Measuring Concurrent Effect of Time and Testing Coverage Using Software Reliability Growth Model with Change Point

Anu G. Aggarwal¹, P.K. Kapur² and Gurjeet Kaur³
Department of Operational Research, University of Delhi, Delhi-110007
¹anuagg17@gmail.com, ²pkkapurl@gmail.com and ³gurjeetkaur85@gmail.com

ABSTRACT
Software reliability growth process does not depend solitary on the testing-time as the software reliability growth factor. One of the other major controlling forces behind the growth development during testing phase is Testing Coverage. TC is a significant measure for software developers as it helps in evaluating the quality of the tested software and determines how much additional effort is needed to improve the reliability of the software. In the last few decades various SRGMs have been proposed in literature that are developed under the assumption that software reliability growth process depends only on testing-time. Later, testing coverage based SRGMs were also proposed in the literature. However, these models failed to take into account the simultaneous effect of Time and Testing Coverage on the growth process. Further, while modeling Software Reliability Models one of the most important factors that is to be considered is the fault detection rate (FDR). FDR helps in measuring the effectiveness of fault detection by test techniques and test cases. In practical situations it is dubious that the stability of the factor can be guaranteed during the whole process of software testing. Indeed, the characteristic of FDR is notably changed. When this change occurs that moment is termed as change point. In this paper we develop a two-dimensional model which measures the concurrent effect of time and testing coverage to remove the faulty sites dormant in the software in which the FDR is changed at change point. The model developed is based on the Cobb Douglas production function. The model developed is also validated on real data sets to show the comparison of the one dimensional time dependent model with proposed two dimensional time and coverage model.

KEYWORDS
Software Reliability Growth Model, Change Point, Cobb Douglas Production Function, Testing Coverage

1. INTRODUCTION
Software reliability engineering is centered on an extremely imperative software attribute which is termed as reliability. Software reliability is defined as the probability of failure-free software operation for a specified period of time in a specified environment [4]. It is one of the attributes of software quality which is commonly accepted as its key factor because it quantifies software failures - which can destroy a powerful system by making it inoperative. So, to quantify reliability, software is tested during the testing phase of software development life cycle.

Software testing is the method of raising the confidence that the software is free of flaws. But a major problem in testing software is that it cannot be made 100% bug free. This is not because programmers are careless or irresponsible, but because the nature of software code is complex and humans have only limited ability to manage complexity. Therefore, although testing helps in assessing and improving quality, but it cannot be performed indefinitely. Hence, time is considered as a very important controlling factor of testing phase. To monitor the relationship between the number of faults removed and time mathematically several Software Reliability Growth Models (SRGMs) are developed in literature [1,5,6,9,11]. But, the postulation that the software reliability growth process depends solitary on the testing-time as the software reliability growth factor essentially is not true. And, one of the other chief software reliability factors that affect the growth of software reliability is Testing Coverage (TC). TC is a significant measure for software developers as it helps in evaluating the quality of the tested software and determines how much additional effort is needed to improve the reliability of the software. Technically, it is defined as the ratio of the number of potential fault-sites sensitized by the test divided by the total number of potential fault-sites under consideration. There have been plenty of coverage measures proposed in literature, such as function coverage, statement coverage, branch coverage, data flow coverage, and so on. Different measures have their advantages and disadvantages when analyzing coverage [7,15]. However, in this paper we have restricted ourselves to statement coverage only. Statement coverage (also known as line coverage and basic block coverage) tells about each execution. It reports the total number of statements (blocks) that have been executed by the test data. Its great advantage is that it is insensitive to some control structures. A testing coverage based SRGM was proposed by Malaiya [16]. Inoue and Yamada [12] also developed SRGM with coverage which used a testing-coverage function to describe a time-dependent behavior or of a testing-coverage attainment process with the testing-skill of test-case designer.
Although, these above mentioned developed models proved to be more accurate than the only time governing SRGMs, but they failed to incorporate the concurrent effect of time and Testing Coverage. Recently, Inoue and Yamada [13] proposed a two dimensional software reliability growth model that considered the simultaneous effect of time and Testing-Coverage. However their modeling framework was not directly based on using mean value functions to represent of fault removal process. They discussed software reliability assessment method by using two dimensional Weibull-type SRGM. Another limitation of their proposed model was that it did not consider the phenomenon of change point.

