Available online at www.sciencedirect.com
ScienceDirect
Reliability Engineering and System Safety 92 (2007) 707–718
RELIABILITY ENGINEERING & SYSTEM SAFETY
www.elsevier.com/locate/ress

Bayesian risk-based decision method for model validation under uncertainty

Xiaomo Jiang, Sankaran Mahadevan*

Box 1831-B, Vanderbilt University, Nashville, TN, 37235, USA

Received 9 November 2005; received in revised form 6 March 2006; accepted 23 March 2006
Available online 23 May 2006

Abstract

This paper develops a decision-making methodology for computational model validation, considering the risk of using the current model, data support for the current model, and cost of acquiring new information to improve the model. A Bayesian decision theory-based method is developed for this purpose, using a likelihood ratio as the validation metric for model assessment. An expected risk or cost function is defined as a function of the decision costs, and the likelihood and prior of each hypothesis. The risk is minimized through correctly assigning experimental data to two decision regions based on the comparison of the likelihood ratio with a decision threshold. A Bayesian validation metric is derived based on the risk minimization criterion. Two types of validation tests are considered: pass/fail tests and system response value measurement tests. The methodology is illustrated for the validation of reliability prediction models in a tension bar and an engine blade subjected to high cycle fatigue. The proposed method can effectively integrate optimal experimental design into model validation to simultaneously reduce the cost and improve the accuracy of reliability model assessment.

Keywords: Bayesian risk; Decision making; Bayesian statistics; Model validation; Reliability prediction

1. Introduction

A computational model is often designed to represent a physical system under specific assumptions. Before this mathematical model can be used for real world applications, model verification and validation must be conducted to demonstrate its correctness and to quantify the confidence of how accurately it represents the real system. Generally, the validation process involves comparing the model prediction with measurements in a real system, theoretical results, expert intuition, or any combination of them. During the comparison, decisions need to be made by decision makers with certain preferences, based upon the available information and prior knowledge. Ignoring this issue may result in costly and inaccurate model validation. The focus of this study is to develop a decision theoretic method for the effective assessment of computational models.

Two key challenges in the validation process are the design of validation experiments and the definition of validation metrics [1]. Under the constraints of time, money, and other resources, validation experiments often need to be optimally designed for a clearly defined purpose, namely computational model assessment. This is inherently a decision theoretic problem where a utility function needs to be first defined so that the data collected from the experiment provides the greatest opportunity for performing conclusive comparisons in model validation. Recently, a Bayesian cross entropy-based methodology has been developed by Jiang and Mahadevan [2] to achieve this objective.

Given the experimental observations, the definition of a proper validation metric becomes another critical challenge of model validation under uncertainty. A validation metric is a measure of agreement between model predictions and experimental observations. Currently the widely used method is the graphical validation through visually comparing graphs of prediction and observation. This is a qualitative approach where the numerical error or quantitative uncertainties in both the predictions and

*Corresponding author. Tel.: 615 322 3040; fax: 615 322 3365.
E-mail addresses: xiaomo.jiang@vanderbilt.edu (X. Jiang), sankaran.mahadevan@vanderbilt.edu (S. Mahadevan).

0951-8320/$ - see front matter
doi:10.1016/j.ress.2006.03.006

experimental results cannot be considered. In addition, such a comparison provides little indication of how the agreement varies over space and time. Generally, two types of quantitative approaches can be pursued to develop quantitative model validation metrics: hypothesis testing-based and decision-based methods. Hypothesis testing-based methods include classical and Bayesian approaches. The classical hypothesis testing approach is a well-developed statistical method of accepting or rejecting a model based on an error statistic. It has been argued that the classical hypothesis testing methods may be difficult to interpret and sometimes misleading [3,4].

Recently, the Vanderbilt team has developed a Bayesian hypothesis testing-based model validation methodology [5,6]. One important difference between the two approaches is that, the Bayesian approach focuses on model acceptance whereas classical hypothesis testing focuses on model rejection. In the latter, not having enough evidence to reject a model is not the same as having enough evidence to accept the model. The differences between classical and Bayesian hypothesis testing have been discussed in detail by many researchers (e.g., [7,8]).

This paper pursues a decision-based approach for model validation. Let d_0 denote a decision to accept the null hypothesis $H_0 : y = y_0$, where y_0 and y are, respectively, the predicted and actual values of a physical quantity of interest, and d_1 denote a decision to accept the alternative hypothesis $H_1 : y \neq y_0$. The utility function u(d_i, y) is defined to choose d_i based on a decision rule. Given the observed data Y, a decision d_0 is made if and only if $E[u(d_0, y) - u(d_1, y)|Y] > 0$. The decision-theoretic approach has been applied to structural reliability analysis [9,10]. These studies treated model selection in structural reliability analysis as a decision-making problem, with the purpose of minimizing the risk of overlooking important prior information and engineers' preferences. However, literature reviews demonstrate that this approach has not yet been widely pursued for computational model validation, which is the focus of this study.

Decision-based validation methods may also be approached through classical or Bayesian statistics. In the classical approach, given the experimental observations, a hypothesis testing is conducted in terms of the conditional probabilities of Type I error (reject a correct model) and Type II error (accept a wrong model). An expected loss function is defined based on the conditional error probabilities. Either a squared loss function or an absolute error metric is usually chosen as the difference in the loss function [11]. The task of a decision is to minimize the expected loss. Balci and Sargent [12] presented a classical hypothesis testing-based cost-risk decision analysis to validate a simulation model of a real system, considering the model user's risk, model builder's risk, acceptable validity range, budget, sample sizes, and cost of data collection. It should be noted that, similar to the classical hypothesis testing methods, the classical decision-based approach may also be misleading and has difficulties in properly interpreting the error probabilities.

In the Bayesian approach, the task of deciding between H_0 and H_1 is conceptually more straightforward. One merely needs to calculate the posterior probabilities $\alpha_0 = Pr(H_0|Y)$ and $\alpha_1 = Pr(H_1|Y)$, or their likelihoods given experimental data. An expected loss function is then defined as a function of α_0 and α_1. Its conceptual advantage is that α_0 and α_1 are the actual probabilities of the hypotheses in light of the observed data and prior knowledge. The Bayes decision approach is therefore pursued in this study for model validation.

An important issue in Bayesian decision theory is to select a proper decision rule for model assessment. There are four commonly-used decision rules, namely, maximum likelihood (ML), minimum probability of error (MPE), maximum a posteriori (MAP), and Bayes risk decision criteria. The ML criterion is the simplest decision rule, which is based merely on the conditional probabilities of the quantity of interest given each of hypotheses, i.e. $Pr(Y|H_0)$ and $Pr(Y|H_1)$. A decision is made to accept H_0 if $Pr(Y|H_0) > Pr(Y|H_1)$, and conversely. The ML criterion makes use of only the conditional probabilities of the quantity of interest and ignores the prior information about H_i.

In the MPE criterion, a decision *region*, which encompasses all experimental data to favor this decision, is selected if it minimizes the total probability of error resulting from two incorrect decisions

$$Pr(error) = Pr(H_0)Pr(H_1|H_0) + Pr(H_1)Pr(H_0|H_1), \quad (1)$$

where $Pr(H_i)$ $(i = 0,\ 1)$ is a prior assumed to the ith hypothesis occurring, $\Pr[H_1|H_0]$ is the probability of deciding H_1 when H_0 is true (type I error), and $\Pr[H_0|H_1]$ is the probability of deciding H_0 when H_1 is true (Type II error). The prior information is utilized in the computation of the conditional probabilities of error.

In the MAP criterion, a decision is made to accept H_0 if and only if $Pr(H_0|Y) > Pr(H_1|Y)$. Based on Bayes theorem, $Pr(H_i|Y) \propto Pr(Y|H_i)Pr(H_i)$, it is easy to derive that the MAP is equivalent to the MPE criterion given validation data Y. Note that the MAP is also identical to the ML criterion if $Pr(H_0) = Pr(H_1)$ is assumed.

