Chapter-3

Unified Framework for modeling Fault removal using Change point phenomenon

Model unification is an insightful investigation for the study of general models without making many assumptions. In the literature various software reliability models have been proposed incorporating change-point concept. To the best of our knowledge these models have been developed separately. This chapter is based on unification schemes involving change in fault removal rate at some point of time. In the first section we have proposed a general framework for deriving several software reliability growth models with change-point concept based on non-homogeneous Poisson process (NHPP). The proposed framework helps in assessment of already existing change-point models along with three new models. Real data sets have been used for the validation of the models presented. In the next Section we have considered fault detection-correction as separate processes in software reliability growth modeling. We have proposed a general framework for deriving several software reliability growth models for fault detection-correction process incorporating change-point concept based on non-homogeneous Poisson process (NHPP). Some existing change-point models along with new models have been derived from the proposed general framework. The performance and application of the proposed models has been demonstrated using real data set. Software statistical package for social sciences (SPSS) has been used to estimate the unknown parameters. Comparison criteria of MSE, $R^2$, Bias, variation and Root mean squared prediction error are used to validate the proposed models.

This chapter is based on the following papers:


3.1 A Unified Modeling Framework Incorporating Change-Point for Measuring Reliability Growth during Software Testing

The plethora of SRGM makes model selection a tedious task. To reduce this difficulty, unified modeling approaches have been proposed by many researchers. These schemes prove be very successful for obtaining several existing SRGM without making many assumptions. There are some, but only a few, model unification schemes in the literature. The work done in this area has been discussed in the first chapter in detail. Apart from the probabilistic approach, Huang et al. (Huang 2003) explained the deterministic behavior of the NHPP models, namely mean value function of time by introducing several kinds of mean operation. Pham et al. (1997) solved a generalized differential equation by which the mean value function in the NHPP model is governed and proposed an NHPP with a generalized mean value parameter. Inoue et al. (2002) applied infinite server queuing theory to the basic assumptions of delayed S-shaped SRGM i.e. fault detection phenomenon consists of successive failure detection and isolation processes considering time distribution of fault isolation process and obtained several NHPP models describing FDP (Fault Detection Process) as a two-stage process. Dohi et al. (2004) developed a unification method for NHPP models describing test input and program path searching times stochastically by an infinite server queuing model. Kapur et. al. (2008) discussed unification scheme based on Cumulative Distribution Function for the detection/correction times incorporating two types of imperfect debugging. Recently, Kapur et. al. (2009) discussed three different approaches to unify a wide range of software reliability growth models. Though these three schemes have been derived under different sets of assumptions with varied backgrounds but they are proved to be mathematically equivalent.

Using the concept of change point we have derived some reliability growth models in this section.

3.1.1 Notations

\( m(t) \) the mean value function or the expected number of faults detected or removed by time \( t \).

\( a \) constant, representing the initial number of faults lying dormant in the software when the testing starts.

\( b \) fault removal/correction rate.

\( \lambda(t) \) intensity function for NHPP models or fault detection rate per unit time.

\( F(t) \) distribution functions for fault removal/correction times.
A Generalized Modeling Framework in Software Reliability and Related Problems

\[ s(t) \] time dependent fault removal/correction rate per remaining faults.

\[ k, \beta \] constant, learning parameter in logistic function.

\[ \alpha, \lambda \] Shape and scale parameters before change-point and \( \alpha_2, \lambda_2 \) are shape and scale parameters after change-point for gamma distribution.

\[ \mu, \sigma^2 \] Mean and variance before change-point and \( \mu_2, \sigma^2_2 \) mean and variance after change-point for normal distribution.

3.1.2 Basic Assumptions

The intensity function \( \lambda(t) \) (or the mean value function \( m(t) \)) is the basic building block of all the NHPP models existing in the software reliability engineering literature.

The proposed models are based upon the following basic assumptions:

1. Fault removal phenomenon is modeled by an NHPP.
2. Software is subject to failures during execution caused by faults remaining in the software.
3. Each time a failure is observed, an immediate debugging effort takes place to find the cause of the failure in order to remove it.
4. The fault detection rate may change at some time moment (called change-point).
5. Fault removal process is not incorporating any kind of imperfect debugging.

