(12)

RADC-TR-86-148
Final Technical Report
September 1986

RELIABILITY GROWTH PREDICTION

The Analytic Sciences Corporation

R. K. Gates, G. J. Gibson and K. K. McLain

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, NY 13441-5700

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-86-148 has been reviewed and is approved for publication.

APPROVED: Eugene Fiorentino

EUGENE FIORENTINO
Project Engineer

APPROVED: W S Tuthill

W. S. TUTHILL, COLONEL, USAF
Chief, Reliability & Compatibility Division

FOR THE COMMANDER: John A. Ritz

JOHN A. RITZ
Plans & Programs Division

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (RBET) Griffiss AFB NY 13441-5700. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b RESTRICTIVE MARKINGS: N/A

A176 128

2a SECURITY CLASSIFICATION AUTHORITY: N/A

2b DECLASSIFICATION DOWNGRADING SCHEDULE: N/A

3 DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.

4 PERFORMING ORGANIZATION REPORT NUMBER(S): TR-5054-1-1

5 MONITORING ORGANIZATION REPORT NUMBER(S): RADC-TR-86-148

6a NAME OF PERFORMING ORGANIZATION: The Analytic Sciences Corporation

6b OFFICE SYMBOL (If applicable):

7a NAME OF MONITORING ORGANIZATION: Rome Air Development Center (RBET)

6c ADDRESS (City, State, and ZIP Code): One Jacob Way, Reading MA 01867

7b ADDRESS (City, State, and ZIP Code): Griffiss AFB NY 13441-5700

8a NAME OF FUNDING/SPONSORING ORGANIZATION: Rome Air Development Center

8b OFFICE SYMBOL (If applicable): RBET

9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: F30602-84-C-0032

8c ADDRESS (City, State, and ZIP Code): Griffiss AFB NY 13441-5700

10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT ACCESSION NO
62702F	2338	02	1C

11 TITLE (Include Security Classification): RELIABILITY GROWTH PREDICTION

12 PERSONAL AUTHOR(S): R. K. Gates, G. J. Gibson, K. K. McLain

13a TYPE OF REPORT: Final

13b TIME COVERED: FROM 1984 TO 1986

14 DATE OF REPORT (Year, Month, Day): September 1986

15 PAGE COUNT: 110

16 SUPPLEMENTARY NOTATION: N/A

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
14	04	07
02	01	09

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number): Reliability; Operations Research; Statistical Modeling

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

A reliability growth data base on nine avionic systems and thirty equipment items was compiled for the study. Several growth models were investigated and three were selected for application to the data: The Duane, AMSAA and IBM models. Goodness-of-fit tests for the models were applied and relationships between growth model parameters and equipment characters/program attributes were investigated. Both the Duane and AMSAA models were found to yield reasonably good fits to the data sets. Both models, however, were found to have limited utility as predictive tools. Anomolies in the growth data, such as burn-in effects and time delays in incorporating corrective actions tended to mask out any relationships between the growth parameters and equipment characteristics or program attributes. The IBM model was found to provide a more workable methodology for growth prediction. A prediction procedure was developed, based upon IBM model, which uses engineering information available at the start of the growth program. Guidelines for conducting reliability growth tests and examples of the reliability growth prediction

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT: ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21 ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL: Eugene Fiorentino

22b TELEPHONE (Include Area Code): (315) 330-3476

22c OFFICE SYMBOL: RADC (RBET)

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

procedures are provided in the report.

ABSTRACT

Candidate models for predicting the reliability growth that could be expected to occur during the development of an electronic system are evaluated. Development programs which conducted reliability growth testing are reviewed. The reliability data generated during these tests are analyzed relative to the candidate models. A procedure for predicting reliability growth based on equipment characteristics and program attributes is described, along with guidelines for conducting reliability growth tests.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

The Reliability Growth Prediction Study was conducted to develop a methodology for predicting the probable reliability growth characteristics for avionics electronic equipment based on specific equipment characteristics and development program attributes. A secondary objective was to provide guidelines for the selection of reliability growth program elements.

A data base of historical programs which had conducted reliability growth testing was assembled. Nine different avionic systems and 30 equipment items (Line Replaceable Units) comprised the data base. The data base included characterizations of the development programs and equipment items, as well as failure/time histories for the equipments as observed during growth testing.

Several reliability growth models were investigated to provide a framework for reliability growth prediction. Three models were selected for analysis of the historical data: the Duane model; the Army Material Systems Analysis Activity (AMSAA) model; and the IBM model. Data analyses included statistical tests of the goodness-of-fit of the models to the data and development of relationships between growth model parameters and equipment characteristics/program attributes.

Both the Duane and AMSAA models were found to yield reasonably good fits to most of the data sets. However, both were found to have limited utility as predictive tools because of the empirical nature of the model parameters. Characteristics of the failure-time history data tended to mask out any underlying relationships between the parameters and equipment characteristics and/or program attributes.

The IBM model was found to provide a more workable methodology for reliability growth prediction because its parameters lend themselves more easily to an engineering interpretation. Based on these interpretations and analyses on a limited number of data sets, a reliability growth prediction procedure was developed. The procedure allows prediction of expected reliability growth based on information that is available at the start of a development program, namely, equipment complexity and maturity, equipment operating hours to be accumulated during testing, and the test environmental profile relative to the operational profile.

An important conclusion is that reliability growth prediction and planning should focus more on classification of failure modes and corrective action identification, rather than on the slope of a line fitted to cumulative failure-time data plotted on a log-log scale. It is also recommended that verification of corrective action effectiveness be a major element of reliability growth planning.

1. INTRODUCTION

1.1 BACKGROUND

There has been an increased emphasis in recent years on the use of reliability growth testing in Full Scale Development programs for electronic systems. This emphasis has resulted from the recognition that one of the most cost-effective ways to achieve high operational reliability is to mature a system design before committing to production and deployment of the system. Reliability growth testing subjects prototype systems to a prolonged period of operation in an environment designed to surface failure modes and, hence, permit the identification of corrective actions, such as design changes or improved production processes, for incorporation into production systems. In most cases, the growth test includes an environmental profile (e.g., temperature, vibration, on-off cycling) which is an acceleration of the mission environment, the purpose being to stimulate equipment failures within the limited time available for testing.

A fundamental question relating to reliability growth testing is "What is the degree of reliability growth achieved during such a test?" Considerable research has been performed addressing this question, the research generally being within the following two categories:

- Formulation of mathematical models which are purported to represent the reliability growth process
- Case studies of reliability growth programs conducted for specific systems.

These two areas have come together in case studies in which specific growth models have been applied to data generated during a reliability growth test program. The two most commonly applied models in such studies are:

- The Duane Model
- The Army Material Systems Analysis Activity (AMSAA) Model.

Both of these models are consistent with the frequently observed phenomenon that measured cumulative failure rate versus cumulative operating time closely plots as a straight line on log-log paper. The principal difference is that the Duane model considers reliability growth to be deterministic in nature, whereas the AMSAA model is probabilistic and also recognizes the possibility that a reliability growth program may be conducted in a series of stages. Both models have become institutionalized to an extent in MIL-STD-1635 and MIL-HDBK-189, respectively.

Both the Duane and AMSAA models are parametric, a principal parameter being reliability growth rate, and the model parameters must be estimated to apply the model in a particular reliability growth program. To date, the model parameter estimation process has been accomplished by fitting the model to reliability growth data, using either statistical or "eyeball" techniques, and, in effect, letting the fit determine the values of the parameters. This is entirely satisfactory for measuring reliability growth during the progress of the test, or for predicting subsequent reliability growth once a sufficient amount of data have been generated to support an adequate fit. However, this estimation method is of limited value in predicting reliability growth before testing has started. Clearly, such a prediction capability is desirable in order to determine whether reliability growth testing should be conducted for a specific system development program, i.e., if it is cost-effective, and, if so, the extent of testing that is necessary to achieve program objectives. All that really is known today are the boundaries of the parameters (e.g., $\alpha = 0.1$ to 0.7 for the Duane model reliability growth rate), that the reliability growth rate is somehow related to the "aggressiveness" of the reliability growth program, and some unsubstantiated "rules-of-thumb" for establishing the initial conditions of the growth models.

1.2 OBJECTIVES

The purpose of the Reliability Growth Prediction Study was to develop a procedure for predicting the reliability growth which could be expected in an equipment development program based on the characteristics of the equipment undergoing reliability growth testing and the attributes of the development program. As the study proceeded, this objective focused on assessing the applicability of currently accepted reliability growth <u>measurement</u> models (i.e., Duane, AMSAA) to the reliability growth <u>prediction</u> problem. Additional objectives included the development of guidelines for (1) identifying what type of electronic equipment programs are most amenable to imposition of the reliability growth philosophy and (2) specifying reliability growth in equipment development contracts.

1.3 APPROACH

To accomplish the above objectives, TASC first conducted an extensive literature review of both candidate reliability growth models and documented case histories. A survey of electronic equipment development programs which imposed reliability growth testing was conducted. For these programs, a data base was developed which included for each program:

- Characterization of the equipment
- Characterization of the reliability growth program (environmental profile for testing, test length, etc.)
- Failure/time histories observed in the reliability growth test
- Identification of all failure modes surfaced in the test and classification as design-related, manufacturing process-related, or random
- All corrective actions identified as a result of the growth test.

The data base encompassed nine different electronic systems, each developed by a different contractor, and 30 equipment items (Line Replaceable Units). All were avionics development programs. These data were organized into a data base structure suitable for statistical analysis.

The first step of the analysis was to examine the failure/time histories using statistical methods. These analyses included the following:

- Fitting of the candidate growth models to the data using least-squares techniques
- Testing the goodness-of-fit of each model to each data set
- Formulation of alternative models relating reliability growth model parameters to equipment characteristics and program attributes
- Multiple linear regressions to identify the best combination of predictive factors and tests of statistical significance.

These statistical analyses of the failure/time histories were supplemented by less rigorous examinations of the failure mode, failure classification and corrective action data which drew upon the insights of individuals involved in the development of the equipment and the conduct of the reliability growth test. These two forms of analysis provided the basis for an assessment of the true nature of the reliability growth achieved in each program and the equipment characteristics and program attributes which appeared to be most strongly related to the achieved growth.

The analyses and assessments above were the basis for developing a procedure for predicting achievable reliability growth. Guidelines for identifying equipment development programs most suitable for reliability growth testing were prepared which illustrate how achievable reliable growth and the cost of conducting a growth program can be balanced against downstream life-cycle cost savings. In the course of the development program survey, a number of "lessons-learned" were accumulated relative to how future reliability growth testing could be improved to better specify and assess reliability growth. These lessons-learned were translated into a set of guidelines pertaining to how the specification of reliability growth in equipment development programs should be structured.

1.4 OVERVIEW OF REPORT

The remainder of this report is organized as follows. Chapter 2 reviews reliability growth models and their applicability to reliability growth prediction. Chapter 3 describes the results of the development program survey and the data base constructed from this survey. The analysis of the reliability growth data is presented in Chapter 4. Chapter 5 describes a procedure for predicting reliability growth in terms of equipment characteristics and program attributes. Guidelines for selecting equipments which are promising candidates for reliability growth testing and for specifying reliability growth are presented in Chapter 6. A summary of the study findings is contained in Chapter 7.

2. RELIABILITY GROWTH MODELS

The definition of reliability growth for purposes of this study is the positive improvement in a reliability parameter over a period of time. A generally accepted metric of reliability is Mean Time Between Failure (MTBF), or its inverse, failure rate. A reliability growth model is an analytic relationship describing MTBF as a function of time and, possibly, other physical and/or empirical parameters. This relationship can be continuous (i.e., progressive reliability improvement as time accumulates) or discontinuous (e.g., characterized by discrete incremental improvements at specific points in time), or a combination of both. As some form of reliability growth model must underly a prediction methodology, a survey was performed to identify candidate models and their suitability to prediction. Initially, emphasis was placed on the Duane and the Army Material Systems Analysis Activity (AMSAA) models, because these are the most widely recognized and commonly used models. Some difficulties in applying these two models to the prediction problem resulted in the exploration of alternative models and techniques, in particular, the IBM model.

2.1 DUANE MODEL

J.T. Duane (Ref. 1), in 1964, discovered that an empirical linear relationship existed between the logarithm of cumulative failure rate and the logarithm of cumulative test time for various equipments that had been through engineering development. Codier (Ref. 2), in 1968, translated Duane's postulate into one describing the association of the logarithm of cumulative MTBF to the logarithm of cumulative test time. Using this approach, the value of the slope of the line describing this interdependence was called the growth rate α. Higher values of α were considered representative of rapid growth while lower ones were indicative of poorer development efforts.

When using the Duane approach, cumulative MTBF is plotted vs cumulative test time on log-log paper and, if the data appear linear, the postulate may apply. The instantaneous MTBF is derived as the cumulative MTBF divided by $(1-\alpha)$ as shown in Section 2.1.1. Figure 2.1-1 is a Duane plot for an airborne radar showing cumulative MTBF versus test time and the instantaneous MTBF line, which runs parallel to it.

A-25951

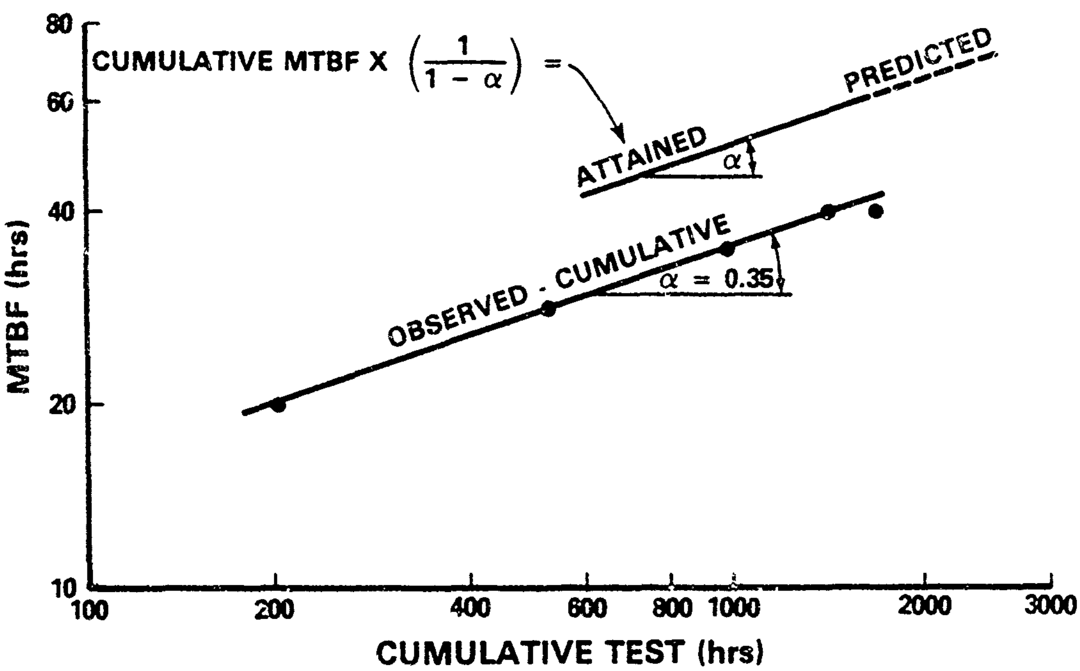

Figure 2.1-1 Duane Plot for Reliability Growth of an Airborne Radar

2.1.1 Mathematical Description

Underlying the Duane model is the hypothesis that as long as reliability improvement efforts continue, the empirical relationship in Eq. 2.1-1 describes the change in cumulative failure rate over time.

$$\lambda_{CUM} = KT^{-\alpha} \tag{2.1-1}$$

where

λ_{CUM} = cumulative failure rate

K = some constant which Duane felt "will depend on equipment complexity, design margin, and design objective for reliability"

T = cumulative operating or test time

α = parameter describing rate of change in MTBF in some broad sense.

Inverting Eq. 2.1-1 yields

$$MTBF_C = \frac{1}{K} T^{\alpha} \tag{2.1-2}$$

where

$MTBF_C$ = cumulative mean time between failure (MTBF).

The following two relationships also hold when failure data are plotted

$$\lambda_{CUM} = \frac{N}{T} \tag{2.1-3}$$

$$MTBF_C = \frac{T}{N} \tag{2.1-4}$$

where

N = number of failures observed in T hours of operation or test

Using Eq. 2.1-3, the current or instantaneous MTBF at time T, denoted by $MTBF_I$, can be derived as follows:

$$\frac{N}{T} = \lambda_{CUM}$$

$$N = T\lambda_{CUM}$$

$$N = TKT^{-\alpha} \quad \text{(from Eq. 2.1-')}$$

$$N = KT^{1-\alpha}$$

$$\frac{dN}{dT} = (1-\alpha)\ KT^{-\alpha}$$

$$\lambda(T) = (1-\alpha)\ KT^{-\alpha} \qquad \text{or}$$

$$MTBF_I = \frac{MTBF_C}{(1-\alpha)} \qquad (2.1\text{-}5)$$

The straight line nature of the logarithm of cumulative MTBF when plotted against cumulative operating time is apparent by taking logarithms of both sides of Eq. 2.1-2.

$$\log MTBF_C = \log \frac{1}{K} + \alpha \log T \qquad (2.1\text{-}6)$$

2.1.2 Applicability to Prediction

The Duane model is a two parameter model, the two parameters being α and K. Therefore, to use this model as a basis for predicting the reliability growth which could be expected in an equipment development program, procedures must be defined for estimating these two parameters as a function of equipment characteristics and program attributes. It is important to note that each of these parameters is empirical in nature, i.e., while they can be estimated for a given data set using curve-fitting methods, there exists no underlying theory for the Duane model which would provide a basis for a priori estimation of α and K.

