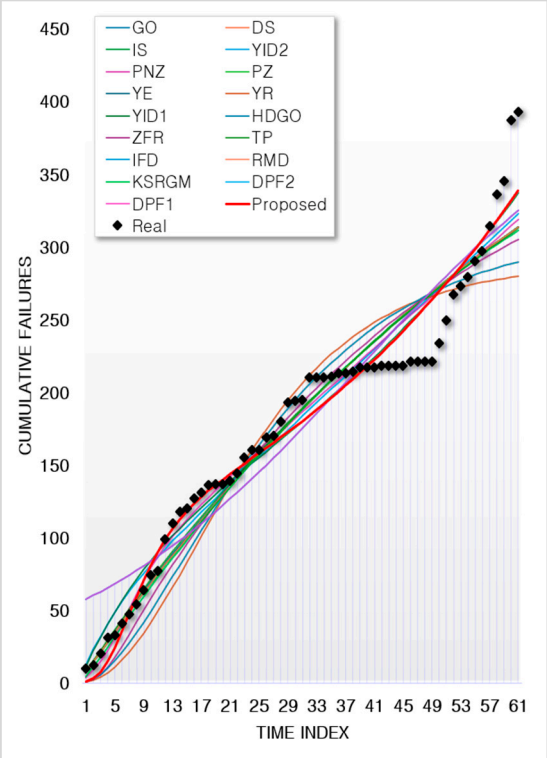



**Figure 2.** Mean value functions of all models using Dataset 1.

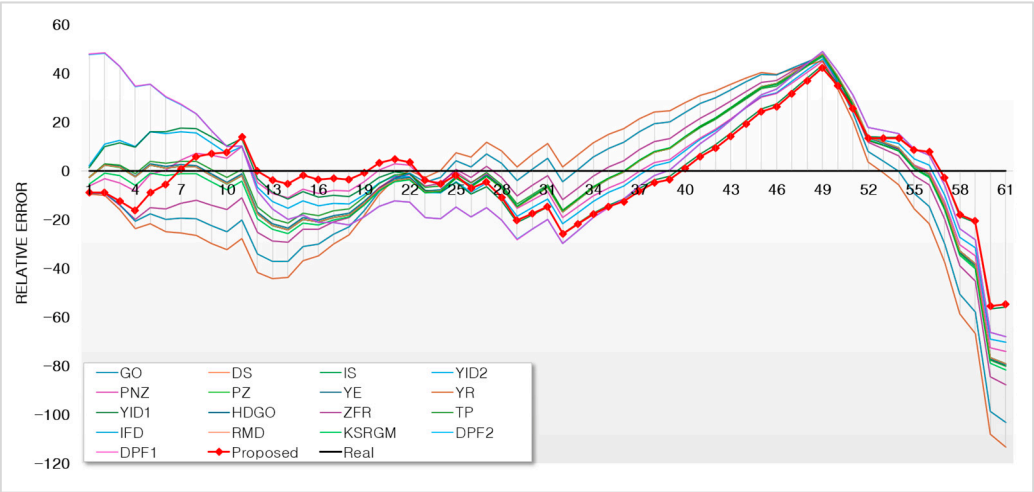



**Figure 3.** Relative error values of all models using Dataset 1.

**Table 2.** Parameter estimation of all models using Dataset 1.

| No. | Model | $\widehat{a}$ | $\widehat{b}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\gamma}$,$\widehat{c}$ | $\widehat{p}$ | $\widehat{q}$,$\widehat{d}$ | $\widehat{m_0}$,$\widehat{h}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | GO | 598.2855 | 0.0122 | | | | | | |
| 2 | DS | 312.2216 | 0.0704 | | | | | | |
| 3 | IS | 598.2825 | 0.0122 | | $6.40 \times 10^{-6}$ | | | | |
| 4 | YID1 | 119.1352 | 0.0989 | 0.0201 | | | | | |
| 5 | YID2 | 62.3701 | 0.2069 | 0.0743 | | | | | |
| 6 | PNZ | 61.9090 | 0.3586 | 0.0713 | 6.0383 | | | | |
| 7 | PZ | 597.5685 | 0.0122 | 119.3517 | $1.09 \times 10^{-4}$ | 0.0128 | | | |
| 8 | YE | 987.6803 | | 0.3524 | 0.0062 | 3.4597 | | | |
| 9 | YR | 325.9055 | | 0.2922 | 0.0013 | 7.2990 | | | |
| 10 | HDGO | 598.2855 | 0.0122 | | | 1.4042 | | | |
| 11 | ZFR | 69.9825 | 0.8464 | 14.8877 | 2.6522 | 0.1322 | 2.8157 | | |
| 12 | TP | 1.1663 | 1.3657 | 0.0095 | 0.0019 | 0.1080 | 1.4289 | 1.4289 | |
| 13 | IFD | 312.2209 | 0.0704 | | | | | $2.30 \times 10^{-7}$ | |
| 14 | RMD | 24.2342 | 0.0122 | 25.6740 | $1.11 \times 10^{-4}$ | | | | |
| 15 | KSRGM | 419.8170 | 9.4990 | 0.2251 | | | 0.0019 | | |
| 16 | DPF1 | 433.2963 | $1.07 \times 10^{-4}$ | | | 0.9441 | | | 63.5059 |
| 17 | DPF2 | 434.0194 | $1.14 \times 10^{-4}$ | | | $6.77 \times 10^{-4}$ | | | 63.4017 |
| 18 | Proposed | 12,867.0328 | $1.35 \times 10^{-4}$ | $1.66 \times 10^{-6}$ | | 0.2846 | | | 0.2144 |

Table 3 lists the criteria derived from the estimated parameters listed in Table 2. It is evident that the MSE, PRV, RMSPE, MAE, MEOP, TS, PIC, and PC values for the proposed model were the lowest at 353.2984, 18.1330, 18.1585, 14.3778, 14.1255, 8.9496, 19,859.7117, and 167.2473, respectively. Furthermore, the R2 and adjusted R2 values were the highest, with values of 0.9597 and 0.9560, respectively, and were the closest to one among all models. Regarding the AIC and BIC criteria, although they do not represent the lowest values, they ranked second among all models with values of 616.3507 and 626.9051, respectively. The YID1 model exhibited the lowest AIC and BIC values (598.8362 and 605.1689, respectively).

**Table 3.** Comparison of the criteria of all models using Dataset 1.

| No. | Model | MSE | $R^2$ | adj_$R^2$ | AIC | BIC | PRV | RMSPE | MAE | MEOP | TS | PIC | PC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | GO | 560.2415 | 0.9326 | 0.9303 | 694.6249 | 698.8467 | 23.4245 | 23.4706 | 16.6753 | 16.3974 | 11.5678 | 33,174.25 | 187.7373 |
| 2 | DS | 947.0918 | 0.8861 | 0.8822 | 932.5013 | 936.7230 | 30.1465 | 30.5113 | 23.1454 | 22.7596 | 15.0404 | 55,998.41 | 203.2257 |
| 3 | IS | 569.9011 | 0.9326 | 0.9291 | 696.6252 | 702.9578 | 23.4245 | 23.4706 | 16.9628 | 16.6753 | 11.5678 | 33,144.26 | 185.6594 |
| 4 | YID1 | 384.9814 | 0.9545 | 0.9521 | 598.8362 | 605.1689 | 19.2796 | 19.2910 | 15.8311 | 15.5628 | 9.5076 | 22,418.92 | 174.2836 |
| 5 | YID2 | 484.3438 | 0.9427 | 0.9397 | 657.1854 | 663.5180 | 21.6301 | 21.6378 | 17.0639 | 16.7747 | 10.6642 | 28,181.94 | 180.9420 |
| 6 | PNZ | 460.1982 | 0.9465 | 0.9427 | 655.9478 | 664.3913 | 20.9078 | 20.9090 | 14.7160 | 14.4623 | 10.3050 | 26,311.29 | 177.0298 |
| 7 | PZ | 590.4541 | 0.9326 | 0.9265 | 700.7640 | 711.3184 | 23.4282 | 23.4745 | 17.5683 | 17.2601 | 11.5698 | 33,140.43 | 181.6275 |
| 8 | YE | 573.9109 | 0.9333 | 0.9285 | 697.8741 | 706.3176 | 23.3142 | 23.3493 | 17.0047 | 16.7115 | 11.5080 | 32,792.92 | 183.3231 |
| 9 | YR | 1224.2446 | 0.8577 | 0.8476 | 1108.6568 | 1117.1003 | 33.5738 | 34.0946 | 27.6614 | 27.1845 | 16.8078 | 69,861.94 | 204.9148 |
| 10 | HDGO | 569.9009 | 0.9326 | 0.9291 | 696.6250 | 702.9576 | 23.4245 | 23.4706 | 16.9628 | 16.6753 | 11.5678 | 33,144.25 | 185.6594 |
| 11 | ZFR | 740.1705 | 0.9170 | 0.9078 | 780.7066 | 793.3719 | 25.7849 | 26.0435 | 20.9628 | 20.5884 | 12.8377 | 40,781.38 | 185.3873 |
| 12 | TP | 589.0623 | 0.9351 | 0.9266 | 700.5562 | 715.3323 | 23.0108 | 23.0249 | 17.3099 | 16.9952 | 11.3479 | 31,879.37 | 176.7071 |
| 13 | IFD | 963.4242 | 0.8861 | 0.8801 | 934.5055 | 940.8381 | 30.1465 | 30.5113 | 23.5445 | 23.1454 | 15.0405 | 55,968.61 | 200.8853 |
| 14 | RMD | 579.8862 | 0.9326 | 0.9278 | 698.6240 | 707.0675 | 23.4243 | 23.4703 | 17.2598 | 16.9622 | 11.5677 | 33,133.51 | 183.6183 |
| 15 | KSRGM | 602.1717 | 0.9300 | 0.9250 | 710.8406 | 719.2841 | 23.7877 | 23.9157 | 17.6973 | 17.3922 | 11.7879 | 34,403.79 | 184.6930 |

**Table 3.** *Cont.*

| No. | Model | MSE | $R^2$ | adj_$R^2$ | AIC | BIC | PRV | RMSPE | MAE | MEOP | TS | PIC | PC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | DPF1 | 778.9282 | 0.9095 | 0.9030 | 792.0447 | 800.4882 | 27.1768 | 27.2022 | 24.6802 | 24.2547 | 13.4068 | 44,478.91 | 192.0283 |
| 17 | DPF2 | 778.1558 | 0.9096 | 0.9031 | 791.6941 | 800.1376 | 27.1635 | 27.1887 | 24.6666 | 24.2414 | 13.4001 | 44,434.88 | 192.0000 |
| 18 | Proposed | 353.2984 | 0.9597 | 0.9560 | 616.3507 | 626.9051 | 18.1330 | 18.1585 | 14.3778 | 14.1255 | 8.9496 | 19,859.71 | 167.2473 |

The 95% confidence intervals for the datasets were computed as follows:

$$\left[\widehat{m}(t) - z_\alpha \sqrt{\widehat{m}(t)}\ ,\ \widehat{m}(t) + z_\alpha \sqrt{\widehat{m}(t)}\right]$$

where $z_\alpha$ is defined as the $100(1-\alpha)/2$th percentile of the standard normal distribution [5].

