Validation assessment with uncertain model inputs

Liang Zhao*, Zhanping Yang

Institute of Electronic Engineering, China Academy of Engineering Physics, 621900 MianYang, China

Received 14 July 2014, www.cmnt.lv

Abstract

This paper presents a validation assessment method to measure the discrepancies between the model predictions and experimental observations under both aleatory and epistemic uncertainty. The model inputs considered in the paper are sparse point data or interval data, which leads to uncertain parameters for the distribution of the model inputs. A likelihood based method is used to represent the stochastic model inputs and it yields a single probability distribution which integrates the aleatory and epistemic uncertainty of model inputs. This representation of model inputs provides an advantage in computation efficiency for the conventional double loop sampling requirement in uncertainty propagation is collapsed into a single loop sampling. An area based validation metric is extended to compare the probabilistic model predictions obtained from uncertainty propagation with the empirical distribution function of the experimental observations, it reflects an objective quantification of the entire discrepancies between predictions and observations. The confidence interval for the validation metric, which just depends on the amount of experimental observations and confidence level is also developed. A numerical example is used to illustrate the proposed method.

Keywords: validation assessment; likelihood based method; validation metric; confidence interval

1 Introduction

Model based computer is applied in various engineering disciplines, examples range from nuclear reactor certification to understanding of the cosmos. The uncertainty and confidence in model prediction are attracting attention increasingly. Model validation has been advocated as a necessary procedure when the model is used for risk assessment with high consequence systems [1-3]. The fundamental concept of model validation has been intensively argued by professional committees [4-8], and the consensus among researchers is that a rigorous model validation should explicitly account for various uncertainties. The uncertainties can be broadly classified into aleatory and epistemic. For instance, a precise probability distribution indicates the aleatory uncertainty, and the probability distribution with uncertain parameters implies the existence of epistemic uncertainty. This paper concerns the question of how to measure the discrepancies between the model predictions and experimental observations under both aleatory and epistemic uncertainty. The question is usually called "validation assessment" [3], it is a basic process in model validation procedure, the performance of the model can be judged from the measurement.

Several approaches have recently been suggested for validation assessment, including significance testing [9-14], Bayesian method [9, 13, 14], mean based comparison [3, 15, 16] and area based method [17]. All of these approaches have drawbacks. For example, the significance testing to validation is primarily focused on identifying the evidence against a certain hypothesis, and the Bayesian method is mostly focused on the belief that the model is correct. They are rather different from the goal of validation assessment, which is interested in the objective quantification of model accuracy. Instead of making a "accept" or "reject" statement with hypothesis, the mean-based comparison method measures distance between the mean of model predictions and the estimated mean of experimental observations. The limitation of this method is that it only takes the central tendency of predictions and observations into account, while the distribution of predictions contain amount of detail which may be represented insufficiently with a comparison of means. With the aim of measuring the discrepancies of the entire distributions between predictions and observations, the area based method uses the area between the prediction distribution and the observation distribution as a validation metric, but it does not provide a confidence level of the metric due to the lack of sufficient observations. Besides, of the aforementioned drawbacks, there is a serious limitation about theses method that they are only suitable for validation assessment under aleatory uncertainty, and cannot be used directly when model inputs are quantities with epistemic uncertainty.

Typical epistemic uncertainty regarding a model input can be a stochastic quantity with uncertain distribution parameters [18], this is usually due to sparse point data or interval data. Several methods such as evidence theory [19], second order probability method [20], fuzzy sets [21], etc. have been proposed for quantification of the epistemic uncertainty, but the results of these methods which are usually in forms of distribution bounds cannot

* ***Corresponding author*** e-mail swjtu_zhaoliang@126.com

be used in validation assessment directly. This paper extends the area-based validation assessment method to account for epistemic uncertainty arising from sparse or interval data with model inputs. A likelihood based approach [22] is used to construct a single probability distribution for model inputs with uncertain distribution parameters. With the likelihood based representation of uncertainty, the model predictions can be characterized as a single probability distribution after uncertainty propagation, it facilitates the comparison between model predictions and probabilistic experimental observations that are the essential of area based validation metric. The confidence interval of the validation metric is provided by calculating the infimum and supremum of the discrepancy between the cumulative probability distribution function (CDF) of model predictions and the possible experimental empirical distribution functions (EDF) which are bounded by the Kolmogorov - Smirnov limit theorem.