There does exist a considerable body of historical data indicating observed values of α in prior programs. This experience has at least established boundaries for α, namely that it ranges from a minimum value of 0.1 to a maximum value of 0.7. However, other than the observation/assertion that where it falls in this range is somehow related to the "aggressiveness" of the contractor's reliability improvement program, there is no basis for estimating α in advance. Instead, the only apparent groundrule is to base an advanced estimate of α on prior corporate experience with similar equipments.

The parameter K is even more ambiguous. A cursory examination of Eq. 2.1-6 reveals that 1/K is something of an intercept point with respect to the log-log plot of cumulative MTBF versus time. In particular, at T = 1 hour, $MTBF_C = 1/K$. Hence, K could be interpreted as the inverse of the cumulative MTBF after one hour of operation. This is a meaningless quantity for prediction purposes prior to testing; there is no reason to postulate that the MTBF at one hour is any different from the MTBF at two hours although the Duane model implies otherwise.

Hence, a principal difficulty in using the Duane model for prediction is where to initialize the growth curve. In most applications of the Duane model, the initialization point is determined afterward. That is, the failure versus time data are laid down on a log-log scale and the initialization point is self-determined based on where the best-fitting line falls.

One generally applied "rule-of-thumb" is that the ordinate (MTBF) should be initialized at 10% of the MIL-HDBK-217 predicted MTBF. An open question is what is the abscissa value (initial time) corresponding to this initial MTBF. Clearly, it is not zero as Eq. 2.1-2 is undefined at t = 0. An examination of Duane plots of actual data in the literature (Refs. 2-4) reveals initialization points ranging from t = 1 hour to t = 100 hours. This uncertainty is extremely significant when using the Duane model for predictions as the following simple example will demonstrate.

Suppose that a system with a predicted MTBF of 100 hours is subjected to a reliability improvement program during which 2000 hours of testing will be performed. Suppose, also, that the initial MTBF is 10% of the predicted MTBF ($MTBF_I$ = 10 hours) and that it is somehow known that a reliability growth rate (α) of 0.5 can be achieved during the test. Then if it is assumed that the initial time is one hour, the MTBF at the conclusion of testing is

$$\begin{aligned} MTBF_{2000} &= 10 \times (2000/1)^{0.5} \\ &= 447 \text{ hours} \end{aligned}$$

However, if the time is initialized at 100 hours, then

$$\begin{aligned} MTBF_{2000} &= 10 \times (2000/100)^{0.5} \\ &= 45 \text{ hours} \end{aligned}$$

Clearly, the Duane model is highly volatile relative to this initialization assumption, an undesirable characteristic for a predictive model.

The above discussion is not meant to completely disparage the Duane model, but only to point out that to use it for prediction, it is necessary to have an initial set of time and failure data to initialize the model and estimate the parameters. Once initialized, the model has proven to be very useful for predicting subsequent reliability growth. However, as illustrated in Section 4.1, it is quite difficult to estimate the parameters in the absence of such initial data, i.e., based solely on equipment characteristics and program attributes.

2.2 AMSAA MODEL

Crow (Ref. 5) approached the reliability growth problem by stating that, within a particular phase of a development program, the failure rates would change over time as prototype items were redesigned or reworked because of problems uncovered during test. An assumption is that each test phase possesses a different level of growth, depending on the test and fix philosophy during that test. Crow showed that if this were the case, conditions would exist whereby the phenomena could be described by a nonhomogeneous Poisson process.

2.2.1 Mathematical Description

Within a particular test phase design modifications would occur at cumulative test times S_1, S_2,...S_i, as shown in Fig. 2.2-1, with i=4. It is assumed that, even though there may be more than one prototype on test, the basic configurations of the equipment are the same, therefore the failure rate between modifications is constant. Let λ_i represent the constant failure rate during the i^{th} time period $[S_{i-1}, S_i]$ between modifications, as in Fig. 2.2-2. Based on the constant failure rate assumption, the number of failures, N_i, has the Poisson distribution with mean $\lambda_i(S_i-S_{i-1})$. That is,

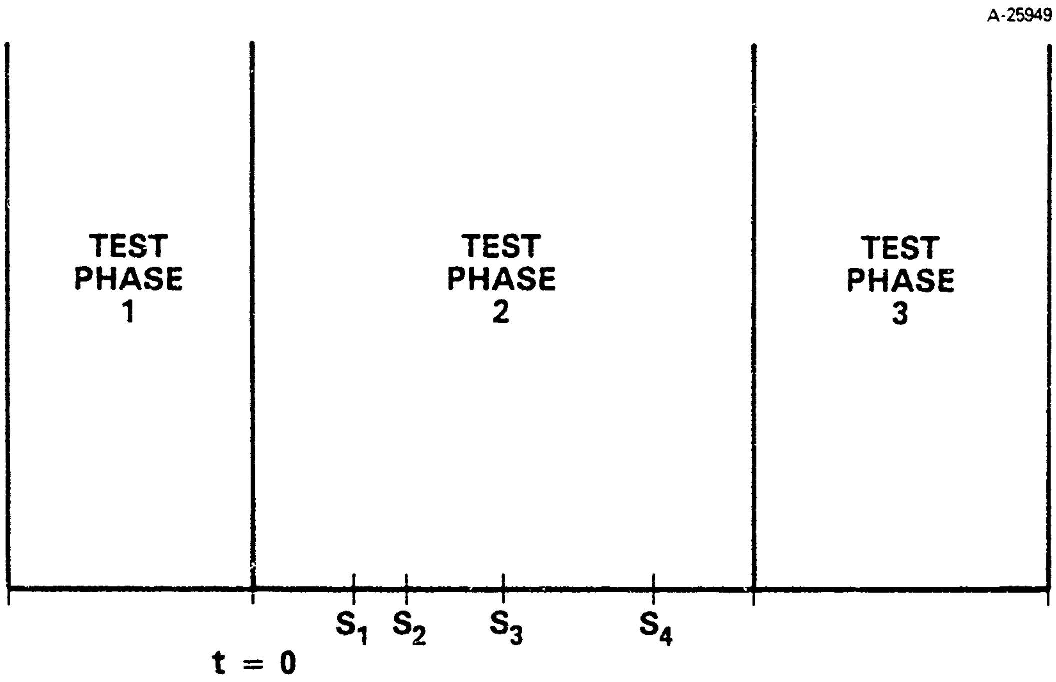

Figure 2.2-1 Phase-by-Phase Reliability Growth

$$\text{Prob}\ (N_i = n) = \frac{[\lambda_i(S_i - S_{i-1})]^n\ e^{-\lambda_i(S_i - S_{i-1})}}{n!} \qquad (2.2\text{-}1)$$

Let N(t) be the total number of system failures by time t. When the failure rate is constant, N(t) follows a homogeneous Poisson process with mean λt. When the failure rate changes with time then, under certain conditions, N(t) is said to follow a nonhomogeneous Poisson process. In this particular situation, N(t) would follow such a process with mean value function

$$\theta(t) = \int_0^t \rho(y)dy \qquad (2.2\text{-}2)$$

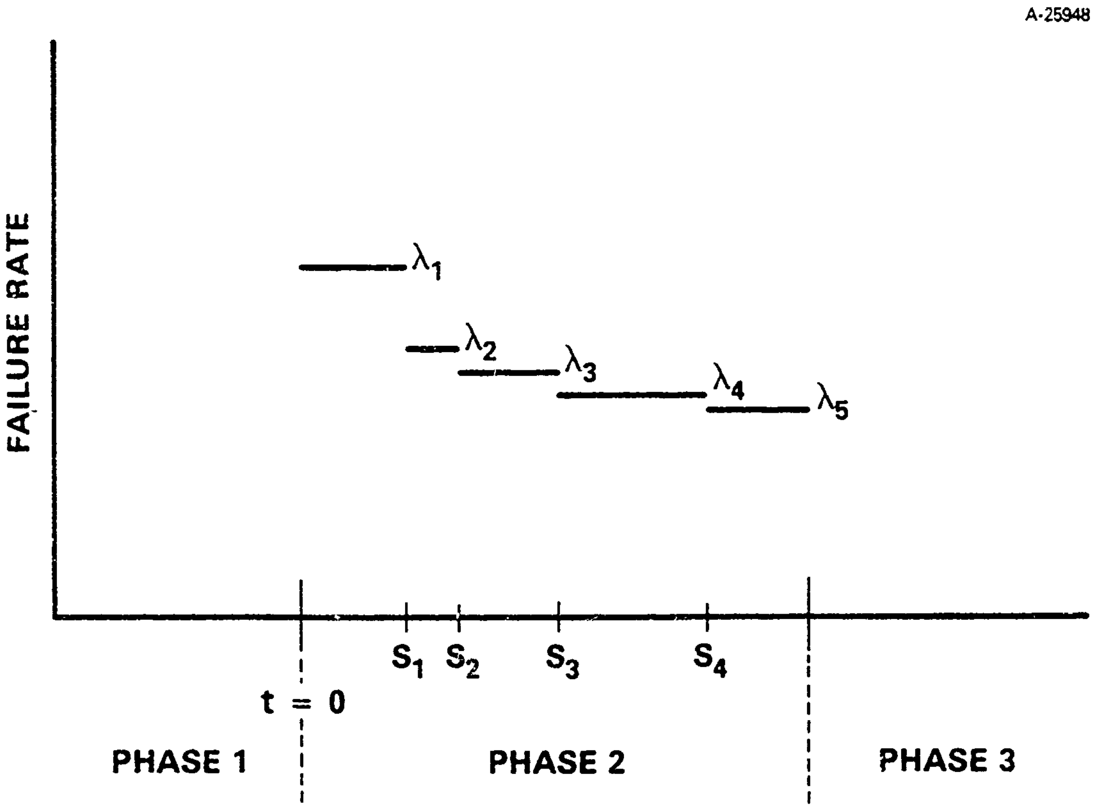

Figure 2.2-2 Failure Rates Between Modifications

where

$$\rho(y) = \lambda_i, \; y_i \varepsilon [S_{i-1}, S_i).$$

Therefore, for any t,

$$\mathrm{Prob}[N(t)=n] = \frac{[\theta(t)]^n e^{-\theta(t)}}{n!} \qquad (2.2\text{-}3)$$

where n = 0, 1, 2,...

The AMSAA model assumes that $\rho(t)$ may be approximated by the parametric form $\rho(t) = \lambda\beta t^{\beta-1}$ (Fig. 2.2-3), $t>0$, $\lambda>0$, $\beta>0$, which is recognized as a Weibull failure rate function.

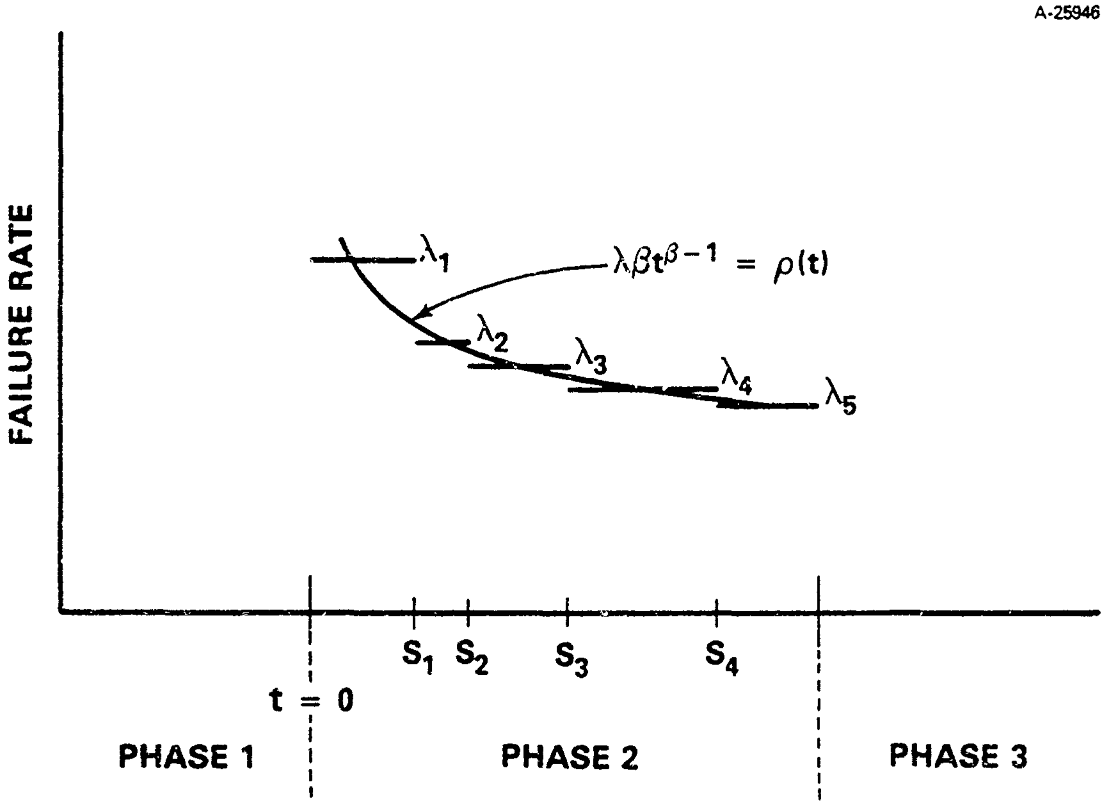

Figure 2.2-3   Parametric Approximation to Failure Rates Between Modifications

This implies that the mean number of failures by time t is $\theta(t)=\lambda t^{\beta}$. The function $(\rho(t))^{-1} = (\lambda\beta t^{\beta-1})^{-1} = m(t)$ represents the current or instantaneous MTBF of the system at time t.

2.2.2 Applicability to Prediction

The Duane and AMSAA approach are related in that they make use of the underlying observed linear relationship between the logarithm of cumulative MTBF and that of cumulative test time, but the Duane is empirical only and does not provide a capability to test whether the change in MTBF observed over time is significantly different from what might be seen due to random error, whereas the AMSAA model allows for such assessments.

The AMSAA approach allows for development of hypothesis testing procedures to determine growth presence in the data ($\beta<1$ implying growth in MTBF, $\beta=1$ implying constant MTBF, $\beta>1$ implying wearout). Confidence bounds on the parameter β can also then be calculated.

However, relative to applying the AMSAA model for prediction, the same problems exist as with the Duane model (see Section 2.1.2). That is, while it is possible to estimate the parameters λ and β for a given set of failure-time data using statistical procedures, it is difficult to relate these parameters to equipment characteristics or program attributes a priori in the absence of such data.

2.3 IBM MODEL

The IBM model (Ref. 6) is based on the assumption that at any point in time there are two types of equipment failures:

- Random failures which occur at a constant failure rate, λ_o
- Non-random design, manufacturing and workmanship defects.

It is further assumed that the random failure rate component is always present at the same level in the equipment and cannot be reduced through testing and corrective action. The non-random component, on the other hand, is progressively reduced as testing is conducted, the defects are surfaced, and corrective actions identified.

2.3.1 Mathematical Description

The IBM model assumes that the number of non-random defects present in the system at the start of testing is fixed, but unknown. If N(t) denotes the number of non-random defects remaining after t hours of testing, it is assumed that the rate of change of N(t) with respect to test time is proportional to the number of remaining defects, that is

$$\frac{dN(t)}{dt} = K_2 \, N(t) \qquad (2.3\text{-}1)$$

Thus,

$$N(t) = e^{-K_2 t + C} \qquad (2.3\text{-}2)$$

The number of defects at time zero is a constant, namely

$$N(0) = e^{C} \qquad (2.3\text{-}3)$$

which is denoted by K_1.

It follows that if V(t) is defined to be the expected cumulative number of failures (random and non-random) after t hours of testing, then

$$V(t) = \lambda_o t + K_1(1 - e^{-K_2 t}) \qquad (2.3\text{-}4)$$

The overall failure rate, $\lambda(t)$, is then given by

$$\lambda(t) = \frac{dV(t)}{dt} = \lambda_o + K_1 K_2 e^{-K_2 t} \qquad (2.3\text{-}5)$$

Thus as t grows large, the non-random defect component becomes small and $\lambda(t)$ approaches λ_o as expected.

2.3.2 Applicability to Prediction

Because of the nature of the relationship in Eq. 2.3-4, the IBM model is not as conducive to the plotting of failure data as are the Duane and AMSAA models. However, by manipulating the terms of Eq. 2.3-5 and taking logarithms, the following relationship can be derived:

$$\ln [\lambda(t) - \lambda_o] = \ln K_1 + \ln K_2 - K_2 t \qquad (2.3\text{-}6)$$

Based on this expression, the non-random component of the instantaneous failure rate would plot as a straight line on a log-linear scale. This means that if one possesses a set of failure data for which it is possible to discriminate random failures and non-random defects (i.e., as a consequence of the failure analyses), then the parameters λ_o, K_1 and K_2 could be estimated by the following plotting procedure:

1. Group the non-random defects into discrete intervals of time $(0, t_1)$, (t_1, t_2), ...
2. Estimate the instantaneous non-random failure rate for each time interval as $\lambda_i = N_i/(t_i - t_{i-1})$, where N_i is the number of non-random failures observed in time interval i
3. Plot the estimates of λ_i versus time on a log-linear scale and fit a straight line to the data points
4. The negative slope of the line is the estimate of K_2
5. The point at which the line intercepts the ordinate, say y_o, is equal to K_1K_2. Thus $K_1 = y_o/K_2$

6. Estimate λ_o as M/t where M is the total number of random failures observed in test time t.

Of course, if one could not discriminate between random and non-random failures in the data set, then, in theory, the three parameters of the model could still be estimated using iterative procedures. However, a very large data set -- larger than usually would be generated is development testing -- would be required to estimate the parameters with fidelity.