Figure 4 shows the 95% confidence interval for Dataset 1. The proposed model provided a good fit during the early and middle testing phases. However, deviations were observed in the later period when several observations fell outside the estimated 95% confidence interval. The deviations observed in the later testing phase indicate the possible presence of structural changes in the failure detection process. In future studies, incorporating change-point detection methods into the proposed NHPP framework will enable the model to adapt to regime shifts by partitioning the testing period into multiple phases with distinct parameter values. Such an extension is expected to further improve model flexibility and predictive accuracy, particularly in long-term or dynamically evolving software systems.

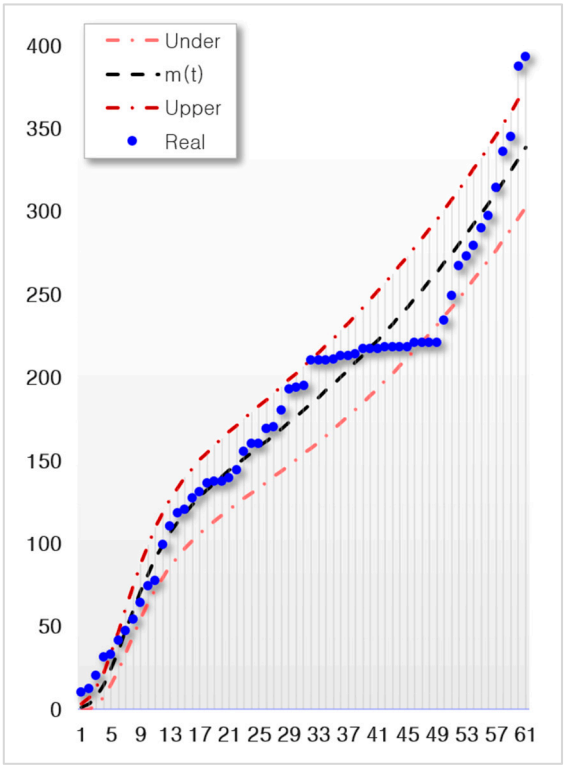



**Figure 4.** 95% confidence intervals of the proposed model using Dataset 1.

Table 4 lists the results of the multi-criteria calculations based on the values reported in Table 3. The proposed model achieved superior overall performance across all multi-criteria metrics, with RED = 0.0396, MCDM = 0.4675, MCDMR = 0.0033, and MCDMM = 0.2345. These findings indicate that the proposed model consistently surpassed all the models included in the study. To enhance visual comprehension, Figure 5

presents a three-dimensional plot of RED, MCDMR, and MCDMM for the top six of the 18 models assessed using Dataset 1. The RED metric is shown in place of MCDM because both MCDMR and MCDMM originate from the MCDM metric. Thus, RED serves as an independent and supplementary metric for comparative visualization.

**Table 4.** Comparison of multi-criteria calculations of all models using Dataset 1.

| No. | Model | RED | MCDM | MCDMR | MCDMM |
|---|---|---|---|---|---|
| 1 | GO | 0.0662 | 0.6088 | 0.0253 | 0.3746 |
| 2 | DS | 0.1185 | 0.8669 | 0.0815 | 0.7312 |
| 3 | IS | 0.0676 | 0.6119 | 0.0271 | 0.3775 |
| 4 | YID1 | 0.0431 | 0.4958 | 0.0059 | 0.2613 |
| 5 | YID2 | 0.0549 | 0.5636 | 0.0157 | 0.3267 |
| 6 | PNZ | 0.0480 | 0.5342 | 0.0096 | 0.2953 |
| 7 | PZ | 0.0785 | 0.6187 | 0.0363 | 0.3842 |
| 8 | YE | 0.0654 | 0.6107 | 0.0250 | 0.3756 |
| 9 | YR | 0.1374 | 1.0325 | 0.1102 | 1.0325 |
| 10 | HDGO | 0.0664 | 0.6119 | 0.0259 | 0.3775 |
| 11 | ZFR | 0.0963 | 0.7167 | 0.0542 | 0.5062 |
| 12 | TP | 0.0682 | 0.6091 | 0.0267 | 0.3724 |
| 13 | IFD | 0.1207 | 0.8717 | 0.0841 | 0.7389 |
| 14 | RMD | 0.0702 | 0.6152 | 0.0294 | 0.3807 |
| 15 | KSRGM | 0.0873 | 0.6303 | 0.0438 | 0.3981 |
| 16 | DPF1 | 0.1111 | 0.7676 | 0.0699 | 0.5822 |
| 17 | DPF2 | 0.1063 | 0.7671 | 0.0653 | 0.5815 |
| 18 | Proposed | 0.0396 | 0.4675 | 0.0033 | 0.2354 |

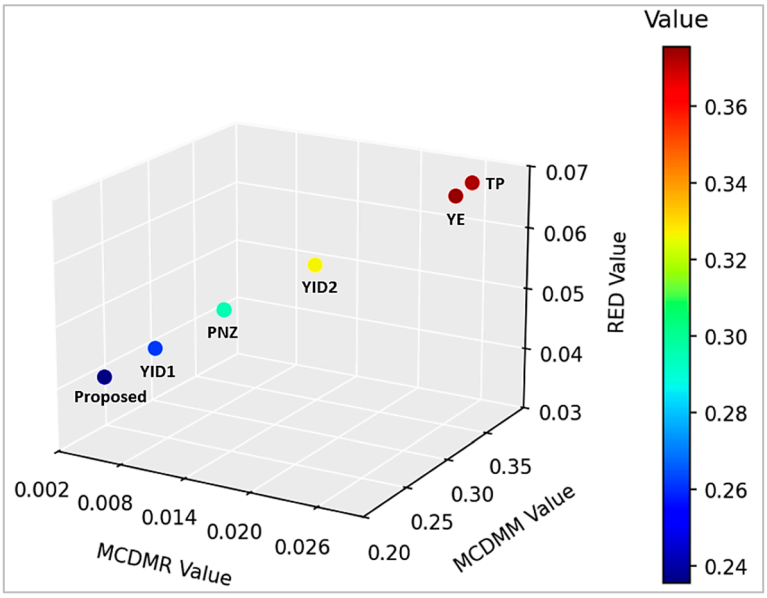



**Figure 5.** 3D plot of MCDMR, MCDMM, and RED of models using Dataset 1.

### *4.3. Results of Dataset 2*

Table 5 lists the estimated parameter values for each model obtained using Dataset 2. The proposed model contains five parameters: $a$, $b$, $\alpha$, $c$, and $m_0$. Each parameter in the proposed model was estimated as follows: $\hat{a} = 360.8328$, $\hat{b} = 0.0160$, $\hat{\alpha} = 0.000433$, $\hat{c} = 0.8760$, and $\widehat{m}_0 = 0.0524$. Figure 6 shows the real cumulative number of failures, along with the corresponding estimates from all 18 models for Dataset 2. The black dots represent the real cumulative failures and the red line represents the estimates from the proposed model. The other colored lines indicate the estimates from existing comparison models. Figure 7 presents the relative error trajectories for all models based on Dataset 2, which shows the predicted values closest to the observed data compared to the other models.

