THE EFFECT OF DISCOUNTING FAILURES AND WEIGHTING DATA ON THE ACCURACY OF SOME RELIABILITY GROWTH MODELS

W. Max Woods; Naval Postgraduate School, Monterey

Key Words: Reliability growth, Failure discounting, weighting data

ABSTRACT

Failure discounting is the practice of removing fractions of previous failures after corrective action has been taken and no failures for the same cause reoccur in subsequent testing. This paper analyzes the effect of two parametric failure discounting methods on the accuracy of three discrete and two continuous reliability growth models. It makes similar comparisons for two data weighting methods. Graphs are used to make comparisons on the accuracy of these models without discounting or weighting, with discounting only, and with weighting only. The accuracy comparisons are made using Monte Carlo methods. The results show that cumulative growth models such as the AMSAA and Maximum Likelihood models have greater bias than the noncumulative regression models for the cases simulated. The results also show that the cumulative models appear to be more sensitive to failure discounting and thus more susceptible to yielding optimistic estimates of reliability than the regression type models when failure discounting is employed. Failure discounting applied too frequently (e.g., after each successful test) can adversely affect the accuracy of any of the models analyzed.

INTRODUCTION

As a product is developed, cycles of test-analyze-fix actions yields groups of test data. Each data group corresponds to tests on similar hardware that differ from the previous group of tests due to changes that have been made to correct determined failure causes. It would appear to be reasonable to remove fractions of previous failures for which corrective actions have been taken when subsequent tests show that no failures for the same causes have occurred. Removing fractions of failures in this manner is labeled **failure discounting**. If performed under a specific scenario, the effect of failure discounting on the accuracy of specific reliability growth models can be analyzed using Monte Carlo simulations. Weighting data is an alternative to failure discounting. It has been employed in numerous statistical settings. In this paper we compare the accuracy of several reliability growth models and assess the effect of failure discounting and weighting on their accuracy.

FAILURE DISCOUNTING METHODS

Two failure discounting methods are used in this paper. For each failure cause, the Fraction Discount (FD) method removes an additional fraction from the current fractional value of a previous failure, each time that an additional number, N, of tests are accumulated without a failure for the same cause. After M successful tests without failure for the same cause, the current value of the failure is $(1-F)^{I(N/M)}$ where $I(N/M)$ is the integer value of N/M.

The number of failure causes and associated probabilities of each failure occurrence are input parameters to the computer program which maintains a running tab on the number of successful tests since the previous failure for each failure cause.

For each failure cause, the Upper Confidence Bound (UCB) discounting method takes the upper confidence bound for the probability of failure of that cause to be the fraction of failure that remains. The confidence level is denoted by $100\gamma\%$. The UCB is computed for each failure cause using attributes data. The UCB procedure is applied to obtain a new fractional failure value at the end of each additional group of N successful tests without failure for the same cause. UCB(γ) is given by

$$UCB(\gamma) = 1-(1-\gamma)^{1/M}$$

where M is the total number of successful tests without failure for the specified cause at the time the discounting is applied.

When a failure for the same cause reoccurs, the associated fractional failure is restored to 1.

DISCRETE MODELS

1. Background

A system under test has several failure causes. Attributes testing is performed until a predetermined number, r, of failures occur at which time the failure causes are determined. Corrective action is taken and the next phase of testing is performed. R_K denotes system reliability in phase K. The number of tests $N_K(r)$ in phase K has a negative binomial distribution which we designate by $N_K(r) \approx NB(r, R_K)$. We drop the subscript K and suppose $N(r)$ is $NB(r,R)$. In this case, it is well known that the MLE $\hat{R}^*$, for R, and the MVUBE $\hat{R}'$ for R are given by

$$\hat{R}^* = \frac{N(r)-r}{N(r)} \tag{1}$$

$$\hat{R}' = \begin{cases} \dfrac{N(r)-r}{N(r)-1} & \text{if } N(r) > r \\ 0 & \text{if } N(r) = r \text{ when } r > 1. \end{cases} \tag{2}$$

