Department of Defense Handbook
Reliability Growth Management

This handbook is for guidance only.
Do not cite this document as a requirement.

FOREWORD

1. This handbook is approved for use by all Departments and Agencies of the Department of Defense (DoD).

2. Reliability growth management procedures have been developed to improve the reliability of DoD weapon systems. Reliability growth techniques enable acquisition personnel to plan, evaluate and control the reliability of a system during its development stage. The reliability growth concepts and methodologies presented in this handbook have evolved over the last few decades by actual applications to Army, Navy and Air Force systems. Through these applications, reliability growth management technology has been developed to the point where considerable payoffs in system reliability improvement and cost reduction can be achieved.

3. This handbook provides procuring activities and development contractors with an understanding of the concepts and principles of reliability growth, advantages of managing reliability growth, and guidelines and procedures to be used in managing reliability growth. It should be noted that this handbook is not intended to serve as a reliability growth plan to be applied to a program without any tailoring. This handbook, when used in conjunction with knowledge of the system and its development program, will allow the development of a reliability growth management plan that will aid in developing a final system that meets its reliability requirements and lowers the life cycle cost of the fielded systems.

4. Because of the brevity of the handbook, detailed development of underlying theory and estimation procedures are not provided. More extensive details may be found in the literature cited.

5. Comments, suggestions, or questions on this document should be addressed to the U.S. Army Materiel System Analysis Activity (AMSAA), ATTN: RDAM-LR ,
392 Hopkins Road Aberdeen Proving Ground MD, 21005-5071, or emailed to
amsaa.reltools@us.army.mil. Since contact information can change, you should verify the currency of the information above using ASSIST Online database at is
https://assist.daps.dla.mil/online/.

CONTENTS

FIGURES

TABLES

1. SCOPE.

1.1 Purpose.
This guide provides an understanding of the concepts and principles of reliability growth. Guidelines and procedures to be used in managing reliability growth are also presented. This guide is not intended to serve as a reliability growth plan to be applied to a program without any tailoring. When used in conjunction with knowledge of the system and its acquisition program, it will allow for the development of a reliability growth management plan that results in a final system that meets its requirements and lowers the life cycle costs of the fielded system.

1.2 Application.
This guide is intended for use on systems/equipment during the development phase by both producer and consumer personnel.

2. APPLICABLE DOCUMENTS.

2.1 General.
The documents listed below are not necessarily all of the documents referenced herein, but are those needed to understand the information provided by this handbook.

2.2 Government Documents.
The following Government documents, drawings, and publications form a part of this document to the extent specified herein. Unless otherwise specified, the issues of these documents are those cited in the solicitation or contract.

> *DOD Guide for Achieving Reliability, Availability, and Maintainability*, August 3, 2005.

2.3 Non-Government publications.
The following documents form a part of this document to the extent specified herein.

GEIA-STD-0009, "Reliability Program Standard for Systems Design, Development, and Manufacturing," August 01, 2008.

IEEE Std 1332-1998, "IEEE standard reliability program for the development and production of electronic systems and equipment," 1998.

3. DEFINITIONS.

3.1   Reliability.
Reliability is the probability that an item will perform its intended function for a specified period of time under the conditions stated in the Operational Mode Summary/Mission Profile (OMS/MP).

3.2   Operational Mode Summary/Mission Profile.
An OMS/MP projects the anticipated mix of ways a system will be used for each moment of time to include both peacetime and wartime. It also includes the percentage of time the system will be exposed to each type of environmental condition and movement terrain.

3.3   Reliability Growth.
Reliability growth is the positive improvement in a reliability parameter over a period of time due to implementation of corrective actions to system design, operation or maintenance procedures, or the associated manufacturing process.

3.4   Reliability Growth Management.
Reliability growth management is the management process associated with planning for reliability achievement as a function of time and other resources, and controlling the ongoing rate of achievement by reallocation of resources based on comparisons between planned and assessed reliability values.

3.5   Repair.
A repair is the refurbishment of a failed part or replacement of a failed part with an identical unit in order to restore the system to be fully mission capable.

3.5.1   Fix.
A fix is a corrective action that results in a change to the design, operation and maintenance procedures, or to the manufacturing process of the item for the purpose of improving its reliability.

3.5.2   Failure Mode.
A failure mode is the failure mechanism associated with a potential or observed failure. Failures due to an individual failure mode may exhibit a given failure rate until a corrective action (termed a fix) is made in the design, operation, maintenance, or manufacturing process that mitigates the failure mechanism.

3.5.2.1 A-mode.
An A-mode is a failure mode that will not be addressed via corrective action.

3.5.2.2 B-mode.
A B-mode is a failure mode that will be addressed via corrective action. One caution is that a B-mode failure corrective action developed and implemented during the test program may not be fully compliant with the planned production model. Such corrective actions are typically

2

referred to as interim, short-term or non-tactical fixes. While such fixes may appear to improve the reliability in test, the final production (i.e. long-term or tactical) fix would need to be tested to assure adequacy of the corrective action.

3.6   Fix Effectiveness Factor (FEF).
A FEF is a fraction representing the reduction in an individual initial mode failure rate due to implementation of a corrective action.

3.7   Growth Potential (GP).
The GP is a theoretical upper limit on reliability which corresponds to the reliability that would result if all B-modes were surfaced and fixed with the realized failure mode FEF values.

3.8   Management Strategy (MS).
MS is the fraction of the initial system failure intensity (rate of occurrence of failures) due to failure modes that would receive corrective action if surfaced during the developmental test program.

3.9   Growth Rate.
A growth rate is the negative of the slope of the graph of the cumulative failure rate versus the cumulative test duration for an individual system plotted on log-log scale. This quantity is a metric that reflects the rate at which the system's reliability is improving as a result of implementation of corrective actions. A growth rate between (0,1) implies improvement in reliability, a growth rate of 0 implies no growth, and a growth rate less than 0 implies reliability decay. This concept of growth rate only pertains to growth models that assume a linear relationship between the expected cumulative failure rate and the cumulative test duration when plotted on a log-log scale.

3.10   Poisson Process.
A Poisson process is a counting process for the number of events, *N(t)*, that occur during the interval [0,*t*] where t is a measure of time. The counting process is required to have the following properties: (1) the number of events in non-overlapping intervals are stochastically independent; (2) the probability that exactly one event occurs in the interval $[t, t + \Delta t]$ equals $\lambda_t \cdot \Delta t + o(\Delta t)$ where λ_t is a positive constant, which may depend on t, and $o(\Delta t)$ denotes an expression of $\Delta t > 0$ that becomes negligible in size compared to Δt as Δt approaches zero; and (3) the probability that more than one event occurs in an interval of length Δt equals $o(\Delta t)$. The above three properties can be shown to imply that *N(t)* has a Poisson distribution with mean equal to $\int_0^t \lambda_s \, ds$, provided λ_s is an integrable function of s.

3.10.1   Homogeneous Poisson Process (HPP).
A HPP is a Poisson process such that the rate of occurrence of events is a constant with respect to time *t*.

3.10.2   Non-Homogeneous Poisson Process (NHPP).
A NHPP is a Poisson process with a non-constant recurrence rate with respect to time *t*.

3

3.11 Idealized Growth Curve.
An Idealized Growth Curve is a planned growth curve that consists of a single smooth curve portraying the expected overall reliability growth pattern across test phases.

3.12 Planned Growth Curve.
A Planned Growth Curve is a plot of the anticipated system reliability versus test duration during the development program. The Planned Growth Curve is constructed on a phase-by-phase basis and as such, may consist of more than one growth curve.

3.13 Reliability Growth Tracking Curve.
A reliability growth tracking curve is a plot of a statistical representation of system reliability consistent with the test data used to portray demonstrated reliability versus test duration.

3.14 Reliability Growth Projection.
Reliability growth projection is an assessment of reliability that can be anticipated at some future point in the development program. The rate of improvement in reliability is determined by (1) the on-going rate at which new failure modes are surfaced, (2) the effectiveness and timeliness of the corrective actions, and (3) the set of failure modes that are addressed by corrective actions.

3.15 Exit Criterion (Milestone Threshold).
An Exit Criterion is the reliability value that needs to be exceeded in order to enter the next test phase. Threshold values are computed at particular points in time, referred to as milestones or major decision points, which may be specified in terms of cumulative hours, miles, etc. Specifically, a threshold value is a reliability value that corresponds to a particular percentile point of an order distribution of reliability values. A reliability point estimate based on test failure data that falls at or below a threshold value (in the rejection region) indicates that the achieved reliability is statistically not in conformance with the idealized growth curve.

3.16 Notations.
Symbols used in formulas within this document include the following:
3.17

k – total number of potential B-modes

m – number of surfaced B-modes

T – total duration of a developmental test

$N(t)$ – number of failures by time t

$\rho(t)$ – expected failure intensity by time t

t_1 – length of the initial test phase

M_1 – average initial MTBF over initial test phase

M_G – goal MTBF

α – growth rate

φ_i – average failure rate for test phase i

MS – management strategy

μ_d – average fix effectiveness

$M(t)$ – number of B-modes surfaced by time t

$\mu(t)$ – expected value of M(t)

$\mu_k(t)$ – μ(t) based on assumption of k initial B-modes

$\Lambda_U(t)$ - system failure intensity for unsurfaced modes

$\Lambda(t)$ - system failure intensity, after mode mitigation

N_i – number of failures for failure mode i

N – the total number of failures (for all modes)

λ – initial system failure intensity

λ_i– true but unknown initial failure rate for mode i

$\hat{\lambda}_i$ - standard estimate of λ_i

$\tilde{\lambda}_i$ - the Stein estimate of λ_i

θ_S - true but unknown Stein shrinkage factor

$h(t)$ - expected rate of occurrence of new B-modes at time t

$h_k(t)$ - h(t) based on assumption of k initial B-modes

d_i - true but unknown fix effectiveness for mode i

$\rho_k(t)$ - approximation of ρ(t)

$MTBF(t)$ - MTBF at time t

$MTBF_k(t)$ - MTBF(t) based on assumption of k initial B-modes

4.   INTRODUCTION.

This handbook provides an abbreviated synopsis on methodology and concepts to assist in reliability growth planning and a structured approach for reliability growth assessments.

4.1   Why.
Reliability growth management procedures were developed to help guide the materiel acquisition process for new military systems. Generally, these systems require new technologies and represent a challenge to the state of the art. Striving to meet these requirements represents a significant portion of the entire acquisition process and, as a result, the setting of priorities and the allocation and reallocation of resources such as funds, manpower and time are often formidable management tasks.

4.2   What.
Reliability growth management procedures are useful for determining priorities and allocating resources. These techniques will enable the manager to plan, evaluate and control the reliability of a system during its development stage. The reliability growth concepts and methodologies presented in this guide have evolved through actual applications to Army, Navy and Air Force systems.

4.3   Layout.
This guide is written as an overview for both the manager and the analyst. The fundamental concepts are covered in section 1, with only the bare essential details regarding the implementation of these concepts discussed in sections 2 and 3.

5

4.4   Reliability Growth Planning.

Reliability growth planning addresses program schedules, amount of testing, resources available, and the realism of the test program in achieving the requirements. The planning is quantified and reflected in the construction of a reliability growth planning curve and the necessary supporting reliability activities. This curve establishes interim reliability goals throughout the program.

4.5   Reliability Growth Assessment.

To achieve reliability goals, it is essential that periodic assessments of reliability be made during the test program (usually at the end of a test phase) and compared to the planned reliability growth values.

4.6   Managing Reliability Growth.

4.6.1   Commitment and Involvement.

The essence of reliability growth management is commitment and involvement in all aspects of planning, evaluating, and controlling the reliability growth effort. Management controls the resources, and therefore directly affects the reliability growth effort. Of significant importance is the need for management to adequately resource reliability improvement up-front.

4.6.2   Controlling Reliability Growth.

Assessments provide visibility of achievements and focus on deficiencies while there is still time to affect the system design. By making appropriate decisions with regard to the timely incorporation of effective fixes into the system, commensurate with attaining the milestones and requirements, management can control the growth process.

4.6.3   Management's Role.

The various techniques associated with reliability growth management do not, in themselves, manage. The planned growth curve and milestones are only targets. Reliability will grow to these values only with the incorporation of an adequate number of effective fixes into the system. This requires dedicated management attention to reliability growth. In addition to how appropriately the system is tested, there are at least four planning elements under management control, including:

a) Management Strategy, MS, or the fraction of system initial failure rate addressed by corrective actions;
b) Rate at which failure modes are surfaced;
c) Turnaround time for analyzing and implementing corrective actions; and
d) Fix Effectiveness Factor, FEF, or the fraction reduction in the rate of occurrence of modes after corrective action.

High level management of reliability growth decisions in the following areas may be necessary in order to ensure that reliability goals are achieved:

a) Revise the program schedule;
b) Increase testing;
c) Fund additional development efforts;

6

d) Add or reallocate program resources; and
e) Stop the program until interim reliability goals have been demonstrated.

4.7   Basic Reliability Activities.
Reliability growth management is part of the system engineering process, but does not take the place of the other basic reliability program management structure and activities, such as:
- Reliability Engineering
- Apportionment
- Failure Modes and Effects and Criticality Analysis (FMECA)
- Stress analysis
- Laboratory component level testing
- Highly Accelerated Life Testing (HALT)
- Highly Accelerated Stress Testing (HASS)
- Environmental Stress Screening (ESS)
- Physics of Failure (PoF)
- Critical Items List/Analysis
- Software reliability assessment
- Failure Reporting and Corrective Action System (FRACAS)
- Fault Tree Analysis (FTA)
- Data collection and test monitoring
- Scoring and Assessment of RAM data

4.8   Benefits of Reliability Growth Management.
The following benefits can be realized by the utilization of reliability growth management:

4.8.1   Finding Unforeseen Deficiencies.
The initial prototypes for a complex system with major technological advances will invariably have significant reliability and performance deficiencies that cannot be foreseen in the early design stages. This is also true for prototypes that are "simply" the integration of existing systems.

4.8.2   Designing-in Improvement through Surfaced Problems.
Even if some potential problems can be foreseen, their significance might not. Prototypes are subjected to a development testing program to surface the problems that drive the rate of occurrence of failures (failure intensity) so that the necessary improvements in system design can be made. The ultimate goal of the development test program is to meet the system reliability and performance requirements.

4.8.3   Reducing the Risks Associated with Final Demonstration. Experience has shown that in many cases, programs that rely solely on a final demonstration to determine compliance with the reliability requirements do not achieve the reliability objectives given the allocated resources. Emphasis on reliability performance prior to the final demonstration using quantitative reliability growth could substantially increase the chance of passing a final demonstration, or could even replace a final demonstration.

4.8.4   Increasing the Probability of Meeting Objectives.
This can be achieved by setting interim reliability goals to be met during the development testing program and making the necessary allocation and reallocation of resources to attain these goals. A comprehensive approach to reliability growth management throughout the development program organizes this process.

4.9   Reliability Growth Process.

4.9.1   Basic Process.
Reliability growth is the result of an iterative design process. As the design matures, it is investigated to identify actual or potential sources of failures. Further design effort is then spent on these problem areas. The design effort can be applied to either product design or manufacturing process design. The iterative process can be visualized as a simple feedback loop, as shown in Figure 1. This illustrates that there are four essential elements involved in achieving reliability growth:

  a) Failure mode discovery;
  b) Feedback of problems identified;
  c) Failure mode root cause analysis and proposed corrective action; and
  d) Approval and implementation of proposed corrective action.
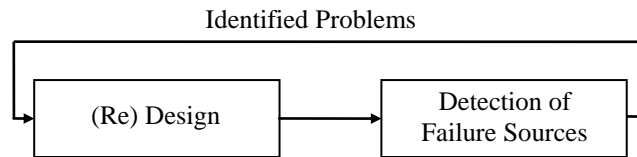
FIGURE 1. Reliability Growth Feedback Model.

Furthermore, if failure sources are detected by testing, another element is necessary:

  e) Fabrication of hardware.

Following redesign, detection of failure sources serves as verification of the redesign effort. This is shown in Figure 2.
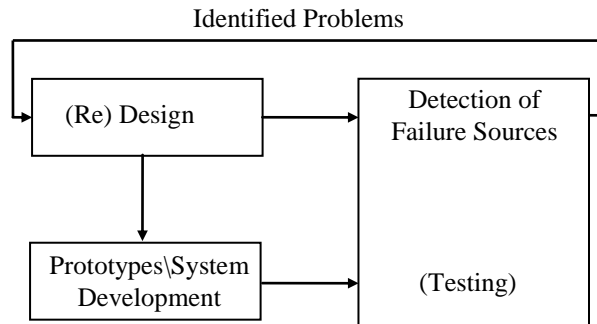
FIGURE 2. Reliability Growth Feedback Model

8

4.9.2   Classifying the Failure Modes.
When a system is tested and failure modes are observed, management can make one of two decisions: either not fix the failure mode or fix the failure mode. Therefore, the Management Strategy (MS) places failure modes into two categories: A-modes and B-modes. A-modes will not have corrective action taken, for example if failure modes are associated with commercial off-the-shelf (COTS) or legacy systems. B-modes, on the other hand, will be addressed via corrective action. Note that a failure mode may be initially classified as an A-mode, but subsequent conditions may change, causing management to reclassify it as a B-mode and address it via corrective action.

4.9.3   Decreasing the Failure Rate.
Growth is achieved by decreasing the failure rate. Since A-modes will not be addressed via corrective action, the failure rate for A-modes will not change. Thus only the B-mode corrective actions can accomplish growth. However, a corrective action that is developed and implemented for a B-mode will rarely totally eliminate the mode's failure rate. As a result, a metric is used to determine the fraction decrease in a mode's failure rate after corrective action implementation, known as the fix effectiveness factor (FEF). FEFs vary according to the commodity or technical area. Note that if an FEF is 0.70, on average, then the failure rate remaining would be 0.30 (or 1 – FEF) of the initial mode failure rate.

4.9.4   Attaining the Requirement.
An important question is: Can the requirement be attained with the planned Management Strategy and Fix Effectiveness Factor? In part, this can be answered by considering the growth potential (GP), which is the maximum reliability that can be attained with the system design, MS, and FEF. This upper limit on reliability, which may never actually be achieved in practice, is attained when all B-modes are found and their corrective actions are incorporated into the system with the specified FEF.

4.9.5   Factors Influencing the Growth Rate.
The rate at which reliability grows depends on how rapidly failure mode discovery, failure analysis, fabrication of systems, and retesting/verification is accomplished. That is, the rate at which a system's reliability is improved is a function of:
   a) The rate at which failure modes are surfaced during testing;
   b) The turnaround time associated with analyzing/implementing corrective actions:
       i.   Time associated with performing root cause analysis
       ii.   Time associated with the corrective action review and approval process
       iii. Time associated with physical implementation of approved corrective actions
   c) The fraction of initial failure rate addressed by corrective actions - MS; and
   d) The fraction by which the failure rate of fixed modes is reduced - FEF.

4.10   Reliability Growth Management Control Processes.
There are two basic ways to evaluate the reliability growth process – assessment and monitoring. The assessment approach is to quantitatively assess the current reliability based on information from the detection of failure sources. This approach is results oriented. The monitoring approach is to monitor activities in the process to assure that they are being accomplished in a timely manner and that the level of effort and quality of work are in compliance with the program plan.

9

The monitoring approach is activities oriented and is used to supplement the assessments. In the early stages of a program, the monitoring approach may be relied on entirely due to the lack of sufficient objective information. Each of these methods complements the other in controlling the growth process.

Figures 3 and 4 illustrate the assessment and monitoring management processes, respectively, in a skeleton form. The representation of an actual program or program phase may be considerably more detailed.

Identified Problems
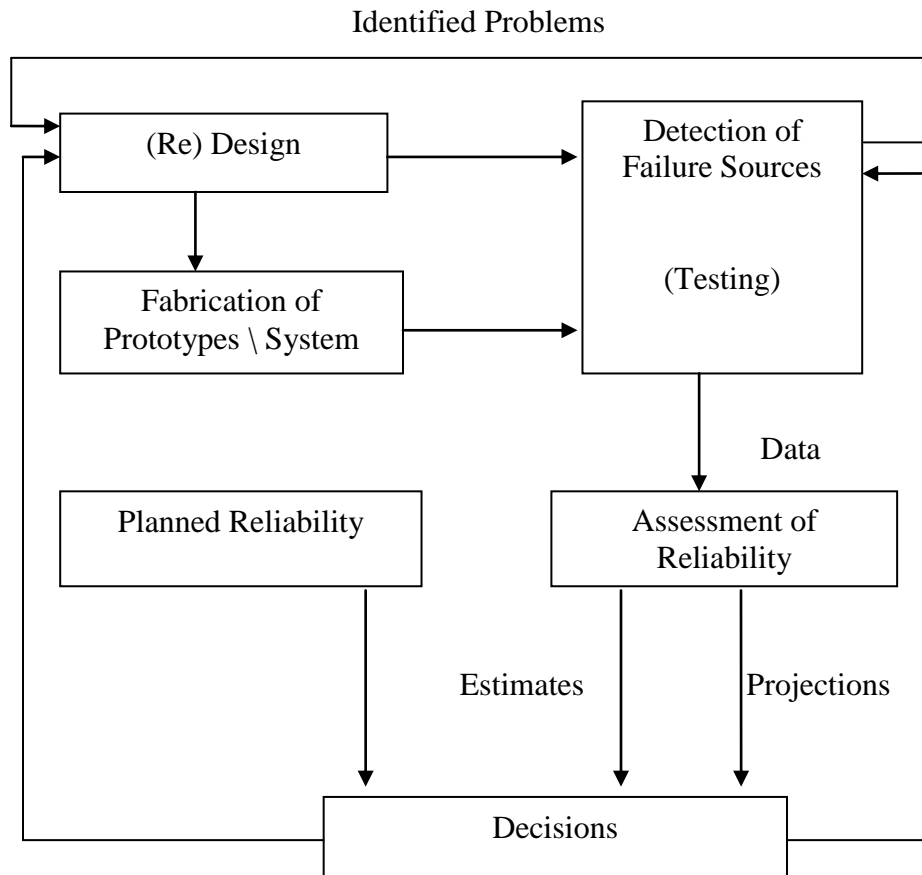
FIGURE 3. Reliability Growth Management Model (Assessment Approach).

10
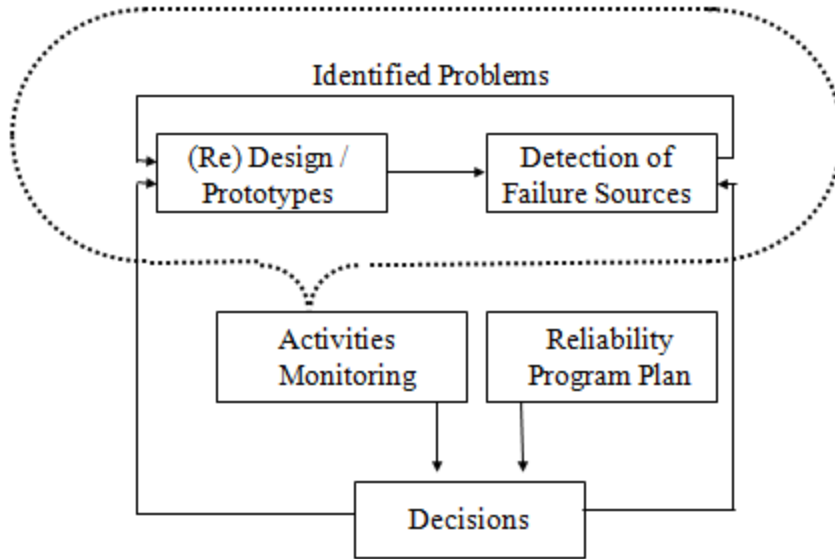
FIGURE 4. Reliability Growth Management Model (Monitoring Approach).

4.10.1 Assessment Approach.

Figure 3 illustrates how assessments may be used to control the growth process. Reliability growth management differs from conventional reliability program management in two major ways. First, there is a more objectively developed growth standard against which assessments are compared. Second, the assessment methods used can provide more accurate evaluations of the reliability of the current system configuration. A comparison between the assessment and the planned value will suggest whether the program is progressing as planned. If the progress is falling short, new strategies should be developed. These strategies may involve the reassignment of resources to work on identified problem areas, adjustment of the schedule, or a re-examination of the validity of the requirement. Figure 5 illustrates an example of both the planned reliability growth and assessments.
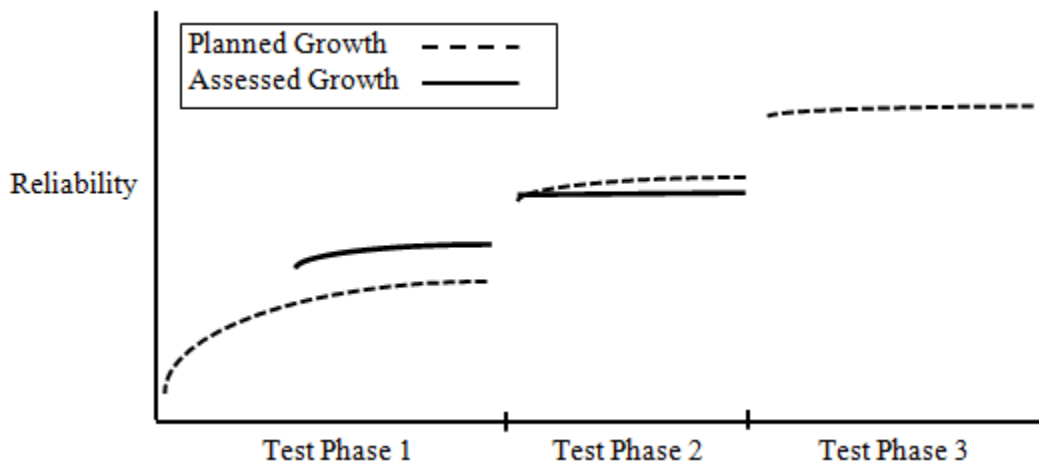
FIGURE 5. Example of Planned Growth and Assessments.

11

4.10.2 Monitoring Approach.
Figure 4 illustrates how monitoring growth activities may be used to control the growth process. This activity is a valuable complement to reliability assessments for a comprehensive approach to reliability growth management. Standards for level of effort and quality of work accomplishment must, of necessity, rely heavily on the technical judgment of the evaluator. Monitoring is intended to assure that the activities have been performed within schedule and meet appropriate standards of engineering practice. It is not intended to second-guess the designer, e.g., redo his stress calculations. A good example of a monitoring activity is the design review, which is a planned monitoring of a product design to assure that it will meet the performance requirements during operational use. Such reviews of the design effort serve to determine the progress being made in achieving the design objectives. Perhaps the most significant aspect of the design review is its emphasis on technical judgment, in addition to quantitative assessments of progress.

4.11 Factors Influencing the Growth Curve's Shape.
Such things as the current stage of the development program, the current test phase, the system configuration under test, the timing of corrective actions, and the units of measure for test duration all influence the growth curve's shape.

4.11.1 Stages of the Development Program.
Generally, any system development program is divided into stages, with different objectives for each stage. The names and objectives for each stage in a given development program need not be the ones given here. These stages are given as representative of a typical development program:

   a) Proposal: What are the requirements, can they be met, and if so, how and at what estimated cost?
   b) Conceptual: Experimental prototypes may bear little resemblance to the actual system. They are for proof-of-principle.
   c) Validation: Prototypes are built and tested to achieve the performance and reliability objectives for the system.
   d) Engineering and Manufacturing Development (EMD): Systems are built as though they are in production and are tested to work out final design details and manufacturing procedures.

Quantitative reliability growth management can be used during the Validation and EMD stages of the program. The different nature of the testing occurring during these stages may differ enough to cause different rates of growth to occur. The amount of difference will determine if they may be treated as part of the reliability growth planning curve.

4.11.2 Test Phases.
During the Validation and EMD stages, it is likely that testing will be broken up into alternating time periods of active testing, followed by corrective action periods (CAPs). Each period of active testing can be viewed as a testing phase. Safety related failure modes and failure modes that are readily understood and easily mitigated may be incorporated into the system during a test phase. Thus reliability growth may occur during a test phase. However, the most significant

12

growth will occur due to groups of failure modes that are scheduled for implementation in the CAP at the conclusion of a test phase. Within a development stage, it is likely that other types of testing will be occurring (e.g., performance testing). If these other tests follow the intended OMS/MP well enough, and if corrective actions are made on the basis of these tests, then the information gathered may be incorporated into the reliability growth test database. These would then contribute to the reliability growth testing phases. Due to the CAPs, it is to be expected that the reliability will grow from one phase to the next. The reliability growth planning curve should reflect this.

4.11.3 Test Phase Reliability Growth.
Based on the activities and objectives of the program, the reliability growth plan should indicate for each test phase the levels of reliability that are expected to be achieved, whether reliability is constant or growing, the objective at the end of the test phase, and whether corrective actions are delayed or incorporated in the test phase. There are three responses that can be made to each identified failure mode:

a.   Incorporate a corrective action during the test phase;
b.   Incorporate a corrective action after the test phase; or
c.   Incorporate no corrective action.

Figure 6 illustrates the effect of deferring corrective action from the test phase to a CAP. As more corrective actions are deferred, the effectiveness is reduced due to the inability to detect ineffective corrective actions and newly introduced failure modes. Thus some allowance should be made for the lesser effectiveness of delayed corrective action. It is especially important to schedule adequate calendar time for the CAP at the end of the test phase. The calendar time must be of sufficient duration to accommodate the expected number of delayed B-modes whose fixes are scheduled to be implemented during the CAP.
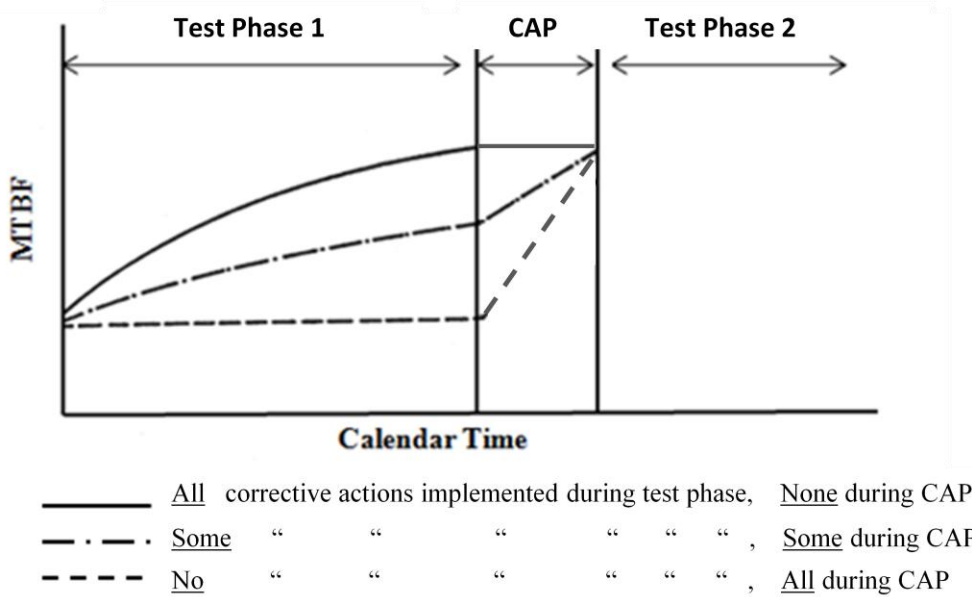
FIGURE 6. Effect of Deferring Corrective Action.

13

When working in terms of test time, a distinct effort involving one or more corrective actions will be shown as a vertical jump. It must be recognized, however, that a certain amount of calendar time is required to achieve the jump. This calendar time, covering test time and calendar time for corrective action to configuration 1, may be completely distinct from the calendar time used for testing, as illustrated in Figure 7. Time constraints may require that at least some of the calendar time is concurrent with the previous test phase, as illustrated in Figure 8. Overlapping corrective action and test in this fashion may yield a less effective corrective action, since it is started somewhat prematurely. The jump in MTBF due to the fixes implemented during the CAP will typically be largely determined by the collective B-mode failure intensity addressed and the average FEF realized during the CAP.
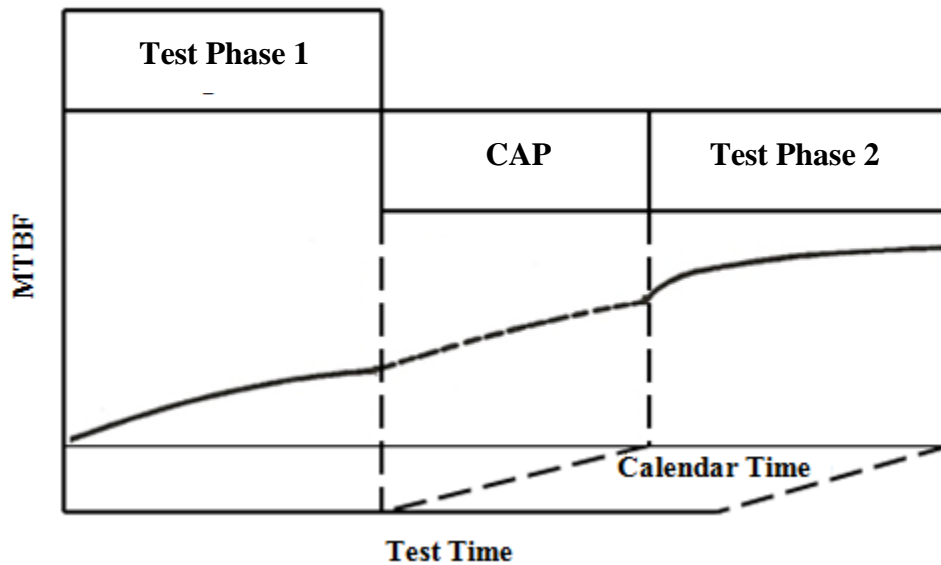
FIGURE 7. Calendar Time Accounting for Test Time and Time Required for Corrective Action
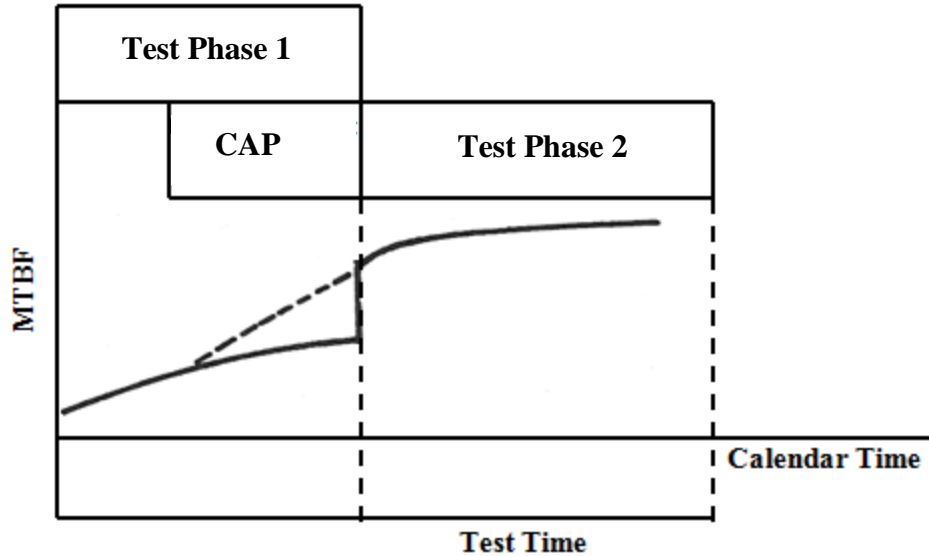
FIGURE 8. Calendar Time Accounting for Only Test Time

4.11.4 System Configuration.
In an absolute sense, any change to the design of a system (e.g. hardware, software, training procedures, maintenance procedures) constitutes a new configuration. For our purposes, a specific design will be termed a new configuration if there has been one significant design change or enough smaller design changes that cause an obviously different failure rate for the system. It is possible that two or more testing phases could be grouped together for analysis based on the configuration tested in these phases being substantially unchanged. It is also possible that one design change is so effective at increasing reliability that a new configuration could occur within a test phase. System configuration decisions can also be made on the basis of engineering judgment.

4.11.5 Timing of Fixes.
Fixes are intended to reduce the rate at which the system fails. Repairs make no change in system failure rate. The time of insertion of a fix affects the pattern of reliability growth.

4.11.5.1 Test-Fix-Test.
In a pure test-fix-test program, when a failure is observed, testing stops until a corrective action is implemented on the system under test. When the testing resumes, it is with a system that has incrementally better reliability. The graph of reliability for this testing strategy is a series of small increasing steps, with each step stretching out longer to represent a longer time between failures. Such a graph can be approximated by a smooth curve, as shown in Figure 9.

15
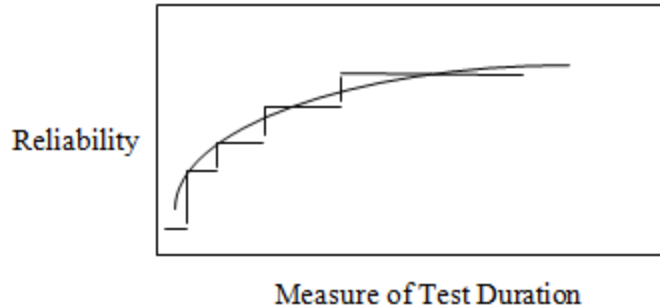
FIGURE 9. Graph of Reliability in a Test-Fix-Test Phase.

A pure test-fix-test program is impractical in most situations. Testing is likely to continue with a repair, and the fix will be implemented later. Nevertheless, if fixes are inserted in the test phase on a non-instantaneous basis but as soon as possible while testing is still proceeding, the stair-step like reliability increases and the shape of the approximating curve will be similar, but rise at a slower rate. This is due to the reliability remaining at the same level that it was at when the failure happened until the fix is inserted. Thus the steps will all be of longer length, but the same height. Continuing to test after the fix is inserted will serve to verify the effectiveness of the corrective action.

4.11.5.2 Test-Find-Test.
During a test-find-test program, the system is tested to determine failure modes. However, unlike the test-fix-test program, fixes are not incorporated into the system during the test. Rather, the fixes are all inserted into the system at the end of the test phase and before the next testing period. Since a large number of fixes will generally be incorporated into the system at the same time, there is usually a significant jump in system reliability at the end of the test phase. The fixes incorporated into the system between test phases are called delayed fixes. See Figure 10.
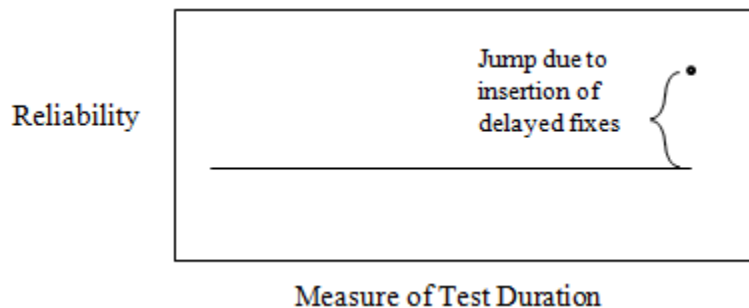
FIGURE 10. Graph of Reliability in a Test-Find-Test Program.

4.11.5.3 Test-Fix-Test with Delayed Fixes.
The test program commonly used in development testing employs a combination of the previous two types of fix insertions. That is, some fixes are incorporated into the system during the test, while other fixes are delayed until the end of the test phase. Consequently, the system reliability

16

will generally be seen as a smooth process during the test phase and then jump due to the insertion of the delayed fixes. See Figure 11.
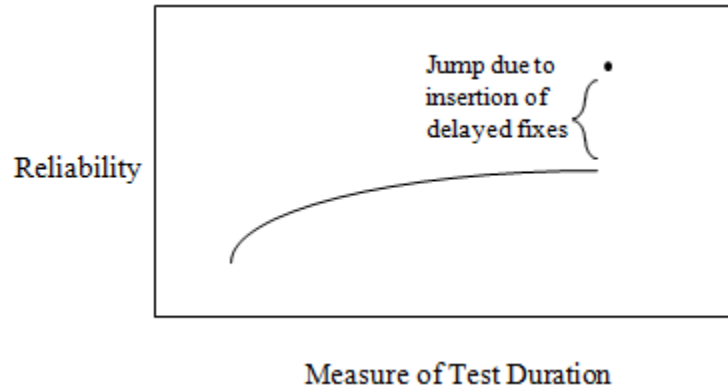
FIGURE 11. Graph of Reliability in a Test-Fix-Test Program with Delayed Fixes.

4.11.5.4 Example of Possible Growth Patterns Resulting from Varying the Timing of Fixes.

In order to reach the goal reliability, the development testing program will usually consist of several major test phases. Within each test phase, the fix insertion may be carried out in any one of the three ways discussed above. As an example, suppose that testing were conducted during the Validation and EMD stages of the program. Each stage would have at least one test phase, implying a minimum of two test phases for the program. In this case, there would be $3^2 = 9$ general ways for the reliability to grow during the development test. See Figure 12. Note that a development stage may consist of more than one distinct test phase. For example, testing may be stopped at several points during the EMD stage to allow for CAPs, during which delayed fixes are incorporated into the system. In such a case, testing would be comprised of a sequence of test phases, with each test phase separated by a CAP.
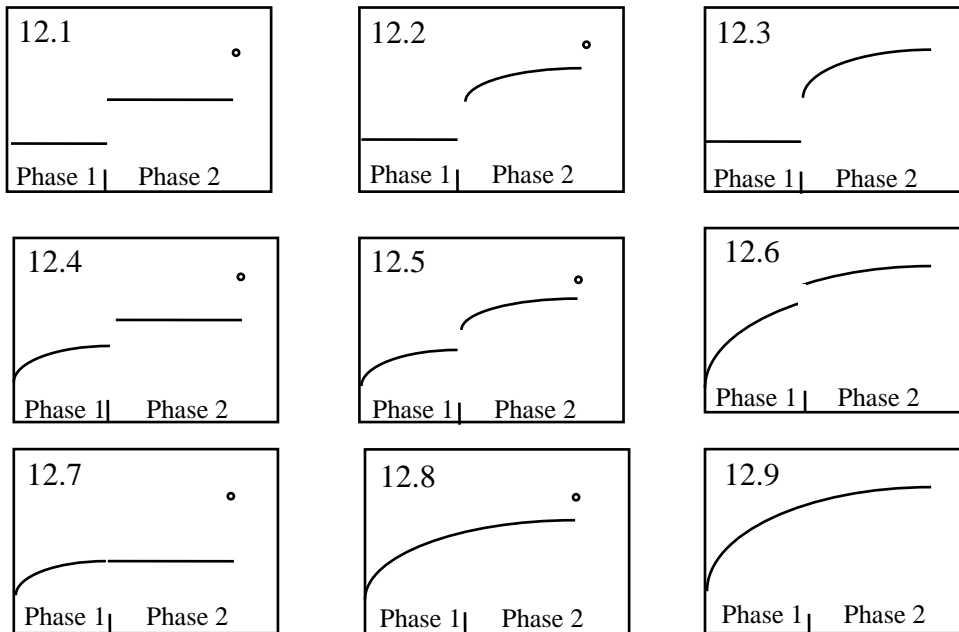
FIGURE 12. The Nine Possible General Growth Patterns for Two Test Phases.

Row 1 shows Phase 1 as having all fixes delayed until the end of the testing phase. Row 2 shows Phase 1 as having some fixes inserted during test and some delayed. Row 3 shows Phase 1 as having all fixes inserted during test, with none delayed. Column 1 shows Phase 2 as having all fixes delayed until the end of the testing phase. Column 2 shows Phase 2 as having some fixes inserted during test and some delayed. Column 3 shows Phase 2 as having all fixes inserted during test, with none delayed. Figures 12.1 and 12.9 represent the two extremes in possible growth test patterns.

4.11.5.5 Statistical Advantages of Test-Fix-Test.
There are some distinct statistical advantages to following a complete test-fix-test program:
  a) The estimated value of reliability at any point along the smooth growth curve is an instantaneous value. That is, it is not dragged down by averaging with the failures that accrued due to earlier (and hopefully) less reliable configurations.
  b) Confidence limits about the true value of reliability can be established.
  c) While the impact of the jumps in reliability can be assessed using a mix of some engineering judgment (this will be discussed in the section on Reliability Growth Projection) and test data, the estimate of reliability in a test-fix-test program is based solely on data.
  d) The effectiveness of corrective actions is continuously assessed in the estimate of reliability.

4.11.6 Growth Curve Re-initialization.
The differences in the growth curves between phases shown in Figures 12.5 and 12.6 represent the difference mentioned in Section 4.11.1 (Stages of the Development Program). Underlying Figure 12.6 is the assumption that the testing environment and engineering efforts are the same across test phases, thus the continuation of the same growth curve into the succeeding phase, after the jump for delayed fixes. In Figure 12.5 a factor influencing the rate of growth has substantially changed between the phases and is reflected in a new growth curve for the succeeding phase. This is called re-initializing the growth curve. It must be emphasized that re-initialization is only justified if the testing environment is so different as to introduce a new set of failure modes, or the engineering effort is so different as to be best represented as a totally new program.

