Performance Analysis of Software Reliability Models using Matrix Method

RajPal Garg¹, Kapil Sharma², Member IEEE, Rajive Kumar³, R. K. Garg⁴

Abstract—This paper presents a computational methodology based on matrix operations for a computer based solution to the problem of performance analysis of software reliability models (SRMs). A set of seven comparison criteria have been formulated to rank various non-homogenous Poisson process software reliability models proposed during the past 30 years to estimate software reliability measures such as the number of remaining faults, software failure rate, and software reliability. Selection of optimal SRM for use in a particular case has been an area of interest for researchers in the field of software reliability. Tools and techniques for software reliability model selection found in the literature cannot be used with high level of confidence as they use a limited number of model selection criteria. A real data set of middle size software project from published papers has been used for demonstration of matrix method. The result of this study will be a ranking of SRMs based on the Permanent value of the criteria matrix formed for each model based on the comparison criteria. The software reliability model with highest value of the Permanent is ranked at number – 1 and so on.

Keywords—Matrix method, Model ranking, Model selection, Model selection criteria, Software reliability models.

ACRONYM

NHPP Non-homogeneous Poisson process
SRM Software reliability model
MLE Maximum likelihood estimation

NOTATION

$m(t)$ Mean value function
$\lambda(t)$ Intensity function
$m_i$ Total number of failures observed at time $t_i$ according to the actual data
$\hat{m}(t_i)$ Expected number of failures at time $t_i$ estimated by a Model

I. INTRODUCTION

THE software development process becomes increasingly time-consuming and expensive due to the complexity of software systems. In the mean time, the need for the highly reliable software system is ever increasing. How to enhance the reliability of the software systems and reduce the cost to an acceptable level becomes the main focus of the software industry. Methods of applying reliability and cost models to the software development practice are highly desired [1].

The effects of this process, by which it is hoped software is made more reliable, can be modeled through the use of Software Reliability Models, hereafter referred to as SRMs. Ideally, these models provide a means of characterizing the development process and enable software reliability practitioners to make predictions about the expected future reliability of software under development. Such techniques allow managers to accurately allocate time, money, and human resources to a project, and assess when a piece of software has reached a point where it can be released with some level of confidence in its reliability. Various models have been proposed to characterize software reliability and its dependence on a number of factors related to the product or the software process, some of these are presented in Xie [2], Goel and Okumoto [3], Lyu [4], and Musa and Okumoto [5].

With the rapid development of computer technology, wide use of computers to control all military and civil systems and increasing demand of high quality software products, software reliability has become the primary concern and it is must to evaluate software reliability accurately and carefully to determine the system reliability. The software reliability models also called as counting models, those represent cumulative number of failures and have time dependent failure intensity function are considered the best ways to measure software reliability. Many mathematical models called SRMs (software reliability models) have been developed. The techniques for achieving and demonstrating high reliability are available, through the various reliability models. But how do we use them and do we need this diversity of models? To answer this question we have to compare models and study how they relate to each other. First of all we have to study the models and select those are best suited for our environment, techniques and applications according to the assumptions made by the models. Having done this we are probably left with either a number of models or none at all. If we are left with none, we have to take an approach similar to the one...
presented in Wohlin [6], i.e. to develop a model which is tailored to the environment and the techniques used. Let us suppose that we still have at least two possible models.

In this paper we proposed a matrix method for performance analysis of sixteen different NHPP software reliability models based on a set of seven contributing model selection criteria. The remainder of the paper is organized as follows: Section II reveals the existing literature for different types of software reliability models, models’ selection criteria and selection methodologies etc. In section III, parameter estimation technique and various comparison criteria are identified. The evaluation of the SRM based on individual criteria for a data set comprising three releases is given in section IV. The matrix method and model demonstration with the help of case study to develop a procedure mingling various comparison criteria for comprehensive ranking of the alternative NHPP software reliability models are described in Section V. Finally, the conclusions are given in Section VI.