The paper is organized as follows. Section 2 describes the likelihood based method to represent the epistemic uncertainty with an unique probability distribution for model inputs. Section 3 derives the validation metric based on area method, accounting for both aleatory and epistemic uncertainty. Furthermore, this section presents a confidence interval for the validation metric, the interval is associated with the amount of experimental observations. Section 4 demonstrates the proposed method with a numerical example. Finally, Section 5 offers some concluding remarks.

2 Likelihood based representation of epistemic uncertainty

Consider model input X that has a probability density function (PDF) $f_X(x|\boldsymbol{P})$ where $\boldsymbol{P}$ denotes the distribution parameters. With the distribution type is known, the PDF is conditioned on the choice of $\boldsymbol{P}$. The likelihood function $L(\boldsymbol{P})$ is defined as the probability of observing data x given $\boldsymbol{P}$ [22]. When the information regarding X is available with independent point data x_i(i=1,2,⋯,n), the likelihood for $\boldsymbol{P}$ can be calculated as Eq.(1) with a finite precision ε:

$$\begin{aligned} L(\boldsymbol{P}) &\propto \prod_{i=1}^{n} P_i\left(X \in \left(x_i-\varepsilon, x_i+\varepsilon\right) | \boldsymbol{P}\right) \\ &= \prod_{i=1}^{n} \int_{x_i-\varepsilon}^{x_i+\varepsilon} f_X(x|\boldsymbol{P})\mathrm{d}x \\ &\propto \prod_{i=1}^{n} f_X(x_i|\boldsymbol{P}) \end{aligned} . \quad (1)$$

Here the independence assumption in Equation (1) is modest in statistical analysis.

Since the likelihood in Equation (1) is actually integrated in an infinitely small interval instead of calculating at data points, it is straightforward to apply the definition to X in the form of interval. When the information regarding X is available with intervals [a_i, b_i] (i=1,2,⋯,m), the likelihood for $\boldsymbol{P}$ of PDF $f_X(x|\boldsymbol{P})$ can be expressed as:

$$\begin{aligned} L(\boldsymbol{P}) &\propto \prod_{i=1}^{m} P_i\left(X \in \left(a_i, b_i\right) | \boldsymbol{P}\right) \\ &= \prod_{i=1}^{m} \int_{a_i}^{b_i} f_X(x|\boldsymbol{P})\mathrm{d}x \end{aligned} . \quad (2)$$

Accounting for both point values and intervals in the available data, the likelihood function of $\boldsymbol{P}$ can be represented as:

$$L(\boldsymbol{P}) \propto \left[\prod_{i=1}^{n} f_X(x_i|\boldsymbol{P})\right]\left[\prod_{i=1}^{m} \int_{a_i}^{b_i} f_X(x|\boldsymbol{P})\mathrm{d}x\right]. \quad (3)$$

It is popularly known that the $\boldsymbol{P}$ can be estimated by maximizing the likelihood function. However, the validation assessment is interested in the usage of entire likelihood function to construct the PDF of $\boldsymbol{P}$ rather than maximizing the likelihood. Thereby, consider the joint probability density of distribution parameters $\boldsymbol{P}$, denoted by $f_P(\boldsymbol{P})$, it can be calculated as Equation (4) using Bayes theorem with an uniform prior PDF $f_P'(\boldsymbol{P})=h$ (over the whole range of $\boldsymbol{P}$).

$$f_{\boldsymbol{P}}(\boldsymbol{P}) = \frac{L(\boldsymbol{P})h}{\int L(\boldsymbol{P})h\mathrm{d}\boldsymbol{P}} = \frac{L(\boldsymbol{P})}{\int L(\boldsymbol{P})\mathrm{d}\boldsymbol{P}} . \quad (4)$$

