Struct Multidisc Optim (2015) 51:573–583
DOI 10.1007/s00158-014-1155-2

RESEARCH PAPER

A framework of model validation and virtual product qualification with limited experimental data based on statistical inference

Byung C. Jung · Jungho Park · Hyunseok Oh · Jisun Kim · Byeng D. Youn

Received: 31 December 2013 / Revised: 22 June 2014 / Accepted: 29 July 2014 / Published online: 15 October 2014

Abstract Virtual testing is a recent engineering development trend to design, evaluate, and test new engineered products. This research proposes a framework of virtual testing based on statistical inference for new product development comprising of three successive steps: (i) statistical model calibration, (ii) hypothesis test for validity check and (iii) virtual qualification. Statistical model calibration first improves the predictive capability of a computational model in a calibration domain. Next, the hypothesis test is performed with limited observed data to see if a calibrated model is sufficiently predictive for virtual testing of a new product design. An area metric and the u-pooling method are employed for the hypothesis test to measure the degree of mismatch between predicted and observed results while considering statistical uncertainty in the area metric due to the lack of experimental data. Once the calibrated model becomes valid, the virtual qualification process can be executed with a qualified model for new product developments. The qualification process builds a design decision matrix to aid in rational decision-making for product design alternatives. The effectiveness of the proposed framework is demonstrated through the case study of a tire tread block.

Keywords Model validation · Virtual qualification · Model calibration · Validity check

B. C. Jung
Department of System Dynamics, Korea Institute of Machinery and Materials, Daejeon 305-343, South Korea

J. Park · H. Oh · J. Kim · B. D. Youn (✉)
Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul 151-742, South Korea
e-mail: bdyoun@snu.ac.kr

Nomenclature

D	critical value of area metric
e	observation error
L	likelihood function
l	number of the known variables
p	number of the unknown variables
q	number of the controllable variables
U_m	area metric
y	observed model
α	significance level for a hypothesis test
$\boldsymbol{\beta}$	known model variable vector
δ	prediction error
$\boldsymbol{\zeta}$	operation variable vector
$\boldsymbol{\Theta}$	hyper-parameter vector
$\boldsymbol{\theta}$	unknown model variable vector
μ	friction coefficient
υ	pressure exponential parameter
φ	contact pressure
ψ	predicted model

1 Introduction

Increased customer expectations have resulted in new product developments at an ever increasing pace. The product development process is traditionally conceived of as a cost-intensive and time-consuming process because it requires repeated product prototyping and testing to improve product performances and reliability. The left hand side box in Fig. 1 shows a product development process from survey of customer requirement to product recycling. As products are becoming more complex with a shorter product lifecycle, virtual product testing using computer simulation has become more important to design and evaluate a new engineered product as shown in Fig. 1. However, it is still challenging to build highly

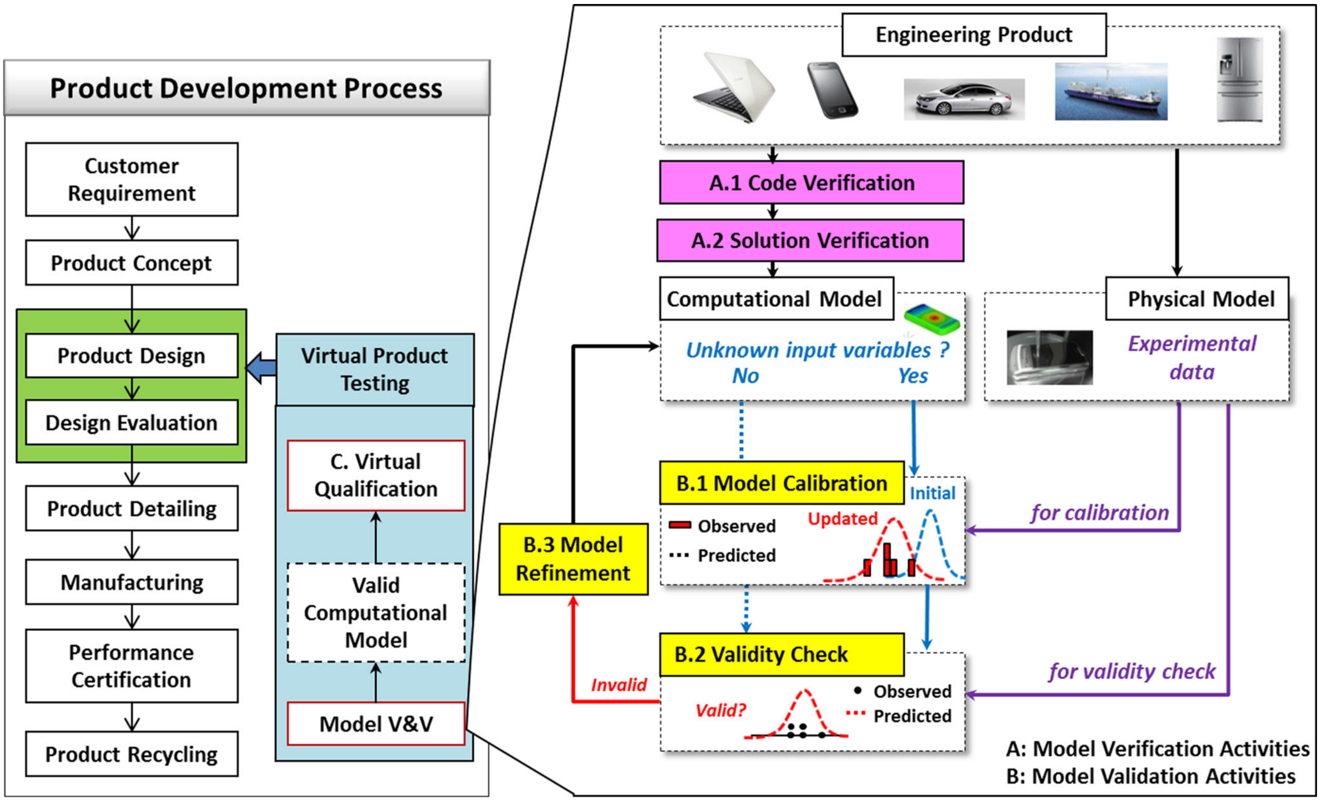


Fig. 1 A framework of model validation and virtual production qualification (Youn et al. 2011)

predictive computational models because of our limited knowledge about the models. To overcome this challenge, considerable attention has been paid to develop verification and validation (V&V) methodology that improves and assesses predictive capability of computational models.