In development programs r may be equal to 1. In this case

$$\hat{R}' = \begin{cases} 1 & \text{if } N(1) > 1 \\ 0 & \text{if } N(1) = 1 \end{cases}$$

This is not a satisfactory estimate of R. A more desirable estimator for R is obtained by writing $R = 1-e^{-A}$. Then

$$\hat{R} = 1-e^{-\hat{A}}, \text{ where} \tag{3}$$

$\hat{A}$ is unbiased for A. For the case r=1 (see Reference 9)

$$\hat{A} = \begin{cases} 0 & \text{if } N(1) = 1 \\ 1+\frac{1}{2}+\ldots+\frac{1}{N(1)-1} & \text{if } N(1) > 1 \end{cases} \tag{4}$$

For the case $r > 1$, the minimum variance unbiased estimator for A is (Reference 9),

$$\hat{A}(r) = \begin{cases} 0 & \text{if } N(r) = r \\ \frac{1}{r}+\frac{1}{r+1}+\ldots+\frac{1}{N(r)-1} & \text{if } N(r) > r \end{cases} \tag{5}$$

Table 1 displays values of $\hat{R}^*$ and $\hat{R}$ for values of $N(1)$.

TABLE 1. VALUES OF $\hat{R}^*, \hat{R}, N$

N	$\hat{R}^*$	$\hat{R}$
5	.800	.875
10	.900	.941
20	.950	.971
150	.993	.996

It is shown in Reference 9 that

a) $E(\hat{R}^*) = 1+\frac{1-R}{R}\ln(1-R)$

b) $\hat{R}^* < \hat{R}$ for all $N > 1$

c) $E(\hat{R}^*) < E(\hat{R}) < R$

d) $MSE(\hat{R}^*) < MSE(\hat{R}^*)$ where MSE is mean squared error.

Consequently, $\hat{R}$ appears to be a better estimator for R than the MLE, $\hat{R}^*$.

2. Exponential Regression (ER) Model

In this model R_K denotes system reliability in phase K. The ER model is

$$R_K = 1-\exp\{-(\alpha+\beta K)\}. \quad (6)$$

R_0 denotes system reliability prior to any corrections.

Let r_K = number of failures in phase K (fixed in advance). Let N_{jK} denote the number of tests between failure number $j-1$ and j in phase K. An unbiased estimator for $\alpha+\beta K$ using the j^{th} sequence of these tests is

$$y_{jK} \equiv \begin{cases} 0 & \text{if } N_{jK}=1 \\ 1+\frac{1}{2}+\ldots+\frac{1}{N_{jK}-1} & \text{if } N_{jK}\geq 2 \end{cases} \quad (7)$$

for $j=1,2,\ldots,r_K$. Therefore

$$\bar{y}_K = \left(y_{1K}+\ldots+y_{r_K K}\right)/r_K \quad (8)$$

is unbiased for $(\alpha+\beta K)$. Alternatively, one can use the minimum variance unbiased estimator y_k' given by (reference 9)

$$y_k' \equiv \frac{1}{r_K}+\frac{1}{r_K+1}+\ldots+\frac{1}{N_K(r_K)-1} \quad (9)$$

For r_K = 1, $\bar{y}_K$ and y_k' are identical. For $r_K \geq 3$, $\bar{y}_K$ and y_K have nearly equal mean square errors. The least squares estimates $\hat{\alpha}_K$ and $\hat{\beta}_K$ for α and β at the end of testing in the K^{th} phase are given by

$$\hat{\beta}_K = \frac{\sum_{j=0}^{K}(j-\bar{K})y_j}{\sum_{j=0}^{K}(j-\bar{K})^2} \quad (10)$$

$$\hat{\alpha}_K = \bar{\bar{y}}_K - \hat{\beta}_K\bar{K}$$

where $\bar{\bar{y}}_K = \left(\sum_0^K y_j\right)/(K+1)$ and $\bar{K} = \left(\sum_0^K j\right)/(K+1) = K/2.$

The resulting ER estimate for R_K is given by

$$\hat{R}_K = \begin{cases} 1-\exp\{-(\hat{\alpha}_K+\hat{\beta}_K K)\} & \text{for } K \geq 1. \\ 1-\exp\{-y_0\} & \text{for } K=0. \end{cases} \quad (11)$$

The term y_j denotes either unbiased estimator $\bar{y}_j$ or y_j' from phase j defined in equations (8) and (9).