II. LITERATURE REVIEW

Early work in the field of software reliability focused around proposing new models. Over the past 30 years, many SRGMs have been proposed for estimation of reliability growth of products during software development process [2, 4, 7-16]. Each model could be shown to work well with a unique data set, but no model appeared to do well on all data sets. Many researchers like Musa et al. [7] have shown that some families of models have, in general, certain characteristics that are considered better than others; for example, that the geometric family of models tends to have better predictive quality than other models. These and other attempts, Schick and Wolverton [17] and Sukert [18], to compare different models have led to an evolution from proposing a new model to proposing techniques for finding the best model for each individual application from among the existing models. Ideally we would like to be able to select, before starting, which model we should use.

This has proven to be a very difficult, almost impossible task. Brocklehurst et al. [19] suggest that it is the very nature of software failures that has made the model selection process in general a difficult task. Software failures are caused by hidden design flaws and not by the psychological sciences that will someday show us how to select the model beforehand. Today we must evaluate different models, compare them, and choose the best.

Goel and Okumoto [3] published a paper describing a nonhomogeneous Poisson process model from the finite exponential class of models. This was one of the first nonhomogeneous Poisson process models proposed. Goel and Okumoto validated this model by showing that it predicted well on a unique data set.

Goel [20] and others started describing processes for which each model would be tested to see how well the model fits the data and predicts the future events. The assertion was that different models predict well only on certain data sets and that by comparing the predictive quality of different models, it is possible to select the best one for a given application.

Abdel-Ghaly et al. [21] compared the predictive quality of 10 models using five different methods of comparison. They showed that different methods of model selection result in different models being chosen. Also some of their methods were rather subjective as to which model was better than others. Clearly a simple and objective method to select models is needed.

Khoshgoftaar and Woodcock [22] proposed a method to select a reliability model among various alternatives using the log-likelihood function. They apply the method to the failure logs of a project. The method selected an S-shaped model as the most appropriate one.

III. PARAMETER ESTIMATION AND COMPARISON CRITERIA

Since computers are being used increasingly to monitor and control both safety critical and civilian systems, there is a great demand for high-quality software products. Reliability is a primary concern for both software developers and software users. Research activities in software reliability engineering have been conducted and a number of NHPP software reliability models have been proposed to assess the reliability of software. In fact, software reliability models based on the NHPP have been quite successful tools in practical software reliability engineering. These models consider the debugging process as a counting process characterized by its mean value function. Software reliability can be estimated once the mean value function is determined. Model parameters are usually estimated using either the maximum likelihood method or regression. Different models have been built upon different assumptions. The sixteen NHPP software reliability models, as mentioned in Table 1, are considered for comparison and ranking in this research paper.

A. Parameter Estimation

Once the analytic expression for the mean value function is derived, it is required to estimate the parameters in the mean value function, which is usually carried out by using Maximum likelihood Estimation technique.

B. Comparison Criteria

A model can be judged according to its ability to reproduce the observed behavior of the software, and to predict the future behavior of the software from the observed failure data. A detailed study of the available literature reveals that the following twelve quantitative criteria are being used for comparison of software reliability models for different data sets.

1. The Bias is defined as [23], [24]:

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

   (1)

   It is the sum of the difference between the estimated curve and the actual data.
The mean square error (MSE) measures the deviation between the predicted values with the actual observations and is defined as [31]:

\[ \text{MSE} = \frac{1}{k-p} \sum_{i=1}^{k} (m_i - \hat{m}(t_i))^2 \]  

\[ (2) \]

2. The mean absolute error (MAE) is similar to MSE, but the way of measuring the deviation is by the use of absolute values. It is defined as [32]:

\[ \text{MAE} = \frac{1}{k-p} \sum_{i=1}^{k} |m_i - \hat{m}(t_i)| \]  

\[ (3) \]

3. The mean error of prediction (MEOP) sums the absolute value of the deviation between the actual data and the estimated curve and is defined as [33]:

\[ \text{MEOP} = \frac{1}{k-p+1} \sum_{i=1}^{k} |\hat{m}(t_i) - m_i| \]  

\[ (4) \]
5. The accuracy of estimation (AE) can reflect the difference between the estimated numbers of all errors with the actual number of all detected errors. It is defined as [32]:

$$AE = \left| \frac{M_a - \hat{a}}{M_a} \right|$$

where $M_a$ and $a$ are the actual and estimated cumulative number of detected errors after the test, respectively.

