ABSTRACT

Title of dissertation: METHODOLOGY FOR ASSESSING RELIABILITY GROWTH USING MULTIPLE INFORMATION SOURCES

Martin Wayne, Doctor of Philosophy, 2013

Dissertation directed by: Professor Mohammad Modarres
Department of Mechanical Engineering

The research presented here examines the assessment of the reliability of a system or product utilizing multiple data sources available throughout the different stages of its development. The assessment of the reliability as it changes throughout the development of a system is traditionally referred to as reliability growth, which refers to the discovery and mitigation of failure modes within the system, thereby improving the underlying reliability. Traditional models for assessing reliability growth work with test data from individual test events to assess the system reliability at the current stage of development. These models track or project the reliability of the system as it matures subject to the specific assumptions of the models.

The contributions of this research are as follows. A new Bayesian reliability growth assessment technique is introduced for continuous-use systems under general corrective action strategies. The technique differs from those currently in the literature due to the allowance for arbitrary times for corrective actions. It also provides a probabilistic treatment of the various parameters within the model, accounting for the uncertainty present in the assessment. The Bayesian reliability

growth assessment model is then extended to include results from operational testing. The approach considers the posterior distribution from the reliability growth assessment of the prior for the operational reliability assessment. The developmental and operational testing environments are not a priori assumed to be equivalent, and the change in environments is accounted for in a probabilistic manner within the model. A Bayesian reliability growth planning model is also presented that takes advantage of the reduced uncertainty in the combined operational assessment. The approach allows for reductions in the amount of demonstration testing necessary for a given level of uncertainty in the assessment, and it can also be used to reduce high design goals that often result from traditional operating characteristic curve applications. The final part of this research involves combining various sources of reliability information to obtain prior distributions on the system reliability. The approach presents a general framework for utilizing information such as component/subsystem testing, historical component reliability data, and physics-based modeling of specific component failure mechanisms.

METHODOLOGY FOR ASSESSING RELIABILITY GROWTH USING MULTIPLE INFORMATION SOURCES

by

Martin Wayne

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2013

Advisory Committee:

Professor Mohammad Modarres, Chair
Professor Gregory Baecher
Dr. Paul Ellner
Professor Ali Mosleh
Assistant Professor Monifa Vaughn-Cooke

TABLE OF CONTENTS

ACRONYMS

AMSAA	Army Materiel Systems Analysis Activity
AMPM	AMSAA Maturity Projection Model
CAP	Corrective Action Period
DoD	Department of Defense
DFR	Design-for-Reliability
DT	Developmental Testing
FEF	Fix Effectiveness Factor
HPP	Homogeneous Poisson Process
MLE	Maximum Likelihood Estimate
MS	Management Strategy
MTBF	Mean Time Between Failure
MTTF	Mean Time to Failure
NHPP	Non-Homogeneous Poisson Process
OC	Operating Characteristic
OT	Operational Testing
PM2	Planning Model Based on Projection Methodology
ROCOF	Rate of Occurrence of Failures
SPLAN	System Planning Model
SSPLAN	Subsystem Planning Model
SSTRACK	Subsystem Tracking Model
TAFT	Test-Analyze-Fix-Test

NOTATION

a	Beta distribution parameter
α	Gamma distribution parameter (pseudo failures)
$\tilde{\alpha}$	Posterior Gamma distribution parameter from developmental reliability growth testing
b	Beta distribution parameter
β	Gamma distribution parameter (pseudo test time)
$\tilde{\beta}$	Posterior Gamma distribution parameter from developmental reliability growth testing
d_i	assessed FEF for mode i
ε	Mean value of degradation between developmental and operational testing
γ	MTBF Degradation between developmental and operational test environments
K	total number of failure modes assumed in system
λ_A	prior mean failure intensity of system for A-modes
λ_B	prior mean failure intensity of system
λ_{DT}	failure intensity of total system during developmental testing
λ_{GP}	growth potential failure intensity of system
λ_i	failure intensity of failure mode i
λ_I	prior mean failure intensity of system
λ_{OT}	failure intensity of total system during operational testing
λ_s	failure intensity of total system
m	total number of failure modes observed during testing
μ_d	average FEF value
M_I	Initial MTBF
M_{GP}	MTBF growth potential
M_R	MTBF requirement
n_i	total number of failures observed for mode i during testing
$n_{i,1}$	total number of failures observed for mode i before corrective action implementation
n_{OT}	total number of failures observed during operational testing
p	total number of systems under test
$t_{i,j}$	j^{th} failure occurrence time for mode i
T	total test time
T_j	total test time on j^{th} system
T_{DT}	total developmental test time
T_{OT}	total operational test time
v_i	time of corrective action for mode i

- CHAPTER 1 -

1 INTRODUCTION, RESEARCH OBJECTIVES, AND OVERVIEW OF DISSERTATION

1.1 Introduction

Reliability growth in complex systems has been widely studied for many years, and it continues to be an area of great interest. A recent Defense Science Board Task Force report [1] cited a significant increase in the number of military systems that are being rated as not operationally suitable, and the main reason cited in the report is poor reliability. Many systems are simply failing to achieve required levels of reliability, and nearly half of U.S. Army systems over a ten-year period failed to meet operational reliability requirements. A major recommendation from the task force was to implement a robust reliability growth program, including periodic reporting on reliability growth program progress. This is due in part to the lack of reliability growth programs on most developmental systems, but also in part to the realization that traditional DFR principles [2],[3] alone are simply not sufficient when designing highly complex systems.

In response to this recommendation, the U.S. Department of Defense has recently implemented guidance to address this problem. Directive Type Memorandum 11-0003 [4] calls for military programs to establish comprehensive reliability programs, to include DFR techniques along with a reliability growth strategy. Reliability growth curves are to be documented in the Systems Engineering Plan at early milestones, and reliability growth should be tracked against planned thresholds. Additional requirements include reporting on the improvement necessary

to achieve the desired reliability requirement during Operational Testing (OT). The U.S. Army has also implemented policy [5] to more specifically address the reliability problem, with language speaking to the importance of DFR techniques, an early reliability threshold report, and contractual requirements for reliability growth planning curves. These recent advances highlight not only the importance of achieving reliability in modern defense systems, but also the inherent connection between DFR and traditional system-level reliability growth.

Reliability growth is generally divided into the three areas of planning, tracking, and projection [6]. Reliability growth tracking and projection models have typically been developed to assess the system reliability under specific assumptions regarding testing, data collection, and corrective action implementation. The differences between tracking and projection lie in the information that is used to develop the model. Tracking models generally use only failure data collected during Developmental Testing (DT) to estimate the reliability improvement during the test, whereas projection models use engineering assessments of the effectiveness of planned corrective actions to observed failure modes, also known as fix effectiveness factors (FEF). This allows for more flexibility in the projection framework, as the models can project the anticipated improvement that will result from corrective actions that may not yet be implemented. The use of FEFs has also been shown to provide very reasonable estimates of reliability improvement when they are assigned correctly [7]. Reliability growth planning models are then typically transformed versions of the assessment models that can be used to plan an appropriate reliability growth program before system level reliability are available. Historically a number

of models have been developed for both continuous and discrete complex systems, with particular emphasis on reliability growth tracking. Continuous systems are those whose test durations are measured over continuous time periods such as hours or miles. Discrete systems are those whose test durations are measured in discrete demands such as trials or shots. A review of a number of reliability growth models is presented in Chapter 2.

Reliability is defined as the ability of an item to function under given conditions for a period of time [8]. Concurrent with the development of the system, a number of reliability tests are conducted, with the purpose of discovering and mitigating failure modes within the system and thereby improving the underlying reliability. The system design is continually evolving throughout these tests, such that combination of the data for overall reliability assessment is generally not practical. The testing environment also tends to be more benign than actual operational usage conditions, with resulting reliability estimates that are overly optimistic when compared to operational failure rates. This process is intended to mature the reliability of the system to a sufficient level for operational use, which is termed the reliability requirement. Proper specification of the reliability requirement is vitally important to the overall reliability program, and factors such as operational effectiveness and life-cycle operating costs should be considered in this process. The requirement will be assumed to be appropriate for the purposes of the research presented here. The reliability growth models discussed above are intended to assess the reliability as it evolves throughout the DT program.

Many systems under development are then subject to a reliability demonstration test [6] in which the assessed reliability from the test results is compared to the original reliability requirement for the system. The demonstration test can be performed in a number of ways [9], but within the United States Department of Defense (DoD) it is generally a full system test with a fixed configuration that is meant to mimic the intended operational environment of the system to the maximum extent possible. The fixed configuration testing allows for application of the exponential distribution and its associated estimators to assess the system reliability. For planning the appropriate resources and test lengths, an Operating Characteristic (OC) curve [6],[10] is typically used to manage the overall test approach. This allows for the consideration of not only statistical confidence, but also the power or probability of acceptance of the test.

The OC Curve for the demonstration test determines the probability of "passing" the demonstration test as a function of the true reliability of the system, where passing is defined by observing less than or equal to the maximum number of allowable failures. The maximum number of allowable failures is chosen such that the desired lower statistical confidence bound on the reliability estimate is greater than or equal to the requirement. For continuously operating systems this is defined to be the largest non-negative integer k such that the inequality in (1) holds.

$$\sum_{i=0}^{k} \frac{\left(\frac{T}{M_R}\right)^i \exp\left[-\frac{T}{M_R}\right]}{i!} \leq \alpha \qquad (1)$$

For the inequality in (1), T is the total demonstration test length, M_R the reliability requirement of the system, and α is the significance of the test. The underlying

assumption of the approach is that the system failures occur according to a Homogeneous Poisson Process (HPP). The times between successive failures are then exponentially distributed, which is generally justified by the constant configuration of the system. As such, the reliability requirement of the system is commonly specified as a "mean time between failure" metric. The OC curve itself is then calculated by

$$OC(M) = \sum_{i=0}^{c} \frac{\left(\frac{T}{M}\right)^{i} \exp\left[-\frac{T}{M}\right]}{i!}, \tag{2}$$

where M is the true but unknown reliability of the system and c is the maximum number of allowable failures in the test determined using (1). An example of the curve is shown in Figure 1.1.

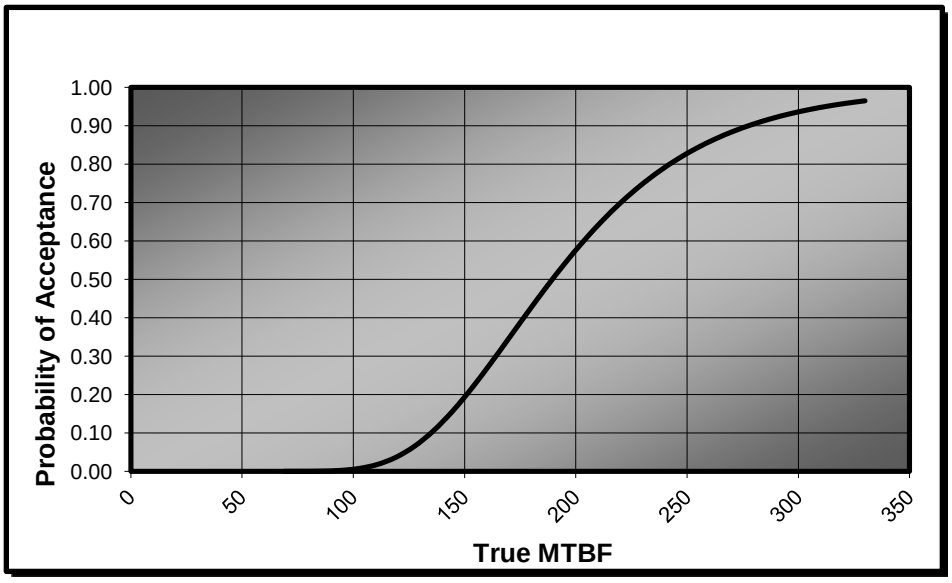


Fig 1.1 OC Curve Example

In order to have a high probability of successfully passing the demonstration test, it is necessary to enter the test with a value of M that is above the value of M_R. The value of M is then treated as the design goal for the system. A drawback of this approach is that the higher design goals resulting from the statistical confidence measures may be unrealistic, cost-prohibitive, or technologically unachievable. The issue is further exacerbated when considering an increased emphasis to place new systems into operational usage faster while under tighter budget constraints. As seen in (1), when T is not much larger than M_R, the number of maximum allowable failures will also be small. This often results in test plans in which no failures or a single failure are allowable. Because of the difficulty of achieving this result, the value of M in (2) must then be significantly larger than M_R, which again results in the previously mentioned problems with high reliability design goals.

As seen in (2), the reliability information available from the developmental test program and reliability growth models is not utilized in the OC curve calculations. Other sources of information related to the reliability of the system are also available throughout its development. Modeling and simulation in the form of Physics-of-Failure analysis is generally conducted on important or high cost components, and component or subsystem level testing is also often conducted prior to beginning system level testing. Historical information on the failure rate of certain components within the system may also be available. Use of these data, along with the demonstration test results, could serve to lower the reliability design goals and result in more efficient and cost effective reliability development programs.

- CHAPTER 1 -

The research proposed here examines the assessment of the reliability of a system or product utilizing multiple data sources available throughout the different stages of its development. The various stages of development involve maturation of the design through the discovery and mitigation of failure modes that are initially present in the system. The result is a constantly evolving design with concurrently evolving reliability. The assessment of the reliability as it changes throughout the development of a system is traditionally referred to as reliability growth, which refers to the discovery and mitigation of failure modes within the system, thereby improving the underlying reliability. A number of approaches have historically been developed to model the growth in reliability throughout system development, each with their own applications, assumptions and limitations.

Traditional models for assessing reliability growth, such as those found in Chapter 2, work with test data from individual test events to assess the system reliability at the current stage of development. Depending on the specific technique and associated assumptions, they track or project the reliability of the system as it matures. Tracking involves the estimation of the reliability of a system during its development based on test data alone, similar to standard parametric estimation techniques. Projection provides the expected improvement that will be realized when corrective actions for observed failure modes are implemented. A more in-depth review of some selected reliability growth models is provided in Chapter 2 along with review of more general system reliability assessment approaches.

1.2 Research Objective

The objective of this research is to develop a model framework for assessing

reliability and reliability growth utilizing data from a variety of potential sources. The model and methodology utilizes data collected throughout the development of the system. It also addresses any differences that may exist in the test environments in which the data are collected, while also providing a probabilistic result that indicates the amount of uncertainty in the assessed values. Additional data sources, such as historical information, component level test data, and modeling and simulation results will also be leveraged in the methodology. The general performance of the model is compared to current techniques, and the relative merits of the approach are discussed.

1.3 Research Goals

There are three main goals of the research presented here. These goals directly address the limitations of the current reliability growth paradigm discussed in Section 2. They are as follows:

1. Provide a reliability growth assessment methodology that utilizes data from throughout the development of the system while rigorously accounting for differences in the test environments that may exist.
2. Provide additional management metrics that will provide additional information beyond the assessed reliability to program managers in order to better inform decision-making.
3. Provide a methodology for combining early design activities such as modeling and simulation and component/subsystem testing to provide a prior distribution on the reliability of a system. The prior distribution can then be

updated when reliability growth testing and reliability demonstration testing are completed.

1.4 Research Contributions

The contributions of this research are as follows. A new Bayesian reliability growth assessment technique is introduced for continuous-use systems under general corrective action strategies utilizing data from reliability growth testing throughout the development and maturation of the system. These techniques differ from those currently in the literature due to their allowance for arbitrary times for corrective actions. They also provide a probabilistic treatment of the various parameters within the model, accounting for the uncertainty present in the assessment. The Bayesian formulation also allows for sequential updating as additional phases of developmental reliability growth testing are completed. The capability to combine data across multiple test phases is also a significant enhancement to currently available reliability growth techniques. To date, many reliability growth models are limited to utilizing data from a single test phase. An additional feature unique to Bayesian models that consider multiple failure modes is the capability to account for the failure rate of unobserved failure modes. The model allows practitioners to model complex systems that are comprised of large numbers of failure modes, recognizing that only a relatively small number of failure modes can be observed during system testing.

The Bayesian reliability growth assessment model is then extended to include results from operational testing. The approach considers the posterior distribution from the reliability growth assessment of the prior for the operational reliability assessment. The DT and OT environments are not a priori assumed to be equivalent

though. System reliability generally degrades when transitioning from a developmental test environment to an operational test environment, and a degradation factor is added to the model to account for this difference. The degradation factor is treated probabilistically within the Bayesian framework, and the resulting marginal posterior distribution on the system failure intensity is used to provide an assessment of the operational reliability. The approach reduces the uncertainty in the operational reliability assessment when compared to the classical approach using the HPP while still accounting for the uncertainty in both the prior failure intensity and the degradation between test environments. New statistical risks for reliability demonstration are also developed which serve as alternatives to those traditionally available through OC curve analysis.

A Bayesian reliability growth planning model is also presented that takes advantage of the reduced uncertainty in the operational assessment. The approach allows for reductions in the amount of demonstration testing necessary for a given level of "confidence" in the assessment, and it can also be used to reduce high design goals that often result from traditional OC curve applications.

The final part of this research involves combining various sources of reliability information to obtain prior distributions on the system reliability to aid in later reliability growth modeling. The approach presents a general framework for utilizing information such as component/subsystem testing, historical component reliability data, and physics-based modeling of specific component failure mechanisms.

The methods presented here also allow for estimation of various management metrics such as initial system failure intensity, projected failure intensity after failure

mode mitigation, growth potential failure intensity, and the rate of occurrence of unobserved failure modes. These types of metrics provide reliability program managers additional information with which to assess the maturity of the system under development. At the same time they can also aid in quantifying the risk that may exist in achieving the reliability goals of the development program.

It should also be noted that the techniques proposed here could serve more generally to assess the reliability of a system even when reliability growth is not being realized through a corrective action process. The model framework is also flexible enough to use as a general data fusion process for assessing the overall system reliability even when reliability growth is not present.

1.5 Research Overview

The rest of this thesis is organized as follows. Chapter 2 provides a literature review of reliability growth models and system reliability approaches that are available in the literature. Chapter 3 presents a Bayesian reliability growth projection model for continuous-type systems that can be applied to multiple systems under test with arbitrary corrective action strategies. Chapter 4 develops a Bayesian method for assessing reliability using a combination of developmental reliability growth test data and operational reliability demonstration data. Chapter 5 provides a Bayesian approach for reliability growth planning using the modeling concepts developed in Chapters 3 and 4. Chapter 6 presents a Bayesian framework for developing prior distributions for system level reliability growth testing using a combination of historical data, component/subsystem testing, and physics-based modeling. Chapter 7 presents a case study that applies the assessment approaches in Chapters 3, 4 and 6 to

a complex system. Chapter 8 discusses areas for future work involving extensions of the research presented here to discrete or one-shot systems. Chapter 9 provides conclusions regarding the research conducted. A brief overview of each chapter is presented below.

1.5.1 Chapter 2 – Literature Review

A number of reliability growth models and reliability assessment techniques have been developed and presented in the literature. Each has been developed for the purpose of providing reliability practitioners with the capability to estimate and manage an overall reliability program for a specific situation of concern. This research covers not only the development of new reliability growth models, but also the more general probabilistic modeling of reliability. For this reason the literature review in Chapter 2 is divided into three major components, the first involving reliability growth models, the second discussing reliability demonstration testing, and the third involving more general reliability assessment techniques. The first section on reliability growth is further divided into sections on reliability growth planning, tracking, and projection. The second section covers reliability testing, and the third section focuses on data combination techniques for reliability assessment.

1.5.2 Chapter 3 – A Bayesian Model for Complex System Reliability Growth Under Arbitrary Corrective Actions

Chapter 3 presents a new method for projecting the reliability growth of a complex continuously operating system. The model allows for arbitrary corrective action strategies, and it differs from other models of this type by using all available

data rather than failure mode first occurrence times only. It also differs from other reliability growth projection models, in that it provides a complete inference framework via the posterior distribution on the system failure intensity. A unique feature of this approach relative to other Bayesian techniques is the analytic expression for the failure intensity contribution from unobserved failure modes. Expressions for the estimating the initial failure intensity, growth potential failure intensity, and the cumulative number of failure modes expected in future testing are also developed. Extensions to the basic framework are also developed. The first accounts for multiple systems under test, and the second develops the posterior distribution while allowing for uncertainty on the FEF values that are assessed. Two separate goodness-of-fit procedures are presented for assessing the appropriateness of the underlying model assumptions.

The main assumptions of the approach are:

1. The system is comprised of a large number of failure modes that are serial in nature; the occurrence of any failure mode results in failure of the system.
2. Failure modes generate failures independently of one another.
3. The failure intensity, or rate of occurrence of failure, for each mode is constant both before and after a corrective action is implemented.
4. The resulting failure intensity after corrective action will be reduced from the initial value according to the assigned FEF.
5. Corrective actions to failure modes do not introduce new failure modes into the system.
6. Failure mode failure intensities have a common Gamma prior distribution.

7. Reliability testing will stress the system in an operationally relevant manner.

1.5.3 Chapter 4 - Assessing Reliability Growth Using Developmental and Operational Test Data

Chapter 4 presents a new Bayesian reliability assessment model to mitigate the problem of reliability demonstration with fixed configuration testing alone. The approach allows for the combination of developmental reliability growth test data with operational reliability test data within a Bayesian probabilistic framework. The reliability data available throughout DT can provide a substantial amount of reliability information that can inform the assessment of reliability in OT. The result is less uncertainty in the assessed reliability than would be present with operational test data alone. Previous data combination methods do not explicitly model traditional reliability growth, which refers to the discovery and mitigation of failure modes within the system. This process ultimately improves the underlying reliability of the system. These models generally work with test data from individual test events to assess the system reliability at the current stage of development. They do not generally consider information that may be available from previous testing, which is particularly useful when individual test events are limited in size. Differences in the test environments and stressors must always be considered when combining data from different test events, and the model explicitly accounts for the degradation that usually exists when moving from developmental to OT. Interval procedures and model assessment techniques are also presented to aid in practical application of the proposed method. Analogous OC curve results are also developed, which can be

used to help plan reliability demonstration testing. The approach developed in this chapter is used as the basis for the reliability growth planning model in Chapter 5.

1.5.4 Chapter 5 – Reliability Growth Planning Using Combined Developmental and Operational Test Data for Reliability Demonstration

Reliability growth planning is a specific subset of reliability growth modeling that is useful for designing reliability test programs. The models are developed and used before actual test data are available, but they are generally based on a corresponding assessment technique. More recent planning models require the user to provide certain inputs that are important to the overall reliability growth program. These inputs include parameters such as the Management Strategy (MS), FEF, initial Mean Time between Failure (MTBF) value, and MTBF requirement.

System level reliability growth planning culminates in a reliability demonstration test to assess whether or not a new acquisition system has met its reliability requirement. This demonstration is statistical in nature, and therefore requires management and understanding of the associated statistical risks of the planned test. Traditional models consider the reliability demonstration event separately from the reliability growth, using an OC curve to manage these risks. This often results in reliability design goals that are significantly higher than the reliability requirements themselves. In many practical cases these goals are more than double the actual requirements, but they are purely a result of the statistical estimator being employed to assess the reliability.

This chapter presents a new reliability growth planning model that explicitly combines developmental and operational data from different test events for reliability

demonstration. The model is a natural extension to the traditional reliability growth planning models that are currently used. Differences in the test environments and stressors must always be considered when combining data from different test events, and the model explicitly accounts for the degradation that usually exists when moving from developmental to OT. The proposed approach explicitly models the uncertainty in both the system failure intensity and the degradation between test environments within a Bayesian framework, allowing for narrower uncertainty intervals and reduced reliability goals for statistical demonstration. The approach directly addresses many of the existing issues with the traditional OC curve based reliability growth planning.

1.5.5 Chapter 6 – Development of Prior Information Using Lower-Level Data Sources

Chapter 6 presents a general methodology for developing prior information on the system reliability by combining lower level information that may be available before full system-level testing has been conducted. System level data collected under operationally relevant testing is the most desirable information source for assessing system reliability, but during early stages of development this type of information may be unavailable. There are often other sources of reliability related information that are available in the early stages of development of the system though, and in these cases it is possible to utilize these information sources to develop an early assessment of the system reliability. When viewing the process of reliability assessment across the various stages of development of the system, assessment in this

manner serves as prior information that can be updated with the reliability growth models in Chapters 3 and 4 when system-level test data are available.

The reliability growth models are developed for complex repairable systems, where components and/or subsystems fail and are replaced throughout the life of the system. The process of component failure and replacement is assumed to result from renewal processes across multiple potential failure mechanisms, which can then be used to develop estimates of component failure intensities via a Homogeneous Poisson Process. A system level representation of the reliability structure, such as a reliability block diagram or fault tree, is then used as a means of combining the information on the components and subsystems. The approach reduces the complex system structure into an equivalent series representation of failure modes, which aligns with the model assumptions in Chapters 3 and 4. The connection between component or redundant block failures and their corresponding failure modes is also developed.

Uncertainty distributions are developed for each failure mode using data for the individual components or those within a redundant block. Bayesian posterior distributions are used when component data are available, and a more general uncertainty distribution is developed when only physics-of-failure model results are available. When the component information that is available does not accurately represent the reliability of the component within the new system (e.g. benign testing, historical data from similar system, etc.), a probabilistic technique is provided to account for the degraded reliability and additional uncertainty that is present due to

additional failure modes or mechanisms that are not accounted for in the data or modeling.

1.5.6 Chapter 7 – Reliability Assessment Throughout Development: A Case Study

Chapter 7 presents a case study that combines the techniques of Chapters 3, 4, and 6 for a military system. The approach in Chapter 6 is first used to develop the prior distribution on the system reliability before system-level testing is conducted. Historical data on the subsystems and components of the system are used along with stress-life modeling of fatigue on the driveshaft within the driveline of the system.

The prior distribution is then updated with test results from developmental reliability growth testing. The posterior distribution after DT is developed using the model in Chapter 3. The model in Chapter 4 is then used to update the DT posterior distribution with results from a limited operational test. The case study demonstrates the practical use of the reliability assessment technique throughout development that is the major intent of this research.

1.5.7 Chapter 8 – Future Work

Chapter 8 presents opportunities for future work in this area. A discussion of the approach as applied to discrete-type (i.e. one shot devices) is presented. A Beta prior distribution is used as the discrete analogue to the Gamma distribution, and a potential discrete extension to the reliability growth model in Chapter 3 is provided. A brief discussion of the analogous approach for combining developmental and operational test data in Chapter 4 is also provided, although analytic results are not developed. The chapter closes by briefly discussing areas for further research

involving probabilistic modeling of failure mechanisms using information theory concepts.

1.5.8 Chapter 9 - Conclusion

Chapter 9 provides a general discussion and conclusions for the work presented throughout the thesis. The contributions of the research are restated and compared to the original objectives. A summary of each chapter is also provided.

2 LITERATURE REVIEW

A number of reliability growth models have been developed and presented in the past. Each has been developed for the purpose of providing reliability practitioners with the capability to estimate and manage an overall reliability program for a specific situation of concern. This research covers not only the development of new reliability growth models, but also the more general probabilistic assessment of system reliability. For this reason the literature review is divided into three main sections. The first reviews various reliability growth models in the literature, the second provides an overview of some recent papers on reliability demonstration testing, and the third discusses some recent system reliability assessment techniques. The first section on reliability growth is also further divided into sections on reliability growth planning, tracking, and projection.

2.1 Reliability Growth planning Models

2.1.1 Duane Model (1964)

The Duane Model [11] was developed based on the observation that changes to a system design to improve reliability resulted in a specific functional relationship when examining the cumulative failure rate of the system with respect to cumulative test time. This relationship was seen to be linear when examined on a Log-Log scale, and is commonly known as the Duane Postulate. The negative of the slope of the line is referred to as the “growth rate”, as it provides an indication of the rate at which the system reliability is improving. The functional form underlying the Duane Model is utilized as a fundamental assumption in a number of reliability growth models that

are discussed later in this chapter. The original intent of the model was for reliability growth tracking, but it is included under planning due to its importance in the development of a number of other reliability growth planning and tracking models.

2.1.2 Selby and Miller's Reliability Planning Management Model (1970)

Selby and Miller developed the Reliability Planning Management Model [12] in 1970. The model provides an approach for planning and managing reliability programs for complex systems, and the basic concept behind the model is that the reliability growth of the system follows the Duane Postulate. The Duane Postulate is a common assumption underlying many reliability growth models in the literature, but the model is the first known use of the Duane postulate for reliability growth planning.

2.1.3 Military Handbook 189 Model (1982)

The reliability growth planning model presented in MIL-HDBK 189 [13] is based on the Power Law Non-Homogeneous Poisson Process (NHPP) Model first presented by Crow [14] in 1974. As discussed below in the section on reliability growth tracking, the Power Law NHPP was the first stochastic application of the Duane Model. The MIL-HDBK 189 Model is the second reliability growth planning model to be based on the Duane Postulate, and its purpose is to provide a plan for reliability improvement over a period of multiple phases of DT. The model is one of the earliest to outline an approach for reliability growth management during a system reliability growth program, where the achieved reliability growth can be compared to the planned values from the model.

The model develops an idealized reliability growth curve for a test-analyze-fix-test (TAFT) process. This process depicts the pattern of reliability improvement when corrective actions are applied to observed failure modes when the failure modes are discovered during the test. The idealized growth curve is defined by a number of parameters that are common in reliability growth planning models. These include the initial Mean Time Between Failures (MTBF), the length of the initial test phase, the goal MTBF at the end of the test program, the growth rate, and the total amount of testing in the entire reliability growth program. The model also provides incremental reliability steps that depict the planned reliability targets for the various phases of the developmental test program. The steps are taken as the average value of the idealized curve over the developmental test phases.

The AMSAA System Planning Model (SPLAN) [6] is a later extension of the MIL-HDBK 189 Model. The SPLAN model requires only four of the five input parameters mentioned above, solving automatically for the remaining input.

2.1.4 AMSAA Subsystem Planning Model (1992)

The AMSAA Subsystem Planning Model (SSPLAN) [15] developed reliability growth planning curves at the system or subsystem level, where the MTBF requirement is to be demonstrated with a desired level of statistical confidence. The model determines testing and MTBF requirements for the subsystems of interest in order to have the system MTBF meet the desired requirement with confidence. OC curve analysis is also provided [16], where the consumer and producer risks are developed in terms of the model parameters.

2.1.5 AMSAA Planning Model Based on Projection Methodology (2006)

The AMSAA Planning Model Based on Projection Methodology (PM2) [17] is the most recent reliability growth planning model to be developed through work at AMSAA. The model was developed with the purpose of providing a reliability growth plan to aid in the management of developmental reliability programs for complex systems. The main difference of this model the other planning models previously discussed is that it is independent of the NHPP assumption. The model is instead based upon the "doubly-stochastic" process developed in the AMSAA Maturity Projection Model (AMPM) [18], and it uses parameters that can be directly influenced by reliability program management. These include parameters such as the initial MTBF, MS, goal MTBF, total test time, average FEF, the number and placement of Corrective Action Periods (CAPs), and the planned developmental test hours. An additional input metric is the average lag-time associated with implementing corrective actions for observed failure modes. PM2 is thought to be the first planning model to consider the impact of this lag time, which can significantly impact the reliability improvement that is actually realized during system development.

As with other planning models, PM2 also provides an idealized curve along with incremental reliability steps for specific test phases. An additional difference with PM2 is that the incremental steps are developed while applying the corrective action lag time, which results in them falling entirely below the idealized reliability growth curve. These steps are based on the realization that a majority of corrective actions to failure modes observed during test events are made during CAPs between

the test phases. United States Army Policy [5] has recently specified PM2 as the preferred reliability growth planning model for developmental reliability programs for complex military systems.

2.1.6 Crow Extended Model for Reliability Growth Planning (2010)

The Crow Extended Planning Model [19] is a modified and improved version of the MIL-HDBK 189 model, and is based on the projection model of the same name [20]. It utilizes many of the advancements first developed in the PM2 model, and the input parameters are mainly the same. These include parameters such as the initial MTBF, MS, goal MTBF, total test time, average FEF, the number and placement of CAPs, the planned developmental test hours, and the average lag-time associated with implementing corrective actions for observed failure modes. An additional input known as the discovery Beta is also required, which describes the rate at which new correctable modes will be discovered during testing. Because of the inherent connection to the NHPP associated with the MIL-HDBK 189 Model, a discovery Beta values less than one indicates reliability growth is occurring.

2.1.7 Hall Discrete Planning Model Based on Projection Methodology (2011)

The Planning Model Based on Projection Methodology – Discrete (PM2-Discrete) [21] was developed as an analogue to the PM2 model developed for continuous systems. The model is based on underlying reliability growth projection methodology developed by Hall [22] to address the lack of projection models for discrete one-shot type systems. The model provides a number of uses for reliability program managers, such as determining planned reliability achievement for available

program resources, serving as a baseline target which realized reliability values can be compared against, and quantifying the feasibility of a test program for achieving final reliability goals. The model also provides a series of useful metrics associated with the reliability growth of the system.

2.2 Reliability Growth Projection Models

2.2.1 Corcoran, Weingarten, Zehna Model (1964)

Corcoran, Weingarten and Zehna [23] developed the first model for estimating reliability after corrective action. The approach was developed with consideration to estimating reliability in the final stage of development of an "expensive item." The reliability projection is suitable in cases where corrective actions are installed at the end of a test consisting of N independent trials. The trial outcomes are assumed to follow a multinomial distribution with parameters N (total number of trials), q_0 (unknown success probability), and p_i (unknown failure probability for failure mode i = 1, ..., k). The assumption of a multinomial model implies that at most one failure mode can occur per trial. An exact expression for the system reliability is presented, and comparisons of various estimators are provided.

2.2.2 AMSAA Crow Projection Model (1982)

The AMSAA Crow Projection Model [24] uses the NHPP interpretation of the Duane Postulate to describe the rate of occurrence of failure modes in the system. The intent of the model is to project the growth in reliability that would be seen at the beginning of the next phase of testing following implementation of planned corrective

actions. In this regard, the model assumes that all corrective actions are delayed until the end of the current test phase. The model is also one of the first to introduce the concept of the reliability growth potential, which is the theoretical upper limit on reliability that can be achieved via the test-fix-test reliability growth paradigm. This concept is an important factor that governs reliability growth programs in general, and it is commonly considered and monitored for current reliability growth programs in the U.S. Department of Defense. Two separate goodness-of-fit procedures are available, a Cramer Von-Mises test and Chi-Squared test, but no interval procedures have been published.

2.2.3 AMSAA Maturity Projection Model (1995)

Instead of a direct NHPP assumption, the AMSAA Maturity Projection Model (AMPM) [18] uses a "doubly-stochastic" process to describe the underlying behavior of the system failure intensity. The model assumes that the system is comprised of a number of failure modes, with the collection of mode failure rates being realizations from a common Gamma distribution. The time between failures for each mode is then assumed to be Exponential. AMPM is the first projection model to allow for arbitrary corrective actions, as the corrective actions can occur during the test or be delayed until after the test. Because of this it uses only the first occurrences of each failure mode to develop failure intensity estimates. The AMPM is also the underlying methodology on which the PM2 reliability growth planning model [17] is founded.

In addition to the system level failure intensity, the model also provides estimates for the expected number of observed failure modes in later testing, the rate of occurrence of new failure modes, and the percent of the initial failure intensity

comprised of the modes that have been surfaced. Goodness of fit procedures are available using the expected number of failure modes, but no confidence intervals have been developed to date.

2.2.4 Clark Projection Model (1999)

The projection model proposed by Clark [25] was developed due to the recognition that many programs do not achieve significant reliability growth until late in the program near production. The proposed reasoning for this occurrence is the lack of focus on reliability early in the development of a new system. The Clark model is an extended version of the AMSAA Crow projection model [24] that has two main differences. The first is that the original model is modified to allow for arbitrary corrective actions, and the second is the addition of an inherent failure rate term that allows for decisions to be made regarding future reliability investment. If the current reliability is too close to the maximum possible value, it may not be cost effective to continue to invest in further reliability improvement through test-fix-test processes.

2.2.5 AMSAA Maturity Projection Model – Stein (2004)

The AMSAA Maturity Projection Model – Stein (AMPM-Stein) [26] was developed as an extension to AMPM [18]. The extension does limit one of the original assumptions of the model, as the corrective actions in this case must be delayed until after the test. The model uses the same underlying theoretical structure as the original AMPM, but additional data are used to develop the model estimates. All of the data, both first and repeat occurrence times, are used to develop shrinkage

estimates, or Stein estimates, [27] to develop the model. The benefit of the approach is that the use of the additional data provides an increase in the accuracy of the resulting estimates. The shrinkage estimation minimizes the mean square error value, which provides an immediate connection to Bayesian modeling using squared error loss functions. As with the original AMPM, goodness of fit procedures are available, but no confidence interval methods have been reported to date.

2.2.6 Crow Extended Model (2004)

The Crow Extended Model [20] was developed to model arbitrary corrective action strategies using the previously existing AMSAA-Crow NHPP modeling framework. The Extended Model is a straightforward combination of the AMSAA Tracking Model [28] and the AMSAA Crow Projection Model [24]. Failure Modes are classified using the traditional A-mode and B-mode distinction, where A-modes are those failure modes that will not be addressed via corrective action. The B-modes are further divided into BC-modes and BD-modes, with BC-modes having corrective actions implemented during the test phase and BD-modes having the corrective action delayed until after the test is complete. The model uses all A, BC, and BD-mode failures in the AMSAA Tracking Model to get an estimate of the reliability growth that occurs during the test. The BD-mode failure intensity is then estimated using the maximum likelihood estimate n_{BD}/T for n_{BD} failures in test time T. Because the BD-mode corrective actions are delayed, their growth contribution during the test must be subtracted from the Tracking Model estimate and replaced with a more appropriate estimator. The BD-mode failure intensity after corrective action is then estimated with the AMSAA Crow Projection Model [24]. The overall result for the Extended

Model then subtracts the BD-mode maximum likelihood estimate n_{BD}/T from the Tracking Model result and replaces it with the AMSAA Crow Projection Model result.

The model has been shown through simulation study [7] to provide extremely optimistic results when a large proportion of corrective actions are delayed. There is also a logical discrepancy by treating the A and BD-modes together with the BC-modes with the AMSAA Tracking Model, which assumes that reliability growth is occurring because failure modes are being addressed during the test. The attempt to overcome the issue by subtracting out the BD-mode contribution leaves a bias in the model that can provide systemically optimistic results.

2.2.7 Hall Discrete Projection Model (2008)

The discrete reliability growth projection model proposed by Hall [29],[30] is a discrete counterpart to the AMPM-Stein Model [26]. The model uses Stein-estimation procedures [27] to develop shrinkage estimates for the failure intensity of unobserved failure modes in the system. All corrective actions are delayed until the end of the current test phase, and more than one failure mode can occur on a given trial during the test.

Analogous to the AMPM Model [18], the discrete method proposed by Hall uses a geometric likelihood for the first occurrence trial of an observed failure mode, and the mode probabilities of failure are assumed to be a realization from an underlying Beta distribution. Both Method of Moments and Maximum Likelihood Estimators are provided, and results are developed for systems with a known number of failure modes and those assumed to be complex with a large number of modes. A

number of associated management metrics are also presented, such as the expected number of new failure modes to be observed during additional testing, the rate of occurrence of new failure modes, and the reliability growth potential of the system. Model performance is also studied via Monte Carlo simulation, and results indicate that performance is reasonable with small errors in the projection estimates.

2.2.8 Bayesian Methodology for Discrete Reliability Growth (2009)

The discrete reliability growth methodology presented by Hall and Mosleh in [31] was developed as an additional estimation procedure to those first presented in [22] and [29]. The approach again uses the underlying theoretical assumption of mode failure probabilities as realizations from an underlying Beta distribution. Additional assumptions include a Binomial distribution for observed failures during test, with all corrective actions delayed until the end of the testing.

The Bayesian inference in the model is used only to estimate the parameters of the underlying Beta distribution. Squared error loss is used along with a constant prior, and numerical methods are used to evaluate the resulting posterior. Simulation methods are also used to generate uncertainty distributions on each of the management metrics developed in [29] and [30].

2.2.9 Hall, Ellner, Mosleh Discrete Reliability Growth Projection Model (2010)

The model presented in [22] uses the underlying assumption of mode failure probabilities as realizations from a Beta distribution. The model differs from the previously presented discrete models though, as it allows for arbitrary corrective actions to occur either during or directly after the test. Only failure mode first

occurrence trials are used, along with the corresponding FEF for each failure mode. Because only the first occurrence of each failure mode is used, the Geometric distribution is used to model the mode first occurrences. Goodness-of-fit procedures are presented in order to validate the model assumptions.

Maximum likelihood estimates are developed for the parameters of the Beta distribution, with results given for a finite number of failure modes and a complex system consisting of a large number of failure modes. A number of management metrics and model equations are also developed, such as the reliability growth potential, the expected number of new failure modes, and the fraction of the initial failure probability surfaced during the testing.

2.3 Reliability Growth Tracking

2.3.1 Weiss's Reliability Growth Model (1956)

Weiss [32] provided an early method for modeling the reliability growth of guided-missile systems. The approach assumed Poisson failures for the system, with the system's Mean Time to Failure (MTTF) changing over successive tests as failure modes were observed and mitigated. The approach is flexible enough to allow for both increasing and decreasing reliability as a function of time. The distribution of the time-to-failure estimator is provided to allow for calculations on the uncertainty.

2.3.2 Lloyd and Lipow's Reliability Growth Model (1962)

Lloyd and Lipow [33] provide a reliability growth model for system containing a single failure mode. Discrete test events are used, with corrective actions made to

the system after any failures. The approach allows for corrective actions to be successful with a specified probability. The reliability of the system at the n^{th} trial is given as

$$R(n) = 1 - a\exp\left[-b(n-1)\right], \tag{1}$$

where a and b are model parameters that are estimated from the data. Other functional forms of potential reliability growth models are also discussed.

2.3.3 Cox and Lewis's Reliability Growth Model (1966)

Cox and Lewis [34] proposed an early NHPP model. The functional form is defined by

$$m(t) = \exp[at + b], \tag{2}$$

where a and b are estimated from the test data. Reliability growth occurs when a is less then zero. Goodness-of-fit procedures are also presented for model assessment.

2.3.4 Barlow and Scheuer Reliability Growth During a Development Testing Program (1966)

Barlow and Scheuer [35] present a reliability growth model involving K stages of DT for a system. A trinomial approach is used to define success and failure, with failure further divided into inherent and assignable cause categories. Inherent failures are defined as "those which reflect the state-of-the-art and whose elimination would require an advancement thereof", and assignable cause failures are defined as failures which can be corrected through design or operational corrective actions. The reliability of the system in the i^{th} stage of testing is given by $r_i = 1 - q_0 - q_i$, where q_0

is the probability of an inherent failure and q_i is the probability of assignable cause failure at the i^{th} stage. Estimates of the failure probabilities and reliability are presented using the trinomial framework. A caution is also provided regarding interpretation and use of the resulting estimates. The authors point out that assignable cause failures may "mask" inherent failures and vice versa, and recommend that the reliability estimate is the only one that should be “trusted”.

2.3.5 Pollock’s Bayesian Reliability Growth Model (1968)

Pollock’s model for reliability growth [36] provides estimated improvements in reliability due to corrective actions that are implemented for the system. The model assumes a decreasing trend in the failure rate for each system, and the probability of successful repair is specified in order to quantify the probability of transitioning from a failed state to a repaired state. The projections are developed for both continuous and discrete systems with separate prior distributions used for before and after corrective actions. Separate estimation procedures are also given for before and after testing and corrective actions have been performed.

The model contains concepts that are similar in application and intent to the FEF values that are used in more recent reliability growth projection models, although they are somewhat more restrictive in the treatment of corrective actions. Pollock’s μ is equivalent to the remaining failure rate after a corrective action has been applied, which is given by $(1-d)\lambda$ for FEF d and failure rate λ. The parameter a is the repair probability, or the probability of the failure rate decreasing to μ. When applied together, these terms do allow for the possibility of non-effective corrective

actions to be made to the system. They do model a binary process though, as the corrective action is either successful or it is not, and levels of imperfect corrective actions cannot be considered. Current approaches using FEF values now allow for varying degrees of corrective action effectiveness represented by the assigned value of the FEF for the specified failure mode.

2.3.6 Littlewood and Verrall Bayesian Reliability Growth Model (1973)

The Bayesian reliability growth model proposed by Littlewood and Verrall in [37] was originally developed for computer software. The concepts underlying the model are introduced within the software development context, but they are easily applicable to a wide variety of situations. The model assumes a Gamma-Exponential process, with Exponential time between failures and failure rates distributed via the Gamma distribution.

The scale parameter of the Gamma is the focus of the reliability growth in the model. The parameter is allowed to vary with time, which implicitly accounts for the effectiveness of corrective actions that are implemented. Estimation of the Gamma scale parameter is also discussed, and numerical procedures are provided along with a Kolmogorov-Smirnov goodness of fit approach that eliminates that need for numerical integration procedures.

2.3.7 Crow's Reliability Growth Tracking Model (1974)

Crow's tracking model [14] is a probabilistic extension of the Duane model [11]. The same log-linear relationship of the Duane model is used to describe the

underlying functional form of the tracking model. The probabilistic extension treats this relationship as a NHPP. The mean value function of the NHPP is given as

$$\mu(t) = \lambda t^{\beta} . \tag{3}$$

The assumption of an NHPP for the number of failures allowed for the development of statistical estimators and associated goodness-of-fit tests, which gave more statistical rigor to the repairable systems reliability growth that was covered by the Duane model. The model is also known as the "Weibull Process" due to the mathematical property that the time to first failure in the assumed NHPP is a Weibull random variable. Another common name is the "Power Law Process", which is a direct result of the functional form of the NHPP.

This approach to reliability growth tracking became the foundation for a number of later models, and the same underlying NHPP form was used for an assortment of planning, tracking, and projection models. Crow continued to expand upon the original development, adding confidence bounds on the NHPP parameters and associated MTBF, and extensions to include both time-truncated and failure-truncated testing. Straightforward extensions were later provided to account for multiple identical systems under test.

2.3.8 Smith's Bayesian Note on Reliability Growth during Development Testing (1977)

This short paper [38] provides a Bayesian method for estimating the final reliability that is achieved after multiple Binomial tests have been conducted. Uniform prior distributions are used on the successive Binomial parameters. The

marginal distribution for the probability of success on the last configuration of the system is developed. The result is a convex combination Beta functions, and a numerical example is provided for comparison with other previously developed techniques.

2.3.9 Fard and Dietrich's Bayes Reliability Growth Model for Development Testing (1987)

Fard and Dietrich [39] provide a correction to the model proposed by Smith [38]. The assumptions surrounding the test program are the same as those originally used by Smith: the development consists of a series of binomial tests, the reliability is assumed to be non-decreasing, and the prior distributions for each of the successive reliabilities are uniform. An example is provided for comparison, and a simulation study is also presented to examine the performance of the corrected estimator. The model is shown to perform as good or better than Read's Barlow Sheuer model, and the mean squared error decreases significantly as the number of test stages is increased.

2.3.10 Engelhart and Bain's Prediction Intervals for the Weibull Process (1978)

Engelhart and Bain [40] provide a short paper on prediction intervals for the model form proposed by Crow in [14]. The intent of the paper is to provide statistical inference on the k^{th} future observation of the model process, to include point estimates and confidence intervals. Results are derived with the intent of avoiding numerical integration. For values other than $k = 1$, simplified approximations are provided based on available Chi-Squared percentile calculations. An example

application is also provided, with comparisons between the assumed Weibull process and an exponential process model.

2.3.11 Barlow Scheuer Reliability Growth from Bayesian Viewpoint (1978)

Weinrich and Gross [41] present the Barlow-Scheuer Reliability Growth Model using Bayesian methods for estimation. As in the Barlow-Scheuer Model [36], failures are assigned to two categories: inherent failures and assignable-cause failures. As in the original version of the model, this assignment is similar to current reliability growth models that divide failure modes into A and B modes, where A modes are those inherent failure modes that will not receive a corrective action and B modes are those failure modes for which a corrective action will be implemented when observed. Dirichlet priors are used for the three possible outcomes of assignable cause failure, inherent failure, or success.

The model assumes that testing occurs in independent stages, but differs from the original version in the treatment of assignable cause failures. When a corrective action is implemented for an assignable cause failure, the original Barlow-Sheuer model redistributed the failure probability into the success category. The Bayesian version is somewhat more conservative in its approach though, as it allows for the assignable cause probability to be redistributed among the other causes of failure along with the success. This treatment still assumes that the corrective action is perfect, with the failure probability for the specific failure reduced to zero after the corrective action occurs.