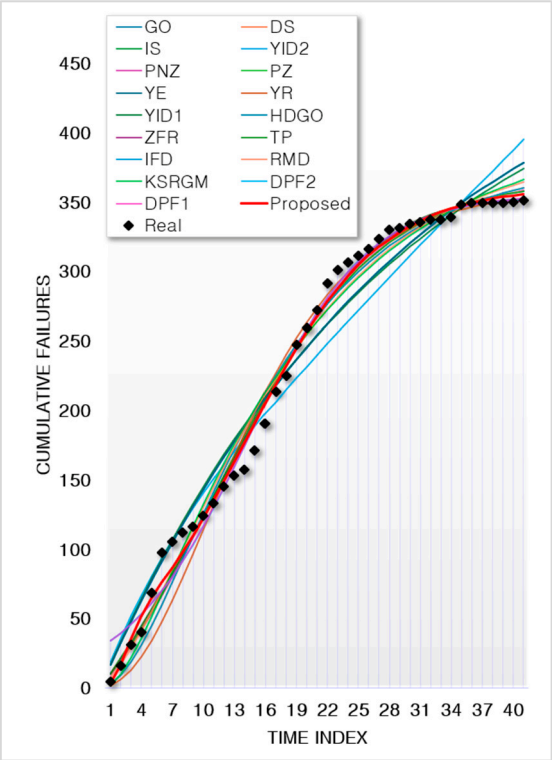



**Figure 6.** Mean value functions of all models using Dataset 2.

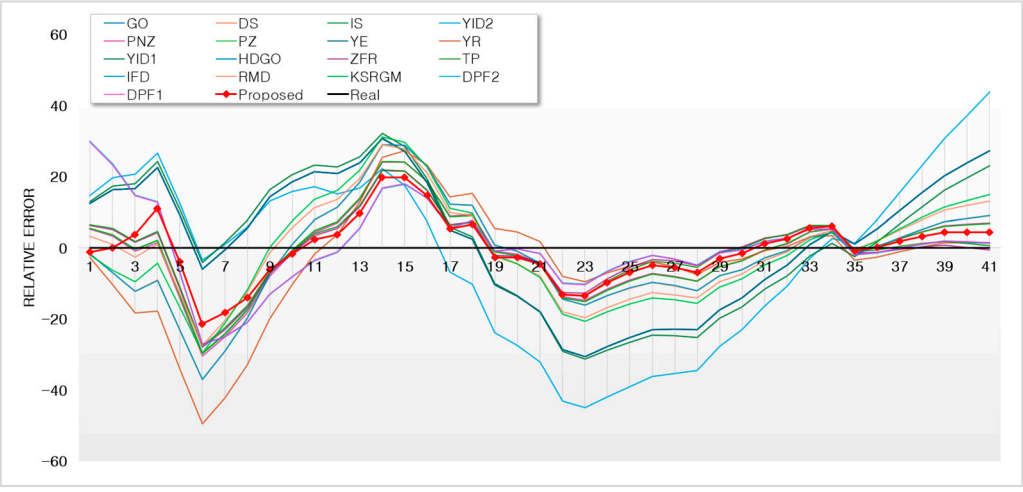



**Figure 7.** Relative error values of all models using Dataset 2.

**Table 5.** Parameter estimations of all models using Dataset 2.

| No. | Model | $\widehat{a}$ | $\widehat{b}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\gamma}$,$\widehat{c}$ | $\widehat{p}$ | $\widehat{q}$,$\widehat{d}$ | $\widehat{m}_0$,$\widehat{h}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | GO | 514.3245 | 0.0324 | | | | | | |
| 2 | DS | 377.3729 | 0.1182 | | | | | | |
| 3 | IS | 367.5242 | 0.1283 | | 4.1668 | | | | |
| 4 | YID1 | 437.2101 | 0.0395 | 0.0025 | | | | | |
| 5 | YID2 | 142.3315 | 0.1359 | 0.0528 | | | | | |
| 6 | PNZ | 360.3713 | 0.1320 | $5.12 \times 10^{-4}$ | 4.3499 | | | | |
| 7 | PZ | 363.0820 | 0.1283 | 13,358.3774 | 4.1688 | 4.4249 | | | |
| 8 | YE | 514.7884 | | 23.4370 | $4.05 \times 10^{-5}$ | 34.1083 | | | |
| 9 | YR | 457.5276 | | 0.4968 | 0.0042 | 3.0521 | | | |
| 10 | HDGO | 514.3245 | 0.0324 | | | 23.2129 | | | |
| 11 | ZFR | 3109.8237 | 0.0776 | 32.2384 | 0.0081 | 0.1086 | 8.8694 | | |
| 12 | TP | 4669.8256 | 0.0771 | 29.1394 | 25.0806 | 39.7926 | 15.6781 | 2.4267 | |
| 13 | IFD | 377.2722 | 0.1185 | | | | | $1.35 \times 10^{-4}$ | |
| 14 | RMD | 333.2178 | 0.1042 | 1.1687 | 0.1042 | | | | |
| 15 | KSRGM | 7.9806 | 0.8942 | 0.9803 | | | 3.9655 | | |
| 16 | DPF1 | 356.0542 | 0.0006 | | | 1.2300 | | | 31.4924 |
| 17 | DPF2 | 355.6816 | 0.0018 | | | 2.8973 | | | 31.7638 |
| 18 | Proposed | 360.8328 | 0.0160 | $4.33 \times 10^{-4}$ | | 0.8760 | | | 0.0524 |

Table 6 lists the criteria derived from the estimated parameters listed in Table 5. It is evident that the values of MSE, AIC, BIC, PRV, RMSPE, MAE, MEOP, TS, PIC, and PC for the proposed model were the lowest, with values of 87.4928, 269.7000, 278.2678, 8.8697, 8.8737, 7.5787, 7.3739, 3.3982, 3199.7414, and 83.7026, respectively. Moreover, the $R^2$ and adjusted $R^2$ values were the highest, with values of 0.9939 and 0.9931, respectively, which were the closest to one among all models. This outcome indicates that the proposed model demonstrated superior performance in fitting Dataset 2.

**Table 6.** Comparison of criteria of all models using Dataset 2.

| No. | Model | MSE | $R^2$ | adj_$R^2$ | AIC | BIC | PRV | RMSPE | MAE | MEOP | TS | PIC | PC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | GO | 345.9104 | 0.9741 | 0.9727 | 335.5469 | 338.9741 | 18.1973 | 18.3606 | 16.6390 | 16.2230 | 7.0328 | 13,570.51 | 115.0766 |
| 2 | DS | 205.2793 | 0.9846 | 0.9838 | 294.8815 | 298.3087 | 14.0317 | 14.1445 | 11.3382 | 11.0548 | 5.4177 | 8085.89 | 104.9013 |
| 3 | IS | 135.1312 | 0.9901 | 0.9893 | 283.2564 | 288.3971 | 11.3130 | 11.3298 | 9.1318 | 8.8977 | 4.3389 | 5194.98 | 94.9328 |
| 4 | YID1 | 375.9467 | 0.9725 | 0.9703 | 337.8801 | 343.0208 | 18.8114 | 18.8963 | 17.5245 | 17.0752 | 7.2371 | 14,345.97 | 114.3737 |
| 5 | YID2 | 636.7465 | 0.9535 | 0.9497 | 391.2470 | 396.3877 | 24.5318 | 24.5934 | 22.3049 | 21.7330 | 9.4186 | 24,256.37 | 124.3852 |
| 6 | PNZ | 141.2056 | 0.9900 | 0.9888 | 285.8218 | 292.6760 | 11.4033 | 11.4281 | 9.3896 | 9.1425 | 4.3766 | 5277.94 | 94.0042 |
| 7 | PZ | 142.6395 | 0.9901 | 0.9887 | 287.2522 | 295.8200 | 11.3125 | 11.3299 | 9.6384 | 9.3779 | 4.3389 | 5185.02 | 92.5004 |
| 8 | YE | 364.7168 | 0.9740 | 0.9712 | 339.5860 | 346.4403 | 18.2017 | 18.3634 | 17.5399 | 17.0784 | 7.0338 | 13,547.86 | 111.5590 |
| 9 | YR | 290.9413 | 0.9793 | 0.9770 | 312.2254 | 319.0797 | 15.9533 | 16.3940 | 12.0716 | 11.7540 | 6.2822 | 10,818.16 | 107.3780 |
| 10 | HDGO | 355.0133 | 0.9741 | 0.9719 | 337.5469 | 342.6877 | 18.1973 | 18.3606 | 17.0768 | 16.6390 | 7.0328 | 13,550.51 | 113.2851 |
| 11 | ZFR | 112.2783 | 0.9924 | 0.9911 | 281.8465 | 292.1279 | 9.8992 | 9.9115 | 8.1049 | 7.8797 | 3.7957 | 3977.74 | 86.7054 |
| 12 | TP | 116.5919 | 0.9924 | 0.9908 | 283.4963 | 295.4913 | 9.9483 | 9.9549 | 8.5550 | 8.3106 | 3.8123 | 4010.79 | 85.9503 |
| 13 | IFD | 211.1144 | 0.9846 | 0.9833 | 296.9876 | 302.1283 | 14.0449 | 14.1590 | 11.6426 | 11.3440 | 5.4233 | 8082.35 | 103.4098 |
| 14 | RMD | 189.3334 | 0.9865 | 0.9850 | 297.2531 | 304.1074 | 13.2307 | 13.2337 | 11.7089 | 11.4007 | 5.0679 | 7058.67 | 99.4302 |
| 15 | KSRGM | 227.7581 | 0.9838 | 0.9820 | 304.7429 | 311.5972 | 14.4896 | 14.5141 | 13.1389 | 12.7932 | 5.5584 | 8480.38 | 102.8485 |

**Table 6.** *Cont.*

| No. | Model | MSE | $R^2$ | adj_$R^2$ | AIC | BIC | PRV | RMSPE | MAE | MEOP | TS | PIC | PC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | DPF1 | 141.2447 | 0.9899 | 0.9888 | 343.3786 | 350.2329 | 11.4240 | 11.4301 | 8.7849 | 8.5537 | 4.3772 | 5279.39 | 94.0094 |
| 17 | DPF2 | 141.2570 | 0.9899 | 0.9888 | 343.9011 | 350.7554 | 11.4243 | 11.4306 | 8.7520 | 8.5217 | 4.3774 | 5279.84 | 94.0110 |
| 18 | Proposed | 87.4928 | 0.9939 | 0.9931 | 269.7000 | 278.2678 | 8.8697 | 8.8737 | 7.5787 | 7.3739 | 3.3982 | 3199.74 | 83.7026 |

Figure 8 shows the estimated mean value function and its corresponding 95% confidence interval. The proposed model accurately captured the overall failure trends. In the early stages, nearly all actual failures were within the confidence interval, except at two time points (in the 1st and 6th week). Given the stochastic nature of the NHPP-based failure process, such occasional deviations were expected and did not undermine the overall adequacy of the model fit.

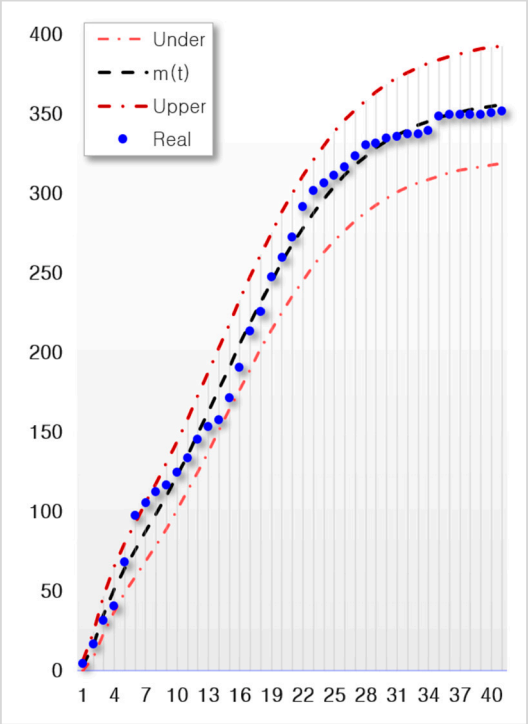



**Figure 8.** 95% confidence intervals of the proposed model using Dataset 2.

Table 7 lists the results of the multi-criteria evaluation calculated based on the values reported in Table 6. Similar to Dataset 1, the proposed model achieved superior overall performance across all multi-criteria metrics, with RED = 0.0336, MCDM = 0.3963, MCDMR = 0.0023, and MCDMM = 0.1697. Figure 9 presents a three-dimensional plot of RED, MCDMR, and MCDMM for the top six models among the 18 models assessed using Dataset 2. The top-performing models for Dataset 2 included the proposed model, ZFR, TP, IS, PZ, and PNZ, which differ from the top six models identified for Dataset 1 (the proposed model, YID1, YID2, YE, PNZ, and TP).