Although it has not been as widely utilized, nor its validity clearly established, the IBM model has distinct advantages as a reliability growth prediction tool relative to the Duane and AMSAA models. These characteristics are as follows:

- It distinguishes between random and non-random failures and thus permits consideration of corrective actions (the only way by which real reliability growth is achieved) in the data analysis
- Its parameters are not completely empirical; i.e., they have a physical significance and it is thus possible to formulate hypotheses relating the parameters to equipment characteristics and program attributes.

To illustrate the latter point, consider the following plausible conjectures relative to estimating the three parameters a priori, before any testing is performed:

- λ_o, being the random failure rate component, is equivalent (or proportional) to the predicted failure rate of the system per MIL-HDBK-217

- K_1, the number of non-random defects initially resident in the equipment, is proportional to λ_o but scaled down to account for equipment maturity and other reliability program elements imposed to eliminate or minimize defects

- K_2, the rate at which defects are surfaced during testing, is related to the conditions (e.g., time compression, environmental stresses imposed) of the test.

The major point here is that if the IBM model can be demonstrated to validly represent the reliability growth process, then procedures for estimating its parameters for prediction purposes can be based on a combination of statistical methods and common-sense engineering reasoning, rather than statistical techniques alone.

2.4 OTHER RELIABILITY GROWTH MODELS

There are many, many other forms of mathematical models for reliability growth that have been proposed in the literature. These include the whole family of discrete growth models, which treat reliability growth in the context of a series of trials. One example is the Auto-Regressive-Integrated-Moving-Average (ARIMA) model developed by Singpurwalla (Ref. 7). The ARIMA model is a time series approach to modeling the change in MTBF over time. It essentially looks at the sequence or series of MTBF readings over time and derives a statistical relationship between the MTBF at a given point in the sequence and the previous known MTBFs. The model can be used to track the MTBF over time and to predict the $(k+1)^{th}$ MTBF given the known previous MTBFs. Statistical algorithms are available which will automatically calculate the values of the time series parameters and thus completely specify the ARIMA model for a given sequence of MTBFs.

Other discrete growth models are described in Appendix B of MIL-HDBK-189. While some analyses based on the ARIMA model were conducted in the course of this study, it was found that it, and the other discrete growth models, were not generally suitable for the type of failure data generated in the equipment development programs included in the survey. In particular, the data were representative of a single test-analyze-fix program for each equipment item, as opposed to a series of trials; hence, the continuous growth models were deemed more appropriate.

MIL-HDBK-189 also describes various other continuous growth type models which were not evaluated in this study. A previous RADC reliability growth study (Ref. 8), conducted by Hughes Aircraft Company, evaluated six of these models by fitting each of them to a significant number of data sets. The major findings from the Hughes study, which are relevant to the current study, are repeated below:

- Although the Duane model was seldom the best fitting model, it almost always fit the data
- The IBM model fit airborne data the best
- Each of the other four models was found to be the best fit to the data for specific combinations of environment, equipment type and aggressiveness of reliability program.

In summary, the study did not identify that any particular model is "best" across the board, but that the Duane model is generally applicable and that the IBM model is a viable candidate for airborne equipment, which is the type of equipment analyzed in the current study.

3. DEVELOPMENT PROGRAM SURVEY

A survey was conducted of historical equipment development programs in which reliability growth testing was performed. The programs were reviewed and analyzed to identify the degree of reliability growth that was achieved during test, as well as the equipment characteristics and program attributes which appeared to enhance reliability growth during development and those which appeared to have a negligible effect. This chapter describes the development program survey in terms of the programs included and the data base constructed to support the analysis. Chapter 4 describes the analyses of the program data.

3.1 LITERATURE SEARCH

Initially, a review of the reliability literature was conducted to identify documented data sources and potential sources for the acquisition of data. This search was effective in identifying equipment programs for which reliability growth testing had been performed and the findings from analysis of the reliability growth data, but unsuccessful in acquiring the raw data generated during the program. Generally, the reliability growth data presented in the literature consist, at best, of Duane plots of failure versus time data resulting from the test. While useful information, such data sets were not judged to be appropriate for inclusion in this study because it was not possible to ascertain what sort of screening, or censoring of the data had occurred, nor was visibility into random versus non-random failures and corrective actions incorporated achievable. Follow-up with the authors did not produce the original data for reasons of corporate sensitivity or loss of the data over time.

Nevertheless, the literature survey was of value in that it revealed various trends of thought relative to reliability growth testing, the degree of acceptance of different growth models, as well as some of the pitfalls involved in analyzing reliability growth data.

3.2 PROGRAMS SURVEYED

The programs included in the survey were programs in which TASC, as an independent advisor to the program office, acquired all relevant data generated in the course of the reliability growth program, and conducted analyses of the data to support program decisions. These programs are summarized below.

3.2.1 Omega Navigation Set

The Omega Navigation Set program was conducted to develop and acquire radio navigation equipment to replace the outdated LORAN-A system on the Air Force's C-130 series of aircraft. It was a pioneering program for its time in that it was one of the first programs to emphasize life-cycle cost as a major decision criterion and in its use of the Reliability Improvement Warranty.

The Omega program was also unique in its imposition of reliability improvement testing while still in a competitive phase. In particular, pre-production prototypes from each of three manufacturers competing for the production contract were subjected to a Government-conducted Combined Stress Reliability

Test (CSRT) performed in an environmental test chamber at the Air Force Flight Dynamics Laboratory. Each set was subjected to stresses in this test designed to surface failure modes and enable the respective manufacturer to identify and incorporate corrective action. The set consisted of three Line Replaceable Units (LRUs)

- Receiver-Processor Unit
- Control-Display Unit
- Antenna Couple-Unit.

TASC established and maintained a data base of all failures observed in test, including the time of failure, equipment configuration at failure and subsequent failure resolution.

3.2.2 Common Strategic Doppler

The Common Strategic Doppler (CSD) program developed a Doppler navigation capability for the Air Force's KC-135 and B-52 aircraft series. The CSD consisted of two LRUs:

- Doppler Velocity Sensor
- Ground Speed Drift Indicator.

As for the Omega program, reliability improvement testing was conducted during the competitive phase. Two sets from each of two manufacturers were subjected to a Combined Environment Reliability Test (CERT), again performed in the Flight Dynamics Laboratory environmental chamber.

The CERT was somewhat different from the Omega CSRT in that rather than being an accelerated test (i.e.\ more severe than the mission environment), it was designed to simulate the mission environment. Thus, the observed reliability in test would be representative of that to be expected in the field. However, it could also be considered to be a reliability growth test in that the two contractors had technical representatives on-site to witness the test and return failed units for analysis and corrective action. A total of 1700 hours was accumulated for one contractor and 2500 hours for the other. TASC maintained a comprehensive data base of all test events and performed independent reliability and reliability growth assessments.

3.2.3 Offensive Avionics System

The Offensive Avionics System (OAS) was developed to replace the aging Bomb Navigation System in the B-52 aircraft. The overall OAS program actually entailed seven equipment development programs, each with a different contractor. The programs were for the following equipments:

- Attitude Heading Reference System
- Electronic Altimeter Set
- Control Display Set
- Radar Set Group
- Data Transfer Set
- Avionics Control Unit
- Interface and Control Units.

A total of 23 LRUs comprised the OAS.

The program was conducted under a compressed schedule which included concurrency of Full Scale Engineering Development and production. Because of the reliability risk inherent in concurrency, an Independent Review Team evaluated the program and recommended reliability improvement testing. In response, the program office embarked on a Test-Analyze-Fix (TAF) program for the OAS. The testing was conducted in environmental facilities at each of the respective contractor plants. In most cases, two test articles for each LRU were included in the TAF program; in a few cases four were tested. The target was to accumulate 2400 hours of test on each equipment; the actual test times were close to this target, some being higher and some lower.

TASC, under contract to the Program Office, established and maintained an OAS TAF data base and performed an independent assessment of the reliability growth achieved for each equipment (Ref. 9). The OAS TAF data base is the most recent and comprehensive of the programs surveyed. To illustrate, Table 3.2-1 indicates the depth of data collected. Therefore, in the subsequent analyses this set of data was used as the principal source.

3.3 DATA BASE ORGANIZATION

The data acquired from the development programs were organized into a data base to support the analyses. The data were assembled by system (manufacturer) and, within each system, by Line Replaceable Unit. A total of twelve systems and 30 LRUs were included as listed in Table 3.3-1. Two sets of data were constructed for each system and LRU:

- Equipment characteristics and program attributes
- Reliability growth test data (failure times, failure classification, corrective action).

TABLE 3.2-1

OAS/TAF DATA ELEMENTS

DATA ELEMENT	DESCRIPTION
LRU	Identification of UUT
Serial Number	Serial number of operating/failed LRU
Operating Time	Total on-hours of the LRU and of the LRU type (especially at time of failure)
Cycle Time	Time into the cycle when a discrepancy or failure is noted
Test Conditions	Environmental conditions at the time of a discrepancy or failure
Discrepancy	Observed symptom, test operator actions and BIT indications
Maintenance Activity	Description of all maintenance actions leading to on-site repair of the LRU, identify removal/replacement of any subassembly (SRU), module, and/or component, include nomenclature (item name) and serial number of replacement item Also identify any adjustments performed and relationship to failed item.
Malfunction	Identify the function or parameter within the LRU that was performed incorrectly or not-at-all as a direct result of the item failure, and its relation to the observed symptoms If the failure was discovered during performance verification tests and not during functional checkout/operation, identify why BIT did not detect the failure
Maintenance Time	The following maintenance times shall be recorded. a) on-site time when bench checkout (repair) was initiated, b) on-site time when repair was completed, c) on-site time when performance verification tests were successfully completed.
Contractor Representative	Name and signature of contractor representative submitted report.
Failure Analysis	As a minimum, the following information shall be provided: <u>Module/Circuit/Component Failure</u> - The faulty module/circuit component shall be identified to the extent corresponding to depot-level maintenance and to the degree that corrective action is proposed/performed (i e , repair of part vs circuit/module redesign) <u>Cause of Failure</u> - Identify the relationship between the failure and the environmental, operating conditions at the time of failure; include pertinent variations in other circuits which could have caused the failure in the failed circuit. Provide sufficient detail to substantiate the failure mechanism (i.e., overdrive, overvoltage, overtemperature/excessive heat dissipation - as relates to environmental conditions) <u>Corrective Action</u> - Identify the performed and/or proposed corrective action to avoid recurrence of the failure. Identify how the modification is to enhance the reliability design of the failed item.

3.3.1 Equipment Characteristics and Program Attributes

Electronic equipments, when entering full scale engineering development, possess varying levels of design maturity, i.e., some parts or assemblies represent proven technology, while others represent state-of-the-art advancement. The device itself may be highly complex and intended to perform a complicated

TABLE 3.3-1

SYSTEMS AND LRUS COMPRISING DATA BASE

SYSTEM/LRU	USAGE
A*	Provide aircraft attitude and heading data
A1	Attitude data supply
A2	Data conversion
B	Provides measure of absolute altitude from 0-5000 feet
B1	Transmitter/Receiver
B2	Data converter to absolute height
C	Provide information on weapon status
C1	Input transformation for video display
C2	Records and produces mission data on 35-millimeter film
C3	Digital computer, command, query, record weapon information
C4	CRT display
C5	Keyboard for data entry into system
D	Radar
D1	Switch assembly
D2	Receiver/Transmitter
E	Receives and transmits data, stores data on mag tapes
E1	Selects processor, programs for load
E2	Tape transport
F	Two processors, aircraft navigation computational capability
G	Provides necessary data conversions
G1	Main interface to radars
G2	Interface unit
G3	Interface to weapon systems
G4	Interface to computational subsystems
G5	Controls for weapon launch
G6	Control computer
G7	Controls critical navigation
G8	Provides signal interface between system and weapons
H	Navigation equipment
H1	Receiver processor unit
H2	Control display unit
I	Navigation equipment
I1	Receiver processor unit
I2	Control display unit
J	Navigation Equipment
J1	Receiver processor unit
J2	Control display unit
K	Doppler navigation
L	Doppler navigation

*Alphabetic characters, e.g. A, refer to systems.
Alphabetic characters followed by numbers refer to LRUs.
A1 means LRU 1 of system A.

array of mission functions, or somewhat simple in design with a minimal set of performance requirements. The usage time period may be brief, but with environmental variability, or long, but under relatively benign operational surroundings. It may be powered in a cyclical manner, or run continuously over constant time periods. Any combination of these characteristics, i.e., maturity, complexity, environment, and usage are possible. Each particular mixture will influence, among other things, testing philosophy, failure isolation and identification, and design change implementation. This implies that reliability growth will also be affected.

Whatever the characteristics of the equipment, system development from a cost/engineering/test point of view must be managed to facilitate eventual entrance into the military inventory. Resources (time, money, manpower) are allocated for testing environments, total test duration, and parts/subassembly/assembly screening. These allocations will affect the type of tests and the time involved in corrective action implementation, and, hence reliability growth. Funding levels, testing types and parts screening are labeled as program attributes.

Information on each system/LRU was organized by characteristics, attributes, and specifications. These summaries, called profiles, were arranged as follows:

1. Characteristics

   A. Usage - intended use of equipment

   B. Environment - using MIL-STD 217D definitions, i.e., aircraft uninhabited fighter, etc.

   C. Complexity - measured by looking at no. of shop replaceable units (SRUs) and equipment cost

D. Maturity - reflected by the % of off-the-shelf technology (whether parts or designs) used by the system

II. Attributes

A. Type Test - identifies the type of test conducted, whether test-analyze-fix (TAF), operational, field development, etc, along with testing environments, levels, temperatures, and conditions

B. Stress Environment - specifically labels test environment as one intended to stimulate failure or simulate operational conditions

C. Reliability Funding Level - percentage of total contract award dollars clearly identified for reliability test and evaluation

D. Parts Control - quality assurance program in place, if any, prior to start of test

III. Specifications

A. MIL-STD 785B tasks imposed contractually or, if program was prior to MIL-STD 785B, reliability-centered contractual requirements

B. MIL-STD 217 prediction

C. Growth test planning - usage of MIL-STD-1635 or MIL-HDBK-189

The following is representative of such a profile.

Program C

I. Characteristics

A. Usage - Perform data insertion, control supervision, and recording of information

B. Environment - Aircraft inhabited bomber

C. Complexity -

	#SRU's	COST
System	84	255K
C1	24	114K
C2	10	47K
C3	28	64K
C4	18	24K
C5	4	6K

D. Maturity -

	% Off Shelf Components
System	26-50
C1	10-25
C2	51-75
C3	10-25
C4	51-75
C5	51-75

II. Attributes

A. Type Test - TAF (contractually required for 2400 hours)
- Cycles defined as follows:

Equipment	Temp. (°C)	Cycle on Time (hrs)	Max Vib. g/rms	Vib. on Time (min.)
C1	-54, 70	4.0	6.0	340.0
C2	-54, 55	4.0	6.0	340.0
C3	-54, 71	4.0	6.0	340.0
C4	-54, 55	4.0	6.0	440.0
C5	-54, 55	4.0	6.0	440.0

- Chamber test

B. Stress Environment - Stimulation

C. Reliability Funding Level - >1% of contract award

D. Quality Assurance Program in conjunction with -

MIL-Q-9858A	(Quality Program Reqmt's)
MIL-C-45662A	(Calibration System Reqmt's)
MIL-I-6870D	(Nondestructive Inspection Program Requirements for Aircraft and Missile Materials and Parts)
MIL-STD-1535A	(Supplier Quality Program Requirements)
MIL-STD-1520A	(Corrective Action and Disposition System for Non-Conforming Material)

Quality Assurance Status Report Required

III. Specifications

A. Failure Review and Corrective Action System in place prior to start of test

B. MIL-STD-217-B Prediction (MTBF) -

	Pred.	Reqm't
System	359.0	250.0
C1	1091.0	1000.0
C2	3752.0	3300.0
C3	1480.0	1400.0
C4	6150.0	3800.0
C5	3316.0	2900.0

C. Growth planning done in accordance with MIL-STD-1635

3.3.2 Reliability Growth Test Data

For each system, as well as for each LRU comprising the system, a data set representing the results of the reliability growth test was constructed. An extract from the data base presenting these data sets for two LRUs is provided in Table 3.3-2 in order to illustrate the format. Each row entry corresponds to a discrete test failure event with the following data contained for each failure event:

TABLE 3.3-2

RELIABILITY GROWTH TEST DATA EXTRACT

MANUFACTURER	LRU	CUM ETI	FAILURE COUNT	CUM MTBF	REPORT NO.	FLAG	CAUSE
	G1	4.9	1	4.9	23	1	D
		17.8	2	8.9	4	1	D
		33.4	3	11.1	6	1	D
		75.3	4	18.8	12	1	D
		84.0	5	16.8	13	1	D
		215.0	6	35.8	25	1	D
		219.0	7	31.3	28	1	D
		262.3	8	32.8	36	-	NR
		313.0	9	34.8	41	21	D
		534.0	10	53.4	57	0	R
		722.0	11	65.6	66	10	M
		1503.0	12	125.3	98	0	R
		1866.0	13	143.5	109	21	D
		1977.0	14	141.2	115	0	R
		2502.0	15	166.8	125	10	M
	G2	12.0	1	12.0	2	0	M
		247.5	2	123.8	17	3	D
		333.5	3	111.2	21	-	NR
		349.0	4	87.3	27	3	D
		352.0	5	70.4	31	3	D
		858.0	6	143.4	67	25	M
		1380.0	7	197.1	91	28	M
		1785.0	8	148.1	108	0	R
		2007.0	9	223.0	116	25	M
		2355.0	10	235.5	121	19	M

- Equipment manufacturer (name deliberately omitted)
- LRU (Refer to Table 3.3-1) - The LRU which experienced the failure
- CUM ETI - Cumulative Elapsed Time Indicator reading at the time of the failure event
- FAILURE COUNT - Cumulative number of failure events for the LRU
- CUM MTBF - (CUM ETI)/(FAILURE COUNT)

- REPORT NO. - Identifies the number of the contractor failure report written as a result of the failure event (available in hard copy)
- FLAG - Identifies the failure mechanism as distinct or as identical to one already observed in test
- CAUSE - Identifies the failure as being due to a design defect (D), a manufacturing/workmanship defect (M), non-relevant (NR) or random (R).