The ER model has compared favorably with other discrete RG models in a study by Corcoran and Read (3). It was found to be superior in some respects to an AMSAA discrete model (reference 5) by Thailieb (8).

Equations (8) and (9) assume that $N_K(r_K)$ is the number of Bernoulli trials to r_K failures. The number r_K of system failures gets reduced under failure discounting and can become a fraction or mixed fraction as more successful testing is performed. Failure discounting occurs within a phase and across phases. When this happens, and the current phase is K, the values of $N_j(r_j)$ are adjusted in all previous phases to get *integers* r_j^* and an $N_j^*(r_j^*)$ so that

$$\frac{r_j^*}{N_j(r_j^*)} = \frac{r_j}{N_j(r_j)}$$

where r_j^* = smallest integer greater than r_j the current discounted (fractional) value of a failure.

3. An MLE Model

Some RG models accumulate all of the test data into one group to form an estimate $\hat{R}_{TT}$ for the current system reliability. Both the AMSAA discrete and continuous models do this. Other models in use today also have this feature. It is appealing to accumulate data when failures are discounted, because the act of discounting failures suggests that the data has been adjusted so that it all comes from the same population.

The model labeled MLEFD in this paper estimates current reliability by the expression

$$\hat{R}_K^* = \frac{N_{TK}-F_{TK}}{N_{TK}}$$

where N_{TK} = total number of tests across phases 1, 2, ..., K and

F_{TK} = current sum of all failures across phases 1, 2, ..., K.

F_{TK} may be adjusted to account for failure discounting. They are adjusted the same way in this model as discussed for the ER Model for simulation purposes.

WEIGHTING METHODS

Two weighting methods were applied to the discrete models. In both methods

W_K = weight allotted to data in phase K.

In Method I $\quad W_K = (1/\hat{\sigma}_K^2)/\sum_{i=0}^{K}(1/\hat{\sigma}_i^2).$

In Method II $\quad W_K = \hat{\sigma}_K^2/\sum_{i=0}^{K}\hat{\sigma}_i^2.$

In the discrete models $\hat{\sigma}_K^2 = r_K\hat{R}_K/(1-\hat{R}_K)$, consequently, if R_K increases in K, Method I assigns greater weight to early data and Method II assigns greater weight to more recent data. Equations for $\hat{\alpha}_K$ and $\hat{\beta}_K$ using weighted data are supplied in reference (6).

RESULTS FOR DISCRETE MODELS

Figures 1 through 5 display results of some of our computer simulations. There are 10 test phases in each case simulated. For each model, each scenario (set of input parameters) was replicated 500 times to obtain 500 values for each of the ten phase reliability estimates. The values of reliability along the vertical axis of the curves are the averages of these 500 values; that is, $\bar{\hat{R}}_1$, $\bar{\hat{R}}_2$, $\bar{\hat{R}}_3$, ..., $\bar{\hat{R}}_{10}$. Sample standard deviations were also computed. One standard deviation curve is provided in this report, but the computer program always makes these computations. For example in Figure 1, No Discounting, the average reliability estimate, $\bar{\hat{A}}_6$ in phase 6 for the EXP REG model is about .59. The true reliability assigned to the system for this phase of testing was about .64. the value of .64 is obtained by multiplying the probabilities assigned to the five failure causes for phase 6. All cases displayed in Figures 1 through 5 have five failure causes in each of 10 phases. System go-no-go tests are performed until the first failure occurs (r = 1), at which time testing in that phase terminates. The symbol I in the case title below each graph denotes the discount interval; i.e., the number of successful tests required between discounting. F denotes the discount rate; the fraction of the current failure value removed each time a discounting is performed. Figure 1 includes a graph of the sample standard deviations of the 500 values of $\hat{R}_i$ for i = 1, 2, ..., 10 for each of the three models. Standard deviation curves like this one were computed for each of the 150 cases simulated. The expression CI = .8 refers to the 80% confidence interval discounting method. Figures 1-4 display the optimistic effects of discounting frequently I = 1, I = 3 for a cumulative model such as the MLEFD model. They also give a clear indication that the ER model is not affected nearly as much by failure discounting as in the MLEFD model.

