

DRAFT

DETC2008-49663

BAYESIAN RELIABILITY ANALYSIS WITH EVOLVING, INSUFFICIENT, AND SUBJECTIVE DATA SETS

Pingfeng Wang¹

Byeng D. Youn²

Zhimin Xi³

Department of Mechanical Engineering, University of Maryland at College Park
College Park, MD 20742, USA

Artemis Kloess⁴

R&D Center, General Motors Corporation, Warren, MI, 48090, USA

ABSTRACT

A primary concern in product design is ensuring high system reliability amidst various uncertainties throughout a product life-cycle. To achieve high reliability, uncertainty data for complex product systems must be adequately collected, analyzed, and managed throughout the product life-cycle. However, despite years of research, system reliability assessment is still difficult, mainly due to the challenges of evolving, insufficient, and subjective data sets. Therefore, the objective of this research is to establish a new paradigm of reliability prediction that enables the use of evolving, insufficient, and subjective data sets (from expert knowledge, customer survey, system inspection & testing, and field data) over the entire product life-cycle. This research will integrate probability encoding methods to a Bayesian updating mechanism. It is referred to as Bayesian Information Toolkit (BIT). Likewise, Bayesian Reliability Toolkit (BRT) will be created by incorporating reliability analysis to the Bayesian updating mechanism. In this research, both BIT and BRT will be integrated to predict reliability even with evolving, insufficient, and subjective data sets. It is shown that the proposed Bayesian reliability analysis can predict the reliability of door closing performance in a vehicle body-door subsystem where the relevant data sets availability are limited, subjective, and evolving.

1. INTRODUCTION

In the last three decades, engineering analysis and design methods have advanced to improve reliability of an engineering product system while considering uncertainties in the system. However, little attention has been made to data modeling with evolving, insufficient, and subjective data sets. In this paper, we refer the data which are not static but evolve with time as “evolving data”, refer the data which are not sufficient to fully characterize random behavior as insufficient data and, similarly, refer the data which are pertaining to or perceived only by individuals as subjective data. To be clear, aleatory uncertainty is defined as the uncertainty which arises because of natural unpredictable variation in the performance of the system under study whereas epistemic uncertainty is defined as the uncertainty which is due to a lack of knowledge about the behavior of the system that is conceptually resolvable. More specifically, aleatory uncertainties are considered to be represented by statistical distributions whereas epistemic uncertainties are considered to be represented by limited data sets in this paper. Most probabilistic analysis and design approaches still depend on the assumed probabilistic models of system inputs without engaging raw data. The research that predicts product reliability with evolving, insufficient, and subjective data sets is strongly in demand for engineering

¹ Graduate Student, pfwang@umd.edu

² Assistant Professor, bdyoun@umd.edu, Corresponding Author

³ Graduate Student, zxi@umd.edu

⁴ Staff Researcher, artemis.kloess@gm.com

analysis and design.

To ensure the reliability of the product system, diverse design methodologies have been developed, such as Reliability-Based Design Optimization (RBDO) [Enevoldsen and Sorensen, 1994; Yu et al., 1997; Youn and Choi, 2003, 2004], Possibility-Based Design Optimization (PBDO) [Du and Choi, 2006, Choi, et. al, 2006], Evidence-Based Design Optimization (EBDO), and Bayesian RBDO [Youn and Wang, 2008]. Most such research activities have focused on how to assess reliability effectively by simply assuming probabilistic models of random system inputs without engaging raw data [Youn and Choi, 2003; Du and Chen, 2004; Youn et al., 2008]. However, less attention has been paid on how to, within a probabilistic framework, model random physical quantities with evolving, insufficient, and subjective data sets. Bayesian approaches have been widely used in many engineering and science fields where data is progressively accumulated. For example, Bayesian reliability analysis has been applied to series systems of Binomial (safe or fail) subsystems and components [Fickas, et al., 1988], to reliability assessment of power systems [Yu, et al., 1999], to the effectiveness of reliability growth testing [Quigley and Walls, 1999], to robust tolerance control and parameter design in the manufacturing process [Rajagopal, 2004], and to input uncertainty modeling by Chung et al. (2004). Two advanced Bayesian (maximum likelihood and parsimony) methods have been compared for molecular biology applications [Merl et al., 2005]. Bayesian updating has been implemented using Markov Chain Monte Carlo simulation for structural models and reliability assessment [Beck and Au, 2002]. Dynamic object oriented Bayesian networks have been proposed for complex system reliability modeling by Weber and Jouffe (2006). Despite numerous efforts, it has been a great challenge to model uncertain product performances while considering evolving, insufficient, and subjective data sets. To overcome the challenge, this research integrates probability encoding methods to a Bayesian updating mechanism. It is referred to as Bayesian Information Toolkit (BIT). Likewise, Bayesian Reliability Toolkit (BRT) is created by incorporating reliability analysis to the Bayesian updating mechanism. In this research, both BIT and BRT are integrated to predict reliability even with evolving, insufficient, and subjective data sets. With the effort in developing both BIT and BRT, the Bayesian Information, Reliability and Design (BIRD) software is developed by incorporating them with the Bayesian RBDO that the authors have developed [Youn and Wang, 2008].

In this paper, the proposed approach is applied for reliability prediction of door closing performance in a vehicle door system. The vehicle door system is of special concern due to its frequency of use and its engineering challenge with respect to design, assembly, and operation. A considerable amount of engineering effort is spent conducting hardware-based or analytical experiments to generate information for supporting engineering decisions during the vehicle development process. At the conceptual stage, the uncertainty characterization of this information is largely based on expert

judgment and data from current or past designs. As the design matures, analysis results and test data are collected to quantify the uncertainty; however, the data is usually of limited sample size. The door seal design engineer, for example, needs to know the requirements for a door seal system that isolates the passenger compartment from the external environment, while simultaneously allowing the door to be closed with minimal effort. A door system design must satisfy a multitude of functional and engineering requirements. The functional requirements are deduced from the voice of the customer, and include, for example, excellent exterior appearance/fit, interior quietness, protection from water leaks and dust intrusion, and an easy to open/close door. The functional requirements must be translated into measurable engineering requirements and the engineering solutions should be simple and include manufacturing restrictions. Due to the inherent uncertainties associated with the voice of the customer, manufacturing processes, material properties etc., engineers must seek an appropriate performance evaluation metric and corresponding method that can incorporate and evaluate the effect of those uncertainties.

NOMENCLATURE

R^B	=	Bayesian reliability
Φ	=	standard Gaussian cumulative distribution function
$F(x)$	=	cumulative distribution function
$F^{-1}(x)$	=	inverse cumulative distribution function
$f_x(x)$	=	probability density function
$f_{(\cdot \cdot)}$	=	conditional probability density function
p_{fs}	=	probability of system failure
G_i	=	function of the i^{th} constraint
C	=	cost function
$B(\alpha, \beta)$	=	beta function with parameters α and β

2. BAYESIAN INFORMATION, RELIABILITY, AND DESIGN (BIRD) TOOLKIT

This section presents the integration of probability encoding methods and reliability analysis to the Bayesian updating mechanism.

2.1 REVIEW OF BAYESIAN UPDATING TECHNIQUES

As mentioned earlier, the evolving, insufficient, and subjective data sets can be obtained through either measurement or survey during the product life cycle. To make use of the valuable information for product performance evaluation and design, BIT employs a Bayesian updating technique. This subsection gives a brief review of the Bayesian updating technique.

