TEST CAMPAIGN DESIGN FOR MODEL UNCERTAINTY REDUCTION

By

Kyle Scott McLemore

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for

the degree of

MASTER OF SCIENCE

in

Civil Engineering

May, 2012

Nashville, Tennessee

Approved:

Professor Sankaran Mahadevan

Professor Ronald Schrimpf

To all of my teachers, both formal and informal

ACKNOWLEDGMENTS

I am greatly appreciative of my research advisor, Dr. Sankaran Mahadevan, whose support and guidance has been unwavering throughout my time at Vanderbilt. Dr. Mahadevan's influence has been fundamental to my continued development as an engineer and a person. I would also like to thank Dr. P.K. Basu who was instrumental in my attendance and success at Vanderbilt.

This research received financial support of the National Aeronautics and Space Administration's Jet Propulsion Laboratory in Pasadena, California. I would like to thank the technical monitors Dr. Lee Peterson and Dr. Case Bradford of the California Institute of Technology for their insight and direction throughout the project. Also, I would like to thank Bell Helicopter and Federal Aviation Administration for their support with the helicopter rotor mast crack growth problem.

I would like to give special recognition to Dr. Shankar Sankararaman for the incredible support he has provided me over the past two years. Dr. Sankararaman's patience, mentorship, and friendship have been invaluable to me and my research.

Finally, I would like to thank the members of the Risk and Reliability Research Team at Vanderbilt for their help and support during this process. In particular, I would like to thank Dr. Vadiraj Hombal and Adam Wolfe for their help with the helicopter rotor mast problem.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Motivation

Throughout the design, fabrication, certification, and operation of an engineering system, measures are taken to reduce the uncertainty regarding the system's performance. In complex and expensive engineering systems, such as those used in the aerospace industry, activities during system tests and periodic inspections are essential to the system's lifespan. Uncertainty reduction regarding system performance is vital to the minimization of risk. The goal in a testing campaign is to develop a sufficiently complete understanding of the performance of the system in question in the most economically efficient manner. Tests need to be performed and designed with adequate fidelity and resolution so that the results provide meaningful information that can be used to reduce the uncertainty of the full system performance. Information gained from lower level testing data is used to assess full system performance through the use of computer model simulations. These models need to be rigorously verified against numerical benchmarks, and systematically validated by a hierarchy of component and subsystem tests. Research into quantification of margins and uncertainties (QMU) has the goal of enabling this overall capability [1].

Currently, test campaign design procedures for uncertainty reduction are developed based on prior experience and ad hoc rules. However, using these approaches, it is difficult to quantify the test campaign's return on investment in terms of risk reduction. Thus, a systematic mathematical approach to test campaign design provides an attractive alternative to current procedures. Bayesian statistics provide the necessary tools for a mathematical approach to test campaign design because of their ability to include uncertainties due to natural variability and lack of knowledge. This research uses a Bayesian framework to develop rigorous analytical methodologies for decision making about the selection and design of tests and inspections for uncertainty reduction in complex engineering systems. Topics explored include: Test resource allocation and design of hierarchical system calibration tests, test design for the purpose of integrated computational materials engineering (ICME), and inspection type decision combination optimization during the system's operational life.

1.2 Preliminaries

Prior to the implementation of a test campaign methodology, both aleatory and epistemic sources of uncertainty must be identified. Aleatory uncertainty includes natural variability in the model inputs and model parameters and leads to variability in the model outputs. Epistemic uncertainty (due to lack of knowledge) includes testing measurement errors as well as uncertainties due to the presence of sparse and interval data in the inputs and outputs. Epistemic uncertainty has been addressed using both probabilistic [2] and non-probabilistic approaches [3].

Prior to the implementation of the resource allocation and test design methodologies, models representing the behavior of materials, components, subsystems, and the overall system of interest must be developed. Researchers have investigated probabilistic as well as non-probabilistic approaches for uncertainty quantification, model verification, and model validation. Probabilistic approaches include sampling methods [4, 5], reliability methods [6, 7], Bayesian methods [8, 9], entropy–based methods [10, 11], etc. Non-probabilistic methods include evidence theory [12, 13, 14], convex models of uncertainty [15], Zadeh's extension principle [16], fuzzy theory [2], etc. While probabilistic methods have been used to address natural variability, non-probabilistic methods have been used to address epistemic uncertainty due to interval data; however, probabilistic approaches to address interval data have also been recently developed [3, 17].

Test campaigns can be developed for both calibration and validation of system models. This thesis focuses on model calibration, but future work could extend the proposed methodologies to validation. A rigorous approach to model verification and validation should explicitly account for the various sources of uncertainty such as physical variability, information uncertainty, model error, measurement error, etc. and develop a robust metric that can quantitatively judge the performance of the model and assess the confidence in the model prediction. Statistical confidence intervals [18, 19], classical statistical hypothesis testing [20, 21], Bayesian hypothesis testing [22], and reliability-based approaches [23] have been investigated for the purpose of validation. The Bayesian hypothesis testing approach [22, 24, 25] is found to be effective as it can account for both aleatory and epistemic uncertainties [26] and directly compute the

probability that the model is correct based on the available validation data. Jiang and Mahadevan [27] established a relationship between the Bayesian validation metric and the risk associated with decision making in model validation.

Recently, Urbina et al [28] developed a Bayesian network approach that serves as a foundation for the development of the selection methodologies in this research. This thesis extends the Bayesian network approach to efficiently make decisions regarding testing and inspections of realistic engineering problems, including multi-level, multi-disciplinary systems. The Bayesian network is an ideal choice for this purpose because it can:

1. Connect (i) component-level, subsystem-level, and system-level models; (ii) corresponding model inputs, parameters, outputs, and errors; and (iii) test data and measurement errors at different levels.
2. Include different types of uncertainty - natural variability, data uncertainty, and model uncertainty.
3. Solve both types of uncertainty quantification problems: forward and inverse.

Once properly developed system models, model parameters, testing data, and errors have been connected through a Bayesian network, this tool can be used to solve a variety of engineering design and resource allocation problems. The following section will outline the specific applications highlighted by this research.

1.3 Review of research and organization of thesis

The overall goal of the presented research is to develop rigorous analytical methodologies for the selection and design of tests and inspections on engineering systems at the material, component, subsystem, and full system levels for the purpose of reducing the uncertainty in full system performance. Each methodology uses a Bayesian network to provide a unified framework that connects information between all levels and all physics involved in each problem. This thesis is organized into seven chapters. Chapter I provides an introduction to the scope of the presented research and outlines the organization of this thesis. Chapter II provides a brief discussion of the key underlying statistical and mathematical processes used in this research, which include: Bayesian analysis, global sensitivity analysis, Gaussian process surrogate modeling, Latin hypercube sampling, and the Kullback-Leibler distance. Four methodologies, each solving a different type of system testing issue are presented in Chapters III-VI, which are briefly outlined in the remainder of this section.

Chapter III develops a methodology for the allocation of testing resources in hierarchical systems. Using a Bayesian network, models, testing data, and errors can be connected in multi-level multi-physics systems. The proposed methodology seeks to select the optimal combination of material, component, subsystem, and full system tests so that the uncertainty in the full system performance is minimized. The proposed methodology is shown on two example problems: (1) a multi-disciplinary thermal vibration problem, and (2) a multi-level multidisciplinary simplified space telescope mirror.

Chapter IV builds upon the methodology developed in Chapter III in order to design the selected tests in an adaptive way. By seeking to select the input settings of the model calibration tests as well as the types of tests to be performed based on the results of previously performed tests, this proposed methodology adapts future tests to previously acquired data. This approach results in a sequence of designed tests that seeks to minimize the uncertainty in performance prediction of the full multi-level multi-physics system. The methodology is demonstrated on the same multi-disciplinary thermal vibration problem shown in Chapter III.

Chapter V extends the methods developed in Chapter IV to address the issue of materials design. Materials have traditionally been treated as a constraint in the design of complex engineering systems. ICME seeks to include materials design in the engineering design process. This goal comes with numerous computational challenges including the issue of uncertainty quantification. This chapter proposes a methodology to optimize the variables of a manufacturing process such that the performance of the manufactured component is optimized with a certain level of confidence. This methodology is demonstrated using a helicopter rotor mast subjected to both bending and torsion loading. The problem optimizes the depth of shot peening (a manufacturing process) that best extends the useful life of the component given an initial defect size and loading scenario.

Chapter VI develops a methodology to make decisions about inspection types. Digital twins are computer models that seek to exactly replicate a complex engineering system and include all uncertainties about the state of the system. Uncertainties are reduced by calibrating the digital twin with test and inspection data. The methodology proposed in this chapter seeks to select the best inspection fidelity such that the digital

twin is updated in the most economical manner. The methodology is presented using a helicopter rotor mast crack growth problem.

Chapter VII provides conclusions about the methodologies developed and potential directions for future research on resource allocation for uncertainty reduction.

CHAPTER II

BACKGROUND INFORMATION

Chapter II will provide a brief introduction to the key underlying statistical and mathematical processes that are essential to the formulation of the test campaign design methodologies developed in this thesis: Bayesian analysis, global sensitivity analysis, Gaussian process surrogate modeling, Latin hypercube sampling, and the Kullback-Leibler distance.

2.1 Bayesian networks

Bayesian analysis updates a prior assumption about the distribution of a random variable based on observed instances of that variable or functions of that variable. A Bayesian network [29, 30] is a graphical representation of the relationship between various uncertain quantities in a system. With the Bayesian network, the outputs of component and subsystem level tests can be related to the inputs of the overall system.

Each uncertain quantity is represented as a node and successive links are connected to each other using unidirectional arrows that express dependence in terms of conditional probabilities. Disconnected nodes imply independence between the corresponding random variables. Figure 1 shows a conceptual Bayesian network that aids in uncertainty quantification across multiple levels of models and observed data. Circles correspond to uncertain variables and squares represent observed data. A solid

line arrow represents a conditional probability link, and a dashed line arrow represents the link of a variable to its observed data if available.

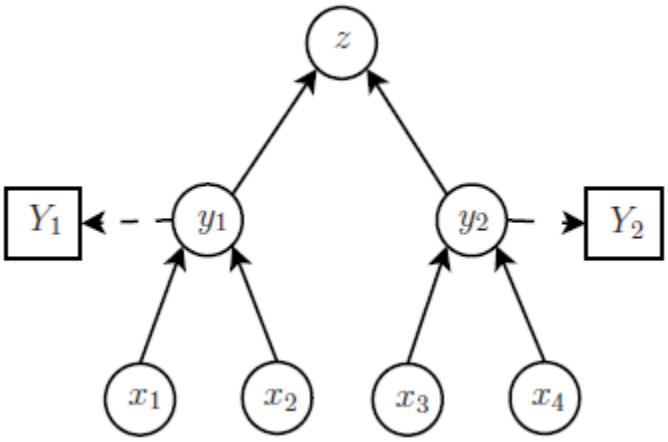

Figure 1: Bayesian Network Illustration

In Figure 1, a system level output z is a function of two subsystem level quantities y_1 and y_2; in turn, y_1 is a function of subsystem-level input x_1 and model parameter x_2, and similarly, y_2 is a function of subsystem-level input x_3 and model parameter x_4. For example, in a beam deflection study, the applied force is an input, the elastic modulus is a model parameter, while the deflection is measured and a model is built to predict the deflection. Experimental data Y_1 and Y_2 are available for comparison with the respective model predictions y_1 and y_2.

In the forward problem, the probability distributions of the inputs (x_1 and x_3) and model parameters (x_2 and x_4) are known or assumed, and these distributions are used to calculate the probability density function (PDF) of y_1 and y_2, which in turn are used to calculate the PDF of the system-level output z as:

$$f_Z(z) = \int f_Z(z|y_1, y_2)\, f_{y_1}(y_1|x_1, x_2) f_{y_2}(y_2|x_3, x_4) f_{x_1}(x_1) f_{x_2}(x_2) f_{x_3}(x_3) f_{x_4}(x_4) dx_1 dx_2 dx_3 dx_4 \quad (1)$$

Equation 1 can be solved using methods of uncertainty propagation such as Monte Carlo simulation (MCS), first-order reliability method (FORM), second-order reliability method (SORM), etc. [7].

In the inverse problem, the probability densities of the model parameters (x_2 and x_4 in Figure 1) can be updated based on the observed data (Y_1 and Y_2) using Bayes theorem as:

$$f(x_2, x_4|Y_1, Y_2) = C\ L(x_2, x_4) f'(x_2) f'(x_4) \quad (2)$$

The joint posterior density is given by $f(x_2, x_4|Y_1, Y_2)$. The likelihood function, $L(x_{2,} x_4)$, is calculated as the probability of observing the given data (Y_1, Y_2), conditioned on the parameters being updated, i.e. x_2 and x_4. The likelihood function accounts for the uncertainty in the inputs x_1 and x_3. For details of the likelihood function, refer to [31, 32].

A prior distribution ($f'(x_2)$ and $f'(x_4)$) is a representation of all of the knowledge known about a parameter before collecting any additional data. Prior distributions of the random variables of interest can be either informative or non-informative. If a significant amount of information is known about the behavior of a variable, an informative prior can be used and will assist the analysis. For instance, the parameters of common materials may have well known distributions, manufactured

products may have specified nominal values and tolerance levels, or an expert may have a high level of confidence that a parameter falls in a certain interval range. This concept of the prior distribution is heavily scrutinized by detractors of Bayesian statistical methods because the assumptions made about the prior distribution affect the result of the Bayesian analysis. Poor assumptions of prior information will ineffectively bias the resulting posterior distributions. However, the inclusion of useful information into a prior distribution increases the effectiveness of the analysis and eases the computational difficulty. The prior distribution describes the subjective knowledge about the system and the associated uncertainty. This prior knowledge is updated in the Bayesian analysis process when new information becomes available in the form of testing or inspection data. This leads to the reduction of uncertainty in the system-level prediction. For more information about prior distributions, including non-informative priors, refer to [9].

2.2 Global sensitivity analysis

Global sensitivity analysis [33, 34, 35] is a powerful tool used to determine the relative contribution of each source of uncertainty to the overall uncertainty of the model outputs. Unlike local sensitivity analysis, global sensitivities average the conditional variance $Var_{X_{\sim i}}(Y|X_i)$ over the entire distribution of X_i. Global sensitivity is generally expressed in the form of two indices, the first order effects index and the total effects index. The first effects index of a particular variable is always between zero and one, is indicative of the variable's contribution to the system variance without considering its interaction with other variables, and is calculated as:

$$S_{i,F} = \frac{Var_{X_i}\left[E_{X_{\sim i}}(Y|X_i)\right]}{Var(Y)} \tag{3}$$

where X_i is the random variable of interest, $X_{\sim i}$ is all random variables excluding X_i, and Y is the model output.

The sum of the total effects indices of all X_i is always greater than or equal to one. The total effects index for a particular variable is indicative of the variable's contribution to the system variance including its interaction with other variables and is calculated as:

$$S_{i,T} = \frac{E_{X_{\sim i}}\left[Var_{X_i}(Y|X_{\sim i})\right]}{Var(Y)} \tag{4}$$

In the context of test design, global sensitivity analysis provides a quantitative measure of which variables are important to the overall model output. This information aids in the process of determining which model parameters to update within the Bayesian network and eliminates the need for tests which do not update "important" parameters. However, global sensitivity analysis does not give any indication of the expected reduction of uncertainty in the full system output before performing a given test. Nor does it determine what value of test input settings are most likely to maximize the information gained from performing a calibration test. Thus, global sensitivity analysis is a necessary tool in the development of the test design optimization problem, but is not sufficient to provide a solution.

2.3 Gaussian process surrogate modeling

The proposed procedure for test resource allocation requires several hundred thousand evaluations of the component-level, subsystem-level, and system-level models. If these are physics-based models implemented in time-consuming computer codes (e.g. finite element models), then this is a computationally expensive task. Hence, a few runs of the actual model are used to train an inexpensive, efficient surrogate model in this research. A Gaussian process surrogate model has been used for this purpose. Different types of surrogate modeling techniques (conventional polynomial response surface, polynomial chaos expansion [36], support vector regression [37], relevance vector regression [38], and Gaussian process interpolation [39]) have been investigated in the literature. This thesis uses the Gaussian process (GP) surrogate model, which is a powerful technique based on spatial statistics for interpolating data and is increasingly being used to build surrogates to expensive computer simulations for the purposes of optimization and uncertainty quantification [39, 40, 41]. The GP model is (1) not constrained by polynomial-type functional forms; (2) capable of representing highly nonlinear relationships in multiple dimensions; and (3) produces an estimate of the prediction uncertainty which depends on the number and location of training data.

The basic idea of the GP model is that the response values Y evaluated at different values of the input variables $\boldsymbol{X}$, are modeled as a Gaussian random field, with a mean and covariance function. Suppose that there are m training points, $x_1, x_2, x_3 \ldots x_m$ of a d-dimensional input variable vector, yielding the output values $Y(x_1), Y(x_2), Y(3) \ldots Y(x_m)$. Let Q denote the m x m matrix of correlations among input variables at the training points. Under the assumption that the parameters

governing both the trend function ($f^T(x_i)$ at each training point) and the covariance (λ) are known, the expected value and variance of the Gaussian process at a prediction location x^* is calculated as:

$$E\big(Y(x^*)\big) = f^T(x^*)\beta + u^T(x^*)Q^{-1}(Y - F\beta)$$
$$Var\big(Y(x^*)\big) = \lambda(1 - u^T Q^{-1} u) \tag{5}$$

In Equation 5, u represents the vector of correlations between the prediction point, x^*, and the training points, β represents the vector of trend coefficients, and F represents the vector of trend functions. For details of this method, refer to [40,41,42].

Adaptive techniques can be used to select training points for the GP model, in order to construct the response surface to a desired level of accuracy or precision. Since the GP model is capable of estimating the variance in the model output, a variance minimization algorithm proposed by McFarland [41] identifies the next training point at the input variable value which corresponds to the largest variance. By repeating this algorithm, training points are adaptively identified until the estimated variance is below a desired threshold. Additionally, training point design has been extended to minimize bias [43]. After the Gaussian process surrogate model is constructed and its parameters are estimated, it is used for Bayesian updating and test resource allocation optimization.

2.4 Latin hypercube sampling

Basic Monte Carlo simulation may not be affordable for a problem where a large number of samples are required to obtain a reasonable estimate of the distribution. In

these cases, advanced sampling techniques are used to reduce the required number of function evaluations. Latin hypercube sampling [44] is one such technique which ensures that each segment in the range of an input random variable is represented in a manner proportional to its probability. In Latin hypercube sampling, each input dimension is divided into a number of segments equal to the number of samples such that each segment has an equal probability of occurrence. When samples are taken, they will be distributed such that the same number of samples will be taken from each segment. Clearly, there are many combinations of samples that would meet these criteria, and the combinations become more numerous as the dimensionality increases. In order to ensure that the entire domain is reasonably covered, optimum symmetric sample design procedures that seek to maximize the minimum distance between any two samples have been developed [45]. Latin hypercube sampling is used extensively in this research to reduce the necessary number of function evaluations that must be performed.

2.5 Kullback-Leibler distance

The pursuit of data that leads us to a better understanding of the performance prediction of the full system leads to an important question: How can the amount of information gained be described in a quantitative manner? Two metrics are used in this research to quantitatively judge the uncertainty reduction in a variable after additional information is gained: (1) variance reduction and (2) the Kullback-Leibler distance. In a variance reduction technique, the testing campaign that corresponds to the posterior output variable with the lowest variance is chosen as optimal. This technique is valid if the "true" value of the calibration parameter is a single deterministic value. This is the

case when the calibration tests being performed use the same component as the component to be used in the full system. However, this metric is not valid if the components being used in the full system are not the exact same specimen as is being tested. In this case, the calibration parameters will in reality be distributions. When this is the case, the model calibration parameters cannot be updated directly and the distribution parameters of the calibration parameters must be updated. When this is the case, an information-theoretic approach, the Kullback-Leibler distance, can be used.

The Kullback-Leibler (KL) distance is an information-theoretic distance developed by Kullback and Leibler [46] used to measure the similarity between two probability distributions. The KL distance in units of 'nats' is given as:

$$d_{KL}(g,h) = \int g(y) ln\left[\frac{g(y)}{h(y)}\right] dy, \tag{6}$$

where $g(y)$ is the PDF of the prior distribution and $h(y)$ is the PDF of the posterior distribution of interest. Note that Equation 6 is asymmetrical, meaning that $d_{KL}(g,h) \neq d_{KL}(h,g)$.

The higher the KL distance, the less similar the posterior distribution is from the prior distribution, which corresponds to a larger amount of information gain. If the KL distance is equal to zero, the two functions are identical. Therefore, in order to maximize the information gained by acquiring new testing data, the proposed methodology in Chapter IV seeks to select the type of test and the test settings that maximize the expected KL distance between the prior model output PDF and the posterior output PDF.

CHAPTER III

ALLOCATION OF TESTING RESOURCES IN HIERARCHICAL SYSTEMS

3.1 Introduction

Typically, a multi-level multi-physics system has several parameters that influence the overall system-level output, and the uncertainty in these parameters can be updated by tests at multiple levels of the system and multiple types of physics coupling. When the posterior distributions of the parameters are propagated through the system model to calculate the overall system-level output, the posterior variance of the overall system-level prediction can be computed. With more acquisition of data, a decreasing trend can be observed in the variance of the system-level output.

Two types of questions need to be answered: (1) What type of tests to do (which component, isolated physics, etc.)? and (2) How many repetitions of each type? Each type of test has a different testing cost and an associated reduction in the variance of system-level prediction. Further, the same type of test may need to be repeated on nominally identical specimens of the same component or subsystem. Such repetition is performed in order to account for the effect of natural variability across nominally identical specimens; while each repetition may have the same monetary cost, the associated reduction in the variance of system-level prediction may be different.

The test conducted on one subsystem is assumed to be statistically independent of another test on another subsystem; in other words, one type of test is independent of any other type. Further, for a given type of test, the repetitions across multiple replicas are

also considered to be independent. It is assumed that a model is available to predict the quantity being measured in each type of test; the model may have several outputs but only that output which is measured is of concern.

The key idea of the proposed methodology is to use all available component-level models and test data to quantify the uncertainty in the overall system level performance prediction. This methodology combines two types of inverse problems (model calibration and test resource optimization) and forward uncertainty propagation. The probability distributions of the model parameters are updated using the Bayesian network after collecting data through testing, and the updated distributions are propagated through the component and system models to recalculate the variance in the system performance prediction. An optimization-based procedure is then used to aid in test resource allocation by taking into consideration the reduction in variance due to testing, as well as the costs involved in testing, thereby facilitating efficient cost-benefit analysis.

3.2 Global sensitivity analysis in resource allocation

The first step of the proposed resource allocation methodology is to use sensitivity analysis to identify those parameters that have a significant influence on the variance of the overall system-level prediction. Once the "important" parameters are identified, only those tests that aid in reducing the uncertainty in these important parameters should be included in the analysis. For example, consider a system-level output that is very sensitive to the uncertainty in the parameters of sub-system-I but not sensitive to the parameters of sub-system-II, then it is logical to perform more sub-system-I tests than sub-system-II tests. Note that this procedure for test identification is only a preliminary

approach. This approach can answer the question - "which tests to do?" In order to answer the question, "how many tests to do?", it is necessary to quantify the decrease in variance that may be caused due to a particular test. The effect of a particular test on variance reduction can be quantified by using Bayesian updating. Therefore, the proposed resource allocation methodology first uses sensitivity analysis for selection of calibration parameters and then uses Bayesian updating to quantify the effect of a test on the variance of system-level prediction.

3.3 Optimization formulation

In order to solve the resource allocation problem and identify the number of tests to be performed for each type, the optimization problem can be formulated in two ways, as explained below.

In the first formulation shown in Equation 7, the goal is to minimize the variance of the system-level output subject to satisfying a budget constraint.

$$\begin{aligned} &\underset{N_{test}}{\text{Minimize}}\, E\big(Var(R)\big) \\ &s.t \sum_{i=1}^{q} (C_i\, N_i) \leq Total\ Budget \\ &N_{test} = \left[N_1, N_2 \ldots N_q\right] \end{aligned} \tag{7}$$

In Equation 7, q refers to the number of different types of possible tests. The cost of the i^{th} $(i = 1\ to\ q)$ type of test is equal to C_i, and N_i (decision variable) denotes the number of repetitions of the i^{th} type of test. Let D_i denote all the data collected through

the i^{th} type of test. Let N_{test} denote the vector of all N_i's and let $\boldsymbol{D}$ denote the entire set of data collected from all q types of tests.

