9-9-2010

Quantifying uncertainty in reliability block diagrams

Bea Yu

Bea Yu

Candidate

Mathematics

Department

This thesis is approved, and it is acceptable in quality and form for publication:

Approved by the Thesis Committee:

,Chairperson

QUANTIFYING UNCERTAINTY IN RELIABILITY BLOCK DIAGRAMS

BY

BEA YU

B.S., PHILOSOPHY, UNIVERSITY OF TEXAS, 1996
M.S., CHEMISTRY, UNIVERSITY OF NEW MEXICO, 2007
M.S., MATHEMATICS, UNIVERSITY OF NEW MEXICO, 2010

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Mathematics

The University of New Mexico
Albuquerque, New Mexico

July, 2010

ACKNOWLEDGMENTS

I would like to thank my applied mathematics advisor, Dr. Helen Wearing, for orchestrating the effort to turn this project from my work, based in statistics, into an applied math thesis. Since we started working together, she has been insightful, consistent, supportive and positive; in sum, a terrific adviser.

I also thank my committee member and statistics advisor, Dr. Curtis Storlie for agreeing to oversee this project during a very busy time in his career. He provided guidance as well as interesting leads in the project, and, being that my acquaintance with statistical theory spans no more than a year, I could not have followed it through to completion without his direction and oversight.

I am grateful that Dr. Jim Ellison has generously agreed to take time out of his summer to sit on my committee. And to Mr. Dan Telford, bearer of cryptic, thought-provoking questions, for planting the seed that grew into this project, my thesis and a useful technique for conducting reliability analysis.

AFOTEC, the Air Force Operational Test and Evaluation Center, supported me financially for the year spanning the development and codification of this analysis project. At AFOTEC, I was given the opportunity to learn statistics and apply what I learned to real-world, technical problems that have repercussions in social and political realms. The past year has been a rich and dynamic one for me both intellectually and in the development of my career.

Last, but not least, I thank my family for patiently supporting me through this and my chemistry MS program. I know it seemed like I would be perpetually in school and perpetually short on time and sleep. Of course, there is still the tempting specter of a PhD looming in the background…

QUANTIFYING UNCERTAINTY IN RELIABILITY BLOCK DIAGRAMS

BY

BEA YU

ABSTRACT OF THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
Mathematics

The University of New Mexico
Albuquerque, New Mexico

July, 2010

Quantifying Uncertainty in Reliability Block Diagram Models

BY

Bea Yu

B.A., Philosophy, University of Texas at Austin, 1996
M.S., Chemistry, University of New Mexico, 2007
M.S., Mathematics, University of New Mexico, 2010

ABSTRACT

Reliability analysis yields statistically derived technical system performance estimates. Traditional reliability analysis employs classical statistical techniques predicated upon asymptotic properties of large data sets. Not uncommonly, however, medium to small data sets constrain analysis efforts for high risk systems characterized by significant danger or cost. This paper outlines a general reliability analysis paradigm to contend with small to medium data sets. Preliminary sensitivity analysis using scatter plots and tests for non-randomness reveals component-level drivers in system-level performance measures. Comprehensive data collection efforts targeting all available, high-quality information sources decrease and allow analysts to estimate uncertainty in model parameters describing driving component performance. Bayesian analysis accumulates these data into posterior distributions summarizing all available performance knowledge about driving components. Sampling-based uncertainty propagation methods then transform component-level posterior distributions into system-level parent and sampling distributions. Reliability metric point-estimates and credible intervals estimate the

system reliability and benchmark the quality of the estimates, respectively. An operational reliability assessment of the B-2 Radar Modernization Program (B2-RMP) modernized radar system demonstrates the mechanics of the analysis paradigm applied to real data. Results from analysis including uncertainty explicitly modeled in all B-2 RMP components benchmark results from analysis including uncertainty modeled for driving components only.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Overview

"Beware of false knowledge; it is more dangerous than ignorance." George Bernard Shaw

"Information is a source of learning. But unless it is organized, processed, and available to the right people in a format for decision making, it is a burden, not a benefit." William Pollard

1.1 Introduction

Mathematical models of real systems allow us to interpolate and extrapolate system characteristics through time and space. However, models are, by definition, approximations of what is real, and as such never include all information impacting system properties. Model approximations can be known and deliberate, known and unavoidable or unknown and unavoidable [1]. Known and deliberate approximations are made in the interest of efficiency when their impact on results is minor relative to some standard of accuracy and precision. Known and unavoidable approximations are made due to an understood lack of knowledge about the system or due to inherent stochasticity in the system. Unknown and unavoidable approximations occur when the system is so poorly understood that analysts are unaware of the extent to which they lack important knowledge about it. Known approximations can be explicitly defined and their effect on model results can often be quantified. When model results inform decision making processes that entail significant environmental, human health or financial risk, accompanying definitions and estimates of the approximations used in the model help decision makers benchmark the quality of the information contained in the results.

The development and procurement of technical systems are often associated with non-trivial resource expenditures. These systems can also pose potential hazards to human health and the environment. As a consequence, technical system performance estimates play an important role in industrial and sociopolitical decision making processes. This paper focuses on methods to quantify the approximations involved in the estimation of technical system performance metrics using reliability block diagram models and demonstrates an approximation analysis in a case study of the B-2 modernized radar operational reliability assessment.

1.2 Reliability Analysis

Less than a century ago, the field of technical system reliability analysis arose in the interest of assessing variation in airplane operational safety as a function of engine number [2]. Since then, mass-production and rapidly evolving technological sophistication in high-value/ high-consequence systems has inspired concurrent evolution in reliability analysis sophistication and rigor. Lines of inquiry in reliability investigations fall into one of three idiosyncratic domains: human reliability, software reliability and hardware reliability [2]. Operations research analysts explore how reliability estimates in these three domains combine to produce an overarching estimate of human-in-the-loop, technical system reliability.

A subdivision within hardware reliability techniques occurs along the lines of physics based versus actuarial-based approaches [2]. Physics-based approaches model the

strength, S, of a technical item and the load it will bear, L, as random variables. The reliability of the item is defined as the probability that S is greater than L:

$$R = Pr(S>L)$$

Modifications of this equation to incorporate multiple dimensions including time are standard. In the actuarial approach, all reliability information is subsumed in a time-to-failure (or a time-between failure if the system is repairable, cf. below) random variable, Y (X if the system is repairable); a proxy for explicitly modeled physical variables such as load and strength.

Technical systems undergo periods of reliability growth and decay. System reliability during these periods is often determined using reliability distributions with time-dependent parameters [3]. When not experiencing reliability growth or decay, a system is said to be in a steady-state. Steady-state reliability regimes are easier to model since reliability distributions during these periods have time-independent parameters. During steady-state periods in the life history of a non-repairable system, the time-to-failure distribution, F, is the generator of all actuarial-based reliability metric estimates and distributions [2,3]. In particular, the reliability distribution of a non-repairable system, R, is defined as $R = 1 - F$. Thus, actuarial-based reliability measures the likelihood a system will not fail during a particular time interval.

System reliability analysis explores how component-level reliability combines to affect system-level reliability. This includes structural and compositional considerations. The counter-intuitive result that a system composed of highly reliable components may still exhibit low reliability is a consequence of emergent properties associated with the structure, rather than the composition, of the system [2]. Somewhat more intuitively, the location and dispersion of component-level time-to-failure or time-between-failures distributions describing the random variable Y or X for each component, i, also influence system-level reliability.

Repairable and non-repairable system reliability investigations differ in complexity and descriptive metrics. Common, non-repairable system reliability metrics include the reliability function and mean time-to-failure (MTTF or $E(Y)$, where Y is the random variable representing time-to-failure events). Repairable system reliability analysis includes all metrics and distributions found in non-repairable system reliability along with others that summarize system repair characteristics and operational capacity given that the system is inoperable for time intervals due to repair events. Availability, mean time-between-failures (MTBF or $E(X)$, where X is the random variable representing the time the system is up), maintainability, and mean down-time (MDT or $E(D)$, where D is the random variable representing the time the system is down) are typical metrics of repairable system reliability [2,3]. Mean-time-between-failures is a measure of the time the system is operating (e.g. not being repaired). Reliability metric nomenclature can be misleading and care should be taken to investigate the formulae defining the metrics to

ensure proper interpretation. Since repairable system models incorporate two random varibles, *X* and *D,* the complexity of the analysis involved scales accordingly.

Actuarial-Based Reliability Analysis Using Reliability Block Diagrams

Reliability block diagrams (RBDs) are commonly used models in actuarial-based reliability investigations. They provide both abstraction, leading to ease of solution and system representation, and solution accuracy for a broad class of systems. System models using RBDs incorporate independent, component-level random variable distributions, such as time-to-failure distributions. The resolution of the system model can be set arbitrarily based on how many component distributions it includes. Component reliability distributions are represented schematically by blocks. The architecture of the RBD represents how block random variable distributions combine recursively to determine a random variable distribution of the system as a whole.

The basic recursive relationships are represented with series and parallel block structures (cf. Figure 1 below). In the case where the component distributions are time-to-failure distributions, a series structure including a set of blocks fails if one or more of the blocks in the structure fail.

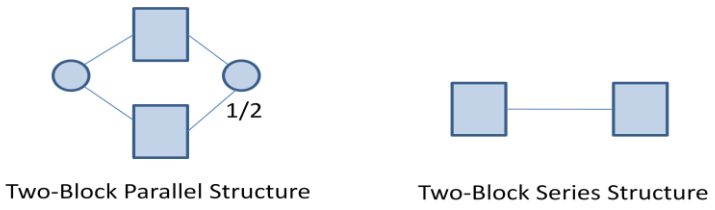


Figure 1: Basic Recursive Structures in an RBD

The analyst sets failure behavior of a parallel structure as a function of component block failure to reflect how redundancy has been engineered into the system to improve reliability. For instance, a parallel structure containing three blocks may fail if one, two or three component blocks fail concurrently. Frequently, a fraction denotes the failure behavior of a parallel structure. The denominator of this fraction is the total number of blocks in the parallel structure while the numerator is the number of blocks that must operate concurrently for the structure to operate. RBD models usually display these fractions in an obvious manner in relation to each parallel structure. A parallel structure that fails when only one component block fails is isomorphic to a series structure containing the same number of blocks.

In a well-formed RBD, connections between the components and their positions relative to one another clearly manifest the manner in which individual component reliability combines to affect system reliability as a whole. RBD components are generally defined so that component random variables are independent. Some RBD software packages allow users to define dependencies between components. When this is not sufficient, other models, such as Markov Chains, are more appropriate [2].

Figure 2 depicts a small RBD model. For simplicity, assume this is a non-repairable system and that the F_{ja}, $j=1,2,..,5$ are time-to-failure distributions. Note the ½ following each of the two parallel structures in this diagram. This indicates that the parallel structure remains operating if one out of the two constituent components is operating.

Appendix A outlines the process of computing the analytical time-to-failure distribution for this system.

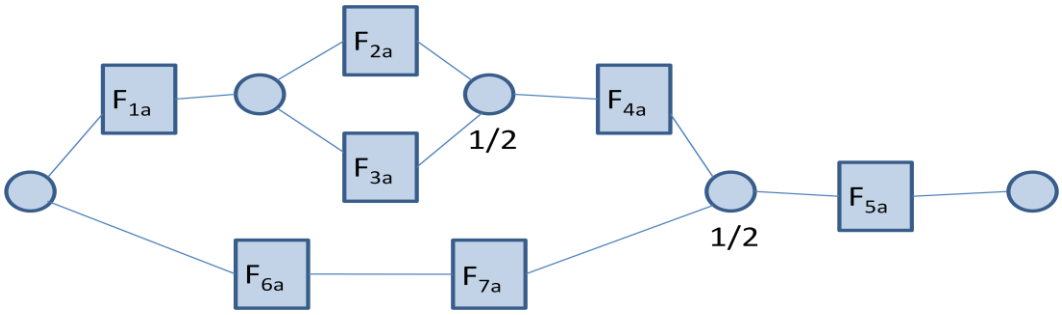


Figure 2: A simple RBD composed of series and parallel structures

Monte-Carlo Simulation Applied to RBDs

For simple RBDs like the one depicted in Figure 2, it is possible to analytically define the system random variable distributions as functions of block random variable distributions. While it is in principle possible to evaluate these functions, it is often impractical for systems with many components and parallel architectures. Even defining a functional relationship between block random variables and system random variables is impractical for many classes of sufficiently complex systems. Monte Carlo (MC) simulation is an attractive alternative to explicit evaluation of RBD system-level random variable distributions. The goal of generic MC simulation is to estimate an unknown parameter or unknown parameters, such as the mean and standard deviation, of some distribution. In the MC community, this distribution is called the "parent distribution" [4]. The parent distribution for a RBD is the system-level random variable distribution of interest, e.g. the system time-to-failure distribution, denoted F_a.

Using a random sampling technique, one MC run generates one point in the parent distribution. Multiple runs constitute an MC simulation and result in an estimate of the parent distribution. Thus, with infinite time and/or resources, it is theoretically possible to exactly form a parent distribution from block random variable distributions. However, practical limits on resources dictate that the MC generated parent distribution will always be approximate. Such an approximation amounts to a known error due to finite, random sampling in the algorithm.

From the approximate parent distribution, it is possible to calculate an estimate of its unknown parameter(s) using a function (usually a linear combination of the points comprising the parent distribution) called an "estimator" [4,5]. The estimator, being a function of parent distribution data generated from randomly selected input distribution data, is itself a random variable. Repeated runs generate a distribution of the estimator, called "the sampling distribution." A sampling distribution is the distribution of a parameter corresponding to a parent distribution. Figure 3 schematically represents the flow of generic MC simulation applied to an RBD model.

Thus, MC simulations generate estimations of unknown parameters that have a known, quantifiable error. Sampling distribution dispersion arises from the randomly, finitely determined parent distribution. This, in turn, is a consequence of random, finite sampling data from the input distributions used to define the parent distribution. However, according to The Central Limit Theorem, as the number of points used to generate the

sampling distribution increases, the sampling distribution becomes approximately normal with mean equal to the parameter of interest [5].

Commercial software programs that use RBDs with Monte Carlo Methods implement various algorithms [17]. It is important to consult user manuals to understand software specific algorithms. Different Monte Carlo algorithms entail different standard error in the sampling distributions. The solutions for some systems may diverge beyond an acceptable tolerance due to algorithmic differences [6]. Determining which package to use under such circumstances would entail more detailed analysis of the interplay between the system of interest and the structure of the different software algorithms.

The ultimate goal of the RBD/MC simulation is to find unbiased, minimum variance estimators of the parameters of the system-level parent distributions such as time-between-failures distributions, F_a [4]. Generally, the user defines a relevant time span for the simulation based on system properties, such as estimated obsolescence, or segregated reliability regimes that should be simulated separately so that distribution parameters are approximately time-independent, such as short growth period intervals and steady-state. During this time span, the simulation samples randomly from the block distributions. Directly and indirectly, these samples constitute data points that combine to form system-level random variable distributions, i.e. the parent distributions. At the end of the simulation time span, system quantities such as reliability are computed as averages of these distributions over time as well as averages over system state space, e.g. the time-between- failures event space.