2.3.12 Langberg and Proschan's Relability Growth Involving Dependent Components (1979)

Langberg and Proschan [42] provide a theoretical discussion on modeling reliability growth when dependency may exist between components within the system. The approach involves transforming the problem involving dependency into an equivalent model with independence. Consistency between the limiting forms of the dependent and independent approaches is demonstrated, but no practical examples are provided.

2.3.13 Goel and Okumoto Time Dependent Error-Detection Rate Model for Software Reliability and Other Performance Measures (1979)

Goel and Okumoto [43] provide a NHPP model for modeling the occurrence of software failures. The objective of the approach is to provide a parsimonious model with parameters that have an underlying physical interpretation while also yielding a quantitative assessment of software performance. The main assumption is a standard NHPP, and the mean value function is motivated by assuming proportionality to the expected number of unobserved failures in the code. This assumption results in a difference equation that can be easily solved to find the mean value function of

$$\mu(t) = a\left(1 - e^{-bt}\right). \tag{4}$$

The parameter b is equal to the ratio of detected failures to remaining failures, which fits the original intent of the model.

Various performance measures of interest are derived along with MLEs for the model parameters. An example application is provided, and comparisons are made between the proposed NHPP and the Jelinski-Moranda model [44]. The results show that the NHPP fits the example failure data well, and the performance measures are also shown to be more conservative than the Jelinski-Moranda approach.

2.3.14 Crow's Discrete Reliability Growth Tracking Model (1983)

Crow's discrete reliability growth tracking model [45] can be considered as the discrete analogue of the continuous version [14] first developed in 1974. The model uses the same NHPP assumption for the cumulative number of failures as the continuous version. This allows for development of expressions that describe the change in probability of failure from configuration to configuration. Statistical estimators and associated goodness-of-fit metrics are developed for the model. Results are also provided for both grouped data (i.e. where each configuration involves multiple test trials) and trial-by-trial data.

2.3.15 Littlewood's Rationale for a Modified Duane Model (1984)

Littlewood [46] provides a discussion of the Duane Model and its associated Power Law NHPP. Two undesirable properties are noted: the first being that the rate-of-occurrence of failiures (ROCOF) is infinite at time $t = 0$, and the second being that the ROCOF is zero as the time t goes infinity. A straightforward modification to handle the second problem is proposed. It involves the addition of a constant term to represent the rate of occurrence of failures that cannot be removed from the system, and no further development is provided.

The paper instead focuses on the first problem of an infinite ROCOF at time zero. The proposed approach is to model the occurrence of independent exponentially distributed faults with rates that are realizations of a Gamma distribution. No estimation procedures or associated goodness-of-fit procedures are presented. The resulting NHPP is a modified Power Law with a finite value of the ROCOF at time zero. This approach appears to be similar in concept to Littlewood and Verral [37], in that it examines the reliability on a failure or fault mode basis. The NHPP is used in this case to describe the behavior between the faults rather than the failures themselves. An additional interesting point in the paper is the further discussion involving the use of the NHPP to model the occurrences of failures in a reliability growth setting. It is noted that the process of observing and then correcting a failure does not result in independent increments, and therefore the assumption of an NHPP for failure occurrences in reliability growth is not appropriate.

2.3.16 Robinson and Dietrich's Nonparametric Bayes Reliability Growth Model (1989)

Robinson and Dietrich [47] present a nonparametric approach to assessing the reliability growth of a system using Bayesian methods. The main model assumptions are Exponential times between failure and non-decreasing reliability over successive stages of testing. Each stage of testing is broken into distinct test periods. General expressions for the moments of the resulting posterior distribution on the failure intensity are provided, along with an additional recursive estimation formula. A numerical example is provided, and simulation studies are presented to compare the

model performance to the Power Law model and a previously developed nonparametric technique.

2.3.17 Bayes Inference for Power Law Non-Homogeneous Poisson Process (1989)

Guida, Calabria, and Pulcini provide a Monte Carlo simulation study of the power law NHPP using Bayesian procedures [48]. Multiple cases are presented, each assuming different non-informative and informative priors for the α and β parameters in the NHPP parameterized as

$$\lambda(t) = \left(\frac{\beta}{\alpha}\right)\left(\frac{t}{\alpha}\right)^{\beta-1}. \tag{5}$$

The first non-informative case uses a joint Jeffrey's prior for both of the parameters, and a second non-informative case uses a uniform prior for β and a Jeffrey's prior for α. Informative priors for the NHPP parameters are developed by using a Gamma distribution to describe the expected number of failures in the time interval given by

$$m(T) = \left(\frac{T}{\alpha}\right)^{\beta}. \tag{6}$$

The simulation comparison results indicate that the Bayesian methods provide better estimates than the classical maximum likelihood estimators. A number of cases were examined, and the benefits of the Bayesian methods were observed even in cases where there was a modest amount of prior information or even weak prior information.

2.3.18 Singpurwalla and Soyer's Non-Homogeneous Autoregressive Processes for Tracking (Software) Reliability Growth, and Their Bayesian Analysis (1992)

Singpurwalla and Soyer [49] present a reliability growth model where the life between successive modifications to the system is assumed to be log-normally distributed. The reliability growth or decay is described by a power law non-homogeneous autoregressive process given as

$$X_t = X_{t-1}^{\theta_t} \varepsilon_t, \tag{7}$$

where θ_t implies growth or decay and ε_t is log-normally distributed.

Two Bayesian formulations are considered, which relax constraints on the existence of least-squares estimators for higher order processes (i.e. greater than one). The first approach assumes exchangeability of coefficients in the prior distribution, while the second imposes an autoregressive relationship between the successive values of θ_t. The two approaches are compared using software reliability data, and likelihood ratio results show that the exchangeable prior approach is preferred over the autoregressive approach.

2.3.19 Mazzuchi and Soyer's Reliability Assessment and Prediction During Product Development (1992)

Mazzuchi and Soyer [50] present a Bayesian assessment approach for discrete type systems. The underlying motivation for the approach is a test-fix-test approach involving multiple identical trials until a failure is observed. Upon observation the

failure is addressed through a modification to the system that results in increased reliability.

The trials can therefore be modeled with a geometric likelihood, with the additional assumption that the sequence of reliability values resulting from successive modifications is non-decreasing. An ordered Dirichlet distribution is used for the prior over the sequence of reliability values, resulting in Beta marginal distributions. Posterior results are developed along with prior predictive analysis. Prior formulation through expert elicitation and feedback is also discussed in the context of the ordered Dirichlet approach.

2.3.20 Fries Discrete Learning Curve Model (1993)

Fries [51] presents a discrete reliability growth tracking model that utilizes a learning curve approach to model the improvement in reliability over time. Derivations and comparisons of the approach are made, along with those of the model proposed by Crow in [45]. Approximate MLE procedures are presented, along with extensions that separate failures into two categories: those for which the root-cause is understood, and those for which the underlying cause remains unknown. Examples are used to demonstrate the application of the approach.

2.3.21 Fakre-Zakeri and Slud's Mixture Models for Reliability of Software with Imperfect Debugging: Identifiability of Parameters (1995)

Fakre-Zakeri and Slud [52] present a mixture model approach for modeling the reliability of software when bugs are instantaneously detected and removed from the system. The model is developed in a general manner, treating bug failure rates as

realizations from a common mixing distribution. Many of the models in literature are shown to be examples of this structure that vary in the details of the mixture model or assumptions surrounding the correction of the bugs. The number of bugs is assumed to fixed and unknown, but the only observable data during test is the overall result from a superposition of independent counting processes for each of the bugs. Debugging, or corrective actions for each observed bug, is assumed to remove bugs imperfectly or even introduce new bugs into the system with a certain probability *p*.

The paper discusses results from a probabilistic viewpoint, paying particular attention to the identifiability of the parameters of the potential probabilistic models. It is shown that the mixture model assumption allows for identifiable parameters, including the probability of introducing a new bug, when the mixture model is the widely used Gamma-Exponential type. No statistical estimators or goodness-of-fit approaches are given, but references to many of the commonly used specific mixture model forms are given.

2.3.22 AMSAA Subsystem Tracking Model (1996)

The AMSAA Subsystem Tracking Model (SSTRACK) [6] is a generalization of Crow's reliability growth tracking model [14] that explicitly handles data from multiple subsystems. The same underlying Power Law NHPP assumption is used for each of the subsystems. The subsystems are assumed to be serial in structure, and the results are combined using the Lindstrom-Madden method [33] for combining data. Confidence intervals and goodness-of-fit procedures are provided along with statistical estimates of the model parameters.

2.3.23 Erklani, Mazzuchi, and Soyer's Bayesian Computations for a Class of Reliability Growth Models (1998)

Erklani, Mazzuchi, and Soyer [53] present a Bayesian approach for a broad class of reliability growth models, including both continuous and discrete approaches. The model first proposed by Mazzuchi and Soyer [50] is presented, along with an extension of this approach to exponential and Weibull data. The reliability changes are assumed to non-decreasing for the discrete version of the model, and the continuous data case is shown to be a special case of this assumption by reparameterizing the likelihood. This allows for the use of the same ordered Dirichlet prior distribution for each case.

Development of the priors through expert elicitation is discussed, and difficulties in the various posterior calculations are also noted. These difficulties provide motivation for the use of Markov Chain Monte Carlo methods for posterior analysis, and Gibbs sampling is demonstrated for each of the proposed approaches. Numerical examples are given along with comparisons between each of the potential model approaches.

2.3.24 Walls and Quigley's Building Prior Distributions for Bayesian Reliability Growth Modeling (2001)

Walls and Quigley [54] discuss an approach for eliciting expert opinion to develop appropriate prior distributions to support reliability growth modeling. Suggestions are given for eliminating possible bias in the elicitation, along with practical guidance for using the approach within a reliability growth program. The

elicitation process consists of a five-stage approach. The stages span the entirety of the process from development of data collection methods to mathematical aggregation of the results in the reliability growth analysis within a feedback loop. Specific guidance is also provided for three different roles within the process: the assessor who collects and analyzes the reliability growth data, the management expert who will make decisions involving the reliability program, and the technical expert who can provide judgment on the various aspects of the system.

2.3.25 Yu, Tian, and Tang's Bayesian Predictive Analyses for Nonhomogeneous Poisson Processes with Power Law (2007)

Yu, Tian, and Tang [55] present a Bayesian approach to modeling reliability growth using the Power Law NHPP. Their work focuses on the predictive capability of the model, rather than the statistical inference itself. The motivation cited is the increasing demand for expensive and highly reliable systems in the commercial and military industries. This paper is in many ways a Bayesian analogue to the work first completed by Engelhart and Bain [40].

Results are developed for both a single system under test and two systems under test. Posterior and predictive distributions are also developed for two cases, with focus on the probability of observing a future number of failures during a test period. The first assumes the shape parameter of the Power Law NHPP is known, and the second assumes both parameters are unknown. Examples are given in order to compare the performance of the model assumptions. The probability of observing at most k failures is shown to converge for relatively small values of k for each assumed treatment of the shape parameter.

2.3.26 Li, Chang, and Chen's Building Reliability Growth Models Using Sequential Experiments (2010)

Li, Chang, and Chen [56] present a sequential Bayesian approach to model the reliability improvement of a system over a series of reliability growth tests. Failure times are treated as Weibull random variables, with the shape parameter $0 < \beta < 1$to match the Weibull Process NHPP under reliability growth conditions. A reduced bias adjustment procedure is used to sequentially develop empirical Bayes estimates of the parameters of the Weibull likelihood for the first test phase, and the posterior results are used for the prior distribution in successive phases of testing. Comparisons are made with Crow's reliability growth tracking model [14], and the proposed approach is shown to provide results that are as good as Crow's model while using much smaller datasets.

2.3.27 Xing, Wu, Jiang, and Liu's Dynamic Bayesian Evaluation Method for System Reliability Growth Based on In-Time Correction (2010)

Xing, Wu, Jiang, and Liu [57] present a discrete reliability growth approach that allows for important information from the development process to be incorporated into the model. As in other papers, the motivation for the model is the development of highly reliable systems under difficult fiscal and time constraints. The Power Law process for discrete reliability growth developed by Crow in [45] is used again here to describe the growth process, and results are derived for both trial-by-trial and configuration-by-configuration development.

A Bayesian updating process is developed, using maximum entropy to develop the prior distribution on the system reliability by assuming that the first two moments are known. A Beta distribution is then used to approximate the prior and allow for simple conjugate updating through the underlying Beta-Binomial process.

2.3.28 Quigley and Walls Reliability Inference Mixing Bayes and Empirical Bayes (2011)

Quigley and Walls [58] present an approach to assessing reliability by combining Bayes and empirical Bayes methods. The system is assumed to be undergoing design changes due to identified failure modes and possible new innovation. Relevant operational data is assumed to exist as well, such as in organizations that develop families of similar products. The method assumes that the system is sufficiently complex, in that it contains a large number of potential engineering concerns, or failure modes. The system can also be divided into mutually exclusive and conditionally independent concerns. System failure occurs when the first concern is realized during test, indicating a serial relationship between the modes.

Subjective Bayes techniques are used to construct prior distributions for the number of concerns in the system, and empirical Bayes is used to develop the distribution for the time to occurrence of a specific concern. The empirical Bayes approach pools the failure data across multiple classes of concerns to assess the covariance structure between the classes of concerns. A unique distribution for each class is then provided to use in the system reliability calculation. An aerospace design example is provided to illustrate the approach.

2.3.29 Bichon, McFarland, and Mahadevan's Surrogate Models for Reliability Analysis with Multiple Failure Modes (2011)

Bichon, McFarland, and Mahadevan [59] present a method for computing the reliability of a system with multiple failure modes through a surrogate model. The approach considers the overall reliability of a system with multiple failure modes in either a series or parallel configuration. A response function is also assumed for each failure mode, where the mode is considered to have occurred when the response value surpasses a predefined threshold. A Gaussian process regression model is then used as a surrogate in a global reliability analysis approach that considers the overall behavior of the various response functions for the failure modes.

2.3.30 Strunz and Herrmann's Planning, Tracking, and Projecting Reliability Growth: A Bayesian Approach (2012)

Strunz and Herrmann [60] present a unique Bayesian approach to reliability growth tracking using lower level data aggregation for liquid rocket engines. A Beta prior distribution is used for each of the functional nodes in the functional block diagram that describes the system behavior. The number of equivalent trials and failures are developed through consideration of two main failure mechanisms and the relationship between the testing profiles and mission profiles. Acceleration factors are also employed to increase the accuracy of the prior distribution.

Markov Chain Monte Carlo techniques are used to develop the posteriors for each node, subsystem, and system defined in the block diagram. The technique is

demonstrated on a liquid rocket reliability test program that includes a contractual reliability requirement.

2.3.31 Pievatolo, Ruggeri, and Soyer's Bayesian Hidden Markov Model for Imperfect Debugging (2012)

Pievatolo, Ruggeri, and Soyer [61] consider software failures within an imperfect debugging process. The introduction of bugs into the software is considered to be a latent or unobservable process, which allows for the use of latent variables within a hidden Markov model. The states within the model are the bug rates of occurrence at specified times in the testing.

A Bayesian approach is used, where each row of the state transition matrix follows a Dirichlet distribution and the bug rates of occurrence follow Gamma prior distributions. The initial states are also assumed to follow a uniform prior distribution, meaning that each is equally likely a priori. The joint posterior of the states, bug occurrence rates, and transition probabilities is calculated through Gibbs sampling. An approach for estimating the unknown number of states is also provided utilizing the marginal likelihood is also presented.

2.3.32 H. Okamura, et. al's Software Reliability Growth Models with Normal Failure Time Distributions (2013)

Okamura, Dohi, and Osaki [62] present a model for software reliability growth that models failure times using the Normal distribution. The authors discuss the problem of modeling software reliability growth with the large number of models that are available in the literature. The model selection process is reduced to one of

choosing the appropriate failure time distribution from a number of well known statistical candidates, of which the Normal distribution has not been previously considered. The truncated Normal and the LogNormal are both developed for software failure times within the reliability growth model, and the Expectation maximization algorithm is demonstrated for estimating the parameters for each case. Both individual failure data and grouped data are presented, and a numerical example is provided to demonstrate the approach.

2.3.33 Wang, et. al.'s Discrete Nonhomogeneous Poisson Process Software Reliability Growth Models (2013)

Wang, Wu, Lu, and Li [63] present an approach to modeling software reliability growth by considering test coverage along with the failures that occur in the process. The testing is considered as a set of discrete test cases, and the coverage assumes that test units are independent with execution probability p for a single test unit. The Beta distribution is used to consider the variability on p, and maximum likelihood methods are presented to estimate the parameters of the distribution.

The reliability growth model is then developed by assuming that the number of faults detected is proportional to the test coverage obtained for each test case. A more restrictive assumption in the initial model is that the debugging process for a fault is perfect. This assumption is relaxed in an extended version of the model that is also developed, and maximum likelihood estimates are developed for both approaches. Numerical examples demonstrating the capability of the approach are also presented.

2.4 Reliability Demonstration Methods

2.4.1 Yadav, et. al.'s Reliability Demonstration Test Planning (2006)

Yadav, Singh, and Goel [64] present an approach to demonstrating the reliability of a new system. The approach is a systematic method for identifying the reliability critical elements within the system, and then deciding the appropriate reliability demonstration methods to be used for each element. The system is broken into 3 dimensions: physical, functional, and time. The system level reliability is then allocated into each of the critical elements in the system. The items are ranked via the criticality from the FMEA or previously available warranty or field data. A 6-step approach is then identified that takes the developer through the process of decomposition, allocation, prior information collection, testing, and assessment. The approach considers the last stages as iterative, and they are to be repeated along with any necessary design modifications until the reliability has been sufficiently demonstrated. The assessment itself is Bayesian, and the Gamma distribution is used as the prior for the rate parameter of the Weibull. Continuous updating of the assessed reliability is discussed, and the approach advocates the use of all prior information, including previous test data and other qualitative information.

2.4.2 Fan and Chang's Bayesian Zero Failure Reliability Demonstration Test of High Quality Electro-explosive Devices (2009)

Fan and Chang [65] present a Bayesian reliability demonstration test for electro-explosive devices. The devices are assumed to be highly reliable, meaning that reliability demonstration using standard test methods may not be feasible. The

lifetime of the product is assumed to follow the Exponential distribution, and the failure rate is related to temperature through an acceleration factor given by

$$\lambda_w = \alpha_0 e^{\alpha_1 w}, \tag{8}$$

where w is the temperature and the α values are unknown parameters.

The statistical risks for the test are also presented to provide the ability to plan for a test that is likely to be passed for a specified temperature. A simulation study is also presented to examine the posterior assurance associated with the approach. The sensitivity of the method with respect to the choice of the prior distribution is also discussed, and the statistical risks are shown to be insensitive to the choice of the prior.

2.4.3 Guo and Liao's Methods of Reliability Demonstration (2012)

Guo and Liao [66] present a discussion of reliability demonstration testing procedures. They compare the binomial test based on the number of failures observed during the demonstration test with the failure time approach using the distribution of the time to failure of the product under test. The binomial test is assumed to be for one-shot devices, or for those systems whose mission length is equal to the length of the demonstration test. A method for conversion of the requirement to an equivalent one for a different test length is also provided.

The failure time test approach is presented, with the Weibull distribution used to model the time to failure. A comparison of the two test approaches is made, and the equivalence between the two methods is presented. The failure time approach assumes that the β parameter is given prior to the test. The results from each method

are shown for different test conditions, with the demonstrated reliability and corresponding median rate provided.

2.4.4 Elsayed's Overview of Reliability Testing (2012)

Elsayed [67] provides a general overview of a number reliability tests. Eight different types of reliability testing are discussed, along with reliability estimation and prediction methods. The tests presented are: highly accelerated life testing, reliability growth testing, highly accelerated stress screening, reliability demonstration testing, reliability acceptance testing, burn-in testing, built-in-self testing, and accelerated life testing. An overview of various estimation techniques is also provided, each of which could applied within the context of multiple reliability tests.

The paper concludes by discussing the design of accelerated life test plans. Mechanical, electrical, and environmental stresses are considered, and the application of the stressors within the test is also discussed. The remaining consideration involves the proportion of units that are to be subjected to the chosen stress levels. The design of the test plan is presented in the context of an optimization problem involving minimization of the variance associated with the problem.

2.4.5 Crow's Demonstrating Reliability Growth Requirements with Confidence (2012)

Crow [68] presents a method for combining developmental and operational test data in the context of demonstrating a reliability growth requirement. The approach follows that of Miller [69], which is based on the MIL-HDBK 189 model [14] originally developed by Crow. The technique assumes a reliability growth

program is followed by a fixed configuration reliability demonstration test. The existence of the growth program is used to increase the probability of successfully demonstrating the reliability requirement. The problems that exist with demonstrating a requirement with a lower confidence bound in a fixed configuration test are discussed, and the high design goals that result from this approach are cited as problematic for developers.

The approach provides a means for combining developmental reliability growth data with data from the demonstration test. Lower confidence bounds are provided along with operations characteristic curve results for the combined estimator. The results are also shown to reduce program risk and cost, and a numerical example is provided.

2.4.6 Cotroneo et. al.'s Combining Operational and Debug Testing for Improving Reliability (2013)

Cotroneo, Pietrantuono, and Russo [70] present an approach for combining debug testing with OT for software systems. The method is presented as a means of overcoming the limitations associated with OT of highly reliable systems. Various problems with OT are discussed, including lack of knowledge of the true operational profile of the system and the effectiveness of the operational test in understanding reliability.

Debug testing is assumed to stress the system under exceptional and unexpected inputs, providing a contrast to OT that stresses the system with representative operational profiles. The failure region domain is defined, and probabilities of observing a failure region within the operational profile are also

specified. The overall probability of system failure in OT is then defined using a Bernoulli distribution for each execution of operational test inputs. The combination of the two test methods is used to increase the likelihood of observing and correcting high and low occurrence failures in the system. An expression for the probability that the combined test strategy is better than OT alone is derived, along with associated confidence bounds on the probability. The approach is examined through simulated test cases, each of which focused on the operational profile, the number of test cases, and the relative amount of debug versus operational test cases. The paper concludes with a case study demonstrating the increased performance of the combined method on a space system.

2.4.7 Hill, et. al.'s Acquisition and Testing, DT/OT Testing (2013)

Hill, Gutman, Chambal, and Kitchen [71] discuss developmental and OT within the acquisition process in the U.S. Department of Defense. Different approaches for assessing test results via statistical estimators are discussed, along with the underlying assumptions associated with each. The assumption of a "good" or "bad" system is presented, along with reasons and relevant scenarios for each approach.

The paper presents an approach that uses both of the assumptions simultaneously by allowing for two separate requirements. Objective and Threshold parameters are defined, where the system is to be designed to the objective in DT and tested to the threshold in OT. The system is assumed to be good during verification of design to the objective in DT. During OT the system assumed to be bad, and the minimal acceptable performance of threshold is then verified with high confidence.

The approach is very similar to current reliability growth approaches within the Defense Department, where the system reliability is assessed with 2-sided confidence intervals in DT and assessed with a lower confidence bound during OT.

2.5 System Reliability Assessment

2.5.1 Hamada et. al.'s Bayesian Approach for Combining Multilevel Failure Information in Fault Trees (2004)

Hamada, et. al. [72] provide a method for combining corresponding to different levels within a fault tree for a system. The data are assumed to be discrete, and the use of the combined data provides an increase in accuracy and precision of the reliability estimates, thus reducing the aggregation error that can exist for such a problem.

Basic, top, and intermediate events in the fault tree are considered. Prior information about the basic events in the fault tree is described by a Beta distribution, and the data for such events are assumed to follow a Binomial likelihood. Equivalent trials and failures are calculated for the prior information on intermediate and top-level events, and the resulting posteriors can be seen to provide information on the basic events through the fault tree structure. The relative strength or weakness of the data for each event is also considered through the coefficient of variation assigned to each. Weak data have a fairly large coefficient of variation, while strong data have a relatively small value. The approach also discusses data collection within the context of reducing the variation in the posterior distribution. The method is developed as a means of defining optimal testing and data collection for the system.

2.5.2 Reese, Hamada, and Robinson's Assessing System Reliability by Combining Data from Different Test Modalities (2005)

Reese, Hamada, and Robinson [73] present an approach to combining discrete data from different test modalities to assess the overall reliability of a two component parallel system. The approach considers the combination of four types of data: test data from specially produced samples, nondestructive testing, laboratory testing, and flight testing. The nondestructive tests are performed at the component level, while the remaining tests are performed at the system level. The fidelity of the data sources is also considered, where fidelity is defined as the amount of agreement with field usage conditions.

A logistic regression approach is used to relate the different test modalities, where the desired component probability of failure, p, from flight testing is defined by

$$\log\left(\frac{p}{1-p}\right) = \mu. \tag{9}$$

The probability of failure values, p_i, from the additional test modalities are then related to the desired value via the regression model

$$\log\left(\frac{p_i}{1-p_i}\right) = \mu + \lambda_i \ , \ i \ = \ 1, \ 2, \ 3. \tag{10}$$

The l_i values act as explanatory variables in the regression model, and they are related through the assumption that they follow a common Gamma distribution. A hierarchical model is then developed where the m parameter is assigned a Normal

prior distribution and the parameters of the Gamma have additional hierarchical Gamma priors. Markov Chain Monte Carlo procedures are then used to solve for the desired posterior distribution of the system failure probability.

2.5.3 Groen and Droguett's Competing Failure Mode Modeling (2005)

Groen and Droguett [74] provide a method that considers multiple failure modes within a system. The approach is a Bayesian framework for assessing time to failure distributions for products. Each product is defined as a combination of operating environment and system design or configuration, and the probabilistic nature of the Bayesian framework allows for combination of data across products. The combination methods use adjustment factors to account for differences in the operating environment or changes to the design. The factors can be treated as deterministic quantities, or uncertainty distributions can be used to characterize the variability associated with each factor. The use of factors allows for data from similar products to be used in the reliability assessment, and while not explicitly stated, it could also allow for reliability growth to be considered.

A major assumption of the approach is that the product is comprised of different failure modes, and the analysis can be performed using data at the failure mode level or product level. There is no assumed inherent connection between the failure modes other than the link that is provided by the system/product level reliability. Weibull likelihoods are used to characterize the time to failure for specific failure modes, and the aging behavior of the mode is assumed to be constant between products with the same mode. The failure rates for each of the modes are then connected serially within the system, allowing for a straightforward summation to

obtain the system level assessment. Markov Chain Monte Carlo methods are used to obtain the posterior distribution of the failure rate for the system. A numerical example considering 2 failure modes is provided. An earlier version of a similar approach is also presented by Groen, et. al. in [75].

2.5.4 Wilson, Graves, Hamada, and Reese's Advanced in Data Combination, Analysis, and Collection for System Reliability Assessment (2006)

Wilson, et. al. [76] provides an overview of various data combination methods for both component and system reliability assessment that exist in the literature. Two examples of component reliability assessment are given. The first involves degradation and failure time data, and the second uses pass-fail Bernoulli and quality data to estimate the reliability of a component.

System reliability is assessed through a combination of life data and degradation data for components within the system. A logistic regression approach similar to that in [73] is used to transform binary test data for a component into the continuous domain, which then allows for a combined reliability function that includes additional Weibull lifetimes and degradation data from a Lognormal process. A Nonhomogeneous Poisson Process model for clustered supercomputer failures is also developed, along with an approach to combining component life data within a fault tree structure.

The review discusses the fully Bayesian implementation of each example, with complex hierarchical models developed and solved through Markov Chain Monte Carlo methods. The paper ends by discussing the various methods for

representing the system, to include Bayesian networks and flowgraph models. Optimal resource allocation in the spirit of [72] is also discussed.

2.5.5 Yadav, Choudhary, and Bilen's Complex System Reliability Estimation Methodology in the Absence of Failure Data (2008)

Yadav, Choudary, and Bilen [77] present a methodology for estimating system reliability in cases where failure data are initially unavailable. The approach is comprehensive in nature, including physics-based fundamentals to establish the mathematical model of the system. Uncertainty is applied at all levels of the system. The approach involves establishing a transfer function to describe the acceptable performance of the system, and the overall system reliability is calculated by computing the ratio of the number of trials that fall within the acceptable range to the total number of trials.

Environmental factors are considered within the model through the reliability estimate adjustor factor, which considers the difference between the initial failure rate and the true failure when the environment is considered. The uncertainty on the variables within the model is accounted for through the use of fuzzy logic, and the initial reliability estimate is then updated through standard Bayesian model updating. A case demonstrating the approach for a hazardous gas detection system is also presented.

2.5.6 Pan's Reliability Prediction Using Accelerated Life Data and Field Failures (2009)

Pan [78] presents a Bayesian approach to system reliability assessment that integrates field failure data and accelerated life test data. The accelerated life test results are assumed to not adequately represent the field failure of the system because of variation in the field usage that exists. A conjugate prior framework is utilized to simplify the process, and a calibration factor is used to account for the differences in the accelerated test and the field conditions. Uncertainty on both the failure rate and the calibration factor is considered, and the posterior is updated using the accelerated life test results via Markov Chain Monte Carlo methods. The calibration factor is not treated as a nuisance parameter, and the joint posterior for both model parameters is developed. An example application involving temperature accelerated testing of an electronic device is presented.

2.5.7 Wilson, Anderson-Cook, and Huzurbazar's Case Study for Quantifying System Reliability and Uncertainty (2011)

Wilson, Anderson-Cook, and Huzurbazar [79] present a case study for system reliability assessment that quantifies the uncertainty present in the estimate. The approach is motivated by the limitations in obtaining full system level test results due to cost, practicality, or technical permissibility. Subsystem and component data are used to develop the system level reliability estimate, and the uncertainty from each of the data sources is aggregated to provide uncertainty at the system level. A reliability block diagram is used to define the system reliability structure.

Discrete test data, continuous test data, and multi-level data are all demonstrated within the proposed framework. Components with no available test data are used in the approach. Markov Chain Monte Carlo methods are used to evaluate the system level posterior distribution, where the system reliability is obtained through the structure defined by the reliability block diagram. The approach provides a proof-of-concept that can be adapted to other more complex systems while still accounting for the appropriate uncertainty that exists in the assessment.

2.6 Conclusion

This chapter reviewed a number of reliability growth models and reliability assessment techniques that are found in the literature. Various models for reliability growth planning, tracking, and projection have been presented in order to define the current state-of-the-art in reliability growth. Many modeling approaches are available for both continuous and discrete systems, with a number of classical and Bayesian approaches available. A number of more general methods for system reliability assessment were also reviewed.

From reviewing the literature, there is a general lack of reliability growth approaches that consider data from throughout the developmental program of the system. This is particularly true for reliability growth projection models, which are the current preference in the U.S. Department of Defense. The models apply to single test phases only, with no way of updating results from test phase to test phase. The models involving arbitrary corrective action strategies can also be improved, as the current state-of-art involves using only the first observed time or trial of occurrence for a given failure mode.

There are also limited approaches for combining data from different types of testing within a reliability growth program. Current approaches for data combination from different test modalities involve various types of testing in combination with reliability demonstration or OT, but there are limited options that consider reliability growth testing combined with operational reliability demonstration testing. The failure mode based options currently available in the literature also consider only finite numbers of known failure modes in the system, with no allowance or accounting for unobserved failure modes. These approaches are also developed specifically for time-to-failure distributions, with no extensions to consider reliability for complex repairable systems.

Finally, there is a decided disconnect between the current reliability growth approaches in the literature and the reliability assessment methods involving reliability engineering efforts. The use of component and subsystem data for system reliability assessment occurs in various papers, but none of these discuss an approach that connects the results to reliability growth modeling approaches. The use of physics-based results within these approaches is also limited.

- CHAPTER 3 -

3 A BAYESIAN MODEL FOR COMPLEX SYSTEM RELIABILITY GROWTH UNDER ARBITRARY CORRECTIVE ACTIONS[1]

3.1 Introduction

3.1.1 Background

The work presented in this chapter differs from the reliability growth projection models reviewed in Chapter 2, as it is a Bayesian reliability growth projection model that explicitly applies FEFs. The model assumes the same piecewise-Exponential probabilistic process modeled by Ellner and Wald's AMSAA [18] while also providing a complete inference framework via the posterior distribution on the system failure intensity. A Bayesian posterior distribution is first developed for each failure mode in the system, allowing for the application of FEFs as appropriate. The likelihood is also generalized to allow for arbitrary corrective actions to occur at any point both during and after the test period. An additional extension that is unique amongst the class of Bayesian reliability growth models is the ability to estimate the posterior distribution of the failure intensity due to failure modes that have not yet been observed, including the important metric of growth potential failure intensity. The model provides an analytic expression that is a useful metric for reliability program managers, as it indicates the relative contribution of any remaining failure modes in the system along with the theoretical lower bound on the failure intensity that can be achieved via the test-fix-test reliability growth paradigm. Goodness-of-fit procedures to assess model appropriateness are a vital part of any

[1] Material from Chapter 3 has been submitted for publication to IEEE Transactions on Reliability.

reliability growth model, and two methods are presented here. The first is a visual method involving the number of observed failure modes, and the second is a more formal test for goodness of fit. An additional feature unique to the proposed model is the ease with which extensions can be made. Extensions of the basic framework are presented to allow for both multiple systems under test and uncertainty on the assessed FEF values.

Use of a model such as that proposed here can be extremely informative to reliability program managers and engineers. This is particularly true for programs that are early in the development process, as the results from early reliability growth testing can be used to shape future resource investment for the program. They can also provide an indication when further reliability growth testing may provide diminishing returns, thus leading to a more efficient developmental reliability program.

3.1.2 Chapter Overview

This chapter is organized as follows. Section 3.2 develops the model framework, including model assumptions, data requirements, and a useful management metric. Section 3.3 discusses prior development while presenting empirical Bayes estimators, and Section 3.4 provides two procedures for assessing the goodness of fit of the model. Section 3.5 develops extensions to the basic model to handle multiple systems under test and uncertainty associated with the assessed FEF values. Simulation performance comparisons of the basic model are presented in Section 3.6, and an example application is presented in Section 3.7. Conclusions are provided in Section 3.8.

3.2 Methodology

3.2.1 Model Assumptions

When projecting the reliability growth of a complex system, it is common to treat the system failure intensity as a sum of failure intensities (or failure probabilities) from independent failure modes. Modeling the system as a combination of failure modes enables the use of FEF values, which mathematically quantify the fractional reduction in the failure intensity for a given failure mode after a corrective action has been implemented. The assumptions for the model are as follows:

1. The system is comprised of a large number of failure modes that are serial in nature; the occurrence of any failure mode results in failure of the system.
2. Failure modes generate failures independently of one another.
3. The failure intensity, or rate of occurrence of failure, for each mode is constant both before and after a corrective action is implemented.
4. The resulting failure intensity after corrective action will be reduced from the initial value according to the assigned FEF.
5. Corrective actions to failure modes do not introduce new failure modes into the system.
6. Reliability testing will stress the system in an operationally relevant manner.

3.2.2 Data Requirements

Reliability growth projection models fall into one of two categories. The first treats all corrective actions as delayed, while the second allows for corrective actions to be either delayed or occur during the test itself. As stated previously, the model developed here falls into the latter category by allowing for arbitrary corrective actions. It differs from previous models in this category [18],[20], as it uses the cumulative times of failure for each failure mode, along with the time of any associated corrective actions and their assessed FEFs to estimate the failure intensity of the system. We stress the importance of assigning realistic FEF values for failure mode corrective actions. The projected improvement in reliability will be erroneous if the assigned FEF values are unrealistic. Proper consideration of the root cause of the failure mode and the type of corrective action is essential to assignment of a realistic FEF value. An additional utility of reliability growth projection models of this type involves sensitivity analysis on the FEF values. If known reliability targets are desired after corrective actions have been implemented, sensitivity analysis on the FEF values can determine the levels that are necessary to achieve the desired targets. Application of the model in this manner can then help to drive resource expenditure during the corrective action process.

Beyond the FEF values, the approach proposed here only requires information that is commonly collected during reliability growth testing. Proper configuration management is vitally important to ensure that reliability estimates are representative of the actual system as the design progresses and matures, and the data requirements for the model fall within a well executed configuration management process.

Assumption 6 is also vitally important, as any estimates of reliability under non-operational conditions should be treated with extreme caution if they are to be used in any programmatic resourcing decisions.

3.2.3 Failure Mode Posterior Distribution

For the i^{th} failure mode in the system, assume that n_i failures are observed in test length T with times $\left(t_{i,1},t_{i,2},...,t_{i,n_i}\right)$. Further assume that the failures are divided such that $n_{i,1}$ occur before a corrective action is implemented, and n_i - $n_{i,1}$ failures occur after the corrective action. For a corrective action that occurs at time v_i, let d_i represent the FEF resulting from the corrective action. From assumptions 2-4 the likelihood can then be expressed as

$$l\left(t_{i,1},t_{i,2},...,t_{i,n_i},n_i,n_{i,1} \mid \lambda_i\right) = \left(1-d_i\right)^{n_i-n_{i,1}} \lambda_i^{\,n_i} \exp\left(-\lambda_i\left[v_i + \left(1-d_i\right)\left(T-v_i\right)\right]\right). \quad (1)$$

Note that the form of the likelihood in (1) does not contain the actual failure times. They are part of a telescoping sum, and are replaced by the counts of failures before and after the corrective action.

For the prior on the individual mode failure intensity, we use the Gamma[α,β] distribution parameterized as

$$p\left(\lambda_i\right) = \frac{\lambda_i^{\,\alpha-1}}{\Gamma\left(\alpha\right)\beta^{\alpha}} \exp\left(-\frac{1}{\beta}\lambda_i\right). \quad (2)$$

The choice of a Gamma prior distribution is not only mathematically convenient; it has also been used commonly in past work to model mode-to-mode variation that

may be present in the system. The failure intensities for the collection of failure modes found in a complex system are shown to be adequately modeled as a random realization from a Gamma distribution in both [18] and [26]. Using a Gamma distribution in this way also recognizes what may be referred to as the "vital few, trivial many" property among mode failure intensities. This property acknowledges that each failure mode provides a different contribution to the overall system failure intensity, with a relatively few number of failure modes being significant enough to be observed in test.

Standard posterior distribution calculations result in the posterior distribution for the failure mode failure intensity given by

$$p\left(\lambda_i \mid t_{i,1}, t_{i,2}, \ldots, t_{i,n_i}, n_i\right) = \frac{\lambda_i^{\alpha+n_i-1}}{\Gamma(\alpha+n_i)\left[\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right]^{-(\alpha+n_i)}} \exp\left[-\lambda_i\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)\right] \tag{3}$$

Note that the $n_{i,1}$ term in (1) only applies to the constant term $(1-d_i)$ and is therefore canceled from the posterior distribution. From (3) the distribution of the failure mode failure intensity is then

$$\lambda_i \sim Gamma\left[\alpha+n_i, \left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)^{-1}\right]. \tag{4}$$

Note that if no corrective action is attempted, d_i is zero and (4) then reduces to the traditional Gamma posterior commonly found in many references that discuss Bayesian statistics. See for example Section 7.4 in [81]. From (4) we can also leverage properties of the Gamma distribution to develop the posterior for the reduced failure intensity after a corrective action has been implemented.

$$(1-d_i)\lambda_i \sim Gamma\left[\alpha + n_i,\ (1-d_i)\left(\frac{1}{\beta} + v_i + (1-d_i)(T-v_i)\right)^{-1}\right] \qquad (5)$$

3.2.4 Complex System Posterior Distribution

We develop the posterior for a complex system by first assuming a finite number of modes, *K*, and then examining the results as *K* becomes large. From assumption 1 the failure intensity after corrective actions for the entire system can be found by summing the individual mode failure intensities. If we denote the number of observed modes as *m* and let $n = (n_1, n_2, \ldots, n_m)$ be the vector of observed failures for each of the observed failure modes, we can represent the mean of the system posterior after corrective actions as

$$E[\lambda_s \mid n] = E\left[\sum_{i=1}^{K}\lambda_i \mid n\right] = \sum_{i=1}^{K}E[\lambda_i \mid n] = \sum_{i=1}^{K}\left(\frac{(1-d_i)(\alpha+n_i)}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T-v_i)\right)}\right)$$

$$= \sum_{i=1}^{m}\left(\frac{(1-d_i)(\alpha+n_i)}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T-v_i)\right)}\right) + (K-m)\left(\frac{\alpha}{\frac{1}{\beta}+T}\right) \qquad (6)$$

It is useful to reparameterize the expression in (6) to aid in computing the limiting form for complex systems. Note that the prior mean for system failure intensity can be expressed as

$$\lambda_B = K\alpha\beta\,. \qquad (7)$$

Reparameterizing the prior Gamma distribution in terms of the prior system mean and the β parameter then allows the result in (6) to be expressed as

$$E[\lambda_s \mid n] = \sum_{i=1}^{m}\left(\frac{(1-d_i)\left(\frac{\lambda_B}{\beta K}+n_i\right)}{\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)}\right)+(K-m)\left(\frac{\frac{\lambda_B}{\beta K}}{\frac{1}{\beta}+T}\right)$$

$$=\sum_{i=1}^{m}\left(\frac{(1-d_i)\left(\frac{\lambda_B}{\beta K}+n_i\right)}{\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)}\right)+\left(\frac{1}{\beta}-\frac{m}{K\beta}\right)\left(\frac{\lambda_B}{\frac{1}{\beta}+T}\right). \qquad (8)$$

The result for a complex system consisting of a large number of failure modes can be examined by taking the limit of (8) as *K* becomes large while holding the prior mean in (7) and the β parameter constant. This yields the expression shown in (9).

$$E[\lambda_s \mid n] = \sum_{i=1}^{m}\left(\frac{(1-d_i)n_i}{\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)}\right)+\frac{\lambda_B}{1+\beta T} \qquad (9)$$

The variance of the system posterior after corrective actions is similarly found to be

$$Var[\lambda_s \mid n] = \sum_{i=1}^{m}\left(\frac{(1-d_i)^2 n_i}{\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)^2}\right)+\frac{\lambda_B}{\beta\left(\frac{1}{\beta}+T\right)^2}. \qquad (10)$$

The results in (9) and (10) can also be used to calculate the incremental reduction in the system failure intensity. The individual FEF for a specific observed failure mode should not be included in the posterior estimate until the corrective action for that

mode has been implemented. This is analogous to using the posterior result in (4) before the corrective action and the result in (5) after the corrective action. Examination of (9) and (10) shows that this is accomplished by setting the d_i in the numerator to 0 before the corrective action has been implemented. We also point out that when no corrective actions are applied during the test, the result in (9) is a Bayesian alternative to the traditional failure intensity point estimate for a HPP.

When all corrective actions occur at the same time with the same level of fix effectiveness, the posterior distribution for the system with a finite number of failure modes will be exactly Gamma. This is not usually the case during reliability growth testing, as most failure modes will have different underlying root causes and different levels of fix effectiveness. For this reason it is necessary to examine the posterior distribution through simulation to understand if the Gamma will be a reasonable approximation under arbitrary corrective action strategies. Because the individual failure mode posteriors are Gamma from (4) and (5), it is straightforward to simulate failure intensities from the individual mode posteriors. A histogram of the posterior results under an arbitrary corrective action strategy is shown in Figure 3.1. The simulated system was comprised of 50 failure modes and tested for 2000 hours. The corrective action strategy allowed for 5% of the observed failure modes to be corrected during the test, while the remaining 95% were corrected when the test was completed. The average FEF value for the corrective actions was 0.7. The dashed line represents a Gamma distribution fit to the data using a standard method-of-moments based approach with parameters given by (11) and (12). The parameters are

developed from the mean and variance in (9) and (10), and the form of the Gamma distribution follows (2).

$$\tilde{\alpha} = \frac{E\left[\lambda_s \mid n\right]^2}{Var\left[\lambda_s \mid n\right]} \tag{11}$$

$$\tilde{\beta} = \frac{Var\left[\lambda_s \mid n\right]}{E\left[\lambda_s \mid n\right]} \tag{12}$$

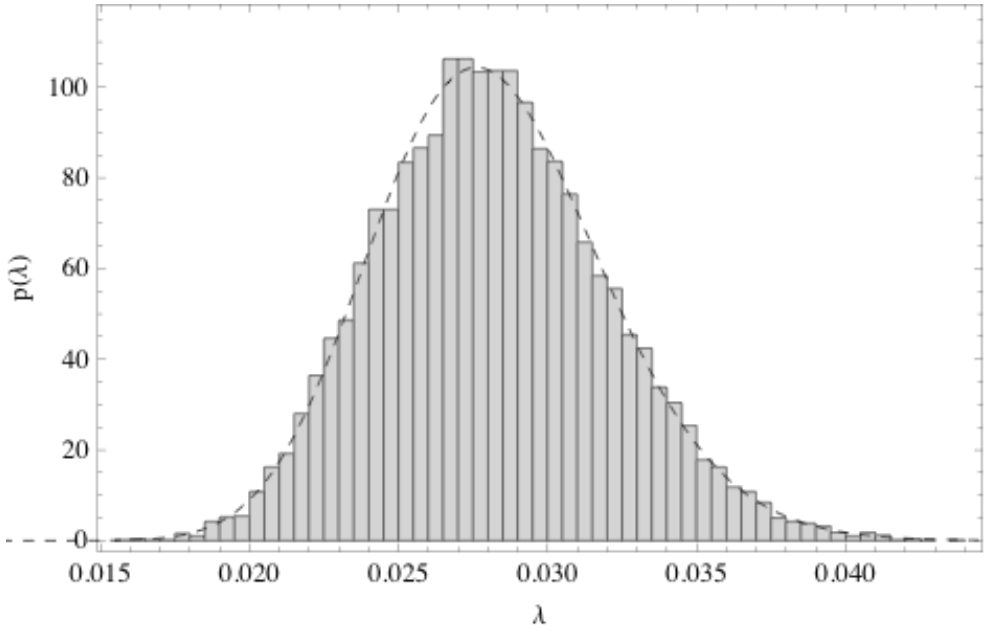


Fig 3.1. System Level Posterior Distribution with Approximate Gamma Distribution Overlaid

The results shown in Figure 3.1 indicate that the Gamma distribution defined by (11) and (12) is indeed a reasonable approximation. The approximate conjugate relationship that is found through (11) and (12) is also beneficial for updating the system level results with data from additional test events. This allows for reliability assessment throughout the system's development to use all of the previously available data, rather than being strictly based on the most current test event. The updating

process acts to decrease the uncertainty associated with the posterior distribution while also accounting for reduced failure intensity resulting from the reliability growth process. Note that configuration changes beyond those involving corrective actions to observed failure modes may cause the conjugate combination to be unrealistic, and any differences in the test events such as test article configuration and environment should always be considered before this technique is practically employed.

3.2.5 Initial Failure Intensity

The initial system failure intensity at the start of the test can be estimated easily using results from the previous section. Using the individual failure mode posterior in (4) instead of (5) provides failure intensity estimates prior to any corrective actions being implemented. This amounts to estimating the initial step in a piecewise Exponential process. Applying (4) in the same manner as previously developed yields the system level posterior mean and variance of the initial failure intensity shown in (13) and (14). Note that the failure intensity for any step in the piecewise Exponential process can be estimated in a similar manner by simply applying the FEF for the failure mode that has been corrected. This is equivalent to using the mode posterior in (5) in place of (4) for the failure mode associated with the corrective action.

$$E\left[\lambda_{s,\,initial} \mid n\right] = \sum_{i=1}^{m}\left(\frac{n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T-v_i)\right)}\right) + \frac{\lambda_B}{1+\beta T} \tag{13}$$

$$Var\left[\lambda_{s,\,initial} \mid n\right] = \sum_{i=1}^{m}\left(\frac{n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T-v_i)\right)^2}\right) + \frac{\lambda_B}{\beta\left(\frac{1}{\beta}+T\right)^2}. \qquad (14)$$

Posterior interval estimates for the initial failure intensity can be found using the same techniques developed in Section 3.2.4.

3.2.6 Growth Potential Failure Intensity

Crow introduced the concept of reliability growth potential after development of the AMSAA Crow Projection Model [82]. The growth potential represents the theoretical upper bound on reliability, and hence lower bound on failure intensity, that can be achieved within the test-fix-test reliability growth paradigm. This theoretical bound can be found by applying the average FEF to the remaining failure intensity due to the unobserved failure modes. It can be estimated using the posterior distribution for the failure intensity in (9) and (10).

The posterior mean growth potential is given by

$$E\left[\lambda_{s,\,GP} \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T-v_i)\right)}\right) + (1-\mu_d)\frac{\lambda_B}{1+\beta T}, \qquad (15)$$

where μ_d is the average FEF that should be expected for the remaining failure modes in the system. The value of μ_d can be estimated by computing the arithmetic average of FEF values for the observed failure modes. The variance is then similarly given by

$$Var\left[\lambda_{s,GP} \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)^2 n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T-v_i)\right)^2}\right) + (1-\mu_d)^2 \frac{\lambda_B}{\beta\left(\frac{1}{\beta}+T\right)^2}. \quad (16)$$

Posterior interval estimates for the growth potential failure intensity can be found using the same techniques developed in Section 3.2.4.