Alternatively, the resource allocation problem can be formulated by minimizing the cost required to decrease the variance of the system-level output below a threshold level, as:

$$\begin{aligned} &\underset{N_{test}}{\text{Minimize}} \sum_{i=1}^{q} (C_i N_i) \\ &s.t.\ E\big(Var(R)\big) \leq Threshold\ Variance \\ &N_{test} = \left[N_1, N_2 \dots N_q\right] \end{aligned} \tag{8}$$

In this chapter, the first formulation (Equation 7) is pursued for resource allocation because the threshold level for the variance is assumed to be unknown. Using $\boldsymbol{D}$, the model parameters are calibrated and the system-level response, $R(\boldsymbol{D})$, is computed. The optimization in Equation 7 calculates the optimal values of N_i, given the cost values C_i, such that the expected value of variance of the system-level prediction, $E\big(Var(R)\big)$, is minimized, while the budget constraint is satisfied.

This optimization formulation uses $E\big(Var(R)\big)$ as the objective function because R is a function of $\boldsymbol{D}$, which is not available before testing. Hence, random realizations of the test data set, $\boldsymbol{D}$ are generated; each random realization is used to compute $Var(R|\boldsymbol{D})$, and the expectation over such random realizations is calculated to be the objective function, as:

$$E\big(Var(R)\big) = \int Var(R|\boldsymbol{D})\, f(\boldsymbol{D}) d\boldsymbol{D} \tag{9}$$

where $f(\boldsymbol{D})$ is the density considered for the test data. Assuming that one type of test is performed independently of another (i.e. a subsystem-level test is independent of a material-level test), Equation 9 can be written as:

$$\begin{aligned} &E\big(Var(R)\big) \\ &= \int Var\big(R|D_1, D_2 \dots D_q\big) f(D_1) f(D_2) \dots f\big(D_q\big)\, dD_1 dD_2 \dots dD_q \end{aligned} \tag{10}$$

where $f(D_i)$ is the density considered for the data obtained through the i^{th} test. Before any testing is done, all prior knowledge regarding the model parameters, and the mathematical models constitute the only information available for the calculation of $f(D_i)$. Therefore, $f(D_i)$ is calculated as:

$$f(D_i) = \int f(y_i|\boldsymbol{\theta}_i) f'(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \tag{11}$$

where y_i represents the output of the mathematical model corresponding to the i^{th} type of test, $\boldsymbol{\theta}_i$ represents the underlying parameters, and $f'(\boldsymbol{\theta}_i)$ represents the prior knowledge regarding those parameters. Note that Equation 11 is simply an uncertainty propagation problem, where the other sources of uncertainty (such as physical variability in inputs, solution approximation errors, data uncertainty) can also be included in the computation of $f(y_i|\boldsymbol{\theta}_i)$.

Equations 9–11 are implemented using a numerical algorithm, where a finite number of realizations of **D** are generated and $E\big(Var(R)\big)$ is computed over these realizations. Then, $E\big(Var(R)\big)$ can be minimized using the optimization in Equation 7, and the ideal combination of tests can be identified.

Note that an inequality constraint (for the budget), and not an equality constraint, is considered in Equation 7. This means that the optimal solution which minimizes $E\big(Var(R)\big)$ need not necessarily exhaust the budget. Consider the simple case where there are two possible test types ($C_1 = 2$ and $C_2 = 3$), and the budget is equal to 6 cost units. There are two test combinations which exhaust the budget: (1) [$N_1 = 3$, $N_2 = 0$], and (2) [N_1=0, $N_2 = 2$]. Suppose that these two combinations lead to a value of $E\big(Var(R)\big)$ which is *greater* than that achieved through the test combination [$N_1 = 1$, $N_2 = 1$]. Then, obviously the combination [$N_1 = 1$, $N_2 = 1$] must be selected because it achieves the goal of reducing $E\big(Var(R)\big)$ even though it may not exhaust the budget.

3.4 Solution to the optimization problem

Equation 7 is a complicated integer optimization problem, where Bayesian updating and forward propagation need to be repeated for each random realization of the test data in order to evaluate the objective function, thus increasing the computational cost several fold. In spite of the use of Gaussian process surrogate models to replace the expensive system model, high computing power is still needed to solve the optimization problem.

Integer optimization is sometimes solved using an approximation method, where the integer constraint is first relaxed, and the integers nearest to the resulting optimal

solution are used in further solution of the original (un-relaxed) problem. Unfortunately, this approach is not applicable to the solution of Equation 7, since the objective function (system-level prediction variance) is defined and computed only for integer-valued decision variables (number of tests). It is meaningless to have a non-integer number of tests.

A multi-step procedure for solving the optimization problem is proposed in this chapter. Within each step, the global optimal solution is computed using an exhaustive search process, whereas across steps, a greedy algorithm is pursued. The step size is chosen in cost units, and additional steps are added until the budget constraint is satisfied. Let the size of the first step be equal to ϕ^1 cost units; the globally optimal testing combination for this cost (=ϕ^1) is denoted by N_{test}^1, and is calculated using exhaustive search, as:

$$\begin{aligned} &\underset{N_{test}^1}{\text{Minimize}}\, E\big(Var(R)\big) \\ &s.t. \sum_{i=1}^{q} (C_i N_i^1) \leq \phi^1 \\ &N_{test}^1 = \left[N_1^1, N_2^1 \dots N_q^1\right] \end{aligned} \tag{12}$$

The optimization procedure in the second stage is dependent on the optimal solution from the first stage, i.e. N_{test}^1. In general, the optimization for the j^{th} stage, given the solution in the previous stage (i.e. N_{test}^{j-1}), is performed for cost = ϕ^1. Note that $\sum_j \phi^j = Total\ Budget$. The j^{th} optimization is formulated as:

$$\underset{N_{test}^{j,new}}{\text{Minimize}}\, E\big(Var(R)\big)$$

$$s.t. \sum_{i=1}^{q} \big(C_i N_i^{j,new}\big) \leq \phi^1 \tag{13}$$

$$N_{test}^{j} = N_{test}^{j-1} + N_{test}^{j,new}$$

$$N_{test}^{j,new} = \left[N_1^{j,new}, N_2^{j,new} \dots N_q^{j,new}\right]$$

As seen in Equation 13, the decision variables for the j^{th} stage are $N_{test}^{j,new}$, i.e. those tests which need to be performed in the j^{th} stage; therefore the total number of tests is equal to the sum of $N_{test}^{j,new}$ and N_{test}^{j-1}, i.e. the optimal number of tests in the previous stage. The same procedure is repeated until no additional test can be performed with the budget constraint satisfied.

The selection of step size for a given budget is an important issue. The true global optimal solution can be calculated by considering *one step* whose size is equal to the entire budget. However, due to the large number of possible testing combinations, this approach may be computationally infeasible.

In a practical problem, several steps are considered, and the step sizes must be chosen judiciously based on (1) the costs of each type of test; (2) time required for each Bayesian update; (3) number of random realizations of data needed to compute $E\big(Var(R)\big)$; and (4) the test combinations that are suitable for the chosen step size; a very small step size may not even include an expensive type of test.

3.5 An illustrative example

This subsection presents a simple illustrative example, only to demonstrate the decrease of variance with testing. In order to focus on this objective, simple mathematical relationships are chosen (even the system-level response has no coupling), and measurement errors are assumed to be negligible. Other features such as coupled system response, measurements errors, solution approximation errors (while replacing the underlying physics-based model with a Gaussian process approximation), etc. are considered later in Sections 3.7 and 3.8.

The Bayesian network for this problem is exactly the same as that in Figure 1. There are four independent quantities and three dependent quantities; the numerical details of this problem are specified in Table 1. The notation $N(\mu, \sigma)$ is used to represent a normally distributed quantity with mean μ and standard deviation σ. Two types of tests (on two different lower levels) can be done and this information is used to update the uncertainty in the system-level response based on the tests.

Table 1: Numerical Details: Resource Allocation Illustration

Quantity	**Type**	**Description**
x_1 (input)	Independent	N(100,5)
x_2 (parameter)	Independent	N(50,10)
x_3 (input)	Independent	N(10,1)
x_4 (parameter)	Independent	N(15,4)
y_1	Dependent	Model: $y_1 = x_1 + x_2$
y_2	Dependent	Model: $y_2 = x_3 + x_4$
z	System-level response	Model: $z = y_1 - y_2$
Quantity to Measure	**Cost**	**No. of Tests**
y_1	10	N_1
y_2	5	N_2

Probability distributions are assumed to be available for the inputs x_1 and x_3; if this information was not available, and only sparse and/or interval data was available for the inputs, then the likelihood-based method developed in [3] can be used to construct a probability distributions for them. The variance of z before conducting any test (i.e. by propagating the above distributions of x_1, x_2, x_3, and x_4 through the models) is 142 units. The objective is calculate the number of tests on y_1 and y_2 (N_1 and N_2), that will lead to a minimum variance in z, subject to a total budget of \$50. Since there are only two parameters, global sensitivity analysis is not necessary, and hence, both x_2 and x_4 are chosen for calibration. The proposed optimization methodology is used for this purpose; five different stages are considered and the available budget in each stage is considered to be \$10. The results of test prioritization are given in Table 2 and Figure 2.

Table 2: Resource Allocation: Results

Cumulative Cost	$\boldsymbol{N_1}$	$\boldsymbol{N_2}$	$\boldsymbol{E\left(Var(z)\right)}$
\$10	**1**	**0**	**62.0**
	0	2	127.0
\$20	2	0	53.0
	1	**2**	**46.6**
\$30	**2**	**2**	**37.6**
	1	4	46.1
\$40	**3**	**2**	**34.0**
	2	4	37.6
\$50	**4**	**2**	**32.5**
	3	4	33.8

At the end of the optimization procedure, the optimal combination is found to be 4 tests on y_1 and 2 tests on y_2. Further, this solution was verified by considering all other combinations (exhaustive search) of N_1 and N_2 and computing the corresponding $E\left(Var(z)\right)$; for this illustrative example, this verification is numerically affordable.

However, for practical examples, a few random values of N_{test} = [N_1, N_2] (if not all) can be considered and it can be verified if the estimated solution is really optimal.

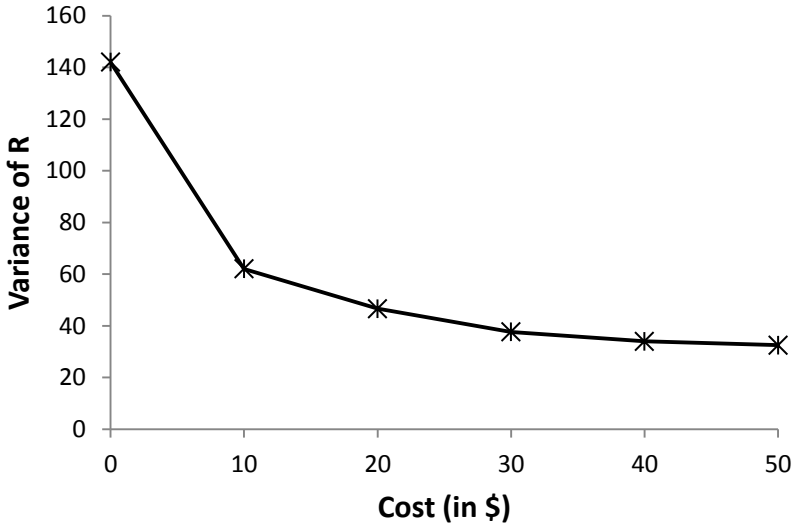

Figure 2: Variance vs. Cost: Resource Allocation Illustration

3.6 Summary of proposed methodology

The various steps of the proposed methodology are summarized below:

1. Construction of the Bayesian network: The first step is to identify the various component-level, subsystem-level, and system-level models. Each model has an output quantity and correspondingly, a test can be performed to measure this quantity. All the models are connected through the Bayesian network, and the data available across the nodes is also indicated. The model errors, if available, can also be included in the Bayesian network. Though solution approximation errors can be calculated prior to testing and included in the Bayes network, model form error cannot be calculated before testing. It must be noted that the Bayesian network, due to its acyclic nature, cannot account

for feedback coupling between models. When the system-level response is a coupled physics-based solution, the overall solution is directly included in the Bayesian network instead of considering the individual physics separately.

2. Sensitivity Analysis: The next step is to perform global sensitivity analysis and identify the "important" parameters that significantly contribute to the uncertainty in the system-level response. Then, those tests which can aid in the reduction of uncertainty in these "important" parameters are selected for consideration in the optimization for test resource allocation.

3. Bayesian updating: The third step is to perform Bayesian updating and calibrate parameters for a particular realization of measurement data. Then, this needs to be repeated by generating multiple realizations of measurement data in order to compute the expected value of variance, as in Equation 9. (Due to the required computational expense, the original physics models can be replaced with Gaussian process surrogates. Though this does not lead to analytical calculation of the posterior, it increases the computational efficiency several fold.) The most important aspect of Bayesian updating is the construction of the likelihood function, which is based on the difference between the model prediction and experimental observations, caused due to measurement and model errors. The solution approximation error due to the use of GP surrogate model is stochastic and is included. The true model form errors and measurement errors cannot be estimated *before* any actual testing is

performed. In this chapter, since the realizations of measurement data are generated based on the model itself, model form errors are not included. However, measurement errors are used in the construction of the likelihood function in the numerical examples in Sections 3.7 and 3.8.

4. Resource allocation optimization: The final step is to perform the resource allocation optimization using the multi-step procedure developed in Sections 3.2 and 3.3. It may be useful to verify that the resultant solution is actually optimal by computing $E(Var(R))$ for few other N_{test} values.

The following sections implement the proposed test resource allocation methodology to multidisciplinary and multi-level problems. Two different types of configurations are considered in order to emphasize the philosophical differences involved in model development and testing of such systems.

In a multi-disciplinary system, the overall system-level output is calculated using a multi-physics simulation. Each of the individual physics is tested. The models governing the individual physics may interact with each other through feedback coupling, i.e. the output of one is the input of the other and vice-versa. From a hierarchical point of view, both these models are at the same level of hierarchy whereas the system-level prediction is at a higher-level. For example, in fluid-structure analysis, the displacement (output of structural analysis) and pressure (output of fluid analysis) fields are inputs to the fluid analysis and structural analysis respectively. Due to the acyclic nature of the Bayesian network, it is not possible to compute the system-level response by explicitly

accounting for this feedback; instead, the coupled solution is directly computed and used. If there is only feed-forward coupling between two disciplinary analyses then each individual analysis can be separately included in the Bayesian network. The tests are always performed for individual physics without coupling. Section 3.7 discusses resource allocation for such a coupled multi-physics thermal-structural problem, representative of vibrations in solar arrays of telescopes and spacecraft booms. The tests performed for individual thermal and structural physics are used to calibrate underlying parameters, which are then used to compute the coupled system-level response.

On the other hand, in a multi-level system, the complexity of the model and underlying phenomenon increase along the hierarchy. The model used for system-level prediction is at the highest level of hierarchy and each subsequent model is at a lower hierarchy. There is a set of parameters common to the models at all levels. These parameters are calibrated using data at the lower levels (where the models and the physical phenomena are simpler relative to the system-level), and the calibrated quantities are used to predict the system-level response. For example, consider two types of tests: (1) axial test on a coupon; and (2) bending test on a beam; either/both of these tests may be used to estimate the modulus, and then predict the deflection in a thick plate, when all three (coupon, beam, and plate) are made of the same material.

Section 3.8 discusses resource allocation for a multi-physics multi-level problem, where both features (tests conducted for individual physics and tests of simpler components or conditions) are used for the calibration of parameters and prediction of the system-level response.

3.7 Multidisciplinary system

3.7.1 Description of problem

This coupled-physics thermal vibration example illustrates a laboratory experiment which can be used to study and simulate the behavior in solar arrays of telescopes and spacecraft booms [47]. The test comprises of a vertically mounted storable, tubular, extendable member (STEM) which is fixed at the top end and a heat source that applies heat to one side of the member. A STEM is a flat plate with residual stresses that when unrolled (much like a tape measure) curves to create a thin-walled circular tube. This allows long telescope or spacecraft booms to be retracted and easily stored. The tube and the mass are initially at rest and a constant heat flux is applied on one side along the length of the tube. The application of the heat flux causes an increase in the temperature on the incident surface while the unheated side remains at the initial temperature. The temperature gradient causes the beam to bend away from the lamp, due to the thermal moment. The displacement of the beam, in turn, changes the distribution of temperature along the length of the beam, leading to a change the temperature gradient and the thermal moment, which in turn affects the flexural behavior. Thus the combination of heat transfer and flexural mechanics leads to oscillations of the beam. The setup of this experiment is shown in Figure 3.

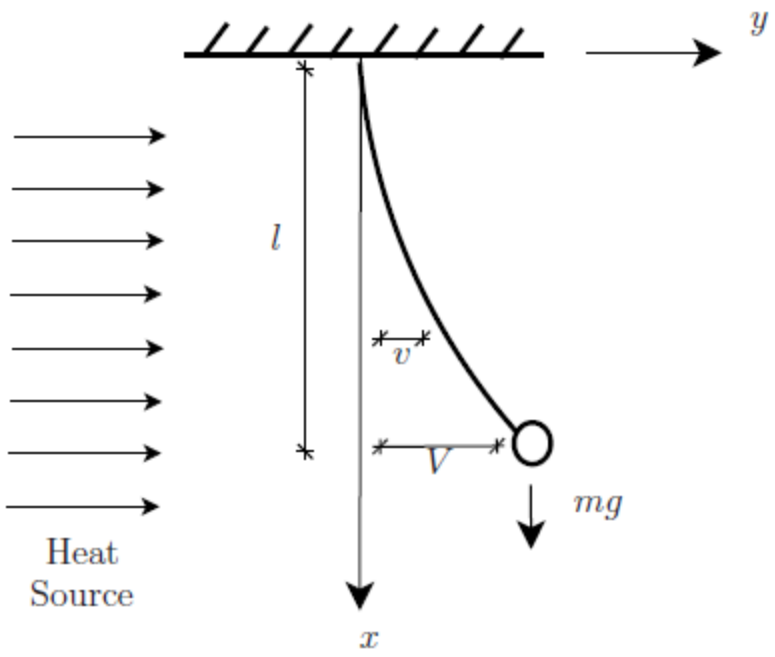

Figure 3: Thermally Induced Vibrations

The temperature at the tip mass (T_m) is given by the following differential equation;

$$\frac{\delta T_m}{\delta t} + \frac{T_m}{\tau} = \frac{T^*}{\tau}\left(1 - \frac{\upsilon(x,t)}{\beta^*}\right) \tag{14}$$

In Equation 14, $\upsilon(x,t)$ represents the displacement of the beam as a function of length and time. Thornton [47] explains how to calculate the parameters T^*, τ, and β^* as a function of the incident solar flux (S).

The displacement $v(x,t)$ can be related to the displacement of the tip mass $V(t)$ as:

$$v(x,t) = \left(\frac{3x^2}{2l^2} - \frac{x^3}{2l^3}\right) V(t) \tag{15}$$

The tip mass displacement $V(t)$, in turn, depends on the forcing function as follows:

$$\ddot{V} + 2\xi\omega_0\dot{V} + \left(\omega_0^2 + \frac{6g}{5l}\right) V = \frac{F(t)}{m} \tag{16}$$

In Equation 16, ξ is the damping ratio, and ω_0 is the angular frequency. The forcing function $F(t)$ depends on the thermal moment which in turn depends on the temperature, thereby causing coupling between the thermal equation and the structural equation. These relations are shown in the following equations:

$$F(t) = -\frac{3}{l^3}\int_0^l \int_0^x M(u,t)\,du\,dx \tag{17}$$

$$M(x,t) = \int E\alpha T_m(x,t)\cos(\Phi) y\, dA \tag{18}$$

In Equation 18, E in the elastic modulus, α is the coefficient of thermal expansion, Φ is the angle of incident flux on the cross section, y is the distance from the center of the cross section and the integral is over the area of the cross section A. Refer to Thornton [47] for a detailed description of this problem.

The objective of the testing campaign for this problem is to reduce the overall uncertainty about the response of the output variable (R), which is defined as the ratios of two displacement amplitudes at different time instances when the incident solar flux (S) is equal to 2000 W/m^2. If $R < 1$, the oscillations are not increasing and the system is stable. Conversely, if $R > 1$ the oscillations are increasing, commonly referred to as flutter, and the system is deemed to have failed. While a Gaussian process surrogate model is constructed to calculate R, individual physics predictions are performed using the above physics-based models.

Table 3: Calibration Quantities: Thermal Vibration Problem

Symbol	**Quantity**	**Property**	**Prior CoV**
E	Elastic modulus	Structural	0.1
c	Specific heat	Thermal	0.1
ξ	Damping	Structural	0.1
r	Radius	Geometric	0.03
e	Emissivity	Thermal	0.1

The calibration parameters need to be estimated during test data; four different types of tests are considered, as shown in Table 4. The total budget available for testing is assumed to be \$2000. It is assumed that the entire multi-disciplinary system cannot be tested.

Table 4: Types of Tests: Thermal Vibration Problem

Test type	**Physics**	**Calibrate**	**Input-Output**	**Cost**	**No. of tests**
Material-level	Thermal	c	Heat-Temperature rise	\$100	N_{m1}
Material-level	Structural	ξ	Amplitude decay	\$100	N_{m2}
Subsystem-level	Thermal	c, e, r	Heat-Temperature rise	\$500	N_T
Subsystem-level	Structural	ξ, E	Acceleration	\$500	N_F

The calibration quantities, the model predictions, and the test data are connected through a Bayesian network, as shown in Figure 4.

In the Bayesian network in Figure 4, "Temp" refers to temperature, "Accn" refers to the acceleration, "Disp" refers to the displacement, and "Amp" refers to the amplitude of vibration. Measurement errors (ϵ) are assumed too have a standard deviation that is equal to ten percent of the model prediction. This Bayesian network is used for uncertainty quantification, Bayesian updating, and resource allocation.

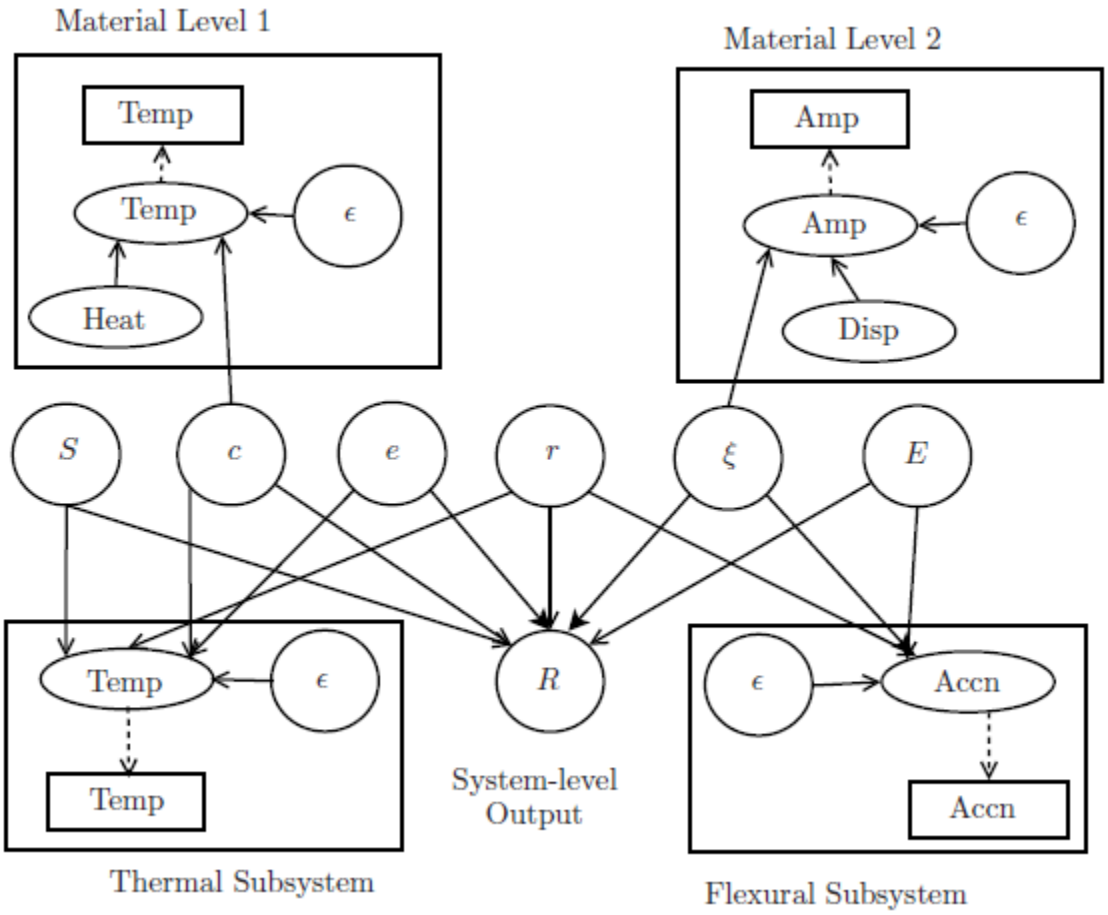

Figure 4: Bayesian Network: Thermal Vibration Problem

3.7.2 Resource allocation

The objective is to calculate the number of tests that lead to a maximum reduction in the variance of R. Let N_{test} denote the number of tests, where N_{test} = [N_{m1}, N_{m2}, N_F, N_F]; where N_{m1} is the number of material level temperature tests, N_{m2} is the number of material level pluck tests, N_F is the number of flexural subsystem tests, and N_T is the number of thermal subsystem tests. Let $\boldsymbol{D}$ = [D_{m1} , D_{m2} , D_F , D_T] denote the test measurements. The optimization problem for resource allocation can be formulated as shown in Equation 19.

$$\begin{aligned} &\underset{N_{test}}{\text{Minimize}}\, E\big(Var(R)\big) \\ &s.t.\ 100(N_{m1} + N_{m2}) + 500(N_F + N_T) \leq 2000 \\ &N_{test} = [N_{m1}, N_{m2}, N_F, N_T] \end{aligned} \tag{19}$$

The above optimization is solved using the multi-stage optimization procedure discussed in Sections 3.2 and 3.3. Four stages and a budget of $500 for each stage are considered, thereby accounting for the total budget of $2000. Each stage has 8 options (as opposed to two in the mathematical example in Section 3.5); only the optimal solution in each stage is shown.