Figure 5 displays two sample results of simulation on the ER model a) without weighting and b) with weighting by methods I and II described earlier. In general, the weighted simulation runs indicate that method II is superior to method I for weighting data for the ER model.

The graphs in Figure 1-4 were selected from Chandler (2) which includes a listing of the computer program that was used to perform the simulations depicted in Figures 1-4 and a set of users instructions for his computer program. The graphs in Figure 5 were selected from Markiewicz (6) which includes a listing of her computer program with users instructions.

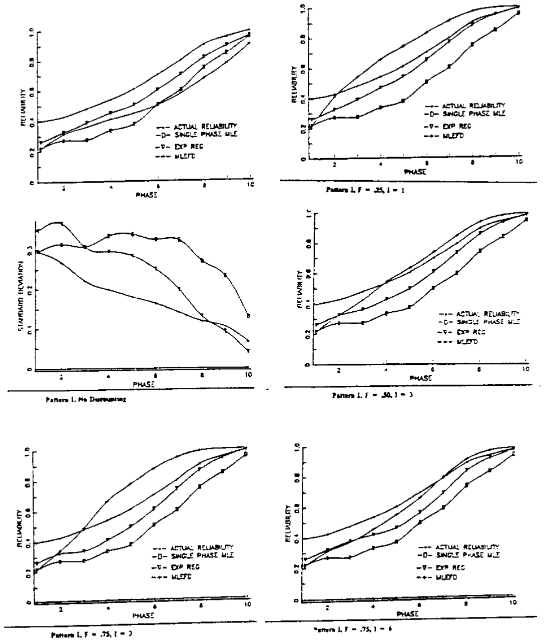


Figure 1

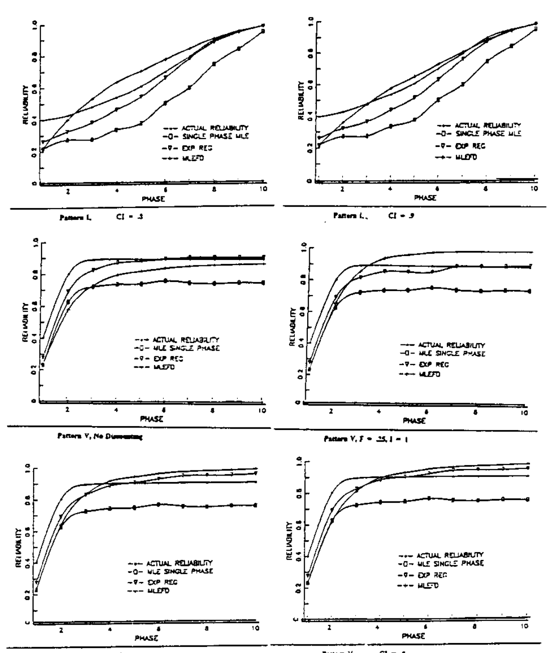


Figure 2

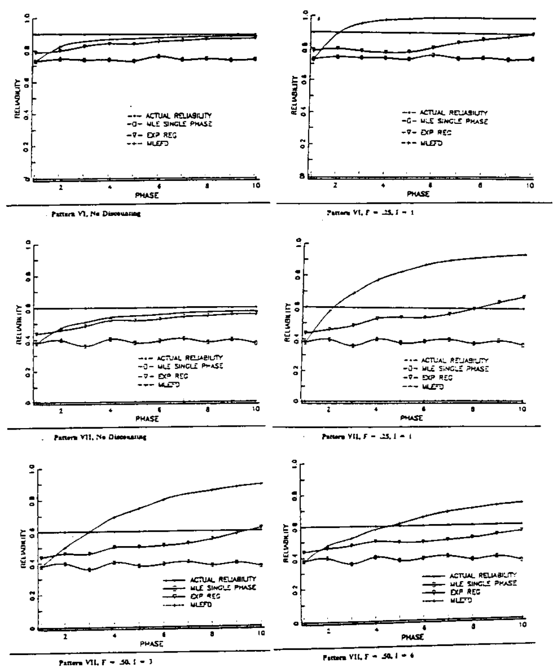


Figure 3

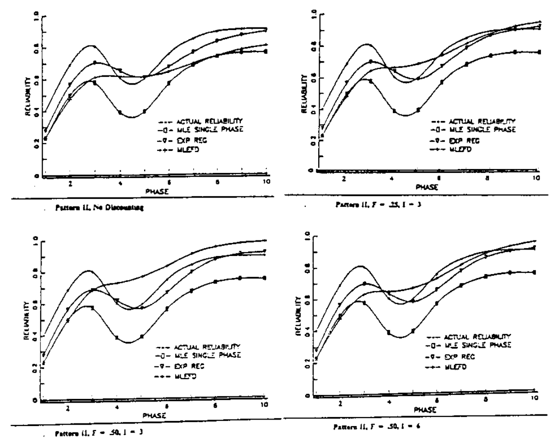
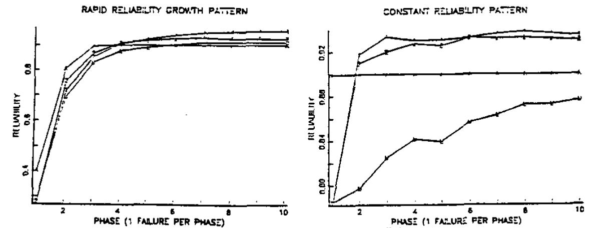


Figure 4



Weighting Methods One and Two and Rapid Reliability Growth Pattern

Weighting Methods One and Two and the Constant Reliability Pattern

Figure 5

CONTINUOUS MODELS

Two continuous RG models were analyzed. One model is the well-known AMSAA instantaneous model

$$\lambda_{TT} = (1-a)b(TT)^{-a} \tag{12}$$

where TT is the total test time. This is a cumulative RG Model. All test time is accumulated to obtain the instantaneous estimate of the failure rate. Equations for $\hat{a}$ and $\hat{b}$, the estimates of a and b, in this standard AMSAA model are as follows

$$\begin{aligned} 1-\hat{a} &= N \Big/ \left(\sum_{i=0}^{K-1} \log(X_k / X_i) \right) \\ \hat{b} &= N / (X_k)^{1-\hat{a}} \end{aligned} \tag{13}$$