The FLAG and CAUSE entries are very important relative to reliability growth assessment. The FLAG permits discrimination of new failure modes from repetitive failures. For example, the first seven failure events for LRU G1 were all due to the same design defect. The CAUSE permits discrimination between random failure (CAUSE = R) and defects which could be eliminated through corrective action (CAUSE = D or M). Of the fifteen failure events observed in testing of LRU G1, one was non-relevant (e.g., due to a chamber problem), seven were due to a single design defect, two were due to a second design defect, two were due to a manufacturing defect and three were random.

The important point to note is that to merely plot the cumulative MTBF versus the cumulative test time as a method for assessing reliability growth (which would indicate a significant MTBF growth of from five to 167 hours) would be to mask out some very relevant information, namely the number of failure modes and correctable defects surfaced.

4. DATA ANALYSIS

4.1 OBSERVATIONS

Before presenting the various analyses performed on the development program survey data base, some less rigorous, but germane, findings will be presented. These observations, which are a result of both TASC's first-hand participation in the reliability growth testing for the programs surveyed and of the data analysis performed, will serve to illustrate some of the peculiarities and potential misinterpretations of reliability growth data.

4.1.1 Time Constraints

It is difficult to argue the point that reliability improvement results only from the identification of defects and the incorporation of corrective actions. In an ideal world, testing would be stopped as soon as a failure is observed and not resumed until the failure is analyzed and a corrective action identified and incorporated into the test article. This, of course, is impractical and not cost-effective. Each of the programs surveyed by TASC was under some sort of pressure to complete testing within a reasonable period of calendar time. Testing couldn't be halted whenever a failure occurred. As a consequence, corrective actions were, at best, incorporated into test articles several hundred operating hours after the failure orginally surfaced, and often, were not incorporated at all but instead were identified for incorporation during production. Hence, the results of corrective actions were only minimally reflected in the test data.

4.1.2 Reliability Improvement or Burn-In?

This issue has been belabored in the literature (Ref. 10, for example). The question is whether reliability growth as measured on test data is a consequence of true reliability improvement of the system, or just reflects the burning in of the serialized units undergoing test. The argument goes that a given serialized unit will have a certain number of bad parts due to random lot selection and as these parts fail and are replaced with better parts, the apparent reliability improves. There is no doubt that this effect is present, to some degree, in any set of data from a reliability growth program performed in FSED on a finite number of test articles.

4.1.3 Cumulative Data

The standard procedure for using the Duane approach to track reliability growth, is to plot cumulative MTBF, (i.e., cumulative time/cumulative failures) on a log-log scale. Working exclusively with cumulative data can be quite misleading when measuring reliability growth. A simple example illustrates this point. Specifically, suppose that the system undergoing test is characterized by having a number of initial defects (i.e., infant mortality), but a constant failure rate after removal of these defects. Five failures are observed in the first 100 hours of testing, and a failure is observed every 100 hours thereafter (to convey the point, randomness in the test data is eliminated). Table 4.1-1 illustrates how the data from 1000 hours of testing would be presented cumulatively.

Clearly, the data imply that reliability growth is taking place, even though the true MTBF is constant at 100 hours after the first 100 test hours. Furthermore, if these data are plotted in a log-log scale, they appear to lie on a straight line with a slope (α), of 0.43.

TABLE 4.1-1
EXAMPLE OF CUMULATIVE DATA

CUMULATIVE HOURS	NUMBER OF FAILURES	CUMULATIVE MTBF (HOURS)
100	5	20
200	6	33
300	7	43
400	8	50
500	9	56
600	10	60
700	11	64
800	12	67
900	13	69
1000	14	71

The point of this example is that it is not necessarily remarkable -- nor a validation of the Duane model -- that failure data tends to appear linear when plotted on a log-log scale. A high incidence of early failures, and the progressive diminishing of their impact through cumulative plotting, may very well be the underlying cause for such behavior, particularly when dealing with small data sets (e.g., less than fifteen failures).

4.2 STATISTICAL ANALYSES OF GROWTH DATA

The first method of investigation utilized in this study was an analysis of the failure-time history data generated in the different reliability growth testing programs. These analyses were based only on the temporal characteristics of the failure data, and did not discriminate failures by type or corrective actions. Specifically, the data set for each equipment consisted of the sequence (T_1,T_1), $(T_2,T_2/2)$, ... $(T_N,T_N/N)$ where T_i is the cumulative test time when the ith failure was observed and T_i/i is the observed cumulative MTBF.

4.2.1 Methodology

Both the Duane and AMSAA models were fit to the data. An International Mathematical and Statistical Library (IMSL) regression analysis subroutine was used to calculate least squares estimates of the growth model parameters, e.g., α and K for the Duane model.

The goodness-of-fit was then assessed. The best fitting model was used to produce a series of predicted cumulative MTBFs, i.e., $\widehat{MTBF}_1$, $\widehat{MTBF}_2$, ... $\widehat{MTBF}_N$, which could then be compared to the observed values $T_1/1$, $T_2/2$, ... T_N/N. Three measures were used to quantify the prediction error. These error measures are given in Table 4.2-1. The criteria used to determine significance of model fit were the Percent Variation Explained (Duane model) and the Cramer Von Mises Statistics (AMSAA model).

The Duane and AMSAA models can be used to predict the instantaneous MTBF at the end of a test interval. In order to measure the error associated with these instantaneous MTBF predictions, the failure time history data were divided into equal width subintervals and the instantaneous Duane and AMSAA predictions computed at the midpoint of the subintervals were compared to the observed MTBF computed over each subinterval. Figure 4.2-1 illustrates this technique. The bar heights correspond to the cumulative MTBF over the corresponding interval of test time. The bar lengths are chosen heuristically as the minimum length required to divide the total test time into subintervals such that each subinterval has at least one failure

TABLE 4.2-1

PREDICTION ERROR MEASURES

VARIABLE	GENERIC NAME	FUNCTIONAL FORM	INTERPRETATION
R_1	Average Absolute % Error	$R_1 = \sum_{t=1}^{N} \frac{\hat{MTBF}_t - MTBF_t}{MTBF_t} \frac{100}{N}$	A model fits well if $R_1 \leq 25$
R_2	Relative Variability	$R_2 = \frac{\sum_{t=1}^{N} (MTBF_t - \hat{MTBF}_t)^2 / N-2}{\sum_{t=1}^{N} (MTBF_t - \overline{MTBF})^2 / N-1}$	Low values of R_2 indicate a good fit.
MSE	Mean Squared Error	$MSE = \sum_{t=1}^{N} (\hat{MTBF}_t - MTBF_t)^2 / N$	Low values of MSE indicate a good fit.

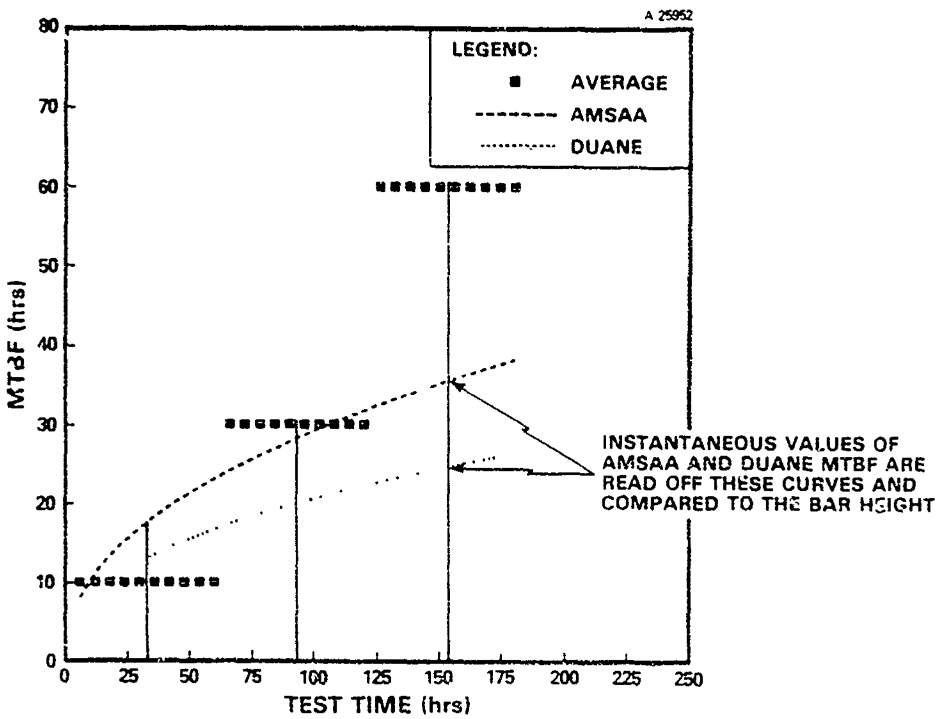

Figure 4.2-1 Estimating Instantaneous Error

(i.e., the cumulative MTBF over each subinterval is not infinite). The error in the instantaneous MTBF prediction for the ith subinterval (i=1,...,M; M=total number of subintervals) is:

$$e_i = \text{(observed MTBF over subinterval)} - \text{(model-calculated MTBF interval midpoint)}$$

The error measures R_1, R_2, and MSE defined in Table 4.2-1 were applied to the instantaneous errors (e_i; i=1,...,M) to assess the overall prediction error and significance of fit.

Next, it was attempted to correlate the estimated growth rates (i.e., α for the Duane model and β for the AMSAA model) to the equipment characteristics and program attributes listed in Section 3.3.1. A series of multiple linear regression analyses were performed using various combinations of the characteristics and attributes as the independent variables and the estimated growth rate as the dependent variable. Both single and multiple variable linear forms were examined, although the multiple variable forms examined were primarily limited to two-variable models. The general form of the two-variable model was:

$$\alpha = P_1X_1 + P_2X_2 \qquad (4.2\text{-}1)$$

where

α = Growth rate for the equipment as estimated by the Duane model fit

X_1, X_2 = Independent variables (characteristics/attributes)

P_1, P_2 = Constants yielding the best relation of α to X_1 and X_2 using a least-squares criterion

An IMSL program was used to accomplish the regression. For each independent variable (or pair of independent variables) tested, the goodness of the regression model was measured by the percentage variation in the growth rate parameter explained by the model, an output of the IMSL program.

4.2.2 Results of Statistical Analysis

Duane Model Fits - The Duane model was fit to all equipments with failure-time histories in the data base. Table 4.2-2 displays the estimates for the Duane intercept (log(1/K)) and slope (α) for each equipment. A key indicator of the regression fit to the Duane relationship between the cumulative MTBF and the cumulative test time is the Duane Percent Variation explained column in Table 4.2-2. The higher the variation explained, the greater the amount of variation in the data that can be accounted for by the fitted Duane model and, hence, the more significant the fit.

The following observations can be drawn from Table 4.2-2:

- After eliminating the data sets with less than five data points, approximately half the data sets display a reasonable fit (% variation > 75%)
- With one exception (Equipment C2), the equipments displaying a negative growth rate have either a small number of data points or do not display a good Duane fit (% variation < 25%)
- For those equipments which both have a significant number of data points and display a good fit, the growth rate ranges from a low of 0.097 to a high of 0.644.

The general conclusion is that, while there are some anomalies, the Duane model provides a reasonably good fit to most of the data sets and the estimated growth rates are within the generally accepted range of 0.1 to 0.7.

TABLE 4.2-2

DUANE PARAMETER ESTIMATES FOR EQUIPMENTS

EQUIPMENT	DUANE GROWTH RATE	DUANE INTERCEPT (LAMBDA)	DUANE % VARIATION EXPLAINED	NO. OF DATA POINTS
A1	.520	2.578	88.6	3
B1	.663	10.500	97.3	3
B2	-2.087	22.388	4.3	7
C1	.340	2.538	78.4	21
C2	-.419	8.300	55.9	20
C3	.486	2.732	74.7	6
C4	.097	4.254	14.4	31
C5	.644	1.679	97.4	5
D2	.374	1.664	94.8	31
E1	.141	2.977	19.4	50
E2	.474	6.546	87.9	59
E3	.528	1.965	97.7	7
G1	.594	.398	98.6	14
G2	.572	1.248	96.1	9
G3	-.253	5.793	7.5	18
G5	-.213	6.546	17.4	9
G6	-4.056	38.018	96.5	3
G7	.540	.151	90.3	31
G8	-.174	6.664	23.5	13
H1	.259	.523	47.0	23
H2	-.024	1.471	.7	21
H3	.368	.855	94.5	11
H4	.251	1.246	55.4	10
I1	.196	.694	31.3	16
I2	.353	.563	48.3	19
I3	.585	-.137	98.2	11
I4	.416	.572	90.7	9
I5	.541	-.016	97.7	4
J1	-.109	2.714	9.6	16
J2	.237	1.109	53.6	13
J3	.286	1.605	77.1	11

<u>AMSAA Model Fits</u> - The AMSAA model was similarly applied to all equipments with failure time histories in the data base. Table 4.2-3 displays estimates for the AMSAA beta and lambda parameters and Fig. 4.2-2 is a histogram of the AMSAA growth rate. These parameters were estimated using the maximum likelihood technique. The measure of fit for the AMSAA

TABLE 4.2-3

AMSAA PARAMETER ESTIMATES EQUIPMENTS

EQUIPMENT	AMSAA GROWTH RATE	AMSAA LAMBDA	CRAMER VON MISES STATISTIC	NO. OF DATA POINTS
A1	.849	.0040	.098	3
B1	.899	.0027	.140	3
B2	5.779	.0000	.265	7
C1	1.049	.0048	.224	21
C2	1.600	.0001	.046	20
C3	1.408	.0001	.235	6
C4	.852	.0189	.249	31
C5	.703	.0100	.080	5
D2	.630	.1669	.105	31
E1	1.060	.0119	.122	50
E2	.794	.0882	.446	59
E3	.790	.0123	.144	7
G1	.412	.5543	.050	14
G2	.639	.0616	.036	9
G3	.640	.1228	.741	18
G4	1.247	.0001	.094	2
G5	.813	.0159	.215	9
G6	5.917	.0000	.073	3
G7	.501	.6235	.268	31
G8	.947	.0055	.173	13
H1	.723	.5686	.530	23
H2	.579	.9931	.537	21
H3	.693	.2774	.039	11
H4	.652	.3347	.204	10
I1	.634	.7940	.667	16
I2	.778	.2940	.353	19
I3	.503	.6956	.059	11
I4	.542	.5209	.096	9
I5	.870	.3404	.090	4
J1	.843	.1744	.253	16
J2	.588	.5344	.194	13
J3	1.115	.0275	.258	11

model is the Cramer Von Mises statistic. Values of the Cramer Von Mises statistic less than 0.28 generally indicate a good fit to the data.

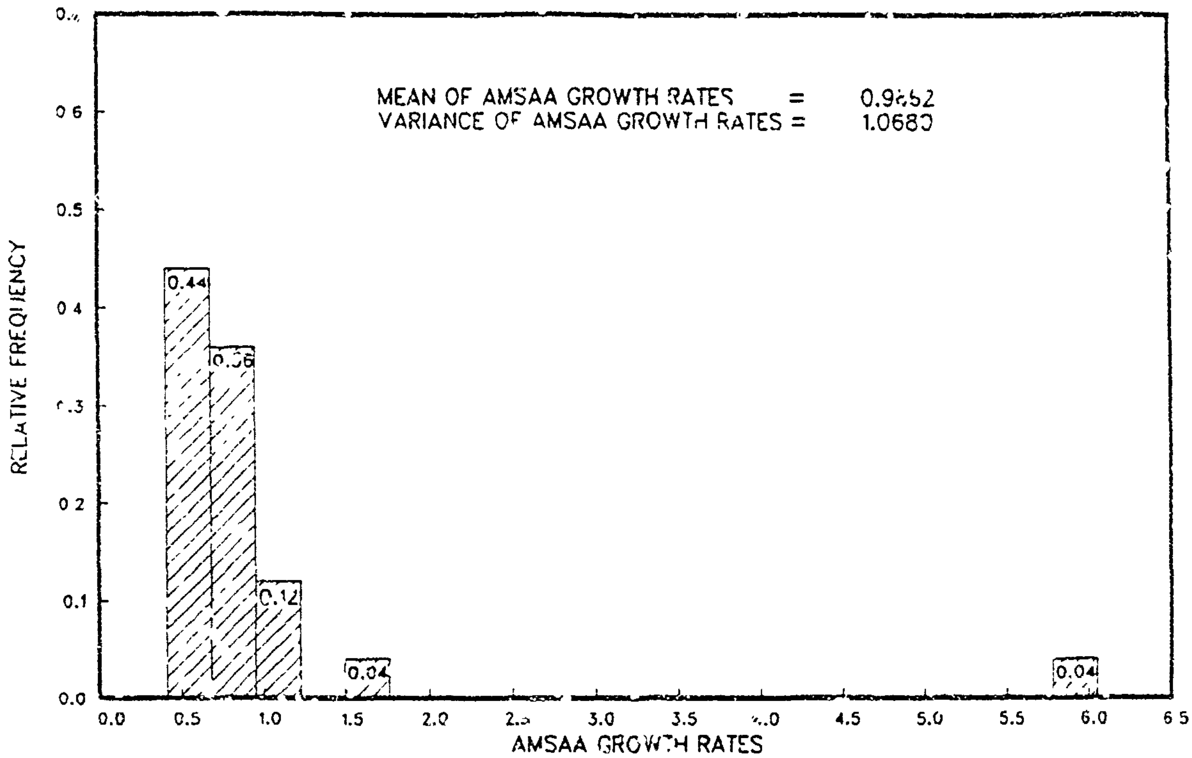

Figure 4.2-2 AMSAA Growth Rates for all Equipments

Based on the Cramer Von Mises statistic as the criterion, a good fit was observed on six equipment data sets, and a marginal fit was observed on eight other equipments. As was the case with the Duane model, some equipments exhibited negative reliability growth ($\beta > 1$), but in all such cases either the number of data points was small or the fit was not significant (or both). For those equipments exhibiting a good fit, the AMSAA growth parameter (β) ranged from 0.579 to 0.794.