Note that Table 5 expresses the expectation of variance of R in terms of percentage of the variance before any testing; this variance is equal to 5.69 x 10−7; since R is a ratio, this variance is dimensionless.

Table 5: Resource Allocation Results: Thermal Vibration Problem

Stage No.	N_{m1}	N_{m2}	N_F	N_T	$E(Var(R))$ **(in %)**
No tests	0	0	0	0	100.0
Stage 1: \$500	1	4	0	0	74.6
Stage 2: \$1000	1	4	1	0	51.4
Stage 3: \$1500	1	4	1	1	44.8
Stage 4: \$2000	1	9	1	1	44.2

For a \$2000 budget, it is seen that one temperature test, nine pluck tests, one thermal subsystem test, and one flexural subsystem test are required to achieve the maximum reduction in the variance of R. The results show that while it is useful to do all the tests, repeating the pluck test which calibrates structural damping is not only cheap but also leads to effective decrease in the variance of R. The decrease of variance with cost is shown in Figure 5.

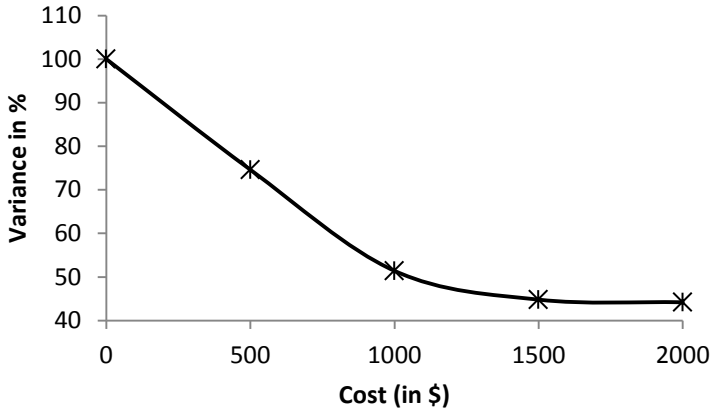

Figure 5: Cost vs. Variance: Thermal Vibration Problem

It is seen that the reduction in variance using the last \$1000 (i.e. from \$1000 to \$2000) was much smaller when compared to the reduction in variance using the initial \$1000. Such information is very useful for budgeting purposes, since all the above

computation (and practical resource allocation) is done before any test is actually conducted.

3.8 Multi-level, multi-disciplinary system

3.8.1 Description of problem

A simplified space telescope mirror problem is considered as an example of a multi-level, multidisciplinary system. As shown in Figure 6, it consists of three components - leg, mirror and plate, which are integrated to form the overall system, which can also be decomposed into various pieces as shown in Figure 6.

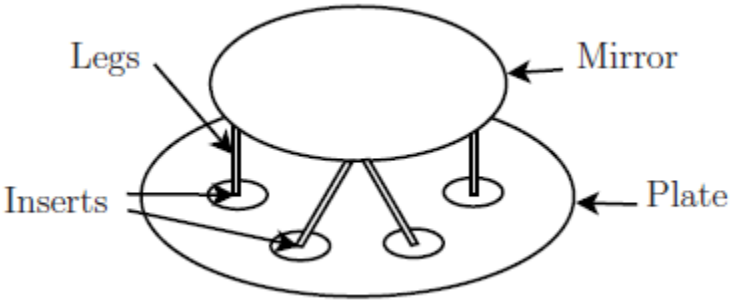

Figure 6: Simplified Space Telescope Mirror Problem

The system is tested under two types of physics – mechanical (due to gravity loading) and thermal (due to solar flux), that interact with each other, and affect the optical performance of the mirror. Eight different types of tests are considered, as tabulated in Table 6. It is assumed that the full system test under combined mechanical and thermal loading cannot be performed.

Table 6: Types of Tests: Telescope Mirror

Test No.	Component	Physics	Cost Units
1	Only leg	Mechanical	1
2	Only mirror	Mechanical	3
3	Only plate	Mechanical	1
4	Entire assembly	Mechanical	15
5	Only leg	Thermal	10
6	Only mirror	Thermal	30
7	Only Plate	Thermal	10
8	Entire Assembly	Thermal	150

The system was simulated using the Sierra multi-physics mechanics simulation suite developed at Sandia National Laboratories [48]. The thermal and the structural properties of the system affect the overall optical performance of the telescope mirror. The thermal and structural meshes were independent, with different and programmable mesh densities. In each case, as appropriate, the Sierra code Aria was used for thermodynamics, heat transfer, and radiation modeling, and the Sierra code Adagio was used for solid mechanics and quasi-static transient dynamics. For the purpose of this study, the optical system output was simply taken to be the deformation of the mirror at the center of the mirror. For this study, each Sierra simulation was wrapped within a DAKOTA [49] script to generate input-output data, which were later used to build Gaussian process surrogate models. Nine different surrogate models are constructed; eight of them to make predictions corresponding to the tests in Table 6, and the remaining model is for system-level prediction. The overall system output (denoted by R) is chosen to be the deflection of mirror relative to the center of the plate under both gravity and solar flux; this is equal to the sum of individual deflections under gravity sag and solar flux. Deflection is here a proxy for performance metrics such as wave-front error, focus drift, or other system-level characteristics that cannot be well-represented without

coupling the structural, thermal, and optical models. The test data and the Gaussian process models can be connected through a Bayesian network, as shown in Figure 7.

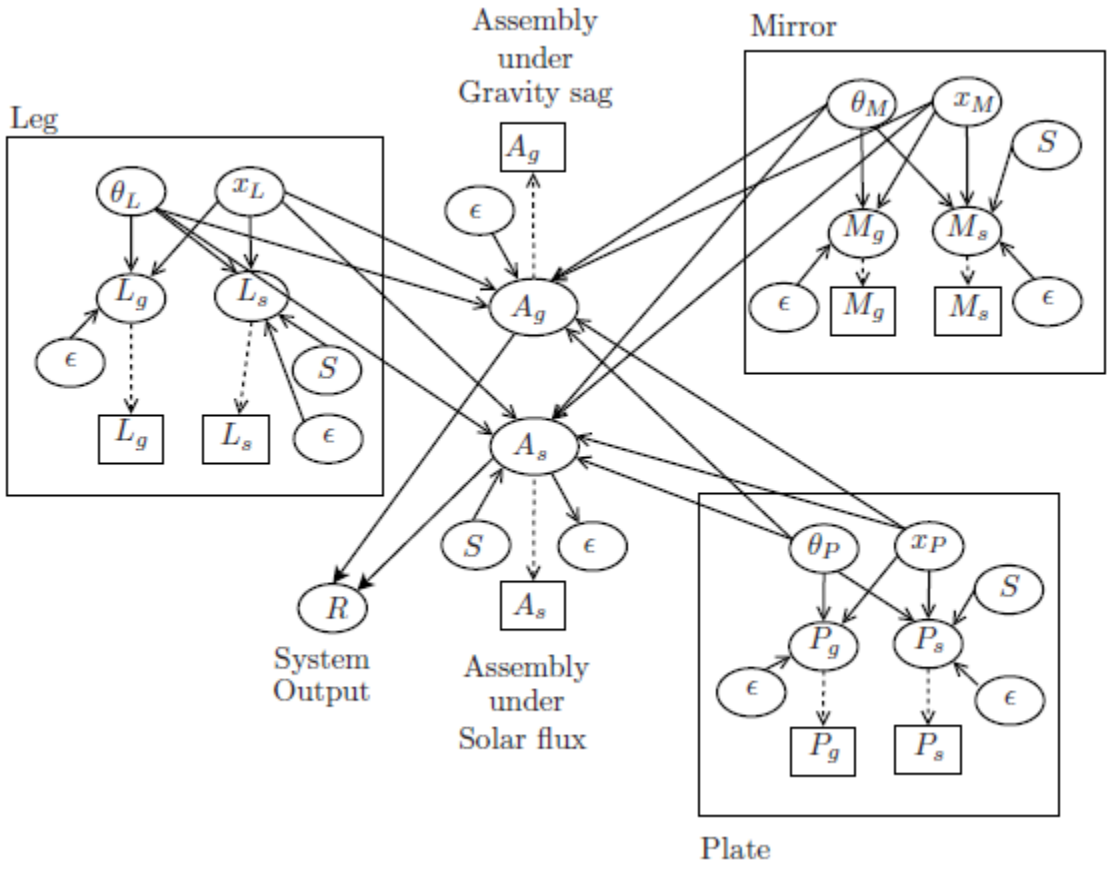

Figure 7: Bayesian Network: Telescope Mirror Problem

Consider the Bayesian network in Figure 7. The quantities θ_L, θ_M, θ_P and x_L, x_M, x_P refer to the model parameters and inputs of the leg, mirror, and plate components respectively; note that each of these quantities is a vector since a component may have more than one parameter/input. Each model parameter vector consists of the corresponding component's density, Young's modulus, Poisson's ratio, coefficient of thermal expansion, thermal conductivity, specific heat, and emissivity. Thus, this

example demonstrates the scalability of the proposed methodology by considering eight possible types of tests and twenty-one different parameters. The model predictions (for leg, mirror, plate, and assembly) are denoted by L_g, M_g, P_g, A_g (mechanical loading due to gravity sag) and L_s, M_s, P_s, A_s (thermal loading due to solar flux) respectively. Though the same symbol ϵ has been used to denote the difference between model prediction and observation throughout the Bayesian network, the statistics of ϵ is different for different tests, and equal to ten percent of the model prediction.

3.8.2 Sensitivity analysis

Sensitivity analysis is very important in this example because there are possible calibration parameters. The sensitivities of the system-level output R to all the parameters – previously mentioned twenty-one model parameters (θ_L, θ_M, θ_P) – are quantified using sensitivity analysis based on DAKOTA [49]. The "important" parameters based on the results of sensitivity analysis are tabulated in Table 7.

Table 7: Sensitivity Analysis for Coupled System: Telescope Mirror

Model Parameter	**Importance Measure [49]**	**Cumulative Importance [49]**
Mirror Young's Modulus	0.8021	0.8021
Leg Emissivity	0.0277	0.8298
Mirror Poisson's Ratio	0.0235	0.8533
Mirror Density	0.0112	0.8645

Further, the input solar flux has an importance measure of 0.12; however this is not a model parameter. From Table 7, it can be seen that four model parameters, along with the solar flux, account for more than 98% of the variance of the system output.

These four parameters, i.e. mirror elastic modulus, leg emissivity, mirror Poisson's ratio, and mirror density, are chosen to be calibrated through testing.

3.8.3 Test resource allocation

The goal of the resource allocation problem is to select tests that minimize the variance of the overall system output (R) under both gravity sag and solar flux. There are four types of tests that can be useful to calibrate the aforementioned "important" parameters; these tests are the gravity sag assembly level test (number of tests = N_{GSA} and each test costs 15 units), the solar flux assembly level test (number of tests = N_{SFA} and each costs 150 units), the solar flux leg component test (number of tests = N_{SFL} and each costs 10 units), and the gravity sag mirror component test (number of tests = N_{GSM} and each costs 3 units). In each test, the displacement of the mirror under the given loading is measured; correspondingly four different Gaussian process surrogate models are constructed to obtain model predictions. Also, a total budget of 150 cost units is assumed to be available. The optimization for test resource allocation is written as:

$$\begin{aligned} &\underset{N_{test}}{\text{Minimize}}\ E\big(Var(R)\big) \\ &s.t.\ 15N_{GSA} + 150N_{SFA} + 10N_{SFL} + 3N_{GSM} \leq 150 \\ &N_{test} = [N_{GSA}, N_{SFA}, N_{SFL}, N_{GSM}] \end{aligned} \tag{20}$$

The results of test resource allocation are given in Table 8. Similar to the previous sections, this is a multi-stage optimization. In each stage, 30 cost units are considered, and there are seven options to exhaust a budget of 30 cost units in each stage,

and the optimal solution in each stage is shown. Table 8 presents the variance in terms of percentage of the variance of R before testing (which was equal to 1.33 x 10^{-12} m^2). N_{test} is the vector of (N_{GSA}, N_{SFA}, N_{SFL}, N_{GSM}).

Table 8: Multi-Stage Optimization: Telescope Mirror

N_{test}	Cost Units	$E(Var(R))$ (in %)
(0,0,1,6)	28	12.3
(1,0,1,11)	58	8.7
(3,0,1,11)	88	7.4
(4,0,1,16)	118	6.6
(5,0,1,21)	148	6.1

The results of the test resource allocation optimization recommend 5 assembly–level tests under gravity sag, 1 leg component test under solar flux, and 21 mirror component tests under gravity sag in order to minimize the system output variance. The decrease in variance with cost, based on the optimal solution in each stage, is shown in Figure 8.

If the assembly–level tests (one under gravity sag and one under solar flux) were alone performed, then the variance decreases to 33.6% of the original value, at a cost of 165 units. Hence, it is evident that the proposed methodology achieves better performance (higher reduction in variance) at a lower cost.

It is seen from Figure 8 that there is little improvement in the system variance after testing worth 58 cost units. At that point, the results recommend performing 1 assembly–level test under gravity sag, 1 leg component test under solar flux and 11 mirror component tests under gravity sag. Hence, subsequent tests do not significantly aid in the reduction of variance in R. If the alternate optimization formulation (Equation 8) with a threshold variance lower than 5% of the prior variance had been chosen for

resource allocation, then it may have been impossible to satisfy the variance constraint. Therefore, the optimization formulation in Equation 7 may be preferred, since it provides an estimate of the variance with cost.

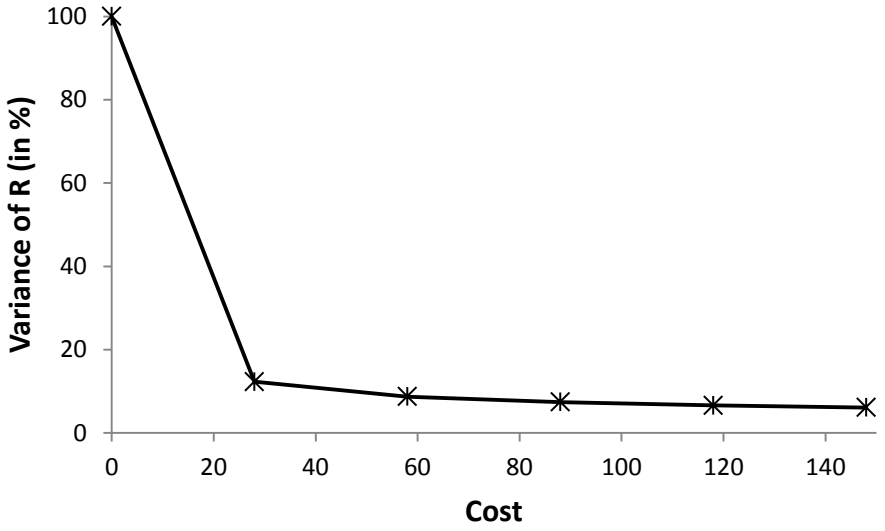

Figure 8: Variance vs. Cost: Telescope Mirror Problem

3.9 Conclusion

Testing at the component, subsystem and system levels is important in the context of uncertainty quantification in multi-level systems. When the systems are multi-disciplinary, it is important to conduct tests for both individual and combined physics. But rarely is it feasible to conduct every imaginable test, either due to schedule or cost considerations. This chapter developed an analytical procedure to aid in deciding which tests to conduct, especially for complex and expensive systems. A Bayesian network is used to connect multiple models and test data at different levels, and also include the various sources of error and uncertainty. The steps of the proposed methodology can be summarized as follows: (1) connect models and experiments at multiple levels efficiently through a Bayesian network; (2) systematically account for and include natural

variability, data uncertainty, and solution approximation errors; (3) predict the overall uncertainty in the system model prediction; and (4) optimize resource allocation for test selection and identify the most effective tests to reduce overall system model uncertainty.

A lower level test can easily isolate individual components and hence, the model parameters can be effectively updated, leading to a significant reduction in the variance of the system-level prediction. However, such a test would not account for interactions between higher level models and the corresponding parameters. In contrast, a higher level test would include the effects of interaction between multiple subsystem-level and component-level models. However, the calibration of parameters across multiple models may be difficult and may not lead to a significant reduction in the variance of the system-level prediction. The proposed test resource allocation procedure trades off between lower level tests and higher level tests by accounting not only for the resultant reduction in variance of the system-level prediction but also the testing costs.

Future work needs to address three major issues. The first deals with computational effort. As the number of calibration variables increases and the number of types of tests increases, the numerical difficulties involved in the numerical solution of the optimization problem increase; efficient numerical methods need to be developed for this purpose. The second deals with test design; having identified the number of tests, the next step would be to design them in order to maximize the uncertainty reduction in the system level prediction. This is addressed in the next chapter. Finally, though the numerical examples presented in this chapter considered different features (multiple levels of complexity and coupled physics interactions) representative of practical

applications, it is necessary to further investigate the extension of the proposed methodology for realistic engineering problems.

CHAPTER IV

DESIGN OF MODEL CALIBRATION TESTS

4.1 Introduction

Testing is a complicated and essential part of product development in any engineering system. Highly complex engineering systems, such as those used in the aerospace industry, often must be designed to perform under extreme conditions in extraordinary locations. Often, the testing of such systems at their actual usage conditions is dangerous, expensive, and impractical. In such situations, it is necessary to predict the performance of these engineering systems at their actual usage conditions by performing tests under controlled laboratory settings and, through computer modeling, use the information gained to reduce the uncertainty regarding full system performance. Existing practices determined the type of tests to be performed and the manner in which a test is to be performed based on legacy and prior experience. However, inefficiencies in these methods lead to wasteful spending and large performance uncertainties. The optimal design of tests at the material, component, and subsystem levels will more efficiently quantify the margins and risks associated with the performance of a complicated engineering system.

In Chapter III, the optimization objective sought to minimize the variance of the model output. This approach is valid if the calibration parameters can be thought of as single values with some degree of uncertainty that the test campaign should seek to reduce. This would be the case if the component being tested was the same component to

be used in the full system and thus its material and geometric properties would have a single true value. This chapter considers a different case where the components being tested in system calibration tests are nominally identical but not the same component that will be used in the full system. In this case it is desirable to calibrate the true distribution of a model parameter for the family of components. A variance reduction method is not sufficient for this purpose because the objective may seek to unfairly reduce the uncertainty in a model parameter. Thus, this chapter employs an information-theoretic objective function (the Kullback-Leibler distance presented in Section 2.5) which seeks to gain the maximum information about the model output.

The goal in a testing campaign is to develop the best possible understanding of the performance of the system in question in the most economically efficient manner. Tests need to be performed and designed with adequate fidelity and resolution so that the results provide meaningful information that can be used to reduce the uncertainty regarding the full system performance. This relationship between lower level testing data and full system performance must be made through the use of computer model simulations. These models need to be rigorously verified against numerical benchmarks, and systematically validated by a hierarchy of component and subsystem tests. Research into quantification of margins and uncertainties (QMU) has the goal of enabling this overall capability [1]. This leads to two concerning questions: (1) what combination of tests to perform and (2) at what settings to perform each test? Question (1) was addressed in depth in Chapter III and will be expanded upon in Section 4.2. Answering question (2) will be the primary focus of this chapter.

The goal of the test design methodology is to select – in an adaptive manner – the optimal combination of system calibration tests for a given cost budget and the most effective design input settings at which to perform each test. The presented methodology's ability to adapt as each piece of testing data is collected provides a significant advantage to the overall testing process. The methodology takes into consideration all previously collected test data when designing the next test in the sequence. Tests and test input settings are chosen based on their corresponding expectation of information gain on the full system output as measured by the Kullback-Leibler distance. The methodology is demonstrated using a thermally induced vibration problem.

4.2 Test combination selection

Chapter III developed a Bayesian network approach to system test resource allocation that selects the combination of tests that best minimizes the expected variance of the overall system output. This chapter utilizes this tool to select the types of tests that will be designed. Ideally, every possible combination of tests and test settings would be compared and the combination whose resulting data is most likely to lead to the maximum gain of information about the full system level output would be selected. However, a large number of possible testing sequences exist for even a simple problem making this computationally prohibitive. Instead, the proposed methodology uses a complicated integer optimization formula to select a set of tests for a predetermined cost step (of size ϕ) whose resulting data ($\boldsymbol{D}$) is most expected to lead to a maximum gain of information about the full system output (R) over the entire possible range of test input

settings (α). The function $d_{KL}(A, B)$ refers to the Kullback-Leibler distance from distribution A to distribution B.

The optimization is formulated as follows:

$$\begin{gathered} \underset{N_{test}}{\text{Maximize}}\ E\left(d_{KL}\left(f'(R), f(R)\right)\right) \\ s.t. \Sigma(C_i N_i)\ \leq\ \phi(i = 1\ to\ q) \\ N_{test} = \left[N_1, N_2 \ldots\ N_q\right] \end{gathered} \tag{21}$$

In Equation 21, C_i is the cost of the i^{th} type of test, N_i is the number of repetitions of the i^{th} type of test, and q is the total number of types of tests available for selection. The expected value of the KL distance between the current and updated distribution of the full system output is calculated as:

$$E\left(d_{KL}\left(f'(R), f(R)\right)\right) = \int d_{KL}\left(f'(R), f(R|\boldsymbol{D})\right) f(\boldsymbol{D}) d\boldsymbol{D} \tag{22}$$

where,

$$f(D_i) = \iiint f(y_i|\mu_\theta \sigma_\theta, \alpha) f(D_i|y_i, \mu_\theta \sigma_\theta, \alpha) f'(\mu_\theta \sigma_\theta) f(\alpha) d\mu_\theta d\sigma_\theta d\alpha \tag{23}$$

$f(D_i)$ is the density considered for the data obtained through the i^{th} test. Before any testing is done, all prior knowledge regarding the model parameters, and the mathematical models constitute the only information available for the calculation of

$f(D_i)$. y_i is the model output corresponding to the i^{th} type of test. $\boldsymbol{D}$ refers to the complete set of data from all performed tests, $\boldsymbol{D} = [D_1, D_2 \ldots D_q]$. In Equation 22, $f'(R)$ refers to the PDF of the full system output at its current state (i.e. already updated with all previously collected testing data). $f'(R)$ is only equal to the original prior distribution if no tests have been performed. Similarly, in Equation 23, $f'(\mu_\theta \sigma_\theta)$ refers to the PDFs of the distribution parameters of the model calibration parameters at the current state. $\mu_\theta \sigma_\theta$ refers to the joint probability distribution of the distribution parameters of the model calibration parameters.

The expression $f(D_i|y_i, \mu_\theta \sigma_\theta, \alpha)$ in Equation 23 represents the test measurement error for the i^{th} test. The test measurement errors used in this thesis are assumed to be normal with zero mean and some standard deviation. Since the test measurement errors are symmetric, and the objective functions used for test resource allocation and design take an expectation over the range of possible test data, test errors are omitted from the calculation of D_i. Using this assumption, Equation 23 can be rewritten as:

$$f(\boldsymbol{D_i}) = \iiint f(y_i|\mu_\theta \sigma_\theta, \alpha) f'(\mu_\theta \sigma_\theta) f(\alpha) d\mu_\theta d\sigma_\theta d\alpha \qquad (24)$$

The step size, ϕ (in monetary units), is of critical importance to the success of the methodology. In a practical problem, each step size must be chosen judiciously based on (1) the cost of each type of test; (2) the time required to solve the optimization in Equation 21 (due to the potentially large number of test combinations); (3) the total budget allotted to testing. The solution to the optimization problem posed in Equation 21

is a set of tests that will be designed in a given step using the methodology proposed in Section 4.3.