Let X be a random variable with probability density function $f(x, \theta)$, $\theta \in \Omega$. From the Bayesian point of view, θ is

interpreted as a realization of a random variable Θ with a probability density $f_{\Theta}(\theta)$. The density function expresses what one thinks about the occurring frequency of Θ before any future observation of X is taken, that is, a prior distribution. Based on

Bayes' theorem, the posterior distribution of Θ given a new observation X can be expressed as

$$f_{\Theta|X}(\theta|x) = \frac{f_{X,\Theta}(x,\theta)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta) \cdot f_{\Theta}(\theta)}{f_X(x)} \quad (1)$$

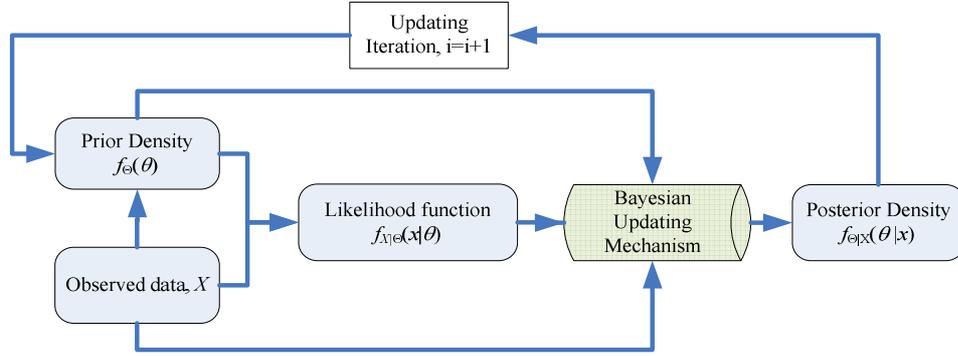


Figure 1 Process of Bayesian Updating

The Bayesian approach is used for updating information about the parameter θ . First, a prior distribution of Θ must be assigned before any future observation of X is taken. Then, the prior distribution of Θ is updated to the posterior distribution as the new data for X is employed. The posterior distribution is set to a new prior distribution and this process can be repeated with an evolution of data sets. This updating process can be briefly illustrated in Fig.1 [M. Rausand and A. Høyland, 2003].

Let us consider a normal inference model as one example to illustrate the Bayesian updating process. For the normal distribution with a known standard deviation σ , the likelihood function for the parameter θ , mean value of the normal distribution, is expressed as

$$p(X|\theta) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x_i - \theta)^2\right] \quad (2)$$

Suppose a conjugate prior distribution for the mean value parameter follow a normal distribution with the known parameter of standard deviation σ . The prior distribution has its mean, u , and variance, τ^2 . Then due to the conjugate property, the posterior distribution can be obtained through the Bayesian updating process which also follows the normal distribution with the mean and variance as [Andrew Gelman, et al]:

$$\tilde{u} = \frac{\bar{X} \cdot N \cdot \sigma^{-2} + u \cdot \tau^{-2}}{N \cdot \sigma^{-2} + \tau^{-2}}, \quad \tilde{\tau}^2 = \frac{\sigma^2 \cdot \tau^2}{\sigma^2 + N \cdot \tau^2} \quad (3)$$

Conjugate models of Bayesian updating are quite useful for uncertainty modeling with the type of data sets we discuss, since the prior and posterior distributions are given in a closed form. However, it is found that the Bayesian updating results often depend on selection of the prior distribution models in the conjugate models. To eliminate the dependency, a non-

conjugate Bayesian updating model is developed using Markov Chain Monte Carlo (MCMC) methods. This is, however, more computationally intensive.

2.2 BAYESIAN INFORMATION TOOLKIT (BIT)

As mentioned earlier, a great challenge exists in dealing with evolving, insufficient, and subjective data sets while performing reliability analysis. BIT is developed by integrating probability encoding methods to the Bayesian updating technique.

Probability Encoding Methods

To systemically extract and quantify subjective information that comes from individual judgment about uncertain quantities, the probability encoding [C.S.Spetzler, et al 1975; Wallsten and Budescu, 1983; Winkler, 1967] methods are employed in BIT. The methods employ an interview process and most are based on questions for which the answers can be represented as points on a cumulative distribution function. The different encoding methods used vary according to whether they ask a subject to assign probabilities (P), values (V), or both. The three basic types of encoding methods are listed below.

- P -Methods require the subject to respond by specifying points on the probability scale while the values are fixed.
- V -methods require the subject to respond by specifying points on the value scale while the probabilities remain fixed.
- PV -methods ask questions that must be answered on both scales jointly; the subject essentially describes points on the cumulative distribution.

Probability encoding consists of a set of questions that the subject responds to either directly by providing numbers or

indirectly by choosing between simple alternatives or bets. In the direct response mode, the subject is asked questions that require numbers as answers which will be given in the form of either values or probabilities depending on the method being used. In the indirect response mode, the subject is asked to choose between two or more bets. The bets are adjusted until the subject is indifferent to choosing between them. This indifference can then be translated into a probability or value assignment. Besides choosing between bets, another procedure is to ask the subject to choose between events defined on the value scale for the uncertain quantity, where each event represents a set of possible outcomes for the uncertain quantity. Subjective data in examples shown in Section 3 are obtained using the direct probability encoding procedure.

The following part of this section is a probability encoding example by using PV method. Two subjects are questioned about tomorrow's temperature with respect to both highest temperature value and corresponding probabilities. Results are shown in table1. By using this PV method, we can obtain a curve similar to the CDF of the temperature distribution, as shown in Figure 2, and this format of data can then be used to model the temperature distribution by Bayesian Updating technique.

Table 1 Results of the temperature survey

Subject I		Subject II	
Temp.	Prob	Temp	Prob.
22	0.10	23	0.08
25	0.16	25	0.10
26	0.40	27	0.25
27	0.55	28	0.60
28	0.75	29	0.85
29	0.90	30	0.95
30	0.95	31	0.99
31	1.00	33	1.00

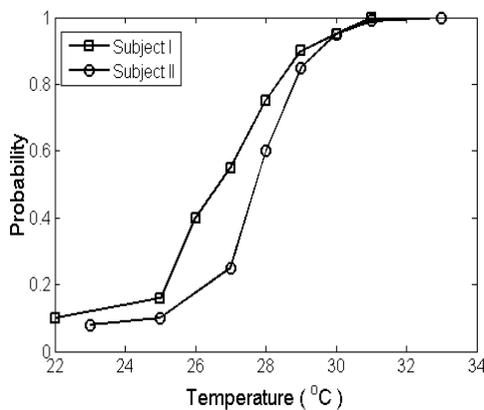


Figure 2 Results of the temperature survey

2.3 BAYESIAN RELIABILITY TOOLKIT (BRT)

In many engineering applications, outcomes of events from repeated trials can be a binary manner, such as occurrence or nonoccurrence, success or failure, good or bad, etc. In such cases, random behavior can be modeled with a discrete probability distribution model. In addition, if the events satisfy the additional requirements of a Bernoulli sequence, that is to say, if the events are statistically independent and the probability of occurrence or nonoccurrence of events remains constant, they can be mathematically represented by the *binomial distribution* [Haldar and Mahadevan, 2000]. In other words, if the probability of an event occurrence in each trial is r and the probability of nonoccurrence is $(1-r)$, then the probability of x occurrences out of a total of N trials can be described by the probability mass function (PMF) of a Binomial distribution as

$$\Pr(X = x, N | r) = \binom{N}{x} r^x (1-r)^{N-x} \quad x = 0, 1, 2, \dots, N \quad (4)$$

where the probability of success identified in the previous test, r , is the parameter of the distribution.