**Table 7.** Comparison of multi-criteria calculations of all models using Dataset 2.

| No. | Model | RED | MCDM | MCDMR | MCDMM |
|---|---|---|---|---|---|
| 1 | GO | 0.1109 | 0.8809 | 0.0733 | 0.6162 |
| 2 | DS | 0.0772 | 0.6283 | 0.0355 | 0.3392 |

**Table 7.** *Cont.*

| No. | Model | RED | MCDM | MCDMR | MCDMM |
|---|---|---|---|---|---|
| 3 | IS | 0.0500 | 0.4971 | 0.0134 | 0.2352 |
| 4 | YID1 | 0.1227 | 0.9207 | 0.0887 | 0.6674 |
| 5 | YID2 | 0.1533 | 1.3032 | 0.1382 | 1.3032 |
| 6 | PNZ | 0.0541 | 0.5050 | 0.0167 | 0.2408 |
| 7 | PZ | 0.0548 | 0.5058 | 0.0170 | 0.2415 |
| 8 | YE | 0.1189 | 0.8983 | 0.0831 | 0.6387 |
| 9 | YR | 0.0962 | 0.7433 | 0.0549 | 0.4484 |
| 10 | HDGO | 0.1124 | 0.8893 | 0.0753 | 0.6269 |
| 11 | ZFR | 0.0390 | 0.4410 | 0.0058 | 0.1977 |
| 12 | TP | 0.0417 | 0.4482 | 0.0084 | 0.2026 |
| 13 | IFD | 0.0800 | 0.6343 | 0.0380 | 0.3442 |
| 14 | RMD | 0.0743 | 0.5997 | 0.0330 | 0.3159 |
| 15 | KSRGM | 0.0885 | 0.6680 | 0.0459 | 0.3775 |
| 16 | DPF1 | 0.0639 | 0.5203 | 0.0260 | 0.2693 |
| 17 | DPF2 | 0.0662 | 0.5202 | 0.0277 | 0.2694 |
| 18 | Proposed | 0.0336 | 0.3963 | 0.0023 | 0.1697 |

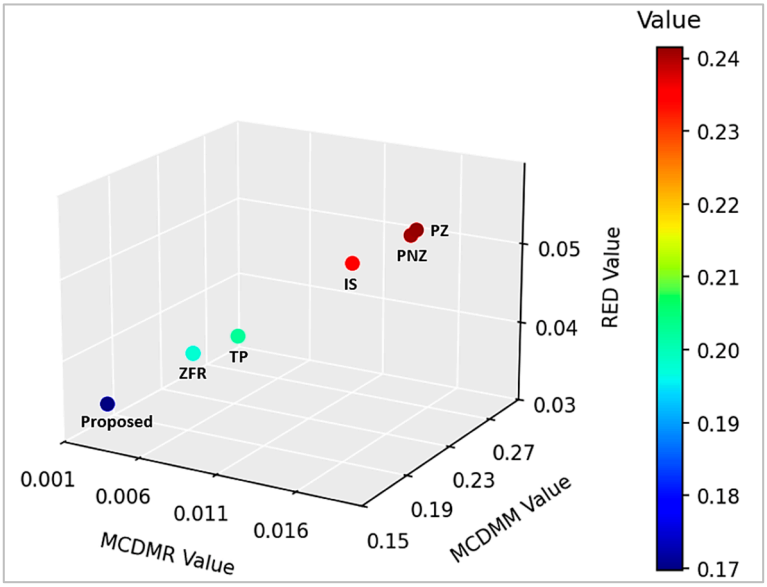
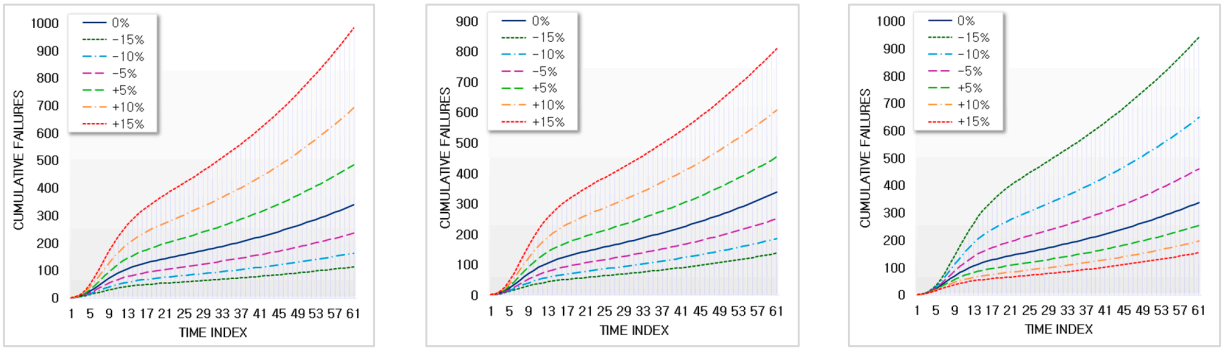


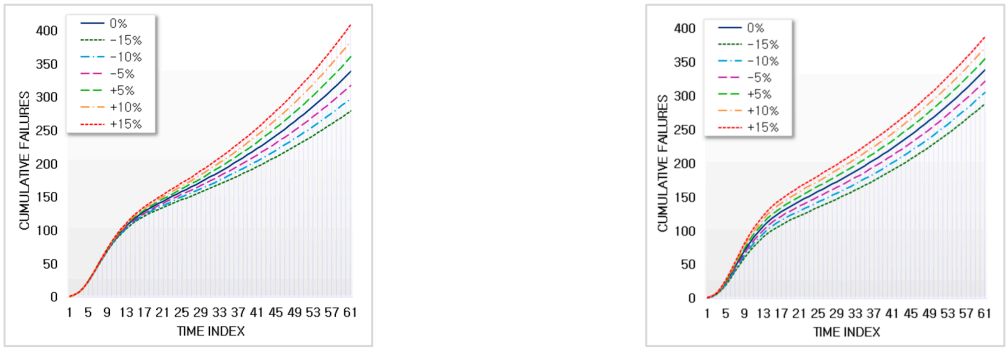

**Figure 9.** 3D plot of MCDMR, MCDMM, and RED of models using Dataset 2.

### *4.4. Sensitivity Analysis*

To examine the effect of the parameters on the behavior of the proposed model, a one-at-a-time sensitivity analysis was conducted. This analysis helped identify the significance of the model parameters and assess the robustness of the proposed model to parameter variations [46,47]. In this analysis, with other parameters held constant, the parameter $m_0$ was adjusted by $\pm 5\%$, $\pm 10\%$, and $\pm 15\%$ around the estimate. The resulting graphs for the two datasets are shown in Figures 10 and 11.

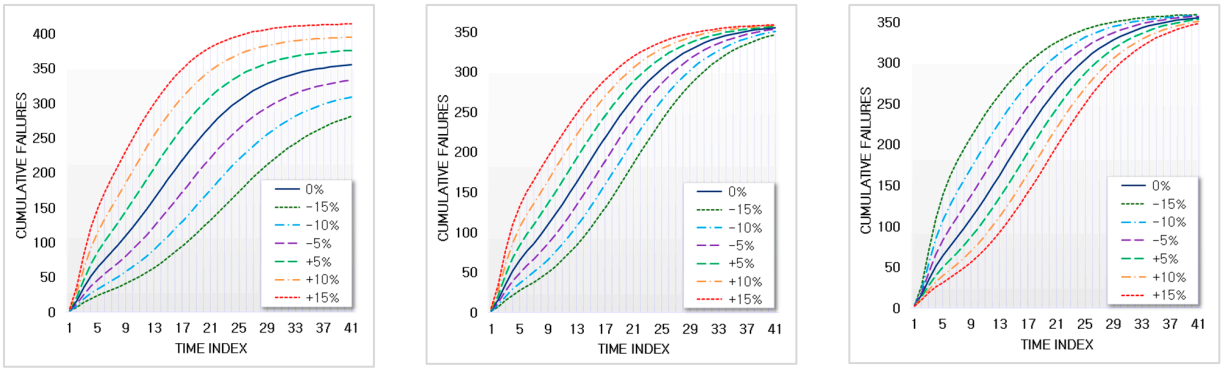



(**a**) (**b**) (**c**)

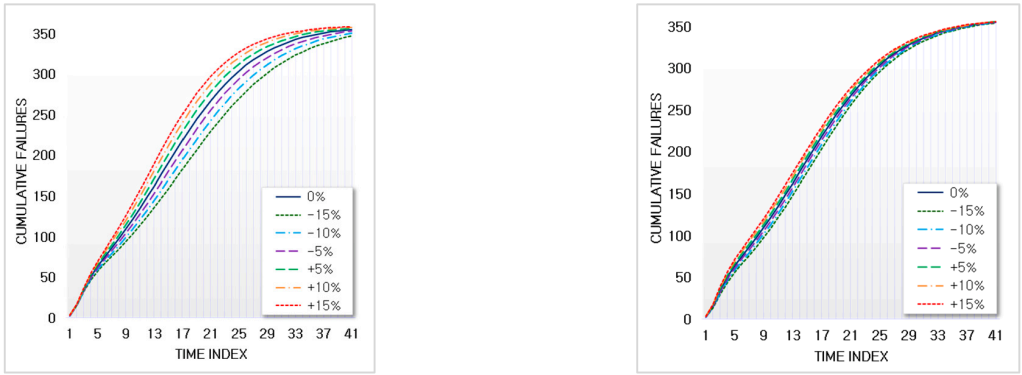



(**d**) (**e**)

**Figure 10.** Sensitivity analysis of changes in the parameters for Dataset 1. (**a**) parameter $a$, (**b**) parameter $b$, (**c**) parameter $c$, (**d**) parameter $\alpha$, and (**e**) parameter $m_0$.




(**a**) (**b**) (**c**)




(**d**) (**e**)

**Figure 11.** Sensitivity analysis of changes in the parameters for Dataset 2. (**a**) parameter $a$, (**b**) parameter $b$, (**c**) parameter $c$, (**d**) parameter $\alpha$, and (**e**) parameter $m_0$.