6. The noise is defined as [34]:

$$Noise = \sum_{i=1}^{k} \frac{\lambda(t_i) - \hat{\lambda}(t_i,c)}{\hat{\lambda}(t_i,c)}$$

7. The predictive-ratio risk (PRR) is defined as [35]:

$$PRR = \sum_{i=1}^{k} \frac{\hat{m}(t_i) - m_i}{\hat{m}(t_i)}$$

which measures the distance of model estimates from the actual data against the model estimate.

8. The variance is defined as [23], [24]:

$$Variance = \frac{1}{k-1} \sum_{i=1}^{k} (m_i - \hat{m}(t_i) - Bias)^2$$

which is standard deviation of prediction bias.

9. The Root Mean Square Prediction Error (RMSPE) is a measure of the closeness with which the model predicts the observation. It is defined as [23], [24]:

$$RMSPE = \sqrt{Variance + Bias^2}$$

R square (Rsq) can measure how successful the fit is in explaining the variation of the data.

10. It is defined as [32]:

$$Rsq = 1 - \frac{\sum_{i=1}^{k} (m_i - \hat{m}(t_i))^2}{\sum_{i=1}^{k} (m_i - \bar{m})^2/n}$$

11. The sum of squared errors (SSE) is defined as [30]:

$$SSE = \sum_{i=1}^{k} (m_i - \hat{m}(t_i))^2$$

12. The Theil statistic (TS) is the average deviation percentage over all periods with regard to the actual values. The closer, Theil’s Statistic is to zero, the better the prediction capability of model. It is defined as [36]:

$$TS = \frac{\sum_{i=1}^{k} (\hat{m}(t_i) - m_i)^2}{\sum_{i=1}^{k} m_i^2} \times 100\%$$

In equations 1 to 12, above, $k$ represents the sample size of the data set and $p$ is the number of parameters.

The comparison criteria, Root Mean Square Prediction Error (RMSPE) is a combination of the comparison criteria ‘bias’ and ‘variance’. The criteria MSE, MAE and MEOP are used to measure the deviation whereas the criteria AE and SSE measure the errors. In order to avoid the replication of the criteria and in order to investigate the effectiveness of software reliability models, a set of seven distinct comparison criteria namely mean absolute error (MAE), accuracy of estimation (AE), noise, predictive-ratio risk (PRR), Root Mean Square Prediction Error (RMSPE), R square (Rsq) and Theil statistic (TS) are proposed to compare models quantitatively.

IV. MODEL EVALUATION AND COMPARISON

In order to evaluate and compare the models, failure data considered by [37] is used in this research paper. As reported by the researchers, this data set is from the testing process on a middle-size software project. Table 2 shows failure data. First column present failure time in weeks and second column presents cumulative number of failures. The values of the parameters for these sixteen NHPP SRMs have been estimated using the MLE technique and confidence bounds of 95%. The estimated values of the parameters have been provided in Table 3.
The values of the seven comparison criteria considered in this research paper have been obtained, using relevant equations (Eqs. 3, 5, 7, 9, 10 and 12) The estimated and optimal values of the parameters are given in Table 4.

From the comparison of rankings of the sixteen SRMs based on the values of all these seven criteria as given in Table 4, it is observed that the ranking of the SRMs varies with respect to the release and criteria. No single model is best suitable for all comparison criteria. In order to avoid this problem it is proposed to apply matrix methodology to analyze the performance and rank the SRMs based on all these seven criteria taken collectively.