The main shortcoming of the above three decision criteria, however, is that they do not account for the consequence or risk of making a wrong decision. The Bayes risk criterion may be a proper alternative to overcome this drawback, which will be developed in this paper for the assessment of computational models. In the Bayes risk criterion, given the experimental data, the ratio of likelihoods of the null and alternative hypotheses is compared with a decision threshold which is a function of the cost information and the priors of two hypotheses. The experimental data is assigned to any of two decision regions (one is to favor the model and another to reject the model), whenever it reduces the total risk (cost) in model validation. The challenge of the Bayes risk approach is to compute the likelihood probability of each hypothesis given experimental data available on one or more intermediate quantities. This difficulty may be overcome by using a Bayes network approach.

The purpose of this paper is to develop a Bayesian decision risk-based method for effective assessment of computational models. The Bayes factor, which is the likelihood ratio of the null and alternative hypotheses, is computed as the validation metric. Then the Bayes risk criterion is computed based on the metric and used as a decision rule during the model validation. In order to compute the likelihood of each hypothesis, the Bayesian network [13] and the Marko chain Monte Carlo (MCMC) technique are utilized to update the distributions of the intermediate and final quantities given validation experimental data on one or more nodes. An expected risk or cost function is defined as a function of the decision costs, and the likelihood and prior of each hypothesis. The minimum risk is obtained through correctly assigning the experimental data to decision regions based on the comparison of the likelihood ratio with a decision threshold. The methodology is implemented for measured data obtained from two types of tests: multiple pass/fail and system response value measurement tests, and is illustrated for the validation of reliability prediction models in a tension bar and an engine blade subjected to high cycle fatigue.

2. Bayesian Decision-Based Validation Method

Within the context of binary hypothesis testing, i.e., the point null hypothesis $\left(H_0 : y = y_0\right)$ accepts the model vs. an alternative hypothesis $\left(H_1 : y \neq y_0\right)$ rejects the model. Each time a validation experiment is conducted, one of the four possible scenarios, $\Pr[H_i|H_j]$ $(i = 0,\ 1;\ j = 0,\ 1)$, may happen. Thus, the loss function for model validation, referred to as Bayes risk R, is defined to be the expected cost of validation experiments, which is obtained by averaging the decision cost over two probabilities: the prior probability of the hypothesis and the probability of a particular action to be taken [14]:

$$R = \sum_{j=0}^{1}\sum_{i=0}^{1}\left(c_{ij}\pi_j Pr[H_i|H_j]\right), \tag{2}$$

where $\pi_j = Pr[H_j]$ $(j = 0,\ 1)$ represents the prior probability of each hypothesis (note that $\pi_1 = 1 - \pi_0$ for this binary hypothesis testing problem), $\Pr[H_i|H_j]$ = the probability of deciding H_i when H_j is true, and c_{ij} = the cost of deciding H_i when H_j is true (decision consequence).

Let Z represent the entire experimental data set, and Z_0 and Z_1 represent two mutually exclusive subsets of Z such that $Z_0 \cup Z_1 = Z$, and $Z_0 \cap Z_1 = \varnothing$, where $\cup$ and $\cap$ represent union and intersection, respectively, and $\varnothing$ represents an empty set or null space. Thus, Z_0 and Z_1 represent two decision regions corresponding to two hypotheses H_0 and H_1. Every possible experimental output Y belongs to either decision region, and the problem to be solved here is the assignment of each experiment to Z_0 or Z_1. Fig. 1 shows a schematic illustration of the decision-based validation method.

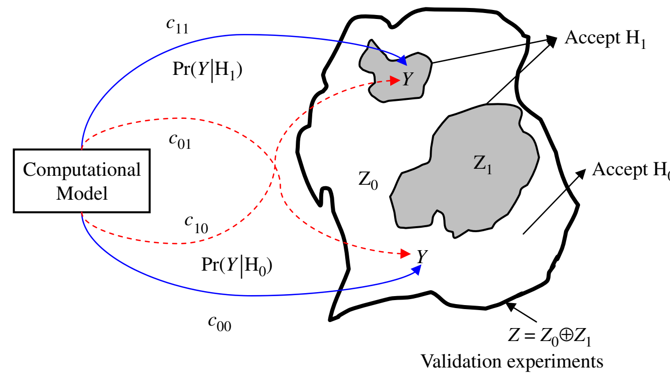

Fig. 1. Schematic illustration of Bayesian decision-based model validation methodology.

2.1. Bayes risk rule

Assuming that the experimental output Y has a probability density function (or likelihood function) under each hypothesis, i.e., $Y|H_0 \sim f(y|H_0)$ and $Y|H_1 \sim f(y|H_1)$. Thus, $Pr[H_i|H_j] = \int_{Z_i} f(y|H_j)\,\mathrm{d}y$ $(i = 0,\ 1;\ j = 0,\ 1)$, and the Bayes risk in Eq. (2) becomes

$$R = \sum_{j=0}^{1}\sum_{i=0}^{1}\left(c_{ij}\pi_j \int_{Z_i} f(y|H_j)\,\mathrm{d}y\right). \tag{3}$$

That is,

$$R = \int_{Z_0}[c_{00}\pi_0 f(y|H_0) + c_{01}\pi_1 f(y|H_1)]\,\mathrm{d}y + \int_{Z_1}[c_{10}\pi_0 f(y|H_0) + c_{11}\pi_1 f(y|H_1))]\,\mathrm{d}y. \tag{4}$$

In Eq. (4), the terms $\int_{Z_0} f(y|H_0)\,\mathrm{d}y$ $(i = 0,\ j = 0)$ and $\int_{Z_1} f(y|H_1)\,\mathrm{d}y$ $(i = 1,\ j = 1)$ are the probabilities of correctly accepting and rejecting the model, respectively, and may be denoted as P_{A} and P_{R}. The terms $\int_{Z_1} f(y|H_0)\,\mathrm{d}y$ $(i = 1,\ j = 0)$ and $\int_{Z_0} f(y|H_1)\,\mathrm{d}y$ $(i = 0,\ j = 1)$ are the probabilities of Type I error (reject a correct model) and Type II error (accept a wrong model), respectively, and may be denoted as P_{I} and P_{II}. Therefore, the decision problem is equivalent to assigning experimental data to the corresponding decision regions Z_0 and Z_1 such that the risk (cost) will be minimized.

In the entire experimental data space Z, observing that

$$\int_Z f(y|H_0)\,\mathrm{d}y = \int_Z f(y|H_1)\,\mathrm{d}y = 1 \tag{5}$$

and

$$\int_{Z_0} f(y|H_i)\,\mathrm{d}y + \int_{Z_1} f(y|H_i)\,\mathrm{d}y = \int_Z f(y|H_i)\,\mathrm{d}y, \quad i = 0 \text{ or } 1. \tag{6}$$

Using Eqs. (5) and (6), and due to the relationships $Z_0 \cup Z_1 = Z$ and $Z_0 \cap Z_1 = \emptyset$, Eq. (4) is rewritten as

$$R = c_{10}\pi_0 + c_{11}\pi_1 + \int_{Z_0} [\pi_1(c_{01} - c_{11})f(y|H_1) - \pi_0(c_{10} - c_{00})f(y|H_0)]\,\mathrm{d}y. \tag{7}$$

The integral in Eq. (7) represents the risk resulting from assigning some of experimental data points to the decision region Z_0. In order to minimize the Bayes risk, only the appropriate data points should be assigned to the decision region Z_0 such that the corresponding integrand in Eq. (7) is minimized. It is assumed that the total risk (cost) resulting from a correct decision is always less than the total risk resulting from a wrong decision. Therefore, the experimental data Y should be assigned to Z_0 (accept the model) whenever

$$c_{00}\pi_0 f(y|H_0) + c_{01}\pi_1 f(y|H_1) < c_{10}\pi_0 f(y|H_0) + c_{11}\pi_1 f(y|H_1). \tag{8}$$

That is,

$$\Lambda(y) = \frac{f(y|H_0)}{f(y|H_1)} > \frac{\pi_1(c_{01} - c_{11})}{\pi_0(c_{10} - c_{00})} = \eta, \tag{9}$$

where $\Lambda(y)$ is the likelihood ratio, referred to as Bayes factor, and η is the acceptable threshold for the Bayes factor. The threshold η can be related to the prior densities of the two hypotheses and the costs of deciding H_i when H_j is true ($i = 0, 1$; $j = 0, 1$). It is a lower bound for accepting the model, and an upper bound for rejecting the model. From Eq. (8), the relationship $\pi_0(c_{10} - c_{00})f(y|H_0) > \pi_1(c_{01} - c_{11})f(y|H_1)$ is obtained. This yields a negative integration in Eq. (7) and thus leads to minimum risk.