4.3 Test input setting selection

Once a particular type of test has been selected to be performed, the testing input settings, α, must be designed. It is desirable that the resulting data from the test provides as much new information about the distribution of the model output (R) as possible. The new information gained by performing the test is measured using the KL distance metric discussed previously. Thus, the selection of best test input settings for a single test is achieved by the following optimization:

$$\underset{\alpha}{\text{Maximize}}\ E\left(d_{KL}\big(f'(R), f(R)\big)\right) \tag{25}$$

where,

$$E\left(d_{KL}\big(f'(R), f(R)\big)\right) = \int d_{KL}\big(f'(R), f(R|\boldsymbol{D})\big) f(\boldsymbol{D}) d\boldsymbol{D} \tag{26}$$

and,

$$f(D_i) = \iint f(y_i|\mu_\theta \sigma_\theta, \alpha) f'(\mu_\theta \sigma_\theta) d\mu_\theta d\sigma_\theta. \tag{27}$$

The optimization posed in Equation 25 requires performing a large number of Bayesian updates for every step of the optimization procedure. Because of this significant computational expense, two simplifications are implemented: (1) the value $E\left(d_{KL}\left(f'(R), f(R)\right)\right)$ at each α is estimated through a first order approximation and (2) a surrogate model relating α to this approximation of $E\left(d_{KL}\left(f'(R), f(R)\right)\right)$ is created. A first order approximation of $E\left(d_{KL}\left(f'(R), f(R)\right)\right)$ is achieved by calculating each $f(D_i)$ using the most likely value of the current joint distribution of the model calibration parameters, $f'(\mu_\theta \sigma_\theta)$, instead of integrating over the entire distribution of μ_θ and σ_θ. The most likely values of the model calibration parameters are denoted as μ_θ^* and σ_θ^*. With this procedure, Equation 12 would be simplified to the following form:

$$f(D_i) = f(y_i | \mu_\theta^* \sigma_\theta^*, \alpha) \tag{28}$$

where,

$$\mu_\theta^* \sigma_\theta^* \; is \; obtained \; by \; \underset{\mu_\theta \sigma_\theta}{\text{Maximize}} \; f'(\mu_\theta \sigma_\theta). \tag{29}$$

The development of a surrogate model to relate α to $E\left(d_{KL}\left(f'(R), f(R)\right)\right)$ varies in computational difficultly depending on the range, sensitivity, and dimensionality of α. In the example problem posed in Section 4.6, Gaussian process surrogate models are used to develop this relationship. Once the surrogate model is created it can be used to solve the optimization problem shown above.

The settings of a laboratory test generally must be restricted to a certain number of significant digits due to user and equipment restrictions. This simplifies the domain of a test input setting to a set of discrete values. This discrete optimization problem can be solved through a simple enumeration. However, if α has a very large range or high dimensionality, more advanced optimization techniques may be necessary. The methods shown in Sections 4.2 and 4.3 are used together in the following section to adaptively design the model calibration testing sequence.

4.4 Test design optimization process

Now that methods have been developed to select both the types of tests to be performed and the input settings at which to perform them, this section will develop a procedure to combine them and develop a test campaign for the calibration of the model parameters relevant to the full system. This section develops a double loop optimization procedure that selects combinations of component, material, subsystem, and full system tests in the outer loop and then designs the input settingss of each selected test in the inner loop.

In Figure 9, j represents the test combination selection step. For each test combination selection step, j, a set of tests to be designed is selected using the method described in Section 4.2. The term ϕ^j represents the size of the cost step for the j^{th} test combination step. The solution to the test combination optimization for the j^{th} step is denoted as N^j and is a vector of the number of repetitions of each test type to be performed ($N^j = \left[N_1^j, N_2^j \ldots N_q^j\right]$ where q is the total number of possible test types).

Once a set of tests, N^j, is developed, a single test (not a single test type, but one individual test) is selected from N^j and its input settings, α, are designed using the method shown in Section 4.3. The test is then performed and the resulting data is used to update the previous probability distribution of the model parameters. The updated distribution of the model parameters is then used as the prior distribution for designing the next test in N^j. This process continues until every test in N^j has been designed. The total number of tests designed in the j^{th} step is equal to the summation of the elements of N^j.

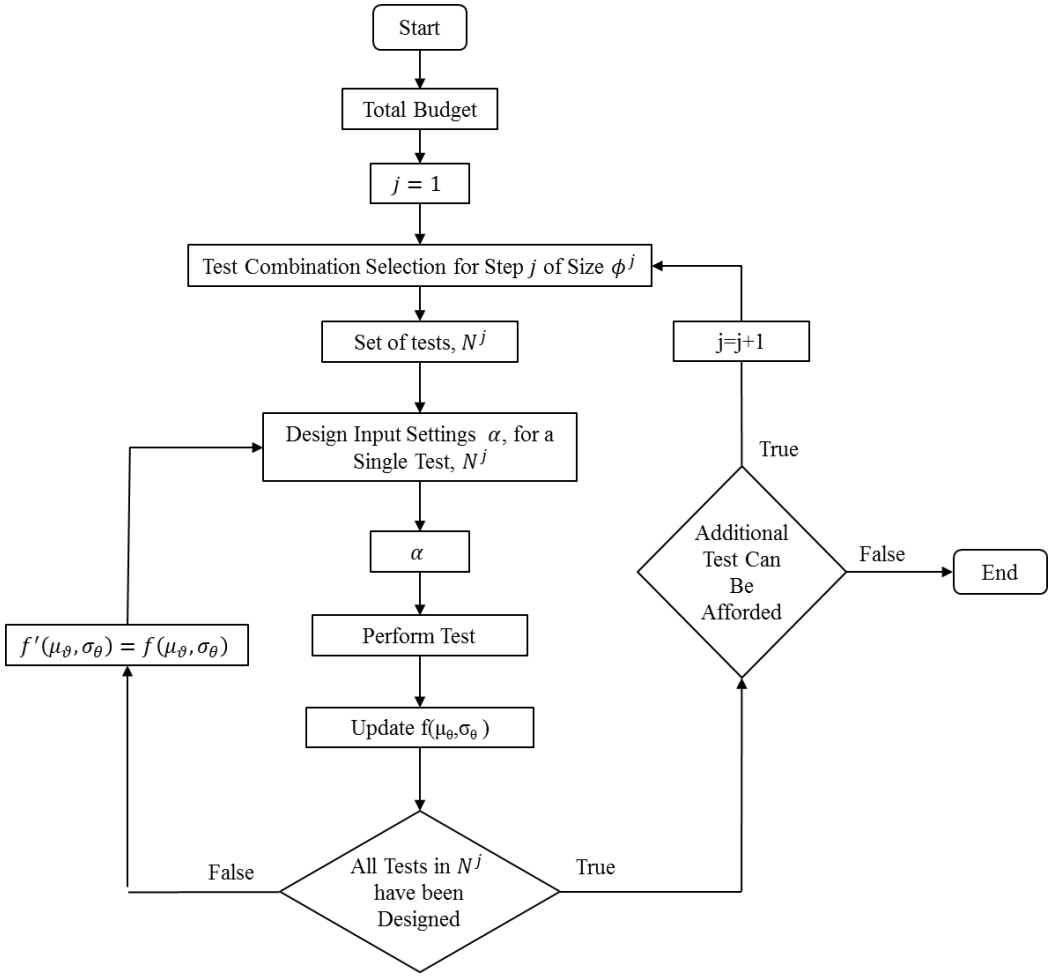

Figure 9: Test Design Flowchart

Figure 9 shows the adaptive test design methodology to calibrate relevant model parameters for a given total budget. "Total Budget" is given in monetary units and represents the upper bound on the cost of the test campaign. The "Total Budget" does not need to be met, but cannot be exceeded. The outer loop is repeated until no additional test can be performed without exceeding the total budget, at which point the test campaign is completed. The phrase "additional tests can be afforded" implies that the cost of the least expensive calibration test is less than the amount of unspent funds in the total budget. If this is true, an additional test selection step $(j + 1)$ can be run to select a new combination of tests to be designed. If this statement is false, no additional tests can be afforded and the optimization is complete. Also, this metric could be modified to include some sort of information gain threshold where if the information gain from the previous test selection step $(j - 1)$ to the current test selection step (j) were sufficiently small, it would be determined that spending additional funds on calibration tests is not economical and the optimization would be complete.

The result of this procedure is a series of calibration tests that is optimized, in an adaptive manner, to gain the most information about the full system response. In this way, the test campaign can most efficiently reduce the uncertainty of the overall system performance. This methodology is illustrated in a simple numerical example in Section 4.5.

4.5 An illustrative example

This section will provide a simple numerical example only for the purposes of illustrating the test design methodology. The system used for this demonstration has

exactly the same Bayesian network as shown in Figure 1 and the model parameters and system tests have the same properties as shown in Table 1. Unlike the illustrative example in Chapter III, measurement error will be included in this example. The assumed measurement errors are given in Table 9.

Table 9: Measurement Errors: Illustrative Example

Test	**Measurement Error**
y_1	N(0,5)
y_2	N(0,1.5)

The input settings for the calibration tests and their corresponding ranges of possible values are shown in Table 10. It will be assume that significant digit limitations only allow test settings to be specified as integer values.

Table 10:Test Input Settings: Illustrative Example

Setting	**Minimum**	**Maximum**
x_1	85	115
x_3	7	13

For the purposes of test design, a "reality" for the model parameters must generated in order to simulate testing data. Samples of the model parameters are taken from this "reality" and used with the mathematical models to produce samples of the full system output. The "reality" for this example problem will be x_2=60 and x_4=10. The test combination selection step size, ϕ, will be set at 10 cost units. After a set of tests for a given test combination selection step is chosen, all tests in the set are designed and simulated in the adaptive sequential method described in this chapter. The model parameter distributions updated with the data collected from the test simulations are then

used to select the next combination of tests to be designed. This iterative process is continued until the budget ($50) is exhausted or no additional tests can be afforded.

The results of the test combination selection and test design are given in Table 11 and Table 12 respectively.

Table 11: Test Combination Selection Results: Illustrative Example

Cumulative Cost	N_1	N_2	$E\left(d_{KL}(f'(z), f(z))\right)$
$10	**1**	**0**	**7.00**
	0	2	0.86
$20	**2**	**0**	**0.58**
	1	2	0.50
$30	3	0	0.14
	2	**2**	**0.63**
$40	**3**	**2**	**0.18**
	2	4	0.02
$50	**4**	**2**	**0.08**
	3	4	0.02

Table 12: Test Design Results: Illustrative Example

Step	**Test Type**	**Input Setting**	**Measured Output**
1	1	$x_1 = 90$	$y_1 = 149.1$
2	1	$x_1 = 98$	$y_1 = 157.2$
3	2	$x_3 = 7$	$y_2 = 14.8$
	2	$x_3 = 9$	$y_2 = 18.5$
4	3	$x_1 = 88$	$y_1 = 147.4$
5	3	$x_1 = 93$	$y_1 = 146.3$

The distance from the updated posterior distribution to the generated reality versus the cost of testing is shown Figure 10 and the PDFs of the prior, posterior, and "real" distributions are shown in Figure 11.

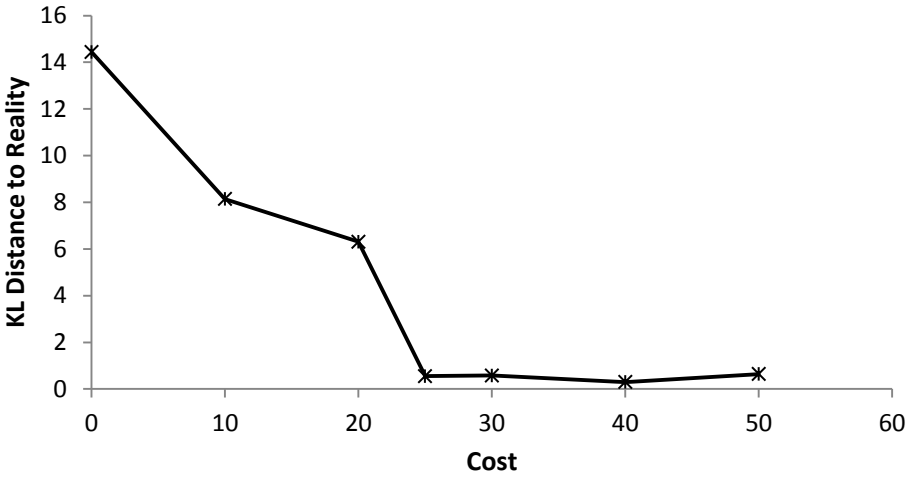

Figure 10: KL Distance to Reality versus Testing Cost: Illustrative Example

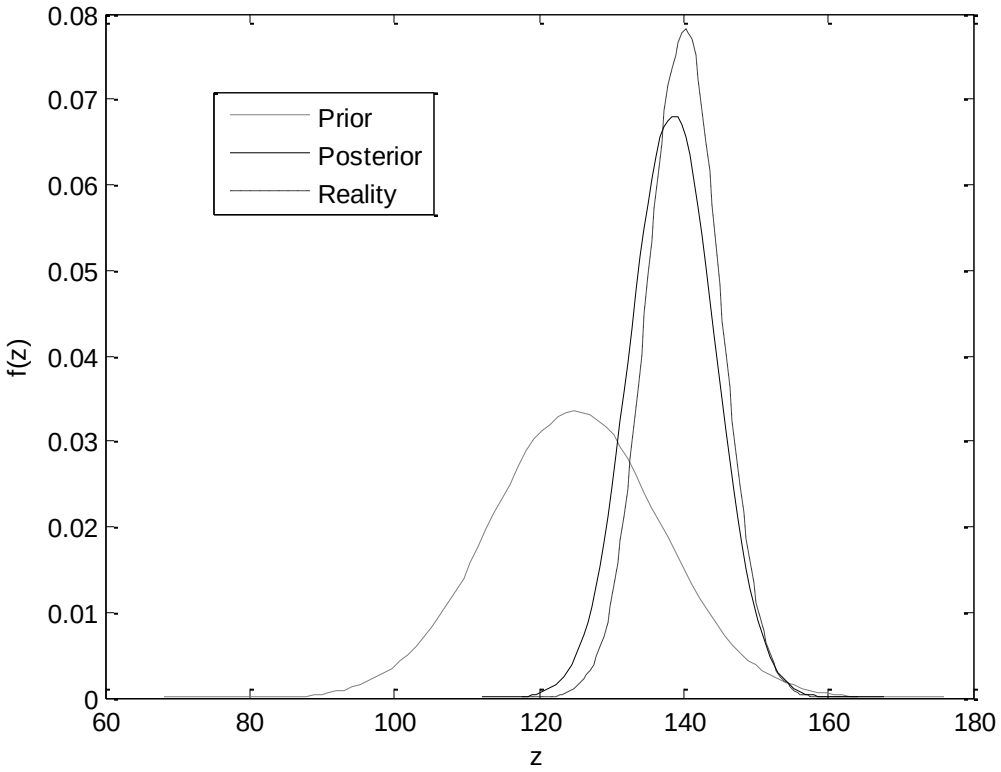

Figure 11: Test Design Results: Illustrative Example

The results show an adaptive optimal solution that performs 4 N_1 tests and 2 N_2 tests with the design parameters shown in Table 12. The results show the updated distributions converge quickly towards the true solution. It can be seen in Figure 10 that little additional information is obtained after a cost of \$25. Therefore, it may not be economical to continue performing calibration tests at that point.

4.6 Summary of proposed methodology

The proposed methodology seeks to select the types of calibration tests and the input values to each calibration test that best gains information about the performance of the full system output. This methodology can be summarized in 3 steps.

1. Construction of the Bayesian network: Much like the methodology developed in Chapter III, the first step is to identify the various component-level, subsystem-level, and system-level models. Each model has an output quantity and correspondingly, a test can be performed to measure this quantity. All the models are connected through the Bayesian network, and the data available across the nodes is also indicated. The model errors, if available, can also be included in the Bayesian network.

2. Test combination selection: The selection of the optimal combination of tests to be performed is done in a manner similar to that in Chapter III. For a given budget (or in this case step size) the prior distributions of the model parameters and the range of possible testing inputs are used to generate

random realizations of testing data for a given combination of tests. Each piece of data that is generated is then used to update the prior distributions and thus the distribution of the overall system output. The test combination that best gains information about the overall output (measured as expectation of the KL distance from the prior to the posterior) is selected as the optimal combination and its tests are performed.

3. Test input selection: For each test that is selected to be performed, the input settings that best gain information about the full system output should be used. First Gaussian process surrogate models are created that relate the input settings to the KL distance from the prior distribution to the posterior distribution developed using data generated from a given input setting. This is done by selecting training points from the domain of test input settings and the current most likely values of the model parameter's distribution parameters and using the test data generated to update the full system output distribution from the previous step. The KL distance from the previous distribution to the updated distribution is the output value of the surrogate model. (This is what is meant by saying the methodology is adaptive. The selection of input settings for the next test is based on the full system output distribution resulting from the previous step.) Once all the training points have been evaluated and the surrogate model has been constructed, a simple brute force enumeration (where every possible discrete input setting combination is evaluated using the surrogate model) is used to find the value of input settings

that maximizes the KL distance from the current distribution to the resulting prior distribution. This process continues until no further tests can be afforded.

4.7 Thermal vibration problem

4.7.1 Description of problem

The test design methodology will be demonstrated using a similar multi-physics thermally induced vibration problem as presented in Chapter III. Refer to Section 3.6 for details on the geometry and physics of the problem.

The objective of the testing campaign for this problem is to reduce the overall uncertainty about the response of the output variable (R), which is defined as the ratios of two displacement amplitudes at different time instances when the incident solar flux (S) is equal to 2000 W/m^2. If $R < 1$, the oscillations are not increasing and the system is stable. Conversely, if $R > 1$ the oscillations are increasing, commonly referred to as flutter, and the system is deemed to have failed.

Using a sensitivity analysis, five model parameters were selected for this example and their distribution parameters will be calibrated using a Bayesian updating procedure. All model parameters are assumed to have a normal distribution and their means and standard deviations will have normal prior distributions that will be updated as testing data is collected. The prior distributions of the calibration quantities are shown in Table 13. Note that Table 13 specifies a prior distribution for the mean and standard deviation of each model calibration parameter as opposed to Table 3 from the previous chapter

which only specifies single values for the mean and standard deviation of each model calibration parameter. This is a consequence of the chapter's modified approach in which we are calibrating the model parameters of a family of components as opposed to calibrating the model parameters of a single component as we did in Chapter III.

Table 13: Calibration Quantities: Thermal Vibration Problem

Sym.	Model Parameter	Property	Mean of Mean	CoV of Mean	Mean of Standard Deviation	CoV of Standard Deviation
E	Elastic Modulus	Structural	$193x10^{9}$	0.10	$193x10^{8}$	0.10
c	Specific Heat	Thermal	502.4	0.10	50.24	0.10
ζ	Damping Ratio	Structural	$1x10^{-3}$	0.10	$1x10^{-4}$	0.10
r	Radius	Geometric	$1.71x10^{-3}$	0.03	$5.1x10^{-5}$	0.10
e	Emissivity	Thermal	0.75	0.10	0.075	0.10

Since it is assumed that the entire multi-disciplinary thermal vibration problem cannot be tested, the presented model calibration quantities are to be updated using both material-level and subsystem-level tests. Four different types of tests are considered; the details of each test are shown in Table 14.

Table 14: Types of Tests: Thermal Vibration Problem

Test Level	Physics	Calibrates	Inputs	Outputs	Cost	No. of Tests
Material	Thermal	c	Heat	Temperature	\$100	N_{m1}
Material	Structural	ζ	Initial Disp., No. of Cycles	Displacement	\$100	N_{m2}
Subsystem	Thermal	E, ζ, r	Initial Disp., Time	Displacement	\$500	N_T
Subsystem	Structural	c, e	Heat Flux	Temperature	\$500	N_F

Each of the four types of tests has associated test input settings whose values will be selected from a given range of possible values in the test design methodology. The

test input settings relevant to the thermal vibration problem are detailed in Table 15. Let Ψ_{m1}, Ψ_{m2}, Ψ_T, and Ψ_F denote a single material level temperature test, material level pluck test, thermal subsystem test, and flexural subsystem test respectively.

Table 15: Test Input Settings: Thermal Vibration Problem

Input Setting	Symbol	Relevant Test	Minimum	Maximum
Heat Energy	Q	Ψ_{m1}	10	100
Initial Displacement	X_0	Ψ_{m2}	0.005	0.015
No. of Cycles	n	Ψ_{m2}	1	400
Measurement Time	t	Ψ_T	0.1	4
Initial Displacement	V_0	Ψ_F	0.005	0.015
Heat Flux	S	Ψ_F	100	2000

The calibration quantities, test input settings, model predictions, test data, and corresponding errors are connected through a Bayesian network as shown in Figure 4.

4.7.2 Test design

The objective of the test design methodology is to select the combination of tests and the test input settings that gain the most information about the output of the overall system. The total budget for the problem is set at $2500 for the entire testing campaign. Given this budget and the costs of the individual tests presented in Table 14, an intermediate step size of $500 makes good sense for the test combination selection portion of the problem because, at that dollar amount, only eight unique combinations (shown in Table 16) of the four possible system tests exist.

Table 16: Test Combinations for $500 Step Size: Thermal Vibration Problem

$\mathbf{N_{m1}}$	$\mathbf{N_{m2}}$	$\mathbf{N_T}$	$\mathbf{N_F}$
5	0	0	0
4	1	0	0
3	2	0	0
2	3	0	0
1	4	0	0
0	5	0	0
0	0	1	0
0	0	0	1

Using a step size of $500, the optimization presented in Equation 21 takes the following form:

$$\begin{gathered} \underset{N_{test}}{\text{Maximize}}\ E\left(d_{KL}\left(f'(R), f(R)\right)\right) \\ s.t. \$100(N_{m1} + N_{m2}) + \$500(N_T + N_F)\ \leq \$500 \\ N_{test} = [N_{m1}, N_{m2}, N_T, N_F] \end{gathered} \tag{30}$$

When the optimization problem presented in Equation 30 is solved for each step of $500, a set of tests to be performed is generated. Since the total budget is $2500 and the intermediate step sizes are $500, 5 test selection steps will be needed to exhaust the budget. Once the set of tests is selected, the input settingss of all the tests in the set are designed in the adaptive, sequential manner shown in Section 4.3. For the purposes of this example, actual laboratory tests were not performed, but instead test results are simulated.

Since actual laboratory tests were not performed for this research, for the purposes of the example problems, a system "reality" will be generated. This "reality" will be a set of distributions of the model parameters that are meant to represent the

actual state of the family of systems being tested in the example problem. The "reality" is used along the mathematical system models, the designed test input settings, and assumed measurement errors to simulate testing data. Results of two test campaign designs for two different "realities" (one where the "reality" is similar to the prior and one where it is significantly different) are shown in the following subsections.

4.7.2.1 Example 1: Reality similar to prior

In the first thermal vibration example problem, the simulated data is close to the prior distribution of the output variable, R. The results of the testing campaign design are shown in Table 17 and the prior, posterior, and "real" distributions are shown in Figure 12. Note that the distributions in Figure 12 are obtained by propagating parameter distributions through the system model; thus, the "reality" distribution is not obtained by directly measuring the system output.

The results show that the adapted optimum testing configuration for a $2500 budget and the observed results consists of 11 N_{m1} tests, 9 N_{m2} tests, 1 N_T test, and 0 N_F tests with the input values shown in Table 17. The probability distribution functions shown in Figure 12 show that the posterior distribution obtained by updating the prior distribution with the simulated data visually appears to be moving in the direction of the simulated "reality". However, it can be seen in Figure 13, which shows the KL distance from the current distribution to the simulated "reality" as a function of the cumulative cost of the tests performed, that the posterior distribution's distance to the "reality" increases and then decreases. This outcome is a result of the general similarity of the

prior distribution to the "reality" and it would be expected that the acquisition of more data would lead to a better result.