The growth potential failure intensity can indicate whether additional test-fix-test reliability should be pursued for the program in question. If the resulting growth potential failure intensity estimates are higher than the desired objective, resources should be expended for reliability improvement initiatives outside of the traditional test-fix-test approach.

3.2.7 Failure Modes Observed During Follow-on Testing

A particularly useful management metric can also be developed from the posterior distribution for the system: the distribution of the number of failure modes observed in a future time interval. These metrics are helpful for reliability program managers who must make decisions regarding future resource allocation for the program, particularly when planning to observe and mitigate new failure modes in future testing. These metrics can also be used to visually examine the validity of the underlying model assumptions for both the prior and posterior distributions, and this is discussed in more detail in Section 3.4. The posterior expected number of failure modes is found by first defining the indicator function for the i^{th} unobserved failure mode as

$$I_i(t) = \begin{cases} 1, \text{ mode occurs by time } t \\ 0, \text{ otherwise} \end{cases}. \tag{17}$$

The posterior predicted mean of $I_i(t)$ is found by examining the posterior probability that the unobserved failure mode is observed by some time $t > T$. The likelihood of observing a failure mode by time $t > T$ is given from Assumption 3 in Section 3.2.1 as

$$p\left(I_i(t) = 1 \mid \lambda_i, n_i, t > T\right) = 1 - e^{-\lambda_i (t-T)} \tag{18}$$

Using the posterior distribution on the mode failure intensity in (4) with the n_i, v_i, and d_i set to 0 for an unobserved failure mode, the unconditional marginal distribution for $I_i(t)$ can be found by calculating the joint distribution of (4) and (18) and then marginalizing with respect to λ_i. The unconditional expected value is given by

$$E\left[I_i(t) \mid t > T\right] = p\left(I_i(t) = 1 \mid t > T\right) = 1 - \frac{1}{\left(1 + \frac{t-T}{\frac{1}{\beta} + T}\right)^{\alpha}}. \tag{19}$$

Summing over all K - m unobserved modes in the system yields

$$E\left[\sum_{i=m+1}^{K} I_i(t) \mid t > T\right] = (K-m)\left[1 - \frac{1}{\left(1 + \frac{t-T}{\frac{1}{\beta} + T}\right)^{\alpha}}\right] \tag{20}$$

and taking the limit as K becomes large results in

$$\lim_{K\to\infty} E\left[\sum_{i=m+1}^{K} I_i(t) \mid t > T\right] = \frac{\lambda_B}{\beta}\log\left(1+\frac{t-T}{\frac{1}{\beta}+T}\right). \tag{21}$$

Because $I_i(t)$ is a Bernoulli random variable, summing over the unobserved modes yields a Binomial random variable. Taking the limit as K becomes large then results in a Poisson random variable with the mean shown in (21). Figure 3.2 shows a plot of the posterior distribution for the number of modes for 500 hours beyond test time T = 500 for a sample data set. The solid line indicates the posterior mean, while the dashed lines indicate 0.05 and 0.95 percentiles respectively (90% probability interval).

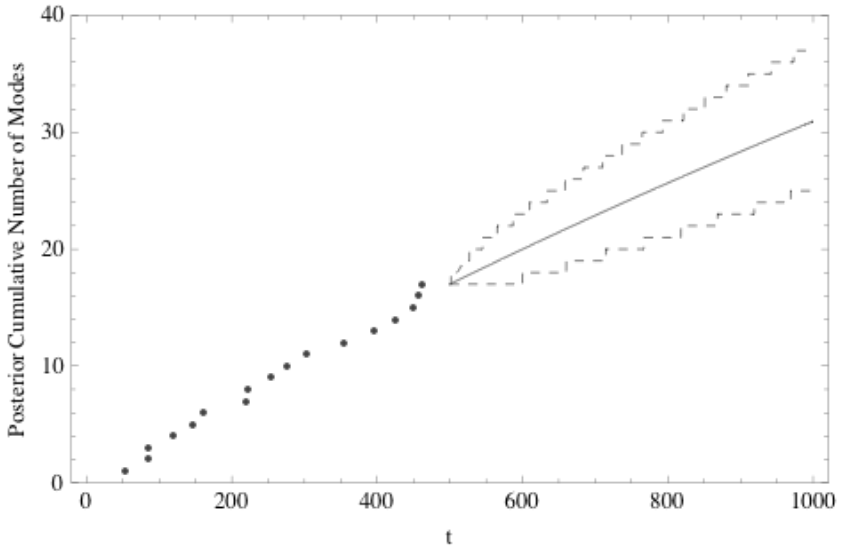


Fig 3.2. Posterior Number of Failure Modes in Follow On Testing

3.3 Empirical Bayes Estimators

We first point out that it is entirely possible to parameterize the Gamma prior through other means such as historical data or elicitation of experts. A five-stage approach for developing priors using expert elicitation is outlined in [80]. Although not fully Bayesian in their application, empirical Bayes estimates are presented here for completeness. These methods present a useful alternative for situations when relevant prior information is not available and excessive subjectivity is undesirable. The posterior probability intervals that result from using this approach will not represent the full uncertainty in the estimation, and this should be considered when making decisions with the model results in this context.

Empirical Bayes estimates are developed in [81] and extended in [26], but the previous results are limited to those cases where there are no corrective actions occurring. The results contained in Section 7.7 in [81] can be generalized to handle arbitrary corrective actions though, noting that the use of the Poisson likelihood will provide the same results as those for the Exponential likelihood. Finite results are developed and then extended to handle the limiting case for complex systems. From the distribution in (4), the resulting maximum likelihood equations for estimating the Gamma parameters are given as

$$\hat{\lambda}_B = \frac{\sum_{i=1}^{m} \frac{n_i}{\frac{1}{\hat{\beta}} + v_i + (1-d_i)(T-v_i)}}{1 - \frac{1}{\hat{\beta} K} \sum_{i=1}^{K} \frac{1}{\frac{1}{\hat{\beta}} + v_i + (1-d_i)(T-v_i)}} \quad \text{and} \quad (22)$$

$$m + \sum_{i \in obs} \sum_{j=1}^{n_i - 1} \frac{1}{1 + \frac{j(K\hat{\beta} - c)}{\sum_{i=1}^{m} \frac{n_i}{\frac{1}{\hat{\beta}} + v_i + (1 - d_i)(T - v_i)}}} = \frac{\left(\sum_{i=1}^{m} \frac{n_i}{\frac{1}{\hat{\beta}} + v_i + (1 - d_i)(T - v_i)} \right) \left(\sum_{i=1}^{K} \log\left[1 + \hat{\beta}(v_i + (1 - d_i)(T - v_i))\right] \right)}{K\hat{\beta} - c} \tag{23}$$

where c is defined as

$$c = \sum_{i=1}^{K} \frac{1}{\frac{1}{\hat{\beta}} + v_i + (1 - d_i)(T - v_i)}. \tag{24}$$

Results analogous to those in (22) and (23) can be found by taking the limit with respect to K.

$$\hat{\lambda}_B = \frac{1 + \hat{\beta}T}{\hat{\beta}T} \sum_{i=1}^{m} \frac{n_i}{\frac{1}{\hat{\beta}} + v_i + (1 - d_i)(T - v_i)} \tag{25}$$

$$m = \frac{1 + \hat{\beta}T}{\hat{\beta}^2 T} \log\left[1 + \hat{\beta}T\right] \left(\sum_{i=1}^{m} \frac{n_i}{\frac{1}{\hat{\beta}} + v_i + (1 - d_i)(T - v_i)} \right) \tag{26}$$

Setting the v_i equal to T in (25) and (26) will provide results equivalent to those in [26]. The equation in (26) will also have a unique solution iff $\sum_{i=1}^{m} n_i > m$.

3.4 Model Assessment and Goodness of Fit

For assessing the utility of the model and appropriateness of the underlying assumptions, we present two approaches. The first uses the distribution for the observed number of failure modes to provide a visual indication of the validity of the

model assumptions. The second is a more formal Bayesian Chi-Squared Test for goodness of fit.

3.4.1 Prior Predicted Cumulative Number of Failure Modes

The distribution for the prior number of failure modes that should be observed by some time *t* can be found similarly to the posterior results shown in Section 3.2.6. By considering the prior distribution in place of the posterior, the expected value for a single failure mode in (19) is modified slightly to be

$$E\left[I_i(t)\right] = p\left(I_i(t) = 1 \mid t \le T\right) = 1 - \frac{1}{(1+\beta t)^{\alpha}}. \tag{27}$$

Note that in this context *t* is any time in the interval (0,*T*). Summing over all failure modes and taking the limit as *K* becomes large yields

$$\lim_{K\to\infty} E\left[\sum_{i=1}^{K} I_i(t)\right] = \frac{\lambda_B}{\beta}\log(1+\beta t). \tag{28}$$

The distributional results from Section 3.2.7 will also apply in this case, and the distribution on the prior number of failure modes will be Poisson with the mean given by (28). An example plot showing the cumulative observed number of failure modes and the prior predicted distribution as a function of time *t* is shown in Figure 3.3. The solid line represents the mean cumulative number of failure modes, and the dashed lines represent the 5^{th} and 95^{th} percentiles of the prior predicted distribution at time *t* respectively. The points are the cumulative number of observed failure modes by time *t*.

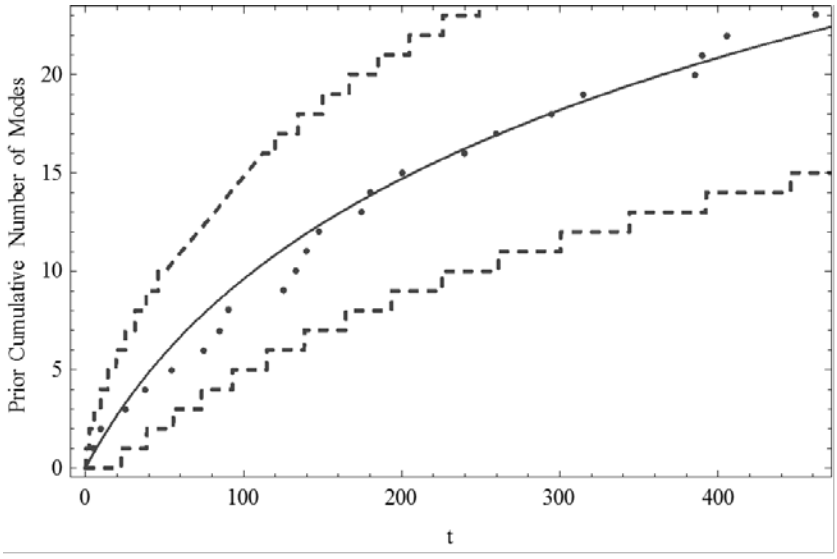


Fig 3.3. Comparison Plot of Observed Failure Modes and Prior Predicted Cumulative Number of Failure Modes Distribution vs Test Time *t*

Although it is not a formal goodness of fit test, using a plot like Figure 3.3 can be useful for indicating any differences between the observed data and the model assumptions. The results in (27) and (28) use both the prior and the likelihood, so reasonable agreement between the distribution and the observed data should indicate that the model assumptions are reasonable.

3.4.2 Bayesian Chi-Square Test

The Bayesian Chi-Square test developed by Johnson in [83] can be used for additional assessment of the model assumptions. The Chi-Square test statistic requires the data to be conditionally independent observations from a probability density function. Assumption 3 in Section 3.2.1 states that the time between successive failures will be piecewise-Exponentially distributed with a rate that depends on the application of specific corrective actions. This allows us to treat the

inter-arrival times between the failures as an observed sequence of Exponential random variables, where the rate parameter of the distribution will be reduced by the appropriate FEF after each successive corrective action. The failure free periods prior to corrective actions, and at the ends of time truncated tests, can be handled using the appropriate Exponential distribution and randomization procedures as defined in [84]. The addition of the failure free periods result in a sample size of *n*, where *n* is one plus the sum of the number of failures and the number of corrective actions that occur during the test.

Following the method outlined in [83], define $\tilde{\theta}$ as a sampled value from the posterior distribution, which is the Gamma distribution developed in Section 3.2. As described in Section 3.2, the mean and variance defined in (9) and (10) will need to be modified accordingly to account for the timing of any corrective actions that occur during the test phase. The required distributions can be found using the same approach discussed in Section 3.2 for estimating each step of the piecewise Exponential process. This is again accomplished by setting the FEF in the numerator to zero until after the corrective action has been implemented.

The unit interval is then divided into q sub-intervals such that $0 \equiv a_0 < a_1 < ... < a_{q-1} < a_q \equiv 1$. Next let y_i be the times between successive failures and F_j be the corresponding Exponential distributions with the appropriate failure rate parameters. Define $z_j(\tilde{\theta})$ to be a vector of length q whose q^{th} j^{th} element is zero unless

$$F_j(y_j \mid \tilde{\theta}) \in (a_{q-1}, a_q). \tag{29}$$

When the condition in (29) holds, the q^{th} element of z_j is set to one. When applying the randomization procedures defined in [84] to the failure free intervals, a random number is drawn from the interval $\left[1 - F_j\left(T - y_j \mid \tilde{\theta}\right), 1\right]$ in place of the F_j in (29). Note that the time prior to corrective actions is handled in the same manner, with T replaced by the corrective action time v. Next define

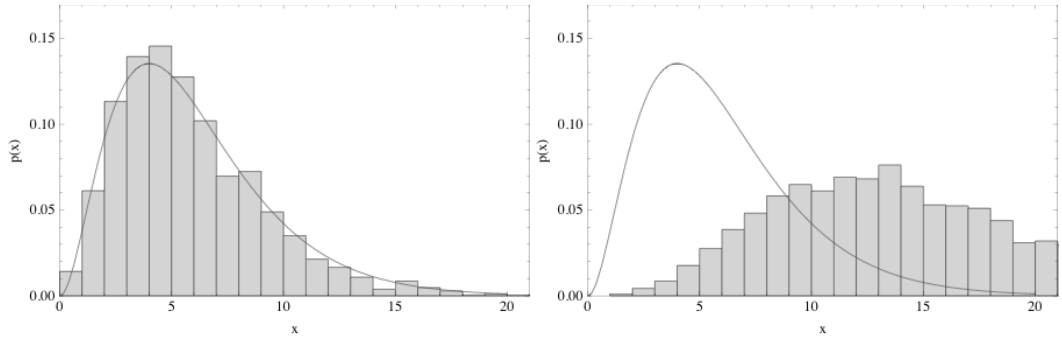
$$m\left(\tilde{\theta}\right) = \sum_{j=1}^{n} z_j\left(\tilde{\theta}\right) \tag{30}$$

as a vector of length q, where each element of the vector represents the total number of points in the corresponding subinterval of the unit interval. The Chi-Square statistic from [83] is defined as

$$R^B\left(\tilde{\theta}\right) = \sum_{k=1}^{q}\left[\frac{m_k\left(\tilde{\theta}\right) - np_k}{\sqrt{np_k}}\right]^2, \tag{31}$$

where the p_k are defined as the differences between the a_q values that divide the unit interval. The statistic in (31) is asymptotically Chi-Squared with q-1 degrees of freedom under repeated sampling of both the data and the posterior parameter of interest. The data cannot be repeatedly sampled in practice, but the utility of the statistic is that poorly specified models will not follow the proper Chi-Squared Distribution and will likely have a higher number of samples in the tail of the distribution. Figure 3.4 shows an example of this concept for the Piecewise-Exponential model developed above. Figure 3.4(a) contains 2000 samples of failure data using the mean of the prior distribution as the true value of the failure intensity, while Figure 3.4(b) samples the same failure data with a prior mean that is a factor of

ten less than the true failure intensity. The data in both cases were simulated using a single corrective action that occurred midway through the test. The solid line shows the reference Chi-Squared Distribution, which has six degrees of freedom for this example. The differences can be seen immediately, and the poorly specified (prior) model shows an extremely poor fit to the reference distribution.



(a) (b)

Fig 3.4. Chi-Squared Distribution Example for (a) Properly specified prior distribution, (b) Misspecified prior distribution

To assess the appropriateness of the model in practical cases, repeated sampling of the data is generally not practical. Random samples can instead be drawn from the posterior distribution while applying the appropriate FEF values for any corrective actions that have occurred during the test. The statistic in (31) can be calculated for each of these samples, and the proportion that exceeds a specified critical value from the Chi-Squared Distribution can be reported and compared to the desired size of the test. While not a formal test, large proportions exceeding the desired critical value indicate problems with the model fit. For the example in Figure

3.4, the proportion of samples greater than the 0.8 percentile are 0.19 and 0.61 for cases (a) and (b) respectively, which again demonstrates the lack of fit for the second example. A number of additional examples applying the test statistic in (31) are presented in [84].

It is worth noting that the Gamma distribution has reasonable flexibility for modeling different behaviors in the data. Use of the empirical Bayes procedures outlined in Section 3.3 should also help to mitigate any problems with the parameters of the prior. Lack of fit in these cases is likely due to problems with the likelihood, with system specific factors causing the rate of occurrence of failures to be non-constant between successive corrective actions. This may occur in situations where system usage involves cycling between severe and benign environmental usage conditions.

3.5 Extensions to Basic Model

A particular benefit of the Bayesian framework established Section 3.2.4. is the ease with which various extensions to the basic model can be made. A number of possible extensions exist, but two very useful cases are presented here. The first involves an extension to allow for data collected from multiple test articles, which is often common in developmental reliability growth testing. The second includes uncertainty on the FEF. The FEF values are generally assessed based on engineering judgment and it is therefore desirable to allow for some uncertainty regarding their values. Though not presented here, it is also possible to develop marginal posterior distributions for the FEF values themselves if desired.

3.5.1 Multiple Systems Under Test

For situations when there are multiple identical systems undergoing the same DT, it is desirable to combine the failure data that are observed in order to more accurately assess the overall reliability of the system. This situation can present practical implementation problems for models that use failure mode first occurrence times such as [18]. It can be difficult to account for all testing that is accrued across the systems without artificially scaling the occurrence times of the failure modes, but the current model framework can easily handle this situation through a straightforward extension of the likelihood shown in (1). For a single failure mode, the extended version of the likelihood for multiple systems under test is given by

$$l\left(t_{1,i,1}, t_{1,i,2},\ldots, t_{1,i,n_{i,1}}, t_{2,i,1},\ldots, t_{p,i,n_p} \mid \lambda_i\right) = \left(1-d_i\right)^{\left(\sum_{j=1}^{p} n_{i,j} - n_{i,j,1}\right)} \lambda_i^{\sum_{j=1}^{p} n_{i,j}} * \exp\left(-\lambda_i\left[\sum_{j=1}^{p} v_{i,j} + \left(1-d_i\right)\left(\sum_{j=1}^{p} T_j - \sum_{j=1}^{p} v_{i,j}\right)\right]\right) \quad (32)$$

The previous notation is extended in (32) as follows:

1. There are p systems under test for test time T_j.
2. $t_{i,j,l}$ is the l^{th} observed time of failure for i^{th} failure mode on the j^{th} system.
3. There are $n_{i,j}$ occurrences of the i^{th} failure mode on the j^{th} system.
4. The $v_{i,j}$ are the corrective action times for the i^{th} failure mode on the j^{th} system.

The likelihood in (32) can also handle a number of common situations that occur in DT, such as dependencies due to preemptive corrective actions on certain systems due to failure modes being observed on previously tested systems. We point out that

setting the $v_{i,j}$ to zero easily handles this situation, which is again difficult for many traditional reliability growth projection models to handle appropriately.

The resulting system level posterior mean and variance can then be found similarly to the basic model in Section 3.2. The posterior mean and variance, respectively, are found to be

$$E[\lambda_s \mid n] = \sum_{i=1}^{m}\left(\frac{(1-d_i)\sum_{j=1}^{p} n_{i,j}}{\left(\frac{1}{\beta}+\sum_{j=1}^{p} v_{i,j}+(1-d_i)\left(\sum_{j=1}^{p} T_j-\sum_{j=1}^{p} v_{i,j}\right)\right)}\right)+\frac{\lambda_B}{1+\beta\sum_{j=1}^{p} T_j} \tag{33}$$

$$Var[\lambda_s \mid n] = \sum_{i=1}^{m}\left(\frac{(1-d_i)^2\sum_{j=1}^{p} n_{i,j}}{\left(\frac{1}{\beta}+\sum_{j=1}^{p} v_{i,j}+(1-d_i)\left(\sum_{j=1}^{p} T_j-\sum_{j=1}^{p} v_{i,j}\right)\right)^2}\right)+\frac{\lambda_B}{\beta\left(\frac{1}{\beta}+\sum_{j=1}^{p} T_j\right)^2}. \tag{34}$$

Note that the expressions in (33) and (34) are simple extensions to those in (9) and (10), with $v_i = \sum_{j=1}^{p} v_{i,j}$, $n_i = \sum_{j=1}^{p} n_{i,j}$, and $T = \sum_{j=1}^{p} T_j$. The same substitutions apply when calculating additional quantities of interest such as the growth potential failure intensity and the empirical Bayes estimators in Section 3.3.

For the posterior distribution on the number of failure modes observed in follow on testing the likelihood of observing a failure mode on at least one of the *p* systems by time *t* is updated as

$$p\left(I(t)=1 \mid \lambda, n, t>T\right) = 1-e^{-p\lambda(t-T)}. \tag{35}$$

The unconditional expected value in (19) is then modified to be

$$E\left[I(t) \mid t > T\right] = p\left(I(t) = 1 \mid n\right) = 1 - \frac{1}{\left(1 + \frac{p\left(t - T_{\max}\right)}{\frac{1}{\beta} + \sum_{j=1}^{p} T_j}\right)^{\alpha}}, \tag{36}$$

where T_{max} is defined as the largest of the T values. The result in (20) is then updated as

$$\lim_{K \to \infty} E\left[\sum_{i=m+1}^{K} I_i(t) \mid t > T\right] = \frac{\lambda_B}{\beta} \log\left(1 + \frac{p\left(t - T_{\max}\right)}{\frac{1}{\beta} + \sum_{j=1}^{p} T_j}\right). \tag{37}$$

For the prior number of failure modes used in the visual goodness of fit in Section 3.4, the likelihood of observing a failure mode on at least one of the p systems by time t is updated as

$$p\left(I(t) = 1 \mid \lambda, n, t > T\right) = 1 - e^{-p\lambda t}. \tag{38}$$

The expression in (28) is then modified to be

$$\lim_{K \to \infty} E\left[\sum_{i=1}^{K} I_i(t)\right] = \frac{\lambda_B}{\beta} \log\left(1 + \beta p t\right). \tag{39}$$

When comparing the observed failure modes to the prior prediction, note that the time of first occurrence for a failure mode should be the minimum of the occurrence times observed over all of the systems under test. Modifications to the Chi-Square test in Section 3.4.2 are not necessary for multiple systems, as (33) and (34) can be used directly in place of (9) and (10). The procedure deals with time between failures, so

the data from each system under test can be used within the existing framework. Note that because each system may have unique corrective action implementation during the test, additional random samples may be necessary to cover the additional systems. Caution should also be exercised to ensure that the appropriate distribution function corresponding to the corrective action interval is used.

3.5.2 Uncertain FEF

The model extension for including uncertainty on the FEF involves the addition of a prior distribution on the FEF parameter *d* in (1). The posterior distribution is then given by

$$p(\lambda \mid n) = \frac{p(d)\,p(\lambda)\,l(n \mid \lambda)}{\int_0^1 \int_0^\infty \int_0^1 p(d)\,p(\lambda)\,l(n \mid \lambda)\,\partial d\,\partial\lambda}\,\partial d. \tag{40}$$

The likelihood and prior distributions from Section 3.2 are used for the failure intensity. The Beta distribution given in (41) is used for the prior distribution on the FEF.

$$p(d) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\, d^{(a-1)}(1-d)^{(b-1)} \tag{41}$$

A useful method for describing the Beta prior can be found by specifying the mean and variance for the FEF and then developing the corresponding Method-of-Moments estimators.

The posterior for the FEF could also be found in a similar manner to that in (40). The outermost integration with respect to *d* could instead be calculated with

respect to λ, which would result in the desired marginal posterior on the FEF. The technique should be used cautiously though, particularly in cases where little information is available after the corrective action has been completed. The approach could be useful in managing the reliability growth program, as it would provide an indication of the overall effectiveness of the corrective action process.

The resulting system level posterior mean and variance can then be found similarly to the basic model in Section 3.2. For *m* observed modes with *m'* corrected, the posterior mean and variance are found to be

$$E[\lambda_s \mid n] = \sum_{i=1}^{m'} \frac{n_i}{\left(\frac{1}{\beta}+T\right)} \frac{{}_2F_1\left[n_i+1, a_i, a_i+b_i+n_i-n_{i,1}, \frac{T-\nu_i}{\frac{1}{\beta}+T}\right]}{{}_2F_1\left[n_i, a_i, a_i+b_i+n_i-n_{i,1}, \frac{T-\nu_i}{\frac{1}{\beta}+T}\right]} + \sum_{i=m'+1}^{m} \frac{n_i}{\left(\frac{1}{\beta}+T\right)} + \frac{\lambda_B}{1+\beta T} \tag{42}$$

$$Var\left[\lambda_s \mid n\right] = \sum_{i=1}^{m'} \left[\frac{n_i(n_i+1)}{\left(\frac{1}{\beta}+T\right)^2} \frac{{}_2F_1\left[n_i+2,\, a_i,\, a_i+b_i+n_i-n_{i,1},\, \frac{T-\nu_i}{\frac{1}{\beta}+T} \right]}{{}_2F_1\left[n_i,\, a_i,\, a_i+b_i+n_i-n_{i,1},\, \frac{T-\nu_i}{\frac{1}{\beta}+T} \right]} - \right.$$

$$\left. \left\{ \frac{n_i}{\left(\frac{1}{\beta}+T\right)^2} \frac{{}_2F_1\left[n_i+1,\, a_i,\, a_i+b_i+n_i-n_{i,1},\, \frac{T-\nu_i}{\frac{1}{\beta}+T} \right]}{{}_2F_1\left[n_i,\, a_i,\, a_i+b_i+n_i-n_{i,1},\, \frac{T-\nu_i}{\frac{1}{\beta}+T} \right]} \right\}^2 \right] +$$

$$\sum_{i=m'+1}^{m} \frac{n_i}{\left(\frac{1}{\beta}+T\right)^2} + \frac{\lambda_B}{\beta\left(\frac{1}{\beta}+T\right)^2} \qquad (43)$$

${}_2F_1(a,b,c,z)$ is the integral form of the hypergeometric function given by

$${}_2F_1(a,b,c,z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt. \qquad (44)$$

Additional parameters of interest such as the growth potential failure intensity and the number of modes observed in follow-on testing are developed similarly by substituting the posterior parameters in (42) and (43) for the original values in (9) and (10).

3.6 Simulation Performance

A simulation was developed to examine the performance of the basic model under varying conditions. Each replication of the simulation consists of the following steps:

1. Set input variables: prior parameters, test length, number of failure modes, mean and coefficient of variation of FEFs, and corrective action strategy.
2. Simulate mode failure intensities from prior Gamma distribution.
3. Simulate random FEF values for each failure mode from a Beta distribution.
4. For each failure mode, simulate random failures with the appropriate failure intensity according to the corrective action strategy.

The corrective action strategy in the simulation randomly determines when a corrective action is made for each observed failure mode. The corrective action decision is made by first choosing the probability that a corrective action will occur during the test. A uniform random number is then drawn for comparison to the probability for each observed failure mode. Model estimates were developed assuming that the exact FEF values were known. The “true” results from the simulation were then compared to the model estimates to gauge the robustness of the technique while also indicating the relative error that should be expected when applying this type of model. The coverage properties of the posterior probability intervals are also examined within the context of the simulation. Comparisons are

made between the proposed approach and two of the more popular reliability growth projection models: the AMPM [18] and the Crow-Extended Model [20]. The comparisons are made only for the point estimate (posterior mean) due to the fact that interval results are not yet published for the AMPM.

A number of different cases were examined using the simulation, with five presented here. Table 3.1 contains the input values for each case. The probability of corrective action is the probability that an observed failure mode is addressed during the test. K is the number of modes in the system, λ_B and β are the parameters of the Gamma distribution, α is found from equation (7), T is the length of the test, μ_d and cv are the mean and coefficient of variation, respectively, of the FEF distribution. Heuristic examination of the simulation convergence shows that 1000 replications were sufficient for each case. Table 3.2 shows the average result for each case, including the posterior mean and 80% lower probability bound, mean relative error, and coverage properties of the probability interval. Empirical distributions of the relative error for each case are provided in Figure 3.5.

TABLE 3.1

SIMULATION INPUTS

Case	Prob. of Corrective Action	K	β	λ_B	α	T	μ_d	cv
1	0.2	5000	0.001	0.002	0.0004	5000	0.7	0.1
2	0.1	2000	0.0001	0.02	0.1	5000	0.7	0.1
3	0.2	1000	0.001	0.02	0.02	2500	0.7	0.1
4	0.1	500	0.0001	0.002	0.04	5000	0.7	0.1
5	0.2	250	0.001	0.02	0.08	2500	0.7	0.1

TABLE 3.2

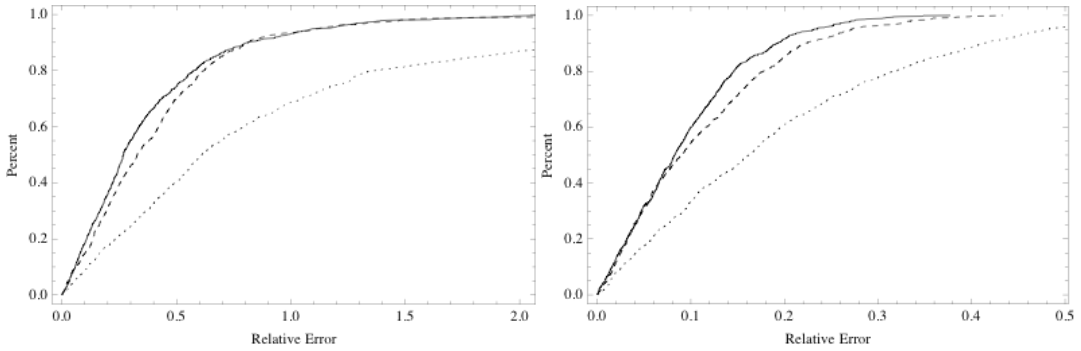
SIMULATION RESULTS

Case	Number Failures	True MTBF	Posterior Mean	Posterior 0.80 Lower Probability Bound	Abs Rel Error	Coverage	AMPM	Abs Relative Error	Crow Extended	Abs Rel Error
1	12	1498.3	1311.0	1008.3	0.38	0.82	1576.4	0.42	2216.0	1.13
2	99	67.0	67.9	64.4	0.10	0.65	64.2	0.11	66.2	0.18
3	48	105.4	105.4	94.1	0.15	0.75	106.3	0.16	106.0	0.34
4	11	676.5	695.2	589.3	0.29	0.70	660.7	0.31	878.6	0.68
5	48	107.8	107.6	95.8	0.16	0.75	109.8	0.17	118.3	0.34

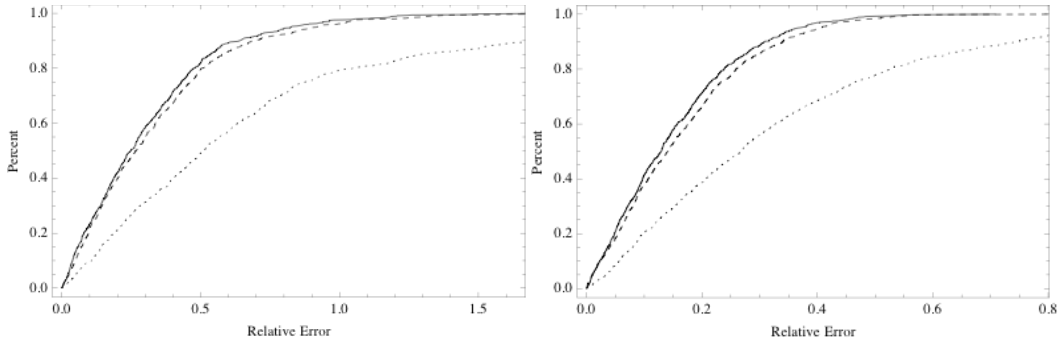
The results in Table 3.2 indicate that the model performs well on average. To aid in examining the model performance beyond just the average comparison, Figure 3.5 shows the empirical distribution of the relative error for each of the test cases. Higher slopes in the plots mean that larger proportions of samples fall closer to the true simulation value, which indicate more accuracy in the model result. The results in Figure 3.5 show that the proposed model outperforms both the AMPM and Crow Extended for accuracy. Additional simulation comparisons over a broader set of test cases are presented in [7], and the results indicate that AMPM outperforms the Crow Extended Model with respect to accuracy and relative error distributions. By

extension, the proposed approach should therefore perform well over a broad range of cases.

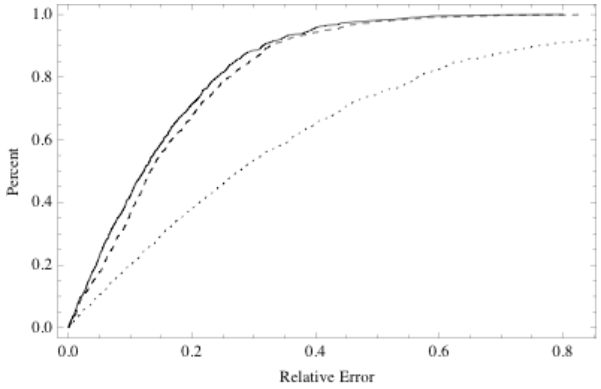
The classical coverage of the probability bounds are shown to be slightly non-conservative in most cases, but this is an unfortunate consequence of simulating with finite values of *K*. We can see this by examining the coverage under non-informative prior conditions such as those of a Jeffrey's prior. These cases should generally behave close to the desired classical properties, but explicit use of the Jeffrey's prior is not possible due to difficulties in convergence of the limit for complex systems. The posterior result will converge to a Jeffrey's based posterior as α becomes small and β becomes large though, which can be seen in Case 1 where *K* is large. Choosing *K* large enough will also act to drive α toward zero and cause behavior more in line with a non-informative prior. Because *K* can be considered arbitrarily large for a complex system, the resulting coverage properties for a complex system should match those of classical methods.



(a) (b)



(c) (d)



(e)

Fig 3.5. Empirical Distribution of Relative Error Comparisons for Cases 1-5 (a-e): Proposed Model (solid), AMPM (thick dashed), Crow-Extended (dotted)

3.7 Example Application

This section presents an example application of the model using data from a developmental test. Twenty-five systems were each tested for 175 hours, and 13 failure modes were observed across all of the systems. Across all systems, 43 total failures were observed. The occurrence times for each mode are shown by system in Table 3.3, along with the associated times of corrective action and FEFs. All

corrective actions were implemented after the test, with the exception of mode 8, which was implemented for all systems at 100 hours.

TABLE 3.3

SAMPLE FAILURE MODE DATA

Mode	FEF	System Number	Occurrence Time	Corrective action time
1	0.5	19	106.3	175
2	0	7	107.3	NA
		15	100.5	
		25	10.6	
3	0	3	79.4	NA
		6	67.3	
		6	67.1	
		6	70.8	
		6	162.3	
		14	102.7	
		17	100.8	
		17	126.3	
		22	13.9	
		22	64.1	
		22	84.4	
		23	39.2	
		23	48.5	
		23	45.6	
		23	53.3	
		23	68.7	
4	0	20	97.0	NA
		25	36.4	
5	0.8	21	70.8	175
6	0	3	148.1	NA
		7	70.6	
		8	118.1	
		22	18.0	
7	0.7	15	37.8	175
8	0.8	6	74.3	100
		7	65.7	
		7	93.1	
9	0.5	13	90.8	175
10	0.5	13	99.2	175
11	0.5	17	130.8	175
12	0	5	169.1	NA

		15	114.7	
		16	6.6	
		17	154.8	
		17	5.7	
		17	21.4	
		23	125.4	
		23	140.1	
13	0	7	102.7	NA

The empirical Bayes estimates in (23) and (24) are found to be

$$\hat{\lambda}_B = 0.0101 \tag{40}$$

$$\hat{\beta} = 0.0016 \tag{41}$$

Applying these estimates within the basic model in (9) and (10) yields the posterior mean and eighty percent upper probability bounds in (42) and (43) respectively.

$$\lambda_s = 0.0087 \tag{42}$$

$$\lambda_{UB} = 0.0098 \tag{43}$$

The resulting posterior approximation is plotted in Figure 3.6. Note that of the 7 failure modes that were addressed with corrective actions, the average FEF value was 0.61. This would indicate that the improvement in reliability from the corrective actions will not be significant. The point estimate of the failure intensity for a standard HPP is just 4375/43 = 0.0098, which is only slightly higher than the projected failure intensity after corrective actions in (42). This comparison not only provides a "consistency check" of the proposed model, but it also indicates the results that can be expected when the FEF values are somewhat moderate.

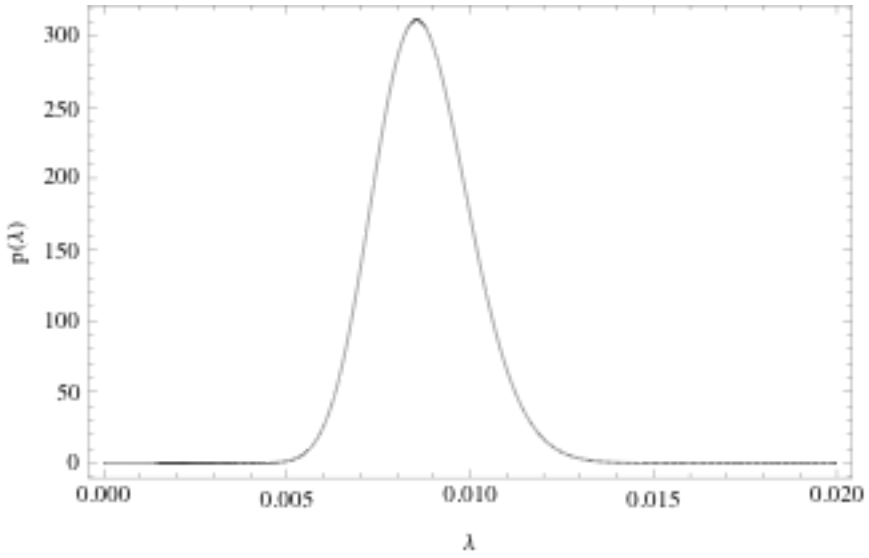


Fig 3.6. Approximate Posterior Distribution for Sample Problem

Goodness of fit can be checked using the visual and Chi-Squared tests developed in Section 3.4, with the appropriate modifications to handle multiple systems as discussed in Section 3.5.1. Figure 3.7 shows the failure mode first occurrence times overlaid on the distribution of the prior cumulative number of modes. The prior mean is the solid line, and the dashed lines represent a 90% probability interval. The visual results indicate that the data agree well with the model assumptions. Ten thousand samples of the Chi-Squared test statistic in (29) showed one sample greater than the critical value for a Chi-Squared variable with 6 degrees of freedom. The six degrees of freedom result from the $n = 43+25+2 = 70$ data points while using $n^{0.4}$ groups for the calculation. This small proportion provides additional evidence of agreement with the model assumptions.

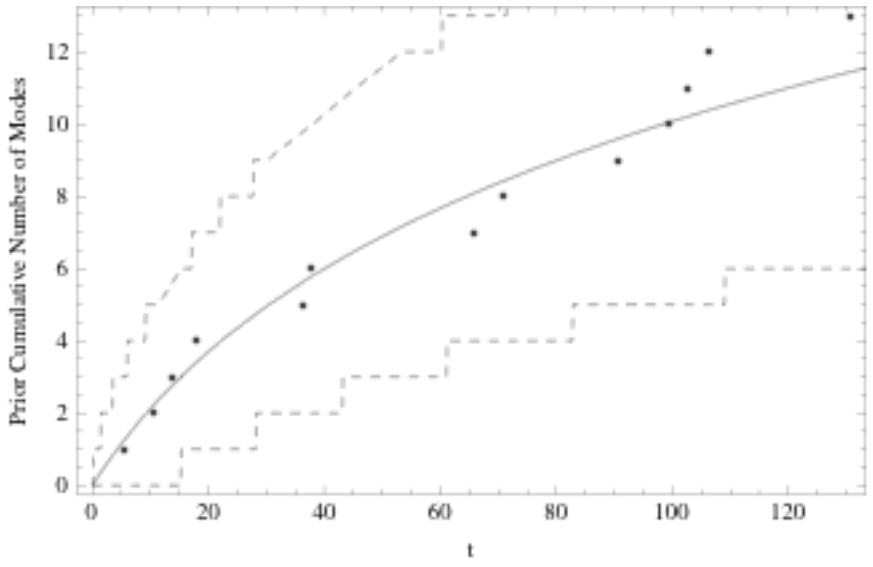


Fig 3.7. Comparison of Observed Failure Modes and Prior Predicted Distribution for Cumulative Number of Failure Modes for Sample Problem

Figure 3.8 also shows the distribution for the cumulative number of modes observed in follow-on testing of 250 hours. The mean is again the solid line, and the dashed lines represent a 90% probability interval.

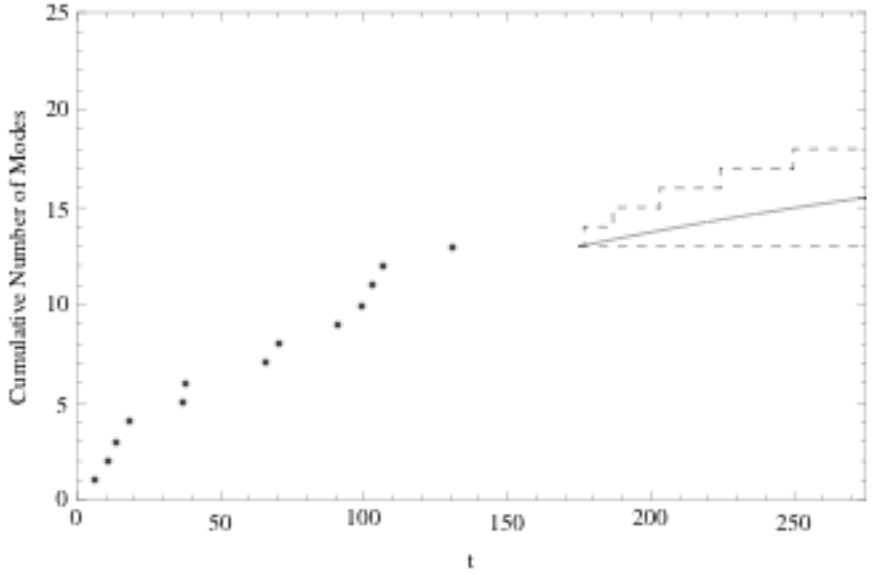


Fig 3.8. Distribution for Cumulative Number of Failure Modes in Additional Testing for Sample Problem

3.8 Conclusions

This chapter has presented a new method for projecting the reliability growth of a complex continuously operating system. The model allows for arbitrary corrective action strategies, and it differs from other models of this type by using all available data. It also differs from other reliability growth projection models in that it offers a complete inference framework via the posterior distribution on the system failure intensity. A unique feature of this approach with respect to other Bayesian techniques is the analytic expression for the failure intensity contribution from unobserved failure modes. Expressions for estimating the initial failure intensity, growth potential failure intensity, and the cumulative number of failure modes expected in future testing are also developed. Two separate goodness-of-fit procedures are presented for assessing the appropriateness of the underlying model assumptions. Extensions to the basic framework are also developed. The first accounts for multiple systems under test, and the second develops the posterior distribution while allowing for uncertainty on the FEF values that are assessed.

Taken as a whole, the results are useful for reliability program managers making decisions regarding the investment of resources to improve reliability. The model can be used to estimate the resulting reliability after corrective actions are applied, and the associated metrics can help to aid in future decisions. Resource planning can be made based on the posterior distribution for the number of failure modes that will be seen in future testing. The use of the growth potential failure

intensity can also indicate whether additional test-fix-test reliability should be pursued for the program. If the resulting growth potential failure intensity estimates are higher than the desired objective, resources should be expended for reliability improvement initiatives outside of the traditional test-fix-test reliability growth approach.

4 ASSESSING RELIABILITY GROWTH USING DEVELOPMENTAL AND OPERATIONAL TEST DATA[2]

4.1 Introduction

4.1.1 Background

The approach presented here is conceptually the same as those in [68]-[70] and [85]. The major difference being that the Bayesian formulation easily allows for uncertainty to be included throughout the problem, to include the degradation scale factor between developmental growth and operational demonstration testing. This provides an additional uncertainty in the results that will more closely match the practical situation, in which limited information on the actual degradation may be available. The model also provides a complete inference framework via the posterior distribution, which includes both developmental and operational reliability information. These types of results are particularly useful in planning reliability growth programs, where different test events, both growth and constant configuration, may be used throughout. The results can also be extended to develop a Bayesian reliability growth planning model, which is the subject of Chapter 5.

4.1.2 Chapter Overview

The methodology of the approach is presented in Section 4.2. It includes model assumptions, data requirements, and the development of the initial framework for combining data from two different tests. The presentation of the methodology is

[2] A condensed version of the material in Chapter 4 was accepted for publication in 2014 RAMS Proceedings. A full detailed version will be submitted for journal publication when Chapter 3 is accepted for publication.

broken into two separate sections: developmental reliability growth assessment, and operational reliability assessment using the developmental reliability growth assessment posterior results as prior information. Straightforward generalizations of the initial framework that are appropriate for multiple systems under test are also presented. Analogous OC curve results for the proposed model are developed in Section 4.3. A Bayesian goodness of fit of test to assess the appropriateness of the overall model assumptions is presented in Section 4.4. Section 4.5 discusses the performance of the proposed model, and conclusions are presented in Section 4.6.

4.2 Methodology

The development of the methodology considers the combination of developmental and operational test data for a single system while accounting for differences in the underlying failure intensity that may exist between the two test events. The method allows for reliability growth through the developmental test, where one or more failure modes may be mitigated through the implementation of a corrective action. The operational test is assumed to be a constant configuration test, and the difference in the two test environments is modeled explicitly. More specifically, a decrease in reliability (or increase in failure intensity) between the test phases is considered probabilistically. The methodology is then extended to include multiple test phases and multiple systems under test.

4.2.1 Assessing Reliability Growth in DT

This section presents a method for modeling the reliability growth of a system during a developmental test phase. The results from this procedure are then used as

prior information when assessing the reliability from the operational test in Section 4.2.2. The approach is the same as that developed in Chapter 3, but an outline of the method is presented here for completeness. The assumptions for the reliability growth test data are as follows:

1. The system is sufficiently complex; i.e. a large number of failure modes exist in the system.
2. Failure modes generate system failures independently.
3. Each occurrence of a failure mode results in a system failure.
4. The failure intensity for a given failure mode is constant both before and after a corrective is implemented.
5. No new failure modes are induced by corrective actions.

Using the same notation as in Chapter 3, we assume for failure mode i that n_i failures are observed in test length T with times $\left(t_{i,1},t_{i,2},...,t_{i,n_i}\right)$. Further assume that the failures are divided such that $n_{i,1}$ occur before a corrective action is implemented and n_i - $n_{i,1}$ failures occur after the corrective action. For a corrective action that occurs at time v_i, let d_i represent the FEF resulting from the corrective action. From assumptions 2. and 4. the likelihood can then be expressed as

$$l\left(t_{i,1},t_{i,2},...,t_{i,n_i},n_i,n_{i,1}\mid\lambda_i\right)=\left(1-d_i\right)^{n_i-n_{i,1}}\lambda_i^{n_i}\exp\left(-\lambda_i\left[v_i+\left(1-d_i\right)\left(T_{DT}-v_i\right)\right]\right). \quad (1)$$

We note that the likelihood involves a telescoping sum of the individual failure times, and the simplified form in (1) reduces to a function of the number of failures. The individual failure times are therefore suppressed in future notation.