For Dataset 1, the results indicate that the proposed model shows different sensitivity levels across parameters. Parameters $a$, $b$, and $c$ show relatively high sensitivity, whereas $\alpha$ and $m_0$ exhibit comparatively lower sensitivity. Higher values of parameters $a$ and $b$ result in higher cumulative failures, indicating that these parameters positively influence the fault growth process. In contrast, parameter $c$ exhibits a reversed sensitivity pattern: increasing this parameter decreases the cumulative failures, whereas decreasing $c$ increases it. Thus, demonstrating a negative relationship with the growth behavior of the model. On the other hand, the relatively smaller deviations observed for $\alpha$ and $m_0$ indicate that the proposed model is less sensitive to changes in these parameters and therefore maintains stable behavior under moderate change.

Compared with Dataset 1, Dataset 2 results show noticeably smaller separations among the sensitivity curves, particularly for parameters $\alpha$ and $m_0$. This indicates that the cumulative failure predictions in Dataset 2 are less affected by parameter perturbations and therefore exhibit greater robustness and stability. In Dataset 1, parameter variations produce larger deviations throughout the testing process, especially in the later stages, implying stronger dependence of model predictions on parameter estimation accuracy. In contrast, the curves in Dataset 2 tend to converge toward similar final cumulative failures, suggesting that the model is more constrained and less sensitive to moderate parameter changes. Overall, these results imply that although $a$, $b$, and $c$ remain dominant parameters in both datasets, the proposed model displays more stable predictive behavior for Dataset 2 than for Dataset 1.

The observed differences in sensitivity behavior between the two datasets may be attributed to their distinct software development and failure generation characteristics. Dataset 1 corresponds to an open-source software environment collected over a longer period and may involve iterative development, repeated updates, and evolving software conditions, which can introduce more variable failure patterns and stronger parameter interactions. Therefore, changes in parameter values can propagate more strongly throughout the testing process, leading to larger deviations in cumulative failure behavior. In contrast, Dataset 2 represents a closed-type industrial software environment collected during a relatively controlled testing phase, where the failure process appears more stable and structured. As a result, parameter perturbations have comparatively smaller effects on cumulative failures, producing more robust and consistent model behavior. These findings suggest that the characteristics of the underlying software environment can influence the sensitivity and stability of model parameters.

### *4.5. Prediction Analysis*

To further evaluate the predictive capability of the proposed model, a prediction analysis was performed by estimating model parameters using 90% of the dataset and predicting the remaining 10%. Table 8 shows that the estimated parameter values obtained using 90% of Dataset 1 differ substantially in magnitude from 100% estimation. Proposed model's parameter $\hat{a}$ decreases from 12,867.0328 to 610.8339, while $\hat{m}_0$ increases from 0.2144 to 5.7205, with noticeable variations also observed in other parameters. However, such differences do not necessarily imply instability in the proposed model. Since the model is nonlinear and contains multiple interacting parameters, different parameter combinations can generate similar cumulative failures. When only 90% of the data are used for estimation, the fitting process is primarily influenced by the earlier and middle stages of the failure process and excludes the final 10% of data. Therefore, the model compensates for the missing information of later stages of data by adjusting parameter values to preserve the overall growth behavior. This results in substantial variations in individual parameter estimates even though the overall failure trajectories remain relatively consistent. From

Figures 12 and 13, the results indicate that the proposed model provides a satisfactory fit during the training phase while maintaining reasonable predictive performance in the testing phase. In the prediction part, the real cumulative failures show a noticeable increase following a relatively stable period. The proposed model is able to follow this increasing trend more effectively than several existing models, many of which tend to underestimate the later-stage growth and produce comparatively flatter prediction curves.

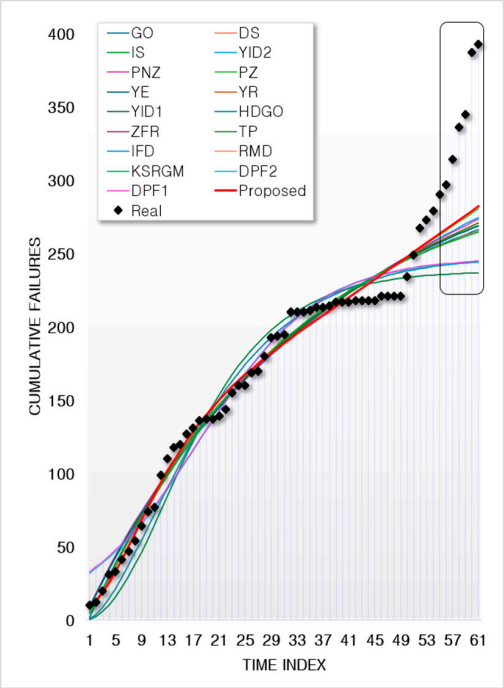



**Figure 12.** Estimated values for the training data and predicted values for the 10% of Dataset 1. The gray box denotes the 10% prediction region.

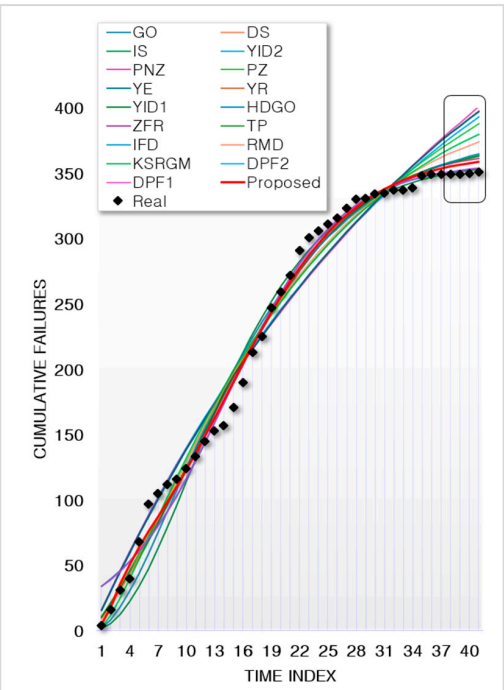



**Figure 13.** Estimated values for the training data and predicted values for the 10% of Dataset 2. The gray box denotes the 10% prediction region.

**Table 8.** The estimated parameter values of the model using 90% of Dataset 1.

| No. | Model | $\widehat{a}$ | $\widehat{b}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\gamma}$,$\widehat{c}$ | $\widehat{p}$ | $\widehat{q}$,$\widehat{d}$ | $\widehat{m_0}$,$\widehat{h}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | GO | 329.8213 | 0.0276 | | | | | | |
| 2 | DS | 249.2065 | 0.0974 | | | | | | |
| 3 | IS | 329.8194 | 0.0276 | | $1.91 \times 10^{-5}$ | | | | |
| 4 | YID1 | 244.9560 | 0.0389 | 0.0050 | | | | | |
| 5 | YID2 | 246.9122 | 0.0383 | 0.0054 | | | | | |
| 6 | PNZ | 114.1839 | 0.1900 | 0.0261 | 3.2520 | | | | |
| 7 | PZ | 308.4236 | 0.0308 | 1.2828 | $7.09 \times 10^{-5}$ | 6.9260 | | | |
| 8 | YE | 387.3673 | | 1.0987 | 0.0068 | 3.2245 | | | |
| 9 | YR | 264.4911 | | 1.1496 | 0.0022 | 2.0003 | | | |
| 10 | HDGO | 329.8213 | 0.0276 | | | 0.8492 | | | |
| 11 | ZFR | 8.0333 | 0.6828 | 1.1368 | 0.0868 | 1.2422 | 0.1127 | | |
| 12 | TP | 9.6583 | 0.4432 | 0.1427 | 0.0466 | 1.7437 | 0.0388 | 0.0314 | |
| 13 | IFD | 249.2064 | 0.0974 | | | | | $6.75 \times 10^{-8}$ | |
| 14 | RMD | 195.8394 | 0.0321 | 1.5851 | 0.6026 | | | | |
| 15 | KSRGM | 5.3141 | 5.5630 | 0.9827 | | | 0.3468 | | |
| 16 | DPF1 | 246.0358 | $1.68 \times 10^{-9}$ | | | $3.73 \times 10^{-6}$ | | | 33.1517 |
| 17 | DPF2 | 246.9666 | $4.44 \times 10^{-4}$ | | | 0.0058 | | | 34.3154 |
| 18 | Proposed | 610.8339 | $7.91 \times 10^{-4}$ | $3.38 \times 10^{-5}$ | | 0.1487 | | | 5.7205 |

Table 9 presents the comparison of goodness-of-fit criteria and preSSE values for models using 90 of Dataset 1. The results indicate that the proposed model demonstrates superior overall performance across most evaluation measures. The proposed model achieves the smallest values for AIC, BIC, PRV, RMSPE, TS, PIC, PC, and preSSE, while also attaining the highest R2 value among all comparing models. For MSE, MAE, and MEOP, the model provides the second-smallest values, slightly behind PNZ model. Although the PNZ model exhibits marginally better performance in these error-based criteria, the proposed model performs better across broader range of criteria and provides smallest preSSE value, indicating stronger predictive capability for unseen data. Furthermore, to provide a more comprehensive assessment of model performance and reduce potential bias associated with relying on individual criteria, multi-criteria decision-making approaches were also employed. Results can be seen in Table 10. The proposed model achieved the best overall performance under all methods, including RED, MCDM, MCDMR, and MCDMM. These results further support the robustness and stability of the proposed model by demonstrating that its superior performance is consistently maintained even when preSSE is included.

It is noteworthy that when the complete Dataset 1 was used, the proposed model achieved only the second-smallest values for AIC and BIC. This can be attributed to the penalty components of AIC and BIC, which evaluate both fitting accuracy and model complexity. Due to the nonlinear structure and parameter interactions within the proposed model, a slightly larger complexity penalty may be imposed compared with other models. However, the prediction results based on 90% of data show superior results across multiple criteria. These suggest that proposed model exhibits stronger predictive capability and better generalization performance.

Compared with Dataset 1, Table 11 shows that only relatively small changes are observed across the estimated parameters for Dataset 2. This suggests that the estimation

process for Dataset 2 is less sensitive to the exclusion of the final 10% of data. Unlike Dataset 1, where substantial parameter adjustments occurred to compensate for missing information, Dataset 2 exhibits greater parameter stability and consistency.