3.1.3. Model Development

Various time domain models have appeared in the literature, which describe the stochastic failure process by an NHPP. These models differ in their failure intensity function \( \lambda(t) \), and hence \( m(t) \). The NHPP models can be further classified into finite failure and infinite failure categories. Finite failure NHPP models assume that the expected number of faults detected given infinite amount of testing time will be finite, whereas the infinite failures models assume that an infinite numbers of faults would be detected in infinite testing time Farr (1996). Let ‘\( a \)’ denote the expected number of faults that would be detected given infinite testing time in case of finite failure NHPP models. Then, the mean value function of the finite failure NHPP models can also be written as (Gokhale et al 1996, Musa et al 1987):
\[ m(t) = aF(t) \]  \hspace{1cm} (3.1)

where \( F(t) \) is a distribution function.

From Equation (3.1), the instantaneous failure intensity \( \lambda(t) \) in case of the finite failure NHPP models is given by:

\[ \lambda(t) = aF'(t) \]  \hspace{1cm} (3.2)

The above equation can be rewritten as:

\[ \lambda(t) = [a - m(t)] \frac{F'(t)}{1 - F(t)} = [a - m(t)] s(t) \]  \hspace{1cm} (3.3)

Here \( s(t) \) is the fault detection/removal rate per remaining fault of the software, or the rate at which the individual faults manifest themselves as failures during testing. \([a - m(t)]\) denotes the expected number of faults remaining in the software at time \( t \).

Since \([a - m(t)]\) is monotonically non-increasing function of time, the nature of the overall failure intensity, \( \lambda(t) \), is governed by the nature of failure occurrence rate per fault \( s(t) \).

Now, Equation (3.3) can be written as:

\[ m'(t) = \frac{dm(t)}{dt} = [a - m(t)] \frac{F'(t)}{1 - F(t)} = [a - m(t)] s(t) \]  \hspace{1cm} (3.4)

Incorporating change-point concept, \( s(t) \) i.e. fault detection/removal rate per remaining fault of the software can be written as

\[ s(t) = \begin{cases} \frac{F_1'(t)}{1 - F_1(t)} & t \leq \tau \\ \frac{F_2'(t)}{1 - F_2(t)} & t > \tau \end{cases} \]

where \( F_1'(t) = \frac{d}{dt} F_1(t) \) and \( F_2'(t) = \frac{d}{dt} F_2(t) \)

Further, incorporating the change-point concept in modeling, Equation (3.4) becomes:
For \( 0 \leq t \leq \tau \)

\[
\frac{dm(t)}{dt} = [a - m(t)] \frac{F_1(t)}{1 - F_1(t)}
\]

Solving the above equation with initial condition at \( t=0 \), \( m(0)=0 \), we get

\[ m(t) = aF_1(t) \]

For \( t > \tau \)

\[
\frac{dm(t)}{dt} = [a - m(t)] \frac{F_2(t)}{1 - F_2(t)}
\]

Solving above equation with initial condition at \( t=\tau \), \( m(\tau) = m(\tau) \), we get

\[ m(t) = a \left[ 1 - \frac{(1 - F_1(\tau))(1 - F_2(t))}{(1 - F_2(\tau))} \right] \]  

(3.5)

i.e.

\[
m(t) = \begin{cases} 
aF_1(t) & \text{for} \ t \leq \tau \\
a \left[ 1 - \frac{(1 - F_1(\tau))(1 - F_2(t))}{(1 - F_2(\tau))} \right] & \text{for} \ t > \tau 
\end{cases}
\]

(3.6)

We now show the derivation of some of the existing models incorporating change-point concept from the proposed unified approach. Also, we will propose some new models incorporating change-point on the similar lines.

