Software Reliability Modeling With Different Type of Faults Incorporating Both Imperfect Debugging and Change Point

Subhashis Chatterjee, Ankur Shukla

Department of Applied Mathematics, Indian School of Mines, Dhanbad, India

Abstract— In past four decades, many nonhomogeneous Poisson process (NHPP) based software reliability growth models (SRGMs) have been proposed to measure and assess the reliability growth of software. During the testing process, the faults which causes failure are detected and removed. One common assumption of many traditional SRGMs is that the fault removal rate is constant. In practical, the fault removal rate increases with time as learning and maturity of software engineer increases. Hence, time variant fault removal rate has been considered in this study. A complex software system may contain different category of faults. Some of faults can be easily detected and removed and some of faults required more effort to be detected and removed. Therefore, in this article, a NHPP based SRGM has been proposed which incorporates mainly two type of faults, major and minor. The concepts of imperfect debugging and change point have also been incorporated in the proposed SRGM. The parameters of the proposed SRGM is estimated using Statistical Package for Social Sciences (SPSS) software and validation of the proposed SRGM has been done using real life data set.

Keywords— Non-homogeneous Poisson process; Software reliability; Software reliability growth model; Imperfect debugging; Change point

I. INTRODUCTION

In this fast changing world, with the rapid growth of technology a remarkable change has been taken place in the area of computer science. This development in the area of computer science has made the modern society more computer dependent. Software plays a major role in functioning of a computer. Therefore, in this automated world, the demand of quality software is increasing for safe performance of any automated system. Accordingly, it becomes the great concern to control and maintain the reliability of the software for the software industries. According to Musa [1], ‘Software reliability is defined as the failure-free operation of a software under specified environment and specified time’. The reliability of software can be measured by SRGMs. SRGMs are developed based on failure data obtained during software testing phase. In past four decades, many SRGMs considering different assumptions and concepts, have been developed to estimate the reliability growth of the software products during the development process [2-12]. One such assumptions is, fault removal rate is, constant during the testing process [3, 4, 5, 8, 13-14]. In reality, it is observed that fault removal rate can't be constant, it changes due to change in testing strategy, testing skill, testing effort, complexity and size, defect density, etc. [8]. Actually, fault removal rate increases with time as learning and maturity of software engineers increases, and become constant at the end of the testing process [5, 8, 15].

On the other hand, different types of faults may be possible in a complex software system. It is very important to measure the characteristics and severity of the faults to improve the fault removal process. Therefore, this issue should be taken into consideration to study the behaviour and severity of faults. The faults present in a software can be classified as major and minor faults. Minor faults are those faults which can be easily detected and removed, and major faults are those faults which are difficult to detect and remove. Hence, it is obvious that major faults require more effort to be detected and removed compare to the minor faults. Initially, fault removal rate of major faults are less in compare to the minor faults, which may become similar to the minor faults at the end of testing [5, 8].

Due to effect of different factors like testing environment, testing strategy, testing team constitution and efficiency, test case effectiveness, resources, etc., fault removal rate and fault introduction rate get affected and they changes when above factors are changed. These points at which changes are possible is known as ‘change point’. Various SRGMs on the concept of change point have been proposed by many researchers in past few years [8, 16-19].

Previously, Many SRGMs has been proposed considering that the debugging process is perfect, i.e., no fault will be spawned. But in reality, it is observed that faults may be introduced when detected faults are removed during the debugging process. This process is known as imperfect debugging. Considering this phenomenon many SRGMs has been developed [4, 5, 8, 17, 20]. The concept of imperfect debugging has also been incorporated in the proposed SRGM. In this paper, a NHPP based SRGM has been developed considering two types of faults based on severity. In addition, imperfect debugging phenomenon has been considered with fault introduction and time dependent fault removal rate.
Different fault removal and introduction rate has been considered for different types of faults to analyse the realistic behaviour of the faults. Furthermore, to study the effect of different factors such that: testing strategy, testing environment, testing effort, resource allocation, the concept of change point has been considered for different parameters.

Rest of the article is organised as follows: The assumption and formulation of the proposed SRGM in section 2. Parameter estimation of proposed SRGM and different comparison criteria are presented in section 3. Numerical analysis and model validation has been done in section 4. Finally, section 5 concludes the work.

II. MODEL DEVELOPMENT

This section gives a brief idea about NHPP based SRGMs and proposed SRGM.

A. NHPP Based SRGMs

Let N(t) be a counting process representing the cumulative number of faults at time t with the mean value function, m(t), which represent the s-expected number of faults by time t. Goel-Okumoto [13] proposed a SRGM by assuming that N(t) follows NHPP and the failure intensity $\lambda(t)$ is proportional to the number of remaining faults, i.e.,

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

\[ \lambda(t) = \frac{dm(t)}{dt} = b(a - m(t)) \]  

where $a$ denotes the total number of faults present in software before testing and $b$ denotes the proportionality constant which is defined as fault removal rate.

Mean value function, $m(t)$, can be defined as:

\[ m(t) = \int_0^t \lambda(t) dt \]  

Many SRGMs are proposed with the assumptions of Goel-Okumoto SRGM [3-5, 8]. The proposed SRGMs has been also integrated the above assumptions.