**Table 9.** The comparison of goodness-of-fit criteria and preSSE for models using 90% of Dataset 1.

| No. | Model | MSE | $R^2$ | adj_$R^2$ | AIC | BIC | PRV | RMSPE | MAE | MEOP | TS | PIC | PC | preSSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | GO | 130.7695 | 0.9770 | 0.9761 | 482.3163 | 486.3310 | 11.3166 | 11.3288 | 9.2807 | 9.1088 | 6.2982 | 7038.7 | 130.2022 | 45,591.6 |
| 2 | DS | 212.6581 | 0.9626 | 0.9612 | 605.4479 | 609.4625 | 14.2879 | 14.4443 | 11.8173 | 11.5984 | 8.0317 | 11,378.8 | 143.0878 | 69,229.2 |
| 3 | IS | 133.2843 | 0.9770 | 0.9757 | 484.3168 | 490.3388 | 11.3166 | 11.3288 | 9.4591 | 9.2807 | 6.2982 | 7011.7 | 128.8617 | 45,591.8 |
| 4 | YID1 | 131.6011 | 0.9773 | 0.9760 | 471.9568 | 477.9788 | 11.2350 | 11.2569 | 9.5996 | 9.4185 | 6.2584 | 6924.2 | 128.5312 | 40,958.3 |
| 5 | YID2 | 131.9469 | 0.9772 | 0.9759 | 474.5731 | 480.5951 | 11.2517 | 11.2717 | 9.5671 | 9.3866 | 6.2666 | 6942.2 | 128.5994 | 42,001.9 |
| 6 | PNZ | 110.7927 | 0.9813 | 0.9798 | 456.3807 | 464.4100 | 10.2287 | 10.2292 | 8.5840 | 8.4190 | 5.6868 | 5722.4 | 122.3552 | 36,777.8 |
| 7 | PZ | 134.2612 | 0.9777 | 0.9755 | 509.5165 | 519.5532 | 11.1497 | 11.1497 | 9.4211 | 9.2364 | 6.1985 | 6780.5 | 125.5119 | 47,979.2 |
| 8 | YE | 135.5389 | 0.9771 | 0.9752 | 482.2200 | 490.2493 | 11.2982 | 11.3138 | 9.6963 | 9.5098 | 6.2899 | 6984.4 | 127.4960 | 44,073.4 |
| 9 | YR | 321.1721 | 0.9457 | 0.9413 | 704.5037 | 712.5330 | 17.1116 | 17.4109 | 15.1548 | 14.8633 | 9.6824 | 16,451.7 | 149.4953 | 78,480.4 |
| 10 | HDGO | 133.2843 | 0.9770 | 0.9757 | 484.3163 | 490.3383 | 11.3166 | 11.3288 | 9.4591 | 9.2807 | 6.2982 | 7011.7 | 128.8617 | 45,591.6 |
| 11 | ZFR | 135.7180 | 0.9779 | 0.9752 | 498.0881 | 510.1321 | 11.0973 | 11.0974 | 9.4388 | 9.2500 | 6.1694 | 6714.9 | 124.0914 | 48,971.2 |
| 12 | TP | 129.3718 | 0.9794 | 0.9763 | 483.1026 | 497.1539 | 10.7235 | 10.7237 | 9.4115 | 9.2194 | 5.9617 | 6272.8 | 121.3124 | 43,903.8 |
| 13 | IFD | 216.7479 | 0.9626 | 0.9604 | 607.4487 | 613.4707 | 14.2880 | 14.4443 | 12.0445 | 11.8173 | 8.0317 | 11,351.8 | 141.5042 | 69,229.3 |
| 14 | RMD | 130.5121 | 0.9779 | 0.9762 | 494.7016 | 502.7309 | 11.1022 | 11.1023 | 9.0865 | 8.9118 | 6.1722 | 6728.1 | 126.5322 | 48,883.1 |
| 15 | KSRGM | 131.5843 | 0.9777 | 0.9760 | 498.3783 | 506.4076 | 11.1478 | 11.1478 | 9.1418 | 8.9660 | 6.1975 | 6782.8 | 126.7409 | 49,259.4 |
| 16 | DPF1 | 241.7294 | 0.9591 | 0.9558 | 653.0446 | 661.0739 | 15.0952 | 15.1093 | 12.4739 | 12.2340 | 8.4000 | 12,400.2 | 142.2492 | 69,714.2 |
| 17 | DPF2 | 241.4569 | 0.9592 | 0.9559 | 648.0188 | 656.0481 | 15.0776 | 15.1007 | 12.4968 | 12.2564 | 8.3952 | 12,386.3 | 142.2205 | 68,881.5 |
| 18 | Proposed | 111.0687 | 0.9816 | 0.9797 | 451.6903 | 461.7269 | 10.1411 | 10.1411 | 8.6396 | 8.4702 | 5.6378 | 5620.9 | 120.7710 | 35,690.2 |

**Table 10.** Comparison of multi-criteria calculations of all models using 90% of Dataset 1.

| No. | Model | RED | MCDM | MCDMR | MCDMM |
|---|---|---|---|---|---|
| 1 | GO | 0.0710 | 0.6437 | 0.0319 | 0.3848 |
| 2 | DS | 0.1090 | 0.8837 | 0.0751 | 0.6967 |
| 3 | IS | 0.0771 | 0.6471 | 0.0380 | 0.3882 |
| 4 | YID1 | 0.0666 | 0.6376 | 0.0279 | 0.3776 |
| 5 | YID2 | 0.0694 | 0.6398 | 0.0308 | 0.3799 |
| 6 | PNZ | 0.0453 | 0.5721 | 0.0061 | 0.3112 |
| 7 | PZ | 0.0714 | 0.6489 | 0.0325 | 0.3914 |
| 8 | YE | 0.0765 | 0.6483 | 0.0369 | 0.3891 |
| 9 | YR | 0.1382 | 1.1360 | 0.1210 | 1.1360 |
| 10 | HDGO | 0.0758 | 0.6471 | 0.0369 | 0.3882 |
| 11 | ZFR | 0.0691 | 0.6468 | 0.0295 | 0.3880 |
| 12 | TP | 0.0549 | 0.6216 | 0.0161 | 0.3605 |
| 13 | IFD | 0.1106 | 0.8889 | 0.0772 | 0.7040 |
| 14 | RMD | 0.0608 | 0.6394 | 0.0221 | 0.3806 |
| 15 | KSRGM | 0.0662 | 0.6434 | 0.0278 | 0.3851 |
| 16 | DPF1 | 0.1231 | 0.9451 | 0.0939 | 0.7919 |
| 17 | DPF2 | 0.1190 | 0.9426 | 0.0888 | 0.7876 |
| 18 | Proposed | 0.0445 | 0.5679 | 0.0044 | 0.3065 |

**Table 11.** The estimated parameter values of the model using 90% of Dataset 2.

| No. | Model | $\widehat{a}$ | $\widehat{b}$ | $\widehat{\alpha}$ | $\widehat{\beta}$ | $\widehat{\gamma}$,$\widehat{c}$ | $\widehat{p}$ | $\widehat{q}$,$\widehat{d}$ | $\widehat{m}_0$,$\widehat{h}$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | GO | 610.6080 | 0.0257 | | | | | | |
| 2 | DS | 384.1783 | 0.1155 | | | | | | |
| 3 | IS | 375.8720 | 0.1207 | | 3.7742 | | | | |
| 4 | YID1 | 579.2169 | 0.0275 | $1.41 \times 10^{-4}$ | | | | | |
| 5 | YID2 | 400.9737 | 0.0401 | 0.0093 | | | | | |
| 6 | PNZ | 255.8614 | 0.1348 | 0.0159 | 2.3571 | | | | |
| 7 | PZ | 357.3141 | 0.1702 | 0.0047 | 11.5369 | 0.0382 | | | |
| 8 | YE | 613.1622 | | 2.7913 | $1.42 \times 10^{-4}$ | 64.5798 | | | |
| 9 | YR | 452.4093 | | 0.2316 | 0.0041 | 6.7760 | | | |
| 10 | HDGO | 610.6079 | 0.0257 | | | 0.7840 | | | |
| 11 | ZFR | 319.9559 | 0.0913 | 5.7026 | 2.1972 | 0.2145 | 3.0665 | | |
| 12 | TP | 32.0857 | 0.1027 | 17.6234 | 0.9834 | 4.8480 | 1.0322 | 0.9465 | |
| 13 | IFD | 384.1783 | 0.1155 | | | | | $2.78 \times 10^{-9}$ | |
| 14 | RMD | 331.7689 | 0.0951 | 1.2317 | 0.0951 | | | | |
| 15 | KSRGM | 81.0221 | 1.3347 | 0.8185 | | | 0.2213 | | |
| 16 | DPF1 | 357.5731 | $3.09 \times 10^{-5}$ | | | 0.0652 | | | 31.5799 |
| 17 | DPF2 | 357.5927 | $4.73 \times 10^{-4}$ | | | $1.06 \times 10^{-5}$ | | | 31.5694 |
| 18 | Proposed | 365.0628 | 0.0119 | $4.17 \times 10^{-4}$ | | 0.7966 | | | 0.1770 |

Table 12 results indicate that the proposed model achieves superior overall performance, producing the best results for all twelve goodness-of-fit criteria among the compared models. This demonstrates that the proposed model effectively captures the overall software failure growth process and provides a highly accurate representation of the observed data during the training stage. However, despite its good fitting performance, the proposed model achieves only the fifth-smallest preSSE value in the prediction analysis. Since the preSSE evaluates prediction performance only on the remaining 10% of data, it is highly sensitive to the characteristics of the final testing stage rather than reflecting the model's overall fitting capability. The later part of Dataset 2 exhibits a nearly saturated growth behavior with only minor changes in cumulative failures. Therefore, models capable of producing stronger flattening behavior in the tail region naturally obtain lower preSSE values. Models that show good results in preSSE, such as YR, PZ, DPF1, and DPF2 incorporate additional structural characteristics like delayed S-shaped behavior, logistic fault detection mechanisms, evolving fault content and dependent failure effects. These mechanisms provide additional flexibility in controlling the asymptotic behavior and adjusting curvature of the reliability growth process during the late testing stage.