Among various works on model V&V, the survey articles of AIAA (1998), Oberkampf et al. (2004), Thacker et al. (2004), Babuska and Oden (2004), ASME (2006), Xiong et al. (2009), and Youn et al. (2011) explain the state-of-the-art concepts and processes in detail. Based on the references, authors have redefined model V&V activities with five steps as shown in Fig. 1: (A.1) code verification, (A.2) solution verification, (B.1) model calibration, (B.2) validity check and (B.3) model refinement.

Model verification deals with the relationship between the mathematical model and its programmed implementation in the code (computational model) with two activities: (1) code verification and (2) solution verification (Oberkampf et al. 2004; Roache 1998). The goal of the code verification is to confirm that the mathematical model works as intended by eliminating programming and implementation errors. The solution verification is to evaluate the accuracy of the discrete solution of the mathematical model by estimating the numerical errors due to discretization approximations.

In this paper, model validation is defined as an activity to improve and assess the accuracy of computational results in comparison with experimental data through model calibration, validity check and model refinement as highlighted with yellow boxes in Fig. 1. Model calibration is an activity to adjust a set of unknown model variables associated with a computational model while maximizing the agreement between the predicted and observed outputs (Xiong et al. 2009; Trucano et al. 2006; Cho and Jung 2007; Datta 2005). In a deterministic sense, model calibration is thought of as the adjustment of a few model variables to minimize the discrepancy between the predicted and observed results. However, the deterministic approach is not appropriate since it cannot predict engineered product behaviors under various inherent uncertainties in manufacturing tolerance, material properties, loading condition, boundary condition, and other physical properties. Statistical model calibration, on the contrary, determines the probability distributions of unknown model input variables to maximize the agreement between the predicted and observed responses in a statistical sense (Campbell 2006). It can employ maximum likelihood estimation (MLE) (Youn et al. 2011; Xiong et al. 2009) and Bayesian statistics (Chen et al. 2008; Kennedy and O'Hagan 2002; Higdon et al. 2008; Liu et al. 2008; Wang et al. 2009).

Validity check is an activity to statistically and quantitatively determine the degree of the validity of computational model by comparing the observed (or experimental) results with predicted (or computational) ones. Although some validity check methods including a confidence interval approach (Hills and Truncano 2002; Chen et al. 2004), Bayesian method (Chen et al. 2008; Zhang and Mahadevan 2003), mean-based comparisons (Oberkampf and Barone 2006) and u-pooling method (Ferson et al. 2008) have been developed, it is still a challenge to consider statistical uncertainty in a validity check metric due to a limited sample size of experiments.

The model validation stops only when acceptable agreement between experimental and computational results is achieved. If the validity check turns out to be invalid, model refinement should be performed by polishing a mathematical or computational model of an engineered product through reconsideration of its physical behavior (e.g., governing equations and related mathematical expressions). The feedback information collected from model calibration and validity check can be used for model refinement. Finally, if the model validation builds a valid computational model successfully, a virtual qualification can be performed for new product designs.

Although researches on model validation have been studied in various engineering field, there is still open discussion about systematic approaches for successful virtual qualification to enhance, evaluate and employ computational models statistically developed with limited experimental data. This paper thus aims at proposing a framework of virtual product testing that includes statistical model calibration for unknown model parameters, validity evaluation under limited experimental data, and virtual qualification of new designs. This paper is organized as follows. Section 2 presents the three-step framework of the virtual testing as: (1) statistical model calibration, (2) hypothesis test for validity check, and (3) virtual qualification. The merits of the proposed framework are demonstrated with one case study in Section 3: a tire tread block simulation for tire friction modeling.

2 A statistical framework of virtual testing

This section presents the three-step framework of the virtual testing: (i) statistical model calibration, (ii) hypothesis test for validity check, and (iii) virtual qualification.

2.1 Statistical model calibration

Model calibration is not trivial because a computational model contains many unknown model variables, such as material properties and boundary conditions. This difficulty underscores the need of a systematic approach for statistical model calibration. The authors proposed statistical model calibration techniques with three sequential steps in Ref.(Youn et al. 2011): (1) model calibration planning, (2) model variable characterization and (3) model calibration execution.

1) Model calibration planning: Calibration experts first identify calibration resources such as performance of interest (POI), required computational models, simulation tools, experimental tests, and modeling details of an engineered product based on time and available budget. Next, the model variables must be carefully examined, and classified as known and unknown model variables. In general, a computational model carries many random model variables. If the variability of model variables can be directly characterized with observed data, the variables are categorized as known model variables (Jung et al. 2009). Otherwise, the variables are grouped as unknown model variables. Sensitivity studies can help to reduce the number of the model variables by leaving out unimportant variables for model calibration. For a complex system, it is recommended to divide the model calibration problems into several sub-problems based on the POIs of an engineered product. Finally, the statistical model calibration requires uncertainty propagation (UP) analysis such as the eigenvector dimension reduction (EDR) method (Youn et al. 2008; Choi et al. 2007) to develop the statistical responses of a computational model.
2) Model variable characterization: Material properties and physical parameters categorized as known model variables should be statistically characterized with experimental data tested with multiple specimens. Three steps were employed as below.
   - Step 1: To obtain optimum distribution parameters for candidate distribution types (e.g., normal, lognormal, Weibull and gamma distributions) using one of the point estimation methods. Maximum likelihood method is used in this paper (Modarres et al. 1999).
   - Step 2: To perform a quantitative hypothesis test for the candidate distributions such as Chi-Square Goodness-of-Fit (GoF) test and Kolmogorov-Smirnov (K-S) GoF test (Modarres et al. 1999).
   - Step 3: To select the distribution with the maximum p-value from the accepted distributions at step 2.
3) Model calibration execution: Model calibration execution adjusts a set of unknown model variables so that the agreement is maximized between the predicted and observed results. The relationship between the observed model (y) and the predicted model (ψ) can be defined as (Kennedy and O'Hagan 2001)

$$y = \Psi(\beta, \theta, \zeta) + \delta(\zeta) + e(\zeta) \quad \beta \in \Omega^l, \theta \in \Omega^p, \zeta \in \Omega^q \qquad (1)$$

