SANDIA REPORT
SAND2005-5980
Unlimited Release
Printed November, 2005

Validation of the Thermal Challenge Problem Using Bayesian Belief Networks

John M. McFarland and Laura P. Swiler

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.osti.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd.
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online order: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND 2005-5980
Unlimited Release
Printed November, 2005

Validation of the Thermal Challenge Problem Using Bayesian Belief Networks

John McFarland
Ph.D. Student
Permanent address: Reliability and Risk Engineering and Management
VU Station B 351831
2301 Vanderbilt Place
Nashville, TN 37235-1831

Summer 2005 address: Validation and Uncertainty Quantification Processes
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0828

Laura P. Swiler
Optimization and Uncertainty Estimation Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0370

Abstract

The thermal challenge problem has been developed at Sandia National Laboratories as a testbed for demonstrating various types of validation approaches and prediction methods. This report discusses one particular methodology to assess the validity of a computational model given experimental data. This methodology is based on Bayesian Belief Networks (BBNs) and can incorporate uncertainty in experimental measurements, in physical quantities, and model uncertainties. The approach uses the prior and posterior distributions of model output to compute a validation metric based on Bayesian hypothesis testing (a Bayes' factor).

This report discusses various aspects of the BBN, specifically in the context of the thermal challenge problem. A BBN is developed for a given set of experimental data in a particular experimental configuration. The development of the BBN and the method for "solving" the BBN to develop the posterior distribution of model output through Monte Carlo Markov Chain sampling is discussed in detail. The use of the BBN to compute a Bayes' factor is demonstrated.

Page Left Blank

Table of Contents

1. Introduction 6
2. Background 7
   2.1. Bayesian Inference 7
   2.2. Continuous Case 7
   2.3. Hypothesis Testing 8
   2.4. Bayesian Belief Networks 10
   2.5. Markov Chain Monte Carlo methods 11
3. BBN Development and Implementation 14
   3.1. Thermal Challenge Problem Description 14
   3.2. Development of Bayesian Network Representation 14
   3.3. Markov Chain Monte Carlo Analysis 17
4. Results 19
5. Summary 23
6. References 25
7. Distribution 26

Validation of the Thermal Challenge Problem Using Bayesian Belief Networks

1. Introduction

This report investigates the use of a validation methodology based on Bayesian Belief Networks (BBNs), applied to a specific problem of interest to Sandia National Laboratories. The validation methodology has been developed by Dr. Mahadevan and Ramesh Rebba at Vanderbilt University [Mahadevan and Rebba, 2005]. The idea is to use Bayesian statistics to compare the prior density of the model-based predictions with the posterior density of these predictions, after they are updated based on the experimental results. The methodology can account for uncertainties in both the experimental results and the model, including several sources of uncertainty that may affect the model prediction. The ratio of the relative likelihoods of the null hypothesis (that the data supports the proposed model) to the alternative hypothesis (the data does not support the proposed model) is called the Bayes' factor. If the Bayes' factor is greater than one, the data is said to favor the null hypothesis that the model prediction is true. If the Bayes' factor is smaller than one, the data is said to favor the alternative hypothesis that the model prediction is not true. A unique aspect of the methodology outlined in this paper is that the Bayes' factor is calculated as the result of a Bayesian Belief Network. The BBN shows the relationships between experimental results, the model prediction, and various parameters associated with the problem.

In this report, we develop a BBN for the thermal challenge problem. This problem has been developed by Kevin Dowding, Martin Pilch, and Richard Hills at SNL [Dowding et al., 2005] as a realistic test problem with features that represent a real-world situation facing experimentalists and modelers when developing and interpreting validation experiments. The thermal challenge problem involves transient heat conduction through the slab. The reader is presented with a variety of experimental configurations involving different heat flux initial conditions, different slab lengths, etc. Both experimental results and model results (based on the governing differential equations) are presented, and the reader must make an assessment of the accuracy and validity of the model as applied to the existing experimental data, and for a new experimental configuration which is outside the parameter space of the first set of experiments. The model has specifically been designed to have incomplete physics in the sense that the thermal parameters are considered constant when in fact they vary as a function of temperature.

The outline of the report is as follows: Section 2 covers the basics of Bayesian inference, Bayesian Belief networks, and the Markov Chain Monte Carlo (MCMC) methods which are used to calculate posterior densities when closed form solutions are not available. Section 3 discusses the thermal challenge problem in more detail, and presents the BBN for this model validation problem. Section 4 provides the results of testing the proposed validation metric on a particular experimental observation, and Section 5 summarizes the work and provides some thoughts about future research directions.

2. Background

2.1. Bayesian Inference

Bayes' rule relates the posterior density of a parameter to the likelihood function and the prior density of that parameter. In the discrete case, the Bayesian formulation for the posterior probability density function *h* is:

$$h(\theta \mid \mathbf{x}_1,...,\mathbf{x}_N) = \frac{f(\mathbf{x}_1 \mid \theta)...f(\mathbf{x}_N \mid \theta)g(\theta)}{\sum_\theta f(\mathbf{x}_1 \mid \theta)...f(\mathbf{x}_N \mid \theta)g(\theta)}, \tag{1}$$

where $\mathbf{x}_1,\ldots\mathbf{x}_N$ are independent and identically distributed observable random vector variables with probability mass function $f(\mathbf{x}|\theta)$ [Press, 1989]. Note that $f(\mathbf{x}|\theta)$ denotes the mass function of random vector $\mathbf{x}$ conditional upon another variable $\Theta = \theta$. Θ is assumed to be unobservable, and θ denotes the numerical value at which Θ is conditioned. In this case, we are assuming that Θ is discrete, and $g(\theta)$ is the probability mass function. The posterior probability density function of θ for a given set of observed data is $h(\theta|\mathbf{x})$.

Note that the denominator of (1) only depends on the x_i's and not on θ. Bayes formula is often written as:

$$h(\theta \mid \mathbf{x}_1,...,\mathbf{x}_N) \propto L(\mathbf{x}_1,...,\mathbf{x}_N \mid \theta)g(\theta), \tag{2}$$

where $L(\mathbf{x}_1,\ldots\mathbf{x}_N|\theta) = f(\mathbf{x}_1|\theta) *\ldots* f(\mathbf{x}_N|\theta)$ = the likelihood function of the data given the parameter θ for independent data. The expression (2) is a statement that the posterior distribution is proportional to the likelihood times the prior distribution.

