A Comparison of the Robustness of Reliability Growth Assessment Techniques

Martin Wayne, USAMSAA  
Paul Ellner, PhD, USAMSAA

Key Words: AMPM, Crow-Extended Model, Reliability Growth

SUMMARY & CONCLUSIONS

Reliability assessment techniques constitute an important element of the reliability growth program. This paper examines the accuracy and robustness of two widely used reliability growth assessment techniques under a number of realistic corrective action processes. These methods are also compared to a newly developed assessment approach. The new approach provides a more robust assessment across a broader spectrum of cases. These include various corrective action processes as well as cases in which the number of failure modes in the system is not large compared to the number of failure modes surfaced during testing.

The results indicate that each of the techniques perform acceptably well under test conditions that adhere to the assumptions on which they are based. The performance can be severely degraded when the assumptions are violated, and this is the case with the Crow-Extended Model. The AMSAA Maturity Projection Model (AMPM) and the newly developed alternate technique have a somewhat limited number of assumptions though. They appear to be more robust to the various test conditions examined here, with the alternate being the most useful across the broadest spectrum of cases.

INTRODUCTION

This paper extends earlier concepts shown in [1]. As motivation for this work, we look to recent Defense Science Board recommendations that robust reliability growth programs be utilized for complex weapons system developmental programs. This recommendation is due in part to the failure of many systems to meet their reliability requirements at the conclusion of their developmental programs. Such shortfalls in reliability translate into significant impacts on operational mission suitability and logistics costs. This renewed emphasis on reliability growth highlights the importance of understanding the robustness of the various models.

1.1 Purpose

Reliability assessment techniques constitute an important element of any reliability growth program. This paper examines the accuracy and robustness of several widely used reliability growth assessment techniques under a number of realistic corrective action processes.

INTRODUCTION

Section 2 briefly discusses the reliability growth models and assessment techniques that are considered. The assumptions associated with each model and technique are discussed. Section 3 contains comparison results from the simulation test cases. The simulated Mean Time Between Failures (MTBF) is compared to those of the three assessment techniques. The empirical relative error distribution is also shown for select cases, allowing for comparisons of the techniques in a more practical setting.

Section 4 provides a general discussion of the robustness of the three models. Results from the comparison tests are used to provide some knowledge of the appropriateness of an individual technique. Section 5 discusses some potential areas for future work involving reliability growth modeling.

2 RELIABILITY ASSESSMENT TECHNIQUES

Before discussing the performance of the existing reliability growth models, we provide a brief overview of their capabilities and assumptions.

2.1 Crow Extended Model

The model proposed by Crow in [2] is used widely in both government and industry alike, and it consists of two main parts: the reliability growth tracking model [3] and the reliability growth projection model [4]. The underlying assumption of each of the models is that of a nonhomogeneous Poisson process with mean value function represented by

$$\mu(t) = \omega e^{\beta t}.$$  \hspace{1cm} (1)

The parameter $\omega$ is used in the place of the customary $\lambda$ in order to avoid any conflicting terminology that may occur later.

The tracking model assumes that all corrective actions are implemented instantaneously in what is known as a test-fix-test environment, and $\beta < 1$ corresponds to reliability growth. Maximum likelihood estimates (MLE) for $\omega$ and $\beta$ can be developed from the cumulative failure times, and they are given in [2].

The projection model distinguishes between A modes and B modes, where B modes are the only modes that will receive a corrective action. The underlying assumption is that all corrective actions applied to B modes are delayed until the end of the test.
of the test, and the reliability is assessed after the delayed corrective actions have been implemented. MLE for the projection model are based only on the first occurrence times for each of the B failure modes.