For the prior distribution on the failure intensity, it is useful to consider a distribution that will adequately reflect the failure intensity for a given failure mode. The failure intensities for the collection of failure modes found in a complex system are shown to be adequately modeled as a random realization from a Gamma distribution [18], [26]. This form of the prior probability distribution recognizes what is commonly referred to as the "vital few, trivial many" property among failure mode failure intensities. This property acknowledges that each failure mode provides a different contribution to the overall system failure intensity, with a relatively few number of failure modes being significant enough to be observed in test. We assume the prior distribution to be a Gamma(α,β) parameterized as

$$p(\lambda_i) = \frac{\lambda_i^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} \exp\left(-\frac{1}{\beta}\lambda_i\right), \tag{2}$$

where $\alpha > 0$ and $\beta > 0$. Because of the conjugate nature of (2), standard techniques yield the posterior distribution for a single failure mode to be the Gamma distribution parameterized in (3).

$$p(\lambda_i \mid n_i) = \frac{\lambda_i^{\alpha+n_i-1}}{\Gamma(\alpha+n_i)\left[\frac{1}{\beta} + v_i + (1-d_i)(T_{DT}-v_i)\right]^{-(\alpha+n_i)}} * \exp\left[-\lambda_i\left(\frac{1}{\beta} + v_i + (1-d_i)(T_{DT}-v_i)\right)\right] \tag{3}$$

Note also that if no corrective action is attempted, the d_i are equal to zero, and (3) reduces to the Gamma posterior that is commonly found through conjugate Gamma-Exponential methods.

The posterior estimate for the system level failure intensity can be found by summing the individual mode posterior estimates and taking the limit as the number

of modes becomes large. The result was shown in Chapter 3 to be well approximated by a Gamma distribution with mean and variance given in (4) and (5). We use the notation λ_{DT} to denote the system level failure intensity from DT and distinguish between later assessments of the desired operational failure intensity. The variable *n* refers to the vector containing the number of failures for each of the *m* observed failure modes. In taking the limit with respect to the number of modes, the mean and variance in (4) and (5) are expressed in terms of the prior system level mean λ_B and the original β parameter from the prior Gamma distribution.

$$E\left[\lambda_{DT} \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T_{DT} - v_i)\right)}\right) + \frac{\lambda_B}{1+\beta T_{DT}} \tag{4}$$

$$Var\left[\lambda_{DT} \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)^2 n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T_{DT} - v_i)\right)^2}\right) + \frac{\lambda_B}{\beta\left(\frac{1}{\beta} + T_{DT}\right)^2}. \tag{5}$$

This results in

$$\lambda_{DT} \sim Gamma\left[\tilde{\alpha},\ \tilde{\beta}\right], \tag{6}$$

where the approximate Gamma can be easily determined using the mean and variance to yield the parameters in (7) and (8).

$$\tilde{\alpha} = \frac{E\left[\lambda_{DT} \mid n\right]^2}{Var\left[\lambda_{DT} \mid n\right]} \tag{7}$$

$$\tilde{\beta} = \frac{Var\left[\lambda_{DT} \mid n\right]}{E\left[\lambda_{DT} \mid n\right]} \tag{8}$$

Note that both the prior and posterior distributions on the system level failure intensity in DT are Gamma, which is a conjugate relationship. This allows for ease in modeling multiple developmental test phases within the framework already established. The posterior distribution results in (7) and (8) can be used as prior parameters for a follow-on phase of DT, and this updating process can be continued with each additional phase of DT. Updating in this manner assumes consistency between successive test phases, and the posterior defined by (7) and (8) may not be an appropriate prior if this assumption is significantly violated. In these cases it is important to consider the model goodness of fit procedures described in Chapter 3 to ensure that the results are reasonable. Equations for empirical Bayes estimation of the unknown prior λ_B and β parameters in (4) and (5) are also provided in Chapter 3.

4.2.2 Assessing Reliability in OT

For assessing the reliability from an operational test, the Gamma posterior distribution from Section 4.2.1 can be used to develop the prior distribution. The conjugate relationship of the Gamma-Exponential can easily be leveraged in this context, but the degradation in the system reliability must also be considered. This degradation is traditionally considered in terms of a decrease in the system Mean Time Between Failure (MTBF). Assuming a $100\gamma\%$ degradation in the MTBF (or a corresponding increase in system failure intensity) leads to the relationship between the developmental and operational failure intensities shown in (9)-(11), where the DT and OT subscripts denote the corresponding MTBF and failure intensity values.

$$MTBF_{OT} = (1-\gamma) MTBF_{DT} \tag{9}$$

$$\frac{1}{\lambda_{OT}} = (1-\gamma)\frac{1}{\lambda_{DT}} \tag{10}$$

$$\lambda_{DT} = (1-\gamma)\lambda_{OT} \tag{11}$$

Utilizing the form of the distribution in (6) and properties of the Gamma distribution [86] leads to

$$\lambda_{OT} = \frac{\lambda_{DT}}{(1-\gamma)} \sim Gamma\left[\tilde{\alpha}, \frac{\tilde{\beta}}{(1-\gamma)}\right]. \tag{12}$$

The prior distribution in (12) is conditioned on the γ parameter, so we can also express (12) as

$$\lambda_{OT} \mid \gamma \sim Gamma\left[\tilde{\alpha}, \frac{\tilde{\beta}}{(1-\gamma)}\right]. \tag{13}$$

The expression in (13) is now a reasonable prior distribution for use with the test data from the operational test phase. For n_{OT} failures in test time T_{OT} at times $\left(t_{OT,1}, t_{OT,2}, \ldots, t_{OT,n_{OT}}\right)$ the likelihood is a simplified version of (1) shown in (14).

$$l\left(t_{OT,1},\ t_{OT,2},\ \ldots\ ,\ t_{OT,n_{OT}}\ ,\ n_{OT} \mid \lambda_{OT}\right) = \lambda_{OT}^{\ n_{OT}} \exp\left(-\lambda_{OT} T_{OT}\right) \tag{14}$$

A simple conjugate relationship would yield a posterior distribution that is conditional on γ. The true value of γ is unknown though, so the desired unconditional posterior is given by

$$p\left(\lambda_{OT} \mid n_{OT}\right) = \int_{\Gamma} \frac{p(\gamma)\, p\left(\lambda_{OT} \mid \gamma\right) l\left(t_{OT,1}, t_{OT,2}, ?, t_{OT,n_{OT}}, n_{OT} \mid \lambda_{OT}\right)}{\iint_{\Lambda,\Gamma} p(\gamma)\, p\left(\lambda_{OT} \mid \gamma\right) l\left(t_{OT,1}, t_{OT,2}, ?, t_{OT,n_{OT}}, n_{OT} \mid \lambda_{OT}\right) \partial\lambda_{OT}\, \partial\gamma}\, \partial\gamma, \tag{15}$$

where the Λ and Γ in (15) denote the support of the prior distributions on λ and γ, and the degradation is assumed to be independent of the failure intensity. Note that the expression in (15) treats the degradation factor as a nuisance parameter by finding the joint posterior distribution and then calculating the resulting marginal distribution of interest. We again suppress the individual failure times in notation for the posterior distribution.

When developing the prior distribution on the degradation between test phases, detailed information is generally not available. For this reason we demonstrate the use of the Maximum Entropy principle [87],[88] to provide a repeatable approach that allows for consistency in application. We further assume that only average MTBF degradation values are available. The average values can be determined through examination of historical performance on similar systems, or through comparison of the potential failure modes that exist in the system with the developmental test environment. Maximizing the entropy subject to the assumed mean value of the MTBF degradation γ and a range of (0,1) results in the prior distribution for γ being a truncated Exponential distribution given by

$$p(\gamma) = \frac{\mu \exp(-\mu\gamma)}{1 - \exp(-\mu)}, \tag{16}$$

where μ is the solution to

$$\frac{1}{\mu} - \frac{\exp(-\mu)}{1 - \exp(-\mu)} = \varepsilon \tag{17}$$

for mean degradation value ε. Examining equation (17) reveals that μ will be zero when the mean degradation is equal to 0.5. This presents no real problem in practice though, as the mean value of the degradation is not likely to be known with high

precision. Perturbing the mean slightly to 0.49 will allow for a positive solution, and the end result can be shown to be insensitive to this level of difference. To aid in analytic calculations of (15), the truncated Exponential distribution in (16) is approximated with a Beta distribution. The parameters of the Beta distribution can be found by equating the mean and second moment about the origin of the two distributions, which results in the system of equations given by (18) and (19).

$$\frac{a}{a+b} = \frac{1}{\mu} - \frac{\exp(-\mu)}{1-\exp(-\mu)} \tag{18}$$

$$\left(\frac{a}{a+b}\right)\left(\frac{a+1}{a+b+1}\right) = \frac{\exp(-\mu) + 2\left[-\frac{1}{\mu}\exp(-\mu) - \frac{1}{\mu^2}\exp(-\mu) + \frac{1}{\mu^2}\right]}{1-\exp(-\mu)} \tag{19}$$

A comparison plot of the two prior distributions is shown in Figure 4.1. The mean in the example is set at ε = 0.20, and the priors are seen to match very closely on the (0,1) interval.

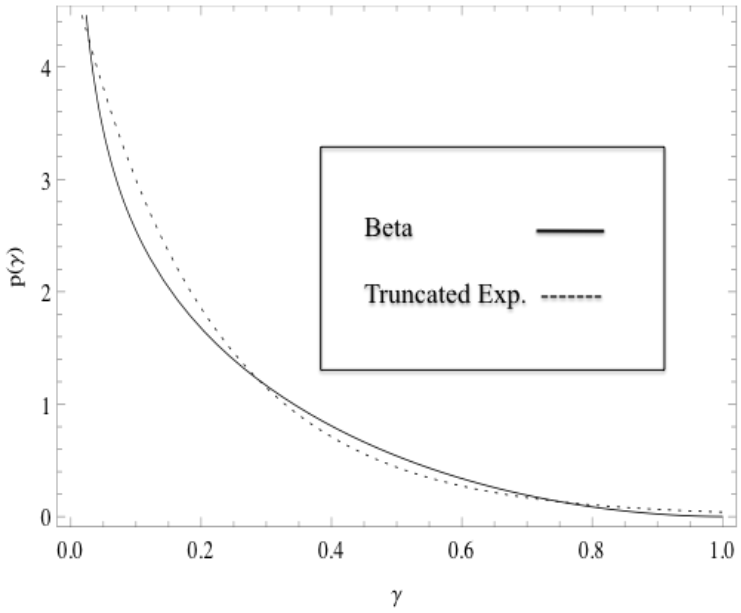


Fig. 4.1: Comparison of Truncated Exponential and Beta Prior Distributions

We point out that the model is developed using only the Beta approximation as the prior distribution. While the maximum entropy approach in (16)-(19) provides a repeatable framework that will be consistent between different users with the same mean degradation, other relevant information should also be used whenever possible to develop the Beta prior.

For the system failure intensity under squared error loss, the Bayes estimate of the failure intensity is just the mean of the posterior distribution in (15). Utilizing the Beta prior on the degradation, the mean is found to be

$$E\left[\lambda_{OT} \mid n_{OT}\right] = \left(\frac{\tilde{\alpha} + n_{OT}}{\frac{1}{\tilde{\beta}} + T_{OT}}\right) \frac{{}_2F_1\left[\tilde{\alpha} + n_{OT} + 1, a, a + b + \tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}} + T_{OT}}\right]}{{}_2F_1\left[\tilde{\alpha} + n_{OT}, a, a + b + \tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}} + T_{OT}}\right]} \tag{20}$$

where ${}_2F_1(a,b,c,z)$ is the integral form of the hypergeometric function given by

$${}_2F_1(a,b,c,z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt. \tag{21}$$

The function given in (21) can be evaluated by most standard mathematical software, and numerous numerical procedures in many programming languages are also available. Note that the mean of the posterior Gamma distribution with no degradation is given by

$$E\left[\lambda_{OT}\right] = \frac{\tilde{\alpha} + n_{OT}}{\frac{1}{\tilde{\beta}} + T_{OT}}. \tag{22}$$

The degradation between the two test environments results in the additional factor containing the Hypergeometric functions. The ratio acts as a scale parameter for the usual posterior mean to account for the differences in the two test environments. The posterior variance can be developed similarly as

$$Var\left[\lambda_{OT} \mid n_{OT}\right] = \left(\frac{\left(\tilde{\alpha}+n_{OT}\right)\left(\tilde{\alpha}+n_{OT}+1\right)}{\left(\frac{1}{\tilde{\beta}}+T_{OT}\right)^{2}}\right)\frac{{}_2F_1\left[\tilde{\alpha}+n_{OT}+2,a,a+b+\tilde{\alpha},\frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+T_{OT}}\right]}{{}_2F_1\left[\tilde{\alpha}+n_{OT},a,a+b+\tilde{\alpha},\frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+T_{OT}}\right]} - \left[\left(\frac{\left(\tilde{\alpha}+n_{OT}\right)}{\left(\frac{1}{\tilde{\beta}}+T_{OT}\right)}\right)\frac{{}_2F_1\left[\tilde{\alpha}+n_{OT}+1,a,a+b+\tilde{\alpha},\frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+T_{OT}}\right]}{{}_2F_1\left[\tilde{\alpha}+n_{OT},a,a+b+\tilde{\alpha},\frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+T_{OT}}\right]}\right]^2 \tag{23}$$

The mean of the posterior distribution in (22) is a scaled Gamma mean. This provides evidence that the overall posterior may be well approximated by a Gamma distribution with parameters given by (7) and (8), substituting the mean and variance from (20) and (23). To confirm this notion, Figure 4.2 shows a histogram developed using a simple Metropolis Random Walk to generate samples from the posterior in (15). The dashed line is the approximate Gamma developed using (20) and (23), which confirms that the Gamma provides a reasonable description of the posterior distribution.

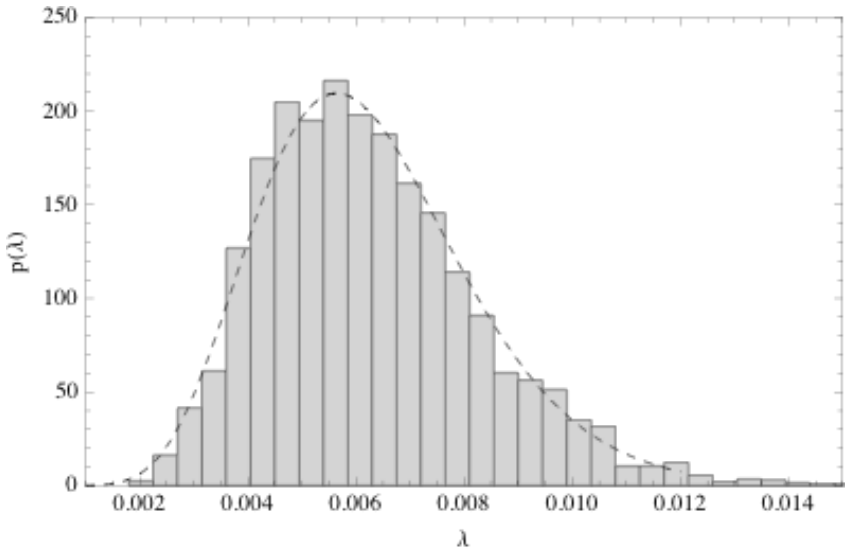


Fig. 4.2: Approximate Gamma Overlaid on True Posterior Distribution

4.2.3 Generalization for Multiple Systems under Test

The results in Sections 4.2.1 and 4.2.2 pertain to a single system under test. A particularly useful property of the Bayesian framework is the ease of extension that is possible. This section presents one such extension to address the common situation involving multiple identical systems under test.

The extension of the reliability growth model in Section 4.2.1 was presented in Chapter 3. The posterior distribution can again be modeled sufficiently with a Gamma distribution, where the parameters are developed using the appropriate posterior mean and variance along with the techniques described in Section 4.2.1. This allows us to utilize the same approach as in Section 4.2.2 for developing the prior distribution on λ. For the extension of the likelihood in (14), assume there are p

systems under test. Next assume that for a given system j, $n_{OT,j}$ failures are observed in test length T_j with times $\left(t_{OT,j,1},\ldots,\ t_{OT,j,n_{OT,j}}\right)$. The likelihood is then given as

$$l\left(t_{OT,1,1},\ldots,t_{OT,p,n_{OT,p}},n_{OT,1},\ldots,n_{OT,p} \mid T_{OT,1},\ldots,T_{OT,p}\lambda_{OT}\right)=\prod_{j=1}^{p}\lambda_{OT}{}^{n_{OT,j}}\exp\left(-\lambda_{OT}T_{OT,j}\right)$$
$$=\lambda_{OT}{}^{\sum_{j=1}^{p}n_{OT,j}}\exp\left(-\lambda_{OT}\sum_{j=1}^{p}T_{OT,j}\right) \quad (24)$$

The likelihood again involves a telescoping sum of the individual failure times, and the simplified form reduces to a function of the total number of failures across each of the systems. Substituting (24) into the posterior defined in (15) results in the updated mean and variance shown in (25) and (26), which are easily found by substituting $\sum_{j=1}^{p}n_{OT,j}$ for n_{OT} and $\sum_{j=1}^{p}T_{OT,j}$ for T_{OT}.

$$E\left[\lambda_{OT} \mid n_{OT}\right]=\left(\frac{\tilde{\alpha}+\sum_{j=1}^{p}n_{OT,j}}{\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p}T_{OT,j}}\right)\frac{{}_2F_1\left[\tilde{\alpha}+\sum_{j=1}^{p}n_{OT,j}+1,a,a+b+\tilde{\alpha},\frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p}T_{OT,j}}\right]}{{}_2F_1\left[\tilde{\alpha}+\sum_{j=1}^{p}n_{OT,j},a,a+b+\tilde{\alpha},\frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p}T_{OT,j}}\right]} \quad (25)$$

$$Var\left[\lambda_{OT} \mid n_{OT}\right]=\left(\frac{\left(\tilde{\alpha}+\sum_{j=1}^{p}n_{OT,j}\right)\left(\tilde{\alpha}+\sum_{j=1}^{p}n_{OT,j}+1\right)}{\left(\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p}T_{OT,j}\right)^2}\right)\frac{{}_2F_1\left[\tilde{\alpha}+\sum_{j=1}^{p}n_{OT,j}+2,a,a+b+\tilde{\alpha},\frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p}T_{OT,j}}\right]}{{}_2F_1\left[\tilde{\alpha}+\sum_{j=1}^{p}n_{OT,j},a,a+b+\tilde{\alpha},\frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p}T_{OT,j}}\right]}-$$

$$\left[\left(\frac{\left(\tilde{\alpha}+\sum_{j=1}^{p} n_{OT,j}\right)}{\left(\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p} T_{OT,j}\right)}\right)\frac{{}_2F_1\left[\tilde{\alpha}+\sum_{j=1}^{p} n_{OT,j}+1, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p} T_{OT,j}}\right]}{{}_2F_1\left[\tilde{\alpha}+\sum_{j=1}^{p} n_{OT,j}, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+\sum_{j=1}^{p} T_{OT,j}}\right]}\right]^2 \quad (26)$$

The likelihood in (24) can also be considered as a single system with test time equal to the total test time across all of the systems. Using this representation, the system level posterior distribution is again sufficiently represented by a Gamma distribution with the parameters defined in (7) and (8).

4.3 Demonstration Testing

As mentioned previously, demonstration test planning with classical OC-curve analysis presents a number of issues for practical application. The method involves calculating the consumer and producer risks associated with the planned demonstration test, where demonstration is considered to be successful if the appropriate lower confidence bound on the MTBF is greater than or equal to the required MTBF [6]. The consumer risk, denoted by α, is the risk of accepting that the MTBF of the system meets its requirement when it truly does not. It is determined by the desired statistical confidence level, resulting in a maximum number of failures that can be observed during a successful test. Producer risk is the probability of not accepting that the MTBF of the system meets its requirement when it truly does. It can be found by considering the power or probability of a successful demonstration the test, where a successful test is defined as observing less than or equal to the

allowable number of failures during the demonstration test. The producer risk can then be recognized as the complement of the probability of a successful test. The approach is conditional on the value of the system failure intensity entering the test, which is not known in practice. This is due to two sources of uncertainty: the first is that involving the system failure intensity at the end of the developmental test program, and the second is the amount of degradation that will occur when transitioning from a developmental test environment to an operational test environment.

The model framework outlined in Section 4.2.2 explicitly addresses the uncertainty present prior to entering the test by considering uncertainty on both the system failure intensity and the associated degradation factor between the prior DT and the operational demonstration test. Figure 4.3 shows an example of the Gamma posterior distribution of the failure intensity relative to the requirement for a consumer risk of 0.20. As shown in the figure, consumer risk of 0.20 corresponds to 0.20 posterior probability of the failure intensity being greater than the requirement. When considering the consumer risk in this setting, the appropriate upper probability bound from the Gamma posterior distribution defined by (20) and (23) can be used to determine the maximum number of allowable failures for the desired level of consumer risk.

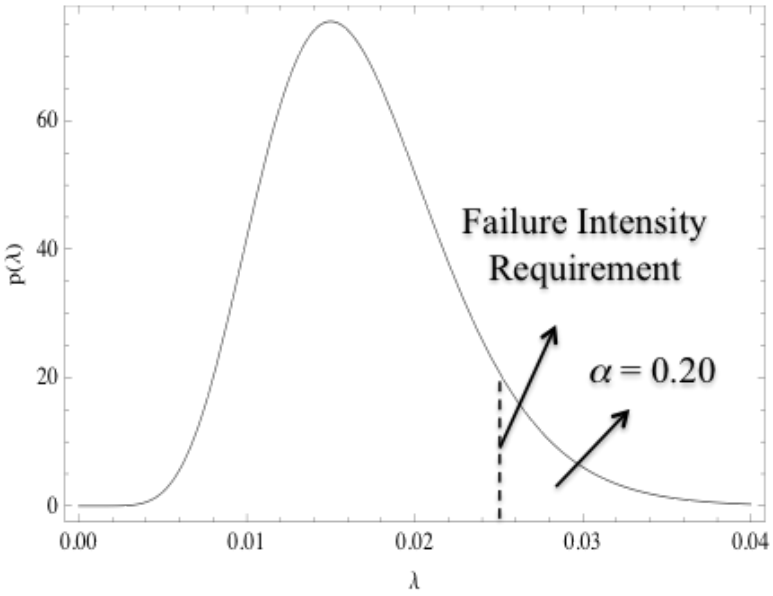


Fig 4.3: Graphical Representation of Consumer Risk ($\alpha = 0.20$)

When the maximum number of failures allowed for a successful demonstration test, n_{max}, has been determined, the full unconditional probability of a successful test can also be calculated. The Gamma prior defined in (13) and the Beta prior defined by (18) and (19) account for the uncertainty in the failure intensity and degradation respectively. The resulting probability is given by

$$p = \sum_{i=0}^{n_{\max}} \left(\frac{{T_{OT}}^{i}}{i!} \right) \frac{\Gamma(a+b)\Gamma(\tilde{\alpha}+i)\Gamma(b+\tilde{\alpha})}{\Gamma(b)\Gamma(\tilde{\alpha})\Gamma(a+b+\tilde{\alpha})\tilde{\beta}^{\tilde{\alpha}} \left(\frac{1}{\tilde{\beta}} + T_{OT} \right)^{\tilde{\alpha}+i}} * {}_2F_1\left[\tilde{\alpha}+i, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}} + T_{OT}} \right] \tag{29}$$

The expression in (29) can be developed as a straightforward extension of the classical probability of acceptance defined in [6] as

$$p = \sum_{i=0}^{n_{OT}} \frac{\left(\lambda_{OT} T_{OT}\right)^{i} \exp\left(-\lambda_{OT} T_{OT}\right)}{i!}. \tag{30}$$

Equation (29) can be developed from (30) by substituting the DT-OT relationship from (11) and (12) for the operational failure intensity in (30) and then marginalizing with respect to the distributions in (13), (18), and (19). This result represents a more complete description of the probability of a successful test and the corresponding producer risk that exists.

When the reliability demonstration event is used to develop reliability design goals, use of the posterior distribution in (20) and (23) along with (29) will result in lower goals for the same consumer and producer risks when compared to the classical OC curve based results. These design goals can be thought of as reliability "demonstration margins", which serve as overall programmatic risk indicators for planned development and testing programs. The proposed assessment approach using combined developmental and operational test data can therefore be seen to directly reduce the programmatic risks that may exist due to reliability demonstration with operational test data alone. Note that the reduced goals are a direct result of the additional information provided by the development reliability growth test data. The reliability growth model in Section 4.2.1 and Chapter 3 provides a substantial amount of information on the failure intensity of the system during DT. This information is useful even when a conservative approach is used to assign the prior on the degradation factor, such as with the maximum entropy approach in Section 4.2.2.

4.4 Model Assessment

The validity of the assumptions and relative "goodness" of the model are important considerations when utilizing any model for inference. A chi-squared goodness of fit test [83],[84] can be used to provide an indication of the appropriateness of the chosen model. The Bayesian chi-squared test statistic developed in [83] utilizes the properties of the cumulative distribution function corresponding to the likelihood function used in the development of the posterior distribution along with samples from the posterior distribution. Application of this test to developmental reliability growth test data was presented in Chapter 3 and will not be presented in this chapter. For assessing the combined developmental and operational test model, the method is as follows.

Following the method outlined in [84], the posterior distribution is the Gamma distribution developed in Section 4.2.2 with mean and variance in (20) and (23). The unit interval is then divided into q sub-intervals such that $0 \equiv a_0 < a_1 < ... < a_{q-1} < a_q \equiv 1$. For failure times ($t_{OT,1}$,$t_{OT,2}$,…,$t_{OT,nOT}$), let y_j be the times between successive operational test failures for j = 1,…,n_{OT}. Also let F_j be the corresponding cumulative Exponential distribution with the failure rate parameter $\tilde{\lambda}$ sampled from the Gamma posterior. Next define $z_j(\tilde{\lambda})$ to be a vector of length q whose $j^{th}q^{th}$ element is zero unless

$$F(y_j \mid \tilde{\lambda}) \in (a_{q-1}, a_q), \tag{31}$$

in which case the q^{th} element of z_j is set to one. For the failure free interval at the end of the operational test, a random number is drawn from the interval $\left[1-F\left(T_{OT}-t_{OT,n_{OT}} \mid \tilde{\lambda}\right), 1\right]$ in place of the F_j in (31) [84]. Next define

$$m\left(\tilde{\theta}\right)=\sum_{j=1}^{n_{OT}+1} z_j\left(\tilde{\theta}\right). \tag{32}$$

The expression in (32) is a vector of length q, where each element of the vector represents the total number of points in the q^{th} subinterval of the unit interval. The summation in (32) also involves n_{OT} + 1 terms because of the failure free interval at the end of the test. The chi-square statistic from [83] is defined as

$$R^B\left(\tilde{\theta}\right)=\sum_{k=1}^{q}\left[\frac{m_k\left(\tilde{\theta}\right)-\left(n_{OT}+1\right)p_k}{\sqrt{\left(n_{OT}+1\right)p_k}}\right]^2, \tag{33}$$

where the p_k are defined as the differences between the a_q values that divide the unit interval. The statistic in (33) is asymptotically chi-squared with q -1 degrees of freedom under repeated sampling of both the data and the posterior parameter of interest. As described in Chapter 3 and [83], the utility of the statistic is that poorly specified models will not follow the proper chi-squared distribution and will likely have a higher number of samples in the tail of the distribution. The proportion of samples of (33) that fall beyond the desired critical value of the chi-squared distribution can be used to assess model fit. For example, 50% of the sampled values of R_B falling above the 0.95 quantile of the chi-squared distribution would indicate a problem with model fit [84].

4.5 Performance Comparisons

Performance of the proposed method is examined using two separate approaches. The first involves comparing the relative error distributions for the proposed method and the classical estimate from multiple replications of simulated tests. The second compares the Bayes posterior risk of each method to examine the sensitivity of the prior distribution on the DT-OT degradation factor.

For the relative error comparisons, a simulation was developed to examine the behavior of the proposed model. The simulation uses an input parameterization for the Gamma distribution and then generates random failure intensities for the specified number of failure modes. For each realized value of the failure mode developmental failure intensity, random failures are then generated for a developmental test phase of a desired length. Corrective actions are applied according to an arbitrary corrective action strategy, with random FEF values for each mode drawn from a Beta distribution. The failure intensities are then scaled through the application of a degradation factor that is randomly sampled from an additional Beta distribution. This determines the true operational failure intensity, from which random failures are then generated for a second operational test of desired length. The true underlying system level failure intensity (and corresponding MTBF) is then known from the sum of the realized values in the simulation, and there are realized failures from one or more of the test phases that can be used in the model framework to estimate the system level failure intensity. One thousand replications were determined to be sufficient based on sufficient convergence of the mean result as a function of the number of replications.

Table 4.1 contains ten cases that were examined for comparison purposes. The cases were chosen to cover the possible scenarios with different amounts of testing and reliability. Table 4.2 contains the average results over the one thousand replications for each case. MTBF values are reported in the table instead of the failure intensities to provide a more intuitive understanding of the inputs for each case. The average relative error values for both the Bayesian technique proposed here and the classical estimate using only the data from the operational test event are also included. The average absolute relative error is a useful measure of performance for models of this type, and it provides an indication on the general ability of the model to provide a reasonable estimate of the desired system failure intensity. The absolute relative error for a single case is defined as

$$\text{Rel error} = \frac{\left|\hat{\lambda} - \lambda\right|}{\lambda}, \tag{34}$$

where λ is the true operational system failure intensity from the simulation, and $\hat{\lambda}$ is the model estimate resulting from the simulated data. Note that the use of the absolute relative error may provide a conservative indication of the performance of the estimators. The failure intensities are generally small values, so seemingly minor differences may actually be large percentage values and result in high relative error values.

TABLE 4.1

COMPARISON CASES FOR PROPOSED MODEL

Case	Number of Failure Modes (K)	β	Initial DT MTBF	DT Length	OT Length	Mean Degradation
1	500	0.0001	100	1000	2000	0.2
2	500	0.0001	400	1000	2000	0.2
3	500	0.0001	100	2000	2000	0.2
4	500	0.0001	400	2000	2000	0.2
5	500	0.0001	100	2000	500	0.2
6	500	0.0001	400	2000	500	0.2
7	500	0.0001	100	5000	500	0.2
8	500	0.0001	400	5000	500	0.2
9	500	0.0001	100	10000	500	0.2
10	500	0.0001	400	10000	500	0.2

TABLE 4.2

COMPARISON RESULTS FOR SIMULATION CASES

Case	True MTBF	Bayes MTBF (Post Mean)	0.80 Lower Probability Bound	Rel Error (Bayes)	Classical Pt Est	0.80 Lower Confidence Bound	Rel Error (Classical)	DT failures	OT failures
1	71.4	72.4	65.5	0.16	74.7	61.9	0.15	11.6	46.1
2	292.8	299.6	239.6	0.27	384.5	231.5	0.32	4.3	7.1
3	77.8	81	74.4	0.15	82.2	67.5	0.15	20.6	29.1
4	313.4	318.3	263.2	0.29	389.4	236.2	0.31	6.9	6.8
5	77.5	84.7	76.6	0.19	92	57.4	0.31	20.5	7.3
6	313.9	319	253.7	0.35	238.4	159.4	0.6	6.8	1.9
7	89.9	97.1	88.8	0.15	110.8	65	0.33	49.3	7
8	372.4	393.2	333.7	0.27	235.6	179.1	0.7	13.4	1.6
9	110.3	116.6	107.6	0.13	137.8	79.8	0.37	99.1	5.5
10	445.2	465	402.7	0.22	250.2	190.2	0.72	25.1	1.3

The results averaged over 1000 replications in Table 4.2 are informative of the general performance of the two models on the average, but the distribution of the

relative error is also of interest. Examining the empirical distribution of the relative error will provide additional information regarding the performance and utility of the model. Figures 4.4 - 4.7 show the empirical distribution of the relative error for the four of the test cases listed in Table 4.1. The dashed and solid lines represent the relative error distributions for the Classical and proposed Bayesian method, respectively.

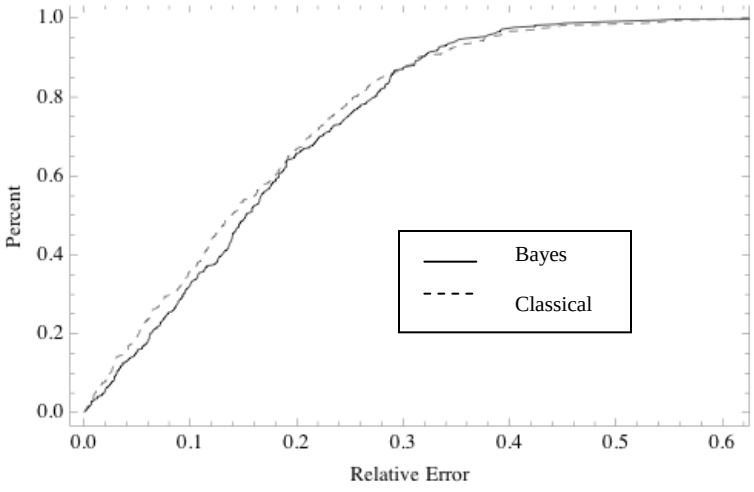


Fig. 4.4: Case 1 Empirical Distribution of Relative Error

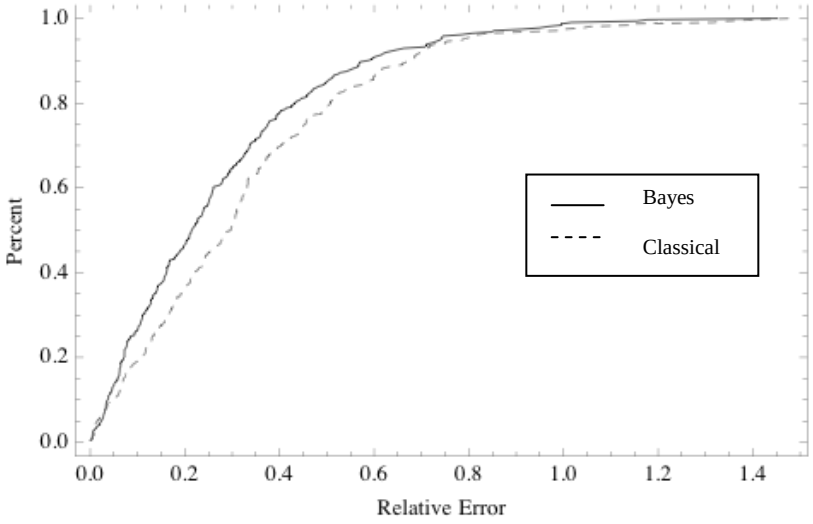


Fig. 4.5: Case 4 Empirical Distribution of Relative Error

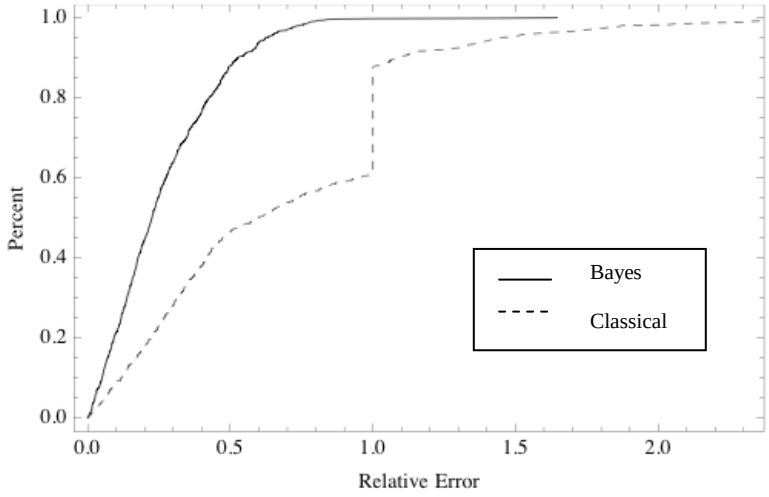


Fig. 4.6: Case 8 Empirical Distribution of Relative Error

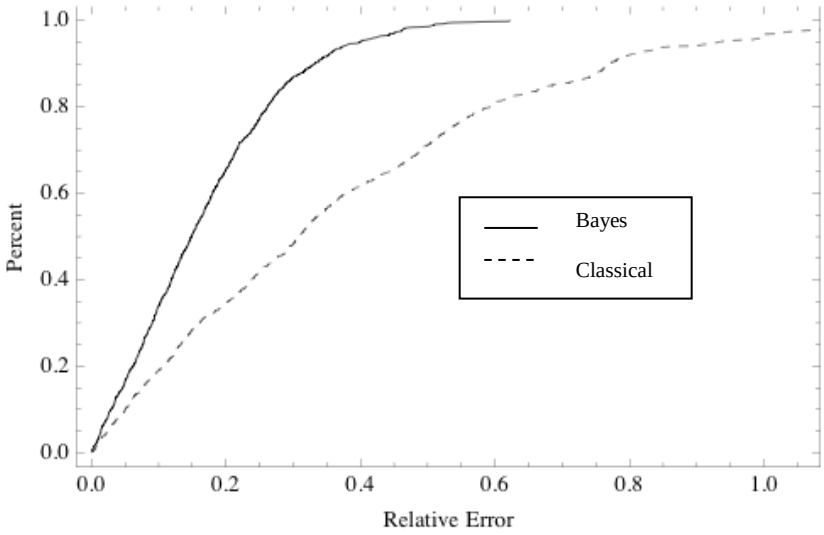


Fig. 4.7: Case 9 Empirical Distribution of Relative Error

The distributions for the relative error in Figures 4.4 and 4.5 demonstrate that the proposed Bayesian model and the classical estimate provide similar results for those cases where there is a relatively large amount of OT. Figures 4.6 – 4.7 show that the Bayesian model using the combined data from both tests provides more accurate estimates of the total system failure intensity than the classical estimate in cases where the amount of OT is limited. For example in Figure 4.6 for Case 8, the Bayesian model provides an estimate that is within 30% of the true value approximately 80% of the time, while the corresponding classical estimate is within 30% roughly 50% of the time. Expanding the relative error to 40% shows nearly 95% of the Bayesian estimates to be within this threshold, while only 65% of the classical estimates fall within this range. This behavior is to be expected though, as more information is being used in developing the combined Bayesian estimate.

Further simulation runs were also made in order to understand the sensitivity and performance of the proposed model with respect to the mean values of the

degradation parameter that are used. Table 4.3 shows the average results over 1000 replications when the chosen mean degradation is optimistic (true mean = 0.2, model input = 0.1), and Table 4.4 shows the average results over 1000 replications when the chosen mean is conservative (true mean = 0.2, model input = 0.3). As the results in tables show, the proposed Bayesian method still performs well and provides lower mean absolute relative error values in many cases even when the mean of the degradation is misspecified. This is particularly true when the amount of OT is lower than the amount of DT, further highlighting the utility of the proposed method.

TABLE 4.3
COMPARISON RESULTS FOR OPTIMISTIC MEAN DEGRADATION (TRUE MEAN = 0.2, MODEL INPUT = 0.1)

Case	True MTBF	Bayes MTBF (Post Mean)	0.80 Lower Probability Bound	Rel Error (Bayes)	Classical Pt Est	0.80 Lower Confidence Bound	Rel Error (Classical)	DT failures	OT failures
1	71.2	83.5	75.8	0.19	73.5	61	0.15	11.5	33.4
2	296.8	294.2	238.9	0.32	370.8	222.1	0.26	4.8	7.2
3	77.4	94.5	87.5	0.19	80.1	65.9	0.15	20.9	47.1
4	332.2	373.6	311.6	0.27	436.6	252.8	0.3	6.6	6.6
5	77.2	103.1	94.2	0.25	93.6	58	0.31	20.7	7.2
6	310.4	380.8	309.1	0.32	234.5	160.9	0.64	6.8	2
7	91.8	119.7	110.9	0.23	108.1	64.9	0.32	50	6.1
8	373.1	481	414.4	0.28	250.7	176.7	0.66	13.4	1.4
9	110.4	143.7	133.9	0.23	139.1	78.7	0.38	99.4	5.2
10	438.1	576	505.9	0.26	234.3	190.9	0.75	24.9	1.4

TABLE 4.4
COMPARISON RESULTS FOR CONSERVATIVE MEAN DEGRADATION
(TRUE MEAN = 0.2, MODEL INPUT = 0.3)

Case	True MTBF	Bayes MTBF (Post Mean)	0.80 Lower Probability Bound	Rel Error (Bayes)	Classical Pt Est	0.80 Lower Confidence Bound	Rel Error (Classical)	DT failures	OT failures
1	72	63.4	57.1	0.26	75.7	62.6	0.16	11.6	32.2
2	293.2	249.4	200.4	0.41	384.4	224.8	0.32	4.8	8.2
3	76.8	64.8	59.2	0.26	80.3	66	0.15	20.7	38.4
4	307.7	267.6	219.7	0.41	397.7	235.3	0.32	6.6	8.6
5	76.6	61.8	55	0.34	95.7	57.9	0.33	20.5	7.5
6	317	252.8	195	0.59	252.5	167.5	0.63	6.9	1.6
7	92	71.7	64.3	0.33	114.1	67	0.34	49.8	6.1
8	366.9	288	234.1	0.47	247.3	170.9	0.67	13.2	1.8
9	110	83.7	75.2	0.34	144.6	78	0.37	100.1	5.2
10	441.6	336	277.3	0.44	248.3	185.6	0.73	25.1	1.4

The Bayesian posterior risk provides an additional method for examining the performance of the proposed method. The use of the posterior risk places the problem of reliability assessment within a decision theoretic framework, and it naturally aligns with the decision of whether a system's reliability is sufficient or not. In this context we continue to use the squared error loss function [89]. Also, in order to rectify potential issues resulting from comparing Bayesian and classical methods within this framework, we note that the classical point estimate and confidence bounds for the exponential distribution can be derived equivalently from the Bayesian posterior distribution resulting from the use of the Jeffrey's prior distribution [90]. We also point out that this prior need not be improper. Finite bounds can be used for the support of the prior, and the posterior can be examined with respect to the limits of the chosen bounds. The development of this result is as follows.

For the Jeffrey's prior development, the distribution will be proportional to the inverse of the failure intensity [89] such that

$$p(\lambda_{OT}) \propto \frac{1}{\lambda_{OT}}. \tag{35}$$

This yields the prior distribution given by (36).

$$p(\lambda_{OT}) = \frac{1}{\lambda_{OT} \log\left(\frac{l_2}{l_1}\right)},\ l_1 < \lambda_{OT} < l_2 \tag{36}$$

The posterior distribution that results from using (36) as a prior along with the likelihood in (14) is given by

$$p(\lambda_{OT} \mid T_{OT}) = \frac{\frac{1}{\lambda_{OT} \log\left(\frac{l_2}{l_1}\right)} {\lambda_{OT}}^{n_{OT}} \exp(-\lambda_{OT} T_{OT})}{\int_{l_1}^{l_2} \frac{1}{\lambda_{OT} \log\left(\frac{l_2}{l_1}\right)} {\lambda_{OT}}^{n_{OT}} \exp(-\lambda_{OT} T_{OT}) \partial \lambda_{OT}}. \tag{37}$$

Because the desired range for the failure intensity is the positive real line, taking the appropriate limits in the denominator of (37) then yields

$$p(\lambda_{OT} \mid T_{OT}) = \frac{{\lambda_{OT}}^{n_{OT}-1} \exp(-\lambda_{OT} T_{OT})}{\Gamma(n_{OT}) \frac{1}{{T_{OT}}^{n_{OT}}}}. \tag{38}$$

The posterior distribution in (38) is a Gamma(α,β) distribution of the same form as in (2), which can then be used to develop the desired classical results.

The Bayesian posterior risk under quadratic loss for a given prior distribution is just the variance of the posterior distribution that results [89]. The posterior risk for the classical estimator is then just the variance of the Gamma distribution in (38), which is given by

$$Var(\lambda_{OT} \mid T_{OT}) = \frac{n_{OT}}{T_{OT}^{\ 2}}. \tag{39}$$

For the proposed Bayes method, the variance of the posterior was defined in (23). Examining (23) and (39), we can compare the posterior risks by varying the number of observed failures n_{OT} for fixed test length T_{OT} along with the prior information regarding the failure intensity and DT-OT degradation. The prior Gamma parameters in (23) can be developed by specifying a prior mean and variance on the failure intensity, or equivalently a prior mean and beta parameter. The prior Beta parameters can be developed by specifying the mean DT-OT degradation and using the maximum entropy methods described in Section 4.2.2. Examining the posterior risk under these conditions allows for straightforward comparison of the two methods, and it also allows for ease in further examining the sensitivity of the proposed Bayesian model to the mean of the DT-OT degradation.

Table 4.5 contains four cases for comparison. The prior Gamma parameters for the failure intensity, prior Beta parameters for the DT-OT degradation, and test length are fixed for each case. The cases correspond to a prior MTBF value of 250 hours, and the Beta prior parameters in cases 1 and 3 correspond to a mean degradation of 0.05. The Beta parameters in Cases 2 and 4 correspond to a Uniform distribution, which assumes no known prior information about the degradation. Note also that the test length is chosen to represent two separate cases: the first being where many multiples of the MTBF are available for testing, and the second being testing for only twice the value of the MTBF. Larger numbers of failures would be expected in Cases 1 and 2, and relatively few failures would be expected in Cases 3

and 4. The plots in Figures 4.8 – 4.11 show the resulting posterior risk comparisons plotted as a function of the number of failures observed during the operational test.

TABLE 4.5

POSTERIOR RISK COMPARISON CASES

Case	α	β	a	b	Test Length
1	40	0.0001	0.9	17.1	5000
2	40	0.0001	1	1	5000
3	40	0.0001	0.9	17.1	500
4	40	0.0001	1	1	500

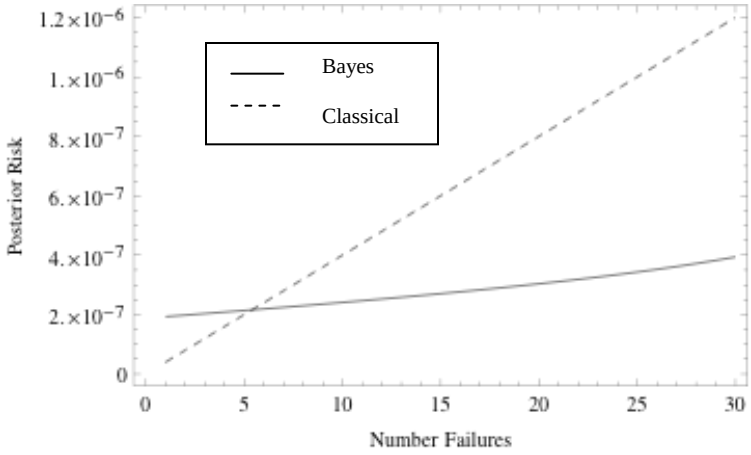


Fig. 4.8: Posterior Risk Comparison Case 1

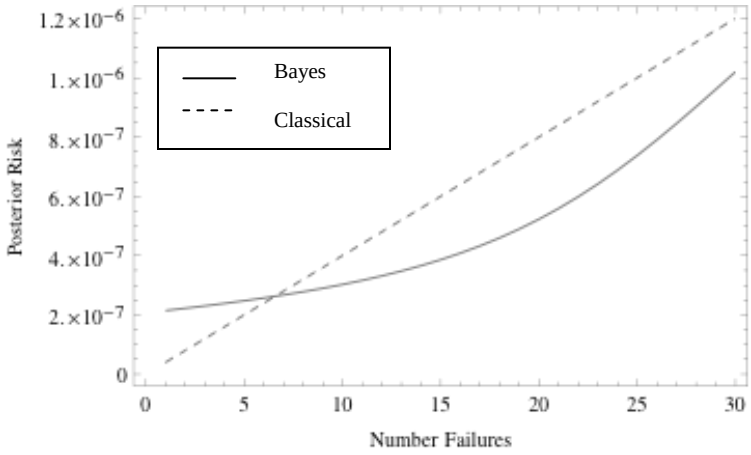


Fig. 4.9: Posterior Risk Comparison Case 2

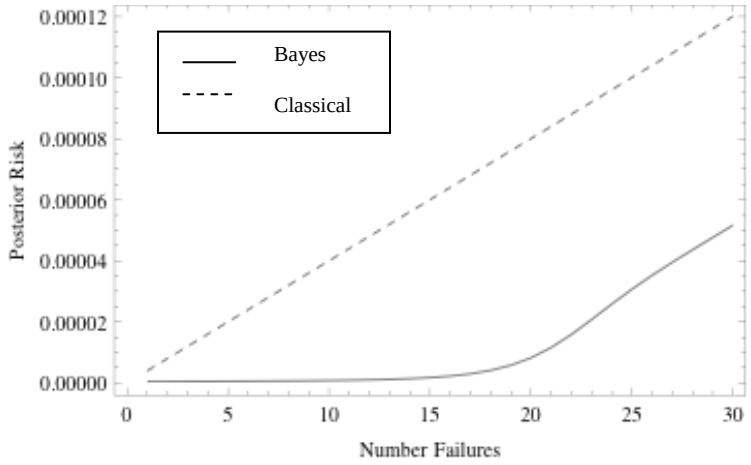


Fig. 4.10: Posterior Risk Comparison Case 3

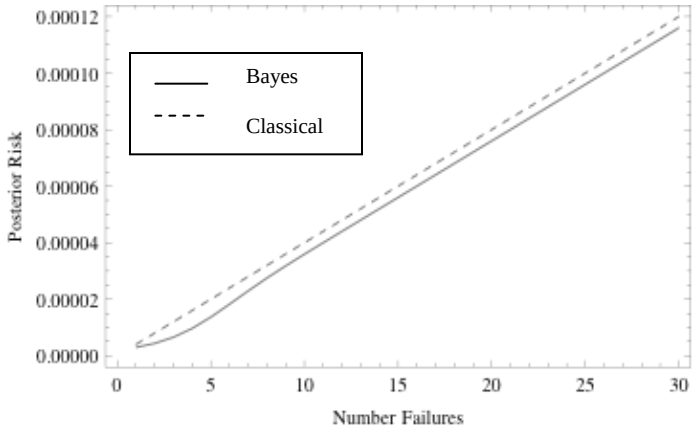


Fig. 4.11: Posterior Risk Comparison Case 4

The results in Figures 4.8 - 4.11 indicate that the proposed Bayesian method provides a generally lower variance, and therefore lower posterior risk, than the classical approach. This is especially true when examining the plots for reasonable numbers of failures that would be expected during the test lengths for each of the cases. For Cases 1 and 2 in Figures 4.8 and 4.9 respectively, the classical risk is below the proposed Bayesian model when there are approximately 6 or less failures. It is likely that more than 6 failures will be observed in a 5000 hour test when the system MTBF is 250 hours, so the proposed will be lower risk in most practical applications. The proposed Bayesian approach also provides lower risk in those cases when the mean DT-OT degradation is treated as unknown and a uniform distribution is used for the uncertainty. These results are a further indication of the utility of the proposed model for reliability assessment.