4.11.7 Shape Changes Due to Calendar Time.
Reliability growth is often depicted as a function of test time for evaluation purposes. It may be desirable to portray reliability growth as a function of calendar time. This can be accomplished by determining the number of units of test duration that will have been completed at each measure point in calendar time and then plotting the reliability value that corresponds to the completed test duration above that calendar point. This is a direct function of the program schedule. Figure 13 shows the reliability growth of a system as a function of test time and calendar time.
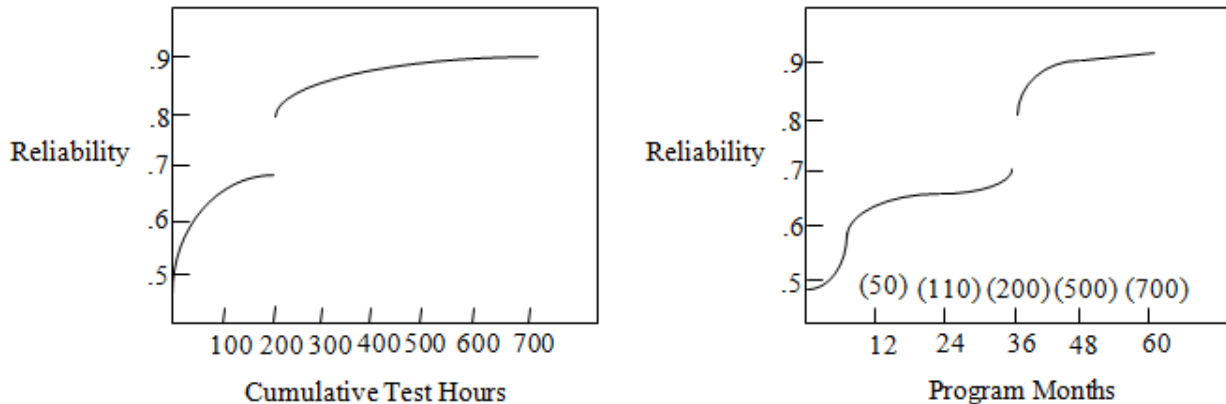
FIGURE 13. Comparison of Growth Curves Based on Test Duration Vs Calendar Time.

4.12   Reliability Growth Programmatic Concepts.

4.12.1 Levels of Consideration for Planning and Controlling Growth.   Planning and controlling reliability growth can be divided along both a program basis and an item under test basis. The appropriate level of consideration can vary at different times during the development. In addition, systems may be classed as to their usage.

   a) Program considerations:

      i.   Global: This approach treats reliability growth on a total basis over the entire development program.

      ii.   Local: This approach treats reliability growth on a phase-by-phase basis.

b) Item Under Test considerations:

      i.   System Level: The entire system as it is intended to be fielded is tested.

      ii.   Subsystem Level: The obvious meaning is the testing of a major and reasonably complex portion of the whole system (e.g., an engine for a vehicle). Sometimes, the subsystem would seem to be an autonomous unit, but because the requirement is for this unit to operate in conjunction with other units to achieve an overall functional goal, it is really only part of "the system" (e.g., radar for an air defense system).

c) Usage of System – continuous and discrete models:

      i.   Continuous models are those that apply to systems for which usage is measured on a continuous scale, such as time in hours or distance in miles.

      ii.   Discrete models are those that apply to systems for which usage is measured on an enumerative or classificatory basis, such as pass/fail or go/no-go. For discrete models, outcomes are recorded in terms of distinct, countable events that give rise to probability estimates.

4.12.2 Analysis of Previous Programs.

Analysis of previous similar programs is used to develop guidelines for predicting the growth during future programs. Such analysis may be performed on overall programs, individual program phases, or both. Of particular interest are the patterns of growth observed and the effect of program characteristics on initial values and other planning model parameters.

4.13   Reliability Growth Planning Concepts.

4.13.1 Planned Growth Curve.

The planned growth curve should portray a picture over the program Validation and EMD stages of a feasible reliability growth path, from an achievable initial reliability to a goal reliability that supports demonstrating the reliability requirement. It is an essential part of the reliability growth management methodology and is important to any reliability program. The planned growth curve is constructed early in the development program, generally before hard reliability data are obtained and is typically a joint effort between the program manager and contractor. Its primary purpose is to provide management with achievable reliability benchmarks at any point in the Validation and EMD program stages and to provide a basis for evaluating the actual progress of the reliability program based upon generated reliability data. The planned growth curve can be constructed on a phase-by-phase basis, as shown in Figure 14.
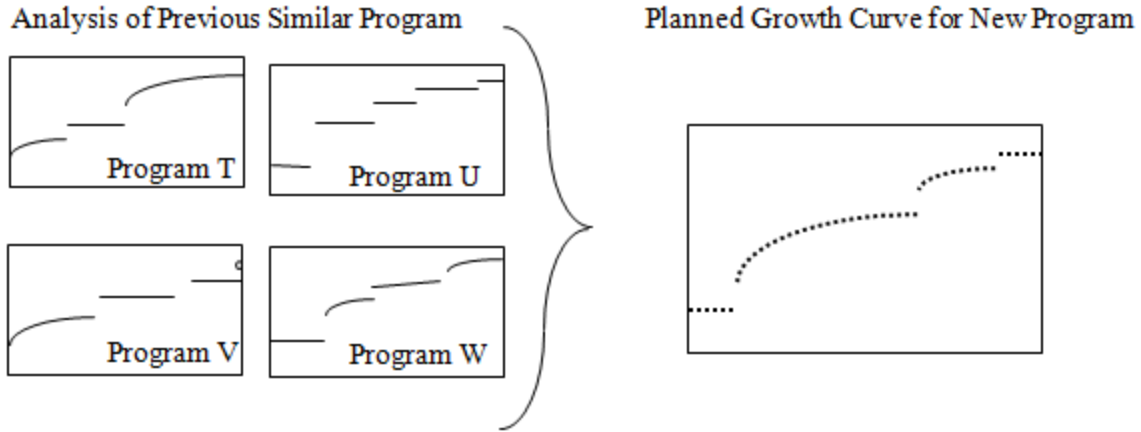
FIGURE 14. Development of Planned Growth Curve on a Phase by Phase Basis.

4.13.2 Idealized Growth Curve.
An Idealized Growth Curve is a planned growth curve that consists of a single smooth curve based on initial conditions, planned Management Strategy, and other growth model parameters. This curve is a strict mathematical function of the input parameters across the measure of test duration (e.g., time, distance, trials), thus the name "Idealized." No program can be expected to assume this exact mathematical ideal shape, but it is useful in setting interim goals. See Figure 15.
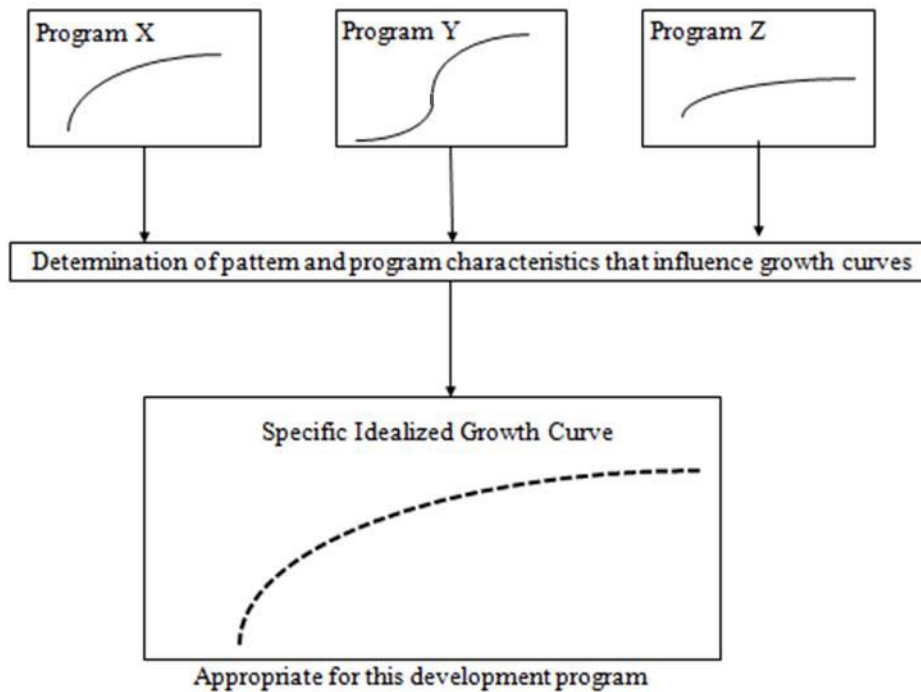
FIGURE 15. Global Analysis Determination of Planned Growth Curve.

21

4.13.3 Other Planning Considerations.
It is important for sufficient testing to be planned and for the testing to be reflective of the OMS/MP. In reliability demonstration testing, the concept of operating characteristic (OC) curves has been used in planning test time and allowable failures. Recall that for a fixed configuration demonstration test, the discrimination ratio – the reliability associated with the producer (contractor) risk, β, over the reliability associated with the consumer (Government) risk, α – has often been used as a guide to determine test time. As a general rule of thumb, the MTBF discrimination ratio of the contractor design-to MTBF to the government requirement MTBF (to be demonstrated with confidence) is generally around 2-3.

This concept is extended to developing reliability growth planning curves where the growth curve follows the Duane failure pattern, i.e., power law expected number of failures. In particular, a system planning curve and associated test duration can be constructed such that if growth occurs in accordance to the planning curve for the planned test duration, then with a prescribed probability, growth test data will be generated that provide a γ statistical lower confidence bound (LCB) that will meet or exceed the technical requirement (TR).

For reliability growth, the ratio of interest is the contractor's goal MTBF, M_G, to the MTBF technical requirement, *TR* (which is to be demonstrated with confidence). A given reliability growth curve has an associated consumer (Government) and producer (contractor) risk. These risks, along with multiple other testing, program, and reliability growth parameters are used to select the best reliability growth curve for the program. Such reliability growth consumer and producer risks are of interest when a program wishes to demonstrate an MTBF value with confidence based on reliability growth test data and have a reasonable chance of doing so. Such a demonstration should not be confused with a possibly mandated Initial Operational Test (IOT) demonstration of an operational required MTBF. This kind of demonstration is typically conducted at the conclusion of the developmental growth test and is conducted with mature production units.

Reliability growth potential MTBF, M_{GP}, is a theoretical upper limit on reliability which corresponds to the reliability that would result if all B-modes were surfaced and fixed with the assumed assessed FEF. It can be shown that

$$M_{GP} = \frac{M_I}{1 - (MS)\mu_d}$$

where M_I is the initial MTBF, *MS* is the Management Strategy, and μ_d is the average FEF. These planning parameters are termed consistent provided $M_G < M_{GP}$.

4.13.4 Threshold.
A threshold is a value in the rejection region of a statistical test of hypothesis, which indicates that an achieved or demonstrated reliability below the value is not in conformance with the idealized growth curve. A threshold value is not a LCB on the true reliability; it is used simply to conduct a test of hypothesis. Threshold values are computed at particular points in time, referred to as milestones, which are major decision points. The Threshold Model can be used to compare a reliability point estimate, which is based on actual failure data from a growth test,

22

against a theoretical threshold value. The test statistic in this procedure is the point estimate of the MTBF achieved at the milestone calculated from the test data. If this estimate falls at or below the threshold value, this would raise a red flag and indicate that the achieved reliability is statistically not in conformance with the idealized growth curve. At that point, management might want to take action to restore reliability to a higher level, perhaps through restructuring the program, a more intensive corrective action process, a change of vendors, additional low-level testing, etc.

4.14   Reliability Growth Tracking Concepts.

4.14.1 Demonstrated Reliability.
A demonstrated reliability value is based on actual test data and is an estimate of the current attained reliability. The assessment is made on the system configuration currently undergoing test, not on an anticipated configuration, nor a prior configuration. This number allows for the effects of introduced fixes into the system as its calculation incorporates the trend of growth established, to date, over the current test phase (or possibly the combined test phases).

4.14.2 Reliability Growth Tracking Curve.
The reliability growth tracking curve is the curve that best fits the data being analyzed. It is typically based on data solely within one test phase. This is due to the fact that between test phases, there is often a CAP during which a group of corrective actions are implemented which significantly increase the reliability. In this commonly encountered situation, the tracking model will not usually statistically fit the data over the two phases bracketing the CAP. However, in the instances where the tracking model is in adequate conformance with the test data from several phases, it may be used to track growth over the combined test phases. Whatever period of testing is used to form a database, this curve is the statistical best representation from a family of growth curves of the overall reliability growth of the system. It depicts the trend of growth that has been established over the database. Thus, if the database covers the entire program to date, the right end point of this curve is the current demonstrated reliability. Figure 16 depicts this reliability growth tracking curve. To the left of the line is the demonstrated reliability using data to date, and to the right of the line is the extension of the planning curve for the expected reliability growth.
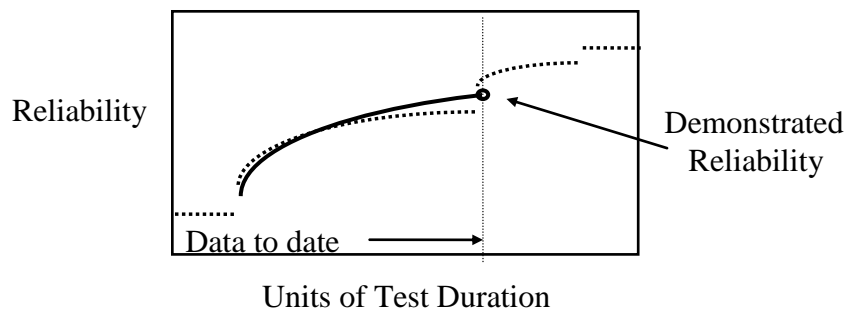
FIGURE 16.   Reliability Growth Tracking Curve.

23

4.15   Reliability Growth Projection Concepts.

4.15.1 Extrapolated Reliability.
Extrapolating a growth curve beyond the currently available data shows what reliability a program can be expected to achieve as a function of additional test duration, provided the conditions of test and the engineering effort to improve reliability are maintained at their present levels (i.e., the established trend continues) in the absence of a significant group of delayed corrective actions.

4.15.2 Projected Reliability.
A reliability projection is an assessment of reliability that can be anticipated at some future point in the development program. The projection is based on the achievement to date and engineering assessments of future program characteristics. Projection is a particularly valuable analysis tool when a program is experiencing difficulties, since it enables investigation of program alternatives. See Figure 17.
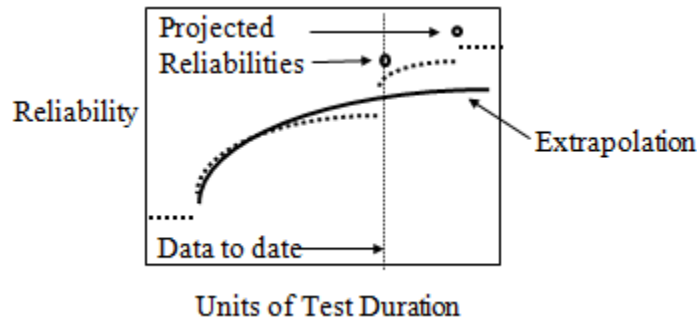
FIGURE 17. Extrapolated and Projected Reliabilities.

4.16   Models Covered in this Handbook.
There are 3 types of reliability growth models covered in this Handbook – planning, tracking, and projection.

The planning models include:
   a) AMSAA Crow Planning Model
   b) System Level Planning Model (SPLAN)
   c) Subsystem Level Planning Model (SSPLAN)
   d) Planning Model Based on Projection Methodology (PM2) - Continuous
   e) Planning Model Based on Projection Methodology - Discrete
   f) Threshold Model

The tracking models include:
   a) AMSAA Reliability Growth Tracking Model – Continuous (RGTMC)
   b) AMSAA Reliability Growth Tracking Model – Discrete (RGTMD)
   c) Subsystem Level Tracking Model (SSTRACK)

24

The projection models include:
   a) AMSAA-Crow Projection Model (ACPM)
   b) Crow Extended Reliability Projection Model
   c) AMSAA Maturity Projection Model (AMPM)
   d) AMSAA Maturity Projection Model Based on Stein Estimation (AMPM-Stein)
   e) Discrete Projection Model (DPM)

4.17   Sources for Models Covered in this Handbook.
For access and/or details on computational programs available for the reliability growth planning, tracking and projection models presented in this handbook, it is suggested that the reader visit websites or contact companies that may offer these capabilities. Potential sources for these, or similar tools, include AMSAA **www.amsaa.army.mil**, ReliaSoft Corporation(www.reliasoft.com),, Reliability Information Analysis Center (RIAC, www.theriac.com), and Relex (www.relex.com).

25

5. RELIABILITY GROWTH PLANNING.

5.1 Introduction.

The goal of reliability growth planning is to optimize testing resources, quantify potential risks, and plan for successful achievement of reliability objectives. The growth plan can serve as a significant management tool in scoping out the required resources to enhance system reliability and improve the likelihood of demonstrating the system reliability requirement. Critical aspects underlying this process include addressing program schedules, amount of testing, resources available, and the realism of the test program in achieving its requirements. Planning activities include establishing test schedules, determining resource availability in terms of facilities and test equipment, and identifying test personnel, data collectors, analysts and engineers. Additionally sufficient calendar time during the program should be planned to analyze, gain approval and implement corrective actions. Planning is quantified and reflected through a reliability growth program plan curve. This curve may be used to establish interim reliability goals throughout the test program. Two significant benefits of reliability growth planning are:

a) Can perform trade-offs with test time, initial reliability, final reliability, confidence levels, requirements, etc to develop a viable test program.
b) Can assess the feasibility of achieving a requirement given schedule and resource constraints by using historical values for parameters.

5.1.1 Basic Model Approaches Covered.

The planning models covered in this handbook are based on two basic approaches – the power law and the AMSAA Maturity Projection Model (AMPM). The power law approach uses an assumed cumulative relationship between the expected number of discovered failures and test duration. The AMPM approach uses an assumed cumulative relationship between the expected number of discovered B-modes and the test duration, which gives rise to a reliability growth relationship between the expected system failure intensity and the cumulative test duration.

5.1.2 Planning Models Covered.

The reliability growth planning models presented in this handbook include:

a) AMSAA Crow Planning Model
b) System Level Planning Model (SPLAN)
c) Subsystem Level Planning Model (SSPLAN)
d) Planning Model Based on Projection Methodology (PM2) Continuous
e) Planning Model Based on Projection Methodology (PM2)-Discrete
f) Threshold Model

Models a), b), and c) use the power law approach, whereas models d) and e) use the AMPM approach. Model f), the Threshold Model, is not a growth model per se, but rather a program or methodology to develop interim goals to ascertain whether the program is proceeding in accordance with the planned growth curve.

5.1.3 Planning Model Limitations.

The foremost limitation associated with developing reliability growth planning models is that the testing utilized for reliability growth planning should be reflective of the OMS/MP. If the test environment during development reasonably simulates the mission environment stresses, then it

26

may be feasible to use the growth test data to statistically estimate the demonstrated reliability. Such use of the growth test data could eliminate or supplement a follow-on fixed configuration reliability demonstration test for compliance testing.

5.1.4   Demonstrating Reliability Requirements with Statistical Confidence.
The adaptation of Operating Characteristic (OC) curve methodology in the development of growth curves allows one to plan for demonstrating requirements with stated confidence, typically 80%. The demonstrations are typically conducted in a fixed configuration test.

In broad terms, the consumer (Government) risk is the probability of accepting a system when the true reliability is below the TR and the producer (contractor) risk is the probability of rejecting a system when the true reliability is at least the contractor's target value (which is set above the TR).

For the non-growth case (constant mature configuration), the parameters defining the reliability demonstration test consist of the test duration, T_{DEM} , and the allowable number of failures, c. The "acceptance" or "passing" criterion is simply $f_{obs} \leq c$, where f_{obs} denotes the observed number of failures.

The probability of observing i failures in T_{DEM} is distributed as Poisson and thus the probability of acceptance (observing c or fewer failures in T_{DEM}) is

$$P(Accept) = Prob(A; M, c, T_{DEM}) = Prob(f_{obs} \leq c)$$
$$= \sum_{i=0}^{c} Prob(f_{obs} = i) = \sum_{i=0}^{c} e^{-T_{Dem}/M} \frac{(T_{DEM}/M)^i}{i!}$$

where $M = MTBF$.

To ensure that "passing the demonstration test" is equivalent to demonstrating the *TR* with at least confidence level γ (e.g., $\gamma = 0.80$ or $\gamma = 0.90$), c must be chosen such that

$$f_{obs} \leq c \quad \Leftrightarrow \quad TR \leq \ell_{\gamma}(f_{obs})$$

where *TR > 0* and $\ell_{\gamma}(f_{obs})$ denotes the value of the 100 γ percent LCB when $f_{obs} = c$ failures occur in the demonstration test of length T_{DEM} . That is, c is chosen to be the largest non-negative integer k that satisfies the inequality

$$\sum_{i=0}^{k} e^{-T_{DEM}/TR} \frac{(T_{DEM}/TR)^i}{i!} \leq 1 - \gamma$$

Recall that the OC curve associated with a reliability demonstration test is the graph of the probability of acceptance, i.e., *Prob (A;M,c,T_{DEM})* given as a function of the true but unknown constant MTBF, *M,* as depicted in Figure 18.

27

$T_{DEM} = 2800$

$C = 23$
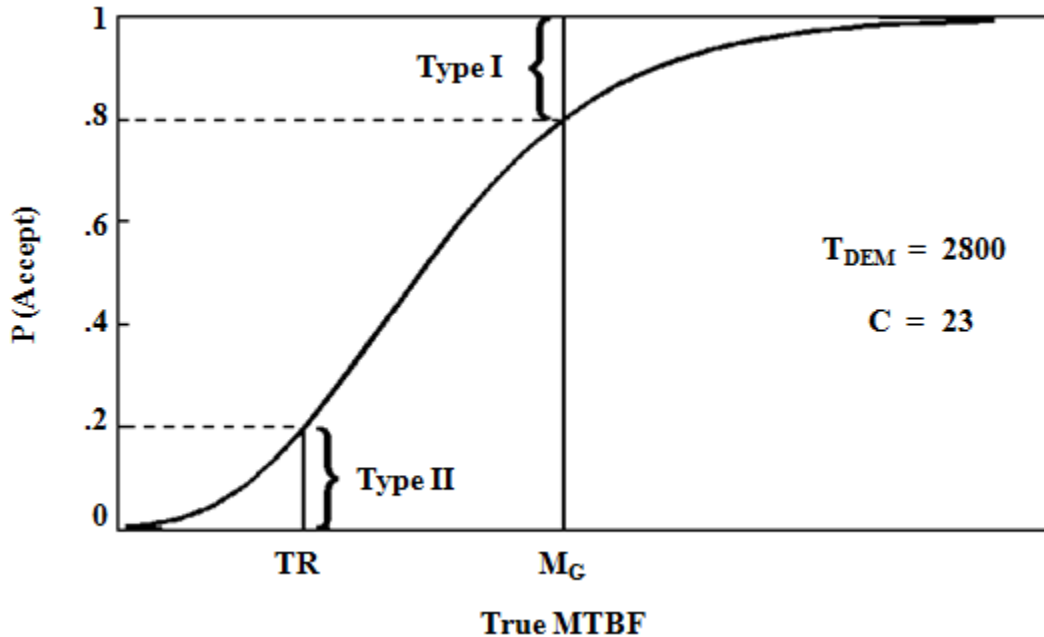
FIGURE 18. Example OC Curve for Reliability Demonstration Test.

The consumer (Government) risk associated with this curve, called the Type II risk, is defined by

$$\text{Type II} \triangleq \text{Prob}(A; TR, c, T_{DEM})$$

Thus, by the choice of c,

$$\text{Type II} \leq 1 - \gamma$$

For the producer (contractor) to have a reasonable chance of demonstrating the *TR* with confidence γ, the system configuration entering the reliability demonstration test must often have an MTBF value of M_G (referred to as the contractor's developmental goal or target MTBF) that is considerably higher than the *TR*. The probability that the producer (contractor) fails the demonstration test, given that the system under test has a true MTBF value of M_G, is termed the producer (contractor) risk, or Type I risk. Thus

$$\text{Type I} = 1 - \text{Prob}(A; M_G, c, T_{DEM})$$

If the Type I risk is higher than desired, then either a higher value of M_G should be attained prior to entering the reliability demonstration test or T_{DEM} should be increased. If T_{DEM} is increased, then c may have to be readjusted for the new value of T_{DEM} to remain the largest non-negative integer that satisfies the Type II inequality.

The discrimination ratio, M_G/TR is commonly used as an aid to determine test plans for the non-growth situation.

5.1.5 Planning Areas.

There are two key planning areas: elements under management control and potential risk elements during the planning phase. Elements under management control include:

  a) Management Strategy (MS): the fraction of system initial failure rate addressed by corrective actions;
  b) Rate at which failure modes are surfaced;
  c) Turnaround time for analyzing and implementing corrective actions; and
  d) Fix Effectiveness Factor (FEF): the fraction reduction in the rate of occurrence of modes after corrective action.

The potential risk elements during the planning phase include:

  a) Initial MTBF (M_I);
  b) Ratio of M_I to final developmental goal MTBF, M_G;
  c) Total test time, T.

Table I provides historically-based AMSAA estimates for the ratio of M_I to M_G and for a collection of system average FEFs (Ellner, Trapnell (1990)).

TABLE I. Historical growth parameter estimates

Parameter	Mean/Median	Range
Initial to Mature Ratio – M_I/M_G	0.30/0.27	0.15-0.47
Fix Effectiveness Factor (FEF)[1]	0.70/0.71	0.55-0.85

The growth rate and FEF information in Tables I and II may be used as a guide in determining the reasonableness of these parameters based on this historical data. However, past experience has shown that to achieve successful results in the IOT&E (which is conducted at the conclusion of the developmental growth program), it is crucial to enter the reliability growth test conducted in the Engineering and Manufacturing Developmental (EMD) phase with an M_I that yields an MTBF growth potential, M_{GP}, that is sufficiently above M_G. Doing so allows one to achieve a reasonable ratio of M_G to M_{GP}, which is recommended by the U.S. Army Evaluation Center to be in the range of 0.60 to 0.80. For a given MS and average FEF, the M_{GP} is directly proportional to M_I. Thus, to support the AEC recommended upper limit of the M_G to M_{GP} ratio (0.80), the M_I to M_G ratio must be greater than or equal to the mean historically achieved ratio displayed in Table I (0.30). To achieve such an M_I requires conducting a comprehensive set of Design for Reliability activities (J. Hall Jun 2009) prior to entering EMD. Failure to achieve a sufficiently high M_I in past developmental programs has resulted in an unacceptably high percentage of DoD

---

[1] Software fixes may have higher FEFs.

developmental systems failing to meet their reliability thresholds in the IOT&E, even as a point estimate. (J. Hall Jun 2009)

For planning purposes, the *MS* during early or prototype testing could exceed 0.95 for the developmental portion of the system. During subsequent testing, the *MS* for the developmental portion of the system typically needs to be at least 0.90 for a successful reliability developmental program, and often must be near 0.95 to achieve a sufficiently high M_{GP} relative to M_G. For most systems, it is not prudent to plan on achieving a *MS* higher than 0.96 for the developmental portion of the system.

5.1.6   Reliability Growth Planning Checklist.
The following provides a checklist for reviewing reliability growth planning curves:
 a) Goal reliability needs to be sufficiently high to have adequate probability of passing the IOT&E or other reliability demonstration test.
 b) The expected initial reliability, M_I or R_I, should be based on expected maturity and prior information (e.g., from previous or similar systems, technology development, or such information as available).
 c) The ratio of M_I to M_G should not be too low (e.g., less than 0.15).
      i. Desirable to have the ratio above usual historical range of 0.20 to 0.35
         Need to achieve sufficient growth in design phase prior to EMD test phase to increase the ratio of M_I to M_G beyond the historical range.
 d) The expected number of failures associated with the planning curve and test duration should be sufficiently large to allow enough corrective action opportunities to grow from M_I to M_G.
 e) There needs to be sufficient calendar time, facility assets, and engineering personnel to ensure timely implementation of effective corrective actions to surfaced failure modes prior to IOT&E.
 f) If corrective actions are to be implemented at only a few designated points during the development program, then the depicted expected growth pattern should reflect this.

5.2   AMSAA Crow Planning Model.
This section contains only a minimum of the details that were contained in the original version of this handbook (AMSAA Feb 1981). For a more detailed discussion of this model and to reference several examples, it is recommended that the reader refer to the original version of this handbook (AMSAA Feb 1981).

5.2.1   Purpose.   The purpose of the AMSAA Crow Planning Model is to construct idealized system reliability growth curves, identify the test time and growth rate required to improve system reliability, and aid in demonstrating the system reliability requirement as a point estimate.

5.2.2   Assumptions.
The assumptions associated with the AMSAA Crow Planning Model include:
a) Within a test phase, reliability growth can be modeled as a Non-Homogeneous Poisson Process (NHPP) with power law mean value function, $\mu(t) = \lambda t^{\beta}$; and
b) Based on the failures and test time within a test phase, the cumulative failure rate is linear on a log-log scale.

5.2.3   Limitations.
The limitations associated with the model include:
 a) The system must be complex (i.e., number of potential failures should be large enough to comply with NHPP assumption);
 b)   Sufficient opportunities for implementation of corrective actions are required to allow growth to be portrayed as a smooth curve; and
 c) Reliability growth testing should be reflective of the OMS/MP.

5.2.4   Benefits.
 The benefits associated with the AMSAA Crow Planning Model include:
a) Allows for generation of a target idealized growth curve; and
b) Can be utilized for discrete data when there are a large number of trials and low probability of failure.

5.2.5   Planning Factors.
The idealized curve has a baseline value M_I over the initial test phase, which ends at time t_1. M_I is the average MTBF over the first test phase. From t_1 to the end of testing at time T, the idealized curve increases steadily according to a learning curve pattern till it reaches the final reliability requirement, M_F. T and the growth rate α are iterated to develop the plan satisfying the constraints. Subsequent to publishing the original version of this handbook (AMSAA Feb 1981) and prior to the development of SPLAN, a function, *Prob*, was developed that assured a designated probability of observing at least one failure in the initial time t. Subsequently the Management Strategy, *MS*, was also included. For MS=1, the function is:

$$\text{Prob} = 1 - e^{-\left(\frac{t_I}{M_I}\right)}$$

which for *Prob* = 0.95 results in t_1 approximately equal to 3 times M_I. After development of *MS*, *3*(M_I /MS)* was used to satisfy the *Prob* of 0.95. If no significant B-mode corrective actions are planned until the first CAP, then t_I should be the test time until the first CAP.

5.2.6   Background of AMSAA Crow Planning Model.
The original version of this handbook (AMSAA Feb 1981) is based on Duane's work and Crow's more generalized work. Duane analyzed data for several systems and noted that if fixes to improve reliability are incorporated into the design of a system under development, then on a log-log plot, the graph of cumulative failure rate vs. cumulative test time is linear.

5.2.6.1 Duane's Growth Model.
The Duane log-log plot of the straight line and linear regression fit is also known as The Duane Postulate:

$$\log C(t) = \delta - \alpha \log t$$

Taking the anti-log, $C(t) = \lambda t^{-\alpha}$ where $\delta = \ln \lambda$.

Duane's model has two parameters: α and λ. α is the shape parameter, which determines the shape of the growth curve. λ is the scale or size parameter for the curve. With these two parameters, the cumulative number of failures, *N(t)*, the average failure rate, *C(t),* and the

31

instantaneous failure rate, *r(t),* can be calculated for any time *t* within the test. Further, given α and λ, it is possible to solve for *t*, the test time it will take to achieve a specific reliability. This assumes that the factors affecting reliability growth remain unchanged across the development.

5.2.6.2 Drawbacks to Duane's Method.

Duane stated that α could be universally treated as being 0.5, the modal value within his database. This has since been shown to be unrealistic, as per Table II. All Duane MTBF growth curves pass through the origin of the graph on a linear-linear plot, imputing zero reliability at the start of test. The method is also a deterministic estimation of the regression, which makes no allowance for variation.

Table II Historical Growth Rate Estimates

System Type	Mean/Median	Range
One Shot (Missiles)	.046/0.47	0.27-0.64
Time or Distance Based	0.34/0.32	0.23-0.53

5.2.7   Development of AMSAA Crow Planning Model.

Crow explored the advantages of using a NHPP with a Weibull intensity function to model several phenomena, including reliability growth. If system failure times follow the Duane Postulate, then they can be modeled as a NHPP with Weibull intensity function (i.e., based on the NHPP with power law mean value function). To make the transition from Duane's formulae to the Weibull intensity functional forms, β has to be substituted for 1- α. Thus the parameters in the AMSAA Crow Planning Model are λ and β , where β determines the shape of the curve. The physical interpretation of β (called the growth parameter) is the ratio of the average (cumulative) MTBF to the current (instantaneous) MTBF at time *t*.

Even though Crow's growth parameter estimate is still interpreted as the estimate of the negative slope of a straight line on a log-log plot, the estimates of λ and β differ from Duane's procedures in that the estimation procedure is Maximum Likelihood Estimate (MLE), not least squares, thus each model's parameters correspond to different straight lines.

The reliability planning curve may extend over all the test phases or just over one test phase. Typically a smooth growth curve is portrayed which represents the overall expected pattern growth over the test phases. As noted earlier, it can be modeled as a NHPP with power law mean value function (expected number of failures as a function of cumulative test time) *E(N(t))* $= \lambda t^{\beta}$ which is comparable to the global pattern noted by Duane. Taking the derivative we obtain the idealized reliability growth pattern with failure intensity function $\rho(t)$ given by $\rho(t) = \lambda \beta t^{\beta-1}, 0 < \beta < 1$. Thus, as with Duane, it has a singularity at *t* = 0.

The methodology is based on first specifying an idealized curve that satisfies the expected number of failures at the end of each test phase with cumulative test times *t₁, t₂, ..., tₖ*. For planning purposes, the overall growth trend is represented only for *t > t₁*. It simply utilizes a constant or average failure rate, $\emptyset_1 = M_1^{-1}$, over the first test phase. The constant $\emptyset_1$ is chosen such that the expected number of failures is satisfied for *t = t₁*. Doing so, it follows that the MTBF growth trend for *t > t₁* and $\emptyset_1$ is given by,

$$\text{MTBF (t)} = \begin{cases} M_1 & 0 \leq t \leq t_1 \\ M_1 \left(\frac{t}{t_1}\right)^{\alpha} (1 - \alpha)^{-1} & t > t_1 \end{cases}$$

5.2.8   Potential Issues.

In using the previous equations, one needs to be careful not to automatically equate M_1 to the planning parameter, M_I, defined as the initial MTBF. In general $M_I \leq M_1$. The two MTBFs should be equated only if no growth is planned over the first test phase, since M_1 is the planned average MTBF over the initial test phase. The growth rate α is used as a measure of programmatic risk with respect to being able to grow from M_1 to $M_F = \text{MTBF(T)}$ in test time T. The higher α is relative to past experience, the greater the risk of attaining M_F. MTBF(T) is a strictly increasing function of the ratio T/t_1 and can be made as large as desired by making t_1 sufficiently small. One should guard against artificially lowering α by selecting t_1 so small that no significant amount of fix implementation is expected to occur until a corrective action period that is beyond t_1.

A reliability projection concept, growth potential (GP), is useful in considering the reasonableness of the idealized curve. The growth potential, M_{GP} , is the theoretical value that would be reached if all B-modes were surfaced and corrected with the assumed or assessed FEFs. Assuming an average *FEF* of μ_d, a Management Strategy of *MS*, and an initial MTBF of M_I, one can express the GP *MTBF* as $M_{GP} = M_I /(1 - (MS)\mu_d)$.

If the final MTBF on the idealized growth curve is not below the M_{GP} for reasonable planning values of *MS* and μ_d, then even if the growth rate α appears modest, it might not be sustainable over the entire period for which the model has to be applied. Note that even with a reasonable choice for t_I, any value of M_F can eventually be obtained since there is no upper limit implied for M_F. This is true even using a growth rate that appears to be reasonable based on past experience with similar types of systems.

5.2.9   Development of the Planned Growth Curve.

The role of the idealized growth curve is to substantiate that the planned growth follows a learning curve which, based on previous experience, is reasonable and can be expected to be achieved. In general, there are two basic approaches for constructing planned growth curves. The first method is to determine the idealized growth pattern that is expected or desirable, and to use this as a guide for the detailed planned curve. The second method is just the reverse. In this case a proposed planned curve is first developed which satisfies the requirement and interim milestones. The idealized curve is then constructed and evaluated to determine if this learning curve is reasonable when compared to historical experience. If not acceptable a new detailed curve would need to be developed.

5.2.10   Determining the Starting Point.

A starting point for the planned growth curve may be determined from (1) using information from previous programs on similar systems, (2) specifying a minimum level of reliability that

management requires to be demonstrated early in order to have assurance that the reliability goals will be met, and (3) conducting an engineering assessment of the design together with any previous test data that may exist. e.g., bench test, prototype test. The practice of arbitrarily choosing a starting point, such as 10% of the requirement, is not recommended. Every effort to obtain information even remotely relevant to a realistic starting point should have been exhausted before an arbitrary figure can be used. See also the example cited in the Planning Factors paragraph 2.2.5

5.2.11 Development of the Idealized Growth Curve.

During development, management should expect that certain levels of reliability be attained at various points in the program in order to have assurance that reliability growth is progressing at a sufficient rate to meet the requirement. The idealized curve portrays an overall characteristic pattern which is used to determine and evaluate intermediate levels of reliability and construct the program planned growth curve. Growth profiles on previously developed, similar type systems may provide significant insight into the reliability growth process. If the learning curve pattern for reliability growth assumes that the cumulative failure rate versus cumulative test time is linear on log-log scale, then the following method is appropriate for construction of the idealized growth curve.

The idealized curve has the baseline value M_I over the initial test phase which ends at time t_1. The value M_I is the average MTBF over the first test phase. From time t_1 to the end of testing at time T, the idealized curve $M(t)$ increases steadily according to a learning curve pattern till it reaches the final reliability M_F. The slope of this curve on the log-log plot in Figure 19 is the growth parameter α. The parametric equation for $M(t)$ on this portion of the curve is $M(t) = M_I \left(\frac{t}{t_1}\right)^\alpha (1-\alpha)^{-1}$.
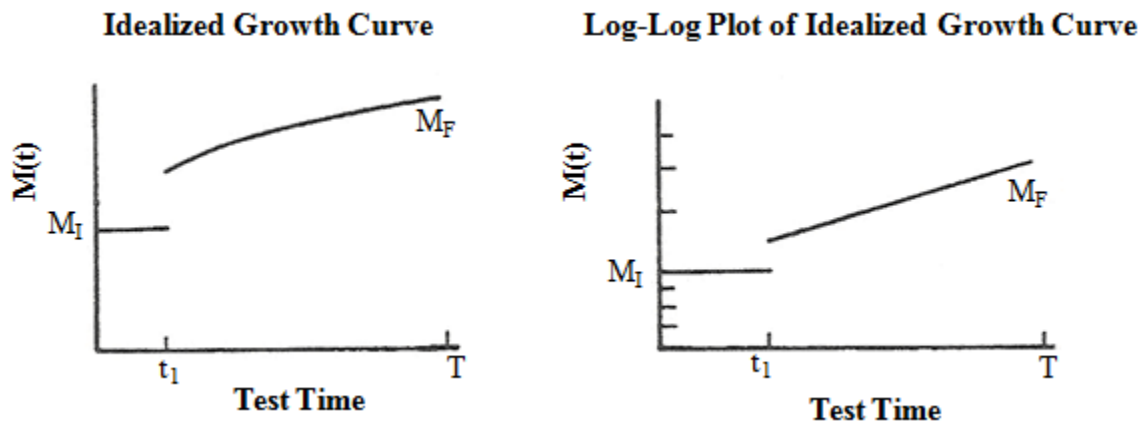
FIGURE 19. Idealized Growth Curve

5.2.12 Equations and Metrics.

This model assumes that the cumulative failure rate versus cumulative test time is linear on log-log scale. It is not assumed that the cumulative failure rates follow the same growth pattern within test phases. In fact, if all fixes are incorporated into the system at the end of a test phase,

then the reliability would be constant during the test phase. Thus, no growth would occur in the test phase.

To illustrate this approach, let t_1, t_2, ..., t_k denote the cumulative test times which correspond to the ends of test phases. It is assumed that $N(t_i)/t_i$ versus t_i, $i = 1, 2, ..., K$, are linear on log-log scale, where $N(t_i)$ is the cumulative number of failures by time t_i. That is, $log\ N(t_i)/t_i$ is linear with respect to $log\ t_i$. This implies that $log\ N(t_i)/t_i$ can be expressed as $log\ N(t_i)/t_i = \sigma - \alpha\ log\ t_i$, where σ and α are, respectively, intercept and slope parameters. Let λ_I denote the initial average failure rate for the first test phase, i.e., $\lambda_I = N(t_1)/t_1$. Since $log\ \lambda_I = \sigma - \alpha\ log\ t_1$, it follows that $\sigma = log\ \lambda_I + \alpha\ log\ t_1$. The cumulative failure rate can be expressed as

$$N(t_i)/t_i = \lambda_I \left(\frac{t_i}{t_1}\right)^{-\alpha}.$$

The idealized growth curve shows that the initial average MTBF over the first test phase is M_I, and that reliability growth from this average begins at t_1. This jump is indicative of delayed fixes incorporated into the system at the end of the first test phase. The idealized curve $M(t)$ is a guide for the average MTBF over each test phase. Further given that

$$M(t) = M_I\ (t\ /\ t_1)^\alpha\ (1 - \alpha)^{-1} \quad for\ t > t_1,$$

then the average failure rate and the average MTBF for the i-th test phase can be determined by $\lambda_i = (N(t_i) - N(t_{i-1}))\ /\ (t_i - t_{i-1})$, and $M_i = 1/\lambda_i$, where $N(t_i) = \lambda_I t_1\ (t_i\ /\ t_1)^{1-\alpha}$. See Figure 20.
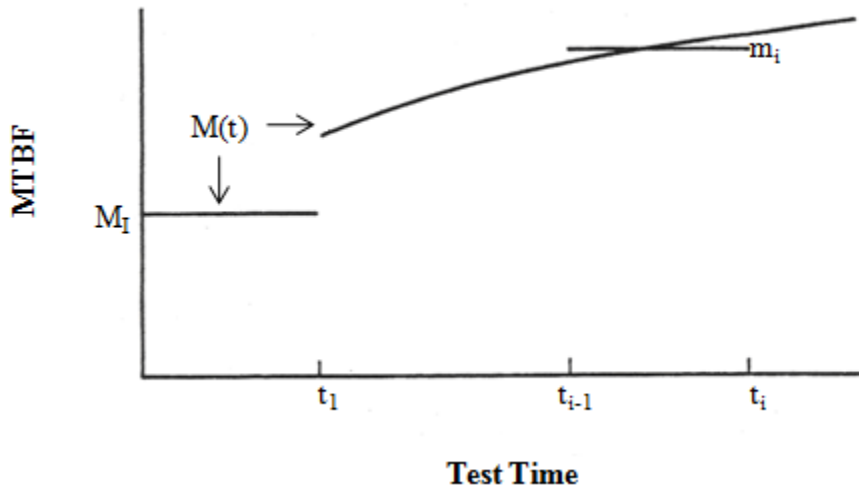
FIGURE 20. Average MTBF over i^{th} Test Phase.

In the application of the idealized growth curve model, the final MTBF value M_F to be attained at time T is set equal to $M(T)$, i.e., $M_I\ (T\ /t_1)^\alpha\ (1 - \alpha)^{-1} = M_F$.

The parameters M_I and t_I of this model have the physical interpretations that M_I is the initial average MTBF for the system and t_1 is the length of the first test phase in the program. The parameter α is a growth rate.

35

5.2.13 AMSAA Crow Planning Model Example.
Specific examples of how to determine the idealized growth curve, test phase growth, and test time needed are not reproduced here but may be seen in the original version of this handbook (AMSAA Feb 1981).

5.3   System Level Planning Model (SPLAN).

5.3.1   Purpose.
The purpose of SPLAN is to construct an idealized system reliability growth curve and determine an associated test duration that has the following property: If the system grows along the idealized curve for the associated test duration then, with a prescribed probability, the system level growth test data realized over the test period will demonstrate a stated MTBF value at a specified statistical confidence level.

The stated system MTBF value to be demonstrated with statistical confidence from the growth test data will be referred to as the technical requirement and denoted by TR in this section and section 5.4.

5.3.2   Assumptions.
The assumptions associated with SPLAN include:
   a) test duration is continuous; and
   b) the pattern of failures during the test period is consistent with a NHPP with power law mean value function.

