AN EXAMINATION OF STATISTICAL RIGOR INFUSED INTO THE KC-46 FLIGHT TEST PROGRAM

THESIS

Sean C. Ritter, Second Lieutenant, USAF

AFIT-ENS-13-M-18

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-13-M-18

AN EXAMINATION OF STATISTICAL RIGOR INFUSED INTO THE KC-46 FLIGHT TEST PROGRAM

THESIS

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Sean C. Ritter, BSME

Second Lieutenant, USAF

March 2013

AFIT-ENS-13-M-18

AN EXAMINATION OF STATISTICAL RIGOR INFUSED INTO THE KC-46 FLIGHT TEST PROGRAM

Sean C. Ritter, BSME
Second Lieutenant, USAF

Approved:

Raymond R. Hill, PhD (Advisor)	Date
Joseph J. Pignatiello, PhD (Reader)	Date

Abstract

The KC-46 program is bringing on-line the replacement aircraft for the KC-135. Although not a new development program, but rather a modification program, there are extensive plans for the flight testing of the KC-46. Recent DoD emphasis mandates the use of statistical design principles for DoD test and evaluation. This project will examine the planned flight test program for KC-46 and reconsider components of that program based on principles of statistical rigor. Of particular focus will be the reliability and maintainability aspects of the flight test program.

Current methodology assumes a constant failure rate in all situations, implying that the underlying failure profile of any component or system is assumed to be exponentially distributed. Use of the Weibull failure distribution is proposed as a more general framework to provide additional insight about the failure profile of the component or system.

To my mother, whose always been the foundation upon which I've been able to succeed.

Acknowledgements

I would like to thank my advisor Dr. Raymond Hill for his continued paitence as I struggled with learning how to do research while writing this thesis. I also want to thank Mr. Barry Mayhew and Mr Dave Gomez at the KC-46 Project Office for their help in trying to understand both what the program office wanted and what I could deliver with an academic thesis.

Sean C. Ritter

Table of Contents

Page

List of Figures

Figure Page

List of Tables

AN EXAMINATION OF STATISTICAL RIGOR INFUSED INTO
THE KC-46 FLIGHT TEST PROGRAM

I. Introduction

1.1 KC-46 at a Glance

KC-46 is the next generation tanker aircraft and is part of a re-capitalization strategy for the KC-135. It is currently slated to eventually replace one-third of the current generation tanker fleet [5:11]. The KC-46 is currently in the Engineering and Manufacturing Development phase and is on the Office of the Secretary of Defense Oversight List. The KC-46 is targeted to satisfy issues identified in the Initial Capabilities Document (ICD) for Air Refueling through 2020 [5:11].

The KC-46 is a split from the Boeing 767-2C. It is based heavily on the commercially available and FAA certified Boeing 767-200ER-IGW incorporating other aspects from the 767-300F, 767-400ER, and the 787-8. Additionally, there is a substantial amount of planned militarization to allow the KC-46 to meet USAF requirements as outlined in the KC-46 System Specifications.

The KC-46 must have the following Key Performance Parameters: Tanker Air Refueling Capability, Fuel Offload versus Radius[1], Civil/Military Communications, Navigation, Surveillance/Air Traffic Management (CNS/ATM), Airlift Capabilities, Receiver Air Refueling Capability, Chemical/Biological Environment Hardening, Net-

[1] As depicted in Figure 3 of the TEMP [5:22]

Ready, Survivability in Hostile Operating Conditions[2], and Capable of performing multiple simultaneous air refuelings.

The KC-46 must have the following Key System Attributes: Formation Capability, Aeromedical Evacuation, Reliability and Maintainability, Operational Availability, and Treaty Compliance Support.

More details on the exact system specifications for the KC-46 are available in the official program documentation. The KC-46 is essentially a two mission system. It combines Airlift and Air Refueling along with all the associated capabilities required of both and that of a US military aircraft.

1.2 The Suitability Mandate

In a 2009 Memo to all Department of Operational Test and Evaluation (DOT&E) staff, Director J. Michael Gilmore stated that suitability must be substantially improved before Initial Operational Test and Evaluation (IOT&E) by [20]:

- Assess at appropriate milestones whether programs meet the requirement to have a reliability growth program and identify for action by DOT&E leadership cases where this requirement is not met.

- Work with developmental testers to incorporate in the Test and Evaluation Master Plan (TEMP) a reliability growth curve or software failure profile, reliability tests during development, an evaluation of reliability growth and reliability potential during development.

- Work with developmental testers to ensure data from the test programs are adequate to enable prediction with statistical rigor of reliability growth potential

[2]Including defensive systems, situational awareness, aircrew night vision devices, and laser eye protection systems, aircraft maneuverability, EMP protection, and night vision and imaging system compatibility

and expected IOT&E results. The rigor should be sufficient to calculate the probabilities of accepting a bad system and rejecting a good system and those probabilities should be used to plan IOT&E.

Dr Gilmore's mandate was the result of several prior efforts to fix a decrease in system reliability and maintainability. In a 2011 presentation to the National Academy of Science, Director Gilmore stated that of 15 systems reported on in FY11, only 6 had met their reliability threshold. Since 1985, of 170 systems, only 30% have met their reliability thresholds [19].

It was not the first time that this reliability shortfall was highlighted. A report by the Defense Science Board Task Force on Test and Evaluation in the DoD remarked that there was substantial room for improvements when it came to modeling Reliability, Availability, and Maintainability (RAM). The focus in this document was knowing in advance what competing designs were doing for RAM, specifically logistics and support costs [21]. This report was produced at a time when Reliability Growth was missing in most acquisition programs due to acquisition reform in the mid-1990s [4].

The consequences of a lack of reliability growth in suitability analysis meant that projects were not ready for IOT&E. For example, additional fixes delayed projects such as the V-22 program for 5 years and required almost a billion additional dollars to get the suitability requirements up to specification [4:22]. The same report also cites the Joint Air-to-Surface Standoff Missile Program as running into similar issues. These sustainment costs account for almost two-thirds of total system costs. The report calls for making RAM improvements by reacquiring reliability personnel and placing a required reliability clause into the contracts and subsequently modify it as needed at all stages of program development. No amount of testing could overcome the shortfall in RAM formulation [4:24].

As a result, a Reliability Working Group chartered by the Director for OT&E and the Deputy Under Secretary of Defense in February of 2008 produced a set of documents that signaled the various service test organizations willingness to begin to adopt new strategies [23]. This was part of the actions recommended by the Defense Science Board report in 2008.

1.3 Problem Statement

As a result of its system complexity, and the diverse operating conditions that the KC-46 is expected to perform in, RAM is essential. This thesis builds upon the implementation strategy in the working group's report [23] by combining statistical rigor into the RAM program used by the KC-46. Proposed is new methodology where variable failure rate is considered through the use of the Weibull failure distribution as a more general framework for analysis of failure profiles. This framework is robust even in presence of censored data.

II. Literature Review

2.1 Basic Reliability Concepts

Many of the principles of reliability and maintainability that apply at a component level also apply at a system level. Two important characteristics in reliability are Mean Time to Failure (MTTF) and the hazard, or failure, rate ($\lambda(t)$). The notation used is consistent with the notation in [18]. These two characteristics are used in defining reliability quantitatively.

MTTF is the expected time before failure of a component or system, defined by MTTF $= \int_{\infty}^{0} tf(t)dt$ [18:26], where $f(t)$ is the underlying failure probability density function for a component or system. The $\lambda(t)$ is defined by the formula, $\lambda(t) = \frac{f(t)}{R(t)}$ where $f(t)$ is divided by the reliability function, $R(t) = 1 - F(t)$ [18:29].

For many systems or components, the failure rate over their lifetime follows a general bathtub curve as shown in Figure 1. Important to note is that this curve depicts failure rate over time and is not representative of the reliability of the system or component. The bathtub curve is characterized by an early period of time where the failure rate is decreasing from some initial rate, known as the burn-in period. The middle portion of the bathtub curve is where the failure rate is approximately constant, known as the useful life. The final portion of the curve is where the failure rate is increasing and is usually beyond the design life of a component or system. Additionally, the hazard rate function for some distributions can approximate this general shape, such as the Generalized Weibull or the Exponentiated Weibull distributions [26:16].

Many failure distributions are estimated based on well-known and characterized distributions. One such failure distribution is the exponential distribution. It is used when failures are expected to be completely random and independent and is

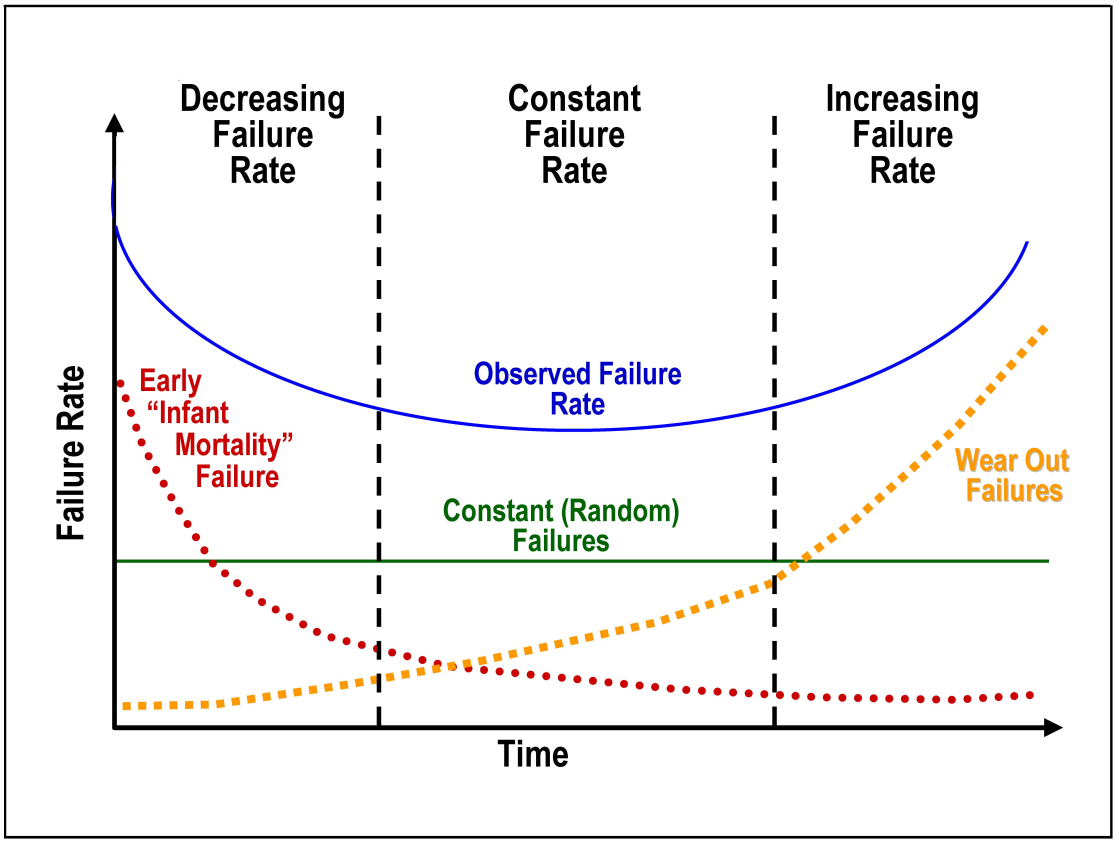


Figure 1. Theoretical Bathtub Curve

the simplest of the failure distributions, exhibiting a constant failure rate [18:44-47]. The exponential distribution is also memoryless. Memoryless means that the remaining time to failure does not depend on the elapsed operating time. However, despite limitations of the memoryless property, the exponential distribution remains useful in practice. An example of the exponential reliability function is shown in Figure 2. Several variations are shown as the key parameter of MTTF (MTTF $= \frac{1}{\lambda}$) varies. Changing the MTTF of an exponential distribution elongates the reliability distribution, but does not change the shape.

Other common failure distributions allow for non-constant failure rates, which allow accurately modeling burn-in and wear-out periods. These distributions include

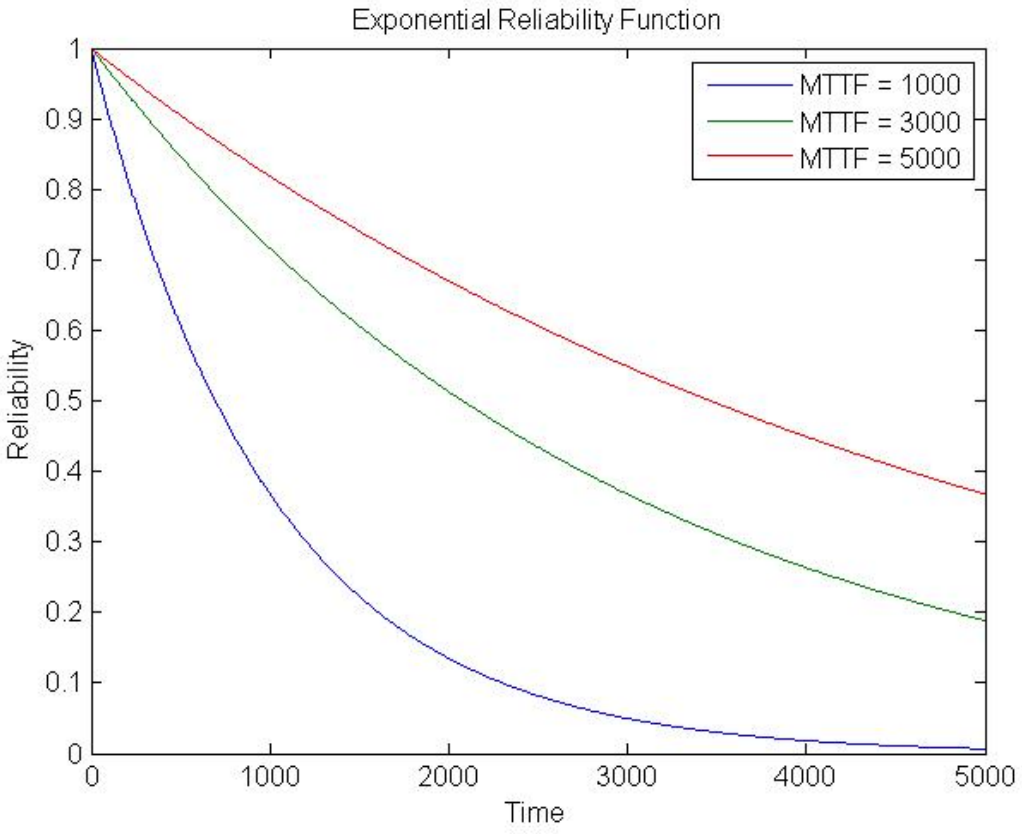


Figure 2. Exponential Reliability Function for several MTTF values

the Weibull, normal, lognormal, and gamma distributions. The Weibull distribution is considered one of the most useful distributions [18:63-74]. The Weibull can model both increasing and decreasing failure rates. Shown in Figure 4, a β less than one results in a decreasing failure rate, while a β greater than one results in an increasing failure rate. In Figure 4, the characteristic life (θ) is held constant. The characteristic life affects the Weibull failure distribution in the same manner as MTTF affects the exponential distribution, elongating the distribution horizontally. The resulting reliability functions are shown in Figure 3. The Exponential distribution is a special case of the Weibull distribution where $\beta = 1$. The simplified formula for MTTF and $\lambda(t)$ in the Exponential and Weibull failure distributions are shown in Table 1.

Table 1. Simplified Formula for MTTF and $\lambda(t)$ for the Exponential and Weibull Failure Distributions

Distribution	MTTF	$\lambda(t)$
Exponential	$\frac{1}{\lambda}$	Constant
Weibull	$\theta\Gamma\left(1+\frac{1}{\beta}\right)$	$\frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1}$

The Gamma distribution can take on shapes very similar to the Weibull. Like the Weibull, the Gamma distribution has two parameters, a shape parameter (γ) and a scale parameter (α). It relates to the exponential distribution through the Erlang-k distribution [18:85], where the Erlang-k is the resulting distribution from the sum of k identical exponential distributions.

The normal distribution is commonly used to model fatigue and wear-out [18:76]. Unlike other reliability distributions, the normal distribution range extends from negative infinity to positive infinity. The lognormal distribution is closely related to the Normal distribution, but is only defined for failure times greater than zero. Neither the lognormal or normal distribution have an analytically defined failure rate [18:83].