2.2. Continuous Case

The formulation is identical to (1), only the parameter θ is now a continuous parameter with prior density $g(\theta)$. An alternative approach to expressing equation (1) is to use the likelihood function $L(\mathbf{x}_1,\ldots\mathbf{x}_N|\theta)$ instead of the conditional probability density functions $f(\mathbf{x}_i|\theta)$. So, one way of expressing Bayes' Theorem in the continuous case is:

$$h(\theta \mid \mathbf{x}_1,...,\mathbf{x}_N) = \frac{L(\mathbf{x}_1,...,\mathbf{x}_N \mid \theta)g(\theta)}{\int L(\mathbf{x}_1,...,\mathbf{x}_N \mid \theta)g(\theta)d\theta}. \tag{3}$$

The Bayesian framework allows one to integrate observed data and prior knowledge. In the case where one has no data or very little data, the posterior distribution is equal to or very close to the prior distribution and we haven't learned that much. In the case where there is a lot of data, and especially in the case where the likelihood function differs from the prior distribution, the posterior distribution is dominated by the likelihood function and it might be better to use a maximum likelihood approach. The ideal case for Bayesian methods are situations where one acquires information over time.

One criticism of Bayesian statistics is the formulation of the prior distribution. Ideally, the prior distribution is supposed to be obtained from subjective judgment and previous experience. In practice, the prior is often chosen from a family of distributions that makes the calculation of the posterior distribution tractable. These families are called "conjugate prior" distributions and will be discussed below in more detail.

2.3. Hypothesis Testing

Bayesian analysis can be useful for hypothesis testing, particularly when comparing two competing hypotheses. For example, consider comparing two competing hypotheses, H_0 and H_1. Let D be a test statistic based on a sample of N observations, $D \equiv D(\mathbf{x}_1,\ldots\mathbf{x}_N)$. Then Bayes' Theorem states that

$$P(H_0 \mid D) = \frac{P(D \mid H_0)P(H_0)}{P(D \mid H_0)P(H_0) + P(D \mid H_1)P(H_!)}, \tag{4}$$

where $P(H_0)$ and $P(H_1)$ denote the prior probabilities of H_0 and H_1 (these probabilities sum to one).

Likewise, for the alternate hypothesis, we have:

$$P(H_1 \mid D) = \frac{P(D \mid H_1)P(H_1)}{P(D \mid H_0)P(H_0) + P(D \mid H_1)P(H_1)}. \tag{5}$$

Taking the ratio of 4 and 5, we have:

$$\frac{P(H_0 \mid D)}{P(H_1 \mid D)} = \left[\frac{P(H_0)}{P(H_1)}\right]\left[\frac{P(D \mid H_0)}{P(D \mid H_1)}\right]. \tag{6}$$

This is interpreted as the posterior odds ratio in favor of H_0 is equal to the product of the prior odds ratio and the likelihood ratio. The "Jeffreys' Hypothesis Testing Criterion" states that if the posterior odds ratio exceeds one, we accept H_0, otherwise we accept H_1. In addition, the ratio of the posterior odds to the prior odds of the hypotheses is known as the "Bayes' factor". The Bayes' factor is also used to compare competing hypotheses. Note that the Bayes' factor is equal to the second term on the right hand side of Eq. (6), and thus it does not depend on the (subjective) prior weighting of the hypotheses. Also note that when the hypotheses are given equal weights, the Bayes' factor is equal to the posterior odds ratio. In Bayesian hypothesis testing, there is no notion of a significance level, which is a typical ambiguity that arises in classical hypothesis testing.

If H_0 and H_1 are simple hypotheses, as in $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$, then the Bayes' factor reduces to the likelihood ratio of the data under the hypotheses, which is also a frequentist test statistic for comparing two simple hypotheses. Instead of a point hypothesis, we could have H_1, for example, specified as a composite hypothesis, such as $H_1 : \theta \neq \theta_0$. In this case, we would also need to specify a probability distribution on the

parameter θ. The corresponding term in the Bayes' factor would then be equal to $P(D \mid H_1) = \int_{-\infty}^{+\infty} f(D \mid \theta) g(\theta) d\theta$, where $f(D \mid \theta)$ is the likelihood function associated with the observation D, and $g(\theta)$ is the prior distribution of θ under H_1. This is known as an averaged or marginal likelihood of the observations under H_1.

Mahadevan and Rebba (2005) formulated a method for applying Bayesian hypothesis testing when we want to assess the predictive capability of a single model. As above, we use the notation that the model response is given by θ, and we have a set of (uncertain) experimental observations characterized by a test statistic D. For the model input parameters corresponding to the observations in D, we denote the model prediction by θ_p. Thus, the null hypothesis, $H_0 : \theta = \theta_p$, is used. When dealing with only one model, the difficulty is in formulating the alternative hypothesis. When no extra information is available, Mahadevan and Rebba propose using the composite hypothesis, $H_1 : \theta \neq \theta_p$, with $g(\theta)$ under H_1 given by the probability distribution of θ based on the simulation model and the uncertainty of the input parameters.

Thus, the Bayes' factor becomes

$$B = \frac{P(D \mid H_0 : \theta = \theta_p)}{P(D \mid H_1 : \theta \neq \theta_p)}, \tag{7}$$

where we have assumed the prior odds ratio for the two hypotheses to be unity. As shown by Mahadevan and Rebba, Eq. (7) reduces to

$$B = \frac{P(\theta_p \mid D)}{g(\theta_p)}, \tag{8}$$

which is the ratio of the posterior probability of θ to the prior probability of θ, evaluated at θ_p.

As a side note, consider what would happen if a non-informative, or diffuse prior distribution were used for the alternative hypothesis, such that we have $g(\theta) \propto c$, where c is an arbitrary constant. From Eq. (7), we have for the Bayes' factor, $B = \frac{P(D \mid \theta = \theta_p)}{\int f(D \mid \theta)(c) d\theta} = \frac{P(D \mid \theta = \theta_p)}{c \int f(D \mid \theta) d\theta}$, which depends on the arbitrary constant, c. Thus, the Bayes' factor will be arbitrarily large or small depending on the constant. This is the typical problem which generally prohibits the use of non-informative priors for hypothesis testing. For a more detailed discussion, refer to Press (2003) and Robert (2001).