Following the above procedure for deciding the region Z_0, it is easy to derive that, in the case of $\Lambda(y) < \eta$, the Bayes risk is minimized only when the experimental data Y is assigned to Z_1 (reject the model). The Bayes risk in Eq. (4) can be rewritten as

$$R = c_{00}\pi_0 + c_{01}\pi_1 + \int_{Z_1} [\pi_0(c_{10} - c_{00})f(y|H_0) - \pi_1(c_{01} - c_{11})f(y|H_1)]\,\mathrm{d}y. \tag{10}$$

From practical considerations, it may be assumed that the cost of a wrong decision is higher than that of a correct decision, i.e., $c_{10} > c_{00}$ and $c_{01} > c_{11}$. This gives two special cases in the Bayes risk approach. First, assume that $c_{10} = c_{01} = 1$, $c_{00} = c_{11} = 0$. The risk in Eq. (4) becomes

$$R = \int_{Z_1} \pi_0 f(y|H_0)\,\mathrm{d}y + \int_{Z_0} \pi_1 f(y|H_1)\,\mathrm{d}y = P_{\mathrm{I}} + P_{\mathrm{II}}. \tag{11}$$

which is equivalent to Eq. (1). In this case, minimizing Bayes risk is identical to minimizing the sum of the probabilities of the two incorrect decisions. Based on Eqs. (5) and (6), Eq. (11) becomes

$$\mathrm{R} = \pi_0 + \pi_1 - \left(\int_{Z_0} \pi_0 f(y|H_0)\,\mathrm{d}y + \int_{Z_1} \pi_1 f(y|H_1)\,\mathrm{d}y \right). \tag{12}$$

Therefore, minimizing Bayes risk in this case is also identical to maximizing the sum of the probabilities of the two correct decisions (i.e., MAP).

Second, besides the cost assumption, assume further that $\pi_0 = \pi_1 = 0.5$ in the absence of prior knowledge of each hypothesis. Then, $\eta = 1$ and the experiment data Y is said to favor the model if and only if $\Lambda(y) > 1$. This is the specific case of the Bayes factor-based model validation method developed by Zhang and Mahadevan [5], and is also identical to the maximum likelihood decision criterion. Clearly, the Bayes risk rule is a comprehensive criterion which incorporates several other decision criteria as special cases.

In practical applications of the Bayes risk approach for model validation, it becomes critical to efficiently compute the likelihoods of each hypothesis given experimental data available on one or more intermediate quantities. This purpose can be achieved by using the Bayes network [13] approach and a MCMC simulation technique. The Bayes network is used because it can provide a straightforward updating process, where the computational model is represented by a network showing the relations among the basic variables and the overall predicted output, which are represented by nodes in the network. Once experimental data is available on any one or more nodes, the posterior statistics of all the nodes can be computed using Bayesian updating. This updating requires complicated analytical integration. The MCMC technique is implemented as an efficient approach to achieve this purpose. Refer to [6] for details about the Bayesian updating.

2.2. Acceptance confidence

Given observations Y, the confidence in the model may be measured by the posterior probability of the null hypothesis $Pr(H_0|Y)$. The relative posterior probabilities of the two hypotheses are obtained using the Bayes theorem as

$$\frac{Pr(H_0|Y)}{Pr(H_1|Y)} = \frac{Pr(Y|H_0)\pi_0}{Pr(Y|H_1)\pi_1} = \Lambda(Y)\frac{\pi_0}{\pi_1}, \tag{13}$$

where $Pr(H_1|Y)$ represents the posterior probability of the alternative hypothesis. For a binary hypothesis testing we have $Pr(H_1|Y) = 1 - Pr(H_0|Y)$. Thus, $Pr(H_0|Y)$ can be derived from Eq. (13) as follows:

$$Pr(H_0|Y) = \frac{\Lambda(Y)\pi_0}{\pi_1 + \Lambda(Y)\pi_0}. \tag{14}$$

Assume that $\pi_0 = \pi_1 = 0.5$ in the absence of prior knowledge of each hypothesis before testing. Eq. (14) can be simplified as

$$Pr(H_0|Y) = \frac{\Lambda(Y)}{1 + \Lambda(Y)}. \tag{15}$$

Thus, Eqs. (14) and (15) quantify the confidence in the model based on the validation experimental data. From Eq. (15), $\Lambda(Y)=0$ indicates zero confidence in the model, and $\Lambda(Y)\to\infty$ indicates 100% confidence. Thus, from the viewpoint of acceptance confidence, the decision-based method yields the same result as the Bayesian hypothesis testing method in [6], since the likelihood ratio is in fact the Bayes factor, as shown in Eq. (9).

3. Implementation of bayesian decision method

Two types of quantities of interest with respect to reliability models, namely, multiple pass/fail tests (Case 1) and system response measurement tests (Case 2), have been considered by Mahadevan and Rebba [6] for model validation based on Bayesian hypothesis testing. In the first case, multiple tests are conducted and the results are reported simply as failure or success with respect to specific performance criteria. In the second case, a response quantity such as strain, displacement, acceleration, etc. is measured, not just failure or success. The Bayes risk decision criterion is developed for both cases of model validation below.

3.1. Pass/fail tests (Case 1)

Assume that the failure probability y of an engineering system is predicted by a model and its value is denoted by y_0. Thus, the point null hypothesis is that the density of y is sharply concentrated at y_0, i.e., $H_0: y=y_0$, and the alternative hypothesis is $H_1: y\neq y_0$. If n identical and independent experiments are performed, and k failures are observed without considering statistical uncertainty, the probability of observing failures is a binomial distribution expressed as

$$Pr(k|y,n)=\binom{n}{k}y^k(1-y)^{n-k}, \tag{16}$$

where $\binom{n}{k}=n!/(n-k)!k!$ is a combination coefficient (the number of ways of selecting k *unordered* outcomes from n possibilities), in which $n!$ is a factorial defined by $n!=n\cdot(n-1)\cdot\dots\cdot 2\cdot 1$. Note that the expression in Eq. (16) is also the likelihood function of y. Thus, under the null hypothesis, the probability $Pr(y|H_0)$ can be estimated by simply substituting y_0 in Eq. (16), i.e., $Pr(y|H_0)=\binom{n}{k}y_0^k(1-y_0)^{n-k}$ Since there is no prior information about y under the alternative hypothesis, it is customary to assume a uniform distribution between [0, 1] for the prior distribution of y under H_1, $f(y|H_1)$. (On the other hand, if we have prior information about the distribution of y based on model output or model calibration, the proposed methodology can easily incorporate this information). Using the uniform prior for y, the likelihood ratio or Bayes factor may be computed as

$$\begin{aligned}\Lambda(y_0)&=\left.\frac{Pr(y|H_0)}{Pr(y|H_1)}\right|_{y=y_0}\\&=\frac{\binom{n}{k}y_0^k(1-y_0)^{n-k}}{\int_0^1\binom{n}{k}y^k(1-y)^{n-k}f(y|H_1)\,\mathrm{d}y}\\&=y_0^k(1-y_0)^{n-k}(n+1)\binom{n}{k},\end{aligned} \tag{17}$$

where

$$Pr(y|H_1)=\int_0^1\binom{n}{k}y^k(1-y)^{n-k}\pi(y|H_1)\,\mathrm{d}y=1/(n+1).$$

When $\Lambda(y_0)$ is greater than the threshold η, the data is said to favor the null hypothesis that the predicted failure probability is true and provides us some confidence in accepting the model prediction based on Eq. (15). However, if $\Lambda(y_0)<\eta$, it may be inferred that the data does not support the null hypothesis. Thus, the likelihood ratio is employed as a decision metric to compare the data and prediction. From Eq. (15), the larger the likelihood ratio, the greater the confidence of accepting the model.