where $\boldsymbol{\beta} = \{\beta_1, \beta_2, \cdot\cdot, \beta_l\}$ is the known model variable vector; $\boldsymbol{\theta} = \{\theta_1, \theta_2, \cdots, \theta_p\}$ is the unknown model variable vector; $\boldsymbol{\zeta}$ is the operation variable vector (e.g., operational conditions, environmental temperature); l, p and q are the number of the known, unknown and operation variables. δ and e are the prediction and observation (experiment) errors, respectively. The uncertainty of unknown model variable vector ($\boldsymbol{\theta}$) can be represented by statistical parameters of a suitable distribution. For example, in case of a normal distribution, the hyper-parameter vector ($\boldsymbol{\Theta}$) is defined as $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\sigma}_{\boldsymbol{\theta}}\}$, which includes mean and standard deviation of $\boldsymbol{\theta}$. Then, $\boldsymbol{\Theta}$ will be the calibration parameter vector in the model calibration execution. The distribution types, such as normal, lognormal, Weibull, etc., of the model variable vector can be assumed or determined based on both historic data and expert opinions. Next, the calibration parameter vector ($\boldsymbol{\Theta}$) will be determined by maximizing the agreement between the predicted and observed results as:

$$\text{maximize } L\left(\boldsymbol{\Theta}\middle|\mathbf{y}\right) = \sum_{i=1}^{n} \log_{10}\left[f\left(y_i\middle|\boldsymbol{\Theta}\right)\right] \qquad (2)$$

where y_i is a component of the random response; n is the number of observed (experimental) data; $f(y_i|\boldsymbol{\Theta})$ is the PDF of y_i for a given value of $\boldsymbol{\Theta}$; and L is a likelihood function (1). The likelihood function is used as the calibration metric to measure the degree of the agreement between the predicted and observed outputs. Of course, other calibration metrics can be employed for the model calibration. Figure 2 shows the procedure of the model calibration execution. After building the PDF of a predicted response (ψ) using UP analysis, the likelihood function is calculated by integrating probability densities over experimental data. The initial calibration parameter vector will be updated until the likelihood function is converged to a maximum value. Thus, the model calibration can be formulated as the unconstrained optimization problem in (2). This paper uses a gradient-based optimizer in MATLAB software to solve the optimization problem.

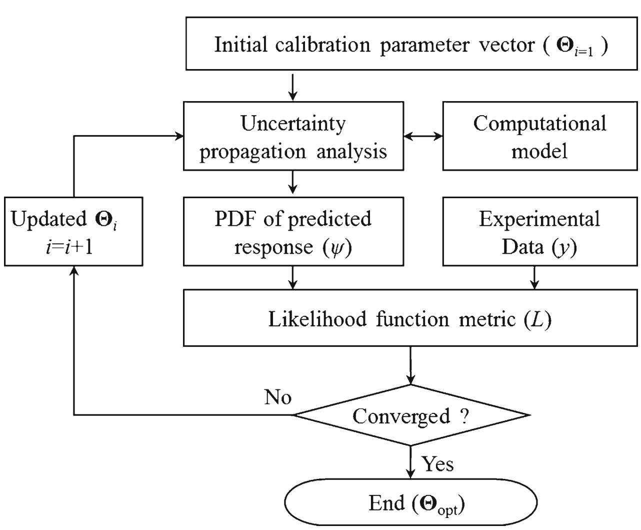


Fig. 2 The procedure of model calibration execution

Figure 3 shows the concept of the likelihood function. In the figure, x-axis indicates a controllable variable and y-axis denotes experimental data. The likelihood function between the experimental data and the response PDFs in Fig. 3a is larger than those of Fig. 3b and c. This is because the mean values of the response PDFs in Fig. 3b are deviated from those of experimental data, and the standard deviations of the PDFs in Fig. 3c are larger than those of experimental data.

The observation error (e) such as random error and biased error can be ignored in model calibration execution with the following assumptions: (i) inherent uncertainties in material properties, loading conditions and boundary conditions are dominant, (ii) experimental tests are well designed so that experimental data highly represent reality. In many cases, it is not feasible to find a true computational model through the model calibration procedure because these errors are unknown that are highly dependent on current knowledge of experimentalists or quality of experiment devices. For instance, it is extremely prohibitive to obtain the experimental random error in case of destructive testing such as tread block test.

The prediction error (δ) can capture the deterministic bias between observed and predicted results that arises from inappropriate model form, coding errors, numerical errors due to discretization approximations, wrong boundary conditions, etc. However, there is confounding between unknown random model variables and prediction error during model calibration execution; that is, they are not statistically identifiable during optimization. If the prior knowledge about the form of prediction error (e.g., linear form or nonlinear form) is given, it is beneficial to consider the prediction error in the calibration process; however, in many cases it is not easy to figure out the error form before the model calibration activity. Should the inappropriate model error form be used in the model calibration process, the calibrated computational model can be severely misrepresented. For example, Xiong et al. (Xiong et al. 2009) considered this error for model calibration; however, extrapolating capability of the calibrated model was not enhanced. To eliminate risk associated with the unknown model form, the prediction error is ignored in the model calibration (especially at first iteration in Fig. 1).

2.2 Hypothesis test for validity check

The validity check of a statistically calibrated model requires many experimental data from multiple samples (or physical products); however, it is impractical to manufacture many prototypes due to expensive manufacturing cost (Youn et al. 2011). The dearth of experimental data poses two challenges for the validity check. First, the experiments for the validity

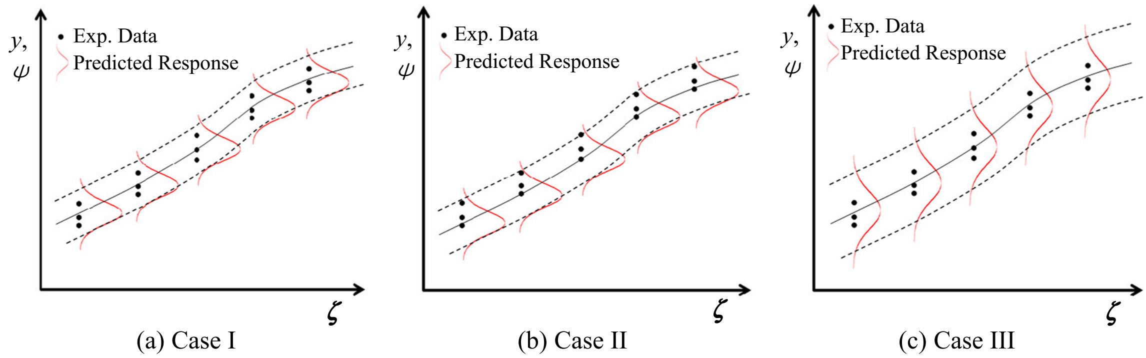


Fig. 3 The concept of the likelihood function in model calibration execution: **a** High likelihood function, **b** Low likelihood function due to deviated mean values, **c** Low likelihood function due to large standard deviations