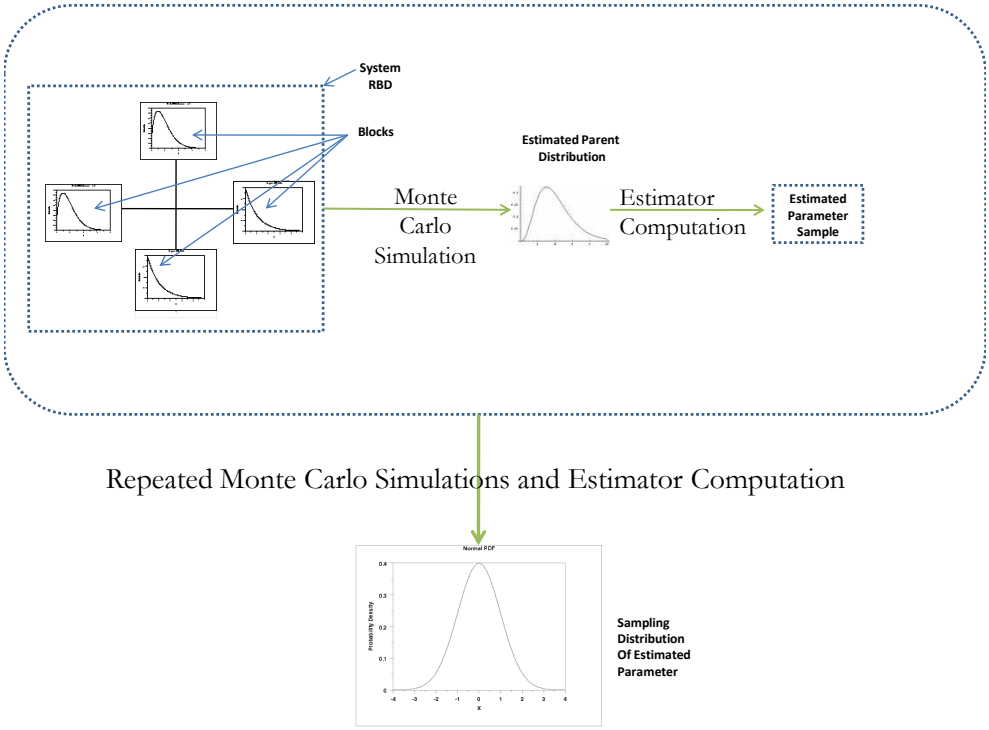


Figure 3: Flow of Monte Carlo analysis applied to RBD

1.3 Approximations in Analysis

As mentioned above, results derived from Monte Carlo simulations are approximate in the sense that they are computed using a finite number of samples from continuous input distributions. The magnitude of Monte Carlo approximation is a function of the number of points used in the estimator computation and the number of estimator points used to define the sampling distribution. Due to time and resource constraints as well as constraints introduced by boundary conditions on equations governing the simulation, it is not always possible to increase these two factors arbitrarily to get MC approximations

down to an acceptable tolerance. In this case, a quantified estimate of the sampling distribution dispersion can benchmark how approximate MC results are.

Equations from probability theory quantify MC approximations when the estimated parameter is a mean [7]. However, approximations from other non-Monte Carlo sources in analysis endeavors are not so readily quantifiable. This poses a serious problem when analytical results guide policy affecting public safety and significant resource expenditures. Under such circumstances, estimates of the accuracy and precision of analytical conclusions inform policy makers about whether such results posses sufficient integrity to be used as a basis for decision making.

All analysis endeavors approximate a physical system with some sort of model. Sets of initial and boundary conditions and descriptions of variables and their respective couplings define models of systems that are not changing, or static, in some reference frame. A model may be as simple as a single deterministic equation or as complex as architectures of nested systems of equations. The amount of detail included in a model is a function of the types of problems for which it is intended to facilitate solution, desired solution accuracy, and efficiency concerns.

However, no matter the level of detail included in an analytical model, it is still an approximation of the physical world. Analytical results will therefore always have some inherent bias and/or dispersion with respect to an unknown true value. A complete model includes an explicit description of all known approximations entailed in its

development as well as the motivation behind and possible limitations arising from these approximations.

There is no unified approach to defining and quantifying approximations resulting in bias and dispersion in modeling and simulation results. Oberkampf et al., attempt to synthesize various and disparate contributions to this end accumulated over many years from a variety of fields [1]. They divide approximation sources into three categories: variability, uncertainty, and error.

Variability is a consequence of inherent randomness in the system being modeled and as such is an irreducible approximation. *Uncertainty* results from a lack of knowledge about the system. Because it is possible to accumulate more knowledge about a system via experimentation or research, uncertainty is a reducible approximation. This paper follows widely used nomenclature replacing the terms *variability* and *uncertainty* with *aleatory uncertainty* and *epistemic uncertainty*, respectively [8]. Dispersion in the random variable distributions characterizing each block in an RBD model is a manifestation of aleatory uncertainty. Greater dispersion in these distributions entails greater aleatory uncertainty in the model of the block. Because an analyst does not know the exact form of the block random variable distributions with arbitrary precision, epistemic uncertainty arises in the block distribution parameters. As a consequence, the parameters of the block distributions themselves can be characterized using distributions. In this case, the parameter random variable distribution dispersion represents the degree of epistemic uncertainty in the value of the parameter.

Error is a recognizable deficiency in the modeling and simulation process that is not due to lack of information. Errors can be isolated, described, and controlled using proper protocols. Errors can arise in RBD models due to approximations made while defining the block variables and their interactions. Since the block random variables must be independent, analysts frequently combine blocks with coupled variables into one block. This approximation decreases the resolution of the model by neglecting the constituent contributions and interactions of the subsumed components, resulting in known errors. Approximations from Monte Carlos simulations, if incorporated into the RBD model, also constitute errors. These are known approximations that can be quantified either analytically or by some other means such as bootstrapping methods [7].

Bayesian Inference and Exploratory Sensitivity Analysis

Implicit in traditional RBD/MC algorithms is the assumption that component-level distribution parameters are known with arbitrary precision. However, epistemic uncertainty in model parameters is almost always present to some extent, especially in low-data scenarios when large deviations from this approximation can result in incomplete, misleading results. There are several approaches to representing and transforming epistemic uncertainty [1,8]. When an analyst uses distributions to represent uncertainty in block parameters, the machinery of probability theory and statistics becomes available to transform the uncertainty in a model.

Modeling block parameters with distributions introduces an additional component-level random variable into the model. Section 1.2 defined *X* and *D* as the random variables describing the aleatory uncertainty in the uptimes (times-between-failures) and downtimes of a repairable system. Now, with the addition of epistemic uncertainty in the model, each influential block has one or more random variables describing the epistemic uncertainty in the parameters of the aleatory uncertainty distributions.

Bayesian statistical inference provides a systematic means to define and update these epistemic uncertainty random variables as more information become available. Under these circumstances, test data can be incorporated with additional information, such as past performance data, subject-matter expert knowledge, software model results and performance data of similar systems. In low data scenarios, including all sources of quality information in analysis can drastically improve fidelity in decision-driving results.

Bayesian Inference requires two basic inputs in the form of distributions. The first, $f(\theta)$, called "the prior distribution," represents information about a parameter, θ, available prior to a data collection event [3,9]. Dispersion in $f(\theta)$ quantifies the level of epistemic uncertainty about the parameter remaining after accounting for available information prior to testing. The prior distribution can encompass past testing and performance information as well as information garnered during a subject matter expert elicitation. The second input distribution, $f(y|\theta)$, called "the likelihood distribution," is a summary of recently collected data not included in the prior distribution determination. Dispersion

in $f(y|\theta)$ represents the amount of aleatory uncertainty surrounding the item being tested.

A conditional probability distribution, computed using *Bayes' Theorem* (cf. equation (1)), incorporates the information contained in the likelihood distribution (component-level mean response and variability) into the component-level parameter information summarized by the prior distribution. The resultant "posterior distribution," $f(\theta|y)$, contains information from both the prior distribution and the test data.

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{\int f(y|\theta)f(\theta)d\theta} \quad (1)$$

The analytical tractability of the conditional probability distribution defining the posterior in (1) is determined by the tractability of the integral in the denominator. This integral can be solved for many types of prior distribution when the likelihood is distributed as a standard exponential with only one uncertain parameter, the scale parameter (cf. Appendix B and C). Frequently, the likelihood distribution can be approximated well by one of the common distribution types. If the prior is then modeled using a distribution called the "conjugate" of the likelihood distribution type, the likelihood and prior distributions are called a "conjugate pair," and the posterior can be obtained analytically. Using conjugate pairs imposes restrictions on the prior distribution type, however.

Tables of conjugate pairs and formulas describing the resultant posterior type as a function of the likelihood variable and parameter(s) and prior variable and parameter(s)

are available in the literature [9]. In the interest of completeness, this paper includes an analytical derivation of several posteriors (cf. Appendices B and C for details). The posterior results given an exponential likelihood for a single observation *y*, i.e. $y \sim Exp(\lambda)$, and gamma prior over λ with parameters, (γ, θ), are summarized below:

$$f(\lambda|y) = \frac{f(y|\lambda)f(\lambda)}{\int f(y|\lambda)f(\lambda)d\lambda} = \frac{(\lambda^n e^{-\lambda t_n})\left(\lambda^{\gamma-1} e^{\frac{-\lambda}{\theta}}\right)}{\int_0^\infty (\lambda^n e^{-\lambda t_n})\left(\lambda^{\gamma-1} e^{\frac{-\lambda}{\theta}}\right)d\lambda} = \frac{\lambda^{\gamma+n-1} e^{-\lambda\left(t_n+\frac{1}{\theta}\right)}}{\int_0^\infty \lambda^{\gamma+n-1} e^{-\lambda\left(t_n+\frac{1}{\theta}\right)}d\lambda}$$

$$= \frac{\lambda^{\gamma+n-1} e^{-\frac{\lambda}{\alpha}}}{\alpha^{\gamma+n}\Gamma(\gamma+n)}, where\ \alpha = \frac{1}{\left(t_n+\frac{1}{\theta}\right)} \qquad (2)$$

As is always the case with conjugate pairs, the posterior here has the same distribution type as the prior, a gamma, with different parameters, shape parameter, $\gamma + n$, and scale parameter, α.

In situations where the conjugate pair approximation is too restrictive, WinBUGS [10], YADAS [3] or other software suites compute posteriors from arbitrary likelihood and prior distributions. These packages use variations of Markov Chain Monte Carlo to approximate the posterior. Although these packages are versatile, the posteriors they produce are approximations, including error associated with the sampling based algorithms used to generate them.

The first step in Bayesian inference involves modeling the input distributions. The likelihood distribution is specified analytically (cf. Appendices B and C). Accumulating

prior information from diverse sources into one prior distribution is less straightforward. One method consists of iteratively applying Bayes' Theorem to distributions formed from different data sources.

Generating accurate and precise prior distributions can be a time and resource intensive process requiring ingenuity and careful planning [8,9,11,12]. Isolating the most influential component-level parameters with a sensitivity analysis circumvents the need to explicitly model priors for non-influential components. Conclusions from a sensitivity analysis conducted on a model before an uncertainty analysis aid analysts in determining how to reduce the dimensionality of the epistemic uncertainty in the model and, hence, the expense entailed by the uncertainty analysis.

Approaches to sensitivity analysis differ depending on the type of model involved and scope of inquiry desired. Saltelli et al., categorize sensitivity analysis approaches according to scope of inquiry [13]. *Factor Screening* involves isolating important factors (inputs) in a system with many factors. *Local Sensitivity Analysis* emphasizes the local impact of factors on a system and is typically accomplished analytically using partial derivatives. *Global Sensitivity Analysis* emphasizes apportioning uncertainty in output to uncertainty in input factors. Alternative sensitivity analysis taxonomies defined in terms of model efficiency (expensive vs. cheap), model type (e.g. linear vs. nonlinear, monotonic vs. non-monotonic), approach (analytical vs. sampling based) or fidelity (using the model directly vs. using a meta-model to approximate the model) also serve as helpful guides in choosing the best approach for specific analytical regimes [8,13].

In the case where the model is a black box, sensitivity analysis techniques accurate for specific model types are inappropriate. The schematic design and recursive nature of, as well as the nested sampling loops in RBDs models including epistemic uncertainty place these models in the black box category. In this case, scatter plots can give insights about implicit relationships between component-level and system-level epistemic uncertainty without any assumptions about the functional form of the relationship.

After approximate relationships have been discerned from scatter plots, tests for non-randomness such as the Common Means Test, the Common Locations Test, the Statistical Independence Test, the Regression Test and the Quadratic Regression Test can quantify and rank inputs according to influence upon output response. Because these tests work well for specific model types, information in scatter plots should inform the weight an analyst gives to specific test results. In tandem, these sensitivity analysis techniques provide an estimate of the global main effects, or drivers, in the system response.

If a sensitivity analysis reveals that output responses are insensitive to changes in one or more input variables, point estimates of these variables can be used in the model in lieu of distributions. For RBD/MC simulations incorporating estimates of epistemic uncertainty in component parameters, this amounts to discovering that system-level epistemic uncertainty is unaffected by epistemic uncertainty in some component-level parameters. Investing resources into characterizing epistemic uncertainty distributions for these parameters is therefore unnecessary.

Bayesian statistical methods provide a means to characterize and iteratively decrease epistemic uncertainty in a parameter as more information becomes available. Augmenting commercial RBD/MC algorithms by using distributions to represent epistemic uncertainty and Bayesian inference to update epistemic uncertainty distributions as new information becomes available is a logical next step in honing reliability analysis. In the area of system reliability analysis, Bayesian methods will most commonly apply at the component-level to update epistemic uncertainty in component-level parameter distributions. As a consequence, uncertainty propagation methods are necessary to make inferences about how these component-level transformations translate to the system-level.

Uncertainty Analysis

RBD/MC commercial software sampling propagates aleatory uncertainty from the component to the system level. An additional sampling layer applied to component parameter distributions representing epistemic uncertainty produces a system-level epistemic uncertainty distribution and a suite of system-level aleatory uncertainty distributions. Figure 4 depicts the flow of the simulation including the additional sampling layer. The system-level epistemic uncertainty distribution summarizes the cumulative effect of epistemic uncertainty in influential component-level distributions. It serves as an important benchmark of the precision in the analytical process.

After forming influential component priors and generating posteriors, samples drawn from each of the influential component parameter uncertainty distributions (the posterior

distributions derived from Bayesian inference) *n* times produce *n* parameter vectors. Each of the *n* parameter vectors is (1 x *m*), where *m* is the number of influential blocks in the RBD model. The vectors comprise randomly-grouped sets of parameter samples, with one sample from each influential block. Each parameter vector defines a set of block distribution parameters to be used as input for a standard RBD/MC algorithm. Non-influential component parameters are fixed at their point-estimate values in the simulation. These point estimates can be obtained using historical data or from subject matter expert estimations.