The reliability assessment for the Crow-Extended Model is then made by combining the results of the tracking and projection models. The tracking model is used first with all cumulative time to failure data. The B modes are then subdivided into BC and BD modes. BC modes are those that will have instantaneous corrective actions and BD modes are those that will have delayed corrective actions. The projection model is applied only to the BD mode times of first occurrence. No A modes are assumed in the projection component because they are accounted for in the tracking portion of the model. The failure intensity of the extended model is given in [2] as

$$\rho_{EM}(t) = \rho_T(t) - \lambda_{BD} + \sum_{i=1}^{K} (1 - d_i) \lambda_i + \tilde{d} \tilde{\alpha}_i \tilde{\beta}_i (t - T)^{\tilde{\alpha}_i - 1}$$

(2)

where $K$ is now the number of BD modes, $\lambda_{BD}$ is the initial failure rate of the BD modes, $\tilde{d}$ and $\tilde{\alpha}_i$ are the respective FEF and rate of occurrence for the $i^{th}$ BD mode. Intuitively, the failure intensity due to the BD modes is subtracted from the tracking assessment and then replaced with the projection assessment. This accounts for the delayed corrective actions that are implemented after the test. See [2] for more details on the model, along with management metrics associated with it.

2.2 AMPM

Unlike the Crow-Extended Model, the AMPM does not assume a nonhomogeneous Poisson process. It also makes no assumptions about when corrective actions are employed, as they are only assumed to have been performed at some time prior to the time of the projection. The foundation of the AMPM is a doubly stochastic process, in that the underlying initial rates of occurrence for individual B failure modes are a realization of a random sample of size $K$ from a Gamma distribution. The failure intensity is given in [5] as

$$\rho_{AMPM}(t) = \tilde{\lambda}_A + (1 - \tilde{d}) \lambda_B + \tilde{d} \frac{K \tilde{\beta}(\tilde{\alpha} + 1)}{(1 + \tilde{\beta} T)^{\tilde{\alpha} + 1}}$$

(3)

where $\tilde{d}$ is the average FEF for the failure modes that are corrected, $\tilde{\alpha}$ and $\tilde{\beta}$ are the parameters from the underlying Gamma distribution. MLEs of $\tilde{\alpha}$ and $\tilde{\beta}$ are found in terms of a fixed $K$ and the B Mode first occurrence times. Note that the $\beta$ parameter in the AMPM is different from that discussed earlier in the Crow-Extended Model, but the overall form of the projection in (3) is similar to that shown in (2).

The limiting behavior with respect to $K$ is also available, leading to two potential forms of the model shown in (5) and (5). $\tilde{d}$ is the average FEF of the $m$ observed B failure modes, and $\tilde{\lambda}_{B,\infty}$ is given in (4). Further details on the model can be found in [5].

$$\lambda_{B,\infty} = \lim_{K \to \infty} K \tilde{\beta}(\tilde{\alpha} + 1) = \frac{m \tilde{\beta}_B}{\ln(1 + \beta_e T)}$$

(4)

$$\hat{\rho}_{AMPM}(t) = \hat{\lambda}_A + (1 - \tilde{d}) \hat{\lambda}_B + \tilde{d} \frac{K \hat{\beta}(\hat{\alpha} + 1)}{(1 + \hat{\beta} T)^{\hat{\alpha} + 1}}$$

(5)

$$\hat{\rho}_{AMPM}(t) = \hat{\lambda}_A + (1 - \tilde{d}) \hat{\lambda}_{B,\infty} + \tilde{d} \frac{\hat{\beta}_B}{(1 + \hat{\beta}_e T)}$$

(6)

2.3 Alternate MLE in AMPM

A potential problem with the AMPM is the practical limitation that exists. The total number of failure modes in the system is usually not known. The infinite form of the model given in (13) can usually be justified for sufficiently complex systems, but this may not always be reasonable. This issue gives rise to the necessity of an estimation procedure that works without requiring knowledge of the total number of failure modes in the system.

Using the fact that the expected value of the initial B mode failure rate, $\lambda_B$, is equivalent to $K \beta(\alpha + 1)$, we can rewrite (5) as

$$\hat{\rho}_{AMPM}(t) = \hat{\lambda}_A + (1 - \tilde{d}) \hat{\lambda}_B + \tilde{d} \frac{\hat{\beta}_B}{(1 + \hat{\beta}_e T)^{\hat{\alpha} + 1}}$$

(7)

Obtaining a reliability assessment independent of $K$ can now be made, provided that suitable estimators (also independent of $K$) can be found for $\hat{\lambda}_B$, $\hat{\alpha}$, and $\hat{\beta}$.