where N = total number of failures and X_i = total test time up to the i^{th} failure. X_k denotes TT.

The second continuous model is not cumulative. It obtains estimates of $\hat{\lambda}_K$ for Phase K using test data in phase K only and uses similar estimates from previous phases to obtain regression estimates $\hat{a}_K$ and $\hat{b}_K$ for a and b. We call it the Modified AMSAA Model. We assume N_i items are tested until r_i fail in phase i. Early in development, this test plan is often the one used (with $N_i = 1$). Let $TT_K = \sum_{i=0}^{K} T_i$ where T_i denotes total test time in test phase i. The model is

$$\lambda_{TT_K} = (1-a)b(TT_K)^{-a} \tag{14}$$

For each phase i compute

$$\hat{\lambda}_i = \begin{cases} .5 / T_i & \text{if } F_i \le 1 \\ [(2F_i - 1)/2F_i] \cdot [F_i / T_i] & \text{if } F_i > 1 \end{cases} \tag{15}$$

where F_i denotes current number of failures in phase i. If no failure discounting is done then $F_i = r_i$.

$$\text{Let } y_i = \ln \hat{\lambda}_i,\ X_i = \ln TT_i,\ \overline{Y}_K = \left(\sum_{0}^{K} Y_i \right) / (K+1)$$

and $\overline{X}_K = \left(\sum_{0}^{K} X_i \right) / (K+1), i = 0.1,,,.K.$ The least squares estimates $\hat{a}_K, \hat{b}_K$ are given by

$$\hat{a}_K = \left[\Sigma X_i Y_i - \overline{Y}_K \Sigma X_i \right] / \left[\overline{X}_K \Sigma X_i - \Sigma X_i^2 \right] \tag{16}$$

$$\hat{b}_K = \frac{1}{1-\hat{a}_K} \exp\left(\overline{Y}_K + \hat{a}_K \overline{X}_K \right)$$

for K = 1, 2, ... The equations for $\hat{a}_K$ and $\hat{b}_K$ can be readily programmed on a hand-held calculator or a microcomputer.