The above discussion of distributions is by no means complete. As suggested by [7:433-446] and [26:2-17] there are a number of statistical distributions that can be used to model a given failure profile. These include discrete distributions such as the geometric, hypergeometric, binomial, Poisson, and less frequently used distributions such as the Generalized Weibull or the Gompertz-Makeham. There is no one distribution for all failure data. Rather, the distribution selected should best represent the data.

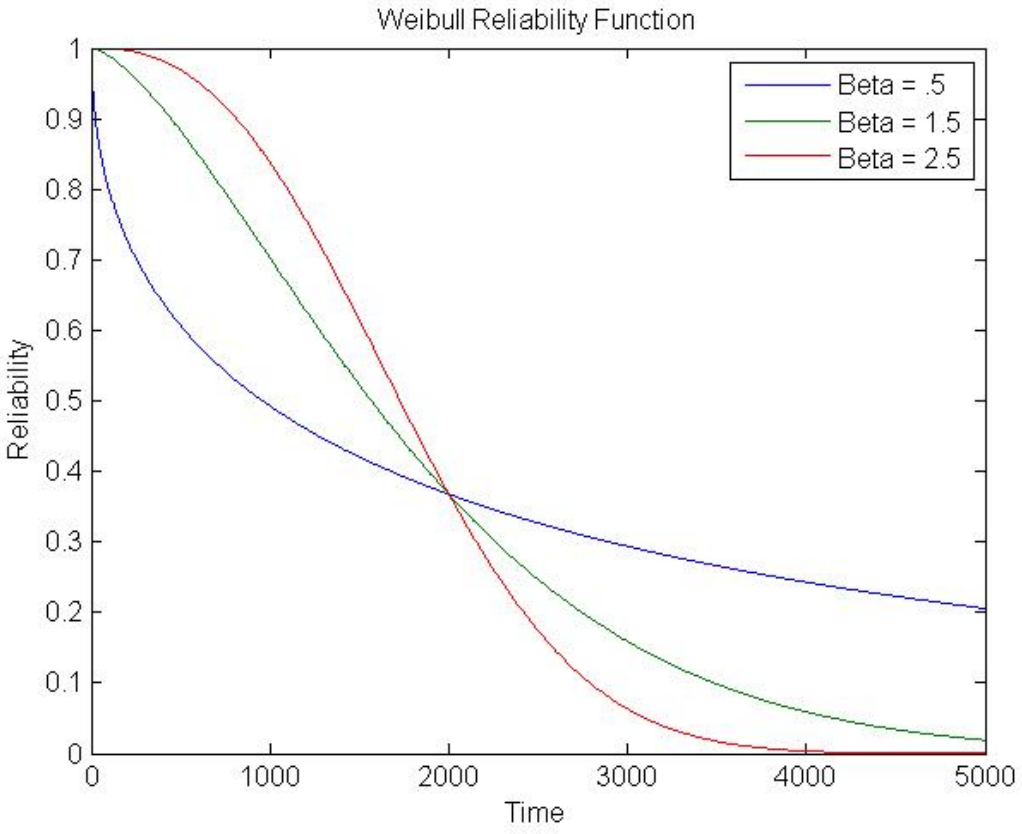


Figure 3. Weibull Reliability Function for several β (shape) parameters and a fixed value of θ

2.2 Basic Empirical Modeling Concepts

2.2.1 Collecting Data.

Data collection is important to modeling the failure profiles of systems or components with unknown failure distributions or for empirically verifying the failure distribution defined for some component or system. However, data collection from testing is not perfect. For example, it is not uncommon for a component or system to survive an entire test without failure or fail for reasons unrelated to the focus of the test. This results in censored data.

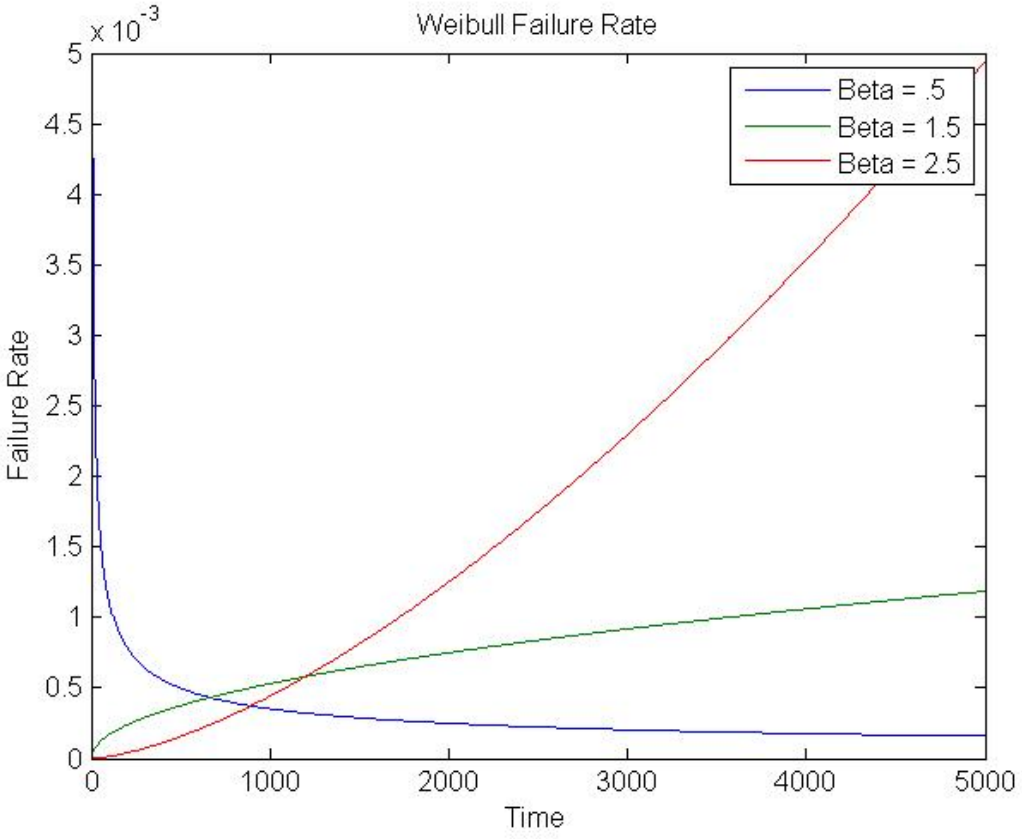


Figure 4. Weibull Failure Rate for several β (shape) parameters and a fixed value of θ

Data censoring can imply that either the unit in question was not run to failure and the test terminated at some time (known as Type I testing), the test was terminated after some pre-determined number of failures or due to a failure from a failure mode different than the failure profile of interest [18:306-308]. For example, a laptop computer is being tested for failures due to the hard drive failing, but in one part of the test the power supply shorts out. Including that power supply failure, unrelated to the hard drive failure of interest, would distort the true hard drive failure profile being estimated.

From the data collected, an empirical distributions can be fit and MTTF and $\lambda(t)$ estimated using the formula given in [18:310]. If the data are incomplete, or censored, then the formula given for MTTF and failure rate estimation are no longer

valid. A point estimate for the reliability at each uncensored failure time can be calculated from the distribution-free model regardless of the presence of censored data. Three common estimation techniques for censored data are given in [18:319-321], the product limit estimator (PLE), Kaplan-Meier form of the PLE (KMPLE), and the rank adjustment method. The PLE, KMPLE, and rank adjustment methods each focus on accommodating censoring in the dataset and providing good fits from data that are not optimal for the empirical distribution fitting effort.

Data censoring enables the extraction of several failure modes from a single test run of a larger system if the failure profiles are independent [18:360]. Any failures due to other reasons are considered censored failures. This analysis can be repeated for every independent failure of interest that was recorded as part of the data collection effort. Furthermore, knowing this improves test economy. Instead of designing a test to look for just one failure mode, it is possible to examine several independent failure modes simultaneously with the results of a single, combined test. This can be useful in system testing when the system involves many components.

2.2.2 Distribution Fitting.

A two step process is used to fit a theoretical distribution to the data. An initial fit is determined from the use of probability plots, which transforms the reliability function and failure times to fit a linear model. If this linear model fits the data "well," there will be little deviation from the line formed by the model and the scatter of the actual data. Different probability plots transform the data differently and result in a better or worse fit depending on the underlying failure profile. The theoretical reliability functions discussed in Section 2.1 all have specific types of graph paper and related transformation functions that can be used to plot data in this manner [18:392-405]. The linear function is estimated via least squares using the transformed data.

The second part of the process of fitting a distribution to sample data is to use the Maximum Likelihood Estimators (MLEs) for known distributions, such as the exponential or the Weibull distribution. In the case of the exponential, a distribution parameter with an MLE is λ. In the case of the Weibull distribution, two distribution parameters β and θ are needed. If the distribution parameters from the linear function and the MLEs are similar, then this is an indication of a good distribution fit.

Once a theoretical distribution has been selected, it must be statistically verified by examining how well the empirical model matches the estimated theoretical model. There are useful statistical tests suggested by [18:435-451] including the Chi Squared Goodness of Fit Test, Bartlett's Test for the Exponential Distribution, Mann's Test for the Weibull Distribution, and the Kolmogorov-Smirnov Test for the normal and lognormal distributions. The Chi Squared and Kolmogorov-Smirnov Tests are also suggested by [6:358-365] and [7:326-332]. Not all statistical tests work in the presence of censored data.

While [18:406] [29:58] suggest that MLEs are the best method for estimation, [3:5-1] [27:136] disagree on the use of MLEs for *small* data sets. In this case, small data sets are defined as those with less than 100 failures. This difference of opinion is not unexpected. Any statistical estimate is improved with more samples. When estimating parameters with small sample sizes, it is important to consider the confidence associated with any estimate, particularly as this applies to the uncertainty associated with the estimated parameter.

2.3 The Basic of Reliability Growth

Reliability growth in the field of engineering development is a set of goals for the change in system reliability over the course of component or system development. It is assessed by testing prototypes of the end design. The failed component or system is

then redesigned, or fixed, to improve its reliability with the goal of reaching a certain target after some amount of test time. The component or system is then tested again.

Reliability growth is closely related to engineering design principles such as Failure Mode, Effects, and Criticality Assessment (FMECA) [18:369] and Environmental Stress Screening (ESS) [7:343]. Any projections made using reliability growth models have no guarantee of being true. Typically, the empirical reliability growth curve is very discontinuous, with gaps or jumps where reliability improved, or degraded, when design changes where made. An example is shown in Figure 5 using the failure rate as the vertical axis. Other common axis values are Mean Time Between Failure (MTBF) or Mean Time Between Maintenance (MTBM).

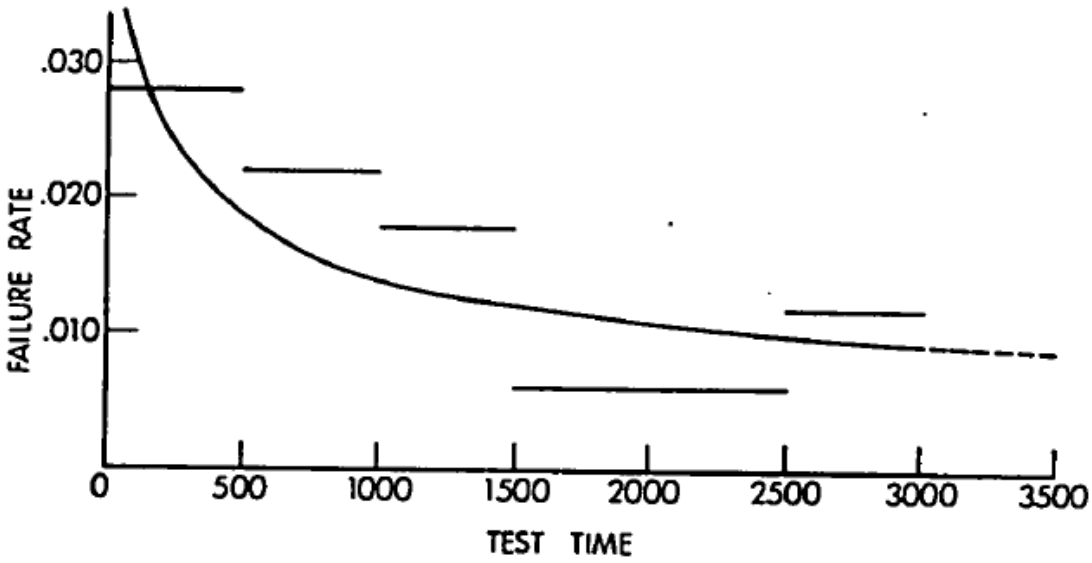


Figure 5. A reliability growth curve using failure rate as an assessment metric [2:71]

In [2:10] the benefits of using reliability growth planning are explicitly laid out in a military context. Initial prototypes of military systems are extremely complex systems and usually involve major technological innovation as part of the requirements of their design. Initial prototypes frequently fail to meet the required reliability requirements. Testing is used to identify problems that may not have been apparent during

development. As testing progresses, failures occur at a component level. Those components are improved, resulting in a decrease in system failures and a corresponding increase, or growth, in system reliability.

From a planning standpoint, reliability growth is used to address the program schedule, amount of testing, resource availability, and realism of the test in achieving the requirements as outlined in the program documentation. This is usually shown as a planning reliability growth curve, which identifies milestones for achievement. The planning growth curve is a guide and can be based on historical information [2:16]. Actual progress is assessed during testing. Failing to meet reliability requirements at one milestone implies that program management may need to take steps to improve reliability.

A number of reliability growth models have been proposed. Both discrete and continuous versions of these have been tabulated in [2:109-129]. However, in more recent literature, most of these models have been replaced by the AMSAA reliability growth model [18:376-381] [7:344-348]. The AMSAA model is used primarily to assess reliability within the program test phases. Often components within a system are assumed to follow an exponential failure distribution. When component failures result in system failures, exponential times are found between system failure occurrences, which implies the failure counts within some period are distributed according to a Poisson distribution. As the component reliability improves (due to the repair process), the system failure Poisson process also changes. This is a non-homogeneous Poisson process (NHPP) for system failure. The AMSAA model employs the NHPP assumption. Furthermore, the AMSAA model is still recommended as *best practice* even with deficient data [3:9-1], such as mixed failure modes or missing data.

Recently, as a result of the DOD mandate discussed in Section 1.2 there have been a number of extensions to the AMSAA model proposed by Crow in [11] [12] [13] [14]

and [15]. These extensions include a methodology to better account for operational testing mission profiles, a redefined failure mode criteria to include failures induced by human factors, a more flexible set of test methodology and accompanying Crow-AMSAA model formulation, and a method to quantify the uncertainty of the point estimates used in the Crow-AMSAA model.

One of the challenges with assessing system reliability in the operational test phase is managing the structure of mission profiles. [11] discusses a methodology for grouping data in accordance with convergence points. Convergence points are based on straight line averages of long-term testing. However, because operational test phases usually have specific goals, this structure changes. The convergence points are taken when the short-term average matches the long-term average. This allows MTBF to be calculated at that point, instead of having to wait until the end of the test phase.

In 2010, [12] suggests that failure mode identification include those failures related to human factors. Specifically, most human-factor influenced failure modes are actually fixed immediately. Furthermore, delayed corrective action does not always occur at the corrective action phase. Test schedules or technology maturity can influence when the corrective action for a particular failure mode is addressed. This expanded Crow-AMSAA model allows for more failures to be identified and fixed during the test program as a whole.

Much of the work in [11] and [12] is replicated and summarized in [14], which incorporates the information from [13]. Crow, in [14], also includes an explanation of the use of the Crow-AMSAA model in the environments of Test-Fix-Test (corrective action is performed immediately upon failure), Test-Find-Test (corrective action is performed at the end of the test phase), and Test-Find-Fix-Test (minor corrective actions are performed immediately, but major corrective actions are performed at the

end of the test phase). Historical documentation of Fix-Effectiveness-Factors (FEFs) are presented for use in prediction models, based on FEFs discovered in past systems. A FEF is the expected decrease in failure mode intensity after a corrective action is implemented, it is an assumed measure of how effective fixes are. Not all corrective actions result in improvements to component or system reliability, thus this value is typically 0.8.

In 2012, [15] proposes another methodology to account for uncertainty in the point estimates of MTBF given by the Crow-AMSAA model. Specifically, each point estimate comes from a Poisson sampling distribution. To truly demonstrate a certain MTBF that always exceeds a threshold, the actual design MTBF has to be greater than the threshold MTBF. The question is, how much greater? The methodology presented attempts to overcome this by using a combination of reliability growth testing and demonstration testing to get a specified confidence on design MTBF. The proposed methodology requires that reliability growth and demonstration test conditions be similar, which may not always be possible and is a limitation.