After calculating the $f_P(\boldsymbol{P})$, the PDF for X can be constructed based on principles of conditional probability and total probability, as represented in Equation (5).

$$f_X(x) = \int f_X\left(x|\boldsymbol{P}\right) f_{\boldsymbol{P}}\left(\boldsymbol{P}\right)\mathrm{d}\boldsymbol{P} . \quad (5)$$

Let Y denotes the model predictions, the PDF of Y represented by $f_Y(y)$ can be obtained using uncertainty propagation analysis. The single PDF of X facilitates the uncertainty propagation for it requires just one level of Monte Carlo sampling to construct the $f_Y(y)$. In contrast, the conventional uncertainty quantification methods such as second order probability method [20] require a two level Monte Carlo sampling: first, draw samples of $\boldsymbol{P}$ from $f_P(p)$, each of them determines a PDF of X: $f_X(x|\boldsymbol{P})$; second, several samples from each $f_X(x|\boldsymbol{P})$ are drawn to calculate a distribution of Y, the process generates a family of $f_Y(y)$ for different $\boldsymbol{P}$ ultimately. The two level Monte Carlo sampling strategy is so computationally costly that it might not always be affordable in practical model application. Furthermore, the family of model output distributions leads to a difficulty in measuring discrepancies between model predictions and experimental observations. The $f_X(x)$ in Eq.(5) can be interpreted as the expected value of $f_X(x|\boldsymbol{P})$, which depends on the choice of $\boldsymbol{P}$. The two level of uncertainties considered in second order probability method are integrated into the single PDF $f_X(x)$, and it can be evaluated numerically. Based on the single PDF for model input, the $f_Y(y)$ can be calculated using a straightforward Monte Carlo method. The resultant single PDF for model prediction brings an advantage in validation assessment. The following section implements the area based method [17] to measure the discrepancies between model predictions and experimental observations. With the likelihood based method explained in section 2, the area based method can account for both aleatory uncertainty and epistemic uncertainty in validation assessment.

3 Validation assessment using area based method

3.1 AREA METRIC

The aforementioned model prediction can be characterized as a CDF, represented as $F_Y(y)$. Consider the experimental observations are provided as a set of point data y_i, (i=1,2,···), the EDF for the data set is represented as $S(y)$, it preserves almost all statistical information in the data set The area based method proposed by Ferson [18] uses the area between the $F_Y(y)$ and the $S(y)$ as the measurement of the discrepancies. The mathematical expression for the area metric can be written as Equation (6):

$$d\left(F_Y, S\right) = \int_{-\infty}^{+\infty} \left|F_Y(y) - S(y)\right| \mathrm{d}y \,, \tag{6}$$

The $d(F_Y, S)$ in Eq.(6) is essentially a special case of the Wasserstein distance. The important merit of the area metric is that $d(F_Y,S)$ measures the discrepancies between the entire distributions from predictions and observations. Since the epistemic uncertainty and aleatory uncertainty in the model inputs have already been quantified with the single PDF $f_X(x)$, the area based method which makes use of the $f_X(x)$ can be applicable in the validation assessment with both epistemic uncertainty and aleatory uncertainty.

3.2 CONFIDENCE INTERVAL FOR THE AREA METRIC

Consider the sample uncertainty in experimental observations, the Kolmogorov-Smirnov statistics can be used to bound the EDF of observations as Equation (7):

$$\begin{aligned} &\overline{S}(y) = \min(1, \max(0, S(y) + D)) \\ &\underline{S}(y) = \min(1, \max(0, S(y) - D)) \end{aligned} \,, \tag{7}$$