Let $\alpha_i = (t_{i,1}, t_{i,2}, \ldots, t_{i,n_{i,1}}, v_i, t_{i,n_{i,1}+1}, \ldots, t_{i,n_{i,T}}, T)$ represent the observed information for the $i^{th}$ failure mode. $t_{i,j}$ is the $j^{th}$ cumulative failure time for the $i^{th}$ failure mode, $n_{i}$ is the total number of failures, $v_i$ is the time when the corrective action is made, $n_{i,j}$ is the number of failures occurring before $v_i$, and $T$ is the total test time. Assuming the time between failures is exponential with rate $\lambda_i$, the log-likelihood for $\alpha_i$ is then

$$L(\alpha_i) = n \log \lambda_i - \lambda_i \sum_{j=1}^{n_{i,j}} (T - v_i)$$

(8)

where $d_i$ is the fix effectiveness factor for the $i^{th}$ failure mode. This likelihood is obtained by realizing that the failures occurring after the corrective action have a reduced rate of $(1 - d_i) \lambda_i$. The MLE for $\lambda_i$ is then given by

$$\hat{\lambda}_i = \frac{n_i}{v_i + (1 - d_i)(T - v_i)}$$

(9)

This provides an estimate for the initial failure rate for an individual B mode, which in turn leads to an estimate for $\lambda_B$ of

$$\hat{\lambda}_B = \sum_{i=obs} \hat{\lambda}_i = \sum_{i=obs} \frac{n_i}{v_i + (1 - d_i)(T - v_i)}$$

(10)

Though not shown here, Method of Moments estimators for $\alpha$ and $\beta$ conditioned on the observed number of B modes do exist in most cases, and these can be used in place of the usual finite K MLE. Because the estimate of $\lambda_B$ is comprised of a sum of estimates for each of the observed failure modes, individual FEF can also be used in place of $\tilde{d}$ in (5).

This leads to an estimate of the overall failure intensity
given by
\[
\hat{\rho}_{\text{AMPM}_t}(t) = \frac{N}{T} + \sum_{i=1}^{n} \left(1 - d_i\right) \frac{n_i}{v_i + (1 - d_i)(T - v_i)}
\]
\[
+ d \frac{n_i}{\left(1 + \beta t\right)^{d+2}}
\]
(11)

3 RESULTS

The same twelve baseline test cases first shown in [1] are used again here and shown in Table 1. The number of A and B modes were 5 and 75 for each case. The mean FEF was 0.7 with a coefficient of variation of 0.1, and 50 replications were used for each case. The probability of a mode being classified a BC or BD was set at 0.80 and 0.20 respectively. Each case was then subjected to varying corrective action strategies, including using a corrective action delay time and a different BD mode classification strategy. Assuming the simulated results from the test bed to be the true parameters, the BD mode classification strategy. Although the simulated model leaves a bias. The A and BD modes have a constant failure rate, which violates the original assumption of a nonhomogeneous Poisson process with decreasing failure rate (reliability growth). Table 2 shows the average number of failures for each mode type along with the corresponding error in the projection. This shows that a sufficient number of BD mode failures will increase the bias term enough to lead to the optimistic reliability projections seen in Figure 1.

Table 2 – Comparison of Results

<table>
<thead>
<tr>
<th>Case</th>
<th>BC Failures</th>
<th>BD Failures</th>
<th>Ratio BC/BD</th>
<th>MTBF Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>54</td>
<td>20</td>
<td>2.70</td>
<td>1.343</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>23</td>
<td>2.09</td>
<td>0.909</td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>21</td>
<td>1.57</td>
<td>-4.496</td>
</tr>
<tr>
<td>4</td>
<td>266</td>
<td>203</td>
<td>1.31</td>
<td>-3.587</td>
</tr>
<tr>
<td>5</td>
<td>268</td>
<td>211</td>
<td>1.27</td>
<td>-2.599</td>
</tr>
<tr>
<td>6</td>
<td>196</td>
<td>238</td>
<td>0.82</td>
<td>-6.142</td>
</tr>
<tr>
<td>7</td>
<td>55</td>
<td>20</td>
<td>2.75</td>
<td>1.421</td>
</tr>
<tr>
<td>8</td>
<td>49</td>
<td>21</td>
<td>2.33</td>
<td>0.922</td>
</tr>
<tr>
<td>9</td>
<td>33</td>
<td>17</td>
<td>1.94</td>
<td>-0.483</td>
</tr>
<tr>
<td>10</td>
<td>270</td>
<td>188</td>
<td>1.44</td>
<td>-3.136</td>
</tr>
<tr>
<td>11</td>
<td>275</td>
<td>176</td>
<td>1.56</td>
<td>-3.301</td>
</tr>
<tr>
<td>12</td>
<td>236</td>
<td>213</td>
<td>1.11</td>
<td>-2.860</td>
</tr>
</tbody>
</table>