While AMSAA remains the model of choice for reliability growth modeling and assessment, there is new research that attempts to overcome some of its shortcomings. One of the key metrics in the AMSAA model is MTBF. [25] suggests a Bayesian based estimation methodology that takes into account test profile characteristics and aggregates all component, subsystem and system level data together to form an estimate that is not based on MTBF.

2.4 Summary

Reliability theory and the concept of failure distributions yield component level insights which can be aggregated up to system level. The methodology to perform

failure distribution analysis is well defined in the literature, including procedures for estimating failure distributions from single or multiply censored data.

III. Methodology

3.1 Past Analysis for Similar Systems

The concept of statistical rigor in operational testing is a relatively new initiative. This new mandate contrasts sharply with how operational test was typically conducted in the past. Two past acquisition efforts, the KC-10 and the KC-135, were examined to gain insight into past practices during operational testing.

3.1.1 KC-10 Operational Testing.

In the 1981 Follow-On Test and Evaluation Plan for the KC-10 [9], the exact methodology for assessment of reliability is unclear. Reliability data was primarily collected from maintenance forms that logged maintenance actions as a result of both planned and unplanned maintenance. Evaluation criteria were established using a three tiered system of threshold, standard, and goal. The threshold represents the minimum value considered acceptable, while the goal represents where the system should be in terms of reliability. The standard tier ranges in rating and action on the system. Figure 6 depicts the tiered criteria used.

This criteria methodology was used throughout the test program, including evaluating the reliability of the system. MTBM was the backbone of the reliability analysis performed and was selected based on similar platforms such as the KC-135, C-141, and DC-10-30CF [9:103-108]. Maintenance data was collected from form records compiled during testing.

As expected in any flight test program, there must be an extrapolation of the empirical results to those achievable in a mature system. In the KC-10 TEMP [9:17], the Logistics Composite Model (LCOM) was suggested to project reliability to maturity. LCOM is a network flow simulation model that uses stochastic probabilities to cap-

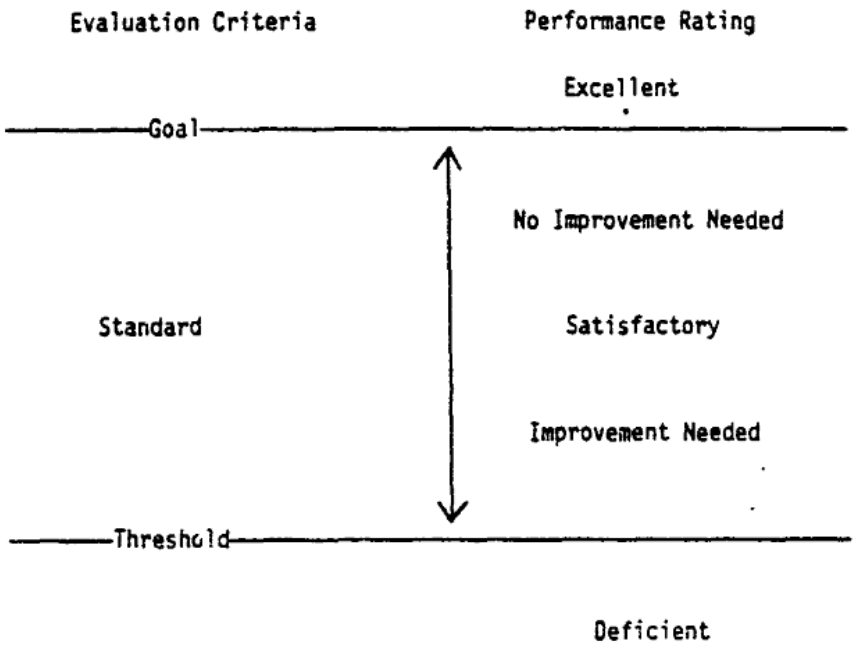


Figure 6. The general test evaluation criteria as given in [9:30]

ture the variable nature of system performance. However, LCOM requires significant user input to provide responses. These user inputs includes estimates for MTBF and MTBM [24:47].

An alternative to simulation model projections of reliability is comparable analysis. Comparable analysis uses any established reliability information on similar parts in similar conditions and combined with field experience, estimates how the new technology should behave over time [24:58]. This information was used to compute the expected value of each node in the LCOM network [24:166].

Ultimately, the results are condensed down to a single evaluation number for use in the evaluation criteria defined by the system. Examples of this are given in [28:63-65]. These condensed results are not projections of future system performance, but

rather are the results of a second phase of testing on a particular troubled system or systems.

3.1.2 KC-135 Operational Testing.

KC-135 Operational Testing has actually been an on-going process, with numerous systems undergoing replacement or upgrade from their original designs. A survey of these operational test plans and reports indicates an evolutionary trend in the analysis methods used in evaluation of failure data. Four reports are examined, providing a glimpse into the methodology used by the KC-135 test teams.

In a 1981 report on the KC-135 Weapons System Trainer, MTBF is defined similar to that for the KC-10 as the average value of time on test over the number of failures [8:40]. This information is derived from similar maintenance documentation as in the KC-10 test program and uses a similar assessment metric. Interestingly, the KC-135 MTBM value is artificially deflated [8:43] by repairs that involved "reinitializing the computer." The reports indicate that when the data are not considered, the reliability estimates fall within the standard and goal levels, but those actual results are not provided.

A 1997 KC-135 test plan does not mention the threshold, standard, and goal levels of acceptance. Instead, the measures of performance (MOPs) are evaluated on a criteria of met, failed, or did not test [10:42]. This alternative assessment metric is degraded from the tiered metric; the tiered metric incorporated variability while the latter is a single number that the system must obtain to pass. However, an explicit definition for subsystem reliability is provided [10:46]. MTBM is used to evaluate suitability but is not explicitly defined in this plan.

A 2004 report on the operational testing of the Global Air Traffic Management system indicates improvements. When evaluating MTBF, data from training time

was incorporated to get a better estimate, but this causes the system to fail to meet its target value [22:74]. However, when the training data was isolated and removed, an empirical instantaneous MTBF met the requirements but with a significant margin of error. The removal of the training data points affects the associated confidence intervals generated, but this point was not mentioned in the report.

More recent test planning of the KC-135 Communication, Navigation, and Surveillance/Air Traffic Management Block 45 Modification in 2011 explicitly lays out a methodology for the analysis of failure data. The assessment metric recommended to evaluate reliability was Mean Time Between Maintenance-Total (MTBM-T). It was evaluated using Equation 3.1 from [17:7], where OH is defined by Equation 3.2. Definitions of the acronyms used are compiled in Table 2.

$$MTBM-T = \frac{\sum OH}{\sum MA_C} \tag{3.1}$$

$$OH = FH + \frac{GH}{1.2} \tag{3.2}$$

Table 2. Acronyms used in Equation 3.1 and Equation 3.2 from [17]

Acronym	Definition
OH	Operating Hours
FH	Flight Hours
GH	Ground Hours
MA_C	Corrective Maintenance Actions for Type 1,2 and 6 failures

Furthermore, this recent test plan adds some statistical rigor to the analysis by considering confidence intervals on the evaluation criteria, as shown in Table 3. This is especially important in small sample size testing, where confidence intervals may be large.

Table 3. Evaluation Criteria for MTBM-T in [17]

Rating	Mean Time Between Maintenance-Total
Satisfactory	Lower 90% confidence bound above target
Marginal	90% confidence bounds contain target
Unsatisfactory	Upper 90% confidence bound below target

3.2 KC-46 Planned Reliability Analysis

The KC-46 TEMP attempts to improve upon the reliability assessment practices used in past programs, by using more quantitative measurements. For example, there is a quantified break rate shown in Equation 3.3. Given clear agreement on what constitutes a break and a mission, this metric is objective. Further, this break rate is assumed constant through the evaluation periods so long as the system remains unchanged.

$$\text{Break Rate} = \frac{\text{Total Mission Breaks}}{\text{Number of Missions Flown}} \tag{3.3}$$

However, the KC-46 requirements are still deterministically based. The TEMP defines a specific required break rate, as well as specific requirements for many of the reliability requirements. Each defines a target value that must be met by the test program.

The KC-46 TEMP calls out ANSI GEIA-STD-0009, an international standard [5:89]. This standard calls explicitly for the:

> *"Engineering analysis and test data identifying the system/product failure modes and distributions."* [1:30,33]

As a result, the KC-46 test program must have the data to support identifying system failure modes and distributions and this data must be agreed upon and ana-

lyzed to estimate the underlying component failure distributions. Section 3.3 suggests a methodology that focuses on this aspect of a reliability evaluation program.

3.3 Improving the Reliability Evaluation Program

This thesis suggests approaches to increase the statistical rigor in KC-46 operational test. These suggestions are: identification and validation of failure distributions, proper utilization of censored data, and using sound statistical methods in assessment of reliability growth.

3.3.1 Identification and Validation of Failure Distributions.

In the current KC-46 methodology, as derived from [5], an assumption is made that system or component failure rate is constant which would imply an exponential failure distribution. Further, the system is not considered to be in the burn-in or wear-out phases as part of the system assessment process and we assume that the system is in normal operating condition. The first part of the proposed methodology takes this assumption and provides the statistical backdrop to for it. It also provides a process for fitting a revised failure distribution in case the assumption is deemed invalid.

We start by considering a non-parametric distribution formed using Equation 3.4 [18:309]. This transforms the cumulative reliability so we may plot $\widehat{R}(t_i)$ versus time. It is a variation of the rank increment method where i_{t_i} is the *ith* failure in an ordered list of failures and n is the total number of failures.

$$\widehat{R}(t_i) = 1 - \frac{i_{t_i} - 0.3}{n + 0.4} \tag{3.4}$$

A probability plot is developed using the transformations in Table 4 for the assumed exponential distribution. These transformed data are modeled using simple

linear regression and plotted as a straight line, when the exponential failure distribution holds.

Table 4. Probability Plotting Functions as given by Ebeling [18:393-398]

Distribution	x(t)	y(t)
Exponential	t	$ln\left[\frac{1}{1-\widehat{F}(t)}\right]$
Weibull	$ln(t)$	$ln\left[ln\left[\frac{1}{1-\widehat{F}(t)}\right]\right]$

Based on the least squares fit, statistical assumptions are verified per [16:129-132]. The estimated parameters are a, the intercept, and b, the slope, of the fitted line, $y = a + bx$. These estimates for a and b also yield parameter estimates for the exponential failure distribution. Equation 3.5 is the relation between the fitted line slope and the estimate for exponential failure rate. This estimate compares to the MLE in Equation 3.6, where r is the total number of failures and T is the total time of the test.

$$\widehat{\lambda} = b \tag{3.5}$$

$$\widehat{\lambda} = \frac{r}{T} \tag{3.6}$$

There are numerous choices for goodness of fit testing. The key limitation is usually sample size and the condition of the data. Bartlett's test is recommended here because it is specifically designed to test whether the failure data are exponentially distributed based on the following hypotheses:

H_0 : Failure times are exponential

H_1 : Failure times are not exponential

A failure to reject H_0 implies the component failure rate is constant, which aligns with the exponential failure distribution assumption. The test statistic is given in Equation 3.7, where t_i is the time of failure of the ith unit and r is the number of failures observed.

$$B = \frac{2r[\ln((1/r)\sum_{i=1}^{r} t_i) - (1/r)\sum_{i=1}^{r} \ln t_i]}{1 + (r+1)/(6r)} \tag{3.7}$$

The test statistic B is compared to the chi-square distribution with $r-1$ degrees of freedom, as in Equation 3.8:

$$\chi^2_{1-\alpha/2,r-1} < B < \chi^2_{\alpha/2,r-1} \tag{3.8}$$

As with most statistical tests, more samples means more accuracy in the test [18:435]. Bartlett's test needs around 20 failure points for an adequate power [18:443].

If we fail to reject the null hypothesis, then the current $MTTF$ used in [5] is a valid estimator for the assumed failure distribution. However, rejecting the null hypothesis requires a change in methodology to obtain an improved answer. The Weibull failure distribution is a good choice here since it models both increasing and decreasing failure rates and includes the exponential distribution as a special case (shape parameter $\beta = 1$).

Under a Weibull failure distribution assumption, the Weibull transformation from Table 4 is used to provide a linear model of reliability versus time. The Weibull distribution has two parameters, the shape parameter (β) (Equation 3.9) and the characteristic life (θ)(Equation 3.10). The MLEs are given in Equation 3.11 for β and Equation 3.12 for θ. Equation 3.11 accommodates time censored data with t_s, the time of the right censored data, and n, the total number of units on test. In the presence of complete, non-censored, data: $n = r$ and t_s is undefined.

$$\widehat{\beta} = b \tag{3.9}$$

$$\widehat{\theta} = e^{-\frac{a}{\widehat{\beta}}} \tag{3.10}$$

$$\frac{1}{r}\sum_{i=1}^{r} \ln\, t_i = \frac{\sum_{i=1}^{r} t_i^{\widehat{\beta}} \ln\, t_i + (n-r)t_s^{\widehat{\beta}} \ln\, t_s}{\sum_{i=1}^{r} t_i^{\widehat{\beta}} + (n-r)t_s^{\widehat{\beta}}} - \frac{1}{\widehat{\beta}} \tag{3.11}$$

$$\widehat{\theta} = \left\{ \frac{1}{r} \left[\sum_{i=1}^{r} t_i^{\widehat{\beta}} + (n-r)t_s^{\widehat{\beta}} \right] \right\}^{\frac{1}{\widehat{\beta}}} \tag{3.12}$$

Mann's Test for the Weibull Distribution can be used to test whether the failure times follow a Weibull failure distribution. The test statistic is calculated using Equation 3.13 where k_1 is the integer portion of $\frac{r}{2}$, k_2 is the integer portion of $\frac{r-1}{2}$, $M_i = Z_{i+1} - Z_i$, $Z_i = \ln\left[-\ln\left(1 - \frac{i-.5}{n+.25}\right)\right]$, i is the ith failure, and n is the total number of failures. The test statistic is compared to a critical value from the F-distribution with $2k_2$ degrees of freedom in the numerator and $2k_1$ degrees of freedom in the denominator.

$$M = \frac{k_1 \sum_{i=k_1+1}^{r-1} [\ln t_{i+1} - \ln t_i / M_i]}{k_2 \sum_{i=1}^{k_1} [\ln t_{i+1} - \ln t_i / M_i]} \tag{3.13}$$

Once the underlying distribution is identified and validated, the hazard rate function for the that particular distribution to be used to estimate the system break rate. An example showing how incorrectly assuming the wrong underlying distribution can lead to inaccurate information is provided in Chapter IV.

3.3.2 Effective Data Censoring.

Data censoring can be used to extract failure data for multiple failure modes among the components in a large system. Any data censored in this manner are known as multiply censored data. Multiply censored data involves test units with different operating, or test, times, and may have even started testing at different times.

As KC-46 completes missions and accrues flight test hours, components within the system accrue operating hours. Tracking component-level hours, in addition, to system hours, provides a means to capture the multiply censored data from which component reliability distributions may be estimated.

Properly using multiply censored data in reliability testing starts at data collection. As the test executes, the failure data must include not only a time, but also a mode of failure. Care must be taken to collect data only for *independent* failure modes among the system components of primary interest. The data are then filtered by failure mode. A failure distribution for each failure mode is constructed from known failures and accounts for the presence of the censored units. Any units that do not fail are listed as censored by time when the test terminates (this is known as type II censoring).

There are three methods for estimating the reliability functions, the PLE, KM-PLE, and the rank adjustment method. PLE assumes that if a unit is censored it has no effect on the reliability of the system. The rank adjustment method assumes that the censored unit affected system reliability and estimates this effect by adjusting its rank. The assumption is made that the censored unit would have failed on or after the censored time. Both the PLE and the rank adjustment method also assume that the last failure has some non zero reliability. The rank adjustment method is

used here as it assumes that all units contribute to system reliability even if they are censored before they can yield information.