For inexpensive models, using MC sampling in the outer sampling layer is simple and yields good results with large sample sizes. Medium to large RBD models entail longer simulation times, imposing limits on system-level sample generation, however. In these cases, strategic sampling from input distributions can survey the input variable space with fewer samples than simple random sampling, thus decreasing the number of model evaluations necessary to map the output space. Stratified sampling and Latin Hypercube sampling are popular alternatives to random sampling when model evaluations are expensive [8,14,15].

Stratified sampling involves partitioning a set into mutually exclusive, exhaustive subsets (strata) [8]. Random samples from these subsets comprise a stratified sample. Partitioning allows an analyst to assure that important events in the set are included in the sample, even if these events have small associated probability.

Latin hypercube sampling (LHS) presents a useful compromise between stratified and simple random sampling in several respects [8,14,15]. Like stratified sampling, LHS involves partitioning a set into mutually exclusive, exhaustive subsets. LHS differs from stratified sampling, however, in that the intervals defining the subsets are equally spaced. Consequently, the analyst only defines the total number of intervals desired, and not their ranges. For most models, LHS estimator convergence improves upon the estimator convergence of simple random sampling [8,14,15]. Furthermore, implementation of LHS requires less time and skill than is required in stratified sampling schemes because it is not necessary to define strata and their associated probabilities [14]. Because model evaluations are expensive for the B-2 modernized radar RBD, the outer sampling loop for the sensitivity and uncertainty analysis presented in Chapter 2 uses LHS.

The output of the entire simulation is a set of sampling distributions of the mean, one generated from each parameter vector. The means of these sampling distributions exhibit dispersion due to the uncertainty in the block parameters as well as sampling error incurred during the simulation. The dispersion of the means can be used to calculate approximate credibility intervals that quantify uncertainty in system parameters (cf. figure 4). However, when model evaluations are expensive and place limitations on system-level data generation, finite sampling error could significantly affect the estimated system-level epistemic uncertainty distribution.

It is possible to estimate the cumulative effect of the error from the two sampling sources. When the parameter of interest is a mean, error due to finite MC sampling executed in

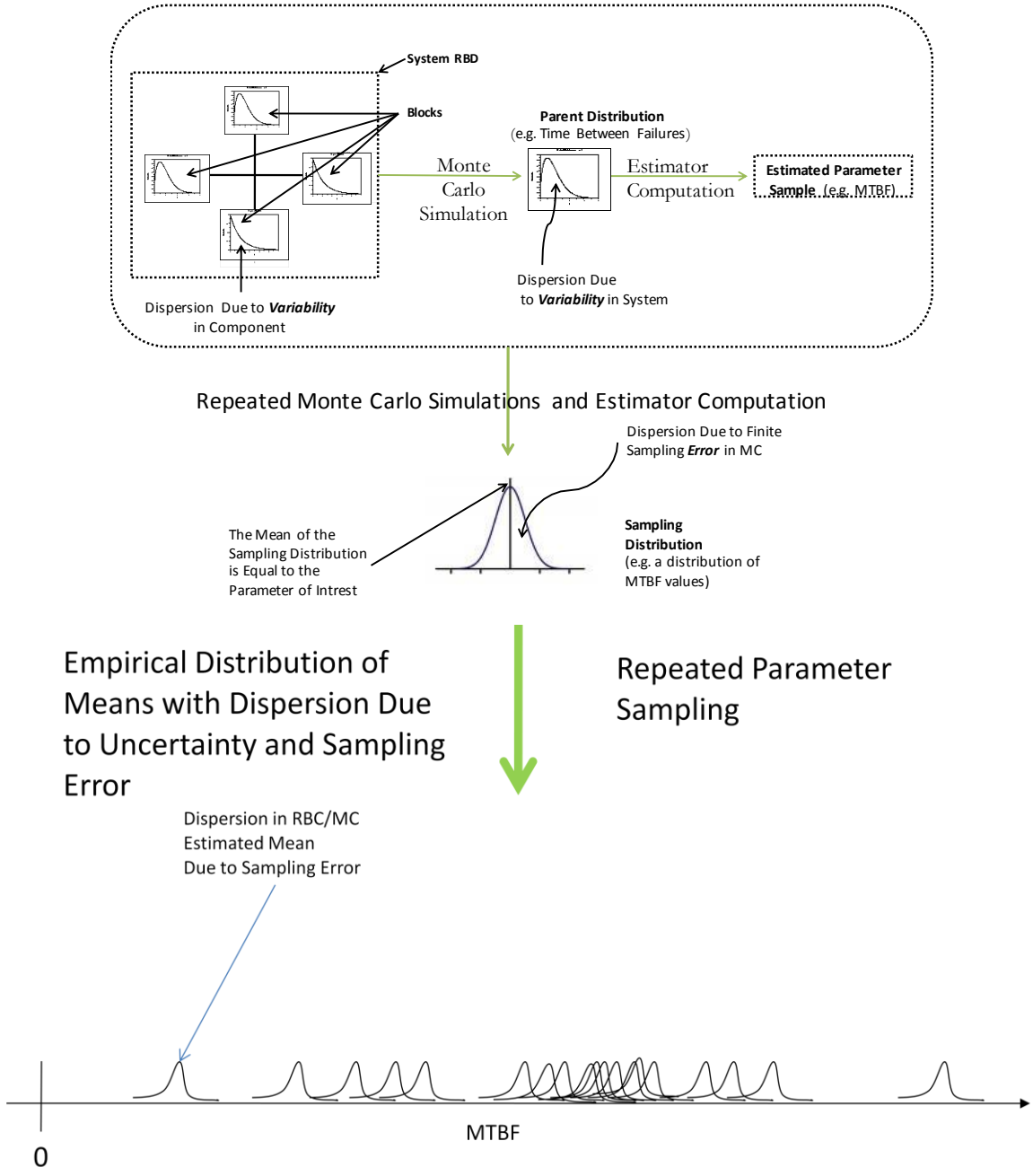


Figure 4: Depiction of the system-level uncertainty analysis conducted by applying an extra layer of sampling from parameter distributions

RBD/MC software is summarized via the standard error of the mean, denoted $se_F(\bar{x})$. In this notation, x is a random variable with distribution F and $\bar{x}$ is the mean of a random

sample taken from the set of events represented by *x*. The standard error of any estimator is a common indicator of statistical accuracy [7]. RBD/MC software suites output a $se_F(\bar{x})$ for all mean estimates. By the Central Limit Theorem, as the number, *k*, of samples increases, the sampling distribution of the mean approaches a normal with standard deviation equal to the $se_F(\bar{x})$ [5,7]. The sampling distribution with mean equal to the estimated mean and standard deviation equal to the $se_F(\bar{x})$ summarizes the error in the mean estimate due to finite sampling in the RBD/MC simulation.

$$se_F(\bar{x}) = var_F(\bar{x})^{1/2} = \sigma_F/\sqrt{k} \qquad (3)$$

A bootstrap method can be used to estimate the cumulative effect of both the MC and the LHS sampling error. In this method, *l* bootstrap sample vectors are generated by sampling with replacement *p* times from an empirical distribution composed of *p* observed values [7]. In the case of RBD models including epistemic uncertainty, an observed value is the system-level metric (usually an approximate mean value), $\hat{\bar{X}}_i, i = 1, \dots, p$, generated by RBD/MC software for each of the p parameter sample vectors. Recall that LHS, stratified sampling or MC methods produced the *p* parameter sample vectors used as inputs for the RBD/MC simulation.

So, we get a sample of size p, $\hat{\bar{X}}_{ik}, i = 1, \dots, p,\ k = 1, \dots, l$, by sampling $\hat{\bar{X}}_i, i = 1, \dots, p$, with replacement *l* times. Further, substituting a random sample from $N\left(\hat{\bar{X}}_i, se_F(\bar{x})_i\right)$ for each $\hat{\bar{X}}_i$ in a bootstrap sample accounts for the inner MC sampling

error. The sampling with replacement of the $\hat{\hat{X}}_i$ accounts for the outer loop sampling error. CDFs formed using these new observed bootstrap sample vectors can then be used to approximate point wise confidence bands around the epistemic uncertainty curve. These bounds estimate the amount of error incurred in the two sampling loops (cf. Appendix D for further details).

In sum, the steps outlined above culminate in an uncertainty CDF, bounded by confidence bands estimating the magnitude of the sampling error in the uncertainty analysis. The uncertainty CDF can be used to compute metric point-estimates as well as credible intervals defining the fidelity of the point-estimates. Additionally, the fidelity of the uncertainty CDF itself is quantified using point wise confidence bands. Because a preliminary sensitivity analysis reduces the dimensionality of the epistemic uncertainty probability space, this analysis paradigm generates results with more information than a point estimate using minimal information from the input space. Of course, there is no free lunch in nature; the cost of this trade-off is the increased effort introduced into the analysis by the sensitivity analysis. The next section applies the process outlined above to real data taken from the B-2 modernized radar program operational reliability tests.

Chapter 2: Operational Reliability Assessment of the B-2 Modernized Radar System

At the Air Force Operational Test and Evaluation Center (AFOTEC), we test and evaluate new weapon system capabilities in operationally realistic battle space environments. Operational test (OT) data are obtained from a production-representative system in an operationally realistic environment. Reliability analysis plays a key role in defining measures of operational capability.

AFOTEC consists of a headquarters in Albuquerque, New Mexico and five detachments spread throughout the U.S. Each detachment specializes in operational evaluation of specific classes of technical systems used in the Air Force. For example, Detachment 5 in California works with mobility, bomber, command and control, intelligence, surveillance and reconnaissance weapon systems. Analysts at AFOTEC headquarters are tasked to support test teams and analysts at the detachments as needed.

Air Force leadership compares results from OT analysis to standards stipulated in contracts between the developer and the Air Force (the purchaser). When OT results indicate that the system performance is lacking significantly, the developer is obliged to discover and modify the source of the shortfall prior to large scale production. OT analysis also provides logistics information to aid decision makers on how to best support the war fighter using these systems in wartime scenarios. Since large sums of tax-payer money as well as the lives of civilian and military personnel hinge on these

considerations, estimates of fidelity in operational reliability analysis results provide crucial information for decision makers.

2.1 The B-2 Radar Modernization Program

The Air Force recently contracted Northrop Grumman Corporation (NGC) to incorporate new technology into the B-2 bomber radar system. This project, deemed the B-2 Modernized Radar Program (B-2 RMP), included the following adaptations to the legacy radar system:

-R/E (receiver exciter) **modified (**6 shop replaceable units) to accommodate radar modernization program (RMP) requirements

-Legacy antenna replaced with **new** AESA (Active Electronically Scanned Array) antenna assemblies: AESA includes all legacy antenna functionalities combined with the functionalities contained in the legacy transmitters. Northrop Grumman Corporation (NGC) lead B-2 RMP engineer, Steve Ruch, considers the legacy antenna and the AESA "totally different technologies."

-Legacy transmitters removed

- WG-SW (Wave-Guide Switch) is **new**

-Electrical power to the AESA provided by **new** PPCU (Prime Power Conditioning Units)

-DMS RFFE **modified** to provide RF signal protection from on-board radar emissions at the new operating frequency (not considered part of the radar system and not modeled in RBD)

-The Ku Band Transponder LRUs **modified** to accommodate the new operating frequency (not considered part of the radar system and not modeled in RBD)

Upon release, Detachment 5 of AFOTEC initiated an operational performance evaluation on several production-representative prototypes of the B-2 modernized radar. Part of this evaluation included reliability testing and analysis to determine the values of several relevant performance metrics. The figure of merit MTBF, E(X), mean time-between-failures, was highlighted by Air Force leadership amongst these as particularly relevant. Sparse operational test data available for the B-2 RMP system due to resource and time constraints limited the accuracy and precision of classical statistical analysis for this system. Consequently, analysts at AFOTEC headquarters were tasked to develop an alternative analysis paradigm to contend with this and other increasingly common low to medium data scenarios.

To leverage the more prolific component-level data, the AFOTEC bomber analysts at Detachment 5 used a reliability block diagram developed by NGC representing the modernized radar system. The RBD underwent an internal accreditation process at AFOTEC prior to being used in this analysis. The final accredited RBD is shown below in Figure 5. As a repairable system, the model of each component, i, incorporates two types of aleatory uncertainty random variable: X_i , the time-between-failures random variable, and D_i, the down-time random variable.

The two aleatory random variables for each component and its mirror image, henceforth called a redundant structure, are defined independently and identically. Component-

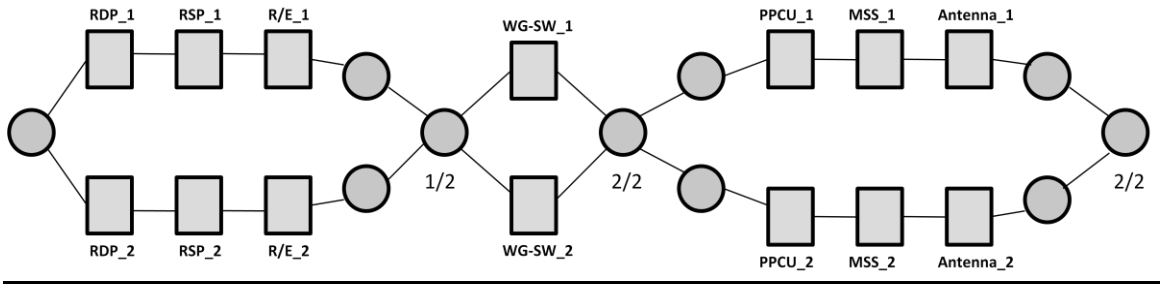


Figure 5: B-2 RMP (Radar Modernization Program) Reliability Block Diagram

	Repair Distribution Type	Mean	Std. Dev.	
Antenna	Lognormal	8	0.5	
MSS	Lognormal	3	0.5	
PPCU	Lognormal	2	0.5	
R/E	Lognormal	4	0.5	CONSTANT OVER ENTIRE ANALYSIS
RDP	Lognormal	2	0.5	
RSP	Lognormal	3	0.5	
WG-SW	Lognormal	5	0.5	

Table 1. Parameter values for component repair distributions

level time-between-failures probability spaces are summarized in the model with exponential distributions, $X_i \sim Exp(\lambda_i)$, while lognormal down-time distributions, $D_i \sim LogN(\mu_i, \sigma_i)$, describe each component's down-time probability space (cf. Tables 1 and 2). AFOTEC bomber analysts and NGC engineers estimated μ_i and σ_i in the model. At headquarters, we determined estimates for $\lambda_i$using legacy data collected by AFOTEC and research and development (R&D) data amassed at NGC.

	Time-Between-Failure Distribution Type	$\lambda_i = 1/\mathrm{MTBF}_i$
RDP	Exponential	1/1879.2
RSP	Exponential	1/2349
R/E	Exponential	1/804
WG-SW	Exponential	1/11,437.80
PPCU	Exponential	1/61,348
MSS	Exponential	1/1566
Antenna	Exponential	1/2412

Table 2. Parameter values for component failure distributions

2.2 Preliminary Sensitivity Analysis

To begin the reliability analysis, we conducted a preliminary sensitivity analysis on the accredited RBD model described above to isolate and characterize the most influential components affecting epistemic uncertainty in B-2 RMP MTBF estimates. We used RAPTOR RBD/MC software (developed at AFOTEC) for the RBD/MC portion of our analysis. RAPTOR provides many unique capabilities necessary for the simulation of Air Force technical systems not found in other RBD/MC software suites.