In contrast, the proposed model's fault content remains constant, and the fault detection rate gradually approaches a nonzero constant value as testing progresses. Although this structure effectively achieves excellent fitting performance, it is primarily designed to represent the complete reliability growth behavior rather than specifically adapting to nearly stationary growth patterns within a limited prediction region.

Lastly, despite achieving only fifth-smallest preSSE value, the proposed model consistently obtained the best overall ranking under the RED, MCDM, MCDMR and MCDMM approaches (Table 13). Since these methods simultaneously consider multiple goodness-of-fit and prediction criteria, the results suggest that the relatively larger preSSE value has limited influence on the overall evaluation. This indicates that the proposed model

maintains superior balance among fitting accuracy, prediction behavior, and reliability growth representation compared with other models.

**Table 12.** The comparison of goodness-of-fit criteria and preSSE for models using 90% of Dataset 2.

| No. | Model | MSE | $R^2$ | adj_$R^2$ | AIC | BIC | PRV | RMSPE | MAE | MEOP | TS | PIC | PC | preSSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | GO | 286.4057 | 0.9781 | 0.9768 | 310.1175 | 313.3394 | 16.5537 | 16.6832 | 15.3852 | 14.9578 | 6.6922 | 10,096.2 | 100.0893 | 6269.1 |
| 2 | DS | 218.2255 | 0.9833 | 0.9824 | 282.1347 | 285.3566 | 14.3931 | 14.5612 | 11.7328 | 11.4069 | 5.8416 | 7709.8 | 95.3314 | 590.2 |
| 3 | IS | 143.0420 | 0.9894 | 0.9884 | 273.3189 | 278.1517 | 11.6100 | 11.6227 | 9.8749 | 9.5928 | 4.6614 | 4917.4 | 86.1123 | 510.3 |
| 4 | YID1 | 297.2725 | 0.9780 | 0.9759 | 310.3196 | 315.1524 | 16.6200 | 16.7522 | 15.9592 | 15.5032 | 6.7199 | 10,161.2 | 98.5480 | 5165.8 |
| 5 | YID2 | 316.9543 | 0.9765 | 0.9744 | 316.9279 | 321.7607 | 17.1683 | 17.2980 | 16.4229 | 15.9536 | 6.9387 | 10,830.4 | 99.6379 | 7210.4 |
| 6 | PNZ | 203.7204 | 0.9853 | 0.9835 | 295.2509 | 301.6945 | 13.6625 | 13.6653 | 13.1370 | 12.7506 | 5.4805 | 6770.7 | 90.2022 | 3918.4 |
| 7 | PZ | 163.5976 | 0.9886 | 0.9867 | 338.2300 | 346.2846 | 12.0516 | 12.0588 | 9.9602 | 9.6584 | 4.8362 | 5280.1 | 84.8606 | 28.4 |
| 8 | YE | 303.9766 | 0.9781 | 0.9754 | 314.1854 | 320.6291 | 16.5600 | 16.6891 | 16.3225 | 15.8425 | 6.6945 | 10,079.2 | 96.8056 | 6289.7 |
| 9 | YR | 325.8885 | 0.9765 | 0.9736 | 303.1594 | 309.6031 | 16.7475 | 17.2696 | 13.3071 | 12.9157 | 6.9316 | 10,802.3 | 97.9540 | 16.8 |
| 10 | HDGO | 294.8294 | 0.9781 | 0.9761 | 312.1175 | 316.9503 | 16.5537 | 16.6832 | 15.8377 | 15.3852 | 6.6922 | 10,078.2 | 98.4077 | 6269.1 |
| 11 | ZFR | 144.0125 | 0.9903 | 0.9883 | 276.9981 | 286.6636 | 11.1173 | 11.1355 | 10.6056 | 10.2742 | 4.4661 | 4507.5 | 81.2588 | 381.3 |
| 12 | TP | 155.8587 | 0.9898 | 0.9873 | 280.3256 | 291.6020 | 11.3939 | 11.3965 | 11.2051 | 10.8436 | 4.5706 | 4717.7 | 80.9884 | 475.0 |
| 13 | IFD | 224.6439 | 0.9833 | 0.9818 | 284.1347 | 288.9675 | 14.3931 | 14.5612 | 12.0779 | 11.7328 | 5.8416 | 7691.8 | 93.7858 | 590.2 |
| 14 | RMD | 185.5763 | 0.9866 | 0.985 | 283.8527 | 290.2964 | 13.0410 | 13.0426 | 12.0526 | 11.6981 | 5.2307 | 6172.0 | 88.6631 | 1632.5 |
| 15 | KSRGM | 218.8016 | 0.9843 | 0.9823 | 286.8795 | 293.3231 | 14.1514 | 14.1619 | 13.2219 | 12.8330 | 5.6797 | 7268.4 | 91.3806 | 2474.7 |
| 16 | DPF1 | 157.7844 | 0.9886 | 0.9872 | 335.1424 | 341.5861 | 12.0185 | 12.0262 | 9.6531 | 9.3692 | 4.8232 | 5254.8 | 85.9862 | 31.2 |
| 17 | DPF2 | 157.7815 | 0.9886 | 0.9872 | 335.1208 | 341.5645 | 12.0184 | 12.0261 | 9.6539 | 9.3699 | 4.8231 | 5254.7 | 85.9858 | 31.4 |
| 18 | Proposed | 95.6735 | 0.9933 | 0.9922 | 258.9323 | 266.9868 | 9.2198 | 9.2218 | 8.2610 | 8.0107 | 3.6984 | 3106.5 | 76.2771 | 216.3 |

**Table 13.** Comparison of multi-criteria calculations of all models using 90% of Dataset 2.

| No. | Model | RED | MCDM | MCDMR | MCDMM |
|---|---|---|---|---|---|
| 1 | GO | 0.1099 | 0.9627 | 0.0785 | 0.8890 |
| 2 | DS | 0.0778 | 0.6846 | 0.0382 | 0.5225 |
| 3 | IS | 0.0493 | 0.5417 | 0.0119 | 0.3406 |
| 4 | YID1 | 0.1156 | 0.9483 | 0.0843 | 0.8692 |
| 5 | YID2 | 0.1300 | 1.0308 | 0.1054 | 1.0180 |
| 6 | PNZ | 0.0807 | 0.7468 | 0.0431 | 0.5492 |
| 7 | PZ | 0.0643 | 0.5761 | 0.0294 | 0.4058 |
| 8 | YE | 0.1172 | 0.9803 | 0.0882 | 0.9217 |
| 9 | YR | 0.1045 | 0.8291 | 0.0756 | 0.7911 |
| 10 | HDGO | 0.1110 | 0.9709 | 0.0793 | 0.9040 |
| 11 | ZFR | 0.0470 | 0.5327 | 0.0103 | 0.3341 |
| 12 | TP | 0.0513 | 0.5543 | 0.0141 | 0.3564 |
| 13 | IFD | 0.0805 | 0.6911 | 0.0416 | 0.5314 |
| 14 | RMD | 0.0695 | 0.6517 | 0.0308 | 0.4501 |
| 15 | KSRGM | 0.0829 | 0.7318 | 0.0446 | 0.5451 |
| 16 | DPF1 | 0.0609 | 0.5691 | 0.0256 | 0.3971 |
| 17 | DPF2 | 0.0593 | 0.5691 | 0.0237 | 0.3971 |
| 18 | Proposed | 0.0354 | 0.4290 | 0.0026 | 0.2339 |

## 5. Conclusions

This study proposed a new NHPP-based SRGM that explicitly accounts for dependent failure behavior with an exponentially decaying fault-detection rate, addressing a key limitation of many classical SRGMs that rely on the assumption of independent failures. By incorporating the failure dependency into the modeling framework, the proposed model accurately reflects realistic software testing and debugging environments characterized by fault interactions, shared components, and imperfect debugging processes. To provide a comprehensive and balanced evaluation of the competing models, this study adopted an integrated multi-criteria decision-making method that incorporates MCDM, MCDMR, MCDMM, and RED. Unlike conventional single-criterion approaches, this framework enables simultaneous consideration of multiple performance and stability criteria.

The effectiveness of the proposed approach was demonstrated through extensive analysis using both closed-type industrial software and OSS failure datasets. Across all evaluated datasets, the proposed model consistently achieved superior performance under all multi-criteria metrics, confirming its robustness. The results indicate that explicitly modeling dependent failures combined with a multi-criteria evaluation strategy leads to more reliable and informative assessments of software reliability growth. In particular, the prediction analysis further demonstrated that the proposed model preserves the overall software failure growth pattern and maintains satisfactory predictive capability even under reduced training conditions. Although prediction performance varied depending on dataset characteristics, the proposed model consistently exhibited good generalization ability. Furthermore, sensitivity analysis revealed that the effects of parameter variations differ across datasets and parameters. Parameters related to fault content and fault detection showed stronger influence on cumulative failure behavior, while other parameters exhibited relatively lower sensitivity.

From a practical perspective, the proposed model can assist software testing managers in estimating software reliability growth and monitoring failure trends during the testing process. The predicted reliability information may support more informed release planning decisions, help identify appropriate testing durations, and assist in allocating testing resources more effectively according to software quality requirements.

Limitations of this study should also be acknowledged. First, parameter estimation was performed using the LSE approach without comparison to alternative estimation methods such as MLE or Bayesian approaches. Second, uncertainty quantification was based on approximate confidence interval estimation rather than more advanced approaches such as bootstrap or likelihood methods. Finally, the multi-criteria framework employed RED, MCDM, MCDMR, and MCDMM methods, while comparisons with alternative decision-making approaches such as AHP and TOPSIS were beyond the scope of the present study.

Future research may extend the proposed model by incorporating change-point mechanisms to capture the structural shifts in software reliability growth caused by changes in testing strategies or operational environments. In addition, the multi-criteria decision method can be further extended by developing a new mean-based MCDM approach capable of handling interval data. This method would enable a more robust model comparison by simultaneously accounting for the performance variability, uncertainty, and robustness.

**Author Contributions:** Conceptualization, I.H.C. and K.Y.S.; methodology, I.H.C.; software, O.-U.O.; validation, I.H.C. and K.Y.S.; formal analysis, K.Y.S. and O.-U.O.; data curation, O.-U.O.; writing—original draft preparation, K.Y.S. and O.-U.O.; writing—review and editing, K.Y.S. and I.H.C.; visualization, O.-U.O.; supervision, I.H.C.; funding acquisition, K.Y.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by research fund from Chosun University, 2026.