4.6 Discussion and Conclusions

This chapter presents a new reliability assessment model that allows for the combination of developmental and operational data from different test events for

continuously operating systems. The model offers an alternative to the traditional reliability assessment that is based on a single test event using only the data collected from the test phase in question. Differences in the test environments and stressors must always be considered when combining data from different test events, and the model explicitly accounts for the degradation that exists when moving from developmental to OT.

Note that the posterior distribution after OT does not explicitly use the amount of DT. It only assumes that the uncertainty on the system failure intensity prior to OT can be represented by a Gamma distribution. This increases the flexibility of the approach by allowing for the use of additional relevant reliability information. Data from lower level testing or analysis may potentially be used along with or in place of developmental reliability growth testing to develop the prior Gamma. If lower level data is used to develop the prior without system level data, it is important that a rigorous examination of the potential failure modes be completed in order to determine a reasonable value for the mean degradation between developmental and OT. For the degradation distribution itself, the use of the Maximum Entropy approach provides a repeatable framework to allow for consistency. The assessment model is developed using only the Beta approximation for prior distribution on the degradation though, and other relevant information should also be used whenever possible to develop the Beta prior on the degradation value.

Also note that the degradation in reliability between the developmental and operational tests is applied through a scaling of the system level failure intensity. While the assumption is a basic approach, it provides the necessary flexibility for the

degradation to occur in many possible ways. For instance, the DT may not have the opportunity to fully exercise certain components of the system, leading to an increase in the failure intensity for a subset of the failure modes during OT. It is also possible that the DT may fully exercise all aspects of the system, but in a more benign manner than the operational environment. The same degradation may then be realized through a smaller individual increase across a larger number of failure modes.

The approach serves as a natural extension of the current approach to reliability demonstration used in the Defense industry while explicitly modeling the additional uncertainty that exists in the problem. Use of the posterior distribution in (20) and (23) along with the probability of a successful test defined by (29) will generally lead to tighter uncertainty intervals and result in lower reliability design goals. This approach can help to directly reduce the programmatic risks that may exist due to reliability demonstration in a constrained environment with operational test data alone.

5 RELIABILITY GROWTH PLANNING USING COMBINED DEVELOPMENTAL AND OPERATIONAL TEST DATA FOR RELIABILITY DEMONSTRATION

5.1 Introduction

5.1.1 Background

This chapter presents an approach to reliability growth planning that serves to mitigate the problems associated with the high reliability goals resulting from traditional reliability demonstration. The proposed method uses a combination of developmental reliability growth test data and operational reliability test data for reliability demonstration within a Bayesian probabilistic framework. Demonstration in this manner was developed in [6]. Degradation between the developmental and operational test phases is accounted for through the probabilistic application of a scale factor. The Bayesian formulation that is used easily allows for uncertainty to be included on the scale factor. This provides an approach that will more closely match the practical situation, in which limited information on the actual degradation may be available. The reliability growth portion of the method is based on the projection model developed in Chapter 3, which is an extension of the models found in [18] and [26].

5.1.2 Chapter Overview

The methodology of the approach is presented in Section 5.2. It includes model assumptions, data requirements, and the development of the initial framework

for combining data from two different tests. The presentation of the methodology is broken into two separate sections: developmental reliability growth assessment, and operational reliability assessment using prior reliability growth assessment results. Section 5.3 presents three associated management metrics that can be helpful in managing a reliability growth program. An example application of the model is presented in Section 5.4, and conclusions are found in Section 5.5.

5.2 Methodology

The methodology is presented in three parts. The first considers the construction of a reliability growth planning curve for the developmental test program. The reliability growth program assumes that the reliability of the system will be increasing due to the discovery and subsequent correction of failure modes inherent in the initial configuration of the system. The second considers reliability demonstration using a combination of developmental and operational test data for a single system while accounting for differences in the underlying failure intensity that may exist between the two test events. The operational test is assumed to be a constant configuration test with a more realistic test environment that causes a difference in the underlying system reliability. Results from the reliability growth portion are used to explicitly model the uncertainty present in the failure intensity, and the difference in the two test environments is modeled explicitly with a probabilistic decrease in reliability (or increase in failure intensity) between the test phases. The third part develops the actual planning curve by combining the results from the first two parts.

5.2.1 Data Requirements

The reliability growth planning model presented here utilizes the same inputs as current models that are found in [13]. The data requirements are:

1. Initial MTBF, M_I
2. MTBF Requirement, M_R
3. Average FEF, μ_d
4. MS
5. Average reliability degradation between developmental and operational test environments, ε
6. Test lengths for planned developmental and OT

The initial MTBF is the mean of the prior distribution in this approach. While it is commonly specified in currently used models, its use in a Bayesian setting may prove to be too informative for some applications. In these cases it is recommended that any initial reliability growth planning curve be updated and reinitialized after the first phase of DT. Empirical Bayes estimates from Chapter 3 can be applied to the data from the test and then utilized in the planning curve construction presented in this chapter.

5.2.2 Modeling Reliability Growth in DT

The assumptions for the reliability growth testing mimic those in Chapters 3 and 4 and are as follows:

1. A large number of failure modes (*K*) exist in the system.
2. Failure modes generate failures independently of one another.
3. Each occurrence of a failure mode results in a system failure.
4. The failure intensity for a given failure mode is constant both before and after a corrective action is implemented.
5. No new failure modes are induced by corrective actions.
6. The resulting failure intensity after corrective action will be reduced from the initial value according to the assigned FEF.
7. Failure mode failure intensities have a common prior Gamma distribution.

For the failure intensity of each failure mode, we use a Gamma(α,β) parameterized as

$$p(\lambda_i) = \frac{\lambda_i^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} \exp\left(-\frac{1}{\beta}\lambda_i\right), \ \alpha > 0, \ \beta > 0 \qquad (1)$$

The system level failure intensity can then be approximated by a Gamma distribution with mean and variance given in (2) and (3). We use the notation λ_{DT} to denote the system level failure intensity from DT and distinguish between later assessments of the desired operational failure intensity. Because the results in Chapters 3 and 4 make no distinction between A-modes and B-modes, we also substitute λ_I in place of λ_B to facilitate later classification of different types of failure modes. The variable *n* refers to the vector containing the number of failures for each of the *m* observed failure modes.

$$E\left[\lambda_{DT} \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T_{DT} - v_i)\right)}\right) + \frac{\lambda_I}{1+\beta T_{DT}} \tag{2}$$

$$Var\left[\lambda_{DT} \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)^2 n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T_{DT} - v_i)\right)^2}\right) + \frac{\lambda_I}{\beta\left(\frac{1}{\beta} + T_{DT}\right)^2}. \tag{3}$$

In this chapter we extend these results for reliability growth planning by making three additional assumptions that are common in reliability growth planning. First, we assume that a portion of the observed failure modes will be considered as A-modes, which are defined as failure modes that will not receive corrective actions when observed during testing. The failure intensity due to A-modes is defined from the inputs as

$$\lambda_A = (1 - MS)\lambda_I \tag{4}$$

B-modes are those failure modes that will be corrected when observed during the reliability growth process. λ_B is defined from the inputs as

$$\lambda_B = MS\lambda_I \tag{5}$$

Note that $\lambda_A + \lambda_B = \lambda_I$. We next assume that all failure modes will be corrected with an average level of fix effectiveness, denoted as μ_d. All corrective actions are also assumed to be delayed until the end of the test length T_{DT}.

To develop the planning model, we use assumption 4 and marginalize with respect to the observed number of failures during testing of length T_{DT}. This is equivalent to substituting the expected number of failures for the observed number of

failures in (2) and (3). These modifications result in the updated expressions for the mean and variance given in (6) and (7).

$$E\left[\lambda_{DT}\right] = \frac{\lambda_A T_{DT}}{\left(\frac{1}{\beta} + T_{DT}\right)} + \frac{\left(1-\mu_d\right)\lambda_B T_{DT}}{\left(\frac{1}{\beta} + T_{DT}\right)} + \frac{\lambda_I}{1+\beta T_{DT}}$$
$$= \frac{\lambda_A T_{DT}}{\left(\frac{1}{\beta} + T_{DT}\right)} + \left(1-\mu_d\right)\left(\lambda_B - \frac{\lambda_B}{1+\beta T_{DT}}\right) + \frac{\lambda_I}{1+\beta T_{DT}} \tag{6}$$

$$Var\left[\lambda_{DT}\right] = \frac{\lambda_A T_{DT}}{\left(\frac{1}{\beta} + T_{DT}\right)^2} + \frac{\left(1-\mu_d\right)^2\lambda_B T_{DT}}{\left(\frac{1}{\beta} + T_{DT}\right)^2} + \frac{\lambda_I}{\beta\left(\frac{1}{\beta} + T_{DT}\right)^2}. \tag{7}$$

The expression in (6) is composed of three main parts. The first is the failure intensity due to observed failure modes that will not be corrected. The second is the remaining failure intensity for failure modes that have been observed during the test and corrected with the average level of fix effectiveness. The third is the remaining failure intensity due to failure modes that have not yet been observed during the test. When no A-modes are considered, this expression is also identical to the idealized reliability growth planning curve for B-modes presented in [17]. When both A-modes and B-modes are considered, the difference between the results here and in [17] is due to the treatment of the A-modes. The approach presented here assumes all failure modes have the same common prior Gamma distribution, whereas the approach in [17] assumes a common Gamma for the B-modes and treats the A-modes separately. Although the approach in [17] is not explicitly Bayesian, the use of a Gamma distribution to model variability in the failure intensity from mode to mode

aligns closely with the use of a Gamma prior distribution in Chapter 3. The methodology underlying [17] also uses shrinkage estimation based on mean-squared-error, which provides an additional connection to Bayesian results assuming squared error loss. We also note that the equations represent planned reliability growth when no data are yet available. As such the mean and variance are no longer conditioned on the observed failures, *n*.

The idealized reliability growth planning curve can also be viewed as a function of the DT test time T_{DT}, allowing for (6) and (7) to be written with arbitrary time *t* as

$$E\left[\lambda_{DT} \mid t\right] = \frac{\lambda_A t}{\left(\frac{1}{\beta} + t\right)} + \left(1 - \mu_d\right)\left(\lambda_B - \frac{\lambda_B}{1 + \beta t}\right) + \frac{\lambda_I}{1 + \beta t} \tag{8}$$

$$Var\left[\lambda_{DT} \mid t\right] = \frac{\lambda_A t}{\left(\frac{1}{\beta} + t\right)^2} + \frac{\left(1 - \mu_d\right)^2 \lambda_B t}{\left(\frac{1}{\beta} + t\right)^2} + \frac{\lambda_I}{\beta\left(\frac{1}{\beta} + t\right)^2}. \tag{9}$$

For the distribution of λ_{DT}, results in Chapter 3 show that the system level failure intensity is well approximated by a Gamma distribution when the corrective actions are applied in an arbitrary fashion throughout or after the test. The approximation is not necessary for planning though. When all corrective actions are delayed until the end of the test with the same average level of fix effectiveness, the system level failure intensity will be exactly Gamma distributed. We then have

$$\lambda_{DT} \sim Gamma\left[\tilde{\alpha},\ \tilde{\beta}\right], \tag{10}$$

where the Gamma parameters can be defined using the mean and variance as in (11) and (12).

$$\tilde{\alpha} = \frac{E\left[\lambda_{DT} \mid t\right]^2}{Var\left[\lambda_{DT} \mid t\right]} \tag{11}$$

$$\tilde{\beta} = \frac{Var\left[\lambda_{DT} \mid t\right]}{E\left[\lambda_{DT} \mid t\right]} \tag{12}$$

5.2.3 Overview of Reliability Demonstration Using Combined Developmental and Operational Test Data

The reliability demonstration approach presented here is the same as that developed in Chapter 4, but small modifications specific to reliability growth planning are necessary. When combining data from DT and OT, degradation in reliability is generally assumed between the two test events [13]. This degradation is traditionally considered in terms of a decrease in the system MTBF. Assuming $100\gamma\%$ degradation in the MTBF (or a corresponding increase in system failure intensity) leads to the relationship between the developmental and operational failure intensities shown in (13) and (14). The DT and OT subscripts denote the corresponding MTBF and failure intensity values.

$$MTBF_{OT} = (1-\gamma) MTBF_{DT} \tag{13}$$

$$\lambda_{DT} = (1-\gamma)\lambda_{OT} \tag{14}$$

Traditional techniques for reliability growth planning [13] treat the failure intensity from DT as a deterministic value, while also assuming a deterministic value for the γ parameter in (14). The approach presented here and in Chapter 4 extends these concepts by considering the uncertainty that is present for both of these parameters. The uncertainty can then be modeled explicitly when calculating the statistical risks

of the demonstration test. From the Gamma distribution defined by (11) and (12) in Section 5.2.2, we have

$$\lambda_{DT} = (1-\gamma)\lambda_{OT} \sim Gamma\left[\tilde{\alpha},\ \tilde{\beta}\right]. \tag{15}$$

Because the failure intensity, λ_{OT}, is the true parameter of interest, we condition on γ and utilize properties of the Gamma distribution [86] to obtain the required distribution.

$$\lambda_{OT} \mid \gamma \sim Gamma\left[\tilde{\alpha}, \frac{\tilde{\beta}}{(1-\gamma)}\right]. \tag{16}$$

From Chapter 4 the uncertainty on the degradation parameter γ can also be modeled using maximum Entropy principles [87],[88] to arrive at an approximate Beta(*a*,*b*) distribution. We assume an average degradation value for planning purposes, which can be determined by examining historical performance on similar systems or examining the potential failure modes that exist in the system in developmental and operational test environments. Maximizing the entropy subject to the assumed mean value of the MTBF degradation γ and a range of (0,1) results in the prior distribution for γ being a truncated Exponential distribution given by

$$p(\gamma) = \frac{\mu \exp(-\mu\gamma)}{1-\exp(-\mu)}, \tag{17}$$

where μ is the solution to

$$\frac{1}{\mu} - \frac{\exp(-\mu)}{1-\exp(-\mu)} = \varepsilon \tag{18}$$

for mean degradation value ε. The Beta parameters can then be found by equating the means and second moments of the two distributions, which results in the system of equations given by (19) and (20).

$$\frac{a}{a+b} = \frac{1}{\mu} - \frac{\exp(-\mu)}{1-\exp(-\mu)} \tag{19}$$

$$\left(\frac{a}{a+b}\right)\left(\frac{a+1}{a+b+1}\right) = \frac{\exp(-\mu) + 2\left[-\frac{1}{\mu}\exp(-\mu) - \frac{1}{\mu^2}\exp(-\mu) + \frac{1}{\mu^2}\right]}{1-\exp(-\mu)} \tag{20}$$

Equation (18) will result in $\mu = 0$ when the mean degradation is set to 0.5. For planning purposes the mean degradation is not generally known with high precision, and this difficulty is easily overcome by slightly perturbing the mean degradation to obtain a non-zero solution. The posterior mean failure intensity from Chapter 4 is

$$E\left[\lambda_{OT} \mid n_{OT}\right] = \left(\frac{\tilde{\alpha} + n_{OT}}{\frac{1}{\tilde{\beta}} + T_{OT}}\right) \frac{{}_2F_1\left[\tilde{\alpha} + n_{OT} + 1, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}} + T_{OT}}\right]}{{}_2F_1\left[\tilde{\alpha} + n_{OT}, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}} + T_{OT}}\right]} \tag{21}$$

where ${}_2F_1(a,b,c,z)$ is the integral form of the hypergeometric function given by

$${}_2F_1(a,b,c,z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt. \tag{22}$$

The function given in (22) can be evaluated using standard numerical procedures. The posterior variance can be developed similarly as

$$Var\left[\lambda_{OT} \mid n_{OT}\right] = \left(\frac{\left(\tilde{\alpha}+n_{OT}\right)\left(\tilde{\alpha}+n_{OT}+1\right)}{\left(\frac{1}{\tilde{\beta}}+T_{OT}\right)^2}\right)\frac{{}_2F_1\left[\tilde{\alpha}+n_{OT}+2, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+T_{OT}}\right]}{{}_2F_1\left[\tilde{\alpha}+n_{OT}, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+T_{OT}}\right]} - \left[\left(\frac{\left(\tilde{\alpha}+n_{OT}\right)}{\left(\frac{1}{\tilde{\beta}}+T_{OT}\right)}\right)\frac{{}_2F_1\left[\tilde{\alpha}+n_{OT}+1, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+T_{OT}}\right]}{{}_2F_1\left[\tilde{\alpha}+n_{OT}, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}}+T_{OT}}\right]}\right]^2 \quad (23)$$

The posterior distribution is again well approximated by a Gamma distribution. The distribution described by (21) and (23) can then be used to construct the reliability growth planning curve, which is presented in Section 5.2.4.

The posterior distribution after OT only assumes that the prior distribution on the system failure intensity can be represented by a Gamma distribution. This increases the flexibility of the approach by allowing for the use of additional relevant reliability information for assessment, and it also provides necessary flexibility for reliability growth planning. A number of reliability improvement activities are likely to occur concurrently with system level testing, the combination of which will help to mature the system to its end reliability target. The planning curve is merely a roadmap though, and in practice growth will not usually follow the curve exactly throughout the entire developmental program. This does not negatively impact planning models in general, as it is important only that the system achieves the

targeted posterior distribution represented by the end of planning curve. As mentioned in Chapter 4, the proposed approach only requires that the system ends its developmental growth program with sufficiently mature reliability and a failure intensity that can be represented by the appropriate Gamma distribution.

5.2.4 Constructing the Reliability Growth Planning Curve

Construction of the reliability growth planning curve begins by considering the statistical risks associated with the reliability demonstration. The consumer risk, or probability of a system with insufficient reliability succeeding in the demonstration test, can be mitigated by considering the Gamma distribution defined by (21) and (23). The distribution describes the uncertainty that is present in the operational assessment of the failure intensity and corresponding MTBF, and it is desirable to have a small probability that the MTBF of the system is below that of the MTBF requirement. For desired consumer risk α, define $\hat{\lambda}_{1-\alpha}$ as the $(1-\alpha)^{th}$ percentile of the Gamma distribution in (21) and (23). This leads to the desired inequality given by

$$\hat{\lambda}_{1-\alpha} \leq \frac{1}{M_R}. \tag{24}$$

The left side of the inequality in (24) is just a percentile of the Gamma distribution from (21) and (23), and closer observation reveals this term to be a function of two unknown parameters: the number of failures observed in the operational test, n_{OT}, and the underlying β parameter. The inequality in (24) is equivalent to requiring that the $(1-\alpha)^{th}$ percentile of the MTBF is greater than or equal to the system's MTBF requirement. Figure 5.1 shows an example of the relationship for a consumer risk of 0.20.

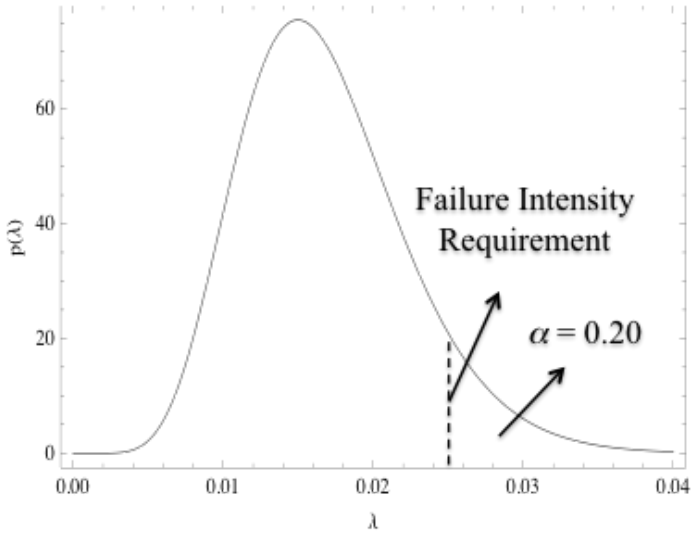


Fig. 5.1: Graphical Representation of Consumer Risk ($\alpha = 0.20$)

The producer risk is defined as the probability of a system with sufficient reliability not succeeding in the demonstration test. The probability of a successful test and the associated producer risk defined in Chapter 4 also address the uncertainty associated with operational assessment of the failure intensity and corresponding MTBF. The uncertainties that are present for the system failure intensity and the associated degradation factor between the DT and the demonstration test are both considered explicitly. It is also desirable to mitigate the producer risk in the reliability growth planning curve development. For desired producer risk 1-*p* this yields the inequality in (25).

$$p \le \sum_{i=0}^{n_{OT}} \left(\frac{T_{OT}{}^{i}}{i!} \right) \frac{\Gamma(a+b)\Gamma(\tilde{\alpha}+i)\Gamma(b+\tilde{\alpha})}{\Gamma(b)\Gamma(\tilde{\alpha})\Gamma(a+b+\tilde{\alpha})\tilde{\beta}^{\tilde{\alpha}} \left(\frac{1}{\tilde{\beta}} + T_{OT} \right)^{\tilde{\alpha}+i}} {}_2F_1 \left[\tilde{\alpha}+i, a, a+b+\tilde{\alpha}, \frac{\frac{1}{\tilde{\beta}}}{\frac{1}{\tilde{\beta}} + T_{OT}} \right]. \qquad (25)$$

The right side of the inequality in (25) is the probability of a successful test, and the left side is the probability of a successful test resulting from the desired producer risk. The inequality therefore forces the reliability growth plan to have a probability of a successful demonstration test at least as high as the desired value, resulting in a producer risk that is at most the desired value. Closer examination of (25) also reveals it to be a function of two unknown parameters: the number of failures observed in the operational test, n_{OT}, and the underlying β parameter. Simultaneously solving the inequalities in (24) and (25) will yield the two unknown parameters. There are a number of possible solutions that can be found, but the pair with the smallest value of β and/or largest n_{OT} can be shown to provide the lowest risk. This is because the β value is indicative of the steepness of the resulting reliability growth planning curve, and smaller values will indicate less aggressive reliability growth.

The solutions from (24) and (25) may appear at first to be a dramatic change from the current approach with classical OC curves [13]. The models currently used for reliability growth planning also appear on the surface to use only operational test data in the operational reliability assessment and corresponding statistical risk calculations. But the use of the degradation factor approach in (13) and (14) is merely an implicit approach for using the developmental test results directly in the risk calculations. The deterministic treatment of both the failure intensity and the degradation in these calculations would also appear to be undesirable. Considering the uncertainty that is actually present leads directly to the inequalities in (24) and (25), with the additional benefit of lower reliability design goals. The approach not only treats the problem within a more complete analytic framework, but it also helps

to directly reduce the programmatic risks that may exist due to statistical demonstration in a constrained environment.

When the (β, n_{OT}) values are found from the inequalities, the idealized reliability growth planning curve can be constructed using the mean in (6). The idealized curve represents the growth in reliability that would occur if corrective were implemented for failure modes immediately after they were observed and prior to further testing. In practice repairs are made to the system and testing is continued, with corrective actions implemented after the test itself in time periods referred to as CAPs. This allows for more efficient testing while also providing the time for root-cause-analysis that is necessary for robust corrective actions. With this in mind, the reliability targets for each successive phase of DT can then be constructed using the idealized curve. The steps for each of the test phases are taken as the value of the idealized curve at the start of each test. Note that this also allows for lag times for corrective actions to be employed as in [17]. Not all failure modes may be corrected prior to the start of the next test phase, so the appropriate point on the idealized growth curve can be chosen to represent any desired lag times.

5.3 Management Metrics

A number of useful reliability growth planning metrics can also be developed within the proposed model. They include the number of B-modes surfaced in testing, the failure intensity for B-modes not yet observed during testing, and the fraction of the initial failure intensity attributed to B-modes already observed in testing. The metrics are comparable to those found in [17], and they also include the uncertainty distribution that results from the Bayesian approach. They are specific to B-modes

because the B-modes are the failure modes that will require root cause analysis and corrective action development. These activities will require additional resources in the form of manpower and time, and the metrics are therefore useful to managers looking to properly resource their reliability growth programs. These results can be easily extended for only A-modes or all failure modes by substituting λ_A or λ_I respectively, in place of the λ_B.

5.3.1 Number of Modes Surfaced in Testing

The number of B-modes surfaced in testing can be developed using the same approach as the prior predicted cumulative number of failure modes presented in Chapter 3. First define the indicator function

$$I_i(t) = \begin{cases} 1, \text{ mode i occurs by time } t \\ 0, \text{ otherwise} \end{cases}. \tag{26}$$

The mean of $I_i(t)$ is found by examining the probability that an unobserved failure mode is observed by some time t. The likelihood of observing a B-failure mode by time t is given from Assumption 4 in Section 5.2.2 as

$$p\left(I_i(t) = 1 \mid \lambda_B\right) = 1 - e^{-\lambda_B t} \tag{27}$$

Using the prior Gamma distribution on the mode failure intensity, the unconditional marginal distribution for $I_i(t)$ can be found through standard techniques. The unconditional expected value is given by

$$E\left[I_i(t)\right] = p\left(I_i(t) = 1\right) = 1 - \frac{1}{(1+\beta t)^{\alpha}}. \tag{28}$$

Summing over all K modes in the system yields

$$E\left[\sum_{i=1}^{K} I_i(t)\right] = K\left[1 - \frac{1}{(1+\beta t)^{\alpha}}\right], \tag{29}$$

and taking the limit as K becomes large results in

$$\lim_{K\to\infty} E\left[\sum_{i=1}^{K} I_i(t)\right] = \frac{\lambda_B}{\beta}\log[1+\beta t], \tag{30}$$

where λ_B is defined by (5). Because $I_i(t)$ is a Bernoulli random variable, summing over all failure modes yields a Binomial random variable, and taking the limit as K becomes large yields a Poisson random variable with the mean shown in (30). Therefore if we denote m as the number of B-modes observed by time t, m will be Poisson distributed with mean

$$E[m] = \frac{\lambda_B}{\beta}\log(1+\beta t). \tag{31}$$

5.3.2 Failure Intensity for Unobserved Failure Modes

From the results in Section 5.2.2 the failure intensity for unobserved failure modes at test time t, $\lambda_{B,unobserved}$, is just the third part of the expressions developed in (6) and (7). This yields the mean and variance in (32) and (33), with λ_B again defined by (5).

$$E\left[\lambda_{B,Unobserved} \mid t\right] = \frac{\lambda_B}{1+\beta t} \tag{32}$$

$$Var\left[\lambda_{B,Unobserved} \mid t\right] = \frac{\lambda_B}{\beta\left(\frac{1}{\beta}+t\right)^2}. \tag{33}$$

5.3.3 Fraction of Initial Failure Intensity Attributed to Observed Failure Modes

The formula for the fraction of the initial B-mode failure intensity attributed to observed failure modes at test time t, $\theta(t)$, is discussed in [17] and can be represented as

$$\theta(t) = \frac{\sum_{i=1}^{m} \lambda_i}{\lambda_B}, \tag{34}$$

where λ_B is defined by (5). The mean of (34) can be expressed as

$$E[\theta(t)] = E\left[\frac{\sum_{i=1}^{m} \lambda_i}{\lambda_B}\right] = \frac{\sum_{i=1}^{m} \frac{\alpha + n_i}{\frac{1}{\beta} + t}}{\lambda_B}. \tag{35}$$

Results in Chapter 3 show that as $K \to \infty$, $\alpha \to 0$ and using similar methods as in Section 5.2.2 yields

$$\lim_{K \to \infty} E[\theta(t)] = \frac{\sum_{i=1}^{m} \frac{n_i}{\frac{1}{\beta} + t}}{\lambda_B} = \frac{\beta t}{1 + \beta t} \tag{36}$$

The variance follows similarly, and is found to be

$$Var[\theta(t)] = \frac{\beta^2 t}{(1 + \beta t)^2}. \tag{37}$$

Because the failure modes are assumed to be independent by Assumption 2 in Section 5.2.2, the distribution will be asymptotically Normal via the Central Limit Theorem.

5.4 Example Application

This section provides an example application of the proposed planning model, along with a comparison with the standard reliability growth planning model [17] using the same inputs. The main difference between the two models is the demonstration test assessment and the associated statistical risks that result. The model proposed here explicitly considers the uncertainty that is present, while the traditional approach using PM2 relies on the classical OC curve analysis and treats both the failure intensity and the DT-OT degradation parameter as known values. As described in Section 5.2.4, all corrective actions are assumed to occur in CAPs between the developmental test phases. In order to simplify the example, no corrective action lag times are assumed, although they can easily be included in both the proposed model and PM2. The inputs for the example follow those in Section 5.2.1 and are as follows:

1. M_I = 180 hours
2. M_R = 220 hours
3. μ_d = 0.7
4. MS = 0.95
5. ε = 0.2 (mean DT-OT degradation)
6. $T_{DT,1}$ = 1000 hours, $T_{DT,2}$ = 1500 hours, $T_{DT,3}$ = 1000 hours, $T_{DT,4}$ = 1500 hours, $T_{DT,5}$ = 1000 hours, $T_{DT,6}$ = 1500 hours
7. $T_{DT} = \sum_{i=1}^{6} T_{DT,i}$ = 7500 hours

8. T_{OT} = 2000 hours
9. Consumer risk = 0.2
10. Producer risk = 0.3

Numerical methods can be used to determine the (β, n_{OT}) pair that satisfies the desired producer and consumer risk values. Multiple solutions to the inequalities in (24) and (25) are possible, but a contour plot can be used to direct the numerical method that is used to gain the solution. Figure 5.2 shows a contour plot for strict equality in (24) and (25).

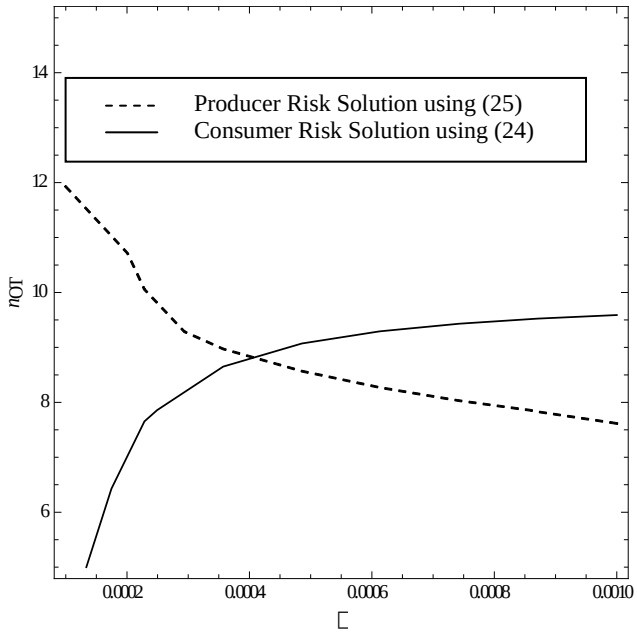


Fig. 5.2: Contour Plot for Solutions to (24) and (25)

The value of n_{OT} must be an integer, and the contour plot indicates that the desired solution lies somewhere near n_{OT} = 9 and β = 0.0004. More precise numerical examination of the inequalities results in β = 0.00046 and n_{OT} = 9 as the solution for the example.

The idealized planning curve that results from using these values is shown in Figure 5.2. Note that the plot in Figure 5.3 only depicts the idealized curve and the associated reliability steps for each planned test phase. It is also possible to plot upper and lower probability bounds for both the idealized curve and the steps by using the Gamma distribution defined in (10)-(12). The probability bounds are left off of the example in Figure 5.3 in order to simplify comparisons between the two approaches.

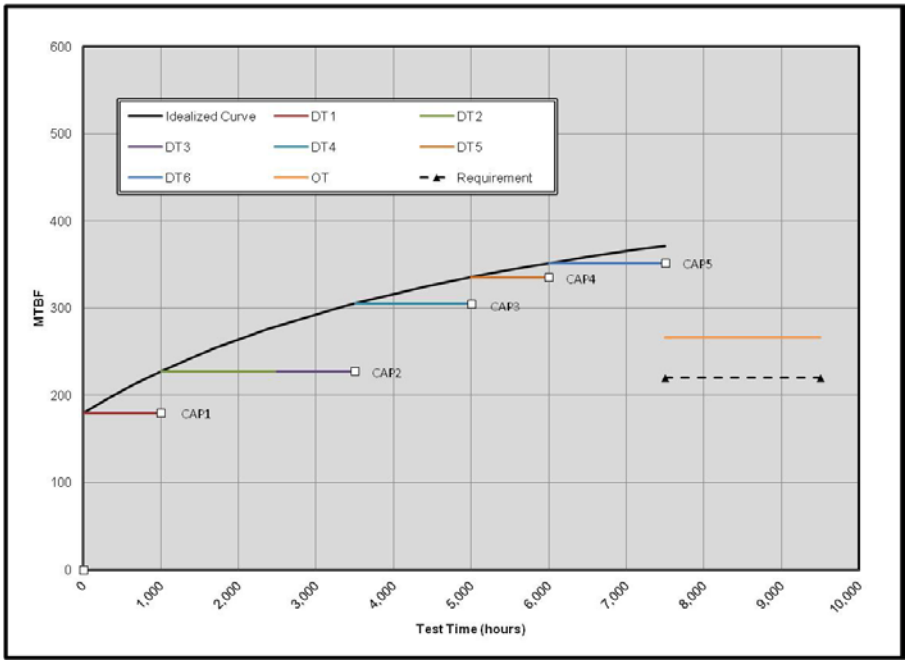


Fig. 5.3: Proposed Bayesian Planning Curve for Example Case

The corresponding curve for the PM2 model is shown in Figure 5.4. The right side of each curve provides an immediate comparison between the traditional OC curve approach and the combined DT-OT demonstration in the Bayesian model. For the MTBF requirement of 220 hours, the idealized planning curve at the end of the

DT period is approximately 372 hours in the Bayesian model. The corresponding point in PM2 is 462 hours. The estimated value for the demonstration test, or initial operational test (IOT), is 267 hours in the Bayesian model and 370 hours in PM2. These values are used to set reliability design targets in the development of new systems, and the differences indicate a significant reduction in the margin that is needed in order to demonstrate the MTBF requirement.

Further comparison reveals that the initial MTBF in the Bayesian model is 180 hours versus 220 hours in PM2. These values are model inputs, but they both correspond to a goal MTBF to growth potential MTBF ratio of approximately 0.70. The growth potential MTBF, M_{GP}, is a function of Mi, MS, and μ_d and is defined as

$$M_{GP} = \frac{M_i}{1 - MS * \mu_d} \tag{38}$$

This ratio provides an indication of risk associated with the overall reliability growth goal. Attempts to grow too near the reliability growth potential often involve significant risk. This is due to diminishing returns in the test-fix-test reliability growth process, making growth more difficult once the initial dominant failure modes are observed and mitigated. The initial MTBF value is also important, as it can be used to drive DFR activities that are to be completed before full system-level testing begins. These activities should be conducted in order to improve the reliability prior to system-level testing, helping the program to achieve the initial targets on the reliability growth planning curve. Achieving the first step is vitally important, as low initial MTBF values will in turn lower the growth potential MTBF in (38) and introduce significant risk to the overall reliability program. The Bayesian model allows for the initial MTBF to be 40 hours lower than in PM2 while maintaining the

same ratio between the goal MTBF and growth potential MTBF. This result helps to directly reduce the risk associated with achieving the initial target on the planning curve. While both models are inherently using the same prior information through the specification of the initial MTBF and the solution for the β parameter, the relative strength of the prior assumptions may be cause for concern in practice. In these cases it is recommended that the planning curve be redeveloped after the first phase of testing is completed. The empirical Bayes procedures developed in Chapter 3 can be used to estimate the λ_B and β parameters directly from the failure data collected during the test. These estimates can then be used directly to reinitialize the original curve based on actual system performance. This will provide a more realistic indication of the expected future reliability performance during the DT program. While not presented in this example, reduced test lengths are also a potential benefit of the methodology developed here. For a fixed developmental reliability target, the amount of testing necessary to demonstrate a reliability requirement at a given consumer risk value would be reduced when compared the traditional OC curve application.

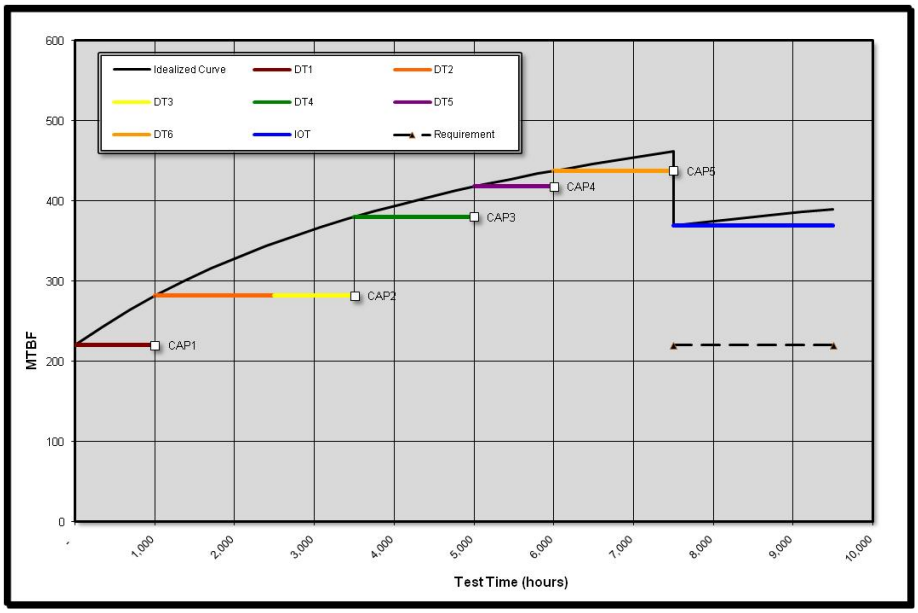


Fig. 5.4: Planning Curve for Example Case Using PM2 Model

5.5 Conclusions

This chapter has presented a new reliability growth planning model that allows for the combination of developmental and operational data from different test events for reliability demonstration. The approach serves as a natural extension of current reliability growth planning models used in the Defense industry while explicitly modeling the additional uncertainty that exists in the problem. The uncertainty that is present in the failure intensity at the end of DT is modeled along with the planned degradation that occurs between the developmental and operational test environments. Considering the additional information along with its uncertainty allows for tighter uncertainty intervals and lower reliability design goals, which are a significant benefit of the proposed approach. The approach considers reliability growth planning within a more complete analytic framework while also helping to

directly reduce the programmatic risks that may exist due to statistical demonstration in a constrained environment.

6 DEVELOPMENT OF PRIOR INFORMATION USING LOWER-LEVEL DATA SOURCES

6.1 Introduction

6.1.1 Background

The reliability growth models in Chapters 3-5 describe the reliability of the system using data collected through complete system-level testing. These data are perhaps the most desirable for assessing system reliability, but during early stages of development this type of information may be unavailable. In early stages of development of the system, there are often other sources of reliability related information that are available, and in these cases it is possible to utilize these information sources to develop an early assessment of the system reliability. When viewing the process of reliability assessment across the various stages of development of the system, assessment in this manner serves as prior information that can be updated with the reliability growth models in Chapters 3 and 4 when system-level test data become available. The approach also serves as a means of connecting early engineering activities involving component modeling and characterization with the full system-level testing and reliability growth modeling that occurs later in the system's life cycle.

6.1.2 Chapter Overview

Section 6.2 of this chapter discusses the decomposition of the system into subsystems, components, and redundant blocks. It also draws connections between

failures in the system and the failure modes that are used in the models in Chapters 3 and 4. Section 6.3 develops the posterior distributions for individual failure modes. The use of component/subsystem data and physics-based model results is discussed, along with probabilistic scaling to account for additional failure mechanisms that may be present in the operational environment. Conclusions are presented in Section 6.4.

6.2 System Level decomposition

The assumptions for the reliability growth models in Chapters 3 and 4 are as follows:

1. The system is comprised of a large number of failure modes that are serial in nature; the occurrence of any failure mode results in failure of the system.
2. Failure modes generate failures independently of one another.
3. The failure intensity, or rate of occurrence of failure, for each mode is constant both before and after a corrective action is implemented.
4. The resulting failure intensity after corrective action will be reduced from the initial value according to the assigned FEF.
5. Corrective actions to failure modes do not introduce new failure modes into the system.

As shown in Chapter 3, these assumptions along with a common prior Gamma distribution for all of the failure modes in the system lead to a posterior distribution on the projected system failure intensity. The mean and variance of the posterior are given by

$$E\left[\lambda_{DT} \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T_{DT} - v_i)\right)}\right) + \frac{\lambda_B}{1+\beta T_{DT}} \qquad (1)$$

$$Var\left[\lambda_{DT} \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)^2 n_i}{\left(\frac{1}{\beta} + v_i + (1-d_i)(T_{DT} - v_i)\right)^2}\right) + \frac{\lambda_B}{\beta\left(\frac{1}{\beta} + T_{DT}\right)^2}. \qquad (2)$$

Examination of the expressions in (1) and (2) shows them to be a function of the prior parameters λ_B and β. The parameters describe the prior distribution of the failure intensity of the complex system. Empirical Bayes procedures were developed in Chapter 3 to provide estimates of the parameters using the system level test data, but alternate estimates derived from lower level data are also possible.

The lower level reliability information must be combined to obtain a posterior distribution on the system level failure intensity, which then serves as the prior for the system-level reliability growth assessment. A key consideration in the use of the lower-level reliability information is the representation of the system reliability structure. As described in [72] and [73], a reliability block diagram or fault tree is typically used to describe the underlying structure of the system reliability as defined by the various components and subsystems within the system. For the technique proposed here, a reliability block diagram will be used without loss of generality. The concepts could be easily extended to other system reliability models such as fault trees or event trees.

6.2.1 System-Level Failure Intensity

To connect the reliability growth assumptions to the system reliability model using a block diagram or fault tree, it is helpful to first consider a representation of the system as defined in Assumptions 1 and 2. The reliability block diagram of this structure is shown in Figure 6.1. The overall system is just a serial connection of the failure modes that exist in the system, where the occurrence of any failure mode causes failure of the system.

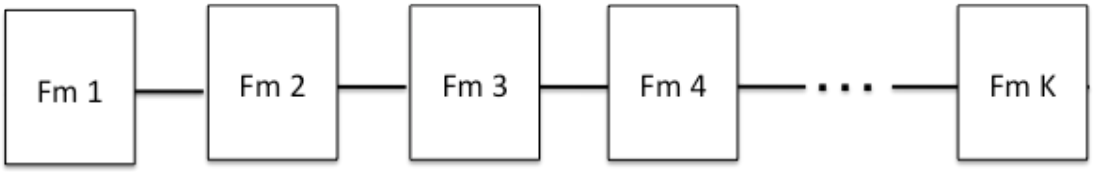


Fig 6.1. Complex System Structure as Series of *K* Failure Modes

By recognizing that that each failure mode can occur in one of two ways, the failure mode structure in Figure 6.1 can be mapped to the structure of subsystems and components within the system. A failure mode can occur as a result of a component failure that causes failure of the system, or a failure mode can occur due to failure of two or more components within a redundant block in the system structure. This allows for the block diagram in Figure 6.1 to be equivalently represented by that shown in Figure 6.2, where component or redundant block failures each lead to specific failure modes within the system. The first two modes in the example are caused by specific component failures, while the third mode is caused by failure of any pairing of components 3, 4, and 5. It is important in developing this

representation that the structure be reduced to its simplest form. For example, if a component can contribute to two failure modes involving redundancy, the structure of the system should be modified such that the two failure modes are combined into a single mode. The redundancy can then be appropriately modeled as described in Section 6.3.2.

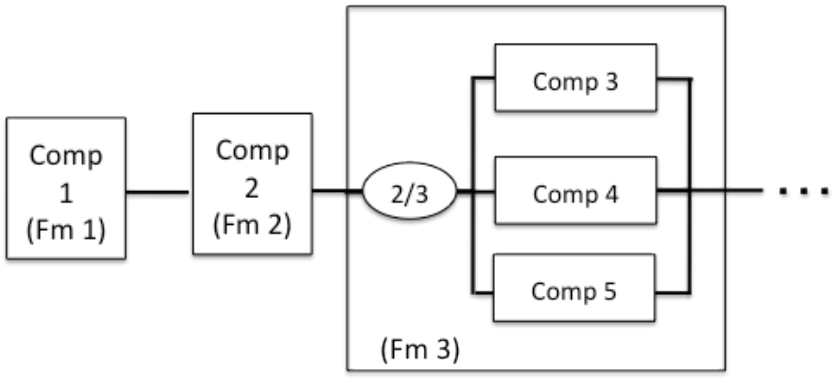


Fig 6.2. Serial Failure Modes Represented as Component or Subsystem Failures

The system level structure in Figure 6.2 can also be represented using subsystems, where one or more components or redundant blocks comprise an individual subsystem. This extends the flexibility of the system-level description when estimating the posterior distribution, as data for individual subsystems can be used in place of estimating the posteriors for each individual mode failure intensity within the subsystem. The system level failure intensity can be represented as

$$\lambda_s = \sum_{i=1}^{p} \lambda_i = \sum_{i=1}^{p} \sum_{j=1}^{Fm_i} \lambda_{i,j}\,, \tag{3}$$

where p subsystems are present in the system and each subsystem contains Fm_i failure modes. This decomposition allows for the failure modes from a complete subsystem to be modeled equivalently as a single failure mode, as the subsystem

failure intensity is just the sum of the contributions from each of the modes within the subsystem. The posterior distribution for each of the failure modes linked to components, redundant blocks, or subsystems can then be developed using any relevant reliability data that are available. A discussion of the use of component life data, historical data, and physics-based modeling is presented in Section 6.3.

When the posterior distributions of the failure intensities for the necessary failure modes are defined, the system reliability structure can be used to develop the posterior distribution for the system level failure intensity. Monte Carlo methods can be used to generate samples from the posterior distributions for each of the mode failure intensities. The resulting samples from each of the failure modes can then be summed across the system due to the serial representation of the system reliability structure. The computational complexity involved is relatively small. As demonstrated in Section 6.3, the posterior distributions for the mode failure intensities will consist mainly of Gamma distributions. For distributions that are more complex, such as those that involve physics-based models, the required samples from the distribution on the failure intensity are already available via the original distribution development, so there is no additional computational burden. The reduction of the system structure to a serial representation may seem overly simplistic, but this approach is very relevant for complex systems. The serial structure results in a constant failure intensity for the system, which is also supported in asymptotic theory for complex repairable systems and superimposed renewal processes such as Drenick's Theorem [91].

As discussed in Chapter 3, sums of Gamma distributed random variables will again be approximately Gamma distributed. This will generally be the case with the approach discussed here, as many of the mode posterior distributions that result will be well approximated by the Gamma. It is worth examining the potential impact of the approximate Gamma distributions on the distribution of the system-level failure intensity. The worst possible scenario would involve a system composed of 2 subsystems, with data available only at the subsystem level. A large difference in the β parameters for the Gamma posteriors for each subsystem may negatively impact the resulting system level posterior. The availability of more data on components and other failure modes should serve to minimize the impact from modeling at the subsystem level, so this scenario should suffice. The resulting posterior distribution for the system is shown in Figure 6.3. The first subsystem used a Gamma(1, 50^{-1}) posterior for the mode failure intensity, and the second used a Gamma(5, 250^{-1}).