Table 17: Test Design Results: Thermal Vibration: Example 1

Test Type	**Test Input**	**Test Input**	**Test Output**
Step 1			
Ψ_{m2}	X_0=0.0150	n=162	X_n= 0.008018
Ψ_{m1}	Q=99.3		$\Delta\, Temp$ = 1.4350
Ψ_{m2}	X_0=0.0144	n=184	X_n= 0.005594
Ψ_{m2}	X_0=0.0150	n=183	X_n= 0.006521
Ψ_{m2}	X_0=0.0145	n=264	X_n= 0.004251
Step 2			
Ψ_{m1}	Q =50.9		$\Delta\, Temp$ = 0.5909
Ψ_{m1}	Q =62.9		$\Delta\, Temp$ = 0.7423
Ψ_{m1}	Q =100.0		$\Delta\, Temp$ = 1.1836
Ψ_{m1}	Q =33.9		$\Delta\, Temp$ = 0.4570
Ψ_{m1}	Q =85.0		$\Delta\, Temp$ = 1.0165
Step 3			
Ψ_{m2}	X_0=0.0150	n=194	X_n= 0.004380
Ψ_{m2}	X_0=0.0150	n=213	X_n= 0.005903
Ψ_{m2}	X_0=0.0133	n=330	X_n= 0.002105
Ψ_{m2}	X_0=0.0150	n=200	X_n= 0.004798
Ψ_{m2}	X_0=0.0144	n=317	X_n= 0.002647
Step 4			
Ψ_{m1}	Q =88.4		$\Delta\, Temp$ = 1.0177
Ψ_{m1}	Q =76.1		$\Delta\, Temp$ = 0.8811
Ψ_{m1}	Q =91.5		$\Delta\, Temp$ = 0.9703
Ψ_{m1}	Q =40.0		$\Delta\, Temp$ = 0.4356
Ψ_{m1}	Q =75.9		$\Delta\, Temp$ = 0.8220
Step 5			
Ψ_T	V_0=1926	t = 4.0	$\Delta\, Temp$ = 1.5665

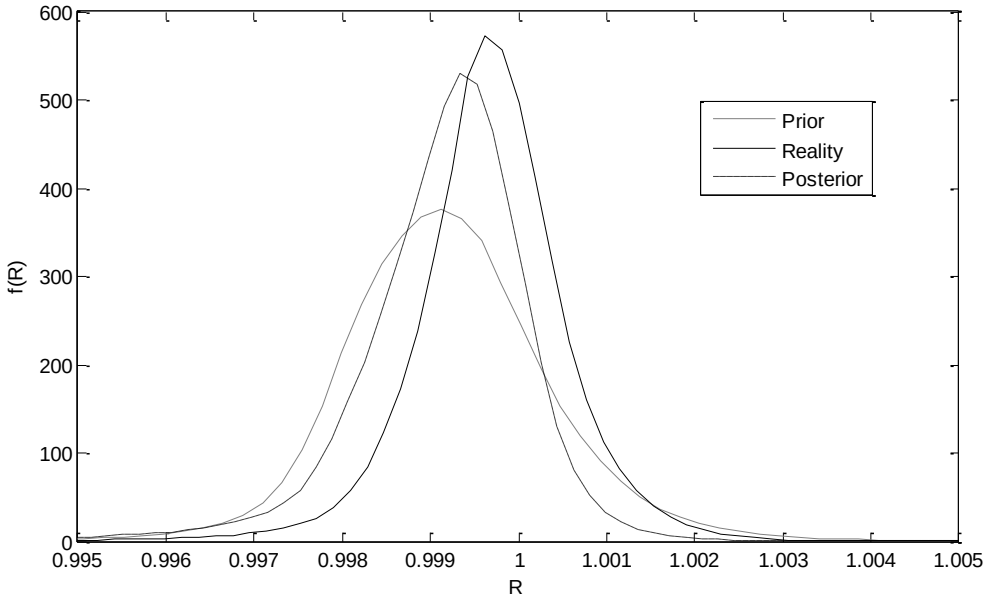

Figure 12: Test Design Results: Thermal Vibration Problem: Example 1

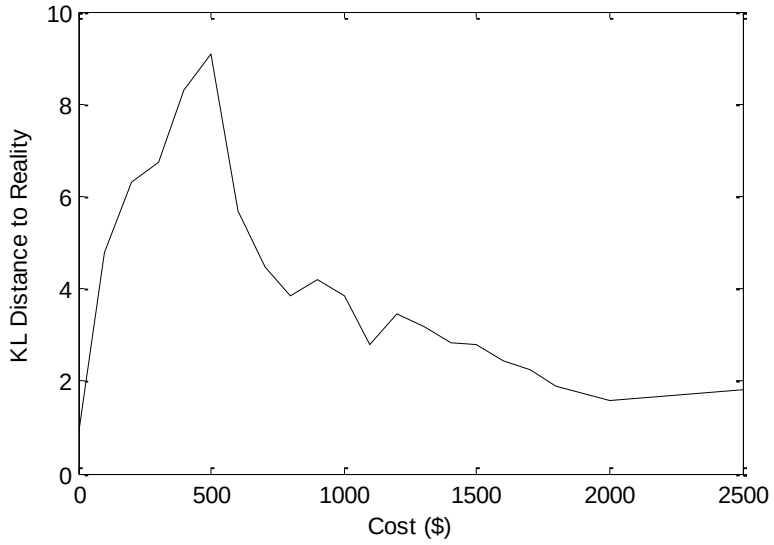

Figure 13: KL Distance from posterior to Reality versus Cost: Thermal Vibration Problem: Example 1

The oscillations in Figure 13 are due to the use of the KL distance as the objective function. Since the objective is to move the posterior distribution as far away from the

current distribution as possible, the distributions tend to jump from side to side as additional data pulls it towards its steady state.

4.7.2.2 Example 2: Reality dissimilar from prior

In this thermal vibration example problem, unlike the previous example, the generated "reality" is significantly different from the prior distribution of R. The results of the testing campaign design are shown in Table 18 and the prior, posterior, and "real" distributions are shown in Figure 14.

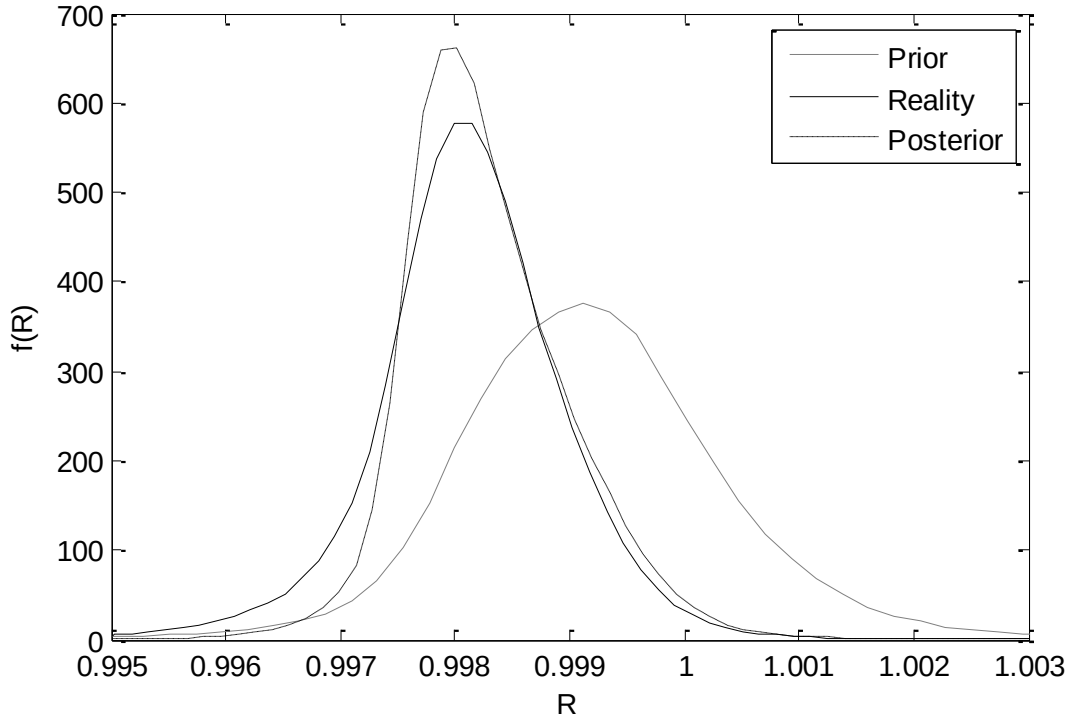

Figure 14: Test Design Results: Thermal Vibration Problem: Example 2

Table 18: Test Design Results: Thermal Vibration: Example 2

Test Type	Test Input	Test Input	Test Output
Step 1			
Ψ_{m2}	X_0=0.0150	n=162	X_n= 0.004067
Ψ_{m1}	Q=48.8		$\Delta\, Temp$ = 0.4855
Ψ_{m2}	X_0=0.0122	n=167	X_n= 0.004765
Ψ_{m2}	X_0=0.0150	n=144	X_n= 0.005184
Ψ_{m2}	X_0=0.0143	n=154	X_n= 0.005113
Step 2			
Ψ_{m2}	X_0=0.0150	n=151	X_n= 0.004152
Ψ_{m2}	X_0=0.0150	n=144	X_n= 0.005213
Ψ_{m2}	X_0=0.0150	n=184	X_n= 0.004796
Ψ_{m2}	X_0=0.0150	n=135	X_n= 0.006233
Ψ_{m1}	Q=72.6		$\Delta\, Temp$ = 0.8608
Step 3			
Ψ_{m1}	Q=63.3		$\Delta\, Temp$ = 0.6761
Ψ_{m1}	Q=81.6		$\Delta\, Temp$ = 0.9442
Ψ_{m1}	Q=100.0		$\Delta\, Temp$ = 1.2654
Ψ_{m1}	Q=58.8		$\Delta\, Temp$ = 0.6341
Ψ_{m1}	Q=99.3		$\Delta\, Temp$ = 0.5909
Step 4			
Ψ_{m2}	X_0=0.0150	n=138	X_n= 0.004281
Ψ_{m2}	X_0=0.0150	n=178	X_n= 0.004125
Ψ_{m2}	X_0=0.0150	n=127	X_n= 0.006175
Ψ_{m2}	X_0=0.0148	n=173	X_n= 0.005538
Ψ_{m2}	X_0=0.0133	n=142	X_n= 0.005177
Step 5			
Ψ_{m1}	Q=39.3		$\Delta\, Temp$ = 0.4546
Ψ_{m1}	Q=92.5		$\Delta\, Temp$ = 1.040
Ψ_{m1}	Q=100.0		$\Delta\, Temp$ = 1.1017
Ψ_{m1}	Q=76.1		$\Delta\, Temp$ = 0.8373
Ψ_{m1}	Q=100.0		$\Delta\, Temp$ = 1.4020

The results show that the adapted optimum testing configuration for a $2500 budget and the observed results consists of 16 N_{m1} tests, 9 N_{m2} tests, 0 N_T tests, and 0 N_F tests with the input values shown in Table 18. The results shown in Figure 14 validate the effectiveness of the methodology. Figure 15 shows the KL distance from the current distribution to the simulated "reality" as a function of the cumulative cost of the tests performed for both a designed test campaign and for a test campaign where the

types of tests performed were the same but the test input settings were not designed but chosen at random. Figure 15 shows a clear advantage to designing the input settings of the tests performed calibration test campaign.

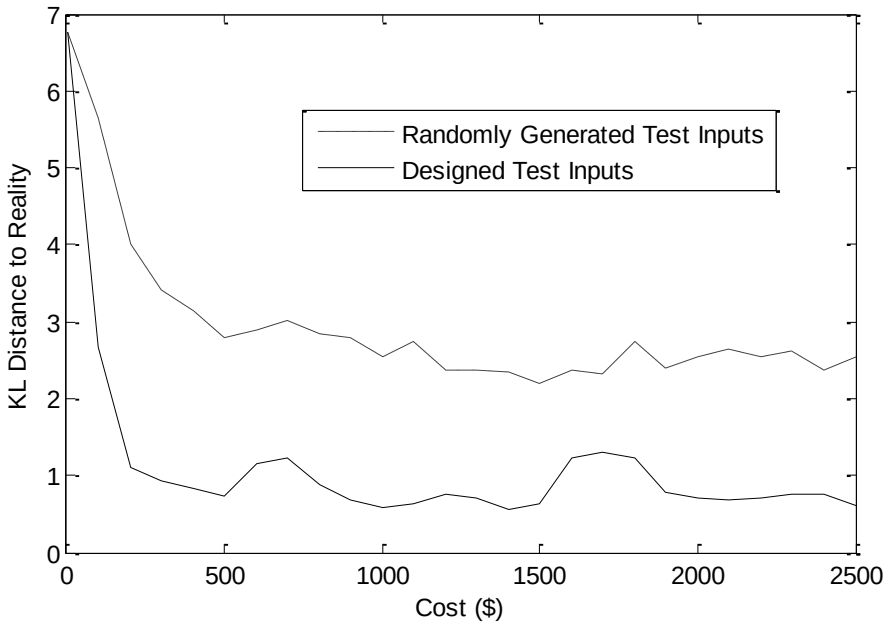

Figure 15: KL Distance from Posterior to Reality versus Cost: Thermal Vibration Problem: Example 2

4.8 Conclusion

The proper design of model calibration tests is a vital part of model development for complicated engineering systems. Relating test input settings to expectations of the resulting information gain allows for the development of a testing campaign that is economically efficient and technically productive. The use of a Bayesian network for the connection of model calibration parameters, test settings, test data, model errors, and testing errors, provides a rigorous framework for the propagation of information through multiple levels of complexity and across multiple physics to predict the performance of

the full system. The steps in the proposed methodology can be summarized as follows: (1) connect models and experiments at multiple levels efficiently through a Bayesian network; (2) systematically account for and include natural variability, data uncertainty, and solution approximation errors; (3) develop, in a sequential adaptive manner, a combination of test types and test input settings that best gains information about the output of the full system.

Future work may include the elimination of simplifying assumptions including making a first order approximation of the expectation of the KL distance during the test setting selection process and the use of a greedy search optimization. Also, optimizing the test type combinations and the test setting together instead of separately would lead to a more accurate solution.

The proposed optimization procedure provides an adaptive approach to the selection of test type combinations and the design of test input settings. By using the current, updated distributions of the calibration parameters, the selection of the test settings of the next test is dependent on the information gained by all the previously performed tests. This flexibility allows for an optimized solution to be achieved that has significant advantage over selecting test inputs by random or ad hoc methods.

CHAPTER V

DESIGN OF TESTS FOR CALIBRATION OF MANUFACTURING OPTIMIZATION MODELS

5.1 Introduction

Materials engineering has historically been a crucial discipline to the development of new engineering products and the advancement of technological fields [50]. In the modern era, new materials are critical to advancements in high-tech fields such as computing, electronics, defense, and space transportation. However, the process of developing new materials is slow in comparison to engineering design in other disciplines and often requires significant trial and error. The materials to be used in a complicated engineering system are typically selected from a set of available materials that have been tested in order to understand their material properties. The cost of this materials testing is often expensive and time consuming, causing a significant barrier between the material's development and the material's use. This reality removes materials engineering from the design process of advanced engineering systems forcing the materials to be treated as constraints. An emerging field, integrated computational materials engineering (ICME) [50], is attempting to rectify this discrepancy in development time in order to return materials engineering to the design process. The success of this movement would result in materials that are specifically tailored to the objectives of an engineering system and would likely lead to significant advances in the capabilities of modern engineering products.

A number of significant challenges exist in the implementation of ICME to realistic engineering problems. Challenges include but are not limited to, the development of models for a range of various manufacturing variables, the development of an implementation structure to integrate various levels of models and information (in this research Bayesian networks are used for this process, however the construction of a Bayesian network is not a simple task; often the structure of the network and the relationship between nodes is not fully known), handling of numerous and large databases, and the proper selection and design of calibration and validation experiments [50]. All of these challenges contain aspects that require the use of advanced uncertainty quantification techniques. The connection of uncertainties from the design and manufacturing of a material to the output of the system in which the material is to be used would allow for the reduction of full system performance uncertainties by improving and adjusting material manufacturing variables. This truly includes materials design in the system optimization process.

This chapter aims to develop a methodology for the inclusion of manufacturing design variables in the optimization procedure for test resource allocation and design. The methodology will build on the two methodologies developed in Chapter III and Chapter IV but will now include the calibration of the parameters of the manufacturing optimization models in addition to the parameters of the system model. The proposed methodology is demonstrated using a single component fatigue crack growth problem where the degree of shot peening that is used in the component's manufacturing is optimized.

5.2 Manufacturing process optimization

The methodology developed in this chapter will be similar to that developed in Chapter IV with one major difference; the objective function for the test design optimization will be some function of the manufacturing process optimization. The relationship from manufacturing variables to a system level output can be made in a multi-level process where first a model relates the manufacturing variables to the material properties of the system and then additional models relate the system material properties to the system output. This multi-level process will be denoted by the relationship $R = h(\theta_m)$, where R is the output of the system, θ_m is a vector of the manufacturing variables, and h is the function relationship between them. In a realistic system h will be a complicated function with a great deal of uncertainty. The goal of the test campaign will be to calibrate the parameters of the function h, which will be denoted as θ_h, in order to obtain a specified goal in the optimization of the manufacturing variables. Once h has been developed, the manufacturing process optimization must determine what characteristics of R are most desirable to the system's intended use. Note that the quantity R in the formulations below is generic, and can represent different outputs (system reliability, number of cycles to failure, stress, etc.). It can be either deterministic or stochastic.

From a reliability perspective, the value of θ_m which results in the smallest probability of failure for the system would be chosen as the solution to the optimization problem (denoted θ_m^*). If the output, R, is the reliability of the system, this can be shown as:

$$\underset{\theta_m}{\text{Maximize}} \; R \tag{31}$$

In the above equation, R is a deterministic quantity. However, if R is stochastic, an approach of reduced performance uncertainty could also be explored where the θ_m value that corresponds to the lowest variance of R is chosen as optimal. The corresponding optimization can be written as:

$$\underset{\theta_m}{\text{Minimize}} \; Var(R|\theta_m) \tag{32}$$

When R is stochastic, one could try to maximize or minimize a number of attributes of the distribution of R such as its mean, variance, or a percentile value, and may include some weighting function related to the cost of the chosen manufacturing process. (A common option is robust design, where the mean of R is maximized and variance of R is minimized). Section 5.6 considers an example problem with R as the component life, which is treated stochastic.

If function evaluations are expensive, surrogate models relating θ_m to R for a given value of θ_h are developed to ease computation. Once the optimization objective is selected, the parameters of the optimization models can be calibrated using tests at multiple system levels. Section 5.3 discusses the design of these tests in order to accomplish the overall goal of the calibration procedure.

5.3 Formulation of calibration test optimization

The goal of this chapter is to select the combination of tests and the input setting at which each test is to be performed in order to best calibrate the distribution parameters of the manufacturing optimization model parameters (e.g. $\mu_{\theta h}$ and $\sigma_{\theta h}$). The development of an objective function for the optimization of model calibration tests is not a simple task and must reflect the desired benefits from the optimization of the manufacturing variables. The development of this function is complicated by the fact that θ_m is not a random variable but a deterministic test setting. Because of this, the output of the full system output, R, is not a single distribution as it was in Chapter III and Chapter IV, but has a distribution corresponding to each value of θ_m. Two general approaches exist for handling this issue: (1) optimize some expectation of a characteristic of R over the entire range of θ_m and (2) only consider the answer to the manufacturing optimization problem, θ_m^*, when optimizing the calibration tests.

Option (1) is concerned with reducing uncertainty over the entire range of θ_m. While this approach is likely to converge to the global optimum of the manufacturing optimization, it will also needlessly reduce uncertainty in portions of the range of θ_m that will not be the solution to the manufacturing optimization. With this option, since the goal is uncertainty reduction, objective functions, such as K-L distance or variance minimization, should be used. The optimization formulations for K-L distance maximization and variance minimization are shown in Equations 33 and Equation 34 respectively. The function $d_{KL}(A, B)$ refers to the Kullback-Leibler distance from distribution A to distribution B.

$$\underset{N_{test},\alpha_{test}}{\text{Maximize}}\ E\left(d_{KL}\left(f'(R), f(R)\right)\right)$$
$$s.t\ \sum_{i=1}^{q}(C_i N_i) \leq Total\ Budget \tag{33}$$

$$\underset{N_{test},\alpha_{test}}{\text{Minimize}}\ E\left(Var(R)\right)$$
$$s.t\ \sum_{i=1}^{q}(C_i N_i) \leq Total\ Budget \tag{34}$$

where,

$$N_{test} = \left[N_1, N_2 \dots N_q\right]$$
$$\alpha_{test} = [\alpha_1, \alpha_2 \dots \alpha_p] \tag{35}$$

In Equation 35, q refers to the number of different types of possible tests, and p refers to the total number of tests performed. The cost of the i^{th} $(i = 1\ to\ q)$ type of test is equal to C_i, N_i (decision variable) denotes the number of repetitions of the i^{th} type of test, and α_{test} (decision variable) represents the vector of test input variables for all tests performed. Let D_i denote all the data collected through the i^{th} type of test. Let N_{test} denote the vector of all N_i's and let $\boldsymbol{D}$ denote the entire set of data, $\boldsymbol{D} = \left[D_1, D_2 \dots D_q\right]$, collected from all q types of tests.

The expected value of the KL distance between the current and updated distribution of the full system output is calculated as:

$$E\left(d_{KL}\left(f'(R), f(R)\right)\right) = \iint d_{KL}\left(f'(R|\theta_m), f(R|\boldsymbol{D}, \theta_m)\right) f(\boldsymbol{D}) d\boldsymbol{D} d\theta_m \quad (36)$$

where,

$$f(D_i) = \iint f(y_i|\mu_{\theta h}, \sigma_{\theta h}, \alpha) f'(\mu_{\theta h}, \sigma_{\theta h}) d\mu_{\theta h} d\sigma_{\theta h} \quad (37)$$

y_i represents the output of the mathematical model of the i^{th} type of test. The expectation of the variance can be calculated as:

$$E\big(Var(R)\big) = \iint Var(R|\boldsymbol{D}, \theta_m) f(\boldsymbol{D}) d\boldsymbol{D}\, d\theta_m \quad (38)$$

Option (2) involves an optimization objective function that is a function of the solution to the manufacturing variable optimization function θ_m^* instead of the entire range of θ_m. Here the problem is a nested optimization, where the solution to the manufacturing variable optimization problem is nested within the calibration test optimization since N_{test} and α_{test} are used to generate a data set, $\boldsymbol{D}$, which in turn help to calibrate the distribution parameters of θ_h which are used to solve the manufacturing optimization problem which produces θ_m^* and R. This option will focus on obtaining a localized solution more quickly than Option (1). Option (2) is conducive to having the objective function of the calibration test design optimization be the same as the objective

function for the manufacturing variables. For example, using the objective function given in Equation 31, the calibration test design optimization could be formulated as:

$$\underset{N_{test},\alpha_{test}}{\text{Maximize}}\ E\ (R)$$
$$s.t\ \ \sum_{i=1}^{q}(C_i\,N_i) \le Total\ Budget \tag{39}$$

where,

$$R = arg\max_{\theta_m}\Big(h\big(\theta_m, \theta_h(N_{test}, \alpha_{test})\big)\Big) \tag{40}$$

where,

$$f(R|\theta_m) = \int f(R|\boldsymbol{D}, \theta_m) f(\boldsymbol{D}) d\boldsymbol{D} \tag{41}$$

and,

$$f(D_i) = \iint f(y|\mu_{\theta h}, \sigma_{\theta h}, \alpha) f'(\theta_h) d\mu_{\theta h} d\sigma_{\theta h} \tag{42}$$

Once the optimization has been formulated, it can be solved using they multi-step adaptive optimization method shown in Section 5.4. Note that multiple sets of possible data, $\boldsymbol{D}$, are generated, leading to multiple posterior distributions of θ_h and therefore

multiple values of R using Equation 40. Equation 39 therefore maximizes the expectation of R.

5.4 Solution to the optimization problem

The multi-step adaptive test design procedure developed in Chapter IV can be used to solve any of the optimization problems posed in Section 5.3. In this section, we will briefly review the procedure and use it to show the solution to the optimization presented in Equation 39 and Equation 40.

5.4.1 Test combination selection

The proposed methodology uses a integer optimization formula to select a set of tests for a predetermined cost step (of size ϕ) whose resulting data ($\boldsymbol{D}$) is most expected to lead to the full system output (R) which best satisfies the objective function over the entire possible range of test input settings (α).