The preliminary sensitivity analysis focused on how epistemic uncertainty in the component-level time-between-failures distribution parameter, λ_i, affected epistemic uncertainty in the system-level MTBF only. We held the repair distribution parameters constant throughout the analysis. We modeled the epistemic uncertainty in the seven redundant-structure scale parameters with a uniformly distributed random variable, Λ_i, centered at $\hat{\bar{\lambda}}_i$ with range [0, 2 $\hat{\bar{\lambda}}_i$], where $\hat{\bar{\lambda}}_i$ is the estimated mean of λ_i determined using legacy and R&D test data (cf. Table 3), e.g. $\Lambda_i \sim Unif(0, 2\hat{\bar{\lambda}}_i)$.

Then, using the R function for Latin Hypercube sampling, we produced 100 randomly grouped (1x7) parameter vectors with components from these distributions. Each vector served as the time-between-failures distribution parameter input for one RBD/MC simulation. We set the termination point of the simulation at 131,534.1 simulated hours to ensure that the $\text{se}_{\text{F}}(\bar{x})_i$ was no larger than approximately one percent of the system-level MTBF. The entire process produced 100 estimates of system-level MTBF as a function of component parameter, λ_i, changes.

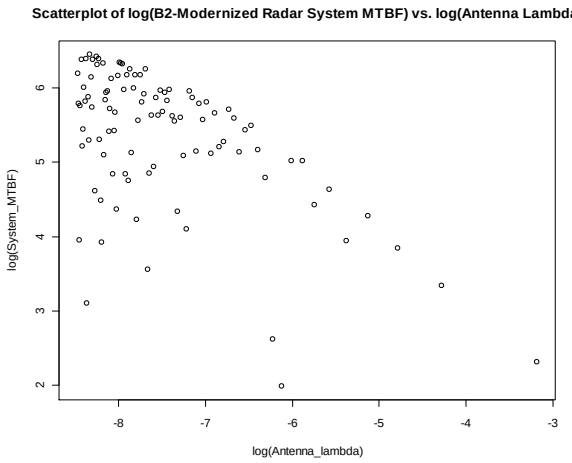
Although one of the simplest sensitivity analysis techniques, scatter plots are robust in that they do not require any assumptions about the functional form of the relationships being explored [8,13]. Because the λ_i samples spanned a broad range including very small numbers, we used scatter plots of mean system-level response $log_e(MTBF)$ versus $log_e(\lambda_i)$ for each of the seven redundant structures as a first step in determining possible relationships between system and component-level uncertainty (cf. Figure 6 below). As

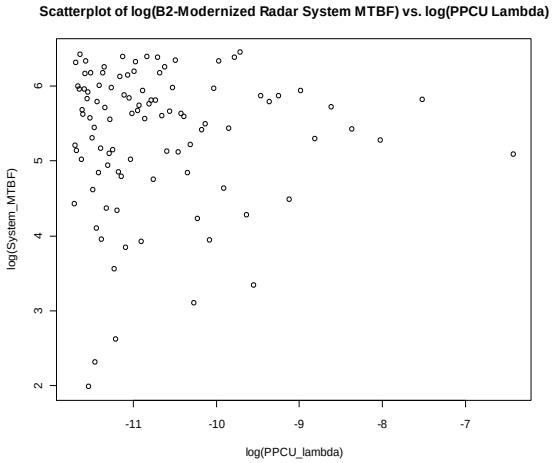
	Component	Mean	Interval
Uniform Distribution Parameters	Antenna	1/2412	[0,1/1206]
	MSS	1/56376	[0,1/28188]
	PPCU	1/61,348	[0,1/30674]
	R/E	1/804	[0,1/402]
	RDP	1/46980	[0,1/23490]
	RSP	1/37584	[0,1/18792]
	WGSW	1/11437.8	[0,1/5718.9]

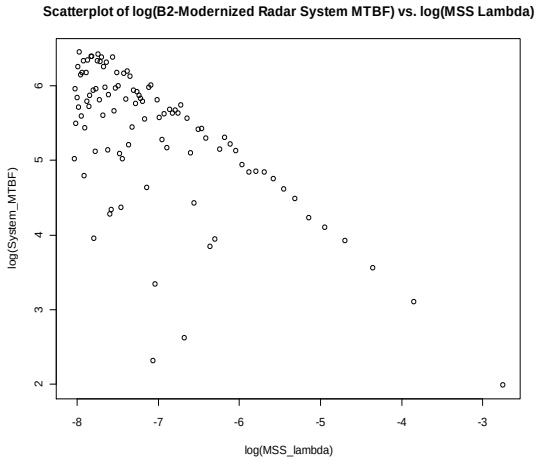
Table 3: Component uniform distribution parameters

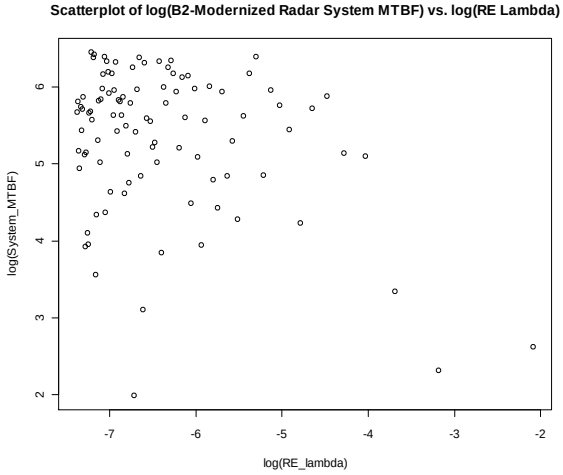
demonstrated in Figure 6, the MSS and Antenna $\lambda's$ appear to result in systematic changes in the system MTBF over the entire range of parameter values. The other component types demonstrate little influence on the system MTBF over the entire sample range.

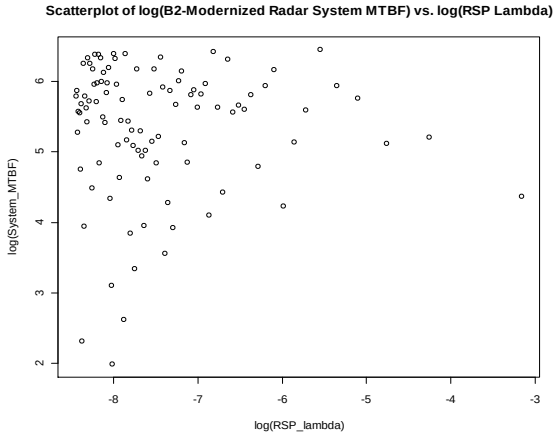
The approximate relationships demonstrated in the scatter plots allowed us to apply quantitative, model-dependent tests for non-randomness to the data in a more informed manner. There appeared to be no obviously non-monotonic relationships. Furthermore, it seemed that the system-level MTBF, as a function of λ_i could be reasonably approximated using a linear or quadratic function of these variables. The common means test (CMT), common locations test (CLT), test for statistical independence (SI), regression test (REG) and quadratic regression test (QREG) are accurate for determining

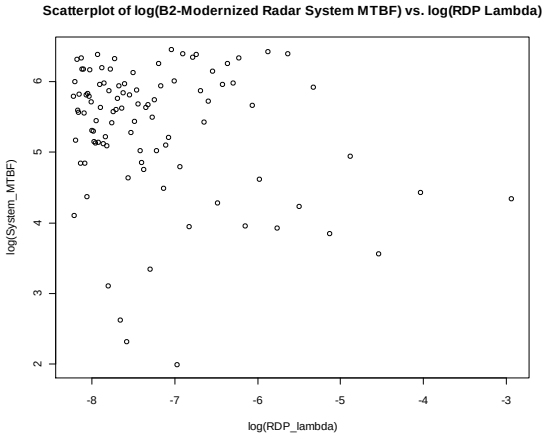

Scatterplot of log(B2-Modernized Radar System MTBF) vs. log(Antenna Lambda)
log(System_MTBF)
log(Antenna_lambda)


Scatterplot of log(B2-Modernized Radar System MTBF) vs. log(PPCU Lambda)
log(System_MTBF)
log(PPCU_lambda)


Scatterplot of log(B2-Modernized Radar System MTBF) vs. log(MSS Lambda)
log(System_MTBF)
log(MSS_lambda)


Scatterplot of log(B2-Modernized Radar System MTBF) vs. log(RE Lambda)
log(System_MTBF)
log(RE_lambda)


Scatterplot of log(B2-Modernized Radar System MTBF) vs. log(RSP Lambda)
log(System_MTBF)
log(RSP_lambda)


Scatterplot of log(B2-Modernized Radar System MTBF) vs. log(RDP Lambda)
log(System_MTBF)
log(RDP_lambda)

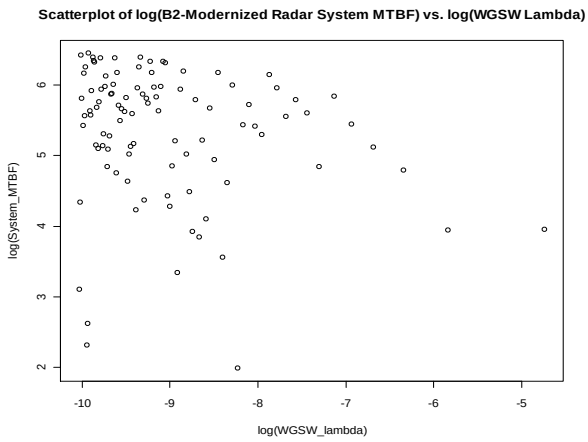


Figure 6: Scatterplots of log_e(System MTBF) vs. $log_e(\lambda_i)$, for i = 1, 2, …, 7 for each of the seven redundant structures

general relationships, failing only in pathological cases. REG is accurate when identifying linear relationships, while QREG is accurate when identifying quadratic relationships.

Non-randomness tests such as the common means test (CMT), the common locations test (CLT), and the test for statistical independence (SI) involve partitioning the scatter-plot abscissa (as well as the ordinate for the SI test) into disjoint classes and determining if the data across the classes share common statistics (e.g. the mean for the CMT, the median for the CLT or the chi-square test for contingency tables for SI). If they do, it is less likely there is a relationship between the dependent and independent variable [8]. The regression (REG) and quadratic regression (QREG) tests involve fitting a surface (linear in the case of REG, quadratic in the case of QREG) to the data and testing to see if the coefficient of a predictor variable is significantly larger than zero. If it is, there is a significant relationship between that predictor and the response variable [8].

For each test, p-values quantify the level of non-randomness in the predictor/response variable relationship. A small p-value corresponds to low level of randomness in the relationship (e.g. small p-value indicates a significant relationship between the predictor and response variable). As can be seen by the small associated p-values in Table 4, the CMT, CLT, SI, REG, and QREG tests confirm that the MSS and antenna λ dispersion exerts significant influence on the system MTBF. Although not apparent in the scatter plot, the non-randomness tests isolated the RE λ as influential as well. The REG and QREG tests designated the WG-SW λ as a driver in the model.

The small number of components in the system as well as the high-level of symmetry facilitates a comparison of these quantitative results with qualitative, more intuitive notions of how specific blocks should influence system reliability measures. The first group of redundant blocks: RSP, RDP and R/E, are paired in a one-out-of-two configuration, meaning that only one of the two redundant blocks needs to be operational to keep the system operational. Thus, the system should not be very sensitive to individual downing events in these blocks.

The subsequent two groups, including the WG-SW, PPCU, MSS and antenna components, are two-out-of-two redundant structures, meaning blocks in both parallel strings must operate if the system is to remain operational. Individual, component-level downing events in these groups result in a system-level downing event, e.g. the system should be more sensitive to component-level failures for these blocks. The fact that the system MTBF is insensitive to PPCU and WG-SW parameters is possibly due to some

interplay between their repair distribution parameters and the portion of the parameter space from which we sampled.

Common Means Test (CMT)				**Common Locations Test (CLT)**			**Statistical Independence Test (SI)**	
Input	MSS	Antenna	RE	MSS	Antenna	RE	MSS	Antenna
P-val	0	0	0.0366	0	0	0.0353	0	0.00008

Regression Test (REG)					**Quadratic Regression Test (QREG)**			
Input	MSS	Antenna	RE	WGSW	MSS	Antenna	RE	WGSW
P-val	0.0016	0.0029	0.0181	0.0393	0	0	0.0193	0.0410

Table 4: P-values from five different non-randomness tests

The preliminary sensitivity analysis conducted via an accredited reliability block diagram of the B-2 RMP indicates that the most influential failure distribution parameters on the system-level mean response (MTBF) are those of the MSS and the Antenna. These two components appear to drive epistemic uncertainty in B-2 RMP system-level MTBF and thus, indirectly, other reliability measures (cf. reference [2] for equations relating MTBF

to other reliability measures). Accurate, well-defined and precise descriptions of reliability measures for such drivers directly entail more informative, higher fidelity system-level measures.

2.3 Component-Level Epistemic Uncertainty Modeling and Minimization with Bayesian Inference

Prior Determination

For the purposes of benchmarking the preliminary sensitivity analysis results, we modeled distributions representing epistemic uncertainty in both driving and non-influential components. This allowed us to compare results from the full model, including epistemic uncertainty distributions for all components, with those from the reduced model, containing epistemic uncertainty distributions for driving components only. Because NGC modeled all component failure distributions as exponential, we modeled all observations constituting the likelihood distributions derived from test data as exponentials. In order to facilitate calculations, we modeled each component-level prior distribution on λ as a gamma, the conjugate prior of an exponential likelihood. To this end, we used a formula cited by NIST on its Bayesian reliability website to form gamma priors from legacy data [16]. According to the NIST formula, the gamma prior shape parameter is defined as the number of recorded failures in the historical data, the scale parameter the reciprocal of the total time during which failure data were logged.

The NIST formula can be derived from first principles by taking a uniform prior on $[0,b)$ (an uninformative prior) and forming a posterior using the legacy data as a likelihood in the limiting case where $b \to \infty$ (cf. Appendix B). When new test data defining another likelihood distribution becomes available, the posterior formed above can be inputted as the prior to be updated using Bayes' Theorem. Thus, Bayesian analysis enables iterative model "learning" as more information from test data becomes available.

However, this method assumes that legacy and RMP components are equivalent; an assumption that is not valid for new or modified RMP components. For these components, the prior parameters can be informed by legacy data only to the extent that the RMP components resemble the legacy components. NGC R&D (research and development) test data for the new and modified components provides additional reliability information for new and modified components.

To reconcile R&D and legacy data sources for the new and modified components, weighted averages of historical data and R&D estimates for the new and modified components would most accurately serve as prior parameter estimates. The weights would reflect an estimate of the deviation of the RMP components from the legacy components. That is, the larger the technological deviation of the RMP components from legacy components, the larger the weight attributed to the R&D data in the prior parameter estimates.