Where $\overline{S}(y)$ and $\underline{S}(y)$ refer to the upper bound and lower bound of the experimental EDF respectively, and the D denotes the critical value for the Kolmogorov-Smirnov statistic [23], it just depends on the confidence level and the amount of experimental observations. After calculating the $\overline{S}(y)$ and $\underline{S}(y)$ with specified confidence level (1-α)×100%, the confidence interval for the area metric can be expressed as Equation (8):

$$\left[\inf_{S(y)\in K} \int_{-\infty}^{+\infty} \left|F_Y(y) - S(y)\right| \mathrm{d}y,\ \sup_{S(y)\in K} \int_{-\infty}^{+\infty} \left|F_Y(y) - S(y)\right| \mathrm{d}y\right], \tag{8}$$

where inf denotes the infimum, sup denotes the supremum, and the K represents the region encompassed by $\overline{S}(y)$ and $\underline{S}(y)$, it contains all possible EDFs of the experimental observations. The task of calculating the infimum or supremum taking the amount of experimental observations into account is sometimes challenging, intelligence optimization algorithms such as genetic algorithm can be used to handle this problem. The genetic algorithm randomly generates a population of possible experimental EDF constrained by Equation (7), then evaluates the fitness of each EDF using the area metric. For the infimum calculation task, a small area between possible EDF and model prediction CDF indicates high fitness for this EDF, whereas the supremum calculation task pursues large area as high fitness. The set of parents forming the next generation can be selected based on the fitness of individual EDF, where the high fitness members have more chances of being chosen. The mutation and crossover operation are applied to the parents to create next generation of possible experimental EDF. The selection, mutation and crossover are repeated until the maximum fitness in the population meets the criterion.

The area metric provides a quantitative measurement of the discrepancies between model predictions and experimental observations. The choice of the threshold for the metric is another question in model validation, which is usually called “adequacy decision” [3]. The adequacy decision considers that whether the validation metric is “great” enough to draw a conclusion that the discrepancies are significant. This question is derived from the concept that the threshold should be separated from the validation metric [9, 16, 17]. If the area metric is presented below the chosen threshold, one can conclude that the model is adequate for the intended use. Furthermore, consider the sampling uncertainty of the experimental observations, the upper bound of the confidence interval for area metric can be used to compare with the threshold, it claims that whether the model is adequate with (1-α)×100% confidence. Usually, the threshold quite depends on the intended use of the model for different requirement of model accuracy. Note that the distribution of model predictions calculated on the basis of $f_X(x)$ is not parametrically available, and the value of $d(F_Y,S)$ is expressed in physical units rather than statistical units. It implies that the engineers or project managers may be more rational than the mathematicians who developed the validation assessment method to decide the choice of the threshold. As stated in section 1 earlier, the paper concerns only the question of how to measure the discrepancies between the model predictions and experimental observations, the choice of the threshold for the metric is out of the research scope.

4 Numerical example studies

A numerical example is presented in this section to illustrate the proposed validation assessment method. Consider a material layer slab with thickness L, volumetric heat capacity ρC and thermal conductivity k, it is exposed to a heat flux, the temperature (T) at a specified time (t) after exposure to the heat flux (q) needs to be predicted. The model for the temperature prediction can be written as:

$$T(k,\rho C,T_0,L,q;l,t) = T_0 + \frac{qL}{k}\left[\frac{kt/\rho C}{L^2} + \frac{1}{3} - \frac{l}{L} + \frac{l^2}{2L^2} - \sum_{n=1}^{6}\frac{2}{\pi^2 n^2}\exp\left(-\frac{n^2\pi^2 kt}{L^2\rho C}\right)\cos\left(n\pi\frac{l}{L}\right)\right], \quad (10)$$

where T_0 is the initial ambient temperature, it is fixed at 25ºC, l is the location within the material. The controllable inputs (L, q) are assumed to be known exactly, i.e. L=0.0127m, q=1000W/m^2, while the (ρC, k) are uncertain inputs. For the sake of illustration, suppose that the probability distribution of the volumetric heat capacity ρC and thermal conductivity k are described as normal by experts but with uncertain distribution parameters. The observations of ρC are available as three point data {4.52E+05J/m^{3}ºC, 4.10E+05J/m^{3}ºC, 4.02E+05 J/m^{3}ºC}, and the data about k is available in the form of intervals as [0.0601W/mºC, 0.0604W/mºC], [0.0545 W/mºC, 0.0547W/mºC]. Let the purpose of the model be to predict the temperature at the surface of the layer (l=0m) after exposure to the heat flux for 500 seconds (t=500s). Experiment in the circumstance consistent with the intended use of the model is carried out four times repeatedly, the observations are four point data as {210.6ºC, 214.6ºC, 186.9ºC, 219.0ºC}. It is required to measure the discrepancies between the model predictions and experimental observations.