The optimization with the objective function from Equation 39 is formulated as follows:

$$\begin{aligned} &\underset{N_{test}}{\text{Maximize}} \;\; E(R) \\ &s.t. \;\; \Sigma(C_i N_i) \leq \phi \;\; (i = 1 \; to \; q) \\ &N_{test} = \left[N_1, N_2 \ldots N_q\right] \end{aligned} \tag{43}$$

where,

$$f(D_i) = \iiint f(y_i|\mu_{\theta h}, \sigma_{\theta h}, \alpha) f'(\mu_{\theta h}, \sigma_{\theta h}) f(\alpha) d\mu_{\theta h} d\sigma_{\theta h} d\alpha \quad (44)$$

The step size, ϕ (in monetary units), is of critical importance to the success of the methodology. In a practical problem, each step size must be chosen judiciously based on (1) the cost of each type of test; (2) the time required to solve the optimization in Equation 43; (3) the total budget allotted to testing. The optimization is solved numerically by generating realizations of $\boldsymbol{D}$ from the prior distribution of $\mu_{\theta h}$, $\theta_{\theta h}$, and the range of α for each possible test combination. The test combination that is expected to best satisfy the objective function is chosen. The solution to the optimization problem posed in Equation 43 is a set of tests that will be designed in a given step using the methodology proposed in Section 5.4.2.

5.4.2 Test input setting design

Once a particular type of test has been selected to be performed, the test input settings, α, to be used must be designed. The value of α should be selected in order to best satisfy the objective function. Using the objective function from Equation 39, the test input setting optimization is formulated as:

$$\underset{\alpha}{\text{Maximize}}\ E(R) \quad (45)$$

where,

$$f(D_i) = \iint f(y_i|\mu_{\theta h}, \sigma_{\theta h}, \alpha) f'(\mu_{\theta h}, \sigma_{\theta h}) d\mu_{\theta h} d\sigma_{\theta h} \quad (46)$$

The optimization posed in Equation 45 requires performing a large number of Bayesian updates for every step of the optimization procedure. Because of this significant computational expense, two simplifications are implemented: (1) the value of R at each α is estimated through a first order approximation and (2) a surrogate model relating α to this approximation of R is created. A first order approximation of R is achieved by fixing D_i using the most likely value of the current joint distribution of the parameters of the model calibration parameters, $f'(\mu_{\theta h}, \sigma_{\theta h})$, to create a distribution. Then using the most likely value of the resulting distribution of θ_h (which is normal with mean $\mu^*_{\theta h}$ and standard deviation $\sigma^*_{\theta h}$) is used with the tests mathematical models to develop D_i. With this procedure, Equation 46 is simplified to the following form:

$$D_i = y_i, \text{ where } y_i \ maximize \ f(y_i|\mu^*_{\theta h}, \sigma^*_{\theta h}, \alpha) \quad (47)$$

where,

$$\mu^*_{\theta h}, \sigma^*_{\theta h} \ maximize \ f'(\mu_{\theta h}, \sigma_{\theta h}). \quad (48)$$

The development of a surrogate model to relate α to an approximation of R varies in computational difficultly depending on the range, sensitivity, and dimensionality of α. In the example problem posed in Section 5.6, Gaussian process surrogate models are used to develop this relationship. Once the surrogate model is created it can be used to solve

the optimization problem shown above. Often (as discussed in Chapter IV), due to restraints in the significant digits of the test input settings, the optimization can be solved quickly through a simple enumeration. However, if α has a very large range or high dimensionality, more advanced optimization techniques may be necessary.

The methods shown in Section 5.4 are used in the manner shown by Figure 9 to solve the full test design optimization problem. A double-loop process is utilized to adaptively select and design each test in sequence. First a set of tests for a given step size is selected and then each one of the selected tests is designed sequentially. The process repeats itself until no more tests can be afforded or a target value of the objective is reached.

5.5 Summary of proposed methodology

The proposed methodology can be summarized in four key points.

1. Construction of the Bayesian network: The proper development of the Bayesian network for the manufacturing optimization problem is not trivial. The conditional relationships between manufacturing input parameters and the model parameters that control the full system output can be complicated, contain a great deal of uncertainty, and may be all together unknown. The models not only of the full system but also the optimization models must be rigorously calibrated and validated. This issue is considered one of the significant challenges to the implementation of ICME and should be the subject of significant future work.

2. Formulate manufacturing process optimization: Once the Bayesian network has been developed, the optimization for the manufacturing process should be developed in a manner that selects the manufacturing input parameters which correspond to the most favorable full system output distribution, R. Numerous objective function possibilities exists for this purpose, including selecting the distribution with the minimum variance or selecting the distribution that shows the maximum reliability.

3. Formulate the calibration test optimization: Separate from the manufacturing process optimization, the optimization of calibration tests may pursue one of two goals. (1) Optimize some expectation of a characteristic of R over the entire range of the manufacturing variables, θ_m. This approach is used to improve the quality of the optimization as it calibrates model parameters over the entire range of θ_m including portions that will likely not be the solution to the optimization problem. And (2) only consider the answer to the manufacturing optimization problem, θ_m^*, when optimizing the calibration tests. This approach is useful for selecting a localized answer that focuses on regions of θ_m that will affect the outcome of the optimization. Option (2) is solved using a nested optimization where the optimization of the manufacturing variables is nesting within the optimization of the calibration tests. Option (2) will be utilized in the example problem that follows in Section 5.6.

4. Solve the optimization problem: The solution to the optimization problem can be obtained using the multi-step adaptive methodology for test design proposed in Chapter IV. This process first selects a set of tests to perform given a step size (in monetary units) and then designs the tests selected. This process continues until no additional tests can be performed while satisfying the total budget constraint.

The result of the proposed methodology is an adaptively chosen set of designed tests which efficiently reduces the uncertainty in the parameters of the full system model and the parameters that relate manufacturing variables to the full system model in order to benefit the manufacturing optimization process. This methodology is demonstrated in Section 5.6 with a fatigue crack growth problem using a component that undergoes a shot peening treatment (a manufacturing process).

5.6 Numerical Example: Shot peening optimization

5.6.1 Description of problem

The example shown in this section is concerned with optimizing the depth of shot peening, SP, (a manufacturing process) in order to best extend the useful life of a helicopter rotor mast subjected to both torsion and bending loads. The objective of this example is to select the set of model calibration tests and corresponding test input settings that best updates the shot peening optimization such that the 95% lower confidence bound on the remaining useful life of the structure given an initial flaw size is maximized.

5.6.1.1 Shot peening

Peening is a manufacturing process designed to work-harden metals by firing small objects at areas of the component where stress concentrations are expected to be high. Shot peening [51] is a form of peening where small metal balls are fired at high speeds to create small indentations on the exterior of a metal component. Shot peening increases the yield strength and failure strength of a material which should retard failure due to crack growth. Shot peened indentations can be varied in size and depth based on the type of shot and the speed at which it is fired. Larger indentations have more positive effects on the material properties of the component; however, on small components (like the one considered in this problem) large shot peening depths, SP, can significantly reduce the effective area of the member, lowering its load capacity. Thus, an optimized indentation depth would most efficiently extend the life of a cracked component. In this example, SP values will be allowed to range from $0.00in$ to $0.02in$. This optimization will be performed in the remainder of the section.

5.6.1.2 Helicopter rotor mast

The component (shown in Figure 16) to be considered is a two radius hollow cylinder which is assumed to experience surface cracking in the fillet region (a common location for crack initiation).

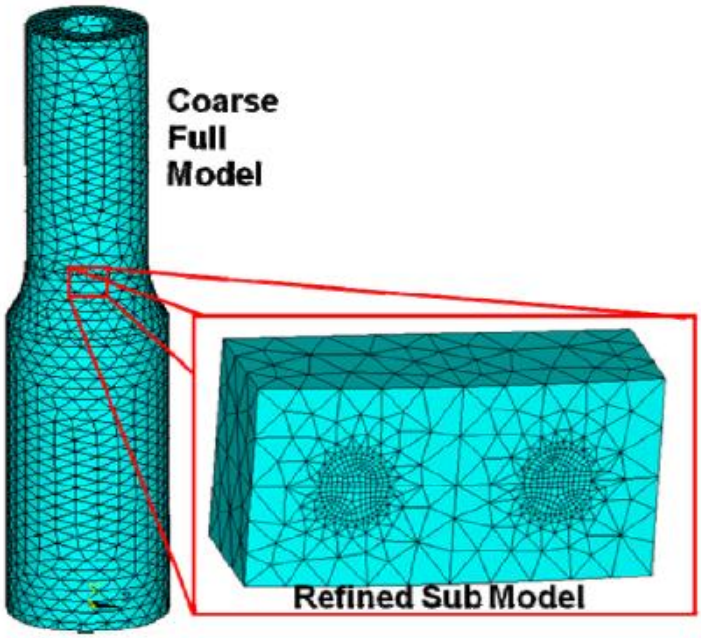

Figure 16: Helicopter Rotor Mast Model

The crack growth in this example uses linear elastic fracture mechanics (LEFM), and assumes a planar crack. The commercial finite element software ANSYS (version 11.0) is used to calculate the mode I, mode II, and mode III stress intensity factors. A sub-modeling technique is used to facilitate computational efficiency in finite element analysis, as shown in Figure 16. The crack growth problem is solved in a multi-level manner. First the entire structure is modeled with a coarse mesh and then the region surrounding the crack is modeled using a refined mesh. The boundary conditions of the sub-model are obtained from the solution of the full model [52].

Paris' law [53] is used for crack growth analysis. Paris' law gives the growth of a crack per cycle, da/dN, as:

$$\frac{da}{dN} = C(\Delta K)^n \tag{49}$$

In Equation 49, ΔK is the range of the stress intensity factor (SIF) which is computed by the finite element model, and C and n are fitting parameters of the crack growth model which are properties of the material. The fitting parameters, C and n, will be treated as random variables and updated as new crack damage inspection data becomes available. All cracks are assumed to be elliptical (characterized by a crack length and a crack depth). Multiple runs of FEA are conducted, accounting for different combinations of input variables - crack sizes along x-axis, aspect ratios, bending moments (torsion is assumed to be proportional to bending), and the SIFs at the crack tip are calculated. The crack lengths at any depth of shot peening were interpolated using a linear model fitted to calculated crack sizes at $SP = 0$ and $SP = 0.005$.

The rotor mast component is made with a 4340 Steel Alloy which has the assumed material properties shown in Table 19. The assumed geometric properties of the component are shown in Table 20. Some of the material and geometric properties are related to the shot peening depth SP.

Table 19: Assumed Material Properties: Shot Peening Problem

Steel Alloy 4340	
Modulus of Elasticity	29,700 ksi
Poisson Ration	0.29
Failure Stress (σ_f)	$e_0 + SP * e_1$
n	$a_0 + SP * a_1$
C	$b_0 + SP * b_1$

Table 20: Assumed Geometric Properties: Shot Peening Problem

Length	6 in
Moment Arm (l_a)	2.5 in
Inside Radius (r_i)	0.3 in
Outside Radius (narrow section) (r_e)	$0.6in - SP * d_0$
Outside Radius (wide section)	0.8 in

The problem has twelve calibration parameters, the mean and the standard deviations of: a_0, a_1, b_0, b_1, d_0, and e_1. Their prior distributions are shown in Table 21. σ_f is the failure stress of the material, l_a is the length of the moment arm or the distance from the crack to the location where the bending load is applied, r_i is the inner radius of the hollow cylindrical component, and r_e is the effective outer radius of the component at the crack location. In Table 19 the quantity e_0 is a constant distribution that will not be calibrated. Its distribution is $N\sim(108\,ksi, 10.8, ksi)$. The physical meaning for not calibrating e_0 is that it is assumed σ_f (which is equal to e_0 when $SP = 0$) has a well-known distribution for 4340 steel that is not shot peened.

Table 21: Prior Distributions of Calibration Parameters: Shot Peening Problem

Model Parameter	**Distribution Type**	**Mean**	**Standard Deviation**
a_0	Normal	$N\sim(3, 0.09\ )$	$N\sim(0.09, 0.009\ )$
a_1	Normal	$N\sim(20, 2\ )$	$N\sim(2, 0.2\ )$
b_0,	Normal	$N\sim(3.408 * 10^{-10},$ $3.408 * 10^{-11})$	$N\sim(3.408 * 10^{-11},$ $3.408 * 10^{-12}\ )$
b_1	Normal	$N\sim(-3.146 * 10^{-8},$ $3.146 * 10^{-9})$	$N\sim(3.146 * 10^{-9},$ $3.146 * 10^{-10})$
d_0	Normal	$N\sim(0, 5 * 10^{-4})$	$N\sim(5 * 10^{-4}, 5 * 10^{-5})$
e_1	Normal	$N\sim(3.5 * 10^{6}, 3.5 * 10^{5}\ )$	$N\sim(3.5 * 10^{5}, 3.5 * 10^{4}\ )$

The rotor mast component will be subjected to a bending, B, and torsion, T, load for a certain duration (measured in number of rotations). The average bending and

torsion loads experienced over the duration are random variables with the distributions given in Table 22.

Table 22: Loading Parameters: Shot Peening Problem

Load	**Distribution Type**	**Mean**	**Standard Deviation**
Bending	Normal	$N\sim(2.794\ kips,$ $0.1956\ kips\ )$	$0.1956kips - BSG * 0.0279$
Torsion	Normal	$N\sim(3.788kips,$ $0.3788kips\ )$	$0.3788kips - TSG * 0.0541$

The system failure criterion, g, is a function of crack size, component geometry, failure stress, and applied loading. Cracks are assumed to be semi-elliptical and planar with length l_c and depth d_c. The stresses due to bending, σ_B, and torsion, σ_T, are assumed to follow simple mechanical formulas given as:

$$\sigma_B = \frac{B l_a r_e}{I}$$
$$\sigma_T = \frac{T r_i r_e}{J} \tag{50}$$

where I and J respectively are the area moment of inertia and the torsional moment of inertia of the component's cross section. Not that it is always assumed that the direction of bending is oriented in such a manner that the crack is at a maximum distance away from the neutral axis. Assuming a semi-elliptical crack, the area of the crack, A_c, is as follows:

$$A_c = \frac{\pi l_c d_c}{2} \tag{51}$$

Using a conservative assumption that the crack is a void of area A_c a distance r_e away from the neutral axis, the denominators of Equation 50 can be estimated as:

$$\begin{aligned} I &= \pi(r_e^4 - r_i^4)0.25 - A_c r_e^2 \\ J &= \pi(r_e^4 - r_i^4)0.50 - A_c r_e^2 \end{aligned} \tag{52}$$

Using a safety factor of 1.667, the failure criterion can be calculated as follows:

$$g = \sigma_f - 1.667\sqrt{\sigma_B^2 + \sigma_T^2} \tag{53}$$

If $g < 0$ the system is assumed to have failed. The full system output variable, R, is calculated as the number of cycles required to reach failure, N_f, assuming an initial crack size of $l_c = 0.10\ in$ and $d_c = 0.05in$. The Bayesian network for this problem is shown in Figure 17.

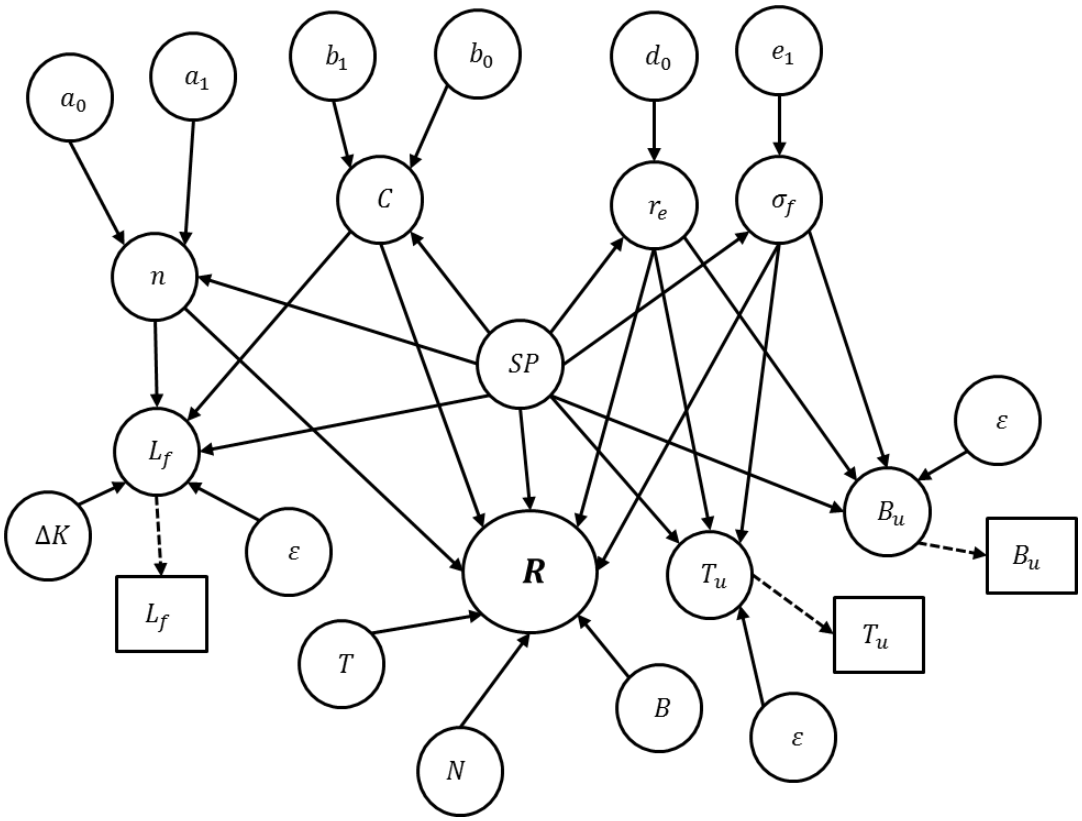

Figure 17: Bayesian Network: Shot Peening Problem

The calibration parameters will be updated using three component level tests: (1) a pure bending test, (2) a pure torsion test, and (3) an isolated crack growth test. It is assumed that the cost of all three tests is equal to 1 cost unit. The total budget for the test is set at 4 cost units, so four tests can be afforded. Each test is assumed to have a measurement error that is normal with a mean of zero and some specified standard deviation.

Pure Bending Test

The pure bending test will test an uncracked specimen with some level of shot peening to failure. The test input is the depth of shot peening, SP, used on the specimen

and the output is the bending load, B_u, that causes failure. The expression to calculate B_u is:

$$B_u = \sigma_f \frac{r^4{}_e - r^4{}_i}{r_e} \frac{\pi}{4l_a} \tag{54}$$

This test calibrates the parameters, d_0 and e_1, since σ_f is related to SP through e_1 (see Table 19) and r_e is related to SP through d_0 (see Table 20).

Pure Torsion Test

The pure torsion test will test an uncracked specimen with some level of shot peening to failure. The test input is the depth of shot peening, SP, used on the specimen and the output is the torsion load, T_u, that causes failure. The formula to calculate T_u is:

$$T_u = \sigma_f \frac{r^4{}_e - r^4{}_i}{r_e} \frac{\pi}{2r_i} \tag{55}$$

The test also calibrates the parameters of, d_0 and e_1.

Controlled Crack Growth Test

The controlled crack growth test will test a rotor mast specimen with an initial crack of length 0.1 inches some level of shot peening over 10,000 cycles. The test inputs are the depth of shot peening, SP, used on the specimen and the stress intensity factor, ΔK (which will be allowed to range from 20 to 30 $ksi\sqrt{in}$), and the output is the final length of the crack, L_f. The equation relating the two quantities is given as:

$$L_f = 0.1\ in + 10{,}000C(\Delta K)^n \tag{56}$$

The test calibrates the parameters, a_0, a_1, b_0, and b_1 since C and n are related to SP (see Table 19). It is assumed that ΔK does not change during the 10,000 cycles.

5.6.2 Manufacturing optimization formulation

A performance based objective function will be used for the manufacturing optimization problem. The objective function will be defined as the number of cycles, N_f, that corresponds to the point where the CDF is equal to 0.05 (denoted $N_{f,95}$). $N_{f,95}$ can be more formally defined as:

$$\int_{N_{f,95}}^{\infty} f(N_f|\mu_{\theta h}, \sigma_{\theta h}, SP)dN_f = 0.95 \tag{57}$$

Recall that $\mu_{\theta h}$ and $\sigma_{\theta h}$ refers to the means and standard distributions of the calibration variables of the test design problem. Thus, the manufacturing optimization equation can be written as:

$$\underset{SP}{\text{Maximize}} \quad N_{f,95} \tag{58}$$

Once the optimization is formulated, it can be integrated with the calibration test optimization to solve the test design problem. Let the solution to Equation 58 be denoted as θ_m^*.

5.6.3 Test design optimization formulation

The objective function for the calibration test design problem will only depend on the solution to the test design problem θ_m^* as described in Section 5.4. The objective function will seek to maximize the expectation of the same objective function as in the manufacturing optimization problem and is given as:

$$\begin{aligned} &\underset{N_{test},\alpha_{test}}{\text{Maximize}} \;\; E\left(N_{f,95}\right) \\ &s.t \sum_{i=1}^{q} (C_i N_i) \leq Total\ Budget \\ &N_{test} = \left[N_1, N_2 \dots N_q\right] \\ &\alpha_{test} = [\alpha_1, \alpha_2 \dots \alpha_p] \end{aligned} \tag{59}$$

where,

$$N_{f,95} = \underset{SP}{\text{argmax}} \left(h\left(SP, \theta_h(N_{test}, \alpha_{test})\right)\right) \tag{60}$$

where, q refers to the number of different types of possible tests, and p refers to the total number of tests performed and is equal to $N_1 + N_2 + \dots + N_q$. The cost of the i^{th} ($i = 1\ to\ q$) type of test is equal to C_i, N_i (decision variable) denotes the number of repetitions

of the i^{th} type of test, and α_{test} (decision variable) represents the vector of test input settings for all tests performed. Let D_i denote all the data collected through the i^{th} type of test. Let N_{test} denote the vector of all N_i's and let $\boldsymbol{D}$ denote the entire set of data, $\boldsymbol{D} = [D_1, D_2 \ldots D_q]$, collected from all q types of tests.

For this optimization $N_{f,95}$ is defined as:

$$\int_{N_{f,95}}^{\infty} f(N_f|\theta_m)dN_f = 0.95 \tag{61}$$

where,

$$f(N_f|\theta_m) = \int f(N_f|\boldsymbol{D}, \theta_m)f(\boldsymbol{D})d\boldsymbol{D} \tag{62}$$

and,

$$f(D_i) = \iint f(y_i|\mu_{\theta h}, \sigma_{\theta h}, \alpha)f'(\mu_{\theta h}, \sigma_{\theta h})d\mu_{\theta h}d\sigma_{\theta h} \tag{63}$$

where y_i is the output of the mathematical model of the i^{th} calibration test. This optimization problem is solved using the multi-step adaptive algorithm developed in Chapter IV and rehashed in Section 5.4.

5.6.4 Test design optimization solution

The proposed methodology uses a complicated optimization method to select a set of tests for a predetermined cost step (of size ϕ) whose resulting data ($\boldsymbol{D}$) is most expected to lead to the full system output (R) which best satisfies the objective function over the entire possible range of test input settings (α). The solution will be found using an adaptive multi-step procedure that first selects a set of tests to be performed in a given cost step, and then designs the tests is a sequential adaptive way. The resulting data set, $\boldsymbol{D}$, is then used as the starting point for the next step of test combination selections. This process continues until the no more tests can be performed under the total budget constraint.

The test combination selection equation for a single step with $\phi = 2\ cost\ units$ is formulated as follows:

$$\begin{gathered} \underset{N_{test}}{\text{Maximize}} \quad E\left(N_{f,95}\right) \\ 1 * N_B + 1 * N_T + 1 * N_{CR} \ \leq 2 \\ N_{test} = [N_B, N_T, N_{CR}] \end{gathered} \tag{64}$$

where N_B is the number of pure bending tests, N_T is the number of pure torsion tests, and N_{CR} is the number of controlled crack growth tests. This produces six possible test combinations to be compared which are shown in Table 23.

Table 23: Test Combinations: Shot Peening Problem

N_B	N_T	N_{CR}	Cost
2	0	0	2
0	2	0	2
0	0	2	2
1	1	0	2
1	0	1	2
0	1	1	2

.

The optimum combination of additional calibration tests for each of the two steps is shown in Table 24.

Table 24: Optimum Combination of Additional Calibration Tests: Shot Peening Problem

N_B	N_T	N_{CR}	$E(N_{f,95})$
Step 1			
1	0	1	14,327
Step2			
0	0	2	14,336

The result is a set of tests to be designed using the algorithm developed in Chapter IV and the optimization in Equation 65.

$$\underset{\alpha_{test}}{\text{Maximize}} \quad E(N_{f,95}) \tag{65}$$

The results of the test design optimization are shown in Table 25. The results of the corresponding manufacturing variable optimization are shown in Table 26 and Figure 18. Let Ψ_B, Ψ_T, and Ψ_{CR} denote a single bending, torsion, and crack growth test respectively.