3.1 Baseline Cases

Average results for 50 replications are shown for the baseline cases in Figure 1. For the Crow-Extended Model, the results highlight a small bias that is present in the model. As explained in [1], the addition and subtraction that occurs in the

Table 1 – Comparison Cases

<table>
<thead>
<tr>
<th>Case</th>
<th>Test Time</th>
<th>A Mode-Intensity</th>
<th>A Mode-Initial Failure</th>
<th>B Mode-Intensity</th>
<th>B Mode-Initial Failure</th>
<th>Parent Pop</th>
<th>Alpha (A Modes)</th>
<th>Beta (A Modes)</th>
<th>Alpha (B Modes)</th>
<th>Beta (B Modes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>500</td>
<td>0.005</td>
<td>0.195</td>
<td>Gamma</td>
<td>0.0001</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>500</td>
<td>0.005</td>
<td>0.195</td>
<td>Gamma</td>
<td>0.002</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>500</td>
<td>0.005</td>
<td>0.195</td>
<td>Gamma</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>10000</td>
<td>0.005</td>
<td>0.095</td>
<td>Gamma</td>
<td>0.0001</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>10000</td>
<td>0.005</td>
<td>0.095</td>
<td>Gamma</td>
<td>0.002</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10000</td>
<td>0.005</td>
<td>0.095</td>
<td>Gamma</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>500</td>
<td>0.005</td>
<td>0.195</td>
<td>Weibull</td>
<td>0.001</td>
<td>0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>500</td>
<td>0.005</td>
<td>0.195</td>
<td>Weibull</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>500</td>
<td>0.005</td>
<td>0.195</td>
<td>Weibull</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>10000</td>
<td>0.005</td>
<td>0.095</td>
<td>Weibull</td>
<td>0.0001</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10000</td>
<td>0.005</td>
<td>0.095</td>
<td>Weibull</td>
<td>0.02</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>10000</td>
<td>0.005</td>
<td>0.095</td>
<td>Weibull</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1 – Baseline Comparisons
indicates that the standard AMPM forms provide more precise estimates of the true MTBF more often than either the alternate MLE form or the Crow Extended Model. This is most likely due to the number of parameters that are being estimated in each model. The standard AMPM has only two parameters that are estimated, while the other models contain more.

The AMPM makes no assumptions about when corrective actions occur, and the performance of the model should not be degraded under these circumstances. The results validate this statement, and the alternate MLE form is again able to provide a useful estimate even in those cases where the infinite form of the model is not appropriate.

For the error distributions in these cases, we again look at Case 3. Results for this case are shown in Figure 4. Note that data do not fit the model for the Crow Extended, so comparisons in this setting are merely to provide information on what can be expected if the model is used inappropriately. Again the results indicate that the standard AMPM performs the best of the models, with the alternate MLE form performing almost as well as the standard forms. The Crow Extended Model has the largest relative errors as expected. Looking at the same 20% error tolerance as in the baseline cases, we can see that probabilities are generally the same for the AMPM versions, but that the Crow Extended is now within the tolerance approximately 40% of the time. This drop is due to the effect of the delay time on the results, and as mentioned before this information is only provided in an effort to inform a user who may unknowingly misuse the model under these circumstances.