V. METHODOLOGY ADOPTED

To depart from complexity of the formulation of objective and constraint functions that occur when the mathematical programming model is used in a multi-attributes decision problem, a modest attempt is made in this paper to develop a deterministic quantitative model based on matrix operations for the purpose of ranking of software reliability models. Matrix methods have previously been used for power quality evaluation in deregulated power system [38] and optimizing selection of power plants [39]. The brief introduction to the basic concepts matrix operations is presented in this section.

A. Criteria Matrix

Each software reliability model at this stage is characterized by multiple criteria, which need to be converted into a single number index that will be used to rank the software reliability models. The matrices lend themselves easily to mechanical manipulations and are suitable for computer processing. The ratings and the relative aggregated weights of the comparison criteria for a software reliability model give the rankings and the relative aggregated weights of the comparison criteria, respectively. Thus, the criteria matrix is a combination of two matrices namely ‘criteria rating matrix’ and ‘criteria relative weight matrix’.

Criteria Rating Matrix: This is a diagonal matrix whose elements (a ii’s or a i’s) represent the ratings of different comparison criteria for a software reliability model and is represented as follows:

\[
\begin{bmatrix}
    a_{11} & 0 & 0 & \cdots & 0 \\
    0 & a_{22} & 0 & \cdots & 0 \\
    \vdots & \vdots & \ddots & \ddots & \vdots \\
    0 & 0 & \cdots & a_{nn}
\end{bmatrix}
\]

### Table IV

| Database for Estimated and Optimal Values of Attributes for Each Alternate SRM |
|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| AE Rank Noise Rank Rsq Rank TS Rank PRR Rank RMSPE Rank MAE Rank |       |       |       |       |       |       |       |       |       |       |       |       |
| Generalized Goel | 0.034614 | 13 | 15.0034 | 16 | 0.380572 | 13 | 28.4307 | 13 | 11135.07 | 16 | 74.47332 | 13 | 35.15119 | 13 |
| Goel-Oukamoto | 0.007959 | 5 | 2.28572 | 6 | 0.992325 | 2 | 3.05982 | 2 | 0.147447 | 3 | 4.54664 | 2 | 3.843456 | 1 |
| Gompert | 0.01836 | 8 | 3.78193 | 11 | 0.982757 | 4 | 4.74347 | 4 | 0.191337 | 7 | 7.045852 | 3 | 5.853303 | 4 |
| Inflection S-Shaped | 0.001625 | 3 | 2.56645 | 7 | 0.993054 | 1 | 3.01668 | 1 | 0.098526 | 1 | 4.467552 | 1 | 3.942101 | 2 |
| Logistic Growth | 0.025522 | 12 | 4.75978 | 14 | 0.973209 | 6 | 5.90287 | 6 | 0.395383 | 8 | 8.851496 | 4 | 7.692465 | 8 |
| Modified Duane | 0.014315 | 7 | 1.54903 | 3 | 0.775166 | 12 | 1.72873 | 12 | 1.069542 | 9 | 53.44483 | 12 | 27.12543 | 12 |
| Musa-Oukamoto | 0.03768 | 14 | 1.63260 | 4 | 0.981355 | 5 | 4.93265 | 5 | 0.217156 | 5 | 9.39541 | 8 | 6.397615 | 5 |
| Yamada imperfect debugging model 1 | 0.020482 | 11 | 1.49818 | 2 | 0.939523 | 10 | 8.83556 | 10 | 0.318831 | 7 | 24.38592 | 10 | 13.29734 | 10 |
| Yamada Rayleigh | 0.925064 | 16 | 7.8719 | 15 | 0.879863 | 11 | 12.5208 | 11 | 5.095656 | 14 | 27.27232 | 11 | 18.22131 | 11 |
| Delayed S-Shaped | 0.020405 | 9 | 4.21039 | 12 | 0.972967 | 7 | 5.93937 | 7 | 1.892668 | 10 | 8.862182 | 6 | 7.120771 | 6 |
| Yamada imperfect debugging model 2 | 0.001086 | 1 | 1.38165 | 1 | 0.96755 | 9 | 6.50725 | 9 | 0.236762 | 6 | 16.4287 | 9 | 9.250905 | 9 |
| Yamada exponential | 0.904552 | 15 | 3.40447 | 10 | 0.301786 | 15 | 35.545 | 14 | 2.553634 | 12 | 114.2261 | 15 | 58.0289 | 15 |
| P-N-Z Model | 0.009734 | 6 | 1.87603 | 5 | 0.987202 | 3 | 4.08667 | 3 | 0.121221 | 2 | 8.859432 | 5 | 5.80468 | 3 |
| P-Z Model | 0.007371 | 2 | 1.5994 | 8 | 0.1087114 | 14 | 38.50367 | 16 | 31.38161 | 15 | 119.504 | 16 | 67.87029 | 16 |
| Pham Zhang IFD | 0.020407 | 10 | 4.21064 | 13 | 0.972964 | 8 | 5.93968 | 8 | 1.893453 | 11 | 8.862517 | 7 | 7.516719 | 7 |
| Zhang-Teng-Pham | 0.006976 | 4 | 2.86226 | 9 | 0.00392 | 16 | 36.0529 | 15 | 2.677499 | 13 | 104.2807 | 14 | 57.67061 | 14 |
| optimal | 0.001086 | 1.38165 | 0.993054 | 3.01068 | 0.089826 | 4.467552 | 3.843456 | | | | | | |
As these criterion ratings are different for different software reliability models, hence, the criteria rating matrix differs from model to model. The criteria ratings are determined as under:

Case - I: When smaller value of the criterion represents fitting well to the actual data i.e. is the best value:

$$\text{Criteria rating} = \frac{\text{Criterion Maximum Value in the Database} - \text{Criterion Value}}{\text{Criterion Minimum Value in the Database} - \text{Criterion Minimum Value in the Database}}$$

Case - II: When bigger value of the criterion represents fitting well to the actual data i.e. is the best value:

$$\text{Criteria rating} = \frac{\text{Criterion Value} - \text{Criterion Minimum Value in the Database}}{\text{Criterion Maximum Value in the Database} - \text{Criterion Minimum Value in the Database}}$$

Criteria Relative Weight Matrix: The Criteria Relative Weight Matrix is formed on the basis of the aggregated weights of different criteria. The off diagonal elements of this matrix represent the relative weights of the criteria e.g. the element \(a_{ij}\) of this matrix will give the relative weight of \(j^{th}\) criterion in respect of \(i^{th}\) criterion. All diagonal elements of this matrix are zero because there is no significance of comparing a criterion with respect to itself. Mathematically \(a_{ij} = \text{weight of } j^{th} \text{ criteria} / \text{weight of } i^{th} \text{ criteria.}\)

Thus the ‘Criteria Matrix’ corresponding to ‘n’ criteria, in general, is written as:

$$\begin{bmatrix}
0 & a_{12} & a_{13} & \ldots & a_{1n} \\
a_{21} & 0 & a_{23} & \ldots & a_{2n} \\
\vdots \\
a_{n1} & a_{n2} & a_{n3} & \ldots & 0
\end{bmatrix}$$

The results, so obtained, depict that the P-Z model is ranked at number one, Yamada exponential is number two whereas inflection S-shaped and Goel-Okumo models are ranked lowest at 15 & 16 numbers respectively. It is well established that the proposed method is suitable for distinct ranking of the models for any data sets based on a number of criteria taken collectively.