The rank adjustment method is not a complex method. A table with six columns can be created and failure data inserted (ranked by time). The first three columns of the table are the number or other identifier of order, the failure mode, and failure time. Rank increment, rank, and estimated reliability are the remaining three columns. Rank increment changes if there is a censored data point. Rank increment is calculated using Equation 3.14. Rank is calculated each time there is a failure and uses the results of rank increment, calculated using Equation 3.15. Finally, a point estimate for reliability is calculated using Equation 3.4.

$$\text{Rank Increment} = \frac{(n+1) - i_{t_{i-1}}}{1 + \text{number of units beyond present censored unit}} \tag{3.14}$$

$$i_{t_i} = i_{t_{i-1}} + \text{rank increment} \tag{3.15}$$

An example rank adjustment method computation for two failure modes is shown in Table 5.

We next estimate parameters for the hypothesized failure distribution. The discussion here is again limited to the exponential and the Weibull distributions. Equation 3.16, where F is the set of failure indices, C is the set of censored indices and t_i remains the ith failure time, is modified from its original form given in Section 3.3.1 for the presence of censored data.

$$\widehat{\lambda} = \frac{r}{\sum_{i \in F} t_i + \sum_{i \in C} t_i^+} \tag{3.16}$$

Table 5. Sample Rank Adjusted Failure Data [18:321]

Identifier	Failure Time	Failure Mode	Rank Increment	Rank	$\widehat{R}(t_i)$
1	150	A	1		0.933
2	340	B			
3	560	A	$\frac{11-1}{1+8} = 1.111$	$1 + 1.111 = 2.111$	0.826
4	800	A		$2.111 + 1.111 = 3.222$	0.719
5	1130	B			
6	1720	A	$\frac{11-3.222}{1+5} = 1.2963$	$3.222 + 1.2963 = 4.518$	0.594
7	2470	B			
8	4210	B			
9	5230	A	$\frac{11-4.518}{1+2} = 2.16$	$4.518 + 2.160 = 6.679$	0.387
10	6890	A		$6.679 + 2.160 = 8.839$	0.179

The MLE for β is calculated by solving Equation 3.17, where F is the set of failure indices and i is the set of all failure and censored indices, for β using a non-linear solution method.

$$\sum_{i \in F} \frac{\ln t_i}{r} = \sum_{all\ i} t_{t_i}^{\widehat{\beta}} \ln t_i \left[\sum_{all\ i} t_{t_i}^{\widehat{\beta}} \right]^{-1} - \frac{1}{\widehat{\beta}} \tag{3.17}$$

Equation 3.18 is used to estimate θ.

$$\widehat{\theta} = \left[\sum_{all\ i} \frac{t_i^{\widehat{\beta}}}{r} \right]^{\frac{1}{\widehat{\beta}}} \tag{3.18}$$

Significance tests use the very general likelihood ratio test. This test is challenging to do manually when the data are censored. Various statistical packages, such as JMP, implement these tests as well as the parameter estimation routines.

3.3.3 Reliability Growth.

Knowing and characterizing failure distributions is important in characterizing component and, ultimately, system reliability. These characterizations are useful when applying these assessments to reliability growth. The AMSAA reliability growth model serves as a prediction model and an assessment tool with which to view progress towards the reliability predictions provided by that model.

The first step in using reliability growth is to define the assessment metric. For instance, MTBF or MTBM are common metrics. Any metric that gives some numerical evaluation of the reliability of the system at some point in time can be used. Since the metric affects the definition of failure, metric selection will require agreement among the assessment team.

Prior to defining the prediction model, test procedures must be identified, corrective actions defined, and corrective action time lines agreed to. Corrective action time

lines can fall into, Test-Fix-Test, Test-Find-Test, or Test-Fix-Find-Test approaches. The most likely candidate is latter, Test-Fix-Find-Test.

The Test-Fix-Find-Test approach implies that any major corrective actions are delayed until the end of a test phase, while less than major actions are corrected immediately. The immediate corrective action may have a small effect on the overall component and system reliability, but the major corrective actions may have large effects on system reliability, resulting in the discontinous "jump" in the growth curve previously discussed. Test-Fix-Find-Test is a flexible test technique and works well in time constrained testing.

Two kinds of prediction reliability growth curves are used. The first is the idealized reliability growth curve, which features a smooth approximation of the underlying reliability trend. This curve is based on growth targets, usually shown as discontinuous jumps in reliability. An example of a simple curve is shown in Figure 7, where the initial MTBF is one hour and the final goal is an MTBF of 3. This curve is generated using the Planning Model Based on Projection Methodology (PM2) provided by the Army Material Systems Analysis Activity. This model is based upon the AMSAA Projection Model.

The curve uses an assumed fix effectiveness factor (FEF) of 0.8 as suggested by Crow in [13]. Corrective action is assumed to take place at the end of each test phase and no corrections are assumed made in a test phase. It is also assumed that the failure rate is constant during each phase of testing.

The planning curve represents system reliability **goals**. These goals should be met during the test phases and corrective actions employed during system development if the system falls short of its defined reliability goals. Actual progress is tracked by comparison with this planning curve. Failing to meet goals require rethinking

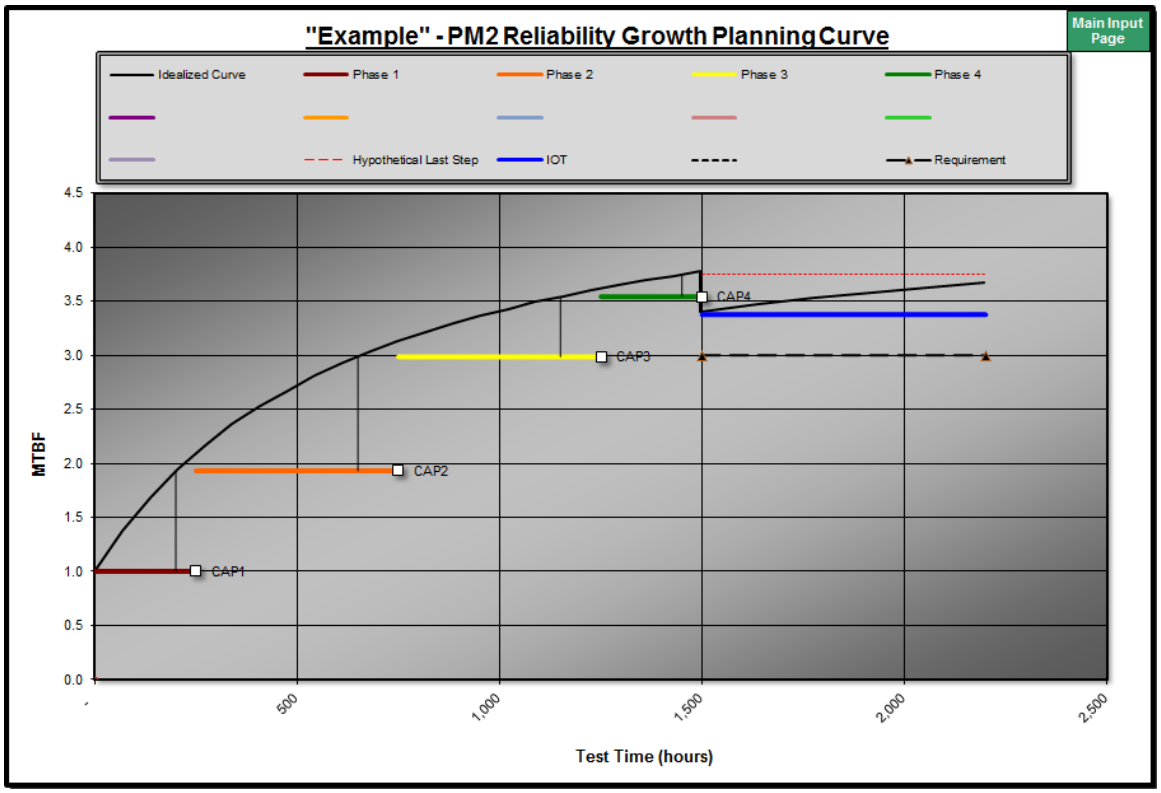


Figure 7. An Example Reliability Growth Planning Curve

the associated planning curve and management strategy for meeting those reliability goals.

Assessment metrics are defined and agreed to early in a program but should be evaluated continuously throughout the development process. Each test phase has the potential to change the metric used, since data from one test phase is typically not used in another. If reliability grows slower than expected, a new prediction model or a new management strategy may be needed. These are decisions made by each program but should be made based on fully understanding the system and how its reliability is being assessed.

3.4 Summary

A brief review of the past and current reliability assessment methodology in tanker programs revealed that assessment methods have improved over time. The goal of improved statistical rigor in the tanker flight test programs is mandated by the OSD. Thus, there is still room for improvement in acquisition program testing and assessment. In Section 3.3, a common assumption of constant failure rate is tested from a statistical perspective. If this assumption is validated, then the current methodology used in tanker test programs is fine. However, if the assumption is false, or at least found quite tenuous, then a methodology incorporating the Weibull distribution is proposed, a methodology that handles both uncensored and censored data.

IV. Notional Implementation

This chapter reinforces the ideas previously discussed. Sections 4.1 and 4.2 address the modeling of component failure distributions and why the Weibull framework should be a preferred approach. Section 4.3 addresses reliability growth model assessment and the potential use of an empirical model of system-level failures.

4.1 Example: Complete Data Analysis

The source data for this notional example is a component failure distribution from a Weibull distribution with $\beta = 0.8$ and $\theta = 4000$. The data are complete and uncensored, and components are run to failure. An incorrect assumption is to assume all components have constant failure rate as found with an assumed exponential failure distribution. While not required to perform Bartlett's test on this hypothesis, a probability plot is generated from which a linear function is fit to find an initial set of estimates for the parameters of the hypothesized exponential distribution. This is shown in Figure 8. These least squares parameter estimates are compared to the MLE for the same data.

The estimated MTTF is 5464 from the linear equation. The MLE calculation, yields an estimated MTTF is 4858. There is a considerable difference between the two values. This is the first indication that the exponential may not be the correct assumption, even though the R^2 value obtained by the least squares analysis is not bad. Bartlett's test test statistic is $B = 132.18$. This is compared to the 95% upper and lower bounds of 73.36 and 128.42 respectively.

There is evidence to reject the hypothesis that the data are exponentially distributed and thus the failure rate is not constant, as expected based on the data generated. The more robust Weibull distribution is now considered. A probability

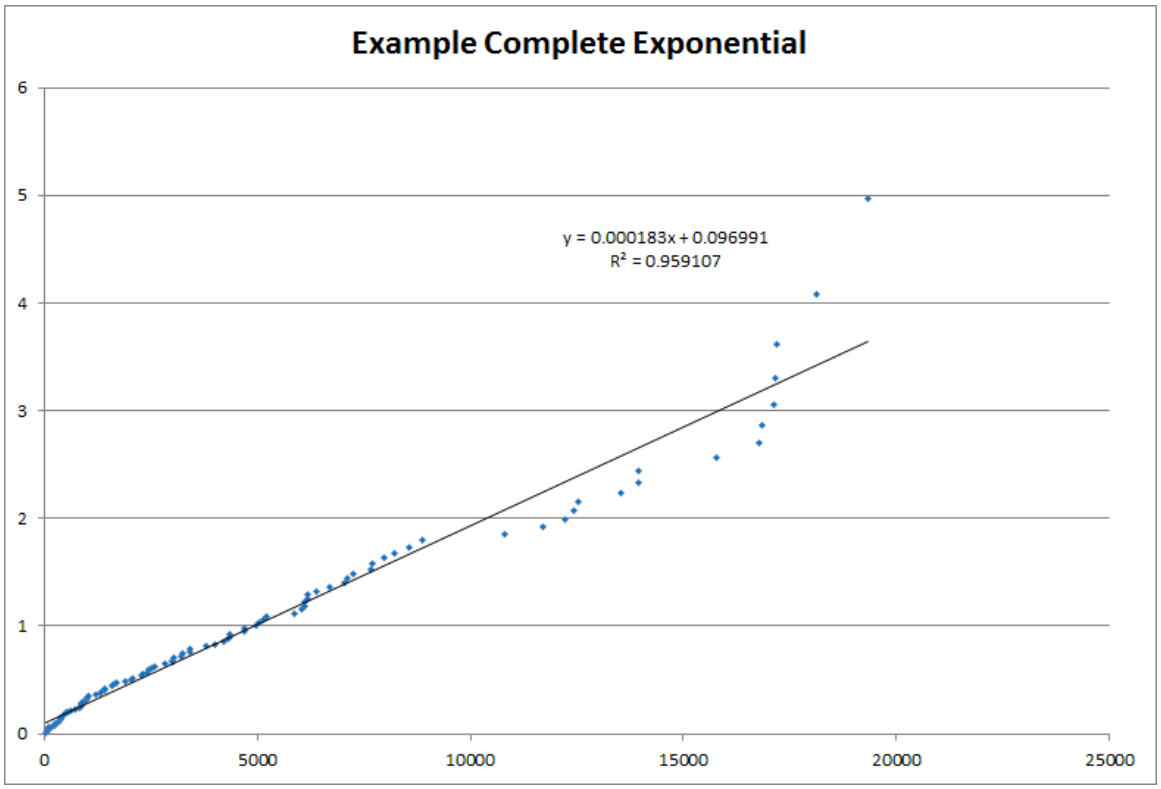


Figure 8. Least Squares fitted Exponential Probability Plot for an Underlying Weibull Distribution

plot is generated using the transformations from Table 4 and the linear line fit. The resulting plot is shown below in Figure 9. There is an improvement to R^2 and the line looks to be a better fit. Table 6 indicates reasonable agreement between the linear and MLE methods.

Table 6. Comparison of Parameter Estimates for the Weibull Distribution hypothesis

	β	θ
Linear	0.814	4508.134
MLE	0.853	4491.438

Mann's test is used to test the Weibull failure distribution hypothesis. The test statistic for Mann's test is 0.39, much lower than the critical value of 0.72. There is

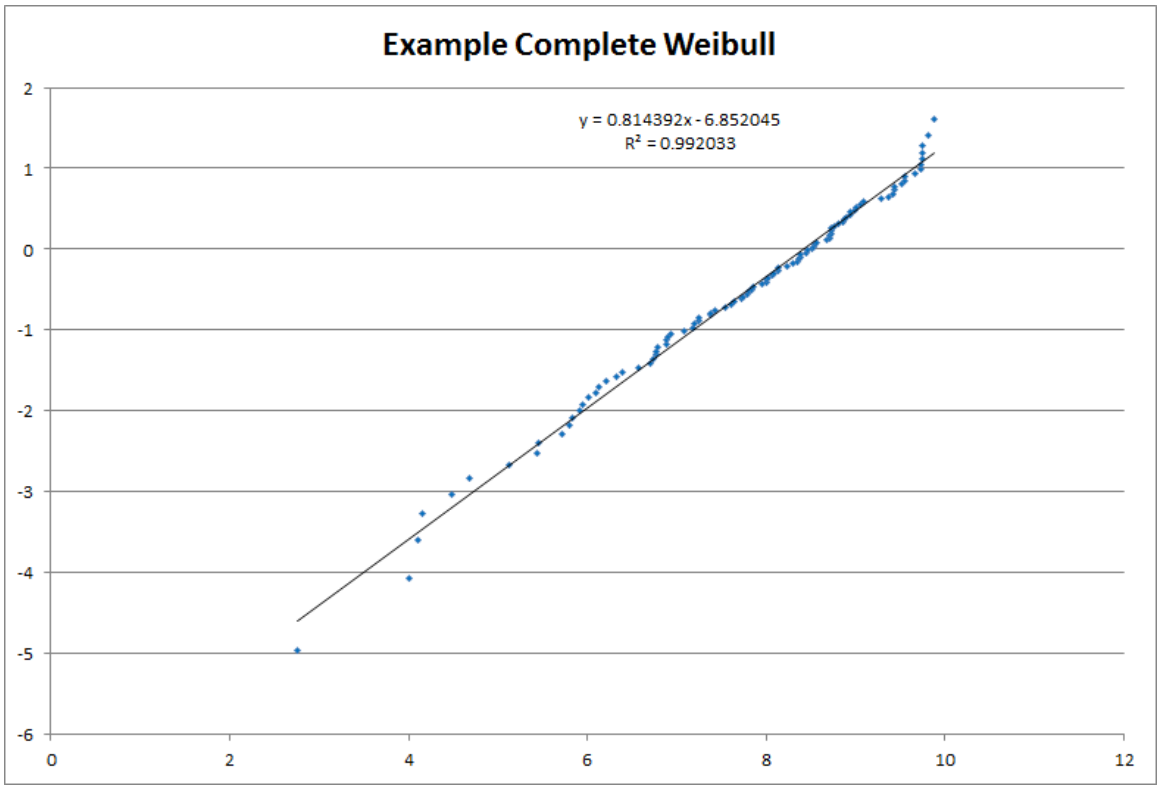


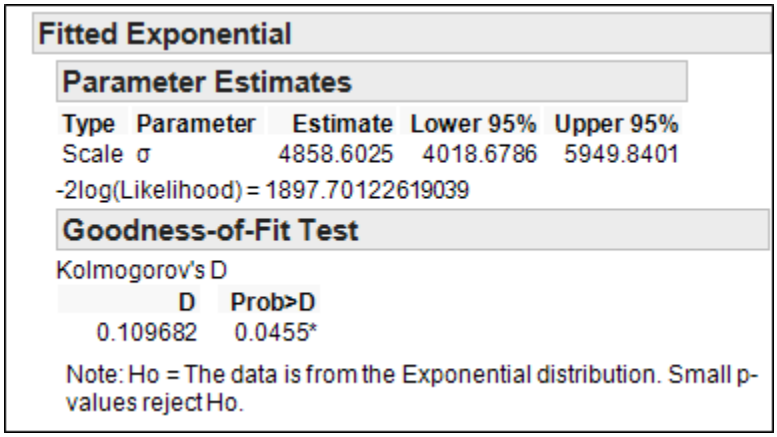
Figure 9. Least Squares fitted Weibull Probability Plot for an Underlying Weibull Distribution

insufficient evidence to reject our hypothesis that the data comes from the Weibull distribution.