Testing a point null hypothesis, as in Eq. (7), is often controversial because it may be possible that the model is rejected even though the difference between the prediction and

the observations is small enough for practical purposes. Although the point null hypothesis is used in this report (for illustration) to analyze the challenge problem, an interval hypothesis can be accommodated with the Bayesian framework. If we want to test $H_0 : \left|\theta - \theta_p\right| < \varepsilon$, the Bayes' factor becomes

$$B = \frac{\int_{\theta_p-\varepsilon}^{\theta_p+\varepsilon} f(D \mid \theta) g(\theta) d\theta}{\int_{-\infty}^{\theta_p-\varepsilon} f(D \mid \theta) g(\theta) d\theta + \int_{\theta_p+\varepsilon}^{+\infty} f(D \mid \theta) g(\theta) d\theta}. \tag{9}$$

Note that by applying Bayes' theorem, Eq. (9) can be reduced to

$$B = \frac{\int_{\theta_p-\varepsilon}^{\theta_p+\varepsilon} f(\theta \mid D) d\theta}{\int_{-\infty}^{\theta_p-\varepsilon} f(\theta \mid D) d\theta + \int_{\theta_p+\varepsilon}^{+\infty} f(\theta \mid D) d\theta}, \tag{10}$$

which is a ratio of areas of the posterior density of θ. This is an entirely different formulation then that of the point hypothesis, and it will not recover the expression of Eq. (8) as ε goes to 0.

2.4. Bayesian Belief Networks

A Bayesian Belief Network (BBN) is a graphical network that represents probabilistic relationships among variables. BBNs enable reasoning under uncertainty. With BBNs, it is possible to articulate expert beliefs about the dependencies between different variables and to propagate consistently the impact of evidence on the probabilities of uncertain outcomes, such as 'future system reliability.'

A BBN is a special type of diagram (called a graph) together with an associated set of probability tables. The graph is made up of nodes and arcs where the nodes represent uncertain variables and the arcs the causal/relevance relationships between the variables. The main use of BBNs is in situations that require statistical inference: in addition to statements about the probabilities of events, the user knows some evidence, that is, some events that have actually been observed, and wishes to infer the probabilities of other events, which have not as yet been observed. A BBN uses conditional probability tables to calculate the probabilities of various possible causes being the actual cause of an event [Jensen, 2001].

A major benefit of Bayesian inference over 'classical statistical inference' (which deals with confidence levels rather than statements of probability) is that it explicitly describes the fact that observation alone cannot predict the probability of unobserved events. In the

Bayesian interpretation, a probability describes the strength of the belief which an observer can justifiably hold that a certain statement of fact is true (subjective probability). The subject, after observing the outcome of an 'experiment' (i.e., collecting new data), updates the belief held before the experiment (the 'prior probability'), producing a 'posterior probability'. The need to assume prior beliefs is a key part of Bayesian inference. The conditional probability tables must be filled in with reasonable estimates, and it is not always easy to obtain sensible prior probabilities, even from domain experts.

Until recently, the computation necessary to calculate the posterior probabilities for a BBN were quite difficult to implement, even for small problems. The general problem of performing such computations is known to be intractable (formally, it is known to be an NP-hard problem). The 1990s has seen the development of many tools which incorporate fairly efficient solution algorithms for BBNs. The computational aspect of calculating posterior distributions in discussed in more detail in the following section 2.5.

In specifying a Bayesian network, each node is given a corresponding probability distribution conditional on its "parent" nodes. A useful relation when analyzing the network with MCMC sampling (see Section 2.5) is that the full conditional density of any node is proportional to the product of all densities in the network which contain that particular variable.

2.5. Markov Chain Monte Carlo methods

Many times the calculation of the posterior density function involves complex integration. To calculate the posterior distribution for higher dimensions, some type of Monte Carlo method is often used to generate samples over which the integrand is calculated. A popular method for doing this is called Markov Chain Monte Carlo (MCMC), where one wants to generate a sampling density that is approximately equal to the posterior density.

The idea behind Monte Carlo Markov Chain is to construct a Markov Chain such that its stationary distribution is exactly the same as the distribution of interest. A stationary distribution of a Markov chain with transition probability matrix P(x,y) is *f* if:

$$f_Y(y) = \sum_x f_X(x) P(x,y)$$

for a discrete state chain. The continuous state equation relates the state of the system after n steps to the state of the system at n-1 steps:

$$f_Y^n(y) = \int_{-\infty}^{\infty} p(x,y)\, f_X^{n-1}(x) dx$$

The point of using MCMC methods is to generate a Markov Chain $\{X_0, X_1, X_2, \ldots\}$ where X_{k+1} only depends on X_k. Note that the Markov chain needs to reach stationarity to produce a useful posterior distribution. The distribution of X_k will approach a stationary form as *k* gets large, but in practice, one has to ignore the first *M* iterations. That is:

$$\frac{1}{n-M}\sum_{k=M+1}^{N} f(x) -> E[f(x)]$$

There are several methods for generating the Markov chain that has a stationary distribution with the properties of interest. Two of the most commonly used are the Metropolis algorithm and Gibbs sampling. The Metropolis algorithm can be used for both univariate and multivariate densities, while Gibbs sampling applies only to multivariate densities. For the purpose of analyzing Bayesian networks, the Metropolis algorithm with component transition kernels is often the most convenient (although this is very similar to Gibbs sampling).

The Metropolis algorithm relies on three basic steps:

1. Generate a candidate perturbation for the uncertain parameter of the current location within the Markov Chain.
2. Define a transition probability associated with the candidate move.
3. Accept the candidate move with a probability equal to the transition probability defined in Step 2.

Step 1 is accomplished with the use of what is known as a proposal density. The proposal density, denoted by $q(X^* \mid X)$, defines a probability density that generates random moves X^* based on the current point X. In theory, the only restriction on the choice of proposal density q is that it is symmetric with respect to its arguments, i.e. the probability of going from X to X^* is the same as that of going from X^* to X. An extension of this algorithm, known as the Metropolis-Hastings algorithm, allows the proposal density to have any form.