B. Proposed SRGM

The proposed SRGM has been developed considering the following assumptions:

1. Software failure follows NHPP.
2. The Software system is subjected to failure caused by remaining faults in the system.
3. Debugging process is imperfect, i.e., it is possible to introduce new faults when detected faults are removed.
4. Fault removal rate increases with time as learning process, testing effort, etc., increases. It is different for different type of faults and it will change due to change in environmental factor, resource allocation, etc. Hence, the fault removal rate $b_y(t)$ will be as follows:

\[ b_y(t) = \begin{cases} b_1(1-q_j t^{-\alpha}), 0 \leq t \leq \tau \\ b_2(1-q_j t^{-\alpha}), t > \tau \end{cases} \]  

where $b_y$, $q_j$, $r_j$ are constant and defined for $i^{th}$ type of faults in $j^{th}$ interval and $r$ is the change point.

5. Due to imperfect debugging new faults are introduced with fault introduction rate $\beta_y(t)$, which is different for different type of faults, defined as:

\[ \beta_y(t) = \begin{cases} \beta_{y1}, 0 \leq t \leq \tau \\ \beta_{y2}, t > \tau \end{cases} \]  

where $\beta_y$ is constant and defined for $i^{th}$ type of faults in $j^{th}$ interval and $r$ is the change point.

From the above assumptions the mean value function for $i^{th}$ type of faults $m_i(t)$ can be obtained solving the following differential equations:

\[ \frac{dm_i(t)}{dt} = b_y(t)[a_i(t) - m_i(t)] \]  

Where $a_i(t)$ is the number of $i^{th}$ type of faults to be eventually detected. The above equations satisfy the initial conditions $a_i(0) = ap_i$ and $m_i(0) = 0$. $p_i$ is the proportion of $i^{th}$ type of faults. Since, fault removal rate can't be negative, hence

\[ 1 - q_j t^{-\alpha} > 0 \]  

\[ t > (q_j)^{1/\alpha} \]  

Hence, this condition must be considered in parameter estimation also.

C. Solution of the Proposed SRGM

Solving the Eqn. (6) and (7) simultaneously with initial conditions $a_i(0) = ap_i$ and $m_i(0) = 0$, we can find the mean value function of the different fault given as follows:

1. Minor Faults (for $i = 1$)

Case I. For $0 \leq t \leq \tau$ mean value function has been shown in eqn. (9)

Case II. For $t > \tau$ mean value function has been shown in eqn. (10)

2. Major Faults (for $i = 2$)

Case I. For $0 \leq t \leq \tau$ mean value function has been shown in eqn. (11)

Case II. For $t > \tau$ mean value function has been shown in eqn. (12)
\[ m_i(t) = \frac{ap_i}{(1 - \beta_i)} \left[ 1 - \exp \left\{ \left(1 - \beta_i\right) \left( t - \frac{q_{i1}}{1 - r_{i1}} t^{\omega_i} \right) \right\} \right] \]  

\[ m_i(t) = \frac{ap_i}{(1 - \beta_i)} \left[ 1 - \exp \left\{ h_i \left(1 - \frac{q_{i1}}{1 - r_{i1}} t^{\omega_i} \right) + h_i \left(1 - \beta_i\right) \left( \frac{q_{i2}}{1 - r_{i2}} \left(t^{\omega_{i2}} - t^{\omega_i} \right) - (t - \tau) \right) \right\} \right] + m(\tau) \frac{(\beta_i - \beta_i)}{(1 - \beta_i)} \]  

\[ m_i(t) = \frac{ap_i}{(1 - \beta_{i2})} \left[ 1 - \exp \left\{ h_i \left(1 - \frac{q_{i1}}{1 - r_{i1}} t^{\omega_i} \right) \right\} \right] \]  

\[ m_i(t) = \frac{ap_i}{(1 - \beta_{i2})} \left[ 1 - \exp \left\{ h_i \left(1 - \frac{q_{i1}}{1 - r_{i1}} t^{\omega_i} \right) + h_i \left(1 - \beta_{i2}\right) \left( \frac{q_{i2}}{1 - r_{i2}} \left(t^{\omega_{i2}} - t^{\omega_i} \right) - (t - \tau) \right) \right\} \right] + m(\tau) \frac{(\beta_{i1} - \beta_{i2})}{(1 - \beta_{i2})} \]  

**D. Mean Value Function of the Proposed SRGM**

The mean value function of the proposed SRGM is defined as follow:

\[ m(t) = \sum_i m_i(t) \]

**III. PARAMETER ESTIMATION AND COMPARISON CRITERIA**

**A. Parameter Estimation**

Least Square method is used to estimate the parameters. The estimated values has been obtained using SPSS. The position of change point is found using 'changepoint' package in 'R' software [21].

**B. Comparison Criteria**

Following comparison criteria has been used for performance analysis of the proposed SRGM.