Unfortunately, the NGC engineers with the expertise to estimate such weights were not contracted to provide this information or any additional information about uncertainty in the component-level drivers. As a consequence, we took a conservative approach here, and the new and modified component priors do not incorporate historical data, only NGC R&D data. We used the NIST formula with the R&D data likelihood to define prior distribution parameters for these components. NGC was the only source of additional information that could decrease and further characterize epistemic uncertainty in driving component parameters. Therefore, the driver parameters do not incorporate additional information to decrease uncertainty and increase the resolution of the models representing uncertainty in these component parameters. AFOTEC legal counsel is in the process of investigating alternatives to existing contracts that would oblige contractors to provide this information for future systems. The following summarizes the considerations we used to form each component-level prior:

The Antenna is a new component incorporating the legacy antenna and transmitter. In NGC R&D testing, it failed one time during a total of 2412 hours. This results in a gamma prior with shape parameter 1 and scale parameter 1/2412.

The PPCU is a new component with no analog in legacy test data. It experienced zero failures during R&D testing. NGC estimated steady state MTBF using the Duane model is 61,348 hrs. The PPCU prior shape parameter in this case is estimated to be 1, the scale to be 1/61,348.

The WG-SW is another new component with no analog in legacy test data. MTBF estimated by NGC using Raptor to be 11436.8. Using 1/11436.8 as the prior scale parameter, the prior shape parameter was set at 1.

	Component	Shape	Scale
Gamma Prior Parameters	Antenna	1	1/2412
	MSS	6	1/9396
	PPCU	1	1/61,348
	R/E	1	1/804
	RDP	5	1/9396
	RSP	4	1/9396
	WGSW	1	1/11437.8

Table 5: Component Prior Gamma Failure Distribution Parameters

The R/E was modified in RMP. NGC encountered one failure for this component in 804 hours of R&D testing which translates to a prior with shape parameter 1 and scale parameter 1/804. Although the R/E appears to have undergone significant modifications, it is considered "modified" rather than "new" by NGC.

The MSS, RDP, RSP are all legacy components. Prior parameters consisted of the total number of failures in legacy data as the shape parameter and the inverse of total time during which failures were logged as the scale parameter (Table 5).

Posterior Determination

Appendix C details the derivation of the formula for defining gamma posteriors from a gamma prior/exponential likelihood conjugate pair when the likelihood data is failure truncated. The resultant posterior shape parameter is simply the sum of the total number of failures in the likelihood data and the shape parameter of the prior. The reciprocal of the scale parameter is the sum of the total test time in the likelihood data and the inverse scale parameter of the prior. These very simple, closed-form posterior parameter equations are a consequence of the gamma prior and exponential likelihood being a conjugate pair. For arbitrary likelihood and prior distribution types, MCMC software or analytical solutions to integral equations provide estimates of posterior parameters.

AFOTEC conducted two sets of operational tests. The first included all B-2 modernized radar components and was time truncated (e.g. the test stopped at a specific time rather than after a specific number of failures). The second included only the antenna and the R/E and was failure truncated (e.g. the test stopped after a stipulated number of failures, in this case the number was one). The posterior parameter formulas derived in Appendix C from a gamma prior/exponential likelihood conjugate pair assume a likelihood formed from failure censored data. Our first data set was time truncated, but included no data on the specific times of each component failure and repair within the test time period. This made it impossible to model the likelihood of that data in as time censored. We therefore made the approximation that this data set was failure truncated and used the formula

	Component	Failures	Test Time		Component	Shape	Scale
	Antenna	2			Antenna	3	1/3725.9
	MSS	0			MSS	6	1/10709.9
Likelihood	PPCU	0		Gamma	PPCU	1	1/62,662
Data: Test 1	R/E	3		Posterior	R/E	4	1/2117.9
	RDP	3	1313.9	Parameters	RDP	8	1/10709.9
	RSP	0			RSP	4	1/10709.9
	WGSW	0			WGSW	1	1/12751.7

Tables 6: Likelihood data for the first operational test and resultant gamma posterior parameters after applying Bayes' Theorem to it and the NIST priors formed from legacy data

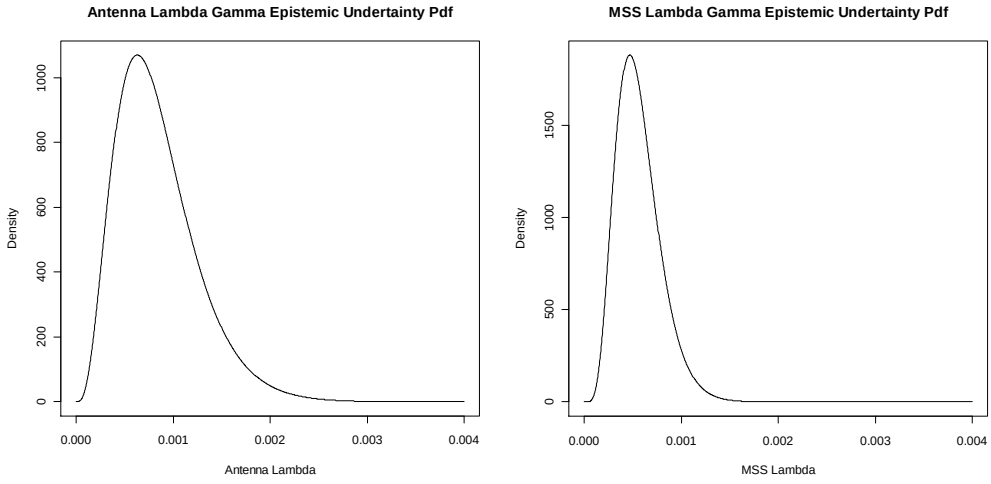
	Component	Failures	Test Time		Component	Shape	Scale
	Antenna	1	1046.5		Antenna	4	1/4772.4
	MSS	N/A			MSS	6	1/10709.9
Likelihood	PPCU	N/A		Gamma	PPCU	1	1/62,662
Data: Test 2	R/E	1	1752.1	Posterior	R/E	5	1/3870
	RDP	N/A		Parameters	RDP	8	1/10709.9
	RSP	N/A			RSP	4	1/10709.9
	WGSW	N/A			WGSW	1	1/12751.7

Table 7: Likelihood data for the second operational test and resultant gamma posterior parameters after applying Bayes' Theorem to it and the priors defined as the posteriors generated from the first test data

derived in Appendix C to define the first set of gamma posterior parameters. It is important to note that, as a consequence of this approximation, the mean time-between-failures for the first data set is overstated. Unfortunately, we cannot quantify this bias due to lack of data from the test team. This situation highlighted the need for

standardization in data collection across the detachments at AFOTEC and is being addressed as part of an organization-wide effort to standardize processes.

To incorporate information from both sets of test results into our component-level uncertainty distributions, we applied Bayes's Theorem twice. We used the NIST gamma priors and the likelihoods of the first test data set to produce component gamma posteriors. These posteriors then became the priors updated by the likelihood of the second set of test data (cf. Tables 6 and 7). The dispersion in the last set of gamma posteriors represents the component-level epistemic uncertainty remaining after incorporating all available information into the model (c.f. Figure 7).



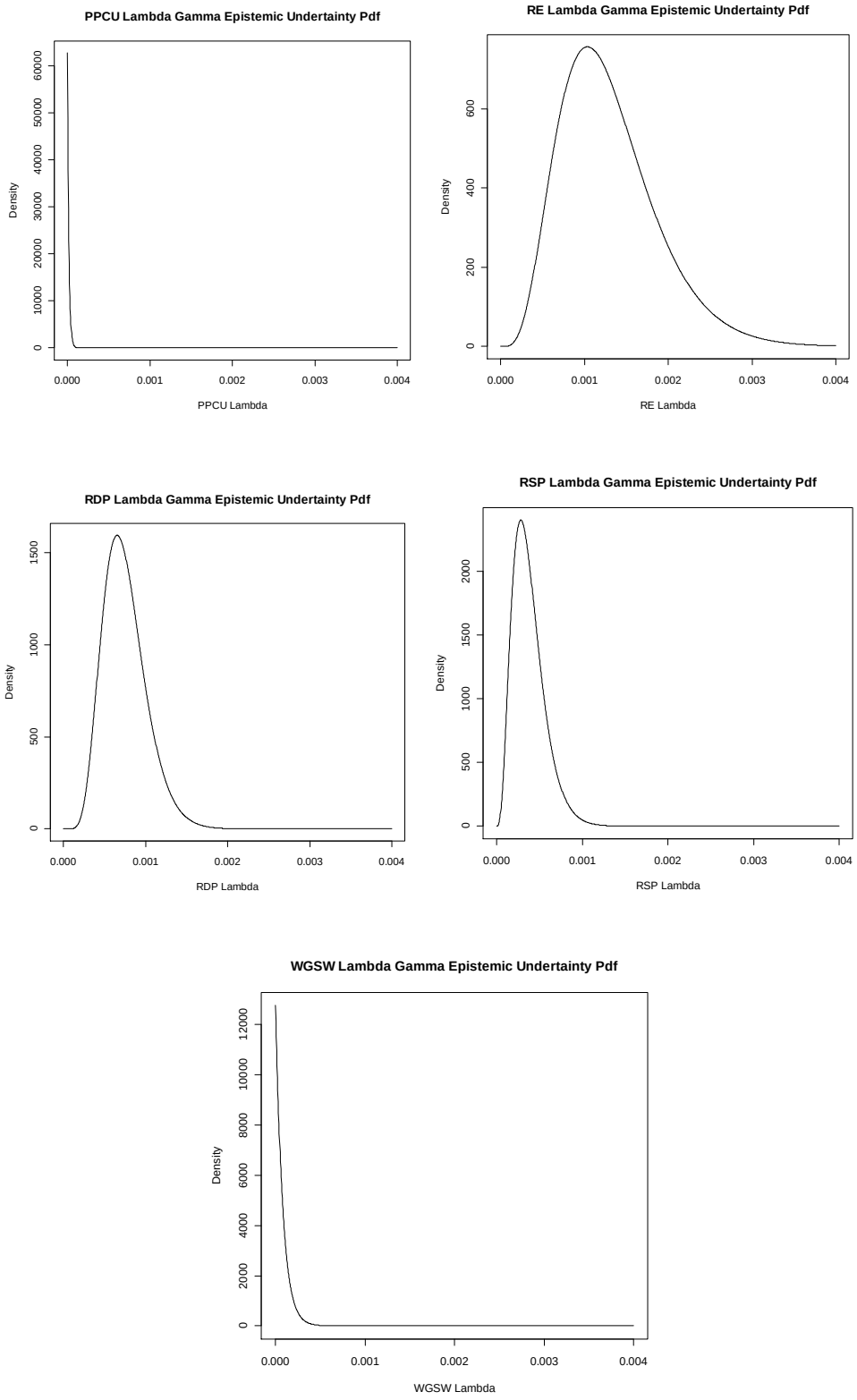


Figure 7: Plots of gamma posterior probability distribution functions summarizing the epistemic uncertainty probability space for each component

2.4 Sampling-Based Uncertainty and Sensitivity Analysis

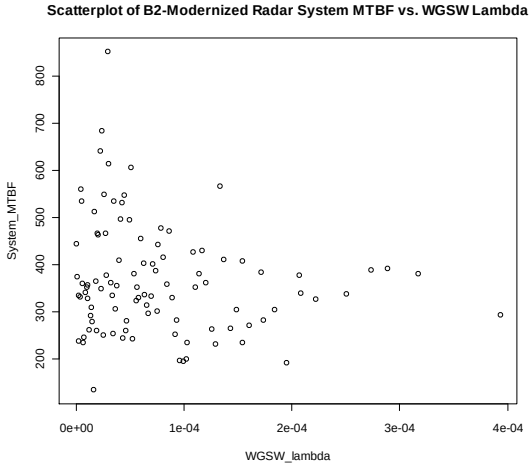
For this portion of the analysis, we produced two sets of 100 sample vectors using LHS, as described in the preliminary sensitivity analysis section. To generate the first set, we sampled from all of the posterior gamma uncertainty distributions defined above. We generated the second set by sampling only from the driver (MSS and antenna) posterior distributions, keeping all other component rate parameters fixed at their point estimates. Each sample vector served as input for one Raptor simulation with run time 131,534.1 hours and repetition level 100. Raptor outputted two sets of 100 B-2 RMP system-level MTBF estimate, one for the full model, the other for the reduced model. The dispersion in the estimates generated for each model represents the epistemic uncertainty in the B-2 RMP system-level MTBF.

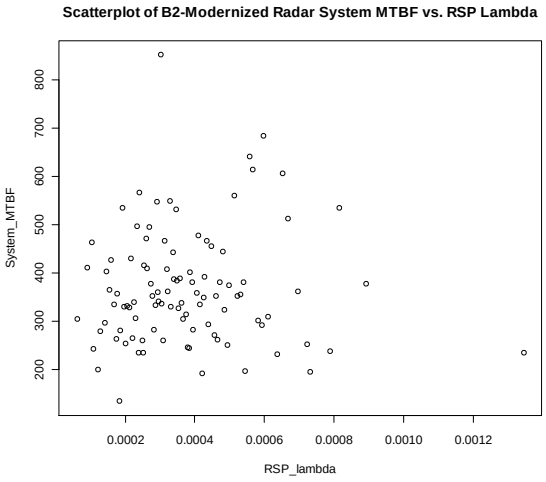
To gauge model sensitivity with the input parameters distributed as gammas for the full model, we produced scatter plots and ran the same non-randomness tests performed in the preliminary sensitivity analysis. The results were not drastically different from those in the preliminary sensitivity analysis, as can be seen in Figure 8 and Table 8, with the exception that the RE appears to be less decisively a third place driver in the system MTBF. The resultant epistemic uncertainty CDF for the full model is shown in Figure 9.