For Step 2 above, the Metropolis algorithm automatically defines the transition probability, often called the acceptance ratio, as $\alpha = \frac{f(X*)}{f(X)}$, where $f(\cdot)$ denotes the target density. Thus, the univariate Metropolis algorithm is given below:

Metropolis Algorithm
Set $i = 0$
Choose a starting value, X_0
Repeat until converged:

1. Sample a candidate X^* from the proposal density function $q(X^* \mid X_i)$
2. Calculate the acceptance ratio $\alpha = \min\left[1, \frac{f(X^*)}{f(X)}\right]$
3. Sample a uniform (0,1) random variable U
4. If $U < \alpha$, set $X_{i+1} = X^*$, else set $X_{i+1} = X_i$.
5. Increment i.

When using the Metropolis-Hastings algorithm, the acceptance ratio is calculated as $\alpha = \min\left[1, \frac{f(X^*)q(X \mid X^*)}{f(X)q(X^* \mid X)}\right]$, and it is not required that q is symmetric.

In general, the user must only specify the starting value and the proposal density, q. Unfortunately, the performance of the algorithm can be sensitive to both of these choices, particularly the choice of proposal density. The most commonly used proposal density is the random walk density, where the candidate point is given by $X^* = X + \xi$, where ξ is a random variable chosen to be symmetric about the origin. The choice of the variance of ξ is critical to the performance of the algorithm. If the moves are very small and the acceptance probability is very high, most moves will be accepted but the chain will take a large number of iterations to converge. If the moves are large, they are likely to fall in the tails of the posterior distribution and result in a low value of the acceptance ratio. One wants to cover the parameter space in a computationally efficient fashion. Many studies have been done on optimal acceptance rates, and the results seem to indicate that 0.45 - 0.5 is the optimal acceptance rate for 1-dimensional problems, whereas 0.23 - 0.25 is the optimal acceptance rate for high-dimensional problems [Gilks et al., 1997].

When the target density is multivariate, candidate moves can be made on all components simultaneously. Choosing a joint proposal density that is a good approximation to the target is often a difficult task, and the added complexity of working with multivariate densities often makes this method impractical. A component-wise scheme discussed by Hastings (1970) allows candidate moves to be made on each component independently. A proposal density is specified for each component of **X**, and the acceptance ratio for a particular move is given by $\alpha_i = \min\left[1, \frac{f(X_i^* \mid \mathbf{X}_{-i})}{f(X_i \mid \mathbf{X}_{-i})}\right]$, where $f(X_i \mid \mathbf{X}_{-i})$ denotes the full conditional density of the ith component. Thus, the components of **X** are sampled sequentially from their respective full conditional densities. This method is very similar to Gibbs sampling, and it is often referred to as Metropolis-within-Gibbs.

Recall that when the model is specified as a Bayesian network, there is a simple result stating that the full conditional density of any variable is proportional to all terms in the network containing that variable. This makes the component-wise Metropolis algorithm a natural choice when analyzing a Bayesian network with MCMC sampling.

Although it can be shown that the Markov chain generated by the Metropolis algorithm will converge to the target distribution for any symmetric proposal density (or any proposal density when using the Metropolis-Hastings algorithm), there is still the issue of how many samples it takes before the chain reaches the stationary distribution. Assessing convergence of an MCMC sampling scheme is a very difficult topic to which much work has been devoted. Although monitoring a function of one of the components (like the mean or variance) can be useful for diagnosing when the chain has not converged, it is not a rigorous method for confirming convergence. One of the most popular methods used in practice is that developed by Gelman and Rubin (1992). They propose running a small number of parallel sampling chains and comparing the across chain variance of a

given parameter to its within chain variances. A simpler, but less rigorous method is given by Geweke (1992), where the output from one chain is split into two parts, e.g. the first 10% and the final 50%. For stationarity, the means of the two samples must be the same. A modified z-test can also be used to compare the two samples. Another simple method often used in practice is to run a small number of parallel chains with overdispersed starting values and monitor how long it takes before the statistical properties of the chains are equivalent.

3. BBN Development and Implementation

3.1. Thermal Challenge Problem Description

The thermal validation challenge problem describes a candidate model for predicting the transient temperature in a slab for 1-D heat conduction. The model is a simple analytical expression, but it can be thought of as representing a possibly expensive finite element or other computer simulation. In addition to the model, a set of hypothetical experimental observations are described. The database of experimental observations is divided into two regimes: a validation regime and an application regime. Eight experiments were conducted in the validation regime, and one in the application regime. The hypothetical purpose of the model and experiments is to assess whether or not regulatory compliance will be achieved within the application regime; that is, within the application regime, the probability of exceeding a threshold temperature must be sufficiently low. This assessment must take into account certain uncertainties present in the model inputs as well as the experimental observations.

In addressing the challenge problem, three objectives are specified:

1. Determine the accuracy of the model for predicting the eight experimental outcomes within the validation regime.
2. Determine the accuracy of the model for predicting the single experimental outcome within the application regime.
3. Assess whether or not regulatory compliance will be achieved within the application regime.

Thus, the first two objectives fall into the category of model validation under uncertainty, and the third is related to prediction. The report addresses only the first two, and will employ the use of Bayesian networks (Section 2.4) in support of computing a Bayesian metric for model validation (Section 2.3).

3.2. Development of Bayesian Network Representation

The first step in addressing the challenge problem using the Bayesian framework is to formulate the Bayesian model, in this case the Bayesian belief network. This involves stating the random variables, describing the conditional relationships between them, and formulating prior probability distributions for each. There are three primary types of random variables associated with this particular problem: input parameters (material properties and boundary conditions), output responses (temperature), and experimentally

observed quantities. The input parameters are q (applied heat flux), k (thermal conductivity of the specimen), ρC (specific heat capacity of the specimen), and L (length of the specimen). The output response, T, is a function of time, and thus for discrete time intervals, T, can be considered as multivariate. For each experiment, measurements are taken on the applied heat flux and the temperature at various time intervals. The random variables associated with these measurements will be noted Dq and DT for data on q and data on T, respectively.