1. **Mean Square Error (MSE)**

   It is defined as [4, 5]:

   \[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \]

   where \( y_i \) and \( \hat{y}_i \) are the observed and predicted faults respectively, \( n \) is the total number of observations.

2. **Bias**

   It is defined as the sum of the deviation of the estimated curve from the actual data, defined as [22, 23]:

   \[ \text{Bias} = \frac{1}{n} \sum_{i=1}^{n} (m(t_i) - m_i) \]

   smaller value of bias is better goodness of fit.

3. **Variance**

   It is defined as follows [22, 23]:

   \[ \text{Variance} = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (m_i - m(t_i) - \text{Bias})^2} \]

   smaller value of variance is better goodness of fit.

4. **Root Mean Square Prediction Error**

   It measure the closeness with which a model predicts the observation, as shown [22, 23]:

   \[ \text{RMSPE} = \sqrt{\text{Variance}^2 + \text{Bias}^2} \]

   lower value of RMSPE is better goodness of fit.

5. **The 100p% Upper and Lower Limit for m(t)**

   It is defined as follows [24]:

   \[ \hat{m}(t) + \eta_p \sqrt{\hat{m}(t)} \text{ and } \hat{m}(t) - \eta_p \sqrt{\hat{m}(t)} \]

   The bounds of \( m(t) \) approximately as follows:

   \[ \hat{m}(t) + \eta_p \sqrt{\hat{m}(t)} = m(t) = \hat{m}(t) - \eta_p \sqrt{\hat{m}(t)} \]

   where \( \hat{m}(t) \) is the estimate of \( m(t) \) and \( \eta_p \) is the \( \frac{(1 + p) \times 100}{2} \) percent point of the standard.
IV. NUMERICAL ANALYSIS AND MODEL VALIDATION

The data set used here for validation of the proposed SRGM has been taken from the article published by Ohba [25]. This data set consists of 328 software faults during 19 weeks. The change point for this data set is found at the position 8th, 10th and 13th week. Let \( p_1 = p \), since \( p_1 + p_2 = 1 \Rightarrow p_2 = 1 - p \), considering this the estimated values of the parameters of the proposed SRGM and other SRGM has been shown in Table 1.

As shown in Table I, it is clear that fault removal rate for minor faults is greater than the major faults and it increases after change point and fault introduction rate for both type of fault is decreases after change point. This establishes the fact that more testing effort is required to remove the major faults than the minor faults from the system. The estimated total number of faults by proposed SRGM is 328.03 which is very close to the actual number of faults i.e. 328. Hence, from the above results it clear that the assumptions of the proposed SRGM is realistic.

<table>
<thead>
<tr>
<th>Estimated Parameters of Some Existing SRGMs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Model</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Kapur et al. model</td>
</tr>
<tr>
<td>Chatterjee et al. Model</td>
</tr>
<tr>
<td>Shyur Model</td>
</tr>
</tbody>
</table>

TABLE 2. Comparison of the Proposed Model with different existing SRGMs

<table>
<thead>
<tr>
<th>Models</th>
<th>MSE</th>
<th>Bias</th>
<th>Variance</th>
<th>RMSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed Model</td>
<td>36.993</td>
<td>0.0758</td>
<td>6.081</td>
<td>6.082</td>
</tr>
<tr>
<td>Kapur Model</td>
<td>133.35</td>
<td>-0.7439</td>
<td>11.839</td>
<td>11.863</td>
</tr>
<tr>
<td>Chatterjee Model</td>
<td>106.56</td>
<td>0.4504</td>
<td>10.595</td>
<td>10.605</td>
</tr>
<tr>
<td>Shyur Model</td>
<td>122.94</td>
<td>0.3483</td>
<td>11.386</td>
<td>11.392</td>
</tr>
</tbody>
</table>

Fig. 1. Estimated cumulative faults using proposed model and the 95% confidence bounds versus time.

Fig 2. Graphical comparison of the proposed SRGM with other existing SRGM

From the Table II, it is clear that the MSE, Bias, Variance, RMSPE of the proposed SRGM is lower in comparison to the other SRGMs [17, 15, 19], which shows that the performance of the proposed SRGM is better in compare to the other SRGMs [17, 15, 19]. Graphical representation of the cumulative number of faults estimated by the proposed SRGM has been shown in Fig. 1 and from
this figure it is clear that the pattern of the estimated faults are very close to the actual faults present in the software. It means no faults are present in the software at end of the testing. Hence, proposed SRGM is better fit for given data set. Fig. 2 represents the graphical comparison of the proposed SRGM with the other existing SRGMs.

V. CONCLUSION

In this paper, a SRGM has been proposed with new time dependent fault removal rate incorporating both the concept of change point and imperfect debugging. Effect of different types of faults on fault removal rate and fault introduction rate has been shown. The proposed model is more realistic and flexible. Real data has been used to validate the proposed model. Experimental results established the fact that, the proposed model has better and accurate prediction capability. Hence, the proposed models can be very helpful for industry and software professionals to improve the quality of software.

ACKNOWLEDGMENT

Authors acknowledge Indian School of Mines, Dhanbad, India, for providing necessary facilities for this work.

REFERENCES