In Eq.(4), the probability of x/N (x occurrences out of N trials) can be calculated when a prior distribution on r is provided. This inference process seeks to update r based on the outcomes of the trials. Given x occurrences out of a total of N trials, the probability distribution of r can be calculated using Bayes' Rule as [Li et al., 2002]

$$f(r | x) = \frac{f(r) f(x | r)}{\int_0^1 f(r) f(x | r) dr} \quad (5)$$

where $f(r)$ is the prior distribution of r , $f(r | x)$ is the posterior distribution of r and $f(x | r)$ is the likelihood of x for a given r . The integral in the denominator is a normalizing factor to make the probability distribution proper. The prior distribution is known for r , prior to the current trials. In this paper, a uniform prior distribution is used to model r bounded in $[0, 1]$. However, it is possible to obtain a posterior distribution with any type of a prior distribution.

For Bayesian reliability predictions, both a prior reliability distribution (r) and the number (x) of safety occurrences out of the total number of test data sets N must be known. If the prior reliability distribution (r) is unavailable, it will be simply modeled with a uniform distribution, $r \sim U(a, b)$ where $a < b$ and $a, b \in [0, 1]$. At an early design stage, it can be modeled using reliability from the previous product designs or expert opinions. Alternatively, if the reliability distribution has been predicted with a data set in a precedent test, this reliability distribution will be used as the prior reliability distribution and updated to a posterior reliability distribution with new test data. In either of these alternative cases, reliability can be modeled with a Beta distribution, the conjugate distribution of the Bayesian binomial inference, since the uniform distribution is a special case of the Beta distribution. The PDF of the Beta distribution is expressed as

$$f(r|x) = \frac{1}{B(\alpha, \beta)} r^{\alpha-1} (1-r)^{\beta-1}, \quad (6)$$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt$$

where $\alpha = x + 1$ and $\beta = N - x + 1$. The posterior distribution, $f(p|x)$, is the Beta distribution and represents the probability distribution of reliability. It is found that the distribution is a function of x and N , the number of safety trials and the total number of trials, respectively.

When only epistemic uncertainties are engaged to assess reliability, its PDF can be modeled using the Beta distribution in Eq. (6) by counting the number of safety occurrences, x . In general, both aleatory and epistemic uncertainties generally appear in most engineering design problems. In such situations, the PDF of reliability can be similarly obtained through Bayesian reliability analysis. To build the PDF of reliability, reliability analysis must be performed at every data point for epistemic uncertainties while considering aleatory uncertainties. Different reliability measures, $R_k = R(x_{e,k})$, are obtained at different sample points for epistemic uncertainties. In Eq.(6), $\alpha = x + 1$ and $\beta = N - x + 1$, where $x = \sum R_k$. Then, the PDF of reliability r with a uniform prior distribution is updated to $R(\mathbf{X}_a, \mathbf{X}_e; \mathbf{d})$ as

$$R(\mathbf{X}_a, \mathbf{X}_e; \mathbf{d}) = f(r|\bar{\mathbf{x}}) = \frac{1}{Beta(\alpha, \beta)} r^{\alpha-1} (1-r)^{\beta-1}$$

where $\alpha = 1 + x$, $\beta = N - x + 1$, $\bar{\mathbf{x}} = \{\mathbf{x}_{e,1}, \dots, \mathbf{x}_{e,N}\}$ (7)

$$x = \sum R_k, \text{ and } R_k = \Pr[g(X_a) \leq 0 | \mathbf{x}_{e,k}]$$

N is the number of finite data sets for epistemic uncertainties. A detailed illustration is given in the following example.

For design optimization, Bayesian reliability must satisfy two requirements: (a) sufficiency and (b) uniqueness. The sufficiency requirement means that the Bayesian reliability must be smaller than an exact reliability realized with a sufficient amount of data for the input uncertainties. Then, Bayesian RBDO provides an optimum design with higher reliability than target reliability, regardless of the data size. To meet the sufficiency requirement, an extreme distribution theory for the smallest reliability value is employed to guarantee the sufficiency of reliability. R_k values are different for different data sets, $\mathbf{x}_{e,k}$, of which each has the same sample size N . Without generating expensive data sets, the extreme distribution theory determines the probability distribution of the smallest R value that guarantees the first requirement. Then, the median value of the extreme distribution uniquely determines Bayesian reliability. To satisfy both requirements, Bayesian reliability is defined as the median value of the extreme distribution for the smallest value derived from the Beta distribution in Eq.(7).

First, based on the extreme distribution theory, the extreme distribution for the smallest reliability value is constructed from the reliability distribution, Beta distribution. For random reliability R with the Beta distribution function, $F_R(r)$, let 1R be

the smallest value among N data points for random reliability, R . Then the Cumulative Distribution Function (CDF) of the smallest reliability value, 1R , can be expressed as [Singiresu S. Rao, 1997]

$$1 - F_{{}^1R}(r) = P({}^1R > r) = P({}^1R > r, {}^2R > r, \dots, {}^N R > r) \quad (8)$$

Since the i^{th} smallest reliability values, iR ($i = 1, \dots, N$), are identically distributed and statistically independent, the CDF of the smallest reliability value becomes

$$F_{{}^1R}(r) = 1 - [1 - F_R(r)]^N \quad (9)$$

Bayesian reliability, R_B , is defined as the median value of the reliability distribution. That is to say, Bayesian reliability is the solution of the nonlinear equation (Eq. (9)) by setting its left hand side to 0.5.

$$R_B = F_R^{-1} \left[1 - \sqrt[N]{1 - F_R(r^m)} \right] = F_R^{-1} \left[1 - \sqrt[N]{0.5} \right] \quad (10)$$

Bayesian reliability analysis can be conducted using the following numerical procedure:

- STEP1 Collect a limited data set for epistemic uncertainties where the data size is N .
- STEP2 Calculate reliabilities (R_k) with consideration of aleatory uncertainties at all epistemic data points.
- STEP3 Build a distribution of reliability using the Beta distribution in Eq. (7) with aleatory and/or epistemic uncertainties.
- STEP4 Construct the extreme distribution in Eq. (9) with the Beta distribution obtained in Step 3.
- STEP5 Determine the Bayesian reliability using Eq.(10).

2.4 BAYESIAN DESIGN TOOLKIT (BDT)

With the effort in developing both BIT and BRT, the Bayesian Information, Reliability and Design (BIRD) software is developed by incorporating them with Bayesian RBDO that the authors have developed [Youn and Wang, 2008]. Although this paper is not focused on Bayesian RBDO, it will be reviewed as one of BIRD modules. Knowing that both aleatory and epistemic uncertainties exist in the system of interest, Bayesian RBDO can be formulated as

$$\begin{aligned} & \text{minimize } C(\mathbf{X}_a, \mathbf{X}_e; \mathbf{d}) \\ & \text{subject to } P_B(G_i(\mathbf{X}_a, \mathbf{X}_e; \mathbf{d}) \leq 0) \geq \Phi(\beta_i), \quad i = 1, \dots, np \quad (11) \\ & \mathbf{d}^L \leq \mathbf{d} \leq \mathbf{d}^U, \quad \mathbf{d} \in R^{nd} \text{ and } \mathbf{X}_a \in R^{na}, \mathbf{X}_e \in R^{ne} \end{aligned}$$

where $P_B(G_i(\mathbf{X}_a, \mathbf{X}_e; \mathbf{d}) \leq 0) = R_i^B(\mathbf{X}_a, \mathbf{X}_e; \mathbf{d})$ is Bayesian reliability and $G_i(\mathbf{X}_a, \mathbf{X}_e; \mathbf{d}) \leq 0$ is defined as a safety event. $C(\mathbf{X}_a, \mathbf{X}_e; \mathbf{d})$ is the objective function; $\mathbf{d} = \boldsymbol{\mu}(\mathbf{X})$ is the design vector; \mathbf{X} is the random vector; β_i is the prescribed reliability target; and np , nd , na , and ne are the numbers of probabilistic constraints, design variables, aleatory random variables, and epistemic random variables, respectively. If the parameters describing a random variable are controllable among all (both

aleatory and epistemic) random variables, they are considered design variables. For instance, a random variable with a normal distribution may have two design variables, mean and standard deviation. It will be shown that the result from Bayesian RBDO asymptotically approaches that from the conventional (or Frequentist) RBDO. In other words, Frequentist RBDO is a special case of Bayesian RBDO because Bayesian RBDO is able to handle aleatory and/or epistemic uncertainties.