This changes the estimate for MTTF from approximately 4858 to approximately 4869. While this change is minor, the background knowledge that failure rate is actually decreasing could indicate that this failure does not need corrective action to improve, only time. It could also indicate that something in the system has changed that has not been accounted for in the test program.

Statistical software packages, such as JMP, are preferred when performing the analysis. JMP, for instance, estimates the parameters for the defined distribution and tests that assumption using a goodness of fit test it finds most suitable. Results for the current example are shown in Figure 10. While JMP uses the Kolmogorov's

D test instead of Bartlett's, it finds sufficient evidence to reject the hypothesis that the distribution is exponential.



Fitted Exponential

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	σ	4858.6025	4018.6786	5949.8401

-2log(Likelihood) = 1897.70122619039

Goodness-of-Fit Test

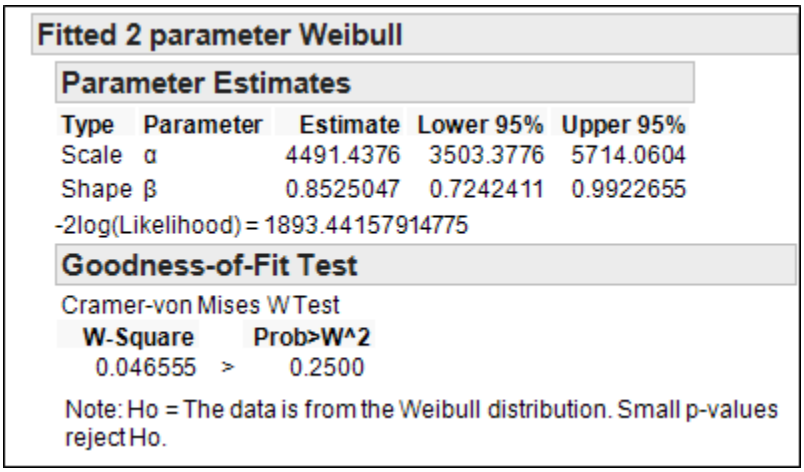
Kolmogorov's D

D	Prob>D
0.109682	0.0455*

Note: Ho = The data is from the Exponential distribution. Small p-values reject Ho.

Figure 10. JMP Distribution Fitting for an Exponential Distribution for an Underlying Weibull Distribution

Figure 11 repeats the JMP analysis but now assuming the more general Weibull failure distribution. The large p-value in Figure 11 means the JMP test results cannot cause a rejection of our assumption of a Weibull distribution.



Fitted 2 parameter Weibull

Parameter Estimates

Type	Parameter	Estimate	Lower 95%	Upper 95%
Scale	α	4491.4376	3503.3776	5714.0604
Shape	β	0.8525047	0.7242411	0.9922655

-2log(Likelihood) = 1893.44157914775

Goodness-of-Fit Test

Cramer-von Mises W Test

W-Square		Prob>W^2
0.046555	>	0.2500

Note: Ho = The data is from the Weibull distribution. Small p-values reject Ho.

Figure 11. JMP Distribution Fitting for a Weibull Distribution for an Underlying Weibull Distribution

This example shows that presuming an exponential distribution can be misleading. The Weibull distribution is more general. In reality, components may be built

and determined to follow some defined failure distribution, such as the exponential. However, actual data will not follow such distributions exactly. This result is further complicated by small sample sizes. Statistical analysis should be conducted using more general assumptions, in this case, a Weibull failure distribution.

4.2 Example: Censored Data Analysis

It is rare to have complete data where all tested components are run to failure, particularly in the high reliability demanded by most government acquisition projects. This introduces censoring into the data analysis. Analysis incorporating data censoring is a powerful tool. Notional examples of singly and multiply censored data are considered.

4.2.1 Singly Censored Data.

For this example, a Weibull random variate with a $\beta = 0.8$ and $\theta = 4000$ is used to generate 100 failure points, with the test time terminated at 5000 time units. This yields 29 censored data points. The component failure distributions are assumed to be exponential, at least initially, which is a common assumption. A probability plot and estimates for the failure rate based on the linear functions and MLEs, are generated. The plot is shown as Figure 12.

The linear function estimate for MTTF of 3773 is very close to the MLE estimate of 3818. The statistical goodness of fit tests used previously do not work in the presence of censored data. The R^2 value of 0.994 does indicate the exponential distribution is a good fit.

A Weibull probability plot of the same data is shown in Figure 13, along with the associated linear regression fit. The R^2 value of this Weibull plot is degraded slightly

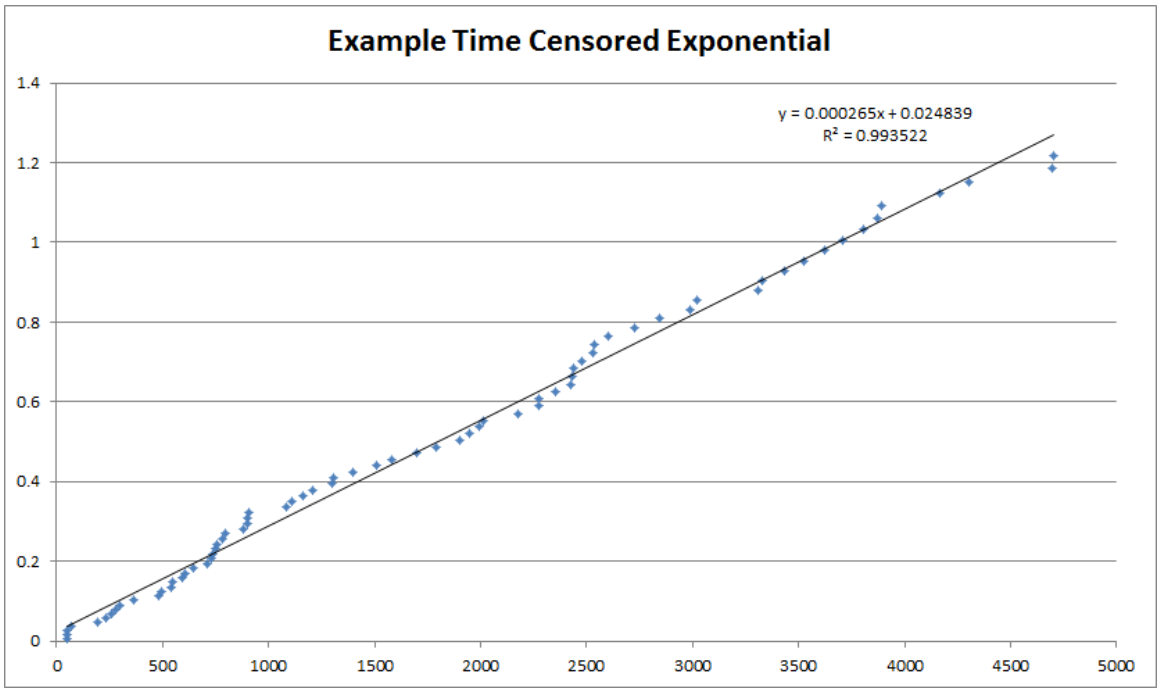


Figure 12. Exponential Probability Plot for Singly Censored Data from an Underyling Weibull Distribution

from that shown in Figure 12. There is also some lack of fit near the left end of the chart, something that does not occur on the exponential plot.

The MLEs for the Weibull coincide with their linear equation approximations as noted in Table 7. In this case, the failure distribution is considered to be exponential despite the known fact that the data were indeed generated from a Weibull distribution.

Table 7. Comparison of Parameter Estimates for the Weibull Distribution hypothesis with Singly Censored Data

	β	θ
Linear	0.966	3636.363
MLE	0.930	3844.778

This example attests to the noise found in empirical data. We know the data are Weibull, but pass a test on exponential. The nearness of the $\beta = 0.8$ actual and

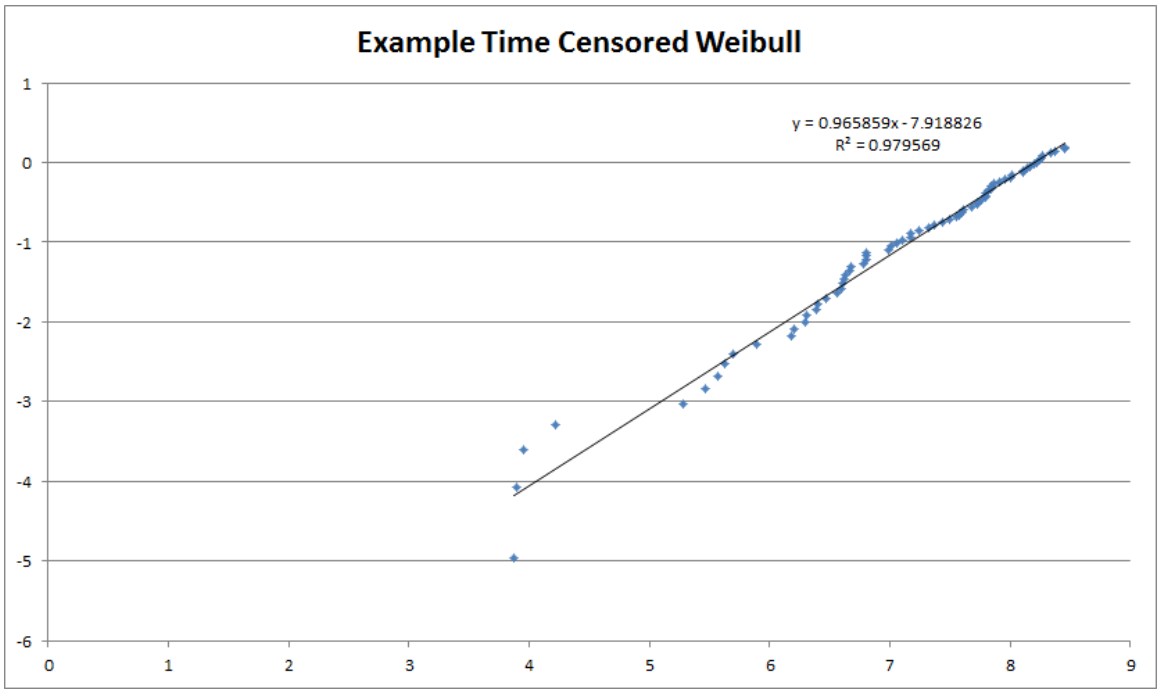


Figure 13. Weibull Probability Plot for Singly Censored Data from an Underyling Weibull Distribution

$\beta = 1.0$ estimated can lead to an erroneous assumption of a constant failure rate among the components. The take away is to use a Weibull in general but be ever cognizant of the range of distributions parameters possible when small sample size data analyzed.

4.2.2 Multiply Censored Data.

Consider now a system with two components having differing failure distributions. Failures are generated from different Weibull distributions with the parameters of $\beta = 0.8$, $\theta = 4000$ and $\beta = 0.9$, $\theta = 3000$, respectively. A random variate is drawn from each distribution and compared for each generated failure. The failure that occurs first becomes the failure in question. Failures are time censored at 5000 time units. The resulting 100 failures are divided into two categories, one for each failure mode. Units at the end of the test are censored for both failure modes.

The first step is to use the rank adjustment method to fit a non parametric reliability function to the data, once for each failure mode. The resulting reliability data, along with the original failure times is then used in the probability plots and linear function estimates of the distribution parameters calculated.

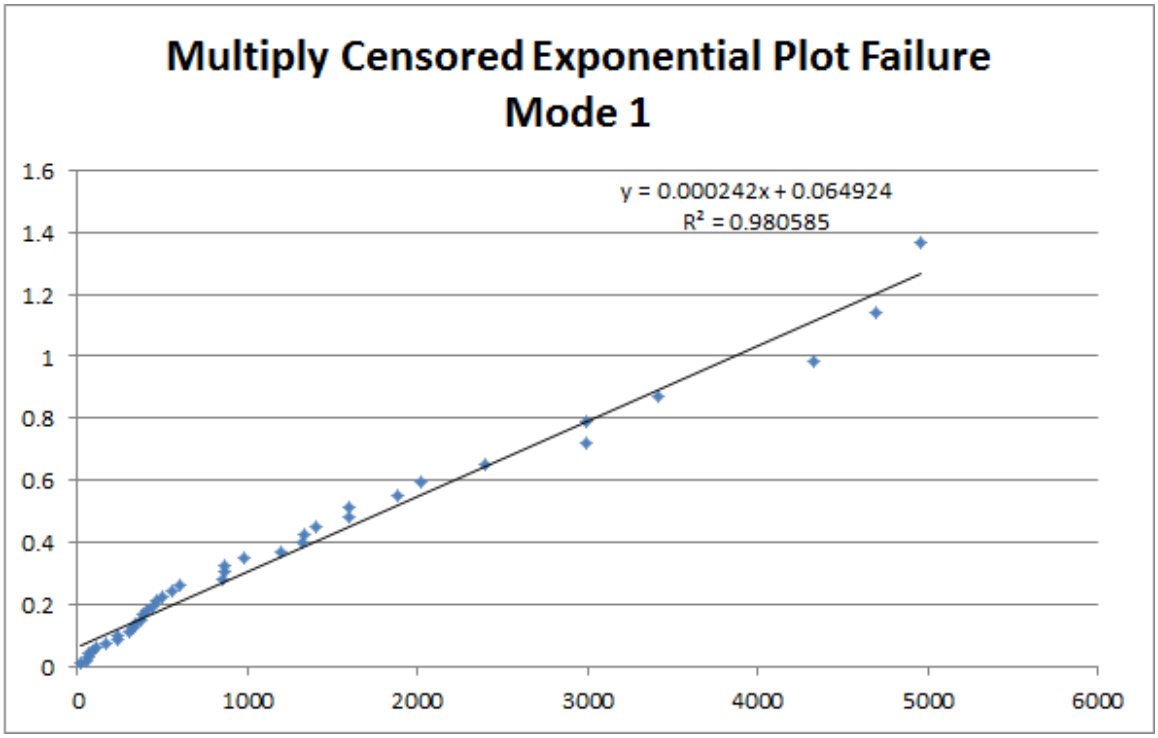


Figure 14. Exponential Probability Plot for Failure Mode 1 Originating from Multiply Censored Data from two Underyling Weibull Distributions

Considering the exponential probability plot in Figure 14 and the Weibull probability plot in Figure 15 it is hard to draw any conclusions. Both plots have very high R^2 values and both plots show relatively small deviations from the fitted lines.