1. Weighting Data

For simulations of the Modified AMSAA Model, two weighting methods were used. One method permits the user to choose a value p in (0,1) for which the computer generates weights $w_i = (w_i^*) / \sum_{1}^{K} w_i^*$ where $w_i^* = p^{\frac{1}{i}}$. The second method computes the weights from the data using $w_i = \hat{\sigma}_i^2 / \sum_{1}^{K} \hat{\sigma}_i^2$ where $\hat{\sigma}_i^2 = 1/\hat{\lambda}_i^2$, $\hat{\lambda}_i$ given by equation 15, and the index i refers to test phase i.

With weighting, equation (16) changes as follows:

$$\begin{aligned} \hat{a}_K &= \left[\sum_{0}^{K} w_i (X_i - \overline{X}_{wK}) Y_i \right] \Big/ \left[\sum_{i=0}^{K} w_i (X_i - \overline{X}_{wK})^2 \right] \\ \hat{b}_K &= \left[\exp(\overline{Y}_{wK} + \hat{a}_K \overline{X}_{wK}) \right] / (1-\hat{a}_K) \end{aligned} \tag{17}$$

where $\overline{X}_{wK}$ and $\overline{Y}_{wK}$ are the averages of the $w_i X_i$ and $w_i Y_i$. The standard AMSAA Model was not evaluated for data weighting.

2. Failure Discounting

Failure discounting applied to the AMSAA Model was very simple. Other discounting methods have been proposed; e.g., see Crow (4). The equations for $\hat{a}$ and $\hat{b}$ in the AMSAA model are given in equation (13).

Discounting of the AMSAA model was performed by replacing N in equation (13) by the sum of all of the current discounted failure values each time $\hat{\lambda}_{TT}$ was computed during the evaluation process. This provides a sequence of estimates, $\hat{\lambda}_{TT_1}$, $\hat{\lambda}_{TT_2}$, ... as testing is completed in each phase and failures are discounted and current estimates $\hat{a}_K$, $\hat{b}_K$ are computed.

The failure discounting method for the Modified AMSAA Model is more complex. For this model, a fraction, f, of a current value of a failure is removed for each specified amount of additional test time, TR, without failure for the same cause. If F was the last updated value of a failure and a total of TSF time has been accumulated since the last update, then the new adjusted value of F, say F_{adj}, is given by

$$F_{adj} = F \cdot (1-f)^{INT(TSF/TR)}$$

where $INT(x)$ is the largest integer less than or equal to x.

RESULTS FOR CONTINUOUS MODELS

Figure 6 shows the results for some failure rate patterns that were simulated. In each case, there were five failure causes and 10 test phases. In each phase, five items were tested until two failed assuming an exponential distribution. The code letter E in Pattern 1E, 2E, 7E refers to discounting parameters f = .50 and TR = 3 which is the greatest level of discounting simulated. Figure 6 shows the results of the simulations for failure rate patterns 1, 2, and 3 under this E level of discounting.

The results in Figure 6 indicate the following:

a) For all of these failure rate patterns, the Modified AMSAA without weighting or discounting appears to track failure rate decrease more closely than the AMSAA model.

b) Failure discounting appears to improve the AMSAA Model slightly for Patterns 1 and 2, but is optimistic for pattern 7 in the later phases.

c) Failure discounting does not consistently improve the Modified AMSAA Model.