3. EXAMPLES

Two examples are employed to demonstrate the feasibility of Bayesian reliability analysis with evolving, insufficient, and subjective data sets: 1) a mathematical example and 2) door closing effort in the vehicle door system.

3.1 A MATHEMATICAL EXAMPLE

Let $G(X_1, X_2) = 3 - X_1^2 X_2 / 20 \leq G_0$ be an inequality constraint with two random variables where X_1 is an epistemic random variable and X_2 is an aleatory random variable, $X_2 \sim N(\mu_2=2.8, \sigma_2=0.2)$. Besides, G_0 is a random parameter which follows a normal distribution $N(2.0, 0.05^2)$ and represents the uncertainty of the target performance. In this mathematical example, the distribution for G_0 is known, however, in most practical cases, this distribution should be determined by observation.

Twenty data values are randomly sampled for X_1 from an assumed normal distribution ($\mu_1 = 2.9, \sigma_1 = 0.2$), as shown in Table 2. The table also shows the corresponding reliabilities $R_k = \Pr [G(X_2) \leq G_0 | X_1(k)]$ for $k=1, \dots, 20$ that are computed from reliability analyses. For example, $X_1(1) = 2.9277$, then $R_1 = P(3 - 2.9277^2 \times X_2 / 20 \leq G_0) = 0.97807$. Figure 3 shows the PDFs of the performance function $G(X_1 = 2.9277, X_2)$ and G_0 . By carrying out the probability analysis for all 20 epistemic data, 20 probability values are then obtained as shown in Table 2. From Table 2, the expected number of safe design points out of the twenty designs can be obtained from the sum of all twenty reliabilities, $x = \sum R_k = 17.4408$. As discussed in the previous section, the reliability can then be modeled with the Beta distribution as $Beta(18.4408, 3.5592)$ at the design point, ($\mu_1=2.9, \mu_2=2.8$). This is graphically shown in Figure 4.

To validate the results, Monte Carlo simulation (10,000 samples) is conducted by assuming X_1 to follow $N(\mu_1=2.9, \sigma_1=0.2)$. It gives the true reliability ($=0.8345$) of the design point. As shown in Figure 4, the true reliability is close to the mean value of the reliability distribution. Therefore, the reliability distribution gives a quite feasible estimate with both aleatory and epistemic uncertainties. In this example, a uniform distribution, $r \sim U(0,1)$, is used as the prior distribution of reliability. Therefore, the reliability distribution appears to be widely distributed, but it can be narrowly distributed if the prior distribution is more precisely given.

X_1	Probability	X_1	Probability
2.9277	0.97807	3.4741	1.00000
2.7605	0.76836	2.9575	0.98709
2.775	0.80247	2.9029	0.96671
3.1006	0.99929	2.9430	0.98323
2.8175	0.88239	2.8196	0.88559
2.5933	0.24267	2.9706	0.98986
3.1047	0.99936	2.7157	0.64237
2.9604	0.98775	2.6738	0.50406
3.1706	0.99986	2.8869	0.95693
2.9354	0.98082	2.8185	0.88392

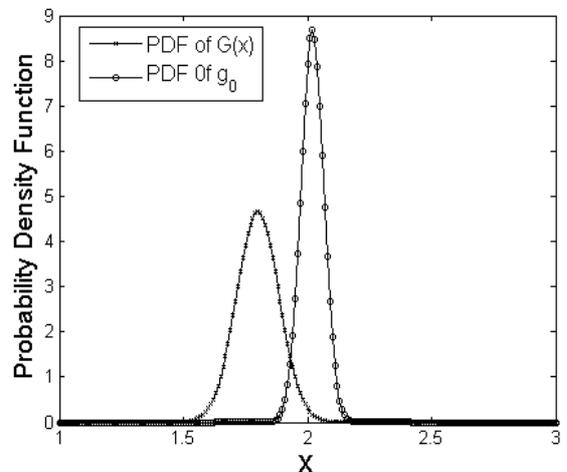


Figure 3 PDFs for $G(X_1, X_2)$ and G_0

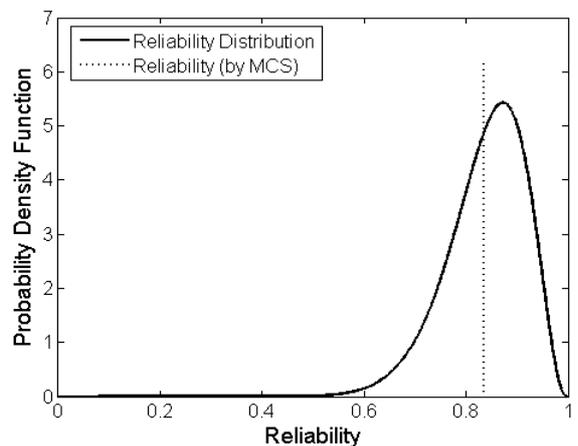


Figure 4 Actual and estimated reliability distribution

Table 2 X_1 samples and probabilities

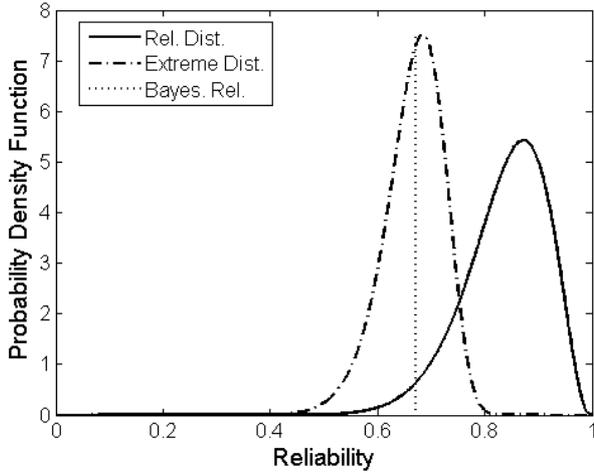


Figure 5 Bayesian reliability

Using Eq. (9), the extreme distribution for the smallest reliability value is obtained as

$$F_{i,R}(r) = 1 - \left[1 - \int_0^r \frac{1}{B(18.7066, 3.2934)} \theta^{17.7066} (1-\theta)^{2.2934} d\theta \right]^{20}$$

From Eq.(10), Bayesian reliability is calculated as $P_B = 0.6884$. The Beta distribution for reliability, its extreme distribution for the smallest reliability value, and the Bayesian reliability are graphically shown in Figure 5.

3.2 BAYESIAN RELIABILITY ANALYSIS FOR A VEHICLE DOOR SYSTEM

The demonstration problem used in this study is the body-door system of a passenger vehicle, as illustrated in Figure. 6. The vehicle door system is of special concern due to its frequency of use and its engineering challenge with respect to design, assembly, and operation. Variation exists in the CLD (Compression Load Deflection) response of the seal, the gap between the body and door, as well as in attaching the door to the car body. Besides the presence of variation, the complexity of the system is high due to the nonlinear seal behavior and the dynamics of door closing. The detail of vehicle door system regarding the problem description, failure mechanism specification, physical model creation and response surface construction can be found from Ref. Kloess et al (2004). The performance measure selected in this study to assess one aspect of door system design is the door closing effort. The measurable quantity for this performance measure is the door closing velocity. A response surface for door closing velocity was created based on results from physics-based models and the performance evaluation criteria were deduced from both expert opinions and voice of the customer information.

