Software reliability growth models: A comparison of linear and exponential fault content functions for study of imperfect debugging situations

Javaid Iqbal

Abstract: The software testing process basically aims at building confidence in the software for its use in real world applications. The reliability of a software system is always important to us. As we carry out the error detection and correction phenomenon on our software, the reliability of the software grows. With an aim to model this growth in the software reliability, many formulations in the form of Software Reliability Growth Models have been proposed. Many of these are based on Non-Homogeneous Poisson Process framework. In this paper, a parallel comparison of the performance of the proposed software reliability growth models is carried out, considering linear and exponential fault content functions for study of imperfect debugging situations. The performance of the proposed models has been compared with some famous existing software reliability models and the proposed models have been validated on some real-world datasets. Three goodness-of-fit criteria that include mean square error, predictive-ratio risk and predictive power are used to carry out the performance comparison of the models. Using these comparison criteria on six actual failure datasets, it is concluded that the proposed Model-2 which always outperforms Model-1, fits the actual failure data better and has better predictive power than other considered SRGMs for at least two data sets.

Subjects: Applied Mathematics; Software Engineering & Systems Development; Computer Science (General)

Keywords: software reliability; reliability growth; software reliability growth model (SRGM); non-homogeneous poisson process (NHPP); learning effect; fault detection rate (FDR); imperfect debugging

ABOUT THE AUTHOR

Javaid Iqbal received his BSc in Mathematics from the University of Kashmir, and Masters in Computer Applications from the same university in 2004. He completed his PhD from the University of Kashmir in 2014. His research interests are Software engineering, Software reliability engineering and reliability modeling, and software performance engineering. He is a member of ACM, CSI and IETE

PUBLIC INTEREST STATEMENT

Software reliability is one of the prime quality concerns. Reliable software is dependable. Software testing helps to improve the reliability of software. This paper explores the assumption of imperfect debugging under two possibilities i.e. linear and exponential fault content functions. This paper demonstrates using real world datasets that under given assumptions how the proposed model with exponential fault content function is better suited to use for the given datasets. The researchers in this field will benefit from the comparison of imperfect debugging situations carried out in this paper.
1. Introduction
The scope and importance of having highly reliable software solutions deployed in every walk of life is an accepted and acknowledged fact. Evaluation of a software before its deployment is a must. Reliability is an important quality characteristic of any software. Software Reliability is defined as the probability that software will provide failure-free operation in a fixed environment for a fixed interval of time (Musa, Iannino, & Okumoto, 1987). The reliability of software applications is measured by use of mathematical formulations called as a software reliability growth model (SRGM) which basically describes the error-detection and correction processes. The SRGMs find their applications in many practical areas which include safety-critical systems like Shuttle’s on-board systems software (Schneidewind & Keller, 1992) and weapon systems (Carnes, 1997).

Over the past three decades many SRGMs have been formulated under Non-homogeneous poisson process (NHPP) modeling framework. Many researchers have used NHPP based SRGMs to capture the reliability growth of a software from the processes of testing and debugging. These models include Goel and Okumoto model (Goel & Okumoto, 1979), Yamada and Ohba model (Yamada, Ohba, & Osaki, 1983), Ohba model (Ohba, 1984) etc. All such models use some underlying assumptions. The general NHPP model is based on the following assumptions: (1) The testing/debugging process follows an NHPP and (2) a debugging procedure is invoked immediately when a software error is detected. Researchers perceive some realistic assumptions to make a model better fit the error-detection and correction process. For example, it is more realistic to consider imperfect debugging than perfect debugging. Goel and Okumoto (1979) proposed an exponential SRGM. The Goel and Okumoto model is one of the earliest NHPP models for software reliability and has been extensively used in literature. This has a concave-shaped mean value function and could not accommodate learning effects very well. As a result, many modifications and generalizations of G–O model have been proposed. Yamada and Ohba (1983) proposed delayed S-shaped SRGM while Ohba (1984) proposed inflection S-shaped SRGM to allow learning effects. These models Goel and Okumoto (1979), Yamada et al. (1983), Ohba (1984) etc. assume perfect debugging. Learning phenomenon has been considered in Yamada, Tokuno, and Osaki (1992), Pham and Zhang (2003) etc. under imperfect debugging assumption. Learning phenomenon and imperfect debugging are two important aspects of many NHPP models.