Consider that model inputs ρC and k are described with epistemic uncertainty, the likelihood based approach explained in section 2 is used to represent the two model inputs with single probability distribution respectively. Note that the normal distribution describing the model inputs ρC and k is parameterized by mean and standard deviation, so the likelihood in Equation (3) is integrated for the mean and the standard deviation together. The entire admissible ranges for the mean and the standard deviation are both $(0,+\infty)$, for the sake of calculation, here we draw finite bounds for the two ranges. For instance, assume that the mean of ρC can vary in (2.5E+05, 5.5E+05). Based on the observations we believe the finite ranges are still wide enough to cover any possible value of the distribution parameters. By combining Equation (4) and (5), the distributions of the ρC and k are calculated as it is shown in Figure 1 and Figure 2.

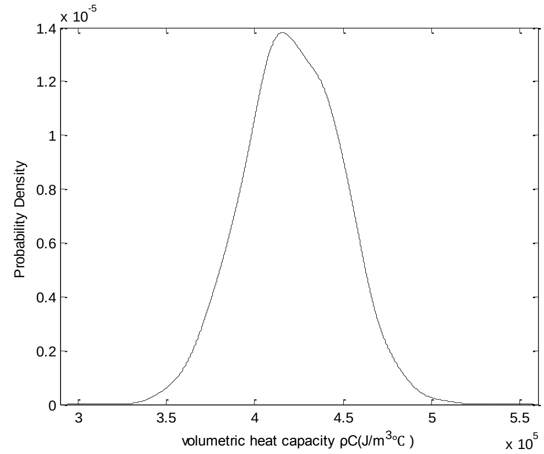

FIGURE 1 Distribution of volumetric heat capacity ρC

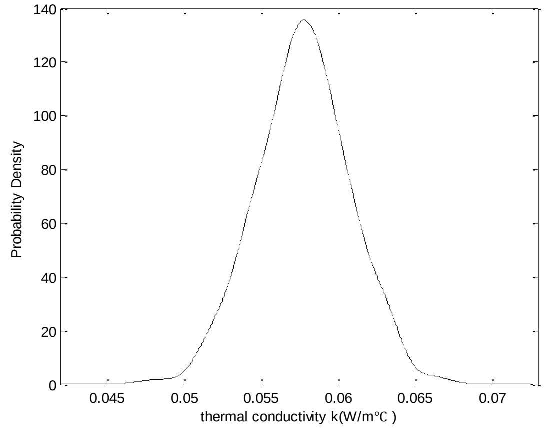

FIGURE 2 Distribution of thermal conductivity k

The distributions presented in Figure 1 and Figure 2 include both epistemic uncertainty and aleatory uncertainty in ρC and k. Note that the integration in Equation (5) is calculated numerically for $f_P(\boldsymbol{P})$ is not parametrically available, so the resultant PDFs for ρC and k are not analytical. An Monte Carlo method is used to propagate the uncertainty of ρC and k through the thermal model in Equation (9), the PDF of the material layer surface temperature at t=500s is shown in Figure 3.

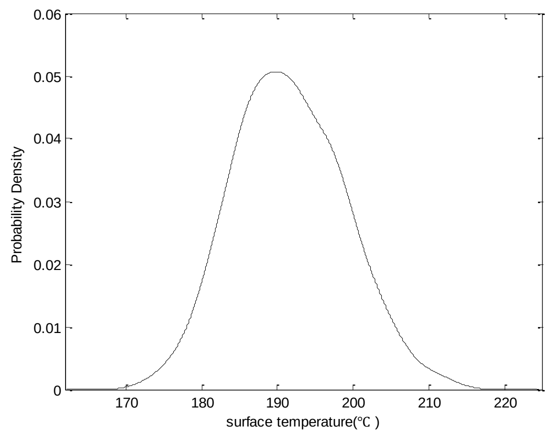

FIGURE 3 Distribution of model predictions

The area metric is used to measure the discrepancies between model predictions and experimental observations. Figure 4 illustrates this area with shaded region between smooth black curve and blue step function. The black curve is the CDF of the model predictions, and the blue step function is the EDF for experimental observations consisting of four point value.