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Permanent Value</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized Goel</td>
<td>0.51102</td>
<td>4</td>
</tr>
<tr>
<td>Goel-Okamoto</td>
<td>0.99994</td>
<td>16</td>
</tr>
<tr>
<td>Gompert</td>
<td>0.96861</td>
<td>13</td>
</tr>
<tr>
<td>Inflection S-Shaped</td>
<td>0.99846</td>
<td>15</td>
</tr>
<tr>
<td>Logistic Growth</td>
<td>0.93988</td>
<td>9</td>
</tr>
<tr>
<td>Modified Duane</td>
<td>0.63657</td>
<td>5</td>
</tr>
<tr>
<td>Masa-Okamoto</td>
<td>0.96011</td>
<td>12</td>
</tr>
<tr>
<td>Yamada imperfect debugging model1</td>
<td>0.85234</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model Name</th>
<th>Permanent Value</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yamada Rayleigh</td>
<td>0.77544</td>
<td>6</td>
</tr>
<tr>
<td>Yamada imperfect debugging model2</td>
<td>0.91554</td>
<td>8</td>
</tr>
<tr>
<td>Yamada exponential</td>
<td>0.15371</td>
<td>2</td>
</tr>
<tr>
<td>P-N-Z Model</td>
<td>0.97131</td>
<td>14</td>
</tr>
<tr>
<td>P-Z Model</td>
<td>0.90058</td>
<td>1</td>
</tr>
<tr>
<td>Pham Zhang IFD</td>
<td>0.94263</td>
<td>10</td>
</tr>
<tr>
<td>Chung-Teng-Pham</td>
<td>0.15922</td>
<td>3</td>
</tr>
</tbody>
</table>
VI. CONCLUSION

This paper addresses the issue of performance analysis of software reliability models. The decision has unrestricted choices in exploring the influences of various different set of model selection criteria to final decision. As soon as a complete set of criteria for SRMs selection, along with the set of alternative SRMs and their level of criteria are formulized, and efficient rationalization process around multi-attribute decision model ‘matrix method’ can be performed. This model allows a decision maker to perform, not just a general analysis, but also other various focused analyses regarding his or her personal preferences.

The proposed method is suitable for ranking of SRMs based on a number of conflicting criteria taken all together with equal or unequal weights. This method uses a simple mathematical formulation and straightforward matrix operation to be performed by computing machines and is capable of solving complex multi-attributes decision problems, incorporating both quantitative and qualitative factors.

REFERENCES


RajPal Garg was born in Haryana, India in 1963. He is pursuing a Doctors Degree in Computer Science and Engineering under the Faculty of Engineering and Technology at the Singhania University, Jhunjhunu (Rajasthan), India. He has obtained his Master of Technology and Master of Science Degrees in Information Technology and Maths from Punjabi
University, Patiala, and Kurukshetra University, Kurukshetra, India in 2003, and 1986 respectively. His research interests include software engineering, and reliability.

**Kapil Sharma** was born in Haryana, India in 1977. He is pursuing a Doctors Degree in Computer Science and Engineering under the Faculty of Engineering and Technology at the M. D. University, Rohtak (Haryana), India. He has obtained his Bachelor of Engineering and Master of Technology Degrees in Computer Science & Engineering and Information Technology from M. D. University, Rohtak, and IASE, Rajasthan, India in 2000, and 2005 respectively. Presently, he is working as a Faculty of the Department of Information Technology at Guru PremSukh Memorial College of Engineering, Budhpur, Delhi, India. His research interests include software engineering, and reliability.

**Rajive Kumar** was born in Hardwar (UP), India in 1963. He has obtained degrees of Master of Science and Master of Philosophy from Roorkee University, Roorkee, India, and Doctor of Philosophy from Indian Institute of Technology, Delhi, India, all in Mathematics, in 1982, 1985, and 1990 respectively. Presently, he is working as Professor in the Department of Mathematics with Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Haryana, India. His research interests include numerical methods and software engineering. Dr. Kumar has published a number of research papers in International journals and conferences. Presently, he is guiding seven research scholars.

**R. K. Garg** was born in Haryana, India in 1970. He has obtained degrees of Bachelor of Engineering with Honours, Master of Engineering, and Doctor of Philosophy in Mechanical Engineering from M. D. University, Rohtak, Panjab University, Chandigarh, and M. D. University, Rohtak, Haryana, India in 1991, 1996, and 2006 respectively. Presently, he is working as Professor in the Department of Mechanical Engineering with Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Haryana, India. His research interests include system design, reliability modeling & analysis, and power & energy engineering. Dr. Garg has published a number of research papers in International journals and conference, and has guided a number of Masters degree dissertations. Presently, he is guiding five research scholars.