In this study, imperfect debugging is considered as an increasing function, with linear and exponential behaviors assumed separately for sake of comparisons. This function is called as total fault content (TFC) function. The motivation behind this paper is to study imperfect debugging by a parallel comparison of the performance of the proposed software reliability growth models that incorporate linear and exponential fault content functions for imperfect debugging and then compare the results with some famous existing models using some real datasets. Indeed, linear and exponential fault content functions have been extensively studied to model imperfect debugging situations. It may be noted that the difference between perfect and imperfect debugging is the nature of TFC function i.e. total number of faults. Perfect debugging has constant TFC while as imperfect debugging has a dynamic TFC which is a function of testing time.

The rest of the paper is organized as: In Section 2 the model development is discussed. The mean value functions of two models are considered, many famous existing models are tabulated, sources of six datasets used in this study are listed and estimated parameter values are tabulated for these six datasets. In Section 3 numerical example are given to illustrate the usefulness of these models and to carry out the performance comparison using three comparison criteria of the models. Finally, discussion and conclusion of the analysis made is presented in Sections 4 and 5 respectively.

2. Model development
The solution of the following differential equation representing relationship between failure intensity and the initial fault content is used to represent a general class of NHPP models (Pham, 2007; Pham, Nordmann, & Zuemei Zhang, 1999):
\[ \frac{dm(t)}{dt} = r(t) \cdot [a(t) - m(t)] \tag{1} \]

where \( a(t) \) is the time-dependent fault content function which represents total number of faults in the system including the initial and introduced faults due to imperfect debugging, \( r(t) \) is the time-dependent fault detection rate function representing faults detected per unit time, \( m(t) \) representing mean value function.

The general solution of differential Equation (1) is represented by the following equation:

\[ m(t) = e^{-B(t)} \cdot \left( m_0 + \int_{t_0}^{t} a(\tau) r(\tau) e^{B(\tau)} d\tau \right) \tag{2} \]

where \( B(t) = \int_{t_0}^{t} r(\tau) d\tau \) with \( m_0 = m(t_0) \) is the marginal condition of (2) and \( t_0 \) is the time at which debugging starts.

With this modeling methodology available in literature, our models are developed under common assumption of an increasing time-dependent fault detection rate function \( r(t) \) that exhibits an S-shaped behavior to capture learning effects and is represented by the following equation, where \( \alpha \), represents autonomous errors and \( \eta \) represents the learning factor.

\[ r(t) = (\alpha + \eta) \left( 1 - \frac{\eta}{ae^{\alpha \eta / t} + \eta} \right) \tag{3} \]

However, as per the intent of this paper, we consider two cases of imperfect debugging. In case of model-1, a linear time-dependent fault content function \( a(t) \) is assumed (Yamada et al., 1992) where \( a \) is the initial fault content and \( \beta \) is the constant fault introduction rate. As is evident, this imperfect debugging situation considers that the current number of faults \( a(t) \) is proportional to the testing time.

\[ a(t) = a(1 + \beta \cdot t) \tag{4} \]

and in case of model-2, an exponential the time-dependent fault content function is assumed (Yamada et al., 1992), where \( a \) is the initial fault content and faults can be introduced exponentially per detected fault.

\[ a(t) = c + a(1 - e^{-\beta t}) \tag{5} \]

These two behaviors of a TFC function have been extensively used to represent imperfect debugging situations in literature available for NHPP models. With the assumed behavior of \( r(t) \), and separately considering each behavior of \( a(t) \), in the generalized mean value function \( m(t) \) for NHPP models in Equation (2), we arrive at the following mean value functions for our models:

Model-1 (MVF):

\[ m(t) = \frac{a}{1 + \frac{\alpha}{\alpha + \beta} \left( 1 - e^{-\beta t} \right)} \left( 1 - \frac{1}{\alpha + \eta} \right) + \beta \cdot t \tag{6} \]
Table 1. Listing of models, their mean value functions and fault detection rate functions