5.3.3   Limitations.
 The limitations associated with SPLAN include:
a)   sufficient opportunities for corrective action implementation are required so growth is portrayed as a smooth curve;
b)   the expected number of failures needs to be sufficiently large;
c)   the portion of testing utilized for reliability growth planning should be reflective of the OMS/MP;
d)   the initial test length must be reasonably small (allowing for reliability growth);
e)   the initial MTBF cannot be specified independent of the length of the initial test phase;
f)   the actual growth test data generated over the test period will typically not satisfy Assumption b) if the period contains one or more CAPs which produce a significant jump in MTBF; and
g)   since the demonstrations discussed in this section and in Section 5.4 are based on developmental growth test data, the TR to be demonstrated with statistical confidence may be more reflective of a hardware/software MTBF than an operational MTBF. The TR will be reflective of an operational MTBF only to the extent that the developmental test period allows potential operational and maintenance based failure modes to occur at rates comparable to those that would occur under tactical use conditions.

5.3.4   Benefits.
The benefits associated with SPLAN include:
a)   allows for generation of a target idealized growth curve;

36

b)   can specify desired probability of achieving the TR with confidence; and
c)   can aid in planning to utilize system growth test data to demonstrate with statistical
confidence a stated MTBF value prior to entering an operational demonstration test.

5.3.5   Planning Factors.

The initial condition planning factors include the following: the test time over the initial test
phase before implementation of corrective actions, t_I; the initial average MTBF, M_I; the
Management Strategy, *MS;* and the probability of observing at least 1 correctable or B-mode
failure, *Prob*, where three of the conditions or factors are chosen and the fourth is determined.
Prob is given by the following equation:

$$\text{Prob} = 1 - e^{-\left(\frac{t_I * MS}{M_I}\right)}$$

As a general rule, the initial time period, t_I, should be at least approximately three times greater
than the ratio of M_I to *MS* to ensure a high probability, say 0.95, of surfacing at least one B-mode
failure by the end of t_I. As discussed earlier, this choice of t_I may not be sufficiently long and
may need to extend over the whole first test phase to ensure that after t_I, one can assume the
AMSAA Crow growth pattern applies. Although this pattern does not apply over the first test
phase, the presented OC curve analysis for reliability growth implicitly assumes this pattern
holds over the whole time interval that is associated with the planning curve.

5.3.6   Reliability Growth OC Curve Analysis.

In the presence of reliability growth, observing *c* or fewer failures is not equivalent to
demonstrating the *TR* at a given confidence level. Both the cumulative times to failure and the
number of failures must be considered when using reliability growth test data to demonstrate the
TR at a specified confidence level, γ. Thus, the "acceptance" or "passing" criterion must be
stated directly in terms of the γ LCB on *M(T)* calculated from the reliability growth data. This
data will be denoted by (*n, s*), where n is the number of failures occurring in the growth test of
duration *T,* and $s = (t_1, t_2, \ldots, t_n)$ is the vector of cumulative failure times. In particular, t_i
denotes the cumulative test time to the ith failure and $0 < t_1 < t_2 \ldots \ldots < t_n \leq T$ for $n \geq 1$. The
random vector (*N, S*) takes on values (*n, s*) for $n \geq 1$.

Following notation from the previous section, the definition of our acceptance criterion is given
by the inequality

$$TR \leq l_\gamma(n, s)$$

where $l_\gamma(n, s)$ is the γ statistical LCB on *M(T)*, calculated for $n \geq 1$. Thus, the probability of
acceptance is given by

$$\text{Prob}\left(TR \leq L_\gamma(N, S)\right)$$

where the random variable $L_\gamma(N, S)$ takes on the value $l_\gamma(n, s)$ when (*N, S*) takes on the value
(*n, s*).

The distribution of (*N, S*), and hence that of $L_\gamma(N, S)$, is completely determined by the test
duration, *T,* together with any set of parameters that define a unique reliability growth curve

37

Thus, the value of the above probability expression also depends on *T* and the assumed underlying growth curve parameters. One such set of parameters is t_I, M_I, and α, together with *T*. In this growth curve representation, t_I may be arbitrarily chosen subject to $0 < t_I < T$. Alternately, scale parameter $\omega > 0$ and growth rate α, together with *T*, can be used to define the growth curve by the equation

$$M(t) = \frac{1}{\omega(\beta t^{\beta-1})}, \quad 0 < t \le T$$

where $\beta = 1 - \alpha$. Note that by the above equation,

$$\frac{1}{\omega} = \big(M(T)\big)\beta T^{\beta-1}$$

Thus, the growth curve can also be expressed as,

$$M(t) = \big(M(T)\big)\left(\frac{t}{T}\right)^{\alpha}, \quad 0 < t \le T.$$

Thus it can be seen that the distribution of (*N, S*), and hence that of L_γ *(N, S)*, is determined by *(α, T, M(T))*.

The Type II or consumer (Government) risk for *M(T)= TR* is at most *1-γ* (for any $\alpha < 1$ and $T > 0$), analogous to the case in the previous section, i.e.,

$$\text{Type II} = \text{Prob}\big(TR \le L_\gamma(N,S)\big) \le 1 - \gamma$$

for any $\alpha < 1$ and $T > 0$, provided *M(T) = TR*.

To emphasize the functional dependence of the probability of acceptance on the underlying true growth curve parameters *(α, T, M(T))*, this probability is denoted by *Prob (A; α, T, M(T))*. Thus,

$$\text{Prob}\big(A; \alpha, T, M(T)\big) \underline{\Delta} \text{Prob}\big(TR \le L_\gamma(N,S)\big)$$

where the distribution of (*N, S*), and hence that of L_γ *(N, S)*, is determined by *(α,T, M(T))*. The *Prob (A; α, T, M(T))* can be shown to only depend on the values of *M(T)/TR* (or equivalently *M(T)* for known *TR*) and *E(N)* [(Broemm, Ellner and Woodworth Sep 2000)]. The ratio *M(T)/TR* is analogous to the discrimination ratio for a constant configuration reliability demonstration test of the type discussed in the previous section. Note *E(N)* denotes the expected number of failures associated with the growth curve determined by *(α, T, M(T))*. More explicitly:

$$E(N) = T/\{(1 - \alpha)M(T)\}$$

As shown in [(Broemm, Ellner and Woodworth Sep 2000)],

38

$$\text{Prob}\big(A; \alpha, T, M(T)\big) = (1 - e^{-\mu})^{-1} \sum_{n=1}^{\infty} \left[\text{Prob}\left(\frac{\chi_{2n}^2}{z_\gamma^2} \geq \frac{1}{2\mu d}\right)\right] e^{-\mu}\left(\frac{\mu^n}{n!}\right)$$

where $\mu \triangleq E(N)$ and $d \triangleq M(T)/TR$. This equation explicitly shows that the probability of acceptance only depends on μ and d. Thus, the probability of acceptance is denoted by *Prob $(A;\mu,d)$* and

$$\text{Type II} = \text{Prob}(A; \mu, 1) \leq 1 - \gamma$$

A discrimination ratio chart for the growth case is shown below in Figure 21. The figure presents three curves for demonstrating the *TR* with a fixed probability of acceptance equal to 0.50 (as a function of the *M(T)/TR* ratio and the expected number of failures). The test duration corresponding to a point (x, y) on a confidence curve in Figure 21 can be shown via the above formula for *E(N)* to satisfy the following:

$$T = (1 - \alpha)\ (TR)\ xy$$

where α denotes the growth rate parameter for the planning curve, $x = E(N)$ and $y = M(T)/TR$. Note also that

$$M(T) = \left(\frac{M_I}{1 - \alpha}\right)\left(\frac{T}{t_I}\right)^{\alpha}$$
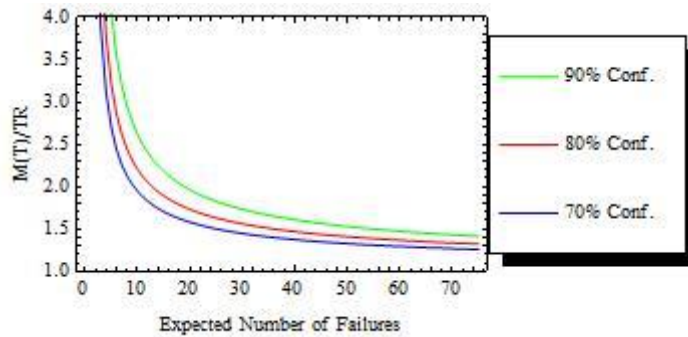
FIGURE 21.   Probability equals 0.50 of demonstrating TR w/% Confidence as a function of *M(T)*/TR and Expected number of failures

By setting the discrimination ratio equal to one in the above expression for the probability of acceptance, one can see that the actual value of the consumer (Government) risk solely depends on μ and is at most 1-γ. To consider the producer (contractor) risk, Type I, let α_G denote the contractor's target or goal growth rate. This growth rate should be a value the contractor feels he can achieve for the growth test. Let M_G denote the contractor's MTBF goal. This is the MTBF value the contractor plans to achieve at the conclusion of the growth test of duration *T*. Thus, if

39

the true growth curve has the parameters α_G and M_G, then the corresponding producer (contractor) risk of not demonstrating the *TR* at confidence level γ (utilizing the generated reliability growth test data) is given by,

$$\text{Type I} = 1 - \text{Prob}(A; \mu_G, d_G)$$

where

$$d_G = M_G/TR \text{ and } \mu_G = T/\{(1 - \alpha_G)M_G\}$$

If the Type I risk is higher than desired, there are several ways to consider reducing this risk while maintaining the Type II risk at or below 1-γ. Since *Prob (A; μ_G, d_G)* is an increasing function of μ_G and d_G, the Type I risk can be reduced by increasing one or both of these quantities, e.g., by increasing *T*. To further consider how the Type I statistical risk can be influenced, we express d_G and μ_G in terms of *TR*, *T*, α_G, and the initial conditions (M_I, t_I). With $\alpha = \alpha_G$ and $M(T) = M_G$,

$$M_G/TR = d_G = \left\{ \frac{M_I}{(1 - \alpha_G)t_I^{\alpha_G}TR} \right\} T^{\alpha_G}$$

and

$$E(N) = \mu_G = \left(t_I^{\alpha_G}/M_I \right) T^{1-\alpha_G}$$

Note for a given requirement *TR*, initial conditions (M_I, t_I), and an assumed positive growth rate α_G, the producer (contractor) risk is a decreasing function of *T*. These equations can be used to solve for a test time *T* such that the producer (contractor) risk is a specified value. The corresponding consumer (Government) risk will be at most 1-γ.

The following section contains two examples of an OC analysis for planning a reliability growth program. The first example illustrates the construction of an OC curve for given initial conditions (M_I, t_I) and requirement *TR*. The second example illustrates the iterative solution for the amount of test time *T* necessary to achieve a specified producer (contractor) risk, given initial conditions (M_I, t_I) and requirement *TR*. These examples use the following equations,

$$M(T) = \left(\frac{M_I}{1-\alpha} \right) \left(\frac{T}{t_I} \right)^{\alpha} \text{ and } E(N) = \frac{T}{(1-\alpha)M(T)}$$

The quantities $d = M(T)/TR$ and $\mu = E(N)$ are then used to obtain an approximation to *Prob (A; μ, d)*.

5.3.7   SPLAN Example 1.

Suppose we have a system under development that has a technical requirement (*TR*) of 100 hours to be demonstrated with 80 percent confidence using growth test data. For the developmental program, a total of 2800 hours test time (*T*) at the system level has been predetermined for reliability growth purposes. Based on historical data for similar type systems and on lower level testing for the system under development, the initial MTBF, M_I, averaged over the first 500 hours (t_I) of system-level testing was expected to be 68 hours. Using this data, an idealized

reliability growth curve was constructed such that if the tracking curve followed along the idealized growth curve, the *TR* of 100 hours would be demonstrated with 80 percent confidence. The growth rate, α, and the final MTBF, *M(T)*, for the idealized growth curve were 0.23 and 130 hours, respectively. The idealized growth curve for this program is depicted below in Figure 22.
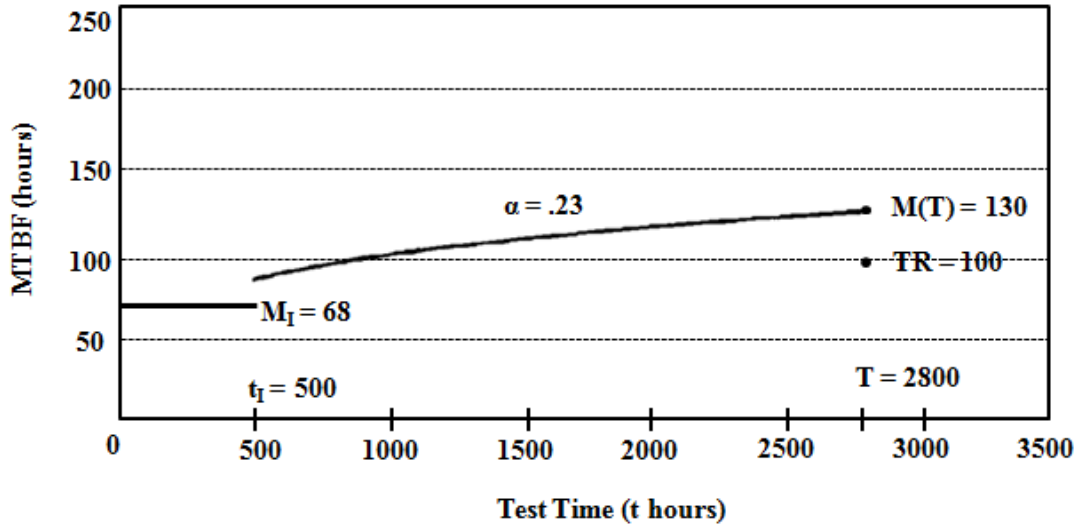
FIGURE 22. Idealized Reliability Growth Curve.

For this example, suppose one wanted to determine the OC curve for the program. For this, one would need to consider alternate idealized growth curves where the *M(T)* vary but the M_I and t_I remain the same values as those for the program idealized growth curve (i.e., $M_I = 68$ hours and $t_I = 500$ hours). In varying the *M(T)*, this is analogous to considering alternate values of the true MTBF for a reliability demonstration test of a fixed configuration system. For this program, one alternate idealized growth curve was determined where *M(T)* equals the TR, whereas the remaining alternate idealized growth curves were determined for different values of the growth rate. These alternate idealized growth curves along with the program idealized growth curve are depicted in Figure 23.

41
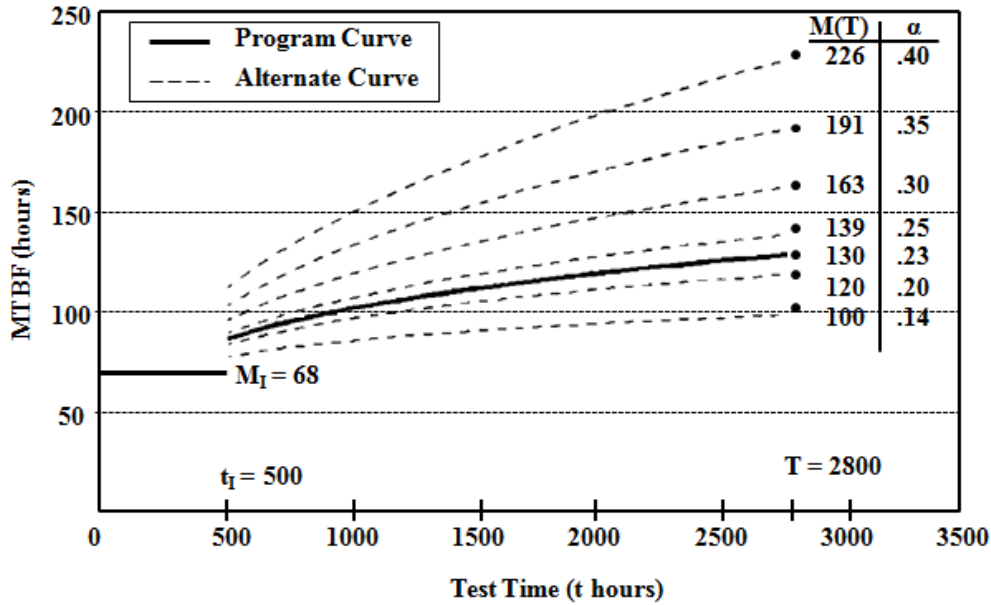
FIGURE 23. Program and Alternate Idealized Growth Curves.

For each idealized growth curve, *M(T)* and the expected number of failures *E(N)* can be found. Using the ratio *M(T)/TR* and *E(N)* as entries in the tables contained in (Ellner and Mioduski Aug 1992), one can determine (by double linear interpolation) the probability of demonstrating the *TR* with 80 percent confidence. This probability is actually the probability that the 80 percent LCB for *M(T)* will be greater than or equal to the TR. These probabilities represent the probability of acceptance (*P(A)*) points on the OC curve for this program which is depicted in Figure 24. The *M(T)*, α, *E(N)*, and *P(A)* for these idealized growth curves are summarized in Table II.

TABLE II. Example 1 - planning data for idealized growth curves.

M (T)	α	E (N)	P (A)
100	0.14	32.6	0.15
120	0.20	29.2	0.37
130	0.23	28.0	0.48
139	0.25	26.9	0.58
163	0.30	24.5	0.77
191	0.35	22.6	0.90
226	0.40	20.6	0.96

42
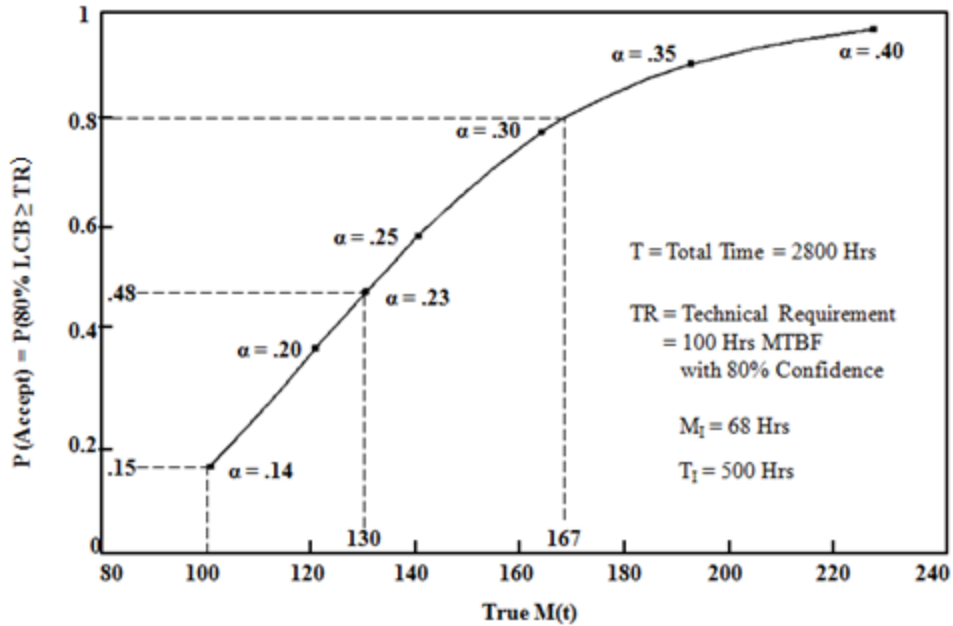
FIGURE 24. Operating Characteristic (OC) Curve.

From the OC curve, the Type I or producer (contractor) risk is 0.52 (1-0.48), which is based on the program idealized growth curve where $M(T) = 130$. Note that if the true growth curve were the program idealized growth curve, there is still a 0.52 probability of not demonstrating the TR with 80 percent confidence. This occurs even though the true reliability would grow to $M(T) = 130$, which is considerably higher than the *TR* value of 100. The Type II or consumer (Government) risk, which is based on the alternate idealized growth curve where $M(T) = TR = 100$, is 0.15. As indicated on the OC curve, it should be noted that for this developmental program to have a producer (contractor) risk of 0.20, the contractor would have to plan on an idealized growth curve with $M(T) = 167$.

5.3.8   SPLAN Example 2.   Consider a system under development that has a technical requirement (*TR*) of 100 hours to be demonstrated with 80 percent confidence, as in Example 1. The initial MTBF, M_I, over the first 500 hours (t_I) of system level testing for this system was estimated to be 48 hours, which (again, as in Example 1) was based on historical data for similar type systems and on lower level testing for the system under development. For this developmental program, it was assumed that a growth rate, α, of 0.30 would be appropriate for reliability growth purposes. For this example, suppose one wants to determine the total amount of system level test time, *T*, such that the Type I or producer (contractor) risk for the program idealized reliability growth curve is 0.20 (i.e., the probability of not demonstrating the TR of 100 hours with 80 percent confidence is 0.20 for the final MTBF value, $M(T)$, obtained from the program idealized growth curve). This probability corresponds to the *P(A)* point of 0.80 (1-0.20) on the OC curve for this program.

To determine the test time *T* which will satisfy the Type I or producer (contractor) risk of 0.20, select an initial value of *T* and (as in Example 1) find $M(T)$ and the expected number of failures *E(N)*. Again, using the ratio *M(T)/TR* and *E(N)* as entries in the tables contained in (Ellner and

43

Mioduski Aug 1992) , one can determine (by double linear interpolation) the probability of demonstrating the *TR* with 80 percent confidence. An iterative procedure is then applied until the *P(A)* obtained from the table equals the desired 0.80 within some reasonable accuracy. For this example, suppose we selected 3000 hours as our initial estimate of *T* and obtained the following iterative results, shown in Table III:

TABLE III. Example 2 - planning data using iterative procedure.

T	M(T)	E(N)	P(A)
3000	117.4	36.5	<0.412
4000	128.0	44.6	<0.610
5000	136.8	52.2	<0.793
5500	140.8	55.8	0.815
5400	140.0	55.1	0.804
5300	139.2	54.4	0.790
5350	139.6	54.7	0.796
5375	139.8	54.9	0.800

Based on these results, we determine *T* = 5375 hours to be the required amount of system level test time such that the Type I or producer (contractor) risk for the program idealized growth curve is 0.20.

5.4   Subsystem Level Planning Model (SSPLAN).

5.4.1   Purpose.
The purpose of SSPLAN is to develop subsystem reliability idealized curves and associated test durations that have the following property: If each growth subsystem grows along its curve for the associated test duration then, with a prescribed probability, the realized subsystem test data will demonstrate a stated system MTBF value at a specified statistical confidence level.

As in Section 5.3, the stated system MTBF value to be demonstrated with statistical confidence from the subsystem test data will be referred to as the technical requirement and denoted by TR. Demonstrating the TR at a specified statistical confidence level provides a measure of assurance that the achieved system MTBF meets or exceeds the TR. Thus the specified confidence level will also be referred to as the assurance level.

For the case where the system is modeled solely as one growth subsystem, SSPLAN is simply SPLAN. In this instance, one is only utilizing system level growth test planning parameters. These inputs can be used to establish a target SPLAN idealized system level planning curve and associated test duration via the analytical formulas presented in Section 5.3. Alternately, the more general SSPLAN simulation procedure discussed in this section can be utilized.

5.4.2   Assumptions.
The assumptions associated with SSPLAN include:
a)   test duration is continuous;

44

b) the system may be represented as a series of independent subsystems; and
c) for each growth subsystem, the pattern of failures over the test period is in accordance with a NHPP with power law mean value function. The mean value functions may differ among the subsystems.

5.4.3   Limitations.
The limitations associated with SSPLAN include:
a) sufficient opportunities for implementation of corrective actions for each subsystem are required to permit portrayal of subsystem growth as a smooth curve;
b) the expected number of failures needs to be sufficiently large;
c) the portion of subsystem testing utilized for reliability growth planning should be reflective of the OMS/MP;
d) problems with subsystem interfaces may not be captured;
e) the initial test interval length should be reasonably small (allowing for reliability growth);
f) an average MTBF over the initial test interval should be specified for each subsystem;
g) assumption c) will typically not be satisfied if a subsystem test period contains one or more CAPs which produce a significant jump in the subsystem's MTBF; and
h) the system TR MTBF may not be reflective of an operational system MTBF (see Section 5.3.3 g).

5.4.4   Benefits.
The benefits associated with SSPLAN include:
a) allows generation of a target idealized growth curve, based on subsystem growth testing;
b) can specify desired probability of achieving a technical requirement with confidence utilizing subsystem growth test data;
c) can aggregate test duration from common subsystems on system variants under test;
d) can reduce the amount of system level testing;
e) can reduce or eliminate many failure mechanisms early in the development cycle where they may be easier to locate and correct;
f) can allow for the use of subsystem test data to monitor reliability improvement;
g) can increase product quality by placing more emphasis on lower level testing; and
h) can provide management with a strategy for conducting an overall reliability growth program.

5.4.5   Planning Factors.
The factors include both system level and subsystem level values. The system level planning factor is the system technical requirement, TR. The subsystem planning factors for developing the system planning curve include: the subsystem initial test time, t_I; the subsystem initial MTBF, M_I; the Management Strategy, MS; and the probability of observing at least one B-mode failure for the subsystem, *Prob*. The three strategies or options are:
a) t_I, M_I, *MS;*
b) t_I, *MS*, *Prob*; or
c) M_I, *MS*, *Prob*.

45

5.4.6   Considerations.
It is important for the subsystem reliability growth process to adhere as closely as possible to the following considerations:
   a) Potential high-risk interfaces need to be identified and addressed through joint subsystem testing,
   b) Subsystem usage/test conditions need to be in conformance with the proposed system level operational environment as envisioned in the OMS/MP
   c) Failure Definitions/Scoring Criteria (FD/SC) formulated for each subsystem need to be consistent with the FD/SC used for system level test evaluation.

5.4.7   Overview of SSPLAN Approach.
SSPLAN provides the user with a means to develop subsystem testing plans for demonstrating a stated system MTBF value prior to system level testing. In particular, the model is used to develop subsystem reliability growth planning curves and associated test times that, with a specified probability, support achieving a system MTBF value with a specified confidence level. More precisely, a probability is associated with the subsystem MTBFs growing along a set of planned growth curves for given subsystem test durations. This probability is termed the probability of acceptance, P_A, which is the probability that the system level TR will be demonstrated at the specified confidence level. The complement of P_A, 1-P_A, is termed the producer (contractor) risk, which is the risk of not demonstrating the system TR at the specified confidence level when the subsystems are growing along their target growth curves for the prescribed test durations. Note that P_A also depends on the fixed MTBF of any non-growth subsystem and on the lengths of the demonstration tests on which the non-growth subsystem MTBF estimates are based.

One of SSPLAN's primary outputs is the growth subsystem test times. If the growth subsystems were to grow along the planning curves for these test times, then the probability would be P_A that the subsystem test data demonstrate the system TR at the specified confidence level. The model determines the subsystem test times by using a specified fixed allocation of the combined growth subsystem final failure intensities to each of the individual growth subsystems.

As a reliability management tool, the model can serve as a means for prime contractors to coordinate/integrate the reliability growth activities of their subcontractors as part of their overall strategy in implementing a subsystem reliability test program for their developmental systems.

5.4.8   Methodology.
The SSPLAN methodology assumes that a system may be represented as a series of $K \geq 1$ independent subsystems, as shown in Figure 25.

$$\boxed{\text{System}} = \boxed{\text{Subsystem 1}} + \; ... \; + \boxed{\text{Subsystem K}}$$
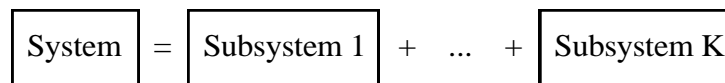
FIGURE 25. System Architecture.

This means that a failure of any single subsystem results in a system level failure and that a failure of a subsystem does not influence (either induce or prevent) the failure of any other

46

subsystem. SSPLAN allows for a mixture of test data from growth and non-growth subsystems. The model assumes that for the growth subsystems, the number of failures occurring over a period of test time follows a NHPP with mean value function

$$E\left[F\left(t\right)\right] \;=\; \lambda\, t^{\beta} \qquad (\lambda,\beta,t>0)$$

where $E[F(t)]$ is the expected number of failures by time t, λ is the scale parameter, and β is the growth (or shape) parameter. The parameters λ and β may vary from subsystem to subsystem and will be subscripted by a subsystem index number when required for clarity. Non-growth subsystems are assumed to have constant failure rates.

5.4.8.1 Mathematical Basis for Growth Subsystems.
The power function shown with the initial conditions (described in this section) provides a framework to describe how SSPLAN develops reliability growth curves. Together, they provide a starting point for describing each growth subsystem's MTBF as a function of the parameters β and t. Since it is not convenient to directly work with for planning purposes, λ is related to an initial or average subsystem MTBF over an initial period of test time. First, we note that the growth parameter, β, is related to the growth rate, α, by the following:

$$\beta \;=\; 1-\alpha \qquad (\beta>0)$$

For planned growth situations, α must be in the interval (0, 1). The initial conditions for the model consist of:

      a.   an initial time period, t_I (for example, the amount of planned test time prior to the implementation of any corrective actions), and

      b.   the initial MTBF, M_I, representing the average MTBF over the interval (0, t_I] .

From this, note that

$$\lambda_I \;=\; \frac{1}{M_I} \qquad (M_I > 0)$$

is the average failure intensity over the interval (0, t_I].

In addition to using both M_I and t_I as initial growth subsystem input parameters, the model allows a third possible input parameter, MS. The product $(MS)\lambda_I$ is the average failure rate due to correctable failure modes over the initial test interval $(0, t_I]$.
The relationship among these three parameters is addressed in the following discussion.

Since reliability growth occurs when correctable failure modes are surfaced and (successful) corrective actions are incorporated, it is desired to have a high probability of observing at least one correctable failure by time t_I (a probability of 0.95 is utilized below). From our assumptions, the number of failures that occur over the initial time period t_I is Poisson distributed with expected value $\lambda_I t_I$. Thus

$$0.95 \;=\; 1-e^{-(ms\times\lambda_I\times t_I)} \;=\; 1-e^{-\left(\frac{ms\times t_I}{M_I}\right)} \quad (0<ms\le1)$$

Based on this, it is evident that specifying any two of the parameters is sufficient to determine the third parameter. Thus, when using SSPLAN, the user has three options when entering the initial conditions for growth subsystems. Caution must be exercised in utilizing the option to solve for t_I. Even if there is a high probability of observing one or more B-modes by the value of t_I obtained from this option, it may not be suitable to use. Such a t_I should only be used if a

significant amount of growth is expected to occur prior to the start of the next test period. In particular, t_I should be chosen large enough so that by the start of the test period, after t_I, one or more corrective actions are expected to be implemented that significantly impacts the subsystem's reliability.

The derivative with respect to time of the expected number of failures function, λt^β, is

$$\rho(t) = \lambda \beta t^{\beta-1}$$

The function $\rho(t)$ represents the instantaneous failure intensity at time t. The reciprocal of $\rho(t)$, shown below, is the instantaneous MTBF at time t:

$$M(t) = \frac{1}{\rho(t)}$$

These equations provide much of the foundation for a discussion of how SSPLAN develops reliability growth curves for growth subsystems. Figure 26 below shows a graphical representation of subsystem reliability growth.
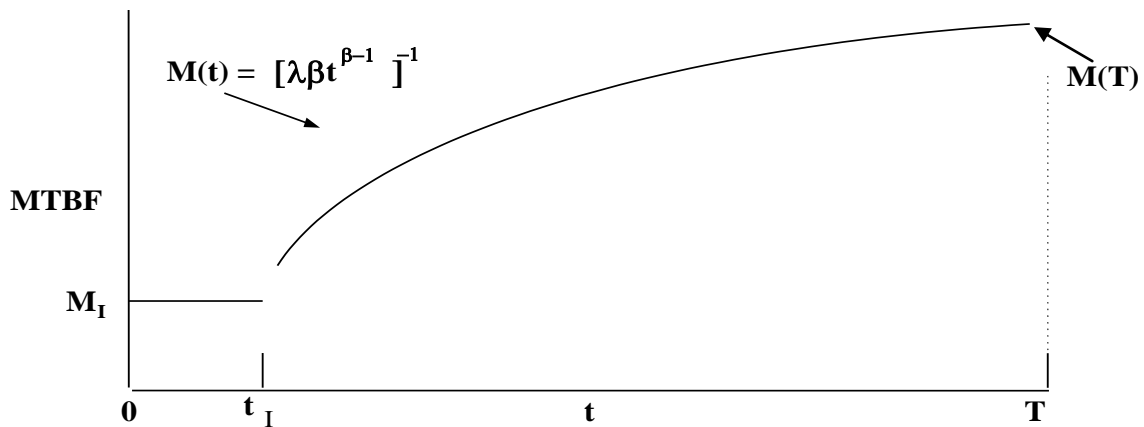
FIGURE 26. Subsystem Reliability Growth in SSPLAN.

5.4.8.2 Mathematical Basis for Non-growth Subsystems.
Based on the constant failure rate assumption, the input parameters that characterize a non-growth subsystem are its fixed MTBF planning value, M, and the planned length of the demonstration test, T, from which the constant MTBF is to be estimated. If a demonstration test is not planned for assessing the MTBF of a non-growth subsystem, then either the inputted non-growth subsystem MTBF can be treated as "known" in calculating a LCB on the system MTBF or can be treated as a statistical estimate based on historical test data generated from a demonstration test of length T.

5.4.9 Algorithm for Estimating Probability of Acceptance.
Rather than using purely analytical methods, SSPLAN uses simulation techniques to estimate the probability of achieving a stated system technical requirement with a specified confidence level. This estimate of the Probability of Acceptance, P_A, is calculated by running the simulation with a large number of trials.

Using the parameters that have been inputted and calculated at the subsystem level, the model generates "test data" for each subsystem for each simulation trial, thereby developing the data required to produce an estimate for the failure intensity for each subsystem. The test intervals and estimated failure intensities corresponding to the set of subsystems that comprise the system provide the necessary data for each trial of the simulation.

The model then uses a method developed for fixed configuration discrete data (the Lindström-Madden Method) to "roll up" the subsystem test data to arrive at an estimate for the final system reliability at a specified confidence level, namely, a statistical LCB for the final system MTBF. In order for the Lindström-Madden method to be able to combine growth subsystem test data and handle a mix of test data from both growth and non-growth subsystems, the model first converts all growth (G) subsystem test data to an "equivalent" amount of demonstration (D) test time and "equivalent" number of demonstration failures so that all subsystem results are expressed in terms of fixed configuration (non-growth) test data. By treating growth subsystem test data this way, a standard LCB formula for time-truncated demonstration testing may be used to compute the system reliability LCB for the combination of "converted" growth and non-growth test data. The combination procedure will be addressed later.

For each simulation trial, if the LCB for the final system MTBF meets or exceeds the specified system TR, then the trial is termed a success. An estimate for the probability of acceptance is the ratio of the number of successes to the number of trials.

The algorithm for estimating the probability of acceptance is described in greater detail by expanding upon the following four topics:
  a. generating "test data" estimates for growth subsystems
  b. generating "test data" estimates for non-growth subsystems
  c. converting growth subsystem data to "equivalent" demonstration data
  d. using the Lindström-Madden method for computing system level statistics

5.4.9.1 Generating Estimates for Growth Subsystems.
There are two quantities of interest for each growth subsystem for each trial of the simulation -
  a. the total amount of test time,

$$ T_{G,i} \quad = \quad \frac{1}{\left[\lambda \beta M_{G,i} \right]^{\left[\frac{1}{(\beta-1)} \right]}} \qquad \left(\beta \neq 1 \right) $$

(Determined by rearranging and combining the relationships for $\rho(t)$ and $M(t)$), and

  b. the estimated failure intensity at that time, $\hat{\rho}_{G,i}\left(T_{G,i}\right)$.

The model generates the estimated failure intensity, $\hat{\rho}_{G,i}(T_{G,i})$, by using, λ, β, $T_{G,i}$ and λt^{β} with $t = T_{G,i}$ to calculate a Poisson distributed random number, $n_{G,i}$, which serves as an outcome for the number of growth failures during a simulation trial. SSPLAN uses the realized value of $n_{G,i}$ and the corresponding conditional estimate for β, say $\hat{\beta}_{G,i}$. The value $\hat{\beta}_{G,i}$ is generated as a

random number from the distribution of the MLE $\hat{\beta}$ conditioned on the realized value of $n_{G,i}$. This conditional distribution of $\hat{\beta}$ is shown below.

$$\hat{\beta} \quad \sim \quad \frac{\left(2\beta n_{G,i}\right)}{\chi^2_{2n_{G,i}}}$$ (Broemm, Ellner and Woodworth Sep 2000)

Next, an estimate for the failure intensity, $\hat{\rho}_{G,i}\left(T_{G,i}\right)$, is obtained for each growth subsystem for each trial of the simulation. This is done via the formula estimate of ith subsystem failure intensity at $T_{G,i}$ which equals $(\hat{\beta}_{G,i} \cdot n_{G,i})/T_{G,i}$. Note the same value for $T_{G,i}$ is used on all the simulation trials.

5.4.9.2 Generating Estimates for Non-growth Subsystems.
There are two quantities of interest for each non-growth subsystem for each trial of the simulation -
    a)   the total amount of test time, $T_{n,i}$, and
    b)   the estimated failure intensity, $\hat{\rho}_{D,i}\left(T_{D,i}\right)$.
$T_{n,i}$ is the length of the demonstration test on which the non-growth subsystem MTBF estimate is based. For each non-growth subsystem, to obtain the estimated failure intensity, $\hat{\rho}_{D,i}\left(T_{D,i}\right)$, the model first calculates the expected number of failures, $T_{D,i}/M_{D,i}$. The expected number of failures is then used as an input parameter (representing the mean of a Poisson distribution) to a routine that calculates a Poisson distributed random number, $n_{D,i}$, which is an outcome for the number of failures in a demonstration test during a simulation trial. An estimate for the failure intensity follows:

$$\hat{\rho}_{D,i}\left(T_{D,i}\right) \quad = \quad \frac{n_{D,i}}{T_{D,i}}$$

5.4.9.3 Converting Growth Subsystem Data to "Equivalent" Demonstration Data.
There are two equivalency relationships that must be maintained for the approach to be valid, namely, the demonstration data and the growth data must yield:
    a.   The same subsystem MTBF point estimate:

$$\hat{M}_{D,i} \quad = \quad \hat{M}_{G,i}$$

    b.   And the same subsystem MTBF lower bound at a specified confidence level γ:

$$LCB_{D,i,\gamma} \quad = \quad LCB_{G,i,\gamma}$$

Starting with the left side of the second equivalency relationship, the LCB formula for time-truncated demonstration testing is:

$$LCB_{D,i,\gamma} \quad = \quad \frac{2T_{D,i}}{\chi^2_{2n_{D,i}+2,\gamma}}$$

Where $T_{D,i}$ is the demonstration test time, $n_{D,i}$ is the demonstration number of failures, γ is the specified confidence level and $\chi^2_{2n_{D,i}+2,\gamma}$ is a chi-squared $100\,\gamma$ percentile point with $2n_{D,i}+2$ degrees of freedom. Using an approximation equation developed by Crow (Broemm, Ellner and Woodworth Sep 2000), the LCB formula for growth testing (the right side of the LCB relationship) is:

50

$$LCB_{G,i,\gamma} \approx \frac{n_{G,i}\,\hat{M}_{G,i}}{\chi^2_{n_{G,i}+2,\gamma}}$$

where $n_{G,i}$ is the number of growth failures during the growth test, $\hat{M}_{G,i}$ is the MLE for the MTBF and $\chi^2_{n_{G,i}+2,\gamma}$ is a chi-squared 100γ percentile point with $n_{G,i}+2$ degrees of freedom.

On equating, separately, the numerator and denominator, the equivalent demonstration data are found as

$$n_{D,i} = \frac{n_{G,i}}{2}$$

and

$$T_{D,i} = \frac{T_{G,i}}{2\hat{\beta}}$$

which SSPLAN uses in converting growth subsystem data to equivalent demonstration data.

5.4.9.4 Using the Lindström-Madden Method for Computing System Level Statistics.
The Lindström-Madden Method is a method for computing the LCB for system reliability when there are two or more subsystems in series. A continuous version of the Lindström-Madden method for discrete subsystems is used to compute an approximate LCB for the final system MTBF from subsystem demonstration (non-growth) and "equivalent" demonstration (converted growth) data. The Lindström-Madden method typically generates a conservative LCB, which is to say the actual confidence level of the LCB is at least the specified level. It computes the following four estimates in order:

    a. The equivalent amount of system level demonstration test time. (Since this estimate is the minimum demonstration test time of all the subsystems, it is constrained by the least tested subsystem.)

    b. The estimate of the final system failure intensity, which is the sum of the estimated final growth subsystem failure intensities and non-growth subsystem failure rates

    c. The equivalent number of system level demonstration failures, which is the product of the previous two estimates.

    d. The approximate LCB for the final system MTBF at a given confidence level, which is a function of the equivalent amount of system level demonstration test time and the equivalent number of system level demonstration failures.

In equation form, these system level estimates are, respectively:

$$T_{D,SYS} = \min T_{D,i} \text{ for } i = 1\ldots K \qquad \hat{\rho}_{sys} = \sum_{i=1}^{K} \hat{\rho}_i$$

where $\hat{\rho}_i = \dfrac{1}{\hat{M}_{D,i}}$ and $\hat{M}_{D,i} =$ the demonstration or equivalent demonstration MTBF estimate for subsystem i.

$$N_{D,Sys} = T_{D,SYS} \times \hat{\rho}_{SYS}$$

$$LCB_\gamma = \frac{2T_{D,SYS}}{\chi^2_{2n_{D,SYS}+2,\gamma}}$$

51

5.4.9.5 Methodology for a Fixed Allocation of Subsystem Failure Intensities.
Let $\rho_{G,SYS}$ denote the sum of all the growth subsystem failure intensities at the conclusion of the subsystem growth tests. The methodology utilizes a fixed allocation, a_i, of $\rho_{G,SYS}$ to each growth subsystem i. Thus $\rho_{G,i}\,(T_{G,i})\ = a_i\,\rho_{G,SYS}$. For this allocation, SSPLAN first determines if a solution exists that satisfies the criteria given by the user during the input phase. Specifically, SSPLAN checks to see if the desired probability of acceptance can be achieved with the given failure intensity allocations and maximum subsystem test times. If a solution does exist, SSPLAN will proceed to find the solution that meets the desired probability of acceptance within a small positive number epsilon.

5.4.9.5.1 Determining the Existence of a Solution.
To determine if a solution is possible, SSPLAN calculates the maximum possible MTBF for each subsystem. The maximum subsystem MTBF is multiplied by its failure intensity allocation to determine its influence on the system MTBF. For example, if a subsystem can grow to a maximum MTBF of 1000 hours and it has a failure intensity allocation of 0.5 (that is, its final failure intensity accounts for half of the total final failure intensity due to all of the growth subsystems), then that particular subsystem will limit the combined growth subsystem maximum MTBF to 500 hours. In other words, the maximum MTBF to which the growth portion of the system can grow $M_{\rho_{sys}}$ is the minimum of the products (subsystem final MTBF multiplied by the subsystem failure intensity allocation) from among all the growth subsystems. The probability of acceptance, P_A, is then estimated using $MTBF_{G,sys}$. If the estimated P_A is less than the desired P_A, then no solution is possible within the limits of estimation precision for P_A, and SSPLAN will stop with a message to that effect.

5.4.9.5.2 Finding the Solution.
On the other hand, if the estimated P_A is greater than or equal to the desired P_A, then a solution exists. If, by chance, the desired P_A has been met (within a small number epsilon) then SSPLAN will use $MTBF_{G,sys}$ as its solution. It is more likely, however, that the estimate corresponding to $MTBF_{G,sys}$ exceeds the requirement, meaning that the program resulting in $MTBF_{G,sys}$ contains more testing than is necessary to achieve the desired P_A. SSPLAN proceeds, then, to find a value for $MTBF_{G,sys}$ that meets the desired P_A within epsilon.

5.4.10 SSPLAN Example.
The following is a successfully implemented subsystem reliability growth planning approach using SSPLAN. The objective is to develop a system planning curve utilizing subsystem level data to address the conduct of trade-off analysis between Entry into Service (EIS) calendar date and EIS expected reliability. It was felt that system reliability measurements start too late in the development cycle, thus, SSPLAN provided an approach making use of subsystem level data to begin the measurement process earlier in development. Key objectives were to predict reliability maturation and govern reliability growth issues from day one utilizing a coordinated subsystem growth strategy. (Chenard, Ellner and J.-L. 2007) (Chenard, Peree and Ellner 2006)

System level inputs include the system TR to be demonstrated with confidence from the subsystem data, statistical confidence level for the LCB on achieved system MTBF, and the specified probability p_0 that the desired assurance level would be realized under the growth

assumption. The system level output is the system target, $M_{targ, sys}$; note that $M_{targ, sys} > TR$, since one wants a reasonable probability that the gamma (the desired confidence or assurance level) LCB on the achieved system MTBF will be at least the system TR (referred to as the system MTBF goal or objective in the following references). (Chenard, Ellner and J.-L. 2007) (Chenard, Peree and Ellner 2006) (Broemm, Ellner and Woodworth Sep 2000)

Growth subsystem inputs include growth rate, initial test period $t_{1,i}$, average MTBF expected over the initial test period, allocation fraction, a_i, of growth subsystem portion of target system failure intensity contributed by growth subsystem i, and maximum subsystem test duration. Subsystem outputs include: subsystem i test duration, target MTBF for subsystem i, and expected number of subsystem failures in test.