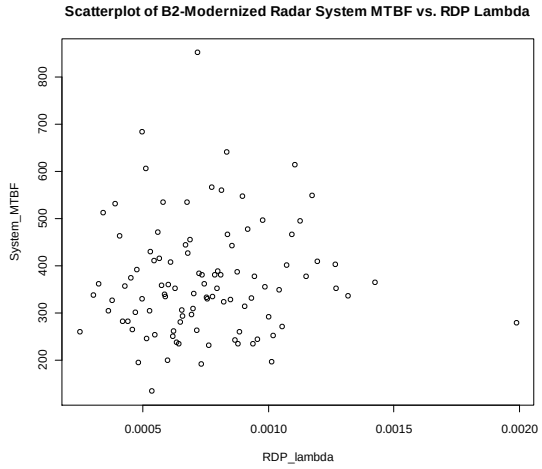
Common Means Test (CMT)			**Common Locations Test (CLT)**			**Statistical Independence Test (SI)**		
Input	MSS	Antenna	MSS	Antenna	RE	MSS	Antenna	RE
P-val	0	0	0	0	0.0488	0	0.0018	0.0488

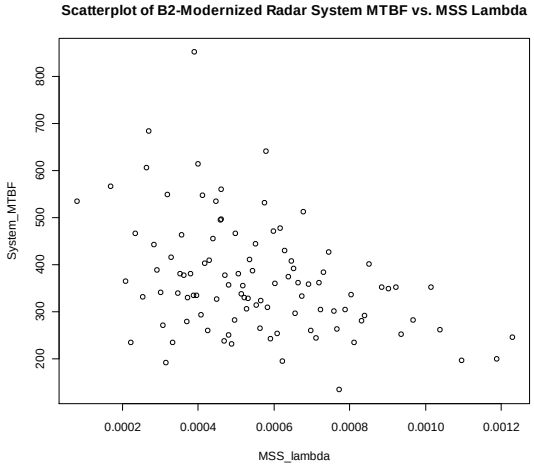
Regression Test (REG)				**Quadratic Regression Test (QREG)**			
Input	MSS	Antenna	WGSW	MSS	Antenna	WGSW	RE
P-val	0	0	0.0164	0	0	0.0224	0.0254

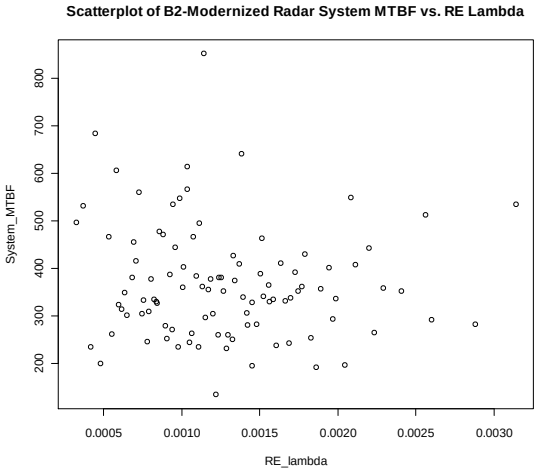
Table 8: Non-randomness test results for the full model with each component λ defined by its gamma posterior distribution

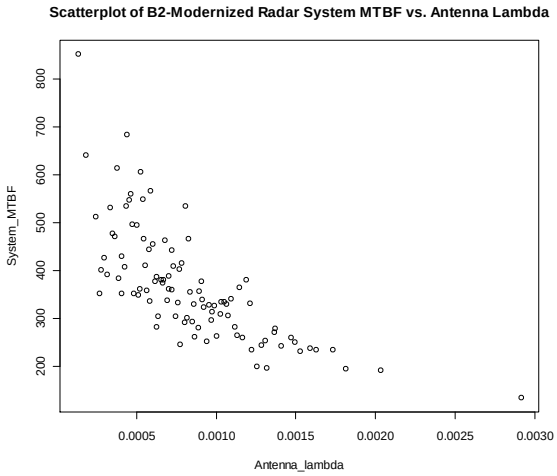

Scatterplot of B2-Modernized Radar System MTBF vs. WGSW Lambda
System_MTBF
WGSW_lambda
0e+00
1e-04
2e-04
3e-04
4e-04
200
300
400
500
600
700
800


Scatterplot of B2-Modernized Radar System MTBF vs. RSP Lambda
System_MTBF
RSP_lambda
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
200
300
400
500
600
700
800


Scatterplot of B2-Modernized Radar System MTBF vs. RDP Lambda
System_MTBF
RDP_lambda
0.0005
0.0010
0.0015
0.0020
200
300
400
500
600
700
800


Scatterplot of B2-Modernized Radar System MTBF vs. MSS Lambda
System_MTBF
MSS_lambda
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
200
300
400
500
600
700
800


Scatterplot of B2-Modernized Radar System MTBF vs. RE Lambda
System_MTBF
RE_lambda
0.0005
0.0010
0.0015
0.0020
0.0025
0.0030
200
300
400
500
600
700
800


Scatterplot of B2-Modernized Radar System MTBF vs. Antenna Lambda
System_MTBF
Antenna_lambda
0.0005
0.0010
0.0015
0.0020
0.0025
0.0030
200
300
400
500
600
700
800

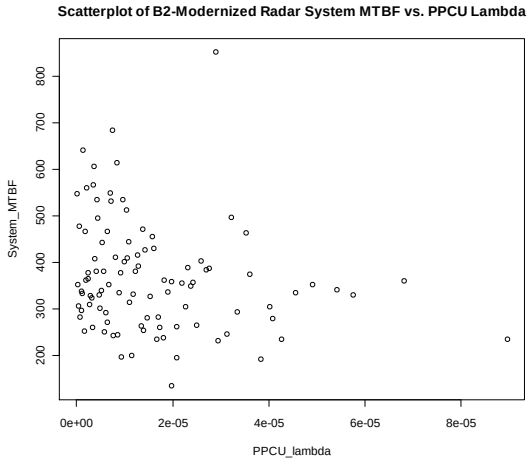


Figure 8: Scatterplots of System MTBF vs. component parameter λ for λ distributed as a gamma posterior defined in Table 7

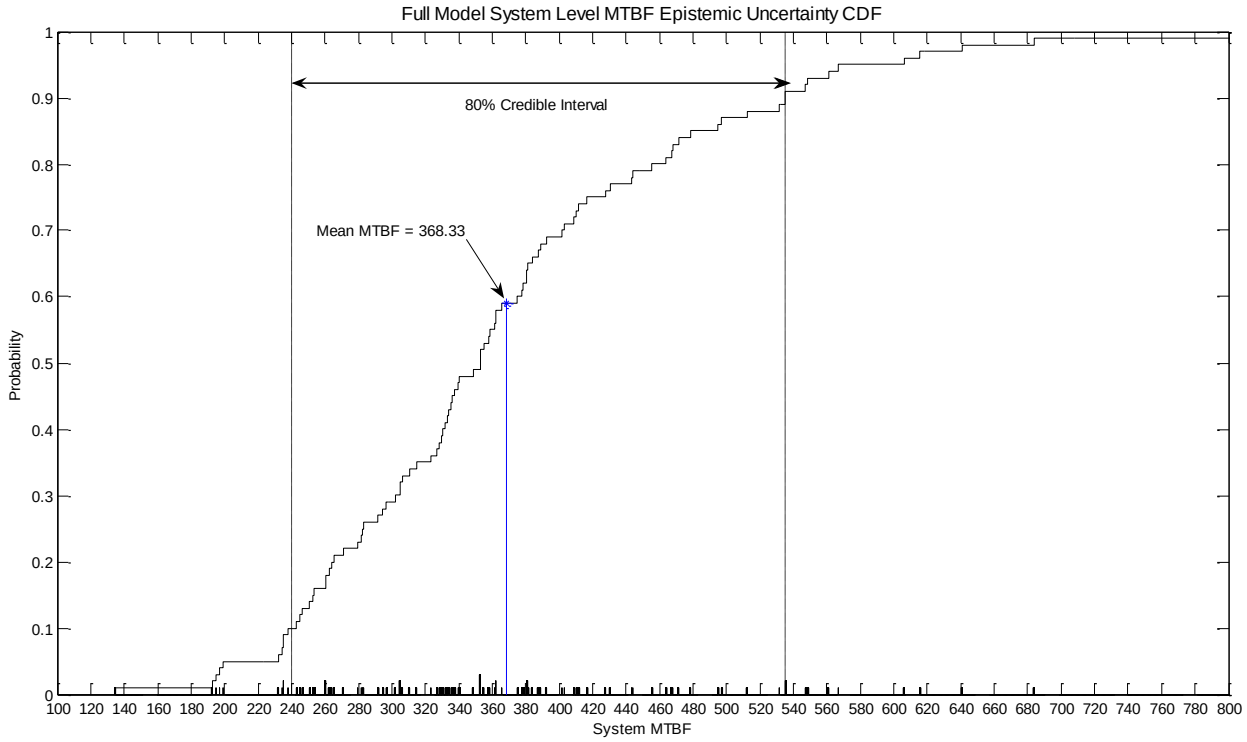


Figure 9: Plot of epistemic uncertainty histogram and CDF from a sample of 100 MTBF estimations

Figure 10 demonstrates the magnitude of the sampling error from the two sampling loops in the simulation run on the full model. The confidence bands can be interpreted as random variable vectors that, when formed repeatedly, will contain the true epistemic uncertainty CDF from the full model 95% of the time.

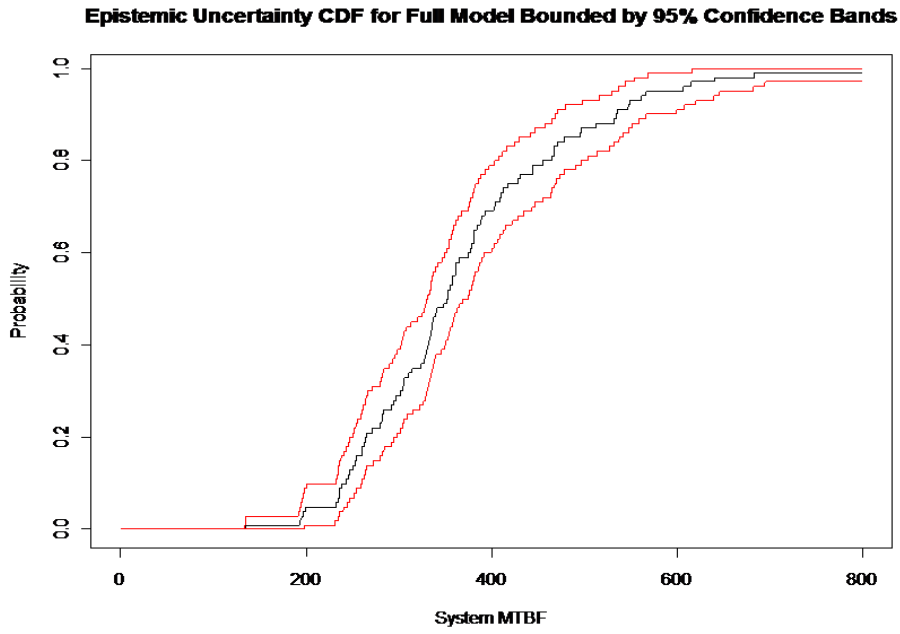


Figure 10: Plot of system uncertainty CDF bounded by 95% point-wise confidence bands

Figure 11 shows the epistemic uncertainty CDF for the driver (reduced) model plotted next to that generated from the full model. Because all non-driving component parameters remained fixed at point-estimate values, the system epistemic uncertainty CDF for the reduced model summarizes the effect of epistemic uncertainty in the antenna and MSS on epistemic uncertainty in the system MTBF. Superimposing the full and reduced model CDFs gave us a qualitative idea of well the driver model captured the true epistemic uncertainty in the system MTBF. We were interested in benchmarking the

preliminary sensitivity results for future reference by checking whether our interpretation of the results in this initial project were sound enough to extend to future projects.

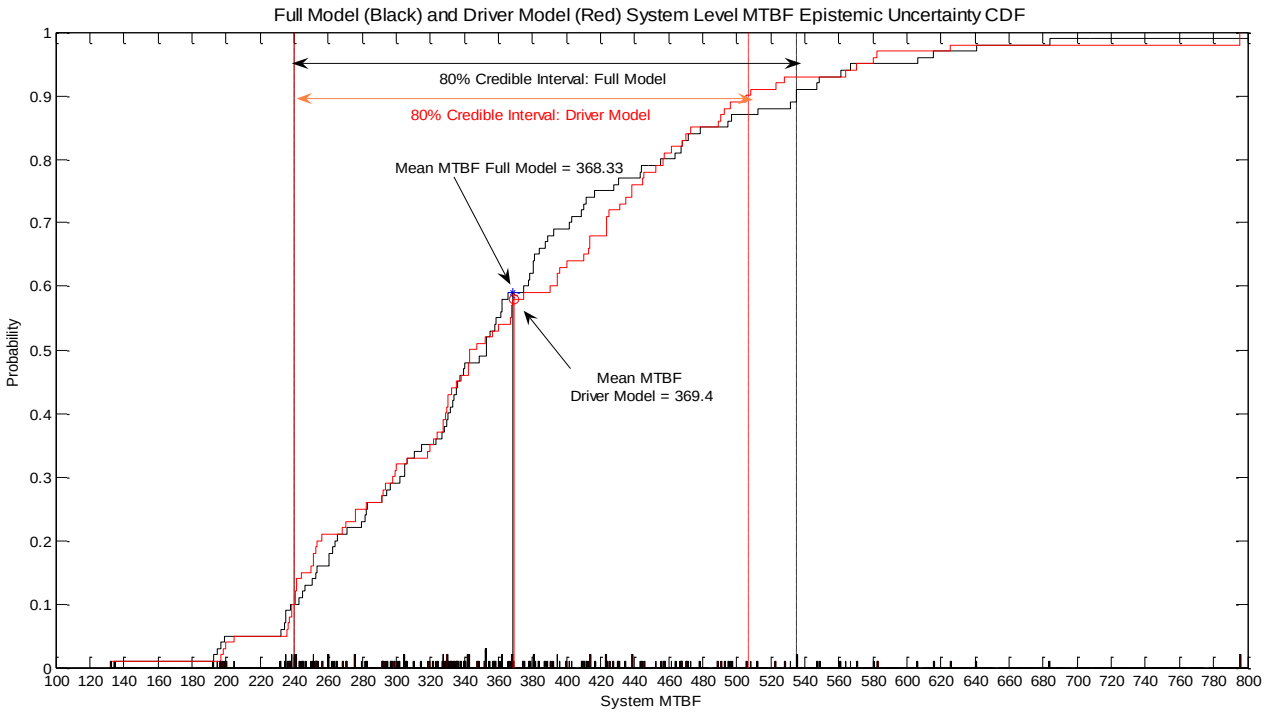


Figure 11: Plot of full model and driver model epistemic uncertainty CDFs

However, the full and reduced model epistemic uncertainty CDFs in Figure 11 are superimposed out of context. As shown by the 95% confidence bands in Figure 11, non-negligible sampling error from the two sampling loops leads to dispersion in our estimate of the full model epistemic uncertainty CDF. Figure 12 demonstrates that the driver model epistemic uncertainty CDF falls within these 95% point-wise confidence bands. Thus, within the 95% confidence bounds defining the sampling error, the two curves are equivalent.