**Data Availability Statement:** The data presented in this study are available in the Apache Issue Tracking System for the Apache jUDDI project and in the published source for the J3/CDS software reliability dataset, cited as References [44,45], respectively. These data were derived from the following resources available in the public domain: the Apache jUDDI project and the J3/CDS software reliability dataset.

**Acknowledgments:** The authors would like to thank Chosun University for its research support in 2026.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Goel, A.L.; Okumoto, K. Time-Dependent Error-Detection Rate Model for Software Reliability and Other Performance Measures. *IEEE Trans. Reliab.* **1979**, *R-28*, 206–211. [CrossRef]
2. Yamada, S. *Software Reliability Modeling: Fundamentals and Applications*; Springer: Tokyo, Japan, 2014. [CrossRef]
3. Yamada, S.; Ohba, M.; Osaki, S. S-Shaped Reliability Growth Modeling for Software Error Detection. *IEEE Trans. Reliab.* **1983**, *R-32*, 475–484. [CrossRef]
4. Haque, A.; Ahmad, N. Key Issues in Software Reliability Growth Models. *Recent Adv. Comput. Sci. Commun.* **2022**, *15*, 741–747. [CrossRef]
5. Pham, H. *System Software Reliability*; Springer: London, UK, 2006. [CrossRef]
6. Yamada, S.; Osaki, S. Software Reliability Growth Modeling: Models and Applications. *IEEE Trans. Softw. Eng.* **1985**, *SE-11*, 1431–1437. [CrossRef]
7. Pham, H.; Nordmann, L.; Zhang, Z. A general imperfect-software-debugging model with S-shaped fault-detection rate. *IEEE Trans. Reliab.* **1999**, *48*, 169–175. [CrossRef]
8. Su, C.; Wang, J.; Wang, X.; Li, Z. Robustness analysis of clustered mine cyber-physical system considering node overload and underload under cascading failure. *Proc. Inst. Mech. Eng. Part O J. Risk Reliab.* **2026**. [CrossRef]
9. Pham, H.; Zhang, X. An NHPP Software Reliability Model and Its Comparison. *Int. J. Reliab. Qual. Saf. Eng.* **1997**, *4*, 269–282. [CrossRef]
10. Pham, H.; Zhang, X. NHPP software reliability and cost models with testing coverage. *Eur. J. Oper. Res.* **2003**, *145*, 443–454. [CrossRef]
11. Wang, J.; Geng, H.; Li, P. A software reliability model for open source big data systems based on Weibull–Weibull distribution. *Sci. Rep.* **2025**, *15*, 14670. [CrossRef] [PubMed]
12. Wang, J.; Li, P.; Hu, J.; Ce, Z. A multi-release reliability model of open source software with fault detection obeying three-parameter lifetime distribution. *Sci. Rep.* **2024**, *14*, 19576. [CrossRef] [PubMed]
13. Hwang, C.L.; Yoon, K. *Multiple Attribute Decision Making*; Springer: Berlin/Heidelberg, Germany, 1981. [CrossRef]
14. Opricovic, S.; Tzeng, G.-H. Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *Eur. J. Oper. Res.* **2004**, *156*, 445–455. [CrossRef]
15. Lee, D.H.; Chang, I.H.; Pham, H. Software Reliability Model with Dependent Failures and SPRT. *Mathematics* **2020**, *8*, 1366. [CrossRef]
16. Yamada, S.; Ohba, M.; Osaki, S. s-Shaped Software Reliability Growth Models and Their Applications. *IEEE Trans. Reliab.* **1984**, *R-33*, 289–292. [CrossRef]
17. Ohba, M. Software reliability analysis models. *IBM J. Res. Dev.* **1984**, *28*, 428–443. [CrossRef]
18. Yamada, S.; Tokuno, K.; Osaki, S. Imperfect debugging models with fault introduction rate for software reliability assessment. *Int. J. Syst. Sci.* **1992**, *23*, 2241–2252. [CrossRef]
19. Yamada, S.; Ohtera, H.; Narihisa, H. Software Reliability Growth Models with Testing-Effort. *IEEE Trans. Reliab.* **1986**, *35*, 19–23. [CrossRef]
20. Hossain, S.A.; Dahiya, R.C. Estimating the parameters of a non-homogeneous Poisson-process model for software reliability. *IEEE Trans. Reliab.* **1993**, *42*, 604–612. [CrossRef]
21. Zhang, X.; Teng, X.; Pham, H. Considering fault removal efficiency in software reliability assessment. *IEEE Trans. Syst. Man Cybern.-Part A Syst. Hum.* **2003**, *33*, 114–120. [CrossRef]
22. Teng, X.; Pham, A. A New Methodology for Predicting Software Reliability in the Random Field Environments. *Reliab. IEEE Trans.* **2006**, *55*, 458–468. [CrossRef]
23. Roy, P.; Mahapatra, D.G.S.; Dey, K. An NHPP software reliability growth model with imperfect debugging and error generation. *Int. J. Reliab. Qual. Saf. Eng.* **2014**, *21*, 1450008. [CrossRef]
24. Kapur, P.K.; Pham, H.; Anand, S.; Yadav, K. A Unified Approach for Developing Software Reliability Growth Models in the Presence of Imperfect Debugging and Error Generation. *IEEE Trans. Reliab.* **2011**, *60*, 331–340. [CrossRef]

25. Kim, Y.S.; Song, K.Y.; Pham, H.; Chang, I.H. A Software Reliability Model with Dependent Failure and Optimal Release Time. *Symmetry* **2022**, *14*, 343. [CrossRef]
26. Armstrong, J.S.; Collopy, F. Error measures for generalizing about forecasting methods: Empirical comparisons. *Int. J. Forecast.* **1992**, *8*, 69–80. [CrossRef]
27. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
28. Akossou, A.Y.J.; Palm, R. Impact of data structure on the estimators R-square and adjusted R-square in linear regression. *Int. J. Math. Comput.* **2013**, *20*, 84–93.
29. Kuha, J. AIC and BIC: Comparisons of Assumptions and Performance. *Sociol. Methods Res.* **2004**, *33*, 188–229. [CrossRef]
30. Pham, H. A New Criterion for Model Selection. *Mathematics* **2019**, *7*, 1215. [CrossRef]
31. Pham, H. On Estimating the Number of Deaths Related to COVID-19. *Mathematics* **2020**, *8*, 655. [CrossRef]
32. Xu, J.; Yao, S. Software Reliability Growth Model with Partial Differential Equation for Various Debugging Processes. *Math. Probl. Eng.* **2016**, *2016*, 1–13. [CrossRef]
33. Pillai, K.; Nair, V.S.S. A model for software development effort and cost estimation. *IEEE Trans. Softw. Eng.* **1997**, *23*, 485–497. [CrossRef]
34. Sharma, K.; Garg, R.; Nagpal, C.K.; Garg, R.K. Selection of Optimal Software Reliability Growth Models Using a Distance Based Approach. *IEEE Trans. Reliab.* **2010**, *59*, 266–276. [CrossRef]
35. Kim, Y.S.; Song, K.Y.; Chang, I.H. Prediction and Comparative Analysis of Software Reliability Model Based on NHPP and Deep Learning. *Appl. Sci.* **2023**, *13*, 6730. [CrossRef]
36. Ghorui, N.; Ghosh, A.; Algehyne, E.A.; Mondal, S.P.; Saha, A.K. AHP-TOPSIS Inspired Shopping Mall Site Selection Problem with Fuzzy Data. *Mathematics* **2020**, *8*, 1380. [CrossRef]
37. Simjanovic, D.; Vesić, N.; Ignjatovic, J.; Randjelovic, B. A novel surface fuzzy analytic hierarchy process. *Filomat* **2023**, *37*, 3357–3370. [CrossRef]
38. Sančanin, B.; Penjišević, A.; Simjanović, D.J.; Ranđelović, B.M.; Vesić, N.O.; Mladenović, M. A Fuzzy AHP and PCA Approach to the Role of Media in Improving Education and the Labor Market in the 21st Century. *Mathematics* **2024**, *12*, 3616. [CrossRef]
39. Pham, H. A Logistic Fault-Dependent Detection Software Reliability Model. *J. Univers. Comput. Sci.* **2018**, *24*, 1717–1730. [CrossRef]
40. Zolfani, S.; Yazdani, M.; Pamucar, D.; Zaraté, P. A VIKOR and TOPSIS focused reanalysis of the MADM methods based on logarithmic normalization. *arXiv* **2020**, arXiv:2006.08150. [CrossRef]
41. Taherdoost, H. Analysis of Simple Additive Weighting Method (SAW) as a MultiAttribute Decision-Making Technique: A Step-by-Step Guide. *J. Manag. Sci. Eng. Res.* **2023**, *6*, 21–24. [CrossRef]
42. Song, K.Y.; Kim, Y.S.; Pham, H.; Chang, I.H. A Software Reliability Model Considering a Scale Parameter of the Uncertainty and a New Criterion. *Mathematics* **2024**, *12*, 1641. [CrossRef]
43. Kim, Y.S.; Song, K.Y.; Pham, H.; Chang, I.H. Non-Homogeneous Poisson Process Software Reliability Model and Multi-Criteria Decision for Operating Environment Uncertainty and Dependent Faults. *Appl. Sci.* **2025**, *15*, 5184. [CrossRef]
44. Tandon, A.; Singh, V.; Sharma, M.; Kumari, M. Entropy based Software Reliability Growth Modelling for Open Source Software Evolution. *Teh. Vjesn.* **2020**, *27*, 550–557. [CrossRef]
45. Nikora, A.P.; Lyu, M.R. Software Reliability Measurement Experience. In *Handbook of Software Reliability Engineering*; Lyu, M.R., Ed.; McGraw-Hill: New York, NY, USA, 1996; p. 255.
46. Hamby, D.M. A review of techniques for parameter sensitivity analysis of environmental models. *Environ. Monit. Assess.* **1994**, *32*, 135–154. [CrossRef] [PubMed]
47. Li, X.; Xie, M.; Ng, S.H. Sensitivity analysis of release time of software reliability models incorporating testing effort with multiple change-points. *Appl. Math. Model.* **2010**, *34*, 3560–3570. [CrossRef]