In this case, given experimental observations Y_i $(i=1, 2, \ldots, N)$ where N is the number of tests, the Bayes risk in Eq. (4) can be expressed as

$$\begin{aligned}R&=\sum_{Z_0}[c_{00}\pi_0 Pr(Y_i|H_0)+c_{01}\pi_1 Pr(Y_i|H_1)]\\&\quad+\sum_{Z_1}[c_{10}\pi_0 Pr(Y_i|H_0)+c_{11}\pi_1 Pr(Y_i|H_1)]\\&=\sum_{Z_0}\left[c_{00}\pi_0 Pr(Y_i|H_0)+c_{01}\pi_1\frac{Pr(Y_i|H_0)}{\Lambda(Y_i)}\right]\\&\quad+\sum_{Z_1}\left[c_{10}\pi_0 Pr(Y_i|H_0)+c_{11}\pi_1\frac{Pr(Y_i|H_0)}{\Lambda(Y_i)}\right],\end{aligned} \tag{18}$$

where the subscript i is the experiment number, and $Pr(Y_i|H_1)=Pr(Y_i|H_0)/\Lambda(Y_i)$ is used since it is usually easier to obtain $Pr(Y|H_0)$ than $Pr(Y|H_1)$. Notice that the capital Y is used to represent the experimental result of the failure probability. The integral item in Eq. (4) is substituted by the summation of binomial probabilities in terms of tested failure probability Y_i obtained from the decision region Z_0. This is feasible because the measured data in validation experiments is usually discrete. Thus, if we know the prior information about the hypotheses and cost of deciding H_i when H_j is true, we can compute Bayes risk (cost) based on the experimental data available. Note that, a minimum Bayes risk will be obtained if and only if validation data are correctly assigned to the region Z_0 when the data accepts the model or Z_1 when the data rejects the model, which will be discussed later.

3.2. Response value measurement tests (Case 2)

The likelihood ratio derived in Eq. (17) is applicable only for the case with uniform prior density probability and binomial pass/fail data. In other situations where only a response quantity may be measured in the tests, it is valuable to derive a more general expression for the likelihood ratio and the corresponding Bayes risk based on the prior and posterior probability density functions (PDF) of the predicted response. Let $f(y)$ be the PDF of the predicted value, and $f(y|H_1)$ the likelihood of the predicted value under H_1. Suppose $f(y|H_1)=f(y)$ is assumed since no information is available for $f(y|H_1)$. Based on the method presented by Mahadevan and Rebba [6] to derive the Bayes factor, the likelihood ratio in this case is computed as

$$\Lambda(y_0) = \frac{Pr(y|H_0 : y = y_0)}{Pr(y|H_1 : y \neq y_0)} = \frac{L(y_0)}{\int L(y) f(y|H_1)\,\mathrm{d}y} = \frac{f(Y|y_0)}{\int f(Y|y_0) f(y)\,\mathrm{d}y}, \tag{19}$$

where $Pr(Y|H_0 : y = y_0) = L(y_0) = \varepsilon f(Y|y_0)$ is the likelihood probability of observing the data under H_0, in which ε is a small positive number, and $Pr(y|H_1 : y \neq y_0) = \int L(y) f(y|H_1)\,\mathrm{d}y = \int \varepsilon f(Y|y) f(y|H_1)\,\mathrm{d}y$. Using the Bayes theorem, $f(y|Y) = f(Y|y)f(y)/\int f(Y|y)f(y)\,\mathrm{d}y$, Eq. (19) becomes [6]

$$\Lambda(y_0) = \left.\frac{f(y|Y)}{f(y)}\right|_{y=y_0}, \tag{20}$$

where $f(y|Y)$ is the posterior PDF of the predicted value given the observed data Y. Thus, the likelihood ratio simply becomes the ratio of posterior to prior PDFs of the response at the predicted value y_0 under the assumption of $f(y|H_1)=f(y)$. If $f(y|H_1) \neq f(y)$, then the likelihood ratio is computed using Eq. (19) with $f(y|H_1)$ instead of $f(y)$ in the denominator. Fig. 2 shows the posterior and prior densities of model prediction y.

Usually the experimental measurement is also discrete in this case, thus the Bayes risk can be easily computed again using Eq. (18) if we know the prior density of the hypotheses and the costs of deciding H_i when H_j is true. Note that the likelihood of observing the data under H_0 needs to be updated using the Bayesian theorem and the MCMC technique, which will be demonstrated in the second example presented in this study. Refer to [6] for the details of the Bayesian updating procedure.

4. Minimizing bayes risk

There are two types of scenarios in the minimization of Bayes risk for model validation. First, given the experimental data available, how to minimize the Bayes risk in the model validation? Second, before model validation, how to optimally design the validation experiments targeted towards minimizing the Bayes risk in the model validation? The two scenarios are addressed below.

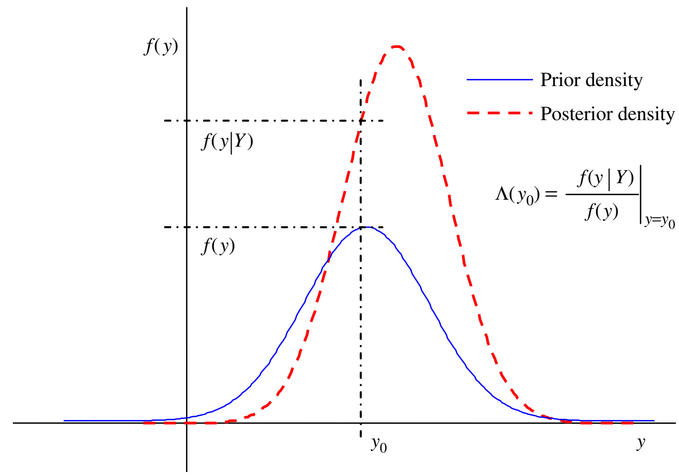

Fig. 2. Likelihood ratio metric Λ for model evaluation.

4.1. Validation data already collected (Scenario 1)

Since the test data are already available for model validation, they should be correctly assigned to the decision regions in order to minimize the Bayes risk. Thus, minimizing the Bayes risk is equivalent to optimizing the assignment of data points to two different decision regions Z_0 and Z_1. This can be easily realized through computing the likelihood ratio $\Lambda(Y^*)$ for each data point Y^*. Based on the above discussion about the Bayes risk, if $\Lambda(Y^*)>\eta$, the experimental data Y^* is said to favor the model and therefore assigned to the decision region Z_0, thus, minimizing the Bayes risk. Similarly, assigning Y^* to Z_1 if $\Lambda(Y^*)<\eta$ will minimize the Bayes risk. The optimization solution can be obtained through computing all possible risks based on Eq. (18) since it is a simple discrete optimization problem with a small data set. If the number of experimental data is very large, any search technique [15] may be used to find the minimum risk-based decision.

Based on the data already collected, if a decision to accept or reject the model can not yet be made using the Bayesian decision rule, either improving the current model or gathering more validation data may be performed under the constraint of the available budget. The model improvement is beyond the scope of this paper. The case of collecting more validation data is discussed below.

4.2. Validation data yet to be collected (Scenario 2)

When the experimental data is not yet available or insufficient for model validation, the experiment needs to be optimally designed to generate the outcome which provides the greatest opportunity for performing conclusive comparisons in model validation. Jiang and Mahadevan [2] developed a Bayesian cross entropy-based methodology to achieve this purpose. In that method, the expected cross entropy, an information-theoretic distance between the distribution of experimental output and that of model prediction, is minimized or maximized to ensure respectively the highest or lowest similarity between the distributions of experimental output and model prediction. The minimization and maximization result in two set of inputs for the experiment.