Prediction Error Analysis - The accuracy of the Fitted Duane and AMSAA models as predictors of reliability growth was tested using the cumulative and instantaneous error analysis techniques described in Section 4.2.1. The results of the cumulative prediction error analysis, in terms of the three error measures defined in Table 4.2-1, are presented in Table 4.2-4. The equivalent results for the instantaneous

TABLE 4.2-4

CUMULATIVE PREDICTION ERROR ANALYSIS

SYSTEM	DUANE			AMSAA		
	R1	R2	MSE	R1	R2	MSE
B	0.883×10^{1}	0.204×10^{0}	0.123×10^{4}	NA	NA	NA
B2	NF	NF	NF	NA	NA	NA
C	0.135×10^{2}	0.392×10^{0}	0.657×10^{2}	NF	NF	NF
C1	0.181×10^{2}	0.412×10^{0}	0.759×10^{3}	0.495×10^{2}	0.152×10^{1}	0.281×10^{4}
C2	0.120×10^{2}	0.681×10^{0}	0.357×10^{4}	0.187×10^{1}	0.129×10^{1}	0.675×10^{4}
C4	NF	NF	NF	NF	NF	NF
D	0.129×10^{2}	0.142×10^{0}	0.673×10^{2}	0.153×10^{2}	0.222×10^{0}	0.105×10^{3}
D2	0.102×10^{2}	0.800×10^{-1}	0.501×10^{2}	0.130×10^{2}	0.144×10^{0}	0.903×10^{2}
E	NF	NF	NF	NF	NF	NF
E1	NF	NF	NF	0.270×10^{2}	0.127×10^{1}	0.127×10^{3}
E2	0.246×10^{2}	0.439×10^{0}	0.158×10^{3}	NF	NF	NF
E3	0.102×10^{2}	0.135×10^{0}	0.292×10^{4}	0.565×10^{2}	0.412×10^{0}	0.891×10^{4}
F	NF	NF	NF	0.490×10^{2}	0.122×10^{1}	0.258×10^{5}
G	0.159×10^{2}	0.141×10^{0}	0.410×10^{1}	NF	NF	NF
G1	0.108×10^{2}	0.385×10^{-1}	0.115×10^{3}	0.197×10^{2}	0.292×10^{-1}	0.875×10^{2}
G2	0.131×10^{2}	0.923×10^{-1}	0.465×10^{3}	0.373×10^{2}	0.958×10^{-1}	0.482×10^{3}
G3	NF	NF	NF	NF	NF	NF
G5	NF	NF	NF	0.405×10^{2}	0.219×10^{1}	0.584×10^{4}
G7	0.213×10^{2}	0.162×10^{0}	0.372×10^{2}	NF	NF	NF
G8	NF	NF	NF	0.248×10^{2}	0.137×10^{1}	0.429×10^{4}
H1	NF	NF	NF	NF	NF	NF
H2	NF	NF	NF	NF	NF	NF
H3	0.805×10^{1}	0.692×10^{-1}	0.974×10^{0}	0.222×10^{2}	0.328×10^{0}	0.462×10^{1}
H4	0.134×10^{2}	0.841×10^{0}	0.383×10^{1}	0.318×10^{2}	0.238×10^{1}	0.108×10^{2}
I1	NF	NF	NF	NF	NF	NF
I2	NF	NF	NF	NF	NF	NF
I3	0.107×10^{2}	0.566×10^{-1}	0.151×10^{1}	0.240×10^{2}	0.100×10^{0}	0.268×10^{1}
I4	0.132×10^{2}	0.138×10^{0}	0.225×10^{1}	0.289×10^{2}	0.393×10^{0}	0.641×10^{1}
J1	NF	NF	NF	NF	NF	NF
J2	0.168×10^{2}	0.460×10^{0}	0.312×10^{1}	0.314×10^{2}	0.902×10^{0}	0.612×10^{1}
J3	0.126×10^{2}	0.361×10^{0}	0.112×10^{2}	NF	NF	NF

NF indicates no significant fit
NA indicates predictions not computable

prediction error analysis are presented in Table 4.2-5. Both models were reasonably accurate predictors of cumulative MTBF as measured by the R1, R2, and MSE values, with the Duane model being more accurate across the board. The AMSAA model demonstrated better accuracy for predicting instantaneous MTBF.

TABLE 4.2-5

INSTANTANEOUS PREDICTION ERROR ANALYSIS

SYSTEM	DUANE			AMSAA		
	R1	R2	MSE	R1	R2	MSE
B	0.285×10^{2}	0.111×10^{1}	0.227×10^{5}	NA	NA	NA
C	0.457×10^{2}	0.171×10^{1}	0.106×10^{1}	NF	NF	NF
C1	0.716×10^{2}	0.147×10^{1}	0.146×10^{5}	0.480×10^{2}	0.157×10^{1}	0.156×10^{5}
C2	0.478×10^{2}	0.144×10^{1}	0.333×10^{5}	0.434×10^{2}	0.129×10^{1}	0.299×10^{5}
D	0.420×10^{2}	0.137×10^{1}	0.190×10^{5}	0.445×10^{2}	0.114×10^{1}	0.158×10^{5}
D2	0.365×10^{2}	0.120×10^{1}	0.157×10^{5}	0.412×10^{2}	0.105×10^{1}	0.138×10^{5}
E1	NF	NF	NF	0.419×10^{2}	0.141×10^{1}	0.483×10^{4}
E2	0.935×10^{2}	0.132×10^{1}	0.726×10^{4}	NF	NF	NF
E3	0.785×10^{2}	0.313×10^{1}	0.149×10^{6}	0.430×10^{2}	0.210×10^{1}	0.994×10^{5}
F	NF	NF	NF	0.468×10^{2}	0.124×10^{1}	0.313×10^{5}
G	0.435×10^{2}	0.130×10^{1}	0.444×10^{4}	NF	NF	NF
G1	0.394×10^{2}	0.122×10^{1}	0.323×10^{5}	0.455×10^{2}	0.917×10^{0}	0.263×10^{5}
G2	0.440×10^{2}	0.118×10^{1}	0.229×10^{5}	0.337×10^{2}	0.911×10^{0}	0.178×10^{5}
G5	NF	NF	NF	0.652×10^{2}	0.130×10^{1}	0.234×10^{5}
G7	0.668×10^{2}	0.216×10^{1}	0.184×10^{5}	NF	NF	NF
G8	NF	NF	NF	0.344×10^{2}	0.966×10^{0}	0.108×10^{5}
H3	0.373×10^{2}	0.114×10^{1}	0.143×10^{3}	0.399×10^{2}	0.107×10^{1}	0.134×10^{3}
H4	0.610×10^{2}	0.199×10^{1}	0.557×10^{2}	0.101×10^{3}	0.351×10^{1}	0.983×10^{2}
I3	0.831×10^{2}	0.262×10^{1}	0.256×10^{3}	0.733×10^{2}	0.208×10^{1}	0.20[illegible]
I4	0.395×10^{2}	0.217×10^{1}	0.459×10^{3}	0.390×10^{2}	0.106×10^{1}	0.223×10^{3}
J2	0.624×10^{2}	0.348×10^{1}	0.117×10^{4}	0.591×10^{2}	0.173×10^{1}	0.580×10^{3}
J3	0.752×10^{2}	0.178×10^{1}	0.643×10^{3}	NF	NF	NF

NF indicates no significant fit
NA indicates predictions not computable

Growth Rate Prediction - Although the linear regression analyses to correlate observed growth rates with equipment characteristics and program attributes were performed for both the Duane and AMSAA models, findings are presented herein only for the Duane model because:

- The reliability growth data analyzed were reflective of a single test for each equipment as opposed to a series of test phases
- The findings from the AMSAA regression analyses were consistent with, but did not augment, those from the Duane model analyses.

Thus, the focus of the subsequent discussion is on the Duane model growth parameter, α, and the factors which seem to explain its observed values for the different equipment failure-time histories.

As described in Section 4.2.1, various multiple linear regression models, assuming different characteristics/attributes as the independent variables and α as the dependent variable, were postulated, the regression constants estimated and then tested for the degree to which the independent variable(s) "explained" the variations in α across the different equipment data sets. As is usually the case when examining real-world data (as opposed to data generated in a statistically designed experiment), the characteristics of the data were not ideal for this type of analysis. The principal data limitations were:

- The large number of potential independent variables relative to the number of equipment data sets

- Lack of sufficient variability across the different equipment items for some of the candidate independent variables.

Nevertheless, the data base was believed to be sufficiently robust that this method would surface any factor, or pair of factors, which were obvious explanatory variables relative to other factors.

The results of the regression analyses did not produce any single linear combination of parameters which explained more than 50% of the variability in the Duane growth parameter estimates. Hence, no such model was identified to be statistically significant.

Further analyses revealed that the variability was actually due to factors other than equipment characteristics or program attributes. In particular, the influence of early failures was investigated. (See Section 4.1.3 for a discussion of how early failures can influence the apparent growth rate). The regression analyses were conducted using time to first failure (i.e., the first failure occurring in the growth test) as an additional candidate independent variable. The analyses yielded an extremely insightful result, namely that the best model for estimating the Duane growth rate was

$$\alpha = -0.0028 \times \text{FTF} + 0.00019 \times \text{TTT} \qquad (4.2\text{-}1)$$

where

α = Duane growth rate parameter

FTF = First time to failure

TTT = Total test time.

From a regression point of view, this model explained better than 87% of the variability in the Duane growth parameter estimates across the data sets and was highly statistically significant.

Since time to first failure is more a single data point subject to randomness than it is a predictable quantity, identification of the strong relationship of it to α did not contribute to the study objective of relating reliability growth to equipment characteristics and program attributes. However, it did serve to explain the difficulties in trying to discern such relationships by examining historical growth curves. It also highlighted the pitfalls involved in working with cumulative data, particularly for relatively small data sets, and the need to pay special attention to the influence of early failures when measuring reliability growth.

Analyses were conducted to determine whether first time to failure is an approximation to initial MTBF at program start, which is usually estimated as 10% of the MIL-STD 217 prediction. Table 4.2-6 relates these concepts by displaying the ratio of first time to failure to the MIL-STD 217 prediction. Note that the data in the table are ranked in order of ascending MIL-STD 217 predictions. Analyses conducted to determine if first time to failure could be statistically related to MIL-STD 217 prediction led to inconclusive results. Figure 4.2-3 displays a histogram of first failure time/MIL-STD 217 prediction. From this histogram it can be seen that the first time to failure cannot be estimated as 10% of the MIL-STD 217 prediction. The variability of these ratios indicates that the average percentage is closer to 6% of the MIL-STD 217 prediction and that the variability is significant.

TABLE 4.2-6

EQUIPMENT FIRST TIME TO FAILURES

EQUIPMENT	(FIRST TIME TO FAILURE/MIL-217 PRED)	FIRST TIME TO FAILURE	TOTAL NUMBER OF FAILURES
C1	.05500	60.0	21
C3	.11892	176.0	6
D2	.00660	11.5	31
G8	.13309	288.0	13
C5	.02654	88.0	5
G7	.00006	.2	31
C2	.07249	272.0	20
B1	.13997	548.0	3
G1	.00097	4.9	14
A1	.04151	214.0	3
G2	.00224	12.0	9
C4	.01431	88.0	31
E1	.00107	7.2	50
G3	.03150	249.0	18
A2	.18698	1485.0	1
G5	.03950	316.0	9
G6	.21409	1836.0	3
G4	.04689	576.0	2
E2	.00078	10.8	59
D1	.03597	666.0	2
B2	.06129	1592.0	7
E3	.00171	68.4	7

Less rigorous analyses of the data did yield some first-level relationships between reliability growth and characteristics/attributes. For example, equipment maturity was found to be a significant factor. The equipment data sets which had a small number of failures and/or did not display a good Duane model fit, tended to be the more mature, in some cases off-the-shelf, items. It can be generally concluded that significant reliability growth will not be achieved on such items. The relationship between growth rate and degree of maturity is less quantifiable, due to the difficulty in defining the latter. For example, one such definition might

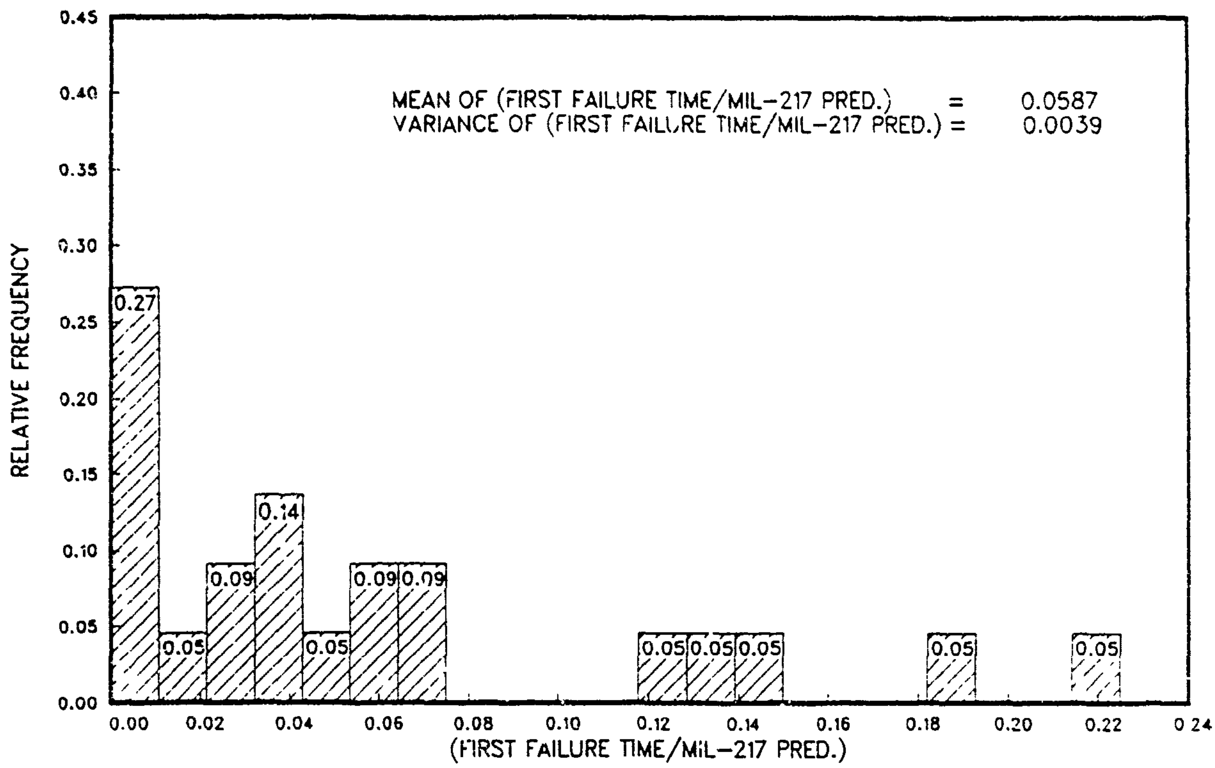

Figure 4.2-3 Histogram of First Failure Time/MIL-217 Prediction

be percentage off-the-shelf technology based on a design assessment; however, analyses revealed no significant relationship between growth rate and this metric as applied to the equipments in the data base.

The type of growth test conducted was also determined to be an influence on reliability growth rate. Referring to Table 4.2-2, items designated by the letters A through G were subjected to a Test-Analyze-Fix program. The test environment was designed to stimulate failures and hence the conditions were accelerated relative to the mission environment (i.e., mission candidates of prolonged, relatively benign operation were reduced or eliminated). The items designated by the letters H through J were subjected to a test in which the conditions reflected a simulation of the mission environment. Eliminating those equipments displaying a small number of data points and/ or an insignificant Duane fit, the average growth rate over the equipments subjected to the accelerated test was approximately 0.5. The average growth rate over those equipments subject to the mission simulation type test was 0.35. By extrapolation, and recognizing that a minimum growth rate of 0.1 is possible even without an agressive corrective action program, it is estimated that a benign test (i.e., equipment operation, but limited environmental stresses) would yield a reliability growth rate on the order of 0.2, providing the contractor is motivated to improve reliability.