<table>
<thead>
<tr>
<th>Model name</th>
<th>MVF and FDR functions</th>
</tr>
</thead>
</table>
| Goel–Okumoto (G–O) (Goel & Okumoto, 1979) | \( m(t) = \alpha (1 - e^{-bt}) \)  
\( a(t) = \alpha \)  
\( r(t) = b \) |
| Yamada–Ohba delayed S-shaped (Yamada et al., 1983) | \( m(t) = \alpha (1 - (1 + bt)e^{-dt}) \)  
\( a(t) = \alpha \)  
\( r(t) = \beta \) |
| Ohba inflection S-shaped (Ohba, 1984) | \( m(t) = \alpha \frac{1 - e^{-\beta t}}{1 + e^{-\beta t}} \)  
\( a(t) = \alpha \)  
\( r(t) = \frac{\beta}{1 + e^{-\beta t}} \) |
| Yamada imperfect debugging (Yamada et al., 1992) | \( m(t) = \alpha \frac{e^{\alpha t}}{1 + e^{\alpha t}}(e^{\beta t} - e^{\gamma t}) \)  
\( a(t) = \alpha e^t \)  
\( r(t) = b \) |
| P–Z model (Pham & Zhang, 2003) | \( m(t) = \frac{1}{1 + e^{-\beta t}} \left( (a + c)(1 - e^{-\beta t}) - \frac{a b}{b-a} (e^{\beta t} - e^{-\beta t}) \right) \)  
\( a(t) = a + c(1 - e^{-\beta t}) \)  
\( r(t) = \frac{\beta}{1 + e^{-\beta t}} \) |
| Chiu–Huang–Lee learning (Chiu, Huang, & Lee, 2008) | \( m(t) = a \left( 1 - \frac{1 + \xi}{2 + \xi e^{\beta t}} \right) \)  
\( a(t) = a \)  
\( r(t) = (a + \eta) \left( 1 - \frac{\xi}{a e^{\beta t} + \alpha} \right) \) |
| Model-1 | \( m(t) = \frac{\alpha}{1 + e^{-\beta t}} \left( 1 - e^{-\beta t} \right) \left( 1 - \frac{\eta}{a e^{\beta t} + \alpha} \right) + \beta \cdot t \)  
\( a(t) = a(1 + \beta \cdot t) \)  
\( r(t) = (a + \eta) \left( 1 - \frac{\alpha}{a e^{\beta t} + \alpha} \right) \) |
| Model-2 | \( m(t) = \frac{\alpha}{a + \eta \cdot e^{-\gamma t}} \left( (a + c)(1 - e^{-\gamma t}) - \frac{a (a + \eta)}{a + \eta - \beta} (e^{-\beta t} - e^{-\gamma t}) \right) \)  
\( a(t) = c + a(1 - \exp(-\beta \cdot t)) \)  
\( r(t) = (a + \eta) \left( 1 - \frac{\alpha}{a e^{\beta t} + \alpha} \right) \) |

Model-2 (MVF):

\[
m(t) = \frac{\alpha}{a + \eta \cdot e^{-\gamma t}} \left( (a + c)(1 - e^{-\gamma t}) - \frac{a (a + \eta)}{a + \eta - \beta} (e^{-\beta t} - e^{-\gamma t}) \right) \tag{7}
\]

Table 1 summarizes the proposed models and other famous existing NHPP models.

Table 2 presents the datasets by labels and the sources of datasets listed.

The parameters estimated for some of the model listed in Table 1 are presented in Table 3 for six datasets used.
3. Numerical applications

In this section, using the comparison criteria like mean square error (MSE), predictive ratio risk (PRR) and predictive power (PP), a parallel comparison of our imperfect debugging models is carried out. The models are also compared to some famous pre-existing models. Six sets of actual software failure data obtained from actual projects have been used. Table lists comparison criteria used in this study. Formula notations include \( m \) is the total number of cumulated errors cumulated between 0 and \( t \), \( m(t) \) is the estimated number of errors at \( t \), obtained from fitting of the mean value function of the proposed model, \( n \) is the number of observations and \( k \) is the number of parameters (Table 4).

Each of these comparison criteria indicates a better fit of model than other when run on same data-set, if the criterion value is smaller. The following Tables 5–7 present criterion (MSE, PRR and PP) value and rank in a value (rank) combination format. Here and onwards, we have used shorter names of models listed in Table 1 and references are also available in Table 1 to avoid repetitions.
Table 3. Comparison of the estimated values of select model parameters