Consider the likelihood ratio $\Lambda(y)$ in Eq. (9) [or Bayes factor in [6]] as a validation metric. Suppose an experiment is conducted with the minimization result, and the experimental output is compared with model prediction. We expect a high value $\Lambda(y)|_{\min}$, where the subscript $_{\min}$ indicates that the likelihood is obtained from the experimental output in the minimization case. If $\Lambda(y)|_{\min}<\eta$, then clearly this experiment rejects the model, since the validation metric $\Lambda(y)$, even under the most favorable conditions, does not meet the threshold value η. On the other hand, suppose an experiment is conducted with the maximization result, and the experimental output is compared with the model prediction. We expect a low value $\Lambda(y)|_{\max}$ in this case. If $\Lambda(y)|_{\max}>\eta$, then clearly this experiment accepts the model, since it is performed under the worst condition and still produces the validation metric to be higher than η. Thus, the cross entropy method provides conclusive comparison as opposed to an experiment at any arbitrary point. Refer to [2] for details about the optimal design of validation experiments.

There may be two special cases. First, if neither $\Lambda(y)|_{\min}<\eta$ nor $\Lambda(y)|_{\max}>\eta$ is satisfied, then the above optimization procedure and corresponding experiment need to be repeated until the confidence measure of model prediction reaches a stable value. The number of tests required is presented in [5]. Second, if both $\Lambda(y)|_{\min}$ and $\Lambda(y)|_{\max}$ are always equal or approximate to a specific decision threshold, i.e., $\Lambda(y)\cong\eta*$, the additional experiments may not help make a conclusive decision to accept or reject the model. This implies that the model needs to be improved and then further validated.

Fig. 3 shows the Bayesian decision-theoretic view of two different decision regions at $\eta=1$. If the likelihood ratio $\Lambda(Y)|_{\max}>1$, the data Y from the maximum cross entropy experiment is said to favor the model and therefore assigned to the decision region Z_0, thus minimizing the Bayes risk. Similarly, assigning Y from the minimum cross entropy experiment to Z_1 if $\Lambda(Y)|_{\min}<1$ will minimize the Bayes risk. Obviously, if the experimental data fall in the shaded zone (Z_0), they are said to favor the model. Otherwise, they are said to reject the model. The relationship $\Lambda(Y)=1$ is referred to as a decision boundary in Fig. 3. Note that, a larger decision threshold (i.e., $\eta>1$) will decrease Z_0 (the decision boundaries in Fig. 3 will move inward), and conversely.

5. Numerical applications

The procedure of applying the Bayesian risk-based decision methodology for model validation is shown in

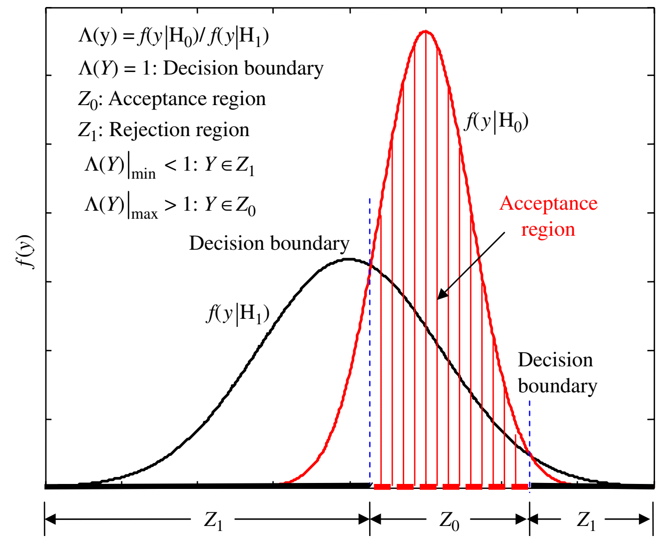

Fig. 3. Bayesian decision-theoretic view of optimal experimental design for model validation at $\eta=1$.

Fig. 4. Two concerns need to be pointed out. First, the methodology integrates the minimum risk of using the current model with the cost of acquiring new information to improve the model. During the model validation, the decision of improving the current model or collecting more data (shaded diamonds in Fig. 4) needs be made by decision makers or engineers based on the available budget and cost information, which are not tackled in this study. Second, if the number of validation data, N_e, is small (e.g., less than 20), the risks of all possible combinations for particular experiments, $\Sigma_{i=0}^{N_e}\binom{N_e}{i}=2^{N_e}$, may be calculated to yield the minimum risk. However, if the number is very large, it is prohibitively costly to compute the risks of all possible combinations. In this case, a search technique is required to find the minimum risk-based decision, which may be pursued in a future study.

In this study, the Bayesian decision methodology is implemented for reliability model validation of two practical problems: a bar subjected to tension stress (Example 1) and high cycle fatigue of rotor blades in aircraft engines (Example 2). Example 1 is used to demonstrate the methodology for multiple pass/fail tests (Case 1). While Example 2 is to demonstrate the methodology for response value measurement tests (Case 2). For the sake of comparison with the Bayesian hypothesis testing approach, assume that the costs of deciding H_i when H_j true are normalized to be $c_{00}=c_{11}=1$ (unit), and $c_{01}=c_{10}=2$ (unit), and the prior densities about the two hypotheses are $\pi_0=\pi_1=0.5$. This gives the threshold $\eta=1$, from Eq. (9).

5.1. Example 1: tension bar

5.1.1. Problem description

The bar subjected to tension stress fails when the applied load (S) exceeds the tensile strength of the bar (R). The

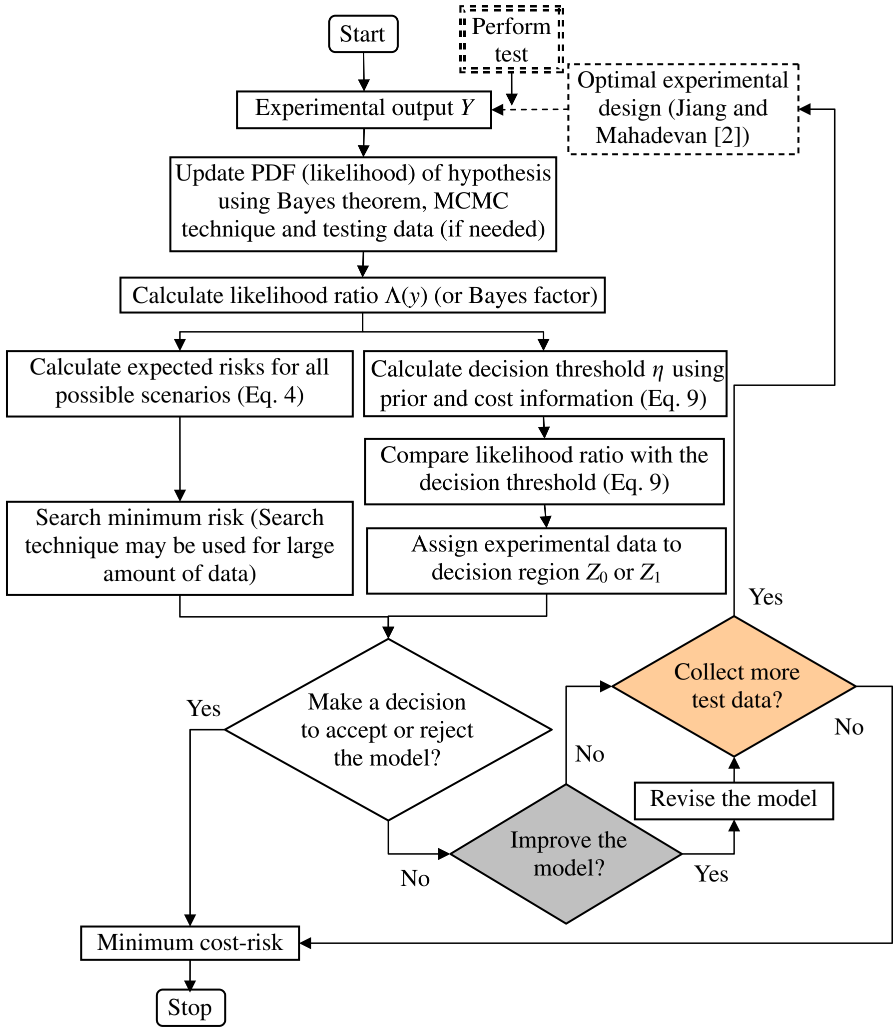

Fig. 4. Application of Bayes risk-based decision methodology for model validation.

limit state function of the bar, g, is defined as

$$g = R - S. \tag{21}$$

Eq. (21) is considered in this example as the reliability computational model of the bar. Assume that R and S are two independent normal variables with mean μ_R and μ_S, and variance σ_R and σ_S. Thus, the reliability index is available as a closed-form equation in terms of the distribution parameters:

$$\beta = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \tag{22}$$

and the failure probability is predicted by

$$y_0 = \Phi(-\beta) \tag{23}$$

where $\Phi(\cdot)$ represents the standard normal cumulative distribution function.