**SRGM-1**

The following exponential distribution function is used to model SRGM-1:

\[ F_1(t) = 1 - \exp(-bt) \quad \text{for} \ t \leq \tau \]  

(3.7)

and
Chapter 3

Unified Framework for modeling Fault removal using Change point phenomenon

Department of Operational Research, University of Delhi

95

\( F_2(t) = 1 - \exp \left( -b_2 t \right) \) for \( t > \tau \) \hspace{1cm} (3.8)

Substituting the value of \( F_1(t) \) and \( F_2(t) \) from Equation (3.7) and (3.8) into Equation (3.6), we get:

\[
m(t) = \begin{cases} a \left[ 1 - \exp \left( -b_1 t \right) \right] & \text{for } 0 \leq t \leq \tau \\ a \left[ 1 - \exp \left( -b_1 \tau - b_2 (t - \tau) \right) \right] & \text{for } t > \tau \end{cases}
\]

Above model has been discussed by Shyur (2003).

**SRGM-2**

Let \( F(t) \) define a two-stage Erlangian distribution function i.e.,

\( F_1(t) = 1 - (1 + b_1 t) \exp \left( -b_1 t \right) \) for \( t \leq \tau \) \hspace{1cm} (3.10)

and

\( F_2(t) = 1 - (1 + b_2 t) \exp \left( -b_2 t \right) \) for \( t > \tau \) \hspace{1cm} (3.11)

Substituting the value of \( F_1(t) \) and \( F_2(t) \) from Equation (3.10) and (3.11) into Equation (3.6), we get:

\[
m(t) = \begin{cases} a \left[ 1 - (1 + b_1 t) \exp \left( -b_1 t \right) \right] & \text{for } 0 \leq t \leq \tau \\ a \left[ 1 - \frac{1 + b_1 \tau}{1 + b_2 \tau} (1 + b_2 t) \exp \left( -b_1 \tau - b_2 (t - \tau) \right) \right] & \text{for } t > \tau \end{cases}
\]

Above model has been discussed by Archana (2007).

**SRGM-3**

Let \( F(t) \) define a logistic distribution function i.e.,
\[ F_1(t) = \frac{(1 - \exp(-b_1 t))}{(1 + \beta \exp(-b_1 t))} \quad \text{for } t \leq \tau \quad (3.13) \]

and

\[ F_2(t) = \frac{(1 - \exp(-b_2 t))}{(1 + \beta \exp(-b_2 t))} \quad \text{for } t > \tau \quad (3.14) \]

In the above distribution function it is assumed that \( \beta \) remains same before and after change-point for simplicity.

Substituting the value of \( F_1(t) \) and \( F_2(t) \) from Equation (3.13) and (3.14) into Equation (3.6), we get:

\[
m(t) = \begin{cases} 
\left[ a \left( 1 - \frac{(1 + \beta) \exp(-b_1 t)}{(1 + \beta \exp(-b_1 t))} \right) \right] & \text{for } 0 \leq t \leq \tau \\
\left[ a \left( 1 - \frac{(1 + \beta)(1 + \beta \exp(-b_2 \tau)) \exp(-b_1 \tau - b_2 (t - \tau))}{(1 + \beta \exp(-b_1 \tau))(1 + \beta \exp(-b_2 \tau))} \right) \right] & \text{for } t > \tau 
\end{cases}
\]

(3.15)

This is same as change-point model given by Kapur et al. (2006)

Above three software reliability growth model SRGM-1, SRGM-2 and SRGM-3 are existing models and special cases of the generalized framework.

Now, we derive some new models based on the proposed framework.

The mean value functions \( m(t) \) corresponding to different forms of distribution functions \( F(t) \) are summarized in Table 3.1.
Table 3.1: Mean value functions for various distributions