The twelve cases here were simulated in the same manner as the baseline, with the exception of having a delay time applied to each of the corrective actions that occurred. This has the effect of changing the classification of some BC modes into BD modes. It also violates two assumptions of the Crow Extended model. The classifications will be dependent on the first occurrence time, which violates the assumption of a nonhomo-geneous Poisson process for the BD modes, and the delay time also violates the tracking model assumptions that the corrective actions are implemented instantaneously.

As seen in [1], large numbers of BD failures can have a dramatic effect on the bias in the Crow Extended Model. It can lead to grossly optimistic results, but it may also be mitigated somewhat if the number of repeat failures for BD modes is small. Hypothesis tests also fail to highlight any potential problems with the data, so extreme caution should be used in these circumstances. The AMPM is not tied to mode
classification, so the performance of the model should remain the same as in the baseline cases under these conditions. The results validate this statement.

For the absolute relative error distributions in the weighted BD Mode cases, we look at Case 3 again. Results are shown in Figure 6. The results indicate that the AMPM versions perform about the same as in the baseline, while the Crow Extended Model meets the tolerance less than 5% of the time. This drop is due to the effect of the weighted mode classification, and this demonstrates the dangers of using the Crow Extended model under these circumstances.

4 DISCUSSION

The results presented here provide a potential user of these reliability growth models with vital information on the robustness of each one to various conditions that may exist. Use of the models without consideration of their limitations can potentially lead to grossly inaccurate projections of the MTBF.

Like most models, if the assumptions of the Crow-Extended Model are violated, the projected MTBF will tend to be inaccurate. Delay times tend to provide conservative results, while weighting the BD modes towards the end of the test may result in overly optimistic ones. Extreme caution should be exercised anytime the model is used under non-ideal conditions, as the resulting assessments may be grossly inaccurate and may lead to poor decisions for the system in question. The distribution of the absolute relative errors also indicates that the results of the Crow Extended Model tend to exhibit larger errors than those of the other models considered here even under ideal circumstances.

As mentioned previously, the infinite form of the AMPM may also provide overly conservative MTBF projections if used in the wrong settings. This issue can be overcome by using the alternate MLE form of the model developed here. The total number of failure modes in the system is not required, and the alternate form of the model still has a reasonably high probability of providing an accurate estimate of the MTBF. As always though, caution should be exercised when choosing a particular version of the AMPM for practical use.

5 FUTURE WORK

Though not presented here, the method of moment estimates used in the alternate MLE form of the AMPM tend to degrade the results when the number of failure modes is small. Additional estimators for $\alpha$ and $\beta$ that use all failure data would be useful, as they would likely increase the performance of the alternate MLE model. Further theoretical properties of the various estimators might also provide useful information on the robustness of each of the techniques.

In the practical application of a reliability growth model, it is usually not possible to estimate the cumulative distribution of the relative error of the model. It may prove useful in these cases to examine the distribution of the estimated MTBF itself through techniques such as bootstrapping. This would provide additional useful information in the practical use of the models.

REFERENCES


BIOGRAPHIES

Martin Wayne, Director USAMSAA
392 Hopkins Road
Aberdeen Proving Ground, MD, 21005, USA
Martin Wayne is an operations research analyst on the Reliability Analysis Team of the Army Materiel Systems Analysis Activity. He provides general statistical reliability support to AMSAA and other DoD organizations, and does research and development in the area of reliability growth. He earned a BS in mathematics from Gettysburg College, a MS in Applied Mathematics from the Johns Hopkins University, and he is currently pursuing his PhD in Applied Mathematics and Statistics at the Johns Hopkins University.

Paul Ellner, PhD
Director USAMSAA
392 Hopkins Road
Aberdeen Proving Ground, MD, 21005, USA

e-mail: reliability.tools@us.army.mil

Paul Ellner is a technical advisor at the U.S. Army Materiel Systems Analysis Activity. He provides general statistical reliability support to AMSAA, and does research and development in the area of reliability growth. Through AMSAA he provides support to other Army and Department of Defense organizations and their associated contractors for complex weapons system developmental programs in the area of reliability growth management and methodology. Dr. Ellner obtained his BS, MS and PhD degrees in mathematics from Rensselaer Polytechnic Institute, Troy, New York.