The governing partial differential equation for one-dimensional heat conduction through the cylinder in this problem is given by:

$$\frac{\partial}{\partial x}(k\frac{\partial T}{\partial x}) = \rho C_p \frac{\partial T}{\partial t} \qquad (0 < x < L)(t > 0)$$

where k and/or ρC_p are functions of temperature for the true model. These properties are constants for the approximate model to be tested. The boundary conditions are a prescribed constant flux on one surface and adiabatic on the other

$$k\frac{\partial T}{\partial x}\bigg|_{x=0} = q_o$$

$$\frac{\partial T}{\partial x}\bigg|_{x=L} = 0$$

A constant initial condition is prescribed

$$T\big|_{t=0} = T_i .$$

The relationship between the underlying random variables is shown by the network in Figure 1. The variable L, the length of the specimen, is not included in the model because for any particular experiment it is known with certainty (i.e., it is deterministic). Note that the variable q denotes the true applied heat flux, which has inherent randomness associated with experiment-to-experiment variability. For the data nodes, note the directionality is from the true value to the observation. Hence, the probability associated with the observation is conditional on the true value. A square is used to represent the node T because for a given set of inputs, the model gives a single, deterministic prediction of T. Hence, conditional on the inputs, T is not a random variable. For this formulation of the model, the network is specified for a particular experimental configuration (a particular value of L, a nominal value of q, and corresponding experimental observations of q and T). Developing a Bayesian model which accounts for all experimental configurations simultaneously is an area for future study.

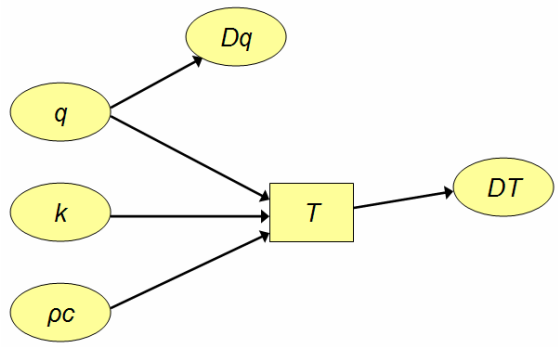

Figure 1. Bayesian network for primary random variables.

The challenge problem specifies two probabilistic components associated with the measurement error of q and T: a "random" component and a "bias" component. Our interpretation of these is as follows: The "random" component describes the probability distribution of the measurement error, ε, based on a particular "target" value of the quantity being measured. The "target" value of the measured quantity is actually equal to the true value plus a random "bias" component. The bias component is assumed to be uniformly distributed on the interval specified. Thus, two additional nodes are added to the network in Figure 1 to account for the bias components. Triangles are used to denote that these bias components are given fixed probability distributions, and thus they are not updated by the addition of experimental observations. The resulting network is given in Figure 2.

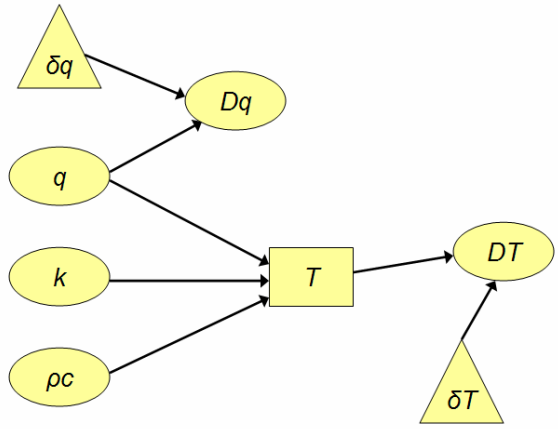

Figure 2. Complete Bayesian network for the challenge problem.

The Bayes' factor proposed by Mahadevan and Rebba, as given in Eq. (8), is used for the purpose of assessing the agreement between the model prediction and the experimental observations. Thus, the probability distribution for the alternative hypothesis, $g(T)$, is constructed by propagating the uncertainties on q, k, and ρC through the simulation model.

Because temperature is a function of time, and experimental observations are made at multiple time instances, the random variable T is in a sense multivariate (a collection of correlated random variables). Evaluating Eq. (8) by discretizing T in accordance with the

experimental observations would require working with the joint probability density function of T. The complexity of such an analysis is beyond the scope of this paper (see for example, Srivastava, 2002). Thus, for illustration, Eq. (8) will be evaluated at one time instance. A similar comparison could be made for each time instance which has a corresponding experimental observation. Such an analysis neglects the correlation structure of T. However, for this particular problem, the correlation structures of both the experimental observations and the model predictions are essentially identical, indicating that univariate comparisons are sufficient.

The formulations of the prior probability distributions for each of the random variables in the network of Figure 2 will now be discussed. Recall that the network corresponds to one particular experimental configuration. We will consider configuration 1, experiment 1. The nominal value of q is 1000 W/m^2. Thus, we assume a prior distribution for q as $f(q) \sim N(q_{nom}, \sigma_q)$, which states that the true value of applied heat flux follows a normal distribution with a mean given by the nominal value (1000 W/m^2) and a standard deviation of σ_q. Specification of σ_q is based somewhat on the stated observations of q, but it is still a subjective decision. Here, σ_q is specified as 3% of the nominal value (30 W/m^2). A more complete model might treat σ_q as a random variable with a suitably vague prior distribution.

The prior distributions for the material properties, k and ρC, are developed based on the material characterization experiment. The material characterization experiment consists of six observations of each property. Two specimens were tested at three different temperatures. The random variables for each property are assumed to follow a normal distribution with parameters estimated for the results of the characterization experiment. Thus, we again assume normal distributions and specify the priors for k and ρC as: $f(k) \sim N(0.0639, 0.0087)$ W/mC and $f(\rho C) \sim N(4.2\times10^5, 0.172\times10^5)$ J/m^{3}C. The purpose of these probability distributions is to account for the unit to unit variability present in the slabs being tested, as well as the uncertainty derived from the temperature-dependence of these material properties.

The prior distributions for the experimental observations are formulated based on the specified experimental uncertainties. In accordance with the previously discussed interpretation of the "random" and "bias" measurement error components, the probability distributions for Dq and DT are given by: $f(Dq \mid q, \delta q) \sim N(q + \delta q, 0.015 q_{nom})$ and $f(DT \mid T, \delta T) \sim N(T + \delta T, 0.005(T - 25))$.