For the door system example in this study, 26 random input

variables are used to specify the uncertainty of the system. Within these 26 random input variables, listed in Table 3, X_5 , X_6 , X_7 , X_{25} and X_{26} are aleatory variables which, for this example, are assigned uniform distributions on different threshold values as shown in the table. Except for these five random input variables, all others are epistemic variables with a total of 79 sets of measurement data. For illustrative purpose, these epistemic data are partially listed in Table 4.

In the following two subsections, we describe the modeling of the performance evaluation criteria, i.e. the marginal velocity, using the Bayesian updating technique introduced in the previous section followed by the Bayesian reliability analysis carried out for the door closure problem.

Modeling of the Marginal Velocity

In this subsection, the marginal velocity which serves as the criteria of the door performance evaluation is modeled by using the Bayesian updating technique based on expert opinion and the customer data. From a hypothetical expert, the door closing velocity values for customer satisfaction should be, for example, within the range of 0 m/s to v_{max} m/s. Customer survey regarding the door closing velocity can be carried out by using the direct customer survey method [C.S.Spezler, et. al., 1975] and illustrative results which show the Customer Rejection Rate (CRR) versus the door closing velocity (normalized by v_{max}) are graphically shown in Fig. 7. For the modeling of the marginal velocity, CRR can be treated as the probability of the marginal velocity being smaller than a given a or $CRR = P(v_m \leq a)$ where v_m is a random marginal velocity and a is within $[0, v_{max}]$ based on expert opinion.

The procedure of marginal velocity modeling can be briefly summarized into three steps. First, based on the customer data, one Bayesian inference model should be specified. For example, if the Bayesian normal inference model is used, the marginal velocity will be modeled as the mean value of the normal distribution which is the conjugate distribution for this model. Second, based on the selected model, the CDF analysis can be carried out for the CDF/ Velocity data. After completing this analysis, the CDF data are then transferred to parameter data for the distribution. Third, with one prior distribution assumed, Bayesian updating can then be carried out with sets of parameter data.

In this study, the Bayesian normal inference model will be used and the marginal velocity will be modeled as the mean value of a Normal distribution. As introduced in the second section of this paper, we suppose that the marginal velocity also follows a Normal distribution, which is the conjugate distribution of the normal inference model. Expert opinion is used in modeling the prior information on the marginal velocity. To properly model the normal distribution with the information from the expert, the six sigma region of the normal distribution is set to the interval, such that $[\mu - 3\sigma = 0, \mu + 3\sigma = v_{max}]$. Although the domain of the normal distribution is $[-\infty, +\infty]$, the contribution out of the bound $[0, v_{max}]$ is negligible.

Normalizing by v_{\max} , the distribution $N(0.5, 0.1667)$ is used for the prior distribution of this model.

As an example to show how the CDF analysis is carried out, we use a set of data, e.g., normalized velocity is 0.43 and customer satisfaction rate is 91.8%, from clinic A's customer data shown in Figure 7. The 8.2% customer rejection rate will be considered as the CDF value corresponding to the velocity value 0.43. As we suppose $\sigma = 0.1667$, then based on the CDF data $Z_{0.082} = 0.43$, we can determine that the mean value of the normal distribution is 0.662. For each set of customer data, the corresponding parameter data is determined by the CDF analysis. Three sets of parameter data are then obtained from the three sets of customer data after the CDF analysis. The Bayesian Normal Inference can be expressed as

$$\mu_1 = \left(\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^N X_i}{\sigma^2} \right) / \left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \right), \quad \sigma_1^2 = \left(\frac{1}{\sigma_0^2} + \frac{N}{\sigma^2} \right)^{-1}$$

where μ_1, σ_1 are parameters for the posterior distribution whereas μ_0, σ_0 are parameters for the prior distributions, X_i is the i th parameter data and σ is the population variance. Based on the Bayesian Normal Inference, the PDFs for the marginal velocity can then be gradually refined by aggregating three different clinic data sets with the normal prior distribution, shown in Figure 8. With the clinic-A data set, the first Bayesian model for the marginal velocity is the posterior distribution I, $N(0.559, 0.068^2)$, shown in Fig. 8. Then this posterior distribution is treated as the prior distribution and combined with the clinic-B data set, to obtain the second Bayesian model, posterior distribution II, $N(0.606, 0.0445^2)$. Similarly with the clinic-C data set, the final Bayesian model is obtained as the posterior distribution III, $N(0.5946, 0.0355^2)$, as shown in Figure 8. Figure 9 shows the PDF and CDF of the final Bayesian model, $N(0.5946, 0.0355^2)$, for the marginal velocity.

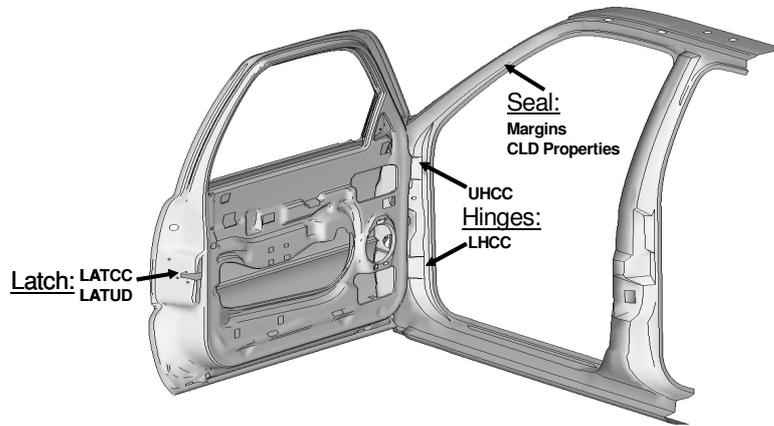


Figure 6 Vehicle Door system

Table 3 Random variables and descriptions

Variable Name	Description	Variable Type
X_1	UHCC- Upper hinge location in cross-car direction	Epistemic
X_2	LHCC- Lower hinge location in cross-car direction	Epistemic
X_3	LATCC-Latch location in cross-car direction	Epistemic
X_4	LATUD-Latch location in up-down direction	Epistemic
X_5	Primary seal CLD property factor	U(0.7, 1.3)
X_6	Auxiliary seal CLD property factor	U(0.7, 1.3)
X_7	Cutline seal CLD property factor	U(0.7, 1.3)
X_8	Primary Seal Margin Region 1	Epistemic
X_9	Primary Seal Margin Region 2	Epistemic
X_{10}	Primary Seal Margin Region 3	Epistemic
X_{11}	Primary Seal Margin Region 4	Epistemic
X_{12}	Primary Seal Margin Region 5	Epistemic
X_{13}	Primary Seal Margin Region 6	Epistemic

X ₁₄	Primary Seal Margin Region 7	Epistemic
X ₁₅	Primary Seal Margin Region 8	Epistemic
X ₁₆	Primary Seal Margin Region 9	Epistemic
X ₁₇	Primary Seal Margin Region 10	Epistemic
X ₁₈	Primary Seal Margin Region 11	Epistemic
X ₁₉	Primary Seal Margin Region 12	Epistemic
X ₂₀	Primary Seal Margin Region 13	Epistemic
X ₂₁	Primary Seal Margin Region 14	Epistemic
X ₂₂	Primary Seal Margin Region 15	Epistemic
X ₂₃	Primary Seal Margin Region 16	Epistemic
X ₂₄	Primary Seal Margin Region 17	Epistemic
X ₂₅	Auxiliary Seal Margin	U(-1, 1)
X ₂₆	Cutline Seal Margin	U(-1, 1)