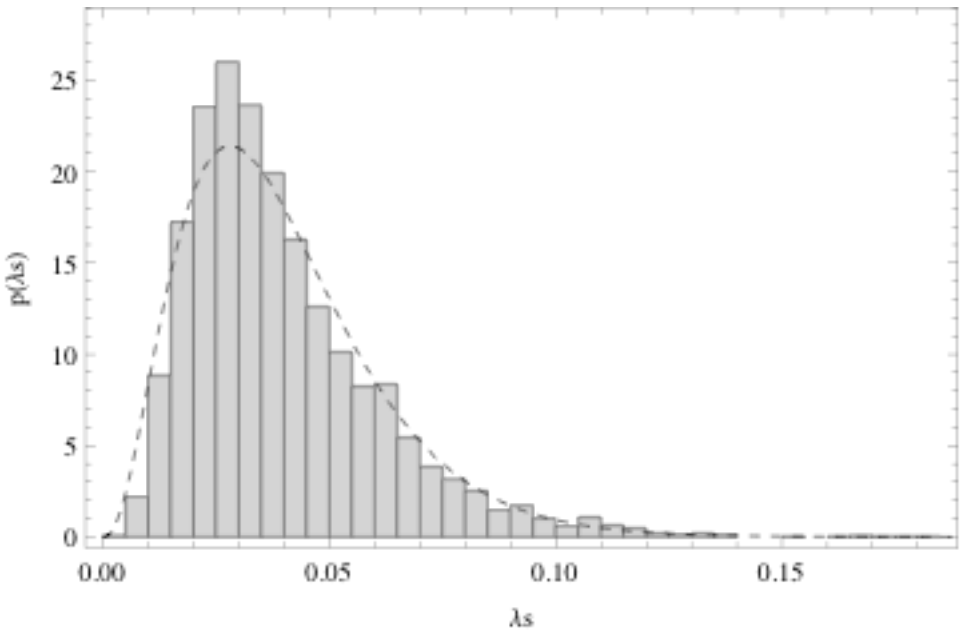


Fig 6.3. System Failure Intensity for 2 Subsystems with Overlaid Gamma Approximation

The results indicate that the system level distribution is still reasonably approximated by a Gamma distribution. The parameters of the system-level distribution can be found through standard method-of-moments based approaches, and the resulting λ_B and β parameters can then be used for the prior parameters in the reliability growth projection model in Chapter 3. This allows for direct connection of lower-level reliability data and information to be used in connection with the system-level reliability growth models.

6.2.2 Accounting for Failures Due to Integration of Components and Subsystems

A potential drawback of the reliability block diagram or fault tree approach for modeling the system reliability is the assumption of independence between the various components and subsystems. For many complex systems, a large number of failures are possible due to interactions between the subsystems, and the methods for assessing system reliability in Section 6.2.1 may not adequately address this issue. The simplest potential solution to the problem is to use engineering knowledge about the potential failure modes in the system and associate any potential integration problems with a specific subsystem or component. There will likely be little to no data available to characterize the integration failure intensity, but the scaling approach in Section 6.3.3 can be used to account for the differences between the estimated or modeled failure intensity and the true operational failure intensity. Historical data from similar systems can also serve as a valuable source of information to help with the application of the scale factor.

A second method for handling the integration failures is to add an integration block directly to the reliability block diagram or system reliability model. This approach can be used to explicitly capture the integration related failure modes that may exist throughout the system in a single block in the model. Historical information from similar can again serve as a valuable source for understanding the potential failure intensity of the integration block, and Maximum Entropy methods can be used to develop the specific distribution given the available information.

Note that in cases where there is little knowledge of the potential integration failure modes, the resulting Maximum Entropy distribution will be uniform. If the integration block is a significant contribution to the overall failure intensity, for example when two major subsystems are combined together, the uniform distribution for the integration failures may influence the posterior on the system-level failure intensity. The system-level result may be look more like a uniform, meaning that it cannot easily be approximated with a Gamma distribution. This is an indication of immaturity with respect to the system-level reliability. More engineering effort should be made to characterize the potential integration failure modes within the system prior to full system-level reliability growth testing. When the reliability becomes more mature, the integration block can be better characterized. The resulting system-level posterior distribution will then be more easily characterized with a Gamma distribution, which then serves as the appropriate prior information for the reliability growth process.

6.3 Failure Mode Posterior Distribution

6.3.1 Failure Mode from Component Failure

When a failure mode is caused by failure of a single component, the mode is caused by the occurrence of one or more failure mechanisms [92]. The failure and subsequent replacement for an individual failure mechanism considers the component as a consumable item that is replaced with a new version upon failure. When modeling the process of failure and replacement in this manner, the failure mechanism can be assumed to induce a renewal process for the component and its associated failure mode [93]. The existence of multiple potential failure mechanisms for the component and associated mode causes a superposition of renewal processes for the failure mode. Theoretical results can be applied to develop a reasonable model of this process. When examining renewal processes, the Elementary Renewal Theorem [93] can be used to describe the behavior of the process over time. Applying the Elementary Renewal Theorem to the failure and replacement process for the failure mechanism states that for a failure process with mean time between successive failures μ and expected number of failures up to time t denoted by $m(t)$,

$$\lim_{t\to\infty}\frac{m(t)}{t}=\frac{1}{\mu}. \tag{4}$$

The limit in (4) states that the rate of occurrence of failures for the failure mechanism being considered converges to $1/\mu$ as t becomes large. The theorem can also be used to state that

$$m(t)\approx\frac{t}{\mu}\ \text{ when } t \text{ is large,} \tag{5}$$

which says that for large t, the expected number of failures from the specific mechanism by time t is the standard rate of occurrence multiplied by the length of time t. An extension of this result is known as Blackwell's Theorem [93]. Blackwell's Theorem generalizes (5) and shows the expected number of failures in a general time interval of length u converges to u/μ as t becomes large.

These results indicate that when the process of failure and subsequent replacement of a component and failure mode is considered over a large enough time period, the only requirement for modeling the process is the mean time between successive failures for the mechanism. This result is true for general renewal distributions even when the life distribution for the mechanism in question is subject to significant wear out or aging over time. The major benefit of this concept is that when multiple failure mechanisms contribute to the occurrence of a failure mode, the failure mode occurs via a superposition of renewal processes. Drenick's Theorem states that a Homogeneous Poisson Process (HPP) with rate $1/\mu$ can then be used to approximate the failure process for the mode over a sufficiently large time period [91]. The use of the limit reduces a potentially complex estimation problem to a straightforward calculation that can be accomplished through the HPP. It is useful to understand the performance of the HPP approximation, as the result assumes a number of failure mechanisms are present for the failure mode. Examining the superposition of renewal processes for 2 failure mechanisms can provide an understanding of how useful the HPP may be in practice. A Weibull(α,β) distribution with density function in (6) can be used to model a failure mechanism with an increasing hazard rate such as one resulting from again or wear-out. Other

distributions with increasing hazard rates could also be used without loss of generality.

$$p(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left(\frac{t}{\alpha}\right)^{\beta}, \quad t > 0, \quad \alpha > 0, \quad \beta > 0 \tag{6}$$

An example of this concept is shown in Figure 6.4. Weibull(1.5, 2) and Weibull(5, 3.2) distributions were used to represent failure mechanisms with increasing hazard rates, and 50 failures were simulated from the corresponding distribution in (6). The plotted results were truncated to represent the time period where the two processes overlapped, resulting in 64 total failures for the superimposed process. The expected number of failures was estimated using the MLE for a HPP [94], and it is shown as the solid line in the plot.

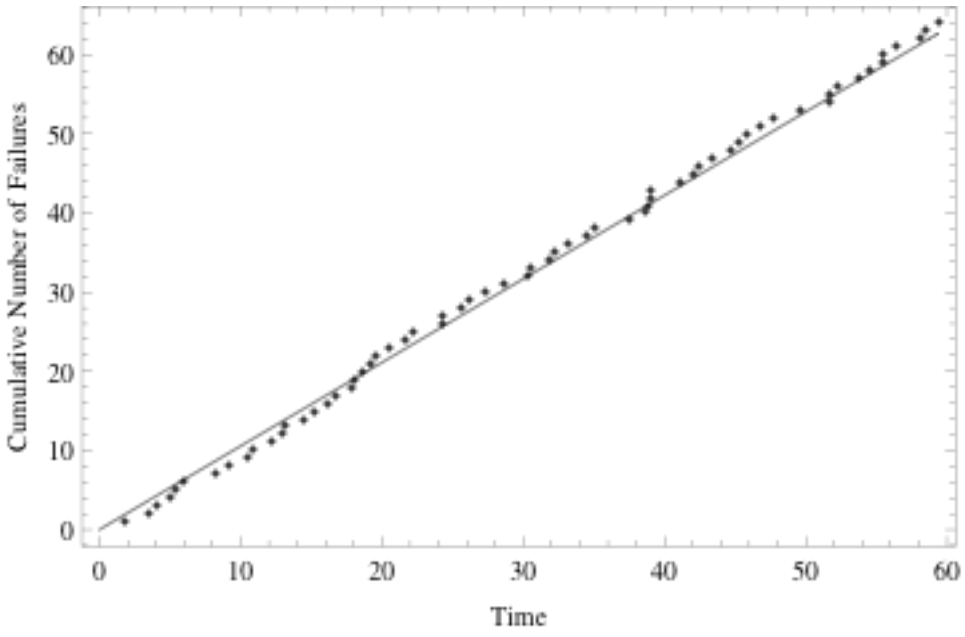


Fig 6.4. Cumulative Number of Failures vs. Time for 2 Superimposed Renewal Processes with Corresponding HPP Expected Number of Failures Overlaid (

The results in Figure 6.4 indicate that the 2 superimposed renewal processes can be well approximated with a HPP. It is also useful to understand the approximation

when the time interval is relatively short, as the theorem from Drenick [91] is asymptotic with respect to complexity and time. Figure 6.5 contains the results for the first 10 samples using the same renewal distributions as in Figure 6.4.

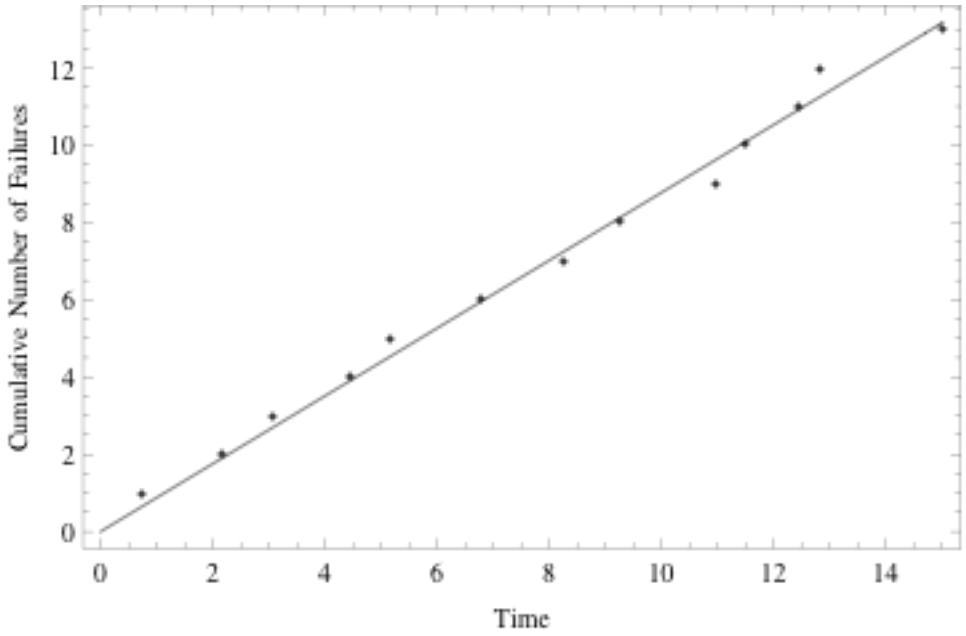


Fig 6.5. Cumulative Number of Failures vs. Time for 2 Superimposed Renewal Processes with Corresponding HPP Expected Number of Failures Overlaid

The results indicate that the HPP approximation is still reasonable even when time and complexity are not large. Caution should nevertheless be used in these cases, as the variation in the process may provide results that are inaccurate. If the failure modes in the system have very little data or are low complexity with respect to the number of contributing failure mechanisms, the HPP should be carefully examined before use.

Component failure data may be directly available from component life testing or from historical failure data collected from field usage. If such data are available

for a component, the results in Sections 6.2 and 6.3 can be used to develop a posterior distribution on the component failure intensity that describes the component renewal process. Assume the n samples of the component life have been collected with failures occurring at times $(t_1,t_2,\ldots,t_n)$. Section 6.3 indicates that failures of the component can be adequately modeled using the HPP, which implies that the Exponential likelihood can be used for the data. The likelihood is given as

$$l(t_1,\ldots t_n \mid \lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda t_i) \tag{7}$$

To be conservative, a non-informative prior distribution can be used for the component failure intensity λ, which is given by the same Jeffrey's prior that was presented in Chapter 4 as

$$p(\lambda) = \frac{1}{\lambda \log\left(\frac{l_2}{l_1}\right)},\ l_1 < \lambda < l_2. \tag{8}$$

The resulting posterior is defined as

$$p(\lambda \mid t_1,\ldots,t_n) = \frac{\frac{1}{\lambda \log\left(\frac{l_2}{l_1}\right)} \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} t_i\right)}{\int_{l_1}^{l_2} \frac{1}{\lambda \log\left(\frac{l_2}{l_1}\right)} \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} t_i\right) \partial\lambda}. \tag{9}$$

Because the desired range for the failure intensity is the positive real line, taking the appropriate limits in the denominator of (9) then yields

$$p(\lambda \mid t_1,...,t_n) = \frac{\lambda^{n-1}\exp\left(-\lambda\sum_{i=1}^{n} t_i\right)}{\Gamma(n)\frac{1}{\left(\sum_{i=1}^{n} t_i\right)^n}}. \quad (10)$$

The posterior distribution in (10) is a Gamma distribution of the same form as used in the reliability growth models in Chapters 3 and 4. Note that if other relevant prior information is available, a Gamma prior distribution can easily be substituted in place of the Jeffrey's prior in (7). As demonstrated in Chapter 3, the posterior will again be a Gamma in this case, which can be used in the same manner when constructing the system level posterior distribution.

When historical reliability information is available for the component, it can be used in a similar manner as the component life test data. If the historical information is for identical components that are used in a different system or operating environment, the data may not provide a completely accurate representation of the failure intensity of the component. The method presented in Section 6.3.4 should be considered in this case, as it provides a probabilistic approach for accounting for the potential differences in the estimated mode failure intensity and the true mode failure intensity.

6.3.2 Failure Mode from Failure of Redundant Block

When a failure mode is the result of failure in two or more components within a redundant block in the system reliability structure, the approach from Section 6.3.1 can be applied for each of the components within the block. A redundant block is considered as a K-of-N structure without loss of generality, as standard parallel

redundancy is a special case of this structure. Each component failure is also assumed to occur from multiple contributing failure modes, allowing for the HPP representation for component failure.

To develop the corresponding failure intensity for the failure mode caused by the K-of-N failure, the mission time T_m can be used to calculate the individual reliability for each component. The reliability for a component in an HPP is given by

$$R_c(T_m) = \exp(-\lambda_c T_m), \tag{11}$$

and the corresponding reliability of the K-of-N block, $R_B(T_m)$, can be found using standard techniques for system reliability analysis in references such as [93] and [94]. The failure intensity for the failure mode, and hence K-of-N block, can then be found by inverting the reliability expression in (11). This yields a failure intensity value of

$$\lambda_i = -\frac{\log(R_B(T_m))}{T_m}, \tag{12}$$

where λ_i is the desired mode failure intensity. The approach in (12) is an approximation that provides consistency with the assumptions of the reliability growth models, but it should be noted that the failure modes from redundant blocks in the system should not provide a large contribution to the overall system failure intensity. In practice the approximation should not influence the overall system level failure intensity, but caution should be used if redundant blocks within the system are thought to provide a significant contribution to the overall system reliability or failure intensity. Sensitivity analysis should be performed in these situations, as the models contained in this work may not be appropriate.

The corresponding posterior distribution for the failure intensity in (12) can be found through Monte Carlo methods. The Gamma posterior described in Section 6.3.1 can be used to generate samples for the component failure intensity, and the resulting distribution for the component reliability in (11) can be found by transforming each of the posterior samples. The transformed samples of the component reliability can then be used to calculate the reliability of the K-of-N block, and the block reliability can be transformed further using (12) to get the mode failure intensity.

6.3.3 Using Physics-Based Model Results

There are a number of examples in the literature involving fusion of different information sources to estimate system reliability; see for example [72] - [76]. None of these approaches consider the use of physics-based modeling of failure mechanisms in the estimation. This information is frequently available early in the design process before much testing has been completed though, and for this reason it is desirable to utilize it to characterize the failure intensity within the renewal process for the specific component in question.

The approach presented here recognizes that physics-of-failure models generally contain material constants that are available in handbooks or other references [92]. Define the physical function for the life of a component as $f(N, a)$, where N is the life and a is a constant parameter related to the underlying physical failure mechanism being considered. These constants are typically estimated from empirical testing under controlled conditions, which imparts some amount of

uncertainty as to their actual value. Because of the uncertainties in the empirical constants, the function *f* is generally regarded as the mean or median value of the predicted life. Additional uncertainty results directly from variation in the manufacturing process that is used to produce the materials. Even if the original test results were available to develop the uncertainty distributions for the material constants, the uncertainty due to manufacturing and production variation would still not be quantified. The principle of Maximum Entropy [87],[88] can therefore be used to develop probability distributions on each of the material constants. This approach is general enough to capture the overarching empirical uncertainty along with the material variation associated with the parameters. A probability distribution on the component life *N* can be found by using the Maximum Entropy derived probability distribution for the constant *a*. Let $p(a)$ denote the probability distribution for *a*. Because $f(N, a)$ will likely be a complex function, an analytical form for the distribution of *N* will generally not be feasible. Monte Carlo simulation can be used instead to develop the necessary distribution. The use of simulation will be even more necessary when considering that most physics-based life calculations involve more than one parameter. The Maximum Entropy probability distribution will generally take a form that allows for straightforward random number generation, and *n* values can be sampled from $p(a)$. The realized values can then be used to calculate *n* realizations of the life *N*, which provides the desired distribution on the component life. Because the original function *f* provides the mean or median life, it is reasonable to use this result to estimate the approximate failure intensity for the component renewal process as described in Section 6.2.1.

To transform the distribution on the cycles to failure N_f into a distribution on the failure intensity for the component renewal process, the operational usage of the component must be understood. In particular, the number of cycles of the applied stress per operating hour or mile must be reasonably known. The distribution on the failure intensity for the renewal process is then defined through the transform

$$\lambda = \frac{c}{N_f}, \tag{13}$$

where λ is the failure intensity and c is the number of cycles per time or mileage increment.

To demonstrate the concept of the approach, consider the stress-life model given by

$$S = AN_f^{\,b}, \tag{14}$$

where S is the applied stress, N_f is the cycles to failure for the component, and A and b are constants specific to the material being considered. As discussed in previous chapters, Maximum Entropy maximizes the uncertainty associated with the problem, subject to any constraints associated with the information available in the analysis. The constraints that are applied in the derivation of the Maximum Entropy distribution take the form of moments for the distribution, and the support of the distribution must also be specified. For the values A and b, we assume only that the support of each variable is known. That is, A will fall in the range (A_1, A_2) and b will fall in the range (b_1, b_2). Applying these basic constraints in the Maximum Entropy approach yields a uniform distribution [87] for each parameter as shown in (15) and (16).

$$p(A) = \frac{1}{A_2 - A_1},\ A_1 < A < A_2 \tag{15}$$

$$p(b) = \frac{1}{b_2 - b_1},\ b_1 < b < b_2 \tag{16}$$

Random realizations can be easily drawn from the distributions in (15) and (16). Rewriting the life model in (14) in terms of the component life N_f yields

$$N_f = \left(\frac{S}{A}\right)^{\frac{1}{b}}. \tag{17}$$

10,000 realizations of each parameter were sampled using (A_1, A_2) = (1800, 2200) and (b_1, b_2) = (-0.12, -0.08). The resulting values were inserted into (17) with applied stress S = 1000. A histogram of the resulting distribution is shown in Figure 6.6.

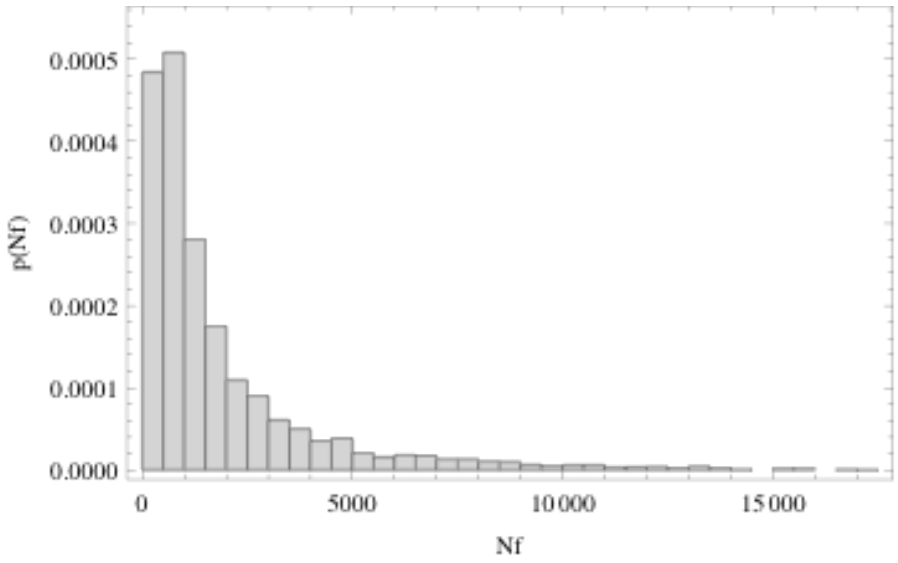


Fig 6.6. Distribution on Component Life

Assuming c = 100 cycles per hour and using the transform defined in (13) yields the distribution on the failure intensity of the renewal process shown in Figure

6.7. This distribution can be used to develop the distribution on the mode failure intensity that is subsequently used to develop the system level failure intensity.

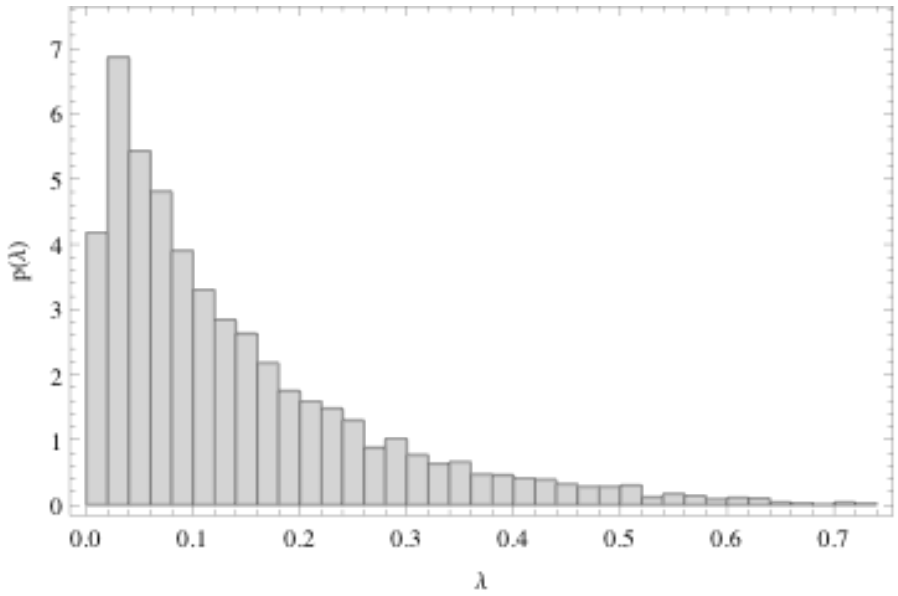


Fig 6.7. Distribution on Failure Intensity for Component Renewal Process

Note that as with other types of component information, the failure mechanism being modeled may not provide a completely accurate representation of the failure intensity of the component. The approach in Section 6.3.4 should be considered in this case, as it provides a probabilistic method for accounting for the potential differences in the estimated failure intensity and the true operational failure intensity.

6.3.4 Scaling to Account Accounting for Additional Failure Mechanisms

This section presents a method for scaling estimates of the failure intensity for those cases where the available data or modeling may not completely represent the known failure mechanisms that contribute to the occurrence of the failure mode. This situation can occur commonly when using historical data from like or identical

components that were used in different operational environments. It can also occur when using the physics-of-failure approaches described in Section 6.3.3, as other failure mechanisms beyond that being modeled may be present. In these cases potentially large contributors to the failure intensity will not be considered, and the result will be an underestimate of the true value.

To account for this issue, we examine the true failure intensity for a failure mode, denoted as $\lambda_{i,true}$. The mode failure intensity can be decomposed into contributions from the *n* individual failure mechanisms as a sum given by

$$\lambda_{i,\,true} = \sum_{j=1}^{n} \lambda_{i,j} \tag{18}$$

When only a subset of the potential failure modes or mechanisms is addressed in the information used to develop the estimate, a scale factor f_m can be used to represent the relationship between the estimated failure intensity, $\hat{\lambda}_i$, and the true failure intensity. This is represented as

$$f_m \lambda_{i,\,true} = \hat{\lambda}_i \;,\; f_m \in (0,\,1) \tag{19}$$

If the number of contributing failure mechanisms is large and only 1 or 2 are considered in the data or modeling, the value of the scale factor f_m will be small. If the information used in the estimate considers most or all of the dominant failure mechanisms, then the value of f_m will be large. Because f_m is unknown in practice, it is desirable to account for the uncertainty associated with its value. This can be accomplished by using Maximum Entropy to assign a probability distribution to the value of f_m. The relationship in (19) leads to the conditional distribution for λ_{true} given by

$$p\left(\lambda_{i,true} \mid f_m\right) = p\left(\frac{\hat{\lambda}_i}{f_m}\right) \tag{20}$$

The unconditional distribution of $\lambda_{i,true}$ is then found by marginalizing with respect to the distribution of f_m as in (21).

$$p\left(\lambda_{i,true}\right) = \int_0^1 p\left(f_m\right) p\left(\hat{\lambda}_i \mid f_m\right) df_m \tag{21}$$

Conservative treatment of the scale factor assumes no information is available other than the support of f_m, which results in a uniform distribution over the specified range [87]. In practical applications, the contribution of the failure mechanism(s) to the overall mode/component failure rate should be examined to determine the range of the scale factor. A FMECA could used to help with this determination, and historical data could also be used when available. Note that an additional assumption of a mean or variance value beyond the range of the factor would change the probability distribution associated with f_m. A detailed discussion of the distributions resulting from different assumptions can be found in [88]. Figure 6.8 depicts an example application of the scale factor assuming a Uniform(0.5,1) distribution for the physics-model example in Figure 6.9 in Section 6.3.3. The application of the scale factor can be seen to shift the distribution to the right, while also increasing the overall variance of the failure intensity. Note that the hyperbolic shape of the distribution also lends itself to being approximated by a Gamma distribution. Though not necessary for the calculations, this point is useful when examining the behavior of the system-level failure intensity.

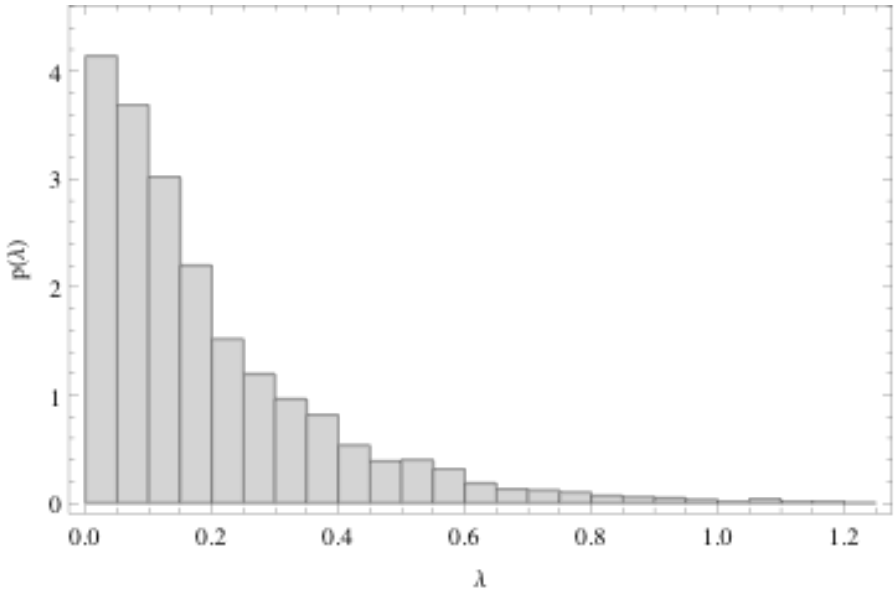


Fig 6.8. Scaled Distribution on Failure Intensity for Component Renewal Process

6.4 Conclusion

This chapter presents a general approach for combing to lower-level sources of reliability data and information to develop a prior distribution on the system-level failure intensity. The reliability growth models in Chapters 3-5 consider the failure intensity of complex repairable systems over time. Each component or redundant block is mapped to the corresponding failure mode that results when a failure occurs. The individual modes are also assumed to occur via multiple potential failure mechanisms, where each failure mechanism induces a renewal process for the component involved. The superposition of the renewal processes from the mechanisms allows for the use of an HPP to approximate the failure process for the failure mode. The HPP provides a simplified approach that easily allows for the use of component or subsystem data when developing the posterior disrtribution for the

mode failure intensity. The result is also extended to blocks of parallel or K-of-N redundancy, where each component follows the HPP and posterior distribution on the failure intensity is found by assuming a constant failure intensity at the redundant block level.

Physics-based modeling results are also used to develop a failure mode posterior distribution. A method is also developed for accounting for additional failure mechanisms that may not be accounted for in the mode posterior. This probabilistic scaling is particularly important when using physics-based results or other component data that may have been collected in specific environment that does not include all potential stressors of the component. The individual failure mode posterior distributions can then be combined using Monte Carlo methods, and the posterior is reasonably represented by a Gamma distribution.

The results of this approach can be utilized when system-level failure data are limited or not yet available. When viewing the process of reliability assessment across the various stages of development of the system, assessment in this manner serves as prior information that can be updated with the reliability growth models in earlier chapters. The approach also serves as a method of connecting early engineering activities that involve component modeling and characterization to the full system-level testing that occurs later in the system's life cycle.

7 CASE STUDY ON RELIABILITY GROWTH ASSESSMENT

7.1 Introduction

7.1.1 Background

This chapter presents a case study depicting the application of the methodology presented throughout this thesis. The approach is demonstrated on a complex military system that is modified through the addition of specific mission equipment. The methodology in Chapter 6 is utilized to develop a prior distribution on the system-level failure intensity. Historical field failure data are used to characterize the components and subsystems that remain unchanged in the modification, and a physics-of-failure model is used to estimate the failure intensity of a specific component due to fatigue. The failure intensity for the mission specific equipment is considered through Maximum Entropy methods. The posterior distribution is then used along with the reliability growth projection model in Chapter 3 to update the posterior distribution after a developmental test event. The results are then subsequently updated after an operational test is conducted using the approach in Chapter 4.

7.1.2 Chapter Overview

This chapter is organized as follows. Section 7.2 presents the development of the prior distribution on the system-level failure intensity. Section 7.3 then presents the updated posterior after DT. Section 7.4 further updates the posterior after OT. Conclusions are presented in Section 7.5.

7.2 Reliability Assessment Prior to DT

A combination of information is available for developing the distribution on the system-level failure intensity prior to DT. Historical failure data is available for the components and subsystems that are common to both the existing system and the modified version. A stress life model was also developed to model fatigue for the driveshaft of the vehicle, as this was an area of concern identified in the existing system. There are nine existing major subsystems within the vehicle, and a tenth subsystem is comprised of the additional mission equipment. The system-level reliability block diagram is represented by each of the subsystems connected in a series. The basic reliability block diagram is shown in Figure 7.1.

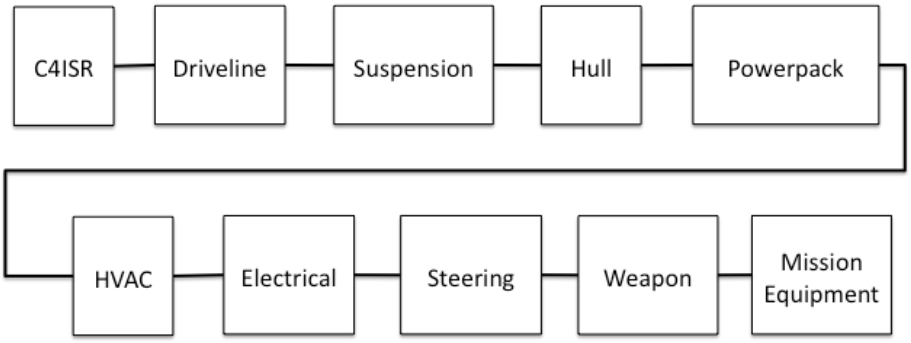


Fig. 7.1. Simplified Reliability Block Diagram of System

Additional detail is also added to consider the contribution of wheel/tire failures within the suspension subsystem and driveshaft failures within the driveline subsystem. The failure definition for this example is any loss of an essential function. The system is an 8-wheeled vehicle designed to carry troops, with one driveshaft per wheel. This creates redundancy within the driveline and suspension subsystems,

where failure is defined by the loss of 5 or more wheels or driveshafts. The reliability block diagram in Figure 7.1 can then be modified to account for the 5-of-8 failure definition within the driveline and suspension subsystems. The updated detailed representation of the two subsystems is shown in Figure 7.2.

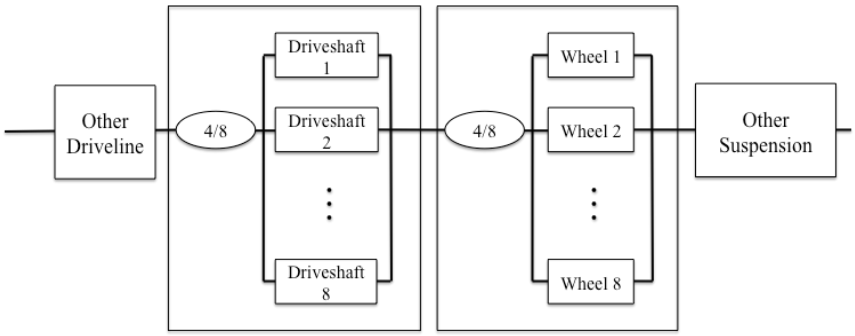


Fig. 7.2. Detailed Reliability Block Diagram of System

The methods in Chapter 6 can then be applied to each of the subsystems and blocks containing redundancy within the system.

7.2.1 Historical Failure Data

A large amount of reliability data is available on the nine subsystems that are common to both versions of the system. The historical reliability data is in the form of operational field failures collected for 3 vehicles over a two-year period. The mileage accrued for each of the vehicles during the two-year period is also collected, and this allows for straightforward application of the methods in Chapter 6. The mileage for each vehicle is shown in Table 7.1. Table 7.2 contains the number of failures for each of the subsystems. Due to the redundancy of the wheels and driveshafts, these failures were separated from the data in Table 7.2. The wheel

failures are shown on a separate line to allow for the failure intensity of the 4-of-8 block to be estimated from the component failures, and the suspension failures do not include the driveshaft.

The estimator described in Section 6.2.1 can be applied at the subsystem level for the data in Table 7.2. The posterior distribution will be a Gamma(n_i,1/T) for n_i subsystem failures during T accumulated miles.

TABLE 7.1

VEHICLE MILEAGE

Vehicle	Mileage
1	5325
2	6532
3	4510
Total	16367

TABLE 7.2

SUBSYSTEM FAILURES

Subsystem	Number Failures
Driveline	6
Wheels	36
C4ISR	15
Hull	54
Powerpack	8
Electrical	16
HVAC	11
Steering	5
Suspension	2
Weapon	11

5000 samples were generated from the Gamma posterior for each of the nine subsystems in Table 7.2. The histograms of the samples are shown in Figures 7.3 - 7.11.

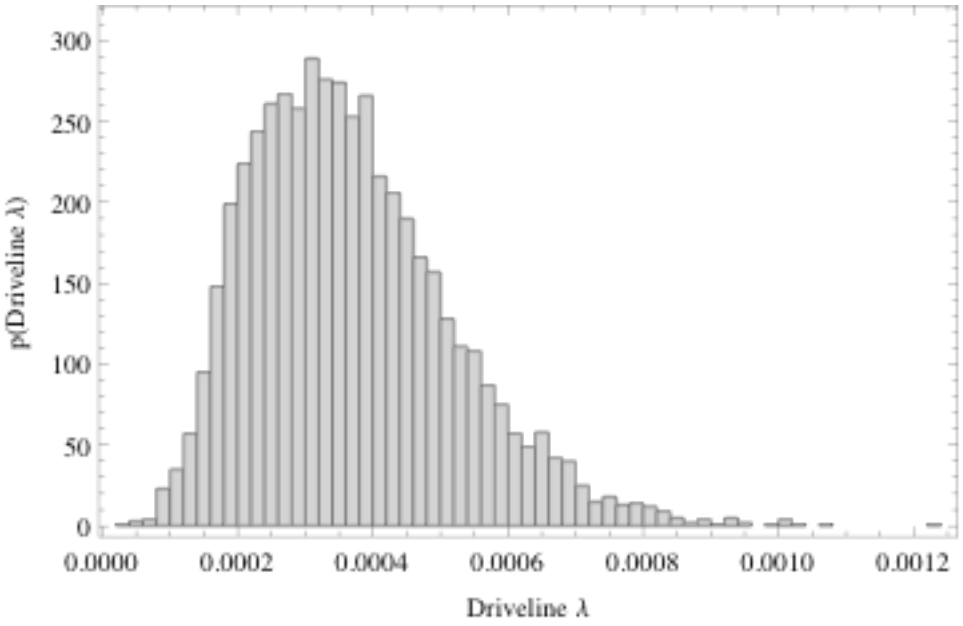


Fig 7.3. Posterior Distribution on Driveline Failure Intensity

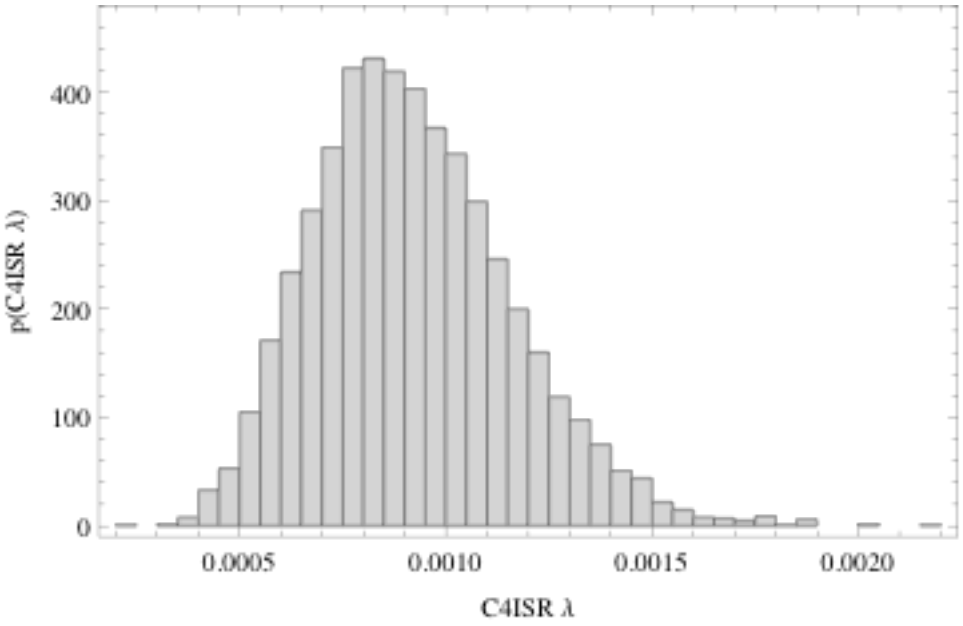


Fig 7.4. Posterior Distribution on C4ISR Failure Intensity

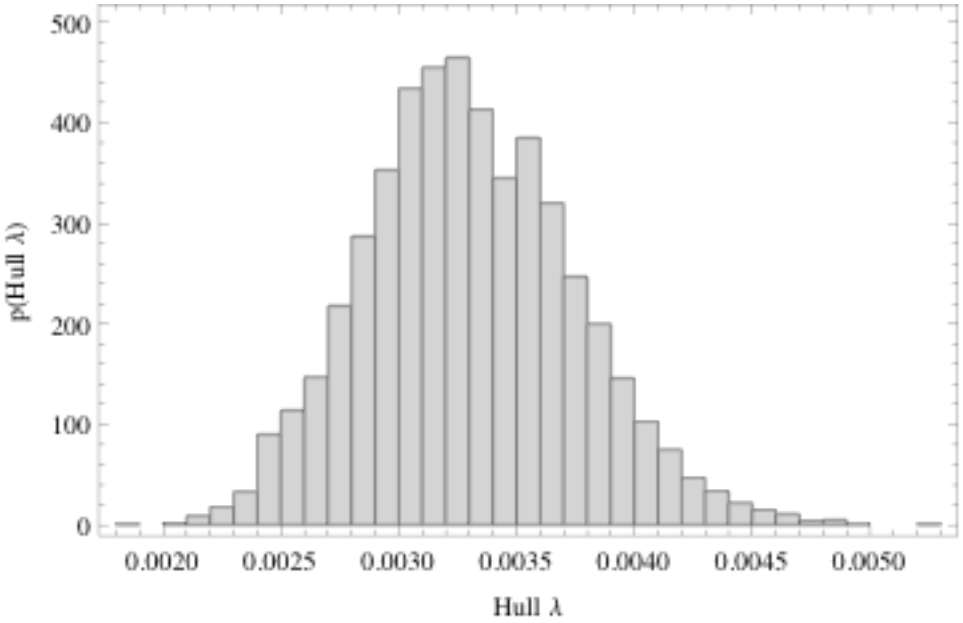


Fig 7.5. Posterior Distribution on Hull Failure Intensity

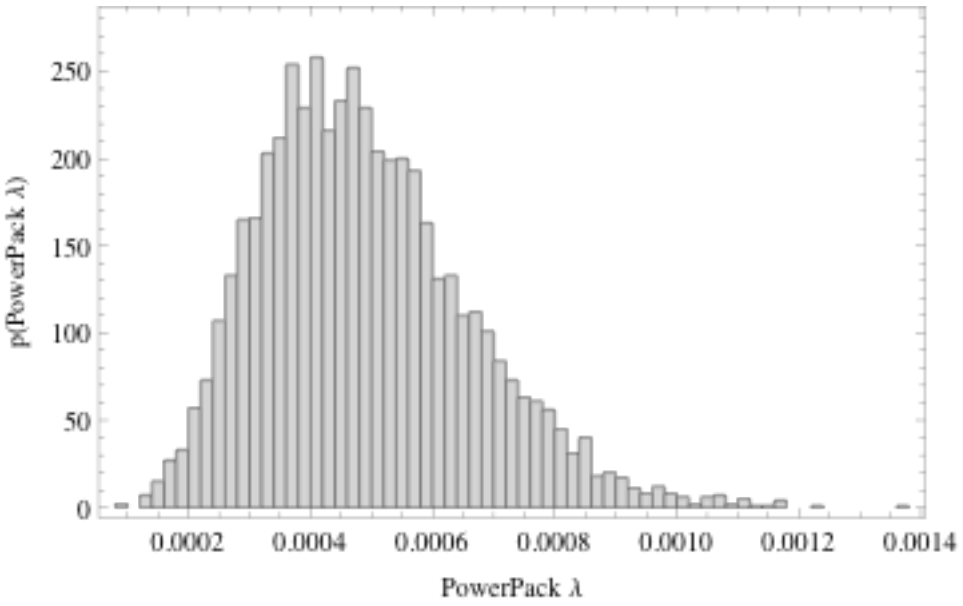


Fig 7.6. Posterior Distribution on PowerPack Failure Intensity

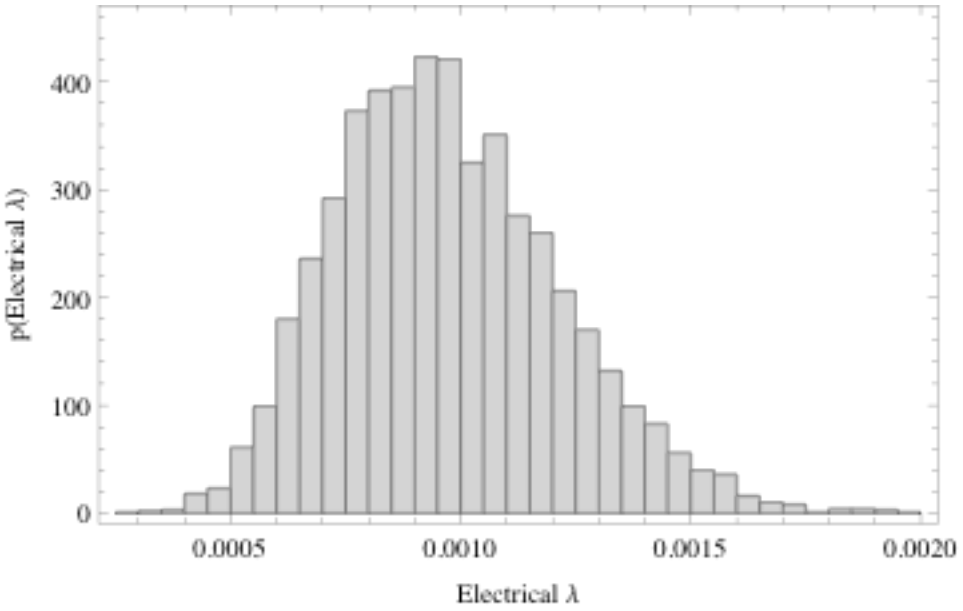


Fig 7.7. Posterior Distribution on Electrical Failure Intensity

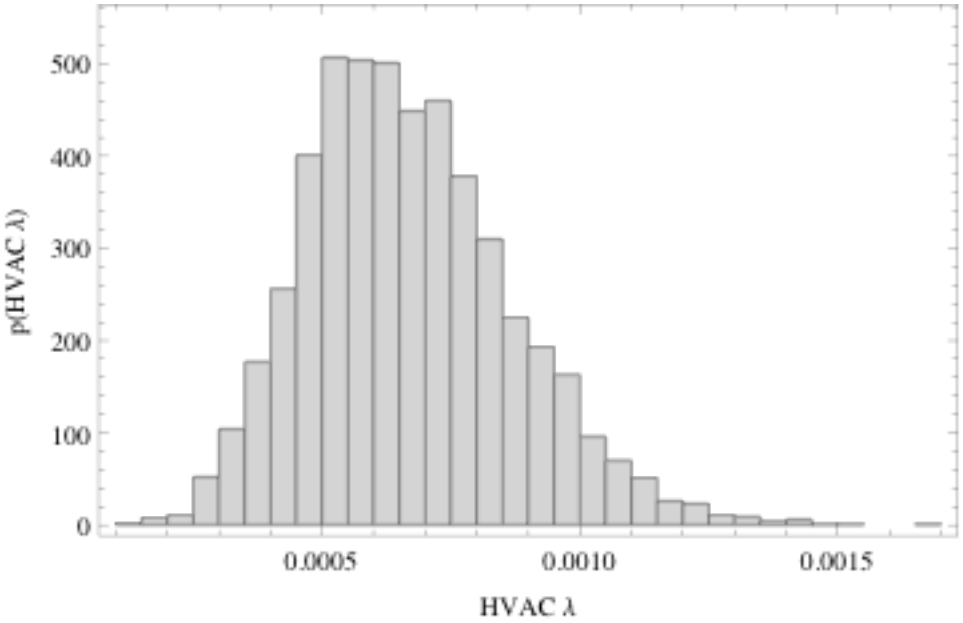


Fig 7.8. Posterior Distribution on HVAC Failure Intensity

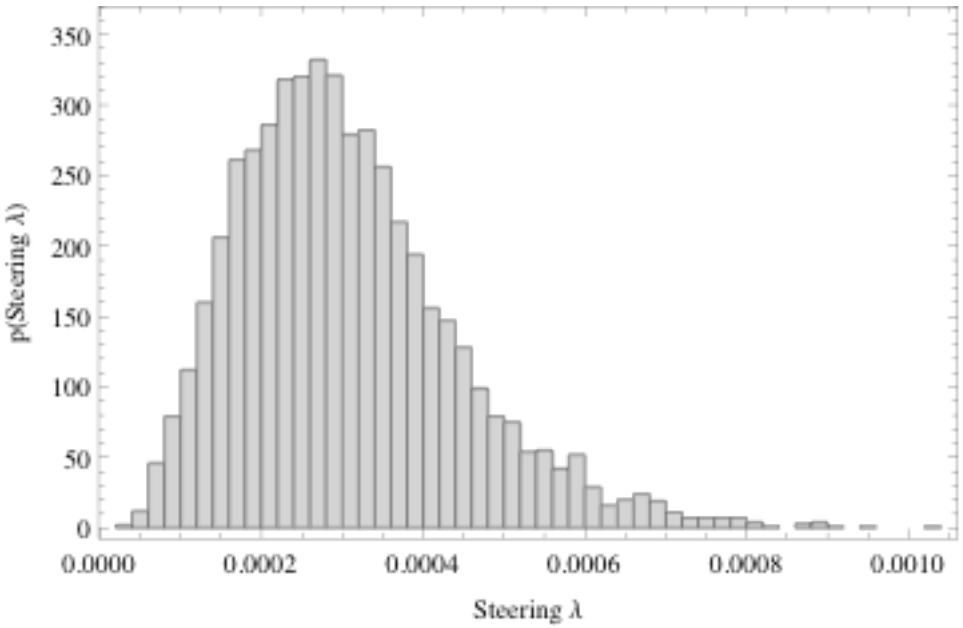


Fig 7.9. Posterior Distribution on Steering Failure Intensity

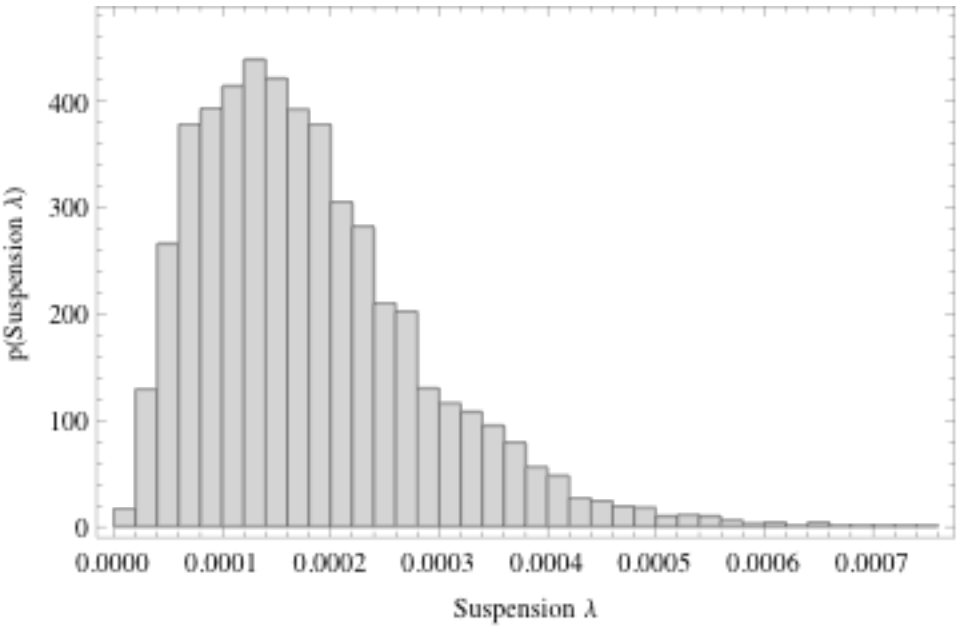


Fig 7.10. Posterior Distribution on Suspension Failure Intensity

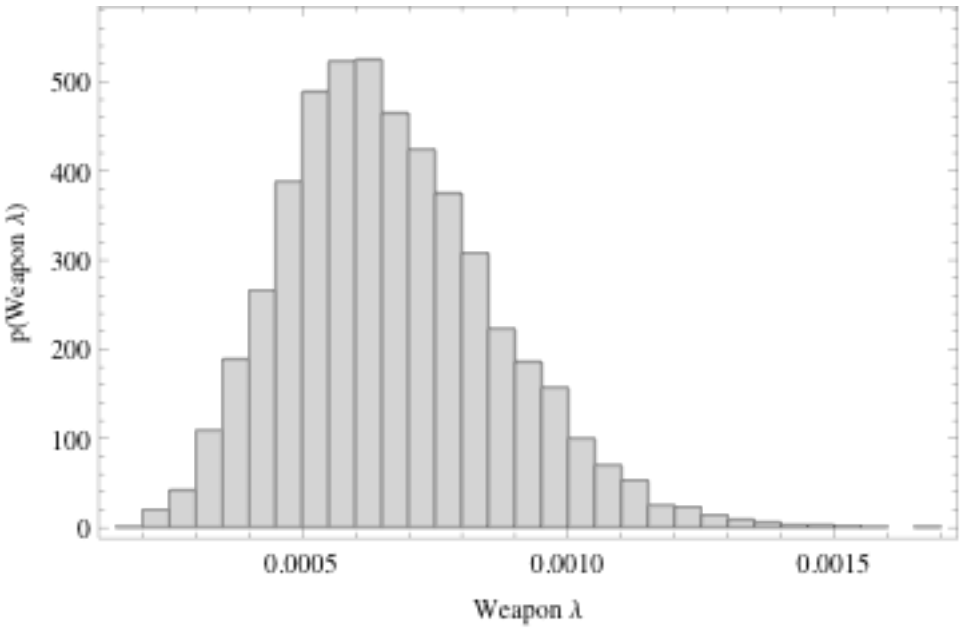


Fig 7.11. Posterior Distribution on Weapon Failure Intensity

As stated in Chapter 6, the Gamma posterior for the wheel failures was used to calculate the reliability for the 4-of-8 redundant block. The block reliability was then used to develop the corresponding failure intensity for the redundant block. The histogram for the wheel component failure intensity is shown in Figure 7.12, and the resulting redundant block failure intensity is shown in Figure 7.13.