check are normally conducted under various operating conditions (or experimental settings) in a validation domain. Given limited experimental data for the validity check, it is beneficial to integrate the evidence from all observation data over the entire validation domain into a single measure of overall mismatch (Ferson et al. 2008). Second, the small sample size of experiments will produce another layer of uncertainty in a validity check metric, of which the effect on model validity must be carefully analyzed. The hypothesis test for validity check proposed in this paper is thus devised to solve these challenges. In the hypothesis test, the null hypothesis (H_0) is defined as the claim that the calibrated model is valid. The null hypothesis can be rejected only if a validity check metric suggests that H_0 is false; otherwise not rejected. Upon the rejection, the calibrated model should be further refined as shown in Fig. 1, which is not the scope of this study.

(a) U-pooling Method

To solve the first challenge, the hypothesis test employs the u-pooling method for the validity check (Ferson et al. 2008). It allows the integration of the evidence from all experimental data under various experimental settings (e.g., environmental temperature, loading, etc.) into a single mismatch metric. In the u-pooling method, the cumulative density, u_i, can be obtained through the transformation of every experimental datum (y_i). The predictive CDF (F_y) of a computational response can define the transformation as

$$u_i = F_{yi}(y_i) \quad (3)$$

where i is the number of experimental data. Under the assumption that the experimental data (y_i) and the corresponding prediction come from a true distribution of y, (i.e., the model is valid), the u_i values obtained using both all the experimental data and the predicted results of a computational model follow a uniform distribution on [0,1]. In other words, if the u_i values follow the uniform distribution, it indicates that the predicted results perfectly agree with the experimental data. Therefore, we can quantify the degree of mismatch between the dispersion of experimental data and the distribution of predicted results by calculating an area (i.e., the area metric (U_m)) between the CDF of the uniform distribution (F_{uni}) and the empirical CDF (F_u) of u_i values corresponding to the experimental data as

$$\begin{aligned} U_m &= area(F_u,\ F_{\text{uni}}) \\ &= \int_0^1 |F_u(u) - F_{\text{uni}}(u)| du \quad 0 \le u \le 1,\ \ 0 \le U_m \le 0.5 \end{aligned} \quad (4)$$

Fig. 4 Calculation of an area metric

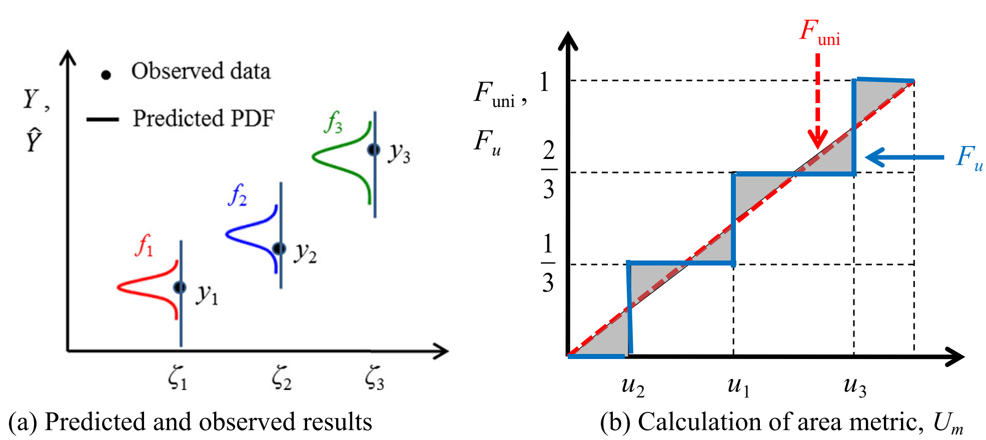
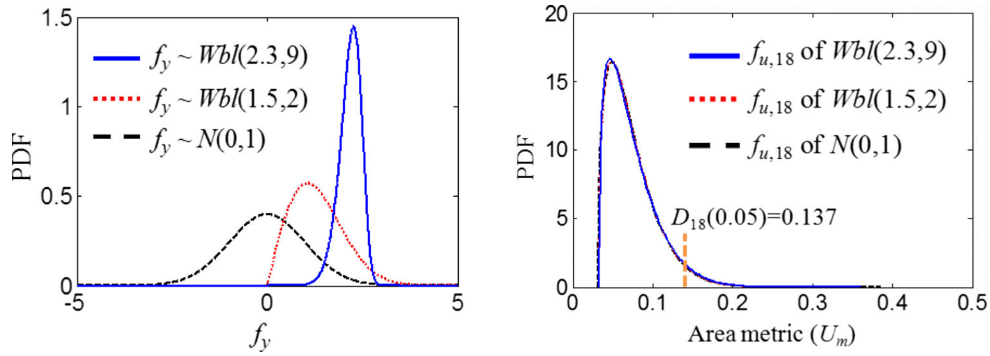


Fig. 5 PDFs of area metric ($f_{u,i}$, i=18) obtained with three different distributions



(a) Three different true distributions

(b) PDFs of the area metric ($f_{u,18}$)

As an example shown in Fig. 4a, there are three experimental data (y_i) and predicted PDFs (f_y) under different operating conditions (or experimental settings). The u_i of each experimental datum is calculated and its empirical CDF is drawn as shown in Fig. 4b. The calculated area of the shaded region in Fig. 4b indicates the area metric, U_m. The smaller the calculated U_m, the closer the predicted PDF to the distribution of experimental data. For example, if the model represents the observed responses well (i.e., the model is valid), U_m gets close to zero when sufficient experimental data exists. Should the model be invalid, U_m is greater than a marginal value of the area metric, which will be defined in Section II.B.(c).