The following is an outline of the procedure to obtain subsystem test durations (all subsystems are growth subsystems).
- $N_i(t)$ is a NHPP with rate of occurrence function $\rho_i(t)$
    - $\rho_i(t) = \lambda_i \beta_i t^{\beta_i - 1}$ where $\beta_i = 1 - \alpha_i$ and $\lambda_i = t_{1,i}^{\alpha_i} / M_{1,i}$ for growth subsystem i
- Steps
    - Use trial value $M_{targ, sys}$ to calculate $\lambda_{targ, i} = a_i \cdot \lambda_{targ, sys}$
    - Obtain trial value T_i by inverting equation $\lambda_{targ, i} = \rho_i(T_i)$
    - For each growth subsystem i simulate NHPP from 0 to T_i
    - Calculate pseudo demonstration test number of failures $n_{D, i} = n_{G,i}/2$ and $T_{D, i} = T_i/(2\beta_{est, i})$
        - $\beta_{est, i}$ is the maximum likelihood estimate of β_i from simulated growth test data
        - Equate point estimate and LCB on $M_{targ, i}$ from pseudo demonstration data to estimates from growth data to obtain pseudo demonstration test data
    - Combine subsystem pseudo demonstration data to obtain approximate LCB_γ on $M_{targ, sys}$
        - SSPLAN applies the Lindström-Madden method adapted for continuous test duration
        - Could use other methods for combining pseudo demonstration test data
    - Repeat last 3 steps prescribed number of times to estimate $Prob (LCB_\gamma \geq TR)$
    - If estimated probability is close to p_0 stop – the current trial T_i values are chosen as the subsystem test durations; otherwise adjust $M_{targ, sys}$ and repeat above steps

Table IV contains the inputs and outputs for this example.

53

TABLE IV. Inputs and outputs for SSPLAN application

MTBF TR	Confidence Level	P$_A$	Number of Subsystems
19.08	0.7	0.7	14

Subsystems	G/N	α	Initial MTBF	Initial Time	Maximum Test Time	Failure Allocation	Expected Failures	Total Test Time	Final MTBF
S 1	Growth	0.44	1558	4673	900000	0.0027	11.81	54033	8168
S 2	Growth	0.32	56	169	300000	0.0755	44.48	8835	292
S 3	Growth	0.34	95	336	80000	0.0447	38.65	12586	493
S 4	Growth	0.34	31	120	20000	0.1393	40.95	4279	158
S 5	Growth	0.33	53	159	16000	0.0803	37.55	6910	275
S 6	Growth	0.33	169	505	70000	0.0253	37.05	21641	872
S 7	Growth	0.33	79	286	200000	0.054	45.1	12341	408
S 8	Growth	0.35	30	109	40000	0.1434	33.97	3396	154
S 9	Growth	0.36	19	70	25000	0.2248	30.85	1937	98
S 10	Growth	0.34	71	214	19000	0.0597	33.06	8059	369
S 11	Growth	0.46	2893	8678	7500000	0.0015	9.81	77904	14703
S 12	Growth	0.34	37	112	11000	0.1142	33.4	4257	193
S 13	Growth	0.33	135	405	64000	0.0316	37.37	17476	698
S 14	Growth	0.43	1396	4189	95000	0.003	12.88	53981	7351

Option	Epsilon	Number of Iterations	Computed P$_A$	Computed Target MTBF
Enter the allocation of failure intensities among subsystems	0.01	500	0.7	22.05

5.5 Planning Model Based on Projection Methodology (PM2)-Continuous.

5.5.1 Purpose.

The purpose of PM2 Continuous is to develop a system-level reliability growth planning curve that incorporates the developmental test schedule and corrective action strategy. The planning curve and associated steps serve as a baseline which reliability assessments may be compared against, possibly highlighting the need for reallocation of resources.

Unlike the AMSAA Crow Planning Model, the PM2 Continuous model does not have a growth rate parameter, nor is there a comparable quantity. Furthermore, PM2 Continuous utilizes planning parameters that are directly influenced by program management, which include:
   a) M_I, the planned initial system MTBF;
   b) MS, the Management Strategy, which is the fraction of the initial failure rate addressable via corrective action;

54

c) M_G, the MTBF goal for the system to achieve at the conclusion of the reliability growth test;

d) μ_d, the planned average FEF of corrective actions;

e) T, the duration of developmental testing; and

f) the average lag time associated with corrective actions.

5.5.2 Assumptions.

The assumptions associated with PM2 Continuous include:

a) Initial B-mode failure rates $\lambda_1,..., \lambda_k$ constitute a realization of an independent random sample $\Lambda_1,..., \Lambda_k$ such that $\Lambda_i \sim$ gamma (α, β) for each $i = 1,\ldots, k$ where the density function is:

$$f_\Lambda(\lambda) = \frac{\lambda^\alpha e^{-\frac{\lambda}{\beta}}}{\alpha!\, \beta^{\alpha+1}}$$

As a rule of thumb, the potential number of B-modes should be at least five times the number of B-modes that are expected to be surfaced during the planned test period.

This assumption models mode-to-mode variation with respect to the initial B-mode rates of occurrence. The assumption for complex systems (i.e., for large k) also gives rise to the form of the function utilized in PM2 Continuous for the number of B-modes that are expected to be surfaced in t test hours. The functional form is reflective of the modeled B-mode initial failure rate variation. This same functional form can be arrived at without this assumption. For one such alternate approach that leads to this functional form, see reference (Ellner and Hall Mar 2006).

b) B- mode first occurrence times $t_1,....., t_k$ constitute a realization of an independent distributed random sample $T_1,.....T_k$ such that $T_i \sim$ Exponential (λ_i) for each $i = 1,\ldots,k;$

c) Each failure mode occurs independently and causes system failure; and

d) Corrective actions do not create new failure modes.

5.5.3 Limitations.

The limitations associated with PM2 Continuous include:

a) The portion of testing utilized for reliability growth planning should be reflective of the OMS/MP;

b) Need to have a realistic test schedule that determines the number of test hours per vehicle per month over the planned testing period; and

c) Need to have CAPs specified within the planned test schedule.

5.5.4 Benefits.

The benefits associated with PM2 Continuous include:

a) can determine the impact of changes to the planned test schedule and associated CAPs;

b) measures of programmatic risk are not sensitive to the length of the initial test phase (which is a limitation of the AMSAA Crow Planning Model);

c) can use different average corrective action lag time for each test phase;

d) provides an MTBF target to track against;

e) can be applied to programs with limited opportunities for implementation of corrective actions; and

f) utilizes planning parameters that are directly influenced by program management.

5.5.5   Overview of PM2 Continuous Approach.
PM2 Continuous reliability growth planning curves primarily consist of two components – an idealized curve, and MTBF targets for each test phase.

5.5.5.1 The Idealized Curve.
The idealized curve may be interpreted as the expected system MTBF at test time $t \in [0, T]$ that would be achieved if all B-modes surfaced by *t* were implemented with the planned average FEF. The idealized curve extends from the initial MTBF, M_I, to the goal MTBF, M_G ,where M_G is greater than the MTBF requirement, M_R. By using an Operating Characteristic (OC) Curve which incorporates consumer (Government) and producer (contractor) risks, the MTBF M_R^+ is calculated as the MTBF needed when entering the IOT&E demonstration test. Then, M_G is the goal MTBF needed at the end of DT and is determined by applying a DT-to-OT reliability degradation factor to M_R^+. Therefore, the relationship between these MTBF values is as follows: $M_G > M_R+ > M_R$. The idealized curve is a monotonically increasing function whose rate of increase depends on a few important assumptions. These assumptions include the levels of MS, the average Fix Effectiveness Factor, μ_d, and the initial and goal MTBFs used to generate the curve.

MS is defined as,

$$MS = \frac{\lambda_B}{\lambda_A + \lambda_B} \text{ whose estimate is } \frac{N_B}{N_A + N_B} \text{ only if all fixes are delayed to T.}$$

In the definition of MS, λ_B and λ_A represent the portion of the initial system failure intensity that program management will and will not address via corrective action, respectively. Note that the initial failure intensity $\lambda_I \equiv M_I^{-1} = \lambda_A + \lambda_B$. Notice that MS does not represent, in general, the fraction of corrected failures or modes (which is a common misconception). In the equation above, N_A and N_B are the number of A-mode failures and B-mode failures observed, respectively.

5.5.5.2 MTBF Targets.
The second component of the PM2 Continuous planning curve includes a sequence of MTBF steps. Since failure modes are not found and corrected instantaneously during testing, PM2 Continuous uses a series of MTBF targets to represent the actual (constant configuration) MTBF goals for the system during each test phase throughout the test program. The rate of increase in the MTBF targets depends on the planning parameters used, and are conditioned explicitly on scheduled CAPs. Unlike the AMSAA Crow Planning Model, the MTBF targets are not computed merely as arithmetic averages of the idealized curve.

5.5.6   Equations and Metrics.
Both the idealized curve and the MTBF targets are generated by the same equation, namely, the expected system failure intensity at test time $t \in [0, T]$ denoted by $\rho(t)$, where

$$\rho(t) \equiv \lambda_A + (1 - \mu_d) \cdot [\lambda_B - h(t)] + h(t)$$

56

In this equation, $\rho(t)$ represents the expected system failure intensity at test time $t \in [0, T]$ if corrective actions to all B-modes surfaced by t are implemented with an average FEF equal to μ_d. Also note that

$$\lambda_A \equiv (1 - MS) \cdot \lambda_I \equiv \frac{(1 - MS)}{M_I}$$

and

$$\lambda_B \equiv MS \cdot \lambda_I \equiv \frac{MS}{M_I}.$$

Under the model assumptions, the rate of occurrence of new B-modes at test time t for complex systems can be well approximated by the following expression:

$$h(t) \equiv \frac{\lambda_B}{1 + \beta \cdot t}$$

where β is a scale parameter that arises from the gamma distribution scale parameter. This parameter solely determines the fraction of $h(0)$ that is due to B-modes surfaced by test time t. For complex systems, the β utilized for planning purposes in the above $h(t)$ formula can be expressed solely in terms of growth planning parameters as follows:

$$\beta = \left(\frac{1}{T}\right)\left(\frac{1 - \frac{M_I}{M_G}}{MS \cdot \mu_d - \left(1 - \frac{M_I}{M_G}\right)}\right)$$

Note that the following planning parameters (which must be tailored on a system-specific basis) are inputs to PM2 Continuous:
a) M_I, the planned initial system MTBF;
b) MS, the Management Strategy, which is the fraction of the initial failure rate addressable via corrective action;
c) M_R, the requirement MTBF
d) Confidence level to demonstrate M_R (i.e. one minus consumer risk)
e) Probability of demonstrating M_R at confidence level (i.e. one minus producer risk)
f) μ_d, the planned average FEF of corrective actions;
g) T, the duration of developmental testing;
h) IOT&E test length;
i) the test phase lag times due to corrective action implementation; and
j) the degradation factor due to transition from a DT to an OT environment.

5.5.6.1 Expected Number of B-Modes.
This metric gives management an indication of the number of potential engineering redesigns that may be required. This is important to consider in determining the amount of resources

needed (e.g. time, personnel and funding). The PM2 Continuous equation for the expected number of B-modes at test time *t* is given by

$$\mu(t) \equiv \left(\frac{\lambda_B}{\beta}\right) \cdot \ln(1 + \beta \cdot t)$$

5.5.6.2 Expected Rate of Occurrence of New B-modes.
The rate of occurrence of new B-modes at test time $t \in [0, T]$, $h(t)$, is mathematically equivalent to the expected failure intensity due to unobserved B-modes (sometimes referred to as latent failure modes). This metric may be used as a measure of programmatic risk. For example, as the development effort is ongoing, one would like the estimate of *h(t)→0*. This condition indicates that program management has observed the dominant B-modes in the system. Conversely, large values of *h(t)* indicate higher programmatic risk of observing additional unseen B-modes inherent to the current system design. Thus, effective management and goal-setting of *h(t)* is a good practice.

5.5.6.3 Fraction Surfaced of the Expected Initial B-mode Failure Intensity.
For complex systems, the portion of the expected initial B-mode failure intensity associated with B-modes observed, i.e. exposed, by test time $t \in [0, T]$ can be expressed as

$$\theta(t) \equiv \frac{\beta \cdot t}{1 + \beta \cdot t}$$

where β is computed in terms of planning parameters (as previously shown). Thus, it is natural to refer to $\theta(t)$ as the exposure function and to β as the exposure parameter. Note that program management can eliminate at most the fraction $\theta(t)$ from the initial B-mode failure intensity of the system (if only B-modes surfaced in the growth test by *t* are addressed) regardless of when corrective actions are implemented, or how effective those corrective actions may be. For instance, say management is currently aware of 20 B-modes in a system. The question is: do those 20 B-modes constitute 9 percent of the initial B-mode failure intensity, or 90 percent? Thus, small values of $\theta(t)$ indicate that further testing is required to find and effectively correct additional B-modes. Conversely, large values of $\theta(t)$ indicate that further pursuit of the development effort may not be economically justifiable (i.e., the cost may not be worth the potential gains). Thus, this metric may be utilized as a measure of system maturity since it quantifies the fraction of the system B-mode failure intensity that program management can impact via corrective action efforts.

5.5.6.4 Reliability Growth Potential.
Reliability Growth Potential MTBF, M_{GP}, is the theoretical upper limit on the reliability of a system that is achieved by finding and mitigating all B-modes at a specified level of fix effectiveness. It is defined by

$$M_{GP} \equiv \lim_{t \to \infty} \rho^{-1}(t) = \frac{M_I}{1 - MS \cdot \mu}$$

Note that the growth potential only depends on three quantities:
   a) M_I, the initial MTBF of the system;
   b) *MS*, the portion of the initial failure intensity that is addressed; and
   c) μ_d, how effectively failure modes are mitigated.

As a rule of thumb, the ratio of the MTBF target, M_G, to M_{GP} should be between 0.6 and 0.8 for a well managed, but not overly aggressive reliability growth program.

The planning parameters M_I, M_G, μ_d, and MS are termed logically consistent provided $M_G < M_{GP}$, i.e.,

$$\frac{M_G}{M_I} < \frac{1}{1-(MS) \cdot \mu_d}$$

The complex system planning formula for β can be written in terms of M_{GP}. One can show,

$$\beta = \left(\frac{1}{T}\right)\left(\frac{\dfrac{M_G}{M_I}-1}{1-\dfrac{M_G}{M_{GP}}}\right)$$

For a logically consistent set of reliability growth planning parameters, one must have $M_I < M_G < M_{GP}$. Note that this ensures that $\beta > 0$.

5.5.7 Plausibility Metrics for Planning Parameters.

If T is chosen to be unrealistically small for growing from M_I to M_G, then the resulting value of β will be unduly large. This would be reflected in the function

$$\theta(t) = \frac{\beta \cdot t}{1+\beta \cdot t}$$

rising towards one at an unrealistic rate. For example, a large β could imply that $\theta(t_0) = 0.80$ for an initial time segment $[0,t_0]$ for which past experience indicates it would not be feasible to surface a set of B-modes that accounted for 80% of the initial B-mode failure intensity. In fact, $\mu(t_0)$ can be calculated to see what the choice of T implies with regard to the expected number of B-modes that account for $\theta(t_0)$. The smaller T is chosen, the larger β will be and the smaller $\mu(t_0)$ will be, holding all the other planning parameters fixed. At some point, as T is reduced, the implied number of B-modes to obtain a given value of $\theta(t_0)$, such as 0.80, should be judged to be unrealistically small based on past experience. An unrealistically large β, and corresponding $\theta(t)$ function, could also arise by choosing M_G to be an excessively high percentage of M_{GP}.

A second potentially useful metric for judging whether the planning parameters give rise to a reasonable value for β is the implied average B-mode failure rate for the expected set of surfaced B-modes over a selected initial reference test period $[0,t_0]$. This is easily calculated as the quantity λ_B minus $h(t_0)$, divided by $\mu(t_0)$, and solely depends on t_0 and β.

A third potentially useful metric to judge the reasonableness of the planning parameters is the expected number of unique B-modes surfaced over an initial test interval $[0,t_0]$ they imply. Prior experience with similar developmental programs or initial data from the current program can serve as benchmarks.

5.5.8 PM2 Continuous Example.

Once the planning parameters are chosen, the approximation for the idealized expected failure intensity, denoted by $\rho(t)$ above, can be used to generate a detailed reliability growth planning curve. For example, suppose a test schedule is laid out that gives the planned number of RAM miles accumulated on the units under test per month. Suppose the test schedule specifies blocks of calendar time for implementing corrective actions. For planning purposes, assume that in order for a failure mode to be addressed in an upcoming CAP, it must occur four months prior to the start of the period (one could use different calendar lag times for each test phase if desired). For this situation, the MTBF could be represented by a constant value between the ends of CAPs and between the start of testing and the end of the first scheduled CAP. For such a test plan, jumps in MTBF would be portrayed at the conclusion of each CAP. The increased MTBF after the jump is given by $M(t_i) = \{\rho(t_i)\}^{-1}$, where t_i denotes the accumulated test time from the calendar date that occurs four months prior to the start of the ith CAP. In such a manner, a sequence of target MTBF steps would be generated that grow from the initial MTBF value to a goal MTBF value.

Figure 27 below depicts a detailed reliability growth planning curve for a complex system for the case where A and B failure mode categories are utilized. The MTBF requirement for the system to demonstrate during IOT&E is $M_R = 65$ hours. IOT&E is an operational demonstration test of the system's suitability for fielding. Such a test is mandated by public law for major DoD developmental systems. In such a demonstration test, it may be required to demonstrate M_R with a measure of assurance. In this example, the measure of assurance is to demonstrate M_R at the 80% statistical confidence level.
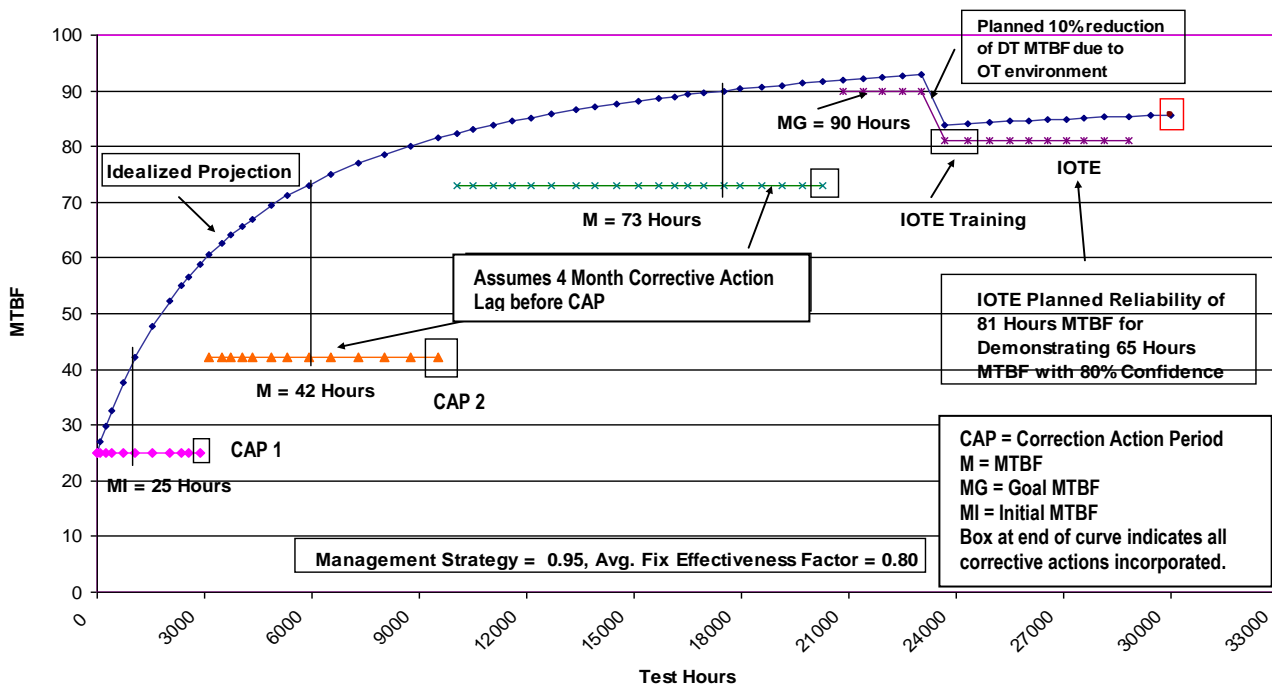
FIGURE 27. PM2 Reliability Growth Planning Curve.

60

The blue curve represents $M(t)=\{\rho(t)\}^{-1}$. Note the value $M(t)$ is the system MTBF one expects to obtain once all corrective actions to B-modes surfaced during test period $[0,t]$ are implemented. The MTBF steps are constructed from the blue curve, the schedule of CAPs, and the assumed calendar average corrective action implementation lag together with the test schedule. Note that the Developmental Test (DT) goal MTBF, M_G, was chosen to be larger than $M_R = 65$ hours, the MTBF to be demonstrated during IOT&E. To have a reasonable probability of demonstrating M_R with 80% confidence, the system must enter IOT&E with an operational MTBF value of M_R^+, which is greater than M_R. Calculating M_R^+ can be accomplished from the IOT&E test length, the desired confidence level of the statistical demonstration, and the specified probability of being able to achieve the statistical demonstration. (See OC curve). After determining M_R^+, one can consider what the developmental goal MTBF, M_G, should be at the conclusion of the development test. The value of M_G should be the goal MTBF to be attained just prior to the IOT&E training period that precedes IOT&E. The goal MTBF, M_G, associated with the development test environment, must be chosen sufficiently above M_R^+ so that the operational test environment (associated with IOT&E) does not cause the reliability of the test units to fall below M_R^+ during IOT&E. The significant drop (degradation) in MTBF that often occurs when transitioning from a DT to an OT environment could be attributable to operational failure modes that were not encountered during DT. On our example, a degradation factor of 10% was used to obtain M_G from M_R^+ ($M_G = \dfrac{M_R^+}{0.90}$).

Figure 28 illustrates the reliability growth planning curve in terms of calendar time and the step function growth pattern as corrective actions are incorporated at planned times in the test program.
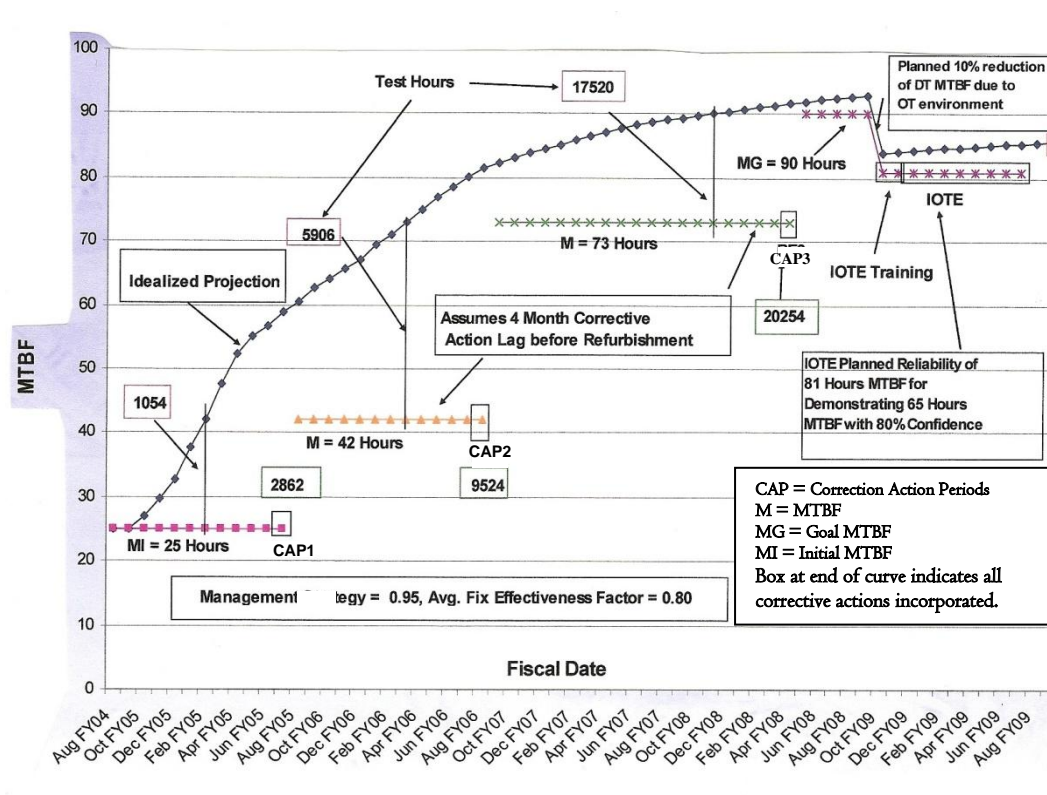
FIGURE 28. PM2 Reliability Growth Planning Curve in Calendar Time

The depiction of growth in Figures 27 and 28 do not preclude the possibility that some fixes may be implemented outside of the CAPs (i.e., during a test phase). These would typically be fixes to maintenance or operational procedures, or may include easily diagnosed and implemented design changes to hardware or software. If fixes are expected to be applied during a test phase, then a portion of the jump in MTBF (or drop in system failure intensity) portrayed in the figures at the conclusion of a test phase CAP would be realized during the test phase (prior to the associated CAP).

However, the majority of the reliability growth would typically be expected to occur due to groups of fixes that are scheduled for implementation in CAPs. These would include fixes whose implementation would involve intrusive physical procedures. For planning purposes, each test phase step in Figures 27 and 28 simply portrays the test phase MTBF that would be expected if no fixes were implemented during the test phase.

5.6   Planning Model Based on Projection Methodology (PM2)-Discrete.

5.6.1   Purpose.
PM2-Discrete is a planning model that possesses measures of programmatic risk and system maturity.

62

PM2-Discrete is the first methodology specifically developed for discrete systems and is also the first quantitative method available for formulating detailed plans in the discrete usage domain. The model has the same conditions of use as the continuous PM2 model, except for the usage domain.

PM2-Discrete utilizes planning parameters that are directly influenced by program management, which include:

a) R_I, the planned initial system reliability;

b) MS, the Management Strategy, which in the discrete case is a value between 0 and 1 that decomposes R_I into the factors R_A and R_B (which are defined below);

c) R_G, the goal reliability for the system to achieve at the conclusion of the reliability growth test;

d) μ_d, the planned average FEF of corrective actions;

e) T, the duration of developmental testing; and

f) average delay associated with corrective actions.

5.6.2 Assumptions.

The assumptions associated with PM2-Discrete include:

a) Initial B-failure mode probabilities of occurrence $p_1, \ldots, p_k$ constitute a realization of an independent and identically distributed (*iid*) random sample $P_1, \ldots, P_k$ such that $P_i \sim Beta(n, x)$ for each $i = 1, \ldots, k$, with Probability Density Function (PDF) parameterization

$$f(p_i) \equiv \begin{cases} \dfrac{\Gamma(n)}{\Gamma(x) \cdot \Gamma(n-x)} \cdot p_i^{x-1} \cdot (1-p_i)^{n-x-1} & p_i \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

With the shape parameters n and x where $\Gamma(x) \equiv \int_0^\infty t^{x-1} \cdot e^{-t} dt$ is the Euler gamma function. The associated mean, and variance of the P_i are given respectively by,

$$E(P_i) = \frac{x}{n}$$

and

$$Var(P_i) = \frac{x \cdot (n-x)}{n^2 \cdot (n+1)}$$

b) The number of trials $t_1, \ldots, t_k$ until B-failure mode first occurrence constitutes a realization of a random sample $T_1, \ldots, T_k$ such that $T_i \sim Geometric(p_i)$ for each $i = 1, \ldots, k$.

c) Potential failure modes occur independently and cause a system failure.

d) Corrective actions do not create new failure modes.

63

5.6.3 Limitations.
The limitations associated with PM2-Discrete include:
   a) The portion of testing utilized for reliability growth planning should be reflective of the OMS/MP;
   b) Need to have a realistic test schedule that determines the number of trials planned over the test period; and
   c) Need to have CAPs specified within the planned test schedule.

5.6.4 Benefits.
The benefits associated with PM2-Discrete include:
   a) PM2-Discrete can determine the impact of changes to the planned test schedule and associated CAPs.
   b) PM2-Discrete can use different average corrective action delay periods for each test phase.
   c) PM2-Discrete provides a reliability target to track against.
   d) PM2-Discrete can be applied to programs with limited opportunities for implementation of corrective actions.
   e) PM2-Discrete offers several reliability growth management metrics of basic interest including:
       i.   Expected number of failure modes observed through trial t;
       ii. Expected reliability on trial t under instantaneous failure mode mitigation;
       iii. Expected reliability growth potential;
       iv. Expected probability of failure on trial t due to a new B-mode; and
       v.   Expected probability of failure on trial t due to a repeat B-mode expressed as a fraction of the initial B-mode probability of failure in the absence of failure mode mitigation.

5.6.5 Equations and Metrics.

5.6.5.1 Expected Reliability.
The idealized curve for PM2-Discrete is generated by plotting the expected reliability on trial t (under instantaneous failure mode mitigation) versus trials. This expression is given by,

$$R(t) \equiv R_A \cdot R_B^{1-\mu_d\left(\frac{t-1}{n+t-1}\right)}$$

where
   - $R_A \in (0,1)$ is the initial system probability that an A-mode does not occur on a trial.
   - $R_B \in (0,1)$ is the initial system probability that a B-mode does not occur on a trial.
   - $\mu_d \in (0,1)$ is the planned average FEF.
   - n is the shape parameter of the beta distribution.

Formulae are required for the model parameters R_A, R_B, and n before $R(t)$ may be utilized in a reliability growth planning context. These formulae are given below and are expressed as a function of the Management Strategy, MS.

In the continuous time domain, MS is defined as the fraction of the initial system failure intensity that would be addressed via the corrective action effort if all B-modes were surfaced in test.

64

$$MS \equiv \frac{\lambda_B}{\lambda_A + \lambda_B} \qquad \text{(continuous parameterization)}$$

where λ_B and λ_A denote the portion of the system failure intensity comprised of failure modes that are, and are not, addressed by corrective action, respectively. Note that λ_B and λ_A are common parameters utilized in the continuous time domain, whereas R_B and R_A are the parameters used in the discrete-use domain. By reparameterizing, the MS for discrete-use systems is analogously defined as

$$MS \equiv \frac{-\ln R_B}{-\ln R_A - \ln R_B} \quad \text{(discrete parameterization)}$$

Note that in both the continuous and the discrete-use domain, the MS is a reliability growth planning parameter that is an input provided by the user, which represents the fractional contribution to λ_I or $-ln\ R(1)$ for the continuous and discrete cases, respectively, of the B-modes. Thus, high values of MS represent high development goals with respect to eliminating the system's propensity to fail.

Using the discrete parameterization of MS, R_B is given as,

$$R_B = R_I^{MS}$$

where R_I is the initial reliability goal input by the user, and R_A is given by,

$$R_A = R_I^{1-MS}$$

Notice that $R_A \cdot R_B = R_I^{1-MS} \cdot R_I^{MS} = R_I$, as desired.

5.6.5.2 Reliability Growth Potential.
The reliability growth potential given by PM2-Discrete is derived as the limit of the expected reliability function.

$$
\begin{aligned}
R_{GP} &\equiv \lim_{t\to\infty} R(t) \equiv \lim_{t\to\infty} R_A \cdot R_B^{1-\mu_d \cdot \left(\frac{t-1}{n+t-1}\right)} \\
&= R_A \cdot R_B^{1-\mu_d} \\
&= R_I^{1-MS} \cdot R_I^{MS \cdot (1-\mu_d)} \\
&= R_I^{1-MS \cdot \mu_d}
\end{aligned}
$$

From a management stand-point, this expression states that the achievable theoretical upper-limit on the reliability of a discrete-use system only depends on three quantities: R_I, MS, and μ_d. Since the growth potential represents the asymptote of the idealized planning curve, it is impossible to achieve a reliability target above the growth potential for a given design and associated planning parameters.

5.6.5.3 Shape Parameter.
Let T_L denote the trial number at the lag time before the last CAP in the growth program (i.e., only B-modes that occur before trial T_L are assumed to be fixed during the last CAP or a prior CAP). $R(T_L)$ is interpreted as the final reliability of the system achieved in the last CAP. Thus, the formula for the shape parameter, n, is found by equating $R(T_L) = R_G$, where R_G is the final

reliability goal for the system to achieve in a developmental test environment. The desired relationship is given by,

$$n = (T_L - 1) \cdot \frac{\ln(R_{GP}/R_G)}{\ln(R_G/R_I)}$$

If the reliability goal, R_G, is chosen higher than R_{GP}, then $n \le 0$ and the idealized curve and metrics do not exist (since n must be finite and positive). This emphasizes the importance for the practitioner to be mindful of the growth potential as well as the values of MS, fix effectiveness, the initial reliability goal, and the final reliability goal when developing a growth program, as not all sets of inputs may be theoretically possible to achieve.

5.6.5.4 Expected Number of Failure Modes.

An important question to answer when developing a reliability growth program is: how many correctable failure modes may the program uncover? The expected number of correctable failure modes observed on or before trial t is expressed as,

$$\mu(t) \equiv n \cdot \ln R_B \cdot [\psi(n) - \psi(n+t)] = \sum_{j=0}^{t-1} \frac{\ln R_B^{-n}}{n+j}$$

where $\psi(x) \equiv \dfrac{\Gamma'(x)}{\Gamma(x)}$ denotes the digamma function, and the parameters R_B and n are given by the formulae above.

5.6.5.5 Expected Fraction Surfaced of B-mode Initial System Probability of Failure.

If a potential reliability growth program is estimated to uncover a dozen B-modes, an important question to answer is: how important are they? In the absence of mitigation, is the expected probability of failure on the next trial due to these surfaced B-modes equal to 9 percent or 90 percent of the initial probability of failure due to a B-mode? A metric that addresses this is defined by,

$$\theta(t) \equiv \frac{1 - R_B^{\frac{t-1}{n+t-1}}}{1 - R_B}$$

In the absence of mitigation, the numerator is the expected probability of failure on trial t due to the occurrence of one or more B-modes that were surfaced before trial t. The denominator is the expected probability of failure on the initial trial due to the occurrence of one or more B-modes.

Note that $\theta(t)$ is simply the fraction of $1 - R_B$ that $1 - R_B^{\frac{t-1}{n+t-1}}$ represents. With this in mind, $\theta(t)$

will be referred to as the "fraction surfaced" metric. The metric equals zero for t=1 and strictly increases with an asymptote of one as t increases. It provides a measure of program management's ability to improve system reliability via the corrective action effort.

A good management practice would be to specify goals for $\theta(t)$ at important program milestones in order to track the progress of the development effort with respect to the maturing design of the system from a reliability standpoint. Small values of $\theta(t)$ indicate that further testing is required

to find and effectively correct additional failure modes. Conversely, large values of $\theta(t)$ indicate that further pursuit of the development effort to increase system reliability may not be economically justifiable (i.e., the cost may not be worth the gains that could be achieved). Note that program management can eliminate at most the portion $\theta(t)$ from the initial system unreliability prior to trial t regardless of when corrective actions are implemented or how effective they are. This follows since this metric is independent of the corrective action process.

5.6.5.6 Expected Probability of Failure Due to a New B-mode.
At the end of a reliability growth program, a reasonable question to ask is: what is the probability that we will find an unknown B-mode on the next trial? Per PM2-Discrete, the expected probability of discovering a new correctable failure mode on trial t is given as

$$h(t) \equiv 1 - R_B^{\frac{n}{n+t-1}}$$

This may be utilized as a measure of programmatic risk. For example, as the development effort is ongoing, one would like the estimate of $h(t) \to 0$. This condition indicates that program management has observed the dominant B-modes in the system. Conversely, large values of $h(t)$ indicate higher programmatic risk of observing additional unseen B-failure modes inherent to the current system design. Effective management and goal-setting of $h(t)$ is a good practice to reduce the likelihood of the customer encountering unknown B-failure modes during fielding and deployment.

5.6.6 PM2-Discrete Example
Once the planning parameters are chosen, a detailed reliability growth planning curve may be generated utilizing the expected reliability on trial t, given by R(t) in Section 5.6.5.1. For example, suppose a test schedule is laid out that gives the planned number of discrete trials accumulated in each test phase and specifies blocks of calendar time for implementing corrective actions. For planning purposes, assume that in order for a failure mode to be addressed in an upcoming CAP, it must occur prior to the last two shots that precede the CAP. These two shots are what is referred to as the corrective action lag (one could use different lags for each test phase if desired). For this situation, the reliability could be represented by a constant value between the ends of CAPs, and between the start of testing and the end of the first scheduled CAP. For such a test plan, jumps in reliability would be portrayed at the conclusion of each CAP. The increased reliability after the jump is given by $R(t_i)$ where t_i denotes the number of trials accumulated, as determined from the test schedule, prior to the start of the ith CAP minus the corrective action lag. In such a manner, a sequence of target reliability steps would be generated that grow from the initial reliability value to a goal reliability value.

Figure 29 below depicts a detailed reliability growth planning curve for a complex system for the case where A and B failure mode categories are utilized. The initial reliability for the system is $R_I = 0.8987$. The reliability requirement for the system to demonstrate during IOT&E is $R_R = 0.9200$. IOT&E is an operational demonstration test of the system's suitability for fielding. Such a test is mandated by public law for major DoD developmental systems. In such a demonstration test, it may be required to demonstrate R_R with a measure of assurance. In this example, the measure of assurance is to demonstrate R_R at the 80% statistical confidence level.

67
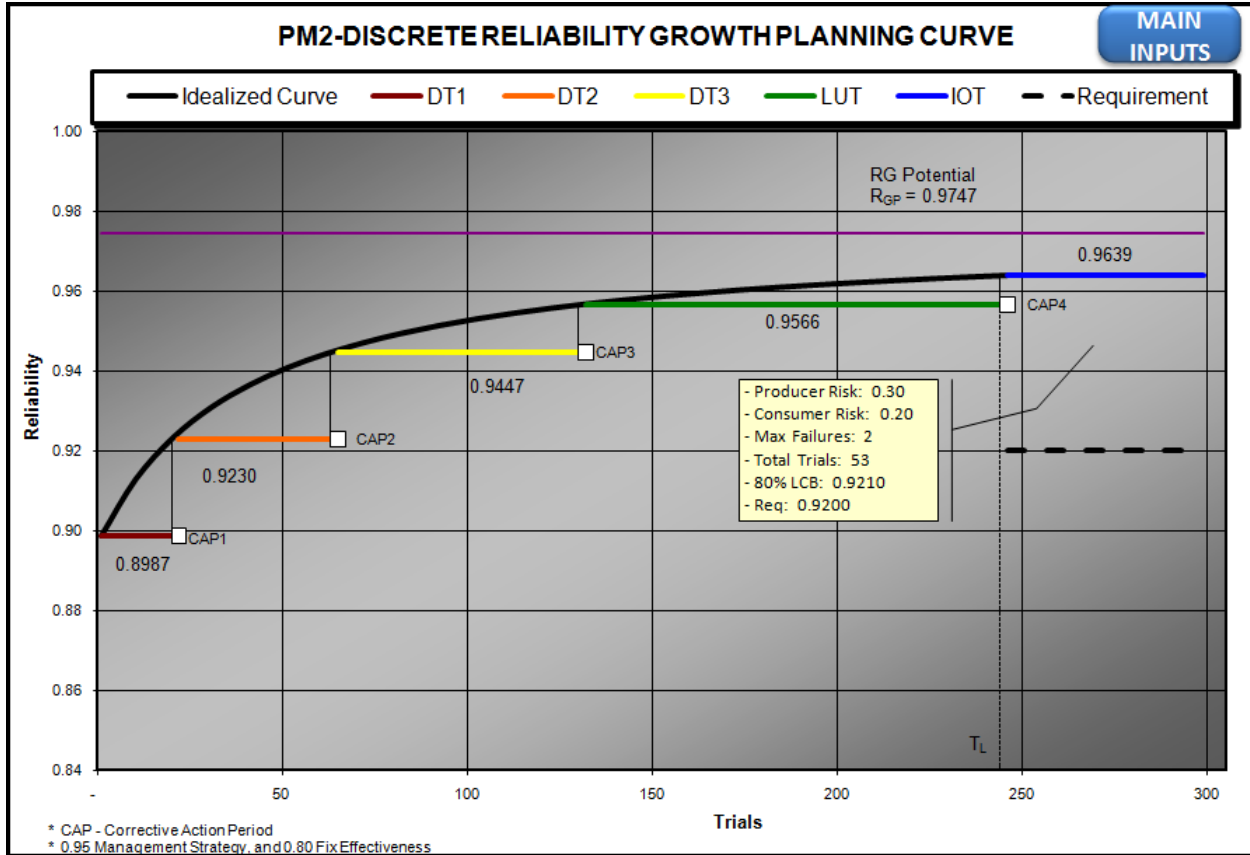
FIGURE 29. PM2-Discrete Reliability Growth Planning Curve.

The black curve represents R(t). Note the value R(t) is the system reliability one expects to obtain once all corrective actions to B-modes surfaced prior to trial t are implemented. The reliability steps are constructed from the black curve, the schedule of CAPs, and the assumed corrective action implementation lag. Note that the goal reliability, R_G, was chosen to be larger than $R_R = 0.9200$, which is the reliability to be demonstrated during IOT&E. To have a reasonable probability of demonstrating R_R with 80% confidence, the system must enter IOT&E with an operational MTBF value of R_R^+, which is greater than R_R. Calculating R_R^+ can be accomplished using an Operating Characteristic (OC) Curve, a statistical procedure for demonstration testing. The OC Curve utilizes the reliability requirement, the number of trials in IOT&E, the desired confidence level of the statistical demonstration, and the specified probability of being able to achieve the statistical demonstration. After determining $R_R^+ = 0.9639$, one can consider what the developmental goal reliability, R_G, should be at the conclusion of the development test. The value of R_G should be the goal reliability to be attained just prior to the IOT&E training period that precedes IOT&E. The goal reliability, R_G, associated with the development test environment, must be chosen sufficiently above R_R^+ so that the operational test environment (associated with IOT&E) does not cause the reliability of the test units to fall below R_R^+ during IOT&E. The significant drop (degradation) in reliability that often occurs when transitioning from a DT to an OT environment could be attributable to operational failure modes that were not encountered during DT. To obtain R_G, the degradation

68

factor, γ, is applied to R_R^+ through the equation $R_G = 1-(1-\gamma)(1-R_R^+)$. This equation follows from assuming $M_G = \frac{M_R^+}{1-\gamma}$ where M_G and M_R^+ denote the expected number of trials to failure (including the failed trial) associated with R_G and R_R^+ respectively. Note $M_G = 1/(1-R_G)$ and $M_R^+ = 1/(1-R_R^+)$. In our example, we used a degradation factor of 0 since we assumed the DT and OT environments were comparable. Therefore, for our example, $R_G = R_R^+ = 0.9639$.

The depiction of growth in Figure 29 does not preclude the possibility that some fixes may be implemented outside of the CAPs (i.e., during a test phase). These would typically be fixes to maintenance or operational procedures, or may include easily diagnosed and implemented design changes to hardware or software. If fixes are expected to be applied during a test phase, then a portion of the jump in reliability (or drop in system probability of failure) portrayed in the figure at the conclusion of a test phase CAP would be realized during the test phase (prior to the associated CAP). However, the majority of the reliability growth would typically be expected to occur due to groups of fixes that are scheduled for implementation in CAPs. These would include fixes whose implementation would involve intrusive physical procedures. For planning purposes, each test phase step in Figure 29 simply portrays the test phase reliability that would be expected if no fixes were implemented during the test phase.

5.6.7 Failure Mode Preclusion Considerations.
In a discrete-use system, there is often a well-defined sequence of stages of operation that comprise the mission of the system. For a missile system, e.g., the mission might be sequentially decomposed into prelaunch ground operations, followed by the launch stage, then flight to target, terminating in the final approach and target destruction. A failure mode in one stage could end the mission and hence preclude potential failure modes from occurring that are associated with later stages. The above planning model assumptions and associated formulas will only be suitable to the extent that such potential preclusion is a second order effect that need not be modeled in the planning methodology. If this is not deemed the case, then to minimize the preclusion effect, one could apply the planning model approach separately to each stage. If this is done then, one has to assign stage developmental reliability goals that multiply to the desired system developmental reliability goal to be achieved by the end of the growth test. This assumes the system can be treated as a serial decomposition of statistically independent stages given proper inputs from the preceding stages. The reliability goal of stage i would represent the probability that stage i completes its functions successfully, given the previous stages are successful.

5.7 Threshold Program.

5.7.1 Purpose.
The purpose of the threshold program is to determine whether the reliability of a system at selected program milestones is failing to progress according to the idealized growth curve established prior to the start of the growth test. The program can be used to compare a reliability point estimate (based on actual failure data) against a theoretical threshold value.

For multiple thresholds, the Threshold Program examines conditional distributions of MTBF estimates, given the previous thresholds were not breached. It provides a hypothesis test of whether growth is occurring along a specified idealized curve. It can be utilized for system or subsystem growth curves.