A greater deviation is found when comparing the linear function estimates to MLEs. The linear function estimate MTTF is 4132.231 while the MLE has a value of 3531.969, under the exponential failure distribution assumption. This is a substantial difference and one that points out that the exponential distribution may not the best fit in this case. The Weibull distribution has linear function estimates of $\beta = 0.855$

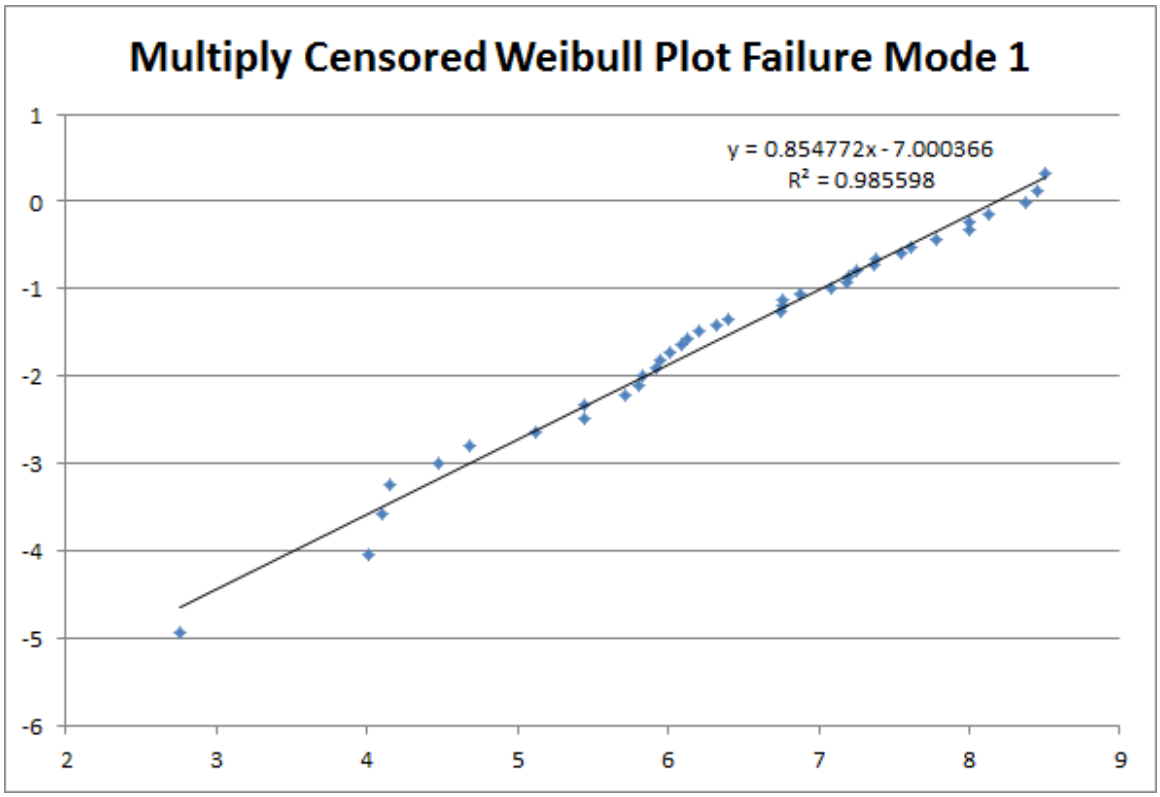


Figure 15. Weibull Probability Plot for Failure Mode 1 Originating from Multiply Censored Data from two Underyling Weibull Distributions

and $\theta = 3603.809$ and MLEs of $\beta = 0.840$ and 3899.246. This results in a MTTF of 3907.393 and an MLE estimate of an MTTF of 4272.989. This is a narrower gap than the exponential distribution. In this case, the first failure mode is correctly identified as a Weibull distributed failure mode.

This can be compared with the JMP output, shown in Figure 16. JMP uses three metrics to assess distribution fit. A good rule of thumb is all should agree and lower is better in each. The JMP output favors neither failure distribution. The model comparison list is sorted by the -2loglikelihood metric.

A nice feature of statistical packages is the confidence bounds provided for parameter estimates. These are beneficial to the analyst considering the range of failure distributions suggested by the empirical data modeled. JMP also provides estimates

Statistics

Model Comparisons

Distribution	AICc	-2Loglikelihood	BIC
Weibull	717.34857	713.22486	722.43520
Exponential	717.27047	715.22965	719.83482

Parametric Estimate - Weibull

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
location	8.2685	0.20956	7.8578	8.6793
scale	1.1904	0.15275	0.8910	1.4898
Weibull α	3899.2459	817.11853	2585.8567	5879.7221
Weibull β	0.8401	0.10780	0.6712	1.1223

Criterion	
-2*LogLikelihood	713.22486
AICc	717.34857
BIC	722.43520

Parametric Estimate - Exponential

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
scale	3531.9695	565.56775	2580.5698	4834.1294

Criterion	
-2*LogLikelihood	715.22965
AICc	717.27047
BIC	719.83482

Figure 16. JMP Life Distribution Output for Failure Mode 1 from two underyling Weibull Distributions

for both hypothesized distributions. These are noticeably wide. The sample size for both failure modes is a total of 100, with 39 failures in mode one, 58 failures in mode two, and three time censored failures. Less than half are in the first failure mode. The MLE estimate is relatively close to the known underlying distribution, but the Weibull shape parameter (α) shows quite the range.

For the second failure mode, the probability plot for the exponential distribution is in Figure 17 and the Weibull distribution plot is in Figure 18. Each plot provides the linear equation and R^2 of the fit.

The exponential distribution has a MTTF of 2564.103 based on the linear equation and a MLE of 2374.945 These correspond to a good fit for the exponential distribution. The Weibull parameter estimates are $\beta = 0.942$ and $\theta = 2281.25$ based on the linear function and $\beta = 0.930$ and $\theta = 2390.981$ based on MLEs. These results correspond to similar MTTF values of 2342.547 and 2466.632, respectively.

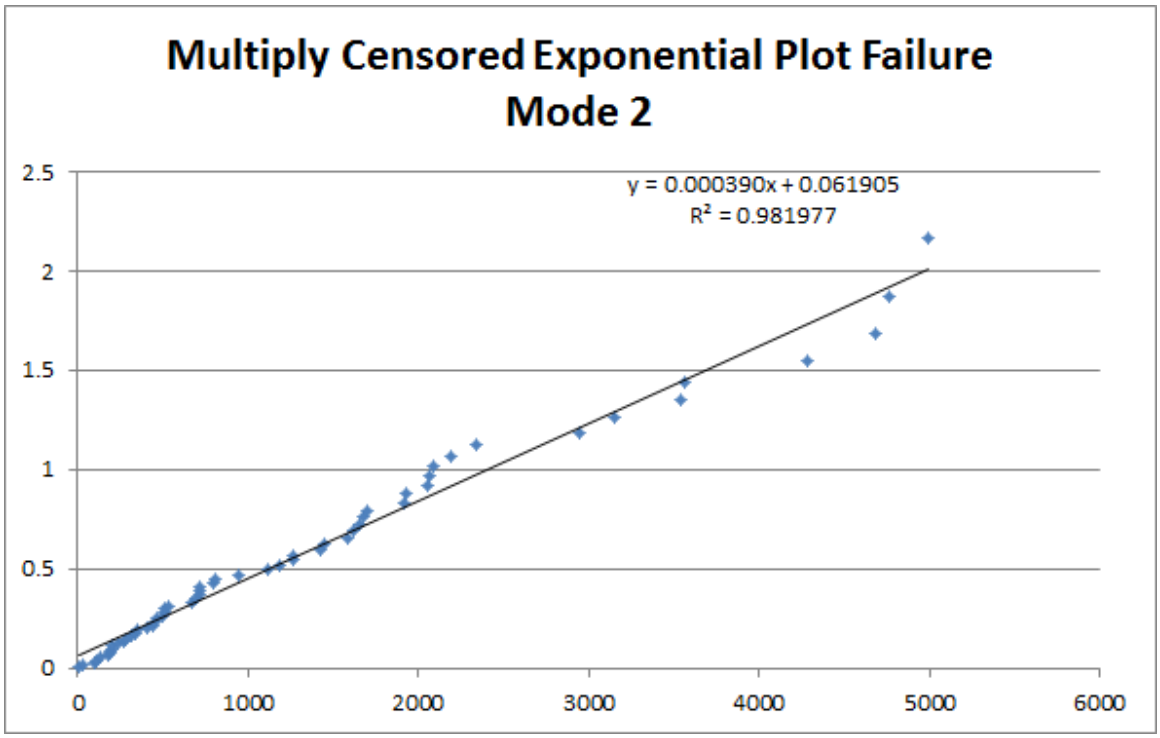


Figure 17. Exponential Probability Plot for Failure Mode 2 Originating from Multiply Censored Data from two Underyling Weibull Distributions

This example demonstrates the uncertainty in empirical modeling. Both the exponential and Weibull are reasonable. The $\beta = 0.9$ is close to the $\beta = 1.0$ associated with an exponential so a failure to clearly distinguish the distributions based on an empirical fit is not surprising.

The JMP report is shown in Figure 19. The JMP report lists the Weibull first even though the exponential is better in two of the three metrics in the model comparison. The model comparison list is sorted by -2loglikelihood.

As the complexity of the analysis increases, so can the effect of noise further obscure the analysis. The takeaway is to ensure parameters are fit using more robust models (such as the Weibull over the exponential) and the uncertainty range of parameter values be explicitly considered for each component evaluated.

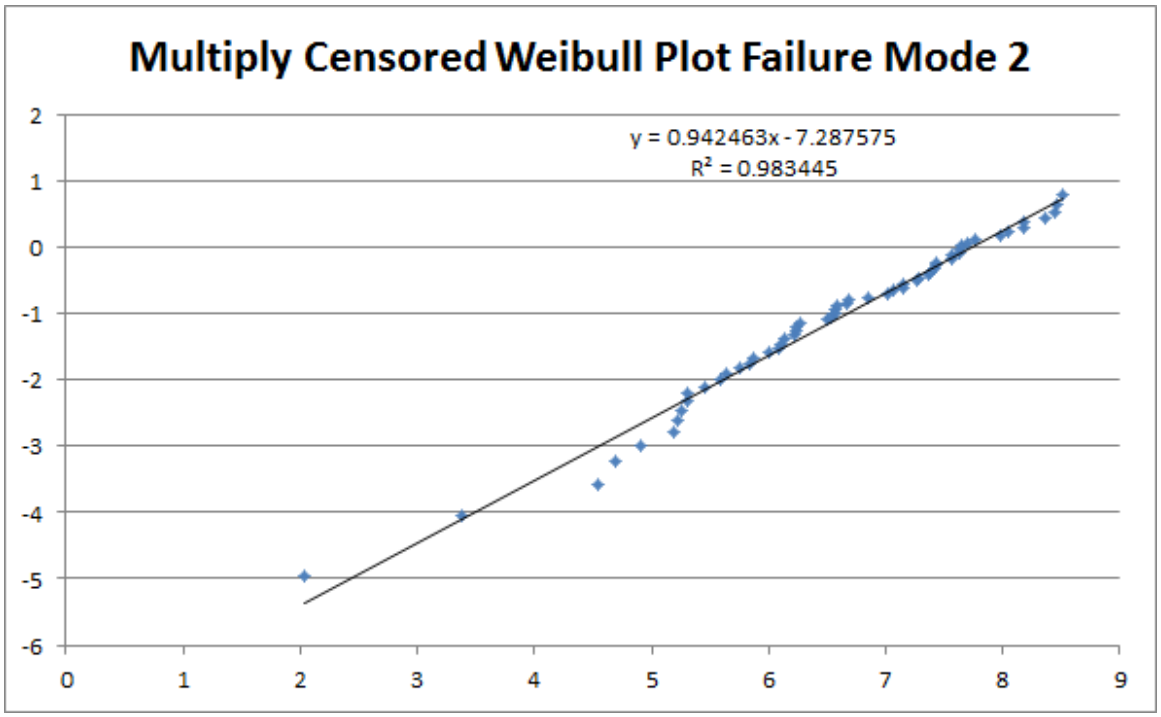


Figure 18. Weibull Probability Plot for Failure Mode 2 Originating from Multiply Censored Data from two Underyling Weibull Distributions

4.3 Example: Applying to Reliability Growth

A simple example of reliability growth is considered. In this example, a new aircraft is entering testing. Six aircraft are used in the test, which is divided into three phases with 500, 500, and 1000 hours of testing in each phase respectively. All systems are run concurrently in the test.

While component failure leads to system failure, for this example we assume the first component failure is system failure and thus generate that failure from a component reliability distribution. This yields the data needed to illustrate the reliability growth example. To cleanly cause changes in system reliability, the failure distribution used changes for each of the three phases in the example test.

In this test, failure is defined as the time that the system is brought off-line for maintenance actions. Based on historical aircraft of similar size and configuration,

Statistics

Model Comparisons

Distribution	AICc	-2Loglikelihood	BIC
Weibull	1021.2515	1017.1277	1026.3381
Exponential	1019.6774	1017.6366	1022.2418

Parametric Estimate - Weibull

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
location	7.7795	0.14169	7.5018	8.0572
scale	1.0749	0.11069	0.8579	1.2918
Weibull α	2390.9810	338.77884	1811.2112	3156.3353
Weibull β	0.9303	0.09580	0.7741	1.1656

Criterion	
-2*LogLikelihood	1017.1277
AICc	1021.2515
BIC	1026.3381

Parametric Estimate - Exponential

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
scale	2374.9450	311.84556	1836.0534	3072.0041

Criterion	
-2*LogLikelihood	1017.6366
AICc	1019.6774
BIC	1022.2418

Figure 19. JMP Life Distribution Output for Failure Mode 2 from two underyling Weibull Distributions

the initial MTBM is estimated to be one flight hour between maintenance, with a goal of four flight hours between maintenance period. When systems undergo "failure," it is assumed that they will be returned to as new condition. Maintenance time is not counted as test time. Furthermore, it is assumed that maintenance testing ensures no immediate failures will occur when the system resumes testing.

Major corrective actions are delayed to the end of each test period and are allocated ten percent of the test phase time to be implemented. A fix effectiveness factor of 0.8 is assumed.

The AMSAA tools are used to generate a planning curve, shown in Figure 20. Of note is that major corrective actions are needed after the first phase of testing to achieve the goal MTBM. This is the curve that is used to evaluate progress towards the reliability goal.

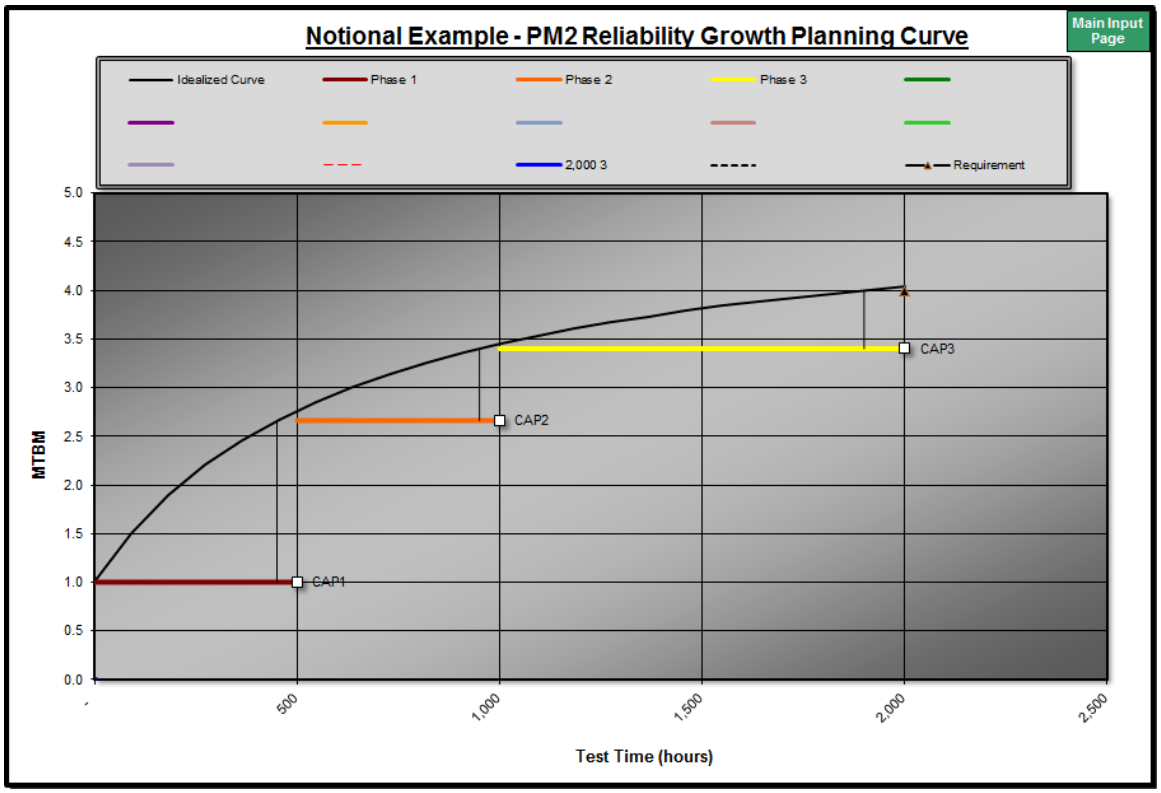


Figure 20. Planning Reliability Growth Curve for AMSAA Example

For the purposes of this example, notional data are used. Since the evaluation is focused on system level MTBM, the data from the test is only censored if the test time exceeds the total allocated time. This yields test data that is multiply censored.