For the purpose of illustration, assume that there is enough information showing that μ_R, σ_R, μ_S, and σ_S have deterministic values equal to 50.0, 5.0, 35.0, and 7.0 Mpa, respectively. Thus, the failure probability is predicted to be $y_p = 0.0406$ based on Eqs. (21)–(23). Assume that the true failure probability of the bar is equal to 5%, and a group of experimental results with different combinations of observed n and k are obtained as shown in Table 1. The test results and their corresponding likelihood obtained by Eq. (16) are also shown in Table 1.

Table 1
A bar with multiple pass/fail tests (Example 1)

No.	1	2	3	4
n	25	50	75	100
k	1	2	4	5
$\Pr(k\mid y_p, n)$ (Eq. 16)	0.3754	0.2762	0.1741	0.1619
Λ (Eq. 17)	9.76	14.08	13.23	16.35
Confidence (%) (Eq. (14))	90.7	93.4	93.0	94.2

5.1.2. Bayes risk analysis

The likelihood ratio for every group of tests is obtained from Eq. (17) and shown in Table 1. It is observed that all likelihood ratios are larger than $\eta(=1)$, i.e., all the four sets of experiment data are to favor the model. Suppose that there are i experiments said to favor the model and the remaining 4 - i experiments said to reject the model, all possible risks occurring to the four group experiments, $\Sigma_{i=0}^{4}\binom{4}{i}=16$, are computed using Eq. (18). The minimum Bayes risk is $R_{\min}=0.5749$ units when all four experimental outputs are assigned to the decision region Z_0. This conclusion is consistent with that obtained by using Bayesian hypothesis testing [5], that is, all four experiments are said to favor the model since the corresponding Bayes factor is larger than one. However, if all the four experiments are wrongly assigned to the decision region Z_1 (reject the model), the maximum Bayes risk is $R_{\max}=1.0281$ units. Obviously, the wrong decision substantially increases the risk in the model validation.

All the risks are plotted in Fig. 5, whose horizontal axis represents the number of experiments assigned to the decision region Z_0 (accept the model). For instance, the number one in the horizontal axis represents that only one experiment out of four favors the model. There are four possible scenarios, depending on which experiment is judged to accept the model, i.e. $\binom{4}{1}=4$. The four dots in the vertical direction corresponding to the number one in the horizontal axis represent the resulting four possible risk values. This indicates the range of risk in the case that only one experiment out of four favors the model. When two out of four experiments favor the model, there are six possible combinations, thus six risk values are computed for this case, and plotted in Fig. 5.

The acceptance confidence of every set of experiments is calculated using Eq. (15) and also shown in Table 1. It is observed that in all cases, the acceptance confidence in the model is larger than 90%. Note that the model is rejected by experiment sets 3 and 4 (Table 1) using classical hypothesis testing in [5]. In that case, since wrong decisions are made to reject the model by experimental data sets 3 and 4, the corresponding lower bound of Bayes risk (cost) for this case, $R=0.7314$ units is obtained by Eq. (18). This is larger than the minimum value 0.5749 (all four experimental data sets judged as supporting the model). The significant increase of costs results from wrongly assigning the experimental data to reject the model.

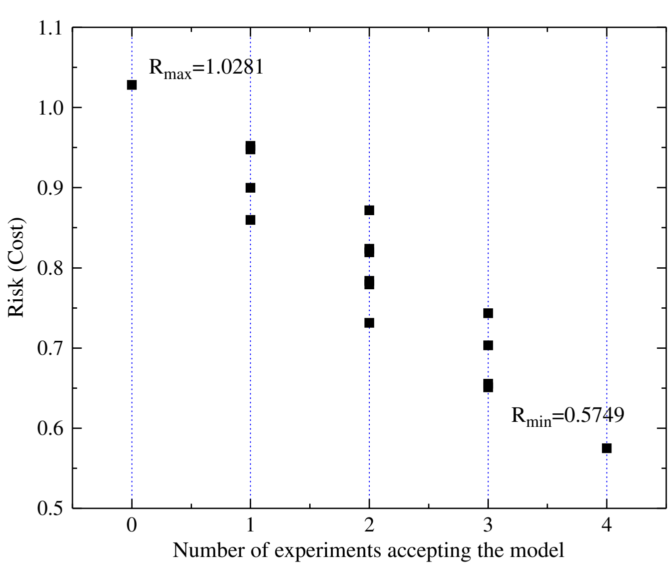

Fig. 5. Bayes risk analysis of reliability model validation for the tension bar (Example 1).

5.2. Example 2: aircraft engine rotor blade

5.2.1. Problem description

The applied dynamic loading and material properties of the rotor blades in aircraft engines are usually random variables. Thus, the failure probability of a single blade under high cycle fatigue can be estimated by a limit state-based reliability prediction model. The blade is assumed to have failed when the actual maximum displacement under dynamic loading exceeds the design or allowable maximum displacement. Generally the blade is modeled as a single degree of freedom oscillator and its dynamics is described by a differential equation consisting of mass, spring, dashpot and with displacement x as follows [16]:

$$F \sin \omega t = m\ddot{x} + c\dot{x} + kx, \tag{24}$$

where $F=$ magnitude of the external harmonic load, $\omega=$ applied load frequency, $m=$ mass of the oscillating body, $c=$ damping constant, and $k=$ stiffness of the spring.

The displacement is computed as

$$x = \frac{F}{k}\frac{1}{\sqrt{[1-(\omega/\omega_{\mathrm{n}})^2]^2 + [2(c/c_{\mathrm{c}})(\omega/\omega_{\mathrm{n}})]^2}}, \tag{25}$$

where $\omega_{\mathrm{n}}=$ natural frequency and $c_{\mathrm{c}}=$ critical damping factor. The performance function for the failure of the blade is thus a function of the natural frequency, damping, load factor and engine speed (all random variables):

$$g = 1 - \left[\frac{p_{\mathrm{d}}}{100}\cdot\lambda_l\lambda_{\mathrm{m}}\cdot\frac{\alpha}{\alpha_{\mathrm{n}}}\cdot\frac{d_{\mathrm{n}}}{D_a}\right]. \tag{26}$$

where λ_{l} and λ_{m} are the load and modal shape factors, respectively, α and α_{n} are the design and nominal amplification factors, respectively, d_{n} is the allowable design displacement or the nominal allowable displacement, D_{a} is the allowable displacement, and p_{d} is the percentage of nominal allowable displacement. The limit state is denoted by the condition $g=0$.