<table>
<thead>
<tr>
<th>Model</th>
<th>Distribution Function (F(t))</th>
<th>Mean Value Function m(t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGM-4</td>
<td>( T \sim \text{Wei}(b, k) ) (Weibull Distribution)</td>
<td>( m(t) = \begin{cases} \left[ a \left[ 1 - \exp \left( -b_1 t^k \right) \right] \right] &amp; \text{for } 0 \leq t \leq \tau \ \left[ a \left[ 1 - \exp \left( -b_1 t^k - b_2 (t^k - \tau^k) \right) \right] \right] &amp; \text{for } t &gt; \tau \end{cases} )</td>
</tr>
<tr>
<td>SRGM-5</td>
<td>( T \sim \phi(\mu, \sigma^2) ) (Normal Distribution)</td>
<td>( m(t) = \begin{cases} a \phi(t, \mu_1, \sigma_1^2) &amp; \text{for } 0 \leq t \leq \tau \ a \left[ \frac{1 - \phi(t, \mu_1, \sigma_1^2)}{1 - \phi(t, \mu_2, \sigma_2^2)} - \phi(t, \mu_2, \sigma_2^2) \right] &amp; \text{for } t &gt; \tau \end{cases} )</td>
</tr>
<tr>
<td>SRGM-6</td>
<td>( T \sim \Gamma(\alpha, \lambda) ) (Gamma Distribution)</td>
<td>( m(t) = \begin{cases} a \Gamma(t, \alpha_1, \lambda_1) &amp; \text{for } 0 \leq t \leq \tau \ a \left[ \frac{1 - \Gamma(t, \alpha_1, \lambda_1)}{1 - \Gamma(t, \alpha_2, \lambda_2)} - \Gamma(t, \alpha_2, \lambda_2) \right] &amp; \text{for } t &gt; \tau \end{cases} )</td>
</tr>
</tbody>
</table>

3.1.4 Model Validation, Comparison Criteria and Data Analyses

To illustrate the estimation procedure and application of the SRGM (existing as well as proposed) we have carried out the data analysis of real software data set. The parameters of the models have been estimated using statistical package SPSS and the change-point of the data sets have been judged by using change-point analyzer.

Data set 1(DS-1)

The first data set (DS-1) had been collected during 35 months of testing a radar system of size 124 KLOC and 1301 faults were detected during testing. This data is cited from Brooks and Motley (1980). The change-point for this data set is 17th month.

The parameter estimation and comparison criteria results for DS-1 of all the models under consideration can be viewed through Table 3.2 and Table 3.3. It is clear from the
table that the value of $R^2$ for SRGM-3 and SRGM-5 are higher and value of MSE, Bias, Variation and RMSPE are lower in comparison with other models and provides better goodness of fit for DS-1.

**Data set 2 (DS-2)**

The second data set (DS-2) had been collected during 19 weeks of testing a real time command and control system and 328 faults were detected during testing. This data is cited from Ohba (1984). The change-point for this data set is 6th week.

The parameter estimation and comparison criteria results for DS-2 of all the models under consideration can be viewed through Table 3.4 and Table 3.5. It is clear from the table that the value of $R^2$ for SRGM-3 and SRGM-5 are higher and value of MSE, Bias, Variation and RMSPE are lower in comparison with other models and provides better goodness of fit for DS-2.

**Table 3.2 : Model Parameter Estimation Results (DS-1)**

<table>
<thead>
<tr>
<th>Models</th>
<th>$a$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\sigma_1$</th>
<th>$\sigma_2$</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$k$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGM-1</td>
<td>2705</td>
<td>.019</td>
<td>.023</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRGM-2</td>
<td>1639</td>
<td>.092</td>
<td>.095</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRGM-3</td>
<td>1321</td>
<td>.213</td>
<td>.211</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRGM-4</td>
<td>1303</td>
<td>.0007</td>
<td>.0007</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRGM-5</td>
<td>1298</td>
<td>-</td>
<td>-</td>
<td>11</td>
<td>15</td>
<td>5</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SRGM-6</td>
<td>1350</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>.683</td>
<td>4.816</td>
<td>.031</td>
<td>.281</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 3.3: Model Comparison Results (DS-1)**

<table>
<thead>
<tr>
<th>Models</th>
<th>$R^2$</th>
<th>MSE</th>
<th>Bias</th>
<th>Variation</th>
<th>RMSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGM-1</td>
<td>.97367</td>
<td>5608.09</td>
<td>-1.28E-05</td>
<td>75.9804</td>
<td>75.9804</td>
</tr>
<tr>
<td>SRGM-2</td>
<td>.98833</td>
<td>2486.007</td>
<td>-3.89E-06</td>
<td>50.5877</td>
<td>50.5877</td>
</tr>
<tr>
<td>SRGM-3</td>
<td>.99930</td>
<td>149.5196</td>
<td>-4.96E-07</td>
<td>12.4063</td>
<td>12.4063</td>
</tr>
<tr>
<td>SRGM-4</td>
<td>.99872</td>
<td>271.789</td>
<td>-22.08</td>
<td>25.0093</td>
<td>33.3660</td>
</tr>
<tr>
<td>SRGM-5</td>
<td>.99922</td>
<td>166.526</td>
<td>-0.0001</td>
<td>13.0928</td>
<td>13.0928</td>
</tr>
<tr>
<td>SRGM-6</td>
<td>.99700</td>
<td>731.7273</td>
<td>-2.85E-07</td>
<td>23.9072</td>
<td>23.9072</td>
</tr>
</tbody>
</table>
### Table 3.4: Model Parameter Estimation Results (DS-2)