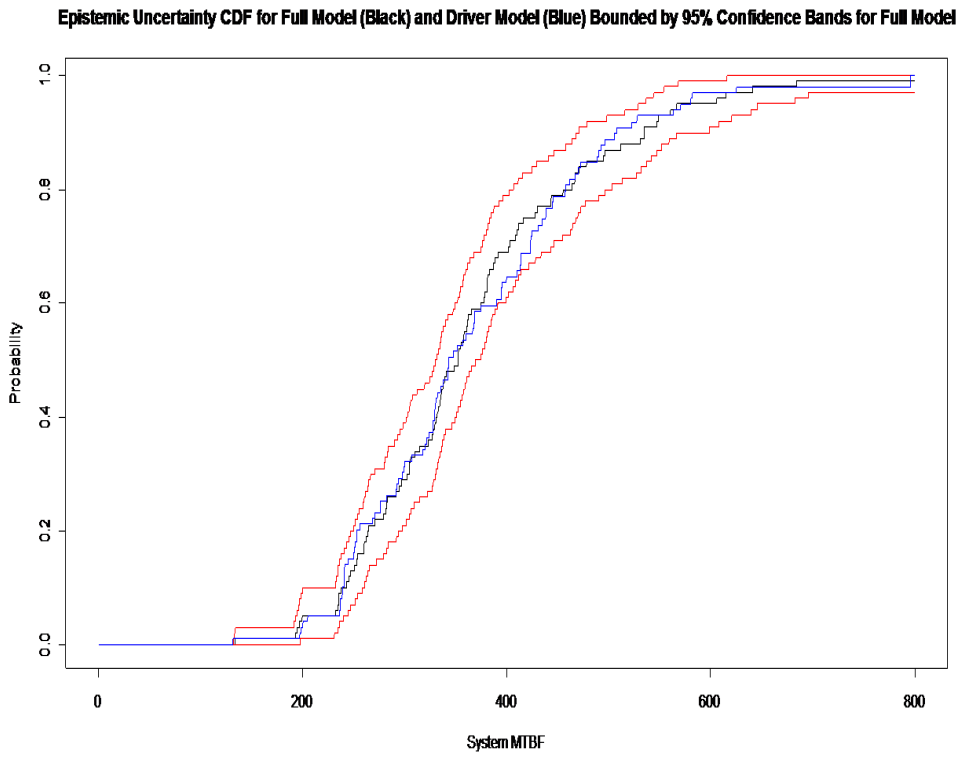


Figure 12: Plot of system uncertainty CDF for the full model (black) and the driver model (blue) bounded by 95% point-wise confidence bands estimating sampling error in full model MTBF determination.

Common statistics describing the central tendency and dispersion in the MTBF uncertainty distribution (e.g. mean and standard deviation,) can be found below in Table 9. Because the MTBF uncertainty distribution is reasonably symmetric, the mean and the 80% credible interval as statistics describe its central tendency and dispersion, respectively. Credible intervals are interpreted as direct probability statements about the variable summarized by a distribution. Note that credible intervals are interpreted differently than classical confidence intervals; many people find the interpretation of the credible interval more intuitively direct than that of a confidence interval.

The mean MTBF for the full model occurs at 368.33 hours while the 80% credible interval spans 240.3 hours to 535.1 hours. In this case, an 80% credible interval means that the true B-2 RMP MTBF lies in the interval [240.3, 535.1] hours with a probability of 80%. The mean of the driver model MTBF, 369.4 hours, is very close to that of the full model. However, the epistemic uncertainty CDF for the driver model underestimates the uncertainty in the MTBF distribution, having an 80% credible interval about 25 hours shorter than that of the full model CDF.

		Full Model	Driver Model
Central Tendency (hours)	**Sample Size**	100	100
	Median (hours)	352.66	345.33
	Mean (hours)	368.33	369.4
Dispersion (hours)	**Standard Deviation** (hours)	118.61	118.36
	80% Credible Interval (hours)	[243.3, 535.1]	[239.74, 506.89]

Table 9: Descriptive statistics for the system MTBF epistemic uncertainty distribution as determined using for full model and the reduced model including epistemic uncertainty in driving component parameters

Because we reduced the dimensionality of the epistemic uncertainty space from 14 to 4, this deviation is not surprising. It is possible that including the RE epistemic uncertainty in the model, the third most influential component isolated in the non-randomness tests,

would result in more accurate dispersion estimates for the reduced model system MTBF. However, since the magnitude of the sampling error was large in the analyses, it is difficult to gauge whether the deviation of the reduced model uncertainty CDF from that of the full model is due to sampling error or due to the reduction itself.

2.5 Conclusion

The May 2008 Defense Science Board report on Developmental Test and Evaluation identified the need for robust reliability, availability, and maintainability programs to address sustainability problems identified during testing. Operational testing for reliability requires appropriate test data to bolster conclusions and provide decision-quality information. As technology evolves and technological complexity increases, analytical techniques intended to describe and understand capabilities must also improve to keep pace. Analytical practices heavily influenced by classical statistical techniques require significant amounts of test data to provide accurate and precise conclusions. Unfortunately, test and analysis costs are often directly proportional to the technological sophistication of the system being characterized. With rising acquisition costs, test and analysis cost will increase as well. The direct competition for limited resources impacts the availability of funds for testing and precipitates the necessity for test organizations to produce quality results with limited data.

The techniques described in the main body of the paper identified a reliability analysis procedure to contend specifically with increasingly common low data scenarios

encountered during OT. The process hones reliability analysis by targeting important drivers in system-level reliability mean responses, by maximizing the use of available developmental test data, non-testing and legacy testing information from diverse sources for these drivers, and by making system-level inferences based upon more prolific component-level information.

Sensitivity analysis techniques isolate important drivers in system-level reliability. When the drivers are known, resources for testing along with the research and analysis can be focused on characterizing these drivers. Resources dedicated to testing the reliability of components driving system-level reliability yield more prolific test data for these components. Resources dedicated to accumulating non-testing or legacy testing data about component-level drivers via elicitation and research yield additional reliability information, thereby increasing the quality and density of information available for analysis.

Bayesian inference provides a means to incorporate information from these diverse sources into one distribution summarizing available information about important component parameters. The method leverages information from similar or legacy components, research and development, and DT data. Uncertainty propagation via layered sampling produces system-level epistemic and aleatory uncertainty distributions summarizing the cumulative effect of component-level epistemic and aleatory uncertainty and mean-responses on the system-level reliability metrics.

Results from applying the Bayesian and sampling process include system-level reliability estimations as well as measures, defined by dispersion in system-level parameter and parent distributions, benchmarking the level of approximation involved in the analysis. In other words, this enhanced analytical process answers two questions:

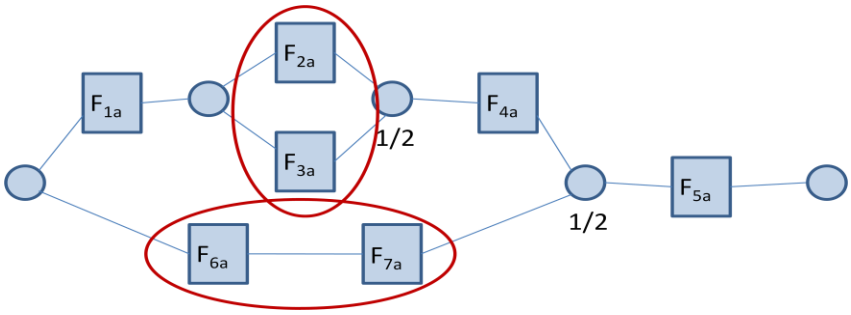
1) Approximately, how reliable is a system?
2) How approximate is this estimate of its reliability?

Standard reliability block diagram algorithms answer only the first question. Answers to both of the questions allow decision makers to determine the risk involved in purchasing and operating a system and, consequently, whether further testing or analysis is necessary.

Appendix A

This appendix demonstrates the analytical determination of the system-level time-to-failure distribution for the RBD depicted in Figure 1 (for further details consult reference [17]).

First, the innermost parallel and series structure distributions combine to form distributions summarizing the failure behavior of the structures.



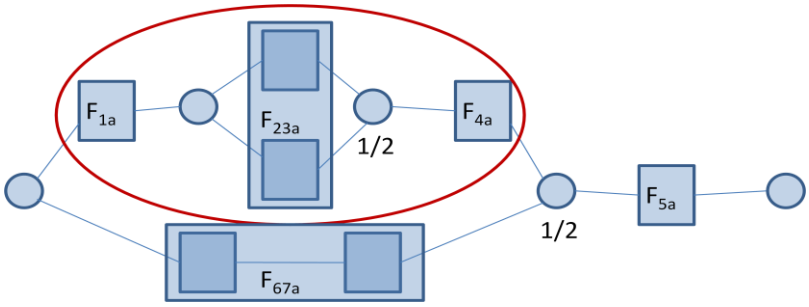
$$F_{23a} = F_{2a}F_{3a}$$

The probability that the parallel structure including blocks 2 and 3 fails is the probability that 2 and 3 fail, e.g. the probability of the event defined by the intersection of block 2 and block 3 time-to-failure event sets.

$$F_{67a} = 1 - R_{67a} = 1 - R_{6a}R_{7a} = 1-(1-F_{6a})(1-F_{7a})$$

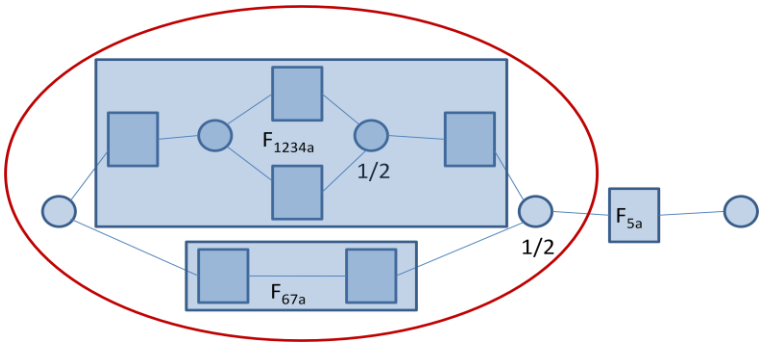
The probability that the series structure including blocks 6 and 7 fails is the probability that 6 or 7 fails, e.g. the probability of the event defined by the union of block 6 and block 7 time-to-failure event sets. Here, R_{67a} denotes the reliability of the series structure containing blocks 6 and 7.

F_1, F_{23} and F_4 combine to form the time-to-failure distribution of the series structure circled below.



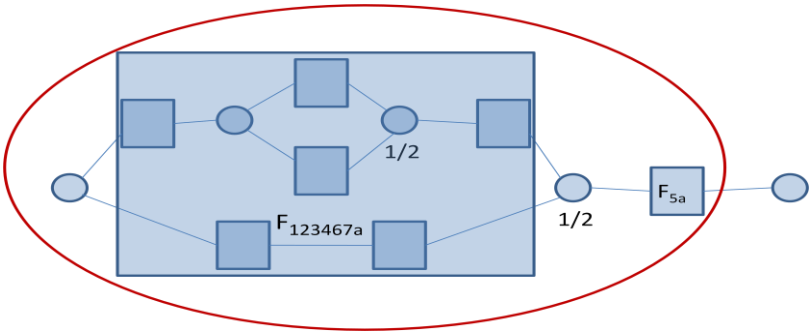
$$F_{1234a} = 1 - R_{1234a} = 1 - R_{1a}R_{23a}(R_{4a} = 1 - (1 - F_{1a})(1 - F_{23a})(1 - F_{4a})$$

The parallel structure time-to-failure distribution including the two block strings combines F_{1234a} with F_{67a}.



$$F_{123467a} = (F_{1234a})(F_{67a})$$

Finally, the system time-to-failure distribution, F_{sa}, combines $F_{123467a}$ and F_{5a} represented in the series structure below.



$$F_{sa} = 1 - R_{sa} = 1 - (1 - F_{123467a})(1 - F_{5a})$$

Appendix B

The NIST webpage on Bayesian Methods in reliability analysis provides a formula for defining a gamma prior from historical data [1]. This appendix details the derivation of this formula. Suppose the historical data describes *n* time increments in which *n-1* failures occur. This would be the case, for example, if a test time is truncated at t_c before the n^{th} failure occurs for some system. Time truncated data is also called Type I censored data [3]. If T_k denotes the time of the k^{th} failure in the time increment [0, t_c], $X_k = (T_k - T_{k-1})$ is the time between the $(k\text{-}1)^{\text{th}}$ and the k^{th} failure. When the X_k are independent and identically distributed (steady-state periods in the system's life history) as exponentials with parameter, λ, the likelihood function summarizing the time-between-failures events is the product of the probability of each time-between-failures event occurring:

$$f(x|\lambda) = \prod_{k=1}^{n-1} f(x_k), \text{ where } f(x_k) = \lambda e^{-\lambda x_k}.$$

Since the last time-between-failures measurement is truncated before a failure occurs, it is the event that $(t_c - t_{n-1})$ is not a time-between-failures event and has associated probability $(1 - F(t_c - t_{n-1}))$, where $F(\cdot)$ is the cdf of the time-between-failures random variable [3].

$$1 - F(t_c - t_{n-1}) = 1 - \int_0^{(t_c - t_{n-1})} \lambda e^{-\lambda x} dx = 1\text{-}1+e^{-\lambda(t_c - t_{n-1})} = e^{-\lambda(t_c - t_{n-1})}.$$

So, the likelihood becomes:

$$f(x|\lambda) = \left(\prod_{k=1}^{n-1} f(x_k)\right) e^{-\lambda(t_c - t_{n-1})} = \lambda^{n-1} e^{-\lambda\left[\sum_{k=1}^{n-1} x_k + (t_c - t_{n-1})\right]} =$$

$$\lambda^{n-1} e^{-\lambda\left[\sum_{k=1}^{n-1} (t_k - t_{k-1}) + (t_c - t_{n-1})\right]} = \lambda^{n-1} e^{-\lambda t_c}.$$

Bayes' Theorem generates a posterior using the likelihood of the historical data. Because no data about the system exists prior to the historical data, the input prior for Bayes' Theorem should contain only the only information we have about λ,the interval defining its domain. Since $\lambda$is defined on the positive real line, its domain is $(0, \infty)$. Forming a prior using this information is not straighforward.