(b) Area metric under epistemic uncertainty

If experimental data are comprehensively collected for the validity check, sampling uncertainty in the area metric (U_m) can be eliminated. In such case the null hypothesis can be rejected unless U_m is zero. In practice, due to limited experimental data, uncertainty exists in the area metric although the predicted and experimental results follow the true distribution of y (i.e., the model is valid). The uncertainty in the area metric can be characterized for a given observation data size (i) as:

Step 1. Determine a virtual sampling size (k) for uncertainty characterization of the area metric.

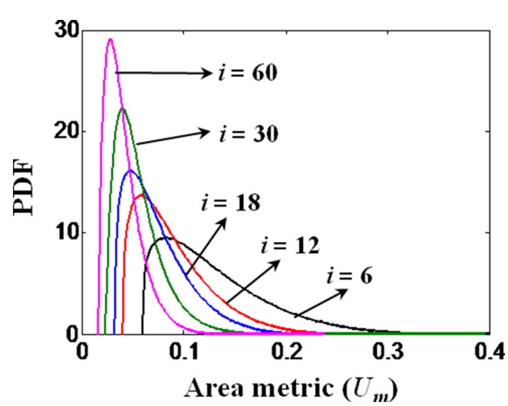


Fig. 6 PDFs of the area metric ($f_{u,i}$): the effect of a data size

Step 2. Generate samples (y_1, y_2, · ·, y_i) randomly for a virtual observation dataset from a true distribution (F_y) of y under the assumption that the predicted and observed data follow F_y (or a computational model is valid).

Step 3. Compute the CDF values (u_1, u_2, · · , u_i) corresponding to y_i using (3).

Step 4. Calculate an area metric value (U_m) using (4).

Step 5. Repeat Steps 2 to 4k times to generate random data (U_{m1}, U_{m2}, · ·, U_{mk}) of the area metric (U_m).

Step 6. Build an empirical probability distribution of the area metric ($f_{u,i}$). The Pearson system (Youn et al. 2008; Xi et al. 2012) is used in this paper.

For example, let us consider three different true distributions (f_y) of y as shown in Fig. 5a. The probability distributions of the area metric ($f_{u,i}$) are characterized with given observation data size (i=18) and virtual sampling size (k=5,000). The $f_{u,18}$ is identically determined irrespective of the shape of a true distribution as shown in Fig. 5b, where N and Wbl indicate normal and Weibull distributions, respectively. It is because a set of u_i values always follow uniform distribution regardless of a true distribution shape. This characteristic allows the proposed hypothesis test to be generally applicable. In addition, $f_{u,i}$ asymptotically converges to zero as the size (i) of experimental data increases, as

Table 1 Statistical information of the PDFs of the area metric

i	Mean	Std. Dev.	Skewness	Kurtosis	Pearson system type
6	0.1296	0.0535	1.1326	4.2585	Type I
12	0.0914	0.0389	1.2337	4.7938	Type I
18	0.0741	0.0319	1.2177	4.6966	Type I
30	0.0569	0.0247	1.4871	6.3728	Type I
60	0.0407	0.0178	1.3245	5.3782	Type I

shown in Fig. 6. In other words, the uncertainty in the area metric decreases as the amount of experimental data increases. Table 1 shows the statistical moments and the Pearson system type of the PDFs in Fig. 6.

(c) Hypothesis test for validity check

The hypothesis test uses a calculated area metric (U_m) using predicted and observed results, and the PDF of the area metric ($f_{u,i}$) for a given observation data size. Because $f_{u,i}$ indicates the plausible frequency of the area metric under the assumption of a valid model, an upper-tailed test can be employed with a predetermined rejection region as

$$U_m > D_i(\alpha) \quad (5)$$

where $D_i(\alpha)$ indicates a critical value of the area metric; α is a significance level, normally set to 0.05. For example, $D_{18}(0.05)$ is 0.137 for the upper-tailed test (see Fig. 5). The null hypothesis will be rejected if and only if the calculated area metric (U_m) falls in the rejection region. In the absence of such evidence, H_0 should not be rejected since it is still quite plausible. The significance level which is referred to as type I error, indicates the probability we reject a computational model although it is valid. It can be calculated as

$$\text{Type I error} = \int_{D_i(\alpha)}^{\infty} f_{u,i}(x)dx \quad (6)$$

In this study, type II error, the probability that we do not reject a computational model when it is invalid, is not considered since quantification of type II error is meaningless since there is no evidence on the degree of invalidity between predicted and true distributions. The increase of type I error (or significance level) results in decrease of the type II error; therefore, it is recommended that higher type I error is employed for validity check of engineered products having high risk on predicted results.

$p(d_i>d_j)$	d_1	d_2	d_3
d_1	0.500	0.85	0.99
d_2		0.500	0.79
d_3			0.500

Fig. 8 Design decision matrix for relative qualification

2.3 Virtual qualification

The virtual qualification is a process to qualify a product design through the use of the valid computational model. This task will be thus executed only if the validity of a computational model is assured through the hypothesis test above. The virtual qualification can be performed in an absolute or relative manner. The absolute qualification can be conducted if a performance of interest (POI) has a strict margin, as shown in Fig. 7a. For example, the design 1 is qualified if its POI exceeds the margin, whereas the design 2 is not. The relative qualification is preferred for a product design if the POI has no strict margin. Then, various product design alternatives can be compared with their PDFs of the POI, as shown in Fig. 7b. For instance, the design 1 is preferred to the designs 2 and 3 if the POI is a larger-the-better type. The virtual qualification can be performed quantitatively by constructing a design decision matrix, as shown in Fig. 8. This matrix can aid in rational decision-making for product design selection. Values in the upper triangular part of the matrix indicate the probability that one design (i: row) is better than the other (j: column, $p(d_i > d_j)$) where p indicates a probability and d_i and d_j indicate i^{th} and j^{th} designs. The design decision matrix provides information for comparison of design alternatives and helps analysts make a rational decision in the product development process.