<table>
<thead>
<tr>
<th>Models</th>
<th>a</th>
<th>b_1</th>
<th>b_2</th>
<th>μ_1</th>
<th>μ_2</th>
<th>σ_1</th>
<th>σ_2</th>
<th>α_1</th>
<th>α_2</th>
<th>λ_1</th>
<th>λ_2</th>
<th>k</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGM-1</td>
<td>614</td>
<td>.040</td>
<td>.044</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SRGM-2</td>
<td>401</td>
<td>.189</td>
<td>.169</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SRGM-3</td>
<td>365</td>
<td>.232</td>
<td>.217</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>SRGM-4</td>
<td>372</td>
<td>.022</td>
<td>.018</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.625</td>
<td></td>
</tr>
<tr>
<td>SRGM-5</td>
<td>351</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>7</td>
<td>1</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>SRGM-6</td>
<td>411</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1.349</td>
<td>1.872</td>
<td>.103</td>
<td>.15</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.5: Model Comparison Results (DS-2)

<table>
<thead>
<tr>
<th>Models</th>
<th>R²</th>
<th>MSE</th>
<th>Bias</th>
<th>Variation</th>
<th>RMSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGM-1</td>
<td>.98812</td>
<td>122.6662</td>
<td>-6.55E-07</td>
<td>11.3789</td>
<td>11.3789</td>
</tr>
<tr>
<td>SRGM-2</td>
<td>99054</td>
<td>97.61703</td>
<td>5.32E-06</td>
<td>10.1508</td>
<td>10.1508</td>
</tr>
<tr>
<td>SRGM-3</td>
<td>99270</td>
<td>75.29769</td>
<td>3.40E-06</td>
<td>8.9152</td>
<td>8.9152</td>
</tr>
<tr>
<td>SRGM-4</td>
<td>99143</td>
<td>88.41156</td>
<td>-2.85E-05</td>
<td>9.6603</td>
<td>9.6603</td>
</tr>
<tr>
<td>SRGM-5</td>
<td>.99300</td>
<td>72.28211</td>
<td>0.0017</td>
<td>8.7348</td>
<td>8.7348</td>
</tr>
<tr>
<td>SRGM-6</td>
<td>.99100</td>
<td>97.23347</td>
<td>-0.0646</td>
<td>10.1311</td>
<td>10.1313</td>
</tr>
</tbody>
</table>
Figure 3.1: Goodness of Fit Curve for DS-1

Figure 3.2: Goodness of Fit Curve for DS-2
3.2 Generalized Modeling Framework for Fault Detection-Correction Process Incorporating Change-point

A common assumption of SRGM is that the detected faults can be removed instantly, while the process to correct these faults is ignored. Firstly Yamada et al. (1983) described testing process as a two-stage process. On the other hand Schneidewind (1975) described the testing process as a two-stage process in which all observed/detected faults are removed/corrected after a constant delay of time. Lo and Huang (2005) has proposed a general framework where some existing NHPP models are re-evaluated from the viewpoint of correction process. Further Xie et al. (2006) emphasized on fault correction process described by delayed detection process with a random or deterministic delay.

3.2.1 Notation

\( m(t) \) the mean value function or the expected number of faults detected or removed by time \( t \).

\( a \) constant, representing the initial number of faults lying dormant in the software when the testing starts.

\( b_1 \) fault detection/correction rate before change-point.

\( b_2 \) fault detection/correction rate after change-point.

\( \lambda(t) \) intensity function for NHPP models or fault detection rate per unit time.