5.7.2 Assumptions.
The Threshold Program assumes that the test duration is continuous and reliability growth is governed by a NHPP with power law mean value function over one test phase.

5.7.3 Limitations.
All the conditions of SPLAN apply, which include:
   a) sufficient opportunities for corrective action implementation are required so growth is portrayed as a smooth curve;
   b) the expected number of failures needs to be sufficiently large;
   c) the portion of testing utilized for reliability growth planning should be reflective of the OMS/MP;
   d) the initial test length must be reasonably small (allowing for reliability growth); and
   e) the initial MTBF cannot be specified independent of the length of the initial test phase.

5.7.4 Methodology.
The program compares a reliability point estimate (based on actual failure data) against a theoretical threshold value. The test statistic in this procedure is the point estimate calculated from the test data. If this estimate falls at or below the threshold value, it would indicate that the achieved reliability is statistically not in conformance with the idealized growth curve. At that point, management might want to take some kind of action to restore reliability to a higher level by restructuring the program, utilizing a more intensive corrective action process, make a vendor change, or perform additional lower level testing.

The threshold program embodies a methodology that is best suited for application during a reliability growth test referred to as the Test-Analyze-Fix-Test (TAFT) program. When a failure is observed under a TAFT program, testing stops until the failure is analyzed and a corrective action is incorporated on the system. Testing then resumes with a system that (presumably) has a better reliability. The graph of the reliability for this testing strategy is a series of small increasing steps that can be approximated by a smooth idealized curve.

Recall that the initial time, T_I, marks off a period of time in which the initial reliability of the system is essentially held constant while early failures are being surfaced. Corrective actions are then implemented at the end of this initial phase, and this gives rise to improvement in the reliability. Therefore, to make threshold assessments during the period of growth, milestones should be established at points in time that are sufficiently beyond T_I.

Note also that reliability increases during test until it reaches its maximum value of M_G by the end of the test at T. Growth usually occurs rapidly early on, and then tapers off toward the end of the test phase. Therefore, in order to have sufficient time to verify that remedial adjustments (if needed) to the system are going to have the desired effect of getting the reliability back on track, milestones must be established well before T.

In practice, it is possible that the actual milestone test time may differ, for a variety of reasons, from the planned milestone time. In that case, one would simply recalculate the threshold based on the actual milestone time.

There are only three inputs necessary to define the idealized curve to build a distribution of MTBF values – the total test time, T; the final MTBF, M_G; and the growth rate, α.. The initial MTBF, M_I, and the initial time period, T_I, are not required because this implementation assumes that the curve goes through the origin. In general, this is not a reasonable assumption to make for planning purposes, but for the purposes of the Threshold Program the impact is negligible, especially since milestones are established sufficiently beyond T_I. If more than one milestone is needed, then the subsequent milestones are conditional in the sense that milestone k cannot be reached unless the system gets through the previous k-1 milestones.

A distribution of MTBF values is developed by generating a large number of failure histories from the parent curve, defined by T, M_G and α. Typically, the number of failure histories may range from 1000 to 5000, where each failure history corresponds to a simulation run. The threshold value is that reliability value corresponding to a particular percentile point of an ordered distribution of reliability values. A percentile point is typically chosen at the 10th or 20th percentile when establishing the rejection region – a small area in the tail of the distribution that allows for a test of hypothesis to be conducted to determine whether the reliability of the system is "off" track. The test statistic in this procedure is the reliability point estimate that is computed from test failure data which is compared to the threshold reliability value.

5.7.5 Threshold Program Example.

The process begins with a previously constructed idealized growth curve with a growth rate of 0.25 and the reliability growing to a final MTBF (requirement) of 70 hours by the end of 1875 hours of test time. These parameters (α, M_G, and T), along with a milestone selected at 1000 hours and a threshold percentile value of 20% were selected. A failure history number was set at 2500 histories. The resulting reliability threshold of approximately 46 hours was computed. Now, suppose that a growth test is subsequently conducted for $T = 1875$ hours. Using the AMSAA RGTMC, an MTBF point estimate is computed based on the first 1000 hours of growth test data. If the resulting MTBF point estimate at the selected milestone is above the threshold value, then there is not strong statistical evidence to reject the null hypothesis that the system is growing according to plan. However, if the resulting MTBF point estimate at the 1000 hour milestone is at or below the threshold value, then there is strong statistical evidence to reject the null hypothesis and a red flag would be raised. This red flag is a warning that the achieved reliability, as computed with the AMSAA RGTMC, is statistically not in conformance with the pre-established idealized growth curve, and that the information collected to date indicates that the system may be at risk of failing to meet its requirement. This situation should be brought to the attention of management, testers, and reliability personnel for possible remedial action to get the system reliability back on track.

6. RELIABILITY GROWTH ASSESSMENT.

6.1   Introduction.

Reliability growth assessment is an area of reliability growth that provides management the opportunity to gauge the progress of the reliability effort (i.e., reliability growth tracking) and project what reliability might be attained at a future time or date for a system (i.e., reliability growth projection). Reliability growth tracking provides a means of estimating the demonstrated reliability from test. Reliability growth projection provides a means of estimating future reliability when all corrective actions are incorporated after testing. For reliability growth tracking, the models covered are all based on the power law approach. For reliability growth projection, the models are based on both the power law and AMSAA Maturity Projection Model (AMPM) approaches.

The objectives of reliability growth tracking include:
   a) Determining if system reliability is increasing with time (i.e., growth is occurring) and to what degree (i.e., growth rate),
   b) Estimating the demonstrated reliability (i.e., a reliability estimate based on test data for the system configuration under test at the end of the test phase). This estimate is based on the actual performance of the system tested and is not based on some future configuration.
   c) Comparing the demonstrated reliability to the threshold value to ascertain that reliability is growing in accordance with planned growth.

The objectives of reliability growth projection include:
   a) Projecting expected reliability at a current or future time or milestone based on
      i. Program planning parameters (MS, FEF)
      ii. Current test results
   b) Analyzing sensitivity of reliability projection to program planning parameters (e.g., FEF)
   c) Determining maturity of the system and/or subsystems

6.1.1 Practical Data Analysis Considerations

There are many tasks that should be performed either before running the models or in conjunction with running the models. For example, a thorough review and analysis of the data should be performed in order to better understand the data, locate potential shortcomings, identify if data from different tests could be aggregated, identify outliers that could affect results, determine if data plots or analyses of the failure modes suggest anything, etc. In some cases the last analysis performed might be running the final model for estimation of point and interval values and looking into projections.

Figure 30 and the subsequent paragraphs are provided as an aid in guiding analysis.
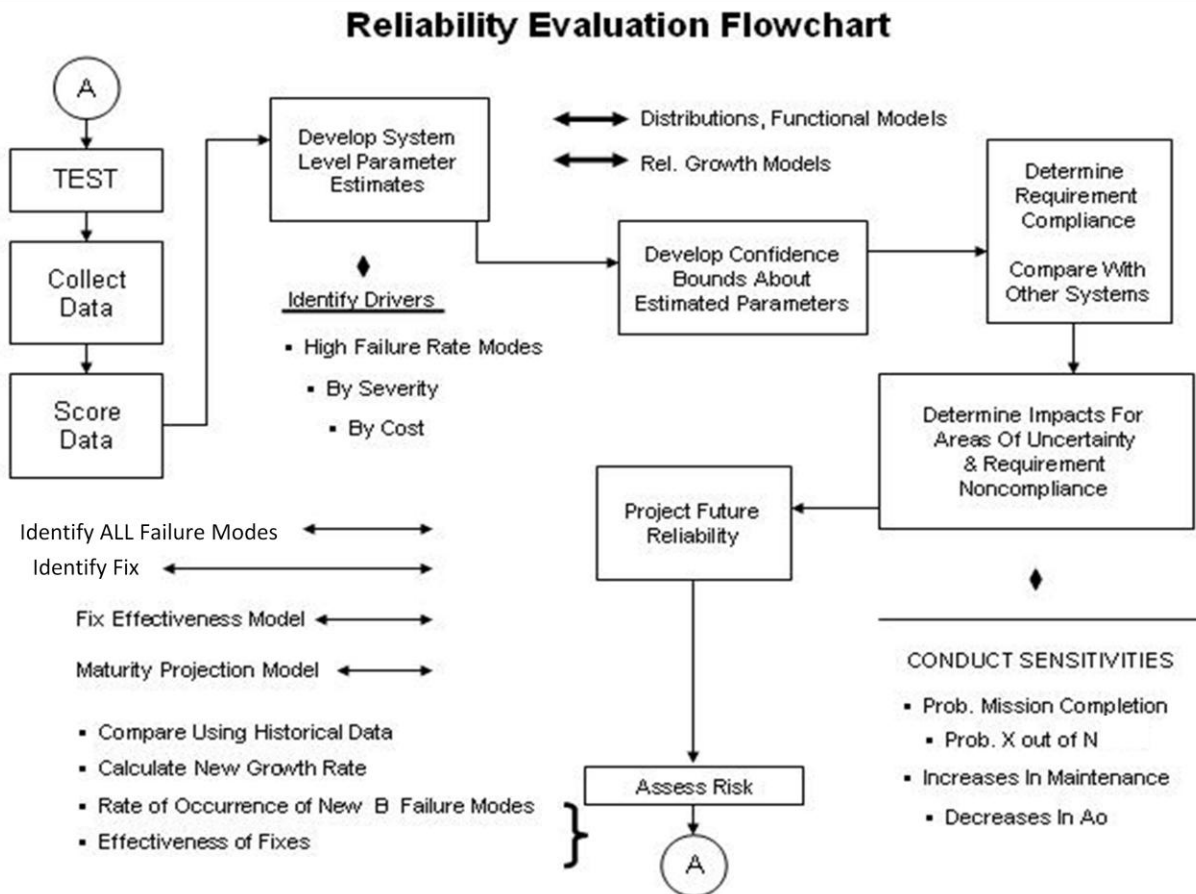
FIGURE 30. Reliability Evaluation Flowchart

The following are suggestions for initial actions that might be taken, a number of which may not seem to directly impact running the tracking models. These often lead to identification of problems and a better understanding of the data, the underlying processes, and subsequently the analysis.

a) Review the data for consistency, omissions, etc. Group the data by configuration or other logical break out, sort the data chronologically, and plot the data. This allows for identification of trends and suspect outliers.
b) Develop functional reliability models, e.g. series versus parallel.

c) Check if data can be aggregated. Look for reasons why data may be different, e.g., different test conditions or system configurations.

d) Compare the data to data from previous test phases, previous tests, or testing of predecessor systems. Determine if improvements have been made and if they are reflected in an improvement in reliability.

e) Identify failure modes and stratification. Identify driver failure modes for possible corrective action. Classify the failure modes by major subsystems.

f) Determine what conditions may have impacted the data. Determine impacts of data analysis on the program.

g) Determine applicability of the growth model. Before using a statistical model, one should determine whether the model is in reasonable agreement with the failure pattern exhibited by the data. This should be done graphically as well as statistically.

    i. Plot cumulative failure rate versus cumulative time on a log-log scale as for the Duane log-log plot or graphically plot cumulative failure (y-axis) vs. cumulative operating time (x-axis), as shown below in Figure 31. Both methods provide simple, effective means to visually assess whether or not a trend exists and whether to model using a HPP (times between failure are independent identically exponentially distributed) or NHPP (times between failure tend to increase or decrease with time or age). Table V is used to illustrate the latter plot for two systems under test, where $TTSF_i$ denotes the time to i^{th} system failure.

    ii. The Laplace test statistic can be used to determine if the times between failures are tending to increase, decrease or remain the same. The underlying probability model for the Laplace test is a NHPP having a log-linear intensity function. This test is appropriate at finding significance when the choice is between no trend and a NHPP log-linear model. In other words, if the data come from a system following the exponential law. This test will generally do better than any other test in terms of finding significance.

h) Apply growth methodology, conduct goodness-of-fit, and calculate MTBF point and interval estimates.

i) Assess the risk. Perform sensitivity analysis. If tracking growth is below the idealized planned growth, determine what growth rate is required to get back on track. Verify that the new growth rate is not too aggressive. Determine if the corrective actions are effective. Verify that the required FEFs do not exceed historical values. Determine if there is sufficient time for implementation of corrective actions and test verification.
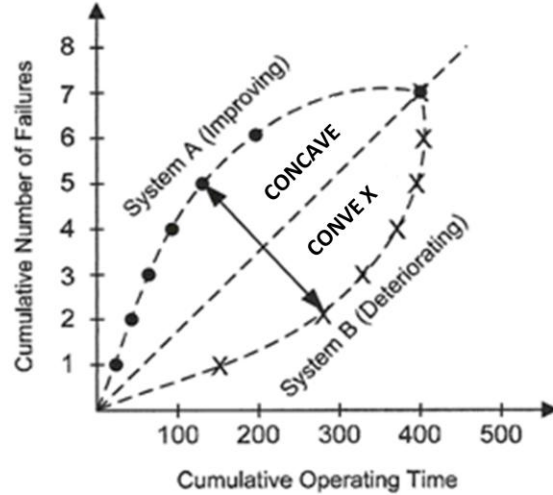
FIGURE 31. Cumulative Failures Vs. Cumulative Operating Time

TABLE V. System arrival times for a NHPP

Failure Order Number (i)	System A Arrival Times (TTSF$_i$)	System B Arrival Times (TTSF$_i$)
1	15	177
2	42	242
3	74	293
4	117	336
5	168	368
6	233	395
7	410	410

6.2 Reliability Growth Tracking.

6.2.1 Introduction.

6.2.1.1 Background.

The reliability growth tracking models presented here are based on an empirical observation, originally made by (Duane 1964). Letting $N(t)$ denote the number of failures by time t, Duane observed that the logarithm of the average cumulative number of failures observed by test time t (i.e., $N(t)/t$) versus the logarithm of the cumulative test time tends to exhibit a linear relationship with slope $-\alpha$ and intercept law: $\ln\left(\frac{N(t)}{t}\right) = \ln\omega - \alpha\ln t$. Taking the inverse logarithm of both sides of this linear equation yields the power law, namely

$$N(t) = \omega t^\beta$$

75

where ω is the scale or size parameter ($\omega > 0$), and ($\beta = 1 - \alpha$) is the shape parameter ($\beta > 0$).

6.2.1.2 Management's Role.
The role of management in the reliability growth tracking process is twofold:
a) To systematically plan and assess reliability achievement as a function of time and other program resources (such as personnel, funding, available prototypes, etc.,) and to control the ongoing rate of reliability achievement by the addition to or reallocation of these program resources based on comparisons between the planned and demonstrated reliability values; and
b) To periodically assess reliability during the test program and compare the results to the planned reliability goals and the planned reliability growth curve.

6.2.1.3 Types of Tracking Models.
Tracking models are distinguished according to the level at which testing is conducted and failure data are collected. Tracking models fall into two categories: system level and subsystem level. System level reliability growth tracking models are further classified according to the usage of the system - continuous and discrete. For continuous models, outcomes are usually measured in terms of time/miles between failures. For discrete models, outcomes are recorded in terms of distinct, countable events that give rise to probability estimates.

6.2.1.4 Test Strategies.
There are three primary test strategies: test-fix-test, test-find-test, and test-fix-find-test. In the test-fix-test strategy, failure modes are surfaced during the test and associated corrective actions are implemented during the test. In the test-find-test strategy, failure modes are surfaced during the test, but all corrective actions are delayed and implemented after completion of the test. The test-fix-find-test strategy is a combination of the previous two approaches. Reliability growth tracking models are appropriate for use in text-fix-test approaches.

6.2.1.5 Benefits.
Reliability growth tracking based on the power law has many significant benefits which prevent the process from being subjected to opinion or bias. It is instead statistically based, and therefore estimation is made on a sound and consistent basis. The following is a list of tracking methodology benefits.
a) Uses all failure data (no purging). The tracking model eliminates the need to purge data, as seen from the estimate of MTBF given in Section 6.2.2.8.1. $\hat{\beta} < 1$ ($\hat{\beta} = 1 - \hat{\alpha}$) for growth, so that the denominator reduces the number of failures in accordance with a growth situation.
b) Statistically estimates the current reliability (demonstrated value) and provides statistical confidence bounds on reliability.
c) Allows for a statistical test of the model applicability through goodness-of-fit tests.
d) Determines the direction of reliability growth from the test data - Positive growth ($\hat{\alpha} > 0$), No growth ($\hat{\alpha} = 0$), Negative growth ($\hat{\alpha} < 0$).
e) Highlights to management shortfalls in achieved reliability compared to planned reliability.

76

f) Provides a metric for tracking progress that may provide a path for early transition into the next program phase.

6.2.1.6 Elements of Reliability Growth Tracking.
Important elements of reliability growth tracking analysis include proper failure classification, test type, configuration control, and data collection. Many of these elements are spelled out in the FRACAS.

6.2.1.7 Reliability Growth Tracking Models Covered.
 The three reliability growth tracking models presented in this handbook include:
   a) AMSAA Reliability Growth Tracking Model – Continuous (RGTMC);
   b) AMSAA Reliability Growth Tracking Model – Discrete (RGTMD); and
   c) Subsystem Level Tracking Model (SSTRACK).

6.2.2   AMSAA Reliability Growth Tracking Model – Continuous (RGTMC).

6.2.2.1 Purpose.
The purpose of the AMSAA RGTMC is to assess the reliability improvement (within a single test phase) of a system during development, for which usage is measured on a continuous scale. The model may be utilized if individual failure times are known, or if failure times are only known to an interval (grouped data).

6.2.2.2 Assumptions.
The assumptions associated with the AMSAA RGTMC are:
   a) Test duration is continuous and
   b) Failures within a test phase occur according to a NHPP with power law mean value function.

6.2.2.3 Limitations.
The limitations associated with the AMSAA RGTMC include:
   a) the model will not fit the test data if large jumps in reliability occur as a result of the applied corrective action implementation strategy;
   b) the model will be inaccurate if the testing does not adequately reflect the OMS/MP;
   c) if a significant number of non-tactical fixes are implemented, the growth rate and associated system reliability will be correspondingly inflated as a result; and
   d) with respect to contributing to the reliability growth of the system, the model does not take into account reliability improvements due to delayed corrective actions.

6.2.2.4 Benefits.
The benefits associated with the AMSAA RGTMC include:
   a) the model can gauge demonstrated reliability versus planned reliability;
   b) the model can provide statistical point estimates and confidence intervals for MTBF and growth rate; and
   c) the model allows for statistical goodness-of-fit testing.

6.2.2.5 Basis for the Model.
The model is designed for tracking system reliability within a test phase, not across test phases. Accordingly, let the start of a test phase be initialized at time zero, and let $0 = t_0 < t_1 < t_2 < \ldots < t_k$ denote the cumulative test times on the system when design modifications are made. Assume the system failure rate is constant between successive t_i's, and let λ_i denote the constant failure intensity during the i-th time interval $[t_{i-1}, t_i)$. The time intervals do not have to be equal in length. Based on the constant failure intensity assumption, the number of failures during the i-th time interval, F_i, is Poisson distributed with mean $\theta_i = \lambda_i (t_i - t_{i-1})$. That is,

$$\text{Prob}(F_i = f) \quad = \quad \frac{(\theta_i)^f e^{-\theta_i}}{f!} \qquad (f = 0,1,2,\ldots)$$

a) If more than one system prototype is tested and the prototypes have the same basic configuration between modifications, then under the constant failure intensity assumption, the following are true: the time t_i may be considered as the cumulative test time to the i-th modification; and
b) F_i may be considered as the cumulative total number of failures experienced by all system prototypes during the i-th time interval $[t_{i-1}, t_i)$.

The previous discussion is summarized graphically in Figure 32.
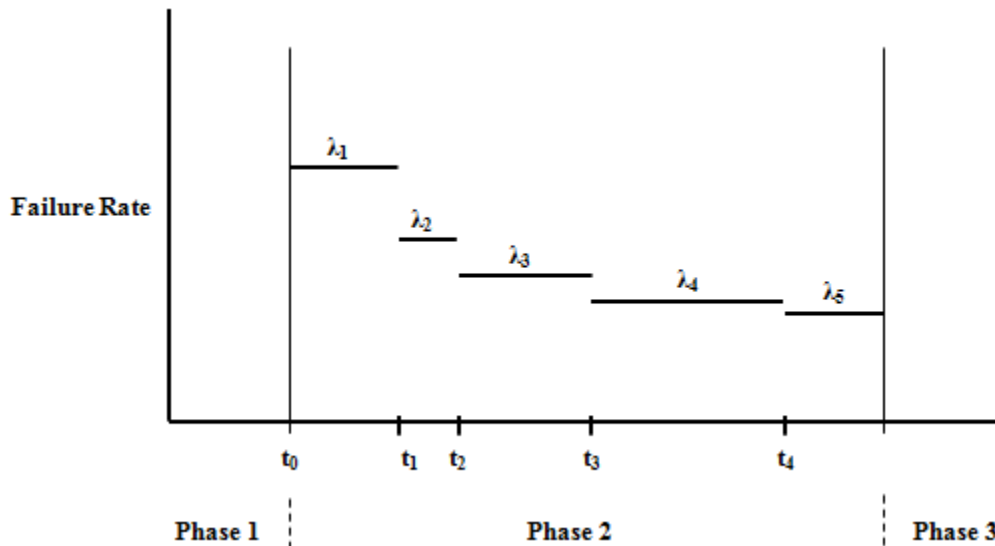
FIGURE 32. Failure Rates between Modifications

When the failure intensity is constant (homogeneous) over a test interval, then $F(t)$ is said to follow a homogeneous Poisson process with mean number of failures of the form λt. When the failure intensities change with time, e.g., from interval 1 to interval 2, then under certain conditions, $F(t)$ is said to follow a NHPP. In the presence of reliability growth, $F(t)$ would follow a NHPP with mean value function $\quad \theta(t) \quad = \quad \int_o^t \rho(y)dy$

78

where $y = \lambda_i$, $y \in [t_{i-1}, t_i)$. From the above equation, for any $t > 0$,

$$Prob[F(t) = f] = \frac{[\theta(t)]^f e^{-\theta(t)}}{f!}, \quad f{=}0,1,2,\ldots$$

The integer-valued process $\{F(t), t{>}0\}$ may be regarded as a NHPP with intensity function $\rho(t)$. If $\rho(t) = \lambda$, a constant failure rate for all t, then a system is experiencing no growth over time, corresponding to the exponential case. If $\rho(t)$ is decreasing with time, $(\lambda_1 > \lambda_2 > \lambda_3 \ldots)$, then a system is experiencing reliability growth. Finally, $\rho(t)$ increasing over time indicates a deterioration in system reliability.

6.2.2.6 Methodology.
The AMSAA RGTMC model assumes that within a test phase, failures are occurring according to a NHPP with failure intensity (rate of occurrence of failures) represented by the parametric function: $\rho(t) = \lambda\beta t^{\beta-1}$ for $\lambda, \beta, t > 0$
where λ is referred to as the scale parameter, β is referred to as the growth or shape parameter (because it characterizes the shape of the intensity function), and t is the cumulative test time. Under this model, the function

$$m(t) = \frac{1}{\rho(t)} = \left(\lambda\beta t^{\beta-1}\right)^{-1}$$

is interpreted as the instantaneous MTBF of the system at time t. When $t = T$, the total cumulative time for the system, then m(T) is the demonstrated MTBF of the system in its present configuration at the end of test. Figure 33 shows the parametric approximation to failure rates between modifications.
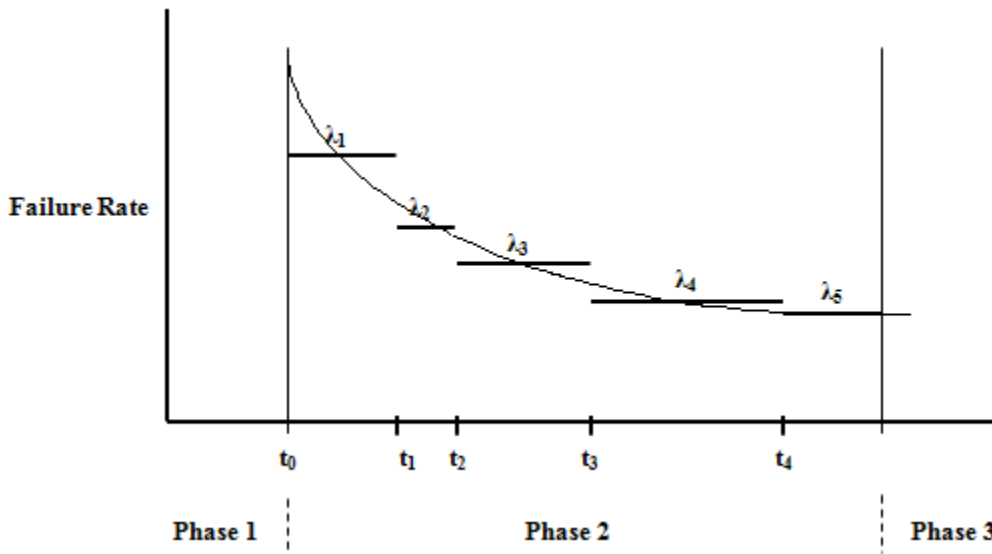
FIGURE 33. Parametric Approximation to Failure Rates between Modifications

79

Note that in Figure 33, the theoretical curve is undefined at the origin. Typically the MTBF during the initial test interval $[0, t_1]$ is characterized by a constant reliability, with growth occurring beyond t_1. Figure 34 shows an example of this type of growth.
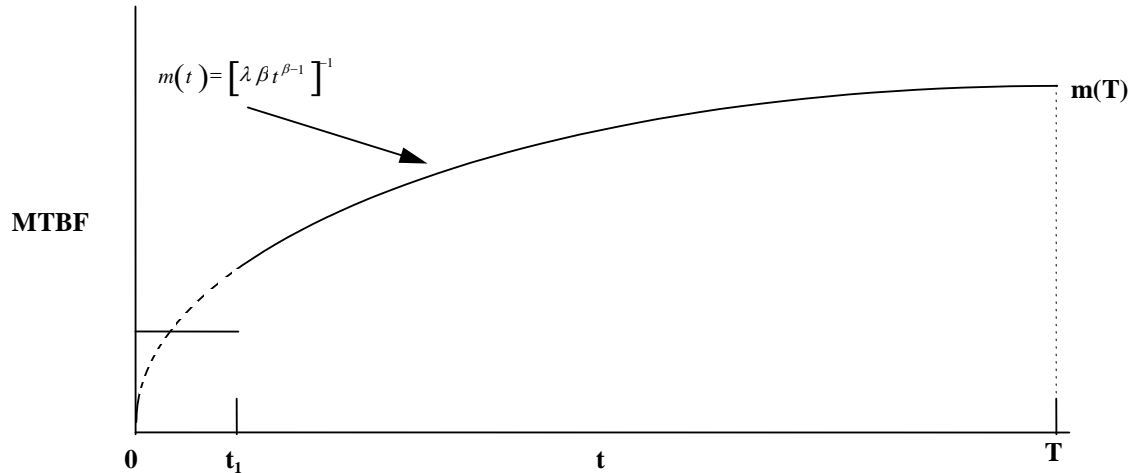
FIGURE 34. Test Phase Reliability Growth based on the AMSAA RGTMC

6.2.2.7 Equations and Metrics.

6.2.2.7.1 Cumulative Number of Failures.
The total number of failures *F(t)* accumulated on all test items in cumulative test time *t* is a Poisson random variable, and the probability that exactly *f* failures occur between the initiation of testing and the cumulative test time *t* is:

$$\text{Prob}[F(t) = f] = \frac{[\theta(t)]^f e^{-\theta(t)}}{f!}$$

in which $\theta(t)$ is the mean value function; that is, the expected number of failures expressed as a function of test time. For a NHPP, the mean value function at time t is just $\theta(t) = \int_0^t \rho(x)dx$. To describe the reliability growth process, the cumulative number of failures is then $\theta(t) = \lambda t^\beta$, where λ and β are positive parameters.

6.2.2.7.2 Number of Failures in an Interval.
The number of failures occurring in the interval from test time t_1 until test time t_2, where $t_2 > t_1$ is a Poisson random variable with mean:

$$\theta(t_2) - \theta(t_1) \quad = \quad \lambda\left(t_2^\beta - t_1^\beta\right)$$

According to the model assumption, the number of failures that occur in any time interval is statistically independent of the number of failures that occur in any non-overlapping interval, and only one failure can occur at any instant of time.

6.2.2.7.3 Intensity Function.
The intensity function is sometimes referred to as a failure rate; it is not the failure rate of a life distribution, rather it is the rate of occurrence of failures for a process, namely a NHPP.

6.2.2.8 Estimation Procedures for the Option for Individual Failure Time Data.
Modeling reliability growth as a NHPP permits an assessment of the demonstrated reliability by statistical procedures. The method of maximum likelihood provides estimates for λ and β, which are used in the estimation of the intensity function $\rho(t)$. The reciprocal of the current value of the intensity function is the instantaneous mean time between failures (MTBF) for the system.

The procedures outlined here are used to analyze data for which (a) the exact times of failure are known and (b) testing is conducted on a time terminated basis or the tests are in progress with data available through some time. The required data consist of the cumulative test time on all systems at the occurrence of each failure as well as the accumulated total test time T. To calculate the cumulative test time of a failure occurrence, it is necessary to sum the test time on every system at the point of failure. The data then consist of the F successive failure times $X_1 < X_2 < X_3 < ... < X_F$ that occur prior to T. This case is referred to as the Option for Individual Failure Time Data.

6.2.2.8.1 Point Estimation.
The method of maximum likelihood provides point estimates for the parameters of the failure intensity function. The maximum likelihood estimate (MLE) for the shape parameter β is:

$$\hat{\beta} = \frac{F}{F \ln T - \sum_{i=1}^{F} \ln X_i}$$

By equating the observed number of failures by time T (namely F) with the expected number of failures by time T (namely $E[F(T)]$) and by substituting MLE's in place of the true, but unknown, parameters we obtain:

$$F = \hat{\lambda} T^{\hat{\beta}}$$

from which we obtain an estimate for the scale parameter λ:

$$\hat{\lambda} = \frac{F}{T^{\hat{\beta}}}$$

For any time $t > 0$, the failure intensity function is estimated by:

$$\hat{\rho}(t) = \hat{\lambda}\hat{\beta}t^{\hat{\beta}-1}$$

and can be written as:

$$\hat{\rho}(T) = \hat{\lambda}\hat{\beta}T^{\hat{\beta}-1} = \hat{\beta}\left(\frac{\hat{\lambda}T^{\hat{\beta}}}{T}\right) = \hat{\beta}\left(\frac{F}{T}\right)$$

where F/T is the estimate of the intensity function for a homogeneous Poisson process.

Finally, the reciprocal of $\hat{\rho}(T)$ provides an estimate of the mean time between failures of the system at the time T and represents the system reliability growth under the model:

$$\widehat{m}(T) = \frac{1}{\hat{\rho}(T)} = \left(\hat{\lambda}\hat{\beta}T^{\hat{\beta}-1}\right)^{-1}$$

6.2.2.8.2 Interval Estimation.

Interval estimates provide a measure of the uncertainty regarding a parameter. For the reliability growth process, the parameter of primary interest is the system mean time between failures at the end of test, $m(T)$. The probability distribution of the point estimate for the intensity function at T, $\hat{\rho}(T)$, is the basis for the interval estimate for the true (but unknown) value of the intensity function at T, $\rho(T)$.

(Crow Jun 1977) developed values to facilitate computation of two-sided confidence intervals for $m(T)$ for time terminated tests by providing confidence coefficients L and U corresponding to the lower bound and upper bound, respectively. These coefficients are indexed by the total number of observed failures F and the desired confidence level γ. The two-sided confidence interval for $m(T)$ is thus:

$$L_{F,\gamma}\widehat{m}(T) \leq m(T) \leq U_{F,\gamma}\widehat{m}(T)$$

These tabulated values may be used to compute one-sided interval estimates (LCBs) for $m(T)$ such that:

$$L_{F,\gamma}\widehat{m}(T) \leq m(T)$$

Since the number of failures has a discrete probability distribution, the interval estimates are conservative; that is, the actual confidence level is slightly larger than the desired confidence level γ.

6.2.2.8.3 Goodness-of-Fit.

For the case where the individual failure times are known, a Cramér-von Mises statistic is used to test the null hypothesis that the NHPP properly describes the reliability growth of the system. To calculate the statistic, an unbiased estimate of the shape parameter β is used:

$$\overline{\beta} = \frac{F-1}{F}\hat{\beta}$$

This unbiased estimate of β is for a time terminated reliability growth test with F observed failures. The goodness-of-fit statistic is:

$$C_F = \frac{1}{12F} + \sum_{i=1}^{F}\left[\left(\frac{X_i}{T}\right)^{\overline{\beta}} - \frac{2i-1}{2F}\right]^2$$

where the failure times X_i must be ordered so that $0 < X_1 \leq X_2 \leq \cdots \leq X_F \leq T$. The null hypothesis that the model represents the observed data is rejected if the statistic C_F exceeds the critical value for a chosen significance level α.

Besides using statistical methods for assessing model goodness-of-fit, one should also construct an average failure rate plot or a superimposed expected failure rate plot (as shown in FIGURE

35). These plots, derived from the failure data, provide a graphic description of test results and should always be part of the reliability analysis.

6.2.2.9 AMSAA RGTMC Example Using Individual Failure Time Data.
The following example demonstrates the option for individual failure time data in which two prototypes of a system are tested concurrently with the incorporation of design changes. Each prototype was tested for 150 hours, for a total of $T = 300$ cumulative test hours.   Table VI shows the time of the failures for each prototype and the cumulative test time at each failure occurrence. There are a total of $F = 45$ failures. Note that the data are from a time terminated test.

TABLE VI. Test data for individual failure time option.
(An asterisk denotes the failed system.)

Failure Number	Prot. #1 Hours	Prot. #2 Hours	Cum Hours	Failure Number	Prot. #1 Hours	Prot. #2 Hours	Cum Hours
1	2.6*	.0	2.6	15	60.5	37.6*	98.1
2	16.5*	.0	16.5	16	61.9*	39.1	101.1
3	16.5*	.0	16.5	17	76.6*	55.4	132.0
4	17.0*	.0	17.0	18	81.1	61.1*	142.2
5	20.5	.9*	21.4	19	84.1*	63.6	147.7
6	25.3	3.8*	29.1	20	84.7*	64.3	149.0
7	28.7	4.6*	33.3	21	94.6*	72.6	167.2
8	41.8*	14.7	56.5	22	104.8	85.9*	190.7
9	45.5*	17.6	63.1	23	105.9	87.1*	193.0
10	48.6	22.0*	70.6	24	108.8*	89.9	198.7
11	49.6	23.4*	73.0	25	132.4	119.5*	251.9
12	51.4*	26.3	77.7	26	132.4	150.1*	282.5
13	58.2*	35.7	93.9	27	132.4	153.7*	286.1
14	59.0	36.5*	95.5	End	132.4	167.6	300.0

By using the 45 failure times listed under the columns labeled "Cumulative Hours" in Table VI we obtain the following estimates derived from RGTMC. The point estimate for the shape parameter is $\hat{\beta} = .826$; the point estimate for the scale parameter is $\hat{\lambda} = .404$; the estimated failure intensity at the end of the test is $\hat{\rho}(T) = .124$ failures per hour; the estimated MTBF at the end of the 300-hour test is $\hat{m}(T) = 8.07$ hours. FIGURE 35. illustrates superimposing a graph of the estimated intensity function atop a plot of the average failure rate (using six 50-hour intervals) which reveals a decreasing failure intensity indicative of reliability growth.
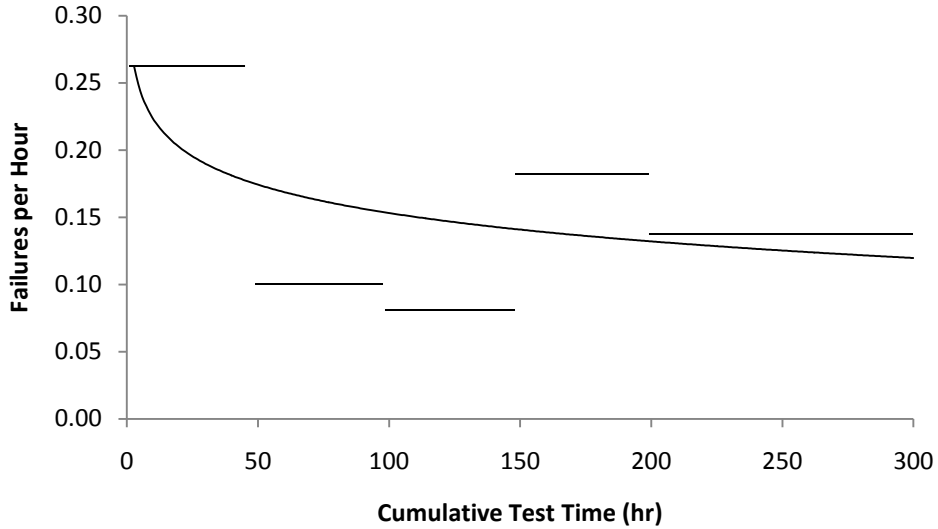
FIGURE 35. Estimated Intensity Function.

For a confidence level of 90 percent, the two-sided interval estimate for the MTBF at the end of the test is [5.7, 11.9]. These results and the estimated MTBF tracking growth curve (substituting t for T) are shown below in FIGURE 36.
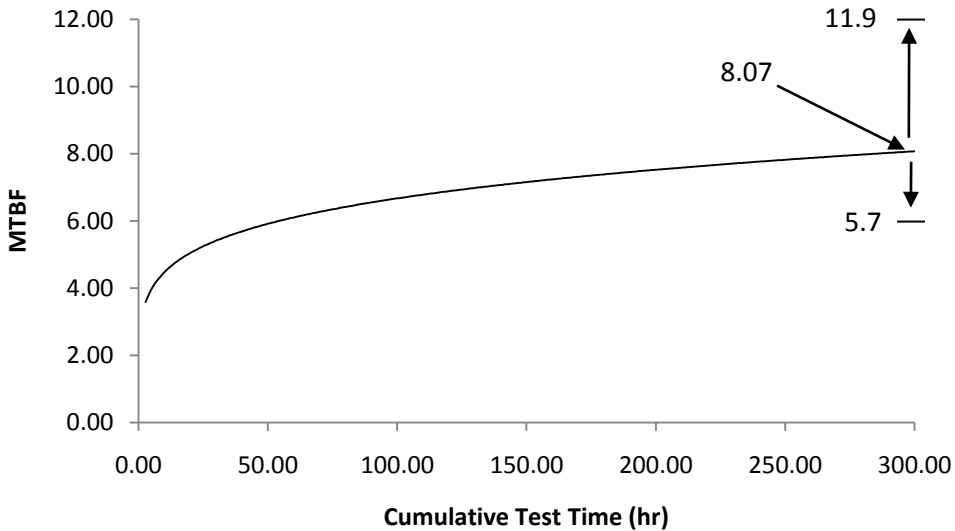
FIGURE 36. Estimated MTBF Function with 90% Interval Estimate at T=300 Hours.

Finally, to test the model goodness-of-fit, a Cramér-von Mises statistic is compared to the critical value corresponding to a chosen significance level $\alpha = 0.05$ and total observed number of failures $F = 45$. Linear interpolation is used to arrive at the critical value. Since the statistic, 0.0915, is less than the critical value, 0.218, there is no statistical evidence against using the AMSAA RGTMC for this data set.

84

6.2.3   Estimation Procedures for the Option for Grouped Data.
Reliability growth parameters can be estimated in accordance with the AMSAA RGTMC even if
the exact times of failure are unknown and all that is known is the number of failures that
occurred in each interval of time provided there are at least three intervals and at least two
intervals have failures. This case is referred to as the Option for Grouped Data. This section
describes the estimation procedures and goodness-of-fit procedures for analyzing such data. In
the following discussion, the words "group" and "interval" are interchangeable.

6.2.3.1.1   Point Estimation.
The required data consist of the total number of failures in each of K intervals of test time. The
first interval always starts at test time zero so that $t_0 = 0$. The groups do not have to be of equal
length. The observed number of failures in the interval from t_{i-1} to t_i is denoted by F_i.

The method of maximum likelihood provides point estimates for the parameters of the model.
The maximum likelihood estimate for the shape parameter β is the value that satisfies the
nonlinear equation given by

$$\sum_{i=1}^{K} F_i \left[\frac{t_i^{\hat{\beta}} \ln t_i - t_{i-1}^{\hat{\beta}} \ln t_{i-1}}{t_i^{\hat{\beta}} - t_{i-1}^{\hat{\beta}}} - \ln t_K \right] = 0$$

in which $t_0 \ln t_0$ is defined as zero. By equating the total expected number of failures to the total
observed number of failures:

$$\hat{\lambda}\, t_K^{\hat{\beta}} = \sum_{i=1}^{K} F_i .$$

Solving for $\hat{\lambda}$, we obtain an estimate for the scale parameter:

$$\hat{\lambda} = \frac{\sum_{i=1}^{K} F_i}{t_K^{\hat{\beta}}} .$$

Point estimates for the intensity function $\rho(t)$ and the mean time between failures function $m(t)$
are calculated as in the previous section that describes the Option for Individual Failure Time
Data; that is,

$$\hat{\rho}(t) = \hat{\lambda}\hat{\beta}\, t^{\hat{\beta}-1} \qquad \hat{\lambda}, \hat{\beta}, t > 0$$

$$\hat{m}(t) = [\hat{\rho}(t)]^{-1} \qquad \hat{\lambda}, \hat{\beta}, t > 0$$

The functions in the above equations provide instantaneous estimates that give rise to smooth
continuous curves, but these functions do not describe the reliability growth that occurs on a
configuration basis representative of grouped data. Under the model option for grouped data, the
estimate for the MTBF for the last group, $\hat{M}_K$, is the amount of test time in the last group divided
by the estimated expected number of failures in the last group.

$$\hat{M}_K = \frac{t_K - t_{K-1}}{\hat{E}_K} ,$$

where the estimated expected number of failures in the last group $\hat{E}_K$ is given by

85

$$\hat{E}_K \quad = \quad \hat{\lambda}\left(t_K^{\hat{\beta}} - t_{K-1}^{\hat{\beta}} \right).$$

The estimated failure intensity for the last group is then

$$\hat{\rho}_K = \frac{1}{\widehat{M}_K}.$$

6.2.3.1.2 Interval Estimation.

Approximate LCBs and two-sided confidence intervals (Crow Jun 1977) may be computed for the MTBF for the last group. A two-sided approximate confidence interval for M_K may be calculated from:

$$L_{F,\gamma}\widehat{M}_K \leq M_K \leq U_{F,\gamma}\widehat{M}_K,$$

and a one-sided approximate interval estimate for M_K may be calculated from

$$L_{F,\gamma}\widehat{M}_K \leq M_K,$$

where F is the total observed number of failures and γ is the desired confidence level

6.2.3.1.3 Goodness-of-Fit.

A chi-squared goodness-of-fit test is used to test the null hypothesis that the AMSAA RGTMC adequately represents a set of grouped data. The expected number of failures in the interval from t_{i-1} to t_i is approximated by

$$\hat{E}_i \quad = \quad \hat{\lambda}\left(t_i^{\hat{\beta}} - t_{i-1}^{\hat{\beta}} \right).$$

Adjacent intervals may have to be combined so that the estimated expected number of failures in any combined interval is at least five. Let the number of intervals after this recombination be K_R, and let the observed number of failures in the i-th new interval be O_i and the estimated expected number of failures in the i-th new interval be $\hat{E}_i$. Then the statistic

$$\chi^2 \quad = \quad \sum_{i=1}^{K_R} \frac{\left(O_i - \hat{E}_i \right)^2}{\hat{E}_i}$$

is approximately distributed as a chi-squared random variable with $K_R - 2$ degrees of freedom. The null hypothesis is rejected if the χ^2 statistic exceeds the critical value for a chosen significance level. Critical values for this statistic can be found in tables of the Chi-Square distribution.

Besides using statistical methods, an average failure rate plot or a superimposed expected failure rate plot should also be constructed. These plots provide a graphical description of test results and should be part of any reliability analysis.

6.2.3.2 AMSAA RGTMC Example Using Grouped Data.

The following example uses aircraft data to demonstrate the option for grouped data. In this example, an aircraft has scheduled inspections at intervals of twenty flight hours. For the first 100 hours of flight testing, the results are shown in Table VII.

TABLE VII. Test data for grouped option.

Start Time	End Time	Observed Number of Failures
0	20	13
20	40	16
40	60	5
60	80	8
80	100	7

There are a total of $F = 49$ observed failures from $K = 5$ intervals. The solution for $\hat{\beta}$ yields an estimate of 0.753. The scale parameter estimate is 1.53. For the last group, the intensity function estimate is 0.379 failures per flight hour and the MTBF estimate is 2.6 flight hours. TABLE VIII shows that adjacent intervals do not have to be combined since all expected numbers of failures are greater than 5. Therefore, $K_R = 5$.

TABLE VIII. Observed versus expected number of failures.