Growth Model Initialization Considerations - The Duane reliability growth models fit to the equipment data sets were also examined from the standpoint of the initialization problem. Referring to Table 4.2-2, the relevant initialization parameter is the Duane Intercept, which can be interpreted as the logarithm of the ordinate (MTBF) at the point T = 1 hour. This being a somewhat difficult quantity to interpret in any intuitive sense, the data were transformed to the observed initialization point in the context of the MIL-STD-1635 rule-of-thumb which suggests that a reliability growth curve be initiated as follows for planning purposes:

- Start the initial MTBF ($MTBF_o$) at 10% of the MIL-HDBK-217 prediction
- Start the initial test time (T_o) at 100 hours, or 50% of the MIL-HDBK prediction, whichever is greater.

50% of the MIL-HDBK-217 prediction was the larger quantity for all equipments. The equivalent initial $MTBF_o$ at time $T_o = MTBF_p/2$ was derived for each equipment as follows:

$$MTBF_o = 1/K\ (MTBF_p/2)^{\alpha} \qquad (4.2\text{-}2)$$

where α and 1/K are the parameter estimates resulting from the Duane model fit (Table 4.2-2*). Table 4.2-7 presents the results of this transformation for the equipment items with positive growth rates and displaying good fits to the Duane model.

Examination of Table 4.2-7 does not suggest a discernible relationship between the initial $MTBF_o$ (or the ratio $MTBF_o/MTBF_p$) and the Duane growth rate or the predicted MTBF. No intuitive relationship to equipment characteristics was identified either; in fact one intuitively plausible relationship was somewhat contradicted. Specifically, it would be plausible to assume that the initial $MTBF_o/MTBF_p$ ratio would be lower for less mature equipments. However, item C3 was one of the less mature equipment items in the data base, and it displayed

TABLE 4.2-7

TRANSFORMED INITIAL MTBF

EQUIPMENT	DUANE GROWTH RATE (α)	$MTBF_p$	T_o ($MTBF_p/2$)	1/K	$MTBF_o$	$MTBF_o/MTBF_p$ (%)
C1	0.34	1091	545	12.165	108	10%
C3	0.486	1480	740	15.36	381	26
C5	0.644	3316	1658	5.36	634	19
D2	0.374	1743	871	5.28	66	4
E3	0.582	40000	20000	7.135	1331	3
G1	0.594	5066	2533	1.49	105	2
G2	0.572	5359	2679	3.48	318	6
G7	0.540	362.5	1812	1.16	67	2

*The Duane Intercept value in Table 4.2-2 is actually log(1/K).

the highest $MTBF_o$ (relative to $MTBF_p$). One significant observation that can be made is that while the $MTBF_o/MTBF_p$ displays considerable variation, the average across all equipments is approximately 9%. In a sense, these data suggest that the MIL-STD-1635 rule-of-thumb is not all that bad, <u>in the absence of any other information</u>.

4.2.3 Summary of Findings

The results of the analyses of the equipment reliability growth data can be summarized as follows.

- Both the Duane and AMSAA models provided good fits to the data sets, with the Duane model providing significant fits to a larger number of equipment data sets
- The estimated growth rates (excluding poor fits) were within the historically observed range ($0.1 < \alpha < 0.7$) in all but a single case
- The significant influence of early failures on reliability growth rate tended to mask out any influence by equipment characteristics or program attributes
- Mature equipments displayed little or no reliability growth
- Equipments subjected to a TAF-type test displayed an average growth rate of 0.5; equipments subjected to a mission simulation-type test displayed an average growth rate of 0.35; it was extrapolated that equipments subjected to a relatively benign operational test would display a growth rate of 0.2
- The MIL-STD-1635 reliability growth test initialization conditions were found to be a reasonable rule-of-thumb in the absence of any historical data.

However, it was also concluded that for a number of reasons (extended burn-in issues, lack of discrimination between random and correctable failures, pitfalls of plotting cumulative data), the Duane growth curve, as conventionally applied, does not always provide an accurate representation of true reliability growth.

4.3 CORRECTIVE ACTION ANALYSIS

The analyses presented in Section 4.2 essentially treated all failures observed in growth testing the same. In particular, failures which resulted in subsequent corrective action by the contractor were not distinguished from "random" failures. Since it is logical to argue that true reliability growth results only from the identification and incorporation of corrective actions (design changes or manufacturing process improvements), additional analyses were conducted on those equipment data sets for which corrective action information was available. This information consisted of a classification of each observed failure as to whether the failure was random or whether a corrective action was identified. It should be noted that data relative to time of incorporation of the corrective actions into the units under test were not generally available. Hence, there was no means of validating the effectiveness of the corrective action.

4.3.1 Failure Mode Classification

For discussion purposes, the terminology established by Crow in Refs. 11 and 12 will be used. Crow discussed a ratio of failure rate due to B failure modes (λ_B) divided by total failure rate, where total is defined to be the failure rate due to A modes plus failure rate due to B modes ($\lambda_A+\lambda_B$).

An A failure mode is one where no engineering fix has been assigned, while a B failure mode is one where an engineering or manufacturing process fix has been designated. Mathematically this ratio, labelled K, is

$$K = \frac{\lambda_B}{\lambda_A + \lambda_B} \tag{4.3-1}$$

From a practical point of view, K represents the percentage of failures that a contractor can adequately define and identify as fixable. In a sense it is reflective of the way in which a contractor works the problems that are uncovered during test. If many of the failure modes that are encountered are defined as random and, therefore, have no corrective actions assigned to them, they will remain inherent within the equipment. The device reliability will be affected accordingly.

Figure 4.3-1 is a histogram of the observed K-factors for the equipment data sets. It can be seen that, for the most part, the K-factor ranged from 25% to 70%. The low-end extreme case was exhibited by two extremely mature equipments which exhibited only Type A failure modes during the test. The high-end extreme case was exhibited by an item for which the observed failures were dominated by recurring failure modes which were deemed fixable.

Because of the wide spread of the K-factor values at the item (LRU) level, the failure data were aggregated at the manufacturer level to gain further insight. This aggregation is reasonable since it is the manufacturer who is analyzing and classifying the failures, as well as identifying the corrective actions. Table 4.3-1 presents the aggregated data for six programs. The data in Table 4.3-1 suggest that reliability growth, as measured by corrective actions identified or by

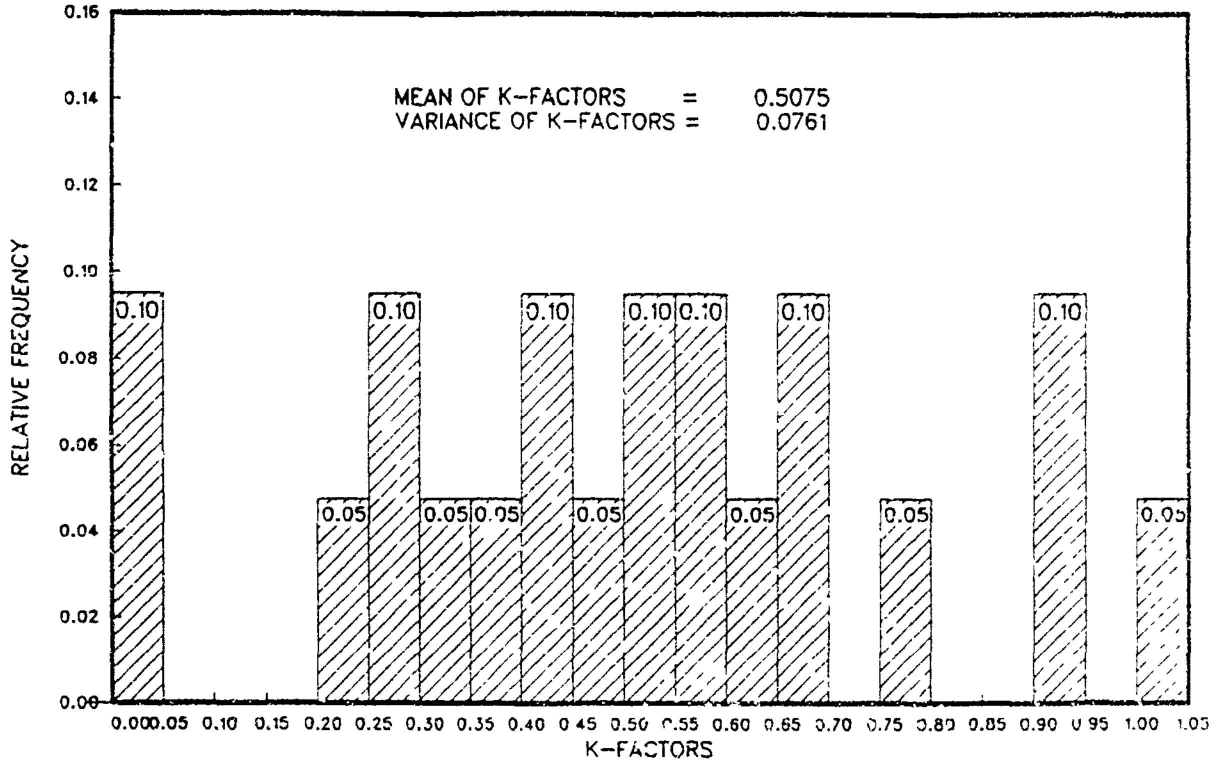

Figure 4.3-1   Histogram of K-Factors for Equipment Data Sets

failure modes eliminated through corrective action, is related to the complexity of the equipment as reflected by predicted MTBF. The MIL-HDBK-217 MTBF prediction is a meaningful measure of complexity, representing as it does both the total number of parts in the equipment and the complexity of the parts themselves. Clearly, more corrective actions were identified for these subsystems with lower predicted MTBFs.

However, complexity is not the only factor influencing the number of corrective actions identified. In order to discern other influencing factors, the data were normalized to an equivalent complexity by multiplying the corrective action count by the ratio of the predicted MTBF to 1,000 hours. The normalized data are shown in Table 4.3-2. These data demonstrate that the highest level of correction action activity

TABLE 4.3-1
RELIABILITY GROWTH TEST PROGRAM DATA SUMMARY

SUBSYSTEM	CHARACTERISTICS	MIL-HDBK-217 PREDICTED MTBF (hrs)	TEST TIME (hrs)	NUMBER OF TEST FAILURES	NUMBER OF CORRECTIVE ACTIONS	NUMBER OF FAILURES ADDRESSED BY CORRECTIVE ACTIONS
A	Mature Technology	2738	2416	8	2	3
B	Mature Technology	2837	2480	10	2	4
C	New Equipment; Several LRUs; High Complexity	359	2912	119	27	55
D	Partially Mature Technology	664	2838	48	7	18
F	Off-the Shelf; Prior TAF	1099	2331	12	1	1
G	New Equipment; Several LRUs	636	2718	100	18	73

TABLE 4.3-2

RELIABILITY GROWTH TEST PROGRAM DATA (NORMALIZED)

SUBSYSTEM	CHARACTERISTICS	NORMALIZED NUMBER OF CORRECTIVE ACTIONS	NORMALIZED NUMBER OF FAILURES ADDRESSED BY CORRECTIVE ACTIONS
A	Mature Technology	5.5	8.2
B	Mature Technology	5.6	11.3
C	New Equipment; Several LRUs	9.7	19.7
D	Partially Mature Technology	4.6	11.9
F	Off-the-Shelf Prior TAF	1.1	1.1
G	New Equipment; Several LRUs	11.4	46

was generated for the least mature equipments, namely Subsystems C and G. Similarly, the least corrective actions were observed on the most mature equipment item in the data base, Subsystem F, which was not only an off-the-shelf item, but had also been subjected to a Test-Analyze-Fix (TAF) program in an earlier aircraft program.

The Type A (random) failures by subsystem observed in test are presented in Table 4.3-3. The Type A MTBF displayed in the test is calculated as the test time divided by the number of Type A failures observed. What is significant is that the Type A MTBF is proportional (within reasonable statistical error bounds) to the MIL-HDBK-217 predicted MTBF. As shown,

TABLE 4.3-3

RANDOM FAILURE COMPONENT OF TEST DATA

SUBSYSTEM	TEST TIME (hrs)	TYPE A FAILURES DURING TEST	TEST MTBF (TYPE A ONLY)	PREDICTED MTBF	RATIO (PREDICTED MTBF/ TEST MTBF)
A	2416	5	483	2038	5.7
B	2488	6	415	2837	6.8
C	2912	69	46	359	7.8
D	2838	30	95	664	7.0
F	2331	11	212	1079	5.2
G	2718	27	101	636	6.3

the ratio of predicted MTBF to Type A test MTBF was consistent across the different subsystems, the average being approximately 6.5.

The fact that the Type A test MTBF was not equivalent to the predicted MTBF can be attributed to the acceleration factor associated with the TAF test. In particular, the TAF tested the equipments at the extremes of the mission environmental profile, and reduced or eliminated periods of relatively benign operation. As a result, random Type A failures occurred at a rate approximately 6.5 times higher than would be expected in the normal mission environment, which the MIL-HDBK-217 prediction reflects. <u>The consistency of the ratio across subsystems, all of which were subjected to equivalent TAF conditions, suggests that a random (Type A) component, which is related to the prediction, exists in the failure data.</u>

4.3.2 Operational Reliability Growth Prediction

The findings that Type B failures are related to complexity and maturity, and that a random failure component related to the MIL-HDBK-217 prediction exists, suggest that the IBM model, described in Section 2.3, might be a valid means of representing the growth data. As discussed in Section 2.3.2, this model has distinct advantages relative to the Duane model for prediction purposes in that:

- It allows discrimination between Type A and Type B failures
- Its parameters are more interpretable in a physical sense than are the Duane parameters.

Investigation of the IBM model is also suggested by Ref. 13. The author of this paper applied the IBM model, as well as several other growth models including the Duane model, to a ground radar in-house test program and concluded that "the IBM model is highly recommended." These considerations motivated a test of the viability of the IBM model for reliability growth prediction.

The data base did not support a completely comprehensive analysis of the IBM model. Referring to Table 4.3-1, it can be seen that only two subsystems (C and G) surfaced a sufficient number of Type B failure modes to support fitting a two-parameter model* to the data. The IBM model was applied to the Subsystem C and Subsystem G as follows:

---

*Including λ_o, the IBM model is actually a three-parameter model; however, λ_o can be independently estimated by extracting the Type A failures from the data set.

- Types A and B failures were distinguished based on the failure classifications in the data base
- The first occurrence of each Type B failure mode was identified, subsequent occurrences of the same failures were screened from the data base
- The procedure described in Section 2.3.2 was applied to estimate the parameters K_1 and K_2.

The results of the IBM model fit to the two data sets are shown in Table 4.3-4.

While definitive conclusions cannot be drawn from an examination of only two data sets, the viability of this modeling approach is supported by the fact that a good fit was obtained for these two sets. It is also significant to note that the fit of the IBM model to the sample data set in Ref. 13 resulted in estimates of the parameters K_1 and K_2 (K_1 = 29.44, K_2 = 0.0003644) which are of the same magnitude as the estimates in Table 4.3-4.

TABLE 4.3-4

IBM MODEL PARAMETERS FOR SUBSYSTEMS

SUBSYSTEM	MODEL PARAMETER		
	λ_o	K_1	K_2
C	0.022	32.3	0.00041
G	0.0049	22.1	0.00064

The estimated values for the parameters K_1 and K_2 provides a basis for developing prediction values in advance of testing based on engineering considerations. First, consider the relation of K_1 to predicted MTBF for the two data sets:

$$C: \quad K_1 \times MTBF_p = 11{,}596$$

$$G: \quad K_1 \times MTBF_p = 14{,}057$$

These results tend to confirm the intuitively plausible assumption that the number of Type B failure modes resident in an equipment at the start of testing (K_1) would be proportional to complexity, which could be measured by predicted MTBF. It is also reasonable to assume that equipment maturity and other reliability program elements imposed in the development program would influence K_1. For subsystems C and G, it was estimated that these considerations accounted for 50% of the Type B failures having been removed prior to the TAF test.

Next, consider the factor K_2. Since subsystems C and G were subjected to very similar TAF programs, one would expect that the rate of surfacing Type B failures (K_2) would be approximately the same. Indeed, the difference in K_2 for the two subsystems is not large. A possible explanation for the difference that does exist is that the manufacturer of Subsystem G was the prime contractor and the manufacturer of Subsystem C was a subcontractor. A prime contractor could be more strongly motivated to aggressively identify corrective actions and improve his hardware than would a subcontractor. Consider the relative values of the ratio of Type B failures to total failures for the two subsystems:

C: $K_2 = 27/(27 + 69) = 28\%$

G: $K_2 = 18/(18 + 27) = 40\%$

Thus, the contractor for Subsystem G identified a higher proportion of failures as being correctable than did the contractor for Subsystem C. This could explain the higher value for the parameter K_2, as well as the higher value for K_1 relative to the predicted MTBF.

While further analyses over a more extensive data base are clearly needed to validate the IBM model and provide fully supported methods to estimate its parameters, the following relationships are suggested for the interim based on a combination of engineering judgement and the data analyses performed:

$$\lambda_o = \lambda_p \times F_A \qquad (4.3\text{-}2)$$

where

λ_p = MIL-HDBK-217 predicted failure rate

F_A = Test acceleration factor, based on an assessment of the degree to which the test environment accelerates the mission environment on which the prediction is based

$$K_1 = 30{,}000 \times F_m \times \lambda_p \qquad (4.3\text{-}3)$$

where

F_m = Maturity factor, based on an estimate of failure modes removed by earlier testing or operation and the imposition of other reliability program elements

$$K_2 = (0.0005/6.5) \times F_A \qquad (4.3\text{-}4)$$

Using these suggested relationships, the IBM model can be applied to predict reliability growth during a reliability improvement development program as described in Chapter 5.