The dynamic amplification factor is described by the equation

$$\alpha = \frac{1}{\sqrt{(1-\beta^2)^2 + (2\xi\beta)^2}}, \tag{27}$$

where $\xi = c/c_{\mathrm{c}}$ is the damping ratio, and $\beta = \omega_{\mathrm{d}}/\omega_{\mathrm{n}}$ is the frequency ratio in which $\omega_{\mathrm{d}} = 13N/60$ is the design

Table 2
Blade problem: Statistics of the variables (Example 2)

Variable	Type	Mean	Std dev
ω_{nom}	Normal	2194 rpm	105 rpm
N	Normal	8800 rpm	50 rpm
b_{nom}	Normal	0.1	0.005
ξ	Lognormal	0.002	0.0005
λ_l	Normal	1	0.25
λ_m	Normal	1	0.05
D_a	Normal	15 mm	0.75 mm
α_n	Deterministic	5.7427	–
p_d	Deterministic	25	–
d_n	Deterministic	15 mm	–

Table 3
Validation data on intermediate quantities in Example 2 (Model prediction $\omega_n = 2220$ rpm, $\beta = 0.9$)

No.	Validation test data	$\Lambda(g)$	Conf. (%)	$f_0(g)$	$f_1(g)$
1	$\beta = 0.6$	1×10^{-5}	0.001	6.9×10^{-6}	0.69
2	$\beta = 0.9$	5.46	84.5	1.818	0.333
3	$\omega_n = 2220$ rpm	2.11	67.8	1.745	0.827
4	$\omega_n = 3316$ rpm	1×10^{-5}	0.001	6.89×10^{-6}	0.689
5	$\beta = 0.87$, $\omega_n = 2220$ rpm	1.6	61.5	1.772	1.107
6	$\beta = 0.6$, $\omega_n = 2220$ rpm	1.2	54.5	1.389	1.157
7	$\beta = 0.87$, $\omega_n = 3316$ rpm	1×10^{-5}	0.001	6.89×10^{-6}	0.689
8	$\beta = 0.6$, $\omega_n = 3316$ rpm	1×10^{-5}	0.001	6.9×10^{-6}	0.69

frequency depending on the rotary speed N in rpm. The Campbell diagram [17] with a slope and a function of rotor speed is used to depict the resonance condition. Thus, the natural frequency ω is computed from the Campbell diagram using the equation [17]

$$\omega_n = \omega_{nom} + b_{nom}(N - 8800). \tag{28}$$

Following the study of Mahadevan and Rebba [6], the statistics of the variables in the above sets of equations are shown in Table 2 and used in this study for Bayes risk analysis in terms of model validation.

5.2.2. Bayes risk analysis

In order to calculate the Bayes factor (the likelihood ratio in this study), the series of quantities involved in computing the performance function g was modeled by Mahadevan and Rebba [6] as different nodes in a Bayes network. With the availability of any experimental validation outcome of any node and the statistics of each node, the statistical distribution function associated with all the nodes in the network, including g, could be updated. Then the likelihood ratio of g is calculated using Eq. (19). In this study, assume that the measured values for two quantities ω_n(2220 and 3316 rpm) and β (0.6, 0.87, and 0.9) are available. Their corresponding predicted values are 2200 rpm and 0.9. Different combinations of the validation data and the corresponding likelihood ratio values for the overall reliability model prediction are shown in Table 3. Fig. 6 shows the prior and posterior densities of g resulting from Bayesian updating with validation data on single and multiple nodes. The decision regions can be easily identified through the decision boundaries with the decision threshold $\eta = 1$.

In order to calculate the Bayesian risk, a Bayesian network (Fig. 7) is designed in this study to update the likelihood of g [i.e., $f_0(g) = L(Y|y)$], given the measured values of ω_n and/or β. In this figure, an ellipse (for example, α) represents a random variable and a rectangle (for example, the experimental value β_{exp}) represents a constant value. The double line arrow represents a logical relationship link between two variables (computational formula) and a single line arrow represents a direct probabilistic relationship link. After any data node is added to the network, the posterior probability densities of all the nodes are computed, including g_{exp} (the experimental value of g). The statistics of the parameters shown in Table 2, the computational models presented in Eqs. (25)–(27), and 10,000 iterations of simulation are used

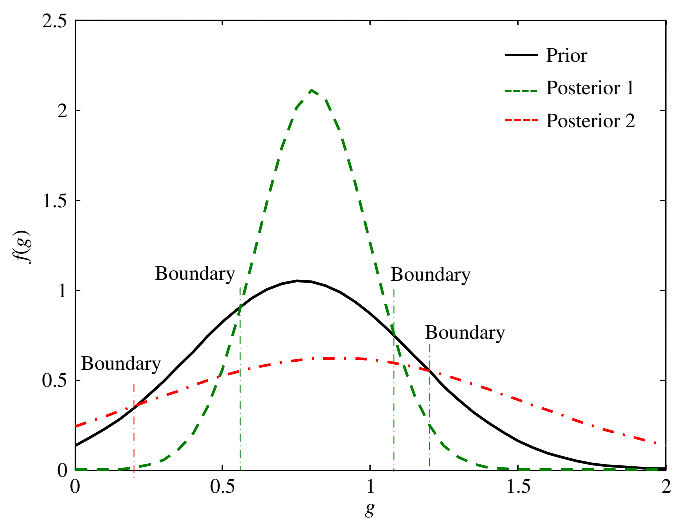

Fig. 6. Prior and posterior distributions of g with different validation data shown in Table 3 (Example 2).

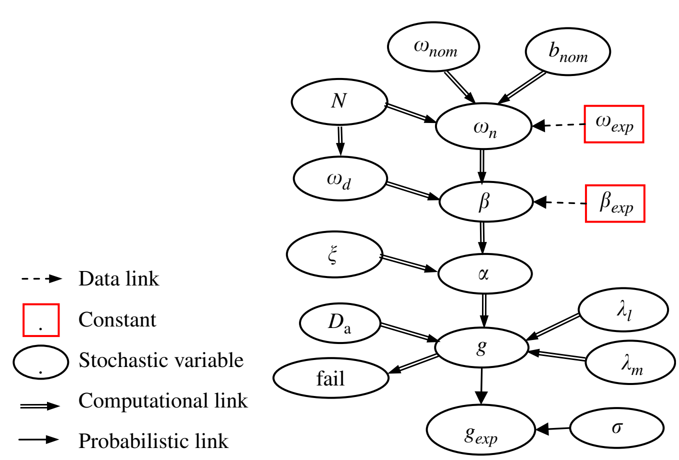

Fig. 7. Bayes Network for updating the PDF of g (Example 2).

in the Bayesian updating. The likelihood values for the overall reliability model prediction in different combinations of the validation data are presented in Table 3. These values are subsequently used to perform the risk analyses for model validation in two cases: $\eta = 1$ and $\eta = 2$.

Case 1: $\eta = 1$.

Suppose that i out of 8 experiments are judged to favor the model and the remaining 8 - i experiments reject the model. For each i, there are $\binom{8}{i}$ possible combinations of the particular experiments that favor or reject the model, and $\binom{8}{i}$ corresponding risk values, based on the cost information, using Eq. (18). All the risk values for $i = 0$–8 are shown in Fig. 8, whose horizontal axis represents the number of experiments assigned to the decision region Z_0 (model acceptance). Similar to Example 1, the dots in the vertical direction show the different risk values corresponding to different combinations of particular experiments that accept or reject the model. The minimum Bayes risk is $R_{\min} = 8.166$ units when $i = 4$, i.e., four experimental outputs whose likelihood ratios are larger than η are assigned to the decision region Z_0, the remaining four experimental outputs whose likelihood ratios are less than η are assigned to the decision region Z_1. This conclusion is consistent with that obtained by using Bayesian hypothesis testing [6]. The correct decision results in the minimum risk. Conversely, the maximum Bayes risk is $R_{\max} = 11.1945$ units when the same eight experimental outputs are wrongly assigned to the different decision regions.

Case 2: $\eta = 2$

In this case, assume that the costs of deciding H_i when H_j is true are $C_{00} = 1$, $C_{11} = 1.2$, $C_{01} = 2$, and $C_{10} = 1.4$ (unit), and the prior information about the hypotheses still are $\pi_0 = \pi_1 = 0.5$. Thus the threshold $\eta = 2$ is obtained

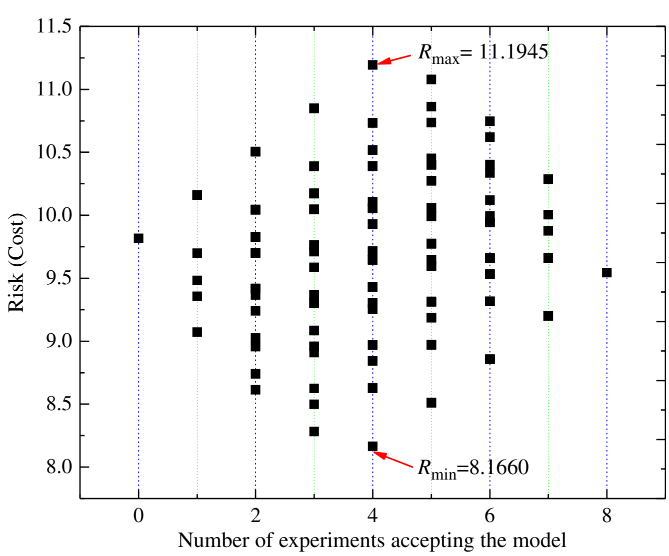

Fig. 8. Bayesian risk-based decision analysis for validation of rotor blade high cycle fatigue reliability model (Example 2: $\eta = 1$).

