Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A.A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

AN UNCERTAINTY CALIBRATION METHOD FOR MULTI-OUTPUT MODEL

Xiaobing Shang
Ju Huo
Tao Chao
Ping Ma
Ming Yang

Control and Simulation Center
Harbin Institute of Technology
West Dazhi Street
Harbin, 150080, CHINA

Bo Liu

No.705 Institute
China Shipbuilding Industry Corporation
High Tech Zone, Jin Ye Road
Xi'an, 710077, CHINA

ABSTRACT

A comprehensive framework, combining the cumulative distribution function and modified Kolmogorov–Smirnov test, is proposed to solve multi-output simulation model's parameter calibration problem with the presence of uncertain parameters. The framework is based on comparing the difference between joint cumulative distribution functions of some observed values and that of simulation sample values. An auxiliary variable method is used to decompose hybrid parameters into sub-parameters. Then the optimal matching values can be found with genetic algorithm according to the index of difference of joint cumulative distribution functions.

1 INTRODUCTION

In the engineering process of aeronautics and astronautics, computer simulation has become an important approach to save cost and improve work efficiency in the engineering model design. Since the foundation of simulation is the mathematical model describing, so it plays an important role to find a method that describing the model accurately. However, in the process of actual engineering development, the model is undetermined with the presence of various uncertainties sourced from lack of knowledge, design and manufacturing defects, environment of the product. It not only leads to an uncertain model, but also may lead to worse output under the influence of above uncertainties. The method of this paper is to solve the challenge of the inverse assessment of model uncertainty, where the model parameters are calibrated simultaneously to reduce parameter epistemic uncertainty. Currently, the widely used methods include interval theory, evidence theory, fuzzy set, possibility theory, and convex method, etc. But most theories only address the situation of singular output. The uncertain model studied in this paper is the multi-output model with the three kind of uncertain variables, and the calibrated object is the epistemic uncertainty parameter in the model. The calibration method of hybrid uncertainty parameter with multi-outputs is studied emphatically. The empirical cumulative distribution function (ECDF) obtained by using the modified two-sample Kolmogorov–Smirnov (K-S) test, is compared with the ECDF of the real data. Then, the optimized matching value can be found, using genetic algorithm. The optimized matching value is seen as approximation of epistemic uncertainty parameters, to eliminate epistemic uncertainty in system caused by lack of knowledge.

2 MULTI-OUTPUT MODEL CALIBRATION METHOD

The ECDF matching approach uses the concept of the two-sample K-S test to compare the ECDF of the observations, with the ECDF for the generated output values using aleatory uncertainties for some realization of the epistemic uncertainty variable. Note that the following formulas is all vector forms, shorted from multi-output.

Step 1: Decompose all uncertain parameters of model, especially the hybrid uncertain parameter. After determining the sub-parameters θ of the model, we should make clear that which sub-parameters are aleatory uncertainty that cannot be calibrated and which sub-parameters are epistemic uncertainty that can be calibrated. So, the uncertainty model can be formulate as:

$$\boldsymbol{Y} = \boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta}) \tag{1}$$

Step 2: Retrieve a given number of realistic target observations from the database. In this paper, the real observation value is replaced by the simulated observation value which is obtained by random uncertainty in the real value condition of the epistemic uncertainty parameter. The corresponding ECDF is calculated after simply eliminating outliers. According to the ECDF function of N independent identically distributed samples, Equation (2) is obtained:

$$F_N\left(\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})\right) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{I}_{\boldsymbol{Y}_i \le \boldsymbol{y}}(\boldsymbol{\theta}) \tag{2}$$

where I is the index function; If $X_i \le x$, then $I=1$, or $I=0$.

Step 3: Based on the double-samples K-S test method, Latin hypercube sampling was used for uncertainty parameters of model. First, getting N' random outputs samples of observations for one particular θ realization. Then an ECDF can be generated by Equation (3). The random stream for generating N' samples of output samples for a particular θ realization is fixed to reduce the noise in objective function calculation:

$$F_i\left(\boldsymbol{f}(\boldsymbol{X}, \boldsymbol{\theta})\right) = \frac{1}{N'}\sum_{i=1}^{N'} \boldsymbol{I}_{\boldsymbol{Y}_i \le \boldsymbol{y}}(\boldsymbol{\theta}), \quad i = 1, \ldots, N' \tag{3}$$

Step 4: A modified K-S test, used here for comparing two ECDFs with N (given observations) and N' (randomly generated using aleatory uncertainty) samples ($D^i_{NN'}$), is given by:

$$D^i_{NN'} = \sum_{x_{1,obs}} \left(F_N - F_i\right)^2 \tag{4}$$

which represents the variance error of the difference between two ECDF curves. In Equation (4), it can be seen that the smaller the index $D^i_{NN'}$, the smaller the difference between the two groups of distribution function curves. The selected value of the corresponding sub-parameters is closer to the real value of the epistemic uncertainty parameter. Therefore, the minimum value of the index is used as the optimal matching value θ^* for the parameter of epistemic uncertainty.

Step 5: For seeking for the minimum value of the index, a genetic algorithm (GA) is used to find optimization. The minimum distance is used as the fitness function and sample points violating the Latin hypercube design restriction will contribute a large penalty value to the fitness function. Then, we can get a set of optimization of sub-parameters. In order to increase the pervasiveness, the random flow needs to be dynamic. That means to repeat step 3 and step 4, for m times. Then the matching sets of optimization values of m – time cycle is composed of the new calibration interval of uncertainty parameters, which shows dynamic calibrated value fluctuating around true value of epistemic parameters. Finally, take the median of the new interval as the best approximation value of the real value of the parameter:

$$\theta_i \in [\theta_{lower}, \theta_{upper}] \tag{5}$$

REFERENCES

Kennedy, M. C., and A. O'Hagan. 2001. "Bayesian Calibration of Computer Models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63(3):425–464.

Arendt, P. D., D. W. Apley, W. Chen, D. Lamb, and D. Gorsich. 2012. "Improving Identifiability in Model Calibration using Multiple Responses". *Journal of Mechanical Design* 134(10):100909.

Oberkampf, W. L., J. C. Helton, C. A. Joslyn, S. F. Wojtkiewicz, and S. Ferson. 2004. "Challenge Problems: Uncertainty in System Response given Uncertain Parameters". *Reliability Engineering & System Safety* 85(1-3):11–19.