Table 4 Data for Epistemic Random Variables

Variables	Data							
	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6	...	Set 79
X ₁	1.62	2.29	1.58	1.58	1.19	1.696667	...	2.16
X ₂	2.82	2.49	1.8	2.1	2.03	1.365	...	1.355
X ₃	2.555	2.1	1.82	1.67	1.75	1.010714	...	1.35
X ₄	-0.38	-0.35	-0.01	-0.01	0.61	-0.43833	...	-0.61
X ₈	1.655	1.235	1.015	0.715	0.71	-0.05583	...	0.559167
X ₉	1.0775	0.7725	0.5925	0.2825	0.115	-0.31375	...	0.39875
X ₁₀	0.5	0.31	0.17	-0.15	-0.48	-0.57167	...	0.238333
X ₁₁	1.24	0.74	0.426667	0.113333	-0.23	-0.02167	...	0.955
X ₁₂	-0.27	-0.31	-0.28	-0.66	-1.29	-0.09167	...	0.278333
X ₁₃	0.03	0.16	-0.205	-0.29	-1.02	-0.3125	...	0.1125
X ₁₄	0.33	0.63	-0.13	0.08	-0.75	-0.53333	...	-0.05333
X ₁₅	0.5	0.79	0.06	0.22	-0.76	-0.345	...	0.135
X ₁₆	0.89	1.01	0.87	0.27	-0.63	0.02	...	0.24
X ₁₇	0.27	0.51	-0.01	-0.21	-1.565	-0.18667	...	0.233333
X ₁₈	-0.35	0.01	-0.89	-0.69	-2.5	-0.39333	...	0.226667
X ₁₉	-0.35	0.01	-0.89	-0.69	-2.5	-0.39333	...	0.226667
X ₂₀	-0.44	-0.53	-1.27	-1.55	-2.93	-0.76667	...	-0.37667
X ₂₁	-0.44	-0.53	-1.27	-1.55	-2.93	-0.76667	...	-0.37667
X ₂₂	0.16	-0.03	-0.7125	-0.8625	-1.6825	-0.17125	...	0.12375
X ₂₃	0.76	0.47	-0.155	-0.175	-0.435	0.424167	...	0.624167
X ₂₄	1.49	0.91	0.56	0.27	0.91	0.075	...	0.535

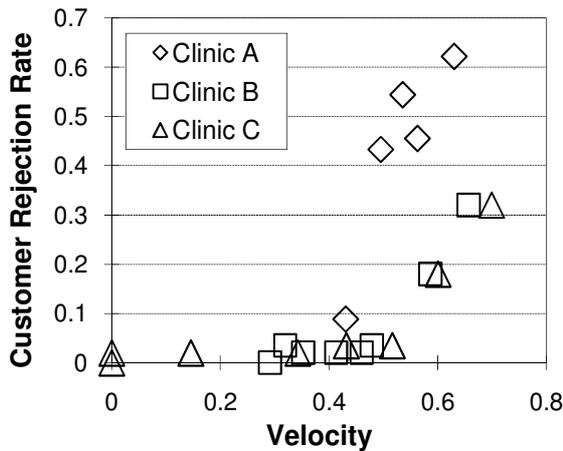


Figure 7 Customer Rejection Rate

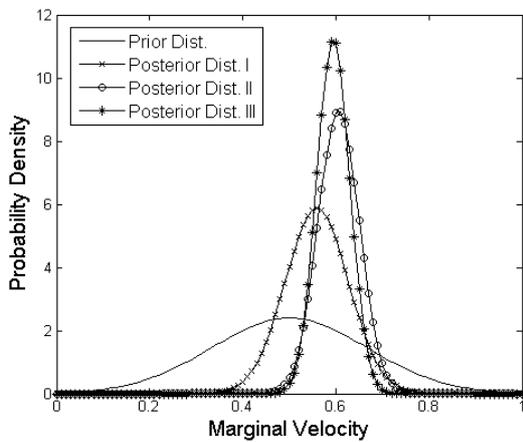


Figure 8 Bayesian Updating for the Marginal Velocity using a Normal Distribution

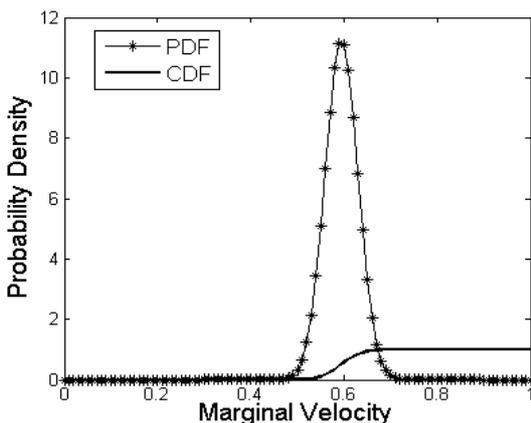


Figure 9 Bayesian Model for the Marginal Velocity using a Normal Distribution

Bayesian Reliability Analyses for a Vehicle Door System

Based on the marginal velocity PDF created, Bayesian reliability analysis is then carried out for the door closing effort problem with both aleatory and epistemic uncertainties. For a given set of input values, the performance response can be obtained from the response surface created based on the physical model [A. Kloess et al, 2004]. Since Bayesian reliability analysis requires the probabilistic performance evaluation for each set of epistemic data, two different approaches, Monte Carlo Simulation (MCS) and Eigenvector Dimension Reduction (EDR) method [Youn, et. al., 2008], are employed in this study to calculate the reliability for each set of epistemic data. EDR method is an efficient and accurate sensitivity free method for reliability analysis. Results for the door closing effort problem in this study from MCS and EDR are compared.

First, for each set of epistemic data, direct Monte Carlo Simulation is used to carry out the reliability analysis. For each aleatory variable (including the variable of marginal velocity), 10,000 samples are generated and used for MCS. Table 5 shows the 55 reliabilities corresponding to the first 55 sets of epistemic data. Based on Table 5, we carried out the Bayesian reliability analysis and obtained the reliability distribution as Beta (53.524, 3.476). Then by the Bayesian reliability definition described in Section 2.3, the extreme distribution of the smallest value for the Beta distribution is constructed and the Bayesian reliability is realized as 0.849185. Figure 10 shows the Beta distribution, extreme distribution and the Bayesian reliability value. With 24 new data sets involved for the epistemic random variables the Bayesian reliability is updated. The updated reliability distribution is Beta (77.1869, 3.8131) and the Bayesian reliability is updated from the original 0.849185 to 0.880935. Table 7 shows the reliabilities corresponding to each set of the new involved data. Figure 11 shows the updated Beta distribution, extreme distribution and the Bayesian reliability.

As we can see from the Monte Carlo Simulation method, the reliability analysis for each set of epistemic data can require a large amount of response performance evaluations depending on the simulation sample size (in this case 10,000). In order to make the calculation of the Bayesian reliability more efficient, the EDR method is used for the probability calculation for each set of epistemic data. By using EDR method, the total number of the response performance evaluation is reduced from 10,000 to $2n+1=13$. Based on the marginal velocity PDF created in subsection 1, the reliability R_i of a certain design (X_a, X_e^i) can be formulated as $R_i = \Pr [V(X_a, X_e^i) - V_t \leq 0]$ where $V(X_a, X_e^i)$ is the performance velocity variable corresponding to a certain design (X_a, X_e^i) , X_a is the aleatory variable set and X_e^i is the i th set of epistemic data, and V_t is the marginal velocity. Totally 55 different reliabilities corresponding to 55 different sets of epistemic uncertainties are realized as shown in Table 6. Based on these results, the reliability distribution is obtained as Beta (53.5076, 3.4924) from Bayesian inference. Then by the Bayesian reliability definition, the extreme distribution of

smallest value for the Beta distribution is constructed and the Bayesian reliability is realized as 0.848752. Figure 12 shows the Beta distribution, extreme distribution and the Bayesian reliability. With 24 new data sets involved, the Bayesian reliability is updated. The updated reliability distribution is Beta (77.1567, 3.8433) and the Bayesian reliability is updated from the original 0.848752 to 0.880363. Table 8 shows the reliabilities corresponding to each set of the new involved data. Figure 13 shows the updated Beta distribution, extreme distribution and the Bayesian reliability.