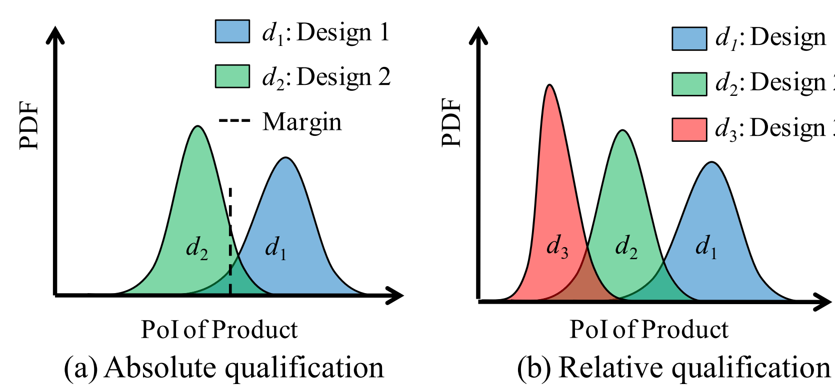
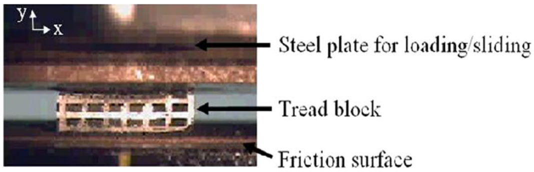
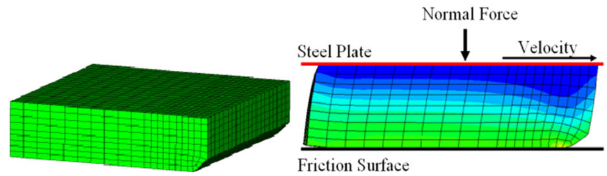


Fig. 7 Absolute and relative virtual qualification methods (d_i indicates ith design)



(a) Tread block tester



(b) Schematic of tread block FE model (Design A, size: 30×30×8mm)

Fig. 9 Tread block tester and tread FE model

3 Case study: tread block problem

3.1 Problem description

The braking and cornering performances of tire are highly related with a friction force between tire and road surface. The consideration of a single tread block, detached from the tire body, enables the investigation of a tire friction behavior. Therefore, the tread block tester (or block FE analysis) in Fig. 9 is widely used to measure (or predict) the friction forces of different tread block designs. As shown in Fig. 9, the top surface of a tread block specimen is fixed to a steel plate and the bottom surface contacts a road surface. While applying a designed normal load to the tread block specimen and pulling it along x-direction in a constant velocity, a friction force is measured.

The commercial FE tool, Abaqus, was used for the transient tread block analysis. As a part of model verification activity (AIAA 1998), a grid (mesh) density was adjusted to reduce a discretization error in the FE model. Realistic description of a friction model is most crucial in a tread block analysis. Generally, the friction model is a function of contact pressure (φ) and sliding velocity (Cho and Jung 2007; Hofstetter et al. 2006). Since the sliding velocity (=1 cm/sec) remains constant during the block test, a contact pressure-dependent model can be employed as

$$\mu = \mu_0 \times \left(\frac{\varphi}{\varphi_{ref}}\right)^{-\upsilon} \tag{6}$$

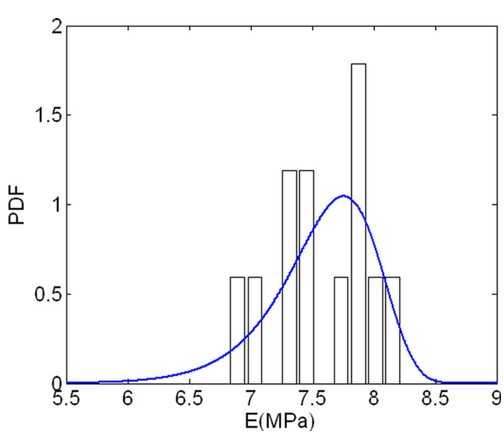


Fig. 10 Variance of elastic modulus: Weibull(7.77,22.07)

where φ_{ref} is a reference contact pressure (=1.0E5 Pa); υ is pressure exponential parameter; μ_0 is a friction coefficient when the pressure is equal to φ_{ref}. The true values of two model parameters, υ and μ_0, vary with the selection of materials and the pair of surfaces in contact.

3.2 Statistical model calibration

The POI is a friction force between a tread block and a road surface. The block model is composed of three random input variables: elastic modulus and two friction model parameters (μ_0, υ). While the elastic modulus is classified as a known model variable, the friction model parameters (μ_0, υ) as a set of unknown model variables. To characterize the elastic modulus, tensile tests (ambient temperature: 23 °C, cross-head speed: 500 mm/min) with eight rubber samples (size: 2.62× 5×20 mm) were performed and Weibull distribution (*Wbl* ~ (7.77, 22.07)) was found to be the most suitable distribution (see Fig. 10).

For the statistical model calibration, twenty-four solid block specimens (see Design A in Fig. 9b) were fabricated. Experiments were executed at four different normal loading conditions (3, 5, 7 and 9 kg/cm^2) that were decided based on tire operating conditions. The response surface models of four different operating conditions (3, 5, 7 and 9 kg/cm^2) were first constructed to reduce a simulation cost. Computational analyses at four-level full factorial design points with three factors (μ_0, υ and elastic modulus) were conducted to generate the response surface models. Figure 11 shows a response surface

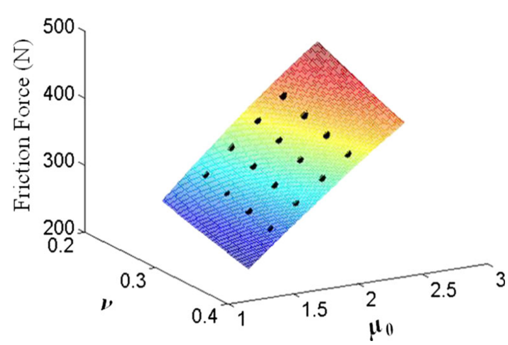


Fig. 11 A surrogate model: 3 kg/cm^2 normal loading

Table 2 Calibrated friction model parameters

Random parameters	Initial vector		Calibrated vector	
	Mean	Std. Dev.	Mean	Std. Dev.
μ_0	1.95	0.195	1.875	0.0246
υ	0.33	0.033	0.258	0.011

model about the 3 kg/cm^2 loading condition. It was observed that the response models look to be linear in the interested region. Next, the statistical calibration was performed by comparing experiment data with the predicted results. The hyper-parameter vector ($\mathbf{\Theta}$) includes the mean and standard deviation of two unknown model variables (μ_0 and υ) that are assumed to follow a normal distribution. Table 2 shows the results of the initial and calibrated hyper-parameter vectors. The predicted friction forces after statistical calibration show good agreement with the experimental data as shown in Fig. 12.