Table 25: Test Design Results: Shot Peening Problem

Test	Input Parameter	Input Parameter	Test Output
Step 1			
Ψ_B	$SP = 0.019$		$B_u = 10{,}645$
Ψ_{CR}	$SP = 0.003$	$\Delta K = 20.26$	$L_F = 0.1287$
Step 2			
Ψ_{CR}	$SP = 0.003$	$\Delta K = 20.62$	$L_F = 0.1303$
Ψ_{CR}	$SP = 0.006$	$\Delta K = 29.70$	$L_F = 0.1180$

Table 26: Manufacturing Optimization Results: Shot Peening Problem

Testing Cost	SP^*	$N_{f,95}$
0	0.0151 in	16262
2	0.0145 in	17805
4	0.0146 in	18000

The two step optimization resulted in the selection and design of four tests. In Step 1 a bending test and a crack growth test are chosen and designed and in Step 2 two crack growth tests are chosen and designed. Figure 18 shows that the 5th percentile value ($N_{f,95}$) of the number of cycles to failure is increasing as more tests are performed. However, it is concerning that such large discrepancy exists between the $N_{f,95}$ values obtained through test design (shown in Table 26**)** and the expectations shown in Table 24. The low expectations of $N_{f,95}$ obtained in the test combination selection process indicate that the model output, R, is expected to decrease for a randomly chosen input setting. This discrepancy indicates that since the calibration parameters were assumed to be distributions, and not single points, that an information-theoretic objective function may have been more suitable.

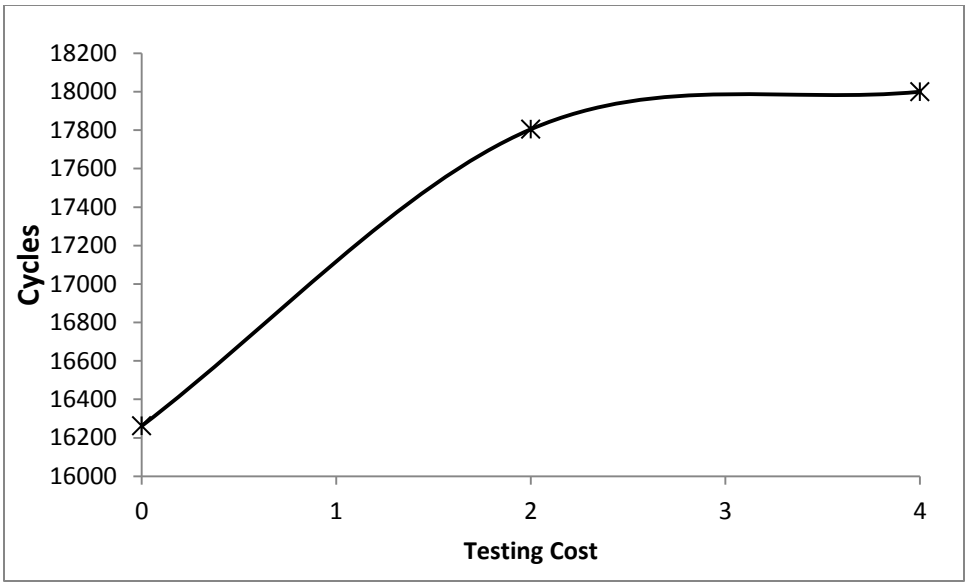

Figure 18: Manufacturing Optimization Results: Shot Peening Problem

5.7 Conclusions

The inclusion of the optimization of new materials in the design of complicated engineering systems has great promise to provide significant advancements for modern technology. The ability to tailor a material for a specific use allows engineers greater flexibility and capability in the design process and thus creates more efficient and powerful systems. The emerging field of ICME seeks to take on the enormous computational challenges involved in the achievement of this idea. Proper implementation of advanced uncertainty quantification techniques is critical to the success of ICME. This chapter proposes a methodology to calibrate the parameters of an optimization model that connects the input parameters of a manufacturing process to the output of the full system of interest. The reduction of uncertainty between the manufacturing variables and the full system output allows for a more robust and meaningful manufacturing optimization. Future work on this topic might include

improved objective functions using an information-theoretic approach and improved modeling to relate manufacturing variables to material parameters.

CHAPTER VI

OPTIMIZATION OF DAMAGE INSPECTION TYPE DECISIONS FOR OPERATIONAL RISK MANAGEMENT

6.1 Introduction

A digital twin [54] is a complex, computational model that mimics all aspects of a particular engineering system, including geometry, experienced loadings, and relevant physics in order to produce performance estimates and facilitate decisions about system usage. A properly developed digital twin should also incorporate the sources of uncertainty related to the current state of the system and future system performance. Information gained on the system including testing results, damage inspection data, load measurement data, and performance results can be used to update the digital twin and thus its reliability and performance predictions.

Manual non-destructive inspections (NDI) [55] are a routine and necessary part of the structural health monitoring process for an engineering system. Inspections on complex engineering systems can be expensive and time consuming, causing vital equipment to remain unusable for significant periods of time. The optimization of the inspection process such that the system is rendered fit or unfit for service as efficiently as possible is desirable to ensure the efficient use of scarce resources.

The inclusion of damage inspection data into the digital twin concept is a type of reliability based optimization [56], and will be the focus of this chapter. This chapter develops a methodology to optimize the scheduling of future system damage inspections

using the data and uncertainties contained within the digital twin and the distribution of possible system loadings. This methodology is demonstrated using a single component fatigue crack growth model.

6.2 Damage inspection

One outcome of a system damage inspection is the updating of the reliability estimate of the system provided by the system's digital twin under a given loading and operating condition. In order for the system to be used, the probability of system failure, p_f, must be less than some allowable probability of failure, p_a, with a specified degree of confidence. If the probability of failure is calculated using a sampling procedure, some degree of uncertainty is associated with the estimate. If Λ is the number of full system output samples taken, the error associated with the probability of failure estimate is [7].

$$COV(p_f) = \frac{\sqrt{\frac{(1-p_f)p_f}{\Lambda}}}{p_f} \tag{66}$$

Using Equation 66, an allowable number of failure samples, $\Lambda_{f,a}$, can be developed for a given number of full-system samples such that if the number of failures observed, which is denoted as Λ_f, is greater than the allowable number, the system is not cleared for use under the specified loading condition. For example, if the acceptable probability of failure is 0.001 with a required confidence of 95% , 100,000 samples of the system output were taken, and 100 were observed to have failed, using Equation 66 the

distribution of p_f would be normal with a mean of 0.001 and a standard deviation of 0.0001. This would mean that the confidence that p_f is less than p_a is only 50%. However, if 84 failures were observed, p_f would be normal with a mean of 0.00084 and a standard deviation of 0.0000916. This would mean that the confidence that p_f is less than p_a slightly over 95%. Therefore, if 84 or fewer system failures are observed in 100,000 samples the system is deemed to have an acceptably low probability of failure.

If the system is not cleared for use, additional inspection data may be collected to reduce the uncertainty of the system, the system may be operated under a less damaging loading condition, or the system may be removed from service.

Three outcomes are possible when a damage inspection is performed: (1) the damage is not detected, (2) the damage is detected but not measured, and (3) the damage is detected and measured. These outcomes introduce two sources of uncertainty to the digital twin updating process: (1) probability of detection, POD, and (2) damage measurement error. POD refers to the inspector's chance of discovering the damage to the system and is often given as some function of the damage severity. Damage measurement error refers to the inspector's inability to measure the precise state of the damage and is given as some function of the fidelity of the inspection being performed. POD and damage measurement errors are used to construct the likelihood functions that are used to update the system's digital twin. Note that damage not being detected is still a significant result of a damage inspection.

In addition to damage inspection, information can be provided to the digital twin by onboard load sensors (such as strain gauges) and damage detection devices. Load sensors reduce the uncertainty regarding the loads experienced by the structure which in

turn reduces the uncertainty regarding the damage accumulated by the system. Damage sensors help to detect flaws in the system, thus significantly increasing the POD. Naturally, both load sensors and damage detection devices have corresponding uncertainties about their own performance. The degree of these uncertainties tends to increase with the number of system uses.

For a given loading scenario, if the information provided to the digital twin, $f'(\boldsymbol{\theta})$, produces the condition where $\Lambda_f \leq \Lambda_{f,a}$, then no damage inspection is necessary and the system is cleared to be subjected to the loading scenario. However, if $\Lambda_f > \Lambda_{f,a}$, a decision must be made regarding which inspection or combination of inspections should be performed in order to best update the digital twin. Three outcomes are possible after an inspection is performed an the collected data is used to update $f'(\boldsymbol{\theta})$: (1) $\Lambda_f \leq \Lambda_{f,a}$ and the system is cleared to be used (this is the ideal scenario), (2) $\Lambda_f > \Lambda_{f,a}$ and additional or higher level inspections exist that, if performed, may clear the system for use, (3) $\Lambda_f > \Lambda_{f,a}$ and no additional or higher level inspections exist that may clear the system for use. While the system being proven unreliable is clearly undesirable, it is still better to come to this conclusion in the most economical way. For example, if a less expensive lower fidelity inspection is selected and its results are inconclusive, it is then necessary to perform a more expensive higher fidelity inspection. If the higher fidelity test was selected instead, it would not have been necessary to perform both inspections, which would have been more economical. On the other hand, if the lower fidelity test is sufficient to clear the system for use, its selection is more economical.

Thus, in order to select the inspection type decision combination that is expected to be the most economical, it is necessary to develop an optimization procedure that

compares the quotient of the expectation of system return and the expectation of inspection cost of each possible inspection type decision combination.

6.3 Optimization procedure

The methodology developed in this chapter will use a sampling-based optimization procedure to determine the most economical inspection type decision combination for the system in question. The methodology takes the current state of the system as understood by the system's digital twin, $f'(\boldsymbol{\theta})$, and updates it with possible inspection data, $\boldsymbol{D}$, which is generated based on $f'(\boldsymbol{\theta})$, for a given inspection type decision combination, $\boldsymbol{I}$. $\boldsymbol{I}$ is a discrete vector containing all of the of test type decisions to be made about a system in the period of interest. The optimization is formulated as:

$$\begin{aligned} &\underset{\boldsymbol{I}_{test}}{\text{Maximize}} \frac{E(V|\boldsymbol{I}_{test})}{E(C|\boldsymbol{I}_{test})} \\ &\boldsymbol{I}_{test} = \left[\boldsymbol{I}_1, \boldsymbol{I}_2 \dots \boldsymbol{I}_q\right] \end{aligned} \tag{67}$$

In Equation 67, V is the return from the system, which could be measured in any number of ways including number of missions performed, number of hours used, or number of cycles completed and C is the cost involved in performing a set of inspections. q is the number of inspection type decision points. An inspection type decision point is defined as a point in time during the system's life span where a decision regarding the type of inspection to be performed needs to be made. The objective function is calculated as:

$$\frac{E(V|\boldsymbol{I}_{test})}{E(C|\boldsymbol{I}_{test})} = \frac{\sum_{k=1}^{n}(V_k|\boldsymbol{D}_k)}{\sum_{k=1}^{n}(C_k|\boldsymbol{D}_k)} \tag{68}$$

In Equation 68, n represents the number of samples of inspection data to be taken of each inspection type decision combination, $\boldsymbol{I}$. V_k represents the system return related to the k^{th} sample, C_k represents the total inspection cost related to the k^{th} sample, and $\boldsymbol{D}_k$ represents the inspection data generated in the k^{th} sample.

As in Chapters III, the analysis is done before any inspections are performed. As a result, $\boldsymbol{D}$ must be generated using all prior knowledge regarding the model parameters and the digital twin. This constitutes the only information available for the calculation of $f(D_i)$, where D_i is the data collected from the i^{th} type of inspection and $\boldsymbol{D} = [D_1, D_2 \ldots D_p]$, where p is the total number of types of inspections. Therefore, $f(D_k)$ is calculated as:

$$f(D_i) = \int f(y_i|\boldsymbol{\theta}) f'(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{69}$$

where y_i represents the output of the digital twin corresponding to the i^{th} type of inspection, $\boldsymbol{\theta}$ represents the underlying parameters, and $f'(\boldsymbol{\theta})$ represents the prior knowledge regarding those parameters. Note that Equation 69 is simply an uncertainty propagation problem, where the other sources of uncertainty (such as physical variability in inputs, solution approximation errors, data uncertainty, POD, and data measurement errors) can also be included in the computation of $f(y_i|\boldsymbol{\theta})$.

The optimization requires the generation of many samples of the outcomes of inspection data over multiple cycles of system usage. Multiple cycles of system usage are required because the value of an inspection may be that fewer additional inspections, or lower quality inspections are required later in the systems life cycle. For example, if a higher fidelity inspection is performed at a cost of 2 cost units before Mission 1, perhaps no inspections are necessary prior to the next two missions, but if a lower fidelity inspection is performed at a cost of 1 cost unit before Mission 1, perhaps lower fidelity inspections are also required before Missions 2 and 3. In this scenario choosing a higher fidelity inspection is optimal even though the lower fidelity inspection more inexpensively clears the system to perform Mission 1. Damage inspection data must be simulated at every point in the system life cycle where an inspection is to be performed; thus, the sampling may need to be extensive. To deal with this problem, Latin hypercube sampling is used to decrease the necessary number of samples.

In each sample, the digital twin propagates its initial distribution of model parameters through several cycles of simulated use, updating the model parameters each time a new piece of inspection data is generated. At points where inspection data is to be generated, a decision must be made about what type of inspection to perform. Within the optimization framework, these decisions are predetermined for each set of samples, $\boldsymbol{I_{test}}$. The cost and return of each sample can then be calculated. By performing many samples, the objective function, shown in Equation 68, can be computed. A set of samples must be developed for each combination of damage inspection decisions. This process is summarized in Figure 19.

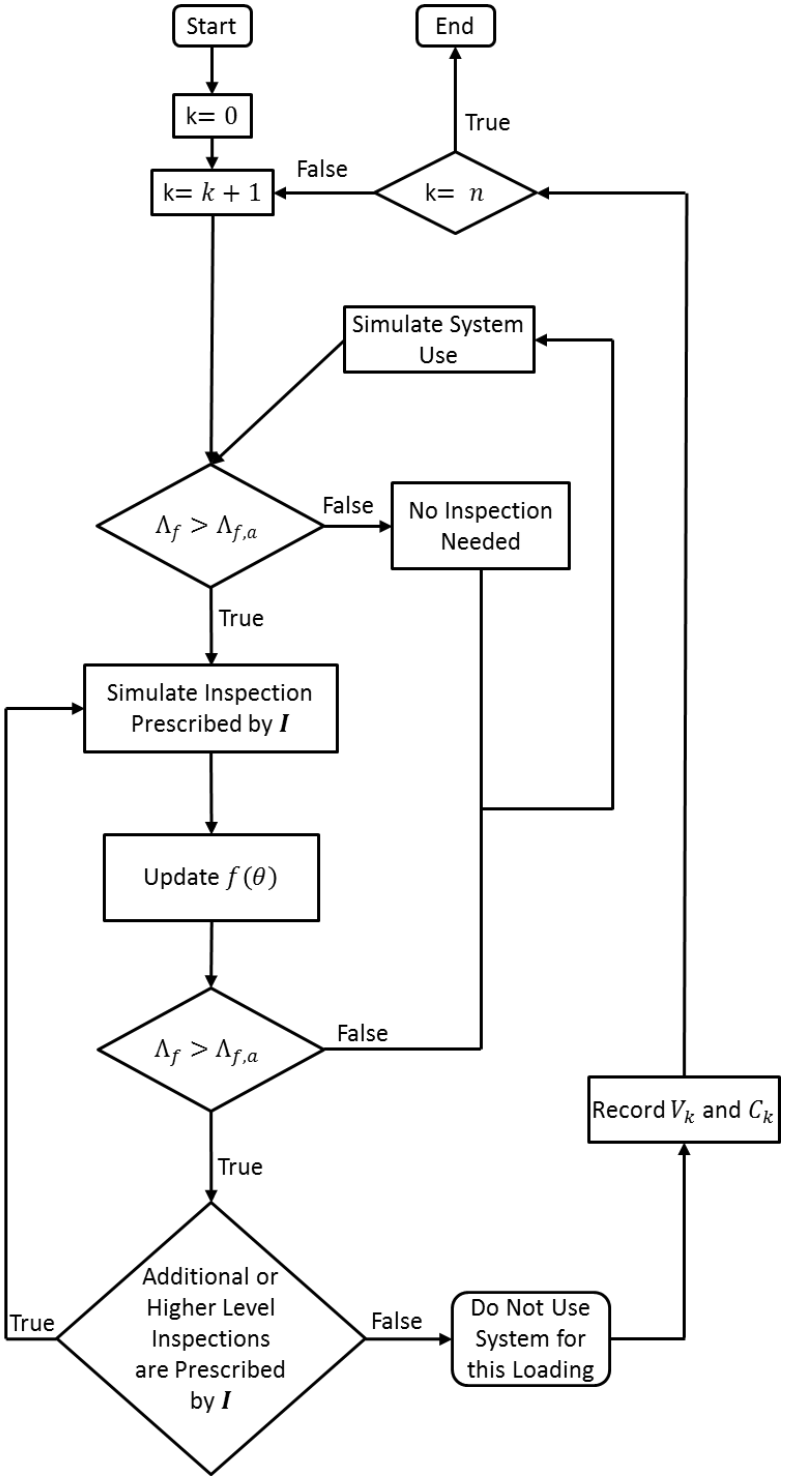

Figure 19: Inspection Sampling Flowchart

It is likely that the number of possible damage inspection type decision combinations is too large to make the optimization computationally affordable. If this is the case an adaptive methodology that is broken into steps with a greedy search method between steps (much like the optimization presented in Chapter III) can be utilized to simplify the computation. The step size, ϕ, is the number of missions to be considered in each step of the multi-step optimization. ϕ should be as large as computationally affordable to maximize the optimization's accuracy. ϕ can also be chosen based on the required intervals between inspections (which is not addressed in this thesis). The selected inspection decision combination for the j^{th} step, I^j, becomes the starting place for the optimization of the next step, ϕ^{j+1}. The multi-step optimization is formulated as follows.

$$\begin{aligned} &\underset{\boldsymbol{I}_{test}^{j,new}}{\text{Maximize}} \frac{E(V|\boldsymbol{I}_{test}^{j,new})}{E(C|\boldsymbol{I}_{test}^{j,new})} \\ &\boldsymbol{I}_{test}^{j} = \boldsymbol{I}_{test}^{j-1} + \boldsymbol{I}_{test}^{j,new} \\ &\boldsymbol{I}_{test}^{j,new} = \left[\boldsymbol{I}_{1}^{j,new}, \boldsymbol{I}_{2}^{j,new} \dots \boldsymbol{I}_{q}^{j,new}\right] \end{aligned} \tag{70}$$

The optimization procedure described in this section determines the ideal decisions for inspection type decisions given the current information contained in the system's digital twin. The single step methodology is presented for a single component fatigue crack growth model in section 6.5.

6.4 Summary of proposed methodology

The methodology developed in the preceding sections uses a sampling technique to select the inspection type decision combination that best maximizes the expectation of the value received from the system of interest. The methodology can be summarized in 4 parts.

1. Digital twin construction: The construction of a proper and useful digital twin is key to the effectiveness of this methodology. Digital twins should accurately represent the engineering system and should incorporate all viable failure mechanisms as well and represent the interactions between different components and subsystems. Digital twins should be rigorously verified, validated and calibrated and should include all uncertainties about the system.

2. Construct Bayesian network: The methodology uses a Bayesian network to connect all the models and the data available in a unified manner. The model errors, if available, can also be included in the Bayesian network. The Bayesian network is used to update the model parameters of the digital twin when new load or inspection data is collected.

3. Inspection development: Inspection procedures should be properly developed so that the information gained from the inspection is relevant to the reduction of uncertainty in the digital twin. Inspection errors must be sufficiently low, the POD of an inspection must be sufficiently high, and the damage being

measured must be critical to the performance or failure of the system. A global sensitivity analysis approach could be used to relate the parameters updated by a given inspection to the full system output in order to determine the relevance of an inspection type. Also, the number of inspections considered in the optimization must be small enough to make the optimization computation affordable.

4. Inspection Optimization: Once the relevant inspections have been selected, the vector $\boldsymbol{I}_{test}$ must be constructed. $\boldsymbol{I}_{test}$ includes all possible combinations of inspection selection decisions possible in the number of missions to be considered. For each combination, a finite number of samples are run using a Latin hypercube sampling method to generate values of damage inspection results. For each sample, the return, V, which is measured by some form of system usage, and the inspection cost, C, is calculated. The optimization output variable is calculated as the expectation of V divided by the expectation of C for all samples of a given inspection combination. The optimal inspection type decision combination is that which best maximizes this quotient.

The result of the proposed methodology is a chosen inspection type decision combination which is best expected to most economically inspect the system. This methodology is demonstrated with a fatigue crack growth model in Section 6.5.

6.5 Fatigue crack growth analysis

Digital twins in a practical sense are complex computer models of complicated engineering systems which are comprised of several functional levels and cover a wide array of physics. A well-developed digital twin evaluates numerous failure modes and models the interactions between different levels and components. For the purposes of this research, a single component digital twin that models fatigue crack growth on a helicopter rotor mast will be used to illustrate the developed methodology for uncertainty reduction.

6.5.1 Description of problem

The model used in this example problem, shown in Figure 16, simulates fatigue crack growth in a helicopter rotor mast subjected to both tensile and bending loads. Note that the model used in this problem is the same as that used in the example problem in Chapter V. For completeness, some of the model properties will be restated in this chapter.

The component is a two radius hollow cylinder which is assumed to experience cracking at the beginning of the transition region (a common place for crack initiation). The objective of this example problem will be to choose a crack damage inspection type decision combination that most economically reduces the uncertainty in the performance of the component.

The rotor mast component is made with a 4340 Steel Alloy which has the material properties shown in Table 27. The geometric properties of the component are shown in Table 28.

Table 27: Mechanical Properties of 4340 Steel: Digital Twin

Steel Alloy 4340	
Modulus of Elasticity	205 GPa
Poisson Ration	0.29
Yield Stress	1.110 GPa
Ultimate Stress	0.710 GPa

Table 28: Geometric Properties: Digital Twin

Length	0.152 m
Inside Radius	7.62 mm
Outside Radius (narrow section)	15.24 mm
Outside Radius (wide section)	20.30 mm

Unlike the examples in previous chapters, this example will incorporate information collected during multiple time steps. In this example, the component's digital twin was first used to determine if the probability of failure under a specified loading condition is sufficiently low. If the failure criterion is met, the component will be "cleared" to perform the "mission". However, if the failure criterion is not met, inspection data must be collected to attempt to clear the component. If the component is cleared to perform a mission, the process repeats itself to attempt to clear the component for the next mission. This process continues until inspection data no longer successfully clears the component at which point it is designated to be retired, repaired, or used at a lower load level.

Each mission consists of an average bending load (B), an average torsion load (T), and a specified duration (D) (measured by number of helicopter blade revolutions). Each of the three loading parameters has nominal values and some degree of uncertainty. The values of the loading parameters are assumed to be Gaussian with a mean and standard deviation as shown in Table 29.

Table 29: Mission Loading Parameters: Digital Twin

Loading Parameter	Mean Value	Standard Deviation
Avg. Bending Load	0.1596 (N-m)	1.596×10^{-3} (N-m)
Avg. Torsion Load	0.1177 (N-m)	8.243×10^{-3} (N-m)
Duration	20,000 cycles	1500 cycles

Once a component is cleared for a mission, the component is submitted to an actual loading which is measured by the strain gauges attached to the component with some degree of uncertainty. This uncertainty increases with time as the confidence in the gauges decreases. The loading parameters measurement uncertainties are given as a function of number of mission run from the beginning of the analysis, N_m. The relationships are given in Equation 71.

$$\begin{aligned} COV(B_{measured}) &= 0.001 * (N_m + 1) \\ COV(T_{measured}) &= 0.01 * (N_m + 1) \\ COV(D_{measured}) &= 0.001 * (N_m + 1) \end{aligned} \tag{71}$$

The relationships between the crack growth model parameters, the component loading parameters, damage inspection data, information gained by strain gauges, and the systems errors and uncertainties can be connected using a Bayesian network as shown in Figure 20.