A comparison of the results from using the two different probability analysis approaches shows that the EDR method maintains good accuracy and at the same time provides a higher computational efficiency compared with MCS. From the analysis results obtained with both MCS and the EDR method,

two points are clear: first, Bayesian reliability increases with the increase of the reliability value corresponding to each set of epistemic data; secondly, the updated Bayesian reliability increases with the addition of more epistemic data into the Bayesian reliability analysis. This is because the Bayesian reliability represents not only the design uncertainty of the system but also the uncertainty due to the limiting information represented by the epistemic uncertainties. As more data is involved, a better understanding of the characteristic of epistemic uncertainties can be expected and consequently a higher Bayesian reliability can be realized. Also, the Bayesian reliability analysis approach proposed in this paper offers a convenient and effective method for the performance evaluation of the problems involving several different types of uncertainty and where uncertainty data are continuously collected.

Table 5 55 reliabilities corresponding to 55 epistemic data sets (by MCS)

Data Set	Rel.								
1	0.9973	12	1.0000	23	0.9987	34	0.9995	45	0.9988
2	1.0000	13	0.9993	24	0.9970	35	0.9998	46	0.2703
3	0.9993	14	1.0000	25	1.0000	36	0.9999	47	0.9987
4	0.9945	15	1.0000	26	0.9951	37	0.9999	48	1.0000
5	0.8265	16	1.0000	27	0.9970	38	0.9974	49	0.9955
6	0.9996	17	0.9999	28	0.9899	39	0.9977	50	0.9937
7	0.9985	18	0.9991	29	0.9998	40	0.9918	51	0.9918
8	1.0000	19	0.9999	30	1.0000	41	0.9007	52	1.0000
9	1.0000	20	0.9993	31	0.9993	42	0.9976	53	0.9994
10	1.0000	21	1.0000	32	1.0000	43	0.9778	54	0.2109
11	1.0000	22	0.9999	33	0.9963	44	0.9730	55	0.4436

Table 6 55 reliabilities corresponding to 55 epistemic data sets (by EDR)

No. Data Set	Rel.								
1	0.9978	12	1.0000	23	0.9991	34	0.9998	45	0.9993
2	1.0000	13	0.9997	24	0.9976	35	0.9998	46	0.2642
3	0.9996	14	1.0000	25	1.0000	36	1.0000	47	0.9992
4	0.9953	15	1.0000	26	0.9963	37	1.0000	48	1.0000
5	0.8243	16	1.0000	27	0.9977	38	0.9982	49	0.9963
6	0.9998	17	0.9999	28	0.9893	39	0.9984	50	0.9944
7	0.9991	18	0.9995	29	0.9998	40	0.9917	51	0.9915
8	1.0000	19	1.0000	30	1.0000	41	0.8938	52	1.0000
9	1.0000	20	0.9996	31	0.9996	42	0.9984	53	0.9997
10	1.0000	21	1.0000	32	1.0000	43	0.9755	54	0.2070
11	1.0000	22	0.9999	33	0.9971	44	0.9702	55	0.4394

Table 7 24 reliabilities corresponding to 24 new data sets (by MCS)

Data Set	Rel.	Data Set	Rel.	Data Set	Rel.
1	0.9929	9	0.9996	17	1.0000
2	0.9999	10	0.9989	18	1.0000
3	0.9995	11	1.0000	19	1.0000
4	0.9993	12	1.0000	20	0.8973
5	0.9993	13	0.8864	21	0.9842
6	0.9994	14	1.0000	22	0.9866
7	0.9996	15	0.9963	23	0.9998
8	0.9999	16	0.9240	24	1.0000

Table 8 24 reliabilities corresponding to 24 new data sets (by EDR)

Data Set	Rel.	Data Set	Rel.	Data Set	Rel.
1	0.9928	9	0.9998	17	1.0000
2	0.9999	10	0.9993	18	1.0000
3	0.9997	11	1.0000	19	1.0000
4	0.9996	12	1.0000	20	0.8915
5	0.9996	13	0.8814	21	0.9835
6	0.9997	14	1.0000	22	0.9867
7	0.9998	15	0.9969	23	0.9999
8	0.9999	16	0.919	24	1.0000

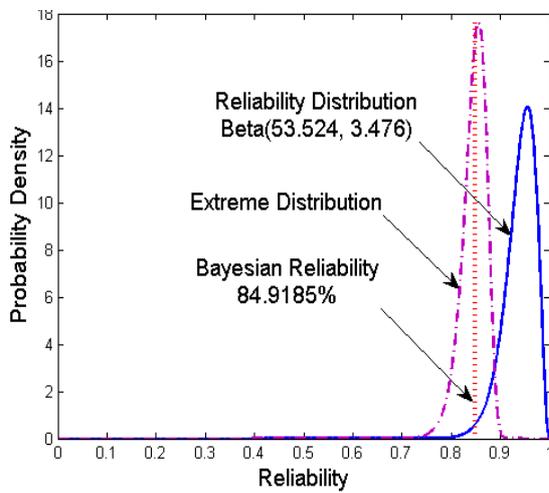


Figure 10 Bayesian Reliability with 55 sets data (by MCS)

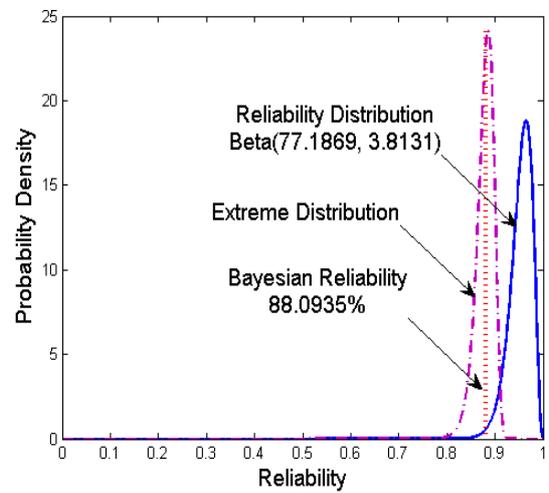


Figure 11 Updated Bayesian Reliability w/ 24 new data sets (by MCS)

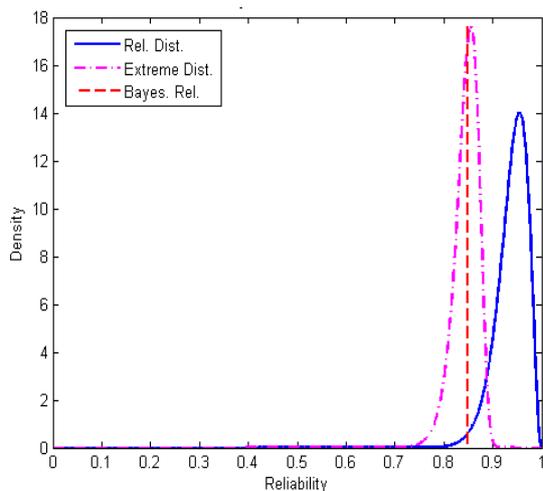


Figure 12 Bayesian Reliability with 55 data sets (by EDR)