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goel–Okumoto (G–O) (Goel &amp; Okumoto, 1979)</td>
<td>a = 13.974, b = 0.138</td>
</tr>
<tr>
<td>Yamada–Ohba Delayed S-Shaped (Yamada et al., 1983)</td>
<td>a = 124.673, b = 0.356</td>
</tr>
<tr>
<td>Ohba inflection S-Shaped (Ohba, 1984)</td>
<td>a = 135.974, b = 0.138, c = 1.000E−5</td>
</tr>
<tr>
<td>Yamada imperfect debugging (Yamada et al., 1992)</td>
<td>a = 89.894, b = 0.289, α = 0.021</td>
</tr>
<tr>
<td>P–Z model (Pham &amp; Zhang, 2003)</td>
<td>a = 115.855, b = 0.741, c = 51.528, α = 0.058, β = 0.239</td>
</tr>
<tr>
<td>Chiu–Huang–Lee learning (Chiu et al., 2008)</td>
<td>a = 135.965, b = 0.138, c = 1.000E−4, α = 0.024, β = 0.343</td>
</tr>
<tr>
<td>Model-1</td>
<td>a = 126.965, b = 0.016, α = 35.358, β = 0.079</td>
</tr>
<tr>
<td>Model-2</td>
<td>a = 125.769, b = 22.914, α = 12.03, β = 0.092, η = 5.599</td>
</tr>
</tbody>
</table>

Table 4. Comparison criteria

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Formula</th>
</tr>
</thead>
</table>
| MSE | \[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (m_i - m_i)^2
\] |
| PRR (Pham, 2007) | \[
\text{PRR} = \sum_{i=1}^{m} \left( \frac{(m_i - m_{i-1})^2}{m_i} \right)
\] |
| PP (Pham, 2007) | \[
\text{PP} = \sum_{i=1}^{m} \left( \frac{(m_i - m_{i-1})^2}{m_i} \right)
\] |

4. Discussion on results

As can be seen from the Table 5, on the basis of MSE criterion Model-2 that uses exponential fault content function outperforms Model-1 that uses linear fault content function. In fact Model-2 performs best among all pre-existing models for datasets 3 and 4 and ranks third for datasets 1 and 5. Rank 4 is the worst Model-2 has got for datasets 2 and 6. This means that from MSE values, Model-2 provides significantly better estimation than any other model for datasets 3 and 4.

The second comparison criterion used is PRR. The PRR values for Model-2 outrank Model-1 as can be seen from Table 6. Here also, Model-2 outranks all other models for datasets 3, 4 and 5. Model-2
ranks second for data-set 6 and ranks third for data-set 1. However, for data-set 2, Model-2 performs worst. Model-1 ranks third for datasets 2 and 4.

On the basis of PP values from Table 7, Model-2 predicts better than all other models for four out of six used datasets i.e. for datasets 3, 4, 5 and 6 and predicts third best for remaining two datasets. This is indicative of significant predictive power of Model-2 over rest of the models. Model-1 ranks second for data-set 6 and third for data-set 4.
5. Conclusion

We present two new software growth models by considering two imperfect debugging situations represented by linear and exponential fault content functions. The mean value functions of the some famous existing and the two considered models have been used to present estimated parameters for six datasets. Three comparison criteria that include MSE, PRR and PP have been used to present a parallel comparison between the two considered models and among some famous existing models. Based on the discussion on analysis of models using these comparison criteria on six actual failure datasets, it is concluded that the proposed Model-2 which always outperforms Model-1, fits the actual failure data better and has better predictive power than other considered SRGMs for at least two data sets (data-set 3 and 4). In fact Model-2 has better predictive power for four out of six data sets considered in this study. This indicates that for the given assumptions, given failure datasets and comparison criteria, Model-2 will perform better than other models for each comparison criterion for at least two datasets, which is significant enough to validate the models.

However, future research needs to be carried out on extending the existing forms of imperfect debugging situations or finding new such forms to better represent realistic imperfect debugging situations.

References

http://dx.doi.org/10.1016/j.jss.2004.02.028

http://dx.doi.org/10.1109/ISSRE.1997.630855

http://dx.doi.org/10.1016/j.ress.2007.11.004

http://dx.doi.org/10.1109/TR.1979.5220566

http://dx.doi.org/10.1109/24.273589

http://dx.doi.org/10.1016/j.jss.2003.10.024

http://dx.doi.org/10.1016/j.jss.2007.10.002


http://dx.doi.org/10.1147/sj.223.0262


http://dx.doi.org/10.1016/j.ejor.2003.03.035

http://dx.doi.org/10.1007/s11633-007-0325-8

http://dx.doi.org/10.1109/24.784276


http://dx.doi.org/10.1109/32.915646

For sake of drawing conclusion for parallel comparison of Model 1 and Model-2, it may be noted that Model-2 always outranks Model-1 for the given six datasets. Therefore for the datasets used in this study, it is observed that exponential imperfect debugging situations are better suited for use of model-2.
Transactions on Software Engineering, 27, 929–948. http://dx.doi.org/10.1109/32.962562