One of the most important factors while modeling SRGMs is the fault detection rate (FDR). FDR helps in measuring the effectiveness of fault detection by test techniques and test cases. Many SRGMs assume that this detection rate remains same throughout the testing phase. However, in practical situations it is dubious that the stability of the factor can be guaranteed during the whole process of software testing. Indeed, the characteristic of the software failure-occurrence or the fault-detection phenomenon is notably changed. When this change occurs that moment is termed as change point. This would result in a software failure intensity function either increasing or decreasing monotonically [2]. The position of the Change Point can be judged by the graph of actual failure data.

In this paper we develop a two-dimensional model which measures the concurrent effect of time and testing coverage to remove the faults lying dormant in the software in which the FDR is changed at change point. The model developed is based on the Cobb-Douglas production function. The Cobb–Douglas functional form [14] of production functions is extensively used to characterize the relationship of an output to inputs. It was proposed by Knut Wicksell (1851–1926), and tested against statistical evidence by Charles Cobb and Paul Douglas in 1900–1928. The function of Cobb-Douglas present a simplified outlook of the economy in which production output is obtained by the amount of labor occupied and the amount of capital invested. While there are many factors influencing economic performance, their model demonstrated remarkable accuracy. The mathematical form of the production function is specified as:

\[ Y = AL^{1-v}K^{v} \]

Where: \( Y \) = total production (the monetary value of all goods produced in a year)

\( L \) = labor input

\( K \) = capital input

\( A \) = total factor productivity

\( v \) = elasticity of labor. This value is constant and is determined by available technology.

The paper is organized as follows: in section 2, two dimensional modeling framework incorporating the concurrent effect of time and coverage with change point is proposed. The parameter estimation and data validation of the formulated model is done in section 3. Finally, conclusions are drawn in section 4.

2. MODELING FRAMEWORK

2.1 NOTATIONS

<table>
<thead>
<tr>
<th>( a )</th>
<th>Initial number of faults in software</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s )</td>
<td>Testing time.</td>
</tr>
<tr>
<td>( u )</td>
<td>Testing Coverage.</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>Time Elasticity to fault removal</td>
</tr>
<tr>
<td>( m(s, u) )</td>
<td>Cumulative number of faults removed by time ( s ) and with coverage ( u )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>Change Point; ( s^\alpha u^{1-\alpha} ) 0 ( \leq \alpha \leq 1 )</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>Change Point; ( s_0^\alpha u_0^{1-\alpha} ) 0 ( \leq \alpha \leq 1 )</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>Fault detection rate per remaining fault before change point.</td>
</tr>
<tr>
<td>( b_2 )</td>
<td>Fault detection rate per remaining fault after change point.</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>Constant before change point</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>Constant after change point</td>
</tr>
</tbody>
</table>

2.2 BASIC ASSUMPTIONS

The two dimensional model proposed in our work incorporating testing time and coverage is based on Non Homogeneous Poisson Process (NHPP). Let \( \{N(s, u), s \geq 0, u \geq 0\} \) be a two-dimensional stochastic process representing the cumulative number of software failures by time \( s \) and with the coverage of \( u \). A two-dimensional NHPP with a mean value function \( m(s, u) \) is formulated as:

\[ \Pr(N(s, u) = n) = \frac{(m(s, u))^n}{n!} \exp((-m(s, u))) ; n = 0, 1, 2… \]

The other assumptions of our model are:

1. Software is subject to failures during execution caused by faults remaining in the software.
2. On a failure, the fault causing that failure is immediately removed and no new faults are introduced.
3. Fault detection rate changes at \( \lambda_0 \).
4. To cater the combined effect of testing time and coverage we use Cobb-Douglas production function of the following form:

\[ \lambda = s^\alpha u^{1-\alpha} \] 0 \( \leq \alpha \leq 1 \] (1)

5. The economy of scale (i.e. \( \alpha \)) remains same before and after change point

2.3 MODEL DEVELOPMENT
Under the above assumptions the differential equation representing the rate of change of cumulative number of faults detected w.r.t. to time and usage is given as:

$$m'(\lambda) = b(\lambda)(a - m(\lambda))$$  
(2)  
Where

$$b(\lambda) = \begin{cases} 1 + \beta_1 \exp(-b_1,1) & \text{for } \lambda \leq \lambda_0 \\ 1 + \beta_2 \exp(-b_2,1) & \text{for } \lambda > \lambda_0 \end{cases}$$  
(3)