For Test Data for Grouped Data Option

Start Time	End Time	Observed Number of Failures	Estimated Expected Number of Failures
0	20	13	14.59
20	40	16	9.99
40	60	5	8.77
60	80	8	8.07
80	100	7	7.58

To test the model goodness-of-fit, a Chi-Square statistic of 5.5 is compared to the critical value of 7.8 corresponding to 3 degrees of freedom and a 0.05 significance level. Since the statistic is less than the critical value, the applicability of the model is accepted.

6.2.4   AMSAA Reliability Growth Tracking Model – Discrete (RGTMD).

Reliability growth tracking methodology may also be applied to discrete data in a manner that is consistent with the learning curve property observed by Duane for continuous data. Growth takes place on a configuration by configuration basis. Accordingly, this section describes model development and maximum likelihood estimation procedures for assessing system reliability for one-shot systems during development.

6.2.4.1 Purpose.

The purpose of the AMSAA RGTMD is to track reliability of one-shot systems during development for which usage is measured on a discrete basis, such as trials or rounds.

6.2.4.2 Assumptions.

The assumptions associated with the AMSAA RGTMD include:

87

a) test duration is discrete (i.e. trials, or rounds);
b) trials are statistically independent;
c) the number of failures for a given system configuration is distributed according to a binomial random variable; and
d) the cumulative expected number of failures through any initial sequence of configurations is given by the power law.

6.2.4.3 Limitations.
The limitations associated with the AMSAA RGTMD include:
a) The MLE solution may occur on the boundary of the constraint region of reliability, which can give an unrealistic estimate of zero for the initial reliability; and
b) Goodness-of-fit tests cannot be performed if there are a limited number of failures.

6.2.4.4 Benefits.
The benefits associated with the AMSAA RGTMD include:
a) Gauging demonstrated reliability versus planned reliability; and
b) Providing approximate LCBs for system reliability (when the MLE solution does not lie on the boundary).

6.2.4.5 Basis for the Model.
The motivation for the AMSAA RGTMD version of the AMSAA RGTMC comes from the learning curve approach for continuous data.

Let t denote the cumulative test time, and let $K(t)$ denote the cumulative number of failures by time t. The cumulative failure rate, $c(t)$, is the ratio

$$c(t) = \frac{K(t)}{t}.$$

While plotting test data from generators, hydro-mechanical devices and aircraft jet engines, Duane observed that the logarithm of the cumulative failure rate was linear when plotted against the logarithm of the cumulative test time, which yields

$$\ln c(t) = \delta - \alpha \ln t.$$

By letting $\delta = \ln \lambda$ for the y-intercept and by exponentiating both sides of the above equation, the cumulative failure rate becomes

$$c(t) = \lambda t^{-\alpha}$$

By substitution,

$$\frac{K(t)}{t} = \lambda t^{-\alpha}$$

Multiplying both sides by t and letting $\beta = 1 - \alpha$, the cumulative number of failures by t becomes

$$K(t) = \lambda t^{\beta}.$$

This power function of t is the learning curve property for $K(t)$, where $\lambda, \beta > 0$.

6.2.4.6 Development of AMSAA RGTMD.
To construct the AMSAA RGTMD, the power function developed from the learning curve property for *K(t)* is used to derive an equation for the probability of failure on a configuration basis. Model development proceeds as follows. Suppose system development is represented by *i* configurations. (This corresponds to *i* − 1 configuration changes, unless corrective actions are implemented at the end of the test phase, in which case there would be *i* configuration changes.) Let N_i be the number of trials during configuration *i*, and let M_i be the number of failures during configuration *i*. Then the cumulative number of trials through configuration *i*, T_i, is the sum of the N_i for all *i*.

$$T_i = \sum_{j=1}^{i} N_j$$

The cumulative number of failures through configuration *i*, K_i, is the sum of the M_i for all *i*.

$$K_i = \sum_{j=1}^{i} M_j$$

The expected value of K_i, $E[K_i]$, is the expected number of failures by the end of configuration *i*. Applying the learning curve property to $E[K_i]$ implies
$$E[K_1] = \lambda T_i^{\beta}.$$

Let the probability of failure for configuration one be f_1. The expected number of failures by the end of configuration one is then

$$E[K_1] = \lambda T_1^{\beta} = f_1 N_1 \Rightarrow f_1 = \frac{\lambda T_1^{\beta}}{N_1}.$$

The expected number of failures by the end of configuration two is the sum of the expected number of failures in configuration one and the expected number of failures in configuration two. This implies

$$E[K_2] = \lambda T_2^{\beta} = f_1 N_1 + f_2 N_2 = \lambda T_i^{\beta} + f_2 N_2 \Rightarrow f_2 = \frac{\lambda T_2^{\beta} - \lambda T_1^{\beta}}{N_2}.$$

By inductive reasoning, a generalized equation for the failure probability, f_i, on a configuration basis is then

$$f_i = \frac{\lambda T_i^{\beta} - \lambda T_{i-1}^{\beta}}{N_i}.$$

For the special case where $N_i = 1$ for all *i*, f_i becomes a smooth curve, g_i. The probability of failure for trial by trial data is

$$g_i = \lambda i^{\beta} - \lambda(i-1)^{\beta},$$

where *i* represents the trial number. Using f_i, the reliability (probability of success) for the *i*-th configuration is then

$$R_i = 1 - f_i,$$
and using g_i, the reliability for the *i*-th trial is

89

$$R_i = 1 - g_i.$$

The equations f_i, g_i, $1\text{-}f_i$, and $1\text{-}g_i$ are the exact model equations for tracking the reliability growth of discrete data using the AMSAA RGTMD.

6.2.4.7 Estimation Procedures for the Option for Grouped Data. This section describes procedures for estimating the parameters and an approximation equation for calculating reliability LCBs.

The MLEs for λ and β allow for point estimates for the probability of failure:

$$\hat{f}_i = \frac{\hat{\lambda}T_i^{\hat{\beta}} - \hat{\lambda}T_{i-1}^{\hat{\beta}}}{N_i} = \frac{\hat{\lambda}\left(T_i^{\hat{\beta}} - T_{i-1}^{\hat{\beta}}\right)}{N_i}$$

and the probability of success (reliability):

$$\widehat{R}_i = 1 - \hat{f}_i$$

for each configuration i.

6.2.4.8 Point Estimation.

Let $\hat{\lambda}$ and $\hat{\beta}$ be the MLEs for λ and β respectively, i.e. let $\left(\hat{\lambda},\ \hat{\beta}\right) = (\lambda, \beta)$ such that (λ, β) maximizes the discrete model likelihood function over the region $0 \leq R_i \leq 1$ for i = 1, …, K. Let $\hat{R}_i$ denote the corresponding estimate of R_i. If $0 < \widehat{R}_i < 1$ for $i = 1, …, K$ then the point $(\lambda, \beta) = \left(\hat{\lambda},\ \hat{\beta}\right)$ satisfies the following likelihood equations:

$$\sum_{i=1}^{K}\left[\lambda T_i^{\beta}\ln T_i - \lambda T_{i-1}^{\beta}\ln T_{i-1}\right]\left\{\frac{M_i}{\left[\lambda T_i^{\beta} - \lambda T_{i-1}^{\beta}\right]} - \frac{N_i - M_i}{\left[N_i - \lambda T_i^{\beta} + \lambda T_{i-1}^{\beta}\right]}\right\} = 0$$

and

$$\sum_{i=1}^{K}\left[T_i^{\beta} - T_{i-1}^{\beta}\right]\left\{\frac{M_i}{\left[\lambda T_i^{\beta} - \lambda T_{i-1}^{\beta}\right]} - \frac{N_i - M_i}{\left[N_i - \lambda T_i^{\beta} + \lambda T_{i-1}^{\beta}\right]}\right\} = 0$$

It is recommended that the MLEs be used only for this case. Situations can occur when the likelihood is maximized at a point $(\hat{\lambda}, \hat{\beta})$ such that $\hat{R}_1 = 0$ and $(\hat{\lambda}, \hat{\beta})$ does not satisfy the above equations. One such case occurs for the trial-by-trial model when a failure occurs on the first trial. To use the model in such an instance, either (1) initialize the model so that at least the first trial is a success or (2) use the grouped version and initialize with a group that contains at least one success. This should typically produce maximizing values $\hat{\lambda}, \hat{\beta}$ that satisfy the above equations with $0 < \widehat{R}_i < 1$ for $i = 1, … K$. Procedure (1) is especially appropriate if performance problems associated with an early design cause the initial failure(s). Since the assessment of the achieved reliability will depend on the model initialization and groupings, the basis for the utilized data and groupings should be considered part of the assessment. A goodness-of-fit test, such as the chi-squared test, should be used to explore whether the model provides a reasonable fit to the data and groupings. If there are insufficient failure data to perform such a test, a binomial point estimate and LCB based on the total number of successes

and trials would provide a conservative assessment of the achieved reliability R_K under the assumption that $R_K \geq R_i$ for $i = 1, \ldots, K$.

The data requirements for using the model include:

1. K number of configurations (or the final configuration)
2. M_i number of observed failures for configuration i
3. N_i number of trials for configuration i
4. T_i cumulative number of trials through configuration i

6.2.4.8.1 Interval Estimation.

A one-sided interval estimate LCB for the reliability of the final (last) configuration may be obtained from the approximation equation

$$ LCB_\gamma \approx 1 - \left(1 - \hat{R}_K\right)\left(\frac{\chi^2_{\gamma,n+2}}{n}\right), $$

where

LCB_γ = an approximate LCB at the gamma (γ) confidence level for the reliability of the last configuration, where γ is a decimal number in the interval (0,1)

$\hat{R}_K$ = a maximum likelihood estimate for the reliability of the last configuration

N = the total number of observed failures (summed) over all configurations i, $(i = 1,..,K)$

$\chi^2_{\gamma,n+2}$ = the gamma percentile point of the chi-squared distribution with $n+2$ degrees of freedom

6.2.4.8.2 Goodness-of-Fit.

Provided there is sufficient data to obtain at least five expected number of failures per group, a chi-squared goodness-of-fit test may be used to test the null hypothesis that the AMSAA RGTMD adequately represents a set of grouped discrete data or a set of trial by trial data. If these conditions are met, then the chi-squared goodness-of-fit procedures outlined previously may be used.

Again, an average failure rate plot or a superimposed expected failure rate plot should be constructed to provide a graphic description of test results as a part of the reliability analysis.

6.2.4.9 AMSAA RGTMD Example Using Grouped Data.

The following example is an application of the grouped data option of the AMSAA RGTMD for a system having four configurations of development test data. See Table IX.

TABLE IX. Test data for grouped option.

Configuration Number, i ($K = 4$)	Observed Number of Failures in Configuration i M_i	Number of Trials in Configuration i N_i	Cumulative Number of Trials Through Configuration i T_i
1	5	14	14
2	3	19	33
3	4	15	48
4	4	20	68

This is represented graphically as:

| ($M_1 = 5$) | ($M_2 = 3$) | ($M_3 = 4$) | ($M_4 = 4$) |

0	14	33	48	68
	($N_1 = 14$)	($N_2 = 19$)	($N_3 = 15$)	($N_4 = 20$)
	T_1	T_2	T_3	T_4

The MLEs for λ and β are 0.595 and 0.780, respectively. Results are presented in Table X.

TABLE X. Estimated failure rate and estimated reliability by configuration

Configuration Number, i ($K = 4$)	Estimated Failure Probability for Configuration i $\hat{f}_i$	Estimated Reliability for Configuration i $\hat{R}_i$
1	.333	.667
2	.234	.766
3	.206	.794
4	.190	.810

A plot of the estimated failure rate by configuration is shown in Figure 37.
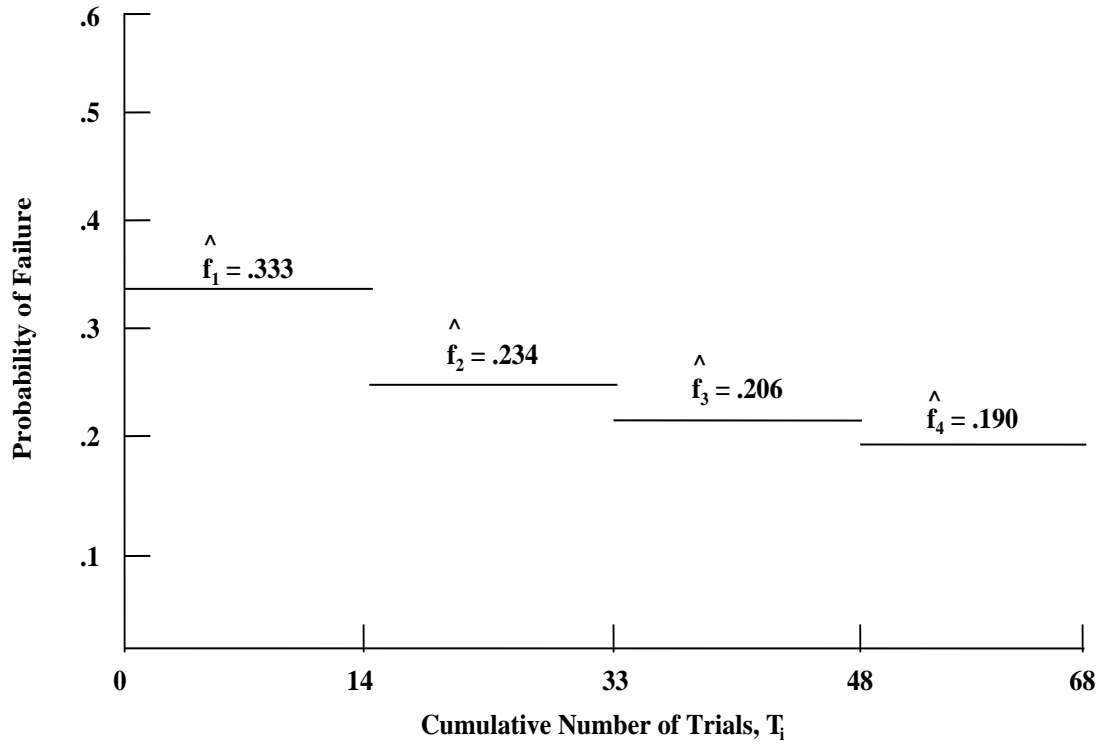
FIGURE 37. Estimated Failure Rate by Configuration.

A plot of the estimated reliability by configuration is shown in Figure 38.
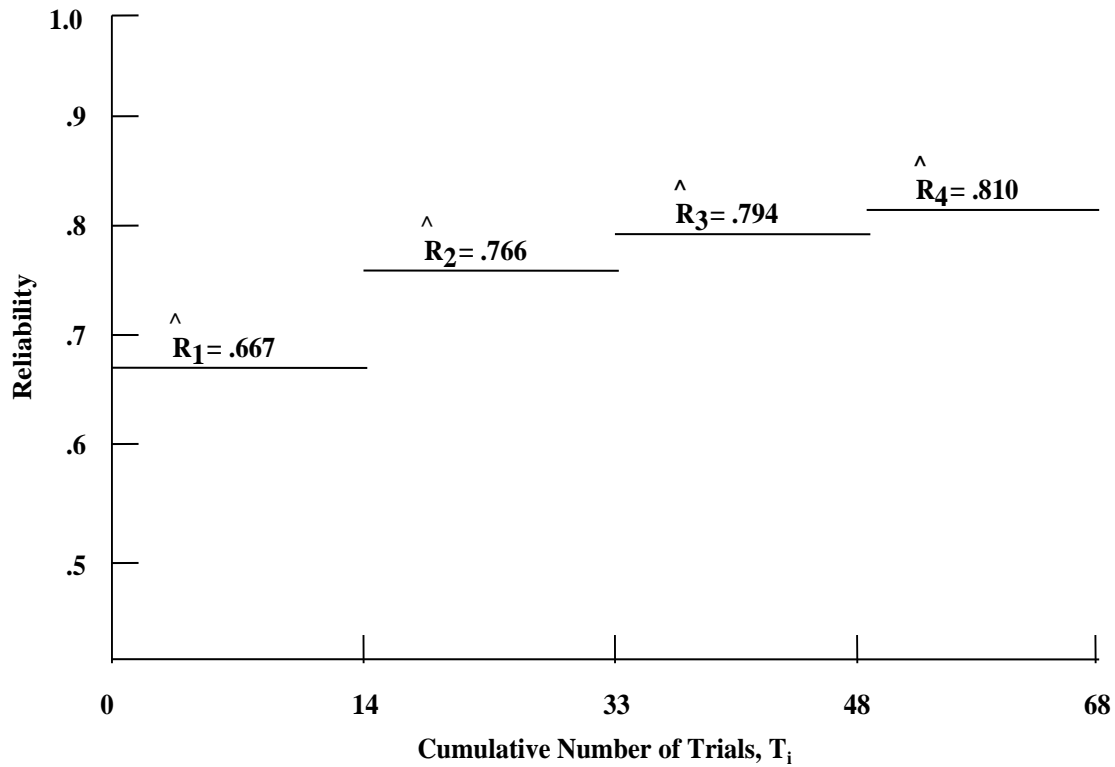
FIGURE 38. Estimated Reliability by Configuration.

93

Table XI presents approximate LCBs for the reliability of the last configuration.

TABLE XI. Approximate lower confidence bounds (LCBs) for final configuration.

Confidence Level	LCB
0.50	0.806
0.75	0.783
0.80	0.777
0.90	0.761
0.95	0.747

6.2.5   Subsystem Level Tracking Model (SSTRACK).

SSTRACK is a tool for assessing system level reliability from lower level test results. The methodology was developed to make greater use of component or subsystem test data in estimating system reliability. By representing the system as a series of independent subsystems, the methodology permits an assessment of the system level demonstrated reliability at a given confidence level from the subsystem test data. This system level assessment is permissible provided the:

a) Subsystem test conditions/usage are in conformance with the proposed system level operational environment as embodied in the OMS/MP;
b) Failure Definitions/Scoring Criteria (FD/SC) formulated for each subsystem are consistent with the FD/SC used for system level test evaluation;
c) Subsystem configuration changes are well documented; and
d) High risk interfaces are identified and addressed through joint subsystem testing.

6.2.5.1 Purpose.

The purpose of SSTRACK is to assess system level reliability from the use of component, or subsystem, test data. SSTRACK is a continuous model, but it may be used with discrete data if the number of trials is large and the probability of failure is small.

6.2.5.2 Assumptions.

The assumptions associated with SSTRACK include:

a) subsystem test duration is continuous;
b) the system can be represented as a series of independent subsystems; and
c) for each growth subsystem, the reliability improvement is in accordance with a NHPP with power law mean value function.

6.2.5.3 Limitations.

The limitations associated with SSTRACK include:

a) the model will not fit the test data if large jumps in reliability occur as a result of the applied corrective action implementation strategy;
b) the model will be inaccurate if the testing does not adequately reflect the OMS/MP;

94

c) if a significant number of non-tactical fixes are implemented, the growth rate and associated system reliability will be correspondingly inflated as a result;

d) with respect to contributing to the reliability growth of the system, the model does not take into account reliability improvements due to delayed corrective actions; and

e) the model does not address reliability problems associated with subsystem interfaces, so high-risk subsystem interfaces should be identified and addressed through joint subsystem testing.

6.2.5.4 Benefits.
The benefits associated with SSTRACK include:
a) can provide statistical point estimates and approximate confidence intervals on system reliability based on subsystem test data;
b) can accommodate a mixture of growth and non-growth subsystem test data;
c) can perform goodness-of-fit test for the NHPP subsystem assumptions;
d) it may allow for reduced system level testing by combining lower level subsystem test results in such a manner that system reliability may be demonstrated with confidence; and
e) it may allow for an assessment of the degree of subsystem test contribution toward demonstrating a system reliability requirement.

6.2.5.5 Basis for the Model.
The SSTRACK methodology supports a mix of test data from growth and non-growth subsystems. Statistical goodness-of-fit procedures are used for assessing model applicability for growth subsystem test data. For non-growth subsystems, the model uses fixed configuration test data in the form of the total test time and the total number of failures. The model applies the Lindström-Madden method for combining the test data from the individual subsystems. SSTRACK is a continuous model, but it may be used with discrete data if the number of trials is large and the probability of failure is small.

6.2.5.6 Methodology.
To be able to handle a mix of test data from growth and non-growth subsystems, the methodology converts all growth subsystem test data to its "equivalent" amount of demonstration test time and "equivalent" number of demonstration failures. This allows all subsystem results to be expressed in a common format in terms of fixed configuration (non-growth) test data. By treating growth subsystem test data this way, a standard LCB formula for fixed configuration test data may be used to compute an approximate system reliability LCB for the combination of growth and non-growth data. The net effect of this conversion process is that it reduces all growth subsystem test data to "equivalent" demonstration test data while preserving the following two important equivalency properties:

The "equivalent" demonstration data estimators and the growth data estimators must yield:
a) the same subsystem MTBF point estimate; and
b) the same subsystem MTBF LCB.

In other words, the methodology maintains the following relationships, respectively:

$$\hat{M}_D \;=\; \hat{M}_G$$

95

$$LCB_{\gamma}(D) \quad = \quad LCB_{\gamma}(G),$$

where

$$\hat{M}_D \quad = \quad \frac{T_D}{N_D}$$

$$LCB_{\gamma}(D) \quad = \quad \frac{2T_D}{\chi^2_{2N_D+2,\gamma}}.$$

Reducing growth subsystem test data to "equivalent" demonstration test data using the following equations closely satisfies the relationships cited above:

$$N_D \quad = \quad \frac{N_G}{2}$$

$$T_D \quad = \quad \hat{M}_G \times \frac{N_G}{2} \quad = \quad \frac{T_G}{2\hat{\beta}}.$$

The growth estimate for the MTBF, $\hat{M}_G$, and the estimate for the growth parameter, $\hat{\beta}$, are described in the sections on point estimation for the AMSAA RGTMC.

To compute an approximate LCB for the system MTBF from subsystem demonstration and "equivalent" demonstration data, SSTRACK uses an adaptation of the Lindström-Madden method by computing the following four estimates:

    a) the equivalent amount of system level demonstration test time, which is a reflection of the least tested subsystem because it is the minimum demonstration test time of all the subsystems;

    b) the current system failure rate, which is the sum of the estimated failure rate from each subsystem i, $i = 1,...,K$;

    c) the "equivalent" number of system level demonstration failures, which is the product of the previous two estimates; and

    d) the approximate LCB for the system MTBF at a given confidence level, which is a function of the equivalent amount of system level demonstration test time and the equivalent number of system level demonstration failures.

In equation form, these system level estimates are, respectively:

$$T_{D,sys} \quad = \quad \min T_{D,i} \quad\quad for\ i \quad = \quad 1..K$$

$$\hat{\rho}_{sys} \quad = \quad \sum_{i=1}^{K} \hat{\rho}_i$$

where

$$\hat{\rho}_i \quad = \quad \frac{1}{\hat{M}_{D,i}}$$

$$\hat{M}_{D,i} \quad = \quad the\ current\ MTBF\ estimate\ for\ subsystem\ i$$

$$N_{D,sys} \quad = \quad \hat{\rho}_{sys} \times T_{D,sys}$$

and

$$LCB_{\gamma} \quad = \quad \frac{2T_{D,sys}}{\chi^2_{2N_{D,sys}+2,\gamma}}.$$

6.2.5.7 SSTRACK Example.
This example is an application of SSTRACK to a system composed of three subsystems: one non-growth and two growth subsystems. Besides showing that SSTRACK can be used for test data gathered from dissimilar sources (namely, non-growth and growth subsystems), this particular example was chosen to show that system level reliability estimates are influenced by:
   a) the least tested subsystem; and
   b) the least reliable subsystem (the subsystem with the largest failure rate).

In this example, Subsystem 1 is a non-growth subsystem consisting of fixed configuration data of 8,000 hours of test time and 2 observed failures.

Subsystem 2 is a growth subsystem with individual failure time data. In 900 hours of test time, there were 27 observed failures occurring at the following cumulative times: 7.8, 49.5, 49.5, 51.0, 64.2, 87.3, 99.9, 169.5, 189.3, 211.8, 219.0, 233.1, 281.7, 286.5, 294.3, 303.3, 396.0, 426.6, 443.1, 447.0, 501.6, 572.1, 579.0, 596.1, 755.7, 847.5, and 858.3.

Subsystem 3 is also a growth subsystem with individual failure time data. In 400 hours of test time, there were 16 observed failures occurring at the following cumulative times: 15.04, 25.26, 47.46, 53.96, 56.42, 99.57, 100.31, 111.99, 125.48, 133.43, 192.66, 249.15, 285.01, 379.43, 388.97, and 395.25.

Table XII shows the pertinent statistics for each subsystem i. It is here that all growth (G) subsystem test data are reduced to equivalent demonstration (D) test data.

TABLE XII. Subsystem statistics.

Statistics ($i = 1,2,3$)	Subsystem 1 (Non-growth)	Subsystem 2 (Growth)	Subsystem 3 (Growth)
$T_{G,i}$	N/A	900	400
$N_{G,i}$	N/A	27	16
$\hat{M}_{G,i}$	N/A	46.53	31.37
$N_{D,i} = \dfrac{N_{G,i}}{2}$	2	13.5	8
$T_{D,i} = \hat{M}_{G,i} \times N_{D,i}$	8000	628.19	250.95
$\hat{M}_{D,i} = \hat{M}_{G,i} = \dfrac{T_{D,i}}{N_{D,i}}$	4000	46.53	31.37
$\hat{\rho}_i = \dfrac{1}{\hat{M}_{D,i}}$	2.50×10^{-4}	2.149×10^{-2}	3.188×10^{-2}

System level statistics are computed by applying the Lindström-Madden method to the equivalent demonstration data from each subsystem.

$$T_{D,sys} = \min T_{D,i(i=1,2,3)} = 250.95$$

$$\hat{\rho}_{sys} = \sum_{i=1}^{3} \hat{\rho}_i = 5.362 \times 10^{-2}$$

$$\hat{M}_{D,sys} = \frac{1}{\hat{\rho}_{sys}} = 18.7$$

$$N_{D,sys} = T_{D,sys} \times \hat{\rho}_{sys} = 13.5$$

$$LCB_{.80} = \frac{(2 \times T_{D,sys})}{\chi^2_{2N_{D,sys}+2,.80}} = 14.32 \quad (\text{confidence level} = 80\%)$$

Table XIII illustrates approximate LCBs for the system reliability (MTBF) for a range of confidence levels.

TABLE XIII. System approximate LCBs.

Confidence Level (in percent)	LCB for System MTBF
50	17.77
55	17.19
60	16.62
65	16.07
70	15.51
75	14.93
80	14.32
85	13.66
90	12.87
95	11.82
98	10.78
99	10.15

There may be cases when not all subsystems function throughout the mission. That is, the OMS/MP may specify that some (or all) subsystems function only a specific percent of the system operating time. In this situation, the term w_i, subsystem i utilization factor (weight) is introduced, where $0 < w_i \leq 1$. The following adjustments are made to the Lindström-Madden method:

$$T_{D,SYS} = \min \frac{T_{D,i}}{w_i} \ (i = 1, \cdots, K)$$
$$\hat{\rho}_{SYS} = \sum_{i=1}^{K} w_i \hat{\rho}_i.$$

For this example, all $w_i = 1$, so the results above apply. If, however, $w_1 = w_2 = 1$, and $w_3 = 0.6$, then

$$T_{D,SYS} = \min \frac{T_{D,i}}{w_i} \ (i = 1,2,3) = \min(8000, 628.19, 418.25)$$
$$= 418.25$$
$$\hat{\rho}_{SYS} = 0.040868$$

$$\hat{M}_{D,SYS} = 24.6$$

$$N_{D,SYS} = T_{D,SYS} \times \hat{\rho}_{SYS} = 17.9$$

$$LCB_{.80} = \frac{2 \times T_{D,SYS}}{\chi^2_{2N_{D,SYS}+2,0.80}} = \frac{836.5}{37.8} = 22.13$$

So, the reduction of operating time for Subsystem 3 results in an increase in MTBF from 18.7 to 24.6 (not unexpected), but the increase in equivalent system test time, $T_{D,SYS}$, from 251 to 418 results in an increase in the degrees of freedom from 29 to 37.8. Subsystem 3 previously had the least reliability, but now Subsystem 2 has the larger contribution to system unreliability. Note that the increase in degrees of freedom results in a more narrow spread between $\hat{M}_{SYS}$ and $LCB_{0.80}$ - from 4.4 to 2.5 – which is a significant reduction.

One lesson to learn from the above examples is to comprehend the entire process. This includes understanding how the system works, the system requirements, the impact that the percentages of time the subsystems operate relative to the system operation has on system reliability, and how to test items and subsystems in order to maximize utility of information and resources. Merely looking at system level testing and overall system numbers is not enough.

6.3   Reliability Growth Projection.

6.3.1   Introduction.
The basic objective of reliability growth projection is to obtain an estimate of the reliability at a current or future milestone based on Management Strategy, planned and/or implemented corrective actions, assessed fix effectiveness, and the statistical estimate of B- mode rates of occurrence. One would then analyze the sensitivity of reliability projection to program planning parameters (e.g., FEF, etc) and determine the maturity of the system or subsystem based on maturity metrics such as MTBF, rate of occurrence of B-modes, or percent of B-mode initial failure rate surfaced.

The reliability growth process applied to a complex system undergoing development involves surfacing failure modes, analyzing the modes, and implementing corrective actions to the surfaced modes. In such a manner, the system configuration is matured with respect to reliability. The rate of improvement in reliability is determined by (1) the on-going rate at which new failure modes are being surfaced; (2) the effectiveness and timeliness of the corrective actions; and (3) the set of failure modes that are addressed via corrective actions.

During testing, failure modes are identified and corrective actions incorporated in testing and/or delayed to the end of the test phase. The delayed corrective actions are usually incorporated as a group during a designated corrective action period (CAP) and the result is generally a distinct jump in the system reliability. A projection model estimates this jump in reliability due to the delayed fixes, called a "projection." These models do not simply extrapolate the tracking curve beyond the current test phase, which would require that the test conditions do not change and that the level of activities promoting reliability growth remain constant, e.g., growth rate, α.

Although there are no hard and fast rules, one needs to surface enough distinct B-modes to allow for statistical estimation of the rate of occurrence of B-modes. That is, there must be enough B-modes so that the graph of the cumulative number of B-modes versus test time appears regular enough and in conformance with the projection model's assumed mean value function so that parameters of this function can be statistically estimated. As suggested throughout this handbook, plotting the cumulative number of observed B-modes verses test time is helpful to obtain a visual comparison of the observed trend. It is also suggested that goodness-of-fit methods be employed.

Another situation in which a projection can be useful is in assessing the plausibility of meeting future reliability milestones, i.e., milestones beyond the commencement of the follow-on test.

6.3.1.1 Reliability Growth Projection Models Covered.
The reliability growth projection models presented in this handbook include:

   a) AMSAA-Crow Projection Model (ACPM);
   b) Crow Extended Reliability Projection Model;
   c) AMSAA Maturity Projection Model (AMPM);
   d) AMSAA Maturity Projection Model Based on Stein Estimation (AMPM-Stein); and
   e) Discrete Projection Model (DPM).

Models a) and b) utilize the power law approach, whereas models c), d), and e) utilize the AMPM approach.

6.3.1.2 Basic Projection Approaches Covered.
There are two basic projection approaches presented, one based on the power law model, the other based on AMPM.

6.3.1.2.1 The Power Law Approach
For the power law approach, the ACPM assumes corrective actions are delayed, but implemented prior to the next test phase. The Crow Extended Reliability Projection Model was developed wherein both delayed and non-delayed fixes are permitted. These projection models may be used to determine reliability "potential" by sensitizing on FEFs.

6.3.1.2.2 The AMSAA Maturity Projection Model (AMPM) Approach.
The AMPM is based on the approach of viewing B-failure mode rates $(\lambda_1, \cdots, \lambda_K)$ as a realization of a random sample $\Lambda = (\Lambda_1, \ldots, \Lambda_K)$ from the gamma distribution, $\Gamma(\alpha, \beta)$. Failure mode first occurrence failure times conditioned on the λ_i are assumed to be independent

100

exponentials with means λ_i for $i = 1,\ldots, K$. This allows one to utilize all the B-mode times to first occurrence observed during Test Phase I to estimate the gamma parameters - α, β.

For the ACPM, a projection applies to the test-find-test management strategy where all fixes are implemented at the end of test. The Crow Extended Reliability Projection Model applies to the test-fix-find-test management strategy wherein some fixes may be implemented in testing and the remaining at the conclusion of testing. In order to provide the assessment and management metric structure for corrective actions during and after a test, two types of B-modes are defined – BC failure modes (corrected during test) and BD failure modes (delayed until the end of the test).

6.3.1.3 Benefits.
The benefits of reliability growth projection include:
   a) Assesses reliability improvements due to surfaced failure modes and corrective actions;
   b) Provides important maturity metrics
      i.   Projections of MTBF
      ii. Rate of occurrence of new B-modes
      iii. Percent surfaced of B-mode initial failure rate;
   c) Projects expected number of new B-modes during additional testing;
   d) Assesses system reliability growth potential; and
   e) Quantifies impact of successful corrective actions and overall engineering and management strategy.

6.3.1.4 Key Assumptions.
Key assumptions for reliability growth projection include:
   a) At the start of test, there is a large unknown constant number, denoted by K, of potential B-modes that reside in the system (which could be a complex subsystem);
   b) Failure modes (both A-modes and B-modes) occur independently;
   c) Each occurrence of a failure mode results in a system failure; and
   d) No new modes are introduced by attempted fixes.

6.3.1.5 Considerations.
When applying a reliability growth projection model, it is important to note whether the estimation procedure for the system failure intensity (rate of occurrence of failures) is only valid when all the corrective actions to surfaced failure modes are delayed to the end of the test phase. Such a fix implementation strategy ensures that the initial B-mode failure rates, λ_i, are constant over the test phase. Such a λ_i can be estimated by simply dividing the test phase duration into the number of failures attributed to mode i. When this is not the case (because a corrective action is implemented for mode i prior to the conclusion of the test phase), the projection model estimation procedure must not rely on such as estimate for λ_i. In such instances, projection methodology that depends on the λ_i remaining constant during the test phase should not be used.

Database considerations for projection methodology include:
   a) Failure mode classification;
   b) Test exposure (i.e. land, water, etc.);
   c) Configuration control; and

d) Engineering assessments of fix effectiveness.

Data Requirements for Projection analysis include:
a) First occurrence time for each distinct correctable failure mode;
b) Occurrence time for each repeat of a distinct correctable failure mode;
c) Number of non-correctable failures;
d) FEF for each correctable failure mode or the average over all; and
e) Total test duration

Note that the Crow Extended Reliability Projection Model has additional data requirements.

6.3.2 AMSAA-Crow Projection Model (ACPM).
This model considers the case where all corrective actions to surfaced B-modes are implemented at the end of the current test phase, prior to commencing a follow-on test phase. Thus all fixes are delayed, making it a test-find-test strategy. The current test phase will be referred to as Phase I and the follow-on test phase as Phase II.

6.3.2.1 Purpose.
The purpose of ACPM is to estimate the system reliability at the beginning of a follow-on test phase by taking into consideration the reliability improvement from delayed fixes.

6.3.2.2 Assumptions.
The assumptions associated with ACPM include:
a) Test duration is continuous;
b) Corrective actions are implemented as delayed fixes at the end of the test phase;
c) Failure modes can be categorized as either A-modes and B-modes;
d) Failure modes occur independently and cause system failure;
e) There are a large number of potential B-modes;
f) The number of B-modes surfaced can be approximated by a NHPP with power law mean value function;
g) The time to first occurrence is exponentially distributed for each failure mode;
h) The number of A-mode failures by test duration t conforms to a homogeneous Poisson process over the test phase; and
i) The number of B-mode failures by test duration t conforms to a homogenous Poisson process over the test phase.

6.3.2.3 Limitations.
The limitations associated with ACPM include:
a) All corrective actions must be delayed;
b) FEFs are often a subjective input; and
c) Projection accuracy can be degraded via reclassification of A-modes to B-modes.

6.3.2.4 Benefits.
The benefits associated with ACPM include:
a) The ability to project the impact of delayed corrective actions on system reliability; and
b) The projection takes into account the contribution to the system failure intensity due to unobserved B-failure modes.

102

6.3.2.5 Overview of ACPM Approach.

The ACPM and associated parameter estimation procedure was developed to assess the reliability impact of a group of delayed fixes. In particular, the model and estimation procedure allows assessment of what the system failure intensity will be at the start of Phase II after implementation of the delayed fixes. Denoting this failure intensity by *r(T)*, where *T* denotes the duration of Phase I, the ACPM assessment of *r(T)* is based on: (1) the A and B-mode failure data generated during Phase I test duration, *T*; and (2) assessments of the FEFs for the B-modes surfaced during Phase I. Since the assessments of the FEFs are often largely based on engineering judgment, the resulting assessment, $\hat{r}(T)$, of the system failure intensity after corrective action implementations is called a reliability projection (as opposed to a demonstrated assessment, which would be based solely on test data).

The ACPM and estimation procedure was motivated by the desire to replace the widely used "adjustment procedure." The adjustment procedure assesses *r(T)* based on reducing the number of failures, N_i , due to B-mode *i* during Phase I to $\left(1 - d_i^*\right)N_i$, where d_i^* is the assessment of the realized FEF for mode i, denoted by d_i. Note $(1 - d_i^*)N_i$ is an assessment of the expected number of failures due to B-mode *i* that would occur in a follow-on test of the same duration as Phase I. The adjustment procedure assesses *r(T)* by $\hat{r}_{adj}(T)$, where

$$\hat{r}_{adj}(T) \;=\; \frac{N_A}{T} + \frac{\sum\limits_{i\in obs}\left(1 - d_i^*\right)N_i}{T}$$

In this equation, *obs* is the index set for all the B-modes that occur during Test Phase I.

6.3.2.6 Methodology.

The ACPM assesses the value of the system failure intensity, *r(T)*, after implementation of the Phase I delayed fixes. This assessment is taken to be an estimate of the expected value of *r(T)*, i.e., an estimate of $\rho_C(T) = E(r(T))$. Crow proposed an approximation for the ACPM approach which can be expressed as

$$\rho_C(T) = \lambda_A + \sum_{i=1}^{K}(1 - d_i)\lambda_i + \mu\_d \cdot h_C(T)$$

where the first two terms are equivalent to $\hat{r}_{adj}$ (T) above.

In this expression, λ_A denotes the assumed constant A-mode failure rate, λ_i denotes the initial B-mode failure rate for mode *i*, μ_d is the assumed common mean of all the FEFs when considered as random variables for the *K* B-modes, and d_i denotes the realized value of the achieved FEF for mode *i* if mode *i* is surfaced. The function $h_C(T)$ represents the rate of occurrence of new B-modes at the end of the test phase.

Based on an empirical study, Crow states that the number of distinct B-modes surfaced over a test period $[0,t]$ can often be approximated by a power function of the form

$$\mu_c(t) \;=\; \lambda\,t^{\beta} \qquad for \quad \lambda,\beta > 0$$

This function is interpreted as the expected number of distinct B-modes surfaced during the test interval $[0,t]$. More specifically, it assumes the number of distinct B-modes occurring over $[0,t]$ is governed by a NHPP with $\mu_c(t)$ as the mean value function. Thus

$$h_c(t) \;=\; \frac{d\,\mu_c(t)}{d\,t} \;=\; \lambda\,\beta\,t^{\beta-1}$$

represents the expected rate at which new B-modes are occurring at test time t.

For estimating $\rho_c(T)$, the first term in the expression λ_A is constant over $[0,T]$ and is estimated by

$$\hat{\lambda}_A \;=\; \frac{N_A}{T}$$

where N_A is the number of A-mode failures over $[0,T]$.

Based on the assumption that all fixes are delayed until Phase I is complete, the failure rate for B-mode i remains constant over $[0,T]$ and is estimated by

$$\hat{\lambda}_i \;=\; \frac{N_i}{T} \qquad (i=1,\cdots,K)$$

where N_i denotes the number of failures during $[0,T]$ attributable to B-mode i. Note

$$E(\hat{\lambda}_i) \;=\; \frac{E(N_i)}{T} \;=\; \frac{\lambda_i\,T}{T} \;=\; \lambda_i$$

The next step to complete the assessment of the expected system failure intensity after incorporation of delayed fixes is to address the rate of occurrence of new B-modes at T, $h_c(T) = \lambda\beta T^{\beta-1}$. The data required to estimate λ and β are: (1) the number of distinct B-modes, m, that occur during $[0,T]$; and (2) the B-mode first occurrence times $0 < t_1 \le t_2 \le \cdots \le t_m \le T$.

The maximum likelihood estimates of λ and β, denoted by $\hat{\lambda}$ and $\hat{\beta}$ respectively, satisfy the following equations:

$$\hat{\lambda}\,T^{\hat{\beta}} \;=\; m$$

$$\hat{\beta} \;=\; \frac{m}{\displaystyle\sum_{i=1}^{m}\ln\!\left(\frac{T}{t_i}\right)}$$

Solving for $\hat{\lambda}$ in the above, the estimate for $h_c(T)$ can be written in terms of m and $\hat{\beta}$ as follows:

$$\hat{h}_c(T) \;=\; \hat{\lambda}\,\hat{\beta}\,T^{\hat{\beta}-1} \;=\; \left(\frac{m}{T^{\hat{\beta}}}\right)\hat{\beta}\,T^{\hat{\beta}-1} \;=\; \frac{m\,\hat{\beta}}{T}$$

Crow notes that conditioned on the observed number of distinct B-modes, m, i.e. $M(T) \;=\; m$, the estimator

$$\bar{\beta}_m = \left(\frac{m-1}{m}\right)\hat{\beta} \qquad m \ge 2$$

is an unbiased estimator of β, i.e.

$$E(\bar{\beta}_m) = \beta$$

Thus estimating $h_c(T) = \lambda\beta T^{\beta-1}$ by using $\bar{\beta}_m$ leads to the estimate

$$\bar{h}_c(T) = \frac{m\bar{\beta}_m}{T}$$

To complete the assessment of the system failure intensity, it is necessary to assess the ACPM expected system failure intensity $\rho_c(T)$

$$\rho_c(T) = \lambda_A + \sum_{i=1}^{K}(1-d_i)\lambda_i + \mu_d\,h_c(T)$$

Based on $\hat{\beta}$ and since $N_i = 0$ for $i \notin obs$

$$\hat{\rho}_c(T) = \frac{1}{T}\left\{N_A + \sum_{i\in obs}(1-d_i^*)N_i + \hat{\beta}\sum_{i\in obs}d_i^*\right\}$$

For $\rho_c(T)$ based on $\bar{\beta}_m$ (provided $m \ge 2$):

$$\bar{\rho}_c(T) = \frac{1}{T}\left\{N_A + \sum_{i\in obs}(1-d_i^*)N_i + \bar{\beta}_m\sum_{i\in obs}d_i^*\right\}$$

Note both estimates of $\rho_c(T)$ are of the form

$$Estimate\ \rho_c(T) = \frac{1}{T}\left\{N^* + (estimate\ \beta)\sum_{i\in obs}d_i^*\right\}$$

where N^* is the "adjusted" number of failures over $[0,T]$. Recall the historically used adjustment procedure assessment for the system failure intensity, after incorporation of delayed fixes, is

given by $\hat{r}_{adj}(T) = \dfrac{N^*}{T}$. Also recall $\bar{\beta}_m = \left(\dfrac{m-1}{m}\right)\hat{\beta} < \hat{\beta}$. Thus

$$\hat{r}_{adj}(T) < \bar{\rho}_c(T) < \hat{\rho}_c(T)$$

Also of interest is an assessment of the reciprocal of $\rho_c(T)$, i.e. $M_c(T) = \{\rho_c(T)\}^{-1}$. For $M_c(T)$,

$$\hat{M}_c(T) < \bar{M}_c(T) < \{\hat{r}_{adj}(T)\}^{-1}$$

6.3.2.7 Reliability Growth Potential.
The expected system failure intensity after incorporation of the delayed fixes decreases to a limiting value as $T \to \infty$. This value is termed the failure intensity growth potential and is denoted by ρ_{GP}. It is given by $\rho_{GP} = \lim_{T\to\infty}\rho_C(T) = \lambda_A + \sum_{i=1}^{K}(1-d_i)\lambda_i$

Its reciprocal is referred to as the MTBF growth potential, which represents a theoretical upper limit on the system MTBF. This limit corresponds to the MTBF that would result if all B-mode failures were surfaced and corrected with specified FEFs. Note that ρ_{GP} is estimated by

$$\hat{\rho}_{GP} \quad = \quad \frac{1}{T}\left(N_A + \sum_{i \in obs}\left(1 - d_i^*\right)N_i\right)$$

If the reciprocal $\left(\hat{\rho}_{GP}\right)^{-1}$ lies below the goal MTBF, then this may indicate that achieving the goal is high risk.

6.3.2.8 ACPM Example.
The following example illustrates application of the ACPM. Data was generated by a computer simulation with $\lambda_A = 0.02$, $\lambda_B = 0.1$, $K = 100$ and the d_i's distributed according to a beta distribution with mean 0.7. The simulation portrayed a system tested for $T = 400$ hours. The simulation generated $N = 42$ failures, with $N_A = 10$ and $N_B = 32$. The 32 B-mode failures were due to $M=16$ distinct B-modes.