Several approaches to forming such noninformative priors exist in the literature. These include using Jeffery's Rule or seeking transformations for which the likelihood is approximately data transformed (e.g. one-to-one transformations on the parameter of interest that result in a likelihood which shifts location only upon changes in the data; dispersion and shape remain constant) [18]. While these methods are rigorous and complete, they require significant effort. Instead, we resort to a simplifying approximation that models the input prior as a uniform distribution:

$$Unif_{[0,b]}(\lambda) = \frac{1}{b}\, \boldsymbol{I}_{[0,b]}(\lambda), where \boldsymbol{I}_{[0,b]} is\ the\ indicator\ function:$$

$$\boldsymbol{I}_{[0,b]} = \begin{cases} 1, if\ \lambda \in [0,b] \\ 0, if\ \lambda \notin [0,b] \end{cases}.$$

In this case, Bayes' Theorem yields:

$$f(\lambda|x) = \frac{f(x|\lambda)f(\lambda)}{\int f(x|\lambda)f(\lambda)d\lambda} = \frac{\left(\lambda^{n-1}\mathrm{e}^{-\lambda t_c}\right)\left(\frac{1}{b}\, \boldsymbol{I}_{[0,b]}(\lambda)\right)}{\int_0^{\infty}\left(\lambda^{n-1}\mathrm{e}^{-\lambda t_c}\right)\left(\frac{1}{b}\, \boldsymbol{I}_{[0,b]}(\lambda)\right) d\lambda}$$

$$= \frac{\left(\lambda^{n-1}\mathrm{e}^{-\lambda t_c}\right)\left(\frac{1}{b}\, \boldsymbol{I}_{[0,b]}(\lambda)\right)}{\int_0^{b}\left(\lambda^{n-1}\mathrm{e}^{-\lambda t_c}\right)\left(\frac{1}{b}\right) d\lambda}.$$

To derive the posterior given under the constraint that we know nothing but the limits of the domain of λ, we take the limit of $f(\lambda|x)$ as $b \to \infty$:

$$\lim_{b\to\infty} f(\lambda|x) = \lim_{b\to\infty} \frac{\left(\lambda^{n-1}\mathrm{e}^{-\lambda t_c}\right)\left(\frac{1}{b}\, \boldsymbol{I}_{[0,b]}(\lambda)\right)}{\int_0^b (\lambda^{n-1}\mathrm{e}^{-\lambda t_c})\left(\frac{1}{b}\right) d\lambda} = \lim_{b\to\infty} \frac{\frac{1}{b}\left(\lambda^{n-1}\mathrm{e}^{-\lambda t_c}\right)\left(\, \boldsymbol{I}_{[0,b]}(\lambda)\right)}{\frac{1}{b}\int_0^b (\lambda^{n-1}\mathrm{e}^{-\lambda t_c})\, d\lambda}$$

$$= \lim_{b\to\infty} \frac{(\lambda^{n-1}\mathrm{e}^{-\lambda t_c})\left(\, \boldsymbol{I}_{[0,b]}(\lambda)\right)}{\int_0^b (\lambda^{n-1}\mathrm{e}^{-\lambda t_c}) d\lambda} = \frac{(\lambda^{n-1}\mathrm{e}^{-\lambda t_c})}{\int_0^\infty (\lambda^{n-1}\mathrm{e}^{-\lambda t_c}) d\lambda}$$

The integral in the denominator can be solved using integration by parts, but this involves some tedious work. Alternatively, multiplying and dividing the integral in the denominator by $(t_c)^n$ puts it into a familiar form:

$$\left(\frac{1}{t_c}\right)^n \int_0^\infty \left((\lambda t_c)^{n-1}\mathrm{e}^{-(\lambda t_c)}\right) d(\lambda t_c) = \left(\frac{1}{t_c}\right)^n \Gamma(n).$$

And we are left with:

$$f(\lambda|x) = \frac{\left(\lambda^{n-1}\mathrm{e}^{-\lambda t_c}\right)}{\left(\frac{1}{t_c}\right)^n \Gamma(n)} = Ga\left(n, \frac{1}{t_c}\right).$$

This posterior can now be used as a prior for new data as it becomes available.

Appendix C

This appendix details the analytical determination of a posterior derived from a likelihood generated from i.i.d. interfailure times, $X_k \sim exp\ (\lambda)$, and a gamma prior. All parameters are general. The gamma is assumed to be a two parameter distribution (γ is the shape and θ is the scale parameter) with a mean equal to the product of its parameters. The X_k are assumed to be failure censored data (Type II censoring), e.g. data collection terminated after the n^{th} failure. For Type II censoring, the likelihood takes the form:

$$f(x|\lambda) = \prod_{k=1}^{n} f(x_k) = \lambda^n \mathrm{e}^{-\lambda\left[\sum_{\mathrm{k}=1}^{n}(t_k - t_{k-1})\right]} = \lambda^n \mathrm{e}^{-\lambda t_n}, \text{with } t_0 = 0.$$

When finding the posterior of a conjugate pair, the posterior distribution type matches the prior distribution type but has modified parameters. In this case, the posterior will be a gamma with parameters updated to reflect the information contained in the exponential likelihood.

Exponential Likelihood: $f(x|\lambda) = \lambda^n \mathrm{e}^{-\lambda t_n}$

Gamma Prior: $f(\lambda) = \frac{\lambda^{\gamma-1} e^{\frac{-\lambda}{\theta}}}{\theta^{\gamma}\Gamma(\gamma)},\ where\ \Gamma(\gamma) = \int_0^{\infty} t^{\gamma-1}\, e^{-t} dt$

Posterior: $f(\lambda|x) = \frac{f(x|\lambda)f(\lambda)}{\int f(x|\lambda)f(\lambda)d\lambda} = \frac{\left(\lambda^n \mathrm{e}^{-\lambda t_n}\right)\left(\frac{\lambda^{\gamma-1} e^{\frac{-\lambda}{\theta}}}{\theta^{\gamma}\Gamma(\gamma)}\right)}{\int_0^{\infty}\left(\lambda^n \mathrm{e}^{-\lambda t_n}\right)\left(\frac{\lambda^{\gamma-1} e^{\frac{-\lambda}{\theta}}}{\theta^{\gamma}\Gamma(\gamma)}\right)d\lambda}$

Since $\theta^{\gamma}\Gamma(\gamma)$ is not a function of λ, we can pull it out of the integral in the denominator and cancel it with $\theta^{\gamma}\Gamma(\gamma)$ in the numerator.

Posterior: $f(\lambda|x) = \frac{\left(\lambda^n \mathrm{e}^{-\lambda t_n}\right)\left(\lambda^{\gamma-1} e^{\frac{-\lambda}{\theta}}\right)}{\int_0^{\infty}\left(\lambda^n \mathrm{e}^{-\lambda t_n}\right)\left(\lambda^{\gamma-1} e^{\frac{-\lambda}{\theta}}\right)d\lambda} = \frac{\lambda^{\gamma+n-1} e^{-\lambda\left(t_n+\frac{1}{\theta}\right)}}{\int_0^{\infty} \lambda^{\gamma+n-1} e^{-\lambda\left(t_n+\frac{1}{\theta}\right)}d\lambda}$

Solution of the integral in the denominator:

$$\int_0^{\infty} \lambda^{\gamma+n-1} e^{-\lambda\left(t_n+\frac{1}{\theta}\right)} d\lambda \quad = \frac{1}{\left(t_n+\frac{1}{\theta}\right)^{\gamma+n}} \int_0^{\infty} \left(\lambda\left(t_n+\frac{1}{\theta}\right)\right)^{\gamma+n-1} e^{-\lambda\left(t_n+\frac{1}{\theta}\right)} d(\lambda\left(t_n+\frac{1}{\theta}\right))$$

$$= \left(\frac{1}{\left(t_n+\frac{1}{\theta}\right)^{\gamma+n}}\right) \Gamma(\gamma+n).$$

Now, we get:

Posterior: $f(\lambda|x) = \frac{\lambda^{\gamma+n-1} e^{-\lambda\left(t_n+\frac{1}{\theta}\right)}}{\left(\frac{1}{\left(t_n+\frac{1}{\theta}\right)}\right)^{\gamma+n} \Gamma(\gamma+n)} = \frac{\lambda^{\gamma+n-1} e^{-\lambda\left(t_n+\frac{1}{\theta}\right)}}{\alpha^{\gamma+n}\Gamma(\gamma+n)} = Ga(\gamma+n, \alpha),$

$$where\ \alpha = \frac{1}{\left(t_n+\frac{1}{\theta}\right)}.$$

As expected, the posterior is distributed as a gamma with shape parameter $= \gamma + n$, and scale parameter $= \alpha = \frac{1}{\left(t_n+\frac{1}{\theta}\right)}$.

Appendix D

The following is code written in R to estimate point-wise confidence bands defining sampling error in the inner MC sampling loop and outer LHS sampling loop that determine the epistemic uncertainty CDF.

```
####Function to find F(X=x.n), the step function of the CDF of a sample, xsample.vec, evaluated at
####x.n


cdfpoint <- function(x.n, xsample.vec){
  return(mean(xsample.vec <= x.n))
 }

####Function to find a vector, cdf.vec, from a vector, x.vec, of x's


Fx <- function(xsample.vec,x.vec){
 cdf.vec <- rep(0,length(x.vec))
 for(i in 1:length(x.vec)){
 cdf.vec[i]<-cdfpoint(x.vec[i],xsample.vec)}
 return(cdf.vec)}

####Script to generate Z(x) matrix, quantiles(t), and pointwise C.I.s at each x


        ##Read in system MTBF data from files on desktop

mydata.data<-read.table(file="C:\\Documents and Settings\\Bea.Yu\\Desktop\\B2data.txt",
header = T)

        ##Read in SEM data for each MTBF determination

sem.data <- read.table(file="C:\\Documents and Settings\\Bea.Yu\\Desktop\\sem.txt", header = T)

        ##Initialize variables

bs.n <- 1000

x.vec <- seq(from = 0, to = 800, by = 0.08)
MTBF.vec<-mydata.data[,8]
SEM.vec<-sem.data[,1]
lowerCI.vec<-rep(0,length(x.vec))
upperCI.vec<-rep(0,length(x.vec))

bs.mat <- matrix(0, nrow=bs.n, ncol=length(MTBF.vec))
bsobs.mat <-matrix(0, nrow=bs.n, ncol=length(MTBF.vec))
```

```
sem.mat <- matrix(0, nrow=bs.n, ncol=length(SEM.vec))
bsobs_x.mat <- matrix(0, nrow=bs.n, ncol=length(x.vec))
Z_x.mat <- matrix(0, nrow=bs.n, ncol=length(x.vec))
quant.mat <- matrix(0, nrow=length(x.vec), ncol=2)

        ##Generate 1000 jostled bootstrap samples by taking a bootstrap sample of the LHS MTBF's,
        ##forming normals with mean=MTBF, std. dev.=SEM, sampling randomly once from each
        ##normal and replacing the MTBF with the random sample from the normal

for(k in 1:bs.n){
                set.seed(k)
                bs.mat[k,] <- MTBF.vec[sample(length(MTBF.vec),replace=TRUE)]
                set.seed(k)
                sem.mat[k,]<-SEM.vec[sample(length(SEM.vec),replace=TRUE)]
                for(j in 1:length(MTBF.vec)){
                                bsobs.mat[k,j]<-rnorm(1,mean = bs.mat[k,j],sd = sem.mat[k,j])
                                }
                }

        ##Generate a step function for the CDF of MTBF uncertainty

F_x.vec <- Fx(MTBF.vec,x.vec)

        ##Generate a step function for the jostled bootstrap samples and subtract this function
        ##evaluated at values in x.vec from the values of F_t.vec evaluated at the same points

for(k in 1:bs.n){
                bsobs_x.mat[k,] <- Fx(bsobs.mat[k,],x.vec)
                Z_x.mat[k,] <- (F_x.vec - bsobs_x.mat[k,])
                }

        ##Produce the 2.5th and 97.5th percentiles to estimate a 95% CI

for(j in 1:length(x.vec)){
                quant.mat[j,] <- quantile(Z_x.mat[,j],probs=c(0.025,0.975),names = FALSE)
                }

upperCI.vec <- (F_x.vec - quant.mat[,1])
lowerCI.vec <- (F_x.vec - quant.mat[,2])

plot(x.vec, F_x.vec, main = "Epistemic Uncertainty CDF for Full Model and Driver Model Bounded by
95% Confidence Bands", xlab = "System MTBF", ylab = "Probability", type = "l")
lines(x.vec,upperCI.vec, col = "red")
lines(x.vec,lowerCI.vec, col = "red")
```

Appendix E

List of Acronyms

AFOTEC – Air Force Operational Test and Evaluation Center
B-2 RMP – B-2 Radar Modernization Program
CDF – Cumulative Distribution Function
CI – Confidence Interval
CLT – Common Locations Test
CMT – Common Means Test
DT – Developmental Test
LHS – Latin Hypercube Sampling
NGC – Northrop Grumman Corporation
NIST – National Institute of Standards and Technology
MC – Monte Carlo
MCMC – Markov Chain Monte Carlo
MSS – Motion Sensing System
MTBF – Mean Time Between Failures (statistic describing the mean uptime for a repairable system)
MTTF – Mean Time To Failure (statistic describing the mean time to failure for a non-repairable system)
OT – Operational Test
PPCU – Prime Power Conditioning Unit
QREG – Quadratic Regression Test for Non-Randomness
RBD – Reliability Block Diagram
RBD/MC – Reliability Block Diagram Model Using Monte Carlo Simulation
R & D – Research and Development
RDP – Undefined acronym for component in modernized radar
R/E – Receiver/Exciter
REG – Regression Test for Non-Randomness
RMP - Radar Modernization Program
RSP - Undefined acronym for component in modernized radar
SI – Statistical Independence Test
WG-SW – Wave-Guide Switch

References

1. Oberkamph, William L.; DeLand, Sharon M. Rutherford, Brian M.; Diegert, Kathleen V.; Alvin, Kenneth F. *Estimation of Total Uncertainty in Modeling and Simulation*; Sandia Report; April 2000

2. Rausand, Marvin and Hoyland, Arnljot. *System Reliability Theory; Models, Statistical Methods, and Applications.* Hoboken : John Wiley & Sons, Inc., 2004.

3. Hamada, Michael S., Wilson, Alyson G., Reese, C. Shane, Martz, Harry F., *Bayesian Reliability*, Springer, New York, 2008.

4. Hammersley, J.M.; Handscomb, D.C. *Monte Carlo Methods*. Halsted Press, New York, 1975

5. Casella, George, Berger, Roger, *Statistical Inference*, The Wadsworth Group, 2002

6. Brall, A., Hagan, W., Tran, H., *Reliability Block Diagram Modeling – Comparisons of Three Software Packages*, Reliability and Maintainability Symposium, 2007

7. Efron, B., Tibshirani, R.J., *An Introduction to the Bootstrap*, Chapman & Hall, 1993.

8. Helton, J.C., Wilson, Johnson, J.D., Sallaberry, C.J., C. Storlie, C.B., *Survey of Sampling-Based Methods for Uncertainty and Sensitivity Analysis*, Reliability Engineering and System Safety, 91, pp.1175-1209, 2006.

9. Robert, Christian P., *The Bayesian Choice*, 2nd ed., Springer, New York, 2007. Ntzoufras, Ioannis, *Bayesian Modeling Using WinBUGS*, John Wiley & Sons Inc., 2009.

10. Lorio, John F., Dvorack, Michael A., Mundt, Michael J., Diegert, Kathleen V., Ringland,James T., Zurn, Rena, Anderson-Cook, Christine, Huzurbazar, Aparna, Wilson, Alyson, and Fatherley, Quinn, *Quantifying Reliability Uncertainty: A Proof of Concept*, Sandia Report, SAND2009-2173, October 2009.

11. O'Hagan, Anthony, Buck, Caitlin E., Daneshkhah, Alisreza, Eiser, J. Richard, Garthwaite, Paul H., Jenkinson, David J., Oakley, Jeremy E., and Rakow, Tim, *Uncertain Judgements: Eliciting Experts' Probabilities*, John Wiley and Sons, Ltd, West Sussex, 2006.

12. Saltelli, A.; Chan, K. and Scott, E.M. *Sensitivity Analysis*. Hoboken : John Wiley & Sons, Inc., 2000.

13. Helton, J.C., Davis, F.J., *Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems,* Reliability Engineering and System Safety, 81, pp. 23-69, 2003.

14. McKay, M.D., Beckman, R.J., Conover, W.J., *A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code*, Technometrics, Vol. 21, No. 2, May 1979.

15. http://itl.nist.gov/div898/handbook/apr/section2/apr25.htm

16. http://www.weibull.com/SystemRelWeb/analytical_life_predictions.htm

17. Box, G.E.P, Tiao, G.C., *Bayesain Inference in Statistical Analysis,* John Wiley and Sons, Inc., 1992