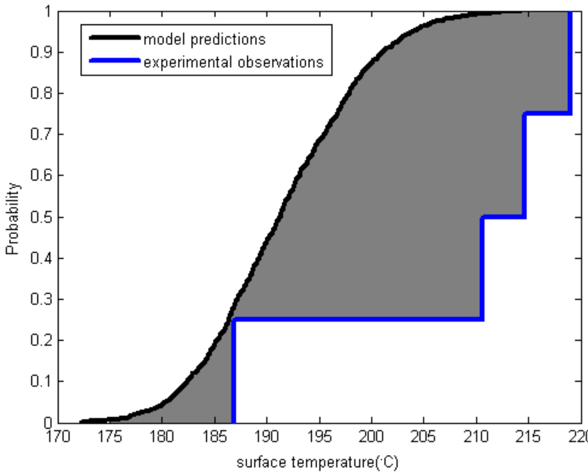

FIGURE 4 EDF of experimental observations against CDF of model predictions

Using Equation (6), the area metric is obtained to be 16.21°C. A genetic algorithm is employed to find the infimum and supremum defined in Equation (8) to get the confidence interval for the area metric. The exact value of the confidence interval with 95% confidence level is [4.13°C, 26.88°C]. The result of the validation assessment provides an objective quantification of the discrepancies between model predictions and experimental observations. As stated early, the threshold for the metric depends on the specific requirement of model accuracy. Here for the sake of illustration, assume the threshold for the discrepancies is 2°C. Since the lower bound of the metric's confidence interval is beyond the threshold, we can conclude with 95% confidence that the accuracy of the model is insufficient, it indicates that the model needs to be improved.

5 Conclusions

This paper presents a framework for validation assessment when there is both aleatory and epistemic uncertainty in model inputs. The likelihood based method is applied to representation of stochastic quantities with uncertain distribution parameters which are due to sparse point data or interval data. The method's result for an uncertain model input is a single probability distribution that facilitates the following uncertainty propagation and validation metric. It provides an obvious advantage in computation efficiency for the conventional double loop sampling strategy is collapsed into a single loop sampling. The probability distribution of model output obtained from uncertainty propagation is compared with the EDF of experimental observations using area based validation metric, it reflects an objective quantification of the entire discrepancies between predictions and observations. A confidence interval for the validation metric which just depends on the amount of experimental observations and confidence level is also developed. It is helpful for the decision making which follows the validation assessment. The numerical example demonstrates the validation assessment framework presented in this paper.

The discussion of the framework in this paper is limited to the univariate case, which implies the model output to be a single response quantity following a statistical distribution. The validation assessment also looks for a metric, which have the flexibility of measuring the discrepancies between predictions and observations in a multivariate case. There are several types of multivariate case such as multiple location for model response or various response quantities at a single location. The validation metric for the multivariate case is required to provide an overall performance measurement for the model. There are still difficulties in aggregating individual metrics accounting for confidence level and correlation among multiple quantities. Future work in this direction will extend the validation assessment to multivariate case.