The B-modes are labeled by the index i. The first occurrence time for mode i is t_i and
$0 < t_1 < t_2 < \cdots < t_{16} < T = 400$.

For each B-mode i, Column 2 in TABLE XIV lists the time of first occurrence, followed by the times of subsequent occurrences (if any). Column 3 lists N_i, the total number of occurrences of B-mode i during the test period. Column 4 contains the assessed FEFs for each of the observed B-modes. Column 5 has the assessed expected number of type i B-mode failures that would occur in $T = 400$ hours after implementation of the corrective action. Finally, the last column contains the base e logarithms of the B-mode first occurrence times. These are used to calculate $\hat{\beta}$.

TABLE XIV. ACPM example data.

B-mode	Failure Times (hrs)	N_i	d_i^*	$(1-d_i^*)N_i$	$\ln t_i$
1	15.04, 254.99	2	.67	.66	2.7107
2	25.26, 120.89, 366.27	3	.72	.84	3.2292
3	47.46, 350.2	2	.77	.46	3.8599
4	53.96, 315.42	2	.77	.46	3.9882
5	56.42, 72.09, 339.97	3	.87	.39	4.0328
6	99.57, 274.71	2	.92	.16	4.6009
7	100.31	1	.50	.50	4.6083
8	111.99, 263.47, 373.03	3	.85	.45	4.7184
9	125.48, 164.66, 303.98	3	.89	.33	4.8321
10	133.43, 177.38, 324.95, 364.63	4	.74	1.04	4.8936
11	192.66	1	.70	.30	5.2609
12	249.15, 324.47	2	.63	.74	5.5181
13	285.01	1	.64	.36	5.6525
14	379.43	1	.72	.28	5.9387
15	388.97	1	.69	.31	5.9635
16	395.25	1	.46	.54	5.9795
Totals		32	11.54	7.82	75.7873

From Table XIV above, the adjustment procedure estimate of $r(T) = r(400)$ is

$$\hat{r}_{adj}(400) = \left(\frac{1}{400}\right)\left(N_A + \sum_{i=1}^{16}(1-d_i^*)N_i\right)$$

$$= \frac{10+7.82}{400} = 0.04455$$

Thus the adjustment procedure estimate of the system MTBF is

$$\{\hat{r}_{adj}(400)\}^{-1} = \frac{400}{17.82} = 22.45$$

As shown in the equation below, the adjustment procedure estimate of system failure intensity after implementation of the fixes is simply $\hat{\rho}_{GP}$, the estimated growth potential failure intensity. Thus

$$\hat{\rho}_{GP} = \hat{r}_{adj}(400) = 0.04455$$

Also, the estimate of the system MTBF growth potential is

$$\hat{\rho}_{GP}^{-1} = \{\hat{r}_{adj}(400)\}^{-1} = 22.45$$

To obtain an estimate with less bias of the system's failure intensity and corresponding MTBF at T=400 hours, after incorporation of fixes to the 16 surfaced B-modes, the ACPM estimation equation is used. This projection is given by

107

$$\hat{\rho}_c(400) = \hat{\rho}_{GP} + \left(\frac{\hat{\beta}}{400}\right)\sum_{i\in obs} d_i^* = 0.04455 + \left(\frac{\hat{\beta}}{400}\right)(11.54)$$

The MLE $\hat{\beta}$ is

$$\hat{\beta} = \frac{m}{\sum\limits_{i=1}^{m}\ln\left(\frac{T}{t_i}\right)} = \frac{m}{m\ln T - \sum\limits_{i=1}^{m}\ln t_i}$$

$$= \frac{16}{16\ln 400 - 75.7873} = 0.7970$$

Thus, the ACPM projection for the system failure intensity, based on $\hat{\beta}$, is

$$\hat{\rho}_c(400) = 0.04455 + \left(\frac{0.7970}{400}\right)(11.54)$$

$$= 0.06754$$

The corresponding MTBF projection is

$$\{\hat{\rho}_c(400)\}^{-1} = 14.81$$

A nearly unbiased assessment of the system failure intensity, for $d_i^* = d_i$, can be obtained by using $\bar{\beta}_m$ instead of $\hat{\beta}$, thus

$$\bar{\beta}_m = \left(\frac{m-1}{m}\right)\hat{\beta} = \left(\frac{15}{16}\right)(0.7970) = 0.7472$$

and the projected system failure intensity based on $\bar{\beta}_m$ is

$$\bar{\rho}_c(400) = \hat{\rho}_{GP} + \frac{\bar{\beta}_m}{T}\sum_{i\in obs} d_i^*$$

$$= 0.04455 + \left(\frac{0.7472}{400}\right)(11.54)$$

$$= 0.06611$$

The corresponding MTBF projection is

$$\{\bar{\rho}_c(400)\}^{-1} = 15.13$$

It is well known that an unbiased estimator for a model parameter ρ is typically biased for estimating the reciprocal of ρ. In particular, the fact that $\bar{\rho}_c(T)$ which uses $\bar{\beta}_m$ is almost unbiased for estimating of $\rho_C(T)$ does not imply that $M_C(T) = \{\rho_C(T)\}^{-1}$ should be estimated by $\bar{M}_C(T)$. In fact the discussion in Section 4.3.5 of the (Broemm, Ellner and Woodworth Sep 2000) suggests that $\widehat{M}_C(T)$ based on $\hat{\beta}$ is a better estimator of $M_C(T)$ than $\bar{M}_C(T)$. Thus, in this example it is recommended that the projected system failure intensity is assessed by $\bar{\rho}_c(400) = 0.06611$, and the projected system MTBF by $\{\hat{\rho}_c(400)\}^{-1} = 14.81$.

6.3.2.9 Goodness-of-Fit for ACPM based on B-Mode First Occurrence Times.
To visually assess whether the B-mode first occurrence time pattern conforms to the ACPM mean value function $\mu C(t)$, it is useful to plot the observed number of unique B-modes that

108

occur on or before each of the B-mode first occurrence times. This realized pattern should be visually compared to the graph of $\hat{\mu}_C(t)$ versus t where $\hat{\mu}_C(t) = \hat{\lambda}\, t^{\hat{B}}$.

Additionally, the fit of $\hat{\mu}_C(t)$ to the observed B-mode first occurrence pattern can be statistically assessed via the Cramér-von-Mises statistic considered in Section 6.2.2.8.3 for the tracking model. For application of this statistic C_F (given in 6.2.2.8.3) to the ACPM simply do the following: (1) Interpret F in the formula as the number of distinct B-modes that occur over the test phase duration T; (2) Interpret X_i as the first occurrence time of B-mode i where $0 < X_1 \le X_2 \le \dots \le X_F \le T$; and (3) Calculate $\bar{B}$ utilized in the formula for C_F by setting $\bar{B} = \left(\frac{F-1}{F}\right)\hat{B}$. In this expression, $\hat{B}$ is obtained from the formula given in Section 6.3.2.6 with $m=F$ and $t_i = X_i$. Then, as in Section 6.2.2.8.3, reject the null hypothesis that $\mu_C(t)$ represents the observed B-mode first occurrence data if the statistic C_F exceeds the critical value for a chosen significance level x.

6.3.2.10 ACPM based for Grouped Data.

At times it is desirable to apply the ACPM to grouped (interval) data. For such an application, the test phase duration T is subdivided into s disjoint test intervals $(z_{i-1}, z_i]$ where $0 = z_0 < z_1 < z_2 < \dots z_s = T$. The ith test interval is $(z_{i-1}, z_i]$ for i=1,….,s. It is required that s≥3. The data recorded for each test interval is the number of distinct B-modes, m_i, that occur in the ith test interval and the test interval duration $z_i - z_{i-1}$.

Such data naturally occurs if the achieved test duration and number of distinct B-modes that occur are collected on a calendar basis, e.g. weekly, over a test period associated with one test phase of duration T. The ACPM assumptions and formulas remain as in Section 6.3.2. However, the estimation procedure for the model parameters λ and β would change. The MLEs for λ and β can be obtained by numerically solving the equation for $\hat{B}$ given in Section 6.2.3.1.1. This equation when applied to the grouped data application of the ACPM should be utilized with K=s (the number of test intervals), $F_i = m_i$, and t_i equal to z_i (the right end point of the ith test interval). Note $\hat{\lambda} = \frac{m}{T^{\hat{B}}}$ where $m = \sum_{i=1}^{s} m_i$. The MLE $\hat{B}$ should always be utilized for the grouped data application of the ACPM and never $\bar{B}$. For this case $\bar{B} = \left(\frac{m}{m-1}\right)\hat{B}$ is <u>not</u> an unbiased estimator of β.

An important instance that may call for utilizing interval data is when there are one or more key system stressors that induce failure modes. Frequently one is attempting to measure system reliability with respect to the mean mission profile (MMP) specified in the OMS/MP. The OMS/MP may identify and quantify key system stressors over several mission profiles and the MMP. For example, the MMP for a tank may specify the duration of a mission measured in miles and mission hours. It may also specify the number of rounds fired from the cannon over the mission time or mission miles. This specified rate of firing is often a key failure mode driver. The MMP may further specify the portions of the mission miles associated with traversing primary roads, secondary roads, and cross-country terrain along with descriptions of each of these categories. The number of mission miles specified per mission hour and the portion of these miles to be realized in each of these categories for the MMP is also frequently a key stressor. The achieved reliability will strongly be influenced by such key failure mode stressors.

In the presence of such stressors, it is often useful to utilize interval test data where every key stressor is adequately represented in each of the test intervals.

For ease of exposition the test duration will be measured in terms of mission hours. Mission miles could be an appropriate alternative measure in some instances. As before, let $0 = z_0 < z_1 < z_2 < ... \ z_s = T$ define the test intervals $(z_{i-1}, z_i]$ for i=1,....,s. Consider the two key stressors discussed above. These intervals will be termed balance test intervals with regard to these two stressors provided the following hold:
(1) The number of rounds that one attempts to fire in each interval closely matches the number of rounds that the MMP implies would be fired for the tested mission hours in the interval; and
(2) The number of mission miles traversed and their allocation to the three categories discussed above in each test interval closely match the mission miles and the category proportions implied by the MMP for the amount of mission test hours accomplished in the test interval. The use of such balanced interval test data with respect to all the identified key stressors in the MMP can significantly contribute to more accurate assessments of projected system reliability associated with the MMP.

If an assessment of the projected reliability for an individual Mission Profile (MP) specified in the OMS/MP is desired, the above considerations still apply but with respect to the identified and quantitatively specified key stressors addressed in the OMS/MP for the particular MP of interest.

The concept of utilizing balanced test intervals also applies to tracking curve model assessments. The interval data for the tracking curve models (also referred to as grouped data) is discussed in Section 6.2.3.

A useful reference for further details regarding balanced test intervals is (Crow 2008).

6.3.2.11 Goodness-of-Fit for ACPM Applied to Interval Data.
For interval (grouped) data over a single test phase discussed in Section 6.3.2.10, a chi-squared goodness-of-fit test can be conducted to test the null hypothesis H_0 that the ACPM adequately represents the data. For this test, the model estimate of the expected number of delayed B-modes first discovered in the ith test interval $(z_{i-1}, z_i]$ is $\hat{E}_i \equiv \hat{\mu}_C(z_i) - \hat{\mu}_C(z_{i-1}) = \hat{\lambda}\left(z_i{}^{\hat{B}} - z_{i-1}{}^{\hat{B}}\right)$

Where necessary, combine adjacent test intervals so that after combination, each of the remaining intervals has a model estimate of at least five for the delayed B-modes first discovered in the interval. Let K_R denote the number of such intervals after combining where necessary and m_i denote the observed number of delayed B-modes first discovered in the resulting interval i for i=1,....,K_R. To utilize this goodness-of-fit test, K_R must be at least three.

Recalculate the $\hat{E}_i$ for the new intervals. Then the test statistic $x^2 = \sum_{i=1}^{K_R} \frac{(m_i - \hat{E}_i)^2}{\hat{E}_i}$ is approximately distributed as a chi-square random variable with K_R-2 degrees of freedom. The null hypothesis is rejected if the x^2 statistic exceeds the critical value for a chosen significance level. Critical values for this statistic can be found in tables of the chi-square distribution.

One should also construct a plot of the cumulative observed number of delayed B-modes versus the model estimate of this quantity at the ends of each test interval. Note the model estimate is given by $\hat{\mu}_C(z_i)$ for test interval $(z_{i-1}, z_i]$ for i=1,….,s. Also, one can plot on a log-log scale the model estimate for the cumulative number of delayed B-modes that occur by z_i divided by z_i versus z_i for i=1,….,s. Under the null hypothesis H_0 the points $(z_i, \frac{\hat{\mu}_C(z_i)}{z_i})$ for i=1,….,s should approximately lie on a line with slope $-(1-\hat{B})$ when plotted using a log-log scale.

6.3.3 Crow Extended Reliability Projection Model.

6.3.3.1 Introduction.
The Crow Extended Reliability Projection Model for a test-fix-find-test management strategy was developed by Crow (Crow 2004) to address the case where some corrective actions are incorporated during test and some corrective actions are delayed and incorporated at the end of the test. This model extends the AMSAA RGTMC for test-fix-test data and the ACPM for test-find-test data. That is, these other two models are special cases of the Crow Extended Reliability Projection Model.

6.3.3.2 Purpose.
The purpose of the Crow Extended Reliability Projection Model is to estimate the system reliability at the beginning of a follow-on test phase by taking into consideration the reliability improvement from fixes incorporated during the test phase and the delayed fixes incorporated at the conclusion of the test phase.

6.3.3.3 Assumptions.
The Crow Extended Reliability Projection Model presented in (Crow 2004) pertains to test data from a single test phase. Test duration is measured on a continuous scale such as time or miles. The model subdivides the B-modes into two categories – BC-modes (fixes incorporated during the test phase) and BD-modes (fixes delayed until the end of the test phase). The model assumes the following:
(i) A-modes and B-modes occur independently and each such occurrence results in a system failure; (ii) the number of failures that occur by test duration t, $N(t)$, is a NHPP with power law mean value function; (iii) the number of distinct BD-modes that occur by t, $M_{BD}(t)$, is a NHPP with power law mean value function (typically having different parameter values than the power law governing $N(t)$); (iv) no new modes are introduced by the corrective actions; (v) the number of BD failures that occur by t, $N_{BD}(t)$, is a HPP; and (vi) the number of A failures that occur by t, $N_A(t)$, is a HPP. The constant failure intensities for the assumed HPPs $N_{BD}(t)$ and $N_A(t)$ are denoted by λ_{BD} and λ_A, respectively.

6.3.3.4 Limitations.
The Crow Extended Reliability Projection Model does not explicitly use BC-mode FEFs. Utilizing justified FEF values based on solid root cause analysis and credible mitigation

111

solutions can enhance the accuracy of MTBF projections. However, the use of FEFs not based on such analysis and mitigation solutions can significantly degrade assessment accuracy.

The use of the power law to model E(N(t)) can cause difficulties. Note assuming $E(N(t))=\lambda t^B$ for some $\lambda, B>0$ implies the system failure intensity $\lambda_S(t)$ is given by $\lambda_S(t)= \lambda B t^{B-1}$. For growth during a test phase, $0<B<1$. Thus, for growth, $\lambda_S(t)$ is a strictly decreasing function such that $\lim_{t \to \infty} \lambda_S(t) = 0$. This can be inconsistent with the assumption that $\lambda_A + \lambda_{BD}$ is a positive constant failure intensity that contributes to $\lambda_S(t)$. If $\lambda_A + \lambda_{BD}$ is a small enough portion of $\lambda_S(T)$ (where T denotes the test phase duration), then this potential inconsistency may not distort the assessment of the projected MTBF (after BC and BD-mode mitigation) significantly. However, clearly problems can be expected if $\hat{\lambda}_A + \hat{\lambda}_{BD} = \frac{N_A + N_{BD}}{T}$ exceeds the tracking assessment $\hat{\lambda}_S(T)$ (where N_A denotes the number of A-mode failures and N_{BD} denotes the number of BD-mode failures, respectively, over the test phase).

6.3.3.5 Benefits.
The benefits associated with the Crow Extended Reliability Projection Model are the same as those for ACPM. Added capability includes pre-emptive fixes at time *T* for failure modes that have not experienced a failure. It is assumed that for these failure modes, an estimate of the failure rate prior to mitigation is already available, either by analysis, analogy, or other test. Details may be found in (Crow 2004).

6.3.3.6 Methodology.
In order to provide the assessment and management metric structure for corrective actions incorporated during and after a test, two types of B-modes are defined: BC-modes, which are corrected during the test phase; and BD-modes, which are delayed until the end of the test phase. A-modes are still defined as failure modes that will not receive a corrective action. These mode classifications define the management strategy and can be changed. Note that the AMSAA RGTMC does not utilize this failure mode designation, which is a practical aspect of the Crow Extended Reliability Projection Model.

During the test phase, only the corrective actions for BC-modes contribute to reliability growth, not the A-modes or the BD-modes. At the end of the test phase, corrective actions for the BD-modes are incorporated, and the reliability increases further, typically as a jump. Estimating this increased reliability with test-fix-find-test data is the objective of the Crow Extended Reliability Projection Model.

For the Crow Extended Reliability Projection Model, the achieved MTBF (before delayed fixes) due to BC corrective actions should be exactly the same as the achieved failure intensity, λ_{CA}, estimated by the AMSAA RGTMC.

Let K denote the number of BD-modes residing in the system, λ_{BD} be the constant failure intensity for the BD-modes, and let *h(T|BD)* be the first occurrence function for the BD-modes (the same as h(T) in the ACPM applied to the BD-modes).

The Crow Extended Reliability Projection Model projected failure intensity is

$$\lambda_{EM} = \lambda_{CA} - \lambda_{BD} + \sum_{i=1}^{K}(1 - d_i)\lambda_i + \mu_d h(T|BD)$$

and the projected MTBF is $M_{EM} = 1/\lambda_{EM}$. This is the MTBF after the incorporation of fixes for the BD-modes.

The failure intensity $\lambda_{CA} = \lambda_S(T)$ is estimated via the tracking model and λ_{BD} is estimated by $\frac{N_{BD}}{T}$, where T is the duration of the test phase and N_{BD} is the number of BD-failures observed over the test phase. Also, for each BD-mode initial failure rate λ_i, $\hat{\lambda}_i = \frac{N_i}{T}$ where N_i is the number of failures associated with mode i observed over the test phase. The average FEF μ_d is assessed as the arithmetic average of the assessed FEFs d_i for the observed BD-modes. The expected rate of occurrence of BD-modes at the end of the test phase, h(T/BD), is estimated from the BD-mode first occurrence times in the test phase using the ACPM procedure. More explicitly, $\hat{h}(T/BD) = \frac{(m_{BD})\hat{\beta}}{T}$ and $\hat{\beta} = \frac{m_{BD}}{\sum_{i=1}^{m_{BD}} \ln \frac{T}{t_i}}$ where m_{BD} denotes the number of distinct BD-modes observed over the test phase and t_i denotes the cumulative first occurrence time for BD-mode i as measured from the start of the test phase. For multiple units under test, t_i would be the sum of the amount of test time accrued by each unit at the time of the first occurrence of BD-mode i on one of the units.

Interval failure mode data using balanced test intervals over the test phase can also be utilized as described in Section 6.3.2.10. The data includes the test duration of each test interval and the number of distinct BD-modes that first occur during the test phase in a test interval. The estimate of λ_{CA} from the tracking model can also be based on interval failure data as described in Section 6.2.3

If it is assumed that no corrective actions are incorporated into the system during the test (no BC-modes), then this is equivalent to assuming that $\beta = 1$ for λ_{CA} and λ_{CA} is estimated by $\hat{\lambda}_{CA} = \hat{\lambda}_A + \hat{\lambda}_B$. In general, the assumption of a constant failure intensity ($\beta = 1$) can be assessed by a statistical test from the data.

In using the Crow Extended Reliability Projection Model, it is important that the classification of a B-mode, with respect to the BC and BD categories, not be dependent on when the mode occurs during the test phase. In some testing programs, modes that occur in the early portion of the test phase tend to have fixes implemented during the test and are thus classified as BC-modes, while those that occur later are not implemented until after the test phase and are thus classified as BD-modes. Under such conditions, the pattern of BD-mode first occurrence times will provide an inaccurate estimate of the failure intensity due to the unobserved BD-modes. This, in turn, would degrade the accuracy of the MTBF projection.

The Crow Extended Reliability Projection Model is illustrated using the data in TABLE XV, where 56 failures due to A, BC, and BD-modes are designated over a test time of $T = 400$ hours. The data in TABLE XV was utilized in (Crow 2004) to illustrate this model. In the following table, TABLE XVI, the BD-mode first occurrences, the frequency of the modes, and the failure

mode FEF assessments, d_i^*, are given. Note that the BD-mode data in TABLE XVI is the same B-mode data provided in TABLE XIV.

TABLE XV. Crow extended reliability projection model example data

Failure i	Time of Occurrence	Failure Mode	Failure i	Time of Occurrence	Failure Mode
1	0.7	BC1	29	192.7	BD11
2	3.7	BC1	30	213.0	A
3	13.2	BC1	31	244.8	A
4	15	BD1	32	249.0	BD12
5	17.6	BC2	33	250.8	A
6	25.3	BD2	34	260.1	BD1
7	47.5	BD3	35	263.5	BD8
8	54.0	BD4	36	273.1	A
9	54.5	BC3	37	274.7	BD6
10	56.4	BD5	38	282.8	BC11
11	63.6	A	39	285.0	BD13
12	72.2	BD5	40	304.0	BD9
13	99.2	BC4	41	315.4	BD4
14	99.6	BD6	42	317.1	A
15	100.3	BD7	43	320.6	A
16	102.5	A	44	324.5	BD12
17	112.0	BD8	45	324.9	BD10
18	112.2	BC5	46	342.0	BD5
19	120.9	BD2	47	350.2	BD3
20	121.9	BC6	48	355.2	BC12
21	125.5	BD9	49	364.6	BD10
22	133.4	BD10	50	364.9	A
23	151.0	BC7	51	366.3	BD2
24	163.0	BC8	52	373.0	BD8
25	164.7	BD9	53	379.4	BD14
26	174.5	BC9	54	389.0	BD15
27	177.4	BD10	55	394.9	A
28	191.6	BC10	56	395.2	BD16

Note that in the failure mode designation in TABLE XV, BC-modes are entirely different than BD-modes. For example, failure mode BC-1 is an entirely different failure mode from BD-1, although both have the similar sub designation "1."

All occurrences of a given B-failure mode have the same letter designators (BC or BD) and the same numerical sub-designation. For example, the BD-failure mode 2 in TABLE XVI has designator BD2 in TABLE XV and appears three times in this table (since $N_2=3$ in TABLE XVI).

114

Failure occurrence times that correspond to an A-failure mode (which may not be identifiable in test) are simply given the designator "A" in TABLE XV.

TABLE XVI. BD Failure mode data and FEFs.

BD Mode j	N_j	First Occurrence	Assessed d^*_j
1	2	15.0	.67
2	3	25.3	.72
3	2	47.5	.77
4	2	54.0	.77
5	3	56.4	.87
6	2	99.6	.92
7	1	100.3	.50
8	3	112.0	.85
9	3	125.5	.89
10	4	133.4	.74
11	1	192.7	.70
12	2	249.2	.63
13	1	285.0	.64
14	1	379.4	.72
15	1	389.0	.69
16	1	395.2	.46
Total	32		11.54

6.3.3.7 AMSAA RGTMC Example Using TABLE XV.

Suppose a development testing program begins at time 0 and is conducted until time *T*. Let *N* be the total number of failures recorded and let $0 < X_1, < X_2, \cdots, X_N < T$ denote the *N* successive failure times on a cumulative time scale. Assume that the AMSAA RGTMC NHPP assumption applies to this set of data. Under the AMSAA RGTMC, the MLEs for λ and β are

$$\widehat{\lambda} = \frac{N}{T^{\widehat{\beta}}} \ , \qquad \widehat{\beta} = \frac{N}{\sum_{i=1}^{N} \ln\frac{T}{X_i}}.$$

Applying these equations to the 56 failure times in TABLE XV results in the following estimates:

$\widehat{\lambda} = 0.2171$ and $\widehat{\beta} = 0.9268$. Thus, the estimated growth rate is 1-$\widehat{\beta}$=0.0732.

While growth is small, hypothesis testing indicates it is significantly different from 0. Thus growth is occurring and the failure intensity is decreasing.

The achieved or demonstrated failure intensity and MTBF at *T* = 400 are estimated by

115

$\hat{\lambda}_{CA} = \hat{\lambda}\hat{\beta}T^{\hat{\beta}-1} = 0.1298$, and $\hat{M}_{CA} = [\hat{\lambda}_{CA}]^{-1} = 7.71$.

This indicates that the "baseline" estimate of reliability at $T = 400$ hours is 7.71 using the AMSAA RGTMC. It is important to note that the AMSAA RGTMC does not assume that all failures in the data set receive a corrective action. Based on the management strategy, some failures may receive a corrective action and some may not.

6.3.3.8 Crow Extended Reliability Projection Model Example Using TABLE XV.
For this example, assume that all the failure times X_j are known.

The first term, $\hat{\lambda}_{CA}$, uses all failure time data in TABLE XV and is calculated in the previous example using the AMSAA RGTMC. This gives
$$\hat{\lambda}_{CA} = 0.1298.$$

For the remaining terms, the FEFs and the BD-mode data given in TABLE XVI are used. Note that the BD-mode data in TABLE XVI is the same as the B-mode data in TABLE XIV, so the calculations are the same as previously shown. That is,

$$m_{BD} = 16, T = 400, \hat{\lambda}_{BD} = 0.0800, \text{ and the assessment of } \mu_d \text{ is } \mu_d^* = 0.72.$$

Also,

$$\sum_{i=1}^{m_{BD}} (1 - d_i^*) \frac{N_i}{T} = 0.0196, \text{ and}$$

$$\hat{h}(T|BD) = \frac{m_{BD}\hat{\beta}}{T} = (16)(0.7970)/400 = 0.0319.$$

This gives $\mu_d^* \hat{h}(T|BD) = 0.0230$. Therefore,
$$\hat{\lambda}_{EM} = \hat{\lambda}_{CA} - \hat{\lambda}_{BD} + \sum_{i=1}^{m_{BD}} (1 - d_i^*) \frac{N_i}{T} + \mu_d^* \hat{h}\left(\frac{T}{BD}\right) = 0.1298 - 0.0800 + 0.0196 + 0.0230 = 0.0924.$$

The Crow Extended Reliability Projection Model's projected MTBF is $\hat{M}_{EM} = 10.82$. The achieved MTBF before the 16 delayed fixes (as calculated in the previous example using the AMSAA RGTMC is estimated by $\hat{M}_{CA} = 7.71$. Therefore, the MTBF grew to 7.71 as a result of corrective actions for BC-modes during the test and then jumped to 10.82 as a result of the delayed corrected actions for the BD-modes after the test.

As pointed out in Section 6.3.3.6 it is important that the classification of a B-mode, with respect to the BD and BC categories, not be dependent when the mode occurs during the test phase. In particular, the Crow-Extended Model requires the assumption that the number of BD-failures that occur by t, denoted by $N_{BD}(t)$, is a HPP. In the following, data from TABLE XV will be used to illustrate how this assumption can be statistically tested.

Let H_0 be the null hypothesis that $N_{BD}(t)$ is a HPP over the test phase. Under H_0, $N_{BD}(t)$ is a Poisson process with a power law mean value function where the power equals one. Thus under H_0: $E(N_{BD}(t)) = \lambda_{BD} t^{\eta_{BD}}$ where $\lambda_{BD} > 0$ and $n_{BD} = 1$.

116

A maximum likelihood estimate (MLE) of η_{BD} is $\hat{\eta}_{BD} = \frac{\eta_{BD}}{\sum_{i=1}^{n_{BD}} \ln\left(\frac{T}{x_{BD,i}}\right)}$ where η_{BD} equals the number of BD-failures over the test phase of duration T and $x_{BD,i}$ denotes the cumulative test phase time to the ith BD-failure.

Under the null hypothesis H_0, it can be shown that the statistic $\frac{2n_{BD}}{\hat{\eta}_{BD}}$ conditioned on $N_{BD}(T) = n_{BD}$ is distributed as a chi-square random variable with $2n_{BD}$ degrees of freedom, denoted by $x^2(2n_{BD})$. Thus for significance level x, reject H_0 if $\frac{2n_{BD}}{\hat{\eta}_{BD}} < x^2_{\frac{x}{2}}(2n_{BD})$ or $\frac{2n_{BD}}{\hat{\eta}_{BD}} > x^2_{1-\frac{x}{2}}(2n_{BD})$.

In the above, $X^2_{\frac{x}{2}}(2n_{BD})$ and $X^2_{1-\frac{x}{2}}(2n_{BD})$ denote the $\frac{x}{2}$ and $1-\frac{x}{2}$ percentiles, respectively, of $X^2(2n_{BD})$. From TABLE XV, $n_{BD}=32$ and thus chi-square tables or statistical software may be used to obtain the $\frac{x}{2}$ and $1-\frac{x}{2}$ percentiles of $X^2(64)$ for a selected significance level x. The results for this test of H_0 using the data of TABLE XV is provided in TABLE XVII below.

TABLE XVII. Results for test of $N_{BD}(t)$

	H_0: $N_{BD}(t)$ is a Homogenous Poisson Process (HPP)			
α	$X^2_{\alpha/2}(2n_{BD})$	$X^2_{1-\alpha/2}(2n_{BD})$	Test Statistic	Test Result
0.10	46.595	83.675	54.76	Do Not Reject Null Hypothesis
0.20	49.996	78.860	54.76	Do Not Reject Null Hypothesis
0.40	54.336	73.276	54.76	Do Not Reject Null Hypothesis

Note H_0 cannot be statistically rejected even at significance level $\alpha=0.40$. Thus, there is no strong statistical evidence against H_0.

In like manner, the null hypothesis H_0: $N_A(t)$ is a HPP can be tested, where $N_A(t)$ denotes the number of A-failures occurring in the test phase by t. In the test of H_0, n_{BD} is replaced by $n_A=N_A(T)$ and $x_{BD,i}$ is replaced by $x_{A,i}$, the cumulative time of occurrence of the ith A-failure in the test phase. Results for the test of this H_0 are displayed in TABLE XVIII.

TABLE XVIII. Results for test of $N_A(t)$

α	$X^2_{\alpha/2}(2n_A)$	$X^2_{1-\alpha/2}(2n_A)$	Test Statistic	Test Result
		H_0: $N_A(t)$ is a Homogenous Poisson Process (HPP)		
0.10	10.851	31.410	11.46	Do Not Reject Null Hypothesis
0.20	12.443	28.412	11.46	Reject Null Hypothesis

TABLE XVIII indicates there is no strong statistical evidence against H_0 in the sense one cannot reject H_0 at significance level α=0.10. However, representing $N_A(t)$ as a HPP is more problematic than for $N_{BD}(t)$ since H_0 in TABLE XVIII is rejected for α=0.20.

The Crow-Extended Model null hypothesis, H_0: $N_{BD}(t)$ is a HPP, can also be tested using interval data for the test phase. Let $0=z_0<z_1<\ldots.<z_p=T$ and let $f_{BD,j}$ denote the number of BD-failures in $(t_{j-1},t_j]$ for j=1,…,p. Under H_0 the expected number of BD-failures in $(t_{j-1},t_j]$ is $e_{BD,j}=(t_j-t_{j-1})\lambda_{BD}$ where λ_{BD} is the constant BD failure intensity for the HPP $N_{BD}(t)$. Note the model estimate of $e_{BD,j}$ under H_0 is $\hat{e}_{BD,j}=(z_j-z_{j-1})(\frac{n_{BD}}{T})$. Combine adjacent intervals, if necessary, to obtain $\hat{e}_{BD,j}\geq 5$ for each interval. Let $\acute{p}$ denote the number of intervals after combining where necessary. In the following, let $f_{BD,j}$ and $\hat{e}_{BD,j}$ respectively denote the number of observed BD-failures and the model estimate of the expected number of BD-failures for interval j after combining the original intervals where required. Also, assume $\acute{p} \geq 2$. Then $X^2=\sum_{j=1}^{\acute{p}}\frac{(f_{BD,j}-\hat{e}_{BD,j})^2}{\hat{e}_{BD,j}}$ is approximately distributed as a chi-square variate with $\acute{p}$-1 degrees of freedom. Thus reject H_0 at significance level α if $X^2 > X^2_{1-\alpha}(\acute{p}-1)$ where $X^2_{1-\alpha}(\acute{p}-1)$ is the $1-\alpha$ percentile of $X^2(\acute{p}-1)$. One can use interval data in a similar manner to test the null hypothesis H_0:$N_A(t)$ is a HPP.

6.3.3.9 Management and Maturity Metrics for the Crow Extended Reliability Projection Model.
In (Crow 2004), he includes a section of 33 management and maturity metrics which are useful for providing management and engineering insights for some practical situations. These are illustrated using the data of TABLES XV and XVI. The metrics are not presented here, but they may be found in this referenced paper.

6.3.4 AMSAA Maturity Projection Model (AMPM).

6.3.4.1 Purpose.
The purpose of AMPM is to provide an estimate of the projected reliability following the implementation of both delayed and non-delayed fixes. The model also provides estimates of the following important reliability growth metrics:
   a) B-mode initial failure intensity
   b) expected number of B-modes surfaced
   c) percent surfaced of the B-mode initial failure intensity surfaced
   d) rate of occurrence of new B-modes

6.3.4.2 Assumptions.
The Assumptions associated with AMPM include:
  a) Test duration is continuous;
  b) Failure modes independently occur and cause system failure;
  c) Failure modes can be classified as either A-modes or B-modes;
  d) Corrective actions are implemented prior to the time at which projections are made; and
  e) Initial B-mode failure rates can be modeled as a realization of a random sample from a gamma distribution.

6.3.4.3 Limitations.
The limitations associated with AMPM include:
  a) FEFs are often a subjective input; and
  b) Projection accuracy can be degraded via reclassification of A-modes to B-modes.

6.3.4.4 Benefits.
The benefits associated with AMPM include:
  a) Corrective actions can be implemented during test or can be delayed until the end of test;
  b) Reliability can be projected for future milestones;
  c) The model can project the impact of delayed corrective actions on system reliability;
  d) The projection takes into account the contribution to the system failure rate due to unobserved B-failure modes;
  e) In situations where there is an apparent steepness of cumulative number of B-modes versus cumulative test time over an early portion of testing after which this rate of occurrence slows, there is methodology to partition the early modes from the remaining modes. These early B-modes must be aggressively and effectively corrected.
  f) Additionally methodology exists to handle cases where there is an early "gap" or if there appears to be a difference in the average FEFs in early or start-up testing versus the remainder of testing (an apparent or visual difference in failure rate in the initial testing).

6.3.4.5 Overview of AMPM Approach.
AMPM addresses making reliability projections in several situations of interest.

The first situation of interest is where all fixes to B-modes are implemented at the end of the current test phase, Phase I, prior to commencing a follow-on test phase, Phase II. The projection goal is to assess the expected system failure intensity at the start of Phase II.

The second situation of interest is where the reliability of the unit under test has been maturing over Phase I due to implemented fixes during Phase I. This case includes the situations where:
  a) all surfaced B-modes in Phase I have fixes implemented within this test phase; or
  b) some of the surfaced B-modes are addressed by fixes within Phase I and the remainder are treated as delayed fixes and fixed at the conclusion of Phase I, prior to commencing Test Phase II.

The third situation of interest involves projecting the system failure intensity at a future program milestone, which may occur beyond the commencement of the follow-on test phase.

119

All three types of projections are based on the Phase I B-mode first occurrence times and whether the associated B-mode fix is implemented within the current test phase or delayed (but implemented prior to the projection time). Additionally, the projections are based on an average FEF with respect to all the potential B-modes, whether surfaced or not. However, this average FEF is assessed based on the surfaced B-modes. For AMPM, the set of surfaced B-modes would typically be a mixture of those addressed with fixes during the current test phase as well as those addressed beyond the current test phase.

For complex systems or subsystems, AMPM utilizes a NHPP with regard to the number of distinct B-modes that occur by test duration *t*. The mean value function for the expected number of B-modes, where not all fixes need be delayed, is given by:

$$(\lambda_B/\beta)\ln(1+\beta t)$$

In this expression, β is a positive constant and λ_B denotes the expected initial failure intensity due to all the B-modes. The constant β is not the β of the power law tracking model (i.e., RGTMC). Statistical tests for this mean value function can be based on a chi-square statistic (Section 6.3.4.9).

The associated pattern of B-mode first occurrence times is not dependent on the corrective action strategy, under the assumption that corrective actions are not inducing new B-modes to occur. Thus, the AMPM assessment procedure is not upset by jumps in reliability due to delayed groups of fixes.

AMPM can also be used to construct a useful reliability maturity metric. This metric is the fraction of the expected initial system B-mode failure intensity, λ_B, surfaced by test duration *t*. (i.e. the expected fraction of λ_B due to B-modes surfaced by *t)*.

6.3.4.6 Development of AMPM.
AMPM provides a procedure for assessing the system failure intensity, $r(t;\underline{\lambda})$, after fixes to all B-modes surfaced by test time *t* have been implemented. AMPM does not attempt to assess the expected system failure intensity, $\rho(t;\underline{\lambda})$, by estimating each λ_i. Instead, the AMPM approach is to view $(\lambda_1,\cdots,\lambda_K)$ as a realization of a random sample $\underline{\Lambda} = (\Lambda_1,\cdots,\Lambda_K)$ from the gamma random variable $\Gamma(\alpha,\beta)$. This allows one to utilize all the B-mode times to first occurrence observed during Test Phase I to estimate the gamma parameters α and β. Thus in place of directly assessing $\rho(t;\underline{\lambda})$, AMPM uses estimates of α and β to assess the expected value of $\rho(t;\underline{\Lambda})$ where

$$\rho(t;\underline{\Lambda}) = \lambda_A + \sum_{i=1}^{K}(1-d_i)\Lambda_i + \sum_{i=1}^{K}d_i\Lambda_i e^{-\Lambda_i t}$$

This assessed value is then taken as the AMPM assessment of the system failure intensity after fixes to all B-modes surfaced over [0,*t*] have been implemented. This approach does away with the need to estimate individual λ_i, which could be particularly difficult in the case where many fixes are implemented prior to the end of the period [0,*T*]. The expected value of $\rho(t;\underline{\Lambda})$ is

$$\rho(t) = \lambda_A + (1-\mu_d)\lambda_{B,K} + \mu_d h(t)$$

120

where $\lambda_{B,K} = E(\Lambda_i)$ and $h(t)=E(\sum_{i=1}^K \Lambda_i\, e^{-t\Lambda_i})$. Both expected values are with respect to Λ. In taking the expected values, it is assumed that d_i does not depend on the magnitude of λ_i.

This expression is more appealing when put in a slightly different form, by subtracting and adding the term *h(t)* on the right side of the equation:

$$\rho(t) \;=\; \lambda_A + (1-\mu_d)\{\lambda_{B,K} - h(t)\} + h(t)$$

This equation expresses the expected failure intensity once all the fixes have been implemented to the B-modes surfaced by test time $t \leq$ T. The first term, λ_A, is simply the assumed constant failure rate due to all the A-mode failures, estimated by N_A/T where N_A denotes the number of A-mode failures that occur during [0, *T*]. To consider the second failure intensity, first consider the final term, *h(t)*. It may be shown that *h(t)* is the expected rate of occurrence of new B-modes at time *t* averaged over random samples. *h(t)* can also be shown to represent the expected rate of occurrence of failures due to the B-modes that have not been surfaced by *t*. Then *h(t)* is the unconditional expected failure intensity due to the set of unsurfaced B-modes at time *t*. Thus, for the second term in the above equation, $\lambda_{B,K} - h(t)$ is the expected rate of occurrence of failures due to the B-modes that have been surfaced by *t*. If these surfaced failure modes are fixed with an average FEF of $\mu_{d,}$ then after mitigation, the expected residual rate of occurrence of failures due to these surfaced B-modes can be approximated by *(1-μ_d)($\lambda_{B,K}$-h(t))*.   The arithmetic average of the individual mode FEFs for the surfaced B-modes can be utilized as an assessment of μ_d. For the case where all fixes are delayed, one can simply estimate $\lambda_{B,K}$ by N_B/T where N_B denotes the number of B-mode failures that occur over the test interval [0, *T*].

6.3.4.7 Estimation Procedures for AMPM using cumulative time data.
This section specifies the procedures to estimate key AMPM parameters and reliability measures expressed in terms of these parameters. Estimation equations will be given for the finite *K* and NHPP variants of AMPM. The model parameter estimators are MLEs.

The parameter estimates are written in terms of the following data:
> m = number of distinct B-modes that occur over a test period of length *T*;

> $\underline{t} = (t_1, \cdots, t_m)$ where $0 < t_1 \leq t_2 \leq \cdots \leq t_m \leq T$ are the first occurrence times of the *m* observed B-modes; and

> n_A = number of A-mode failures that occur over test period *T*.

Holding the test data constant and letting $K \rightarrow \infty$ yields AMPM projection estimates that are appropriate for complex subsystems or systems that typically have many potential B-modes. The AMPM limit estimating equations can also be obtained from MLE equations for the NHPP associated with AMPM. This process has the mean value function

$\left(\frac{\lambda_B}{\beta}\right)(\ln(1 + \beta t))$, where λ_B, *B>0.*

Recall α_K, β_K are the gamma parameters for AMPM, where it is assumed the K initial B-mode failure rates are realized values of a random sample from a gamma random variable, $\Gamma(\alpha_K, \beta_K)$. The MLE for β_K is $\hat{\beta}_K$ where

$$K = \frac{\left(\sum\limits_{i=1}^{m}\ln\frac{1+\hat{\beta}_K T}{1+\hat{\beta}_K t_i}\right)\left(\sum\limits_{i=1}^{m}\frac{1}{1+\hat{\beta}_K t_i}\right) - \left(\frac{m\hat{\beta}_K}{1+\hat{\beta}_K T}\right)\sum\limits_{i=1}^{m}\frac{T-t_i}{1+\hat{\beta}_K t_i}}{\left\{\ln\left(1+\hat{\beta}_K T\right)\right\}\left(\sum\limits_{i=1}^{m}\frac{1}{1+\hat{\beta}_K t_i}\right) - \left(\frac{m\hat{\beta}_K}{1+\hat{\beta} T}\right)T}$$

The MLE for α_K is $\hat{\alpha}_K$, where $\hat{\alpha}_K$ can be easily obtained from $\hat{\beta}_K$ and either equation below. These equations are the maximum likelihood equations for α_K and β_K, respectively.

$$\left(\hat{\alpha}_K + 1\right)^{-1} = m^{-1}\left[K\ln\left(1+\hat{\beta}_K T\right) - \sum\limits_{i=1}^{m}\ln\left(\frac{1+\hat{\beta}_K T}{1+\hat{\beta}_K t_i}\right)\right]$$

$$\hat{\alpha}_K + 1 = \frac{\dfrac{m}{\hat{\beta}_K} - \sum\limits_{i=1}^{m}\dfrac{t_i}{1+\hat{\beta}_K t_i}}{\dfrac{(K-m)T}{1+\hat{\beta}_K T} + \sum\limits_{i=1}^{m}\dfrac{t_i}{1+\hat{\beta}_K t_i}}$$

Using $\left(\hat{\alpha}_K, \hat{\beta}_K, K\right)$ one can estimate all finite K AMPM quantities where the A-mode failure rate, λ_A, is estimated by $\hat{\lambda}_A = {n_A}/{T}$ and the average B-mode FEF, μ_d, is assessed as

$$\mu_d^* = \frac{1}{m}\sum\limits_{i\in obs}d_i^*$$

The limiting AMPM estimates for key projection model quantities are:

$$\hat{\lambda}_\infty = \frac{m\hat{\beta}_\infty}{\ln\left(1+\hat{\beta}_\infty T\right)} \qquad \text{Initial B-mode failure rate}$$

$$\hat{\mu}_\infty(t) = \left(\frac{\hat{\lambda}_{B,\infty}}{\hat{\beta}_\infty}\right)\ln\left(1+\hat{\beta}_\infty t\right) \qquad \text{Expected number of distinct B-modes}$$

$$\hat{h}_\infty(t) \;=\; \frac{\hat{\lambda}_{B,\infty}}{1+\hat{\beta}_\infty t} \qquad\qquad \text{Expected rate of occurrence of B-modes}$$

$$\hat{\rho}_{GP,\infty} \;=\; \hat{\lambda}_A + \left(1-\hat{\mu}_d^*\right)\hat{\lambda}_{B,\infty} \qquad\qquad \text{Failure intensity growth potential}$$

$$\hat{\rho}_\infty(t) \;=\; \hat{\rho}_{GP,\infty} + \hat{\mu}_d^* \, \hat{h}_\infty(t) \qquad\qquad \text{Failure intensity projection equation}$$

$$\hat{\theta}_\infty(t) \;=\; \frac{\hat{\beta}_\infty t}{1+\hat{\beta}_\infty t} \qquad\qquad \text{Fraction surfaced of initial B-mode failure intensity}$$

Note that

$$\hat{\mu}_\infty(T) \;=\; \left(\frac{\hat{\lambda}_{B,\infty}}{\hat{\beta}_\infty}\right)\ln\left(1+\hat{\beta}_\infty T\right) \;=\; m$$

This agrees with intuition in the sense that $\hat{\mu}_\infty(T)$ is an estimate of the expected number of distinct B-modes generated over the test period $[0,T]$, while m is the observed number of distinct B-modes that occur.