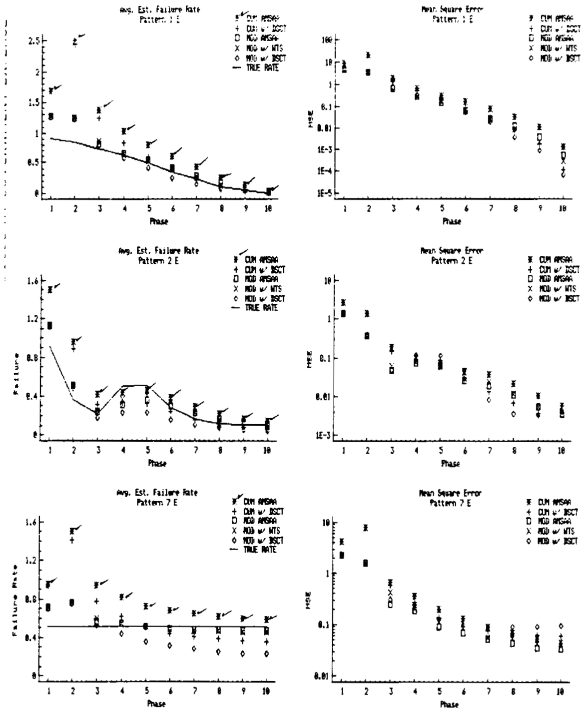


Figure 6

SUMMARY AND CONCLUSIONS

The results of the research reported here indicate the following for the models analyzed:

a) The non-cumulative growth models tended to track different growth patterns better than did the cumulative models.

b) Failure discounting has a greater impact on the cumulative growth models than on the non-cumulative models.

c) Weighting the data can have a significant impact on both cumulative and non-cumulative growth models.

d) Requiring five or more successful tests between failure discounting substantially reduces the hazards of failure discounting.

Computer programs exist that can be used to evaluate the impact of proposed weighting or failure discounting scenarios on the accuracy of reliability growth models. these and other simulation methods can be used to answer "what if" questions pertaining to the accuracy of reliability growth models.

REFERENCES

1. Drake, J. E., *Discrete Reliability Growth Models using Failure Discounting*, M.S. Thesis, Naval Postgraduate School, Monterey, CA, September 1987.

2. Chandler, James D., Jr., *Estimating Reliability with Discrete Growth Models*, M.S. Thesis, Naval Postgraduate School, Monterey, CA, March 1988.

3. Corcoran, W. J. and Read, R. R., *Comparison of Some Reliability Growth Estimation and Prediction Schemes*, United Technology Center Report Addendum UTC2140, 1967.

4. Crow, L. H., "Reliability Growth Prediction from Delayed Fixes," *Proceedings of the Annual Reliability and Maintainability Symposium*, 1983.

5. Crow, L. H., *AMSAA Discrete Reliability Growth Model*, U.S. Army Material Systems Analysis Activity, Aberdeen Proving Ground, MD, 1983.

6. Markiewicz, Pamela A., *Discrete Reliability Growth*, M. S. Thesis, Naval Postgraduate School, Monterey, CA September 1988.

7. Negus, Scott, *Evaluation of a Modified AMSAA Continuous Reliability Growth Model using Failure Discounting and Weighting Factors*, M.S. Thesis, Naval Postgraduate School, Monterey, CA, September 1989 (in preparation).

8. Thalieb, Rio M., An Accuracy Analysis of Army Material Systems Analysis Activity Discrete Reliability Growth Model, M.S. Thesis, Naval Postgraduate School, Monterey, CA, March 1988.

9. Woods, M., Chandler, J. and Drake, J., (1988), *Analysis and Evaluation of Discrete Reliability Growth Models With an Without Failure Discounting*, Naval Postgraduate School Technical Report NPS55-88-013.

BIOGRAPHY

Prof. W. M. Woods (55Wd)
Department of Operations Research
Naval Postgraduate School
Monterey, CA 93943

W. M. Woods has taught numerous courses in reliability, product assurance, statistical quality control and statistics. He has supported DoD agencies in various product quality areas through his research since 1962 when he first joined the Naval Postgraduate School. Some of his methods for system reliability confidence intervals, sampling inspection, reliability growth and CEP analysis are still in use today.