References

[1] Romero V J 2006 Some issues and needs in quantification of margins and uncertainty in complex coupled systems *47th AIAA/ ASME/ ASCE/ AHS/ ASC Structures, Structural Dynamics, and Materials Conference* Newport
[2] Helton J 2009 *Conceptual and Computational Basis for the Quantification of Margins and Uncertainty* Sandia National Laboratories: Albuquerque
[3] Oberkampf W L, Trucano T G 2007 *Verification and Validation Benchmarks* Sandia National Laboratories: Albuquerque
[4] American Institute of Aeronautics and Astronautics 1998 *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations* AIAA: Reston VA
[5] The American Society of Mechanical Engineers 2008 *ASME V&V 20-2008 Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer* ASME: New York
[6] National Aeronautics and Space Administration 2008 *NASA-STD-7009 Standard for Models and Simulations* NASA: Washington DC
[7] Trucano T G, Swiler L P, Igusa T, Oberkampf W L, Pilch M 2006 Calibration, validation, and sensitivity analysis: What's what *Reliab Eng Syst Safety* **91**(10-11) 1331-57
[8] Roy C J, Oberkampf W L 2011 A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing *Comput Method Appl M* **200**(25-28) 2131 – 44
[9] Liu Y, Chen W, Arendt P and Huang H Z 2011 Toward a better understanding of model validation metrics *J Mech Design* **133**(7) 48-60
[10] Hills R G, Trucano T G 1999 *Statistical Validation of Engineering and Scientific Models: Background* Sandia National Laboratories: Albuquerque
[11] Hills R G, Trucano T G 2002 *Statistical Validation of Engineering and Scientific Models: A Maximum Likelihood Based Metric* Sandia National Laboratories: Albuquerque
[12] Buranathiti T, Cao J, Chen W 2006 Approaches for Model Validation: Methodology and Illustration on a Sheet Metal Flanging Process *J Manuf Sci E* **128**(2) 588-97

[13] Rebba R, Mahadevan S 2006 Validation of models with multivariate output *Reliab Eng Syst Safety* **91**(8) 861-71
[14] Rebba R, Mahadevan S 2008 Computational Methods for Model Reliability Assessment *Reliab Eng Syst Safety* **93**(8) 1197-207
[15] Oberkampf W L, Trucano T G, Hirsch C 2004 Verification, Validation, and Predictive Capability in Computational Engineering and Physics *Appl Mech Rev* **57**(3) 345-84
[16] Oberkampf W L, Barone M F 2006 Measures of agreement between computation and experiment: validation metrics *J Comput Phys* **217**(1) 5-36
[17] Ferson S, Oberkampf W L, Ginzburg L 2008 Model validation and predictive capability for the thermal challenge problem *Comput Method Appl M* **197**(29-32) 2408-30
[18] Baudrit C, Dubois D 2006 Practical representations of incomplete probabilistic knowledge *Computational Statistics and Data Analysis* 86-108
[19] Agarwal H, Renaud J E, Preston E L and Padmanabhan D 2004 Uncertainty quantification using evidence theory in multidisciplinary design optimization *Reliab Eng Syst Safety* **85**(1-3) 281-94
[20] Swiler L P, Giunta A A 2007 *Aleatory and epistemic uncertainty quantification for engineering applications* Sandia National Laboratories: Albuquerque
[21] Rao S, Annamdas K 2009 An evidence based fuzzy approach for the safety analysis of uncertain systems *50th AIAA/ ASME/ ASCE/ AHS/ ASC Structures, Structural Dynamics, and Materials Conference* California
[22] Sankararaman S, Mahadevan S 2011 Likelihood based representation of epistemic uncertainty due to sparse point data and/or interval data *Reliab Eng Syst Safety* **96**(7) 814-24
[23] Miller L H 1956 Table of percentage points of Kolmogorov statistics *Journal of the American Statistical Association* **51** 111-21

Authors

Liang Zhao, born in 1983, MeiShan, China

Current position, grades: Ph.D candidate
University studies: Computer science in Southwest Jiaotong University
Scientific interest: Quantification of uncertainty and model validation.
Publications: 5
Experience: 2011-2014: Ph.D in system science, China Academy of Engineering Physics; 2009-2011: System engineer in JEZETEK; 2006-2009: Master in computer science, Southwest Jiaotong University

Zhanping Yang, born in 1966, ChongQing, China

Current position, grades: Researcher, Chief Engineer of Institute of Electronic Engineering
University studies: System science in Beijing Institute of Technology
Scientific interest: Modelling and assessment of complex system
Publications: 27
Experience: 2001-2014: Electronic system design in Institute of Electronic Engineering; 1998-2001: Ph.D in system science, Beijing Institute of Technology