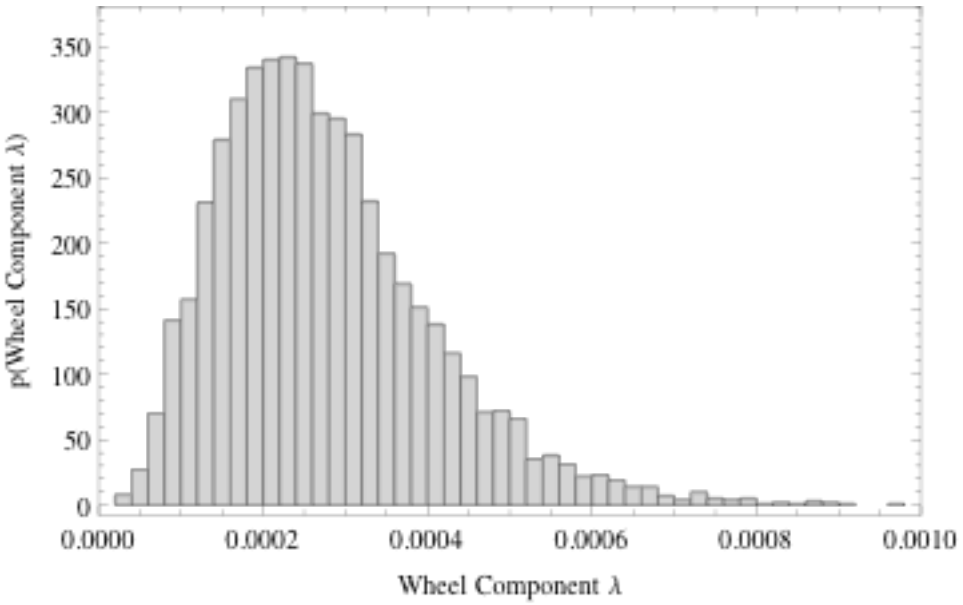


Fig 7.12. Posterior Distribution on Wheel Component Failure Intensity

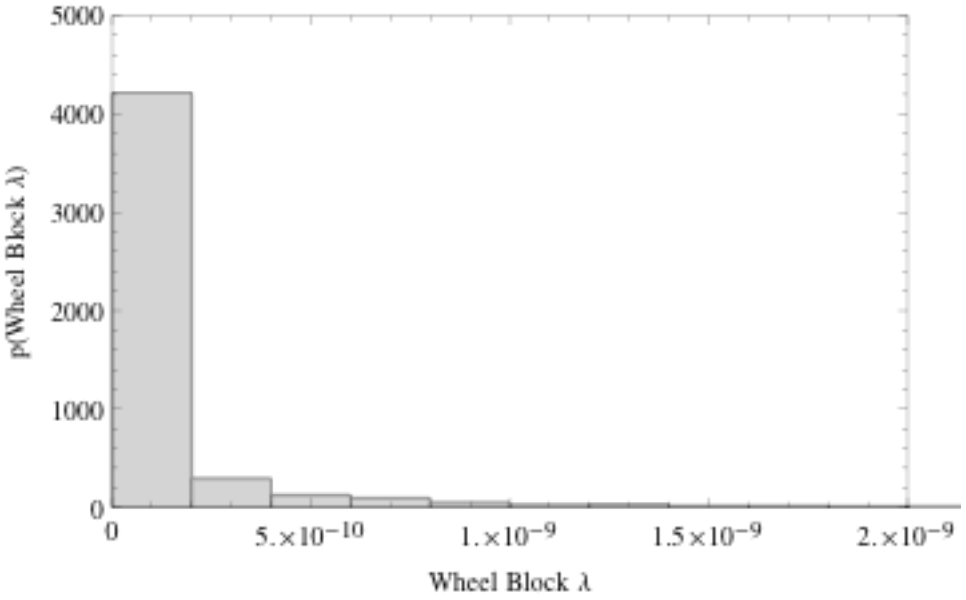


Fig 7.13. Posterior Distribution on Wheel Redundant Block Failure Intensity

7.2.2 Driveshaft Posterior from Fatigue Modeling

The driveshaft of the vehicle was previously identified as a candidate for fatigue modeling, as the vehicle contains 8 wheels and therefore contains 8 driveshafts. The driveshaft is also subject to varying amounts of stress induced through torgue applied when the driver accelerates the vehicle. This presents a potential reliability concern for the vehicle, as the repeated stress could cause failure of the shaft if the design is not robust to the fatigue created by the stressor.

The simple stress-life model in Chapter 6 was used again for the driveshaft. The relationship between the stress and cycles to failure is given as

$$N_f = \left(\frac{S}{A}\right)^{\frac{1}{b}}. \tag{1}$$

where S is the applied stress, N_f is the cycles-to-failure for the component, and A and b are constants specific to the material being considered. As discussed in Chapter 6, Maximum Entropy can be used to develop prior distributions for the material constants in the model. Assuming only the support for each distribution yields a uniform distribution over the defined range of the support [87]. The distributions for the parameters are shown in (2) and (3).

$$p(A) = \frac{1}{A_2 - A_1},\ A_1 < A < A_2 \tag{2}$$

$$p(b) = \frac{1}{b_2 - b_1},\ b_1 < b < b_2 \tag{3}$$

For the steel material of the driveshaft the range of A was chosen as (A_1, A_2) = (1800, 2200), and the range for b was chosen as (b_1, b_2) = (-0.12, -0.08). To account for the variable loading of the stress, the Normal distribution method used by Steinberg [95]

can be used for variable applied stress on the driveshaft. The shaft was instrumented during limited testing over representative terrain to understand the stress applied to the driveshaft during operation. This approach assumes that the testing is a sufficient representation of the operational usage of the system, and that the stressors measured from the test are reasonably represented by a Normal distribution. The 1-σ, 2-σ, and 3-σ values from the testing were given as 271 MPa, 392 MPa, and 514 MPa. These values were used as the applied stress in (1), and 5000 samples were drawn from the distributions in (2) and (3). For each of the 5000 samples, the resulting cycles-to-failure $N_{f,i}$ was calculated for the i^{th} stress value. The cycles-to-failure for the driveshaft was then found by probabilistically combining the results based on the underlying Normal distribution. The 1-σ, 2-σ, and 3-σ values correspond to 0.683, 0.271, and 0.043 percentages of the total loading, which results in the total cycles to failure for the driveshaft of

$$N_f = 0.683 * N_{f,1} + 0.271 * N_{f,2} + 0.043 * N_{f,3} \quad (4)$$

To translate the number of cycles-to-failure into the failure intensity for the driveshaft, the result in (4) must be divided by the number of cycles per mile. This value was measured from the test data as 28,660 cycles per mile.

The last consideration in developing the driveshaft failure intensity is the relative contribution of fatigue to the overall failure of the driveshaft. Examining data on similar systems shows that fatigue accounts for 10-25% of the failure rate on the driveshaft. This information can be used within the scaling approach discussed in Chapter 6. The resulting failure intensity of the driveshaft is found through the expression

$$p(\lambda_{i,true}) = \int_{0.1}^{0.25} p(f_m)\, p(\hat{\lambda}_i \mid f_m)\, df_m \quad , \tag{5}$$

where $\hat{\lambda}_i$ is the estimated failure intensity after transforming the cycles-to-failure in (4). The resulting histogram for the driveshaft failure intensity is shown in Figure 7.14.

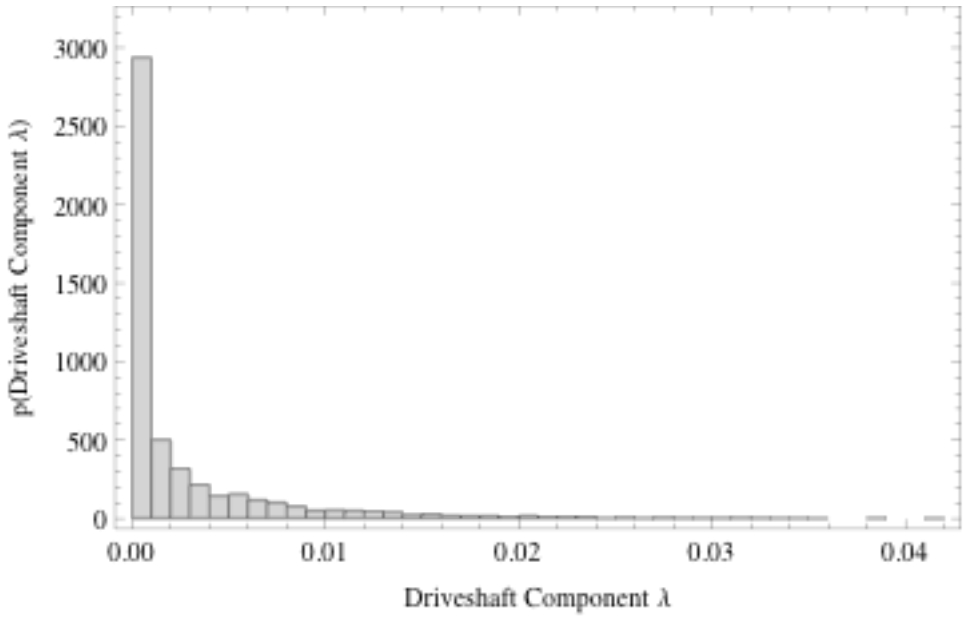


Fig 7.14. Posterior Distribution on Driveshaft Component Failure Intensity

To account for the 4-of-8 redundancy of the driveshaft, the approach used for the wheels can be applied again to the driveshaft. The reliability for the 4-of-8 block was calculated for each of the 5000 samples, and the resulting failure intensity for the block was calculated from the block reliability values. The histogram of the redundant block is shown in Figure 7.15.

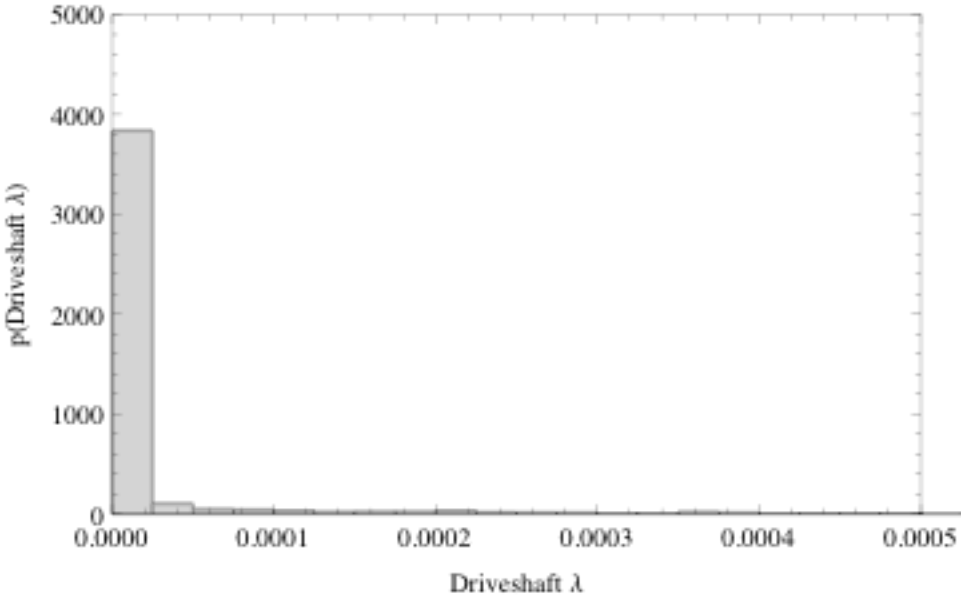


Fig 7.15. Posterior Distribution on Driveshaft Redundant Block Failure Intensity

The use of the stress-life approach is somewhat basic, and more detailed analysis is possible using Finite-Element Methods in combination with more detailed strain-life models. The stress-life approach is more than adequate in this context though. The 4-of-8 redundancy for the component within the larger subsystem and system reduces the relative contribution of the driveshaft to a second order effect, so more detailed modeling is not necessary. If the driveshaft was a larger relative contribution to the overall system level failure intensity, more detailed modeling may be appropriate. The complexity of more detailed models should always be considered though, as additional parameters within the model will increase the uncertainty in the predicted life and resulting failure intensity.

7.2.3 Mission Equipment Posterior

The failure intensity for the additional mission equipment on the modified system cannot be developed from historical data. Conservative estimates can be made by examining similar equipment on existing systems and applying Maximum Entropy to develop the failure intensity distribution. The mean failure intensity for the mission equipment is assumed to 0.01 with a support of (0, 0.03). Maximum Entropy methods will yield a truncated Exponential distribution similar to the developed in Chapter 4. The resulting probability density function is given by

$$p(\lambda) = \frac{\theta e^{-\theta\lambda}}{1 - e^{-0.3*\theta}},\ 0 < \lambda < 0.3. \tag{6}$$

The θ parameter is determined from the constraint on the mean, which results in the equation shown in (7).

$$0.1 = \frac{1}{\theta} - \frac{0.3}{e^{0.3*\theta} - 1}. \tag{7}$$

Generating 5000 samples from the distribution in (6) yields the histogram shown in Figure 7.16. The distribution on the mission equipment failure intensity is conservative when compared to the other subsystems in the system. The uncertainty is also fairly high, which is desirable. The performance of the additional mission equipment may be different than the other similar types of equipment when it is fully integrated into the new system.

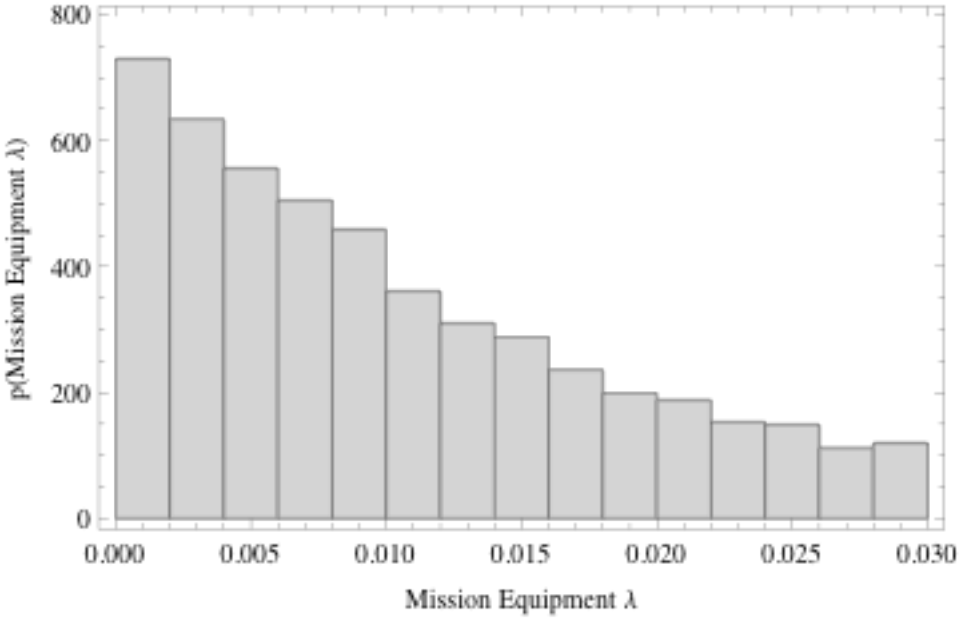


Fig 7.16. Posterior Distribution on Mission Equipment Failure Intensity

7.2.4 System Level Failure Intensity

To develop the posterior distribution on the system level failure intensity, the reliability block diagram in Figure 7.2 can be used. The structure defined by Figure 7.2 allows for summation of the failure intensity posteriors developed in Section 7.2.1 and 7.2.2. Summing the 5000 samples from each of the posteriors for the subsystems and redundant blocks yields the posterior distribution on the system level failure intensity shown in Figure 7.17.

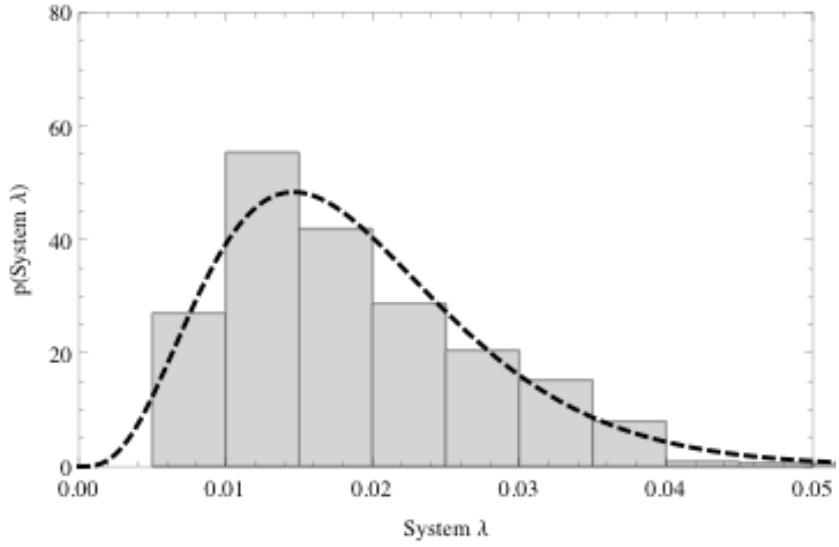


Fig 7.17. Posterior Distribution on System Failure Intensity

Method-of-Moments estimators were used to develop the approximate Gamma distribution represented by the dashed line. The resulting λ_B and β estimates are shown in (6) and (7).

$$\hat{\lambda}_B = 0.01904 \tag{6}$$

$$\hat{\beta} = 0.00442 \tag{7}$$

These values can be used in place of the empirical Bayes estimators within the reliability growth model of Chapter 3. The β value is somewhat large relative to the empirical Bayes estimates that generally result from the reliability growth model in Chapter 3. This is a desirable result, as the β value directly impacts the variance of the Gamma and hence the uncertainty surrounding the knowledge of the system failure intensity. Larger values indicate higher uncertainty in the system level failure intensity, and given that the estimate may not properly account all potential integration failure modes, the conservatism is warranted.

7.3 Reliability Growth Assessment During DT

Developmental reliability growth testing of the vehicle was conducted to identify any unknown failure modes that may have been introduced into the system through the upgraded design. Testing was conducted using 2 vehicles with a fixed configuration. The intent of the testing was to make any necessary corrective actions at the completion of the test event to improve the system reliability prior to additional testing. Any potential reliability incidents during the testing were recorded and later adjudicated prior to assessing the reliability results from the test. The test identified 130 failures comprising 31 failure modes. Cumulative mileage for each vehicle is shown in Table 7.3. The number of occurrences for each failure mode and the assigned FEF is shown in Table 7.4. Due to resource constraints, only the three most prevalent failure modes (14, 15, and 17) were corrected prior to the next phase of testing.

TABLE 7.3

VEHICLE MILEAGE

Vehicle	Mileage
1	4198
2	4275
Total	8473

TABLE 7.4

FAILURE MODE DATA

Mode	Number of Occurrences		FEF
	System 1	System 2	
1	2	1	0
2	1	0	0
3	3	1	0
4	5	3	0
5	1	1	0
6	1	1	0
7	5	1	0
8	1	0	0
9	1	0	0
10	1	0	0
11	2	1	0
12	2	0	0
13	1	1	0
14	11	1	0.7
15	11	1	0.7
16	3	2	0
17	21	8	0.7
18	2	0	0
19	1	0	0
20	3	4	0
21	3	4	0
22	4	1	0
23	0	1	0
24	0	2	0
25	0	2	0
26	0	4	0
27	0	1	0
28	0	1	0
29	0	1	0
30	0	1	0
31	0	1	0

The data were used in the reliability growth projection from Chapter 3. The estimates developed in Section 7.2 were used for the prior parameters. The posterior distribution for the projected MTBF of the system is shown in Figure 7.18. The mean

of the MTBF posterior is found to be 88.8 hrs. The updated posterior distribution on the failure intensity is a Gamma with parameters $\tilde{\alpha}$ = 109.044 and $\tilde{\beta}$ = 0.000102. These parameters can then be used to update the assessment when OT is completed for the system. The Chi-Squared goodness of fit test from Chapter 3 was also applied, and the proportion of statistics greater than the critical value was approximately 0. This indicates that the model provides a reasonable description of the failure intensity of the system.

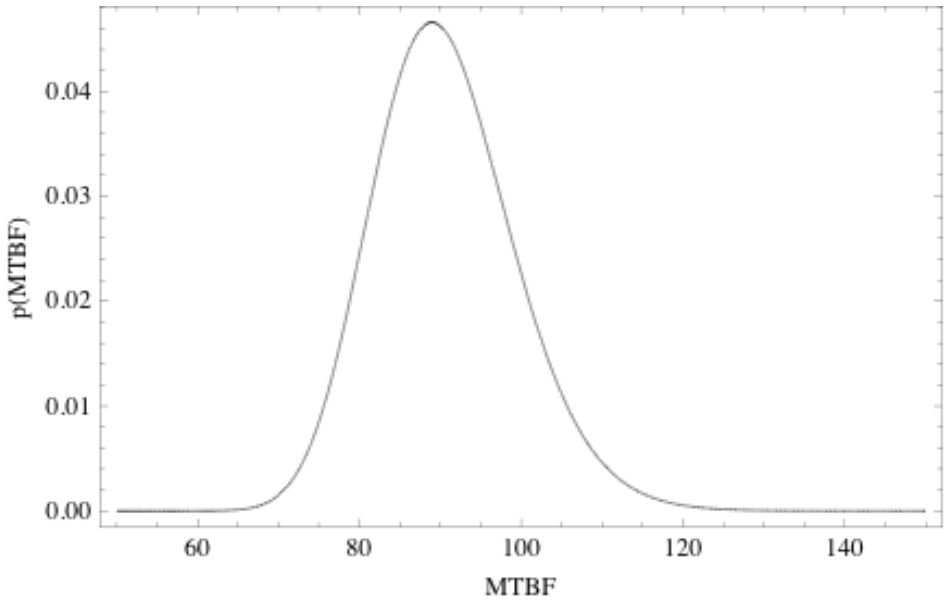


Fig 7.18. Posterior Distribution on System MTBF

7.4 Operational Reliability Assessment

OT was conducted in order to demonstrate the level of reliability for the upgraded version of the system. The test was a fixed configuration test involving a single vehicle with military operators utilizing the system within the context of the

intended operational mission of the vehicle. Incidents were recorded and later adjudicated as to the severity of the incident. Failure is once again defined as loss of an essential function in the system. The results identified 3 failures with 224 miles. Because the reliability growth testing from Section 7.3 was conducted in a rigorous environment with a large number of miles accumulated, the reliability degradation between the reliability growth test and the operational test was determined to be small. The mean degradation for the uncertainty distribution was therefore chosen to be 0.1.

Using the posterior Gamma parameters from the reliability growth model along with operational test results and the assigned degradation yields the updated system level posterior distribution on the MTBF shown in Figure 7.19.

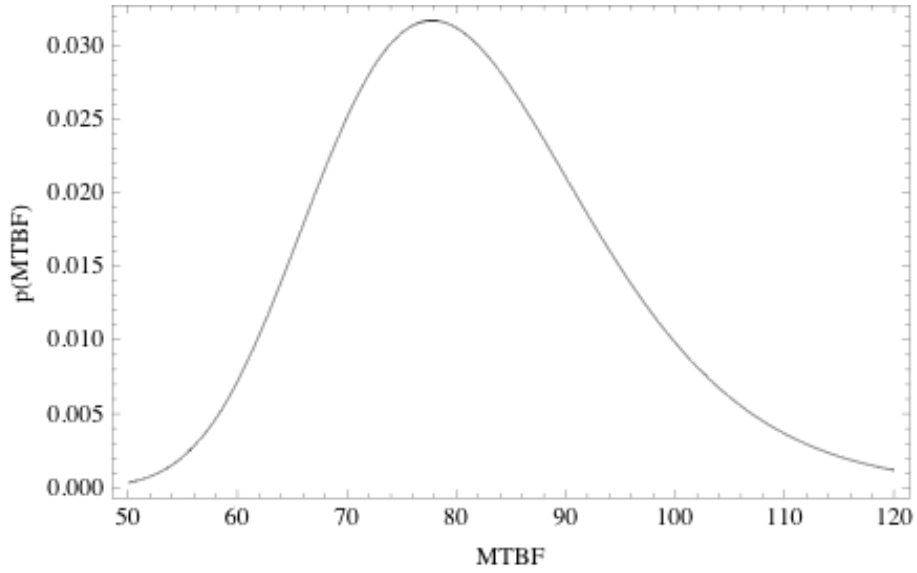


Fig 7.19. Posterior Distribution on System MTBF after OT

The mean MTBF is 79.8 hrs, and the 0.80 probability interval is found to be (66 hrs, 99.5 hrs). For comparison purposes, using just the operational test results alone yields the standard point estimate and 80% confidence interval of 74.7 hrs and (33.5 hrs, 128.4 hrs) respectively. The combined Bayesian method reduces the uncertainty associated with the estimate considerably. To demonstrate additional utility of the method, the estimate resulting from using a non-informative prior distribution on the reliability degradation yields a mean MTBF of 59.5 hrs with a corresponding 80% probability interval of (41.3 hrs, 97.9 hrs). Even when the DT-OT degradation is treated as an unknown value, the resulting interval estimate still has less uncertainty than the operational assessment alone.

7.5 Conclusion

This chapter has presented a case study depicting the application of the methodology presented throughout this thesis. The approach was demonstrated on a complex military system that was modified through the addition of specific mission equipment on an existing vehicle. Historical failure data for nine major subsystems was used along with a physics-based fatigue model. Redundancy was considered for two components within the system, and the methodology in Chapter 6 was applied to develop a posterior distribution on the system-level failure intensity.

The posterior distribution from the lower level data was then updated with test data from a system level reliability growth test. The reliability growth projection model in Chapter 3 was used to update the posterior distribution, considering the impact of corrective actions for 3 of the failure modes that were observed during the test. The results were then subsequently updated after an operational test by applying

the model developed in Chapter 4. The resulting posterior distribution on the failure intensity was shown to have a much lower uncertainty when compared to the traditional reliability assessment using the operational test data alone. The uncertainty was still lower even when a non-informative prior distribution was used for the DT-OT degradation. The case study demonstrates how the methodology in this thesis can be applied throughout the development of a complex system.

8 EXTENSIONS AND FUTURE WORK

8.1 Introduction

8.1.1 Background

This chapter contains a discussion of future work that can extend the research presented in this thesis. The development of analogous methodology for discrete one-shot type systems is an obvious area where additional work is possible. An outline of a potential discrete version of the reliability growth projection model in Chapter 3 is presented here. A brief discussion of the extension of the approach in Chapter 4 to discrete systems is also given. The last extension discussed involves the use of physics-based modeling within the framework for developing estimates of the system-level failure intensity.

8.1.2 Chapter Overview

This chapter is organized as follows. Section 8.2 presents a potential framework for discrete reliability growth under arbitrary corrective actions, including model assumptions and data requirements. Initial analytic results are presented along with empirical Bayes estimates. Section 8.3 discusses the potential extension of the approach to include operational test data, which is a discrete analogue of Chapter 4. Section 8.4 contains a brief discussion on additional work in the probabilistic modeling of failure mechanisms. Section 8.5 discusses the potential extension of general reliability assessment via failure mode modeling.

8.2 Reliability Growth of One-Shot Devices under Arbitrary Corrective Actions

8.2.1 Model Assumptions

The approach outlined here follows that of [21] and [22], which treats the system reliability as a product of reliabilities from independent failure modes. Modeling the system as a combination of failure modes in this manner enables the use of FEF values, which mathematically quantify the reduction in the mode failure probability for a given failure mode after a corrective action has been implemented. As with the model in Chapter 3, the approach allows for arbitrary corrective actions to be made during the test period. The assumptions are as follows:

1. The system is comprised of a large number of failure modes that are serial in nature; the occurrence of any failure mode results in failure of the system.
2. Failure modes generate failures independently of one another.
3. The failure probability for each mode is constant both before and after a corrective action is implemented.
4. Corrective actions to failure modes do not introduce new failure modes into the system.
5. Testing is conducted under operationally relevant conditions and stressors.

8.2.2 Posterior Inference for Single Failure Mode

To develop the model for the system reliability, start by examining the posterior distribution for a single failure mode. For a given test event of T trials with arbitrary corrective actions, assume for the i^{th} failure mode there are n_i failures on trials t = ($t_{i,1}$, …, $t_{i,ni}$) with corrective action on trial v_i and FEF d_i. Further assume

that there are $n_{i,1}$ failures prior to the corrective action. Denoting the failure mode probability of failure as p_i, assumptions 2 and 3 result in the likelihood for a failure mode given by

$$l\left(t \mid p_i, v_i, d_i, T\right) \propto p_i^{n_{i,1}} \left(1-p_i\right)^{v_i - n_{i,1}} \left(1-p_i\right)^{(1-d_i)\left[T - v_i - \left(n_i - n_{i,1}\right)\right]} \left[1-\left(1-p_i\right)^{1-d_i}\right]^{n_i - n_{i,1}} \quad (1)$$

Note that the use of the FEF in (1) is slightly different than that used in previous discrete reliability growth projection models such as [21] and [22]. The FEF has been previously applied by directly scaling the probability of failure for a failure mode. The approach used here yields a failure mode reliability after corrective action of

$$R_{i,new} = R_i^{(1-d_i)}, \quad (2)$$

which provides more consistency between continuous and discrete reliability growth projection models. The continuous models assume a constant failure rate for each failure mode, meaning that the reliability can also be expressed using an Exponential distribution. A log-transform will result in the traditional failure rate remaining after corrective action. Log-transforming the Exponential representation of (2) results in

$$\lambda_{i,new} = \left(1-d_i\right)\lambda_i, \quad (3)$$

which fits the usual definition for FEF in the continuous reliability growth projection models.

For the prior distribution on the mode failure probability p_i we assume a Beta distribution of the form

$$p\left(p_i\right) = \frac{\Gamma\left(a+b\right)}{\Gamma\left(a\right)\Gamma\left(b\right)} p_i^{a-1}\left(1-p_i\right)^{b-1}. \quad (4)$$

The posterior distribution for the mode probability of failure is then given by

$$p(p_i \mid t) = \frac{p(p_i)\,l(t \mid p_i, v_i, d_i, T)}{\int_0^1 p(p_i)\,l(t \mid p_i, v_i, d_i, T)\,dp_i}. \tag{5}$$

To solve for the posterior, substitute $n_{i,2} = n_i - n_{i,1}$ and use a Binomial expansion such that

$$\left[1-(1-p_i)^{1-d_i}\right]^{n_{i,2}} = \sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j (1-p_i)^{(1-d_i)j}. \tag{6}$$

This yields a failure mode posterior distribution of

$$p(p_i \mid t) = \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j\, p_i^{\,a+n_{i,1}-1} (1-p_i)^{b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j)-1}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma[a+n_{i,1}]\,\Gamma[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j)]}{\Gamma[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j)]}}. \tag{7}$$

If all corrective actions are delayed until the end of the test, $n_{i,2} = 0$, $v_i = T$, and the posterior in (7) will simplify to that of standard Beta-Bernoulli conjugate relationships. Also if the failure mode is unobserved during testing, the posterior in (7) reduces to

$$p(p_i \mid t) = \frac{p_i^{\,a-1}(1-p_i)^{b+T-1}}{\frac{\Gamma[a]\Gamma[b+T]}{\Gamma[a+b+T]}} = \frac{\Gamma[a+b+T]}{\Gamma[a]\Gamma[b+T]}\, p_i^{\,a-1}(1-p_i)^{b+T-1}. \tag{8}$$

We are interested in the posterior reliability for the failure mode, so using (7) we can develop the posterior mean reliability as

$$E(1-p_i \mid t) = \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j)+1]}{\Gamma[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j)+1]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j)]}{\Gamma[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j)]}}. \tag{9}$$

For the mean posterior failure mode reliability after corrective action, we similarly have

$$E\left[(1-p_i)^{(1-d_i)} \mid t\right] = \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j+1)\right]}{\Gamma\left[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j+1)\right]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j)\right]}{\Gamma\left[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j)\right]}}. \tag{10}$$

The result in (10) will then be used to construct the posterior mean for the complex system. To support later calculations involving the posterior variance, we also calculate the corresponding second moment for the mean in (10). The second moment follows similarly as

$$E\left[\left\{(1-p_i)^{(1-d_i)}\right\}^2 \mid t\right] = \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j+2)\right]}{\Gamma\left[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j+2)\right]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j)\right]}{\Gamma\left[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j)\right]}}. \tag{11}$$

8.2.3 Posterior Inference for Complex System

For a complex system consisting of a large number of failure modes, the previous results for a single failure mode can be used to develop analogous results for the entire system. From assumption 1 in Section 8.2.1, we have

$$E[R \mid t] = E\left[\prod_{i=1}^{K}(1-p_i)^{(1-d_i)} \mid t\right] = \prod_{i=1}^{m} \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j}(-1)^j \frac{\Gamma\left[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j+1)\right]}{\Gamma\left[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j+1)\right]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j}(-1)^j \frac{\Gamma\left[b+v_i-n_{i,1}+(1-d_i)(T-v_i-n_{i,2}+j)\right]}{\Gamma\left[a+b+v_i+(1-d_i)(T-v_i-n_{i,2}+j)\right]}} * \left[1-\frac{a}{a+b+T}\right]^{K-m} \quad (12)$$

where K is the total number of failure modes in the system and m is the number of failure modes observed during the test. The expression in (12) can be seen as the product of failure mode reliabilities for the entire system, with the left hand term representing the observed failure modes and the right hand term representing the unobserved failure modes. Equation (10) is used for the observed modes, and the mean of the posterior in (8) is used for the unobserved modes. The expression in (12) assumes that the number of failure modes in the system is known, which is not usually practical for complex systems. For this reason we examine the limit of (12) as K becomes large in order to develop an estimate that does not rely on knowing the number of failure modes in the system.

Prior to taking the limit, reparameterize (12) using the prior mean reliability for the system and an additional parameter for the Beta distribution. First let

$$\tilde{n} = a + b, \quad (13)$$

and then let the prior mean reliability for the system be denoted as

$$R_I = \prod_{i=1}^{K} 1 - \frac{a}{a+b}. \quad (14)$$

The a parameter can be expressed using (13) and (14) as

$$a = \tilde{n}\left(1 - R_I^{1/K}\right), \quad (15)$$

which can be seen to go to zero as K becomes large. Reparameterizing (12) in this manner yields

$$E[R \mid t] = E\left[\prod_{i=1}^{K}(1-p_i)^{(1-d_i)} \mid t\right] = \prod_{i=1}^{m} \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} - a + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j + 1)\right]}{\Gamma\left[\tilde{n} + v_i + (1-d_i)(T - v_i - n_{i,2} + j + 1)\right]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} - a + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j)\right]}{\Gamma\left[\tilde{n} + v_i + (1-d_i)(T - v_i - n_{i,2} + j)\right]}} * \left[1 - \frac{a}{\tilde{n} + T}\right]^{K-m}. \quad (16)$$

Taking the limit of (16) with respect to K requires calculation of the limit of the term that represents the unobserved failure modes, which is given by

$$\left[1 - \frac{a}{\tilde{n} + T}\right]^{K-m}. \quad (17)$$

Substituting the expression for a in (15) into (17) yields

$$\lim_{K \to \infty} \left[1 - \frac{a}{\tilde{n} + T}\right]^{K-m} = \exp\left(\frac{\tilde{n}}{\tilde{n} + T} \log R_I\right) = R_I^{\frac{\tilde{n}}{\tilde{n}+T}}. \quad (18)$$

The limiting form for the mean reliability of the complex system is then

$$\lim_{K \to \infty} E[R \mid t] = \lim_{K \to \infty} \prod_{i=1}^{m} \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} - a + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j + 1)\right]}{\Gamma\left[\tilde{n} + v_i + (1-d_i)(T - v_i - n_{i,2} + j + 1)\right]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} - a + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j)\right]}{\Gamma\left[\tilde{n} + v_i + (1-d_i)(T - v_i - n_{i,2} + j)\right]}} \left[1 - \frac{a}{\tilde{n} + T}\right]^{K-m}$$

$$= \prod_{i=1}^{m} \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j + 1)\right]}{\Gamma\left[\tilde{n} + v_i + (1-d_i)(T - v_i - n_{i,2} + j + 1)\right]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j)\right]}{\Gamma\left[\tilde{n} + v_i + (1-d_i)(T - v_i - n_{i,2} + j)\right]}} R_I^{\frac{\tilde{n}}{\tilde{n}+T}}. \quad (19)$$

The second moment of the posterior can be developed through techniques similar to those for the posterior mean. Again distinguishing between observed and unobserved failure modes, the second moment for the system is

$$E\left[R^2 \mid t\right] = \prod_{i=1}^{m} \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[b + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j + 2)\right]}{\Gamma\left[a + b + v_i + (1-d_i)(T - v_i - n_{i,2} + j + 2)\right]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[b + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j)\right]}{\Gamma\left[a + b + v_i + (1-d_i)(T - v_i - n_{i,2} + j)\right]}} \left[\frac{(b+T+1)(b+T)}{(a+b+T+1)(a+b+T)}\right]^{K-m}. \quad (20)$$

Using the same reparameterization in (13)-(15) allows the contribution from unobserved failure modes to be expressed as

$$\left[\frac{(b+T+1)(b+T)}{(a+b+T+1)(a+b+T)}\right]^{K-m} = \left[\left(1 - \frac{a}{\tilde{n}+T+1}\right)\left(1 - \frac{a}{\tilde{n}+T}\right)\right]^{K-m}. \quad (21)$$

Taking the limit with respect to *K* yields

$$\lim_{K\to\infty}\left[\left(1 - \frac{a}{\tilde{n}+T+1}\right)\left(1 - \frac{a}{\tilde{n}+T}\right)\right]^{K-m} = R_I^{\frac{\tilde{n}}{\tilde{n}+T+1}+\frac{\tilde{n}}{\tilde{n}+T}}. \quad (22)$$

Using (22) and taking the limit of (20) yields

$$\lim_{K\to\infty} E\left[R^2 \mid t\right] = \prod_{i=1}^{m} \frac{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j + 2)\right]}{\Gamma\left[\tilde{n} + v_i + (1-d_i)(T - v_i - n_{i,2} + j + 2)\right]}}{\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} + v_i - n_{i,1} + (1-d_i)(T - v_i - n_{i,2} + j)\right]}{\Gamma\left[\tilde{n} + v_i + (1-d_i)(T - v_i - n_{i,2} + j)\right]}} R_I^{\frac{\tilde{n}}{\tilde{n}+T+1}+\frac{\tilde{n}}{\tilde{n}+T}}. \quad (23)$$

Denoting the second moment in (23) as m_2 and the mean in (19) as m, the posterior variance for the complex system reliability is then

$$Var\left[R \mid t\right] = \sigma^2 = \mu_2 - \mu^2. \quad (24)$$

The product of individual Beta random variables does not follow an exact Beta distribution, but a Beta distribution can likely be used as a suitable approximation of the posterior for the system reliability. A standard method-of-moments approach can be used to determine the parameters of the approximate distribution, which involves simultaneously solving the equations in (25) and (26) for the new Beta parameters a_1 and b_1.

$$\mu = \frac{a_1}{a_1 + b_1}. \tag{25}$$

$$\sigma^2 = \frac{a_1 b_1}{\left(a_1 + b_1\right)^2 \left(a_1 + b_1 + 1\right)}. \tag{26}$$

The approximation can be checked by simulating failure modes and corrective actions in a complex system. The Beta approximation should be sufficient, but simulation should also be used to examine the performance of the approximation under arbitrary corrective action strategies.

8.2.4 Extension to Include Multiple Systems Under Test

The results in Section 8.2.3 were developed for a single system being tested. It is often the case that more than one system is tested concurrently during a reliability growth program though, and this section provides straightforward extensions of the results in Section 8.2.3 for multiple systems under test.

For q systems under test, assume each system has likelihood as defined in (1). Assuming independence between the systems, the total likelihood for the i^{th} failure mode is given by

$$l\left(t \mid p_i, \nu_i, d_i, T\right) \propto p_i^{\sum_{j=1}^{q} n_{i,1,j}} \left(1-p_i\right)^{\sum_{j=1}^{q} \nu_{i,j} - n_{i,1,j}} \left(1-p_i\right)^{\left(1-d_i\right)\left[\sum_{j=1}^{q} T_j - \nu_{i,j} - \left(n_{i,j} - n_{i,1,j}\right)\right]} \left[1-\left(1-p_i\right)^{1-d_i}\right]^{\sum_{j=1}^{q} n_{i,j} - n_{i,1,j}}, \tag{27}$$

where the subscript j denotes data for the specific system under test. Examining (27) reveals that the likelihood for multiple systems can be found by summing the data across the individual systems and substituting the terms into (1). This indicates that

the results in Section 8.2.3 can be extended through the same substitution and require no additional model development.

8.2.5 Empirical Bayes Estimators

As with the continuous model in Chapter 3, it is entirely possible to parameterize the Beta prior through other means such as historical data or elicitation of experts. While not fully Bayesian in their application, empirical Bayes estimates can be developed for the discrete systems approach.