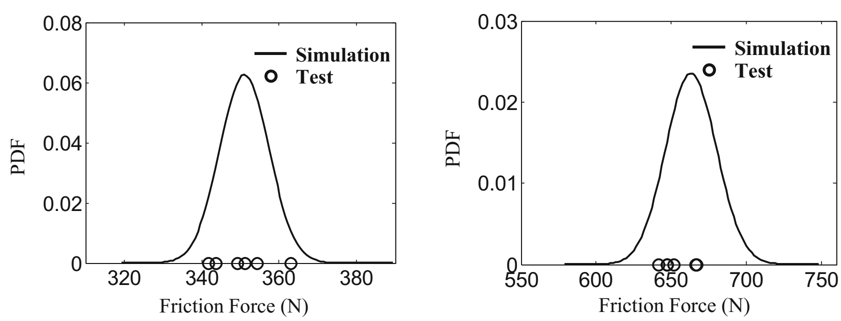
3.3 Hypothesis test for validity check

Twelve tread blocks of two alternative tread designs as depicted in Fig. 13 (six blocks of Designs B and C each) were fabricated and tested under 7 kg/cm^2 normal load for the validity check experiments. Figure 14a shows the measured experiment data and the predicted PDFs of the different designs. The area metric (U_m=0.0525) was calculated with the aggregated eighteen test data (six of Designs A, B, and C each) and the predicted PDFs of three designs as shown in Fig. 14b. It is found that U_m=0.0525 is far less than $D_{18}(0.05)$=0.137 as shown in Fig. 5b. It can be concluded that the calibrated model is valid.

3.4 Virtual qualification

A relative design qualification process was executed with three different designs in Figs. 9b and 13. As shown in Fig. 15, the design decision matrix for the product designs was built with three PDFs in Fig. 14a. It is noted that the designs in the design decision matrix are ordered based on the magnitude of the friction force to make the values in upper triangular matrix larger than 50 %. Based on the design decision matrix, we can quantitatively see that the chance to prefer Design C to Design B is 95.61 %. There are two possible reasons that Design C with a groove along vertical direction (see Fig. 13) turns out to be better than Design B. First, Design C of vertical groove has larger contact area than Design B of longitudinal groove and it increase total friction force. Second, higher contact pressure of Design B at the edge of longitudinal groove locally decrease friction coefficient and friction force and it may decrease total friction force. The virtual qualification using statistical evidence in the design decision matrix can help to make a rational decision making for two reasons as follows:

1. Once a computer simulation model becomes statistically valid, we can use the model proactively in a product design process with a high confidence on the simulated results. Through repeated exercises of the product virtual testing, one can produce more related data and make the virtual testing routinely incorporated into a product design process. This eventually makes an engineering product developed in a reliable and efficient manner.
2. The design decision matrix enables a quantitative and statistical decision making in determining a best design alternative even though a product involves manufacturing variability and operational uncertainty. In other words, this matrix can eliminate a chance to make an erroneous decision due to deterministic quantification of product designs. For example, according to the matrix in Fig. 15, design A must be selected although one physical test may reveal that design B could outperform design A. Eventually the statistical evidence in the design decision matrix helps minimize a confliction between experimental and predicted results in a product development process.



(a) Normal loading: 3kg/cm^2 (b) Normal loading: 7kg/cm^2

Fig. 12 Predicted and observed results after model calibration

(a) Design B (Longitudinal groove) (b) Design C (Vertical groove)

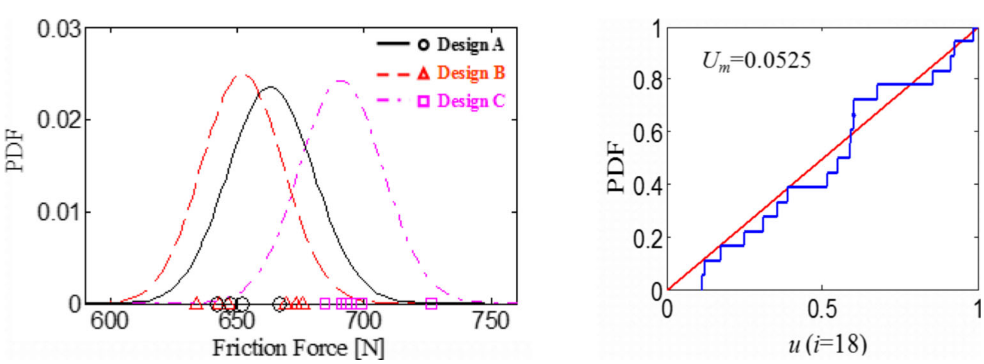
Fig. 13 Two design alternatives (size: 30×30×8 mm, contour indicates simulated contact pressure.)

4 Conclusion

This paper proposed the generic framework for the virtual testing, which is composed of three sequentially executed procedures: (1) statistical model calibration, (2) hypothesis test for validity check, and (3) virtual qualification. The statistical model calibration was used to improve the predictive capability of a computational model in a statistical sense. This model calibration determined the calibrated statistics of the unknown model variables while maximizing the likelihood between predicted and observed data. Secondly, the hypothesis test with limited observed data was proposed for the validity check to see if a calibrated model is sufficiently predictive for virtual testing of a new design. The hypothesis test employed the area metric while taking into account sampling uncertainty in the area metric. The area metric quantifies the degree of mismatch between predicted and observed results. The uncertainty in the area metric due to the lack of experimental data can be characterized with the PDF of the area metric ($f_{u,i}$) for a given observation data size. The calibrated model can be accepted only if the area metric is less than a predefined critical value ($D_i(\alpha)$). Lastly, the design decision matrix of a valid model was proposed to make a decision on virtual qualification of a product. This qualification can provide statistical evidence for rational decision-making on new product designs. The merits of the proposed framework were demonstrated with a tread block problem. The uncertainty in elastic modulus was first characterized with the material test data whereas the statistics of the unknown model variables (μ_0, v) were determined through the statistical calibration. The hypothesis test confirms the validity of the calibrated model. Finally, the virtual qualification helps make a qualified product design decision. It is concluded that the proposed framework can offer a standard guideline for engineers to develop a highly predictable computational model and is more practical and appropriate approach in industry because of limited resources (e.g., time, budget, man-power).