\( F_1(t), G_1(t) \) distribution functions for failure detection and fault correction times respectively before change-point

\( F_2(t), G_2(t) \) distribution functions for failure detection and fault correction times respectively after change-point

\( s(t) \) time dependent fault removal/correction rate per remaining faults

\( k_1, \beta_1 \) constant, learning parameter in logistic function before change-point.

\( K_2, \beta_2 \) constant, learning parameter in logistic function before change-point.

* Convolution.

\( \otimes \) Steiltjes convolution.
3.2.2 Basic Assumptions

The intensity function $\lambda(x)$ (or the mean value function $m(t)$) is the basic building block of all the NHPP models existing in the software reliability engineering literature.

The proposed models are based upon the following basic assumptions:

1. Fault detection-correction process is modeled by an NHPP.
2. Software is subject to failures during execution caused by faults remaining in the software.
3. Each time a failure is observed, an immediate debugging effort takes place to find the cause of the failure in order to remove it.
4. The fault detection-correction rate may be change at some time moment (called change-point).
5. Fault removal process is not incorporating any kind of imperfect debugging.

3.2.3 Model Development

Let ‘$a$’ denote the expected number of faults that would be detected given infinite testing time in case of finite failure NHPP models. Then, the mean value function of the finite failure NHPP models can also be written as (Musa et al 1987):

$$m(t) = aF(t) \quad \text{(3.16)}$$

Where $F(t)$ is a distribution function.

Now, in order to incorporate the two testing processes i.e. failure detection process defined by $F(t)$ distribution function and, fault correction process defined by $G(t)$ distribution function in a unified or generalized modeling framework, Equation (3.16) can be modified as (which can be proved on the lines of Musa et al. (1987):

$$m(t) = a(F \otimes G)(t) \quad \text{(3.17)}$$

Hence,

$$\frac{dm(t)}{dt} = a(f * g)(t) \quad \text{(3.18)}$$

The above Equation (3.18) can be rewritten as follows:
\[
\frac{dm(t)}{dt} = \frac{(f * g)(t)}{1 - (F \otimes G)(t)}\left[a - m(t)\right]
\]  
\[
\frac{dm(t)}{dt} = s(t)\left[a - m(t)\right]
\]

Where \( s(t) = \frac{(f * g)(t)}{1 - (F \otimes G)(t)} \) is the failure observation/detection-fault removal/correction rate.

Incorporating change-point concept, i.e. fault detection/removal rate per remaining fault of the software can be written as

\[
s(t) = \begin{cases} 
\frac{(f_1 * g_1)(t)}{1 - (F_1 \otimes G_1)(t)} & t \leq \tau \\
\frac{(f_2 * g_2)(t)}{1 - (F_2 \otimes G_2)(t)} & t > \tau 
\end{cases}
\]

where \( f_1(t) = \frac{d}{dt}F_1(t) \) and \( f_2(t) = \frac{d}{dt}F_2(t) \)

\( g_1(t) = \frac{d}{dt}G_1(t) \) and \( g_2(t) = \frac{d}{dt}G_2(t) \)

Further, incorporating the change-point concept in modeling, Equation (3.19) becomes:

For \( 0 \leq t \leq \tau \)

\[
\frac{dm(t)}{dt} = [a - m(t)]\frac{(f_1 * g_1)(t)}{1 - (F_1 \otimes G_1)(t)}
\]

Solving the above equation with initial condition at \( t=0, m(0)=0 \), we get

\( m(t) = a(F_1 \otimes G_1)(t) \)

For \( t > \tau \)

\[
\frac{dm(t)}{dt} = [a - m(t)]\frac{(f_2 * g_2)(t)}{1 - (F_2 \otimes G_2)(t)}
\]
Solving the above equation with initial condition at \( t = \tau \), \( m(t) = m(\tau) \), we get

\[
m(t) = a \left[ 1 - \frac{(1 - (F \otimes G_1)(\tau))(1 - (F_1 \otimes G_2)(\tau))}{(1 - (F_2 \otimes G_2)(\tau))} \right]
\]

(3.20)

i.e.

\[
m(t) = \begin{cases} 
  a(F \otimes G_1)(t) & \text{for } t \leq \tau \\
  a \left[ 1 - \frac{(1 - (F \otimes G_1)(\tau))(1 - (F_1 \otimes G_2)(\tau))}{(1 - (F_2 \otimes G_2)(\tau))} \right] & \text{for } t > \tau
\end{cases}
\]

(3.21)

The mean value functions \( m(t) \) corresponding to different forms of distribution functions \( F(t) \) and \( G(t) \) incorporating change-point are summarized below.