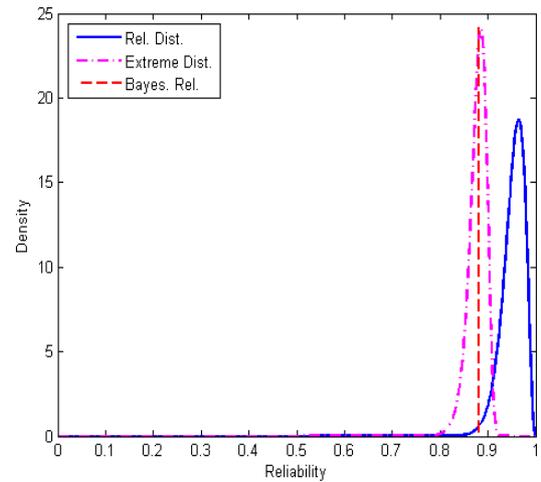


Figure 13 Updated Bayesian Reliability w/ 24 new data sets (by EDR)

4. CONCLUSION

This research presented a new paradigm of reliability prediction that enables the use of evolving, insufficient, and subjective data sets (from expert knowledge, customer survey, system inspection & testing, and field data) potentially over the entire product life-cycle. To predict reliability amidst various uncertainties with evolving, insufficient, and subjective data sets, the Bayesian updating mechanism was integrated with the probability encoding methods and reliability analysis. Such integration created Bayesian Information Toolkit (BIT) and Bayesian Reliability Toolkit (BRT). With both BIT and BRT, the Bayesian Information, Reliability and Design (BIRD) software is developed by incorporating them with the Bayesian RBDO. It was shown that the proposed Bayesian reliability analysis can predict the reliability of the door closing performance in the vehicle body-door subsystem where the relevant data sets availability are limited, subjective, and evolving.

ACKNOWLEDGMENTS

The authors would like to acknowledge that this research is partially supported by US National Science Foundation (NSF) under Grant No. GOALI-0729424, U.S. Army TARDEC by the STAS contract (TCN-05122), and by General Motors under Grant No. TCS02723. Also, the authors would like to give grateful thanks to Dr. John Cafeo from GM R&D Center for reviewing the paper and providing valuable comments.

REFERENCES

- Andrew Gelman, John B. Carlin, Hal S. Stern and Donald B. Rubin, "Bayesian Data Analysis", Second Edition, 2004 by Chapman & Hall/CRC.
- A. Kloess, Z.P. Mourelatos and P.R. Meernik, "Probabilistic analysis of an automotive body-door system", *Int.J. of Vehicle Design*, Vol.34, No.2, pp.101-125, 2004.
- Beck, J.L., and Au, S.-K., "Bayesian Updating of Structural Models and Reliability Using Markov Chain Monte Carlo Simulation," *Journal of Engineering Mechanics*, pp. 380-391, April, 2002.
- Carl S. Spetzler, Carl-Axel S. Stael Von Holstein, "Probability Encoding in Decision Analysis", *Management Science*, Vol. 22, No. 3. (Nov., 1975), pp. 340-358.
- Choi, K.K., Du, L., and Youn Byeng D., "An Inverse Possibility Analysis Method for Possibility-Based Design Optimization," *AIAA Journal*, 2006, Vol.44 No (11), pp 2682-2690.
- Chung, T.H., Mohamed, Y., and AbouRizk, S., "Simulation Input Updating Using Bayesian Techniques," *Proceedings of the 2004 Winter Simulation Conference*, pp. 1238-1243, 2004.
- Du, L., and Choi, K.K., "Possibility-Based Design Optimization Method for Design Problems with both Statistical and Fuzzy Input Data," *ASME Journal*, 2006, Vol.128 No.4, pp 928-935.
- Du, X. and Chen, W., "Sequential Optimization and Reliability Assessment for Probabilistic Design," *ASME Journal of Mechanical Design*, Vol. 126, No. 2, pp. 225-233, 2004
- Enevoldsen, I. and Sorensen, J.D., "Reliability-Based Optimization in Structural Engineering," *Structure Safety*, Vol. 15, pp. 169-196, 1994.
- Fickas, E.T., Martz, H.F., and Waller, R.A., "Bayesian Reliability Analysis of Series Systems of Binomial Subsystems and Components," *Technometrics*, Vol. 30, No. 2, pp. 143-154, May, 1988.
- Haldar, A., and Mahadevan, S., 2000, *Probability, Reliability, and Statistical Methods in Engineering Design*, John Wiley & Sons, Inc., Ch. 4.
- Li, G.Y., Wu, Q.-G., and Zhao, Y.H., 2002, "On Bayesian Analysis of Binomial Reliability Growth," *J. Japan Statist. Soc.*, Vol. 32, No. 1, pp.1-14.
- Marvin Rausand and A. Høyland, "System Reliability Theory: Models, Statistical Methods, and Applications, Second Edition", Ch.13, Wiley-Interscience, ISBN-13: 978-0471471332, 2003.
- Merl, D., Escalante, A., and Prado, R., "Comparison of Bayesian, Maximum Likelihood and Parsimony Methods for Detecting Positive Selection," *AMS2005-03 Report*, 2005.
- Quigley, J., and Walls, L., "Measuring The Effectiveness of Reliability Growth Testing," *Quality and Reliability Engineering International*, Vol. 15, pp. 87-93, 1999.
- Rajagopal, R., "Bayesian Methods for Robustness in Process Optimization," Ph.D. Thesis, Industrial Engineering and Operations Research, Penn. State University, 2004.
- Singiresu S. Rao, 1992, *Reliability-Based Design*, McGraw-Hill, Inc., Ch. 4.
- Wallsten, T. S. and Budescu, D. V., "Encoding Subjective Probabilities: A Psychological and Psychometric Review," *Management Science*, Vol. 29, No. 2, February 1983.
- Weber, P., and Jouffe, L., "Complex System Reliability Modelling with Dynamic Object Oriented Bayesian Networks (DOOBN)," *Reliability Engineering and System Safety*, Vol. 91, pp. 149-162, 2006.

Winkler, R. L., "The assessment of Prior Distributions in Bayesian Analysis," *J. Amer. Statist. Assoc.* Vol. 62 (1967), pp. 776-800.

Youn, B.D., Choi, K.K., and Park, Y.H., "Hybrid analysis method for reliability-based design optimization," *Journal of Mechanical Design, ASME*, Vol. 125, No. 2, 2003, pp 221-232.

Youn, B.D. and Choi, K.K., "Selecting probabilistic approaches for reliability-based design optimization," *AIAA Journal*, Vol. 42, No. 1, 2004.

Youn, B.D. and Wang, P.F., "Bayesian Reliability-Based Design Optimization Using Eigenvector Dimension Reduction Method," In Press, *Struct. Multidisc. Optim.*, 2008.

Youn, B.D., Zhimin, X., and Wang, P.F., "The Eigenvector Dimension-Reduction (EDR) Method for Sensitivity-Free Uncertainty Quantification," In Press, *Struct. Multidisc. Optim.*, 2008.

Yu, X, Choi, K.K., and Chang, K.H., "A Mixed Design Approach for Probabilistic Structural Durability," *Journal of Structural Optimization*, Vol.14, No. 2-3, pp. 81-90, 1997.

Yu, D.C., Nguyen, T.C., and Haddawy, P., "Bayesian Network Model for Reliability Assessment of Power Systems," *IEEE Transactions on Power Systems*, Vol. 14, No. 2, pp. 426-432, May, 1999