5. RELIABILITY GROWTH PREDICTION PROCEDURE

5.1 OVERVIEW

A procedure for predicting the reliability growth that can be expected for an equipment undergoing a reliability growth development program has been developed based on the data analyses described in Chapter 4. The recommended procedure is a departure from conventional reliability growth planning in that the recommended metric for reliability growth is based on the number of correctable failure modes surfaced during the test program, rather than the cumulative MTBF measured during the test. The IBM model provided the mathematical framework for characterizing reliability growth in this manner. The key equipment characteristics and program attributes influencing the degree of reliability growth to be expected include:

- Equipment complexity as measured by the MIL-HDBK-217 predicted MTBF
- Equipment maturity in terms of the number of failure modes removed through prior usage and/or the imposition of other reliability program elements
- Length of the test program in terms of the number of equipment operation hours to be accumulated
- Test environmental conditions relative to the operational environment.

These factors are the basis for estimating the initial reliability (prior to growth testing) and predicting the expected reliability at the conclusion of the reliability growth development test program using the procedures described in the subsequent sections.

5.2 INITIAL RELIABILITY

The first step of the procedure is to estimate the initial reliability. The initial reliability is defined to be that which would be expected for the equipment if it were produced and deployed operationally without being subjected to reliability growth testing. It is assumed that there are both Type A (random) failures and a fixed number of Type B failures in the equipment at the start of the test. The Type A failures are characterized by a constant failure rate which is estimated as the MIL-HDBK-217 prediction, λ_p. The number of Type B failures (i.e., the parameter K_1 of the IBM model) is estimated as

$$K_1 = 30{,}000 \times F_m \times \lambda_p \qquad (5.2\text{-}1)$$

where

F_m = Maturity factor, estimated as the percentage of Type B failures already removed from the equipment.

Experience and engineering judgement must be used to estimate F_m. As an example, if the equipment is a modification of an existing mature equipment for a new application, then the maturity factor should reflect the percentage of the design which is new.

The equipment reliability, as measured by MTBF, prior to testing can then be estimated, based on Eq. 2.3-5 with t=0, as

$$MTBF(o) = [\lambda_p + K_1 \times (0.0005/6.5)]^{-1} \quad (5.2\text{-}2)$$

The constant (0.0005/6.5) was derived in the data analyses presented in Section 4.3.2.

5.3 RELIABILITY GROWTH DURING TESTING

The next step is to predict the reliability improvement, relative to the initial reliability, to be expected through the conduct of a reliability growth test of specific length, as measured by equipment operation hours during test, t, and the defined test environment conditions. Reliability improvement is measured by the number of Type B failures, $N_B(t)$, identified in the test. This can be estimated based on Eq. 2.3-4 as

$$N_B(t) = K_1 \times (1-e^{-K_2 t}) \quad (5.3\text{-}1)$$

K_1 is given by Eq. 5.2-1, and K_2 by

$$K_2 = (0.0005/6.5) \times F_A \quad (5.3\text{-}2)$$

where

F_A = Test acceleration factor.

The numerical value for the test acceleration factor should be based on an assessment of the degree to which the test environment cycle represents an acceleration of the operational environmental cycle which is the basis for the MIL-HDBK-217 prediction. If the test is a Combined Environments Reliability Test (CERT) in which the test profile is a simulation of the operational profile, the test acceleration factor is unity.

If the test profile is designed to reduce periods of operation in benign environments, then the acceleration factor can be estimated as

$$F_A = T_{op}/T_{test} \tag{5.3-3}$$

where

T_{op} = Length of the operational duty cycle (i.e., missing length)

T_{test} = Length of the test cycle.

Assuming that effective corrective actions are incorporated for all Type B failures surfaced during the test (an optimistic assumption; see Section 5.5), the equipment MTBF after performance of the test is predicted as

$$MTBF(t) = F_A/(F_A \times \lambda_p + K_1K_2\ e^{-K_2t}) \tag{5.3-4}$$

Equations 5.2-1 and 5.3-4, in conjunction with the techniques described herein for estimating the various parameters, provide a methodology for predicting reliability improvement of the equipment as a consequence of the growth test.

5.4 EXAMPLE

To illustrate application of the reliability growth prediction procedure described above, consider the following hypothetical example of an avionics equipment to be subjected to reliability growth testing during full-scale development. The following assumptions are made:

- 40% of the equipment is new design; the remainder is comprised of mature, off-the-shelf items
- The MIL-HDBK-217 MTBF prediction is 300 hours (i.e., $\lambda_p = 1/300$)
- A Test-Analyze-Fix (TAF) program is to be conducted during which 3000 hours will be accumulated on the equipment
- The operational cycle for the equipment is a ten hour aircraft mission
- The TAF test profile eliminates the period of operation in a relatively benign environment (e.g., the cruise portion of the mission) resulting in a test cycle of two hours.

The predicted number of Type B failures in the equipment prior to testing (Eq. 5.2-1) is

$$K_1 = 30{,}000 \times (0.4) \times (1/300)$$
$$= 40$$

The initial MTBF (Eq. 5.2-2) is

$$MTBF(o) = [1/300 + 40 \times (0.0005/6.5)]^{-1}$$
$$= 156 \text{ hours}$$

The test acceleration factor (Eq. 5.3-3) is

$$F_A = 10/2$$
$$= 5$$

The rate of surfacing Type B failures during the TAF test (Eq. 5.3-2) is

$$K_2 = (0.0005/6.5) \times 5$$
$$= 0.0003846$$

The number of Type B failures identified during the test (Eq. 5.3-1) is

$$N_B(3000) = 40 \times (1 - e^{-0.0003846 \times 3000})$$
$$= 27$$

The equipment MTBF after incorporation of corrective actions to eliminate those Type B failures identified in the TAF program (Eq. 5.3-4) is

$$MTBF(3000) = 5/(5 \times 1/300 + 40 \times 0.0003846\ e^{-0.0003846 \times 3000})$$
$$= 232 \text{ hours}$$

Hence, the predicted reliability growth is from an initial MTBF of 156 hours to an improved MTBF of 232 hours, approximately a 50% improvement.

5.5 FOLLOW-ON TEST PHASES

The reliability growth prediction procedure described above applies to a single phase of reliability growth testing. If the reliability growth development program is to be conducted in multiple phases (see Section 2.2.1), then the procedure can be applied sequentially to predict the expected incremental reliability growth in each test phase. The principal consideration is that the assumed number of Type B failures at the start of a follow-on test phase should reflect the elimination of these Type B failures identified in prior test phases. For example, if $K_{1,2}$ denotes the equivalent of the parameter K_1 for a second phase of testing, then

$$K_{1,2} = K_1 - N_B(t) \qquad (5.5\text{-}1)$$

where K_1 and $N_B(t)$ are given by Eq. 5.2-1 and Eq. 5.3-1, respectively. Additionally, the distinct characteristics of the test program in each phase should be reflected in the test-related parameters. Specifically, the test length, t, and test acceleation factor, F_A, will in all likelihood be unique for each phase.

If the testing is conducted in multiple phases, an opportunity will be provided to assess the effectiveness of corrective actions incorporated to eliminate Type B failures identified in earlier test phases. The prediction procedure as described yields an upper bound on reliability growth in that 100% fix-effectiveness of Type B failures is assumed. A fix-effectiveness factor can be derived as follows:

$$F_E = 1 - N_{B,R}/N_B \qquad (5.5\text{-}2)$$

where

N_B = Number of Type B failure identified in prior test phases with corrective actions incorporated

$N_{B,R}$ = Number of Type B failures which have recurred after incorporation of corrective action

The fix-effectiveness factor thus provides a means for scaling the upper bound reliability growth prediction to account for imperfect corrective actions.

6. RELIABILITY GROWTH GUIDELINES

6.1 SELECTION OF EQUIPMENT DEVELOPMENT PROGRAMS

It is easily argued that reliability improvement potential exists in all systems, that any reliability improvement is good and, hence, reliability growth testing should be imposed on all equipment development programs. However, one of the findings of this study was that reliability growth testing is not cost-effective for some equipments. Therefore, a rational procedure is needed to determine whether to impose such testing. A four-step procedure is recommended:

1. Determine feasibility of growth testing and associated constraints
2. Assess current reliability and the potential for improvement
3. Determine need for improvement to meet user requirements
4. Estimate cost-effectiveness.

Each step is described below.

6.1.1 Feasibility and Constraints

Availability of Facilities - Meaningful reliability growth testing of avionics equipment can only be conducted when the system can be subjected to combined environmental conditions (temperature, vibration, humidity) at extremes representative of the mission environment. This implies the requirement for a specialized test chamber, preferably at the manufacturer facility or, alternatively, at an independent test laboratory. The chamber must be large enough for all equipments to be tested. Whether all equipments need to be tested together as an integrated system, in order to verify system operation and exercise interfaces, is a consideration. The need for software to simulate external interfaces must also be addressed.

Schedule Constraints - Reliability growth testing requires time, generally at least six months. The urgency of the user's need, and the attendant program schedule, must be evaluated to determine if the time is available. If the schedule is severely constrained, then the objectives of the growth test could be defeated, and it may be more beneficial to improve reliability through the imposition of other reliability program elements (e.g., high reliability parts) during the test program.

Availability of Test Articles - System prototypes must be dedicated to the growth test. If provisions for such prototypes were not included initially, then the program must be reviewed to determine if conflicting demands for test articles (e.g., flight testing) preclude such dedication. The test article should be complete. While some growth tests have been conducted on partial systems, this can limit the ability to verify operational performance. Further, it is often the missing segments still undergoing development which are the "tall-poles" relative to reliability.

Cost Constraints - Obviously, if reliability growth testing is imposed by the program office, the FSD budget must include the associated cost. This includes the cost of the test articles, chamber operations, and engineering labor to analyze test failures and identify corrective actions. If only a limited amount of funds can be targeted for growth testing, it must be determined if a meaningful growth test, e.g., of sufficient length, can be supported. The risk of stretching available funds too thin is a lack of aggressiveness by the contractor to improve the equipment.

6.1.2 Reliability Improvement Potential Assessment

A lesson-learned from the B-52 OAS TAF is that there is minimal benefit derived from imposing reliability growth testing on mature equipments. An assessment should be performed of both the current expected operational reliability of the system, i.e., if growth testing is not imposed, and the degree of improvement that can be achieved through reliability growth testing within the constraints discussed in Section 6.1.1.

Current Reliability - Clearly, the best means for assessing current reliability is operational data, if they exist. In the absence of such data, or equivalent data on analogous systems manufactured by the contractor, the method described in Section 5.2 can be applied. This approach is recommended rather than the Duane 10% rule because of the ambiguities in the Duane initial conditions. It focuses attention on the component of reliability which can be improved through corrective actions, i.e., Type B failures, rather than on burn-in type failures which would be more effectively removed by Environmental Stress Screening in production.

Growth Potential - Section 5.3 provides procedures for predicting the operational reliability growth achievable in a test of specified environmental conditions and length. Using these procedures, the operational reliability at the conclusion of the test can be predicted. The current and final reliability predictions, as estimated using Eqs. 5.2-2 and

5.3-4, respectively, provide the basis for assessing the need for and cost-effectiveness of reliability growth testing as discussed next.

6.1.3 Need Assessment

Assessment of the need for reliability growth is simple and straightforward. The current reliability, MTBF(o), is compared to the user requirement as stated in the Statement of Need (SON), Required Operational Capability (ROC), Mission Element Need Statement (MENS), etc. If the user requirement is not achieved by the current reliability, then reliability improvement is clearly necessary. If the requirement is met, then the imposition of growth testing should be based on a cost-effectiveness analysis.

6.1.4 Cost-Effectiveness

Reliability improvement results in reduced initial logistics support (e.g., spares to meet system readiness requirements) and operations & support (e.g., maintenance labor and materials) costs. Generally, the greater the quantity of systems to be deployed and the higher the utilization, the greater the savings downstream. Both the costs of reliability improvement and the resulting cost reduction can be quantified. The improvement cost includes fixed costs (e.g., test article fabrication) and costs which are a function of test length (e.g., chamber operations). A life-cycle cost model, Ref. 14 for example, can be exercised to quantify the cost savings associated with improving reliability from MTBF(o) to MTBF(t). Hence, the cost of conducting a test of length t to improve reliability from MTBF(o) to MTBF(t) can be traded off against the downstream savings, i.e., LCC (at MTBF(o)) minus LCC (at MTBF(t)), to determine cost effectiveness.

6.2 SPECIFICATION GUIDELINES

The principal recommended modifications to current reliability growth testing and specification practices pertain to an increased emphasis on failure classification (Type A vs Type B) and corrective action tracking. Also, a follow-on test to permit incorporation of all corrective actions into test articles is recommended subsequent to the primary TAF test, the purpose being to verify the effectiveness of corrective actions and measure operational reliability.

6.2.1 Planned Growth Curves

MIL-HDBK 189 discusses the reliability growth that occurs over an entire development program. The overall pattern is described by an idealized growth curve, as shown by the dashed line in Fig. 6.2-1. The solid lines are the planning curves that are relevant to particular test phases within the program itself. This report deals with predicting the growth that should be observed during a Test-Analyze-Fix (TAF) test phase.

When tracking reliability during a major test phase within a development program, such as a TAF, there are three reliability values that are of interest on the planned growth curve. They are:

- Reliability predicted at the beginning of the test phase
- Reliability achieved by incorporating fixes during the test
- Reliability achieved by introducing delayed fixes into the system at the end of the test.

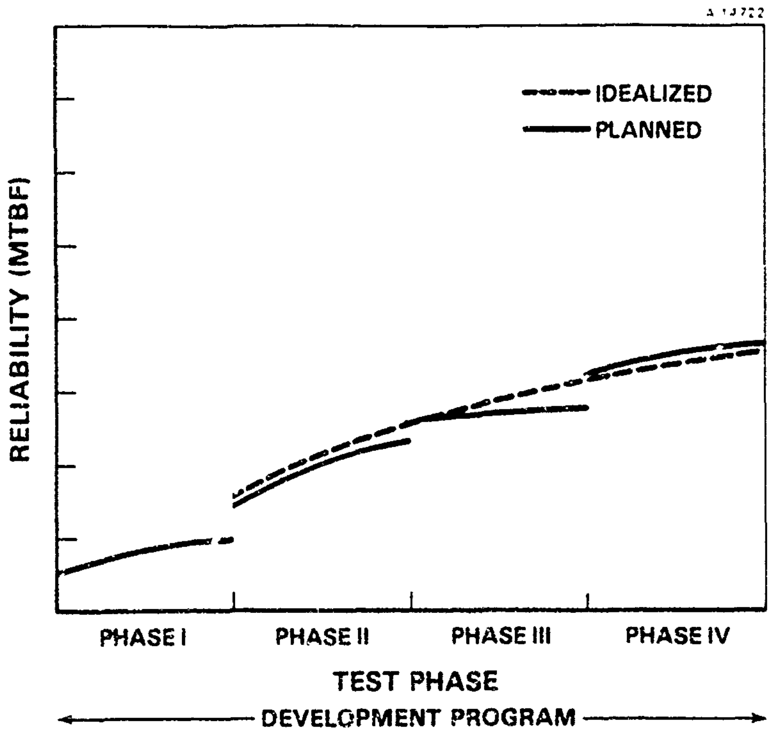

Figure 6.2-1 Development Program Idealized and Planned Growth

The situation is depicted pictorially in Fig. 6.2-2. A planned growth curve taking these considerations into account should be established prior to start of reliability growth testing.

6.2.2 Test Environment

The test environment should be designed to surface equipment design or manufacturing problems which can be corrected in order to improve reliability. There are two forms of test acceleration:

- Time Acceleration
- Environmental Acceleration.

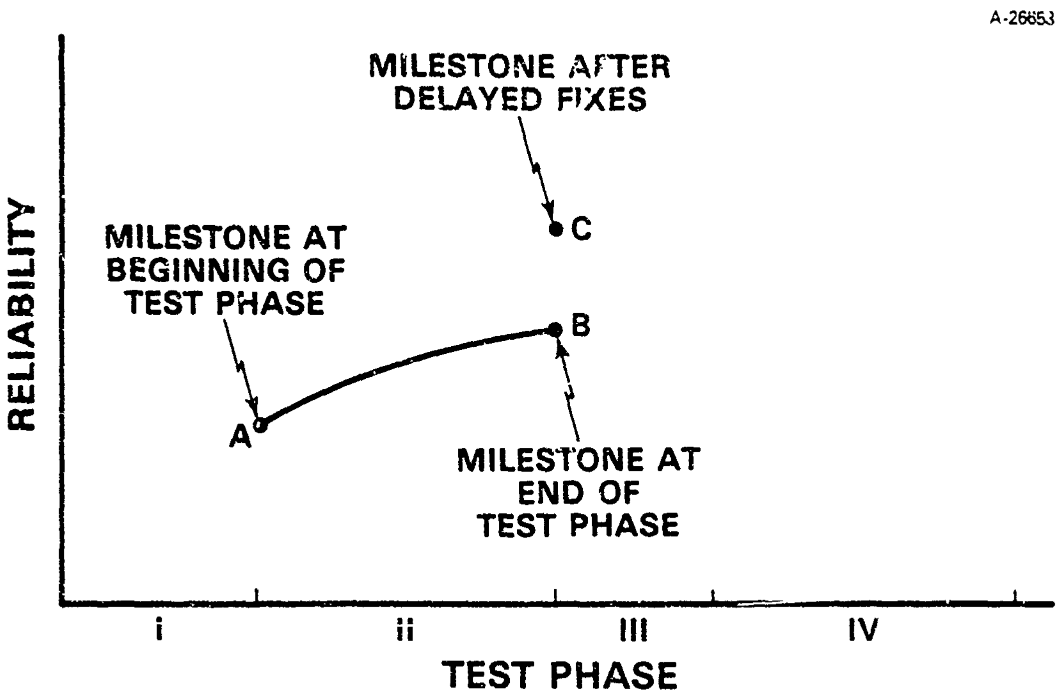

Figure 6.2-2 Reliability Values Associated With A Major Test Phase

In a <u>time-accelerated</u> test, the simulated environmental profile deemphasizes the benign mission phases that are assumed not to generate many failures. The benign phases include extended cruise and ground time. Accordingly, most of the test cycle is devoted to the more stringent environments, which for avionics are associated with takeoff, climbout, and altitude changes.