<table>
<thead>
<tr>
<th>Model</th>
<th>Distribution Function (F(t))</th>
<th>Mean Value Function m(t)</th>
</tr>
</thead>
</table>
| SRGM-1 | \( F_1 : T \sim \exp(b_1) \) \( G_1 : T \sim 1(t) \) \( \text{for } t \geq \tau \) \( F_2 : T \sim \exp(b_2) \) \( G_2 : T - 1(t) \) \( \text{for } t > \tau \) | \( m(t) = \begin{cases} 
  a[1 - \exp(-b_1)] & \text{for } 0 \leq t \leq \tau \\
  a[1 - \exp(-b_1 + b_2 (t - \tau))] & \text{for } t > \tau
\end{cases} \) Shyur (2003) |
|--------|-----------------------------|--------------------------|
| SRGM-2 | \( F_1 : T \sim \exp(b_1) \) \( G_1 : T \sim \exp(b_1) \) \( \text{for } t \leq \tau \) \( F_2 : T - \exp(b_2) \) \( G_2 : T - 1(t) \) \( \text{for } t > \tau \) | \( m(t) = \begin{cases} 
  a[1 - (1 + b_1 t)\exp(-b_1 t)] & \text{for } 0 \leq t \leq \tau \\
  a \left[ \left( \frac{1}{1 + b_2 t} \right)(1 + b_2 t)\exp(-b_1 t(1 - \tau)) \right] & \text{for } t > \tau
\end{cases} \) Archana (2007) |
|--------|-----------------------------|--------------------------|
| SRGM-3 | \( F_1 : T \sim \text{We}i(b_1, k_1) \) \( G_1 : T - 1(t) \) \( \text{for } t \leq \tau \) \( F_2 : T - \exp(b_2) \) \( G_2 : T - \exp(b_2) \) \( \text{for } t > \tau \) | \( m(t) = \begin{cases} 
  a(1 - \exp(-b_1 t^k)) & \text{for } 0 \leq t \leq \tau \\
  a[1 - \exp(-b_1 t^k - b_2 (t^2 - \tau^2))] & \text{for } t > \tau
\end{cases} \) |

Department of Operational Research, University of Delhi
Chapter 3
Unified Framework for modeling Fault removal using Change point phenomenon

For SRGM-3 and SRGM-4 if the value of k and $\beta$ remain same before and after change-point, models reduce to Kapur et al. (2006, 2008).

3.2.4 Model Validation, Comparison Criteria and Data Analyses
To illustrate the estimation procedure and application of the SRGM (existing as well as proposed) we have carried out the data analysis of real software data set. The parameters of the models have been estimated using statistical package SPSS and the change-point of the data sets have been judged by using change-point analyzer.

Data set 1(DS-1)
The first data set (DS-1) had been collected during 35 months of testing a radar system of size 124 KLOC and 1301 faults were detected during testing. This data is cited from Brooks and Motley (1980). The change-point for this data set is 17th month.
The parameter estimation and comparison criteria results for DS-1 of all the models under consideration can be viewed through Table 3.7. It is clear from the table that the value of $R^2$ for SRGM-3 and SRGM-4 are higher and value of MSE, Bias, Variation and RMSPE are lower in comparison to other models and provide better goodness of fit.
Data set 2 (DS-2)
The second data set (DS-2) had been collected during 19 weeks of testing a real time command and control system and 328 faults were detected during testing. This data is cited from Ohba (1984). The change-point for this data set is 6th week.

The parameter estimation and comparison criteria results for DS-2 of all the models under consideration can be viewed through Table 3.8. It is clear from the table that the value of $R^2$ for SRGM-3 and SRGM-4 are higher and value of MSE, Bias, Variation and RMSPE are lower in comparison with other models and provides better goodness of fit.