Case 1: For $\lambda \leq \lambda_0$

Solving (2) with the initial condition $m(\lambda = 0) = 0$ and using equation (1) we get

$$m(s,u) = a \left[ 1 - \frac{1 + \beta_1}{1 + \beta_1 \exp(-b_1,(s^u)^{-\alpha})} \exp(-b_1,(s^u)^{-\alpha}) \right]$$  
(4)

Case 2: For $\lambda > \lambda_0$

Solving (2) with the initial condition $m(\lambda = \lambda_0) = m(\lambda_0)$ and using equation (1) we get

$$m(s,u) = a \left[ 1 - \frac{(1 + \beta_1)(1 + \beta_2 \exp(-b_2,(s^u)^{-\alpha}))}{(1 + \beta_1)(1 + \beta_2 \exp(-b_2,(s^u)^{-\alpha})) \exp(-b_2,(s^u)^{-\alpha})} \right]$$  
(5)

Combining Case 1 and Case 2 we get $m(s,u)$ as:

$$m(s,u) = a \left[ 1 - \frac{1 + \beta_1}{1 + \beta_1 \exp(-b_1,(s^u)^{-\alpha})} \exp(-b_1,(s^u)^{-\alpha}) \right] \quad \text{for } \lambda \leq \lambda_0$$  
(6)  
$$m(s,u) = a \left[ 1 - \frac{(1 + \beta_1)(1 + \beta_2 \exp(-b_2,(s^u)^{-\alpha}))}{(1 + \beta_1)(1 + \beta_2 \exp(-b_2,(s^u)^{-\alpha})) \exp(-b_2,(s^u)^{-\alpha})} \right] \quad \text{for } \lambda > \lambda_0$$

2.4 RELIABILITY EVALUATION

Software evaluation is a very significant phenomenon in quantitative software reliability assessment. The software reliability function signifies the probability that a software failure does not occur in time-interval $(t, t + x)$ $(t > 0, x > 0)$ given that the testing team or the user operation has been going up to time $t$. In two dimensional SRGM, we can assess software reliability in an operation phase where we assume that the testing coverage is not expanded. We can derive the probability that the software failure does not occur in time-interval $[s_\pi, s_\pi + \omega]$ $(s_\pi > 0, \omega > 0)$ given that testing has been going up to $s_\pi$ and the value of testing coverage has been attained up to $u_\pi$ by testing termination time $s_\pi$ as:

$$R(s, u) = \exp \left\{ \left[ m(s + \omega, s, u) \right] / K - m(s, u) \right\}$$  
(7)

Where $K$ indicates the set of parameter estimates of a two-dimension SRGM

3. PARAMETER ESTIMATION AND MODEL VALIDATION

The success of mathematical modeling approach to reliability evaluation depends heavily upon quality of failure data collected. The parameters of the SRGMs are estimated based upon these data. Hence, efforts should be made to make the data collection more explicit and scientific. Usually data is collected in one of the following two ways. In the first case the times between successive failures are recorded. Though this type of data collection is more desirable, it may not be simple. Complication can arise in measuring the testing effort for each fault and it may not be very convenient to note the time at each failure report. The other easier and commonly collected data type is known as the grouped data. Here testing intervals are specified and number of failures experienced during each such interval is noted. The proposed model presented in this paper is non-linear and presents extra problems in estimating the parameters. Technically, it is more difficult to find the solution for non-linear models using Least Square method and requires numerical algorithms to solve it. Statistical software packages such as SPSS, which has been used here for estimating the parameters helps to overcome this problem.

We have carried out the parameter estimation on two data sets. First data set (DS-1) is cited in Malaiya et al [16]. The data set is Coverage Data set with 796 test cases and 9 cumulative numbers of faults removed with block coverage as 95.99%. Second data set (DS-2) is also from Malaiya et al [16]. The coverage data set consists of 9 cumulative faults removal with 1196 test case covering 95.97% of the block coverage. To check the performance of the model estimates we have compared the results with the traditional time dependent flexible SRGM by Kapur and Garg[10]. The goodness of fit measures used are Mean Square Error (MSE), and Coefficient of multiple determination ($R^2$).

DS-1

The Change point for DS-1 is obtained at (44, 87) (see figure 3.1).
The parameter estimates results for DS-1 is tabulated in table 3.1. The comparison result for the data is shown in table 3.2. The 3-D graph of goodness of fit depicting the estimated cumulative number of faults with respect to time and coverage is shown in figure 3.2.