3.3. Markov Chain Monte Carlo Analysis

Because the Bayesian model we are dealing with involves several correlated variables, direct calculation of the posterior probability densities is not practical. For this reason, a Markov Chain Monte Carlo sampling scheme for computing the posterior probability distributions is needed. Recall from Sections 2.4 and 2.5 that the componentwise Metropolis algorithm provides a convenient method for analyzing a Bayesian network.

We first specify the full conditional densities for each of the random variables that we are sampling from. We now apply the result from Section 2.4 that the full conditional density of any variable is proportional to all terms in the network containing that variable. Because T is a deterministic function of q, k, and ρC, we include terms containing T in the full conditional densities of its inputs. Thus, we have:

$$f(q|...) \propto f(q) \times f(Dq|q,\delta q) \times f(DT|T,\delta T)$$
$$f(k|...) \propto f(k) \times f(DT|T,\delta T)$$
$$f(\rho C|...) \propto f(\rho C) \times f(DT|T,\delta T)$$

Next, we specify the proposal density for each component. Random walk proposal densities are used for each. The noise terms, ξ, were chosen to be normally distributed with means of 0. The variance of each noise term was adjusted to provide an average acceptance rate for that variable of approximately 0.5. The componentwise Metropolis algorithm was then implemented as described in Section 2.5. As an example, consider an updating step for the variable k:

Generate a candidate perturbation, ξ_k
The candidate value is $k^* = k_i + \xi_k$
Calculate the acceptance ratio:

$$\alpha = \frac{f(k^*|...)}{f(k_i|...)} = \frac{f(k^*) \times f(DT|T^*,\delta T)}{f(k) \times f(DT|T_i,\delta T)}$$
$$= \frac{\phi(k^*|0.0639,.00837) \times \phi(DT|T^* + \delta T, 0.005(T^* - 25))}{\phi(k|0.0639,.00837) \times \phi(DT|T_i + \delta T, 0.005(T_i - 25))}$$

where $\phi(x|\mu,\sigma)$ denotes the normal PDF evaluated at x with mean μ and standard deviation σ, and T^* is the valued given by the simulation model with the input k^*

Sample a uniform (0,1) random variable u
If $u < \alpha$,
$k_{i+1} = k^*$
T = Model(q_i, ρC_i, k_{i+1})
(T_{i+1} is only updated after a complete scan of all other variables)
Else $k_{i+1} = k_i$

Note that each time one of the input random variables is updated, a new value for T must also be generated. This is done by evaluating the simulation model to compute T whenever one of the input parameters changes. However, sample draws of T are only stored each time the algorithm has parsed all of the random variables.

The above process is repeated for each component to draw a vector of samples. In order to approximate the posterior probability distributions, a large number of samples must be drawn from each component. For this study, 50,000 samples were generated for each variable. The first 1,000 samples were discarded to deemphasize the effect of the starting values (in other words, to allow the chain to reach its stationary distribution).

4. Results

For the purpose of illustration, first consider the results for configuration 1, experiment 1, with temperature predictions compared at time = 100s. A graphical representation of the uncertainties associated with the model predictions and experimental observations of temperature for this particular configuration is given in Figure 3. The curves representing the model predictions were obtained by propagating the uncertainties on q, k, and ρC through the model. For this plot the heat flux, q, was assumed to be normally distributed with a mean given by the experimental observation and a variance equal to the total variance specified by the experimental uncertainty. Note that this probability distribution for q is not the same as the one used in the Bayesian network. Also note that Figure 3 is not the result of a Bayesian analysis, and is only given for the purpose of illustrating the uncertainties present in the problem. A close-up of the uncertainty associated with the observation of T at 100 seconds is given in Figure 4.

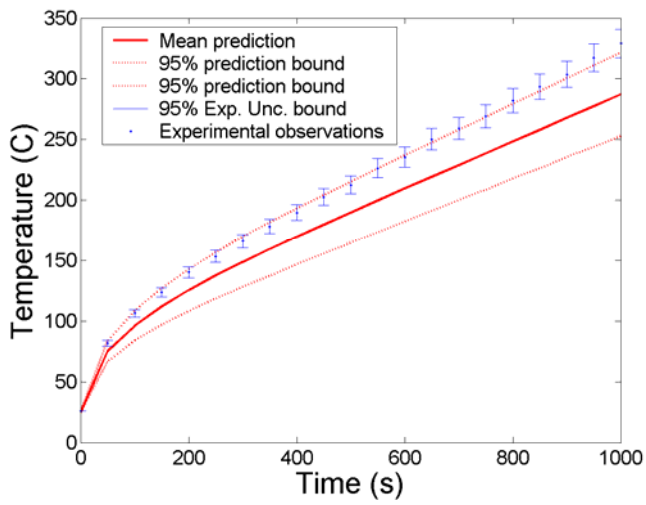

Figure 3. Uncertainties for configuration 1, experiment 1.

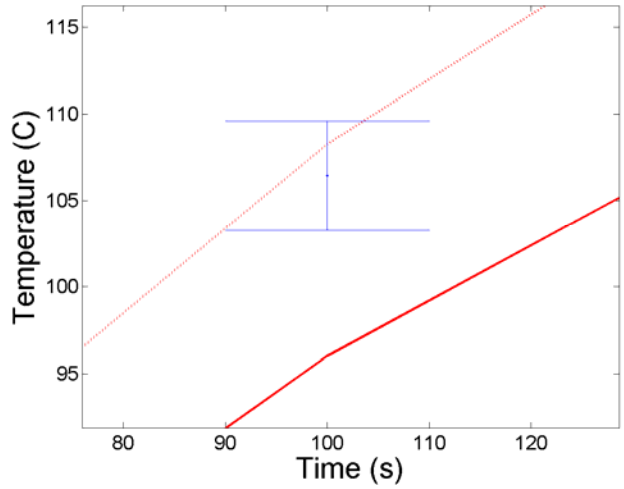

Figure 4. Uncertainty for configuration 1, experiment 1, at 100 seconds.

The graphical depiction of measurement uncertainty in Figures 3 and 4 should be interpreted very loosely. The standard deviation used to compute the 95% confidence bounds on the observations of temperature was calculated as the sum of the standard deviation of the "random" and "bias" components. The standard deviation for the "bias" component came from the assumed uniform distribution. Simply adding these two uncertainties is not correct, and as seen in the following Bayesian analysis, this is a significant overestimate of total measurement uncertainty.