First note that the mean in (19) and the variance in (24) are now expressed in terms of the prior system-level mean R_I and the $\tilde{n}$ parameter. The empirical Bayes estimators for these parameters can be developed by examining the likelihood in (1) and the resulting marginal likelihood when all failure modes in the system are considered. The marginal likelihood for a single failure mode is just the denominator of the posterior distribution in (5) given by

$$p(n_i) = \frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma[a+n_{i,1}]\Gamma[b+v_i-n_{i,1}+\tau_{i,j}]}{\Gamma[a+b+v_i+\tau_{i,j}]}, \tag{28}$$

where $\tau_{i,j} = (1-d_i)(T-v_i-n_{i,2}+j)$ and n_i represents the total number of observed failures for the ith failure mode as in Section 8.2.3. From (28) the total likelihood over *K* modes in the system is given by

$$L(n) = \prod_{i=1}^{K} \frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma[a+n_{i,1}]\ \Gamma[b+v_i-n_{i,1}+\tau_{i,j}]}{\Gamma[a+b+v_i+\tau_{i,j}]}, \tag{29}$$

and the corresponding log-likelihood is given by

$$\mathrm{l}(n) = \log L(n) = \sum_{i=1}^{K} \left\{ \log\Gamma[a+b] - \log\Gamma[a] - \log\Gamma[b] + \log\left[\sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma[a+n_{i,1}]\ \Gamma[b+\nu_i - n_{i,1} + \tau_{i,j}]}{\Gamma[a+b+\nu_i+\tau_{i,j}]} \right] \right\} \tag{30}$$

Again assuming that *m* failure modes are observed during the test, (30) can be represented as

$$\mathrm{l}(n) = \log\left[\binom{K}{m} \prod_{i=1}^{m} \frac{\Gamma[a+b]}{\Gamma[a]\Gamma[b]} \sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma[a+n_{i,1}]\ \Gamma[b+\nu_i - n_{i,1} + \tau_{i,j}]}{\Gamma[a+b+\nu_i+\tau_{i,j}]} \right] + (K-m)\log\left[\frac{\Gamma[a+b]\Gamma[b+T]}{\Gamma[a+b+T]\Gamma[b]} \right] \tag{31}$$

The expression in (31) is a sum of log-likelihood terms for the *m* observed failure modes and the *K*-*m* unobserved failure modes. Also note the addition of the constant

$$\binom{K}{m} \tag{32}$$

to account for the possible ways of observing *m* failure modes from the total population of *K* modes. Reparameterizing (31) in terms of the prior system-level mean R_I and the $\tilde{n}$ parameter then yields

$$\mathrm{l}(n) = \log\left(\binom{K}{m} \prod_{i=1}^{m} \frac{\Gamma[\tilde{n}]\ \Gamma\left[\tilde{n}\left(1-R_I^{\frac{1}{K}}\right)+n_{i,1}\right]}{\Gamma\left[\tilde{n}\left(1-R_I^{\frac{1}{K}}\right)\right]\Gamma\left[\tilde{n}R_I^{\frac{1}{K}}\right]} \sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n}R_I^{\frac{1}{K}}+\nu_i-n_{i,1}+\tau_{i,j}\right]}{\Gamma[\tilde{n}+\nu_i+\tau_{i,j}]} \right) + (K-m)\log\left(\frac{\Gamma[\tilde{n}]\ \Gamma\left[\tilde{n}R_I^{\frac{1}{K}}+T\right]}{\Gamma[\tilde{n}+T]\Gamma\left[\tilde{n}R_I^{\frac{1}{K}}\right]} \right) \tag{33}$$

Taking the limit of (33) as *K* becomes large results in

$$\ell_{\infty}(n) = \lim_{K\to\infty} \ell(n) = \sum_{i=1}^{m} \log\left(\log R_I^{\,-\tilde{n}} \prod_{q=1}^{n_{i,1}-1} (n_{i,1} - q) \sum_{j=0}^{n_{i,2}} \binom{n_{i,2}}{j} (-1)^j \frac{\Gamma\left[\tilde{n} + \nu_i - n_{i,1} + \tau_{i,j}\right]}{\Gamma\left[\tilde{n} + \nu_i + \tau_{i,j}\right]} \right)$$

$$+ \log R_I^{\,\tilde{n}} \left[\psi(\tilde{n} + T) - \psi(\tilde{n})\right] \quad (34)$$

Taking the derivative with respect to the R_I and $\tilde{n}$ will result in the equations for the empirical Bayes estimates.

8.3 Combining Developmental and Operational Test Data for Discrete Systems

For combining developmental and operational test data results for one-shot systems, an approach analogous to that of Chapter 4 can be utilized. A degradation factor, γ, can be employed in the same manner as the FEF value in Section 8.2. The relationship between the DT reliability and the OT reliability can be represented as

$$R_{DT} = R_{OT}^{\,1-\gamma} \quad (35)$$

This would allow the marginal distribution for the operational reliability to be developed in manner analogous to that in Chapter 4. Note the Markov Chain Monte Carlo methods may need to be employed to develop the marginal posterior, as the resulting distribution may not be analytically tractable.

8.4 Developing Prior Information from Physics-based Modeling

The results presented in Chapters 6 and 7 for developing posterior distributions for mode failure rates are an additional area where much further work is possible. In particular, the Steinberg damage accumulation approach could be further explored for situations where the loading does not adhere to the assumption of Normality. Expansion in the area to allow for non-Normal distributions or other

means of damage accumulation from complex loading environments would help to provide more reasonable life estimates, thereby improving the accuracy of the failure mode posterior and the overall system level posterior.

8.5 General Reliability Assessment

The methodology presented throughout this thesis decomposes the reliability of a complex system into the various failure modes that exist within the system. The failure intensity of the various failure modes is also assumed to be constant, which is justified when a sufficient number of failure mechanisms give rise to the specific failure mode. But the constant failure intensity may not be reasonable when a single mechanism causes a failure mode, or in cases when the failure mode results from redundancy between components. Additional work could prove useful in these areas.

In particular the use of alternative likelihoods, such as the Weibull distribution given in (6) in Chapter 6, could prove useful for modeling these types of failure modes. A common prior distribution could potentially be used to allow for a connection between the modes such as the Gamma distribution utilized in this thesis. The approach may provide a more general reliability assessment framework that allows for different likelihoods while still providing an analytic result with significant practical utility.

9 CONCLUSIONS

The methodology presented in this thesis provides a model framework for assessing reliability and reliability growth utilizing data from a variety of potential sources throughout the development of the system. The approach addresses the differences that may exist in the test environments in which the data are collected, while also providing a probabilistic result that indicates the amount of uncertainty in the assessed values. Additional data sources, such as historical information, component level test data, and modeling and simulation results are also leveraged to provide prior assessments of the system reliability. The approach provides a flexible Bayesian framework that can easily be extended to allow for other information sources where appropriate. The only requirement for inclusion is a suitable characterization of the posterior distribution on the failure intensity for the failure mode profile underlying the system. The three main goals of the research were as follows:

1. Provide a reliability growth assessment methodology that utilizes data from throughout the development of the system while rigorously accounting for differences in the test environments that may exist.
2. Provide additional management metrics that will provide additional information beyond the assessed reliability to program managers in order to better inform decision-making.
3. Provide a methodology for combining early design activities such as modeling and simulation and component/subsystem testing to provide a prior distribution on the reliability of a system. The prior distribution can then be

updated when reliability growth testing and reliability demonstration testing are completed.

The research presented in Chapters 3-6 specifically addresses the goals that were identified above. Chapter 2 presented a literature review on reliability growth and assessment techniques that established the current state-of-the-art in reliability growth models. Chapter 3 presented a new method for projecting the reliability growth of a complex continuously operating system. The model allows for arbitrary corrective action strategies, while using all available data rather than failure mode first occurrence times only. A complete inference framework was also provided via the posterior distribution on the system failure intensity. A unique feature of this approach relative to other Bayesian techniques is the analytic expression for the failure intensity contribution from unobserved failure modes. Expressions for the estimating the initial failure intensity, growth potential failure intensity, and the cumulative number of failure modes expected in future testing are also developed. These important metrics provide additional useful reliability information to program managers, which can in turn make better informed decisions regarding resource management and future reliability initiatives.

Chapter 4 presented a new Bayesian reliability assessment model to mitigate the problem of reliability demonstration with fixed configuration testing alone. The approach combined the developmental reliability growth approach in Chapter 3 with operational reliability test data within a Bayesian probabilistic framework. The resulting model reduces the uncertainty in the assessed reliability below that associated with reliability assessment using operational test data alone. Differences

in the test environments and stressors are explicitly considered when combining the data from different test events. The approach developed in Chapter 4 is also used as the basis for the reliability growth planning model in Chapter 5.

Chapter 6 presents a general methodology for combining lower level information that may be available before full system-level testing has been conducted. System-level data collected under operationally relevant testing is the most desirable information source for assessing system reliability, but during early stages of development this type of information may be unavailable. There are often other sources of reliability related information that are available in the early stages of development of the system though, and in these cases it is possible to utilize these information sources to develop an early assessment of the system reliability. When viewing the process of reliability assessment across the various stages of development of the system, assessment in this manner serves as prior information that can be updated with the reliability growth models in Chapters 3 and 4 when system-level test data are available.

Uncertainty distributions are developed for each failure mode using data for the individual components or those within a redundant block. Bayesian posterior distributions are used when component data are available, and a more general uncertainty distribution is developed when only physics-of-failure model results are available. When the component information that is available does not accurately represent the reliability of the component within the new system (e.g. benign testing, historical data from similar system, etc.), a probabilistic technique is provided to account for the degraded reliability and additional uncertainty that is present due to

additional failure modes or mechanisms that are not accounted for in the data or modeling.

Chapter 7 provides a case study demonstrating the application of the overarching model framework to a complex military system. Historical data is used to develop posterior distributions for the major subsystems within the system. The data are also used to develop a posterior distribution on the failure intensity for the 4-of-8 configuration for the wheels of the vehicle. Physics-based fatigue modeling is used to develop a posterior distribution on the driveshaft of the vehicle, and the results are combined to form an overall posterior on the system level failure intensity.

The resulting posterior is then used to form the prior distribution for the reliability growth model in Chapter 3, and the system reliability is assessed after corrective action of 3 failure modes observed during reliability growth testing. The resulting posterior distribution after reliability growth testing is then updated after OT by applying a degradation factor between the two test events.

APPENDIX A: DERIVATIONS

This section contains derivations of the limiting results for the posterior variance and the expected number of failure modes observed in follow on testing from Chapter 3.

Posterior Variance:

From Section 3.2 we write the system level posterior variance as the sum of the variances for the individual failure mode failure intensities.

$$Var\left[\lambda_s \mid n\right] = Var\left[\sum_{i=1}^{K}\lambda_i \mid n\right] = \sum_{i=1}^{K} Var\left[\lambda_i \mid n\right] = \sum_{i=1}^{K}\left(\frac{(1-d_i)^2(\alpha+n_i)}{\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)^2}\right)$$

$$= \sum_{i=1}^{m}\left(\frac{(1-d_i)^2(\alpha+n_i)}{\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)^2}\right) + (K-m)\left(\frac{\alpha}{\left(\frac{1}{\beta}+T\right)^2}\right) \quad (1)$$

Reparameterizing allows us to express the variance as

$$Var\left[\lambda_s \mid n\right] = \sum_{i=1}^{m}\left(\frac{(1-d_i)^2\left(\frac{\lambda_B}{\beta K}+n_i\right)}{\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)^2}\right) + (K-m)\left(\frac{\frac{\lambda_B}{\beta K}}{\left(\frac{1}{\beta}+T\right)^2}\right)$$

$$= \sum_{i=1}^{m}\left(\frac{(1-d_i)^2\left(\frac{\lambda_B}{\beta K}+n_i\right)}{\left(\frac{1}{\beta}+v_i+(1-d_i)(T-v_i)\right)^2}\right) + \left(\frac{1}{\beta}-\frac{m}{K\beta}\right)\left(\frac{\lambda_B}{\left(\frac{1}{\beta}+T\right)^2}\right). \quad (2)$$

Taking the limit of (2) as K becomes large while holding the prior mean and the β parameter constant yields the expression shown in (10) in Chapter 3.

Expected Number of Failure Modes Observed in Follow-On Testing:

Reparameterizing (20) in Chapter 3 results in

$$E\left[\sum_{i=m+1}^{K} I_i(t) \mid t > T\right] = K\left[1 - \frac{1}{\left(1 + \frac{t-T}{\frac{1}{\beta}+T}\right)^{\frac{\lambda_B}{\beta K}}}\right] - m\left[1 - \frac{1}{\left(1 + \frac{t-T}{\frac{1}{\beta}+T}\right)^{\frac{\lambda_B}{\beta K}}}\right] \tag{3}$$

The second term multiplying the m in (3) will go to zero as K becomes large. For the limit of the first term multiplying K we can rewrite it as

$$K\left[1 - \frac{1}{\left(1 + \frac{t-T}{\frac{1}{\beta}+T}\right)^{\frac{\lambda_B}{\beta K}}}\right] = \frac{K\left[\left\{\left(1 + \frac{t-T}{\frac{1}{\beta}+T}\right)^{\frac{\lambda_B}{\beta}}\right\}^{\frac{1}{K}} - 1\right]}{\left(1 + \frac{t-T}{\frac{1}{\beta}+T}\right)^{\frac{\lambda_B}{\beta K}}} \tag{4}$$

The denominator of the exponent in (4) will go to zero as K becomes large. The limit of the numerator is by definition just

$$\log\left(1+\frac{t-T}{\frac{1}{\beta}+T}\right)^{\frac{\lambda_B}{\beta}}, \tag{5}$$

which yields the desired result in Chapter 3.

APPENDIX B: MATHEMATICA CODE FOR RELIABILITY GROWTH MODELS OF CHAPTERS 3 & 4

The results in Chapters 3 and 4 utilize a Mathematica simulation to generate test data for comparing the proposed approach with those currently in the literature. The simulation begins by simulating mode failure rates from a Gamma distribution. Each failure rate is then used to simulate failures within the desired length of testing. Corrective actions are applied with the test according to a probabilistic strategy that allows for certain modes to be corrected immediately, while others have corrective actions delayed until the end of the test. The individual mode failure rates were reduced by the corresponding FEF when the corrective action occurs in the simulation. The resulting failures and associated FEF values are then used within the reliability growth model. The model results can then be compared to the true values that were used to generate the random failures.

The failure rates remaining after the reliability growth test were then scaled according to the amount of reliability degradation that was assumed for the simulation. The scaled mode failure rates were then used to generate failures for the operational test in the same manner as for the reliability growth test. No corrective actions were employed during the OT though, as the model approach assumes a constant configuration. The resulting OT failures were then used within the model approach in Chapter 4, and the results can again be compared to the true values underlying the simulated failure times. The simulation code is as follows, with comment descriptions presented in-line within the code to explain the various calculations that are performed.

```
(* inputs *)
NumReps = 1000;

(* number of modes *)
k = 500;

(* initial gamma parameterization *)
β = 0.0001;
Eλ = 1 / 100;
α = Eλ / (k * β);

(* test lengths *)
Tprior = 10 000;
Tot = 500;

(* probability of fix happening within test phase *)
pfix = 0.1;

(* percentile for lower bound *)
ProbBound = 0.8;

(* true mean dt-ot drop *)
γ = 0.2;
(* find lagrange multiplier from mean *)
λg = λg /. FindRoot[1 / λg - Exp[-λg] / (1 - Exp[-λg]) == γ, {λg, 0.5}];


(* model estimated mean dt-ot drop to check sensitivity of estimate *)
γ1 = 0.15;
(* find lagrange multiplier from mean *)
λg1 = λg1 /. FindRoot[1 / λg1 - Exp[-λg1] / (1 - Exp[-λg1]) == γ1, {λg1, 0.5}];

(* equating 1st and 2nd moments of Beta and
 truncated exponential to get a and b for approximate Beta *)
Clear[a];
Clear[b];
{a, b} = {a, b} /. NSolve[{a / (a + b) * (a + 1) / (a + b + 1) == λg1 / (1 - Exp[-λg1]) *
          (-1 / λg1 * Exp[-λg1] + 2 / λg1 * (-1 / λg1 * Exp[-λg1] - 1 / λg1^2 * Exp[-λg1] + 1 / λg1^2)),
       a / (a + b) == 1 / λg1 - Exp[-λg1] / (1 - Exp[-λg1])}, {a, b}][[1]];


(* Fix Effectiveness Factors - using Beta distribution with mean c and cv =
 v from input section *)
c = 0.7;
v = 0.1;
q = (1 - c - c * (v^2)) / (v^2);
r = ((1 - c) / c) * q;

(* function to generate failures according to failure rate λ,
fix effectiveness factor d, fix time v, total time T, and 1st occurrence time *)
failuretimes[λ_, d_, v_, T_, i_, FOT_] := {t1[[i]] = {0, FOT}; While[Max[t1[[i]]] < T,
```

```
    If[Max[t1[[i]]] ≥ v, Newt = Max[t1[[i]]] - Log[RandomReal[]] / ((1 - d) * λ),
                  Temptime = Max[t1[[i]]] - Log[RandomReal[]] / λ;
                  If[Temptime < v, Newt = Temptime, Newt = v - Log[RandomReal[]] / ((1 - d) * λ)]];
                 If[Newt ≤ T, AppendTo[t1[[i]], Newt], Break[]]];

(* function to simulate failures and estimates *)
model[] := {

  t1 = Table[{0}, {i, 1, k}];

  λ = Table[RandomReal[GammaDistribution[α, β]], {i, 1, k}];
  If[Length[Position[λ, 0]] > 0, λ = ReplacePart[λ, 10^-10, Position[λ, 0]]];
  If[Length[Position[λ, 0.]] > 0, λ = ReplacePart[λ, 10^-10, Position[λ, 0.]]];
  λtrue = Total[λ];

  (* generate degradation values *)
  γ = Table[-Log[1 - RandomReal[] * (1 - Exp[-λg])] / λg, {i, 1, k}];

  (* generate random FEF values *)
  d = Table[Random[BetaDistribution[q, r]], {i, 1, k}];
  (*d=Table[0,{i,1,k}];*)
  AverageFEF = Mean[d];

  (* generate DT first occurrences *)
  FOTdt = Table[temp = RandomReal[ExponentialDistribution[λ[[i]]]];
    If[temp < Tprior, temp, Tprior], {i, 1, k}];

  (* corrective action times *)
  V = Table[If[FOTdt[[i]] < Tprior, If[RandomReal[] < pfix,
       FOTdt[[i]] + RandomReal[] * (Tprior - FOTdt[[i]]), Tprior], Tprior], {i, 1, k}];

  (* simulate DT phase failures -
   use result to determine prior on λ and system level failure rate estimate *)
  For[i = 1, i ≤ k, i++, If[FOTdt[[i]] < Tprior,
    failuretimes[λ[[i]], d[[i]], V[[i]], Tprior, i, FOTdt[[i]]], t1[[i]] = 0]];
  For[i = 1, i ≤ k, i++, t1[[i]] = DeleteCases[t1[[i]], 0]];

  (* number of observed failures *)
  nprior = Table[Length[t1[[i]]], {i, 1, k}];

  (* number of observed failure modes *)
  mprior = Length[DeleteCases[t1, 0]];

  (* true failure rate after reliability growth *)
  λdt = Table[If[nprior[[i]] > 0, (1 - d[[i]]) * λ[[i]], λ[[i]]], {i, 1, k}];

  (* estimate prior parameters for gamma distribution - mle's extended from M&W *)
  If[Total[nprior] > mprior,

   βhat = βhat /. FindRoot[(1 + βhat * Tprior) / (βhat^2 * Tprior) * Log[1 + βhat * Tprior] * Sum[
```

```
        nprior[[i]]
    ------------------------------------------------------, {i, 1, k}] == mprior, {βhat, β}];
    (V[[i]] + (1 - d[[i]]) * (Tprior - V[[i]]) + 1 / βhat)

λhat = (1 + βhat * Tprior) / (βhat * Tprior) * Sum[
        nprior[[i]]
    ------------------------------------------------------, {i, 1, k}];
    (V[[i]] + (1 - d[[i]]) * (Tprior - V[[i]]) + 1 / βhat)

(* generate system level DT results *)
                                                  (1 - d[[i]]) * nprior[[i]]
DTResult = Sum[If[nprior[[i]] > 0, ------------------------------------------------------, 0],
                                   ((1 - d[[i]]) * (Tprior - V[[i]]) + V[[i]] + 1 / βhat)
    {i, 1, k}] + λhat / (1 + βhat * Tprior);
                                                  (1 - d[[i]])^2 * nprior[[i]]
DTResultVar = Sum[If[nprior[[i]] > 0, ------------------------------------------------------,
                                      ((1 - d[[i]]) * (Tprior - V[[i]]) + V[[i]] + 1 / βhat)^2
     0], {i, 1, k}] + λhat / (βhat (1 / βhat + Tprior)^2);

(* updated beta parameter for system level gamma in OT prior *)
Newβ = DTResultVar / DTResult;

DTUB = If[DTResult^2 / DTResultVar > 0 && DTResultVar / DTResult > 0, Quantile[
    GammaDistribution[DTResult^2 / DTResultVar, DTResultVar / DTResult], ProbBound], 0];

(* AMPM calcs for comparison cases *)
(* Beta Infinite *)
BetaInfinite =
 x /. FindRoot[Log[1 + x * Tprior] * Sum[1 / (1 + x * FOTdt[[i]]), {i, 1, mprior}] -
     mprior * x * Tprior / (1 + x * Tprior) == 0, {x, β}];
(* Lambda Infinite *)
LambdaInfinite = mprior * BetaInfinite / Log[1 + BetaInfinite * Tprior];

(* AMPM failure intensity *)
ResultAMPM = (1 - c) * (LambdaInfinite - LambdaInfinite / (1 + Tprior * BetaInfinite)) +
  LambdaInfinite / (1 + Tprior * BetaInfinite);

(* Crow-Extended Calculations for comparison cases *)
(* AMSAA Tracking Model Estimates *)
x = Flatten[DeleteCases[t1, 0]];
βhatCrow = Sum[nprior[[i]], {i, 1, k}] /
  Sum[Log[Tprior / x[[i]]], {i, 1, Sum[nprior[[i]], {i, 1, k}]}];
ρhatT = (Sum[nprior[[i]], {i, 1, k}] * βhatCrow) / Tprior;

(* AMSAA Crow Projection Model Estimates *)
βhatCrowBD = mprior / Sum[Log[Tprior / FOTdt[[i]]], {i, 1, k}];
hhatBD = mprior * βhatCrowBD / Tprior;
                                                  (1 - d[[i]]) * nprior[[i]]
ρhatBD = Sum[If[nprior[[i]] > 0, ------------------------------------------------------, 0], {i, 1, k}] + c * hhatBD;
                                                  Tprior
```

```
(* Crow-Extended Model Estimates *)
ρhat = ρhatT - Sum[nprior[[i]], {i, 1, k}] / Tprior + ρhatBD;


(* updated FEF values for OT *)
d = Table[If[nprior[[i]] == 0, 0, d[[i]]], {i, 1, k}];

(* OT first occurrence times *)
FOTot =
 Table[temp = RandomReal[ExponentialDistribution[(1 - d[[i]]) * λ[[i]] / (1 - γ[[i]]))]];
  If[temp < Tot, temp, Tot], {i, 1, k}];

(* simulate OT failures *)
For[i = 1, i ≤ k, i++, If[FOTot[[i]] < Tot, failuretimes[
    (1 - d[[i]]) * λ[[i]] / (1 - γ[[i]]), 0, Tot, Tot, i, FOTot[[i]]], t1[[i]] = 0]];
For[i = 1, i ≤ k, i++, t1[[i]] = DeleteCases[t1[[i]], 0]];

(* observed OT failures *)
nOT = Table[Length[t1[[i]]], {i, 1, k}];

(* classical pt estimate and confidence bound *)
λestOT = N[Total[nOT] / Tot];
λUCB = Quantile[ChiSquareDistribution[2 * Total[nOT] + 2], 0.8] / (2 * Tot);

(* calculate combined posterior mean failure
 intensity estimate assuming infinite number of failure modes *)
Result = Sum[If[nOT[[i]] > 0, nOT[[i]] / (Tot + 1 / Newβ) * Hypergeometric2F1[a, nOT[[i]] + 1, a + b,
        1 / Newβ / (Tot + 1 / Newβ)] / Hypergeometric2F1[a, nOT[[i]], a + b, 1 / Newβ / (Tot + 1 / Newβ)], 0], {i, 1, k}] +
  DTResult / (1 + Newβ * Tot) * Hypergeometric2F1[a, 1, a + b, 1 / Newβ / (Tot + 1 / Newβ)];


(* posterior variance assuming infinite number of failure modes *)
ResultVar =
 Sum[If[nOT[[i]] > 0, (nOT[[i]] + 1) * nOT[[i]] / (Tot + 1 / Newβ)^2 * Hypergeometric2F1[a, nOT[[i]] + 2,
        a + b, 1 / Newβ / (Tot + 1 / Newβ)] / Hypergeometric2F1[a, nOT[[i]], a + b, 1 / Newβ / (Tot + 1 / Newβ)] -
    (nOT[[i]] / (Tot + 1 / Newβ) * Hypergeometric2F1[a, nOT[[i]] + 1, a + b, 1 / Newβ / (Tot + 1 / Newβ)])^2, 0], {i, 1, k}] +
  DTResult / (1 + Newβ * Tot) * (2 * Hypergeometric2F1[a, 1, a + b, 1 / Newβ / (Tot + 1 / Newβ)] -
    DTResult / (1 + Newβ * Tot));
```

```
        Hypergeometric2F1[a, nOT[[i]], a + b, 1 / Newβ / (Tot + 1 / Newβ)]) ^ 2, 0], {i, 1, k}] +
    DTResult / (Newβ * (1 / Newβ + Tot) ^ 2) * Hypergeometric2F1[a, 2, a + b, 1 / Newβ / (Tot + 1 / Newβ)];

  (* probability bounds *)
  PosteriorApproximateUB =
   Result + Quantile[NormalDistribution[0, 1], ProbBound] * Sqrt[ResultVar];
  PosteriorGammaUB = If[Result ^ 2 / ResultVar > 0 && ResultVar / Result > 0,
    Quantile[GammaDistribution[Result ^ 2 / ResultVar, ResultVar / Result], ProbBound], 0];

  (* number of failure modes surfaced *)
  SurfacedModes = Total[Table[If[nprior[[i]] > 0 || nOT[[i]] > 0, 1, 0], {i, 1, k}]];

  (* coverage calcuations *)
  If[PosteriorApproximateUB > Total[Table[(1 - d[[i]]) * λ[[i]] / (1 - γ[[i]]), {i, 1, k}]],
   BayesCount = 1, BayesCount = 0];
  If[PosteriorGammaUB > Total[Table[(1 - d[[i]]) * λ[[i]] / (1 - γ[[i]]), {i, 1, k}]],
   BayesCountGamma = 1, BayesCountGamma = 0];
  If[λUCB > Total[Table[(1 - d[[i]]) * λ[[i]] / (1 - γ[[i]]), {i, 1, k}]],
   ClassicalCount = 1, ClassicalCount = 0];
  If[DTUB > Total[λdt], DTCount = 1, DTCount = 0],

  (* set bayes values to 0 when n=m *)
  Result = 0;
  PosteriorApproximateUB = 0;
  BayesCount = 0;
  BayesCountGamma = 0;
  DTCount = 0;
  λ = Table[0, {i, 1, k}]];

 (* function output for each replication *)
 DTResult, DTResultVar, DTUB, Total[λdt], Result, ResultVar, PosteriorApproximateUB,
 Total[Table[(1 - d[[i]]) * λ[[i]] / (1 - γ[[i]]), {i, 1, k}]], λestOT, λUCB, Total[nprior],
 Total[nOT], BayesCount, ClassicalCount, DTCount, BayesCountGamma, ResultAMPM, ρhat
}

(* model call *)
CoverageResults = Table[model[], {i, 1, NumReps}];
For[i = 1, i ≤ NumReps, i++, If[CoverageResults[[i, 8]] == 0, CoverageResults[[i]] = 0]];
CoverageResults = DeleteCases[CoverageResults, 0];
CoverageResults = DeleteCases[CoverageResults, Total[Im[CoverageResults]] != 0];

(* output section - prints results along with plot of relative error *)
Print["β = ", β];
Print["α = ", α];

Print["Mean True MTBF = ",
```

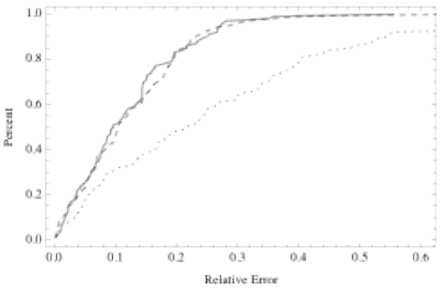
```
  Mean[Table[1 / CoverageResults[[i, 8]], {i, 1, Length[CoverageResults]}]]];
Print["Mean Bayes Post Mean = ",
  Mean[Table[1 / CoverageResults[[i, 5]], {i, 1, Length[CoverageResults]}]]];
Print["Mean Bayes LB = ",
  Mean[Table[1 / CoverageResults[[i, 7]], {i, 1, Length[CoverageResults]}]]];
Print["Mean Pt Est = ", Mean[Table[If[CoverageResults[[i, 9]] == 0,
     0, 1 / CoverageResults[[i, 9]]], {i, 1, Length[CoverageResults]}]]];
Print["Mean LCB = ", Mean[Table[1 / CoverageResults[[i, 10]],
    {i, 1, Length[CoverageResults]}]]];


BayesCov = N[
   Sum[CoverageResults[[i, 13]], {i, 1, Length[CoverageResults]}] / Length[CoverageResults]];
Print["Bayes Normal Coverage = ", BayesCov];
ClassicalCov = N[
   Sum[CoverageResults[[i, 14]], {i, 1, Length[CoverageResults]}] / Length[CoverageResults]];
Print["Classical Coverage = ", ClassicalCov];

Print["Mean Rel Error Bayes = ",
  Mean[Table[Abs[(CoverageResults[[i, 5]] - CoverageResults[[i, 8]])] /
     CoverageResults[[i, 8]], {i, 1, Length[CoverageResults]}]]];
Print["Mean Rel Error Classical = ", Mean[Table[
    Abs[(CoverageResults[[i, 9]] - CoverageResults[[i, 8]])] / CoverageResults[[i, 8]],
    {i, 1, Length[CoverageResults]}]]];

Print["Mean DT Failures = ",
  N[Mean[Table[CoverageResults[[i, 11]], {i, 1, Length[CoverageResults]}]]]];
Print["Mean OT Failures = ",
  N[Mean[Table[CoverageResults[[i, 12]], {i, 1, Length[CoverageResults]}]]]];


Print["Mean DT MTBF = ",
  Mean[Table[1 / CoverageResults[[i, 1]], {i, 1, Length[CoverageResults]}]]];
Print["Mean LB = ", Mean[Table[1 / CoverageResults[[i, 3]],
    {i, 1, Length[CoverageResults]}]]];
Print["Mean True DT = ", Mean[Table[1 / CoverageResults[[i, 4]],
    {i, 1, Length[CoverageResults]}]]];
Print["Mean AMPM = ", Mean[Table[1 / CoverageResults[[i, 17]],
    {i, 1, Length[CoverageResults]}]]];
Print["Mean Crow Extended = ", Mean[Table[1 / CoverageResults[[i, 18]],
    {i, 1, Length[CoverageResults]}]]];

Print["DT Coverage = ", N[Sum[CoverageResults[[i, 15]],
     {i, 1, Length[CoverageResults]}] / Length[CoverageResults]]];
Print["Rel Error DT = ", Mean[Table[Abs[(CoverageResults[[i, 1]] - CoverageResults[[i, 4]])] /
     CoverageResults[[i, 4]], {i, 1, Length[CoverageResults]}]]];
Print["Rel Error AMPM = ", Mean[Table[
    Abs[(CoverageResults[[i, 17]] - CoverageResults[[i, 4]])] / CoverageResults[[i, 4]],
    {i, 1, Length[CoverageResults]}]]];
Print["Rel Error Crow Extended = ", Mean[
   Table[Abs[(CoverageResults[[i, 18]] - CoverageResults[[i, 4]])] / CoverageResults[[i, 4]],
```

```
    {i, 1, Length[CoverageResults]}]]];


(* DT comparison plots for Bayes, AMPM, and Crow Extended *)
SortedresultDT = Sort[Table[Abs[(CoverageResults[[i, 1]] - CoverageResults[[i, 4]])] /
      CoverageResults[[i, 4]], {i, 1, Length[CoverageResults]}]];
SortedresultAMPM = Sort[Table[Abs[(CoverageResults[[i, 17]] - CoverageResults[[i, 4]])] /
      CoverageResults[[i, 4]], {i, 1, Length[CoverageResults]}]];
SortedresultCrowExt = Sort[Table[Abs[(CoverageResults[[i, 18]] - CoverageResults[[i, 4]])] /
      CoverageResults[[i, 4]], {i, 1, Length[CoverageResults]}]];
Show[ListPlot[Table[{SortedresultDT[[i]], i / Length[CoverageResults]},
   {i, 1, Length[CoverageResults]}], PlotStyle → Black, Joined → True,
  Axes → False, Frame → True, FrameLabel → {"Relative Error", "Percent"}],
 ListPlot[Table[{SortedresultAMPM[[i]], i / Length[CoverageResults]},
   {i, 1, Length[CoverageResults]}], PlotStyle → {Black, Dashed}, Joined → True],
 ListPlot[Table[{SortedresultCrowExt[[i]], i / Length[CoverageResults]},
   {i, 1, Length[CoverageResults]}], PlotStyle → {Black, Dotted}, Joined → True]]

(* OT comparison plots for Bayes vs classical *)
SortedresultBayes = Sort[Table[Abs[(CoverageResults[[i, 5]] - CoverageResults[[i, 8]])] /
      CoverageResults[[i, 8]], {i, 1, Length[CoverageResults]}]];
SortedresultClass = Sort[Table[Abs[(CoverageResults[[i, 9]] - CoverageResults[[i, 8]])] /
      CoverageResults[[i, 8]], {i, 1, Length[CoverageResults]}]];
ListPlot[{Table[{SortedresultBayes[[i]], i / Length[CoverageResults]},
   {i, 1, Length[CoverageResults]}], Table[
   {SortedresultClass[[i]], i / Length[CoverageResults]}, {i, 1, Length[CoverageResults]}]},
 PlotStyle → {Black, Dashed}, Joined → True, Axes → False, Frame → True,
 FrameLabel → {"Relative Error", "Percent"}]
MeanTrue = Mean[Table[CoverageResults[[i, 8]], {i, 1, Length[CoverageResults]}]];
VarTrue = Variance[Table[CoverageResults[[i, 8]], {i, 1, Length[CoverageResults]}]];
```

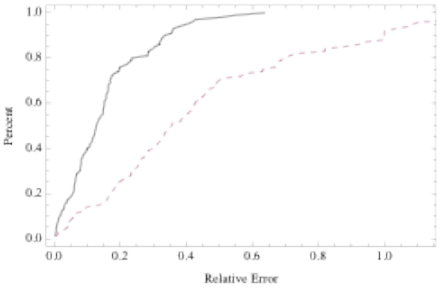
```
β = 0.0001

α = 1.

Mean True MTBF = 142.712

Mean Bayes Post Mean = 159.538

Mean Bayes LB = 146.529

Mean Pt Est = 167.508

Mean LCB = 98.3025

Bayes Normal Coverage = 0.47

Classical Coverage = 0.85

Mean Rel Error Bayes = 0.157244

Mean Rel Error Classical = 0.458104

Mean DT Failures = 98.88

Mean OT Failures = 4.14

Mean DT MTBF = 201.572

Mean LB = 185.984

Mean True DT = 212.764

Mean AMPM = 201.716
```

```
Mean Crow Extended = 189.62

DT Coverage = 0.87

Rel Error DT = 0.12062

Rel Error AMPM = 0.125908

Rel Error Crow Extended = 0.262734
```

Out[114]=



Out[117]=



REFERENCES

[1] U.S. Department of Defense, Defense Science Board Task Force Report on Developmental Test and Evaluation, Washington, D.C., 2008.

[2] P. O'Connor, *Practical Reliability Engineering-Fourth Edition*, John Wiley and Sons, West Sussex, England, 2002.

[3] M. Rhoads, et al., "Design for reliability handbook," AMSAA TR-2011-24, Aberdeen Proving Ground, MD, 2011.

[4] U.S. Department of Defense, Directive-Type Memorandum (DTM) 11-003 – Reliability Analysis, Planning, Tracking, and Reporting, Washington, D.C., 2011.

[5] U.S. Department of the Army, "Reliability of U.S. Army Materiel Systems," Washington D.C., 2011.

[6] P.M. Ellner, et al., "AMSAA reliability growth guide," AMSAA TR-652, Aberdeen Proving Ground, MD, 2000.

[7] M. Wayne and P. Ellner, "Comparing the robustness of reliability growth projection models," IEEE Proceedings of the Annual Reliability and Maintainability Symposium, San Jose, CA, 2010.

[8] ISO8402, Quality Vocabulary, International Standards Organization, Geneva, 1986.

[9] MIL-HDBK-781A, Reliability Test Methods, Plans, and Environments for Engineering, Development Qualification, and Production, 1996.

[10] A.C. Tamhane and D.D. Dunlap, *Statistics and Data Analysis*, Prentice Hall, Upper Saddle River, NJ, 2000.

[11] J.T. Duane, "Learning curve approach to reliability monitoring," IEEE Transactions on Aerospace, 2, 1964.

[12] J.D. Selby and S.G. Miller, "Reliability Planning and Management," Proceedings ASQC/SRE Seminar, Milwaukee, WS, 1970.

[13] U.S. Department of Defense, Reliability Growth Management, "Military Handbook 189C," Washington D.C., 2011.

[14] L.H. Crow, "Reliability analysis for complex, repairable systems in reliability and biometry," SIAM, Philadelphia, PA, 1974.

[15] M. McCarthy, et al. "Developing a subsystem reliability growth program using the subsystem reliability growth planning model," AMSAA TR-555, Aberdeen Proving Ground, MD, 1994.

[16] P.M. Ellner and R. Mioduski, "Operating Characteristic Analysis for Reliability Growth Programs," AMSAA TR-524, Aberdeen Proving Ground, MD, 1992.

[17] P.M. Ellner and J.B. Hall, "An approach to reliability growth planning based on failure mode discovery and correction using AMSAA projection methodology," IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 2006.

[18] P. Ellner and Wald, "AMSAA maturity projection model," IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 1995.

[19] L.H. Crow, "Planning a reliability growth program utilizing historical data," IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 2011.

[20] L.H. Crow, "An extended reliability growth model for managing and assessing corrective actions," IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 2004.

[21] J.B. Hall, "Reliability growth planning for discrete-use systems," IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 2011.

[22] J.B. Hall, P.M. Ellner, and A. Mosleh, “Reliability growth management metrics and statistical methods for discrete-use systems,” Technometrics, 52-4, 2010.

[23] W.J. Corcoran, H. Weingarten and P.W. Zehna, “Estimating reliability after corrective action,” Management Science, 10-4, 1964.

[24] L.H. Crow, “An improved methodology for reliability growth projection,” AMSAA TR-357, Aberdeen Proving Ground, MD, 1982.

[25] J.A. Clark, “Modeling reliability growth late in development,” IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 1999.

[26] P. Ellner and J.B. Hall, “AMSAA maturity projection model based on stein estimation,” AMSAA TR-751, Aberdeen Proving Ground, MD, 2004.

[27] C.M. Stein, “Estimation of the mean of a multivariate normal distribution,” The Annals of Statistics, 9, 1981.

[28] L. H. Crow, “On Reliability Growth Tracking,” AMSAA Interim Note R-30, Aberdeen Proving Ground, MD, 1974.

[29] J. B. Hall and A. Mosleh, “A reliability growth projection model for one-shot systems,” IEEE Transactions on Reliability, 57-1, 2008.

[30] J.B. Hall and A. Mosleh, “An analytical framework for reliability growth of one-shot systems,” Reliability Engineering and System Safety, 98:1751-1760, 2008.

[31] J.B. Hall and A. Mosleh, “Bayesian methods for evaluating discrete reliability growth,” IEEE Proceedings of the Annual Reliability and Maintainability Symposium, Fort Worth, TX, 2009.

[32] H.K. Weiss, “Estimation of Reliability Growth in a Complex System with Poisson-Type Failure,” Operations Research, 4-5, 1956.

[33] D.K. Lloyd and M. Lipow, *Reliability Growth Models; Reliability: Management, Methods, and Mathematics*, Englewood Cliffs, New Jersey: Prentice-Hall, 1962.

[34] D. R. Cox and P. A. Lewis, *The Statistical Analysis of Series of Events*. New York, NY, Wiley and Sons, 1966.

[35] R.E. Barlow and E.M. Scheuer, "Reliability growth during a development testing program," Technometrics, 8-1, 1966.

[36] S.M. Pollock, "A bayesian reliability growth model," IEEE Transactions on Reliability, R17-4, 1968.

[37] B. Littlewood and J.L. Verrall, "A bayes reliability growth model for computer software," J. Royal Stat. Soc. C, 22-3, 1973.

[38] A.F.M. Smith, "A bayesian note on reliability growth during a development testing program," IEEE Transactions on Reliability, R26-5, 1977.

[39] N.S. Fard and D.L. Dietrich, "A bayes reliability growth model for a development testing program," IEEE Transactions on Reliability, R36-5, 1987.

[40] M. Engelhardt, and L.J. Bain, "Prediction intervals for the Weibull Process," Technometrics, 20-2, 1978.

[41] M.C. Weinrich and A.J. Gross, "The Barlow-Scheuer reliability growth model from a Bayesian viewpoint," Technometrics, 20-3, 1978.

[42] N.A. Langberg and F. Proschan, "A reliability growth model involving dependent components," The Annals of Probability, 7-6, 1979.

[43] A.L Goel and K. Okumoto, "Time-dependent error detection rate model for software reliability and other performance measures," IEEE Transactions on Reliability, 28, 1979.

[44] Z. Jelinski and P.B. Moranda, “Software reliability research,” in *Statistical Computer Performance Evaluation*, Academic Press, 1972.

[45] L.H. Crow, “AMSAA discrete reliability growth model,” AMSAA Office Note 1-83, Aberdeen Proving Ground, MD, 1983.

[46] B. Littlewood, “Rationale for a modified Duane model,” IEEE Transactions on Reliability, R33-2, 1984.

[47] D. Robinson and D. Dietrich, “A nonparametric Bayes reliability growth model,” IEEE Transactions on Reliability, R38-5, 1989.

[48] M. Guida, R. Calabria, and G. Pulcini, “Bayes inference for a non-homogeneous poisson process with power intensity law,” IEEE Transactions on Reliability, R38-5, 1989.

[49] N. Singpurwalla and R. Soyer, “Non-homogeneous autoregressive processes for tracking (software) reliability growth, and their Bayesian analysis,” Journal of the Royal Statistical Society B, 54-1, 1992.

[50] T.A. Mazzuchi and R. Soyer, “Reliability assessment and prediction during product development,” IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 1992.

[51] A. Fries, “Discrete reliability-growth models on learning-curve property,” IEEE Transactions on Reliability, 42:303-306, 1993.

[52] I. Fakhre-Zakeri and E. Slud, “Mixture models for reliability of software with imperfect debugging: identifiability of parameters,” IEEE Transactions on Reliability, R44-1:104-113, 1995.

[53] A. Erklani et al., “Bayesian computations for a class of reliability growth models,” Technometrics, 40-1, 1998.

[54] L. Walls and J. Quigley, “Building prior distributions to support Bayesian reliability growth modeling using expert judgment,” Reliability Engineering and System Safety, 74:117-128, 2001.

[55] J. Yu et al., “Predictive analyses for nonhomogeneous poisson processes with power law using bayesian approach,” Computational Statistics and Data Analysis, 51, 2007.

[56] D. Li et al., “Building reliability growth model using sequential experiments and the bayesian theorem for small datasets,” Expert Systems with Applications, 37, 2010.

[57] Y.Y. Xing et al., “Dynamic bayesian evaluation method for system reliability growth based on in-time correction,” IEEE Transactions on Reliability, R59-2, 2010.

[58] J. Quigley and L. Walls, “Mixing Bayes and empirical Bayes inference to anticipate the realization of engineering concerns about variant system designs,” Reliability Engineering and System Safety, 96:933-941, 2011.

[59] B. Bichon, et. al., “Efficient surrogate models for reliability analysis of systems with multiple failure modes,” Reliability Engineering and System Safety, 96:1386-1395, 2011.

[60] R. Strunz and J.W. Herrmann, “Planning, tracking, and projecting reliability growth: a Bayesian tutorial,” IEEE Proceedings of the Annual Reliability and Maintainability Symposium, Reno, NV, 2012.

[61] Pievatolo, et. al., “A Bayesian hidden Markov model for imperfect debugging,” Reliability Engineering and System Safety, 103:11-21, 2012.

[62] H. Okamura, et. al., “Software reliability growth models with normal failure time distributions,” Reliability Engineering and System Safety, 116:135-141, 2013.

[63] S. Wang, et. al., “Discrete nonhomogeneous Poisson process software reliability growth models based on test coverage,” Quality and Reliability Engineering International, 29:103-112, 2013.

[64] O.M. Yadav, N. Singh, and P.S. Goel, ”Reliability demonstration test planning: a three dimensional consideration,” Reliability Engineering and System Safety, 91:882-893, 2006.

[65] T.H. Fan and C.C. Chang, “A Bayesian zero-failure reliability demonstration test of high quality electro-explosive devices,” Quality and Reliability Engineering International, 25:913-920, 2009.

[66] H. Guo and H. Liao, “Methods of reliability demonstration testing and their relationships,” IEEE Transactions on Reliability, 61-1, 2012.

[67] E.A. Elsayed, “Overview of reliability testing,” IEEE Transactions on Reliability, 61-2, 2010.

[68] L.H. Crow, “Demonstrating reliability growth requirements with confidence,” IEEE Proceedings of the Annual Reliability and Maintainability Symposium, Reno, NV, 2012.

[69] G.W. Miller, “An extension to the Weibull process model,” AMSAA TR-341, Aberdeen Proving Ground, MD, 1981.

[70] D. Cotroneo, R. Pietrantuono, and S. Russo, “Combining operational and debug testing for improving reliability,” IEEE Transactions on Reliability, 62-2, 2013.

[71] R.R. Hill, A.J. Gutman, S.P. Chambal, and J.W. Kitchen, “Acquisition and testing, DT/OT testing: the need for two parameter requirements,” Quality and Reliability Engineering International, 29:691-697, 2013.

[72] M. Hamada, H.F. Martz, C.S. Reese, T. Graves, V. Johnson, and A.G.Wilson, “A fully Bayesian approach for combining multilevel failure information in fault tree quantification and optimal follow-on resource allocation,” Reliability Engineering and System Safety, 86:297-305, 2004.

[73] C.S. Reese, M. Hamada, and D. Robinson, “Assessing system reliability by combining multilevel data from different test modalities,” Quality Technology and Quantitative Management, 2-2:177-188, 2005.

[74] F. Groen and E.L. Droguett, “Competing failure mode modeling in a Bayesian reliability assessment tool,” IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 2005.

[75] F. Groen, et. al., “Reliability data collection and analysis system,” IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 2004.

[76] A.G.Wilson, T. Graves, M. Hamada, and C.S. Reese, “Advances in data combination, analysis and collection for system reliability assessment,” Statistical Science, 21-4:514-531, 2006.

[77] O.M. Yadav, N. Choudhary, and C. Bilen, “Complex system reliability estimation methodology in the absence of failure data,” Quality and Reliability Engineering International, 24:745-764, 2008.

[78] R. Pan, “A Bayes approach to reliability prediction utilizing data from accelerated life tests and field failure observations,” Quality and Reliability Engineering International, 25:229-240, 2009.

[79] A.G. Wilson, C.M. Anderson-Cook, and A.V. Huzurbazar, “A case study for quantifying system reliability and uncertainty,” Reliability Engineering and System Safety, 96:1076-1084, 2011.

[80] S.M. Brown, “Development and validation of methodology for fix effectiveness projection during product development,” Ph.D. Dissertation, University of Maryland, College Park, Dept. Of Mechanical Engineering, 2009.

[81] H.F. Martz and R.A. Waller, *Bayesian Reliability Estimation*, John Wiley and Sons, New York, NY, 1982.

[82] L.H. Crow, “Methods for assessing reliability growth potential,” IEEE Proceedings of the Annual Reliability and Maintainability Symposium, 1984.

[83] V.E. Johnson, “A bayesian chi-square test for goodness of fit,” The Annals of Statistics, 32-6, 2004.

[84] M.S. Hamada, et al., *Bayesian Reliability*, Springer Series in Statistics, 2008.

[85] F.J. Samaniego, D.L. Steffey, and H. Tran, “Linear data fusion,” Proceedings of the Fourth Annual Army Conference on Applied Statistics, 1998.

[86] J. Rice, *Mathematical Statistics and Data Analysis*, Duxbury Press, Belmont, CA, 1995.

[87] D.S. Sivia, *Data Analysis – A Bayesian Tutorial*, Oxford University Press, New York, NY, 2006.

[88] E.T. Jaynes. *Probability Theory- The Logic of Science*, Cambridge University Press, Cambridge, United Kingdom, 2003.

[89] P. Bickel and K. Doksum, *Mathematical Statistics*, Pearson Prentice Hall, Upper Saddle River, NJ, 2007.

[90] K.C. Kapur and L.R. Lamberson, *Reliability in Engineering Design*, New York: John Wiley & Sons, Inc, 1977.

[91] R.F. Drenick, "The failure law of complex equipment," Journal of the Society for Industrial Applied Mathematics, 8:680-690, 1960.

[92] J.A. Collins, *Failure of Materials in Mechanical Design: Analysis, Prediction, Prevention*, Wiley & Sons, New York, NY, 1993.

[93] M. Rousand and A. Hoyland, *System Reliability Theory: Models, Statistical Methods, and Applications*, Wiley & Sons, Hoboken, NJ, 2004.

[94] M. Modarres, M. Kaminskiy, and V. Krivtsov, *Reliability Engineering and Risk Analysis: A Practical Guide*, Taylor and Francis, Boca Raton, FL, 1999.

[95] D. Steinberg, *Vibration Analysis for Electronic Equipment*, Wiley & Sons, New York, NY, 1988.