For the first phase of testing, system failure data are exponentially distributed with a MTBM of one flight hour between maintenance period. The relatively low MTBM ensures that there are a substantial number of datapoints as a result of this test. JMP is used for the analysis and is used to verify the planned MTBM. The results associated with this analysis are provided in Appendix A.

The MLE estimates for the Weibull and the exponential fitted distributions are shown in Figure 23. The comparison criterion in JMP favors the Weibull distribution over the exponential. In this case, the MLE estimators and the JMP criterion

comparison are not as insightful as a simple graphical comparison, which is shown in Figure 21 for the exponential and Figure 22 for the Weibull.

The empirical system MTBM yields the first phase result in Figure 20. Estimating the system failure distribution, whose mean estimates system MTBM, suggests using the Weibull.

For the second phase, system failures are generated from a Weibull distribution with a $\beta = 1.5$ and a $\theta = 2$. Unlike all of the examples previously used, this distribution is very clearly not an exponential distribution.

For the second phase data, the graphical plot very clearly favors the Weibull distribution. This is shown in Figure 24 and Figure 25. This corresponds to the JMP criterion which also clearly shows that the Weibull distribution is a significantly better fit in all three criterion. Results are shown in Figure 26. The estimates very closely match the underlying distribution, showing that failure rate is actually increasing. The associated MTBM is 1.845. This is below the target value of 2.6 shown in Figure 20.

Failing to meet MTBM goals derived from the planned reliability growth curve can lead to managerial interventions. Using an estimated system failure distribution can provide some insight into how far apart the empirical and target MTBM are. Note this system failure distribution is simply an empirical model of the data and does not correspond to any derived system failure model.

The third and final test phase is longer, given the reliability is expected to improve substantially. This time, system failure data is generated from a Weibull random variate with a $\beta = 0.8$ and a $\theta = 4$.

The empirical MTBM of 5.08 exceeds the planning requirement. Figure 27, Figure 29, and Figure 28 denote use of the Weibull as the preferred choice if creating a model of the system failure distribution.

While this example is not complete, and the data are notional, the intent was to focus on how to integrate system failure data from system test into reliability growth curve assessment and the use of a Weibull distribution as a general framework for deriving empirical models of system failure distributions.

4.4 Summary and Notes

This chapter uses notional examples to demonstrate the methodology in Chapter III. Examples of complete and censored data highlight the benefits of assuming Weibull failure distributions, due to their more general use, over the usual assumption of an exponential failure distribution.

In generating data to show how these techniques could be implemented, the assumption was made that the underlying distribution was either exponential or Weibull. In reality, the underlying distribution is actually unknown and must be estimated from the data on-hand. If unsatisfied with either the exponential or Weibull fit for the failure data, other distributions are available in JMP, or other similar software packages.

V. Conclusions and Recommendations

This thesis proposes methodology to improve the statistical rigor in the KC-46 Flight Test Program. Specifically, the assumption of constant failure rate is challenged in favor of the more general purpose Weibull distribution. This chapter reviews the technical and non-technical insights as a result of this research and propose topics of future research.

5.1 Non-Technical Insights

The important non-technical insight from this research is that the assumption of constant failure rate can lead to incorrect results. The methodology proposed here, especially when combined with a statistical software package, can be easily implemented. Using such statistical tools in a knowledgeable fashion can greatly improve the reliability analysis associated with the test.

Data censoring, either singly or multiply, will occur and must be accommodated by the analysis. The resulting analysis provides better information to the designers of the system, which in turn allows for improved corrective actions and greater reliability improvements.

5.2 Potential Future Research

5.2.1 The Use of Accelerated Life Data.

One of the biggest limitations in testing highly reliable systems is the lengthy test time needed to get failure data. Frequently, this time is not available or failures are not in sufficient quantity to obtain good estimates. An investigation into the use of accelerated life testing can provide methodology to further "get more with less."

5.2.2 Software Tools.

A limitation in the analysis methods suggested here is that we do not want to conduct manual calculations. Analysis should involve capable statistical tools, the cost of which will be a very small part of any overall test budget.

5.2.3 Investigating Dependant Failure Modes.

A question not addressed in this thesis is the issue of dependent failure modes in testing. Both the current and proposed methodology ignores dependent failure modes. Is there a way to separate out the underlying relationship between dependent failure modes in multiply censored testing? Knowing how to address this could allow for more effective use of censored data but will require a good deal of fundamental research.

5.3 Conclusion

To conclude, this thesis does not cover all of the opportunities to enhance the statistical rigor of the KC-46 specifically or DoD acquisition programs in general. The focus is on an aspect of reliability and maintainability. Understanding failure distributions and the forms we hypothesize for those distributions allows for much better characterization of system reliability which in turn aids system development.

Appendix A. Supplementary Material

This appendix contains the figures showing the results of the notional reliability growth example in Section 4.3 and the Thesis Quadchart.

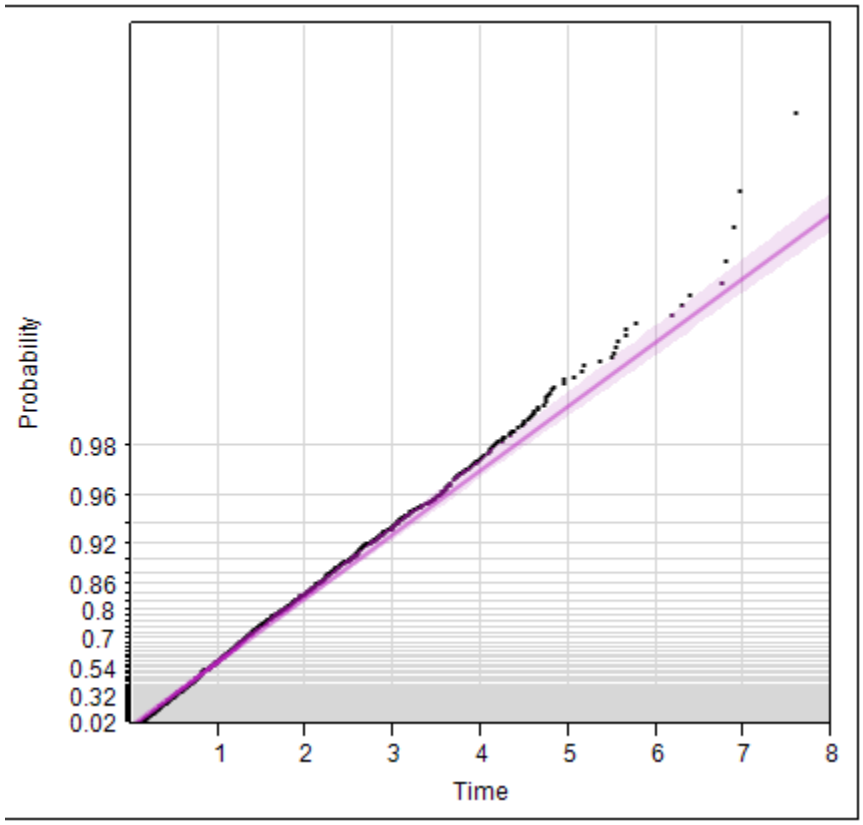


Figure 21. Phase 1 Reliability Growth Exponential Probability Plot

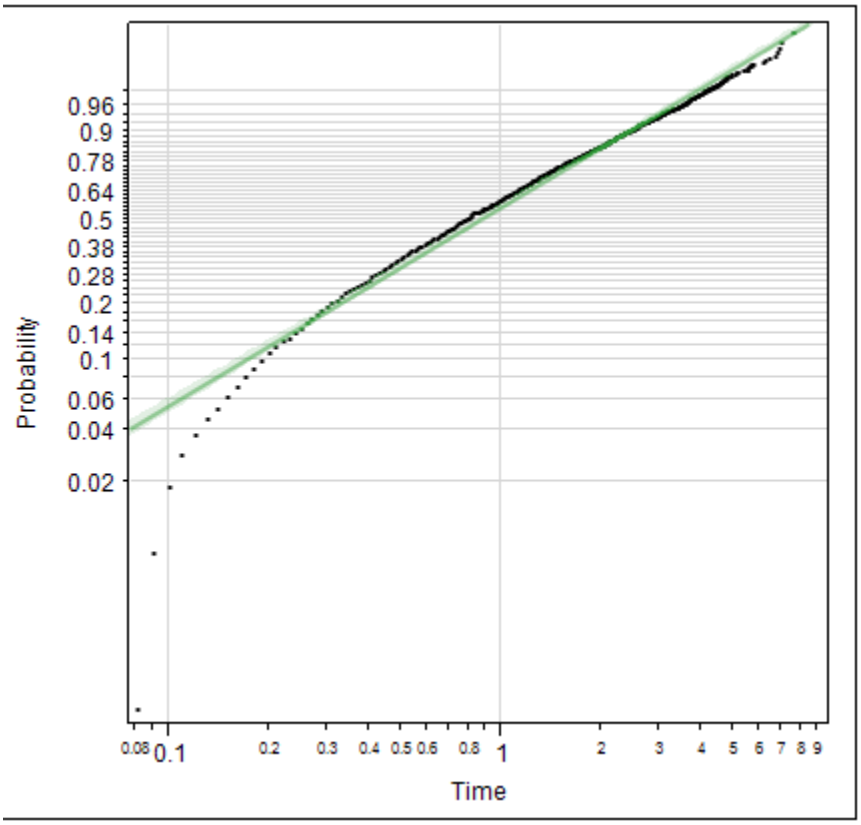


Figure 22. Phase 1 Reliability Growth Weibull Probability Plot

Parametric Estimate - Weibull

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
location	0.1618031	0.01752516	0.1274544	0.1961517
scale	0.8605671	0.01246258	0.8361409	0.8849933
Weibull α	1.1756287	0.02060308	1.1359330	1.2167115
Weibull β	1.1620244	0.01682823	1.1299520	1.1959706

Criterion	
-2*LogLikelihood	5869.5489
AICc	5873.5534
BIC	5885.3546

Parametric Estimate - Exponential

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
scale	1.1110778	0.02138270	1.0699490	1.1537875

Criterion	
-2*LogLikelihood	5968.7848
AICc	5970.7863
BIC	5976.6876

Figure 23. Phase 1 Reliability Growth Failure Distribuion Results

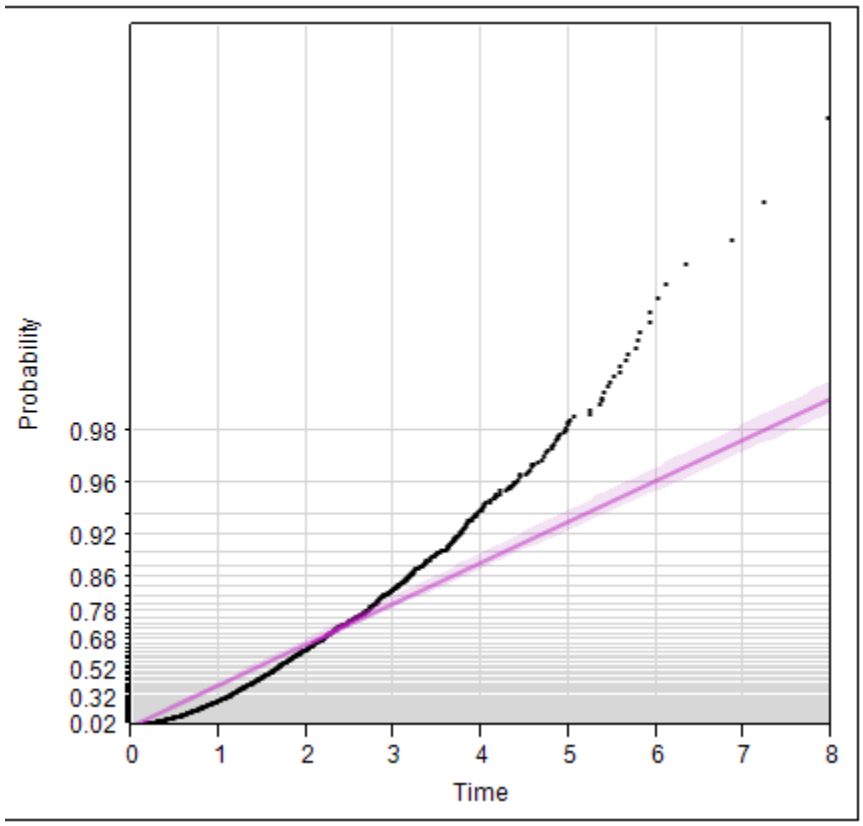


Figure 24. Phase 2 Reliability Growth Exponential Probability Plot

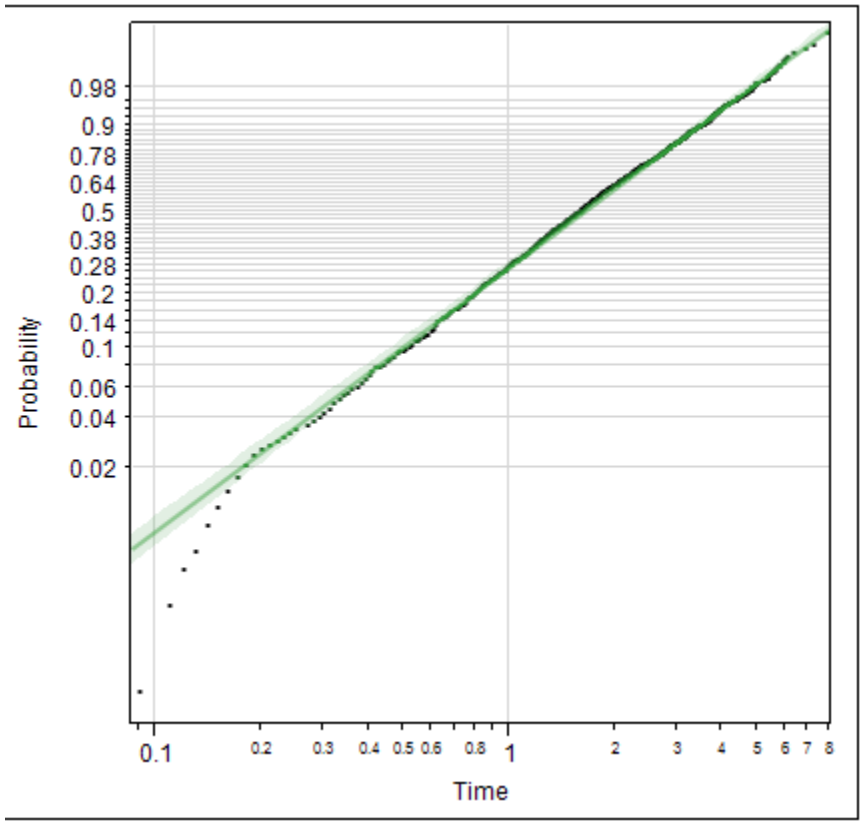


Figure 25. Phase 2 Reliability Growth Weibull Probability Plot

Parametric Estimate - Weibull

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
location	0.7137832	0.01647959	0.6814838	0.7460826
scale	0.6328724	0.01206831	0.6092190	0.6565259
Weibull α	2.0417009	0.03364639	1.9768088	2.1087231
Weibull β	1.5800973	0.03013103	1.5231692	1.6414459

Criterion	
-2*LogLikelihood	4799.9745
AICc	4803.9818
BIC	4814.7843

Parametric Estimate - Exponential

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
scale	1.8314713	0.04525256	1.7448913	1.9223474