If one adopts the view that the "model of reality" for a system or subsystem is AMPM for a finite K, which is large but unknown, then one can consider the limiting AMPM projection estimates as approximations to the AMPM estimates that correspond to the "true" value of K. Over the projection range of $t \geq T$ values of practical interest, the limiting estimates should be good approximations for complex systems or subsystems. In this sense, knowing the "true" value of K is usually unimportant.

6.3.4.8 Estimation for AMPM using Interval Data.
One may wish to utilize interval data for the reasons discussed in Section 6.3.2.10. Let $0 = z_0 < z_1 < \ldots \leq z_s = T$ partition a test phase of interest into intervals. To estimate the AMPM parameters λ_B and β, for each interval $(z_{i-1}, z_i]$ one must record the interval test duration $z_i - z_{i-1}$ and the number of first occurrence B-modes that occur in the interval. The parameters λ_A and μ_d are assessed as in Section 6.3.4.7. The model equations are the same as in the previous section. Only the estimation formulas for λ_B and β change.

The interval data (also referred to as grouped data) estimation procedure for the AMPM will only be addressed for k=∞. For this case the number of unique B-modes that occur by test duration t, denoted by $M_B(t)$, can be viewed as a NHPP with mean value function $\mu(t) = E(M_B(t)) = (\frac{\lambda_B}{\beta})\ln(1+\beta t)$. In the above, $E(M_B(t))$ denotes the expected value of $M_B(t)$.

For interval $(z_{i-1} - z_i]$, the probability of observing m_i new B-modes in the interval is given by the Poisson probability $[e^{-\{\mu(z_i)-\mu(z_{i-1})\}}][\frac{\{\mu(z_i)-\mu(z_{i-1})\}^{m_i}}{m_i!}]$.

Thus the likelihood associated with the B-mode data vector $\vec{m} = (m_1, m_2,\ldots, m_s)$ is

$L(\vec{m}; \lambda_B, \beta)= \prod_{i=1}^{s}[e^{-\{\mu(z_i)-\mu(z_{i-1})\}}][\frac{\{\mu(z_i)-\mu(z_{i-1})\}^{m_i}}{m_i!}]$. The likelihood equations are

$\frac{\partial}{\partial \lambda_B}\ln\{L(\vec{m}; \lambda_B, \beta)\}|(\hat{\lambda}_B, \hat{\beta})=0$ and $\frac{\partial}{\partial \beta}\ln\{L(\vec{m}; \lambda_B, \beta)\}|(\hat{\lambda}_B, \hat{\beta})=0$. The above notation indicates that the partials of $\ln \{L(\vec{m}; \lambda_B, B)\}$ with respect to λ_B and β, respectively, when evaluated at $(\lambda_B, \beta) = (\hat{\lambda}_B, \hat{\beta})$ must equal zero.

These two likelihood equations for the MLEs of λ_B and β, denoted by $\hat{\lambda}_B$ and $\hat{\beta}$ respectively, are equivalent to the following two equations:

$$\hat{\lambda}_B = \frac{m\hat{\beta}}{\ln(1+\hat{\beta}T)} \text{ and } \frac{(1+\hat{\beta}T)\ln(1+\hat{\beta}T)}{T}\sum_{i=1}^{s}\{\frac{m_i}{\ln(1+\hat{\beta}z_i)-\ln(1+\hat{\beta}z_{i-1})}\}\{\frac{z_i}{(1+\hat{\beta}z_i)}-\frac{z_{i-1}}{(1+\hat{\beta}z_{i-1})}\} = m.$$

In the above equations, $m = \sum_{i=1}^{s} m_i$. $\hat{\beta}$ is numerically solved for in the second equation and used to obtain $\hat{\lambda}_B$ via the first equation. Note that the first equation can be written as $\frac{\hat{\lambda}_B}{\hat{\beta}}\ln(1 + \hat{\beta}T)=m$. This equation says the model estimate of the number of distinct B-modes, based on the MLEs $\hat{\lambda}_B$ and $\hat{\beta}$, equals the observed number of distinct B-modes surfaced over the test phase.

6.3.4.9 Statistical and Visual Goodness-of-Fit Procedures for AMPM.
The chi-squared goodness-of-fit test discussed for the AMSAA-Crow interval data case may be applied to test the null hypothesis H_0: $M_B(t)$ is a NHPP with mean value function $\mu(t) = (\frac{\lambda_B}{\beta})\ln(1 + \beta t)$. The test can be applied to H_0 whether λ_B and β are estimated by cumulative B-mode first occurrence times or by using the interval data approach of Section 6.3.4.8.

To utilize a chi-squared test one must select a partition of the test phase $0=z_0<z_1<\ldots\leq z_s=T$. When λ_B and β are estimated from interval data, the partition used to estimate these parameters is the natural starting partition for the chi-squared test. For interval $(z_{i-1}, z_i]$, one must estimate e_i , expected number of new B-modes that occur during $(z_{i-1}, z_i]$. This is estimated by $\hat{e}_i=\hat{\mu}(z_i)-\hat{\mu}(z_{i-1})=\frac{\hat{\lambda}_B}{\hat{\beta}}\{\ln(1 + \hat{\beta}(z_i)-\ln(1 + \hat{\beta}(z_{i-1})\}$. If cumulative B-mode first occurrence data is utilized to estimate λ_B and β, then $\hat{e}_i$ should be based on these estimates. Likewise, if interval data is utilized to estimate λ_B and β, then $\hat{e}_i$ should be based on interval data estimates.

For both data cases, one should combine adjacent intervals where necessary to obtain a set of intervals $(z'_{i-1}, z'_i]$ for i=1,\ldots, s' such that the model estimate of $e'_i=\mu(z'_i)- \mu(z'_{i-1})$ is at least five for i=1,\ldots, s'. To use this chi-squared test one must have $s'\geq3$. The estimate of e'_i will be denoted by $\hat{e}'_i$. It is given by $\hat{e}'_i= \hat{\mu}(z'_i)- \hat{\mu}(z'_{i-1})=\frac{\hat{\lambda}_B}{\hat{\beta}}\{\ln(1 + \hat{\beta}(z'_i)- \ln(1 + \hat{\beta}(z'_{i-1})\}$. The $\hat{\lambda}_B$ and

124

$\hat{\beta}$ are based on the cumulative B-mode first occurrence time data or the interval data associated with the original intervals $(z_{i-1}, z_i]$. With $\hat{e}_i' \geq 5$ for i=1,…, s', the test statistic $X^2(s'-2)=\sum_{i=1}^{s'}\frac{(m_i'-\hat{e}_i')^2}{\hat{e}_i'}$ will be approximately a chi-squared random variable with s'-2 degrees of freedom. In the above equation, m_i' denotes the observed number of new B-modes that occur during $(z'_{i-1}, z'_i]$.

If $X^2(s'-2)> X_{1-\propto}^2(s'-2)$ then H$_0$ would be rejected at significance level $\propto$. In this inequality, $X_{1-\propto}^2(s'-2)$ denotes the $1-\propto$ percentile of a chi-squared random variable with s'-2 degrees of freedom.

For cumulative time data, one should plot the model estimate of the expected number of distinct B-modes that occur by t_i versus the observed number that occur by t_i, for each cumulative B-mode first occurrence time t_i for i=1,…, m. For interval data, one should base the estimates of λ_B and β on the original partition 0=z$_0$<z$_1$<….<z$_s$=T. The corresponding model estimates of $\hat{\mu}(z_i)$ versus the observed number of distinct B-modes by z_i should be plotted for i=1,…, s.

6.3.4.10 A Plausibility Check for AMPM.
Let T denote the length of a test phase of interest. Also let N_A and N_B denote the numbers of failures over the test phase associated with A-modes and B-modes, respectively. Then the total number of failures observed during the test phase is $N = N_A + N_B$. Let ρ(T) denote the system failure intensity after mitigating all the B-modes that are observed in the test phase. Recall $\rho(T) = \lambda_A+(1-\mu_d)\{\lambda_B - h(T)\}+ h(T)$ assuming the observed B-modes are fixed, on average, with fix effectiveness factor μ_d. Let $\hat{\rho}$(T) denote the AMPM assessment of $\rho(T)$. In assessing $\rho(T)$, one is assuming that each surfaced B-mode is mitigated either during the test phase or in the CAP at the conclusion of the test phase. Thus if a follow-on test of length T is conducted under the same test conditions, one would expect to obtain fewer failures than N. Note $\{\hat{\rho}(T)\}$T would be an estimate of the number of failures to expect in such a follow-on test. Thus it is reasonable to expect that $\hat{\rho}$(T) satisfies the following inequality: $\{\hat{\rho}(T)\}$T<N, i.e., $\hat{\rho}$(T)<$\frac{N}{T}$. This inequality can be considered a "plausibility check" for the assessment $\hat{\rho}$(T).

This inequality can also be viewed as a plausibility check on μ_d^*, the assessed value of the average B-mode FEF μ_d. To see this, note that the inequality is equivalent to $\hat{\lambda}_A +(1-\mu_d^*)(\hat{\lambda}_B - \hat{h}(T))+ \hat{h}(T)<\frac{N}{T}$. Replacing $\hat{\lambda}_A$ by $\frac{N_A}{T}$ and $\frac{N}{T}$ by $\frac{N_A}{T} + \frac{N_B}{T}$, upon simplifying one obtains $(1-\mu_d^*) \hat{\lambda}_B+\mu_d^*\hat{h}(T)<\frac{N_B}{T}$. This inequality is equivalent to $\mu_d^*>\frac{\hat{\lambda}_B-\frac{N_B}{T}}{\hat{\lambda}_B-\hat{h}(T)}$.

Assuming $\hat{h}$(t) is a decreasing function, one can show $\hat{h}$(T)<$\frac{N_B}{T}$. This implies there always exists an assessment of μ_d that is less than one which will satisfy inequality $\mu_d^*>\frac{\hat{\lambda}_B-\frac{N_B}{T}}{\hat{\lambda}_B-\hat{h}(T)}$. Such an assessment of μ_d guarantees that $\hat{\rho}$(T) will satisfy the plausibility check, i.e., inequality $\hat{\rho}$(T)<$\frac{N}{T}$.

For the special case where all fixes are delayed until the CAP then any assessment of μ_d greater than zero will ensure that $\mu_d^* > \dfrac{\hat{\lambda}_B - \frac{N_B}{T}}{\hat{\lambda}_B - \hat{h}(\text{T})}$ holds, provided λ_B is estimated by $\dfrac{N_B}{T}$ for this case.

If the FEF assessments d_i^* tend to be biased low relative to the true effectiveness of the fixes, d_i, for the observed B-modes then $\mu^* = \frac{1}{m}\sum_{i \in obs} d_i^*$ will be biased low. This could lead to inequality $\mu_d^* > \dfrac{\hat{\lambda}_B - \frac{N_B}{T}}{\hat{\lambda}_B - \hat{h}(\text{T})}$ not being satisified. This in turn would lead to $\hat{\rho}(\text{T})$ failing the plausibility check, i.e. violating inequality $\hat{\rho}(\text{T}) < \frac{N}{T}$. This situation can arise when short-term expedient fixes are being applied during the test phase that are more effective than the assessments of d_i based on long-term tactical solutions. The short-term fixes may be highly effective at preventing repeat failures for the remainder of the test but are typically not suitable tactical solutions. One commonly encountered short-term "fix" is simply avoiding exercising devices on the system that have known failure modes which one intends to fix at a future date. The motivation for this course of action is to foster test execution in a timely manner. Such a test execution strategy would in essence be employing short-term fixes for the known device failure modes that have one as their realized FEF. This would result in the right hand side of inequality $\hat{\rho}(\text{T}) < \frac{N}{T}$ being artificially reduced. However $\hat{\rho}(\text{T})$ is typically based on engineering judgements d_i^* that reflect the expected effectivity of tactically suitable long-term corrective actions. Such a disconnect between the left and right hand sides of inequality $\hat{\rho}(\text{T}) < \frac{N}{T}$ would tend to result in $\hat{\rho}(\text{T})$ being greater than $\frac{N}{T}$. Additionally, highly effective short-term fixes applied in the test phase would tend to inflate tracking model estimates of the reliability.

Another reason the inequality $\hat{\rho}(\text{T}) < \frac{N}{T}$ may not be satisfied is that the model is not consistent with the data as determined by statistical or visual goodness-of-fit tests. Even when the model appears to be consistent with the pattern of B-mode first occurrence data, a biased estimate $\hat{\lambda}_B$ can be obtained that is substantially larger than λ_B. This could also result in $\hat{\rho}(\text{T})$ not satisfying the inequality $\hat{\rho}(\text{T}) < \frac{N}{T}$.

In cases where the graph of the number of B-modes that occur by t versus t appears steep at the origin (such as displayed in Figure 39), $\hat{\mu}(\text{t})$ will have a steep slope $\hat{\lambda}_B = \hat{h}(0)$ at the origin. Under such conditions the accuracy of $\hat{\lambda}_B$ relative to λ_B could degrade. If $\hat{\lambda}_B$ is significantly greater than λ_B then the growth potential failure intensity estimate $\hat{\rho}_{GP} = \hat{\lambda}_A + (1 - \mu_d^*)\hat{\lambda}_B$ would be unduly large. Since $\hat{\rho}(\text{T}) > \hat{\rho}_{GP}$, such a circumstance could lead to $\hat{\rho}(\text{T})$ not passing the plausibility test, even when the assessment of μ_d is reasonable. To check this possibility, it is useful to calculate $\dfrac{N_B(t_0)}{t_0}$ where $N_B(t_0)$ denotes the number of B-mode failures over $[0, t_0]$, where $0 < t_0 \le \text{T}$. The value t_0 should be chosen as small as possible subject to $N_B(t_0)$ being large enough so that $\dfrac{N_B(t_0)}{t_0}$ provides a stable estimate of the average B-mode failure intensity over $[0, t_0]$. Although one should expect $\hat{\lambda}_B > \dfrac{N_B(t_0)}{t_0}$, if $\hat{\lambda}_B$ significantly exceeds this ratio then $\hat{\lambda}_B$ may be unduly large compared to the value λ_B.

When $\hat{\mu}$(t) and the B-mode cumulative data have a steep slope at the origin it could be an indication of test start-up problems. Many of these early B-modes may be due to operator and maintenance procedure problems. Such steepness would dictate that these initial problems must be fixed with a high average FEF to attain a growth potential MTBF that supports the development test goal MTBF. This suggests that over an initial period $[0, v]$ where the cumulative number of observed B-modes is quickly rising, one may wish to assess an average FEF μ_1^* based on the individual B-mode FEF assessments d_i^* for modes that occur during $[0,v]$. A separate average FEF μ_2^* could be assessed based on the arithmetic average of the d_i^* for the B-modes surfaced over the remainder of the test phase (v,T). The resulting AMPM assessment of ρ(T) utilizing v, μ_1^*, and μ_2^* is discussed in Section 6.3.4.8.2. The plausibility check can then be applied to this $\hat{\rho}$(T) assessment.

If the initial steepness over $[0, v]$ is thought to be due to infant mortality issues one may wish to utilize only the data beyond $[0, v]$. Section 6.3.4.11 discusses two such procedures.

6.3.4.11   Analysis Considerations for Apparent Changes in Failure Mode Rates of Occurrence.

There may be times when there appear to be changes in the rate of occurrence of B-modes, especially during initial periods of testing or near the end of testing. This may be seen through plots of the rate of occurrence of B-modes, wherein the rate is initially high, and after a period of testing, the rate of occurrence slows; or the reverse occurs. Such changes may occur due to a number of reasons: initial assembly or infant mortality, quality or weak areas of the design and operator unfamiliarity with equipment, or refurbishing hardware near the end of a test phase. The following sections address three approaches for dealing with apparent changes in failure mode rates of occurrence.

6.3.4.11.1   The Gap Method.

A projection is a reliability estimate that takes into account the contribution of A-mode failures, B-mode failures, and the effectiveness of the corrective actions that are implemented to B-modes. The statistical reliability projection, based on this information, is significantly influenced by the initial B-mode rate of occurrence. For some systems under development, the initial rate of occurrence of B-modes is steep (see Figure 38). This can be due to problems associated with infant mortality, initial assembly procedures, or operator unfamiliarity with the system. One way of excluding the impact of such start-up problems on the reliability projection is to utilize only the test data beyond an initial time period. This is referred to as "jumping the gap."

In deciding to "jump the gap" over $[0, v]$, one needs to make the following judgements concerning the NHPP model: (1) it provides an adequate representation for the number of distinct B-modes that occur by t for those modes not associated with the special classes of start-up modes; and (2) to a high degree, the start-up modes are confined to the initial segment $[0, v]$ of the test phase. Under such circumstances it is reasonable to estimate the parameters governing the underlying NHPP mean value function μ(t) by maximizing the portion of the NHPP likelihood function associated with the B-mode first occurrence data in (v, T). The resulting model parameter MLEs along with estimates of λ_A and μ_d (based on the data not associated with start-up problems) would be used to assess the failure intensity projection $\rho(T)$. Visual and

127

statistical goodness-of-fit procedures can be conducted as before utilizing the B-mode first occurrence data beyond v.
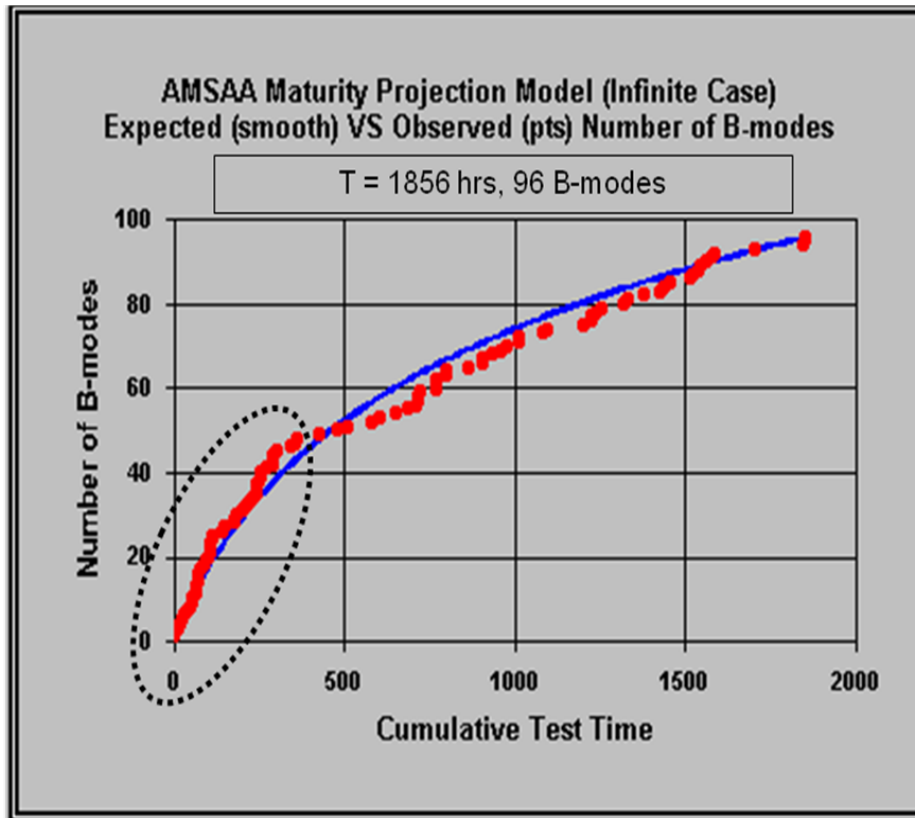

FIGURE 39.   Example Curve for Illustrating the Gap Method.

6.3.4.11.2 The Segmented FEF Method.
The segmented FEF method is another strategy that may be useful for a dataset where the initial rate of occurrence of B-modes is very steep due to start-up problems. With this method, a partition point, *v*, is chosen such that a relatively high average FEF, μ_1, is applied to the B-modes occurring on or before *v*, and a more typical average FEF, μ_2, is applied to the B-modes surfaced beyond *v*. Note that this method is justified only if the early B-modes (those occurring on or before *v*) are aggressively and effectively corrected. Engineering analysis should be the driving force in choosing the initial segment *v*; however, statistical analysis may be useful in viewing and/or estimating *v*. In using this method, the underlying assumption is that the B-modes occurring during the early segment are those whose failure mechanisms are so well understood that they can be assigned very high FEFs.

Recall that the failure intensity projection equation (for single FEF) is given by

$$\rho(t) = \lambda_A + (1 - \mu_d)[\lambda_B - h(t)] + h(t).$$

The failure intensity projection equation for the segmented FEF method is given by

$$\rho(t) = \lambda_A + (1 - \mu_1)\big(\lambda_B - h(v)\big) + (1 - \mu_2)(h(v) - h(t)) + h(t)$$

128

where $0 \leq v < t \leq T$. "v" is a partition point such that FEF μ_1 is applied to B-modes surfaced on or before v, and FEF μ_2 is applied to B-modes surfaced beyond v. Note that when $v = 0$, the last equation reduces to the first equation, and $h(0) = \lambda_B$.

6.3.4.11.3 The Restart Method.
Another approach to analyze the data is to re-initialize the data beginning at the partition point, v, and to re-initialize the test time to zero at v. Thus, failure data prior to v would not be used in the analysis. Note that a repeat of a failure mode occurring prior to v *that occurs after v* may now be the first occurrence of that failure mode and thus included as a "first occurrence."

The rationale for implementing this approach versus the Gap Method may be for practical analysis or engineering concerns of the data, such as significant changes in the system's configuration or possible differences in test conditions.

The basic difference between the Gap Method and the Restart Method is after determining the partition point, v, the Restart Method reinitializes test time to zero at v. And as noted above, first occurrences of modes prior to v having repeats after v would now, with the next occurrence of that mode, be considered first occurrences. The analysis would then be the same as with the normal AMPM application.

6.3.5 AMSAA Maturity Projection Model based on Stein Estimation (AMPM-Stein).

6.3.5.1 Purpose.
The purpose of AMPM-Stein is to provide an estimate of the projected reliability following the implementation of delayed fixes. The model also provides estimates of the following important reliability growth metrics:
   a) B-mode initial failure intensity
   b) expected number of B-modes surfaced
   c) percent surfaced of the B-mode initial failure intensity
   d) rate of occurrence of new B-modes
.
6.3.5.2 Assumptions.
The assumptions associated with AMPM-Stein are the same as those for AMPM, with one exception – all corrective actions must be delayed.

6.3.5.3 Limitations.
The limitations associated with AMPM-Stein include:
   a) FEFs are often a subjective input; and
   b) there must be at least one repeat failure mode.

6.3.5.4 Benefits.
The benefits associated with (Ellner and Hall 2004) include:

a) May provide a more accurate reliability projection than that of ACPM, given that the model assumptions hold (this has been shown via simulation case studies conducted to date);
b) Allows for natural trade-off analysis between reliability improvement and incremental cost;
c) Identified failure modes do not need to be grouped into A-modes and B-modes;
d) Utilizes individual failure mode FEFs that are not required to be independent of the modes' initial failure rate; and
e) Less data is required than ACPM.   Mode first occurrence times are not required for estimating model parameters.

6.3.5.5 Differences in Technical Approach.

There are a few significant differences between the AMPM-Stein approach and the other methods (ACPM and AMPM), which include:
a) For surfaced modes, the AMPM-Stein approach does not require one to distinguish between A-modes and B-modes, other than through the assignment of a zero, and positive FEF, respectively, to surfaced modes.
b) Only FEFs associated with the surfaced modes need to be referenced; and
c) The AMPM-Stein projection is a direct assessment of the realized system failure intensity after failure mode mitigation, whereas the ACPM and AMPM approaches indirectly attempt to assess the realized system failure intensity by estimating the expected value of the mitigated system failure intensity, $r(T)$, where $r(T)$ is viewed as a random variable.

6.3.5.6 Methodology.

AMPM-Stein uses an estimation criterion utilized by Stein in (Stein 1981) to estimate the vector of means for multi-variate normal random vectors. The criterion applied to estimating the vector of initial B-mode failure rates ($\lambda_1, ..., \lambda_k$) produces the estimated vector $(\tilde{\lambda}_1, ..., \tilde{\lambda}_k)$ where

$$\text{Stein Estimate:} \quad \tilde{\lambda}_i = \theta \hat{\lambda}_i + (1 - \theta) \left\{ \frac{1}{k} \sum_{i=1}^{k} \hat{\lambda}_i \right\}$$

for $i = 1,\ldots,k$. In this expression $\hat{\lambda}_i = \frac{N_i}{T}$ and θ is chosen to be the value $\theta_S \in [0, 1]$ that minimizes the expected sum of squared errors $\sum_{i=1}^{k} (\tilde{\lambda}_i - \lambda_i)^2$. The parameter θ_S is referred to as the shrinkage factor. The corresponding projected failure intensity is estimated as:

$$\hat{\rho}(T) = \hat{\lambda}_A + \sum_{i \in obs\,(B)} (1 - d_i^*) \tilde{\lambda}_i + \sum_{i \in obs\,(B)} \tilde{\lambda}_i$$

where d_i^* is the assessment of the true value of FEF d_i. Note that the Stein estimate of observed mode failure rates are reduced by their assessed mode FEF and that the collective failure rate due to unobserved failure modes is estimated by the sum of the Stein failure rate estimates for unobserved modes. It may be shown that

$$\theta_S = \frac{k \text{Var}(\lambda_i)}{k \text{Var}(\lambda_i) + \left(\frac{\lambda}{T}\right)\left(1 - \frac{1}{k}\right)}$$

130

where var $(\lambda_i) = \frac{\sum_{i=1}^{K}(\lambda_i - \bar{\lambda})^2}{k}$, $\lambda = \sum_{i=1}^{k} \lambda_i$, and $\bar{\lambda} = \frac{\lambda}{k}$. Thus, θ_S depends on the unknown values k, λ, and the population variance of the λ_i, $Var(\lambda_i)$. Proceeding as for the AMPM, the λ_i are regarded as a realized random sample from a gamma distribution. Doing so, one can derive an MLE of θ_S, say $\hat{\theta}_{S,K}$, for finite k. The limit for the $\hat{\theta}_{S,K}$, $\lim_{K \to \infty} \hat{\theta}_{S,K}$ is then shown to equal $\hat{\theta}_{S,\infty} = \frac{\hat{\beta}_\infty}{1 + \hat{\beta}_\infty T}$. In this formula, $\hat{\beta}_\infty$ satisfies the equation $\left(\frac{N_B}{\hat{\beta}_\infty T}\right) \ln(1 + \hat{\beta}_\infty T) = m$, the number of observed B-modes. This yields a projected MTBF, $\hat{M}_{S,\infty}$, such that

$$\hat{M}_{S,\infty} = \frac{1}{\hat{\rho}_{S,\infty}(T)} \text{ for } \hat{\rho}_{S,\infty} = \frac{N_A}{T} + \sum_{i \in \text{obs}(B)} (1 - d_i^*)\tilde{\lambda}_{i,\infty} + \sum_{i \in \overline{\text{obs}}(B)} \tilde{\lambda}_{i,\infty}$$

where $\sum_{i \in \overline{\text{obs}}(B)} \tilde{\lambda}_{i,\infty} = (1 - \hat{\theta}_{S,\infty})\left(\frac{N_B}{T}\right), \tilde{\lambda}_{i,\infty} = \hat{\theta}_{S,\infty}\left(\frac{N_i}{T}\right)$ for each $i \in$ obs. In light of the above procedure, the derived projection is termed AMPM-Stein. Simulations conducted by (Ellner and Hall 2004)indicate that the accuracy of the AMPM-Stein projection appears favorable compared to that of the international standard adopted by International Electrotechnical Commission (IEC) and American National Standards Institute (ANSI), even when λ_i were randomly chosen from Weibull or lognormal parent populations. Thus, with regard to these three parent populations, the results obtained in the simulation study were robust, even though the estimation procedure assumes that the parent population of the λ_i is a gamma distribution.

The Stein projection cannot be directly calculated from the data for a set of d_i^*, since k is typically unknown before and after the test and θ_S is a function of $Var[\lambda_i]$, λ, and k. However, as indicated above, approximations to the Stein projection can be obtained that can be calculated from the test data and the assessed FEFs.

Finally, here is a note on the connection between the AMPM estimate for $h(t)$ and the Stein projection. It is shown that $h(t)$ is the expected failure intensity due to the B-modes not surfaced by t. The AMPM estimate for $\hat{h}(t)$ as $k_B \to \infty$ is of the form $\frac{\hat{\lambda}_B}{(1 + \hat{\beta} \cdot t)}$. For $t = T$, this form is compatible for complex systems with the Stein projection expression for the portion of the mitigated system failure intensity attributable to the B-modes not surfaced by T. This is interesting to note since the Stein projection approach does not treat the initial B-mode failure rates as a realization of a random sample from some assumed parent population.

6.3.6   AMSAA Discrete Projection Model based on Stein Estimation (ADPM-Stein).

6.3.6.1 Purpose.
ADPM-Stein is useful in cases where one or more failure modes are (or can be) discovered in a single trial, and catastrophic failure modes have been previously discovered and corrected. ADPM-Stein is an alternative to the popular competing risks approach. However, ADPM-Stein would not be suitable for application to one-shot development program data sets where failure mode preclusion effects significantly alter the expected number of B-modes prior to trial t from that obtained by the model's assumptions.

an overly optimistic assessment. Therefore, a finite and positive estimate for each failure mode probability of occurrence is desired, whether observed during testing or not. One way to do this is to utilize a shrinkage factor estimator given by

$$\tilde{p}_i \equiv \theta \cdot \hat{p}_i + (1 - \theta) \cdot \frac{\sum_{i=1}^{k} \hat{p}_i}{k}$$

where θ (unknown) is referred to as the shrinkage factor, and k denotes the total potential number of failure modes inherent to the system. The optimal value of $\theta \in (0,1)$ can be mathematically expressed as

$$\frac{d}{d\theta} E\left[\sum_{i=1}^{k} (\tilde{p}_i - p_i)^2\right] = 0$$

The solution of θ is given by

$$\theta_k = \frac{Var(p_i)}{\left(\frac{E(p_i)[1 - E(p_i)] - Var(p_i)}{T}\right)\left[1 - \frac{1}{k}\right] + Var(p_i)}$$

Thus, the shrinkage is given in terms of quantities that can be easily estimated for a given k; namely the mean and variance of the p_i.

6.3.6.6 Reliability Growth Projection.
After mitigation to some, or all failure modes observed during testing, the true but unknown system reliability growth is defined as

$$r(T) \equiv \prod_{i \in obs} [1 - (1 - d_i) \cdot p_i] \prod_{j \in obs'} (1 - p_j)$$

where $d_i \in [0,1]$ represents the FEF of failure mode i, the true but unknown fraction reduction in initial mode failure probability i due to implementation of a unique corrective action. In the ADPM-Stein, $(1 - d_i) \cdot p_i$ represents the true reduction in failure probability i due to correction. It will typically be the case that $d_i \in (0,1)$, as $d_i = 0$ models the condition where a given failure mode is not addressed (e.g., an A-mode), and $d_i = 1$ corresponds to complete elimination of the failure mode's probability of occurrence. Notice that the model does not require utilization of the A-mode/B-mode classification scheme, as A-modes need only be distinguished from B-modes via a zero FEF.

6.3.6.7 Estimation Procedures.

6.3.6.7.1 Parametric Approach. Assume that the initial mode probabilities of failure $p_1, \ldots, p_k$ constitute a realization of a random sample $P_1, \ldots, P_k$ from a beta distribution with the parameterization

$$f(p_i) \equiv \frac{\Gamma(n)}{\Gamma(x) \cdot \Gamma(n-x)} \cdot p_i^{x-1} \cdot (1-p_i)^{n-x-1}$$

for $p_i \in [0,1]$, and 0 otherwise; where n is often interpreted as pseudo trials, and x as pseudo failures, and

$$\Gamma(x) \equiv \int_0^\infty t^{x-1} \cdot e^{-t}\, dt$$

is the gamma function. In the presented model these interpretations of n and x are not utilized. However they are useful to clarify the particular parameterization that is being used for $f(p_i)$. The above beta assumption not only facilitates convenient estimation of θ, but models mode-to mode variability in the initial failure probabilities of occurrence. The source of such variability could result from many different factors, including, but not limited to: mode failure mechanism, variation in environmental conditions, manufacturing processes, operating procedures, maintenance philosophies, or a combination of these. Based on the beta assumption with parameterization of $f(p_i)$, the associated mean and variance are given respectively by

$E(P_i) = \frac{x}{n}$,
and

$$Var(P_i) = \frac{x \cdot (n-x)}{n^2 \cdot (n+1)}$$

The equation for estimating θ_k, the shrinkage factor, is now in terms of only three unknowns— the population mean of the failure probabilities, the population variance of the failure probabilities, and k. The first two unknowns are approximated by the expected value and variance terms given above, which are in terms of the two unknown beta shape parameters. The third unknown, k, is treated in two ways. First, one assumes a value of k, which can be done in applications where the system is well understood. Second, one allows k to grow without bound to study the limiting behavior of the model equations. This is suitable in cases where the number of failure modes is unknown and the system is complex.

6.3.6.7.2 Moment-based Estimation Procedure.
Moment estimators for the beta shape parameters, per the special case where all failure probabilities are estimated via the same number of trials, are given by

$$\check{n}_k = \frac{\left(\overline{p}_u - m_u^2\right)}{m_u^2 - \frac{\overline{p}_u}{T} - \left(1 - \frac{1}{T}\right) \cdot \overline{p}_u^2}$$

and
$\check{x}_k = \check{n}_k \cdot \overline{p}_u$,

where $\overline{p}_u \equiv \frac{\sum_{i=1}^k \hat{p}_i}{k}$, and $m_u^2 \equiv \frac{\sum_{i=1}^k \hat{p}_i^2}{k}$ are the unweighted first and second sample moments, respectively. Using the above method of moments (MME) for the beta parameters with the equation for θ_k, the approximation of θ_k can be expressed as

$\check{\theta}_k = \frac{1}{\left(\frac{\check{n}_k}{T}\right)\left(1-\frac{1}{k}\right)+1}$.

Using this, the moment-based shrinkage factor estimate of p_i for finite k is then given by

$$\check{p}_{k,i} = \check{\theta}_k \cdot \hat{p}_i + \left(1 - \check{\theta}_k\right)\left(\frac{N}{k \cdot T}\right)$$

where $N \equiv \sum_{i=1}^k N_i$ is the total number of failures observed in T trials. In the notation above, $\check{p}_{k,i}$ denotes the method of moments estimate of $\tilde{p}_{k,i}$. Let the total number of observed failure modes be denoted by $m = |obs|$, which implies that there are $|obs'| = k - m$ unobserved failure modes. Then the MME-based reliability growth projection for an assumed number of failure modes k is given by

$$\tilde{r}_k(T) = \prod_{i \in obs} \left[1 - (1 - d_i^*) \cdot \check{p}_{k,i}\right] \cdot \left[1 - \left(1 - \check{\theta}_\infty\right) \cdot \left(\frac{N}{T}\right)\right]^{k-m}$$

where d_i^* estimates d_i.

Letting $k \to \infty$, the reliability projection for complex systems or subsystems is given by

$$\check{r}_\infty(T) = \lim_{k \to \infty} \check{r}_k(T) = \prod_{i \in obs}\left[1 - (1 - d_i^*) \cdot \check{p}_{\infty,i}\right] \cdot \exp\left[-\left(1 - \check{\theta}_\infty\right) \cdot \left(\frac{N}{T}\right)\right]$$

where

$$\check{p}_{\infty,i} = \check{\theta}_\infty \cdot \hat{p}_i,$$

$$\check{\theta}_\infty = \frac{T}{\check{n}_\infty + T}, \text{ and}$$

$$\check{n}_\infty \equiv \lim_{k \to \infty} \check{n}_k = \frac{\sum_{i=1}^m \hat{p}_i - \sum_{i=1}^m \hat{p}_{i=1}^2}{\sum_{i=1}^m \hat{p}_{i=1}^2 - \sum_{i=1}^m \frac{\hat{p}_i}{T}}.$$

In the above $\prod_{i \in obs}$ indicates a product of the displayed factors for each observed mode i.

6.3.6.7.3 Likelihood-based Estimation Procedure.
The method of marginal maximum likelihood provides estimates of the beta parameters *n* and *x* that maximize the beta marginal likelihood function.

The finite k likelihood-based estimates $\hat{\theta}_k$, and $\hat{\tilde{p}}_{k,i}$ are obtained analogously to that of the moment based estimators, with appropriate substitution of the MLE in place of the MME. This provides the likelihood-based estimate of system reliability growth

$$\hat{r}_k(T) = \prod_{i \in obs} \left[1 - (1 - d_i^*) \cdot \hat{\tilde{p}}_{k,i}\right] \cdot \left[1 - (1 - \hat{\theta}_k) \cdot \left(\frac{N}{k \cdot T}\right)\right]^{k-m}$$

It may be shown that the limiting behavior of the likelihood-based estimate for one-shot system reliability growth is given by

$$\hat{\tilde{r}}_\infty \equiv \lim_{k \to \infty} \hat{\tilde{r}}_k(T) = \prod_{i \in obs} \left[1 - (1 - d_i^*) \cdot \hat{\tilde{p}}_{\infty,i}\right] \cdot \exp\left[-(1 - \hat{\theta}_\infty) \cdot \left(\frac{N}{T}\right)\right]$$

where

$$\hat{\tilde{p}}_{\infty,i} = \hat{\theta}_\infty \cdot \hat{p}_i$$

$$\hat{\theta}_\infty = \frac{T}{\hat{n}_\infty + T}$$

and $\hat{n}_\infty$ is found as the solution for the beta parameter n by numerically solving the equation below:

$$\sum_{i=1}^{m} \sum_{j=0}^{T-N_i-1} \left(\frac{1}{\hat{n}_\infty + j}\right) = m \cdot \sum_{j=0}^{T-1} \left(\frac{1}{(\hat{n}_\infty + j)} - \frac{1}{(\hat{n}_\infty + j)^2 \cdot \sum_{i=0}^{T-1}\left(\frac{1}{\hat{n}_\infty + i}\right)}\right)$$

Additional details concerning this discrete projection model may be found in (Hall and Mosleh 2008).

There is also a version of the discrete projection model that addresses the more general situation where there may be a mix of delayed and nondelayed corrective actions. The reference for this ADPM is (Hall, Ellner and Mosleh 2010).

The presented ADPM could be considered a discrete version of the AMPM described in Section 6.3.4.

136

6.3.6.7.4 Goodness-of-Fit.

The Goodness-of-Fit of the model can be graphically studied by plotting the cumulative number of observed failure modes versus trials against the estimate of the cumulative expected number of observed failure modes through trial *t* given by

$$\hat{\mu}(t) = \left(\frac{m}{\sum_{j=1}^{T-1} \left(\frac{1}{\hat{n}_\infty + j} \right)} \right) \cdot \left(\frac{\Gamma'(\hat{n}_\infty + t)}{\Gamma(\hat{n}_\infty + t)} - \frac{\Gamma'(\hat{n}_\infty)}{\Gamma(\hat{n}_\infty)} \right)$$

7. NOTES.

7.1 Intended Use.
This handbook provides guidance to help in the management of reliability growth through the acquisition process.

7.2 Superseding Information.
This handbook is in lieu of the materials in MIL-HDBK-189, dated 24 Oct 2000, and the interim revision document MIL-HDBK-00189, dated 10 Sep 2009.

7.3 Subject Term (Keyword Listing).
AMSAA Crow Planning Model (ACPM)
AMSAA Discrete Projection Model (ADPM)
AMSAA Maturity Projection Model (AMPM)
Fix Effectiveness Factor (FEF)
Growth Potential (GP)
Growth Rate
Homogeneous Poisson Process (HPP)
Idealized Growth Curve
Management Strategy (MS)
Non-Homogeneous Poisson Process (NHPP)
Poisson Process
Planning Model Based on Projection Methodology (PM2)
Reliability Growth Planning
Reliability Growth Tracking
Reliability Growth Projection
System Level Planning Model (SPLAN)
Subsystem Level Planning Model (SSPLAN)
Subsystem Level Tracking Model (SSTRACK)

Changes from previous issue: Marginal notations are not used in this revision to identify change with respect to the previous issue due to the extent of the changes.

8. BIBLIOGRAPHY.

AMSAA. "MIL-HDBK-189: Reliability Growth Management." APG, Feb 1981.
Broemm, Ellner, and Woodworth. *AMSAA Reliability Growth Guide.* APG: AMSAA TR-652, Sep 2000.

Broemm, William. "Subsystem Tracking Model (SSTRACK) Demo of System Reliability from Lower Level Testing." APG, March 1996.

Chenard, L., J.-L. Peree, and Paul M. Ellner. "Successful Applications of a Coordinated Subsystem Reliability Growth Planning Approach." *Proceedings of the 30th ESReDA Seminar: Reliability of Safety-Critical Systems.* 2006.

Chenard, L., Paul M. Ellner, and Peree J.-L. "A Successfully Implemented Coordinated Subsystem Reliability Growth Planning Approach." *23rd Annual National Test and Evaluation Conference.* 2007.

Crow, Larry H. "A Methodology for Managing Reliability Growth during Operational Mission Profile Testing." *RAMS.* 2008.

Crow, Larry H. "AMSAA Discrete Reliability Growth Model." APG: Methodology Office Note 1-83, 1983.

—. "An Extended Reliability Growth Model for Managing and Assessing Corrective Actions." Proceedings of RAMS 2004 Symposium. 73-80, 2004.

Crow, Larry H. *An Improved Methodology for Reliability Growth Projections.* APG: AMSAA Technical Report TR-357, Jun 1982.

Crow, Larry H. *Confidence Interval Procedures for Reliability Growth Analysis.* APG: AMSAA Technical Report TR-197, Jun 1977.

Crow, Larry H. *Reliability Analysis for Complex Repairable Systems.* APG: AMSAA Technical Report TR-138, 1974.

—. "Reliability Growth Projection from Delayed Fixes." *Proceedings of RAMS 1983 Symposium.* 1983. 84-89.

—. "Reliability Values in the Development Process." *Proceedings Annual Reliability and Maintainability Symposium.* 1985.

Duane, J.T. "Learning Curve Approach to Reliability Monitoring." *IEEE Transactions on Aerospace, Vol. 2*, 1964: 563-564.

Ellner, Paul M., and Bruce Trapnell. *AMSAA Reliability Growth Data Study.* APG: AMSAA Interim Note IN-R-184, 1990.

Ellner, Paul M., and J. Brian Hall. "AMSAA Maturity Projection Model Based on Stein Estimation." Technical Report TR-751, APG, 2004.

Ellner, Paul M., and J. Brian Hall. *Planning Model Based on Projection Methodology (PM2).* APG: AMSAA Technical Report TR-2006-09, Mar 2006.

Ellner, Paul M., and Lindalee C. Wald. "AMSAA Maturity Projection Model." *1995 Proceedings of Annual Reliability and Maintainability Symposium.* Jan 1995.

Ellner, Paul M., and Robert Mioduski. *Operating Characteristic Analysis for Reliability Growth Programs.* APG: AMSAA Technical Report TR-524, Aug 1992.

Ellner, Paul M., Lindalee C. Wald, and J. Woodworth. "A Parametric Empirical Bayes Approach to Reliability Projection." *Workshop on Reliability Growth Modeling: Objectives, Expectations and Approaches.* The Center for Reliability Engineering, 1998.

Hall, J. Brian, and A. Mosleh. "A Reliability Growth Projection Model for One-Shot Systems." *IEEE Transactions on Reliability, VOL. 57, NO. 1, MARCH 2008* (IEEE), 2008.

Hall, J. Brian., Paul M. Ellner, and A. Mosleh. "Reliability Growth Management Metrics and Statistical Methods for Discrete-Use Systems." *Technometrics, NOVEMBER 2010, VOL. 52, NO. 4*, 2010: 379.

Hall, J.B. *Implementation Guide for U.S. Army Reliability Policy.* APG: ATEC, Jun 2009.
Stein, C.M. "Estimation of the Mean of a Multivariate Normal Distribution"." *The Annals of Statistics*, 1981.

CONCLUDING MATERIAL

Custodians: Preparing Activity:
Army-SY   Army-SY
Navy-EC
Air Force-11

(Project SESS-2010-008)

Review Activities:
Army-CR
Army-AV
Army-MI

NOTE: The activities listed above were interested in this document as of the date of this document. Since organizations and responsibilities can change, you should verify the currency of the information above using ASSIST Online database at is https://assist.daps.dla.mil/online/.

141