This study can be further improved in future by addressing the following issues including

1. In case a calibrated model turns out to be invalid, model refinement should be considered. ASME (ASME 2006) defined model refinement as a process to change the mathematical expressions for building a more realistic model that better represents the physics of the system. Technically speaking, model refinement process includes activities such as identifying the root causes of the model invalidity and minimizing the degree of the invalidity by improving mathematical and computational models. This underscores the need of a systematic methodology for model refinement to develop formal steps and related ideas.
2. As mentioned in Section II.A, the prediction and observation (experiment) errors were not considered in this study. However, it might be interesting to address a model validation procedure including a process to determining the best form of these errors when these errors are not negligible, and how the errors affect the model validation.
3. The distribution type candidates of unknown model variables were decided based on the best of the authors' knowledge. It is of course interesting to investigate how



(a) Predicted and observed results (b) Calculation of U_m

Fig. 14 Predicted and observed results (Normal load: 7 kg/cm^2)

(%)	Design C	Design A	Design B
Design C >	50.00	88.01	95.61
Design A >		50.00	68.48
Design B >			50.00

Fig. 15 Example of design decision matrix (Normal loading: 7 kg/cm^2)

the distribution types of model variables affect calibration results. Moreover, it is important to build the database for the distribution types of the model variables based on pre-knowledge and known information as one understand an engineered system better. The database can help computer model developers determine appropriate distribution types.

Acknowledgments This work was partially supported by a grant from the Energy Technology Development Program (2010101010027B) and International Collaborative R&D Program (0420-2011-0161) of Korea Institute of Energy Technology Evaluation and Planning (KETEP), funded by the Korean government's Ministry of Knowledge Economy, the National Research Foundation of Korea (NRF) grant (No. 2011–0022051) funded by the Korea government, the Basic Research Project of Korea Institute of Machinery and Materials (Project Code : SC1000) supported by a grant from Korea Research Council for Industrial Science & Technology, and the Institute of Advanced Machinery and Design at Seoul National University (SNU-IAMD).

References

AIAA, P.T.C (1998) Guide for the verification and validation of computational fluid dynamic simulations, AIAA Guide

ASME, P.T.C (2006) Guide for verification and validation in computational solid mechanics, ASME, New York

Babuska I, Oden JT (2004) Verification and validation in computational engineering and science: basic concepts. Comput Methods Appl Mech 193(36–38):4057–4066

Campbell K (2006) Statistical calibration of computer simulations. Reliab Eng Syst Saf 91(10–11):1358–1363

Chen W, Tsui K, Wang S (2008) A design-driven validation approach using Bayesian prediction models. J Mech Des 140(2):021101

Chen W, Baghdasaryan L, Buranathiti T, Cao J (2004) Model validation via uncertainty propagation. AIAA J 42:1406–1415

Cho JC, Jung BC (2007) Prediction of tread pattern wear by an explicit finite element model. Tire Sci Technol 35(4):276–299

Choi JH, Lee WH, Park JJ, Youn BD (2008) A study on robust design optimization of layered plate bonding process considering uncertainties. Struct Multidiscip Optim 35(6):531–540

Datta, B. N (2005) Finite element model updating and partial eigenvalue assignment in structural dynamics: recent developments on computational methods, Math Model Anal, pp. 15–27

Ferson S, Oberkampf WL, Ginzburg L (2008) Model validation and predictive capability for the thermal challenge problem. Comput Methods Appl Mech Eng 197(29–32):2408–2430

Higdon D, Nakhleh C, Gattiker J, Williams B (2008) A Bayesian calibration approach to the thermal problem. Comput Methods Appl Mech Eng 197(29–32):2431–2441

Hills RG, Truncano TG (2002) Statistical validation of engineering and scientific models: a maximum likelihood based metric, SAND2001–1783. Sandia National Labs, Albuquerque

Hofstetter K, Grohs C, Eberhardsteiner J, Mang HA (2006) Sliding behaviour of simplified tire tread patterns investigated by means of FEM. Comput Struct 84(17–18):1151–1163

Jung BC, Lee D, Youn BD (2009) Optimal design of constrained-layer damping structures considering material and operational condition variability. AIAA J 47(12):2985–2995

Kennedy MC, O'Hagan A (2002) Bayesian calibration of computer models. J R Stat Soc Ser B 63(3):425–464

Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. J R Stat Soc 63:425–464

Liu F, Bayarri MJ, Berger JO, Paulo R, Sacks J (2008) A Bayesian analysis of the thermal challenge problem. Comput Methods Appl Mech Eng 197(29–32):2457–2466

Modarres, M., Kaminskiy, M., and Krivtsov, V (1999) Reliability engineering and risk analysis; a practical guide, CRC Press

Oberkampf WL, Trucano TG, Hirsch C (2004) Verification, validation, and predictive capability in computational engineering and physics. Appl Mech Rev 57:345

Oberkampf WL, Barone MF (2006) Measures of agreement between computation and experiment: validation metrics. J Comput Phys 217:5–36

Roache PJ (1998) Verification of codes and calculations. AIAA J 36(5): 696–702

Thacker BH, Doebling SW, Hemez FM, Anderson MC, Pepin JE, Rodriguez EA (2004) Concepts of model verification and validation, Technical report, LA-14167. Los Alamos National Lab, Los Alamos

Trucano TG, Swiler LP, Igusa T, Oberkampf WL, Pilch M (2006) Calibration, validation, and sensitivity analysis: what's what. Reliab Eng Syst Saf 91(10–11):1331–1357

Xi Z, Hu C, Youn BD (2012) A comparative study of probability estimation methods for reliability analysis. Struct Multidiscip Optim 45(1):33–52

Xiong Y, Chen W, Tsui K, Apley DW (2009) A better understanding of model updating strategies in validating engineering models. Comput Methods Appl Mech Eng 198(15–16):1327–1337

Youn BD, Xi Z, Wang P (2008) Eigenvector dimension reduction (EDR) method for sensitivity-free probability analysis. Struct Multidiscip Optim 37(1):13–28

Youn BD, Jung BC, Xi ZM, Kim SB, Lee WR (2011) A hierarchical framework for statistical model calibration in engineering product development. Comput Methods Appl Mech Eng 200:1421–1431. doi:10.1016/j.cma.2010.12.012

Wang P, Youn, B. D., Xi Z., Kloess A. (2009) Bayesian reliability analysis with evolving, insufficient, and subjective data sets. J Mech Des 131(11):111008

Zhang R, Mahadevan S (2003) Bayesian methodology for reliability model acceptance. Reliab Eng Syst Saf 80:95–103