The model assessment corresponding to the observation shown in Figure 4 was carried out using the methodology described in Section 3. The objective is to compare the posterior distribution of T to the prior distribution, in order to compute the Bayes' factor as described in Section 2.3. Thus, we calculate the Bayes' factor using Eq. (8), where the numerator is approximated using the Metropolis algorithm, and the denominator is approximated using an uncertainty quantification where the distributions for the input parameters are as specified in Section 3.2. The two distributions are compared at the model prediction, i.e. T = 95.5 C. Note that the corresponding experimental observation is T = 106.43 C which is not close to the model prediction with respect to measurement uncertainty.

The approximated densities for the prior and posterior distributions of T are plotted in Figure 5. Note that the Bayes' factor is approximately equal to 0 because the posterior density is approximately equal to 0 at the temperature predicted by the model. This indicates that given the specified uncertainties, the observations of T and q do not support the corresponding model prediction of T = 95.5 C.

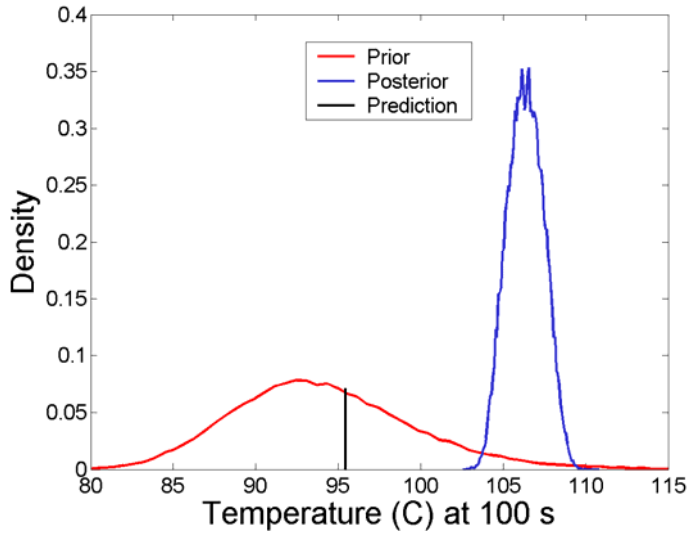

Figure 5. Prior and posterior distributions for config. 1, exp. 1, at 100 seconds.

As a second example, now consider a case where the model prediction and the experimental observation are closer. For configuration 2, experiment 2, the model predicts T = 91.0 C at t = 100s, and the corresponding experimental observation is T = 93.0 C. The approximated prior and posterior densities are shown in Figure 6. The Bayes' factor of Eq. (8) is approximately 0.644, indicating insufficient support for the point null hypothesis that the model prediction is correct. Although the prediction is close to the observation for this case, the Bayes' factor does not accept the model because the measurement uncertainty for temperature is so low. Without considering the ambiguous "bias" term, the standard deviation of measurement error for this observation is 0.34 C.

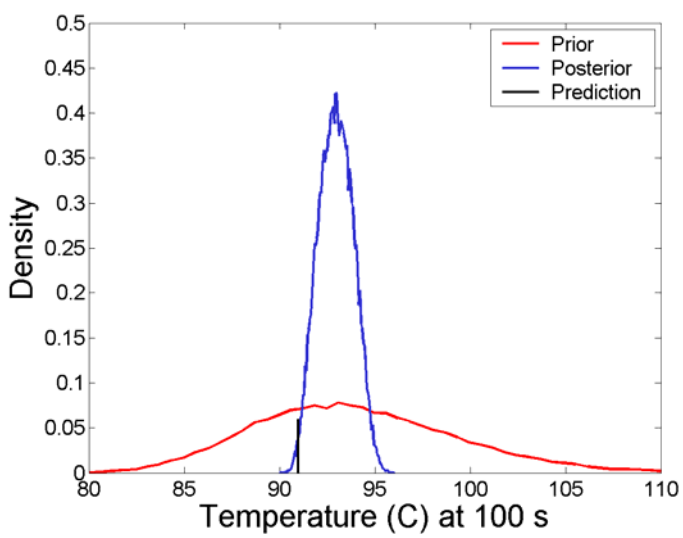

Figure 6. Prior and posterior distributions for config. 2, exp. 2.

Finally, consider the effect of altering the prior probability distribution of T. Note, that the prediction of temperature is very sensitive to the values of the material properties, k, and ρC. Thus, for illustration, the effect of increasing the uncertainty on these variables is considered. The prior standard deviations for both k, and ρC are increased by a factor of 1.5, and we again consider configuration 2, experiment 2. The resulting distributions on T are shown in Figure 7, and the Bayes' factor is estimated as 1.18, indicating a small degree of support for the model prediction. The primary effect is that the prior distribution of T has become flatter, or more diffuse, while the posterior has remained essentially the same.

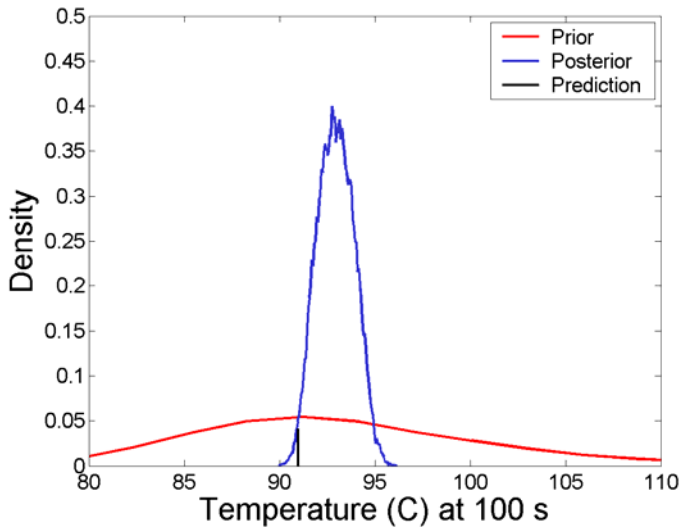

Figure 7. Effect of increased prediction uncertainty for config. 2, exp. 2.