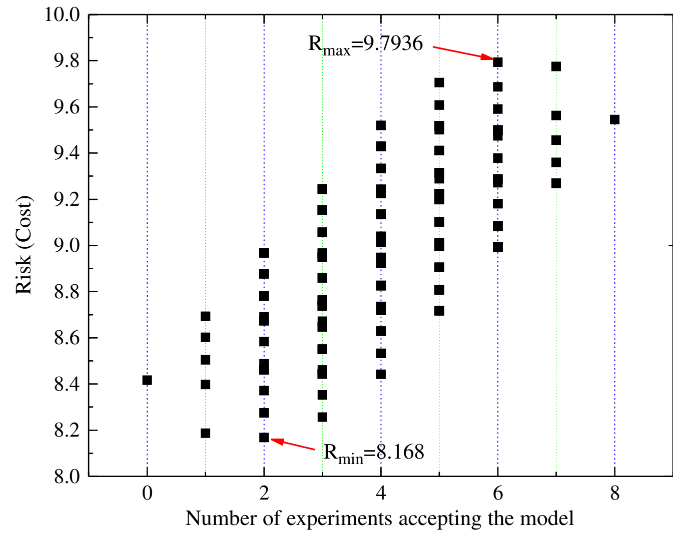

Fig. 9. Bayesian risk-based decision analysis for validation of rotor blade high cycle fatigue reliability model (Example 2: $\eta = 2$).

from Eq. (9) and only the 2nd and 3rd experiments support the model. The 256 possible risk values are computed again using Eq. (18) and plotted in Fig. 9. The minimum Bayes risk is $R_{\min} = 8.168$ units when two experimental outputs whose likelihood ratios are larger than $\eta(= 2)$ are assigned to the decision region Z_0, while the other six experimental outputs whose likelihood ratios are less than η are assigned to the decision region Z_1. Again, the maximum Bayes risk is $R_{\max} = 9.794$ units when all eight experimental outputs are wrongly assigned to the decision regions (i.e., two experiments that support the model are wrongly assigned to Z_1 and six experiments that do not support the model are assigned to Z_0).

It should be pointed out that the decision threshold η depends on both the cost information C_{ij} and the prior of each hypothesis π_i. Only the cost information is changed in this example for the purpose of illustration. However, it is easy to incorporate the engineers' or decision-makers' preferences about both the cost and prior information in the threshold.

6. Conclusions

To the best of the authors' knowledge no paper has been published on the development of a Bayesian decision theory for model validation under uncertainty, incorporating the optimal design of validation experiments. This paper is intended to be the first. In contrast to the traditional methods such as hypothesis testing, the Bayesian decision approach offers a more comprehensive theoretical foundation for model validation by including risk or not. The task of Bayesian decision-based model validation method is clearly identified, that is, to accept or reject the computational model with minimum cost or risk. A decision maker can determine various decision thresholds by specifying the importance of the different cost sources and the prior knowledge about the null and alternative hypotheses. In fact, the Bayes factor validation metric based on Bayesian hypothesis testing is derived in this paper from a generalized risk-based decision perspective.

A correct decision usually results in a small risk in model validation. Therefore, a minimum risk implies the correct decisions made by the engineers or decision-makers. Thus, in practical applications of Bayesian risk-based decision method for model validation, the decisions leading to the minimum risk are the correct ones. This issue has been clearly demonstrated in Figs. 5, 8, and 9.

In summary, the Bayes risk criterion has been demonstrated as a powerful rule which advances the Bayesian decision theory in the context of model validation based on two features. First, it incorporates decision costs (the preferences of a decision maker) into the risk analysis through assigning costs to each correct or incorrect decision and then minimizing the total average cost (loss or risk function). Second, it is a comprehensive criterion that generalizes three other decision rules. The Bayes risk analysis is easily applied for measured data that is already available. If the data is yet to be collected, Bayesian decision theory, i.e., minimizing the risk, and information theory, i.e., minimizing/maximizing the information distance, are integrated to simultaneously minimize the cost of experiments and minimize the model user's risk.

It should be pointed out that, during the model validation, the decision of improving the current model or collecting more data needs be made by decision makers or engineers based on the available budget and the corresponding cost information about model development and additional experiments, which are not tackled in this study. In addition, the number of validation data is small in the two examples presented in this study. Therefore, the risks of all possible combinations for particular experiments have been calculated to yield the minimum risk. However, in the case of a large number of validation data available for model validation, a search technique is required to find the minimum risk-based decision, which may be pursued in a future study.

Acknowledgement

The research was supported by funds from Sandia National Laboratories, Albuquerque, New Mexico (contract no. BG-7732), under the Sandia-NSF Life Cycle Engineering program (project monitors: Dr. Thomas L Paez, Dr. Martin Pilch). The support is gratefully acknowledged.

References

[1] Trucano TG, Pilch M, Oberkampf WL. (2002). General concepts for experimental validation of ASCI code applications. Technical Report SAND2002-0341. Albuquerque, NM: Sandia National Laboratories.
[2] Jiang X, Mahadevan S. Bayesian cross entropy methodology for optimal design of validation experiments. Meas Sci Technol 2006; in press.
[3] Berger JO, Sellke T. Testing a point null hypothesis: the irreconcilability of *P* values and evidence. J Am Stat Assoc 1987;82(397): 112–22.
[4] Marden JI. Hypothesis testing: from *p* values to Bayes factors. J Am Stat Assoc 2000;95(452):1316–20.
[5] Zhang R, Mahadevan S. Bayesian methodology for reliability model acceptance. Reliab Eng Syst Saf 2003;80(1):95–103.
[6] Mahadevan S, Rebba R. Validation of reliability computational models using Bayes networks. Reliab Eng Syst Saf 2005; 87(2):223–32.
[7] Berger JO, Delampady M. Testing precise hypotheses. Stat Sci 1987;2(3):317–52.
[8] Hwang JT, Casella G, Robert C, Wells MT, Farrell RH. Estimation of accuracy in testing. Annu Stat 1992;20(1):490–509.
[9] Grigoriu M, Veneziano D, Cornell CA. Probabilistic modeling as decision making. J Eng Mech-ASCE 1979;105(4):585–96.
[10] Geyskens Ph, Der Kiureghian A. Bayesian assessment and selection of models for structural reliability analysis. In: Frangopol DM, Grigoriu MD, eds., *Proceedings of the 7th ASCE EMD/STD Joint Specialty Conference on Prob Mech Struct Reliab,* Worcester, MA, ASCE, New York; August, 1996. p.566–569.
[11] Schervish MJ. Theory of statistics. NY: Springer; 1995.
[12] Balci O, Sargent RG. A methodology for cost-risk analysis in the statistical validation of simulation models. Commun Assoc Comp Mach (ACM) 1981;24(11):190–7.
[13] Jensen FV, Jensen FB. Bayesian networks and decision graphs. NY: Springer; 2001.
[14] Nowak R, Scott C. The Bayes risk criterion in hypothesis testing. Connexions 2004; http://cnx.rice.edu/content/m11533/1.6/.
[15] Parker RG, Rardin RL. Discrete optimization. New York: Academic Press; 1988.
[16] Annis C. Modeling high cycle fatigue with Markov chain Monte Carlo-a new look at an old idea. In: Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC struct, struct dyn, and mater conf (AIAA-2002-1380). Denver, CO, 2002. p. 1351–1361.
[17] Prohl MA. A general method for calculating critical speeds of flexible rotors. J Appl Mech ASME 1945;67:A142–8.