Criterion	
-2*LogLikelihood	5258.3719
AICc	5260.3744
BIC	5265.7768

Figure 26. Phase 2 Reliability Growth Failure Distribuion Results

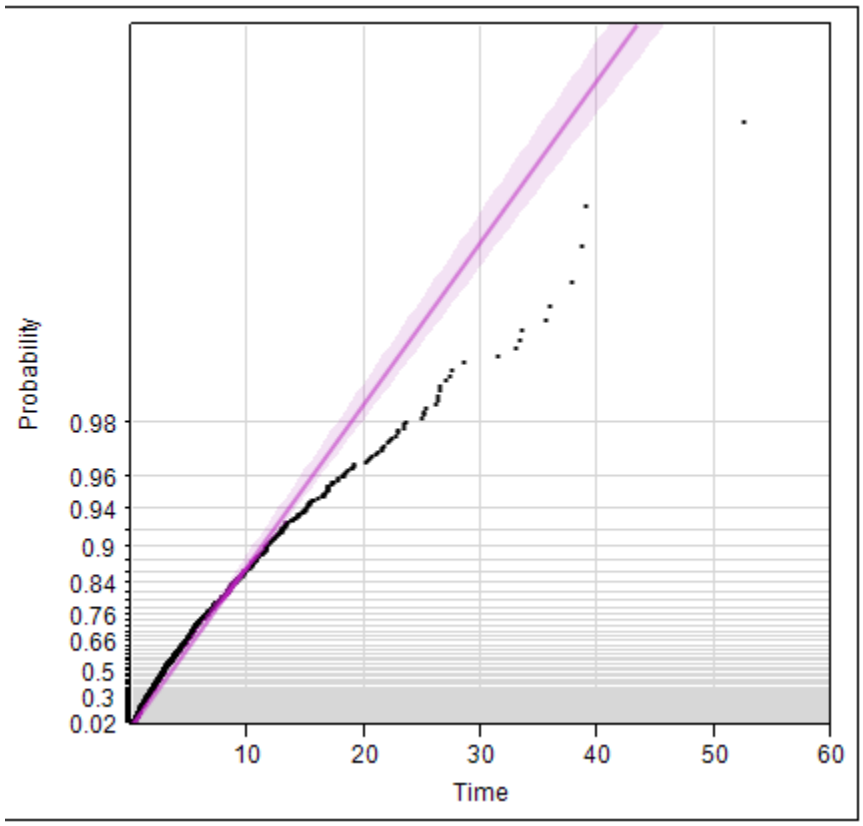


Figure 27. Phase 3 Reliability Growth Exponential Probability Plot

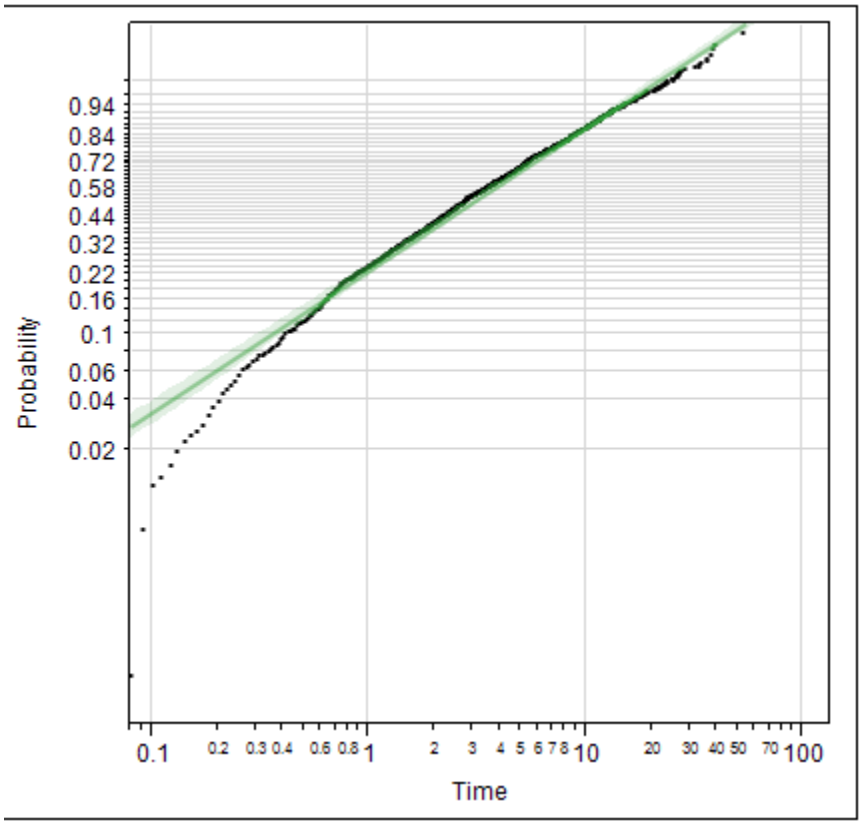


Figure 28. Phase 3 Reliability Growth Weibull Probability Plot

Parametric Estimate - Weibull

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
location	1.4948237	0.03362836	1.4289133	1.5607341
scale	1.1303813	0.02405098	1.0832422	1.1775203
Weibull α	4.4585505	0.14993373	4.1741609	4.7623159
Weibull β	0.8846573	0.01882274	0.8492422	0.9231546

Criterion	
-2*LogLikelihood	6420.4535
AICc	6424.4630
BIC	6434.7423

Parametric Estimate - Exponential

Parameter	Estimate	Std Error	Lower 95%	Upper 95%
scale	4.7580888	0.13399081	4.5025876	5.0280886

Criterion	
-2*LogLikelihood	6455.9318
AICc	6457.9350
BIC	6463.0762

Figure 29. Phase 3 Reliability Growth Failure Distribuion Results

AN EXAMINATION OF STATISTICAL RIGOR INFUSED INTO THE KC-46 FLIGHT TEST PROGRAM

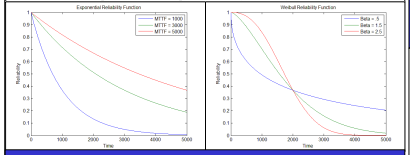
2LT Sean C. Ritter
Advisor: Dr. Raymond R. Hill
Reader: Dr. Joseph J. Pignatiello
Department of Operational Sciences (ENS)
Air Force Institute of Technology

DoD Mandated

- Director of DOT&E, J. Michael Gilmore stated suitability must be improved.
- Only 6 of 15 systems reported on in FY11 met reliability thresholds.
- Requires additional reliability tests and an understanding of failure profiles as they relate to reliability growth
- Part of a long standing process to improve reliability assessment in DoD acquisition

More General Assumptions

- Weibull Failure Distribution more robust
- Shape parameter allows for wider range of distribution shapes
- Includes the exponential as a special case (Beta = 1)



Maximum Likelihood Estimation

Exponential
Complete or Censored Data

$$\widehat{\lambda} = \frac{r}{\sum_{i \in F} t_i + \sum_{i \in C} t_i}$$

Weibull
Multiply Censored Data

$$\sum_{i \in F} \frac{\ln t_i}{r} = \sum_{all\ i} t_i^{\widehat{\beta}} \ln t_i \left[\sum_{all\ i} t_i^{\widehat{\beta}}\right]^{-1} - \frac{1}{\widehat{\beta}}$$

$$\widehat{\theta} = \left[\sum_{all\ i} \frac{t_i^{\widehat{\beta}}}{r}\right]^{\frac{1}{\widehat{\beta}}}$$

Weibull
Complete or Singly Censored Data

$$\frac{1}{r}\sum_{i=1}^{r} \ln t_i = \frac{\sum_{i=1}^{r} t_i^{\widehat{\beta}} \ln t_i + (n-r) t_s^{\widehat{\beta}} \ln t_s}{\sum_{i=1}^{r} t_i^{\widehat{\beta}} + (n-r) t_s^{\widehat{\beta}}} - \frac{1}{\widehat{\beta}}$$

$$\widehat{\theta} = \left\{\frac{1}{r}\left[\sum_{i=1}^{r} t_i^{\widehat{\beta}} + (n-r) t_s^{\widehat{\beta}}\right]\right\}^{\frac{1}{\widehat{\beta}}}$$

t_i= ith failure t_s= singly censored time
$\widehat{\beta}$= estimated Weibull shape parameter
$\widehat{\theta}$= estimated Weibull scale parameter
$\widehat{\lambda}$= estimated exponential failure rate
r = total number of failures
n = total number of test units
F = set of failure indices, C = set of failure indices

Research Purpose

- Incorporate lessons learned from prior tanker test programs (KC-10, KC-135) and KC-46 test program
- Propose improvements to existing methodology to obtain improved estimates

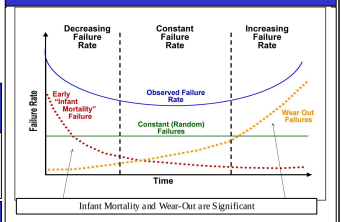
Research Results

- Assuming constant failure rate can lead to erroneous results
- Use the Weibull as a more general framework to analyze the uncertainty in the shape of a failure profile
- Insight gained from knowledge of frequency, and change of frequency, of failure
- Multiple component failures can be tested simultaneously from a single combined system test using data censoring
- Empirical modeling of the general system failure distribution provides insight into reliability growth progress

Reliability Growth Assessment

- A system level failure event can be used to assess progress towards reliability goals given agreement among test team
- An empirical model provides insight into the changing nature of the system within a test phase
- If insufficient progress is being made, either the planning curve, or the management strategy can be adjusted to better model reality

General Bathtub Curve



Non-Parametric Reliability

- Estimate reliability non-parametrically with the formula:

$$\widehat{R}(t_i) = 1 - \frac{i_{t_i} - 0.3}{n + 0.4}$$

- If data are censored, use rank adjustment method to adjust i_{t_i} to account for the censored data point

$$\text{Rank Increment} = \frac{(n+1) - i_{t_{i-1}}}{1 + \text{number of units beyond present censored unit}}$$

$$i_{t_i} = i_{t_{i-1}} + \text{rank increment}$$

Probability Plotting

- Probability plots transform data to fit a linear model ($y = a + bx$) related to a selected distribution

Distribution	x(t)	y(t)
Exponential	t	$\ln\left[\frac{1}{1-F(t)}\right]$
Weibull	$\ln(t)$	$\ln\left[\ln\left[\frac{1}{1-F(t)}\right]\right]$

Exponential
Complete or Censored Data

$$\widehat{\lambda} = b$$

Weibull
Complete or Censored Data

$$\widehat{\beta} = b$$

$$\widehat{\theta} = e^{-\frac{a}{\widehat{\beta}}}$$

Figure 30. QuadChart

Bibliography

[1] "ANSI GEIA-STD-0009".

[2] "MIlitary Handbook on Reliability Growth Management (MIL-HDBK-189)", February 1981.

[3] Abernethy, Robert B. *The New Weibull Handbook on Reliability and Statistical Analysis for Prediciting Life, Safety, Survivability, Risk, Cost, and Warranty Claims*. Robert B. Abernethy, 2006.

[4] Adolph, Charles. *Developmental Test and Evaluation*. Technical report, Defense Science Board, May 2008.

[5] AFRL/LCMC. *Test and Evaluation Master Plan*. Technical report, USAF, 2012.

[6] Banks, Jerry, John S. Carons II, Barry L. Nelson, and David M. Nicol. *Discrete-Event System Simulation*. Pearson Educations Inc, 2010.

[7] Birolini, Alessandro. *Reliability Engineering: Theory and Practice*. Springer, 2010.

[8] Blinn, Major Roger K. *KC-135 Weapon System Trainer Qualification Operational Test and Evaluation Final Report*. Technical report, Air Force Test and Evaluation Center, 1981.

[9] Brenholdt, Col James P. *KC-10A Advanced Tanker/Cargo Aircraft Follow-On Operationla Test and Evaluation Test Plan*. Technical report, Air Force Test and Evaluaton Center, 1981.

[10] Coppa, Mark C. and John J. Scorsone. *C/KC-135 PACER CRAG System Qualification Operational Test and Evaluatio Plan*. Technical report, Air Force Operational Test and Evaluation Center, 1997.

[11] Crow, L.H. "A methodology for managing reliability growth during operational mission profile testing". *Reliability and Maintainability Symposium, 2008. RAMS 2008. Annual*, 48 –53. jan. 2008. ISSN 0149-144X.

[12] Crow, L.H. "The Extended Continuous Evaluation reliability growth model". *Reliability and Maintainability Symposium (RAMS), 2010 Proceedings - Annual*, 1 –6. jan. 2010. ISSN 0149-144X.

[13] Crow, L.H. "Planning a reliability growth program utilizing historical data". *Reliability and Maintainability Symposium (RAMS), 2011 Proceedings - Annual*, 1 –6. jan. 2011. ISSN 0149-144X.

[14] Crow, L.H. "Reliability Growth Planning, Analysis and Management", 2011. URL http://www.reliasoft.com/pubs/2011_RAMS_reliability_growth_planning_analysis_and_management.pdf.

[15] Crow, L.H. "Demonstrating reliability growth requirements with confidence". *Reliability and Maintainability Symposium (RAMS), 2012 Proceedings - Annual*, 1–6. jan. 2012. ISSN 0149-144X.

[16] Douglas C. Montgomery, G. Geoffrey Vining, Elizabeth A. Peck. *Introduction to Linear Regression Analysis*. Wiley, 2006.

[17] Earl, Jack A. *KC-135 Communication, Navigation, and Surveillance/Air Traffic Management Block 45 Modification Reliability and Maintainability Test and Evaluation*. Technical report, Air Force Flight Test Center, 2011.

[18] Ebeling, Charles E. *An Introduction to Reliability and Maintainability Engineering*. Waveland Press Inc., 2010.

[19] Gilmore, Dr. Michael. "Key Issues in Reliability Growth", September 2011. URL http://www.dote.osd.mil/pub/presentations/Gilmore-NAS-presentation-finalv1.pdf.

[20] Gilmore, J. Michael. "Test and Evaluation (T&E) Initiatives", Nov 2009.

[21] Heebner, David R. *Test and Evaluation*. Technical report, Defense Science Board, September 1999.

[22] Reeves, Louise S, Michael Whelan, Sharon L. Cook, and Sheldon Carter. *KC-135 Global AIr Traffic Management Qualifiaction Operational Test and Evaluation*. Technical report, Air Force Operational Test and Evaluation Center, 2004.

[23] RIWG. *Report of the Reliability Working Group*. Technical report, Department of Defense, 2008.

[24] Robinson, Richard S. *Logistics Composite Model Workbook*. Technical report, R/M Systems Inc, 1976.

[25] Strunz, R. and J.W. Herrmann. "Planning, tracking, and projecting reliability growth a Bayesian approach". *Reliability and Maintainability Symposium (RAMS), 2012 Proceedings - Annual*, 1 –6. jan. 2012. ISSN 0149-144X.

[26] Vilijandas Bagdonavicius, Mikhail Nikulin. *Accelerated Life Models: Modeling and Statistical Analysis*. Chapman and Hall/CRC, 2002.

[27] Wasserman, Gary S. *Reliability Verification, Testing, and Analysis in Engineering Design*. Marcel Dekker Inc, 2003.

[28] Wegner, Lt Col Lavern J. and Capt James R. Villines. *KC-10A Follow-On Operational Test and Evaluation, Phase II, Final Test Report.* Technical report, 4201st Test Squadron, 1983.

[29] Wolstenholme, Linda C. *Reliability Modelling, A Statistical Approach.* Chapman and Hall/CRC, 1999.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
21–03–2013	Master's Thesis	Aug 2011 – Mar 2013

4. TITLE AND SUBTITLE

An Examination of Statistical Rigor infused into the KC-46 Flight Test program

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Ritter, Sean C. Second Lieutenant, USAF

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-13-M-18

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of the Secretary of Defense
Attn: Dr. Catherine Warner
1700 Defense Pentagon
Washington D.C. 20301
(703) 697-7247, catherine.warner@osd.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

OSD

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The KC-46 program is bringing on-line the replacement aircraft for the KC-135. Although not a new development program, but rather a modification program, there are extensive plans for the flight testing of the KC-46. Recent DoD emphasis mandates the use of statistical design principles for DoD test and evaluation. This project will examine the planned flight test program for KC-46 and reconsider components of that program based on principles of statistical rigor. Of particular focus will be the reliability and maintainability aspects of the flight test program. Current methodology assumes a constant failure rate in all situations, implying that the underlying failure profile of any component or system is assumed to be exponentially distributed. Use of the Weibull failure distribution is proposed as a more general framework to provide additional insight about the failure profile of the component or system.

15. SUBJECT TERMS

KC-46, Failure Distribution Analysis, Statistical Rigor, Flight Test, Aerial Refueling

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Raymond R. Hill
U	U	U	UU	74	19b. TELEPHONE NUMBER *(include area code)* (937) 255–3636, ext 3469; Raymond.Hill@afit.edu