This result is related to the discussion of Section 2.3 regarding hypothesis testing with non-informative prior information. The null hypothesis will always be accepted at some level, as the prior becomes more and more diffuse. This is because the Bayesian metric is taking into account uncertainty associated with the model prediction. However, if the uncertainty associated with the model prediction is extremely high, attempting to validate it with experimental observations makes little sense. Thus, the Bayes' factor metric is sensitive to the specified prediction uncertainty. When applying this metric, if a model is accepted, it would be wise to confirm that it is not simply because the prediction uncertainty is "too high".

As previously discussed, the methodology outlined in Section 3 constrains the analysis to deal with only one experimental setup at a time. Likewise, for a particular experiment, a univariate comparison methodology was chosen, so that the model predictions of temperature are compared to the experimental observations at each time step independently. Repeated applications of this model validation methodology would give a spectrum of validation assessments corresponding to various conditions. Clearly, an approach which allows for all experiments and all predictions to be compared

simultaneously is desirable. Such an approach is within the capabilities of the Bayesian framework, and is an area for future study.

5. Summary

The purpose of this report is to introduce and demonstrate a candidate metric for assessing the predictive capability of a model based on experimental observations (model validation). It was shown that the statistical metric based on Bayesian hypothesis testing can provide a quantitative assessment of the candidate predictive model, while accounting for both model parameter uncertainty and experimental uncertainty. The method was demonstrated on the Sandia thermal challenge problem by assessing the agreement between one particular experimental observation and the corresponding model prediction.

There are several weaknesses with the methodology implemented in Section 3. First of all, the Bayesian model formulation of Figures 1 and 2 is specific to a particular experimental configuration. That is, the formulation of the model is specific to a particular nominal value of applied heat flux and a particular value for the length of the test specimen. Changing either of these values, as is the case with each of the experimental setups described by the challenge problem, requires altering the Bayesian network. The result is that the model assessment given by the Bayes' factor of Eq. (8) must be computed independently for each experimental observation corresponding to a different experimental setup.

The second weakness of the methodology is that because no other information is available, the alternate hypothesis must be constructed based on the simulation model. Recall that the null hypothesis is the point hypothesis that the true temperature is equal to the value predicted by the model, corresponding to the input conditions of the experiment. The alternative hypothesis is that the temperature has the probability distribution given by a propagation of the input uncertainties through the simulation model (the red curve in Figure 5). However, choosing an appropriate alternative hypothesis is the typical difficulty when assessing the predictive capabilities of a single model, and it may be the case that no better choices are available. Also, because evaluation of the simulation model was built into the formulation of the Bayesian network, computation of the numerator in Eq. (8) requires numerous evaluations of the simulation model (see Section 3.3). For many real world applications, this may be impractical due to the high cost associated with evaluations of the simulation model. However, a surrogate model could be used to reduce the number of evaluations of the simulation model.

The third weakness is the use of a point null hypothesis. It is often argued that the use of a point hypothesis is not warranted because it may cause the model to be rejected when in fact the prediction is close enough to the experimental observation for practical purposes. This may be true, and the formulation of an interval hypothesis for Bayesian hypothesis testing is given in Section 2.3.

Future work in the area of Bayesian model validation may want to focus on some of the issues discussed above. Particularly, a different formulation of the null and alternate hypotheses may allow for a metric which gives a model assessment based on all experimental data. For example, such a formulation might involve building a relatively objective Bayesian model of the experimental data. A nonlinear model could be proposed for T, in which the coefficients are given suitably vague probability distributions. A hierarchical structure might assign several levels of uncertainty to the model to increase the level of objectivity. The model would then be "updated" based on the experimental observations, and compared against the simulation model for the purpose of validation. Such a formulation may also reduce the number of evaluations of the simulation model required to compute the validation metric.

6. References

Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, 1985.

Dowding, K. J., M. Pilch, and R. G. Hills. *Thermal Validation Challenge Problem*. SAND XXX-2005.

Gamerman, D. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall/CRC, Boca Raton, 1997.

Gelman, A. J., B. Carlin, H. S. Stern and D. B. Rubin. *Bayesian Data Analysis*, 2nd edition. Chapman and Hall/CRC, Boca Raton, 2004.

Gelman, A. J. and D. Rubin. "Inference from iterative simulation using multiple sequences (with discussion)," *Statistical Science*. **7**(1992):457-511.

Geweke, J. "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." In: *Bayesian Statistics 4*, Oxford University Press, Oxford, 1992, pp. 169-193.

Gilks, W. R., S. Richardson, and D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice.* Chapman and Hall/CRC, Boca Raton, 1996.

Hastings, W. K. "Monte Carlo sampling methods using Markov chains and their applications," *Biometrica*. **57**(1970): 97-109.

Jensen, F. V. *Bayesian Networks and Decision Graphs*, Statistics for Engineering and Information Science Series, Springer-Verlag, 2001.

Mahadevan, S. and R. Rebba. "Validation of Reliability Computational Models using Bayes Networks," *Reliability Engineering and System Safety*. **87**(2005): 223-232.

Press, S. J. *Bayesian Statistics: Principles, Models, and Applications.* Wiley, 1989.

Press, S. J. *Subjective and Objective Bayesian Statistics: Principles, Methods and Applications*, 2nd edition, Wiley, New York, 2003.

Robert, C. P. *The Bayesian Choice*, 2nd ed. Springer-Verlag, New York, 2001.

Srivastava, M. S. *Methods of Multivariate Statistics*. John Wiley, New York, 2002.

7. Distribution

1	MS1415	Normand Modine	01112
1	MS1056	Sam Myers	01112
1	MS1167	Kevin Horn	01343
1	MS0370	Michael Eldred	01411
1	MS0370	Scott Mitchell	01411
4	MS0370	Laura P. Swiler	01411
1	MS0370	Timothy Trucano	01411
1	MS0828	Kevin Dowding	01533
1	MS0828	Anthony Giunta	01533
1	MS0828	Richard G. Hills	01533
1	MS0779	Jon Helton	01533
1	MS0828	William Oberkampf	01533
1	MS0828	Martin Pilch	01533
1	MS0828	Vicente Romero	01533
1	MS9159	Monica Martinez-Canales	08962
1	MS0748	David G. Robinson	06861
1	MS0829	Floyd Spencer	12337
1	MS0123	LDRD Office	01011
1	MS9018	Central Technical File	08945-1
2	MS0899	Technical Library	09616
1	MS0612	Review & Approval Desk for DOE/OSTI	09612