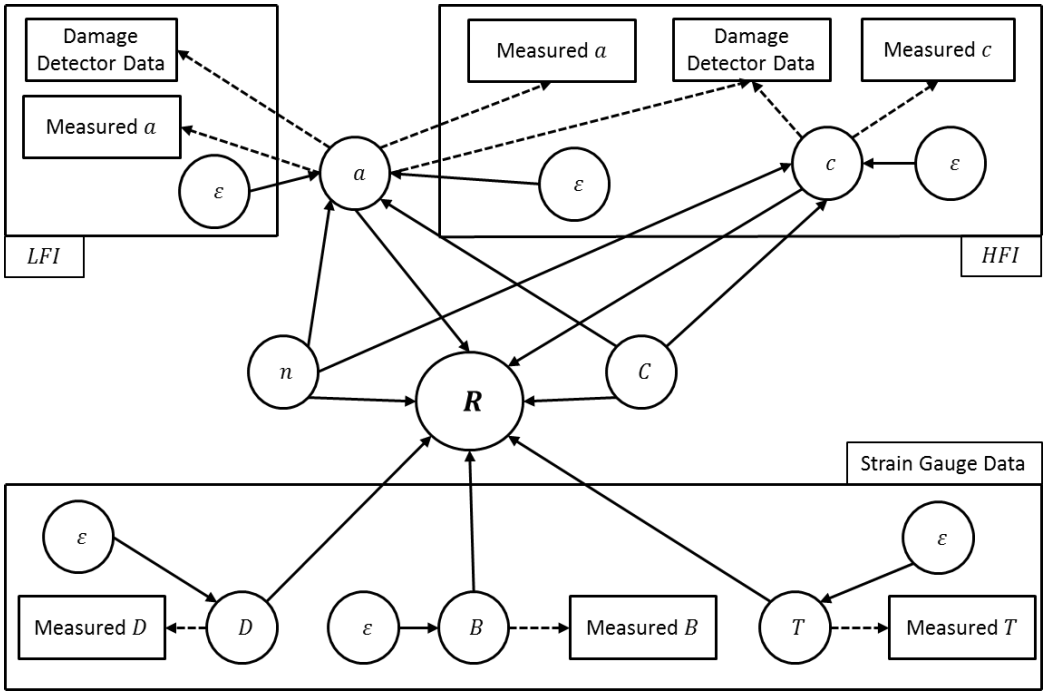

Figure 20: Bayesian Network: Digital Twin

The failure criterion, g, for this problem will be a function of the crack length, CL, and the crack depth, CD, and is given as:

$$g(CL, CD) = 0.635mm^2 - \frac{\pi}{4} CL * CD \tag{72}$$

In Equation 72, the area of a semi-elliptical crack with crack length CL and crack depth CD is compared to an area of 0.635 mm^2. If the area is smaller than 0.635 mm^2 g is positive. Otherwise, g is negative and the component is deemed to have failed. The system is deemed to have an acceptably low probability of failure if $p_f \leq p_a = 0.001$ with a confidence of 95%. If 20,000 samples of g, given the initial crack distribution as understood by the system's digital twin subjected to a prescribed loading condition, are taken, observing 13 (=$\Lambda_{f,a}$) or fewer failures in the 20,000 samples creates a distribution

of the probability of p_f (as calculated by Equation 66) that is normal with a mean of 0.00065 and a standard deviation of 0.00018. Observing 13 or fewer failures in 20,000 samples satisfies the above conditions and the component is cleared to perform a mission under the prescribed loading conditions.

This example problem will seek to determine the optimal inspection type decision combination for the component given the prior condition as understood by the digital twin. The prior condition is shown in Table 30.

Table 30: Parameter Distribution Prior to Optimization: Digital Twin

Parameter	Distribution	μ	σ
a (crack length)	lognormal	0.0987	0.0660
c (crack depth)	lognormal	0.0486	0.0315
n	normal	3.0	0.0090
C	normal	3.408×10^{-10}	3.408×10^{-11}

6.5.2 Inspection optimization

Two types of inspections will be considered in this example: (1) a *High Fidelity Inspection*, HFI, and (2) a *Low Fidelity Inspection*, LFI. It is also possible that no inspection is necessary. That situation will be denoted as NI.

A HFI will use ultrasonic equipment to measure both the length and depth of a crack in the helicopter rotor mast. HFI will have a high POD and low measurement error. If a HFI fails to clear the component to perform the next mission, the component is designated for repair or retirement or downgraded mission as there is not a higher level inspection. The cost of a HFI is assumed to be 3 cost units.

A LFI consists of a visual inspection that can only measure the length of the crack. LFI will have a lower POD and higher measurement error than a HFI. If a low

fidelity inspection fails to clear the component for use, a HFI must then be performed to attempt to clear the component. The cost of a single LFI is assumed to be 1 cost unit. It must be noted that if both LFI and HFI need to be performed because the LFI fails to clear the component, the total cost of this scenario is 4 cost units. The possible inspection scenarios for a single mission are summarized in Figure 21.

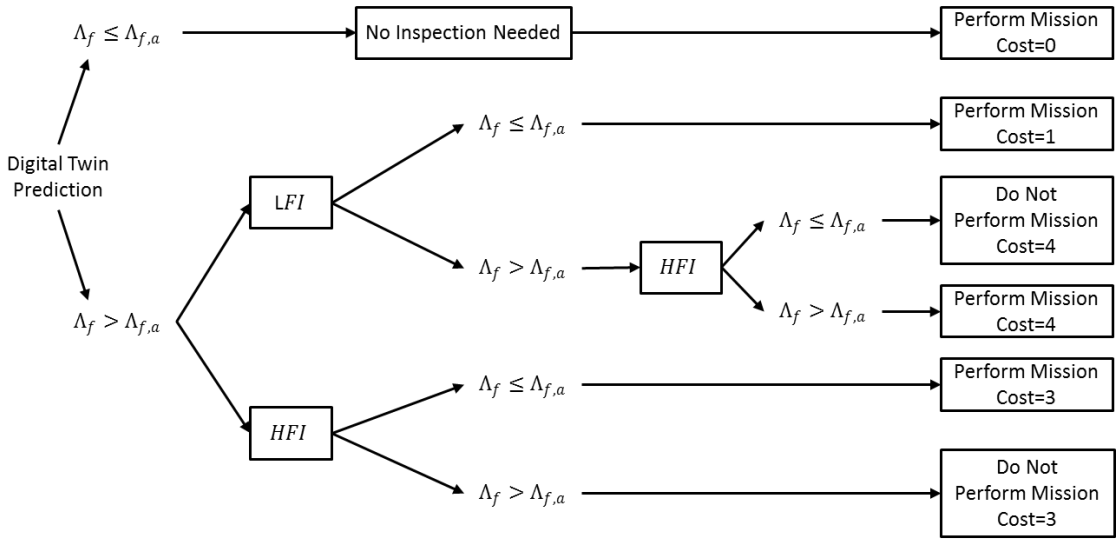

Figure 21: Single Inspection Decision Flowchart: Digital Twin

POD is usually a monotonic function of crack size and is used to represent uncertainty in crack detection. The result of crack detection can be treated as a binary variable [52]

$$I_d = \begin{cases} = 1, Crack\ Detected \\ = 0, Crack\ Not\ Detected \end{cases} \tag{73}$$

The probability of the two possible values of I_d can be expressed in terms of POD

$$Pr(I_d = 1|a) = POD(a) = w(a)$$
$$Pr(I_d = 0|a) = 1 - POD(a) = 1 - w(a) \tag{74}$$

The function $w(a)$ can be obtained either by pure empirical methods [57] or model-assisted methods [58]. In this chapter, a statistical representation is adopted by treating $w(a)$ as a standard normal cumulative distribution function [59]

$$w(a) = \Phi(\alpha + \beta a) = \frac{1}{2}\left[1 + erf\left(\frac{\alpha + \beta a}{\sqrt{2}}\right)\right] \tag{75}$$

The parameters α and β are assumed to be determined using the following equations:

$$\alpha_{LFI} = -1.5$$
$$\beta_{LFI} = 15 - 1.67N_m$$
$$\alpha_{HFI} = 0$$
$$\beta_{HFI} = 10 - 1.00N_m \tag{76}$$

The measurement errors corresponding to the two test types are given as:

$$\varepsilon_{m,LFI} = N\sim(0\text{mm}, 0.254\text{mm})$$
$$\varepsilon_{m,HFI} = N\sim(0\text{mm}, 0.127\text{mm}) \tag{77}$$

where $N\sim(\mu,\sigma)$ denotes a normal distribution with mean μ and standard deviation σ.

In order to solve the optimization problem, all possible decision combinations should be considered for the desired number of missions to be considered. Notice from Figure 21 that only one inspection decision (whether to first perform a LFI or a HFI) needs to be made per inspection cycle. For this example, we will consider an analysis length of four missions. Therefore, only sixteen possible combinations of inspection decisions, $\boldsymbol{I}_{test}$, exist. One hundred samples of each $\boldsymbol{I}_{test}$ are taken and the value of the objective function, $E(V)/E(C)$, is calculated for each. The $\boldsymbol{I}_{test}$ which maximizes the objective function is selected. The results are shown in Table 31.

Table 31: Inspection Optimization Results: Digital Twin

Mission Number				$\boldsymbol{E(V)/E(C)}$
1	2	3	4	
LFI	LFI	LFI	LFI	0.230
LFI	LFI	LFI	HFI	0.234
LFI	LFI	HFI	LFI	0.245
LFI	LFI	HFI	HFI	0.249
LFI	HFI	LFI	LFI	0.248
LFI	HFI	LFI	HFI	0.252
LFI	HFI	HFI	LFI	0.262
$\boldsymbol{LFI}$	$\boldsymbol{HFI}$	$\boldsymbol{HFI}$	$\boldsymbol{HFI}$	**0.265**
HFI	LFI	LFI	LFI	0.199
HFI	LFI	LFI	HFI	0.200
HFI	LFI	HFI	LFI	0.212
HFI	LFI	HFI	HFI	0.214
HFI	HFI	LFI	LFI	0.197
HFI	HFI	LFI	HFI	0.199
HFI	HFI	HFI	LFI	0.210
HFI	HFI	HFI	HFI	0.210

Table 31 shows that given only the information currently contained in the system's digital twin, the ideal inspection type decision combination is LFI, HFI, HFI,

HFI and it is expected to produce 0.265 missions for every cost unit spent on inspections.

6.6 Conclusions

The improvement of digital twin technology will provide great advances in the ability of engineers to understand the capabilities and limitations of complex engineering systems. Uncertainty quantification is a vital part of the development of digital twins and the optimization of uncertainty reduction techniques will help to economically gain information that aids in the engineering decision making process. In this chapter, tools were developed to select non-destructive test type decisions based on the information and uncertainties currently incorporated in a digital twin. This methodology is easily extendable to other aspects of digital twin development such as validation test selection and design as well as calibration test selection and design (as shown in Chapters II and III). Future work might seek to relax the time constraints imposed by this methodology so that decisions can be made about schedule as well inspection type.

The main road block in the implementation of this methodology (as with most aspects of digital twin development and use) is the intense computational effort required. In order to solve the optimization problem posed in Section 6.5, 96 processors were run continuously for over 8 days. It can be imagined that for a more complex system with multiple components, several failure modes, and a variety of inspection types, the computational burden would become significant. For this methodology to have real world practicality, significant computational advancements, either faster computer processing or faster computational algorithms would be required.

CHAPTER VII

CONCLUSIONS AND FUTURE WORK

The increasing complexity of modern engineering systems has created a demand for robust uncertainty quantification techniques. As engineering systems become more complicated, expensive, and dangerous, the need to accurately model their behavior prior to their use or even fabrication continues to increase and the cost of failure becomes increasingly unacceptable. The design of an efficient test campaign for uncertainty reduction, and accelerated model and system certification is essential to the continued development of high-tech systems.

This research developed four methodologies for uncertainty reduction under various engineering conditions. All four methodologies utilized the Bayesian network tool to connect data at different system levels and across multiple physics and engineering processes. The ability to connect data in a unified framework through the Bayesian network allows information gained about the system at the material, component, and subsystem level to be propagated upward to reduce the uncertainty of the full system output. The primary benefit to using a Bayesian network for test campaign design is the ability to calibrate the full system model without testing the full system. This is critically important when full system tests are impractical due to high costs or extreme operating conditions (i.e. space vehicle or high performance aircraft).

Chapter III primarily explored the concept of test resource allocation for uncertainty reduction in hierarchical systems. Using a Bayesian network model,

uncertain quantities, data, and both model and testing errors (all at multiple levels) were connected. Decisions were made about which tests at which levels, and on which variables should be performed. The methodology selected both the tests that should be performed and how many repetitions of each test to do in order to best reduce the uncertainty in the full system output. The methodology was demonstrated on a multi-physics thermal vibration problem, and a multi-physics multi-level telescope mirror problem. The results clearly show the benefit of systematic test selection. The designed methodology used a complex integer optimization coupled with a greedy search algorithm to develop a set of tests to perform that is best expected to reduce the variance of the full system output. Future work into this subject may include increases in computational efficiency to allow for more complicated problems with several levels and many calibration parameters, and the use of robust optimization techniques.

Chapter IV took the methodology developed in Chapter III and extended it to include the design of the test input settings for each calibration test. Chapter IV also incorporated an adaptive methodology in which the tests are designed sequentially and the results from all previous tests are used to select the input settings for the next test. Chapter IV also used an information-theoretic objective function which allows the calibration to update the mean, the uncertainty in the mean, the variance, and the uncertainty in the variance, for each calibration parameter as opposed to a variance reduction technique which converge to a single value for each calibration parameter. The result of the methodology is a multi-step adaptive algorithm to find a sequence of test types and the corresponding test input settings that seek to maximize the information gained on the full system output. Future work could include increases in computational

efficiency and a mechanism for choosing the order in which tests in a given combination should be performed. Also, the simultaneous optimization of both the test type and the test input settings would likely lead to a more accurate solution, but is computationally challenging.

Chapter V adapts the methodologies developed in the previous chapters to solve a test campaign problem for manufacturing optimization. Integrated computation materials engineering is an emerging field that is seeking to rectify the disparity in the development time between complex engineering systems and new engineering materials. This chapter discusses an optimization methodology that could be taken to design calibration tests that would reduce the uncertainty in the manufacturing optimization. The proposed methodology incorporates an optimization of the manufacturing variables in the objective function of the test allocation and design optimizations. The result is a two-level optimization that optimizes some aspect of the full system output as a function of the manufacturing variables. This field has great promise for new research and future work can include a variety of different topics including: the development of models for a range of various manufacturing variables, development of an implementation structure to integrate various levels of models and information, and the handling of numerous and large databases on materials behavior.

Chapter VI explores the optimization of damage inspection type decision combinations using the digital twin of a system. Using a sampling technique, different inspection type decision combinations are compared to determine which one best maximizes the value gained from the system in its remaining useful life. The result of the methodology is a set of inspection decisions that is most efficient given the current state

of the digital twin. Future work in this area requires increased computational efficiency in order to analyze large realistic systems with multiple interacting components and a variety of failure modes.

The continued improvement of advanced uncertainty quantification techniques including the optimization of system test campaigns will accelerate the development of highly complex engineering systems and will play a crucial role in the advancement of modern technology.

REFERENCES

[1] J. C. Helton. Quantification of margins and uncertainties: Conceptual and computational basis. *Reliability Engineering & System Safety*, In Press, Accepted Manuscript, 2011.

[2] SS. Rao and KK. Annamdas. Dempster-Shafer Theory in the Analysis and Design of Uncertain Engineering Systems. *Product Research*, pages 135–160, 2009.

[3] S. Sankararaman and S. Mahadevan. Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data. *Reliability Engineering & System Safety*, 96(7):814 – 824, 2011.

[4] N. Metropolis and S. Ulam. The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341, 1949.

[5] A. Dey and S. Mahadevan. Ductile structural system reliability analysis using adaptive importance sampling. *Structural Safety*, 20(2):137–154, 1998.

[6] O. Ditlevsen and HO Madsen. *Structural reliability methods*. John Wiley and Sons, 1996.

[7] A. Haldar and S. Mahadevan. *Probability, reliability, and statistical methods in engineering design*. John Wiley & Sons, Inc., 2000.

[8] T. Leonard and JSJ. Hsu. *Bayesian methods: an analysis for statisticians and interdisciplinary researchers*. Cambridge Univ Pr, 2001.

[9] PM. Lee. *Bayesian statistics*. Arnold London, UK:, 2004.

[10] A. Caticha and R. Preuss. Maximum entropy and Bayesian data analysis: Entropic prior distributions. *Physical Review E*, 70(4):046127, 2004.

[11] X. Jiang and S. Mahadevan. Bayesian cross-entropy methodology for optimal design of validation experiments. *Measurement Science and Technology*, 17:1895, 2006.

[12] G. Shafer. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton, NJ, 1976.

[13] HR. Bae, RV. Grandhi, and RA. Canfield. An approximation approach for uncertainty quantification using evidence theory. *Reliability Engineering & System Safety*, 86(3):215–225, 2004.

[14] H. Agarwal, JE. Renaud, EL. Preston, and D. Padmanabhan. Uncertainty quantification using evidence theory in multidisciplinary design optimization. *Reliability Engineering & System Safety*, 85(1-3):281–294, 2004.

[15] Y. Ben-Haim and I. Elishakoff. *Convex models of uncertainty in applied mechanics*, volume 112. Elsevier Amsterdam, 1990.

[16] D. Dubois, H. Prade, and EF. Harding. *Possibility theory: an approach to computerized processing of uncertainty*, volume 2. Plenum Press New York, 1988.

[17] K. Zaman, S. Rangavajhala, MP. McDonald, and S. Mahadevan. A probabilistic approach for representation of interval uncertainty. *Reliability Engineering & System Safety*, 96(1):117 – 130, 2011. Special Issue on Safecomp 2008.

[18] WL. Oberkampf and MF. Barone. Measures of agreement between computation and experiment: Validation metrics. *Journal of Computational Physics*, 217(1):5 – 36, 2006. Uncertainty Quantification in Simulation Science.

[19] WL. Oberkampf, TG. Trucano, and C. Hirsch. Verification, validation, and predictive capability in computational engineering and physics. *Applied Mechanics Reviews*, 57:345, 2004.

[20] RG. Hills and TG. Trucano. Statistical validation of engineering and scientific models: Background. *Sandia National Laboratories, SAND99-1256*, 1999.

[21] A. Urbina, TL. Paez, TK. Hasselman, GW. Wathugala, and Yap K. Assessment of model accuracy relative to stochastic system behavior. In *44th AIAA structures, structural dynamics, materials conference*, Norfolk, VA, Apr 7 - 10, 2003.

[22] R. Zhang and S. Mahadevan. Bayesian methodology for reliability model acceptance. *Reliability Engineering & System Safety*, 80(1):95–103, 2003.

[23] R. Rebba and S. Mahadevan. Computational methods for model reliability assessment. *Reliability Engineering & System Safety*, 93(8):1197–1207, 2008.

[24] S. Mahadevan and R. Rebba. Validation of reliability computational models using Bayesian networks. *Reliability Engineering & System Safety*, 87(2):223–232, 2005.

[25] R. Rebba, S. Mahadevan, and S. Huang. Validation and error estimation of computational models. *Reliability Engineering & System Safety*, 91(10-11):1390–1397, 2006.

[26] S. Sankararaman and S. Mahadevan. Model validation under epistemic uncertainty. *Reliability Engineering & System Safety*, In Press, Available Online:–, 2011.

[27] X. Jiang and S. Mahadevan. Bayesian risk-based decision method for model validation under uncertainty. *Reliability Engineering & System Safety*, 92(6):707–718, 2007.

[28] A. Urbina, S. Mahadevan, and T. Paez. Resource allocation using quantification of margins and uncertainty. In *51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, 2011.

[29] Jensen, F., *An introduction to Bayesian networks*, Vol. 210, UCL press London, 1996.

[30] Heckerman, D., "A tutorial on learning with Bayesian networks," *Innovations in Bayesian Net-works*, 2008, pp. 33–82.

[31] Edwards, A., *Likelihood*, Cambridge Univ Pr, 1984.

[32] Pawitan, Y., *In all likelihood: statistical modelling and inference using likelihood*, Oxford University Press, USA, 2001.

[33] Sobol, I., "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," Mathematics and Computers in Simulation, Vol. 55, No. 1-3, 2001, pp. 271–280.

[34] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., Global sensitivity analysis: the primer , Wiley and Sons, 2008.

[35] Sudret, B., "Global sensitivity analysis using polynomial chaos expansions," Reliability Engineering & System Safety, Vol. 93, No. 7, 2008, pp. 964–979.

[36] Ghanem, R. and Spanos, P., "Polynomial chaos in stochastic finite elements," *Journal of Applied Mechanics*, Vol. 57, No. 1, 1990, pp. 197–202.

[37] Boser, B., Guyon, I., and Vapnik, V., "A training algorithm for optimal margin classifiers," *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, 1992, pp. 144–152.

[38] Tipping, M., "Sparse Bayesian learning and the relevance vector machine," *The Journal of Ma- chine Learning Research*, Vol. 1, 2001, pp. 211–244.

[39] Rasmussen, C., *Evaluation of Gaussian processes and other methods for non-linear regression*, Ph.D. thesis, University of Toronto, 1996.

[40] Santner, T., Williams, B., and Notz, W., *The design and analysis of computer experiments*, Springer Verlag, 2003.

[41] McFarland, J., *Uncertainty Analysis for Computer Simulations through Validation and Calibration*, Ph.D. thesis, Vanderbilt University, 2008.

[42] Rasmussen, C., "Gaussian processes in machine learning," *Advanced Lectures on Machine Learning*, 2004, pp. 63–71.

[43] Hombal, V., Mahadevan, S., "Bias minimization in Gaussian process surrogate modeling for uncertainty quantification," *International Journal for Uncertainty Quantification*, Vol. 1(4), 2011, pp 321-349.

[44] McKay, M., Beckman, R., Conover, W., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *American Society for Quality*, Vol. 21, No. 2, 1979, pp. 239-245.

[45] Park, J., "Optimal Latin-hypercube designs for computer experiments," *Journal of Statistical Planning and Inference*, Vol. 39(1), 1994, pp. 95-111.

[46] S. Kullback and R. Leibler. On information and sufficiency. *The Annuals of Mathematical Statistics*, 22(1):79-86,1951.

[47] Thornton, E., *Thermal structures for aerospace applications*, AIAA, 1996.

[48] Edwards, H., "SIERRA framework for parallel adaptive multiphysics applications." Tech. rep., Sandia National Laboratories, 2004.

[49] Eldred, M., Giunta, A., van Bloemen Waanders, B., Wojtkiewicz Jr, S., Hart, W., and Alleva, M., "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis Version 3.0 Reference Manual," Tech. rep., Sandia National Labs., Albuquerque, NM (US) Sandia National Labs., Livermore, CA (US), 2001.

[50] National Material Advisory Board. *Integrated Computational Materials Engineering*. National Academy Press, 2008.

[51] Xiang, Y. and Liu, Y. "Mechanism modeling of shot peening effect on fatigue life prediction", Fatigue and Fracture of Engineering Materials and Structures, 2009.

[52] Ling, Y., Mahadevan, S., "Integration of Structural Health Monitoring and Fatigue Damage Prognosis," *Mechanical Systems and Signal Processes*, Vol. 28, 2012, pp..89-104.

[53] P.C. Paris and F. Erdogan, "A critical analysis of crack propagation laws," *Journal Of Basic Engineering*, vol. 85, 1963, pp. 528-534.

[54] Tuegel, E., Ingraffea, A., Eason, Y. and Spottawood, S., "Reengineering Aircraft Structural Life Prediction Using a Digital Twin," *International Journal of Aerospace Engineering*, Vol. 2011, Article ID 154798.

[55] Zhang, R., Mahadevan, S., "Fatigue Reliability Analysis Using Nondestructive Inspection," *Journal of Structural Engineering*, Vol. 127, Issue 8, 2001, pp. 957.

[56] Stratman, B., *Reliability and Clustering Techniques for Inspection Optimization of Large Populations*, Ph.D. thesis, Vanderbilt University, 2007.

[57] F. Spencer and D. Schurman, "Reliability Assessment at Airline Inspection Facilities, Volume III: Results of an Eddy Current Inspection Reliability Experiment," *DOT/FAA/CT-92/12*, Vol. III, 1995.

[58] K. Smith, B. Thompson, B. Meeker, T. Gray, and L. Brasche, "Model-Assisted Probability of Detection Validation for Immersion Ultrasonic Application," *AIP Conference Proceedings*, 2007, pp. 1816-1822.

[59] A.P. Berens and P.W. Hovey, "Statistical Methods for Estimating Crack Detection Probabilities," *Probabilistic Fracture Mechanics and Fatigue Methods: Applications for Structural Design and Maintenance*, 1983, pp. 79-94.