In an <u>environmentally-accelerated</u> test, the environmental levels employed are more severe than those associated with a typical mission profile (or with the design limits of the equipment). Accelerated environments are used to detect equipment failure modes that would be exhibited only under extreme flight conditions, or would not be observed until the equipment is exposed to normal conditions for the greater periods of time associated with equipment wear-out modes. Environmental acceleration imposes the risk of inducing non-relevant failure modes.

The time and environmental acceleration conditions should then be translated into a test environmental profile as a basis for specifying test chamber operations. Figure 6.2-3 is an example of such a profile.

6.2.3 Length of Test

The length of testing should be chosen so as to generate sufficient experience with specific failure modes so that

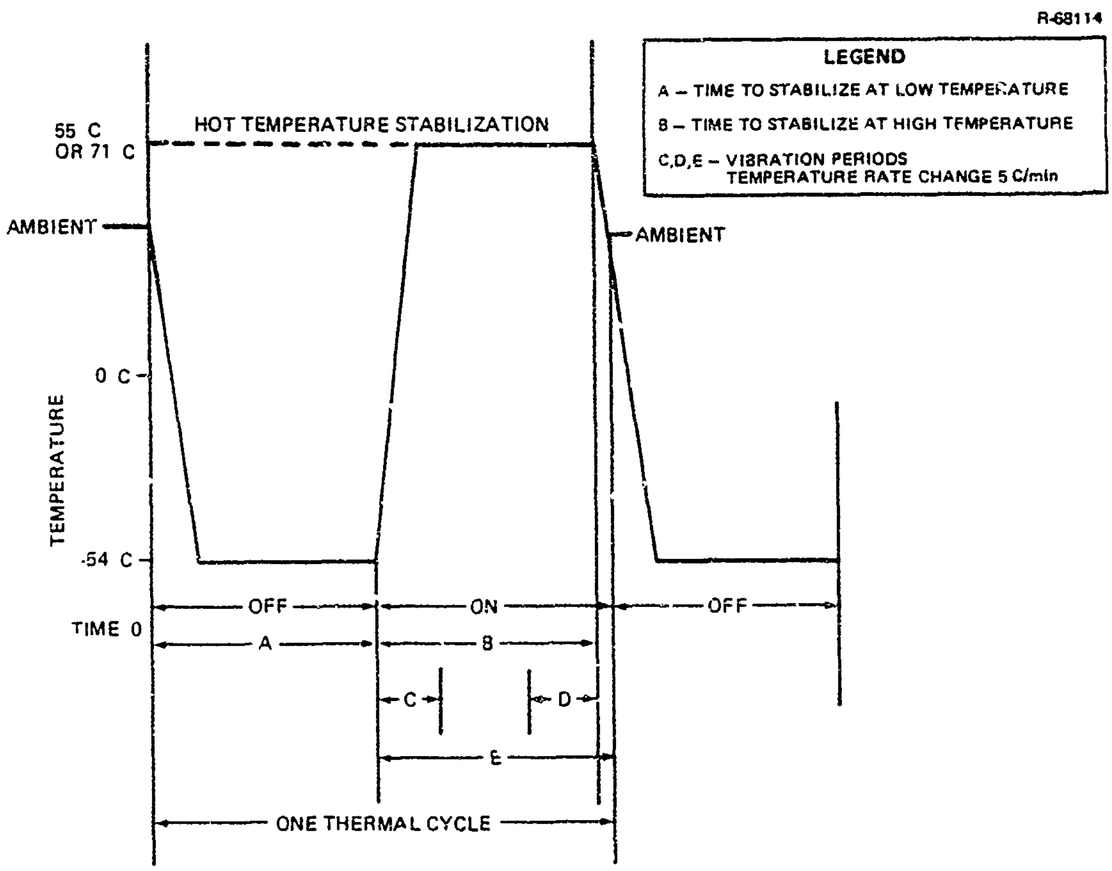

Figure 6.2-3 Representative Test Environmental Profile

corrective action may be identified and implemented. It is very important that the test hours are either long enough to allow corrective actions to be incorporated or flexible enough to permit lengthening the scheduled test time. The contractors should be encouraged to incorporate corrective actions and to verify them in TAF; therefore, they should be allowed control over equipment configurations while being required to provide an audit trail of configurations to the test monitor. The procedure described in Section 6.1.4 can be utilized to specify a test length that is cost-effective.

6.2.4 Data Requirements

Data requirements during test must be specified. Procedures must be written such that no subjective determinations are required on the part of the test operators. Failure criteria should be consistent with those applied in the anticipated operational environment. Detailed procedures should be specified for recording equipment "on time" with each test event noted in the data logs. If elapsed time indicator (ETI) meters are installed in the equipment, these should be used as the "on time" baseline; ETI values should be recorded, as a minimum, at the beginning and end of each test cycle/sortie. Serial numbers of each LRU undergoing test should be recorded prior to each test cycle for the purpose of correlating subsequent equipment performance with configuration.

With the above requirements in mind, the key TAF data elements are identified in Table 6.2-1. This list is considered to be the minimum level of data required to track and assess reliability growth.

With regards to reporting of the above data elements, it is necessary to specify reporting requirements with respect

TABLE 6.2-1

DATA REQUIREMENTS

DATA ELEMENT	DESCRIPTION
LRU	Identification of UUT
Serial Number	Serial number of operating/failed LRU
Configuration	Corrective actions incorporated
Operating Time	Total on-hours of the LRU and of the LRU type (especially at time of failure)
Cycle Time	Time into the cycle when a discrepancy or failure is noted
Test Conditions	Environmental conditions at the time of a discrepancy or failure
Discrepancy	Observed symptom, test operator actions and BIT indications
Maintenance Activity	Description of all maintenance actions leading to on-site repair of the LRU; identify removal/replacement of any subassembly (SRU), module, and/or component; include nomenclature (item name) and serial number of replacement item. Also identify any adjustments performed and relationship to failed item.
Malfunction	Identify the function or parameter within the LRU that was performed incorrectly or not-at-all as a direct result of the item failure, and its relation to the observed symptoms. If the failure was discovered during performance verification tests and not during functional check out/operation, identify why BIT did not detect the failure.
Maintenance Time	The following maintenance times shall be recorded: a) on-site time when bench checkout (repair) was initiated, b) on-site time when repair was completed, c) on-site time when performance verification tests were successfully completed.
Contractor Representative	Name and signature of contractor representative submitted report.
Failure Analysis	As a minimum, the following information shall be provided: Failure Classification - Type A, Type B or non-relevant Module/Circuit/Component Failure - The faulty module/circuit component shall be identified to the extent corresponding to depot-level maintenance and to the degree that corrective action is proposed/performed (i.e., repair of part vs. circuit/module redesign). Cause of Failure - Identify the relationship between the failure and the environmental/operating conditions at the time of failure; include pertinent variations in other circuits which could have caused the failure in the failed circuit. Provide sufficient detail to substantiate the failure mechanism (i.e., overdrive, overvoltage, overtemperature/excessive heat dissipation as relates to envrionmental conditions). Corrective Action - Identify the performed and/or proposed corrective action to avoid recurrence of the failure. Identify how the modification is to enhance the reliability of the failed item

to time. However, the reporting structure should not be so stringent as to prevent the contractor from performing detailed failure analysis in order to meet a reporting requirement. It should be possible, within the framework of the reporting requirements, for specific failure report to be allowed to remain open pending the outcome of analysis. For example, if failure analysis is incomplete, the contractor should submit the required documentation presenting what has been accomplished up to that point and stating that failure analysis is on-going. At some other point in time, e.g., 30 days later, the failure analysis should be readdressed and either closed out or continued with further updates of on-going analysis.

6.2.5 Configuration Tracking

A separate log should be maintained of all corrective actions identified as the result of the failure analyses. This should include for each corrective action:

- The elapsed time at which each failure associated with the corrective action occurred and the serial number of the test item.
- The elapsed time at which the corrective action was incorporated into each unit under test by serial number.

These data will allow estimation of reliability improvement in accordance with the procedure described in Chapter 5, as well as verification of the effectiveness of each corrective action.

6.2.6 Follow-Up Testing

A follow-up to the reliability growth test is recommended allowing sufficient time for incorporation of all corrective actions. The follow-up test should be a verification, as opposed to a growth, test. As such, the environmental profile should be representative of the mission profile, equipment configuration should be controlled and the test should be of sufficient length to allow demonstration of the target MTBF with statistical confidence. MIL-STD-781 provides guidelines for structuring such a test.

7. CONCLUSIONS

7.1 RELIABILITY GROWTH TESTING: PRACTICES AND PITFALLS

Several findings emerged from the study relating to reliability growth testing as it has been performed in the past and the true meaning of reliability growth curves as they are conventionally presented.

7.1.1 Reliability Growth Testing is Conducted in A Time-Constrained Environment

The focus of this study was on reliability growth testing during full-scale development, as opposed to the production and operational phases of the system life cycle. In each of the development programs surveyed in this study, the amount of calendar time available to perform growth testing was limited to less than a year, usually only six months. While it could be argued that the available test time could be expanded by starting testing earlier, the test articles should be reasonably representative of production prototypes. Similarly, there are constraints on the number of prototypes which can be dedicated to reliability growth testing due to funding limitations and other test agency demands. Consequently, the programs surveyed generally entailed intensive testing of a small number of prototypes (usually two) continuously over a limited interval of time.

7.1.2 Reliability Growth as Conventionally Portrayed Can be Misleading

Clearly, true reliability improvement can only result from the incorporation of design changes and other corrective actions. However, the test scenario described above does not allow timely incorporation of corrective actions into the units undergoing test, the reason being that when a failure occurs, the test unit is repaired and entered back into test as soon as possible. The process of failure analysis and corrective action identification is conducted in parallel. At best, corrective actions are not incorporated into the test articles until the tail-end of the test and, more often, are not incorporated until production. Hence, the results of corrective actions are not represented in the failure/time history data generated in the growth test. Still the data do indicate reliability growth. This "apparent" growth is best explained by Swett (Ref. 15) as follows:

> "The weakest part of the test article fails first. The replacement part is a random sample from the spare parts inventory. That means there may be ten-thousand-to-one odds that the replacement is a better part than the one it replaces. So the composite reliability of parts in the test article is being systematically improved each time there is a failure, and observed reliability of the test article continues to grow."

It was also demonstrated in this study that the practice of plotting cumulative data, combined with infant mortality in test articles, can lead to the appearance of growth when, in fact, none is occurring. The implication relative to the objectives of this study is that it is not sufficient to draw conclusions from historical growth curves alone; it is necessary, instead, to factor into consideration the number and nature of the corrective actions generated as a consequence of the test.

7.2 SYSTEM COMPLEXITY, MATURITY AND GROWTH TEST CONDITIONS ARE THE FACTORS MOST STRONGLY RELATED TO RELIABILITY GROWTH

Reliability growth, as measured by corrective actions identified or by failure modes eliminated through corrective action, was found to be related to the complexity of the equipment as reflected by predicted MTBF. The MIL-HDBK-217 MTBF prediction is a meaningful measure of complexity, representing as it does both the total number of parts in the equipment and the complexity of the parts themselves. Additional analyses demonstrated that the highest level of corrective action activity was generated for the least mature equipments included in the program survey and the the least corrective actions were observed on the most mature equipment item. Thus, maturity as characterized by the degree of new hardware development, is related to reliability growth.

The environmental stress levels and time-compression factors associated with the reliability growth test profile were also found to influence reliability growth. A higher degree of growth was observed in these programs for which the test profile was designed to stimulate failures, as opposed to simulating mission conditions.

7.3 RELIABILITY GROWTH PREDICTION

7.3.1 Historically-Observed Relationships were Confirmed

Analysis of the failure-time history data generated in the programs surveyed confirmed that:

- Such data do tend to track to a Duane model

- The Duane reliability growth rate (α) is within the historically observed bounds, $0.1 < \alpha < 0.7$

- While there is considerable variation for specific items, the Duane initialization groundrules given in MIL-STD-1635 are, on the average, reasonable if no other information is available.

The growth rate, α, was found to be strongly influenced by the occurrence of failures early in the test, an influence which tended to mask out any underlying relationships between growth rate and equipment characteristics or program attributes. However, the following quantifiable relationships to test conditions were identified:

- $\alpha = 0.5$, for a TAF-type test designed to stimulate equipment failures

- $\alpha = 0.35$, for a test designed to simulate the mission environment

- $\alpha = 0.2$, for a benign operational test.

These relationships should be accepted in the mean-value sense only, recognizing that there will be variation about the mean for any specific program based on the aggressiveness of the contractor's corrective action program

7.3.2 Alternatives to the Duane and AMSAA Models are Recommended for Reliability Growth Prediction

One of the major conclusions of this study is that the two models (Duane and AMSAA) which are widely used today for reliability growth measurement and management have limitations as predictive tools. The limitations are two-fold:

- The models do not discriminate between the apparent reliability growth due to extended burn-in effects and true reliability growth associated with the identification and incorporation of design changes and other corrective actions

- The parameters of the model are empirical in nature (i.e., determined by data fits) and cannot be intuitively related to equipment characteristics and program attributes.

It was deemed desirable that prediction techniques instead be based on a model whose parameters are more directly related to equipment characteristics and program attributes, and which addresses corrective action potential. A promising candidate, the IBM model, was evaluated.

A procedure for applying the IBM model to reliability growth data was identified, providing that the data include failure classifications (random vs. correctable). The model was applied to two data sets in the data base and found to yield reasonable fits. Intuitively plausible relationships were developed for estimating the parameters of the model based on:

- MIL-HDBK-217 predicted reliability

- Equipment maturity

- Time acceleration factor associated with the growth test profile

- Test length.

While further study will be required to validate the model and refine the constants developed for these relationships, the model provides an interim capability for reliability growth prediction.

7.4 RECOMMENDATIONS

The Reliability Growth Prediction Study has surfaced the need to focus less attention on cumulative failure-time data plotted on a log-log scale and more on failure model classification and corrective action identification and verification. Toward that need, the following specific recommendations are offered:

- Institutionalize procedures for failure classification and corrective action tracking in future reliability development growth testing procedures
- Require performance of a reliability demonstration test, subsequent to the growth test, for verification of corrective action effectiveness
- Develop procedures for centralizing the data generated during such tests
- As data are accumulated, perform further research to validate the reliability growth prediction procedure suggested herein.

REFERENCES

1. Duane, J.T., "Learning Curve Approach to Reliability Monitoring," IEEE Transactions, Aerospace, Vol. 2, 1964.

2. Codier, E.O., "Reliability Growth in Real Life," Proceedings Annual Reliability and Maintainability Symposium, 1968.

3. Selby, J.D., and Miller, S.G., "Reliability Planning and Management - RPM," General Electric Company, Aerospace Electronic Systems, Utica, NY.

4. Herd, G.R., "Comparisons of Reliability Growth Experience," Proceedings - Institute of Environmental Sciences, 1978.

5. Crow, L.H., "Reliability Analysis for Complex Repairable Systems," Reliability and Bimetry, eds. F. Proschan and J. Sufling, Philadelphia, SIAM, pp. 379-410, 1974.

6. Rosner, N., "System Analysis - Non-Linear Estimation Techniques," IBM Corporation, New York, 1965.

7. Singpurwalla, N.D., "Estimating Reliability Growth Using Time Series Analysis," Naval Logistics Research Quarterly, Vol. 25, pp. 109-112, March 1978.

8. Schafer, R.E., Saller, R.B., and Torrey, J.D., "Reliability Growth Study," Hughes Aircraft Company, RADC-TR-75-253, October, 1975.

9. Lawlor, J.E., "Offensive Avionics System Test-Analyze-Fix Reliability Improvement Program," The Analytic Sciences Corporation, Technical Report TR-3406-3, February 1983.

10. Clarke, J.M., "No-Growth Growth Curves," Proceedings of the Annual Reliability and Maintainability Symposium, 1979.

11. Crow, L.H., "Reliability Growth Projection from Delayed Fixes," Proceedings of the Annual Reliability and Maintainability Symposium, 1983.

REFERENCES (Continued)

12. Crow, L.H., "Methods for Assessing Reliability Growth Potential," Proceedings of the Annual Reliability and Maintainability Symposium, 1984.

13. Koo, D., "Growth Test Time - Key to Effective Test Planning," Proceedings of the Annual Reliability and Maintainability Symposium, 1981.

14. Life-Cycle Cost Analysis Program Users Guide, The Analytic Sciences Corporation, Engineering Memorandum EM-2081-6, July 1979.

15. Swett, B.H., "Keynote Address: Reliability Growth Management," Reliability Growth Management, Testing, and Modeling: Seminar Proceedings, Institute of Environmental Sciences, Washington, DC, February 1978.

MISSION
of
Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence (C^3I) activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POs) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.