<table>
<thead>
<tr>
<th>Proposed Two Dimensional Model</th>
<th>One Dimensional KG Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a</strong></td>
<td>10</td>
</tr>
<tr>
<td><strong>b/a</strong></td>
<td>0.031884</td>
</tr>
<tr>
<td><strong>b_2</strong></td>
<td>0.02248</td>
</tr>
<tr>
<td><strong>β_1/β</strong></td>
<td>1.028487</td>
</tr>
<tr>
<td><strong>β_2</strong></td>
<td>0.001</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.523349</td>
</tr>
</tbody>
</table>

Table 3.1 Parameter Estimates for DS-1

<table>
<thead>
<tr>
<th>Proposed Two Dimensional Model</th>
<th>One Dimensional KG Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R^2</strong></td>
<td>0.991</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td>0.06847</td>
</tr>
</tbody>
</table>

Table 3.2 Goodness of Fit Measures for DS-1

Figure 3.2 Goodness of Fit Curve(DS-1)

**DS-1**

The Change point for DS-1 is obtained at (20, 70.5) (see figure 3.3).

The parameter estimates results and goodness of fit measures for DS-2 is tabulated in table 3.3 and 3.4 respectively. The 3-D graph of goodness of fit depicting the estimated cumulative number of faults with respect to time and coverage is shown in figure 3.3.

<table>
<thead>
<tr>
<th>Proposed Two Dimensional Model</th>
<th>One Dimensional KG Model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>a</strong></td>
<td>10</td>
</tr>
<tr>
<td><strong>b/a</strong></td>
<td>0.016475</td>
</tr>
<tr>
<td><strong>b_2</strong></td>
<td>0.022992</td>
</tr>
<tr>
<td><strong>β_1/β</strong></td>
<td>1.576795</td>
</tr>
<tr>
<td><strong>β_2</strong></td>
<td>2.074583</td>
</tr>
<tr>
<td><strong>α</strong></td>
<td>0.288297</td>
</tr>
</tbody>
</table>

Table 3.3 Parameter Estimates for DS-2

Figure 3.2 Actual Failure Data Set (DS-2)

**Figure 3.2 Actual Failure Data Set (DS-1)**

**Figure 3.2 Goodness of Fit Curve(DS-1)**
Table 3.3 Parameter Estimates for DS-2

<table>
<thead>
<tr>
<th></th>
<th>Proposed Two Dimensional Model</th>
<th>One Dimensional KG Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.973</td>
<td>0.947</td>
</tr>
<tr>
<td>MSE</td>
<td>0.2094714</td>
<td>0.4177143</td>
</tr>
</tbody>
</table>

Table 3.4 Goodness of Fit Measures for DS-2

Figure 3.3 Goodness of Fit Curve (DS-2)

Based on the comparison results (large $R^2$, large adjusted $R^2$, small bias and small MSE) it is observed that the proposed model is better than its counterpart in one dimension. This is due to the fact that the estimates are concurrently based on testing time and testing coverage.

CONCLUSION

For software development cycle to remain on track software testing phase plays a crucial role. And the two major factors among several software reliability factors like the test execution-time, the testing-skill, the testing coverage, and so forth that govern the pace of testing progress are testing time and coverage. Therefore their effect together needs to be incorporated in developing an accurate Software Reliability Growth Model. To capture this concurrent effect in determining the cumulative number of faults removed from software, we have proposed in this paper a two dimensional modeling framework. The proposed model reflects a broader framework which accounts for interaction between two dimensions of software reliability metrics namely testing time and coverage. The proposed two dimensional model is based on the Cobb Douglas production function. The model is validated on two real software data sets. The results on goodness of fit measures of the proposed model indicate that our approach outperforms when compared with traditional time dependent flexible SRGM.

FUTURE SCOPE

In this paper it is assumed that the model is developed under the perfect debugging environment. The overcoming of this limitation in modeling forms a scope of future research. We have focused only on two dimensional framework in this work. However, it is known that a software reliability growth process in a testing-phase is influenced by the following several software reliability factors: the test execution-time, the testing-skill and so forth. In future we can explore the possibility of including multi dimensional software reliability growth modeling so as to take care the effect of not only testing coverage but also other testing factors like testing effort, testing time/number of test cases on the fault removal process simultaneously. One of the functional forms that can be used for this purpose is the multi dimension extension of Cobb Douglas function given by:

$$\tau \equiv s_1^{\alpha_1} s_2^{\alpha_2} \ldots s_n^{\alpha_n}$$

where $\sum_{i=1}^{n} \alpha_i = 1$ ; $\alpha_i \geq 0 \ \forall i$

REFERENCES