Table 3.7 and 3.8 are on the next page
### Table 3.7: Model Parameter Estimation and comparison Results (DS-1)

<table>
<thead>
<tr>
<th>Models</th>
<th>a</th>
<th>b₁</th>
<th>b₂</th>
<th>( \beta_1 / k_1 )</th>
<th>( \beta_2 / k_2 )</th>
<th>R²</th>
<th>MSE</th>
<th>Bias</th>
<th>Variation</th>
<th>RMSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGM-1</td>
<td>2705</td>
<td>.019</td>
<td>.023</td>
<td>-</td>
<td>-</td>
<td>.97367</td>
<td>5608.09</td>
<td>-1.28E-05</td>
<td>75.9804</td>
<td>75.9804</td>
</tr>
<tr>
<td>SRGM-2</td>
<td>1639</td>
<td>.092</td>
<td>.095</td>
<td>-</td>
<td>-</td>
<td>.98833</td>
<td>2486.00</td>
<td>-3.89E-06</td>
<td>50.5877</td>
<td>50.5877</td>
</tr>
<tr>
<td>SRGM-3</td>
<td>1303</td>
<td>.004</td>
<td>.007</td>
<td>.856</td>
<td>2.50</td>
<td>.99872</td>
<td>271.78</td>
<td>-8.06E-05</td>
<td>16.72671</td>
<td>16.72671</td>
</tr>
<tr>
<td>SRGM-4</td>
<td>1321</td>
<td>.073</td>
<td>.211</td>
<td>.770</td>
<td>25.3</td>
<td>.99930</td>
<td>149.51</td>
<td>6.14E-05</td>
<td>12.4063</td>
<td>12.4063</td>
</tr>
<tr>
<td>SRGM-5</td>
<td>1460</td>
<td>.165</td>
<td>.163</td>
<td>-</td>
<td>-</td>
<td>.99454</td>
<td>1162.19</td>
<td>0.000285</td>
<td>34.5882</td>
<td>34.5882</td>
</tr>
</tbody>
</table>
Table 3.8: Model Parameter Estimation and comparison Results (DS-2)

<table>
<thead>
<tr>
<th>Models</th>
<th>a</th>
<th>b₁</th>
<th>b₂</th>
<th>$\beta_1/_{k_1}$</th>
<th>$\beta_2/_{k_2}$</th>
<th>$R^2$</th>
<th>MSE</th>
<th>Bias</th>
<th>Variation</th>
<th>RMSPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRGM-1</td>
<td>614</td>
<td>.040</td>
<td>.044</td>
<td>-</td>
<td>-</td>
<td>.98812</td>
<td>122.66</td>
<td>-6.55E-07</td>
<td>11.3789</td>
<td>11.3789</td>
</tr>
<tr>
<td>SRGM-2</td>
<td>401</td>
<td>.189</td>
<td>.169</td>
<td>-</td>
<td>-</td>
<td>99054</td>
<td>97.617</td>
<td>5.32E-06</td>
<td>10.1508</td>
<td>10.1508</td>
</tr>
<tr>
<td>SRGM-3</td>
<td>372</td>
<td>.020</td>
<td>.018</td>
<td>2</td>
<td>1</td>
<td>.99143</td>
<td>88.415</td>
<td>-0.00100</td>
<td>9.66039</td>
<td>9.66039</td>
</tr>
<tr>
<td>SRGM-4</td>
<td>365</td>
<td>.096</td>
<td>.217</td>
<td>.521</td>
<td>4.9</td>
<td>.99270</td>
<td>75.2976</td>
<td>0.00247</td>
<td>8.91520</td>
<td>8.91520</td>
</tr>
<tr>
<td>SRGM-5</td>
<td>362</td>
<td>.338</td>
<td>.287</td>
<td>-</td>
<td>-</td>
<td>.98846</td>
<td>119.05</td>
<td>-0.01</td>
<td>11.21106</td>
<td>11.21106</td>
</tr>
</tbody>
</table>
Chapter 3
Unified Framework for modeling Fault removal using Change point phenomenon

Figure 3.3 Goodness of Fit Curve for DS-1

Figure 3.4 Goodness of Fit Curve for DS-1